

Gesellschaft für Informatik e.V. (GI)

publishes this series in order to make available to a broad public recent findings in informatics (i.e. computer science and information systems), to document conferences that are organized in cooperation with GI and to publish the annual GI Award dissertation.

Broken down into

- seminars
- proceedings
- dissertations
- thematics

current topics are dealt with from the vantage point of research and development, teaching and further training in theory and practice. The Editorial Committee uses an intensive review process in order to ensure high quality contributions.

The volumes are published in German or English.

Information: <http://www.gi.de/service/publikationen/lni/>

ISSN 1617-5468

ISBN 978-3-88579-639-8

The proceedings of the BIOSIG 2015 include scientific contributions of the annual conference of the Biometrics Special Interest Group (BIOSIG) of the Gesellschaft für Informatik (GI). The conference took place in Darmstadt, 09.-11. September 2015. The advances of biometrics research and new developments in the core biometric application field of security have been presented and discussed by international biometrics and security professionals.



Arslan Brömme, Christoph Busch, Christian Rathgeb, Andreas Uhl (Eds.):  
BIOSIG 2015 - 14<sup>th</sup> International Conference of the Biometrics Special Interest Group

245

# GI-Edition

## Lecture Notes in Informatics

**Arslan Brömme, Christoph Busch ,  
Christian Rathgeb, Andreas Uhl (Eds.)**

## BIOSIG 2015

**Proceedings of the 14<sup>th</sup> International  
Conference of the Biometrics  
Special Interest Group**

**09.-11. September 2015  
Darmstadt, Germany**

# Proceedings





Arslan Brömme, Christoph Busch,  
Christian Rathgeb, Andreas Uhl (Eds.)

# **BIOSIG 2015**

**Proceedings of the 14<sup>th</sup> International Conference  
of the Biometrics Special Interest Group**

**09.-11. September 2015 in  
Darmstadt, Germany**

Gesellschaft für Informatik e.V. (GI)

## **Lecture Notes in Informatics (LNI) - Proceedings**

Series of the Gesellschaft für Informatik (GI)

Volume P-245

ISBN 978-3-88579-639-8

ISSN 1617-5468

### **Volume Editors**

Arslan Brömme

GI BIOSIG, Gesellschaft für Informatik e.V.

Ahrstraße 45, D-53175 Bonn

*Email: arslan.broemme@aviomatik.de*

Christian Rathgeb

Hochschule Darmstadt, CASED

Haardtring 100, D-64295 Darmstadt

*Email: christian.rathgeb@h-da.de*

Christoph Busch

Hochschule Darmstadt, CASED

Haardtring 100, D-64295 Darmstadt

*Email: christoph.busch@h-da.de*

Andreas Uhl

University of Salzburg,

Jakob-Haringer Str. 2, A-5020 Salzburg

*Email: uhl@cosy.sbg.ac.at*

### **Series Editorial Board**

Heinrich C. Mayr, Alpen-Adria-Universität Klagenfurt, Austria

(Chairman, mayr@ifit.uni-klu.ac.at)

Dieter Fellner, Technische Universität Darmstadt, Germany

Ulrich Flegel, Hochschule für Technik, Stuttgart, Germany

Ulrich Frank, Universität Duisburg-Essen, Germany

Johann-Christoph Freytag, Humboldt-Universität zu Berlin, Germany

Michael Goedicke, Universität Duisburg-Essen, Germany

Ralf Hofestädt, Universität Bielefeld, Germany

Michael Koch, Universität der Bundeswehr München, Germany

Axel Lehmann, Universität der Bundeswehr München, Germany

Peter Sanders, Karlsruher Institut für Technologie (KIT), Germany

Sigrid Schubert, Universität Siegen, Germany

Ingo Timm, Universität Trier, Germany

Karin Vosseberg, Hochschule Bremerhaven, Germany

Maria Wimmer, Universität Koblenz-Landau, Germany

### **Dissertations**

Steffen Hölldobler, Technische Universität Dresden, Germany

### **Seminars**

Reinhard Wilhelm, Universität des Saarlandes, Germany

### **Thematics**

Andreas Oberweis, Karlsruher Institut für Technologie (KIT), Germany

© Gesellschaft für Informatik, Bonn 2015

**printed by** Köllen Druck+Verlag GmbH, Bonn

## Chairs' Message

Welcome to the annual international conference of the Biometrics Special Interest Group (BIOSIG) of the Gesellschaft für Informatik (GI) e.V.

GI BIOSIG was founded in 2002 as an experts' group for the topics of biometric person identification/authentication and electronic signatures and its applications. Over more than a decade the annual conference in strong partnership with the Competence Center for Applied Security Technology (CAST) established a well known forum for biometrics and security professionals from industry, science, representatives of the national governmental bodies and European institutions who are working in these areas.

The BIOSIG 2015 international conference is jointly organized by the Biometrics Special Interest Group (BIOSIG) of the Gesellschaft für Informatik e.V., the Competence Center for Applied Security Technology e.V. (CAST), the German Federal Office for Information Security (BSI), the European Association for Biometrics (EAB), the ICT COST Action IC1106, the European Commission Joint Research Centre (JRC), the TeleTrust Deutschland e.V. (TeleTrust), the Norwegian Biometrics Laboratory (NBL), the Center for Advanced Security Research Darmstadt (CASED), and the Fraunhofer Institute for Computer Graphics Research (IGD). This years' international conference BIOSIG 2015 is again technically co-sponsored by the Institute of Electrical and Electronics Engineers (IEEE) and is enriched with satellite workshops by the TeleTrust Biometric Working Group and the European Association for Biometrics.

The international program committee accepted full scientific papers strongly according to the LNI guidelines (acceptance rate ~22%) within a scientific double-blinded review process of at minimum five reviews per paper. All papers were formally restricted for the printed proceedings to 12 pages for regular research contributions including an oral presentation and 8 pages for further conference contributions including a poster presentation at the conference site.

Furthermore, the program committee has created a program including selected contributions of strong interest (further conference contributions) for the outlined scope of this conference. All paper contributions for BIOSIG 2015 will be published additionally in the IEEE Xplore Digital Library.

We would like to thank all authors for their contributions and the numerous reviewers for their work in the program committee.

Darmstadt, 09<sup>th</sup> September 2015

Arslan Brömme	Christoph Busch	Christian Rathgeb	Andreas Uhl
<i>GI BIOSIG,</i>	<i>Hochschule</i>	<i>Hochschule</i>	<i>University of</i>
<i>GI e.V.</i>	<i>Darmstadt</i>	<i>Darmstadt</i>	<i>Salzburg</i>

## Chairs

Arslan Brömme, *GI BIOSIG, GI e.V., Bonn, Germany*

Christoph Busch, *Hochschule Darmstadt - CASED, Germany*

Christian Rathgeb, *Hochschule Darmstadt - CASED, Germany*

Andreas Uhl, *University of Salzburg, Austria*

## Program Committee

Harald Baier (CASED, DE)

Oliver Bausinger (BSI, DE)

Thiriamchos Bourlai (WVU, US)

Patrick Bours (GUC, NO)

Sebastien Brangoulo (Morpho, FR)

Ralph Breithaupt (BSI, DE)

Julien Bringer (Morpho, FR)

Arslan Brömme (GI/BIOSIG, DE)

Christoph Busch (CAST-Forum, DE)

Victor-Philipp Busch (Sybuca, DE)

Patrizio Campisi (Uni Roma 3, IT)

Nathan Clarke (CSCAN, UK)

Paul Lobato Correira (LXIT, PT)

Adam Czajka (NASK, PL)

Farzin Deravi (UKE, UK)

Martin Drahansky (BUT, CZ)

Andrzej Drygajlo (EPFL, CH)

Julian Fierrez (UAM, ES)

Simone Fischer-Hübner (KAU, SE)

Lothar Fritsch (NR, NO)

Steven Furnell (CSCAN, UK)

Sonia Garcia (TSP, FR)

Patrick Grother (NIST, US)

Olaf Henniger (Fhg IGD, DE)

Detlef Hühnlein (ecsec, DE)

Robert W. Ives (USNA, US)

Christiane Kaplan (softpro, DE)

Stefan Katzenbeisser (CASED, DE)

Tom Kevenaar (GenKey, NL)

Didier Meuwly (NFI, NL)

Emilio Mordini (CSSC, IT)

Elaine Newton (NIST, US)

Mark Nixon (UoS, UK)

Alexander Nouak (Fhg IGD, DE)

Markus Nuppeney (BSI, DE)

Hisao Ogata (Hitachi, JP)

Martin Olsen (GUC, NO)

Javier Ortega-Garcia (UAM, ES)

Michael Peirce (Daon, IR)

Dijana Petrovska (TSP, FR)

Anika Pflug (CASED, DE)

Ioannis Pitas (AUT, GR)

Fernando Podio (NIST, US)

Raghu Ramachandra (GUC, NO)

Kai Rannenber (Uni FFM, DE)

Christian Rathgeb (CASED, DE)

Arun Ross (MSU, US)

Heiko Roßnagel (Fhg IAO, DE)

Raul Sanchez-Reillo (UC3M, ES)

Stephanie Schuckers (CIU, US)

Günter Schumacher (JRC, IT)

Takashi Shinzaki (Fujitsu, JP)

Max Snijder (EAB, NL)

Luis Soares (ISCTE-IUL, PT)

Luuk Spreeuwiers (UTW, NL)

Elham Tabassi (NIST, US)

Tieniu Tan (NLPR, CN)

Massimo Tistarelli (UNISS, IT)

Ulrike Korte (BSI, DE)  
Bernd Kowalski (BSI, DE)  
Ajay Kumar (Poly, HK)  
Herbert Leitold (a-sit, AT)  
Guoqiang Li (GUC, NO)  
Stan Li (CBSR, CN)  
Paulo Lobato Correira (IST, PT)  
Davide Maltoni (UBO, IT)  
Johannes Merkle (secunet, DE)

Dimitrios Tzovaras (CfRaT, GR)  
Andreas Uhl (COSY, AT)  
Markus Ullmann (BSI, DE)  
Raymond Veldhuis (UTW, NL)  
Anne Wang (Cogent, US)  
Jim Wayman (SJSU, US)  
Peter Wild (UoR, UK)  
Andreas Wolf (BDR, DE)  
Bian Yang (GUC, NO)

## **Hosts**

Biometrics Special Interest Group (**BIOSIG**)  
of the Gesellschaft für Informatik (GI) e.V.  
<http://www.biosig.org>

Competence Center for Applied Security Technology e.V. (**CAST**)  
<http://www.cast-forum.de>

Bundesamt für Sicherheit in der Informationstechnik (**BSI**)  
<http://www.bsi.bund.de>

European Association for Biometrics (**EAB**)  
<http://www.eab.org>

European Commission Joint Research Centre (**JRC**)  
<http://ec.europa.eu/dgs/jrc/index.cfm>

TeleTrusT Deutschland e.V. (**TeleTrust**)  
<http://www.teletrust.de>

Norwegian Biometrics Laboratory (**NBL**)  
[http://www.nislab.no/biometrics\\_lab](http://www.nislab.no/biometrics_lab)

Center for Advanced Security Research Darmstadt (**CASED**)  
<http://www.cased.de>

Fraunhofer-Institut für Graphische Datenverarbeitung (**IGD**)  
<http://www.igd.fraunhofer.de>

# **BIOSIG 2015 – Biometrics Special Interest Group**

**“2015 International Conference of the Biometrics Special Interest Group”**  
09<sup>th</sup> -11<sup>th</sup> September 2015

Biometrics provides efficient and reliable solutions to recognize individuals. With increasing number of identity theft and misuse incidents we do observe a significant fraud in e-commerce and thus growing interests on trustworthiness of person authentication.

Nowadays we find biometric applications in areas like border control, national ID cards, e-banking, e-commerce, e-health etc. Large-scale applications such as the European Union Visa Information System (VIS) and Unique Identification (UID) in India require high accuracy and also reliability, interoperability, scalability, system reliability and usability. Many of these are joint requirements also for forensic applications.

Multimodal biometrics combined with fusion techniques can improve recognition performance. Efficient searching or indexing methods can accelerate identification efficiency. Additionally, quality of captured biometric samples can strongly influence the performance.

Moreover, mobile biometrics is an emerging area and biometrics based smartphones can support deployment and acceptance of biometric systems. However concerns about security and privacy cannot be neglected. The relevant techniques in the area of presentation attack detection (liveness detection) and template protection are about to supplement biometric systems, in order to improve fake resistance, prevent potential attacks such as cross matching, identity theft etc.

BIOSIG 2015 offers you once again a platform for international experts' discussions on biometrics research and the full range of security applications.

## Table of Contents

<b>BIOSIG 2015 – Regular Research Papers .....</b>	<b>13</b>
<b>Karl Ricanek Jr., Shivani Bhardwaj, Michael Sodomsky</b> <i>A Review of Face Recognition against Longitudinal Child Faces.....</i>	<b>15</b>
<b>Ester Gonzalez-Sosa, Ruben Vera-Rodriguez, Julian Fierrez, Pedro Tome, Javier Ortega-Garcia</b> <i>Pose Variability Compensation Using Projective Transformation Forensic Face Recognition.....</i>	<b>27</b>
<b>Andreas Ranftl, Fernando Alonso-Fernandez, Stefan Karlsson</b> <i>Face Tracking using Optical Flow Development of a Real-Time AdaBoost Cascade Face Tracker.....</i>	<b>39</b>
<b>Ning Jia, Victor Sanchez, Chang-Tsun Li, Hassan Mansour</b> <i>On Reducing the Effect of Silhouette Quality on Individual Gait Recognition: a Feature Fusion Approach.....</i>	<b>49</b>
<b>Peter Wild, Heinz Hofbauer, James Ferryman, Andreas Uhl</b> <i>Segmentation-level Fusion for Iris Recognition.....</i>	<b>61</b>
<b>Michael Happold</b> <i>Structured Forest Edge Detectors for Improved Eyelid and Iris Segmentation.....</i>	<b>73</b>
<b>Martin Aastrup Olsen, Martin Böckeler, Christoph Busch</b> <i>Predicting Dactyloscopic Examiner Fingerprint Image Quality Assessments.....</i>	<b>85</b>
<b>Johannes Kotzerke, Stephen A. Davis, Robert Hayes, Luuk J. Spreeuwiers, Raymond N.J. Veldhuis, Kathy J. Horadam</b> <i>Identification performance of evidential value estimation for fingermarks.....</i>	<b>97</b>
<b>Jesse Hartloff, Avradip Mandal, Arnab Roy</b> <i>Privacy Preserving Technique for Set-Based Biometric Authentication using Reed-Solomon Decoding.....</i>	<b>109</b>
<b>Benjamin Tams, Johannes Merkle, Christian Rathgeb, Johannes Wagner, Ulrike Korte, Christoph Busch</b> <i>Improved Fuzzy Vault Scheme for Alignment-Free Fingerprint Features.....</i>	<b>121</b>
<b>Edlira Martiri, Bian Yang, Christoph Busch</b> <i>Protected Honey Face Templates.....</i>	<b>133</b>
<b>Alexandre Sierro, Pierre Ferrez, Pierre Roduit</b> <i>Contact-less Palm/Finger Vein Biometric.....</i>	<b>145</b>

<b>Guoqiang Li, Bian Yang, Christoph Busch</b> <i>A Fingerprint Indexing Scheme with Robustness against Sample Translation and Rotation.....</i>	157
<b>Kribashnee Dorasamy, Leandra Webb, Jules Tapamo</b> <i>Evaluating the Change in Fingerprint Directional Patterns under Variation of Rotation and Number of Regions.....</i>	169
<b>BIOSIG 2015 – Further Conference Contributions.....</b>	181
<b>Christof Kauba, Andreas Uhl</b> <i>Robustness Evaluation of Hand Vein Recognition Systems.....</i>	183
<b>Nahuel González, Enrique P. Calot</b> <i>Finite Context Modeling of Keystroke Dynamics in Free Text.....</i>	191
<b>Heinz Hofbauer, Christian Rathgeb, Johannes Wagner, Andreas Uhl, Christoph Busch</b> <i>Investigation of Better Portable Graphics Compression for Iris Biometric Recognition.....</i>	199
<b>Nalla Pattabhi Ramaiah, Nalla Srilatha, Chalavadi Krishna Mohan</b> <i>Sparsity-based Iris Classification using Iris Fiber Structures .....</i>	207
<b>Pedro Tome and Sébastien Marcel</b> <i>Palm Vein Database and Experimental Framework for Reproducible Research.....</i>	215
<b>Michael Fairhurst, Meryem Erbilek, Marjory Da Costa-Abreu</b> <i>Exploring Gender Prediction from Iris Biometrics.....</i>	223
<b>Dominik Klein, Jan Kruse</b> <i>A Comparative Study on Image Hashing for Document Authentication.....</i>	231
<b>Alaa Darabseh, Akbar Siami Namin</b> <i>On Accuracy of Keystroke Authentications Based on Commonly Used English Words.....</i>	239
<b>Christof Jonietz, Eduardo Monari, Chengchao Qu</b> <i>Towards Touchless Palm and Finger Detection for Fingerprint Extraction with Mobile Devices.....</i>	247
<b>Naser Damer, Alexander Nouak</b> <i>Weighted Integration of Neighbors Distance Ratio in Multi-biometric Fusion.....</i>	255

<b>Soumik Mondal, Patrick Bours</b> <i>Does Context matter for the Performance of Continuous Authentication Biometric Systems? An Empirical Study on Mobile Devices.....</i>	263
<b>Thomas Klir</b> <i>Fingerprint Image Enhancement with easy to use algorithms.....</i>	271
<b>Lisa de Wilde, Luuk Spreeuwers, Raymond Veldhuis</b> <i>Exploring How User Routine Affects the Recognition Performance of a Lock Pattern.....</i>	279
<b>Thomas Herzog, Andreas Uhl</b> <i>JPEG Optimisation for Fingerprint Recognition: Generalisation Potential of an Evolutionary Approach .....</i>	287
<b>Nassima Kihal, Arnaud Polette, Salim Chitroub, Isabelle Brunette, Jean Meunier</b> <i>Corneal Topography: An Emerging Biometric System for Person Authentication....</i>	295
<b>Rig Das, Emanuele Maiorana, Daria La Rocca, Patrizio Campisi</b> <i>EEG Biometrics for User Recognition using Visually Evoked Potentials.....</i>	303
<b>Maximilian Krieg, Nils Rogmann</b> <i>Liveness Detection in Biometrics.....</i>	311
<b>M. Hamed Izadi, Andrzej Drygajlo</b> <i>Discarding low quality Minutia Cylinder-Code pairs for improved fingerprint comparison.....</i>	319
<b>Christian Kahindo, Sonia Garcia-Salicetti, Nesma Houmani</b> <i>A Signature Complexity Measure to select Reference Signatures for Online Signature Verification.....</i>	327



**BIOSIG 2015**

**Regular Research Papers**



# A Review of Face Recognition against Longitudinal Child Faces

Karl Ricanek Jr., Ph.D. Senior Member IEEE, Shivani Bhardwaj, & Michael Sodomsky

I3S Institute – Face Aging Group  
University of North Carolina Wilmington  
601 South College Road  
28403 Wilmington  
ricanekk@uncw.edu  
sb2534@uncw.edu

**Abstract:** It is an established fact that the face-based biometric system performance is affected by the variation that is caused as a result of aging; however, the question has not been adequately investigated for non-adults, i.e. children from birth to adulthood. The majority of research and development in automated face recognition has been focused on adults. The objective of this paper is to establish an understanding of face recognition against non-adults. This work develops a publicly available longitudinal child face database of child celebrities from images in the wild (ITWCC). This work explores the challenges of biological changes due to maturation, i.e. the face grows longer and wider, the nose expands, the lips widen, etc, i.e. craniofacial morphology, and examines the impact on face recognition. The systems chosen are: Cognitec’s FaceVacs 8.3, Open Source Biometric Recognition (SF4), principal component analysis (PCA), linear discriminant analysis (LDA), local region principal component analysis (LRPCA), and cohort linear discriminant analysis. Face matchers recorded low performance: top performance in verification is 37% TAR at 1% FAR and best rank-1 identification reached 25% recognition rate on a gallery of 301 subjects.

## 1. Introduction

The human face is an important feature of identity recognition. The characteristics of the face that makes it a desirable biometric modality is its uniqueness, universality, acceptability, semi-permanence, and easy collectability [RB11]. Because of its potential and possible variety of application, automated face recognition has received a lot of attention over the last two decades. Face recognition can be accomplished from a distance and via non-contact acquisition, which offer an added advantage over most biometric systems and make it more suitable for security and surveillance systems. Face recognition, may play a vital role in identifying children that go missing and in extensive range of access control and monitoring systems, especially to safeguard children. This technology can provide a whole new approach to protect and support latched-key kid and to provide access control for various internet of things across different age groups. It can be used to protect the non-adult population from predators and illegitimate web contents.

Face recognition is a challenging problem, and a great deal of work has been completed for pose correction, illumination variation, and expression to support face recognition in the wild. However, the majority of the work done has been focused on adults and deals with the dynamics of mature faces. The objective of this paper is to review the current state of facial recognition algorithms with a focus on non-adult stages of growth and development, 2 years to 16 years.

Aging with respect to facial recognition system includes variation in shape, size and texture of the face. These temporal changes will cause performance degradation. Hence, state issued id's, e.g. driving license, has to be renewed every 5-10 years. Mathew Turk stated that "developing a computational model of face recognition is quite difficult, because faces are complex, multidimensional, and meaningful visual stimuli" [TP91]; however, when aging information is added to this problem, it becomes infinitely more difficult. The most challenging problem in developing a solution for childhood face recognition is the formation of a useful dataset. This work addresses this primary concern.

## **Contributions**

This paper provides the following contribution to the research community: 1. provides the baseline for face recognition performance for children against a suite of traditional face recognition techniques and investigate the impact of well-defined structural (skeletal) changes of the face on a suite of FR techniques; 2. establishes the first moderate scale publicly available child face database focused across the growth and development period<sup>1</sup>, which is one of the key issue in evaluation; and, 3. provides a methodology framework for investigating the problem of face recognition across childhood.

## **2. Background**

Facial recognition is a complex topic that has been researched very heavily and many attempts have been made to understand the effects aging has on facial recognition systems. However, algorithm performance with respect to human aging: as a subject of the growth and development phase of childhood, has just begun to be fully explored by researchers. One of the biggest issues is the vast amount of data that is required to fully understand the human face and its maturation process. As the face changes over time, the ability to recognize the person becomes more challenging. This is further exacerbated if the person under inspection is not known to the observer.

Anthropological and forensic studies have contributed significantly to show that age related changes of non-adults are different from face aging for adults. Human aging can be studied as a two staged process: first involves the growth and development phase and

---

<sup>1</sup> The FG-Net face database has childhood images of 80 subjects, however many of these images are scanned from photographs. Additionally, CASIA has a twin's dataset that contains a number of child captures but only across a couple of days.

second deals with the effect of maturity as age progress [Rk09]. The performance for non-adult recognition over time spans less than a few weeks, may be on par with adult FR systems; However due to the rapid change that occurs during childhood, temporal displacement has a more profound impact on FR systems.

## **2.1 Adult aging**

Adult aging is dominated by morphological and soft tissue changes i.e. skin texture, wrinkles etc., but some skeletal changes continue to occur [ARP07], [Tk10]. Early adulthood shows the first signs of soft tissue stressing. Hyper-dynamic expressions will start to show wrinkles on the face. Fine facial lines will appear horizontally on the forehead, vertical lines between the forehead and thin lines around the outside corners of the eyes will appear [Ks13]. From ages 40 to 50 there are noticeable changes to skin texture while minimal changes are found in younger years [THB00]. The aging rate of adults differs heavily on the individual, which is not the case for non-adults [ARP07]. These differences can be attributed to genetics and external features. Biological changes in adults alters the shape and texture of a face. As the skull continues to change with age, the eyes appear smaller as they sink in deeper into their orbits. As the skins elasticity begins to degrade wrinkles form, more notably in the eyelids, and the corners of the mouth [Lj74]. These feature begin to sag and change in size thus changing the relationships of the features of the face.

## **2.2 Facial growth and development**

Face aging with respect to children majorly involves craniofacial growth and development. This is the phase which is dominated by facial structural development which causes change in shape and size of face. In Karen T. Taylor's book "Forensic Art and Illustration" she describes the changes of the cranium and face year by year from childhood to young adulthood [Tk10].

These underlying skeletal changes will alter the appearance of the face. Cranial growth will not greatly change the features within the face, but it is the cause of change in proportions between them. During developmental changes the features of the face will remain alike to their original. This growth pattern is known as gnomatic growth [Tk10]. Craniofacial growth rate is affected by factors such as puberty and the growth of permanent teeth and there are jumps in growth rate at these periods, which makes growth a non-linear function. The rates of change of face is maximum in non-adult, particularly between birth and 5 years old [F192]. For this reason, high rate of change in the 0 to 5 years, face recognition technology may not be an appropriate technology for use, and hence, credentialing systems like national id's and passports should abstain from being used on persons in this age group. Maturation is achieved in males between the ages of 12 and 15 years while the same is true for females between 10 years and 13 years [F192]. After maturation the underlying structure of the face will continue to grow, however, not as rapidly.

### 2.3 Aging effect on performance

Growth & development and aging factors have a great impact on the performance of existing system over time [NG14]. The face develops and ages in numerous ways which pose challenges for face processing techniques. Humans have the ability to recognize a person from years ago; however, the person does look fundamentally different.

Early work in the impacts of face recognition by [LT00] established baselines for performance degradation in the problem of aging for adult faces. The work concluded that the performance does decrease as the time between probe and gallery increases, also it shows that older faces tended to be better recognized than younger faces i.e. individuals in the age range of 40–49 years were better recognized than those in the youngest age range, <18 years. Klare and Jain [RB05] concluded the same results. Later work by NIST concluded that recognition becomes easier with advanced age; however, recognition remains a challenge across large time spans for adults.

A very recent publication by NIST [NG14] evaluated the performance of non-adult face recognition on a suite of commercial FR systems for which the report concluded that significant weakness exists for current commercial systems. Further the report indicates that identification accuracy is strongly dependent on subject age. Where older subjects are easier to identify and easily distinguished from other, the opposite is true with children, being very hard to identify. In case if infants both false negative and false positive rates are much higher [NG14].

## 3. Dataset

Data is the primary necessity for exploration of face recognition systems whether through algorithm design or algorithm performance. FR technology performs better with the highly constrained images; however most of the time it is not the real scenario where we use this technology. Also to develop a database of this nature is extremely difficult because of human subject requirements and the nature of capturing or finding images across time, of same individual.

Table 1: Outline of Available Aging Datasets

Database	# Subjects	# Images	Images per subjects	Age Range	Image Quality	Label for Age
VADANA [SRK11]	43	2298	3-300	0-78	24-bit colored, 30 scanned	Yes
FGNET [Cf12]	82	1002	6 -18	0-69	Mostly scanned images	Yes
MORPH (Album1) [RT06]	631	1690	1-6	16-69	Digitally scanned at 300dpi, Grey scaled	Yes
MORPH	13673	55608	1-53	16 - 99	8-bit color	Yes

(Album 2) [RT06]					200x240 JPEG or 400x480 JPEG	
Cross-Age Celebrity Dataset [CCH15]	2,000	163,446	--	16 - 62	24-bit colored images	Yes

The recognized public databases that contain child faces is FG-NET; however they did not offer the sufficient number of subjects to evaluate the face recognition systems for children, also most of the images in the dataset are scanned from photographs, which tend to lose anthropometric measures of faces as well as introduces scanning artifacts that are difficult to decouple. That is the reason only few subjects, around 82 are usable from the FG-NET dataset [Cc10]. Adience dataset [EEH] contains non-adult subjects and label its subjects for different age groups; however it is a cross-sectional dataset and does not provide any longitudinal information of subjects.

To support the objective: In-the-Wild Child Celebrity, or ITWCC, dataset was created. It is the largest longitudinal dataset that has been developed to study the present system performance specifically for the non-adults. ITWCC focuses on having large sets of individuals, where the subject growth and development can be observed. As the dataset’s name ‘In-the-wild’ suggests, the images are collected with unrestricted face and the data corpus is designed to emulate a real-life scenario as shown in figure 1. Images were captured by exploiting the fame of the subjects and gathered through open Internet sources, which are free to use. The data was captured until December of 2013. The criteria used to develop this dataset are as follows: 1) The subject must have at least three images to qualify. 2) The subject must have at least two images less than 16 years of age. 3) The date that the photo was taken must be available.



Figure 1: In-The-Wild Child Celebrity Dataset

In addition to the image, other meta-data is also captured. Age, race, gender, data of the photo, subject name, a unique photo identifier, and a conditional makeup and glasses marker, and the URL of the image is recorded for each entry. This information can further illuminate the difference in gender specific aging variations and occlusion’s effects on facial recognition systems. In-The-Wild Child Celebrity (ITWCC) dataset is composed of 304 subjects and 1705 images. The subject’s age within this dataset range from 5 months to 32 years. The dataset contains 876 female images and 839 male images. The average age of all images is 13.4 years with a standard deviation of 3.4 years. The average age of the first capture for the acquisition into this dataset is 10.2 years with a standard deviation of 3.9 years; furthermore, the average age of final capture is 16.3 years with a standard deviation of 4.467 years.

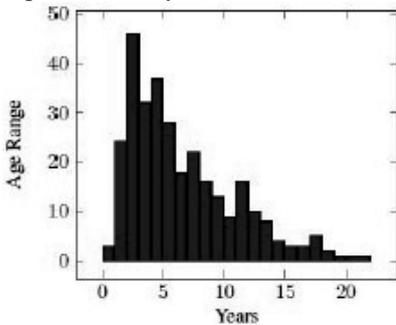


Figure 1: Age Range of Subjects in Year

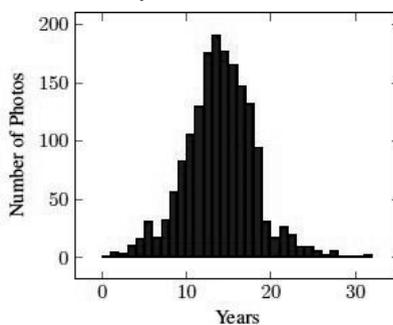


Figure 2: Age of Image

Figure 2 and figure 3, shows the number of subjects with a particular longitudinal age range, i.e. the maximum age less the minimum age and expresses the number of images for each age.

#### 4. Methods

The challenges of using facial recognition techniques on children and adolescent faces were evaluated by running multiple baseline algorithms, an open source matcher and an extensively evaluated commercial system against the ITWCC dataset. Foundational open source algorithms were used because they have been well researched against many different types of dataset, including “in-the-wild” adult dataset, and a commercial system, Cognitec, was chosen due to its strength in nearly all scenarios. Another biometric evaluation toolkit and API is Open Source Biometric Recognition(OpenBR) collaboratory [Kj13]. OpenBR is a collaborative tool that provides a method for researchers to compare algorithms in a controlled environment. Standard face matching systems used were principal component analysis (PCA), linear discriminant analysis (LDA), cohortLDA and local region PCA algorithm (LRPCA). All the foundational algorithms are implemented in open-source environments. Two toolkits implemented by Colorado State University are used in this work: the 2011 Baseline Algorithms, and the CSU Face Identification Evaluation System [Br03], [Ly12], [Pp11].

This work also examines the fully automated face recognition system, Cognitec's, FaceVacs [Gc14]. Each biometric system in this work preprocesses the images to handle illumination, scale, and orientation issues. To achieve this, all the systems use the eye coordinates from image to extract the face and register it into a standardized format and then normalized. Cognitec, OpenBR and CSU's 2011 baselines uses an automated eye detection algorithm to acquire the eye locations. CSU's 2006 biometric toolkit need eye coordinates to be provided and marked by hand for this work. If one of the eye detection algorithms cannot find the eyes in an image the image is considered failure to enroll and is not included into the matching. PCA and LDA do not have any failure to enroll errors.

## 5. Evaluation technique

Two scenarios are considered to understand the difficulty of temporally displaced data. The scenarios were developed to replicate real world situations, where this type of data would be often used. All six techniques: Cognitec FaceVacs, OpenBR, PCA, LDA, LRPCA and cohort- LDA are used independently, to evaluate the performance of FR systems, under both the scenarios. Each technique detect, preprocess, match and finally evaluate the images. All of the matching information was provided to OpenBR's Face Evaluation toolkit [Kj13]. This toolkit evaluates the matching information and then plots the information in a standardized format.

Fundamentally each system will match at least two biometric templates, one being the stored template and the other being the new users, to produce a score which will decide acceptance or rejection. This match score is a standardized number that shows the likeness between two templates. Both genuine users and impostors are used to evaluate a system. Ideally all genuine users should be accepted while all impostors should be rejected. Important metrics to note are as follows: true accept rate (TAR), true reject rate (TRR), false accept rate (FAR), and finally false reject rate (FRR). True accept rate is the ratio of genuine users whom have been accepted, while the true reject rate is the ratio of impostors who have been correctly rejected. The false accept rate is the ratio of genuine users who are rejected and finally the false reject rate is the ratio of impostors who are mistaken as genuine matchers, i.e. the system grants.. A user is rejected or accepted by comparing the match score to a match threshold. The match threshold is an arbitrary number that each system is tuned to achieve the results it requires. Each system evaluated here used the default threshold values for identification matching.

## 6. Experiments

### 6.1 All to all verification

The first experimental scenario designed for this work mimics an access control. The purpose of the All-to-All Verification experiment is to determine how effective face verification performs when matching between temporally displaced non-ideal images. This experiment compares all images within the ITWCC dataset against all other images.

Images of the same individual are matched against the same individual and all others. The Access Control Scenario was conducted to understand how effective, or ineffective, the selected algorithms perform for matching. The entire ITWCC dataset was used in this scenario to generate 2,905,320 matches, with 10,652 genuine matches. Table 2 list the match matrix of all-to-all comparison.

### 6.2 Young to old identification

Experiment two is an identification task to explore how aging will effect identification performance. This experiment attempts to setup a scenario in which an end-user of a photo tagging tool, such as Facebook, Picasa, etc., would begin adding images over a span of time. The ITWCC dataset is used in this experiment similarly to the first experiment; however, only the first image of each person is used for the gallery and all other images for the individual are used as probes: the youngest image is matched to all of its elder images. The average age of the enrolled faces for all 304 subjects was 10.21 years with a standard deviation of 3.98 years. The minimum age of the gallery was 5 months old. The remaining images were then placed in the probe set; the average age of the probe set was 14.43 years of age , which represents the next chronological age image for every subject. Table 2 list the match matrix of old-to-young comparison.

### 6.3 Augmented young-to-old identification

The augmented Young-to-Old identification experiment further extends the last experiment by increasing the gallery size. The gallery is augmented with both the CASIA Twins (1,234 images) and the Labeled Faces in the Wild datasets (13,233 images) [Sz10], [Sz10]. By expanding the gallery with these datasets, the scenario will be closer to a real world situation in which a user would upload additional data to match against. This scenario is expected to be much more challenging for identification across non-ideal images. All of the CASIA Twins and LFW are added to the gallery resulting in a gallery size of 14,764 images (14,467 from CASIA twins and LFE, plus 304 first image ITWCC). LFW does not contain images of celebrities younger than 16 years; however, the data is captured in a similar manner to ITWCC, non constrained, public sourced images. The CASIA Twins dataset does contain child and adolescent data, but it is captured in a slightly less varied means. This dataset is not readily available to the general research community.

Table 2: Experiment Match Matrix  
# - Verification, \*- Identity

Experiment	Genuine Matches	Imposter Matches	Ignore Matches	Total Matches
All to All#	10,652	2,894,668	1,705	2,905,320
Young to Old*	1404	421,200	0	422,904
Aug Young to Old*	1404	20,732,868	0	20,734,272

## 7. Results

To evaluate the performance of the face matchers, True Acceptance Rates (TAR) is compared with the False Acceptance Rates (FAR). Table 3, 4 and 5, shows the performance of FR techniques in All-to-All Verification and the Young-to-Old Identification experiments. It gives an estimation of the accuracy of each algorithm, in particular, how often an impostor gains access to the system vs. how often a true user is accepted.

Table 3: All to All Verification: True Accept Rate at 1% False Accept Rate

		Face Recognition Method					
		Cognitec	SF4	Cohort LDA	LRPCA	LDA	PCA
TAR	@	37%	25%	12.1%	13.5%	12.6%	15%
1%FAR							

Table 4: Rank Identification Performance Results Closed-Set with matching on 304/304 Subjects      Table 5: Rank Identification Performance Results Closed-Set with matching on 304/14,764 Subjects

Algorithm	Young to Old (Exp. #2)			Algorithm	Aug. Young to Old (Exp. #3)		
	Rank-1	Rank-10	Rank-100		Rank-1	Rank-10	Rank-100
Cognitec	25.0%	41.1%	73.7%	Cognitec	0.0%	0.0%	0.6%
S4F	13.7%	32.1%	67.2%	<b>S4F</b>	<b>7.9%</b>	<b>19.2%</b>	<b>32.8%</b>
Cohort LDA	6.6%	19.6%	55.8%	Cohort LDA	3.6%	7.6%	15.9%
LRPCA	6.3%	18.4%	47.9%	LRPCA	4.0%	7.5%	14.8%
LDA	9.9%	23.1%	56.4%	LDA	6.9%	13.6%	26%
PCA	8.4%	23.8%	64.5%	PCA	5.9%	12.0%	25.6%

## 8. Conclusion

This work presents the first study of the impacts of craniofacial morphology for infants through to adolescents on face recognition. This work does not provide a solution; however, it does address the biological phenomenon responsible for making this area of face recognition extremely difficult. In comparison to adult aging, child aging is far more complex due to the changes in the boney structure as well as in the shape and size of the facial components. This work clearly illustrates the difficulty of this problem through the performance metrics against a set of algorithms that have performed reasonably to extraordinarily on other wild datasets.

Three experimental scenarios: Access control, Photo-tagging, augmented photo-tagging were designed to explore the difficulty of non-adult aging. Six algorithms were used to test the hypothesis: Cognitec’s FaceVacs, OpenBR’s S4F, CohortLDA, LRPCA, LDA,

and PCA [Gc14], [Kj13], [Br03], [Ly12], [TP91]. Results on this unique, albeit small dataset, shows that aging on non-adults is a challenging task for facial recognition algorithms.

The most accurate algorithm for verification was Cognitec at a TAR 37.0% at 1.0% FAR and 23.8% at 0.1% FAR. By current standards this level of performance is considered dismal as demonstrated by Klare et al., a true accept rate of 96.3% was achieved on adults with 0-1 years of lapse between images[Cc10]. The identification experiments were far more diasterous with rank-1 identification task ranging from 25% to 6.6% on a closed set of 304 subjects (1704 images). The best performer Cognetic became the worst performer when the gallery was increased to 14,767 subjects by augmenting the gallery with images from LFW and CASIA Twins. Cognetic registered a rank-1, rank-10, and rank-100 performance of less than 1% true match rate. S4F was the best performing algorithm on this test.

We conclude that non-adult facial recognition is a challenge for concurrent face matchers; with the provided dataset, researchers can start exploring the problem space. The authors of this work will continue to augment the dataset and robust the face recognition system for children's.

## 9 References

- [ARP07] A review of the literature on the aging adult skull and face: Implications for forensic science research and applications. In (Forensic Science International), 2007; p. 1–9
- [Bg00] The OpenCV Library. Dr. Dobb's Journal of Software Tools, 2000.
- [BHK97] Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. In (Pattern Analysis and Machine Intelligence): IEEE Transactions , 1997; p. 711–720.
- [Br03] The CSU face identification evaluation system user's guide: version 5.0. Computer Science Department. Colorado State University, 2003.
- [Cc10] Face age estimation using model selection. In (Computer Vision and Pattern Recognition Workshops (CVPRW)): IEEE San Francisco CA, 2010.
- [CCH15] Face Recognition using Cross-Age Reference Coding with Cross-Age Celebrity Dataset. IEEE Transactions on Multimedia. 2015.
- [Cf12] The FG-NET aging database. 2007-11-12. <http://sting.cycollege.ac.cy/~alanitis/fgnetaging/index.Htm> , 2012.
- [EEH] Age and Gender Estimation of Unfiltered Faces. IEEE Transactions on

Information on Forensics and Security.

- [F192] Growth patterns of the nasolabial region: a morphometric study. *The Cleft Palate-Craniofacial Journal*, 1992; pp. 18-324.
- [Gc14] Facevac software developer kit. 2014.
- [Hg07] Labeled faces in the wild: A database for studying face recognition in unconstrained environments.
- [KJ01] Face recognition across time lapse: On learning feature subspaces. In (IEEE Biometrics (IJCB)): International Joint Conference on Biometrics Compendium Washington DC, 2011.
- [Kj13] Open source biometric recognition. In (Biometrics: Theory): Applications and Systems (BTAS) 2013 IEEE Sixth International Conference, 2013.
- [Ks13] Sketch based face recognition: Forensic vs. composite sketches.
- [Lj74] The human face. , Stein and Day, 1974.
- [LT00] Robust face recognition using automatic age normalization. In (IEEE Electrotechnical Conference): MELECON 2000. 10th Mediterranean, 2000.
- [Ly12] Preliminary studies on the good, the bad and the ugly face recognition challenge problem. In (Computer Vision and Pattern Recognition Workshops (CVPRW)): 2012 IEEE Computer Society Conference, 2012.
- [NG14] FRVT: Performance of Face Identification Algorithms. Information Access Division: National Institute of Standards and Technology, 2014.
- [Pp11] An introduction to the good, the bad and the ugly face recognition challenge problem. In (Automatic Face & Gesture Recognition and Workshops): FG 2011. 2011 IEEE International Conference
- [RB05] The Effect of Normal Adult Aging on Standard PCA Face Recognition Accuracy Rates. In (International Joint Conference on Neural Networks): Montreal Canada, July 2005.
- [RB11] What Are Soft Biometrics and How Can They Be Used? In (IEEE Computer); , 2011; pp. 106-108
- [Rk09] Craniofacial aging. In (Wiley Handbook of Science and Technology for Homeland Security), 2009.

- [RT06] Morph: A longitudinal image database of normal adult age-progression.
- [SRK11] VADANA: A dense dataset for facial image analysis. Computer Vision Workshops (ICCV Workshops), IEEE International Conference, 2011; pp. 2175 - 2182.
- [Sz10] A study of multibiometric traits of identical twins. SPIE Defense Security and sensing, 2010.
- [THB00] Comments on Facial Aging in Law Enforcement Investigation. In (Forensic Science Communications), 2000.
- [Tk10] Forensic art and illustration. CRC Press, 2010.
- [TP91] Face recognition using eigenfaces. In (Computer Vision and Pattern Recognition), 1991

# Pose Variability Compensation Using Projective Transformation for Forensic Face Recognition

Ester Gonzalez-Sosa, Ruben Vera-Rodriguez,  
Julian Fierrez, Pedro Tome and Javier Ortega-Garcia  
Biometric Recognition Group - ATVS, EPS, Universidad Autonoma de Madrid  
Avda. Francisco Tomas y Valiente, 11 - Campus de Cantoblanco,  
28049 Madrid, Spain  
{ ester.gonzalezs,ruben.vera,julian.fierrez,pedro.tome,javier.ortega }@uam.es

**Abstract:** The forensic scenario is a very challenging problem within the face recognition community. The verification problem in this case typically implies the comparison between a high quality controlled image against a low quality image extracted from a close circuit television (CCTV). One of the downsides that frequently presents this scenario is pose deviation since CCTV devices are usually placed in ceilings and the subject normally walks facing forward. This paper proves the value of the projective transformation as a simple tool to compensate the pose distortion present in surveillance images in forensic scenarios. We evaluate the influence of this projective transformation over a baseline system based on principal component analysis and support vector machines (PCA-SVM) for the SCface database. The application of this technique improves greatly the performance, being this improvement more striking with closer images. Results suggest the convenience of this transformation within the preprocessing stage of all CCTV images. The average relative improvement reached with this method is around 30% of EER.

## 1 Introduction

Face biometric trait has been established in the biometric recognition field as one of the least intrusive biometric techniques [ARP04]. This is because it does not require any cooperation from the user. Face recognition can be applied to a wide range of different applications, which range from access control, commercial applications, government issued identity documents, up to law enforcement applications.

Although the problem of face recognition under controlled conditions has achieved great enhancements [YST14], there are still challenges to overcome.

The forensic scenario is one of the areas in which face recognition is involved. The crucial issue of this scenario is dealing with the differences of the images to be compared. The most challenging case within the forensic scenario implies a comparison between a high-resolution image, also known as mug shot, against a low-resolution image acquired from a CCTV device. While mug shot images are extracted under controlled conditions of pose, illumination and background, CCTV images are acquired unobtrusively. The CCTV camera is generally a low-resolution device that acquires video without focusing on the subject.

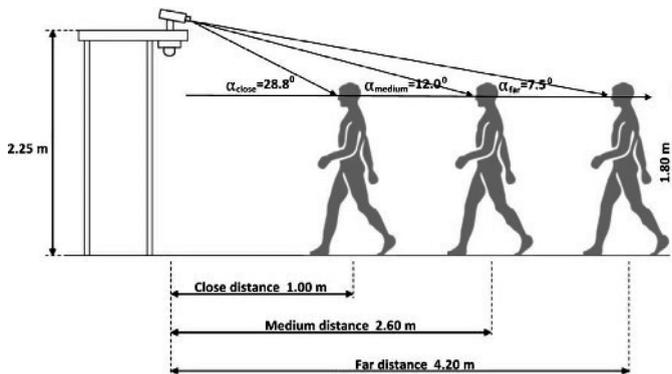


Figure 1: SCface database. There are three different acquisitions distances: close, medium and far. Acquisition angle of each distance calculated for a subject with mean height of 1.80 meters. Figure extracted from [RVRR13]

This fact leads to images with any kind of variation: illumination, pose, expression, occlusion etc. Also, CCTV cameras are commonly situated at ceilings or corners of ceilings, turning towards the floor. As subjects normally walk facing forward, we encounter that the majority of faces detected from the CCTV images suffer from pitch rotation, hindering even more the matching against a frontal image.

Bearing in mind that the last decision in a forensic case is normally done manually by a forensic examiner, any previous automatic procedure to make CCTV images look more similar to mug shot images would be useful to better carry out this manual comparison.

The work developed by *Klontz and Jain* [KJ13] shows a real application of a forensic scenario. In this case, after the terrible incident of the Boston Marathon bombing, an experimental work was carried out to show off the potential capabilities of automatic face recognition system to narrow down the search in this investigation. Results suggest that although state-of-the-art commercial face recognition systems are not yet ready to produce rank-1 results they would help hugely to reduce the number of subjects of the watchlist that are currently being compared manually.

Although there are approaches based on new and more challenging datasets such as [PPB13], neither of them is focused strictly under a forensic point of view. The SCface [KDG11] is a more suitable database for studying the forensic scenario. This database contains both mug shot and CCTV images from three different distances (far, medium and close) from 130 subjects. Fig. 1 shows the situation of the subject for the three different distances. The reader may notice the difference in angle deviation between the position of the camera and the head of the subject, which depends on the height of the subject and the distance to the camera. Hence, for an average height of 1.80 meters, close distance images suffer from a pitch rotation of  $28.8^\circ$ ; medium distance images from  $12.0^\circ$  and far distance images from  $7.5^\circ$ .

The work carried out in [RVRR13] builds a system based on principal component analysis (PCA) and support vector machines (SVM) for the SCface database [KDG11]. Concretely,

three different systems are developed according to the distance to the camera. Among the different experiments carried out in this work, we focus our efforts on one of the most challenging cases in the forensic scenario, which consists of facing mug shot images against CCTV images.

Assessing the results obtained for the mug shot against CCTV images experimental protocol [KDG11], one may think that error should decrease when distance decreases since closer images possess better quality and resolution than images acquired at a far distance. According to those results, this is not the case. We demonstrate in this paper that this is due to the effect of pitch between the camera and the face, which produces errors that directly affect the performance of the system.

In this context the contribution of this paper is to prove the benefits of the use of a projective transformation before comparing mug shot and CCTV images suffering from pitch rotation. This technique leads to frontal images.

The strengths of the proposed technique rely on its simplicity and low computational cost. This technique could be easily used by forensic examiners similar to the work conducted in an *AFIS* (Automated fingerprint identification system [Kom05]). When working with an *AFIS*, forensic examiners first manually mark a set of minutiae points. Then, the *AFIS* system is used to match the feature template associated to this fingerprint sample against a stored database. Likewise, in a hypothetical forensic face recognition case, an examiner could mark a small set of points (e.g. the four points need to define the projective transformation matrix) to ease the task of any face recognition system.

Other related works have tried to compensate general variability sources using probabilistic techniques such as joint factor analysis (JFA) and intersession variability (ISV), reaching promising results [MMM11]. However, our aim with our approach is not to improve the state of the art on face recognition when dealing with general variability sources but to show off the potential of a simple technique to compensate the pitch rotation produced mainly in real forensic caseworks.

This paper is structured as follows. Section 2 presents related work regarding pose compensation and projective transformation. Section 3 describes in depth the SCface database. Section 4 features the preprocessing technique and the projective transformation put forward in this work. Section 5 addresses the experimental protocol followed in our experiments and Section 6 presents the major results obtained in this paper. Finally, Section 7 offers some brief conclusions and future work.

## 2 Related work

### 2.1 Pose compensation techniques

Pose compensation techniques are a matter of growing importance within the face recognition community (see [ZG09] for a survey). Different approaches have been proposed to overcome the difficulties of not having a frontal face. There are general techniques that in-

directly address the pose compensation problem. Approaches based on manifolds or deep belief neural networks are some examples.

As reported in the cited survey, there exist other methods that present algorithms designed specifically to compensate the pose either in 2D or in 3D. Regarding the 2D space, active appearance models and procrustes analyses address the alignment of faces through specific keypoints.

The approach presented in [MVN07] consists of creating a mosaic from frontal and semi profile face images. In this manner, they achieve a more representative subject model without the drawback of storing plenty of images.

3D imaging has produced noteworthy improvements in pose compensation. The most remarkable techniques are based on 3D face models, 3D morphable models and stereo matching.

## 2.2 Projective transformation

Projective transformation has been used in certain applications related to face recognition. The work developed by *Chen and Medioni* [CM01] builds a 3D human face model stemmed from two photographs.

In [HCD14], they manage to estimate the pose of a subject through the projective transformation of the features points of the 3D face model and video sequence. There are also projective transformation-based works to estimate the orientation of the face, useful for human-computer interaction applications [SPD11]. The convenience of using the projective transformation relies on its simplicity.

## 3 Database

This section describes the subset of the SCface database [KDG11] used in our experiments. SCface is a database of static images of human faces with 4.160 images (visible and infrared spectrum) of 130 subjects.

The dataset used in this paper is divided into 6 different subsets: *i*) mug shot images, which are high resolution frontal images and *ii*) five visible video surveillance cameras (CCTV). The images were acquired in an uncontrolled indoor environment with the people walking towards several video surveillance cameras (with different qualities). Further, the images were acquired at three different distances: 1.00 (Close), 2.60 (Medium) and 4.20 (Far) meters respectively (see Fig. 1).

This database is of particular interest from a forensic point of view because images were acquired using commercially available surveillance equipment and under realistic conditions.

There are several landmarks describing the most discriminative parts of the face: eyes,

nose, mouth, eyebrows, etc. In this work, landmarks acquired manually and automatically were used. To extract the landmarks automatically the commercial SDK Luxand Face 4.0<sup>1</sup> was used resulting in a set of 13 points. For the manual approach of landmark detection, a set of 21 facial landmarks were manually tagged by a human bearing in mind the procedure followed by a forensic examiner [JF13].

For this study, the 5 available CCTV images per person and per distance (1950 images in total, 3 distances  $\times$  5 cameras  $\times$  130 persons) plus the 130 mug shot images are considered.

## 4 System description

### 4.1 Preprocessing

First, we obtain the grayscale version of the image. Then, we equalize the grayscale facial image. The face is normalised following the ISO standard<sup>2</sup> with an interpupilar pixel distance (IPD) of 75 pixels by using the eyes coordinates provided (computed either automatically or manually). This step eliminates variations in translation, scale and rotation in horizontal plane, and provides a normalized face in order to compare with a standard size for all faces considered.

### 4.2 Projective transformation

The projective transformation (often called homography) models the geometric distortion that is introduced in a plane when an image is taken with a perspective camera. Under a perspective camera, some geometric properties such as linearity are kept, whereas others such as parallelism are not.

A projective transformation is a two-dimensional transformation that maps two set of points that define a quadrilateral and that belong to two different projective planes.

A projective transformation between two planes is represented as a  $3 \times 3$  matrix acting on homogeneous coordinates of the plane. The general projective transformation  $H$  from one projective plane,  $A$ , to another,  $B$ , is represented as:

$$\begin{bmatrix} b_1 \\ b_2 \\ 1 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ 1 \end{bmatrix} \quad (1)$$

where  $a_1$ ,  $a_2$ ,  $b_1$ ,  $b_2$  are the points of the projective plane  $A$  and projective plane  $B$  res-

<sup>1</sup>Luxand Face SDK, <http://www.luxand.com>

<sup>2</sup>ISO/IEC 19794-5:2011, Information Technology - biometric data interchange formats - part 5: Face image data

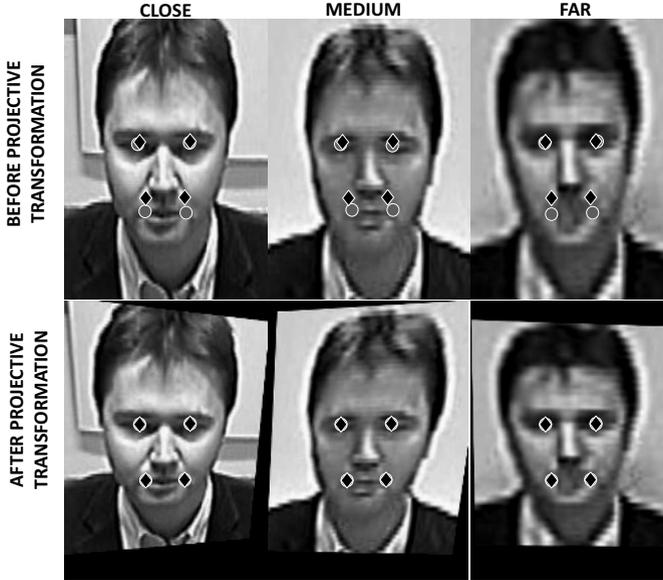


Figure 2: Example of applying the projective transformation to CCTV image from user 1 for all distances. Images presented followed the ISO format. Circles are the landmarks coordinates of the specific image and diamonds are the reference landmarks.

pectively and  $h_{ij}$  with  $i = 1 : 3$  and  $j = 1 : 3$  are the coefficients of the mapping transformation  $H$ .

The procedure followed to obtain the projective transformation was the following. First, we selected four landmarks: 2 eyes centres and 2 mouth vertices. We then average each of these landmarks for a set of mug shots, to obtain a general landmark position. A specific transformation for each CCTV image is then obtained to this reference positioning by solving (1) using digital image warping methods (specifically using the quadrilateral to quadrilateral mapping)<sup>3</sup>.

Finally, we extracted the region of interest of the face from the projected images. Fig. 2 draws an example of the result of applying the projective transformation to the face image for the three different distances: far, medium and close. In the first row, images from the three different distances are plotted, marking with diamonds the reference coordinates and with circles the coordinates of the specific images. Notice the difference of situation between these two sets of points before the transformation. The projective transformation finds the transformation that maps circles to diamonds. The second row of Fig. 2 plots the resulting image after applying this transformation.

<sup>3</sup>The motivation of using mouth and eyes points relies on the fact that projective transformation works only in planar surfaces and, even though the human face is not planar, we may make the assumption that eyes and mouth points are coplanar.

### 4.3 Matching

In what concerns the recognition system itself, principal component analysis (PCA) is applied to the face image over the training set considering the first 200 principal components. Similarity scores are computed in this PCA vector space using a Support Vector Machine (SVM) classifier with a linear kernel.

## 5 Experimental protocol

The database is divided into 3 subsets based on the subject ID: Development (IDS in the range [1-43]), SVM Training (IDS in the range [44-87]), and Test (IDS in the range [88-130]). Each of the sets is comprised of mug shot and CCTV images.

The mug shot versus CCTV images scenario is common in forensic laboratories, and it is very challenging because of the difficulty in finding reliable similarities between probe CCTV images and gallery mug shot images from police records. For this reason, the results obtained in this scenario are especially helpful for the forensic practice.

In this case, each subject model is trained using a single mug shot image (SVM Training Clients) and impostors for the SVM are extracted from the SVM Training set. Then, Test images are taken from the 5 surveillance cameras at 3 different distances: close, medium and far (Test set).

Two different protocols have been defined: *distance-dependent* and *combined protocol*. For the *distance-dependent protocol*, we build the PCA-SVM system for each specific distance: close, medium and far. The analysis of these three configurations is also of great interest for forensics and face biometrics. Additionally, with the *combined protocol* we use jointly all CCTV images regardless of the distance. In this protocol, the PCA matrix transformation is estimated using all images from the three distances belonging to the Development Set. Likewise, the SVM model for each user is modeled having the mug shot image as the positive sample, and the rest of mug shot images and the CCTV images from all distances for the Training set. This latter protocol is more realistic than the *distance-dependent protocol* because a subject to camera distance should be estimated for the other three cases otherwise.

## 6 Experiments

We empirically proved that the projective transformation based on the coordinates of the eyes and the mouth seems to compensate better the pose deviation of the images. Table 1 compares the results obtained for the test set when applying this chosen projective transformation and the baseline system with the original images using manual landmarks. The relative improvement is 16.16%, 31.44% and 39.58% with respect to the baseline method for far, medium and close images respectively. As far, medium and close images suffer

Table 1: Equal Error Rates (EER in %) of the PCA-SVM system on the test set using **manual landmarks**.

Method	FAR	MEDIUM	CLOSE	COMBINED
No Pose Correct.	28.90	31.20	33.10	32.24
Pose Correct.	24.23	21.39	20.00	21.86

Table 2: Equal Error Rates (EER in %) of the PCA-SVM system on the test set using **automatic landmarks**.

Method	FAR	MEDIUM	CLOSE	COMBINED
No Pose Correct.	35.10	31.20	35.40	34.41
Pose Correct.	32.09	24.65	27.33	28.37

from an average pitch rotation of  $7.5^\circ$ ,  $12.0^\circ$ ,  $28.8^\circ$  respectively (for an average height of 1.80 meters), it is straight forward to think that the images with more deviation benefit more from this compensation of pose.

An additional experiment is carried out using automatic landmarks in order to assess the influence of the landmarks detection procedure in conjunction (manual or automatic) with the use of the projective transformation. Table 2 compares the result obtained between the baseline system and the projected images when automatic landmarks are employed. As can be seen, the relative improvement is 8.57%, 20.99% and 22.79% with respect to the baseline method for far, medium and close images respectively. Fig 3 draws the DET curves for the *distance-dependent* and *combined protocol* using manual landmarks.

Comparing results between the transformation defined by manual points and transformations defined by automatic points (Table 1 and Table 2) we conclude that, in both cases, the relative improvement increases when the distance is reduced. However, the improvement is always higher in transformations defined by manual points compared to transformations defined by automatic points. Hence, it may be deduced that the projective transformation is sensitive to the landmarks used. Although automatic points are essential for automatic face recognition systems, they are not so crucial for forensic applications in which the last decision is made by a forensic examiner.

As specified in Section 5, the *combined protocol* is defined with the aim of assessing the influence of this projective transformation in a more realistic scenario in which the distance between the subject and the camera is unknown. The last column of Table 1 refers to this protocol. Specifically, the performance of the original combined system is slightly worse than the average of the three distance-dependent systems. As it was expected, the influence of applying this projective transformation improves also the results of the *combined protocol*, having a relative improvement of 32.19% and 17.51% for manual and automatic points respectively. The average relative improvement for all protocols (far, medium, close and combined) is 29% and 17% for manual and automatic points respectively.

Paying attention now to the results obtained with the images projected, we clearly see

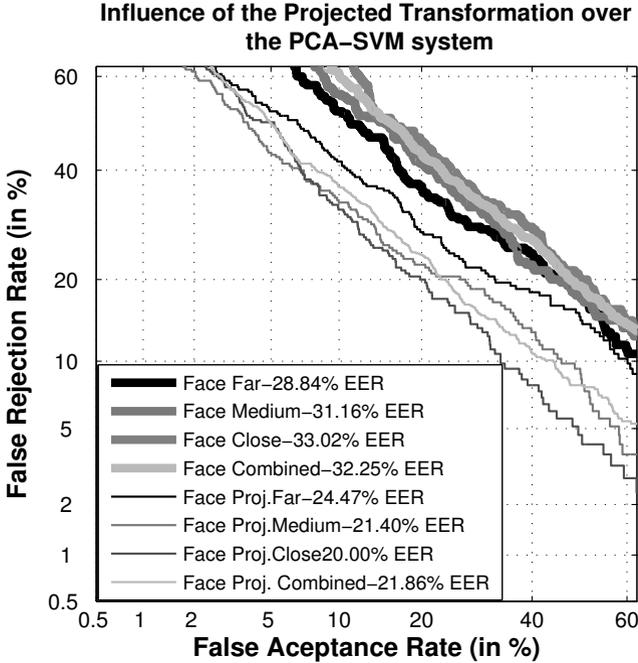


Figure 3: DET curves for the *distance-dependent protocol* and *combined protocol* before and after applying the projective transformation using manual landmarks.

now this improvement of equal error rate when reducing the distance of the subject to the CCTV camera.

## 7 Conclusions and future work

In this work, the specific case of pose compensation has been analysed. It must be noted that the aim with this approach was not to improve the state of the art on face recognition but to show the potential use of a simple technique to compensate the pitch rotation produced mainly in real forensic caseworks.

The relative improvement of this technique is greater for images that suffer higher pitch rotation, such as close images. Concretely, the application of the projective transformation may result in average relative improvements of 29% or 17% for the case of using manual or automatic points respectively. Hence, results suggest that the projective transformation may be used as a preprocessing stage for compensating pitch rotation of CCTV images, especially when comparing them to mug shot images in forensic scenarios. This projective transformation may be easily applied before using COTS face recognition systems, helping this way to narrow down even more the search of suspects to the forensic examiner.

For future experimental work, we aim to use other types of matching techniques and make comparisons with more general pose compensation techniques.

## 8 Acknowledgment

This work has been partially supported in part by Bio-Shield (TEC2012-34881) from Spanish MINECO, in part by BEAT (FP7-SEC-284989) from EU and in part by Cátedra UAM-Telefónica. E. Gonzalez-Sosa is supported by a PhD scholarship from Universidad Autonoma de Madrid.

## References

- [ARP04] A. Jain A. Ross and S. Prabhakar. An introduction to biometric recognition. *IEEE Transactions on Circuits and Systems for Video Technology*, 2004.
- [CM01] Qian Chen and Grard Medioni. Building 3-D Human Face Models from Two Photographs. *Journal of VLSI signal processing systems for signal, image and video technology*, 27(1-2):127–140, 2001.
- [HCD14] C.Liu H. Cheng and A. Dasu. Scale Robust Head Pose Estimation Based on Relative Homography Transformation. *New Mathematics and Natural Computation*, 2014.
- [JF13] R. Vera-Rodriguez F.J. Vega J Fierrez, N. Exposito. Analysis of the variability of facial landmarks in a forensic scenario. In *Proc. of IWBF*, 2013.
- [KDG11] M. Grgic K. Delac and S. Grgic. SCface—surveillance cameras face database. *Multimedia tools and applications*, 2011.
- [KJ13] J. Klontz and A. Jain. A Case Study on Unconstrained Facial Recognition Using the Boston Marathon Bombings Suspects. *Michigan State University, Tech. Rep*, 2013.
- [Kom05] Peter Komarinski. *Automated fingerprint identification systems (AFIS)*. Academic Press, 2005.
- [MMM11] R. Wallace M. McLaren, C.McCool and S. Marcel. Inter-session variability modelling and joint factor analysis for face authentication. In *Proc. of IJCB*, 2011.
- [MVN07] R. Singh M. Vatsa, A. Ross and A. Noore. A mosaicing scheme for pose-invariant face recognition. *IEEE Transactions on Systems, Man, and Cybernetics*, 2007.
- [PPB13] J. Beveridge P.J Phillips and D.S. Bolme. The challenge of face recognition from digital point-and-shoot cameras. In *Proc. of BTAS*, 2013.
- [RVRR13] P. Tome R. Vera-Rodriguez, J. Fierrez and D. Ramos. Identification using face regions: Application and assessment in forensic scenarios. *Forensic science international*, 2013.
- [SPD11] P. Petrov O. Boumbarov S. Panev, I. Paliy and L. Dimitrov. Homography-based face orientation determination from a fixed monocular camera. In *Proc. of IDAACS*. IEEE, 2011.

- [YST14] X. Wang Yi Sun and X. Tang. Deep learning face representation from predicting 10,000 classes. In *Proc. of CVPR*, 2014.
- [ZG09] X. Zhang and Y. Gao. Face recognition across pose: A review. *Pattern Recognition*, 2009.



# Face Tracking using Optical Flow

## Development of a Real-Time AdaBoost Cascade Face Tracker<sup>1</sup>

Andreas Ranftl, Fernando Alonso-Fernandez, and Stefan Karlsson

Embedded and Intelligent Systems Research  
Halmstad University  
Kristian IV:s väg 3, 301 18 Halmstad, Sweden  
andran13@student.hh.se, {stefan.karlsson, fernando.alonso-fernandez}@hh.se

**Abstract.** In this paper a novel face tracking approach is presented where optical flow information is incorporated into the Viola-Jones face detection algorithm. In the original algorithm from Viola and Jones face detection is static as information from previous frames is not considered. In contrast to the Viola-Jones face detector and also to other known dynamic enhancements, the proposed face tracker preserves information about near-positives. The algorithm builds a likelihood map from the intermediate results of the Viola-Jones algorithm which is extrapolated using optical flow. The objects get extracted from the likelihood map using image segmentation techniques. All steps can be computed very efficiently in real-time. The tracker is verified on the Boston Head Tracking Database showing that the proposed algorithm outperforms the standard Viola-Jones face detector.

## 1 Introduction

Viola and Jones introduced a real-time face detector in 2001 [VJ01]. Face detection is performed by applying a classifier on several windows within the image. The windows vary in location and size in order to determine scale and position of the face rather exactly [VJ01].

The Viola-Jones method has neither a way of incorporating temporal constraints nor combining evidence from previous frames to aid the inference. In short, it is a fully static algorithm. Apart from lacking temporal consistency, the cascade classifier also lacks a way to save information about near-positives. Not only may face positions behave erratically, but if a face becomes temporarily distorted so that the very last part of the cascade fails, the detection fails abruptly. This paper investigates a way to extend the Viola-Jones cascade classifier to achieve a likelihood map that is suited for a form of belief propagation over time. For example, it is possible that a face is detected after several likelihood map refreshes even if it does not pass all stages of the cascaded classifier.

---

<sup>1</sup> This paper follows a master thesis which was written within the double degree master program in Embedded and Intelligent Systems of Salzburg University of Applied Sciences, Austria and Halmstad University, Sweden.

The real-time optical flow enhanced AdaBoost cascade face tracker aims at calling the Viola-Jones algorithm at every 20<sup>th</sup> frame. In the frames in between, face detection is done by processing the likelihood map, which is interpolated with optical flow information.

## 2 Face Detection Performed by the Viola-Jones Algorithm

The algorithm developed by Viola and Jones is based on a cascade of classifiers using Haar-like features, built up in an AdaBoost-based training process by both extracting features from face images and non-face images. The algorithm achieves real-time performance through the cascade structure of the classifiers. Each window constitutes a hypothesis, which gets discarded as soon as a stage is not passed. Each classifier is designed to cancel the evaluation of windows which contain no faces as soon as possible. If a window passes all classification stages it is considered to contain a face [VJ01]. The method of Viola and Jones was improved by Lienhart et al. in 2002 by introducing diagonal Haar-like feature sets [LKP03]. Multi-Block Local Binary Patterns (MB-LBP) are the currently used type of features for classification and they are also used in this work [ZH07]. The Viola-Jones algorithm does not preserve information about near positives. Furthermore, it does not consider previously obtained information.

## 3 Optical Flow Enhanced AdaBoost Cascade Face Tracker

When detecting faces within an image sequence with the Viola-Jones algorithm every frame is handled separately. This means that the detection process, by shifting different sized windows over the entire image and evaluating them, is done on every single frame. As there is no temporal information taken into consideration, the resulting face bounding boxes appear to be unstable. The boxes slightly change in size and position although the face does not move, and on occasion the tracking is lost altogether for a few intermediate frames. In order to overcome these problems the proposed algorithm works with a likelihood map that saves information about near positives as well as previously computed data.

### 3.1 Basic Ideas and Flow Chart of the Proposed Face Tracker

In order to track the faces, the algorithm follows the flow chart shown in **Fig. 1**. In the initialization phase the program opens the video input, loads the cascade classifiers and builds the initial likelihood map by utilizing the modified Viola-Jones algorithm. A likelihood map is used due to the fact that it offers the possibility to recognize an undetected face from the Viola-Jones algorithm after several refreshes.

Within the algorithm frame per frame is processed and optical flow computation is done. The current likelihood map gets then warped with the results from optical flow. The employment of optical flow prevents a face from being lost, as long as it has passed a high number of classification stages when establishing or refreshing the likelihood map. Additionally, the size of the tracked face area does not change erratically.

If the frame obtained from the input is a refresh frame, a modified version of the Viola-Jones algorithm is applied to it. There a temporary likelihood map is built which influences the warped likelihood map by recursive filtering. For the purpose of extracting a face from the current frame, a binary likelihood map is created and then segmenting is done by finding edges within the binary image.

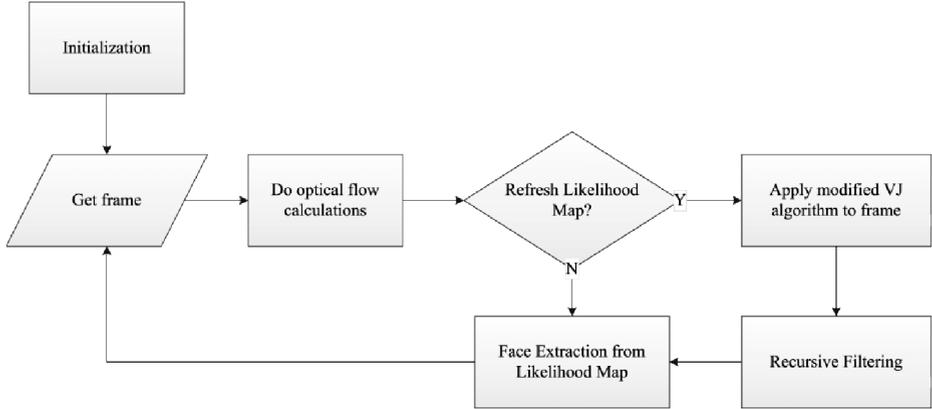


Fig. 1. Flow chart of the developed algorithm.

### 3.2 Likelihood Map Setup

The likelihood map is built from the intermediate results of the Viola-Jones algorithm. In particular, the number of stages of the classification cascade passed by each detection window is used to form it. This means that the likelihood provides for every pixel a value that indicates the probability of being a face located at the respective position in the original frame.

There are different ways of setting up a single 2-D likelihood map from the Viola-Jones results. It is important to ensure that the possible maximum energy of each detection window scale used in the Viola-Jones algorithm is the same. Otherwise different scales would be weighted differently in the resulting likelihood map.

In order to utilize previously obtained information, a new likelihood map  $\mathbf{L}_t$  at time  $t$  is formed by recursive filtering, see (1).  $\mathbf{L}$  describes the current likelihood map obtained from the Viola-Jones algorithm and  $\mathbf{L}_i$  is the interpolated likelihood map at same time  $t$ .

$$\mathbf{L}_t = (1 - \alpha) * \mathbf{L} + \alpha * \mathbf{L}_i \quad (1)$$

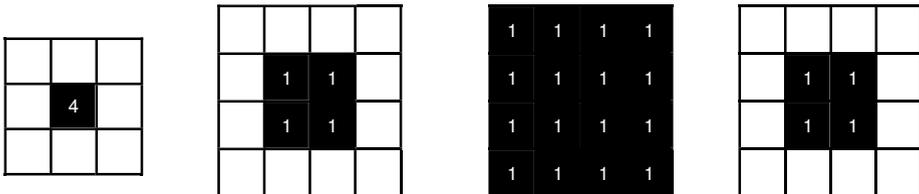
The computation time of building the likelihood map can be lowered by taking advantage of a reject level threshold. This means that only windows, which pass a certain number of stages of the cascaded classifier, contribute to the likelihood map.

#### Window Center Orientated Likelihood Map

One way of building a likelihood map is to add the result of each detection window to the pixel in the likelihood map corresponding to the center pixel of the respective

window, see **Fig. 2** for a  $3 * 3$  pixels example window. By doing so, information about the size of the face is lost. The likelihood value for an odd sized window is calculated by multiplying the number of passed stages by 4. Otherwise the influence on the likelihood map of even and odd sized windows would not be equal.

As a window with even sized height and width has no single center pixel, the  $2 * 2$  pixels center is set to the rejection level value, see **Fig. 2** for an example.



**Fig. 2.** Principle of setting the likelihood value for an odd sized window (left) and for an even sized window (middle left) when using the window center orientated likelihood map approach.

The windows on the right are examples for the area orientated approach. Every pixel of the likelihood map corresponding to a detection window is set to the number of passed stages (middle right). The very right example uses a shrinking factor of 0.5. Values represent the factor with which the number of passed stages of the respective detection window is multiplied.

As previously mentioned, every detection window is weighted the same in the likelihood map. This requires us to compensate the step size caused by a scaling factor used in the detection process. The compensation of the detection window scale is done by weighting the counter of passed stages  $n_{levels}$  with the respective window width  $w_{window}$ , see (2) for an odd sized window.

$$\mathbf{L}_n = \mathbf{L}_{n-1} + n_{levels} * 4 * w_{window} \quad (2)$$

### Window Area Orientated Likelihood Map

Another approach for building a likelihood map is to set every pixel of a detection window to the corresponding Viola-Jones result and then adding all the windows up, see **Fig. 2**. By doing so, we do not have to care about even or odd sized windows and we do not have to care about weighting of results of different window sizes. Also the information about the size of the face is preserved. It could be shown that there is nearly no difference in computation time when using a reject level threshold of 15. Compared to the original Viola-Jones implementation the likelihood map building methods need 10% more computation time.

When setting the respective pixels in the likelihood map to the correct value, a shrinking factor was introduced. This shrinking factor shrinks the area which should be set to the Viola-Jones result. The principle of applying a shrinking factor is based on the fact that detection windows are bigger than the faces outlined by them. When interpolating the likelihood map with the flow map there are only motion vectors for the pixels of the face. Therefore setting the whole window area to the respective Viola-Jones result would lead to an inaccurate likelihood map after interpolating with optical flow results.

### 3.3 Face Extraction from the Likelihood Map

In order to differentiate between face-containing and non-face-containing regions of the image a threshold is applied to the likelihood map. Note that the probabilities which are represented by the pixel intensities of the likelihood map can vary even for the same face region, depending on face detections within neighboring windows and windows with different sizes that detect the same face.



**Fig. 3.** The left image is the original frame. The middle left image represents the generated likelihood map. Different probabilities are shown by different gray shades. The high values for the two faces can be observed clearly in the likelihood map. The image in the middle is the binary likelihood map which is the result of applying a threshold of 65 to the original likelihood map. The middle right image illustrates the segmented likelihood map and the very right image shows the computed bounding boxes of the face tracker marked in the original frame.

In order to label the face regions uniquely and to visualize the outcome, edge detection is performed on the binary likelihood map. As the contours of the face regions are known, the segmentation algorithm searches for the outer points. Each face region in the likelihood map equals roughly a rectangle. The outer points are stretched by the inverse of the shrinking factor  $s$  which was used for downscaling face-containing windows when putting them into the likelihood map. By shrinking the windows with a factor that is small enough, the face-containing regions do not overlap in the likelihood map even when the faces are near to each other. The resizing is done by applying equations (3) and (4) where  $x_{face}$  and  $y_{face}$  represent the coordinates of the upper left corner of the bounding box which is used to mark a face. The variables  $width_{face}$  and  $height_{face}$  are the differences of the max and min values of  $x$  and  $y$  respectively.

$$x_{face} = x_{min} + \frac{width_{face}}{2} - \frac{width_{face}}{2*s} = x_{min} - width_{face} * \frac{1-s}{2*s} \quad (3)$$

$$y_{face} = y_{min} + \frac{height_{face}}{2} - \frac{height_{face}}{2*s} = y_{min} - height_{face} * \frac{1-s}{2*s} \quad (4)$$

## 4 Tracking During Occlusion

Under certain circumstances the developed face tracking algorithm is able to track faces during partial and complete occlusion. Given that the occlusion is encompassed fast enough (e.g. a passing car) we can take advantage of the fact that the utilized Farneback dense flow technique is invariant to very fast motion (see **Fig. 4**).



**Fig. 4.** The optical flow face tracker is able to track faces under partial occlusion (left) and also under complete occlusion (right).

## 5 Results

The proposed face tracking algorithm was implemented in C++ utilizing the OpenCV library in version 2.4.8.0 with activated parallelization. The face tracker was evaluated on the Boston Head Tracking Database [LSA00]. We used the 45 videos which were recorded under uniform light conditions. Ground truth information was made available through the UVAEYES annotations which represent the positions of the eyes (see **Fig. 5** as example) [VG09].

The method with which the presented approach is compared is the original Viola-Jones face detector. For terms of comparison, the Viola-Jones method performs face detection on every single frame. Within the developed optical flow face tracker, the reject level threshold for accepting windows to contribute to the likelihood map was set to 15. The shrinking factor was set to  $1/3$ .

### 5.1 Measurement of Detection Rate and Accuracy

In order to measure the accuracy of the tracking algorithm the Euclidean distance of the computed center point to the real center point of the face is calculated. The real center point  $[x, y]^T$  is computed by utilizing (5) and (6), which calculate the middle point between the eyes ( $x_1, y_1$  and  $x_2, y_2$  respectively) and add 10 pixels in  $y$  direction (origin is upper left corner of the image) for setting the coordinates of the face center. This works quite well as there is not much difference in the sizes of faces.

$$x = x_1 + \frac{x_2 - x_1}{2} \quad (5)$$

$$y = y_1 + \frac{y_2 - y_1}{2} + 10 \quad (6)$$

The evaluated algorithms output a bounding box. The center of this bounding box is set as the face center point. If the Euclidean distance of a center point computed by an algorithm to the real center point is higher than 20 pixels, the detection is classified as a false positive. If there is no detection, it is counted as a false negative.

If there are multiple detections, then all the distances are measured and the nearest one is chosen as valid attempt as long as it is within the threshold defined above. The other false detections cause the number of false positives to increase. If all detections are outside the threshold circle, they are all added to the false positives count.

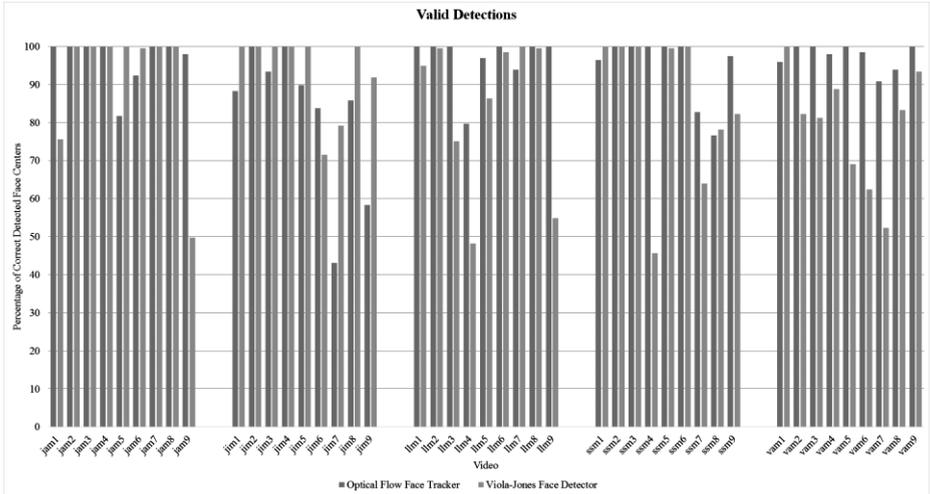


**Fig. 5.** Example frames (23, 92, 100, 150 and 188) of the video jam1.avi from the Boston Head Tracking Database. The red dots indicate the ground truth positions of the eyes. The green dots represent the computed face centers. The blue rectangles represent the face bounding boxes returned by the optical flow face tracker. In frame 23 a refresh is done as also the Viola-Jones bounding box (white) is visible.

## 5.2 Detection rate

The detection rate  $r$  is calculated by summing up the valid detections  $d$  within a video and dividing the obtained sum by the number of frames  $nframes$  (see (7)).

$$r = \frac{1}{nframes} * \sum_{i=1}^{nframes} d_i \quad (7)$$



**Fig. 6.** Illustration of percentage of correctly detected face centers per video.

Taking all videos of the database into consideration the optical flow face tracker shows a better detection rate. The average detection rate of all videos is 79.15% for the optical flow face tracker and 73.71% for the Viola-Jones face detector. **Fig. 6** shows the average detection rate per video of the Boston Head Tracking Database.

## 5.3 Accuracy and Robustness

When measuring the average accuracy of valid detections we observed that the Viola-Jones algorithm (2.56 pixels offset) is on average in all videos more accurate than the developed face tracker (5.99 pixels offset). It should be considered, however, that the

Viola-Jones method loses the face in several frames. These frames are not considered in the computation of the average accuracy of the Viola-Jones algorithm. The average offset in pixels is illustrated in Fig. 7.

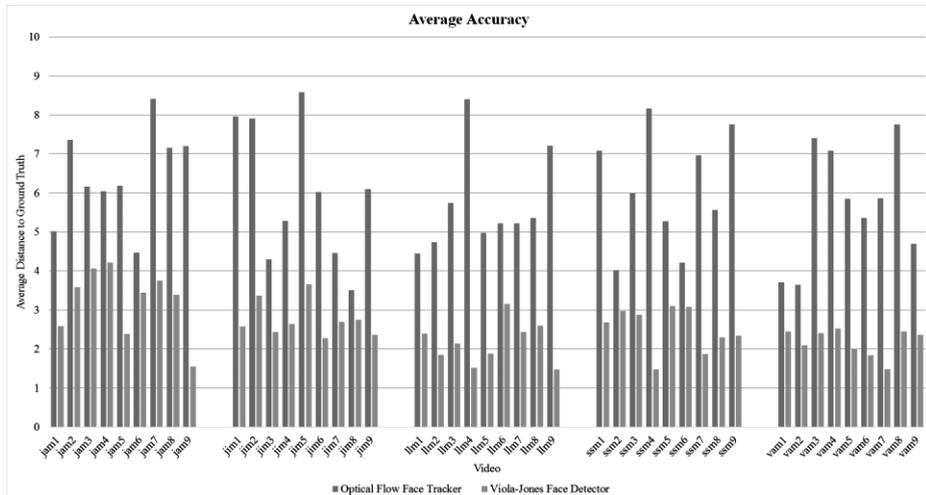


Fig. 7. Average inaccuracy of the optical flow face tracker and the Viola-Jones algorithm.

On average the Viola-Jones algorithm outputs a higher amount of false detections. In particular, the Viola-Jones algorithm returned on average 26 false detections per video and the optical flow face tracker 18 false detections. This is caused by certain head poses that make it impossible for Viola-Jones to detect the face due to the utilized classifier which was trained with upright face images. Therefore these false detections are mostly false negatives.

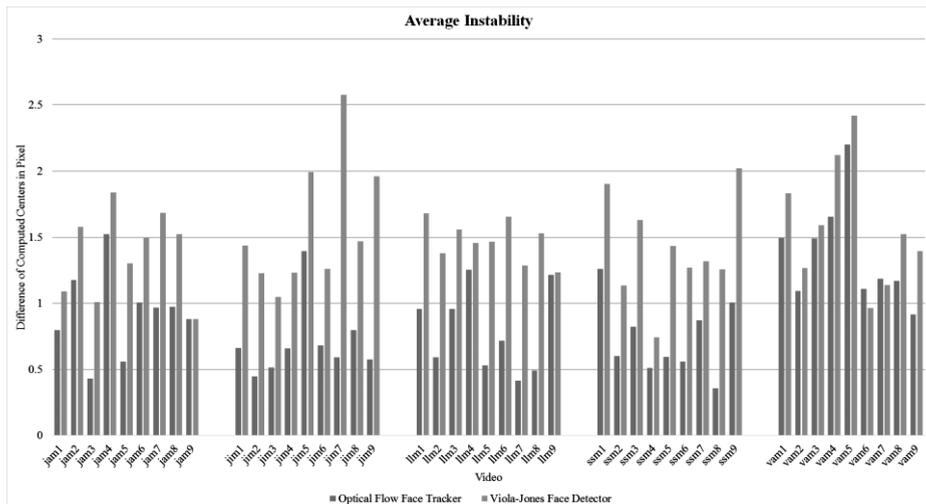
By design the optical flow face tracker will always detect a face as long as it was initialized with one. If the Viola-Jones algorithm does not detect a face in the refreshment frame it is re-executed until it detects one. If Viola-Jones produces a false negative no refresh on the likelihood map is done, which causes the optical flow face tracker to become inaccurate and produce false positives with increasing time. If Viola-Jones outputs a false positive, the error is propagated as this false detection is followed by the face tracker. By increasing the threshold for segmentation of the likelihood map, the false positive rate can be lowered. However, it is possible that the accuracy gets worse. The Viola-Jones detector returned only a few false positives.

### 5.4 Stability

Erratic movements within the face tracking process are unpleasing as faces are only moving relatively slow in practice. The stability is measured by taking the Euclidean distance  $e$  between the face centers of two successive frames with valid detections (see (8)). The distance  $e$  is a direct measure for the instability. Center points are considered as valid if their distance to ground truth is less than 20 pixels.

$$e = \sqrt{(x_t - x_{t-1})^2 + (y_t - y_{t-1})^2} \quad (8)$$

In general the optical flow face tracker shows less erratic movements. The average on all videos equals 0.9 pixel for the optical flow face tracker and 1.5 pixels for the Viola-Jones face detector. **Fig. 8** shows the average instability per video in pixels.



**Fig. 8.** Average instability of both algorithms applied to all videos of the database.

## 5.5 Speed

The speed of the algorithms is measured in computation time in milliseconds. Compared to the Optical Flow Face Tracker, the Viola-Jones algorithm needs less time for computation. It has to be outlined that the average CPU usage is higher when executing the Viola-Jones algorithm. This is caused by the high parallelization of the OpenCV implementation of the Viola-Jones method. In contrast, the computation of the Farneback optical flow is mostly done by one core which causes the average CPU usage to be low, but the computation time to be high. The optical flow computation takes a majority of the computation time of the face tracking algorithm (in average 11.33 milliseconds). The rest of the computation time of the algorithm is much lower (1.67 milliseconds) compared to the Viola-Jones algorithm (6 milliseconds).

The Viola-Jones algorithm and our algorithm have the same complexity due to being the Viola-Jones algorithm a part of the proposed system. However, because the Viola-Jones algorithm is called not at every frame, one can expect a constant factor of speed increase. This speed increase is given by how often the likelihood map should be updated.

## 6 Conclusion

We presented a novel real-time face tracker which utilizes a modified version of the Viola-Jones algorithm for face detection. In contrast to a pure Viola-Jones face detector the developed approach calls a modified Viola-Jones method only at every 20th frame for refreshing a likelihood map. The likelihood values within this map are dependent on the numbers of classification stages which detection windows pass when the classification is done.

In order to track the faces the likelihood map is interpolated with a flow map computed by the Farneback dense optical flow method. The resulting likelihood map of the modified Viola-Jones algorithm contributes to the system's likelihood map by recursive filtering.

Compared to the original Viola-Jones implementation, the likelihood map approach enables faces to be detected even when they do not pass all of the stages of the cascade classifier. Due to the fact that the likelihood map is never discarded completely, a region gets also a high value in the likelihood map if the respective window passes for example 17, 18 or 19 stages within several executions of the modified Viola-Jones method. Another advantage of the developed face tracker is that it can also track faces under partial and complete occlusion.

The developed face tracking algorithm and the original Viola-Jones face detector were evaluated on the Boston Head Tracking Database. The developed face tracker achieved a higher detection rate than the Viola-Jones face detector. Furthermore, the optical flow face tracker showed less erratic movements of detections.

The developed tracker relies on the Viola-Jones algorithm which means that an error of Viola-Jones algorithm during initialization or refresh of the likelihood map gets propagated. By utilizing also other methods for face detection this effect could be minimized (e.g. execute a face detection method with a low computation time only on the extracted face areas in order to check if the tracker has lost the faces or not).

## Bibliography

- [LKP03] Lienhart, R.; Kuranov A., Pisarevsky V.: Empirical Analysis of Detection Cascades of Boosted Classifiers for Rapid Object Detection. In DAGM 25th Pattern Recognition Symposium, 2003.
- [LSA00] La Cascia M., Sclaroff S., Athitsos V.: Fast, Reliable Head Tracking under Varying Illumination: An Approach Based on Registration of Texture-Mapped 3D Models. In IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000.
- [VG09] Valenti R., Gevers T.: Robustifying Eye Center Localization by Head Pose Cues. In IEEE conference on Computer Vision and Pattern Recognition, 2009.
- [VJ01] Viola P., Jones M.: Rapid Object Detection using a Boosted Cascade of Simple Features. In IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2001.
- [ZH07] Zhang, L. et al.: Face Detection Based on Multi-Block LBP Representation. In ICB International Conference, 2007.

# On Reducing the Effect of Silhouette Quality on Individual Gait Recognition: a Feature Fusion Approach

Ning Jia, Victor Sanchez, Chang-Tsun Li

Dept. of Computer Science, University of Warwick, UK  
n.jia@warwick.ac.uk, vsanchez@dcs.warwick.ac.uk, c-t.li@warwick.ac.uk

Hassan Mansour

Mitsubishi Electric Research Laboratories, Cambridge, MA, USA  
mansour.hassan@gmail.com

**Abstract:** The quality of the extracted gait silhouettes can hinder the performance and practicability of gait recognition algorithms. In this paper, we propose a framework that integrates a feature fusion approach to improve recognition rate under this situation. Specifically, we first generate a dataset containing gait silhouettes with various qualities based on the CASIA Dataset B. We then fuse gallery data with different qualities and project data into embedded subspaces. We perform classification based on the Euclidean distances between fused gallery features and probe features. Experimental results show that the proposed framework can provide important improvements on recognition rate.

## 1 Introduction

As the only biometric source that can be acquired at a distance, gait recognition has drawn great attention over the past decade. Various algorithms have been proposed for individual gait recognition, which can be classified as model-based or model-free approaches. Model-based approaches rely on the extraction of parameters from the subjects' body and walking cycle to construct a structural model of human motion [WZW<sup>+</sup>12]. Model-free approaches, on the other hand, rely on spatio-temporal representations of gait, which may be directly obtained from the acquired gait sequences. These approaches, in general, provide high recognition accuracy with low computational cost, as they usually use binary silhouette images to represent gait sequences.

Unfortunately, gait is not as reliable as other biometric traits such as fingerprints and iris [HB06][SPL<sup>+</sup>05]. Factors such as age, clothes, walking surfaces, viewing angles, and health condition may result in poor recognition performance. Furthermore, recognition efficiency may be hindered if the associated gallery and probe gait silhouettes are acquired under different situations. The quality of the gait silhouette can be influenced, for example, by the background environment when capturing gait sequences and the accuracy of the

segmentation method used to detect the gait silhouette. A detailed review of gait databases and algorithms can be found in [MMN<sup>+</sup>15].

The effect of the gait silhouette quality on the performance and practicability of model-free gait recognition algorithms is an important issue that has been less intensively studied and only a limited number of solutions are reported in the literature. Sarkar et al. [SPL<sup>+</sup>05] discuss several cases when gait silhouette segmentation errors occur in the HumanID Gait Challenge Problem data set due to the shadow of the individuals, varying lighting conditions and moving objects in the background. In [LS05], Liu and Sarkar observe that the drop in gait silhouette segmentation quality may lead to a decrease in recognition accuracy. They also observe that if gallery and probe gait sequences are captured under the same conditions, and are segmented by the same method, the recognition accuracy may be high even if the data quality is poor. Zhang et al. [ZPCF10] address the issue of poor recognition accuracy when low-resolution gait silhouettes are used. The authors propose to combine super-resolution with multi-linear tensor-based learning without parameters (SRMTP) to overcome this problem. However, they test their algorithm on artificial dataset, and thus the issue of silhouette quality remains unexplored in practical scenarios. Recently Shaikh et al. [SSC14] propose a partial silhouette-based approach that extracts the hand dynamics as gait signatures, and claim to be efficient on incomplete or distorted silhouettes. However, no result is provided about the performance on low-quality silhouette.

In this work, we consider the situation in which the gait data related to an individual to be recognised (probe data) are not captured under ideal conditions, and therefore the associated gait silhouettes may be noisy and inaccurately segmented; whereas the stored gait data (gallery data) are well segmented, or vice versa. This is a common situation encountered in practice; for example, when the probe data is captured using CCTV cameras at low resolution and poor quality, but the gallery data is previously captured under ideal conditions and it is not feasible to re-capture the probe data under the same ideal conditions. Based on this scenario, we employ various segmentation algorithms to generate different silhouette quality data using sequences from the CASIA Dataset B. A subspace learning method is then used to find a low dimension feature subspace. A fusion strategy is employed to fuse gallery data of different quality levels, and the fused gallery data is then matched with the probe data in the feature space. We employ two popular subspace learning methods, namely, Linear Discriminant Analysis (LDA) [WC11], and Locality Preserving Projection (LPP) [BHK97], to confirm the improvement brought by fusion approach. Experimental results show that the fusion strategy attains high recognition accuracy, making it a promising solution to reduce the negative effects of poor gait silhouette quality on individual recognition.

The rest of the paper is organized as follows. Section 2 briefly reviews the subspace learning methods used in this work. Section 3 details the proposed framework. Section 4 presents the experimental results and related discussions. Finally, Section 5 draws conclusions.

## 2 Subspace learning methods

Gait Energy Image (GEI) [HB06] is an effective gait representation for individual recognition as it reduces computational cost as well as storage space. This work employs such GEIs. Let us assume there are  $N$  gait silhouettes, represented as binary images, in one gait period. A GEI  $G(x, y)$  is defined as  $G(x, y) = \frac{1}{N} \sum_{k=1}^N I_k(x, y)$ , where  $I_k(x, y)$  is the  $k$ th binary image, and  $(x, y)$  denotes the pixel coordinates. Examples of GEIs are shown at the rightmost column of Fig. 2.

Consider  $n$  GEI samples that are stored as  $d$ -dimensional column vectors in a matrix  $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ ,  $\mathbf{x}_i \in \mathfrak{R}^d$ ,  $i \in \{1, 2, \dots, n\}$ . Let  $W$  be the transformation matrix that projects the original space onto an  $r$ -dimensional subspace, where  $d \gg r$ . The new feature matrix in the subspace is denoted as  $\mathbf{Y} = \{\mathbf{y}_1, \dots, \mathbf{y}_n\}$ , where  $y_i \in \mathfrak{R}^r$ . The transformation matrix for each element is given by  $\mathbf{y}_i = W^T \mathbf{x}_i$ ,  $i \in \{1, \dots, n\}$ . Matrix  $W$  varies according to the subspace learning method used.

### 2.1 Dimensionality Reduction: PCA

PCA is used as an approach to avoid singularities in further covariance matrix calculations [WC11], for example in LDA and LPP. PCA seeks a compact representation of patterns in a feature subspace. The columns of the PCA transformation matrix  $W_{PCA}$  are calculated by solving the eigen-decomposition problem  $\lambda_i \mathbf{e}_i = S \mathbf{e}_i$ , where  $\lambda_i$  and  $\mathbf{e}_i$  are the corresponding eigenvalues and eigenvectors, respectively, and  $S = \frac{1}{n} \sum_{i=1}^n (\mathbf{x}_i - \mu)(\mathbf{x}_i - \mu)^T$  is the covariance matrix of the original sample matrix  $X$ , where  $\mu$  is the sample mean,  $\mu = \frac{1}{n} \sum \mathbf{x}_i$ ,  $i \in \{1, 2, \dots, n\}$ . Matrix  $W_{PCA}$  is then composed by column eigenvectors corresponding the  $r$ th highest eigenvalues;  $W_{PCA} = \{\mathbf{e}_1', \mathbf{e}_2', \dots, \mathbf{e}_r'\}$ , where  $\mathbf{e}_j'$ ,  $j \in [1, r]$  is the  $j$ th eigenvector.

### 2.2 Discriminant Analysis: LDA

Compared to PCA, LDA embeds discriminant power between different classes in the feature subspace, which makes it a supervised subspace learning method suitable for multi-class learning problems. Assuming there are  $c$  classes in  $X$ , with  $n_l$  samples in subset  $\mathcal{X}_l$ ,  $l \in \{1, 2, \dots, c\}$ , so that  $n = \sum_{l=1}^c n_l$ ; the within-class scatter matrix  $S_W$  is then defined as:  $S_W = \sum_{l=1}^c \sum_{\mathbf{x} \in \mathcal{X}_l} (\mathbf{x} - \mu_l)(\mathbf{x} - \mu_l)^T$ , and the between-class scatter matrix  $S_B$  is defined as:  $S_B = \sum_{l=1}^c (\mu_l - \mu)(\mu_l - \mu)^T$ , where  $\mu_l$  is the mean of the samples in class  $l$ , and  $\mu$  is the mean of all samples. In order to maximize between-class scatter while minimizing within-class scatter after projection, the following criterion is used:

$$W_{LDA} = \arg \max_W \frac{|W^T S_B W|}{|W^T S_W W|}$$

where  $W_{LDA}$  is the transformation matrix, whose columns are the generalised eigenvectors  $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_r\}$  that correspond to the largest eigenvalues in  $S_B W_{LDA} = \lambda_i S_W W_{LDA}$ .

### 2.3 Manifold Learning Method: LPP

LPP tends to preserve the local data structure after projecting the data onto a subspace [HN03]. It first constructs an adjacency graph  $G$  to model the local structure of the samples. The adjacency graph has  $n$  nodes, with node  $i$  corresponding to  $\mathbf{x}_i$  in  $\mathbf{X}$ . A pair of nodes  $i$  and  $j$  are connected if  $\mathbf{x}_i$  and  $\mathbf{x}_j$  are close in the space. The elements of the weighted similarity matrix  $A$ , which specifies the similarities among nodes in  $G$ , are formulated as follows:

$$A_{ij} = \begin{cases} \exp\left(\frac{-\|\mathbf{x}_i - \mathbf{x}_j\|^2}{t}\right), & \text{if nodes } i \text{ and } j \text{ are connected,} \\ 0, & \text{otherwise.} \end{cases} \quad (1)$$

The heat kernel parameter  $t \in \mathfrak{R}$  can be determined empirically; if  $t$  is very large,  $\exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2/t) = 1$  and matrix  $A$  comprises binary weights. Two possible ways exist to determine if nodes are *close*:

1.  $K$  nearest neighbours: if  $\mathbf{x}_i$  is among the  $K$  nearest neighbours of  $\mathbf{x}_j$ , or vice versa;
2.  $\epsilon$ -nearest neighbours: if  $\|\mathbf{x}_i - \mathbf{x}_j\|^2 < \epsilon, \epsilon \in \mathfrak{R}$ .

The eigen-decomposition problem of LPP is generalized as

$$X L X^\top W_{LPP} = \lambda X D X^\top W_{LPP},$$

where  $D$  is a diagonal matrix with  $D_{ii} = \sum_j A_{ij}$ , and  $L$  is the Laplacian matrix  $L = D - A$ . The Laplacian of the graph is an approximation of the Laplace-Beltrami operator. The transformation matrix  $W_{LPP} = \{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_r\}$ , and  $\{\lambda_1, \lambda_2, \dots, \lambda_r\}$  are the corresponding  $r$  smallest eigenvalues. The feature subspace created by  $W_{LPP}$  can preserve an intrinsic geometric structure of the manifold samples [HN03][HYH<sup>+</sup>05][BN03]. LPP can perform supervised learning by assigning a weight equal to 0 to all between-class similarity matrix values. The total similarity matrix  $A$  is then given as follows:

$$A = \begin{bmatrix} A_1 & \cdots & 0 \\ \vdots & \ddots & 0 \\ 0 & 0 & A_c \end{bmatrix} \quad (2)$$

In our experiments, we employ supervised LPP.

## 3 Proposed Framework

The block diagram of the proposed framework is shown in Fig. 1. We first generate the GEIs with different qualities for the training, gallery and probe data. Using the training data, we compute the transformation matrix corresponding to the subspace learning

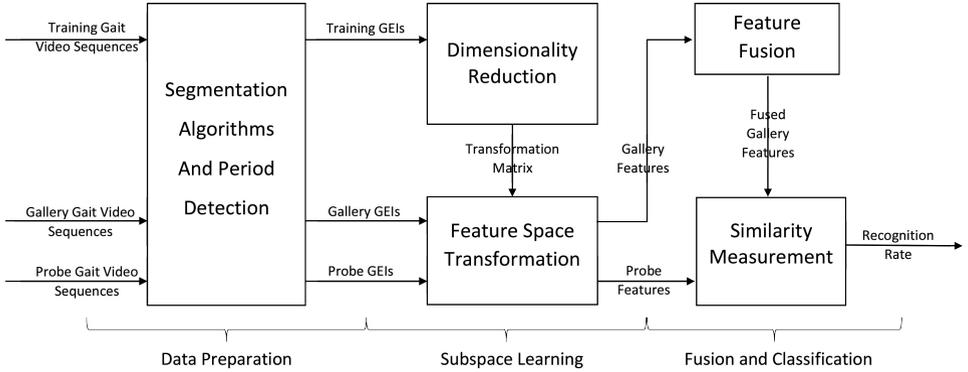


Figure 1: Block diagram of the proposed framework.

Table 1: Gait silhouette quality levels and the corresponding notations

Quality	Segmentation approach used
Q.1	Approach 1: BS method using Otsu’s threshold
Q.2	Approach 2: Normalised BS method plus dilation & erosion
Q.3	Approach 3: BS method with small threshold
Q.4	Approach 4: Frame differentiation plus dilation & erosion
Q.5	Approach 5: GMM & EM method
Q.6	Approach 6: LMedS method

method. We fuse gallery data with a set of weights computed by least square fitting. Fused gallery data and probe data are transformed into fused gallery features and probe features in a lower dimension space. Finally, we measure the similarities between fused gallery features and probe features.

### 3.1 Segmentation algorithms

In order to create gait silhouettes of different qualities, we combine different background subtraction (BS) methods with de-noising methods to create four different silhouette segmentation approaches. We also employ the Gaussian Mixture Model and Expectation Maximization(GMM and EM) method [SPL<sup>+</sup>05], and Least Median of Squares (LMedS) method [WTNH03], as additional segmentation approaches. The segmented silhouettes obtained by each of these approaches is used to generate binary images (and GEIs) at a specific quality. The quality levels and the corresponding segmentation approaches used



Figure 2: Sample gait silhouette binary images and their corresponding GEIs (right -most column) computed at six different qualities for the same subject.

are listed in Table 1. The segmentation approaches are explained in the following paragraphs.

Approach 1: A pixel is marked as foreground if  $|I_t - B_t| > threshold$ , where  $I_t$  refers to an image with both foreground and background objects and  $B_t$  contains only background objects. The threshold is set using Otsu's method[SS04].

Approach 2: The background image is normalized to eliminate the negative effects of noise, i.e.  $|I_t - avgB_t| > threshold$ , where  $avgB_t = B_t / \sum p_{i,j}$ ,  $p_{i,j}$  refers to the value of pixel  $i, j$  in  $B_t$ . The threshold is set using Otsu's method. As the obtained foreground may comprise several disconnected regions, dilation and erosion operations are performed to generate the final foreground.

Approach 3: A small threshold is used in order to introduce a distinct contrast in the segmented silhouettes and to include more background objects in the foreground; namely  $|I_t - B_t| > threshold/3$ . The threshold is set using Otsu's method.

Approach 4: Frame differentiation is used to mark the moving foreground pixels,  $I_t - I_{t-1} > threshold$ , where the threshold is set using Otsu's method. In addition, dilation and erosion operations are used in order to connect the disconnected regions comprising the foreground.

Approach 5: The GMM and EM method, as introduced in the baseline algorithm of Sarkar et al. [SPL<sup>+</sup>05].

Approach 6: The LMedS method, as is introduced in [WTNH03], and provided in the CASIA B dataset.

Six silhouette qualities are generated for each gait sequence, including the original silhouette provided by CASIA B dataset denoted as Q.6, as summarized in Table 1. This different qualities allow representing differences between the quality of the probe and gallery data. For each quality, we compute the corresponding GEIs following the work of Han et al. [HB06]. Fig. 2 shows visual samples of silhouettes and corresponding GEIs computed for the different qualities.

### 3.2 Feature extraction

Different subspace learning methods may be used to project the data onto a feature subspace. In this work, to avoid singularity problems in computation, we employ PCA before implementing LDA, or LPP. The generated transformation matrix  $W_{trans}$  is then  $W_{trans} = W_s^T W_{PCA}^T$ , where  $s \in \{LDA, LPP\}$ .

### 3.3 Similarity computation

The feature sets after projection are:

$$\begin{aligned} \{\hat{\mathbf{g}}\} : \hat{g}_i &= W_{trans} G_i \\ \{\hat{\mathbf{p}}\} : \hat{p}_j &= W_{trans} P_j \end{aligned} \quad (3)$$

where  $i = \{1, 2, \dots, n_1\}$ ,  $j = \{1, 2, \dots, n_2\}$ , and  $n_1, n_2$  are the total number of GEIs in gallery and probe data sets, respectively. The centroid of class  $l$  in  $\{\hat{\mathbf{g}}\}$  is calculated as  $\mathbf{mg}_l = \frac{1}{n_l} \sum_{\hat{\mathbf{g}} \in \hat{\mathbf{g}}_l} \hat{g}_i$ , where  $\hat{\mathbf{g}}_l$  is the set of gallery feature vectors in class  $l$ . The centroid of class  $l$  in  $\{\hat{\mathbf{p}}\}$  is calculated in the same way and is denoted as  $mp_l$ . The classifier is then defined as:

$$D(\mathbf{mg}_l, \mathbf{mp}_i) = \|\mathbf{mp}_i - \mathbf{mg}_l\|, i = 1, 2, \dots, c. \quad (4)$$

If  $D(\mathbf{mg}_l, \mathbf{mp}_i) = \min_{i=1}^c D(\mathbf{mg}_l, \mathbf{mp}_i)$ , the probe feature vector is assigned to the right class label.

## 4 Experimental evaluation

In order to evaluate the framework, we use the gait sequences of CASIA B dataset to generate the GEIs at different qualities. CASIA B dataset comprises video sequences for 124 individuals. The frame size is  $320 \times 240$ , and the frame rate is 25 fps. As this work aims at studying the effect of gait silhouette quality on recognition, other factors that may influence the recognition performance are excluded. Therefore, only normal gait sequences are chosen from CASIA B, without the factors of carrying bags, different clothes, different view angles, etc.

### 4.1 Evaluation without subspace learning

We first evaluate the performance with no dimensionality reduction or subspace learning method. The matching rates in percentage are tabulated in Table 2. Two observations can be drawn from the table:

Table 2: Matching rates between gallery data (G) and probe data (P) without dimensionality reduction and subspace learning. (%)

G \ P	Q.1	Q.2	Q.3	Q.4	Q.5	Q.6
Q.1	<b>85</b>	12	7	10	80	70
Q.2	12	<b>67</b>	17	8	10	35
Q.3	17	15	<b>78</b>	5	17	8
Q.4	15	8	5	<b>38</b>	18	15
Q.5	83	12	7	13	<b>83</b>	63
Q.6	58	25	5	10	43	<b>97</b>

1. The entries in the main diagonal represent the matching results between gallery and probe data of the same quality. These values are usually the highest values in each row which shows that when both gallery and probe data have the same quality, even if the quality is poor, the best matching results are attained. In our experiment, it is shown that silhouettes of quality Q.1 and Q.5 have smaller distances (i.e., the matching rate between them is relatively high), which means the quality is similar, even though the segmentation algorithms are totally different.
2. The entries outside the main diagonal show that the discrepancy in gait silhouette quality between gallery and probe data indeed decreases the recognition accuracy. In some cases, the matching rate between data segmented using the same algorithm can also be very low, which indicates that the segmentation algorithm may be inappropriate for the video source (see for example Q.4 matched with Q.4).

## 4.2 Evaluation with subspace learning

Tables 3 and 4 tabulate the average matching rates in percentage, after using PCA+LDA and PCA+LPP respectively. It is important to recall that in practical cases, the quality of the gallery and probe data may differ. It is then important for recognition algorithms to maintain a high accuracy even in this situation. Therefore, we measure the similarity between each individual in the probe data set against all individuals in the gallery data set for all qualities except for the quality of the probe data. This scenario corresponds to the empty entries in Table 3 and Table 4. These two tables show that by using dimensionality reduction plus a subspace learning method, matching rates can be considerably improved. Note that LPP can effectively deal with poor quality matching, i.e., qualities different from Q.6, while LDA appears to perform better than LPP with high quality matching, i.e., Q.6 data.

Table 3: Matching rates between gallery data (G) and probe data (P) using PCA+LDA (%)

G \ P	Q.1	Q.2	Q.3	Q.4	Q.5	Q.6
Q.1		73.3	63.3	28.3	93.3	95
Q.2	85		86.7	31.7	75	90
Q.3	71.7	73.3		28.3	70	83.3
Q.4	61.7	51.7	71.7		58.3	70
Q.5	95	73.3	63.3	23.3		96.7
Q.6	85	70	66.7	28.3	85	

### 4.3 Evaluation with fusion strategy

We propose to further improve the performance of subspace learning methods by fusing gallery data before matching with probe data features. Specially, we fuse the multi GEI representation of the gallery subjects, which consists of various quality levels, into one GEI representation.

Before feature space transformation, we compute a set of weights to be used in fusion strategy. In this experiment, for each probe data of a specific quality, there are gallery data of 5 different qualities available for fusion. For example, for probe data of quality Q.1, we fuse gallery data of all qualities except quality Q.1. The set of weights for fusion are computed using least squares fitting. Specifically, we aim to find the combination of weights for the gallery data centroids, i.e., gallery data of different qualities, that best match the centroid of the probe data of a specific quality. Let us denote the set of GEI vectors in the gallery set as  $\mathbf{G} = \{\mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_n\}$ , for  $n$  different qualities. Let us also denote the probe GEI vector as  $\mathbf{p}$ . The set of weights  $\mathbf{w} = \{w_1, w_2, \dots, w_n\}$  for the gallery data of  $n$  different qualities is then computed as follows:

$$\mathbf{w} = \arg \min_{\mathbf{w}} \|\mathbf{G} * \mathbf{w}^T - \mathbf{p}\|. \quad (5)$$

These weights are then used to fuse gallery data into  $\mathbf{G}_f$  :

$$\mathbf{G}_f = \sum_i^n \mathbf{g}_i * w_i, i \in \{1, 2, \dots, n\}. \quad (6)$$

After fusion, the gallery data is projected into the feature space, where similarities are measured between probe and the fused gallery features. The results of this experiment are shown in Table 5. In this table, for comparison, we average the column rates of Table 3 and 4, and denote them as 'LDA' and 'LPP' respectively.

Table 4: Matching rates between gallery data (G) and probe data (P) using PCA+LPP (%)

G \ P	Q.1	Q.2	Q.3	Q.4	Q.5	Q.6
Q.1		71.7	75	28.3	96.7	93.3
Q.2	83.3		85	35	83.3	93.3
Q.3	78.3	73.3		26.7	75	83.3
Q.4	66.7	53.3	63.3		66.7	65
Q.5	96.7	68.3	68.3	26.7		93.3
Q.6	88.3	65	66.7	20	91.7	

Table 5: Average matching rate of six different quality gallery data (G) and probe data (P) using PCA+LDA (LDA) and PCA+LPP (LPP), and the matching rate of fused gallery data using PCA+LDA with feature fusion (LDAF) and PCA+LPP with feature fusion (LPPF)(%)

G \ P	Q.1	Q.2	Q.3	Q.4	Q.5	Q.6	Average
LDA	81	72	74	32	79	88	<b>68.2</b>
LDAF	75	78.3	93.3	35	93.3	95	<b>78.3</b>
LPP	84	72	75	31	85	90	<b>69.4</b>
LPPF	78.3	75	90	31.7	95	95	<b>77.5</b>

#### 4.4 Discussions

In cases when the data quality in the gallery set is different from that of the data in the probe set, the performance of recognition algorithms may be poor, making it hard to chose a dependable classifier. The fusion strategy proposed in this work finds the combination of gallery data that has a minimum distance to the probe data. This is done by finding a set of weights using least square fitting, which is efficient and parameter-free. In our experiment design, we assume that the silhouette quality of probe data does not match any of the quality in gallery data, which would be more frequently occurred for practical cases. As show in Table 5, this strategy can considerably improve recognition performance under such circumstances. It is important to mention that by introducing least square fitting, the multi-quality gallery data is fused to best fit the probe data, which is similar to the case where gallery and probe data are equally segmented, i.e. the diagonal data in Table 2.

## 5 Conclusion

This paper presented a framework for GEL-based gait recognition for cases when a discrepancy in quality between gallery and probe data exists. The motivation is to tackle the problem where gallery and probe data are segmented using different algorithms. To this end, we generate GELs with different qualities in order to represent segmentation inaccuracies commonly encountered when dealing with low quality data. To perform recognition,

we study the use of subspace learning methods after dimensionality reduction by PCA. Simulation experiments on the CASIA B dataset using LDA and LPP indicate that gait recognition is indeed affected if the quality of the probe data set differs from that of the gallery data set. Results also suggest that important improvements in matching rate may be attained when subspace learning methods are used, since the feature subspace finds the best projection to match probe with gallery features of the same quality level. The paper also presented a fusion strategy that fuses gallery data of different qualities before feature space transformation. Experiments showed that this fusion strategy, which employs a number of weights, can further improve matching rates.

## References

- [BHK97] P.N. Belhumeur, J.P. Hespanha, and D. Kriegman. Eigenfaces vs. Fisherfaces: recognition using class specific linear projection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 19(7):711–720, Jul 1997.
- [BN03] Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural computation*, 15(6):1373–1396, 2003.
- [HB06] J. Han and B. Bhanu. Individual recognition using gait energy image. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 28(2):316–322, 2006.
- [HN03] Xiaofei He and Partha Niyogi. Locality preserving projections. *Advances in Neural Information Processing Systems(NIPS)*, 16:234–241, 2003.
- [HYH<sup>+</sup>05] Xiaofei He, Shuicheng Yan, Yuxiao Hu, P. Niyogi, and Hong-Jiang Zhang. Face recognition using Laplacianfaces. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 27(3):328–340, March 2005.
- [LS05] Z. Liu and S. Sarkar. Effect of silhouette quality on hard problems in gait recognition. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 35(2):170–183, April 2005.
- [MMN<sup>+</sup>15] Yasushi Makihara, Darko Matovski, MARK S Nixon, John N Carter, and Yasushi Yagi. *Gait Recognition: Databases, Representations, and Applications*. Wiley, 2015.
- [SPL<sup>+</sup>05] S. Sarkar, P.J. Phillips, Z. Liu, I.R. Vega, P. Grother, and K.W. Bowyer. The humanID gait challenge problem: data sets, performance, and analysis. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 27(2):162–177, Feb 2005.
- [SS04] Mehmet Sezgin and BuÁlent Sankur. Survey over image thresholding techniques and quantitative performance evaluation. *Journal of Electronic Imaging*, 13(1):146–168, 2004.
- [SSC14] S.H. Shaikh, K. Saeed, and N. Chaki. Gait recognition using partial silhouette-based approach. pages 101–106, Feb 2014.
- [WC11] A.R. Webb and K.D. Copsey. *Statistical Pattern Recognition*. Wiley, Chichester, third edition, 2011.
- [WTNH03] Liang Wang, Tieniu Tan, Huazhong Ning, and Weiming Hu. Silhouette analysis-based gait recognition for human identification. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 25(12):1505–1518, Dec 2003.

- [WZW<sup>+</sup>12] Chen Wang, Junping Zhang, Liang Wang, Jian Pu, and Xiaoru Yuan. Human Identification Using Temporal Information Preserving Gait Template. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 34(11):2164–2176, Nov 2012.
- [ZPCF10] Junping Zhang, Jian Pu, Changyou Chen, and R. Fleischer. Low-Resolution Gait Recognition. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 40(4):986–996, Aug 2010.

# Segmentation-level Fusion for Iris Recognition

Peter Wild<sup>1,3</sup>, Heinz Hofbauer<sup>2</sup>, James Ferryman<sup>1</sup> and Andreas Uhl<sup>2</sup>

<sup>1</sup> School of Systems Engineering, University of Reading, Reading RG6 6AY, UK.

<sup>2</sup> Dept. of Computer Sciences, University of Salzburg, 5020 Salzburg, Austria.

<sup>3</sup> AIT Austrian Institute of Technology GmbH, 2444 Seibersdorf, Austria.

peter.wild@ait.ac.at, {hhofbaue, uhl}@cosy.sbg.ac.at, j.m.ferryman@reading.ac.uk

**Abstract:** This paper investigates the potential of fusion at normalisation/segmentation level prior to feature extraction. While there are several biometric fusion methods at data/feature level, score level and rank/decision level combining raw biometric signals, scores, or ranks/decisions, this type of fusion is still in its infancy. However, the increasing demand to allow for more relaxed and less invasive recording conditions, especially for on-the-move iris recognition, suggests to further investigate fusion at this very low level. This paper focuses on the approach of multi-segmentation fusion for iris biometric systems investigating the benefit of combining the segmentation result of multiple normalisation algorithms, using four methods from two different public iris toolkits (USIT, OSIRIS) on the public CASIA and IITD iris datasets. Evaluations based on recognition accuracy and ground truth segmentation data indicate high sensitivity with regards to the type of errors made by segmentation algorithms.

## 1 Introduction

Iris recognition challenges for on-the-move and less constrained acquisitions, like the Noisy Iris Challenge Evaluation (NICE) [PA12], and Multiple Biometrics Grand Challenge (MBGC), illustrated the importance of robust iris segmentation in latest-generation iris biometric systems. Iris verification rates as low as 44.6% [RJS<sup>+</sup>12] are reported for unconstrained applications, and image quality has been shown to play a critical role in the segmentation and normalisation process [AFB13]. Normalisation seems to be at the heart of the problem, but combination past feature-extraction (see fusion scenarios in [RNJ06]) is easier and segmentation fusion lacks standardisation. ISO/IEC TR 24722:2007 does not foresee multinormalisation, nor does ISO/IEC 19794-6:2011 define a segmentation-only exchange format: there is a *cropped and masked* data format for normalised textures following IREX K7, but without direct access to segmentation results/parameters. While segmentation algorithms themselves might combine different approaches, iris segmentation fusion as proposed in [UW13] is widely ignored as a means to achieve more robust and accurate segmentation. As a common alternative, multi-algorithm fusion is suggested as a scenario [RNJ06] operating on the same input images. However, the expected increase in accuracy is usually not justifying the cost (in terms of additional processing power). Strong correlation of algorithms combined at system levels due to similar/same normalisation steps, and the dominance of local Gabor-based features (following Daugman's rubbersheet normalisation and original feature extraction [Dau04]) are likely to be reasons for observed little impact on accuracy (compared to combining, e.g. image-based

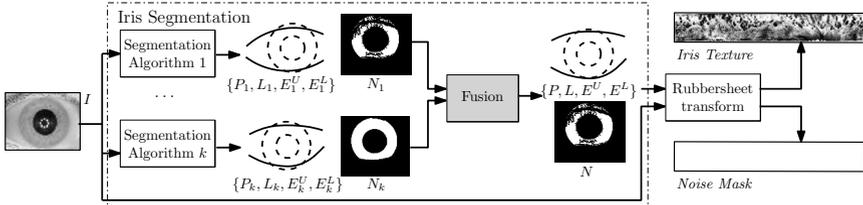


Figure 1: Iris Segmentation Fusion Framework.

and binary features [HRUW12]). Fusion at image data level, such as in [LVGV<sup>+</sup>15] following [SGCL14] reveals promising results, but requires the multiple execution of the iris unwrapping and normalisation process (for each obtained segmentation). Furthermore, given multiple normalised textures after segmentation and unwrapping, it is difficult to determine faulty or highly inaccurate segmentation versions.

The novelty of this work is a thorough analysis of how segmentation-based fusion in iris recognition can help in achieving higher accuracy considering the entire iris processing chain involving feature extraction, which may itself be tolerant to deformation to a certain extent. The latter observation raises the question on evaluation of fusion schemes at this stage, as ground-truth conformance is just one of several impacting factors. Especially the impact of outliers is highlighted in this paper. For this task, (1) a framework of combining segmentation results following Daugman’s rubbersheet model is presented (see Fig. 1); (2) a set of reference fusion methods combining segmentation curves, models, and masks is implemented, and; (3) pairwise combination improvement is analysed on public datasets with regards to both, ground-truth and recognition-accuracy. The following questions are addressed in this paper: (1) Does the combination of automated iris segmentation results yield more accurate result than each of the employed original segmentation algorithms? (2) How does the choice of database and segmentation algorithms impact on iris segmentation fusion? (3) How do outliers impact on overall recognition accuracy and how do ground-truth-based vs. recognition-based evaluations relate to each other?

As an introduction to the topic of multi-segmentation fusion Section 2 reviews related work on iris normalisation, fusion approaches, and segmentation data interoperability. Section 3 presents the proposed framework of segmentation fusion and discusses implementations. An experimental evaluation of proposed techniques is given in Section 4, analysing results with regards to questions outlined in this introduction. Finally, Section 5 concludes this work on segmentation-based fusion for iris biometric systems.

## 2 Related Work

Modern iris recognition algorithms operate on normalised representations of the iris texture obtained by mapping the area between inner and outer iris boundaries  $P, L : [0, 2\pi) \rightarrow [0, m] \times [0, n]$  to “Faberge” or “Rubbersheet” coordinates” (using angle  $\theta$  and pupil-to-limbic radial distance  $r$ ) [Dau04], independent of pupillary dilation:  $R(\theta, r) := (1 - r) \cdot P(\theta) + r \cdot L(\theta)$ . Normalised texture and noise masks  $T, M : [0, 2\pi) \times [0, 1] \rightarrow C$  are obtained ( $C$  is the target color space,  $M = N \circ R, T = I \circ R$  for the original  $n \times m$  image  $I$  and noise mask  $N$ ). The latter usually considers reflections and upper and lower

eyelid curves masking out occlusions, such that  $N(x, y) \neq 0$  if and only if pixel  $(x, y)$  refers to an in-iris location. While normalisation is standardised, there are several iris segmentation approaches for obtaining  $P, L$  and  $N$ . Original approaches employed circular boundary-based segmentation, such as Daugman’s integrodifferential operator [Dau04] and Wildes’ circular Hough Transform (HT) [Wil97]. Today’s advanced iris segmentation techniques are often multi-stage approaches combining various techniques: Active shape models [AS06], clustering-based iris localization [THS10] (e.g. locating the sclera for NICE.I data), AdaBoost-cascade and Pulling-and pushing models [HTSQ09], agent-based methods [LPS09], the Viterbi algorithm at different resolutions [SGSD12], or iterative multi-scale approaches and ellipsopolar transform for elliptical iris models [UW12b]. With the recent focus on visible-range (VR) iris segmentation compared to traditional near-infrared (NIR) segmentation techniques, the robust combination of independent segmentation approaches becomes an interesting aspect. Recently, [HAFW<sup>+</sup>14] compared multiple segmentation algorithms on different VR and NIR datasets based on ground truth information, illustrating the dependence of algorithms on database-specific assumptions and underlining the need for more robust segmentation.

There are not many proposed fusion techniques operating before feature extraction, most of them focusing on data-level fusion: Huang et al. [HMTW03] present a Markov network learning-based fusion method to enhance the resolution of iris images. Hollingsworth et al. [HPBF09] combine high-resolution images from multiple frames to create a single combined representation. Jillela and Ross [JRF11] proposed image-level fusion with Principal Components Transform. Recently, Llano et al. [LVGV<sup>+</sup>15] investigate the positive segmentation impact of PCA-based fusion vs. Laplacian Pyramid and Exponential Mean at image-level, i.e. multiple normalised iris textures are fused retrieved by following different segmentation algorithms. A first fusion approach of segmentation information (i.e. prior to normalisation) with the benefit of single normalisation and potentially simpler treatment and classification of errors than post-normalisation fusion is proposed in [UW13].

This work builds upon the framework of fusion for multiple iris segmentations introduced by Uhl and Wild [UW13], who combined evidence from human (manual) ground truth segmentation as a proof of concept work, but without any tests on automated iris segmentation algorithms and on a single dataset only. Two fusion methods were tested, both achieved higher recognition accuracy independent of the employed feature extraction algorithm (testing 3 approaches). Yet, the type of fusion technique (model-wise or data-wise) did not have a huge impact on accuracy and manual segmentation was reported to be fairly stable with 97.46% vs. 97.64% genuine acceptance rate (GAR), at 0.01% false acceptance rate (FAR), without any severe segmentation outliers [UW13].

The performance on automated segmentation algorithms raises further questions, especially questions related to stability if algorithms fail: Accurate results of a cohort of segmenters might be severely affected by a single segmentation error. Further, evaluations will be extended to ground-truth segmentation information as suggested by the Noisy Iris Challenge Evaluation - Part I (NICE.I), and the F-measure used in [HAFW<sup>+</sup>14]: Errors are estimated from the segmentation result (noise mask)  $N_i$  (or, more specifically, to estimate boundary detection performance an artificial noise mask is constructed rendering the iris using boundary curves  $P_i, L_i$ ) for each image  $I_i$  and compared using a ground truth

mask  $G_i$ . Let  $tp_i, fp_i, tn_i, fn_i$  refer to true / false respectively positive / negative pixel in-iris classifications for image index  $i$  (with dimension  $m \times n$ ), then:

$$E_1 := \frac{1}{k} \sum_{i=1}^k \frac{fp_i + fn_i}{mn}; \quad E_2 := \frac{1}{2} \left( \frac{1}{k} \sum_{i=1}^k \frac{fp_i}{fp_i + tn_i} \right) + \frac{1}{2} \left( \frac{1}{k} \sum_{i=1}^k \frac{fn_i}{fn_i + tp_i} \right) \quad (1)$$

$$\text{F-measure} = F_1 := \frac{1}{k} \sum_{i=1}^k \frac{tp_i}{tp_i + \frac{1}{2}(fn_i + fp_i)} \quad (2)$$

Error rate  $E_1$  refers of the rate of pixel disagreement between ground truth and segmentation noise masks,  $E_2$  accounts for the disproportion between a priori probabilities,  $F_1$  gives a measure of correctly to incorrectly proportions. Augmenting [UW13], this paper evaluates ground truth accuracy (using public IRISSEG-EP [HAFW<sup>+</sup>14]) and recognition impact, including an exhaustive significance analysis, to gain a deeper understanding of reasons for improvement. The McNemar test [McN47] is used for statistical significance analysis. In contrast to [LVGV<sup>+</sup>15] this work does not assume access to multiple source images and unlike [SGCL14] does not rely on multiple normalisations. However, the same open segmentation and recognition algorithms (USIT) are employed for reproducibility.

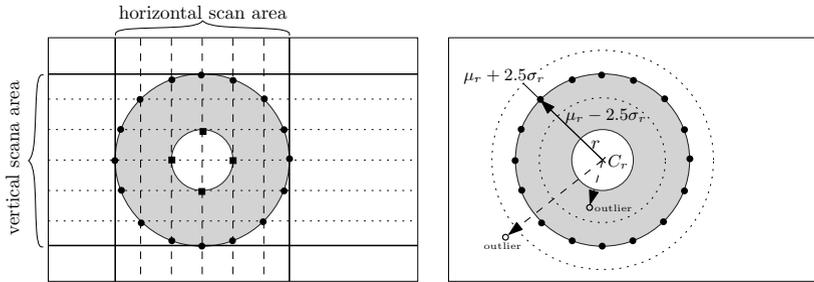
### 3 Multi-Segmentation Fusion Methods

Modern iris recognition algorithms pushed by challenge measures (NICE.I  $E_1, E_2$  as introduced in Sect. 2) focus on the problem of boundary refinement, taking occlusions and reflections into account [AS06, THS10, SGSD12]). For the Faberge mapping however, a robust segmentation of true (potentially occluded) boundaries  $P, L$  is critical, neglecting the presence of noise artifacts. This is to avoid non-linear distortions [UW12b]. While such distortions could possibly be targeted by more sophisticated matching techniques (e.g. by using Levenshtein distance), in identification mode it is more time-efficient to employ fast matching and study more advanced normalisation techniques, or combinations thereof and subject to this paper. In case direct parameterisations of the algorithm are available (e.g. center and radius for circular models, elliptical models, splines, or polygonal boundary representations), the following techniques have been proposed in [UW13]:

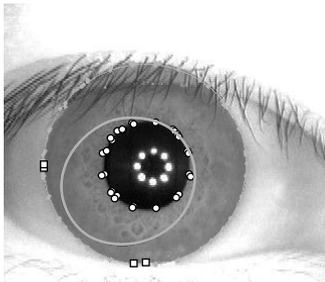
$$\text{Sum Rule: } B(\theta) := \frac{1}{k} \sum_{i=1}^k B_i(\theta); \quad \text{Aug Rule: } B(\theta) := \text{ModelFit} \left( \bigcup_{i=1}^k B_i \right) (\theta) \quad (3)$$

- **Sum-Rule Interpolation:** This fusion rule combines boundary points  $B_i(\theta)$  of curves  $B_1, B_2, \dots, B_k : [0, 2\pi) \rightarrow [0, m] \times [0, n]$  into a single boundary  $B$ , for pupillary and limbic boundaries, in analogy to the sum rule.
- **Augmented-Model Interpolation:** This model combines boundaries  $B_1, \dots, B_k$  within a jointly applied parametrisation model *ModelFit* minimizing the model-error (e.g., Fitzgibbon's ellipse- [FPF99], or least-squares circular fitting), executed separately for inner and outer iris boundaries. Models are combined, not only points.

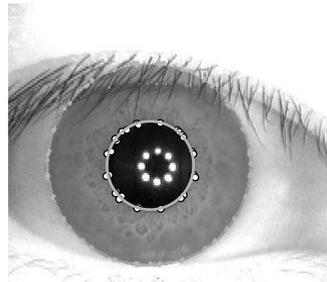
Segmentation masks  $N$  are common intermediate results, as normalisation is usually integrated rather than providing parameterisations of boundaries. A natural approach therefore is the extraction of parameterisations using noise masks, e.g. by employing an elliptical fitting. The following section illustrates the mask scanning process proposed in this work.



(a) Overview over the iris scanning and pruning process.



(b) With outliers



(c) With outliers pruned

Figure 2: Overview over the iris scanning and pruning process with examples.

### 3.1 Scanning Iris Masks

The mask fusion is an augmented-model interpolation based on a scan of the provided iris masks. This fusion method is based on the assumption that the mask is available but the original fitted model for the iris, pupil and eyelid boundaries are not, as would be the case for the IREX K7 specification. It follows the basic outline of the augmented model interpolation but skips the eyelid polygon fits. In a scan it is not necessarily possible to differentiate between iris and eyelid based purely on the mask. The model used for the augmentation is an ellipse fitting based on a scan of the iris mask.

First, the iris boundaries for each axis are determined. Then  $N$  equidistant scan lines are used to generate points along the iris and pupil boundaries. The boundary points of the provided masks are combined and pruned for outliers. Outliers typically happen when the outer mask of an iris is not convex, leading to wrongfully detected pupil boundary points along the iris boundary. The outlier detection is done by using the center of gravity  $C_r$  of all the detected points for a given boundary. The radius for each point from  $C_r$  is calculated and all points are pruned for which the radius has a z-score of greater than 2.5. This is illustrated in Fig. 2 along with the difference of pruned and unpruned iris detection. In order to get a stable outlier detection and correct boundaries a high number of scan lines is desirable, for our experiments  $N = 100$  was used. Furthermore, to properly associate mask transitions with iris or pupil boundaries there should be no extra transitions. Such transitions can be generated by noise exclusions in the mask. If the mask contains holes of this kind, they should be closed by an dilate+erode morphological operation.

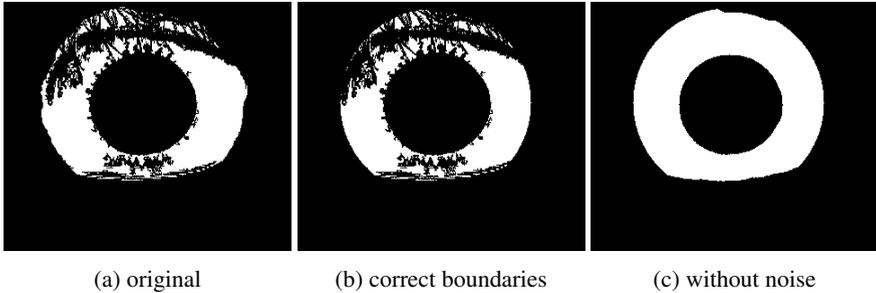


Figure 3: Original OSIRIS mask, corrected version for rubbersheet mapping, and corrected without noise masking.

The implementation of the tested OSIRIS algorithm produces masks which extend over the actual boundaries used for unrolling the iris image, see Fig. 3a, which would produce incorrect masks during the scanning steps. For the experiments we modified OSIRIS to restricted the produced mask to the detected boundaries, see Fig. 3b. In addition, we introduced an option to skip the noise mask, resulting in masks as shown in Fig. 3c. This speeds up the fusion of the masks by allowing to skip the dilate+erode morphological operations. The noise mask is only skipped for mask fusion not for calculation of the OSIRIS scores.

The mask fusion algorithm produces two points clouds, pertaining to the iris and pupil boundary. The actual mask is generated by fitting an ellipse to the point clouds by a least-squares method [FPF99]. The segmentation tool from [HAFW<sup>+</sup>14] is used for unrolling the iris image. It should also be noted that the mask level fusion only generates a mask which is used for unrolling the iris. No noise or occlusion mask is generated and consequently all tests performed on the fusion are performed purely on the unrolled iris image without masking.

## 4 Experimental Study

Addressing the question of multisegmentation fusion performance, we assessed pairwise combinations of the following segmentation algorithms: CAHT [RUW12], a traditional sequential (limbic-after-pupillary) method based on circular HT and contrast-enhancement; WAHET [UW12b], a two-stage adaptive multi-scale HT segmentation technique using elliptical models; OSIRIS [PM07], a circular HT-based method with boundary refinement; IFPP [UW12a] using iterative Fourier-series approximation and Pulling and Pushing methods. The motivation for selecting these algorithms were public availability as open source software for reproducibility, therefore also basing experiments on ground-truth segmentations released with [HAFW<sup>+</sup>14, DBS] and referring to the public CASIA-v4 and IITD iris databases. As feature extractors we used the wavelet transform-based algorithm by Ma *et al.* [MTWZ04] and the local Gabor-filter based algorithm by Masek [Mas]. The results in terms of equal error rate were obtained by using Hamming distance based verification, the tools are provided by and further documented in the USIT package [RUW12].

Table 1: Equal error rate for the segmentation fusion.

(a) Casia v4 Interval database					(b) IIT Delhi database				
Equal-error rate [%] of Masek					Equal-error rate [%] of Masek				
	CAHT	WAHET	OSIRIS	IFPP		CAHT	WAHET	OSIRIS	IFPP
CAHT	1.22	<b>0.92</b>	<b>1.03</b>	1.30	CAHT	1.85	3.60	1.65	<b>1.38</b>
WAHET		1.89	<b>1.02</b>	<b>1.41</b>	WAHET		6.82	3.90	<b>3.70</b>
OSIRIS			1.04	1.44	OSIRIS			1.40	1.94
IFPP				8.10	IFPP				3.87

Equal-error rate [%] of Ma					Equal-error rate [%] of Ma				
	CAHT	WAHET	OSIRIS	IFPP		CAHT	WAHET	OSIRIS	IFPP
CAHT	0.99	<b>0.64</b>	0.84	1.17	CAHT	1.72	4.06	<i>1.95</i>	<b>1.43</b>
WAHET		1.72	0.89	<b>1.22</b>	WAHET		7.43	4.86	<b>4.23</b>
OSIRIS			0.73	1.53	OSIRIS			1.21	2.40
IFPP				8.78	IFPP				4.36

#### 4.1 Impact on Recognition Accuracy

The main motivation for combining segmentation algorithms is to achieve a better overall system recognition accuracy. Whereas segmentation is an integral part of a biometric recognition system, the advantage of a system-based evaluation is that it takes into account that small segmentation errors do not necessarily implicate an impact on recognition accuracy, as the feature extraction (and comparison) algorithm itself tries to extract features invariant under slight transformations (e.g. small shifts, different illumination, etc.). Table 1a gives the results of the evaluation on the CASIA-IrisV4-Interval [DBC] database, and Table 1b gives the results on the IIT Delhi Iris Database [DBI]. The entries along the principal diagonal are the results of the original segmentation algorithms. Fusion results which are an improvement over both fused algorithms are shown in a bold font and fusion results where the fusion performs worse than both individual algorithms are shown in italics.

From Table 1 we can see that segmentation fusion increased performance in 10 out of 24 combination scenarios involving different algorithms and databases. While there is only one case, IFPP fused with WAHET, which consistently increases the performance, there are numerous cases where the fusion improves over both algorithms. In particular, there is only one case, OSIRIS fused with CAHT with feature extraction of Ma on the IITD database, where the combined performance is worse than both solitary performances. Given that all employed segmentation algorithms aim for gradient-based detection rather than employing completely different approaches and thus limiting the fusion potential as any independence assumption is likely violated, the fraction of cases with improvement is rather encouraging and deserves further attention.

In order to verify the statistical significance of results, we conducted McNemar tests [McN47] dedicated to matching pairs of subjects. The test uses the dichotomous trait

Table 2: Results of the McNemar test, reported as the  $X^2$  values. The row gives the single method compared to the fusion as indicated by row  $\times$  column.

(a) Casia v4 Interval database					(b) IIT Delhi database						
$X^2$ statistic for Masek					$X^2$ statistic for Masek						
single method					single method						
CAHT WAHET OSIRIS IFPP					CAHT WAHET OSIRIS IFPP						
fused with	CAHT		24742	8	246149	fused with	CAHT		49180	169	35918
	WAHET	2543		13	247450		WAHET	20317		42328	24
	OSIRIS	1158	22002		243734		OSIRIS	1746	27835		17116
	IFPP	928	8110	3729			IFPP	3193	38721	3655	
$X^2$ statistic for Ma					$X^2$ statistic for Ma						
single method					single method						
CAHT WAHET OSIRIS IFPP					CAHT WAHET OSIRIS IFPP						
fused with	CAHT		28739	135	273347	fused with	CAHT		21271	4614	61327
	WAHET	3993		1649	276351		WAHET	52945		78177	53
	OSIRIS	1620	15752		261445		OSIRIS	368	10149		26311
	IFPP	1438	7076	10532			IFPP	1145	21256	11669	

of correct classification (in relation to the known ground truth). We utilize the  $\chi^2$  approximation with the continuity correction proposed by Edwards [Edw48]. Table 2 reports obtained  $X^2$  values at the EER operating point. Note that a critical value  $X^{2*} \geq 6.64$  indicates a rejection of the null hypothesis — that there is no difference between the two methods — with at least 99% significance. The table gives the comparison of single method in the column, e.g. CAHT(column), with the fusion as indicated by column and row, e.g. CAHT(column) fused with WAHET(row).

## 4.2 Ground-truth Segmentation Accuracy

To understand how fusion influences the segmentation performance we compared the segmentation results to ground truth, which is available from two independent manual segmentations. Fig. 4 gives the F-measure segmentation error introduced in Eq. 2 for IFPP, WAHET and their fusion on the CASIA v4 interval database. The fusion exhibits a closer conformity to the ground truth than each individual segmentation algorithm. Using the outlier detection from [HAFW<sup>+</sup>14] we can further confirm the conformity to the ground truth for the fusion; IFPP had 95 outliers, WAHET had 32 and the fusion only 16.

We compared Sum Rule segmentation fusion performance on “good” versus “bad” segmentations using segmentation consistence between both algorithms as an indicative measure (we used distance of pupillary and limbic centers, as well as the absolute difference in radii using threshold  $\eta = 10$ ). Results indicated, that fusion performance on the “good” set improved accuracy, while averaging performance for the “bad” set with deviating information, rather than consistently eliminating over- and undersegmentation errors. Table 3

Table 3: Fusion for good vs. bad segmentation results.

Segmentation error [%]					Segmentation error [%]						
		$E_1$		$E_2$				$E_1$		$E_2$	
		Good	Bad	Good	Bad			Good	Bad	Good	Bad
CAHT		1.98	2.76	3.02	4.10	CAHT		2.61	5.00	3.48	8.33
WAHET (NIR)		2.30	6.05	3.54	8.90	WAHET (NIR)		2.77	15.31	3.73	20.76
Fusion (Sum Rule)		1.87	3.85	2.87	5.61	Fusion (Sum Rule)		2.40	9.95	3.23	13.84

(a) Casia v4 Interval database                      (b) IIT Delhi database

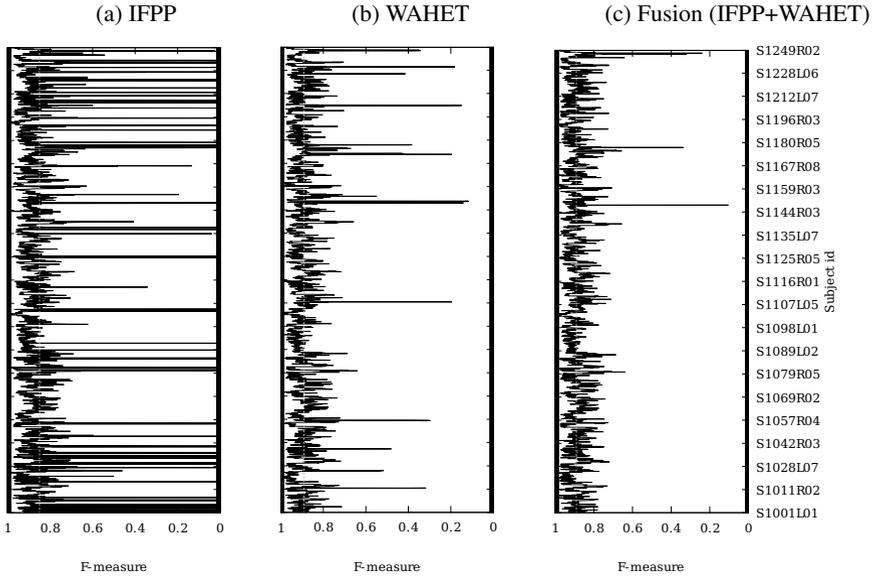
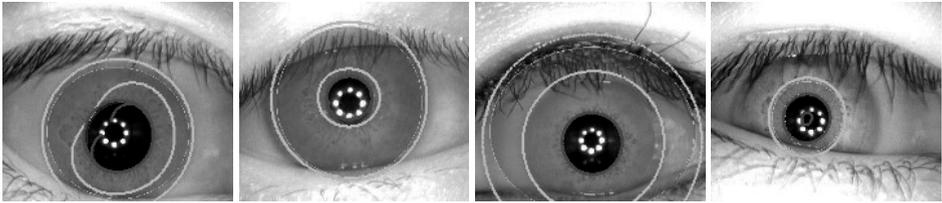


Figure 4: Segmentation comparison with ground truth on CASIA-v4-Interval.

illustrates this observation based on  $E_1$  and  $E_2$  error rates comparing segmentation results on both databases for the CAHT and WAHET combinations (for IITD we used dataset-optimised parameters to increase the set of segmentation-consistent images). Given that small segmentation errors are likely to be tolerated by the feature extraction algorithm, we identify the reduction in outliers as a strong factor in the overall improvement, which is unlikely to be reflected in ground-truth-based evaluations aiming to identify statistically significant improvements over the entire set. In the following some of the outliers will be discussed to make the fusion impact clearer.

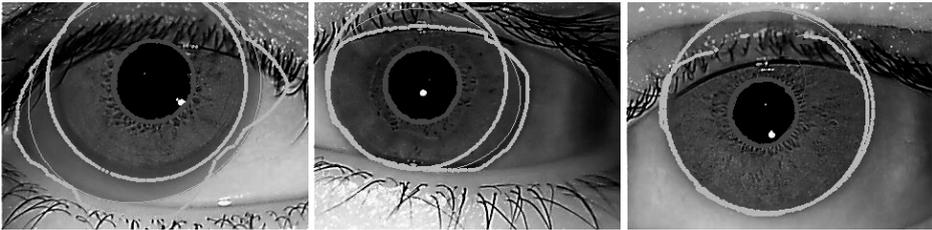
### 4.3 Analysis of Fusion Behaviour

For mask fusion, Fig. 5 shows samples from CASIA v4 interval database. Both the resulting segmentation as well as the point clouds for iris and pupil boundary are given. The



(a) Shape mismatch correction. (b) Boundary mismatch correction. (c) Sample discrepancy due to cut off iris. (d) Matching errors.

Figure 5: Possible effects of combining masks.



(a) Detection flaw. (b) Missed boundary. (c) Pruning failure.

Figure 6: Boundary overestimation and non-convex masks.

correction behaviour is due to the least-squares ellipse fitting valuing the outer boundaries higher. This leads to corrective behaviour when one of the masks has detected the wrong shape (5a) or the wrong boundary, in this case collarette instead of iris (5b). There are however limits to this, like in case the iris boundary being cut off, leading to a shape bias in the fitting process as seen in Fig. 5c. Further, if both original segmentations exhibit the same type of error the fusion can obviously not correct it, see Fig. 5d.

Fundamentally, the mask fusion values boundary points located farther from the center to a greater extent, e.g. Figs. 5a and 5b. As long as the boundary detection of the iris undershoots rather than overshoots the fusion is auto-corrective. A case where the boundary detection overshoots is the OSIRIS fusion with CAHT on the IITD database. The OSIRIS algorithm frequently overestimates the iris boundary. While this is often corrected by the mask provided by OSIRIS the resulting non-convex and miss shaped masks can lead to fusion problems. Examples comprise the cases of detection flaw and corresponding fusion error (Fig. 6a), missed boundary and an almost correct mask (Fig. 6b), and pruning errors due to a non-convex mask which is not sufficiently removed from correct points (Fig. 6c).

Essentially, as long as the boundary estimation is conservative, i.e. underestimates rather than overestimates, the auto-corrective properties of the mask fusion result in an increased performance. The same properties however will reduce the quality of the mask fusion when boundaries are frequently overestimated. Furthermore, non-convexity of the mask can lead to sample points which are attributed to the wrong boundary. These erroneous samples can be pruned to an extent, but non-convex masks always carry the possibility of a deformed pupillary boundary.

## 5 Conclusion and Future Work

This paper analysed multisegmentation fusion using pairwise combinations of CAHT, WAHET, IFPP and OSIRIS iris segmentation algorithms, revealing the autocorrective properties of augmented model fusion on masks in most of the tested cases (best result 0.64% EER for WAHET+CAHT versus 0.99% EER for CAHT only). Evaluations on ground-truth masks and recognition scores indicated, that ground-truth based evaluations are likely to miss corrective behaviour for outliers, which is critical for the overall task. Detailed error-specific analysis revealed case-specific corrective behaviour, which will be a good starting point for future case-specific fusion approaches. Benefits of multisegmentation in contrast to traditional multialgorithm fusion comprise better normalised source images available for feature-independent storage and the ability to focus on the time-consuming segmentation process, where parallelisation and advanced fusion might be most beneficial. Future work will focus on advanced, sequential approaches taking processing time into account.

## Acknowledgements

This work was supported by the EU FASTPASS project under grant agreement 312583 and the Austrian Science Fund, project no. P26630.

## References

- [AFB13] F. Alonso-Fernandez and J. Bigun. Quality factors affecting iris segmentation and matching. In *Proc. Int'l Conf. on Biometrics (ICB)*, 2013.
- [AS06] A. Abhyankar and S. Schuckers. Active shape models for effective iris segmentation. In *Proc. of SPIE*, 2006.
- [Dau04] J. Daugman. How iris recognition works. *IEEE Trans. on Circuits and Systems for Video Technology*, 14(1), 2004.
- [DBC] CASIA-IrisV4 Interval Database. <http://biometrics.idealtest.org/dbDetailForUser.do?id=4>.
- [DBI] IIT Delhi Iris Database. [http://www4.comp.polyu.edu.hk/~csajaykr/IITD/Database\\_Iris.htm](http://www4.comp.polyu.edu.hk/~csajaykr/IITD/Database_Iris.htm).
- [DBS] Iris Segmentation Ground Truth Database – Elliptical/Polynomial Boundaries (IRISSEG-EP). <http://www.wavelab.at/sources/Hofbauer14b>.
- [Edw48] A. Edwards. Note on the “correction for continuity” in testing the significance of the difference between correlated proportions. *Psychometrika*, 13(3):185–187, 1948.
- [FPF99] A. Fitzgibbon, M. Pilu, and Robert B. Fisher. Direct Least Square Fitting of Ellipses. *IEEE Trans. Pat. An. Ma. Int.*, 21(5), 1999.
- [HAFW<sup>+</sup>14] H. Hofbauer, F. Alonso-Fernandez, P. Wild, J. Bigun, and A. Uhl. A Ground Truth for Iris Segmentation. In *Proc. 22nd Int'l Conf. Pattern Rec. (ICPR)*, 2014.
- [HMTW03] J. Huang, L. Ma, T. Tan, and Y. Wang. Learning Based Resolution Enhancement of Iris Images. In *Proc. BMVC*, 2003.
- [HPBF09] K. Hollingsworth, T. Peters, K.W. Bowyer, and P.J. Flynn. Iris Recognition Using Signal-Level Fusion of Frames From Video. *IEEE Trans. Inf. For. Sec.*, 4(4), 2009.

- [HRUW12] H. Hofbauer, C. Rathgeb, A. Uhl, and P. Wild. Image Metric-based Biometric Comparators: A Supplement to Feature Vector-based Hamming Distance? In *Proc. Int'l Conf. Biom. Special Int. Group (BIOSIG)*, 2012.
- [HTSQ09] Z. He, T. Tan, Z. Sun, and X. Qiu. Toward Accurate and Fast Iris Segmentation for Iris Biometrics. *IEEE Trans. Pattern Anal. Mach. Intell.*, 31(9), 2009.
- [JRF11] R. Jillela, A. Ross, and P.J. Flynn. Information fusion in low-resolution iris videos using Principal Components Transform. In *IEEE WS Appl. Comp. Vis. (WACV)*, 2011.
- [LPS09] R. Labati, V. Piuri, and F. Scotti. Agent-based image iris segmentation and multiple-views boundary refining. In *Proc. Int'l Conf. Biom.: Th., Appl. Syst. (BTAS)*, 2009.
- [LVGV<sup>+</sup>15] E. Llano, J. Vargas, M. Garca-Vzquez, L. Fuentes, and A. Ramirez-Acosta. Cross-sensor iris verification applying robust fused segmentation algorithms. In *Proc. Int'l Conf. on Biometrics (ICB), 2015*, pages 1–6, 2015.
- [Mas] L. Masek. Recognition of Human Iris Patterns for Biometric Identification, MSC thesis, Univ. Western Australia, 2003.
- [McN47] Q. McNemar. Note on the sampling error of the difference between correlated proportions of percentages. *Psychometrika*, 12(2):153–157, 1947.
- [MTWZ04] L. Ma, T. Tan, Y. Wang, and D. Zhang. Efficient iris recognition by characterizing key local variations. *IEEE Trans. Image Proc.*, 13(6), 2004.
- [PA12] H. Proença and L. Alexandre. Toward Covert Iris Biometric Recognition: Experimental Results From the NICE Contests. *IEEE Trans. Inf. For. & Sec.*, 7(2), 2012.
- [PM07] D. Petrovska and A. Mayoue. Description and documentation of the BioSecure software library. Technical report, Project No IST-2002-507634 - BioSecure, 2007.
- [RJS<sup>+</sup>12] A. Ross, R. Jillela, J.M. Smereka, V.N. Boddeti, B.V.K.V. Kumar, R. Barnard, Xiaofei Hu, P. Pauca, and R. Plemmons. Matching highly non-ideal ocular images: An information fusion approach. In *Proc. Int'l Conf. on Biometrics (ICB), 2012*.
- [RNJ06] Arun A. Ross, Karthik Nandakumar, and Anil K. Jain. *Handbook of Multibiometrics*. Springer, 2006.
- [RUW12] C. Rathgeb, A. Uhl, and P. Wild. *Iris Recognition: From Segmentation to Template Security*, volume 59 of *Advances in Information Security*. Springer, 2012.
- [SGCL14] Y. Sanchez-Gonzalez, Y. Cabrera, and E. Llano. A Comparison of Fused Segmentation Algorithms for Iris Verification. In *Proc. Ib. Congr. Patt. Rec., (CIARP)*, 2014.
- [SGSD12] G. Sutra, S. Garcia-Salicetti, and B. Dorizzi. The Viterbi algorithm at different resolutions for enhanced iris segmentation. In *Proc. Int'l Conf. Biom. (ICB)*, 2012.
- [THS10] T. Tan, Z. He, and Z. Sun. Efficient and robust segmentation of noisy iris images for non-cooperative iris recognition. *Image and Vision Computing*, 28(2), 2010.
- [UW12a] A. Uhl and P. Wild. Multi-stage Visible Wavelength and Near Infrared Iris Segmentation Framework. In *Proc. Int'l Conf. Image An. Rec. (ICIAR)*, LNCS, 2012.
- [UW12b] A. Uhl and P. Wild. Weighted Adaptive Hough and Ellipsopolar Transforms for Real-time Iris Segmentation. In *Proc. Int'l Conf. on Biometrics (ICB)*, 2012.
- [UW13] A. Uhl and P. Wild. Fusion of Iris Segmentation Results. In *Proc. 18th Ib. Congr. on Pattern Recog., (CIARP)*, 2013.
- [Wil97] R. P. Wildes. Iris recognition: an emerging biometric technology. In *Proc. of the IEEE*, volume 85, 1997.

# Structured Forest Edge Detectors for Improved Eyelid and Iris Segmentation

Michael Happold

Independent Researcher  
217 Hunt Rd  
Fox Chapel, PA 15217, U.S.A.  
mhappold@startmail.com

**Abstract:** Edge detection has long figured into iris segmentation algorithms, often providing a first-pass estimate of the inner and outer iris boundaries. Standard edge detectors, however, generally produce too many extraneous edges inside and outside the iris to be used for simple ellipse fitting to robustly find the iris boundaries. Solutions to this problem have been either to have additional filtering to select relevant edges or to design specialized edge detectors that highlight iris boundary edges and suppress irrelevant edge types. This description holds ever so more for eyelid boundaries, which are often very subtle. An edge detector that will trigger on an eyelid boundary will also likely trigger on almost any slight intensity gradient in an image. We seek to solve this problem by learning specialized edge detectors for each type of relevant boundary in an iris image. Using a fast Structured Random Forest approach developed for learning generalized edge detectors, we train detectors for the iris/sclera, iris/pupil, and eyelid boundaries. The results show that learned edge detectors should become part of the standard toolbox for iris segmentation and eyelid boundary detection.

## 1 Introduction

Edge detection seems like the obvious first step when one initially approaches the problems of iris segmentation and eyelid boundary detection. The human visual system has little trouble picking out these boundaries in all but the most degraded images. What one discovers immediately is that tweaking a threshold to find the right balance between maximizing true edges and minimizing false edges rarely produces satisfactory results and surely does not generalize to novel instances. One could search for the threshold that produces the best results over a large sample of images, but this is hardly a solution as we show in Section 6.

Better approaches than the naïve approach would be either to design specialized edge detectors for each type of boundary or to apply an independent form of filtering to the outputs of the detectors. For example, in [Hu09] the authors develop a radial-suppression edge detector that retains annular edges while suppressing radial edges by employing a non-separable wavelet transform and radial non-maxima suppression. This technique,

however, is not applicable to detecting eyelid boundaries. The Canny edge detector appears as component in numerous techniques. For example in [Ma03], the Canny edge detector is used to find prospective boundaries followed by a circular Hough transform to extract the circles best corresponding to pupil and iris boundaries. Again, this method is inapplicable to eyelid boundaries due to the assumption of a canonical shape. Pre-filtering can also be used to improve the results of edge detection, such as applying a Contourlet transform to smooth the image before employing the Canny edge detector as [ZAS12].

Daugman's integro-differential operator [Da04] has been interpreted as a circular edge detector, but it is really a circular boundary detector because the entire set of edges making up the boundary of the iris is considered when computing the gradient. Its success highlights the importance of including context in a boundary detector. Daugman extended the method to eyelid detection by changing the parameterized shapes to be second order curves with mixed results.

Similar to the iris segmentation algorithms described above, edge-detection-based techniques for eyelid boundary discovery are focused on pre- and post-filtering edge images and specialized detectors. In [Ad08], the image is filtered using anisotropic diffusion before a Canny edge detector is run, with the output then being fitted with a parabolic curve model. Another filtering step is to remove eyelashes before edge detection or parabolic curve fitting. For example, authors in [Al11] use a wavelet transform and a neural network to extract eyelashes before detecting the eyelid.

Given the stark differences between pupil, iris and eyelid boundaries, it makes sense to search for specialized detectors and to employ pre- and post-filtering on the detector outputs. However, there is a better way to design the specialized detectors than trying to make an educated guess at what is needed. Recent research on learning edge detectors can be employed to train up individualized boundary detectors that run quickly and accurately, suppressing irrelevant image gradients. They achieve this spurious edge detection by exploiting the image context of a candidate edge pixel. We make use of Structured Random Forests [Ko11] adapted to produce edge-label patches as output [DZ13] to train these specialized edge detectors. Our detectors can more accurately find the boundaries of the structures of interest than standard techniques, especially in poor quality images, such as those with significant reflections or thick eyelashes.

The rest of the paper is organized as follows. Section 2 discusses the algorithmic flow. Section 3 describes the Structured Forest approach to edge detection, starting with an overview of Random Forests and moving to the modifications needed to adapt them to structural learning. Section 4 gives an overview of the features used. Section 5 describes the training methodology. Section 6 provides results on CASIA [Ca11] and ND-IRIS-0405 [Ph06] datasets for pupil, iris, and eyelid boundary detection, comparing them to adaptive threshold Canny detection. Section 7 gives a summary, our conclusions, and suggestions for future work.

## 2 Algorithmic Flow

The input to our algorithm is the raw grayscale image of the eye, without cropping for the iris, taken by standard infrared IR cameras. The Structured Forest approach operates on multiple feature channels; in the original work of [DZ13], these included color channels and pixel differences. We instead apply standard edge detectors and a combination of histogram equalization and anisotropic diffusion to augment the intensity image. We extract patches of these channels to include context which are then paired with corresponding label patches from hand labelings of the pupil, iris and eyelid boundaries. We then train on these pairs, producing individual classifiers for each boundary type as well as a combined classifier trained on all three boundary types. The classifiers output a label patch for a given image region representing the boundaries. Figure 1 illustrates the flow.

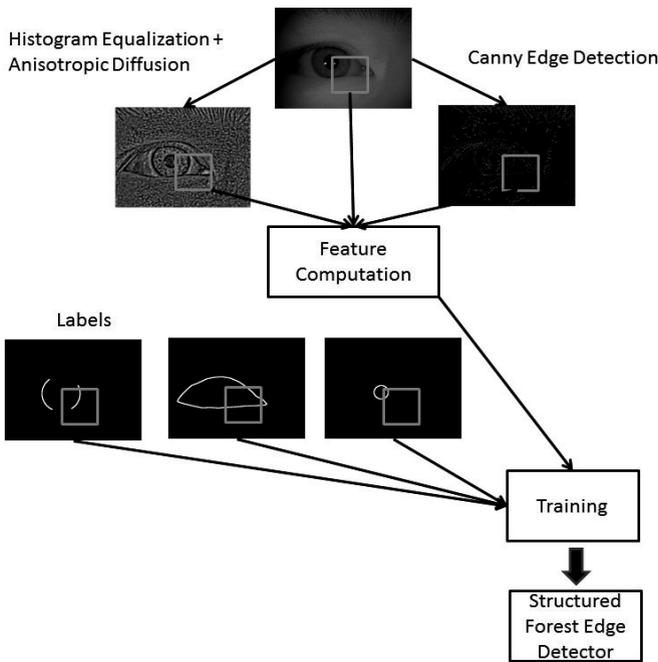


Figure 1: Algorithmic Flow

## 3 Structured Random Forests

The Random Forest framework forms the basis of our learning approach. A Random Forest is an ensemble of Decision Trees, each grown independently and in a manner that results in them being largely decorrelated. The output of the Forest for classification problems is a combination of the individual trees, often through simple majority voting, although other means are possible. Random Forests combine two basic ideas to increase

robustness: Bagging and random feature subset selection. Bagging [Br96], or Bootstrap aggregation, generates new training sets from a base training set by means of randomly sampling examples from the base uniformly and with replacement. By means of Bagging, each Decision Tree is exposed to a different bootstrap sample. Further robustness is introduced by randomly selecting a different subset of the feature space to grow each node of a given tree. Traditionally, the feature that maximally reduces a measure of impurity, whether Gini impurity or entropy, is chosen at each node, but randomness can be introduced at this stage as well, leading to the Extremely Random Tree. Random Forests are fast, well-suited to be adapted to structured learning tasks, and easily understood in their operations. The success of the Structured Random Forest approach of [DZ13] on the Berkeley Segmentation Dataset [Ma01] strongly suggests its applicability to iris and eyelid segmentation.

To adapt the Random Forest framework for structured learning, we need an appropriate splitting function for the tree nodes and a method for producing structured labels at the tree leaves and from the tree ensemble. Following the method of [DZ13], we treat only the output space as structured, the input space being just a bag of features. The insight of [DZ13] that the leaves of the forest's trees can store any type of label is key: we can simply store edge segmentation patches at the leaves that represent the edges predicted in the given image region. What this implies is that the method will only be able to predict edge masks that it has been exposed to, increasing the importance of a representative training set. This seeming limitation turns out to be less a drawback than a benefit for our domain, because the edges that we are seeking are highly constrained in their shape, relative position, and gradient orientation. For example, the boundary between the pupil and the iris is always curved in toward the pupil with a gradient that always points away from the pupil. Limiting the predicted edges to what the forest has been exposed to in the training set will work toward filtering spurious edges, particularly eyelashes and the interior structure of the iris.

Creating a splitting function for structured labels implies designing a measure of information gain, even if approximate, for a complex space. Rather than attempt to create a measure of similarity for this complex space, the Structured Random Forest approach is to map the output space  $Y$  to an intermediate space  $Z$  that is more suited to measuring the similarity between labels. The intermediate space  $Z$  should also have a straightforward mapping to the space of simple labels  $C$ , where  $C = \{1, \dots, k\}$ . Similar complex labels should at the end of the mapping have similar discrete label  $c \in C$ .

Our labels for edge detection in the space  $Y$  are patches of  $N \times N$  binary labels, with  $N$  in this case being 16. The transformation from  $Y$  to  $Z$  presented in [DZ13] such that a Euclidean distance metric can be used in  $Z$  is to create a long binary vector of segment membership comparisons for each pair of pixels (1 if they belong to the same segment, 0 if not). There are 32640 pixel pairs for a  $16 \times 16$  pixel patch, so creating the exact space  $Z$  for every member of  $Y$  would be too expensive for an edge detection algorithm. To reduce the dimensionality, the space  $Z$  is randomly sampled down to 256 dimensions, which gives an added degree of randomization, and is then further reduced to 2 dimensions via Principal Component Analysis (PCA).

The next step is to develop an information gain criterion for splitting nodes in the trees. The method of [DZ13] is used here, which uses standard information theory entropy to compute information gain. PCA discretization yields two classes, but this discretization occurs at each node during training rather than globally, and so the distributions found at each node determine different discretizations, further decorrelating the trees.

Because each leaf could have multiple label patches, we must have a method for combining them into a single prediction. Similarly, the ensemble of trees needs to combine possibly conflicting predictions from each tree. At each leaf, the medoid  $z_k$  is chosen, where  $z_k$  is defined as  $\min_k = \sum_j (z_{kj} - \bar{z}_j)^2$ . This same method is used to combine the outputs of the trees to produce a final label. Using the medoid is what limits the forests from generating novel results; what comes out is the already observed label that best fits the data. As we noted, this apparent limitation is in fact an important constraint in our domain.

## 4 Features

The work of [DZ13] focused on segmenting and detecting edges in color images; thus the features used were primarily derived from the color space, including simple pixel look-ups and color gradient magnitudes. This makes less sense in the iris segmentation domain as most iris imagery is grayscale representation only an infrared channel. There is also no reason why we shouldn't learn from low level edge information produced by standard detectors such as the Canny edge detector. Thus we replaced the color space features with those derived from the raw grayscale image, as well as from the output of a Canny edge detector and from histogram equalization followed by anisotropic diffusion. Anisotropic diffusion [PM87] applied to a histogram equalized image serves suppress noise in the image while preserving the initially faint edges, such as those found along the eyelid, that histogram equalization will strengthen. Formally, anisotropic diffusion is defined as:

$$\frac{\partial I}{\partial t} = \nabla \cdot (c(x, y, t) \nabla I),$$

where  $\nabla \cdot$  is the divergence operator,  $c(x, y, t)$  is the diffusion coefficient that controls its rate, and  $\nabla I$  is the image gradient. In our case, the coefficient was chosen to be

$$c(\|\nabla I\|) = e^{-\|\nabla I\|/K},$$

where  $K$  is a constant that controls how sensitive the diffusion is to edges. An example of this process on a typical iris image from CASIA is found in Figure 2. A comparison with the output of Canny edge detection is shown in Figure 3, where the anisotropic diffusion output is inverted, thresholded, and then small blobs are filtered out. What can be seen relative to the Canny output is suppression of the eyelashes while the eyelid boundaries in the lower left corner of the eye are picked out.

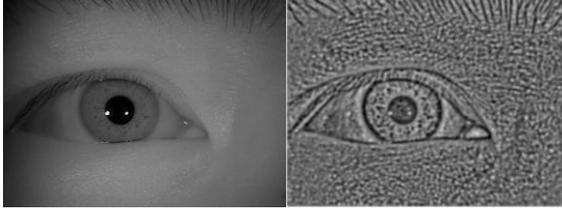


Figure 2: Histogram equalization followed by anisotropic diffusion (R) on an image (L) from the CASIA dataset



Figure 3: Comparison of Canny edge detection (L) and filtered anisotropic diffusion (R). Note the suppression of eyelashes and highlighting of faint eyelid boundaries

Similar to [DZ13], we compute the gradient magnitudes at two scales and the gradient orientations, which we discretize into four bins, but we do this individually for the grayscale intensity image and the output of anisotropic diffusion, producing four magnitude channels and 16 orientation features at each pixel. We add to this the raw intensity and diffused intensity values as well as the binary edge value from the Canny edge detection mask. In total, this gives us 23 features for each pixel in a patch, to which we apply the method of [DZ13] for downsampling and selecting patch features. This consists of downsampling by a factor of 2 in each patch dimension to get to 5888 features per patch. Each feature channel is blurred using a triangle blur and downsampled to  $5 \times 5$ , from which all pairwise differences for that channel are computed.

Thus, for 23 channels, we have  $23 \cdot \binom{5 \cdot 5}{2} = 6900$  additional features, for a total of 12788 for each patch.

## 5 Training and Testing Methodology

Ground truth edge masks for training and testing were created by hand-labeling randomly chosen subsets of the CASIA-Iris-Thousand and ND datasets with simple polygons representing the eyelid and ellipses representing the iris and pupil boundaries. The eyelid polygon is used to trim the iris and pupil ellipses when they overlap the eyelid.

We break up the labeled CASIA-Iris-Thousand dataset into a training and a testing hold-out set; we do the same for the ND dataset. We then train a classifier on the CASIA training data and test on its hold-out set, repeating the same procedure on the ND

dataset. Because cross-dataset validation is an important test of generalizability [To11], we then use the entire CASIA dataset as our training set and test on the entire ND set (and vice-versa). Our evaluation methodology is to generate Precision-Recall curves and Average Precision scores by varying the threshold at which a response from the classifier is considered an edge and checking whether there is a corresponding edge in the ground-truth edge mask. The method for determining the correspondence between a machine-generated edge and a ground-truth edge is described in detail in [Ma04]. It is formulated as a bipartite assignment problem with the distance between a machine-generated edge pixel and a ground-truth edge pixel serving as the weight between the two in the graph. Full details of this method are beyond the scope of this paper and can be found in the above reference. False positives are those machine-generated edge pixels that do not match with any ground-truth within a given distance threshold. False negatives are those ground-truth edge pixels that have no matching machine-generated edge pixels.

This training and testing pipeline is used for both a classifier trained with Canny and anisotropic diffusion features as well as one trained using only intensity as input to the algorithm. We compare both classifier types to the results from a Canny edge detector, this time varying its low threshold [Ca83] to generate Precision-Recall curves.

## 6 Results

On both the CASIA and ND datasets, both classifier types substantially outperform a Canny edge detector, as seen in Figure 4 and Figure 5. The Canny edge detector performance is shown in blue, our classifier using only intensity input is shown in green, and our classifier using Canny edge and anisotropic diffusion input is shown in red.

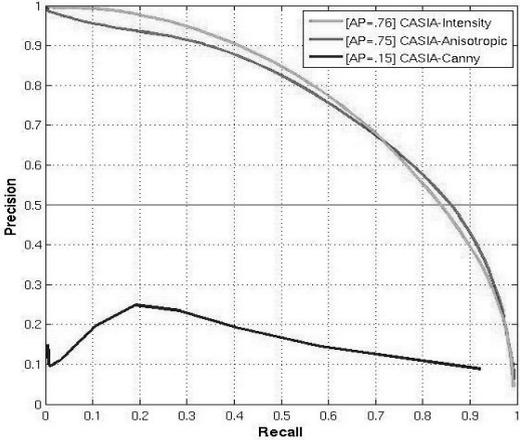


Figure 4: Precision-Recall for the CASIA dataset. Average Precisions are given in the Legend.

The classifier trained on the combination of intensity, Canny edge, and anisotropic diffusion slightly outperforms the classifier trained on intensity alone at the higher recall

levels for the CASIA set, but the difference is insignificant. An example of Canny edge detection versus the two classifier types on a CASIA image is shown in Figure 6. Note the suppression of responses to reflection edges in the learned edge detection (bottom row), and that both classifier types pick out eyelid edges in addition to finding iris and pupil boundaries.

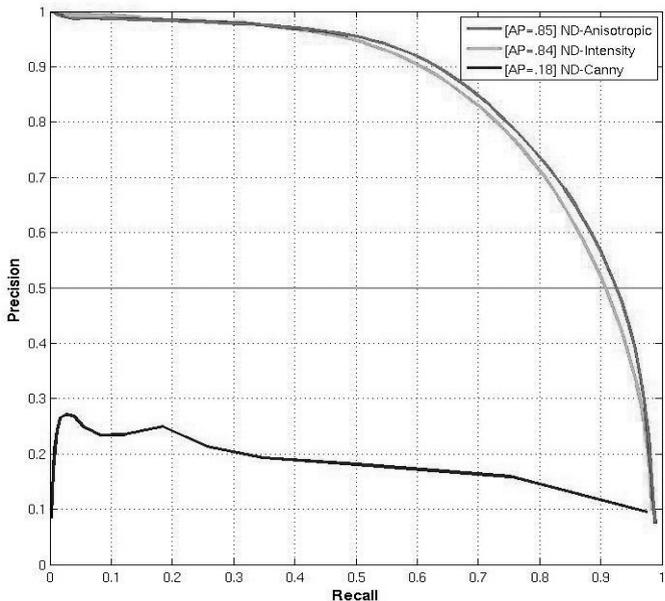


Figure 5: Precision-Recall Curve for the ND dataset

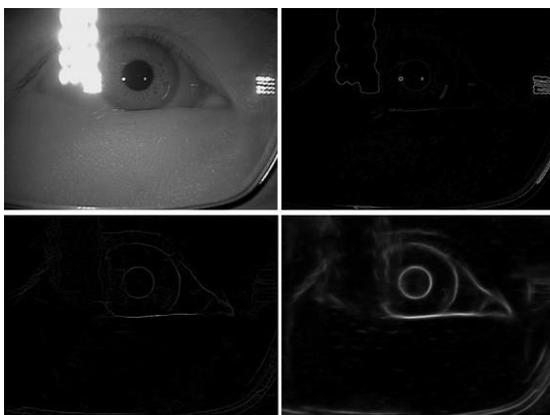


Figure 6: An example of Canny edge detection (Upper Right) and learned edge detection (Bottom Row) on the CASIA, with Intensity, Canny, and Anisotropic Diffusion inputs (Bottom Left) versus purely intensity inputs (Bottom Right).

The Precision-Recall curve for the ND dataset also shows a small gain for adding Canny edge and Anisotropic Diffusion features. An example of results on the ND set is given in Figure 7. Again, both classifiers pick out eyelid boundaries along with the pupil and iris edges while suppressing eyelash and reflection edges.

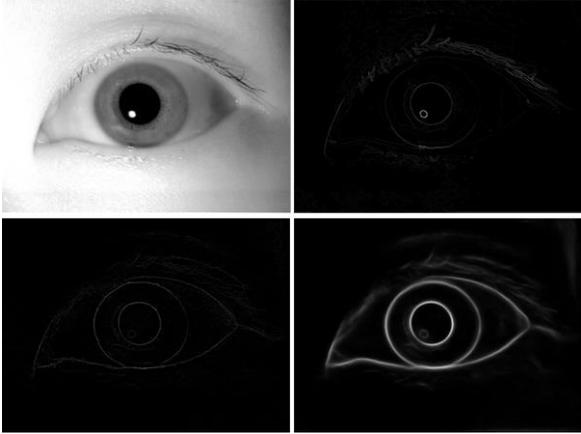


Figure 7: An example of Canny edge detection (Upper Right) and learned edge detection (Bottom Row) on the ND dataset, with Intensity, Canny, and Anisotropic Diffusion inputs (Bottom Left) versus purely intensity inputs (Bottom Right).

Our cross-dataset validation was conducted using CASIA to train and ND to test, and vice versa. There is a substantial drop-off in performance compared to the intra-dataset evaluation for both training-testing schemes, although the results still greatly exceed the accuracy of using the Canny edge detector to find these important boundaries. The benefit of using the additional anisotropic diffusion and Canny inputs over purely intensity becomes evident in the cross-dataset validation.

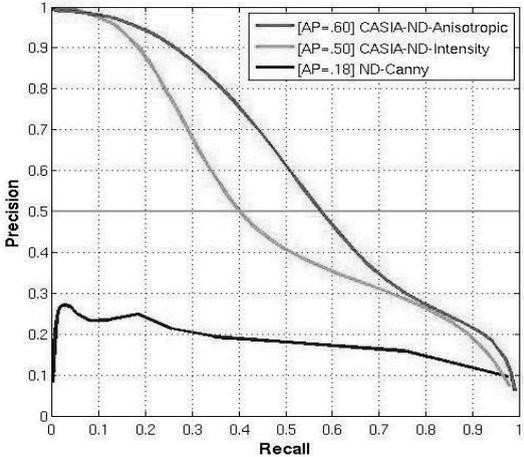


Figure 8: Precision-Recall Curve on the ND dataset for classifiers trained on the CASIA dataset

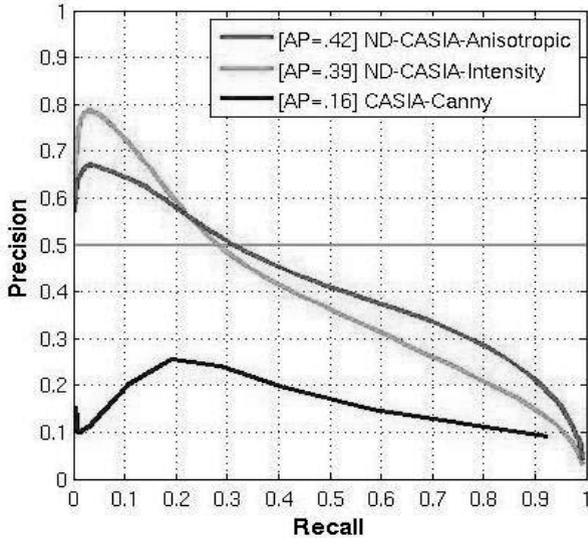


Figure 9: Precision-Recall Curve on the CASIA dataset for classifiers trained on the ND dataset

A comparison of the Average Precisions (APs) for each of the methods on the CASIA and ND datasets is given in Table 1. The two classifier types achieve APs that are 4-5 times greater than the Canny edge detector.

Table 1: Intra-dataset average precisions

Method	Training Set	Test Set (holdout )	Average Precision	Training Set	Test Set (holdout )	Average Precision
Classifier with Intensity inputs	CASIA	CASIA	0.76	ND	ND	0.3
Classifier with Canny, Anisotropic, and Intensity Inputs	CASIA	CASIA	0.75	ND	ND	0.85
Canny	N/A	CASIA	0.15	N/A	ND	0.18

The AP scores for the classifiers drop significantly in the inter-dataset test, though they still outperform the Canny edge detector by 2.5 to 3.5 times, as shown in Table 2. Note that the test sets are the full labeled datasets indicated rather than just their hold-out sets used in the intra-dataset comparison, hence the slight variations in Canny AP scores between the two experimental set-ups.

Table 2: Inter-dataset average precisions

Method	Training Set	Test Set (Full)	Average Precision	Training Set	Test Set (Full)	Average Precision
Classifier with Intensity inputs	CASIA	ND	0.5	ND	CASIA	0.39
Classifier with Canny, Anisotropic, and Intensity Inputs	CASIA	ND	0.6	ND	CASIA	0.42
Canny	N/A	ND	0.18	N/A	CASIA	0.16

## 7 Conclusion

The learned edge detectors described here significantly outperform the Canny edge detector at finding eyelid, pupil and iris boundaries while showing admirable resistance to irrelevant edges such as eyelashes. Using the boundaries generated by this methodology would remedy the weaknesses of edge-based approaches to segmenting the usable iris regions in eye imagery. There is considerable opportunity for improving even further on these results by developing additional features that generalize better across datasets.

## References

- [Ad08] Adam, M. et. al.: Reliable Eyelid Localization for Iris Recognition. In Proc. 10<sup>th</sup> Int. Conf. on Advanced Concepts for Intelligent Vision Problems, 2008.
- [Al11] Aligholizadeh, M. et. al.: Eyelid and Eyelash Segmentation based on Wavelet Transform for Iris Recognition. In Proc. 4<sup>th</sup> Int. Congress on Image and Signal Processing, 2011: pp. 1231-1235.
- [Br96] Breiman, L.: Bagging Predictors. In Machine Learning, 24(2), 1996; pp. 123-140.
- [Ca83] Canny, J.: A Variational Approach to Edge Detection. In Proc. of AAAI-83, 1983; pp. 54-58.
- [Ca11] CASIA Iris Image Database (ver 4.0). <http://www.sinobiometric.com/casiairis.html> (Retrieved 1-10-2011).
- [Da04] Daugman, J.: How Iris Recognition Works. In IEEE Trans. on Circuits and Systems for Video Technology, 14(1), 2004; pp. 21-30.
- [DZ13] Dollar, P.; Zitnick, C.: Structured Forests for Fast Edge Detection. In Proc. IEEE Int. Conf. on Computer Vision, Sydney, 2013; pp. 1841-1848.
- [Hu09] Huang, J. et. al.: A Novel Iris Segmentation using Radial-Suppression Edge Detection. In Signal Processing, 89(12), 2009; pp. 2630-2643.

- [Ko11] Kotschieder, P. et. al.: Structured Class-Labels in Random Forests for Semantic Image Labelling, In Proc. 13<sup>th</sup> Int. Conf. on Computer Vision, 2011; pp. 2190-2197.
- [Ma01] Martin, D. et. al.: A Database of Human Segmented Natural Image and its Application to Evaluating Segmentation Algorithms and Measuring Ecological Statistics. In Proc. 8<sup>th</sup> Int. Conf. on Computer Vision, 2001; pp. 416-423.
- [Ma03] Masek, L.: Recognition of Human Iris Patterns for Biometric Identification. M.S. Dissertation, University of Western Australia, 2003.
- [Ma04] Martin, D. et. al.: Learning to Detect Natural Image Boundaries Using Local Brightness, Color, and Texture Cues. IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 26, no. 5, 2004;; pp. 530-549.
- [PM87] Perona, P.; Malik, J.: Scale-space and Edge Detection Using Anisotropic Diffusion. Proc. IEEE Computer Society Workshop on Computer Vision, 1987; pp. 16-22.
- [Ph10] Phillips, P. et. al.: FRVT 2006 and ICE 2006 Large-Scale Experimental Results. In IEEE Trans. on Pattern Analysis and Machine Intelligence, 32(5), 2010; pp. 831-846.
- [To11] Torralba, A.; Efros, A.: Unbiased look at dataset bias. In Proc. 13<sup>th</sup> Int. Conf. on Computer Vision, 2011; pp. 1521-1528.
- [ZAS12] Zali-Vargahan, B.; Amirani, M.; Seyedarabi, H.: Contourlet Transform for Iris Image Segmentation. In Int. Journal of Computer Applications, 60(10), 2012; pp. 41-44.

# Predicting Dactyloscopic Examiner Fingerprint Image Quality Assessments

Martin Aastrup Olsen<sup>1</sup>, Martin Böckeler<sup>2</sup>, Christoph Busch<sup>2</sup>

<sup>1</sup>Faculty of Computer Science and Media Technology, Gjøvik University College, Norway, [martin.olsen@hig.no](mailto:martin.olsen@hig.no)

<sup>2</sup>da/sec Biometrics and Internet Security Research Group, Hochschule Darmstadt, Darmstadt, Germany, [christoph.busch@h-da.de](mailto:christoph.busch@h-da.de), [martin.boeckeler@googlemail.com](mailto:martin.boeckeler@googlemail.com)

**Abstract:** We work towards a system which can assist dactyloscopic examiners in assessing the quality and decision value of a fingerprint image and eventually a fingerprint. However when quality assessment tasks of dactyloscopic examiners are replaced by automatic quality assessment then we need to ensure that the automatic measurement is in agreement with the examiner opinion. Under the assumption of such agreement, we can predict the examiner opinion. We propose a method for determining the examiner agreement on ordinal scales and show that there is a high level of agreement between examiners assessing the ground truth quality of fingerprints. With ground truth quality information on 749 fingerprints and using 10-fold cross validation we construct models using Support Vector Machines and Proportional Odds Logistic Regression which predicts median examiner quality assessments 35% better than when using the prior class distribution.

## 1 Introduction

Fingerprint sample quality in the context of forensic applications where an examiner following the Analysis, Comparison, Evaluation, and Verification (ACE-V) protocol is part of the initial information gathering phase where the examiner studies the impression to quantify the present discriminating information and assess the quality and completeness [Exp12].

The quality assessment will among other factors such as the completeness of the fingerprint have an impact on the decision value of the impression, which can be one of: Value for Individualization (VID), which is used when the quantity and quality of the information present is deemed sufficient to determine if the impression is from the same source as another, yet unseen, impression; Value for Exclusion Only (VEO) is used when sufficient information is present to determine that the impression is not from the same source as another impression; No Value (NV) is used when the impression is deemed unsuitable. The process of assigning VID, VEO, and NV is inherently subjective and requires training and experience to perform accurately and consistently. Ulery et al. conducted a study on the accuracy and reliability of 169 forensic examiners who assigned VID, VEO, and NV decisions to 100 latents and found that VID decisions were unanimous in 48% of cases for mated pairs and 33% for non-mated pairs [UHBR11]. In a related study on the

repeatability and reproducibility of decision by individual examiners, it was found that 93% of VID, 85% NV, and 55% VEO decisions were repeated by individual examiners when presented twice with the same impression after a 7 month interval [UHBAR12]. These findings, which indicate a high degree of accuracy with respect to VID, and lower accuracy with respect to VEO mirror the findings by Langenburg [Lan09].

It has been demonstrated that when provided with extraneous contextual information, an examiner might change the method of judging and comparing fingerprints [DCP05]. Additionally, the examiners are vulnerable to biasing information such as evidence of confession of a crime, even in cases where the comparison of the impressions is non-difficult [DC06].

Biometric sample quality has successfully been applied in the context of Automated Fingerprint Recognition Systems (AFIS) to reject samples which are likely to contribute negatively to False Non-Match Rates (FNMR). By rejecting those samples before they are enrolled, a high level of biometric performance is achievable and with it higher levels of satisfaction by the users interacting with the biometric system [WGW04].

We are motivated by the successful application of automated quality assessment in AFIS and by the findings of Ulery et al. and Bradford et al., which highlight the subjective nature of the ACE-V protocol, to determine methods which objectively assesses the quality of an impression in the form of a fingerprint or a fingermark. This paper represents one step in this direction and our main objective is to determine methods which predict the quality assessment that a dactyloscopic examiner gives a particular fingerprint. To achieve this, we leverage a dataset which contains ground truth quality labels on inked impressions as assessed by dactyloscopic examiners from the German Federal Criminal Police Office (BKA), and apply quality assessment algorithms identified or developed in the context of NFIQ 2.0 [Nat14].

The rest of the paper is organized as follows: section 2 outlines state of the art methods for objective quality assessment. In section 3 we propose a method for quantifying examiner agreement, section 4 details the ground truth dataset on which we base our experiments that are described in section 5. We discuss our results in section 6 and conclude in section 7.

## 2 Fingerprint quality

There exists a large number of fingerprint image quality assessment algorithms in the literature and several reviews have been made, e.g. in the context of optical and capacitive sensors [AFRM<sup>+</sup>07, AFRM<sup>+</sup>08]; relation between quality assessed by human experts and algorithms and comparison scores [FAMSAFOG05]; and more recently quality assessment using no-reference algorithms have been investigated [EANCR13]; a comprehensive review of biometric sample quality is provided by Bharadwaj et al. [BVS14] and a quality metric for fingermarks has been proposed [YCLJ13].

We have selected a subset based on those features specified in the NFIQ 2.0 quality feature definitions document version 0.5 [NFI12] which we summarize here: Frequency Domain Analysis ( $Q_{\text{FDA}}$ ) uses the magnitude of the dominant frequency as determined by discrete Fourier transform fingerprint image as a local quality value. Local Clarity Score ( $Q_{\text{LCS}}$ )

determines the clarity of the fingerprint image by estimating how well each block in the fingerprint image can be segmented into ridge and valley region. Orientation Flow ( $Q_{OFL}$ ) measures the continuity of ridge flows in the fingerprint image by determining the dominant ridge orientation agreement between one block and its neighbouring blocks. Orientation Certainty Level ( $Q_{OCL}$ ) is a measure of the strength of the ridge orientation within a image block. A high score indicates that the ridge orientation within the block is well defined. Ridge Valley Uniformity ( $Q_{RVU}$ ) measures the consistency between ridge and valley widths within each image block. The widths of ridges and valleys are expected to be similar across the entire fingerprint image. Radial Power Spectrum ( $Q_{RPS}$ ) quantifies the energy concentration within a specified band in the Fourier spectrum. The limits of the band are determined by the expected ridge valley frequency. Image mean ( $Q_{MU}$ ) is the mean value of pixel intensities across the fingerprint image. Image standard deviation ( $Q_{STD}$ ) is the standard deviation of pixel values across the fingerprint image.

The features  $Q_{FDA}$ ,  $Q_{LCS}$ ,  $Q_{OFL}$ ,  $Q_{OCL}$ ,  $Q_{RVU}$  operate on 32 by 32 pixel non-overlapping regions of the image and thus provide a vector of local quality values.  $Q_{RPS}$ ,  $Q_{MU}$ ,  $Q_{STD}$  work on the entirety of the image and provide a scalar value. Based on these two groups we define two sets of features: Set A which contains the mean and standard deviation of each of the local quality vectors of  $Q_{FDA}$ ,  $Q_{LCS}$ ,  $Q_{OFL}$ ,  $Q_{OCL}$ ,  $Q_{RVU}$  giving a total of 10 features; set B contains, in addition to the features in set A,  $Q_{RPS}$ ,  $Q_{MU}$ ,  $Q_{STD}$  for a total of 13 features.

### 3 Quantifying examiner agreement on ordinal scales

To quantify examiner agreement, which is an essential task to determine to which degree examiners agree on what quality means and to compare and judge how well automatic quality prediction will perform against its human counterpart, the following requirements are specified for an examiner agreement coefficient:

1. an unlimited number of assessments on a single fingerprint shall be considered
2. the agreement of assessments shall be weighted according to their distance
3. assessments that belong to the same decision categories shall be assigned with high weights
4. when the assessment scale varies the measure results shall remain consistent

Ad 1 - the coefficient shall be capable of measuring agreement for fingerprints that were annotated by a minimum of 2 examiners and for fingerprints that are annotated by more than 2 examiners. Ad 2 - assessments that are “closer” to each other shall result in higher agreement. Ad 3 - not only the “distance” of single assessments shall be measured, assessments in equal decision categories shall be weighted higher than assessments that are not in the same decision category. Ad 4 - quality assessments in varying assessment scales shall result in the same agreement if the relative distance of the single ratings are the same. For example, a quality assessment of  $x_{11} = 1, x_{12} = 2, x_{13} = 5$   $\{x_{1x} \in \mathbb{N} | 1 \leq x_{1x} \leq 5\}$  and a quality assessment of  $x_{21} = 1, x_{22} = 25, x_{23} = 100$   $\{x_{2x} \in \mathbb{N} | 1 \leq x_{2x} \leq 100\}$  shall result in the same agreement coefficient.

Several common statistical measures like the Percentage agreement  $\bar{P}$  [UHBR12], the Interquartile Range  $IQR$  [ZK99, p.27], the Median Absolute Deviation  $MAD$  [HMT82,

p.220] and the Standard Deviation  $SD$  [ZK99, p.26] were investigated to determine if they fulfil the specified requirements. Table 1 displays how these measures perform on various assessment examples. It is clear that none of the common measures are able to measure examiner agreement sufficiently as they all violate one of the predefined requirements.

		Assessment example							
		1	2	3	4	5	6	7	8
Assessment	excellent, 1	1 2 3	1 2	1 2	1	1	1	1 2 3 4 5	1 2
	very good, 2		3		2	2			3
	good, 3			3	3		2		4
	fair, 4								5
	poor, 5					3	3	6	6
Agreement	$\bar{P}$	1.000	0.333	0.333	0.000	0.000	0.000	0.667	0.067
	$IQR$	0.000	1.000	2.000	2.000	4.000	4.000	0.000	3.000
	$MAD$	0.000	0.000	0.000	1.000	1.000	2.000	0.000	1.500
	$SD$	0.000	0.471	0.943	0.816	1.700	1.633	1.491	1.491
	$CMCA$	1.000	0.839	0.689	0.422	0.166	0.125	0.765	0.208

Table 1: Examples of agreement values computed using  $\bar{P}$ ,  $IQR$ ,  $MAD$ ,  $SD$ ,  $CMCA$  for 8 cases of examiner assessments where the number of experts and their assessments vary. In the top half of the table, each dot represents an examiner assessment ranging from excellent (1) to poor (5). The bottom half of the table shows the agreement value assigned by the 5 metrics for each assessment example.

$\bar{P}$  is equal for assessment examples 2 and 3, and examples 4, 5 and 6 thus violating requirement 2.  $IQR$  is equal for the assessment examples 3 and 4 violating requirements 2 and 3 as also for assessment example 5 and 6 which violates requirement 2.  $MAD$  is equal for the first 3 examples thus violating requirement 2 and 3 and in the 4th and 5th example requirement 2 is violated.  $SD$  violates requirement 3 in examples 2, 3, 7 and 8. Requirement 4 is violated by  $IQR$ ,  $MAD$ ,  $SD$  as the produced values depend on the range of the scale.

None of the measures described above satisfy the task of measuring examiner agreement as specified by our requirements, and hence we propose a new coefficient which does fulfil the specified requirements: Closest-neighbour Median Cluster Agreement ( $CMCA$ ).

$CMCA$  consists of multiple parts which are calculated as follows. Let  $x_j$  be the  $j^{th}$  ascending sorted rating on the fingerprint. The closest neighbour distance  $ND_j \in [0, 1]$ , one part to fulfil requirement 2, of the  $j^{th}$  rating on the fingerprint is:

$$ND_j = \min(|x_j - x_{j-1}|, |x_j - x_{j+1}|) \tag{1}$$

Further, let  $\max i$  the maximum possible rating on the rating scale,  $\min i$  the minimum possible rating on the scale,  $\tilde{x}$  the median of ratings and  $r$  the number of ratings on the fingerprint. The distance consensus  $D \in [0, 1]$  of the fingerprint made to fulfil requirements

2 and 4 is:

$$D = \frac{\left( \sum_{j=1}^r \left( 1 - \frac{ND_j}{\max i - \min i} \right) + \left( 1 - \frac{|x_j - \bar{x}|}{\max i - \min i} \right) \right) - 1}{2r - 1} \quad (2)$$

Let  $c_j$  be the  $j^{\text{th}}$  cluster (ratings in the same decision category) and let  $|c_j|$  be its cardinality. Let  $nc$  be the number of clusters on. The average cluster size difference  $CSD \in [0, r - 2]$ , which compares the size of each assessment cluster against each other and calculates the mean of their size differences is:

$$CSD = \frac{\sum_{j=1}^{nc} \sum_{k=j}^{nc} ||c_j| - |c_k||}{\frac{(nc-1) \cdot nc}{2}} \quad (3)$$

To fulfil requirement 3, the distance consensus  $D$  is multiplied by the cluster size power  $CSP \in [1, r - 1]$ , which is calculated as:

$$CSP = nc - \frac{CSD}{r} \quad (4)$$

The  $CMCA$  on the fingerprint is finally calculated by:

$$CMCA = \begin{cases} 1.0 & , nc = 1 \\ 1.0 - \frac{|x_1 - x_2|}{\max i - \min i} & , r = 2 \\ D^{CSP} & , \text{otherwise} \end{cases} \quad (5)$$

The  $CMCA \in [0, 1]$  over a set of fingerprints  $I$  is calculated as the arithmetic mean over the  $CMCA$  agreements of each fingerprint in  $I$ .

The measurement results of  $CMCA$  applied to the 8 assessment examples are shown in the last row of table 1. Assessment examples 5 and 6, show that  $CMCA$  fulfils requirement 2 by considering the distance of between single assessments. It also fulfils requirement 3, heavy weighted assessments that belong to the same decision category. The example shows that the  $CMCA$  fulfils requirement 1, measuring agreement for any number of assessments on a single fingerprint.

After all, the  $CMCA$  was designed to measure examiner agreement per fingerprint, see eq. (5). Other coefficients like Cohens Kappa [Coh60] or Fleiss Kappa [Fle71] were designed to measure inter-rater agreement over the whole assessment population which is expressed by the arithmetic mean of the set of  $CMCA$ . Nonetheless, Cohens Kappa has the disadvantage that it can only measure pairwise rater agreement on nominal data. In addition, Fleiss Kappa is able to measure inter-rater agreement for more than 2 raters, but was also designed for nominal data, so additional information of the natural order of categories on ordinal data like it is present in our case of fingerprint quality, would not be considered properly.

## 4 Ground truth data

The ground truth data of this paper, representing the human examiner quality assessment part, originates from the NIST special database 14 [Wat93], containing 54 000 fingerprints from live-scan and scanned ink impressions and from the NIST special database 29 [Wat01], containing 4 320 fingerprints. In 2009, a team of 9 dactyloscopic examiners from the BKA annotated for the purpose of a conformance testing study [BLTK09] several fingerprint characteristics, like minutia points, singular points and the overall fingerprint quality which is investigated in this paper. To establish an objective annotation, the examiner team was equipped with a simple graphical user interface that omitted the support of the automatic or semi-automatic minutia extraction functionality of the AFIS. To further increase the objectivity and anonymity of the process, each examiner was assigned with an ID that was not known to other examiners. The examiners rated the overall fingerprint quality within 5 decision categories, ranging from excellent (1), very good (2), good (3), fair (4) to poor (5). A total of 749 fingerprints were annotated by at least 2 examiners. Figure 1 shows the logical partitioning of the annotated samples, where the first level contains the set of 749 fingerprints; the second level shows how many samples were annotated grouped by the number of examiners. The third level shows the number of samples each distinct group of examiners annotated, e.g. there were 713 fingerprints annotated by 3 examiners, which came from two groups of 3 examiners annotating respectively 361 and 352 fingerprints ( $S_2$  and  $S_3$ ). Table 2 shows the number of images that were annotated by each of the 9 examiners. We note that examiners 11 to 16 have each annotated nearly 400 samples, while examiners 17 to 19 have annotated fewer than 20 samples.

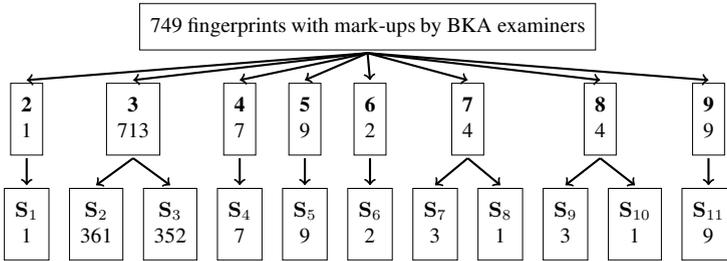


Figure 1: Examiner markup tree. The bold digits in the second tree level display how many examiners annotated a specific number of fingerprints. The last tree level shows how many fingerprints where annotated by distinct groups of examiners.

	Examiner								
	11	12	13	14	15	16	17	18	19
Annotations	396	393	397	378	388	371	10	17	17

Table 2: The number of images annotated by each of the 9 examiners.

Table 3 summarizes the *CMCA* computed on median ground truth quality levels and for all fingerprint images in the ground truth data set. The mean *CMCA* is .88 across all

Quality	n	CMCA				
		min	max	mean	median	std
1	41	.69	1.00	.82	.84	.08
2	306	.42	1.00	.86	.84	.15
3	305	.42	1.00	.90	.84	.10
4	92	.42	1.00	.88	.84	.11
5	5	.84	.84	.84	.84	.00
All	749	.42	1.00	.88	.84	.13

Table 3: Summary statistics of *CMCA* at each median ground truth quality level and across all samples.

fingerprint images, indicating that there is generally consensus between examiners when subjectively assessing the quality level of a fingerprint image. At the individual levels we note some differences in the *CMCA*, in particular that mean *CMCA* at quality level 1 is .82 while for level 3 it is .90.

Figure 2 shows an example fingerprint for each median quality assessment. For median quality levels 1 through 4 the examiners were in agreement with  $CMCA = 1.0$  while for the fingerprint illustrating level 5 where  $CMCA = 0.838$ . We observe that at level 1 (left most figure) the ridge lines are clearly separated around the core, where the examples for levels 4 and 5 appear without clearly defined ridge lines and with blurry or low contrast regions.

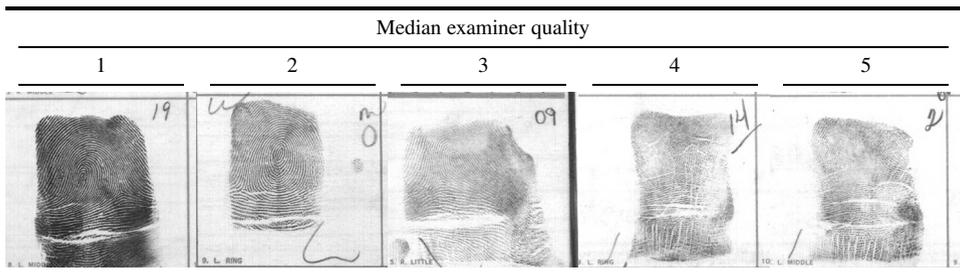


Figure 2: Median examiner quality assessment with example fingerprints. All images are from the NIST special database 14 [Wat93] (file names from left to right: f0000118, f0000109, f0000095, f0000969, f0000968).

To assess whether human examiner quality assessments are indicative of the eventual genuine comparison score of the fingerprint sample, we computed the genuine comparison scores for all samples and grouped them according to the examiner who made the assessment and the quality score that was assigned. Box plots showing the relation between assigned quality and genuine comparison score [id314] for each of the 6 examiners who annotated the most images (see table 2), as well as the median examiner quality are depicted in fig. 3. The plots show that generally a higher quality score is associated with a higher comparison score, however, for some cases we note some irregularities. For Ex-

aminer 11 (left most plot), we note that no images were assigned quality level 5 and that those samples which received a quality level of 1 were involved in comparisons resulting in scores similar to those samples which were assessed as being quality level 2.

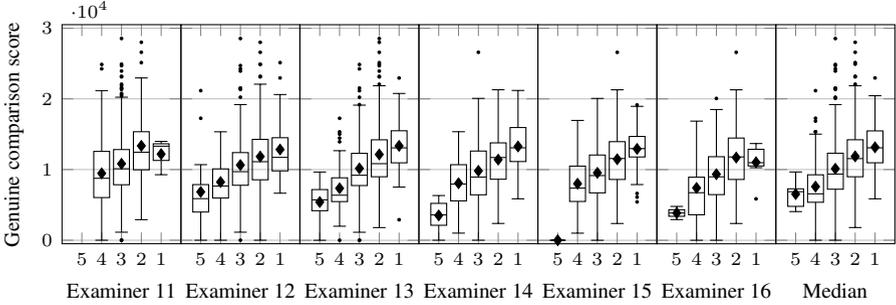


Figure 3: Boxplots of genuine comparison score for each of 5 quality levels assigned by 6 examiners and for the median of assigned quality levels.

## 5 Experiments

Our goal is to predict the quality level that an examiner will assign to a given fingerprint. We note from table 3 that our proposed *CMCA* coefficient indicates a high degree of agreement between examiners as to the assigned ground truth quality levels for the data set.

We perform a series of experiments in order to assess to which degree the quality scores assigned by individual examiners or the median quality score is predicted. From table 3 we note that the distribution of median quality levels is not uniform with the majority of samples being assigned levels 2 or 3 and with only 5 samples in level 5. Due to the low annotation count by examiners 17, 18, and 19 we do not attempt to predict their assessments (see table 2) and instead only assess examiners 11 to 16 individually.

We train our predictive models using Multi-class Support Vector Machine (SVM) [CV95] and Proportional Odds Logistic Regression (POLR) [MZ75] where the response variable is the assigned quality level (either individual examiner or median of examiners), and the explanatory variables are features in sets A or B (see section 2).

The experiments are performed using 10-fold cross validation, i.e., we divide the available data in each experiment randomly into 10 disjunct partitions of equal size. Over the 10 possible permutations we perform training of SVM and POLR on the 9 folds and test the performance on the remaining fold. In the case of SVM we use Radial Basis Function as kernel and perform a grid search for optimal sample influence radius ( $\gamma$ ) and cost ( $C$ ) over  $\gamma \in \{0.001, 0.01, \dots, 1000\}$  and  $C \in \{0.001, 0.01, \dots, 1000\}$  given the training folds. In the case of POLR no parameter optimization is performed.

The predictive capability of the constructed models in each experiment setting is determined by calculating the mean and standard deviation of the  $F_1$  score and Cohen’s Kappa

( $\kappa$ ) [Coh60] over the 10 permutations.

Cohen’s Kappa quantifies the class agreement between the model predictions and the ground truth by taking into account the observed probabilities of the classes. When  $\kappa = 0$  the agreement is equal to that which can be achieved by random chance based on the priors - when  $\kappa = 1$  then the agreement is complete.  $F_1$  is the harmonic mean of precision and recall and the lowest score is achieved when  $F_1 = 0$  and highest when  $F_1 = 1$ .

Experiments were performed using R [R C14] with SVM from e1071 [MDH<sup>+</sup>14]; POLR from MASS [VR02]; cross validation and miscellaneous functions from caret [KWW<sup>+</sup>15] and xtable [Dah14] packages.

## 6 Results

Our analysis of the ground truth data set (section 4) showed that there is a high level of agreement in assessing quality levels across examiners, and that the higher quality levels are associated with higher genuine comparison scores.

Following the protocol described in section 5 we performed a total of 14 experiments and have summarized the mean and standard deviations of  $F_1$  and  $\kappa$  for each of them in table 4. Each line in the table alternates between feature set *A* and *B* in *Set* column with the *Target* column indicating what is being predicted where *Median* indicates that it is the median of examiner quality assessments per fingerprint that is predicted, and *Examiner 11* indicates that it is the quality assessments of Examiner 11 which are predicted. The remaining columns are first grouped by  $F_1$  and  $\kappa$ , next by method *SVM* or *POLR* and finally arithmetic mean (*mean*) and standard deviation (*std*) of testing results across the 10 fold cross validation.

Set	Target	$F_1$				$\kappa$			
		SVM		POLR		SVM		POLR	
		mean	std	mean	std	mean	std	mean	std
A	Median	.58	.05	.59	.06	.31	.07	.33	.07
B	Median	.60	.08	.60	.06	.34	.12	.35	.07
A	Examiner 11	.72	.05	.70	.03	.21	.17	.13	.10
B	Examiner 11	.72	.05	.71	.06	.28	.14	.18	.16
A	Examiner 12	.45	.06	.48	.05	.17	.07	.23	.07
B	Examiner 12	.51	.05	.52	.07	.26	.08	.29	.08
A	Examiner 13	.49	.07	.49	.08	.21	.09	.21	.12
B	Examiner 13	.53	.08	.52	.08	.30	.10	.26	.09
A	Examiner 14	.57	.10	.59	.09	.30	.14	.33	.13
B	Examiner 14	.57	.09	.60	.09	.29	.14	.36	.13
A	Examiner 15	.57	.06	.60	.09	.29	.08	.35	.14
B	Examiner 15	.58	.07	.61	.08	.32	.09	.37	.12
A	Examiner 16	.63	.11	.64	.09	.33	.21	.36	.16
B	Examiner 16	.64	.12	.65	.11	.36	.21	.38	.19

Table 4: Results of experiments in predicting median and individual examiner quality assessment using SVM and POLR on feature sets A and B.

The best prediction results when using Median as target is achieved with POLR and feature

set B when considering either of  $F_1$  and  $\kappa$  as evaluation criteria. We see that the mean  $F_1$  is .60 for both SVM and POLR, but the standard deviation is smaller in the case of POLR with .06 over .08 achieved with SVM. For  $\kappa$  we note a mean of .35 for POLR and .34 for SVM, again with a smaller standard deviation favouring POLR.

We note that both  $F_1$  and  $\kappa$  spans a wide range when using individual examiner quality assessments as target for the predictions. Examiner 12 appears to be hardest to predict with a mean  $\kappa$  of .17 and .23 and  $F_1$  of .45 and .58 for respectively SVM and POLR when using feature set A. Prediction of the scores assigned by Examiner 16 using feature set B gives the highest mean  $F_1$  of .64 and .65 and  $\kappa$  of .36 and .38, however in both cases the standard deviation over the 10 folds is the highest of the experiments performed.

Generally using the global features present in feature set B lead to marginal increases in the mean  $F_1$  score while  $\kappa$  is increased further for both SVM and POLR.

In addition to SVM and POLR listed in table 4 we also used Recursive Partitioning, however that algorithm had difficulty working with the relatively small dataset with 10-fold cross validation and was thus not able to complete all intended experiments.

## 7 Conclusion

In this paper we have made steps towards assisting dactyloscopic examiner in assessing the quality of a given fingerprint or fingermark with the aim of determining its decision value in the ACE-V protocol. We address the objective quality of fingerprints and relation to examiner opinion with a continued goal to extend the quality assessment evaluation to fingermarks which pose the greatest challenge in the forensic evaluation.

We proposed the *CMCA* coefficient as a general method for quantifying examiner agreement on ordinal scales containing any number of categories. Using *CMCA* we have shown that there is a high level of agreement between examiners as to what constitutes a high quality fingerprint and further that the ground truth assessments made by dactyloscopic examiners are indicative of genuine comparison scores.

On our limited dataset containing 749 finger images and using 13 quality features we have constructed a model which predicts examiner quality assessments around 35% better than random chance given the prior quality class probabilities as measured using Cohen's Kappa.

Our future work includes refinement of the quality feature set to improve the predictive capabilities; evaluation of the importance of the individual features to gain insights as to which image covariates are important to examiners; evaluation of the trained model to determine the degree that the assigned quality levels are indicative of biometric performance; and finally how the trained models can be incorporated in ACE-V protocol to assist the decision making of dactyloscopic examiners when working with fingerprints or fingermarks.

## 8 Acknowledgements

We thank the team of dactyloscopic examiners with the German Federal Criminal Police Office for providing ground truth fingerprint information and National Institute of Stan-

dards and Technology for providing the selection of images from SD14 and SD29.

## References

- [AFRM<sup>+</sup>07] Fernando Alonso-Fernandez, Fabio Roli, Gian Luca Marcialis, Julian Fierrez, and Javier Ortega-Garcia. Comparison of fingerprint quality measures using an optical and a capacitive sensor. In *Proc. IEEE Conference on Biometrics: Theory, Applications and Systems, BTAS*, pages 1–6, September 2007.
- [AFRM<sup>+</sup>08] F. Alonso-Fernandez, F. Roli, G.L. Marcialis, J. Fierrez, J. Ortega-Garcia, and J. Gonzalez-Rodriguez. Performance of fingerprint quality measures depending on sensor technology. *SPIE Journal of Electronic Imaging, Special Section on Biometrics: Advances in Security, Usability and Interoperability*, 17(1), January–March 2008.
- [BLTK09] Christoph Busch, Dana Lodrova, Elham Tabassi, and Wolfgang Krodel. Semantic Conformance Testing for Finger Minutiae Data. *Proceedings of the 1st International Workshop on Security and Communication Networks (IWSCN)*, pages 1 – 7, 2009.
- [BVS14] Samarth Bharadwaj, Mayank Vatsa, and Richa Singh. Biometric quality: a review of fingerprint, iris, and face. *EURASIP Journal on Image and Video Processing*, 2014(1), 2014.
- [Coh60] Jacob Cohen. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20(1):37–46, 1960.
- [CV95] Corinna Cortes and Vladimir Vapnik. Support-Vector Networks. *Machine Learning*, 20(3):273–297, 1995.
- [Dah14] David B. Dahl. *xtable: Export tables to LaTeX or HTML*, 2014. R package version 1.7-4.
- [DC06] Itiel E. Dror and David Charlton. Why experts make errors. *Journal of Forensic Identification*, 56(4):600, 2006.
- [DCP05] Itiel E. Dror, David Charlton, and Ailsa E. Péron. Contextual information renders experts vulnerable to making erroneous identifications. *Forensic Science International*, 156(1):74–78, January 2005.
- [EANCR13] M. El Abed, A Ninassi, C. Charrier, and C. Rosenberger. Fingerprint quality assessment using a no-reference image quality metric. In *Signal Processing Conference (EUSIPCO), 2013 Proceedings of the 21st European*, pages 1–5, Sept 2013.
- [Exp12] Expert Working Group on Human Factors in Latent Print Analysis. *Latent print examination and human factors : improving the practice through a systems approach*. US Department of Commerce, National Institute of Standards and Technology, February 2012.
- [FAMSAFOG05] J. Fierrez-Aguilar, L. M. Muñoz-Serrano, F. Alonso-Fernandez, and J. Ortega-Garcia. On the effects of image quality degradation on minutiae- and ridge-based automatic fingerprint recognition. In *Proc. IEEE Intl. Carnahan Conf. on Security Technology, ICCST*, pages 79–82, October 2005.
- [Fle71] Joseph L Fleiss. Measuring nominal scale agreement among many raters. *Psychological bulletin*, 76(5):378, 1971.
- [HMT82] David C. Hoaglin, Frederick Mosteller, and John W. Tukey, editors. *Understanding Robust and Exploratory Data Analysis (Wiley Series in Probability and Statistics)*. Wiley, 1 edition, 12 1982.

- [id314] id3 Technologies. id3 Fingerprint SDK v. 1.4.1. <http://www.id3.eu/>, 2014.
- [KWW<sup>+</sup>15] Max Kuhn, Jed Wing, Steve Weston, Andre Williams, Chris Keefer, Allan Engelhardt, Tony Cooper, Zachary Mayer, Brenton Kenkel, the R Core Team, Michael Benesty, Reynald Lescarbeau, Andrew Ziem, and Luca Scrucca. *caret: Classification and Regression Training*, 2015. R package version 6.0-41.
- [Lan09] Glenn Langenburg. A Performance Study of the ACE-V Process: A Pilot Study to Measure the Accuracy, Precision, Reproducibility, Repeatability, and Biasability of Conclusions Resulting from the ACE-V Process. *Journal of Forensic Identification*, 59(2):219–257, 2009.
- [MDH<sup>+</sup>14] David Meyer, Evgenia Dimitriadou, Kurt Hornik, Andreas Weingessel, and Friedrich Leisch. *e1071: Misc Functions of the Department of Statistics (e1071)*, TU Wien, 2014. R package version 1.6-4.
- [MZ75] Richard D. McKelvey and William Zavoina. A statistical model for the analysis of ordinal level dependent variables. *The Journal of Mathematical Sociology*, 4(1):103–120, 1975.
- [Nat14] National Institute of Standards and Technology. Development of NFIQ 2.0. [http://www.nist.gov/itl/iad/ig/development\\_nfiq\\_2.cfm](http://www.nist.gov/itl/iad/ig/development_nfiq_2.cfm) – accessed January 2015, 2014. Development of NFIQ 2.0.
- [NFI12] NFIQ 2.0 Team. Development of NFIQ 2.0 – Quality Feature Definitions – Version 0.5. [http://biometrics.nist.gov/cs\\_links/quality/NFIQ\\_2/NFIQ-2\\_Quality\\_Feature\\_Defin-Ver05.pdf](http://biometrics.nist.gov/cs_links/quality/NFIQ_2/NFIQ-2_Quality_Feature_Defin-Ver05.pdf) – accessed January 2015, 2012.
- [R C14] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2014.
- [UHBAR12] Bradford T. Ulery, R. Austin Hicklin, JoAnn Buscaglia, and Maria Antonia Roberts. Repeatability and Reproducibility of Decisions by Latent Fingerprint Examiners. *PLoS ONE*, 7(3), March 2012.
- [UHBR11] Bradford T. Ulery, R. Austin Hicklin, JoAnn Buscaglia, and Maria Antonia Roberts. Accuracy and reliability of forensic latent fingerprint decisions. *Proceedings of the National Academy of Sciences*, 108(19):7733–7738, 2011.
- [UHBR12] Bradford T. Ulery, R. Austin Hicklin, JoAnn Buscaglia, and Maria Antonia Roberts. Repeatability and Reproducibility of Decisions by Latent Fingerprint Examiners. *PLoS ONE*, 7(3):e32800, 03 2012.
- [VR02] W. N. Venables and B. D. Ripley. *Modern Applied Statistics with S*. Springer, New York, fourth edition, 2002. ISBN 0-387-95457-0.
- [Wat93] Craig I. Watson. Nist special database 14. Technical report, National Institute of Standards and Technology, 1993.
- [Wat01] Craig I. Watson. Nist special database 29. Technical report, National Institute of Standards and Technology, 2001.
- [WGW04] C. L. Wilson, M. D. Garris, and C. I. Watson. Matching Performance for the US-VISIT IDENT System Using Flat Fingerprints NISTIR 7110. Technical report, NIST, May 2004.
- [YCLJ13] Soweon Yoon, Kai Cao, Eryun Liu, and AK. Jain. LFIQ: Latent fingerprint image quality. In *Biometrics: Theory, Applications and Systems (BTAS), 2013 IEEE Sixth International Conference on*, pages 1–8, Sept 2013.
- [ZK99] Daniel Zwillinger and Stephen Kokoska. *CRC Standard Probability and Statistics Tables and Formulae*. CRC Press, 1 edition, 12 1999.

# Identification performance of evidential value estimation for fingermarks

J. Kotzerke<sup>\*†</sup>, S.A. Davis<sup>\*</sup>, R. Hayes<sup>‡</sup>,  
L.J. Spreeuwiers<sup>†</sup>, R.N.J. Veldhuis<sup>†</sup>, K.J. Horadam<sup>\*</sup>

<sup>\*</sup> School of Mathematical and Geospatial Sciences,  
RMIT University, Melbourne, Australia

<sup>†</sup> Services, Cybersecurity and Safety,  
University of Twente, Enschede, The Netherlands

<sup>‡</sup> Forensic Services Department, Victoria Police, Melbourne, Australia

{johannes.kotzerke, stephen.davis, kathy.horadam}@rmit.edu.au  
robert.hayes@police.vic.gov.au  
{l.j.spreeuwiers, r.n.j.veldhuis}@utwente.nl

**Abstract:** Law enforcement agencies around the world use biometrics and fingerprints to solve and fight crime. Forensic experts are needed to record fingermarks at crime scenes and to ensure those captured are of evidential value. This process needs to be automated and streamlined as much as possible to improve efficiency and reduce workload.

It has previously been demonstrated that is possible to estimate a fingermark's evidential value automatically for image captures taken with a mobile phone or other devices, such as a scanner or a high-quality camera.

Here we study the relationship between a fingermark being of evidential value and its correct and certain identification and if it is possible to achieve identification despite the mark not having sufficient evidential value. Subsequently, we also investigate the influence the capture device used makes and if a mobile phone is an option worth considering.

Our results show that automatic identification is possible for 126 of the 1,428 fingermarks captured by a mobile phone, of which 116 were marked as having evidential value by experts and 123 by an automated algorithm.

## 1 Introduction

Increases in the rate of reported crime are evident in Victoria. Official recorded offences for the year 2012/13 have risen by 3.4% to 406,497, compared to 2011/12 [Vic13]. Forensic experts must travel in many cases to the crime scene and collect the evidence themselves, spending a lot of time travelling. Highly trained specialists such as fingerprint examiners are valuable resources, making streamlining of processes and the search for tools to assist both experts and non-experts in the field a priority. Therefore, we want to determine if fingermarks are of insufficient evidential value as early as possible to en-

sure the marks collected are of sufficient evidential value and to assist in case evidence collection for the specialists. This can be achieved by using mobile phones to capture fingerprints, determine their binary evidential value and transmit the valuable ones directly to the forensics unit; all done automatically either at the scene or at the lab after mark development/enhancement. This task can be performed by regular police officers or professionals with a different area of expertise, thus allowing the fingerprint experts to focus on the analysis of the fingerprints.

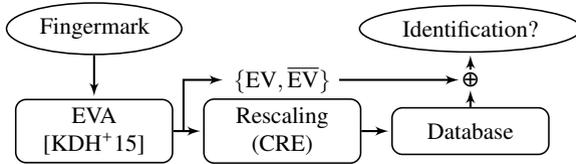


Figure 1: Diagram of the experiment performed. A fingerprint is captured, its evidential value  $\{EV, \overline{EV}\}$  is estimated by the Evidential Value Algorithm (EVA) of [KDH<sup>+</sup>15] and rescaled in the same way as in EVA. The number of correct and certain identifications (ccID) of the mark matched to a reference database is measured w.r.t. the image capture device and evidential value estimation method.

Previously, Kotzerke *et al.* have established that e.g. mobile phone images are suitable to estimate if a fingerprint is of sufficient evidential value (EV) and that an automated algorithm (EVA) can achieve results close to an expert assessment, based on the image quality [KDH<sup>+</sup>15]. Now, we extend this work and investigate the following worst case scenario. Are there any marks, which can be automatically and with certainty identified (against a reference database we collected) but are not of EV according to either the algorithm or the expert assessment from [KDH<sup>+</sup>15]? The proposed experiment is shown in Figure 1.

### 1.1 Background

Fingermarks are of essential value in order to exclude or to identify suspects. Nowadays, law enforcement agencies rely heavily on the fingerprint via automatic systems such as IAFIS and forensic experts [Mal09]. These examiners are expected to follow the Analysis, Comparison, Evaluation, and Verification (ACE-V) protocol [Ash99]. During the analysis phase, they decide if the mark at hand is of value for individualisation (VID), value for exclusion only (VEO) or no value (NV). Those with VID or VEO are EV; those with NV are  $\overline{EV}$ .

However, fingerprints suffer often from low quality due to being smudged or partial, overlap with other marks [FSZ12], or distorted by the surface pattern of the object they are found on [SHAF11]. Their forensic value is difficult to grasp for non-experts. Ulery *et al.* show that accuracy and repeatability varies even for forensic experts and mostly depends on the print quality [UHBR11, UHBR12], especially for borderline decisions. Consequently, Kellman *et al.* use image features to predict “expert performance and subjective

assessment of difficulty in fingerprint comparisons” [KME<sup>+</sup>14].

Most quality measures are used to prevent low quality images from being automatically matched because they tend to produce false minutiae and consequently false matches [AFFOG<sup>+</sup>07]. Therefore, they are suited to operational law enforcement agency setups and only optimised and tested for contact scanners [CDJ05, FKB06, LCCK08, The13] but not fingermarks. This has resulted in various algorithms tuned to a capture resolution of 500 ppi.

On the other hand, fingermarks require robust methods to estimate their quality because all factors mentioned above will vary and influence the quality and its estimate. Yoon and Jain demonstrated in [YJ13] that the current NIST quality estimator reference implementation NFIQ1 is outdated because IAFIS was able to return the print’s mate although it has been classified to have the lowest possible quality. Currently, NFIQ2 [The13] is under development and closing this gap; it is scheduled to be released soon. However, it is still primarily developed for fingerprints captured at a known resolution. In a scenario where the capture resolution is unknown, an estimate based on image features can improve the performance significantly [KDH<sup>+</sup>15].

Despite the need to reject low quality fingermarks for matching to avoid false identifications, the proposed scenario takes place much earlier. It includes the danger of missing a potentially valuable mark, which could solve a criminal case, because according to some algorithm the mark isn’t of EV. Naturally, there is a trade-off involved between the likelihood of missing some important marks and capturing as few marks as possible.

## 1.2 Outline

We investigate how the (estimated) EV of a fingermark influences an identification scenario, the performance achieved w.r.t. the capturing device (scanner, high-quality camera, phone) and capture resolution estimation (CRE) algorithm used (cf. Figure 1) and if any certain identifications would be missed if only marks of EV were to be analysed.

In the following sections we set up the methodology used during the identification scenario (Section 2), elaborate on the databases employed, perform identification experiments to demonstrate the interplay between a mark’s correct and certain identification (ccID) and if it is of EV, and discuss the results (Section 3). Finally, we summarise our findings and their implications and point out the direction for our future research (Section 4).

## 2 Methodology

The main idea behind our experiment is to evaluate if fingermarks, which are not of EV and hence wouldn’t be collected in a crime scene scenario, can be correctly identified with certainty (worst case). We now recap some important concepts relevant to the experiment, which have been introduced in [KDH<sup>+</sup>15].

As already motivated, there are scenarios when the capture resolution for an image is unknown because of an unconstrained setup. Most quality or feature extraction algorithms are optimised towards a certain resolution, most commonly 500 ppi and if the input image deviates from the assumed resolution, the applied algorithm usually falls short. Therefore it is sensible to perform a CRE. The RLAPS algorithm introduced in [KDH<sup>+</sup>15] estimates the inter-ridge spacing of a fingerprint or fingermark image and infers the capture resolution used. The power spectrum is computed and its radial average is determined only around its maximum peak within a certain frequency range. Finally, the assumption of an average inter-ridge spacing of 9 px for an adult is applied and leads to a capture resolution estimate.

Type of distortion	Number of marks taken	Prints of sufficient evidential value				
		Assessor 1	Assessor 2	Assessor 3	Ground truth	EVA
(i) light placement	168	48.2%	48.2%	48.2%	48.2%	54.2%
(ii) smeared	168	3.6%	4.2%	3.6%	3.6%	14.9%
(iii) finger twisted lightly	168	4.2%	4.8%	4.8%	4.8%	11.3%
(iv) strong twist	168	0.0%	0.0%	0.0%	0.0%	6.0%
(v) heavy placement	168	69.6%	65.5%	65.5%	65.5%	64.9%
(vi) partial, heavy placement	168	45.8%	48.2%	48.2%	48.2%	50.6%
(vii) normal	420	47.4%	49.0%	50.0%	49.0%	50.7%
Total	1,428	34.1%	34.5%	34.7%	34.5%	38.66%

Table 1: A breakdown of the 1,428 marks into the categories of distortion (including no deliberate distortion), and the final status of the assessment of the 3 experts in terms of the proportion of marks found to be of EV. Assessor 1’s opinion regarding the marks of categories (iii) and (vii) are respectively 9 and 21 decisions short of the total number. However, the other assessors agree on those marks and therefore a clear decision on ground truth can be made via majority vote. The EV distribution for EVA has been calculated for the mobile phone images which have been rescaled using CRE Global and the fused quality feature set at the decision threshold corresponding to the EER because of its performance (cf. Figure 3 and [KDH<sup>+</sup>15]).

Furthermore in [KDH<sup>+</sup>15], the idea of sufficient evidential value has been introduced and an algorithm to compute it based on image features has been presented. The feature sets of NFIQ2 as specified in the preliminary definition guide [The13], Neurotechnology Verifinger 6.7 [Neu15] and its quality value and the number of minutiae and their Fusion (concatenation of their feature vectors) have been investigated. We refer to this specific estimation algorithm as EVA. For details see [KDH<sup>+</sup>15].

Finally, we would like to clarify the concept of ccID. Assuming that a fingermark is compared against a reference database containing  $N$  unique fingerprints, a verification score  $S_i$  is returned for every comparison. We define a decision as correct and certain if and only if the mark and the print with the highest score are from the same subject *and* if the largest score is larger by factor  $d > 1$  than any other score:

$$\nexists S_j : S_i \leq dS_j, \quad i, j \in [1, N], \quad i \neq j. \quad (1)$$

This would lead to a correct and certain identification. One has to keep in mind that the smaller  $d$  is chosen, the greater the likelihood becomes that a decision is considered to be certain but is in fact due to low verification scores derived from poor quality images.

### 3 Experiments and discussion

In this section we evaluate the ability of an automatic system to perform a certain identification and how many of the fingerprint images are considered to be of EV in the context of different capture devices such as a flatbed scanner, a high-quality camera and a mobile phone and their interplay with different CRE methods (cf. Figure 1).

First, we recapitulate the properties of the fingerprint database and its ground truth from [KDH<sup>+</sup>15] and introduce our own reference database (Section 3.1), then we determine the identification performance and look at those images’ evidential value determined by the experts and an algorithm w.r.t. the use of different CREs and image capture devices (Section 3.2). Finally we discuss our findings and their implications (Section 3.3).

#### 3.1 Databases

Recently, Kotzerke *et al.* have introduced a pseudo fingerprint database [KDH<sup>+</sup>15]; it consists of 1,428 normal and deliberately distorted fingerprints from two males and two females. In order to create the distorted marks, they defined six different distortion categories listed in the first column of Table 1. There are 168 marks per distortion category, the other 420 marks are “normal” and don’t suffer from deliberate distortions (cf. Table 1).

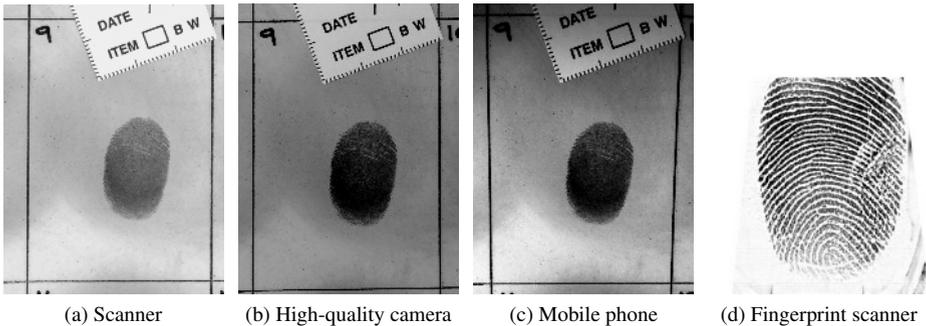


Figure 2: A subject’s right middle finger “heavily” placed on the sheet captured by different devices: scanner (a), high-quality camera (b), mobile phone (c) and fingerprint scanner (d). The first three images (a – c) have been cropped more closely before entering them into the database, the image captured with the fingerprint scanner (d) is used in the reference database and only shown here for reference purposes.

All fingerprints were left on a sheet of paper, brushed with magnetic black powder and laminated afterwards, under the supervision of a fingerprint expert. All sheets were digitised with 3 different capture devices: (i) a flatbed scanner (HP Scanjet G4010, abbr: Scanner), (ii) a high-quality camera (Nikon D3S with a Nikkor  $f/2.8$  60 mm-macro lens attached, abbr: DSLR) and (iii) a mobile phone (Apple iPhone 4S, abbr: Phone). It has to be noted that capture resolution for the mobile phone varies as it has been used in an

unconstrained setup. However, its captures were taken perpendicular to the fingermark sheets and both capturing and lighting conditions were kept as consistent as possible. The high-quality camera was attached to an operational stand setup, which is usually used for police work. The estimated capture resolutions are 1200 ppi (Scanner), 460 ppi (DSLR) and 890 ppi (Phone). More details can be found in [KDH<sup>+</sup>15].

All laminated marks have been assessed by three Victoria Police experts who decided for each mark if it is of EV by undergoing at least a partial markup process. The ground truth is the majority vote of their assessment. The EV distribution can be found in Table 1.

In this research, we created a reference database to match the marks against. For this purpose, we collected all ten fingerprints of the same subjects as found in the fingermark database with a Digital Persona U.are.U 4000 fingerprint scanner. We captured one image per print without any deliberate distortion to imitate a reference scenario (cf. Figure 3.1). Also, we added imposter images with alike characteristics (no deliberate distortion) which were all captured with optical fingerprint scanners similar to the one we employed. Specifically, we used all third prints of FVC2000 DB3 [MMC<sup>+</sup>02a] and FVC2004 DB2 [MMC<sup>+</sup>04] and all sixth prints of FVC2002 DB1 [MMC<sup>+</sup>02b]. This leads to a reference database consisting of 40 genuine and 330 imposter prints. We verified via the cross verification scores that there are no duplicates included.

## 3.2 Experiment

This experiment aims to investigate the relationship between a fingermark, which can be automatically identified with a high certainty and the evidential value assigned to it by experts or EVA (cf. Figure 1).

The EVA is predominantly influenced by (i) the image properties such as capture device used and CRE and hence (ii) the quality features extracted as briefly discussed in Section 2. The feature sets of NFIQ2, Verifinger and their fusion (concatenation of their feature vectors) have been investigated. In this context, we use EVA from [KDH<sup>+</sup>15] to obtain the estimated evidential values for three different CREs (None, Global, RLAPS) and three capture devices (Scanner, DSLR, Phone) for all 1,428 fingermarks. This process is extensively described in [KDH<sup>+</sup>15].

Next, a verification score for every fingermark matched against every print in the reference database is computed. This verification process is performed by a commercial fingerprint extractor and matcher, Neurotechnology Verifinger 7.0. We consider it to be ccID if and only if the highest score returned for one particular fingermark is from the comparison to the same subject's finger and the second highest score multiplied by  $d = 1.5$  is still smaller than the highest one (see Equation 1). The factor  $d$  has been empirically chosen in order to ensure a reliable and certain estimation due to the actual fingermark and fingerprint similarity rather than poor image quality (cf. Section 2).

This experiment is carried out on all fingermarks, which have been digitised and rescaled using different CREs such as no rescaling, a global rescaling factor for each device, or an individual estimate based on image characteristics (RLAPS).

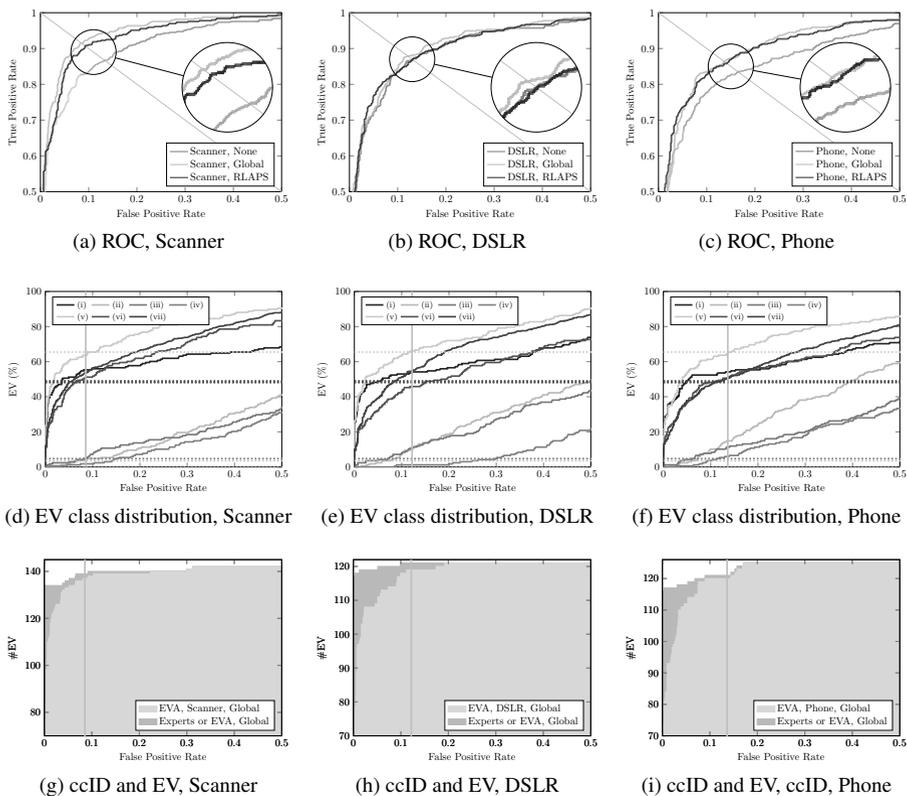


Figure 3: The first row ((a) to (c)) shows the top left corner of the receiver operating characteristics (ROCs) for all capture devices calculated on the Fusion image quality feature set with global rescaling. The colour varies according to the fraction of  $EV \& ccID/ccID$  as the classification threshold moves along the ROC; the smaller the fraction, the lighter the colour (only applicable to “Scanner, None” and “Phone, None”; the fraction equals one across the whole range in all other cases). The second row ((d) to (f)) shows the EV distribution according to the mark’s distortion class which has been computed by EVA (solid) and the experts (dashed). The latter isn’t affected by the decision threshold and hence remains constant. The third row ((g) to (i)) shows the number of ccIDs classified as EV by EVA (light colour), the additional ones by the experts (dark colour) and the ones not classified as EV by EVA but are ccID (white). The gray line is the threshold corresponding to the EER when the operating point moves along the ROC.

Additionally, we check if the ccID fingermarks are considered to be of EV by either the experts or EVA. In case of the algorithm, the decision threshold corresponding to the equal error rate (EER) has been chosen (cf. Figure 3g to 3i).

The step to determine if a fingermark is of EV is performed first and subject to capture device and CRE. Therefore, the verification scores used for identification are calculated afterwards on the already rescaled image (see Figure 1). Verifinger failed to compare 20 query fingermarks to the database because of their very high image resolution; this was

only the case for unrescaled scanner images.

Finally, we applied different decision thresholds (instead of just the one corresponding to the EER) to the evidential value raw scores. The aim is to observe if allowing more false positive errors (and hence collecting more marks in a real world scenario) would lead to a set of marks of being EV according to EVA which is a superset of the experts' decision. The results are shown in Figure 3 and Table 2.

	CRE	Capture device		
		Scanner	DSLR	Phone
ccID	None	4	118	6
	Global	145	122	126
	RLAPS	29	46	36
Experts	Global	133	117	116
EVA <sub>NFIQ2</sub>	Global	137	118	123
EVA <sub>Verifinger</sub>		134	118	119
EVA <sub>Fusion</sub>		137	119	120

Table 2: Number of fingermarks which have been correctly and with certainty identified (ccID) and the amount of those marks which have been classified by experts or EVA to be of sufficient evidential value (EV) w.r.t. capture device (Scanner, DSLR, Phone), CRE Global and quality feature set (NFIQ2, Verifinger, Fusion) if applicable. EVA uses the threshold corresponding to the EER. The EV results for the CREs None and RLAPS are not reported separately due to their much smaller numbers compared to Global (see *ccID*). Please refer to Table 1 for the total number of EV decisions or their distribution amongst the different distortion classes.

### 3.3 Discussion

The experiment shows a strong correlation between the automatically estimated evidential value and if a certain identification is possible to be performed for a particular fingermark. This is partially due to the setup used because both the matching score computation and EV estimation are based on image features.

Further limitations of the matching system used became evident and confirm the findings in [KDH<sup>+</sup>15]. Verifinger is very resolution dependent and requires marks or prints to be in a very narrow capture resolution window (around 500 ppi) with as little variation as possible to perform properly. This is the reason that a global rescaling factor and the high-quality camera images without any rescaling work well. It also explains why there are very few ccIDs when images with very high resolution without (CRE None) or with individual (CRE RLAPS) rescaling are used. Nevertheless, the image quality due to the use of different capture devices is not a major drawback. The mobile phone performs more strongly than the DSLR but falls shy of the scanner, under the condition that the capture resolution is adjusted properly. The difference between the quality feature sets is rather small but should be considered in a real world framework.

Further, we note there are prints which can be automatically identified with certainty but haven't sufficient evidential value according to the experts' assessment. This might be

again due to experimental setup that heavily favours image processing algorithms or the limited size of the test population and database. Additionally, it is worth pointing out that some of the identified fingermarks are only considered to be of evidential value by the experts or the algorithm, but not both.

Encouragingly, only *one* of the 116 marks being identified with certainty and having EV according to the experts was missed by the algorithm in a mobile phone scenario with global rescaling and the NFIQ2 feature set.

Table 1 also indicates that EVA works rather conservatively and tends to flag a fingermark as being of sufficient evidential value slightly more often than an expert who applies other considerations (such as court eligibility) than just image quality. Nevertheless, an expert's accuracy and repeatability can vary mostly due to the print quality [UHBR11, UHBR12] regarding borderline decisions and experts "tend to free the guilty rather than to convict the innocent" [TTM11].

Our results underscore the importance of capturing all fingermarks of sufficient evidential value in the field. They should have VEO if they don't have VID.

## 4 Conclusion and future work

The experiment performed indicates a strong correlation between the fact that a fingermark can be automatically identified with certainty and its inferred evidential value. Therefore, it is sensible to run fingermarks with sufficient evidential value against a reference database to potentially obtain an identification.

Also, our findings indicate that an automatic mobile phone setup is suitable to determine if a fingermark at a crime scene is of sufficient evidential value and could be operated by non-experts. In the case that the capture conditions are unknown, it is sensible to use a capture resolution estimator to improve performance.

In the future, we would like to perform more exhaustive testing on additional and considerably larger databases with different matching systems as well. Also, a fingermark determined to be of EV needs to be evaluated as either VEO or VID. Eventually we would like to test performance in the field.

## Acknowledgment

We thankfully acknowledge the Victoria Police fingerprint examiners who kindly assessed all fingermarks. The research project is funded by the Victoria Police. Also we would like to thank Dr. Arathi Arakala for her valuable input and suggestions. This work forms part of the PhD thesis of the first author.

## References

- [AFFOG<sup>+</sup>07] Fernando Alonso-Fernandez, Julian Fierrez, Javier Ortega-Garcia, Joaquin Gonzalez-Rodriguez, Hartwig Fronthaler, Klaus Kollreider, and Josef Bigun. A comparative study of fingerprint image-quality estimation methods. *Information Forensics and Security, IEEE Transactions on*, 2(4):734–743, 2007.
- [Ash99] David R Ashbaugh. *Quantitative-qualitative friction ridge analysis: an introduction to basic and advanced ridgeology*. CRC press Boca Raton, 1999.
- [CDJ05] Yi Chen, Sarat C Dass, and Anil K Jain. Fingerprint quality indices for predicting authentication performance. In *Audio-and Video-Based Biometric Person Authentication*, pages 160–170. Springer, 2005.
- [FKB06] Hartwig Fronthaler, Klaus Kollreider, and Joseph Bigun. Automatic image quality assessment with application in biometrics. In *Computer Vision and Pattern Recognition Workshop, 2006. CVPRW'06. Conference on*, pages 30–30. IEEE, 2006.
- [FSZ12] Jianjiang Feng, Yuan Shi, and Jie Zhou. Robust and Efficient Algorithms for Separating Latent Overlapped Fingerprints. *Information Forensics and Security, IEEE Transactions on*, 7(5):1498–1510, 2012.
- [KDH<sup>+</sup>15] J. Kotzerke, S. A. Davis, R. Hayes, L. J. Spreeuwens, R. N. J. Veldhuis, and K. J. Horadam. Discriminating fingermarks with evidential value for forensic comparison. In *Biometrics and Forensics (IWF), 2015 International Workshop on*, pages 1–6. IEEE, March 2015.
- [KME<sup>+</sup>14] Philip J Kellman, Jennifer L Mnookin, Gennady Erlikhman, Patrick Garrigan, Tandra Ghose, Everett Mettler, David Charlton, and Itiel E Dror. Forensic Comparison and Matching of Fingerprints: Using Quantitative Image Measures for Estimating Error Rates through Understanding and Predicting Difficulty. *PLoS one*, 9(5):e94617, 2014.
- [LCCK08] Sanghoon Lee, Heeseung Choi, Kyoungtaek Choi, and Jaihie Kim. Fingerprint-quality index using gradient components. *Information Forensics and Security, IEEE Transactions on*, 3(4):792–800, 2008.
- [Mal09] Davide Maltoni. *Handbook of fingerprint recognition*. Springer, 2009.
- [MMC<sup>+</sup>02a] Dario Maio, Davide Maltoni, Raffaele Cappelli, James L. Wayman, and Anil K. Jain. FVC2000: Fingerprint verification competition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(3):402–412, 2002.
- [MMC<sup>+</sup>02b] Dario Maio, Davide Maltoni, Raffaele Cappelli, James L Wayman, and Anil K Jain. FVC2002: Second fingerprint verification competition. In *Pattern recognition, 2002. Proceedings. 16th international conference on*, volume 3, pages 811–814. IEEE, 2002.
- [MMC<sup>+</sup>04] Dario Maio, Davide Maltoni, Raffaele Cappelli, Jim L Wayman, and Anil K Jain. FVC2004: Third fingerprint verification competition. In *Biometric Authentication*, pages 1–7. Springer, 2004.
- [Neu15] Neurotechnology. VeriFinger SDK. <http://www.neurotechnology.com/verifinger.html>, 2015.
- [SHAF11] Nathan J. Short, M.S. Hsiao, A.L. Abbott, and E.A. Fox. Latent fingerprint segmentation using ridge template correlation. In *Imaging for Crime Detection and Prevention 2011 (ICDP 2011), 4th International Conference on*, pages 1–6, 2011.

- [The13] The National Institute of Standards and Technology. NFIQ2 Feature Definitions Document (v0.5). [http://biometrics.nist.gov/cs\\_links/quality/NFIQ\\_2/NFIQ-2\\_Quality\\_Feature\\_Defin-Ver05.pdf](http://biometrics.nist.gov/cs_links/quality/NFIQ_2/NFIQ-2_Quality_Feature_Defin-Ver05.pdf), April 2013.
- [TTM11] Jason M Tangen, Matthew B Thompson, and Duncan J McCarthy. Identifying fingerprint expertise. *Psychological science*, 22(8):995–997, 2011.
- [UHBR11] Bradford T Ulery, R Austin Hicklin, JoAnn Buscaglia, and Maria Antonia Roberts. Accuracy and reliability of forensic latent fingerprint decisions. *Proceedings of the National Academy of Sciences*, 108(19):7733–7738, 2011.
- [UHBR12] Bradford T Ulery, R Austin Hicklin, JoAnn Buscaglia, and Maria Antonia Roberts. Repeatability and reproducibility of decisions by latent fingerprint examiners. *PloS one*, 7(3):e32800, 2012.
- [Vic13] Victoria Police. Crime Statistics 2012/2013. [http://www.police.vic.gov.au/retrievemedia.asp?Media\\_ID=72176](http://www.police.vic.gov.au/retrievemedia.asp?Media_ID=72176), August 2013.
- [YJ13] Soweon Yoon and Anil Jain. Quality assessment of latent fingerprints. [http://biometrics.nist.gov/cs\\_links/quality/NFIQ\\_2/presentations\\_4-26/nfiq2\\_yoon-2013-04-26.pdf](http://biometrics.nist.gov/cs_links/quality/NFIQ_2/presentations_4-26/nfiq2_yoon-2013-04-26.pdf), April 2013.



# Privacy Preserving Technique for Set-Based Biometric Authentication using Reed-Solomon Decoding

Jesse Hartloff\*<sup>1</sup>, Avradip Mandal<sup>2</sup>, and Arnab Roy<sup>2</sup>

<sup>1</sup>SUNY at Buffalo, Buffalo, NY, USA  
hartloff@buffalo.edu

<sup>2</sup>Fujitsu Laboratories of America, Sunnyvale, CA, USA  
{amandal, aroy}@us.fujitsu.com

**Abstract:** In this work, we present a single-factor biometric authentication system that provides template security against an adversarial server while allowing error-tolerant matching. Our approach is to secure templates represented as sets using error-correcting codes and Reed-Solomon decoding. To accomplish this, each element in the set is combined with a random codeword and a secret share is computed using the codeword and a Reed-Solomon based secret sharing scheme. These random codewords provide uncertainty for an attacker, while the genuine user can decode to the correct values for verification. Without a reading from the enrolling biometric the shares will appear random, thus protecting the users biometric. We show implementation results for this system on fingerprints using pairs of minutia points. Our system overcomes many common weaknesses for template security systems including replay attacks, malicious servers, eavesdroppers, and record multiplicity attacks.

## 1 Introduction

The appeal of using biometrics has led to an increase in their use as a means of identification. With biometric-based authentication, users are not required to remember extraneous passwords or carry tokens such as smartcards. All users effortlessly bring their biometrics with them wherever they go making it an ideal candidate for user-friendly authentication.

However, this increase in use leads to privacy concerns when sharing biometric information with various service providers since it can be difficult to tell if they are trustworthy. The problem becomes severe when using the same biometric to enroll in several different systems. If one of them is not using proper privacy protocols, it can allow your biometric to be revealed and used to access other systems.

To achieve security against malicious servers, we construct a client-based system which is an instantiation of a secure sketch [DORS08] based on Reed-Solomon decoding for error-tolerant secure matching and an authentication protocol that prevents replay attacks. We use the secure sketch as part of a fuzzy extractor [DORS08] to bind a secret to the enrolling biometric reading creating a secure template which is sent to the server. We

---

\*Work done while Jesse Hartloff visited Fujitsu Laboratories of America.

note that our primitives do not satisfy the stringent cryptographic requirements outlined in [Boy04]. In Section 4, we independently argue why our scheme still should be secure based on reasonable assumptions.

By having a client-centered system, we gain some valuable security properties. Since the only biometric related information that leaves the client is in the form of a secure template, there is no need to trust the server or to secure the communication channels for enrollment or verification. Also, since the client controls the generation of the template they can alter the protocol if they wish. Changing the system to utilize different modalities or adding a user-specific key can be done by the client without the server even being aware of the clients protocol, thus adding security and flexibility to the system. We prevent replay attacks by utilizing a secure signature of random values for each authentication.

We implemented this system on fingerprints using the publicly available FVC2002-DB1 [MMC<sup>+</sup>02] dataset and report the results in Section 5. To construct a template from a fingerprint, we extract a set of pairs of minutia points and quantize each pair. This provides a set of features that is used to construct the secure templates. These features are combined in a secret sharing scheme such that an attacker cannot attack individual templates points but must correctly match, or guess, a subset of features with size depending on the degree of the underlying polynomial. Since all matching occurs at the client where the fingerprint is read, we have access to the full fingerprint image of the test reading and utilize this during verification.

## 2 Related Work

There have been various systems proposed that provide template security for fingerprints. Possibly the most popular of them is the fuzzy vault construct of Juels and Sudan [JS06]. Similar to our current system, the fuzzy vault binds a secret to a set of values and releases the secret using Reed-Solomon decoding given a set that is sufficiently similar to the one used during enrollment. The security of the fuzzy vault relies on adding many randomly generated chaff points to obfuscate the enrolling data. The fuzzy vault has been implemented in various forms to secure fingerprint templates [BCF12, JA07, NNJ08], however the fuzzy vault is vulnerable to various attacks including record multiplicity, chaff injection, and replay attacks [KY08, MNS<sup>+</sup>10, MMT09, PM09, SB07] and can have very large template sizes due to the addition of enough chaff points for sufficient security. Our current system overcomes all of these shortcomings.

Many protocols combine cryptographic techniques with biometrics to form a secure biometric cryptosystem [BBCdS08, BCI<sup>+</sup>07, Sto10]. However, these schemes have been shown to have vulnerabilities especially when a malicious server is considered instead of the common honest but curious server [SBCS12].

Our work develops a new method that falls under the category of secure sketch [DORS08] which we use as part of a fuzzy extractor [DORS08] to bind a secret to the enrolling fingerprint. A general theoretical framework for an authentication scheme using secure sketch is given in [BDK<sup>+</sup>05]. Our scheme can be considered as a concrete instance of a fingerprint matching scheme that generalizes to the abstract scheme from [BDK<sup>+</sup>05] which utilizes

a secure sketch and an authentication scheme, both of which are treated as black boxes. The scheme also utilizes client-based computation to further protect the biometric data. Another instance of this scheme has been implemented for face biometrics [SLM07].

There are many other proposed systems for fingerprint template security including [BSW07] and [FMC12], both of which report more accurate matching performance than ours. We note that in [BSW07], the system stores partial template information in the clear to compute a robust distance measure that improves matching accuracy. A security analysis shows that the remaining entropy in the template is still sufficient for security, though an attacker is given some information of the enrolling fingerprint. Our proposed system does not reveal any biometric information. The security of [FMC12] is based on two attacks implemented by the authors that attempt to recover information of the enrolling fingerprint from the stored template. Results are shown with parameters for which these attacks have a success rate of 0.

We show that in the proposed system, the enrolling fingerprint information is secure against any attack, including an adversarial server, under the assumption of uniformity of the template points. Addressing this assumption is a source of continuous research.

### 3 System Protocol

Our system consists of two phases - an Enrollment phase which is one-time for a single user and server pair and an Authentication phase which can be executed for every session of a user with the server. There are three parties involved in the Enrollment phase - the user, the server and a Trusted Third Party, while in the Authentication phase only the user and the server are involved. We describe the phases in detail below. The only role of the Trusted Third Party is to verify that the server stores the correct template in tact, thus preventing a man-in-the-middle attack.

#### 3.1 Enrollment Phase

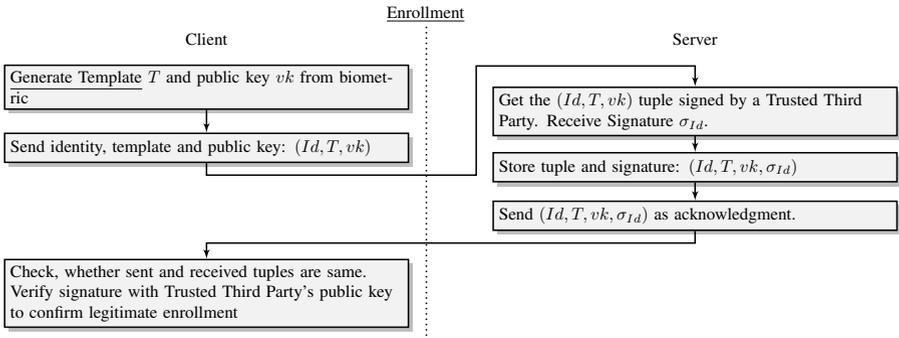


Figure 1: Enrollment process

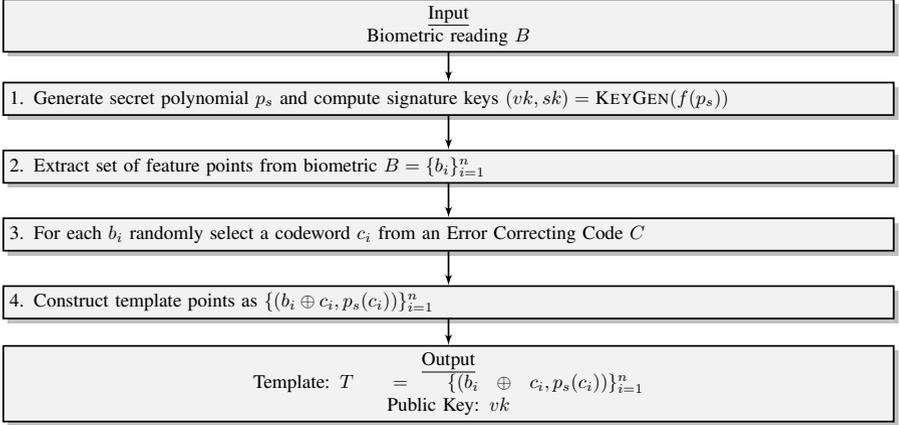


Figure 2: Generate template

Enrollment consists of a user generating a random secret and binding it to her biometric to create a secure template. The secret is used as a seed to generate a public verification key using a signature scheme and will be used to generate the corresponding private signature key during verification. The template and public key are then sent to server for storage.

The template and public key together with the user's identity are then digitally signed by a Trusted Third Party. This signature is sent back to the client for verification. See Figure 1.

**Generating Template.** As part of the enrollment process, the user will generate a secure template using a random secret and a biometric reading by following the steps in Figure 2 which are described below. We assume that the reading is a set of bit-vectors  $b_i$  of identical length, where  $i$  runs from 1 to  $n$ , which is true in our implementation in Section 5. To tolerate errors in reading each bit vector, we use an Error Correcting Code (ECC)  $C$ , regarded as a set of codewords. The length of codewords is chosen to be the same as an individual bit-vector  $b_i$ .

1. User generates a random secret that will be used as a seed to generate a key pair of a secure signature scheme. For use in our system, this secret is encoded as a polynomial  $p_s$ . We use a cryptographic hash function (e.g. SHA-3)  $f$  and a secure signature scheme (e.g. PKCS #1) (KEYGEN, SIGN, VERIFY) in our system. The key pairs are generated as follows, by using  $f(p_s)$  as source of randomness for KEYGEN.

$$(vk, sk) = \text{KEYGEN}(f(p_s))$$

2. Extract a set ( $B$ ) of feature points from a biometric reading, which is a set of bit-vectors  $b_i$  of identical length, where  $i$  runs from 1 to  $n$ . Please see Section 5 for a concrete example.
3. Choose a random codeword  $c_i$  from a code  $C$  for each  $b_i \in B$ . The  $c_i$  values

will be used to hide the template data while allowing some error-correction during verification.

4. Compute  $y_i = b_i \oplus c_i$  and  $\gamma_i = p_s(c_i)$  for each  $i \in [n]$  and store as the secure template  $T = \{(y_1, \gamma_1), \dots, (y_n, \gamma_n)\}$ . The polynomial evaluations will be used as input in a Reed-Solomon decoder to correct for errors.

### 3.2 Authentication Phase

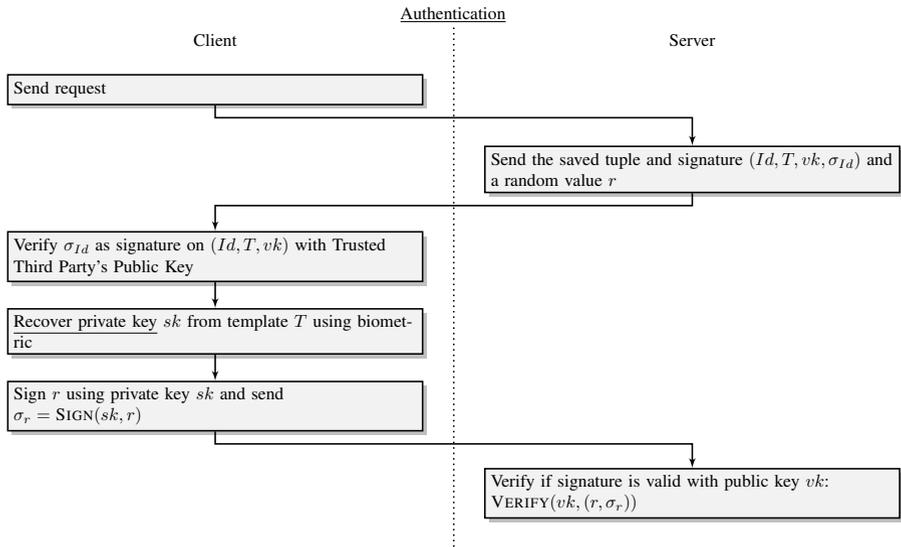


Figure 3: Authentication

To authenticate, the server sends the signed tuple back to the user along with a random value  $r$ . If the user is legitimate, she will be able to recover the secret from the template using her biometric and generate the private signature key to be used to sign  $r$ . The signature on  $r$  is sent back to the server where it is verified using the public verification key. See Figure 3.

**Key Recovery.** During verification, the client must recover the secret signature key from the secure template and use it to sign the random value  $r$  sent by the server. Since the key is recovered at each verification, the user is not required to remember it making this a single-factor system. This signature is sent back to the server to complete the verification process. Below are descriptions of the steps outlined in Figure 4.

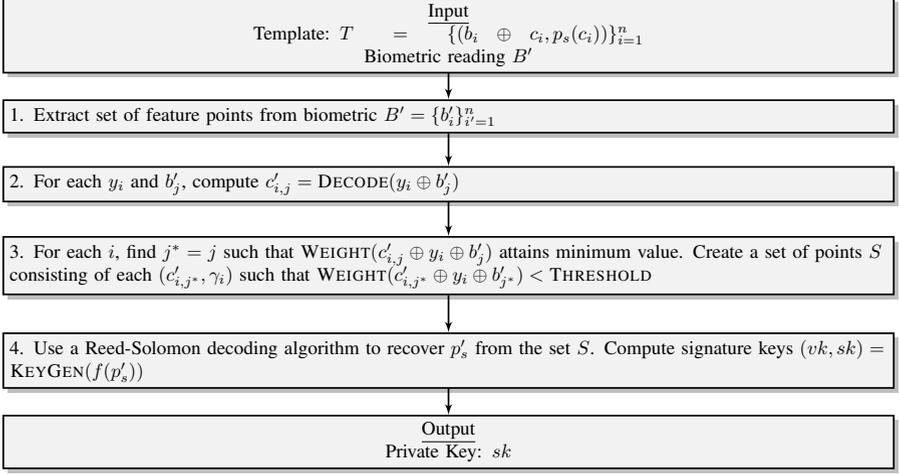


Figure 4: Recover private key

1. Extract set of feature points  $\{b'_j\}_{j=1}^n$  from biometric reading  $B'$  using the same method as for enrollment.
2. Client computes  $c'_{i,j} = \text{DECODE}(y_i \oplus b'_j)$  for each  $(y_i, b'_j)$  pair. Here DECODE outputs the nearest codeword from  $(y_i \oplus b'_j)$ .
3. For each  $i$ , client chooses  $j^* = j$ , such that  $\text{WEIGHT}(c'_{i,j} \oplus y_i \oplus b'_j)$  attains minimum value for  $j$  between 1 to  $n$ . Create a set of points  $S$  consisting of each  $(c'_{i,j^*}, \gamma_i)$  such that  $\text{WEIGHT}(c'_{i,j^*} \oplus y_i \oplus b'_{j^*}) < \text{THRESHOLD}$ .
4. Use Reed-Solomon decoding on  $S$  to recover  $p'$ . By the properties of the Reed-Solomon decoder, if the number of genuine points in the  $S$  minus the number of false points in  $S$  is greater than the degree of  $p_s$ , then  $p' = p_s$ .

## 4 Security

### 4.1 Template privacy against Brute Force Attacker

We first show that the proposed system is secure against a basic brute force attack before proceeding to formally show its security. A brute force attacker against our protocol from the previous section has access to the template  $T = \{(b_i \oplus c_i, p_s(c_i))\}_{i=1}^n$  and public key  $vk$ , where  $(vk, sk) = \text{KEYGEN}(f(p_s))$ . The goal of the attacker is to recover the biometric template  $B = \{b_i\}_{i=1}^n$  (or another biometric template close to  $B$ ). This problem is equivalent to guessing the random codewords  $\{c_i\}_{i=1}^n$  and testing the correctness of guess from  $\{p_s(c_i)\}_{i=1}^n$  and  $vk$ . If an attacker can correctly guess the random polynomial  $p_s$ , it can easily find the codewords from  $p_s(c_i)$  values. If the polynomial  $p_s$  is of degree  $t$ , then

to find out the polynomial  $p_s$ , the brute force attacker has to make correct guesses for  $c_i$  values simultaneously for at least  $(t + 1)$  points. These random guesses can be interpolated to a possible guess for the polynomial as  $p'_s$ . The only way the attacker can check whether the random guesses for  $c_i$  values (equivalently, random guess for the polynomial  $p_s$ ) are correct or not, is by running the KEYGEN algorithm on  $f(p'_s)$  and checking whether the resultant verification key is same as the published verification key  $vk$  or not. If we sample the codewords from a  $(n, k, d)$  error correcting code, then each codeword  $c_i$  has  $k$  bits of entropy and the attacker has to make simultaneous guesses to at least  $(t + 1)$  codewords. Hence our protocol has  $k(t + 1)$  bits of security against brute force attackers.

## 4.2 Formal Security Guarantees

The protocol described in Section 3 provides the following security guarantees:

- **Type-I Security:** It is a secure authentication protocol, i.e., only legitimate users can get authenticated to the server.
- **Type-II Security:** It protects user's biometric data against malicious servers. If multiple servers are authenticating the users using our protocol and one of them is acting maliciously, even then the malicious server cannot authenticate to another server on behalf of any common user (a person who has enrolled to both servers).

To show that our scheme is secure in both the cases, we consider a powerful adversary  $\mathcal{A}$ , which

1. Has access to the template

$$T = \{(b_i \oplus c_i, p_s(c_i))\}_{i=1}^n$$

2. Has access to the public verification key  $vk$ , such that  $(vk, sk) = \text{KEYGEN}(f(p_s))$
3. Can send any  $r$  of its choice<sup>1</sup> to the client and receive  $\sigma_r$ , which is a valid signature of  $r$  (gets verified by the verification key  $vk$ ).

The goal of the above adversary is to come up with a valid message, signature pair  $(\hat{r}, \hat{\sigma}_r)$  which would get verified by the verification key  $vk$ , without querying  $\hat{r}$  to the client<sup>2</sup>. Such an adversary mimics a dishonest server trying to authenticate to another server (Type-II attacker), as well as a powerful man in the middle attacker (Type-I attacker).

If the signature scheme is chosen message secure, then security of our protocol can be based upon the following assumption, which we later argue to be reasonable.

---

<sup>1</sup>In the actual protocol the server also sends the enrolled template  $T$  along with its signature signed by a trusted third party. Client verifies integrity of the template by verifying the signature with trusted third parties public key, before reconstructing its private key based on the template. This forces the attacker to send the same  $T$  to evoke a response.

<sup>2</sup>This actually provides a stronger security guarantee. In the actual protocol execution, the attacker has to come up with a valid signature of some  $\hat{r}$ , chosen by the honest server, not an  $\hat{r}$  chosen by the attacker herself.

**Assumption 1.** If  $f$  is a cryptographic hash function,  $B = \{b_1, \dots, b_n\}$  and  $B' = \{b'_1, \dots, b'_n\}$  are two sets of feature points corresponding to fingerprints of two different individuals ( $Id$  and  $Id'$ ),  $(c_1, \dots, c_n)$  and  $(c'_1, \dots, c'_n)$  are random code-words, and  $p_s, p'_s$  are two randomly selected polynomials, then the following two tuples are indistinguishable to a computationally bounded adversary:

$$(Id, f(p_s), \{(b_i \oplus c_i, p_s(c_i))\}_{i=1}^n) \\ \approx (Id, f(p_s), \{(b'_i \oplus c'_i, p'_s(c'_i))\}_{i=1}^n)$$

The above assumption says that biometric templates corresponding to two different individuals are indistinguishable, as well as it is infeasible to correlate the output of the hash function  $f(p_s)$  and the biometric template  $T$ . We now prove that, under this assumption, the only way the adversary  $\mathcal{A}$  can be successful is to break the security of the signature scheme.

**Theorem 1.** If  $(\text{KEYGEN}, \text{SIGN}, \text{VERIFY})$  is a signature scheme which is existentially unforgeable under chosen message attack (EU-CMA), and Assumption 1 holds true, then adversary  $\mathcal{A}$  can win only with negligible probability.

*Proof.* Using adversary  $\mathcal{A}$ , we construct an EU-CMA adversary  $\mathcal{B}$  against the signature scheme. The EU-CMA challenger will generate verification key  $vk$  and signing key  $sk$  by running KEYGEN. Adversary  $\mathcal{B}$  will receive the verification key  $vk$  from the EU-CMA challenger. The EU-CMA challenger will also provide access to the signing oracle  $\text{SIGN}(sk, \cdot)$  to the adversary  $\mathcal{B}$ . Adversary  $\mathcal{B}$  works as follows:

1. Sample fingerprint  $B' = \{b'_1, \dots, b'_n\}$  from a random individual. Sample random codewords  $\{c'_1, \dots, c'_n\}$  and random polynomial  $p'_s$ . Send  $vk$  and  $T' = \{(b'_i \oplus c'_i, p'_s(c'_i))\}_{i=1}^n$  to the adversary  $\mathcal{A}$ . Assumption 1 says, adversary  $\mathcal{A}$  would not be able to distinguish the simulated  $(vk, T')$  from (public key of  $\text{KEYGEN}(f(p_s)), \{(b_i \oplus c_i, p_s(c_i))\}_{i=1}^n$ ) received in the real protocol.
2. For every signature query  $r$  sent by adversary  $\mathcal{A}$ ,  $\mathcal{B}$  can forward the query to the EU-CMA challenger.
3. In the end, a successful  $\mathcal{A}$  would provide a valid message signature tuple  $(\hat{r}, \hat{\sigma}_r)$  which was not obtained as the response to a query.  $\mathcal{B}$  can send the same tuple to the EU-CMA challenger and provide a valid forgery.

□

**Justification of Assumption 1.** Assumption 1 plays a key role in our security proof. Assuming the hash function  $f$  behaves as a random oracle, Theorem 2 stated below reduces Assumption 1 to the following simpler one (independent of the hash function  $f$ ).

**Assumption 2.** If  $B = \{b_1, \dots, b_n\}$  and  $B' = \{b'_1, \dots, b'_n\}$  are two sets of feature points corresponding to fingerprints of two different individuals ( $Id$  and  $Id'$ ),  $(c_1, \dots, c_n)$  and  $(c'_1, \dots, c'_n)$  are random code-words, and  $p_s, p'_s$  are two randomly selected polynomials, then

- *The following two tuples are indistinguishable:*

$$(Id, \{(b_i \oplus c_i, p_s(c_i))\}_{i=1}^n) \approx (Id, \{(b'_i \oplus c'_i, p'_s(c'_i))\}_{i=1}^n)$$

- *Given  $\{(b_i \oplus c_i, p_s(c_i))\}_{i=1}^n$ , it is hard to output  $p_s$*

Assumption 2 consists of two parts, of which the first one says that if we XOR biometric feature bit-vectors with random codewords, the xored bit-vectors corresponding to two different individuals become indistinguishable. Moreover, indistinguishability continues to hold when we additionally provide random polynomial evaluations corresponding to those random codewords. As a quick sanity check, if we assume the  $b_i$ 's are coming from a uniform distribution and  $p_s$  is a linear polynomial our assumption holds provably. We claim that even if  $p_s$  is a higher degree polynomial and the  $b_i$ 's are coming from an actual fingerprint distribution, our assumption still holds. The second part of the assumption says xoring the random codewords with biometric minutia points, hides the codewords to sufficient degree that it is impossible to recover  $p_s$  given the evaluations  $\{p_s(c_i)\}_{i=1}^n$ .

**Theorem 2.** *If  $f$  is a random oracle, then Assumption 2 implies Assumption 1.*

*Proof.* We show that if there is an Assumption 1 adversary  $\mathcal{A}_1$ , which succeeds with non negligible advantage then we can construct an Assumption 2 adversary  $\mathcal{A}_2$  succeeding with non negligible advantage. From the Assumption 1 challenger,  $\mathcal{A}_2$  receives the identity of an individual  $Id$  and a biometric template  $T$  (using a coin flip,  $T$  was generated by the biometric corresponding to either the individual  $Id$  or from another different individual).  $\mathcal{A}_2$  wins if

1. It can correctly guess whether  $T$  was generated using the biometric corresponding to individual  $Id$ , or
2. It can output the polynomial  $p_s$  used during the generation of  $T$

$\mathcal{A}_2$  works as follows:

1. Sample random  $r$  from the range of  $f$ . Send  $(Id, r, T)$  to  $\mathcal{A}_1$
2. For each different random oracle query  $f(p_i)$  made by  $\mathcal{A}_1$ , answer by sampling a random  $r_i$  from the range of  $f$ . Save  $(p_i, r_i)$  in a table.
3. In the end  $\mathcal{A}_1$  will return its guess, whether  $T$  was generated using  $Id$ 's biometric or not. Send the same guess to Assumption 2 challenger along with a random  $p$  out of  $\{p_i\}$  (random oracle queries made by  $\mathcal{A}_1$ ) as a guess for  $p_s$ .

$f$  being a random oracle,  $(Id, r, T)$  is a valid Assumption 1 challenge, as long as the  $p_s$  used in generation of  $T$  does not belong to the set  $\{p_i\}$ . In that case, whenever  $\mathcal{A}_1$  makes a successful guess,  $\mathcal{A}_2$  also makes a successful guess and wins against the Assumption 2 challenger. In the other case,  $\mathcal{A}_1$  being an efficient adversary can only make polynomially many  $f(p_i)$  queries and one of those  $p_i$ 's is actually  $p_s$ . Hence,  $\mathcal{A}_2$  can successfully guess  $p_s$  with  $1/\text{poly}()$  probability, which is non-negligible.  $\square$

## 5 Experimental Results

As a proof of concept of the feasibility of our system, we implemented a simple fingerprint matching method using pairs of minutia points. These template values consist of the concatenation of the distance between the points, the difference in their orientations, the angle of the line defined by the points, the angle between the orientation and the connecting line, the number of ridges between the points, and the type of each minutia in the pair. Each value is then quantized and the gray code is applied to the quantized values. This encoding has the property that any two consecutive integers will have hamming distance 1. This allows us to treat the integers as bit strings and enables the use of hamming distance to compare similar minutia pairs. After quantization, these template values are 22 bits in length. Since the angle of the line connecting the points is rotation variant, we consider several different rotations of the test fingerprint during verification. We only compute the enrolling template at a single rotation to limit the amount of information and correlation in the template.

We use a (22, 6, 4) randomly generated code to protect the template points for this implementation. This code has 64 codewords which provides 6-bits of entropy that an attacker would have to guess in order to recover the template point. Since there is no feedback for an attack on a single point, an attacker would have to simultaneously correctly guess enough points to recover the secret polynomial. At the ZeroFAR, this polynomial has degree 16 meaning an attacker must simultaneously guess at least 17 points each with 6 bits of entropy resulting in  $6 * 17 = 102$  bits of security against a brute force attack as described in Section 4.1. To decode this polynomial we use the Welch-Berlekamp decoder for the Reed-Solomon code.

The choice of code provides a tradeoff between security and matching accuracy. For this implementation, we only utilize 1 bit of error correction, meaning we could use a more efficient code with more than 64 codewords to increase entropy while still being able to correct from 1 bit errors. However, increasing the number of codewords also increases the number of false matches on the template points since it is more likely that a random value will be within 1 bit of a codeword. Thus, the tradeoff between accuracy and security can be adjusted by altering the size of the code.

To match a template with a fingerprint reading, we first extract a set of pairs from the enrolled template using the method from Section 3.2. To increase the accuracy of genuine matches, we filter out some of the extracted template points by only considering sets of points that form a complete sub-graph of at least 4 minutia points. This reduces the chance that a false match will be considered.

We use the FVC (Fingerprint Verification Competition) style of measuring results with 2800 genuine and 4950 impostor tests. Results are reported for FVC2002-DB1. Minutiae points were obtained using the open-source minutiae extraction method MINDTCT [WGT<sup>+</sup>] published by NIST.

A summary of our matching results can be found in table 1. Since security is a focus of our system, we are concerned with large polynomials that lead to no false accepts.

Degree of Secret Polynomial	FAR (%)	FRR (%)
14	0.08	19.0
15	0.04	20.3
16	0.0	21.4

Table 1: Matching results for FVC2002-DB1. We note that the system should not be used when the FAR is not 0.0 as this compromises security. We include additional values to give more information on the matching performance of the system.

## 6 Conclusion

In this work, we presented a fingerprint matching system that provides template security against an adversarial server by utilizing the entropy of random codewords in conjunction with polynomial based secret sharing. We accomplish this in part by shifting the matching responsibility to the client instead of the server. We also include a signature scheme for verification that prevents replay attacks from potential eavesdroppers. In addition to presenting this novel theoretical system, we provide the results of an implementation based on pairs of minutiae points to show the feasibility of this system in practice.

## References

- [BBCdS08] Manuel Barbosa, Thierry Brouard, Stéphane Cauchie, and SimoMelo de Sousa. Secure Biometric Authentication with Improved Accuracy. In Yi Mu, Willy Susilo, and Jennifer Seberry, editors, *Information Security and Privacy*, volume 5107 of *Lecture Notes in Computer Science*, pages 21–36. Springer Berlin Heidelberg, 2008.
- [BCF12] Julien Bringer, Herv Chabanne, and Mlanie Favre. Fuzzy Vault for Multiple Users. In Aikaterini Mitrokotsa and Serge Vaudenay, editors, *Progress in Cryptology - AFRICACRYPT 2012*, volume 7374 of *Lecture Notes in Computer Science*, pages 67–81. Springer Berlin Heidelberg, 2012.
- [BCI<sup>+</sup>07] Julien Bringer, Herv Chabanne, Malika Izabachne, David Pointcheval, Qiang Tang, and Sbastien Zimmer. An Application of the Goldwasser-Micali Cryptosystem to Biometric Authentication. In Josef Pieprzyk, Hossein Ghodosi, and Ed Dawson, editors, *Information Security and Privacy*, volume 4586 of *Lecture Notes in Computer Science*, pages 96–106. Springer Berlin Heidelberg, 2007.
- [BDK<sup>+</sup>05] Xavier Boyen, Yevgeniy Dodis, Jonathan Katz, Rafail Ostrovsky, and Adam Smith. Secure Remote Authentication Using Biometric Data. In Ronald Cramer, editor, *Advances in Cryptology EUROCRYPT 2005*, volume 3494 of *Lecture Notes in Computer Science*, pages 147–163. Springer Berlin Heidelberg, 2005.
- [Boy04] Xavier Boyen. Reusable cryptographic fuzzy extractors. In Vijayalakshmi Atluri, Birgit Pfitzmann, and Patrick Drew McDaniel, editors, *Proceedings of the 11th ACM Conference on Computer and Communications Security, CCS 2004, Washington, DC, USA, October 25-29, 2004*, pages 82–91. ACM, 2004.

- [BSW07] Terrance E. Boulton, Walter J. Scheirer, and Robert Woodworth. Revocable Fingerprint Biotokens: Accuracy and Security Analysis. In *CVPR*, 2007.
- [DORS08] Yevgeniy Dodis, Rafail Ostrovsky, Leonid Reyzin, and Adam Smith. Fuzzy Extractors: How to Generate Strong Keys from Biometrics and Other Noisy Data. *SIAM J. Comput.*, 38(1):97–139, 2008.
- [FMC12] M. Ferrara, D. Maltoni, and R. Cappelli. Noninvertible Minutia Cylinder-Code Representation. *Information Forensics and Security, IEEE Transactions on*, 7(6):1727–1737, Dec 2012.
- [JA07] Jason Jeffers and Arathi Arakala. Fingerprint Alignment for A Minutiae-Based Fuzzy Vault. In Arathi Arakala, editor, *Biometrics Symposium, 2007*, pages 1–6, 2007.
- [JS06] Ari Juels and Madhu Sudan. A Fuzzy Vault Scheme. *Des. Codes Cryptography*, 38(2):237–257, 2006.
- [KY08] Alisher Kholmatov and Berrin Yanikoglu. Realization of correlation attack against the fuzzy vault scheme. In *Security, Forensics, Steganography, and Watermarking of Multimedia Contents X*, volume SPIE 6819, pages 681900–681900–7, 2008.
- [MMC<sup>+</sup>02] D. Maio, D. Maltoni, R. Cappelli, J.L. Wayman, and A.K. Jain. FVC2002: Second Fingerprint Verification Competition. In *Pattern Recognition, 2002. Proceedings. 16th International Conference on*, volume 3, pages 811–814 vol.3, 2002.
- [MMT09] Preda Mihailescu, Axel Munk, and Benjamin Tams. The Fuzzy Vault for Fingerprints is Vulnerable to Brute Force Attack. In Arslan Brömme, Christoph Busch, and Detlef Hühnlein, editors, *BIOSIG*, volume 155 of *LNI*, pages 43–54. GI, 2009.
- [MNS<sup>+</sup>10] Johannes Merkle, Matthias Niesing, Michael Schwaiger, Heinrich Ihmor, and Ulrike Korte. Security Capacity of the Fuzzy Fingerprint Vault. *International Journal on Advances in Security*, 3(3 & 4):146–168, 2010.
- [NNJ08] Abhishek Nagar, Karthik Nandakumar, and Anil K. Jain. Securing fingerprint template: Fuzzy vault with minutiae descriptors. In *ICPR*, pages 1–4, 2008.
- [PM09] Hoi Ting Poon and Ali Miri. A Collusion Attack on the Fuzzy Vault Scheme. *ISeCure, The ISC International Journal of Information Security*, 1(1):27–34, 2009.
- [SB07] W.J. Scheirer and T.E. Boulton. Cracking Fuzzy Vaults and Biometric Encryption. In *Biometrics Symposium, 2007*, pages 1–6, sept. 2007.
- [SBCS12] K. Simoons, J. Bringer, H. Chabanne, and S. Seys. A Framework for Analyzing Template Security and Privacy in Biometric Authentication Systems. *Information Forensics and Security, IEEE Transactions on*, 7(2):833–841, April 2012.
- [SLM07] Y. Sutcu, Qiming Li, and N. Memon. Protecting Biometric Templates With Sketch: Theory and Practice. *Information Forensics and Security, IEEE Transactions on*, 2(3):503–512, Sept 2007.
- [Sto10] A. Stoianov. Cryptographically secure biometrics. volume 7667, pages 76670C–76670C–12, 2010.
- [WGT<sup>+</sup>] Craig I. Watson, Michael D. Garris, Elham Tabassi, Charles L. Wilson, R. Michael McCabe, Stanley Janet, and Kenneth Ko. User’s Guide to NIST Biometric Image Software (NBIS).

# Improved Fuzzy Vault Scheme for Alignment-Free Fingerprint Features

Benjamin Tams<sup>1</sup>, Johannes Merkle<sup>2</sup>, Christian Rathgeb<sup>3</sup>, Johannes Wagner<sup>3</sup>,  
Ulrike Korte<sup>4</sup>, and Christoph Busch<sup>3</sup>

<sup>1</sup>Institute for Mathematical Stochastics, University of Göttingen, Germany,  
btams@math.uni-goettingen.de

<sup>2</sup>secunet Security Networks AG, Essen, Germany, johannes.merkle@secunet.com

<sup>3</sup>da/sec - Biometrics and Internet Security Research Group, Hochschule Darmstadt,  
Germany, {christian.rathgeb,johannes.wagner,christoph.busch}@cased.de

<sup>4</sup>Federal Office for Information Security, Bonn, Germany, ulrike.korte@bsi.bund.de

**Abstract:** The *fuzzy vault scheme* is one of the most prominent tools for protecting fingerprint templates, typically being minutiae-based. However, there exist two major problems. Firstly, the fuzzy vault scheme is vulnerable to attacks correlating different templates of the same user. Secondly, auxiliary alignment data may leak information about the protected fingerprints which negatively affects security and privacy. In this paper, we tackle both problems. Our implementation uses alignment-free fingerprint features and fusions thereof, thereby removing the need to store alignment parameters. Furthermore, the features are passed through a quantization scheme and then dispersed in a maximal number of chaff, thereby thwarting correlation attacks.

## 1 Introduction

The *fuzzy vault scheme* [JS06] is a biometric cryptosystem considered eligible for protecting fingerprint features, where the features, typically based on *minutiae*, are hidden within a large number of randomly generated *chaff minutiae*. The eligibility of minutiae templates for being protected with the fuzzy vault scheme has been analyzed by Clancy *et al.* [CKL03], and further explored in a series of minutiae-based fuzzy vault implementations [NJP07, NNJ10, MIK<sup>+</sup>11].

There exist, however, two major problems with previous implementations. Firstly, the fuzzy vault is generally vulnerable to correlation attacks, which exploits that in two matching vault records genuine minutiae correlate well as opposed to chaff minutiae [SB07]. This property very clearly violates the *unlinkability requirement* [ISO11]. Even worse, the correlation attack allows to efficiently recover the protected feature data. This vulnerability can be avoided by rounding the minutia data to a rigid grid and use these quantizations to encode genuine vault features; chaff minutiae are encoded by all remaining unoccupied grid points. In this way, the templates contain the same set of points, precisely, the all grid points, which makes correlation attacks impossible.

Secondly, in order to successfully verify a query fingerprint, its minutiae are required to be sufficiently close to the genuine minutiae in the fuzzy vault. This may require a preliminary *alignment step* in which an accurate spatial translation and rotation of the query minutiae are achieved. Many implementations outsource the alignment problem by adjusting the query to auxiliary alignment data published along with the vaults [NJP07, NNJ10]. However, such auxiliary alignment data leak information about the protected fingerprints which conflicts with the irreversibility requirement of effective biometric information protection [ISO11, BBGK08] and may even facilitate correlation attacks.<sup>1</sup>

In this paper, we solve both problems. First, we use alignment-free features, which eliminates the need to store auxiliary alignment data. Secondly, we prevent correlation attacks by applying a quantization scheme to the fingerprint features and filling up the whole feature space with chaff points. Another advantage of using quantized features is the possibility of using the *improved fuzzy vault scheme* by Dodis *et al.* [DRS04] which generates significantly smaller records and can also be secured against correlation attacks [MT13].

As alignment-free features, we use *absolutely pre-aligned minutiae*, *i.e.*, minutiae represented w.r.t. a coordinate system that can be robustly estimated from the fingerprint, as well as three different local minutiae descriptors: *minutia orientation descriptors* [TK03], the *minutia frequency descriptors* [Fen08], and *local minutia structures* [JY00]. All of these minutiae descriptors have already been deployed in fuzzy vault schemes [LYC<sup>+</sup>10, NNJ10]. However, while these schemes solve the alignment problem, they do not use all unoccupied feature points as chaff and are, hence, inherently vulnerable to correlation attacks. A fuzzy vault implementation using another type of alignment-free minutiae descriptors that is also immune against correlation attacks has been presented in [BFPdS14].

The paper is outlined as follows. In Sect. 2 the alignment-free feature types tested in this paper are described. In Sect. 3 we describe the framework of our fuzzy vault. In Sect. 4 the experimental setup is described and results are reported. Finally, in Sect. 5 final discussion are given, conclusions are drawn, and outlook for future research is motivated.

## 2 Analyzed Feature Types

In this section, we outline the local minutiae descriptors deployed in our implementation, *minutia orientation descriptors*, *minutia frequency descriptors*, and *local minutia structures*, as well as appropriate distance and averaging functions used for the quantization based on a  $K$ -mean clustering algorithm [For65]. We also outline the *absolutely pre-aligned minutiae* used as fourth feature type, which is quantized component-wise.

### 2.1 Minutia Orientation Descriptors

*Minutia orientation descriptors* have been proposed in [TK03] and consists of local estimations of the fingerprint's orientation field sampled from locations around a reference

---

<sup>1</sup>Auxiliary alignment data could be avoided by basing the implementation on a comparison-fit approach [MIK<sup>+</sup>11]; however, this requires an existing correlation between genuine vault minutiae — exactly the property assuming vulnerability against correlation attacks.

minutia; the orientation estimations can be represented w.r.t. the orientation at the reference minutia; in this way, the descriptor is independent of the finger’s rotation and translation. More specifically, according to [TK03, NNJ10, LYC<sup>+</sup>10], a minutia orientation descriptor’s sample coordinates are arranged on 10, 16, 22, and 28 equidistant points lying on four concentric circles of radius 27, 46, 63, and 81 around a local coordinate system defined by the reference minutia, *i.e.*, its position defines the coordinate system’s origin and its angle the direction of the coordinate system’s abscissa. Consequently, an orientation descriptor is a 76-length vector with real entries each encoding an orientation measurements relative to the orientation of the reference minutia (see Fig. 1(a) for a visualization). A method for estimating an image pixel’s local orientation can be found in [KW87].

**Dissimilarity Computation** We compute the difference between two orientation angles  $\phi, \varphi \in [0, \pi)$  as  $0.5 \cdot \text{diff}(2\phi, 2\varphi)$  component-wisely where  $\text{diff} : [0, 2\pi) \times [0, 2\pi) \rightarrow [0, \pi)$  denotes the distance of two angles along the unit circle. In summary, given two orientation descriptors  $\omega = (\omega_1, \dots, \omega_{76})$  and  $\omega' = (\omega'_1, \dots, \omega'_{76})$  we may compute their distance as  $\text{dist}(\omega, \omega') = 1/76 \cdot \sum_{i=1}^{76} 0.5 \cdot \text{diff}(2\omega_i, 2\omega'_i)$  where the normalization factor  $1/76$  guarantees that the distance between two orientation descriptors lies in the interval  $[0, \pi/2)$ .

**Averaging** Given a set of orientation descriptors we may determine its arithmetic mean component-wise, where the average should be computed along the unit circle accounting for the fact that orientations are undirected [KW87].

## 2.2 Minutia Frequency Descriptors

*Minutia Frequency Descriptors* have been proposed in [Fen08] and represent the local inter-ridge distances at coordinates placed around the reference minutia. Precisely, a minutia’s frequency descriptor is thus a 76-length vector with real positive entries (see Fig. 1(b) for a visualization). A method for estimating a fingerprint pixel’s local ridge frequency estimation can be found in [Got12].

**Dissimilarity Computation** We compute the distance between two frequency descriptors, of which components consist of the inverse of local inter-ridge distance measurements, as the normalized Euclidean distance of 76-length vectors. Specifically, given two frequency descriptors  $\lambda = (\lambda_1, \dots, \lambda_{76})$ ,  $\lambda' = (\lambda'_1, \dots, \lambda'_{76}) \in (0, 1]^{76}$ , the distance can be computed as  $\text{dist}(\lambda, \lambda') = 1/76 \cdot \sum_{i=1}^{76} |\lambda_i - \lambda'_i|$ .

**Averaging** The mean of a set of frequency descriptors is computed by applying the harmonic mean component-wise. This corresponds to averaging the components’ inter-ridge distances first and then re-obtaining the inter-ridge frequencies by inverting the results.

## 2.3 Local Minutia Structures

*Local Minutia Structures* have been proposed in [JY00] and consist of a six-length vector  $(d_1, d_2, \theta_1, \theta_2, \phi_1, \phi_2)$  derived from a reference minutia and its two spatially nearest minutiae. Here  $(d_1, \theta_1)$  and  $(d_2, \theta_2)$  are the polar coordinates of the closest and second



Figure 1: *Minutia orientation descriptor* (a) and *minutia frequency descriptor* (b) consist of local orientation and ridge frequency estimations, respectively, sampled on 76 coordinates equidistantly spaced around the reference minutia. A *local minutia local structure* (c) is a six-length vector  $(d_1, d_2, \theta_1, \theta_2, \phi_1, \phi_2)$  encoding the constellation of a the reference minutia and its two spatially closest neighboring minutiae.

closest minutia relative to the reference minutia;  $\phi_1$  and  $\phi_2$  denote the angle of the reference minutia formed with the angle of the closest and second closest minutia, respectively. For a visualization we refer to Fig. 1(c).

**Distance Computation** We adopt the similarity measure used in [LYC<sup>+</sup>10] to derive a reasonable distance function for local minutia structures. That is, given two structures  $\mathbf{s} = (d_1, d_2, \theta_1, \theta_2, \phi_1, \phi_2)$  and  $\mathbf{s}' = (d'_1, d'_2, \theta'_1, \theta'_2, \phi'_1, \phi'_2)$ , we set

$$\text{dist}(\mathbf{s}, \mathbf{s}') = |d_1 - d'_1| + |d_2 - d'_2| + \frac{0.3 \cdot 180}{\pi} \cdot (\text{diff}(\theta_1, \theta'_1) + \text{diff}(\theta_2, \theta'_2) + \text{diff}(\phi_1, \phi'_1) + \text{diff}(\phi_2, \phi'_2)). \quad (1)$$

**Averaging** For a set of  $m$  local minutia structures  $\mathbf{s}^{(j)} = (d_1^{(j)}, d_2^{(j)}, \theta_1^{(j)}, \theta_2^{(j)}, \phi_1^{(j)}, \phi_2^{(j)})$  with  $j = 1, \dots, m$ , the average is computed component-wise, where the average of angles is computed along the unit circle. More specifically, we define  $\text{mean}(\{\mathbf{s}^{(j)}\}) = (\overline{d_1}, \overline{d_2}, \overline{\theta_1}, \overline{\theta_2}, \overline{\phi_1}, \overline{\phi_2})$  where, for  $i = 1, 2$

$$\overline{d_i} = 1/m \cdot \sum_j d_i^{(j)}, \quad \overline{\theta_i} = \arg \left( \sum_j \left( \cos(\theta_i^{(j)}) + \sqrt{-1} \cdot \sin(\theta_i^{(j)}) \right) \right), \quad (2)$$

and  $\overline{\phi_i} = \arg \left( \sum_j \left( \cos(\phi_i^{(j)}) + \sqrt{-1} \cdot \sin(\phi_i^{(j)}) \right) \right).$

## 2.4 Absolutely Pre-aligned Minutiae

As additional alignment-free feature we use its minutiae represented w.r.t. an *intrinsic coordinate system*. This coordinate system is derived from a robust *directed reference*

*point* estimation, *i.e.*, a position and orientation of a reference point: It’s position can be used to define a coordinate system’s origin while the direction defines its orientation. Further, the minutia’s angle is measured relatively to the orientation of the reference point. The directed reference point estimation is taken from [TMM15]

### 3 Proposed Construction of the Improved Fuzzy Vault Scheme

#### 3.1 Quantization

Let  $U$  be the universe of all features of the same type (*e.g.*, minutia local structures) and let  $\text{dist} : U \times U \rightarrow \mathbb{R}_{\geq 0}$  be a distance function that measures the similarity between two features of  $U$ . Assume that we are given a system  $\{u_1, \dots, u_K\} \subset U$  to which we refer as the *quantization system*. Now, we may determine the quantization of an  $x \in U$  by computing the index of its closest element of  $\{u_1, \dots, u_K\} \subset U$ ; this essentially corresponds to a rounding procedure. More specifically, we use the integer

$$\text{quant}(x) = \arg \min_{i=1, \dots, K} \text{dist}(u_i, x) \tag{3}$$

as the quantization of  $x$ . Hence, we can easily compute a quantization of a feature  $x \in U$  assuming that we are given a reasonable quantization system and a reasonable distance function. To establish the quantization system for a general type of features, we may perform a cluster analysis. We employ the well-known  $K$ -mean clustering algorithm [For65] (where  $K$  is considered as a parameter) and use the final quantization system. Therefore, it is required to utilize a reasonable distance and averaging function; these have been specified for the individual feature types outlined in section 2.1, 2.2, and 2.3.

In principle, it is also possible to quantize absolutely pre-aligned minutiae with the help of a quantization system. Yet, a more direct way is to quantize minutia representations component-wisely: Given a minutia coordinate and its angle, the coordinate could be rounded to a rigid grid (*e.g.*, rectangular or hexagonal) while its angle can be quantized into a few number of partitions; such an approach has, for example, been used in [TMM15]. In this paper we consider a variation by replacing the quantization of absolutely pre-aligned minutia coordinates by the quantization of their coordinates in polar representation. More specifically, by  $(\alpha, \beta)$  we denote an absolutely pre-aligned minutia’s coordinate represented w.r.t. a directed reference point; this coordinate can be transformed in polar coordinate representation  $(\delta, \Phi)$  where  $\delta = \sqrt{\alpha^2 + \beta^2}$  and  $\Phi = \arctan_2(\beta, \alpha)$ . In this paper, we divide  $\delta$  through a parameter  $\text{distQuanta} > 0$  and use its nearest integer to encode the quantization; a partition of  $\text{phaseQuanta}$  is used to encode the quantization of  $\Phi$  (see Fig. 2 for a visualization); further,  $\text{angleQuanta}$  is used to quantize an absolutely pre-aligned minutia’s angle.

#### 3.2 Fusion

Let  $U_1, \dots, U_N$  be universes of different feature types. Furthermore, assume that each type is *minutia-related*, *i.e.*, its features relate to a single reference minutia. By  $q_1, \dots, q_N$



Figure 2: Visualization of how minutiae can be quantized such that their representation is alignment-free. First, minutiae are represented with respect to a Cartesian coordinate system; then the polar coordinate representation of the minutiae’s position can be quantized component-wisely; in our implementation, we also account for the minutiae’ angles quantizations.

denote respectively the quantizations of  $x_1, \dots, x_N$  each encoded by an integer in the range  $[0, K_1), \dots, [0, K_N)$  (Sect. 3.1). A feature-level fusion of  $q_1, \dots, q_N$  can be encoded by the integer  $q_1 + q_2 \cdot K_1 + \dots + q_N \cdot K_1 \cdots K_{N-1}$  of the interval  $[0, n)$  where  $n = K_1 \cdots K_N$ .

### 3.3 Fuzzy Vault System

Given a fingerprint, a set of integers  $\mathbf{A} \subset \{0, \dots, n-1\}$  can be extracted containing fusions of alignment-free feature quantizations. We call  $\mathbf{A}$  a *feature set*. Let  $\mathbf{F}$  be a finite field of size at least  $n$ , *i.e.*,  $|\mathbf{F}| \geq n$ . Then each element of  $\mathbf{A}$  can be used to encode an element of  $\mathbf{F}$ ; in this paper we do not necessarily distinguish between feature quantizations from  $\mathbf{A}$  and the finite field elements encoding them.

**Enrolment** The fuzzy vault scheme can be used to store  $\mathbf{A}$  in a protected way by hiding the set unless a sufficiently similar *query set*  $\mathbf{B}$  is presented, *i.e.*, a set  $\mathbf{B} \subset \{0, \dots, n-1\}$  with  $|\mathbf{A} \cap \mathbf{B}|$  being sufficiently large.

On enrolment, a cryptographic key encoded by a secret polynomial  $f \in \mathbf{F}[X]$  of degree smaller than  $k$  is chosen uniformly at random. In the original fuzzy vault construction [JS06], we would generate a set of genuine pairs from  $\mathbf{A}$  lying on the graph of  $f$  and hide them among a randomly generated set of chaff pairs containing pairs not lying on the graph of  $f$ . However, in order to prevent correlation attacks, we would need to use all remaining elements of the feature set  $\mathbf{A}$  as chaff points, which would render the vault record very large. Therefore, we apply the improved fuzzy vault scheme by Dodis *et al.* [DRS04] where genuine and chaff points are encoded by a second polynomial. In order to thwart correlation attacks against the improved fuzzy vault [BA13], we apply a record-specific

permutation to the finite field (see [MT13] for details).

More specifically, we first compute a cryptographic hash  $\text{SHA}(f)$  and use it as a seed to generate a pseudo-random permutation  $\sigma : \mathbf{F} \rightarrow \mathbf{F}$ . Then, the polynomial  $V(X) = f(X) + \prod_{a \in \mathbf{A}} (X - \sigma(a))$  is computed and the pair  $(V(X), \text{SHA}(f))$  is published as the vault record.

**Verification of Positive Biometric Claim** Given a vault record  $(V(X), \text{SHA}(f))$  and a query set  $\mathbf{B}$ , the verifier first reconstructs the field permutation  $\sigma : \mathbf{F} \rightarrow \mathbf{F}$  from  $\text{SHA}(f)$  and then builds the set of unlocking pairs  $\mathbf{U} = \{ (\sigma(b), V(\sigma(b))) \mid b \in \mathbf{B} \}$ . It is important to note that if  $b \in \mathbf{A}$ , then  $V(\sigma(b)) = f(\sigma(b))$  and thus  $(\sigma(b), V(\sigma(b)))$  is a genuine pair; otherwise, if  $b \notin \mathbf{A}$ , then  $V(\sigma(b)) \neq f(\sigma(b))$  and thus  $(\sigma(b), V(\sigma(b)))$  is a chaff pair. Consequently,  $\mathbf{U}$  contains exactly  $|\mathbf{A} \cap \mathbf{B}|$  genuine pairs lying on the graph of the secret polynomial  $f$  which can be recovered if  $|\mathbf{A} \cap \mathbf{B}|$  is sufficiently large. The correctness of  $f$  can be verified using  $\text{SHA}(f)$ .

### 3.4 Decoder

In the original version of the fuzzy vault scheme, the use of a Reed-Solomon decoder has been proposed to recover the polynomial on verification; however, extensive experimental investigations suggest that the error-correction capability of a classical Reed-Solomon decoder seems not to be able to result in acceptable verification performances for single-finger systems (*e.g.*, see [NJP07, LYC<sup>+</sup>10, TMM15]). Instead, most implementations work by iterating through all  $k$ -sized subsets of  $\mathbf{U}$  and for each subset compute its interpolation polynomial  $f^*$ ; if  $\text{SHA}(f^*) = \text{SHA}(f)$ , then, with very high reliability,  $f^* = f$  and recovery of  $f$  is considered as successful resulting in an accept decision; otherwise, if for all  $\binom{|\mathbf{U}|}{k}$  iterations  $\text{SHA}(f^*) \neq \text{SHA}(f)$ , then  $f$  could not be discovered resulting in a reject decision.

However, for large  $|\mathbf{U}|$  the above systematic decoding approach easily becomes infeasible; for example, in [NJP07] the unlocking sets can be of size up to  $|\mathbf{U}| = 24$  where  $k = 9$  resulting in a worst-case running time of  $\binom{24}{9} \approx 2^{20}$  polynomial interpolations. Consequently, in [TMM15] a randomized decoding approach has been proposed in which not all but at most  $\text{numDecIts}$  polynomial iterations with randomly selected  $k$ -sized subsets of  $\mathbf{U}$  are performed. In our experiments, we strictly utilized this randomized decoding strategy, though it may be easily modified by selecting larger subsets allowing for some errors which can be corrected with a Reed-Solomon decoder; this could result in a more efficient decoder. As a reasonable choice for  $\text{numDecIts}$  we selected  $2^{16}$ .

## 4 Experiments

### 4.1 Quantization Systems

Using the FVC 2002 DB1 [MMC<sup>+</sup>02] we established quantization systems for the three alignment-free feature types minutia orientation descriptors OD, minutia frequency descriptors FD, and local minutia structures LMS as described in Sect. 2. For each feature type we used the first (among eight) impressions of the first 55 fingers to build a “cloud”

Table 1: Parameters selected on base of a training for different feature-level fusion strategies.

	OD+FD+LMS	APM	APM+OD+FD+LMS
odQuanta	31	1	5
fdQuanta	26	1	1
lmsQuanta	31	1	1
distQuanta	–	19	20
phaseQuanta	–	11	9
angleQuanta	–	10	8
maxFeatures	34		
numDecls	$2^{16}$		

of feature elements; then, for each  $K = 1, \dots, 32$  the feature cloud has been input to our  $K$ -mean clustering implementation which resulted in a candidate for the final  $K$ -sized quantization system. To assess the quality of the quantization system, the *reproducibility rate* has been determined with the help of ground-truth minutia correspondences manually marked between the first and second impressions of the remaining 55 fingers of the FVC 2002 DB1. We repeated the  $K$ -mean clustering procedure 1000 times and selected the system that resulted in the highest reproducibility rate.

## 4.2 Parameters

We performed experiments with our fuzzy vault implementation. Among the feature types OD, FD, LMS, and absolutely pre-aligned minutiae APM, we tested the following three feature-level fusions: 1. OD+FD+LMS; 2. APM only; and 3. APM+ OD+ FD+LMS. For the respective fusions, on base of a previous training, we selected the following parameters which are also listed in Tab. 1:

- **odQuanta**, **fdQuanta**, and **lmsQuanta** denoting the number of clusters/quantization system size for minutia orientation descriptors, minutia frequency descriptors, and local minutia structures, respectively.
- **distQuanta**, **phaseQuanta** denoting the quantization parameters for an absolutely pre-aligned minutia’s coordinate in polar representation where the first and the second correspond to the distance and angular coordinate, respectively; these parameters are only relevant if the fusion contains the feature type absolutely pre-aligned minutiae (Sect. 2.4).
- **angleQuanta** denoting the quantization parameter for minutia angles; this parameter is only relevant if the fusion contains the feature type absolutely pre-aligned minutiae APM (Sect. 2.4).
- **maxFeatures** denoting the maximal number of quantized features protected by a vault; if from a fingerprint more than **maxFeatures** features can be extracted, those relating to the highest minutia quality are selected.
- **numDecls** denoting the number of decoding iterations on verification (Sect. 3.4).

Table 2: Verification performance achievable with our fuzzy vault implementation for different feature-level fusion strategies.

$k$	OD+FD+LMS		APM		APM+OD+FD+LMS	
	GAR (FAR)	GDT (IDT)	GAR (FAR)	GDT (IDT)	GAR (FAR)	GDT (IDT)
1	97% (34%)	3ms (60ms)	99% (92%)	1ms (7ms)	98% (81%)	2ms (18ms)
2	89% (9%)	13ms (106ms)	98% (78%)	2ms (25ms)	98% (58%)	2ms (50ms)
3	79% (2%)	33ms (149ms)	98% (60%)	3ms (60ms)	98% (38%)	4ms (95ms)
4	65% (0.3%)	82ms (202ms)	98% (38%)	4ms (121ms)	97% (21%)	5ms (163ms)
5	44% (0%)	162ms (257ms)	97% (17%)	7ms (192ms)	97% (7%)	9ms (231ms)
6	25% (0%)	268ms (331ms)	96% (7%)	11ms (260ms)	96% (2%)	15ms (303ms)
7	12% (0%)	378ms (416ms)	95% (2%)	19ms (322ms)	94% (0.46%)	29ms (375ms)
8	5% (0%)	486ms (505ms)	94% (0.4%)	30ms (389ms)	91% (0.06%)	51ms (454ms)
9	2% (0%)	599ms (607ms)	92% (0.12%)	49ms (458ms)	88% (0.02%)	86ms (540ms)
10	1% (0%)	714ms (719ms)	89% (0.08%)	77ms (533ms)	82% (0%)	142ms (633ms)
11	0.4% (0%)	834ms (837ms)	85% (0.02%)	118ms (614ms)	76% (0%)	220ms (733ms)
12	0.2% (0%)	978ms (980ms)	80% (0%)	179ms (704ms)	69% (0%)	325ms (855ms)

### 4.3 Evaluation

For each of the three tested fusions and the parameters determined during training, we evaluated the verification performance of our fuzzy vault implementation with the help of the optical scans of right index fingers contained in the MCYT-100 database [OGFAS03]. To measure the genuine acceptance rate **GAR**, we used each of the 100 individual's  $j$ th scans ( $j = 1, \dots, 11$ ) to generate a fuzzy vault-protected record. The remaining scans ( $j' = j + 1, \dots, 12$ ) were used to perform a total of  $11 \cdot 12/2 = 66$  genuine verification attempts per person. Consequently, we performed up to 6,600 genuine verification attempts. To measure the false acceptance rate, for each person (labeled  $i = 1, \dots, 100$ ) we generated a fuzzy vault record using his first scan. The remaining persons' ( $i' = i + 1, \dots, 100$ ) first scans were used to perform impostor verification attempts. In such a way, we ran a total of 4,950 impostor verification attempts. Furthermore, we kept track of the average verification times on genuine and impostor verification that we denote by **GDT** and **IDT**, respectively. The result of our evaluation, conducted on a single core of a 1.9 GHz server, can be found in Tab. 2.

As can be seen from Tab. 2, with a fusion of minutiae orientation descriptors, minutia frequency descriptors, and local minutia structures we reached a **GAR** of 44% at the zero **FAR** for  $k = 5$ . In comparison, merely using absolutely pre-aligned minutiae has the capability of providing the significantly better **GAR** of 80% at a similar **FAR** for  $k = 12$  which can even be slightly improved when combined with minutia orientation descriptors resulting in a **GAR** of 82% for  $k = 10$ . Furthermore, we found that the average decoding times can be performed within an amount of time significantly smaller than a second which makes our implementation feasible to be run in verification mode.

One may argue that, for example in [NJP07, NNJ10], a better genuine acceptance has been reached. We stress, however, that these implementations are vulnerable to cross-matching and information leakage from auxiliary alignment data. When compared with another existing implementation avoiding these problems [TMM15] in which a **GAR** of 79% at the zero **FAR** has been reached, we may conclude that the performance of our

implementation is slightly better.

#### 4.4 Security

A very important aspect of a fuzzy vault implementation is its resistance to recovery attacks, *i.e.*, the effort for an attacker given a vault record to recover the original feature sets or, equivalently, the secret polynomial. Generally, the fuzzy vault can be attacked by a *brute-force attack*, where the attacker repeatedly samples  $k$  points from the vault and tries to interpolate the secret polynomial from these. The expected number of attempts of this attack can be estimated by combinatorial means [MMT09].

In contrast, the *false-accept attack* exploits the specific distribution of the biometric features, by repeatedly simulating (impostor) verifications using the features of randomly chosen (real) fingerprints, *e.g.*, chosen from a biometric database [TMM15]. The success probability of the false-accept attack is equal to the FAR for the parameters used. In general, the attacker can deviate in her simulation from the parameters used in actual operation to optimize her success rate; however, in our fuzzy vault implementation, the number of decoding iterations `numDecIIts` is the only parameter that is not already fixed in the enrolment. It has been proven in [TMM15] the expected number of decoding attempts of the false-accept attack is minimized for `numDecIIts` = 1. Hence, we estimate the security against false-accept attacks using this optimal strategy.

Estimating very high security levels assumes sharp estimations of FARs when they are close to zero. In biometric systems with deterministic verification algorithm, the FAR can only be estimated down to the magnitude of  $1/N$ , where  $N$  is the number of impostor verifications performed in the evaluation. However, the verification of our implementation is probabilistic as soon as the unlocking set contains more than  $k$  points. This property allows us to give heuristic estimates of FARs that are much smaller than  $1/N$ : For each single impostor verification, we compute the success probability based on the size of the unlocking set and the number of correct points contained, and, finally, we estimate the FAR as the mean over all verifications. For details, we refer to [TMM15].

It turns out that for the parameters chosen, the false-accept attack is much more efficient than the brute-force attack and, hence, we estimate the security against recovery attacks by the expected number of attempts required for a false-accept attack, *i.e.*, by the reciprocal of the FAR achieved with `numDecIIts` = 1.<sup>2</sup> Fig. 3 shows a plots of the genuine acceptance rate versus the security level (depending on  $k$ ), for different combinations of features.

Another very important security aspect concerns the risk of correlation attacks on two or more vault records of the same user. Since we use the improved fuzzy vault scheme, which effectively uses all finite field elements as vault points [DRS04], the correlation attack from [SB07] cannot be applied. On the other hand, there are specific correlation attacks against the improved fuzzy vault scheme based on solving systems of polynomial equations [BA13] or deploying the extended Euclidean algorithm [MT13]. However, these attacks only work, if in both vault records the features are represented by the same finite field elements, and, hence, are prevented by our use of a record-specific permutation  $\sigma$  of

---

<sup>2</sup>This estimate is conservative insofar as we neglect the attacker's computational effort for verification.

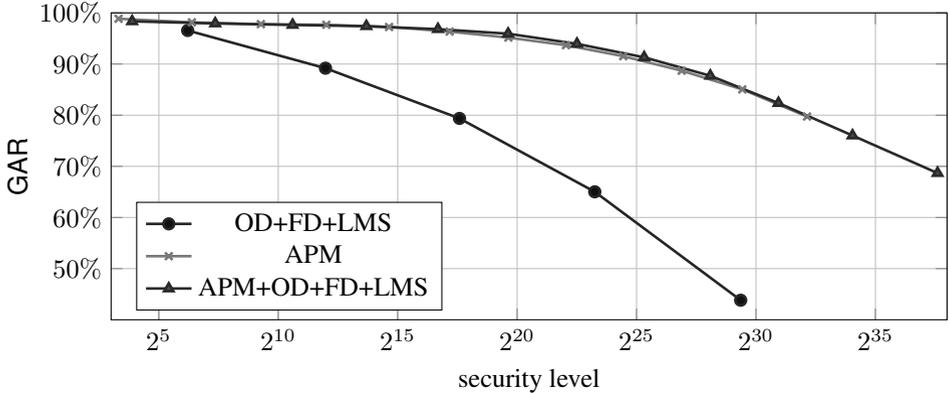


Figure 3: Genuine acceptance rate plotted versus false-accept security.

the field elements (see Sect. 3.3); for further details we refer to [TMM15].

## 5 Discussion

In this work, we designed an implementation of the improved fuzzy vault scheme for three fusions of alignment-free fingerprint feature types. We considered four different feature types one of which is given by absolutely pre-aligned minutiae. The choice of the other three types has been motivated by the work of Li *et al.* [LYC<sup>+</sup>10]; for these feature types, a generic quantization scheme based on the  $K$ -mean clustering algorithm is proposed in the present paper which can be padded with a maximal number of chaff in a fuzzy vault to achieve resistance against linkage attacks.

If quantizations of the feature types from Li *et al.* are fused using the techniques of this paper, we found that the achievable verification performance is clearly inferior as compared to the use of absolutely pre-aligned minutiae (see Fig. 3). Yet, our investigations indicate that, if absolutely pre-aligned minutiae are fused with other alignment-free feature types, verification performance can be slightly improved. From our experiments, we may therefore conclude that absolutely pre-aligned minutiae seem to be an indispensable feature type for the verification performance of a fingerprint-based fuzzy vault. In this view, it seems worthwhile to improve the robustness of existing directed reference point estimation methods during future research. However, this research should be conducted while having in mind that a single finger only seems not be capable of providing sufficient security.

## References

- [BA13] M. Blanton and M. Aliasgari. Analysis of Reusability of Secure Sketches and Fuzzy Extractors. *IEEE Trans. Inf. Forensics Security*, 8(9):1433–1445, 2013.
- [BBGK08] J. Breebart, C. Busch, J. Grave, and E. Kindt. A Reference Architecture for Biometric

- Template Protection based on Pseudo Identities. In *Proc. BIOSIG*, pages 25–37, 2008.
- [BFPdS14] J. Bringer, M. Favre, C. Pelle, and H. d. Saxcé. Fuzzy vault and template-level fusion applied to a binary fingerprint representation. In *BIOSIG 2014*, pages 235–242, 2014.
- [CKL03] T. Charles Clancy, Negar Kiyavash, and Dennis J. Lin. Secure Smartcard-Based Fingerprint Authentication. In *Proc. ACM SIGMM workshop on Biometrics methods and applications*, pages 45–52, New York, NY, USA, 2003. ACM.
- [DRS04] Y. Dodis, L. Reyzin, and A. Smith. Fuzzy Extractors: How to Generate Strong Keys from Biometrics and Other Noisy Data. In *EUROCRYPT*, pages 523–540, 2004.
- [Fen08] J. Feng. Combining minutiae descriptors for fingerprint matching. *Pattern Recognition*, 41(1):342–352, 2008.
- [For65] E. W. Forgy. Cluster Analysis of Multivariate Data: Efficiency versus Interpretability of Classifications. *Biometrics*, 21:768–769, 1965.
- [Got12] C. Gottschlich. Curved Regions Based Ridge Frequency Estimation and Curved Gabor Filters for Fingerprint Image Enhancement. *IEEE Trans. Image Process.*, 21:2220–2227, 2012.
- [ISO11] ISO/IEC JTC1 SC2 Security Techniques. ISO/IEC 24745:2011. Information Technology - Security Techniques - Biometric Information Protection. International Organization for Standardization, 2011.
- [JS06] A. Juels and M. Sudan. A Fuzzy Vault Scheme. *Des. Codes Cryptography*, 38(2):237–257, 2006.
- [JY00] X. Jiang and W.-Y. Yau. Fingerprint minutiae matching based on the local and global structures. *Proc. Int. Conf. on Pattern Recognition ICPR*, 2:1038–1041, 2000.
- [KW87] M. Kass and A. Witkin. Analyzing oriented patterns. *Computer Vision, Graphics, and Image Processing*, 37(3):362–385, 1987.
- [LYC<sup>+</sup>10] P. Li, X. Yang, K. Cao, X. Tao, R. Wang, and J. Tian. An alignment-free fingerprint cryptosystem based on fuzzy vault scheme. *J. Netw. Comput. Appl.*, 33:207–220, 2010.
- [MIK<sup>+</sup>11] J. Merkle, H. Ihmor, U. Korte, M. Niesing, and M. Schwaiger. Performance of the Fuzzy Vault for Multiple Fingerprints. In *Proc. BIOSIG’11*, pages 57–72, 2011.
- [MMC<sup>+</sup>02] D. Maio, D. Maltoni, R. Cappelli, J.L. Wayman, and A.K. Jain. FVC2002: Second Fingerprint Verification Competition. In *Proc. Int. Conf. on Pattern Recognition*, pages 811–814, 2002.
- [MMT09] Preda Mihăilescu, Axel Munk, and Benjamin Tams. The Fuzzy Vault for Fingerprints is Vulnerable to Brute Force Attack. In *Proc. of BIOSIG*, pages 43–54, 2009.
- [MT13] J. Merkle and B. Tams. Security of the Improved Fuzzy Vault Scheme in the Presence of Record Multiplicity. *CoRR abs/1312.5225*, 2013. available online: <http://arxiv.org/abs/1312.5225>.
- [NJP07] K. Nandakumar, A. K. Jain, and S. Pankanti. Fingerprint-Based Fuzzy Vault: Implementation and Performance. *IEEE Trans. Inf. Forensics Security*, 2(4):744–757, 2007.
- [NNJ10] A. Nagar, K. Nandakumar, and A. K. Jain. A hybrid biometric cryptosystem for securing fingerprint minutiae templates. *Pattern Recogn. Lett.*, 31:733–741, June 2010.
- [OGFAS03] J. Ortega-Garcia, J. Fierrez-Aguilar, and D. Simon *et al.* MCYT baseline corpus: a bimodal biometric database. *IEE Proc. on Vision, Image and Signal Processing*, 150(6):395–401, 2003.
- [SB07] W. J. Scheirer and T. E. Boulton. Cracking Fuzzy Vaults and Biometric Encryption. In *Proc. of Biometrics Symp.*, pages 1–6, 2007.
- [TK03] M. Tico and P. Kuosmanen. Fingerprint Matching Using an Orientation-Based Minutiae Descriptor. *IEEE Trans. Pattern Anal. Mach. Intell.*, 25(8):1009–1014, August 2003.
- [TMM15] B. Tams, P. Mihăilescu, and A. Munk. Security Considerations in Minutiae-based Fuzzy Vaults. *IEEE Trans. Inf. Forensics Security*, 10(5):985–998, 2015.

# Protected Honey Face Templates

Edlira Martiri<sup>1</sup>, Bian Yang<sup>2</sup>, Christoph Busch<sup>3</sup>  
{edlira.martiri; bian.yang; christoph.busch}@hig.no

<sup>1,2</sup>Norwegian Biometrics Laboratory, Gjøvik University College, 2815 Gjøvik, Norway

<sup>3</sup>Hochschule Darmstadt - CASED, Haardtring 100, 64295 Darmstadt, Germany

**Abstract.** Most existing biometric template protection schemes (BTPS) do not provide as strong security as cryptographic tools; and furthermore, they are rarely able to detect during a verification process whether a probe template has been leaked from the database or not (i.e., being used by an imposter or a genuine user). By using the “honeywords” idea, which was proposed to detect the cracking of hashed password database, we show in this paper how to enable the detectability of biometric template database leakage.

We add an extra layer of protection since biometric features cannot be renewed. The biometric system design implies that protection mechanisms must satisfy the irreversibility property and to this respect we apply different correlation tests to show the non-distinguishability between genuine and honey templates. In this paper we implement the idea of a honey template protection scheme on faces and evaluate the security and accuracy performance.

## 1 Introduction

From a security perspective, the protection of templates stored in the database of a biometric system is one of the main challenges. An adversary can try to carry out a modification of their contents or even an unauthorized transfer of templates from the database towards another system. He can then generate a pre-image of the template by hill-climbing or brute force attack and since the adversary can create a fake physical characteristic from the biometric template this leads to physical masquerade attacks. All these attacks may occur as in a biometric standalone system, such as an automated border control system, as well as in remote biometric authentication. To prevent the leakage of biometric information, in addition to a better control of database access, other techniques should be implemented to prevent attacks and, even better, warn if such a leak has occurred.

While masquerade attacks are possible to cope with by better anti-spoofing technologies, they are not very possible to be completely prevented. To discourage both the physical and the digital masquerade attacks, we do this by empowering the system with detectability of the leakage of protected templates. Juels in [JR13] proposes the idea of “honeywords” used on passwords and we extend this idea to the biometric templates. In this paper we elaborate further the architecture design for a biometric system using BTPS-based honey templates explained in [YM15] and apply this idea on faces. We also evaluate the security, biometric recognition performance and irreversibility of honey face templates. These are detailed in Section 3 and 4 whereas in section 2 is given the background information and a review of honey objects.

## 2 Biometric templates as honey objects

Honey objects are used in various aspects of system security to deceive internal threats or external intruders (be they people or machines) against unauthorized data access. An example are honeypots mostly used for the detection of outer intruders, which are network machines used to distract adversaries from other more important machines, and honey farms (a network of honeypots) [Ho15] enabling deep research into server-side attacks. Subsequently, honeytokens [Sp03] are mostly implemented against internal threats and another development we find at system level are honeyclients [Na09], the complementary of honeypots, designed to mimic the behavior of a web browser. On the data level we find solutions such as the honeywords and honeydocuments. The honeyword method [JR13] hides the password of a user between  $k$  hash values of random passwords, and honeydocuments [Bo09] is again a trap-based mechanism which uses decoy documents. All these mechanisms serve as a safeguard against adversaries who try to get unauthorized data access. In [Ju14] honey objects must comply to two main properties: (1) *indistinguishability*, honey and real objects must be hard to distinguish from each other (e.g. a real password from a generated password or a database entry of a real patient in a health system from a fake one); (2) *secrecy*, the real object must be secret and camouflaged among the honey objects. In the case of honeywords, if an intruder by some means gets access to the user's set of passwords, he can use/guess only one of them. The system intercepts that a honeyword is used, it will consider this as information leakage and proceed with further steps (set off an alarm and/or update the passwords set).

The honeywords method provides us a systematic way to counter the masquerade attack against protected biometric templates. It resorts to probability (i.e. information-theoretic security) instead of computational complexity based security to cope with the crackable-hash assumption. In the biometric context, most databases are facing the same challenges. In hash based biometric template protection scheme, such as fuzzy commitment [JW99], and secure sketch [SLM07], if the hash is cracked, then the adversary can estimate the pre-image of the biometric features. And for feature-transformation based BTPs (in [RCB01] and [TGN06]), the masquerade attack is even more straightforward. This is because the protected templates, PTs, are compared directly with a distance threshold and the attacker can find a PT's pre-image (biometric feature) with normally less effort than the case finding a pre-image of a hash value. As a result, for every enrolled user in the database, we need to provide a protection mechanism which needs to be applied on all the sweet templates (sugar and honey) and satisfies the abovementioned properties. Firstly, templates must be constructed in such a way that an adversary is not able to distinguish a sugar from a honey one, even if he: breaks the protection mechanisms; uses automatic tools such as classifiers; or tries to visually capture differences of honey and sugar templates pre-images to differentiate them. Secondly, the sugar template must be placed in a random position in the user database entry, or user data file, among the honey templates and this specific index must be known only to the honeychecker. We note that the aim of our approach on biometric templates, as well as the honeywords method, is not to lure the intruders with fake data, but to provide a means to alert the system that an internal or external adversary had access to the users' data and used them back: in other words that there have been system attack, information leakage, and user impersonation (masquerade attack).

## 2. 1. Honey objects database design

The design of a honey objects database is shown in Figure 1 and it can be applied to passwords, biometric templates and other objects which have similar properties with them, in terms of usage and storage. As in [YM15] for the  $i^{\text{th}}$  user, the sugar object  $S_i$  will be created from the user data and  $K-1$  honey objects  $H_{ij}$  ( $j = 1, 2, \dots, K-1$ ) will be generated. These objects will pass through a protection mechanism (like hashing in passwords or one of the BTPS in biometric systems) having as a result a set of  $K$  protected objects: one  $SP_i$  from the sugar object and  $K-1$  protected honey objects  $HP_{ij}$  ( $j=1, 2, \dots, K-1$ ). To hide  $SP_i$  among the other honey objects, its memory address or index is randomly allocated and to the other objects are assigned the remaining addresses or indices. This process is handled by the Order Randomization block, which uses the auxiliary data  $AD_i$  to generate the  $SP_i$ 's index and the indices of the protected honey objects. We define the protected objects as  $PI_{ij}$  ( $j = 1, 2, \dots, K$ ) and the set  $PO_i = \{AD_i, PI_{i1}, PI_{i2}, \dots, PI_{iK}\}$  of user  $i$ , containing the auxiliary data and the randomized protected objects. This set is stored in the database whereas the index  $L_i$  is stored in the Honey Checker Database.

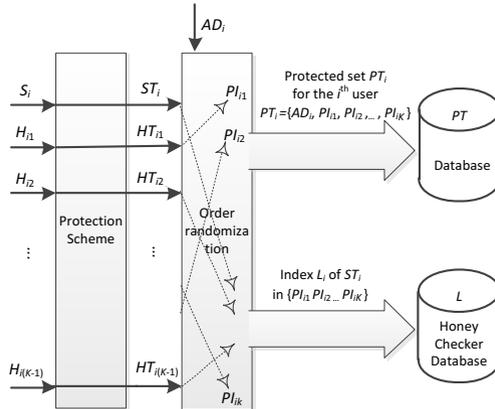


Figure 1. Honey objects database architecture design

In figure 2 we recall from [YM15] the architecture design of a biometric system using honey templates. During enrollment the Application Server converts the plain biometric feature  $B_i$  of user  $i$  to a set of protected templates. This set is defined in the same way as we did with the protected objects  $PO$ , i.e.  $PT_i = \{AD_i, PI_{i1}, PI_{i2}, \dots, PI_{iK}\}$ .  $PT_i$ . It will be stored in the Biometric Database and the index  $L_i$  in the Honey Checker Database. During verification the user  $i$  will show his biometric characteristic and the Application Server will retrieve the  $PI_i^*$  from the Biometric Database and send it to the Biometric Database Server. After comparing  $PI_i^*$  with all the  $PI_{ij}$  ( $j = 1, 2, \dots, K$ ) of user  $i$ , it will send the best-matched template's index to the Application Server. If the index  $idx_i$  matches  $L_i$  the user will grant access to the system, otherwise an alarm will be set off and specific rules will follow, according to the defined security policies of the system. In the latter case, if the user personal identifiers match but the templates' indices do not, the system will consider this attempt as information leakage.

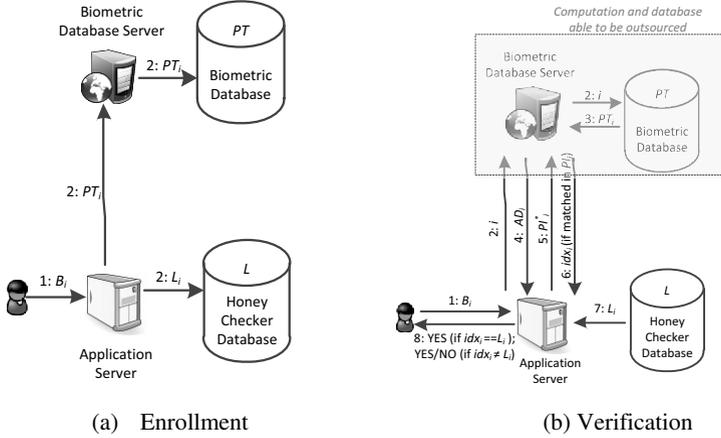


Figure 2. Architecture for a honey templates based biometric system [YM15].

### 3 Honey Face Templates

The architecture of a honey templates based biometric system in figure 2 can be applied to different biometric characteristic. Our first attempt is in the construction and protection of honey face templates. The main challenge, as we mentioned in section 2, for a honey based system is to generate honey objects, in our case synthetic face templates, which cannot be distinguished from real templates. The protection mechanism we have adopted on the faces' feature vectors, is the plain-feature-defined sub-set selection borrowed by the idea in [Ya10]. Recognition performance degradation can be anticipated from the adoption of BTPS as discussed in [Si12]. This degradation caused by adding honey templates can be easily seen as well in our case. To make a quick proof-of-concept evaluation of the proposed honey template based biometric system, we created two small-scale face databases denoted as  $DB_{aux}$  and  $DB_{lst}$  representing an auxiliary database (for purposes of PCA training and the construction of  $AD_i$  used by BTPS) and a testing database, respectively.  $DB_{lst}$  is formed by 40 faces with 10 samples each as a sub-set of the ORL face database [ORL15]. To be fair in performance evaluation,  $DB_{aux}$  is formed by 40 faces from public websites with 1 sample for each face. All these 40 samples in  $DB_{aux}$  were cropped and normalized in their size to the same specification of those ORL samples (*i.e.*,  $92 \times 112$  pixels, 256 gray scales).

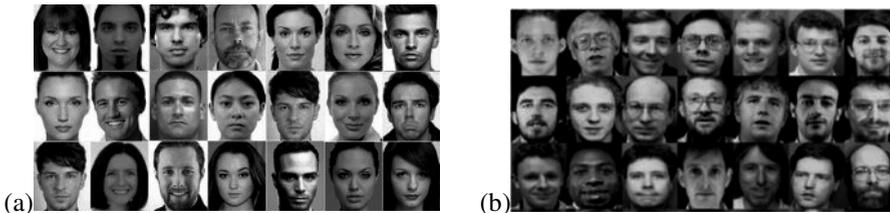


Figure 3: (a) Samples in  $DB_{aux}$  which is used for eigen faces training and used as  $AD_i$  by BTPS. (b) Samples in  $DB_{lst}$  which is used for recognition performance

Suppose we take the first  $N = 20$  PCA coefficients  $\mathbf{c}_{\text{lst}} = (c_{\text{lst}1}, c_{\text{lst}2}, \dots, c_{\text{lst}20})$  (weights of eigenfaces that are trained from  $\text{DB}_{\text{aux}}$  to decompose each of test sample in  $\text{DB}_{\text{lst}}$ ) of a test sample as the sugar biometric feature vector  $SB = \mathbf{c}_{\text{lst}}$ . Define an indicative binary vector  $\mathbf{v} = (v_1, v_2, \dots, v_{20})$  as

$$v_j = \begin{cases} 0 & \text{if } c_{\text{lst}j} \geq 0 \\ 1 & \text{if } c_{\text{lst}j} < 0 \end{cases} \quad (1)$$

According to the indicative binary vectors we can randomly group  $\text{DB}_{\text{aux}}$  into two non-overlapped sub-sets  $\text{DB}_{1\text{aux}}$  and  $\text{DB}_{2\text{aux}}$  with 20 samples each. Then we generate 20 PCA coefficient feature vectors, denoted as  $\mathbf{c}1_{\text{aux}q} = (c1_{\text{aux}q1}, c1_{\text{aux}q2}, \dots, c1_{\text{aux}q20})$  and  $\mathbf{c}2_{\text{aux}q} = (c2_{\text{aux}q1}, c2_{\text{aux}q2}, \dots, c2_{\text{aux}q20})$ , from each of the two sub-sets, respectively, with  $1 \leq q \leq 20$ . From the two groups of feature vectors  $\mathbf{c}1_{\text{aux}q}$  and  $\mathbf{c}2_{\text{aux}q}$  ( $1 \leq q \leq 20$ ), a 20-dimensional mask vector  $\mathbf{m} = (m_1, m_2, \dots, m_{20})$  is constructed as

$$m_j = \frac{s_j}{20} \sum_{q=1}^{20} |(1 - v_q) c1_{\text{aux}qj} + v_q c2_{\text{aux}qj}| \quad (2)$$

where  $s_j$  is defined as a sign value (+1 or -1) in eq. (3) depending on  $A_j$  defined in eq.(4):

$$s_j = \begin{cases} 1 & \text{if } A_j \geq 0 \\ -1 & \text{if } A_j < 0 \end{cases} \quad (3)$$

$$A_j = \sum_{q=1}^{20} ((1 - v_q) c1_{\text{aux}qj} + v_q c2_{\text{aux}qj}) \quad (4)$$

Now with the plain feature vector  $\mathbf{c}_{\text{lst}}$  and the mask vector  $\mathbf{m}$ , we can generate the protected sugar template  $ST = (st_1, st_2, \dots, st_{20})$  as

$$st_j = \text{rem}(\text{rem}(w_1 m_j, 2\sigma) + w_2 c_{\text{lst}j}, w_3 \sigma) \quad (5)$$

where  $\sigma$  is an estimated standard deviation of all feature vectors in  $\text{DB}_{\text{lst}}$ , and the three weights  $w_1, w_2$ , and  $w_3$  can be tuned to achieve the best trade-off between security and recognition performance. The values of the weights  $w_i$  with  $i$  in  $\{1, 2, 3\}$  will be discussed later during the experimental evaluation. “*rem*” is a modular operation but keep the value’s sign. In a brief, the above BTPS relates the mask feature vector generation to the plain feature vector’s components’ signs and thus obtains varied mask feature vectors for each different plain feature vector. For each sugar protected template, we try generating 15 honey protected templates to make the total number of templates  $K = 16$ . The generation process of a honey template is exactly the same as that of a sugar template as described in Eq.(1-5) except that the plain feature  $\mathbf{c}_{\text{lst}}$  is replaced by a random feature vector with each component’s dynamic range  $[-0.5, +0.5]$ .

## 4 Security and irreversibility evaluations

In our BTPS design,  $AD_i$  is assumed to be public and length  $N$  of the PCA feature vector (*i.e.*, how many eigenfaces are used) decides the security level since  $\mathbf{m}$  is constructed by

$N$  selected feature vectors from  $c1_{aux}$  and  $c2_{aux}$ . In our quick proof-of-concept experiment,  $N$  is set to be 20 and therefore the complexity of identifying the correct 20 feature vectors selected is  $2^{20}$ . In practice, we can extend the security depth by increasing  $N$ , or keep the  $DB_{aux}$  as a secret parameter.

Whether an adversary can distinguish the generated sugar protected templates from those honey protected templates is the key criterion for the proposed honey template generation method. We further check if a machine learning algorithm can be used to classify the protected templates and identify the  $ST$  from other  $HT$ s. SVM was trained by half (200 test  $PI$ s) generated protected templates with ground-truth labels (sugar or honey) and evaluated by the other half (the other 200 test  $PI$ s) generated protected templates in our experiments and the classification results are given in Table 1 with various  $w_1, w_2$ , and  $w_3$  settings. It indicates that the proposed honey template generation method can well hide the sugar template among those honey ones – though the FMR and the FNMR are less than 50%, they should be already large enough to discourage the adversary to launch a masquerade attack if  $N$  is large enough. In Table 1, the values (1, 1, 1) of ( $w_1, w_2, w_3$ ) imply that we have not taken into consideration these parameters.

Table 1: Classification of protected templates by SVM for different values of  $w_1, w_2, w_3$ .

$w_1$	$w_2$	$w_3$	FMR	FNMR
1	1	1	0.3433	0.4050
1	1	36	0.2287	0.4000
1	1	72	0.2380	0.4200
1	2	6	0.1470	0.4350
1	2	36	0.1553	0.3800
1	2	72	0.2417	0.4100
2	2	6	0.1150	0.4750
2	4	6	0.1537	0.3800
36	6	6	0.1573	0.3500
72	36	6	0.1767	0.3950

The need for a deep understanding of the security properties of our BTP scheme is crucial and has to be well characterized. In [Ma06] it is shown that all template protection schemes including fuzzy encryption, biometric salting and cancelable biometrics offer limited protection against different attacks. For instance, fuzzy encryption is vulnerable against linkage attacks, biometric encryption to Hill-Climbing attack and helper data scheme is limited especially by the correlation of features. An analysis of security and privacy protection can be seen in [Zh09] where it is implied that if algorithms such as ICA, PCA or LDA are used, the resulting features are more uncorrelated.

Biometric systems design implies that template protection mechanism must be such that the template should not imply the raw data of the extracted features. This is known as irreversibility [ISO11] and it is defined as the difficulty of determining exactly or with tolerance margin from a protected template the biometric sample(s) or features used during enrolment to generate that template [S112].

#### 4.1. Irreversibility evaluation tests and results

In this section we will show the honey-based BTP on faces shows lack of irreversibility. To do this we provide the tests below showing the correlation values between the feature elements and the protected sugar templates; between the excerpt of the feature vectors used to create the honey templates and the honey templates themselves; and between sugar templates and the correspondent honey templates in their protected and unprotected form.

##### *Correlation test between plain feature vectors and protected templates*

Correlation tests are done to measure how strong the relationship between two vectors or variables is. Our first test to this regard is between the plain feature vectors  $c_{\text{tst}}$  and the protected templates  $ST$  of user  $i$ . The general mathematical formula we have used is:

$$\text{corr}(A, B) = \frac{A * B^T}{\sqrt{\text{var}(A) * \text{var}(B)}} \quad (6)$$

where  $\text{var}(x)$  is the variance of the vector  $x$ , and it is interpreted as the covariance of  $x$  with itself. The covariance  $\text{cov}(x, y)$  between vectors  $x$  and  $y$  shows how much the two vectors differ from each other.

The correlation values interval is between 1 and -1. If the correlation is zero or near to zero, means that there is no correlation between the vectors. If the correlation value tends to be 1, this means that there exists a direct correlation between the two vectors and if it tends to be -1 then the two vectors are inversely related (they go in opposite directions).

In our terms, the first test will be applied between the PCA coefficients  $c_{\text{tst}} = (c_{\text{tst}1}, c_{\text{tst}2}, \dots, c_{\text{tst}20})$  (weights of eigenfaces that are trained from  $DB_{\text{aux}}$  to decompose each of test sample in  $DB_{\text{tst}}$ ) and the sugar protected templates  $ST = (st_1, st_2, \dots, st_{20})$  which is expressed in eq. (5). We have evaluated this test for different values of the tuning parameters triplet  $(w_1, w_2, w_3)$ . The results are presented in Table 2 where we can see that there is zero correlation between the two vectors. This means that there can be no information leakage from the protected sugar template.

##### *Correlation test between selected or unselected feature vectors and protected templates*

As we mentioned in section 3.1 the sugar template is constructed by selecting randomly a part of the set of the plain feature vectors. In order to argue the no correlation between the final protected sugar template and these two separate and non-overlapping subsets we performed two correlation tests between the selected vectors used to construct the sugar templates and the protected sugar templates. In this construction, the selected feature vectors are expressed in equation (4), and the unselected portion of the set of feature vectors can be expressed as:

$$A^*_j = \sum_{q=1}^{20} (v_q c1_{\text{aux}qj} + (1 - v_q) c2_{\text{aux}qj}) \quad (7)$$

The results of the two abovementioned tests are presented in Table 2.

Table 2: Correlation values between {plain feature vectors, selected, and unselected vectors} and the protected sugar template

No.	A	B	Corr (A, B)
1	Sugar feature vectors (SB)	Protected Sugar Template (ST)	0
2	Selected vectors for Sugar Template ( $A_j$ )	Protected Sugar Template (ST)	0,06423
3	Unselected vectors for Sugar Template ( $A_j^*$ )	Protected Sugar Template (ST)	-0.06416

The problem of irreversibility between templates must be considered also for the honey protected template case. We can see that between the selected vectors of the sugar feature vectors and the corresponding honey-templates the correlation mean value is very low. Nearly the same correlation situation can be seen between the protected honey templates and the counterpart of the selected vectors in the set of the plain feature vectors. The results are presented in Table 3.

Table 3: Correlation values between selected and unselected vectors and protected honey template

No.	A	B	Corr (A, B)
1	Selected vectors for Sugar Template ( $A_j$ )	Protected Honey Template (HT)	0,0506
2	Not-selected vectors for Sugar Template ( $A_j^*$ )	Protected Honey Template (HT)	0.05004

The correlation between the selected feature vectors that we have chosen to build the honey-templates and the honey-templates themselves is near to zero. The same holds for the correlation between the unselected feature vectors and the constructed honey-templates. The Equal Error Rates for the sugar templates and the honey templates are measured for different values of the three tuning coefficients  $w_1$ ,  $w_2$ , and  $w_3$ .

Table 4. EER for sugar and honey templates for different values of  $w_1$ ,  $w_2$ ,  $w_3$ .

$w_1$	$w_2$	$w_3$	EER <sub>Sugar</sub>	EER <sub>Honey</sub>
1	1	1	0.4078	0.4197
1	1	36	0,3993	0,4080
1	1	72	0,3993	0,4084
1	2	6	0.4249	0.4368
1	2	36	0.4214	0.4353
1	2	72	0.4214	0.4357
2	2	6	0.4365	0.4509
2	4	6	0.4668	0.4771
36	6	6	0.4749	0.4854
72	36	6	<b>0.4809</b>	<b>0.4994</b>

Table 2 and 3 give only correlation values between the selected / unselected feature vectors and the protected templates. Furthermore, Table 4 shows the EER of classifying the correlations of the two cases (selected and unselected) for both sugar templates and honey

templates. SVM is used to do the classification training and testing tasks. The optimal value of  $(EER_{\text{Sugar}}, EER_{\text{Honey}})$  is where the triplet  $(w_1, w_2, w_3)$  is equal to  $(72, 36, 6)$ .

## 4. 2 Performance comparison between protected and unprotected case

### *Unprotected case*

The first comparison is between samples of unprotected sugar templates. As a first score we measure the similarity between sugar feature vectors which belong to the same subject. We have called this  $GEN_{(SB,SB)}$ . This stands for genuine score of the sugar feature vector. A second score we have calculated is the similarity between the first samples of the sugar feature vectors. We have called this  $IMP_{(SB,SB)}$  which stands for imposter score of the sugar feature vector. As a similarity measure we have used the inverse of SSD (Sum of Squared Differences). For these scores we have the following equations:

$$GEN_{(SB,SB)} = \frac{1}{\sum_{i=1, j=1}^{i=40, j=10} (SB_i(j) - SB_i(j+1))^2} \quad (8)$$

$$IMP_{(SB,SB)} = \frac{1}{\sum_{i=1}^{i=40} (SB_i(j=1) - SB_{i+1}(j=1))^2} \quad (9)$$

The accuracy of the unprotected case is the Equal Error Rate (EER-unprotected case) between  $GEN_{(SB,SB)}$  and  $IMP_{(SB,SB)}$  and the results are presented in Table 5.

### *Protected case*

The accuracy test of the unprotected case is repeated for the protected templates also. We have two other scores:  $GEN_{(ST,HT)}$  and  $IMP_{(ST,HT)}$  for the EER evaluation. The  $GEN_{(ST,HT)}$  comparison score is between genuine protected templates and the correspondent 15 protected honey templates. We called this parameter “protected genuine score”. It is calculated as:

$$GEN_{(ST,HT)} = \frac{1}{\sum_{i=1, j=1}^{i=40, j=10} (ST_i(j) - HT_i(j+1))^2} \quad (10)$$

The last comparison score is between first samples of protected templates of the same subject. We called this parameter “protected imposter score”.

$$IMP_{(ST,HT)} = \frac{1}{\sum_{i=1}^{i=40} (ST_i(j=1) - HT_{i+1}(j=1))^2} \quad (11)$$

The accuracy of the protected case is calculated as the EER (protected case) between  $GEN_{(ST,HT)}$  and  $IMP_{(ST,HT)}$  and the results are presented in Table 5. Results for both error rates are measured for different values of the tuning parameters triplet  $(w_1, w_2, w_3)$ , where the best result is highlighted.

## 4. 3. Recognition performance evaluation

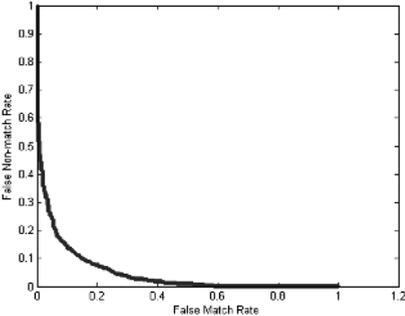
To evaluate the recognition of the proposed BTPS and honey template based biometric

system, we compare the plain template case (without BTPS and without honey templates) and the proposed BTPS and honey template based case. In the plain template case, we use the following testing protocol: for all 40 faces in the testing database  $DB_{\text{test}}$ , we cross compare the 10 PCA feature vectors from each same face and this resulted in 1800 genuine comparison scores; and for all 40 faces we use only the PCA feature vector from the first sample of each face to perform cross comparison and this resulted in 780 imposter comparison scores.

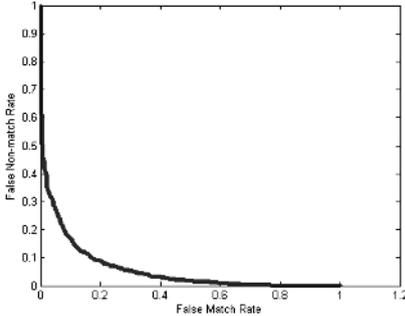
Table 5. EER for unprotected and protected case for different values of  $w_1, w_2, w_3$ .

$w_1$	$w_2$	$w_3$	EER (unprotected case)	EER (protected case)
1	1	1	0.1223	0.3351
1	1	36	0.1223	0.2791
1	1	72	0.1223	0.2801
1	2	6	0.1223	0.2000
1	2	36	0.1223	0.1735
1	2	72	0.1223	0.1580
<b>1</b>	<b>6</b>	<b>36</b>	<b>0.1223</b>	<b>0.1285</b>
2	2	6	0.1223	0.1696
2	4	6	0.1223	0.1975
36	6	6	0.1223	0.2577
72	36	6	0.1223	0.2704

In the proposed BTPS and honey template based case, the highest comparison score out of the 16 comparison scores between the probe’s  $PI^*$  and all 16  $PI$ s in the database is recorded as the final comparison score to verify the probe.



(a) Unprotected templates: EER = 0.1223



(b) BTPS with honey templates: EER = 0.1285

Figure 4. Recognition performance comparison:  $w_1 = 1, w_2 = 6,$  and  $w_3 = 36$ .

From Figure 4 we can see an example of recognition performance comparison between the plain PCA feature vector case and the proposed BTPS and honey template case with the setting  $w_1 = 1, w_2 = 6,$  and  $w_3 = 36$ . We also see performance degradation in other parameter settings. Recognition degradation, as in many other BTPS, can be limited in an acceptable range if the parameters are fine tuned.

## 5 Conclusions and further work

We borrowed the honeywords concept used for detecting leaked passwords in a biometric system and proposed a BTPS based honey template construction method in this paper. The honey biometric protected templates can be used as chaff data to hide the storage address of the sugar (genuine) biometric protected template. Once a honey template is matched by a probe, the system can reasonably conclude with a high probability that the corresponding biometric data entries in the database had been already leaked and a pre-image masquerade attack is launched. This could be very helpful in detecting such data leakage accidents which cannot be achieved by existing biometric template protection schemes.

We proposed in this paper a biometric database construction design and architecture design for a biometric system using such a honey templates idea. It is applicable to verification applications in physical and logical access control, such as ATM, health records, etc. To prove the concept's effectiveness in a practical biometric system, we tested the honey template idea on PCA based face features. Experiments demonstrated the effectiveness of the proposed concepts in both security and recognition performance aspects. An important part of our paper was the security evaluation of our BTPS, focused on the irreversibility of the biometric sugar and honey templates. The low levels of correlations between different feature or template sets show the effectiveness of the scheme.

As a further work we will continue with the system evaluation in terms of unlinkability, as the difficulty of classifying the protected templates over time and across applications [Zh09]. Inuma in [In14] has mathematically proven the relationship between the two notions of irreversibility and linkability in a biometric system. We proved in our paper that if an attacker possesses the set of protected sugar and honey templates, he is not able to recover the feature element and then pretend to be that user. But what if he possesses two sets of protected templates of user  $i$ , coming from two different applications: is the attacker able to identify that these template sets belong to the same characteristic? This process should be computationally hard and we expect to evaluate the notion of linkability in our future work.

We can finally conclude that while use of honey templates in biometrics is still a new direction to explore, we believe both the BTPS method and the honey template construction method have wide room to improve in security and recognition performance aspects in the future. We hope this work can provoke thoughts and discussions in this field.

## 6 Acknowledgement

This research work was partially funded by the European 7<sup>th</sup> Framework Programme project FIDELITY and Competiveness and Innovation Framework Programme project PIDaaS. All information is provided as is and no guarantee or warranty is given that the information is fit for any particular purpose. The user thereof uses the information at its sole risk and liability. The European Commission has no liability in respect of this document, which is merely representing the authors' view.

## References

- [Bo09] Bowen, B.M.; Hershkop, S.; Keromytis, A. D.; Stolfo, S. J.: Baiting inside attackers using decoy documents. In *Security and Privacy in Communication Networks*, pages 51–70, 2009.
- [Ho15] Honeypots. <http://www.honeypots.org/>, accessed: May 2015.
- [In14] Inuma, M.: A relation between irreversibility and unlinkability for biometric template protection algorithms. In *Josai Mathematical Monographs*, vol. 7, 2014, pp. 55-65.
- [ISO11] ISO/IEC 24745 Information technology - Security techniques - Biometric information protection, 2011.
- [JR13] Juels, A.; Rivest, R.: Honeywords: making password-cracking detectable. In *Proceedings of the 2013 ACM SIGSAC Conference on Computer & Communications Security (ACM-CCS'13)*, 2013, pp. 145-160.
- [Ju14] Juels, A.: A bodyguard of lies: the use of honey objects in information security. In *Proceedings of the 19th ACM symposium on Access control models and technologies*. ACM, 2014.
- [JW99] Juels, A; Wattenberg, M.: A fuzzy commitment scheme. In *CCS '99*, ACM, 1999, pp. 28–36.
- [Ma06] Martinez-Diaz, M.; Fierrez-Aguillar, J.; Alonso-Fernandez, F.; Ortega-Garcia, J; Siguenza, A.J.: Hill-climbing and brute-force attacks on biometric systems: A case study in Match-on-card fingerprint verification. In *Proceedings of IEEE Intl. Carnahan Conf. on Security Technology, ICCST, Lexington, USA, 2006*, pp. 151-159.
- [Na09] Nazario, J.: Phoneyc: A virtual client honeypot. In *Proceedings of the 2nd USENIX conference on Large-scale exploits and emergent threats: botnets, spyware, worms, and more*. USENIX Association, 2009.
- [ORL15] The Database of Faces (formerly “The ORL Database of Faces”). <http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html>, accessed: January, 2015.
- [RCB01] Ratha, N.; Connell, J.; Bolle, R.M: Enhancing security and privacy in biometrics-based authentication systems. In *IBM Systems Journal*, 40(3), 2001, pp. 614–634.’
- [Si12] Simoons, K.; Yang, B.; Zhou, X.; Beato, F.; Busch, C.; Newton, E.M.; Preneel, B.: Criteria towards metrics for benchmarking template protection algorithms. In *Proc. of 5th IAPR International Conference on Biometrics (ICB)*, 498-505, 2012.
- [SLM07] Sutcu, Y.; Li, Q.; Memon, N.: Protecting biometric templates with sketch: theory and practice. In *IEEE Transaction on Information Forensics and Security*, 2(3), 2007, pp. 503-512.
- [Sp03] Spitzner, L.: Honeytokens: The other honeypot. In *Symantec Security Focus*, July 2003.
- [TGN06] Teoh, A; Goh, A; Ngo, D.: Random multispace quantization as an analytic mechanism for biohashing of biometric and random identity inputs. In *IEEE Trans. Pattern Anal. Mach. Intell.*, 28(12), 2006, pp.1892–1901.
- [Ya10] Yang, B.; Hartung, D.; Simoons, K.; Busch, C.: Dynamic random projection for biometric template protection. In *Proceedings of the 4<sup>th</sup> IEEE International Conference on Biometrics: Theory Applications and Systems (BTAS)*, 2010.
- [YM15] Yang, B. and Martiri, E.: Using Honey Templates to Augment Hash Based Biometric Template Protection. In *Proc. of the 1<sup>st</sup> International Workshop on Secure Identity Management in the Cloud Environment (SIMICE) at the 39<sup>th</sup> Annual International Computers, Software & Applications Conference (COMPSAC)*, 2015.
- [Zh09] Zhou, X.; Wolthusen, S.; Busch, C.; Kuijper, A.: Feature correlation attack on biometric privacy protection schemes. In *Fifth International Conference on Intelligent Information Hiding and Multimedia Signal Processing, IEEE, 2009*, pp. 1061-1065.

# Contact-less Palm/Finger Vein Biometrics

Alexandre Sierro, Pierre Ferrez, Pierre Roduit

System Engineering  
University of Applied Sciences Western Switzerland  
Route du Rawyl 47  
1950 Sion, Switzerland  
firstname.lastname@hevs.ch

## **Abstract:**

Finger and palm vein recognition, based on near infra-red images of the vein pattern of the finger or the palm, are promising biometric authentication methods. The main advantage of vein recognition over fingerprints is its touch-less nature, making it more robust to spoofing and more comfortable to the user. To this point, vein recognition has mainly been developed by private companies rather than by academic institutions and there are only a relatively limited number of scientific publications on the topic. This paper presents two palm vein and one finger vein imaging prototypes developed in our institution. An image database has also been acquired with each of these three prototypes.

## **1 Introduction**

Fingerprint recognition, available now on several smart phones (e.g. Samsung Galaxy S6, Apple iPhone 6), is one of the few biometric authentication methods available nowadays alongside iris and face recognition. A promising technique, first described in the eighties and subject to academic research since the mid nineties, is vein recognition.

This authentication method is based on the pattern of the blood vessels of a person's hand or finger. This pattern seems not to evolve with time and contains enough discriminant information to be used as a person recognition method. The image of the veins, located about 3 mm under the surface of the skin, is typically acquired using near infra-red (NIR) illumination in reflection mode (Fujitsu [Fuj, WESS05]) or in transmission mode (Hitachi [Hit]). Infra-red LEDs are used to deliver the illumination and the reflected or transmitted light is typically acquired by a CCD or CMOS camera equipped with a filter that only lets infra-red light through. Vein recognition is based on the fact that veins carry haemoglobin absorbing near infra-red light (wavelengths between 700 and 1000 nm). Therefore, when using NIR illumination veins appear as specific and unique black structures that can be used for authentication using standard pattern recognition methods. This authentication method is very interesting and promising, not only because of its non-invasive and user-friendly nature, but also because it is touch-less and therefore doesn't leave marks on the acquisition system, as opposed to finger marks left on the sensor with fingerprint recognition. This

touch-less nature makes vein recognition more robust to spoofing attacks and also more comfortable and hygienic to the user who is then more likely to adopt it. A spoofing attack is a situation in which one person successfully masquerades as another by falsifying data and thereby gaining an illegitimate advantage.

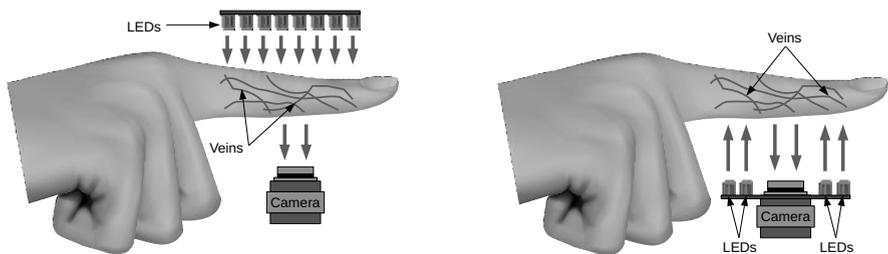
To this point, vein recognition has mainly been developed by private companies rather than by academic institutions. The main reason lies in the difficulty of creating an acquisition system as it requires multi-disciplinary expertise in optics, electronics and computer science. Buying an existing system is not a solution for academic institutions for research purposes given the prohibitive prices, the availability to private companies only and the imposed non-disclosure agreements. As a consequence, there are only a limited number of scientific publications on vein recognition and many questions remain without an answer.

This paper presents two palm vein (one using multi-spectral illumination) and one finger vein acquisition systems developed in our institution, as well as the three related image databases.

## 2 Current State of Research and Development

Since the early 2000s and the work of Hitachi and Fujitsu on finger and palm vein recognition [MNM02, WESS05], the availability of open databases for research remained scarce, as the development of acquisition systems is a difficult task, requiring multiple competencies (optics, electronics, embedded devices). A few acquisition systems were however produced by research teams. For example, Zhang et al. presented in [ZKYW03, ZLYB07] a palm vein system that was successfully used to extract palm structures in the NIR and visible domains. Diestler et al. [DJM<sup>+</sup>11], Lee [Lee12], Kabacinski et al. [KK11] developed their own system to acquire palm vein structures. Moreover, the last cited authors acquired a publicly available image database [CIE]. Other databases of palm vein images are also accessible, such as the Hong Kong Polytechnic University database [MSP] or the National Laboratory of Pattern Recognition database [CAS]. It is noticeable that these last two databases were acquired using multi-spectral illumination. Badawi [Bad06] worked on a closely related system acquiring the vein structure of the back of the hand.

Likewise, several acquisition systems dedicated to finger vein recognition have been developed but only a few accessible databases exist. For example, Ton et al. [TV13] presented such a system. Likewise, Raghavendra et al. [RRSB14] developed a multi-spectral acquisition system to capture finger vein patterns and finger prints. Examples of accessible databases are the Shandong University database [SDU], the Hong Kong Polytechnic University database [HKP], the Chonbuk National University database [MMC], the Sains University database [FV-] and the Idiap Research Institute database [VER]. Finally, Vanoni et al. [VTESM14] presented a review of different databases.



(a) Finger vein acquisition using transmitted light. The NIR LEDs are placed on one side of the finger and the camera is placed on the other side of the finger.

(b) Finger vein acquisition using reflected light. Both the NIR LEDs and the camera are placed on the same side of the finger.

Figure 1: Finger vein recognition using transmitted light (left) and reflected light (right).

### 3 Acquisition Systems

#### 3.1 Working Principles

Most vein imaging systems rely on NIR illumination as the contrast between veins and other tissues is higher than with visible light and they can however rely on cheap CMOS or CCD cameras. Vein recognition systems use either a transmission method or a reflection method.

In the transmission method (Fig. 1a), mostly used to capture finger vein pattern images, the finger is placed between the illumination source and the image sensor. The image sensor captures the light that passed through the finger. As the veins absorb the NIR light, less light will go through the veins and they will appear as a dark pattern.

In the reflection method (Fig. 1b), mostly used to capture palm, back of hand and wrist vein pattern images, the illumination source and the image sensor are located on the same side of the hand. The light enters the finger and is diffused in the tissue. Part of this diffused light is reflected back towards the surface of the finger and is then captured by the image sensor. The image is created by differences in the intensity of the reflected light. As the veins absorb the NIR light, less light is reflected from these areas and the veins will appear as a dark pattern whereas the image will be brighter for the rest of the hand.

Three different prototypes will be presented in this section, the first two dedicated to palm vein recognition and the third one to finger vein recognition. All prototypes use reflection method and are touch-less. The hand is not mechanically guided in front of the camera.



(a) PCB with the 20 LEDs, the camera (middle) and the ultrasound sensor.



(b) The PCB is integrated in a box and covered with the PTFE sheet.

Figure 2: First reflection-based palm vein imaging prototype.

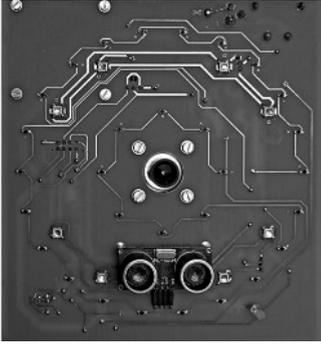
### 3.2 First Palm Vein Prototype

The first palm vein prototype developed was using a unique wavelength for the NIR illumination. Different wavelengths were tested in the NIR, namely 770 nm, 880 nm, and 940 nm. Although 940 nm sits at the limit of the sensibility of the chosen camera (IEEE 1394 Sony ICX618 659x494 CCD camera), this wavelength was selected as it provided the best contrast. The set-up was completed by a 920 nm long-pass filter to block the ambient light that can significantly alter the contrast.

Twenty 940 nm NIR LEDs (TSAL6400) were mounted on a printed circuit board (PCB) for a total power consumption of 2.7 W and a total radiant flux of 700 mW. The LEDs were mounted in the edges of the PCB to provide a uniform illumination of an object situated between 10 and 20 cm in front of the board. Figure 2a shows the PCB. The camera and the long-pass filter can also be seen in the middle of the PCB, as well as an ultrasound distance sensor that was used to cut the lighting when the hand was not in the working range (12.5 to 16.5 cm). To optimize the uniformity of the illumination, a 1mm thick PTFE (Teflon) sheet was added in front of the LEDs. The acquisition device can be seen in Figure 2b.

### 3.3 Second Palm Vein Prototype

As single wavelength acquisition systems can easily be cheated using a vein structure printed on a standard laser printer [TM15], the second prototype was using multi-spectral illumination of the palm vein structure. The increased spoofing difficulty lies in finding or developing an ink responding to the three different wavelengths in the exact same way as human body tissues. The acquisition system is based on three wavelengths, namely blue, far-red, and infra-red. Blue illumination enhance the skin surface structure, whereas far-red and infra-red mainly show the vein structure.



(a) PCB with the 38 LEDs, the camera (middle) and the ultrasound sensor.



(b) The PCB is integrated in a box and covered with the PTFE sheet.

Figure 3: Second reflection-based multi-spectral palm vein imaging prototype.

The ultrasound distance sensor and the CCD camera selected for the first prototype were also used for this second prototype. However, the GigE Vision standard was preferred to IEEE 1394 for more efficient communication and control between the prototype and the computer. Figure 3b shows the second prototype.

**Illumination Optimization** To optimize the image contrast and the uniformity of the illumination, a simulation of the illumination depending on the position of the LEDs was performed. The optimal positioning of the infra-red LEDs consists in two concentric circles of respectively 14 cm and 7 cm of diameter containing 22 NIR LEDs and 8 NIR LEDs evenly positioned, respectively. The CCD camera lies at the center of these concentric circles. As the use of three wavelengths made impossible the inclusion of a long-pass filter to suppress ambient light, the illumination power was instead considerably increased. Four blue LEDs (LZ100B200, total radiant flux of 3.9 W and total power of 13.3 W), four far-red LEDs (LZ100R300, total radiant flux of 2.8 W and total power of 9.6 W) and thirty NIR LEDs (SFH4046, total radiant flux of 1.2 W and total power of 3.4W) were used for the multi-spectral illumination. The acquisition cycles through the three wavelengths at fifteen frames per second. Figure 3a shows the PCB with the different LEDs, the camera and the ultrasound sensor.

Figure 4 shows three images of the same palm captured at the different wavelengths. With the blue illumination (Figure 4a), the main visible feature is the skin structure, whereas with the far-red and the NIR illumination (Figures 4c and 4b), the main feature is the vein pattern.

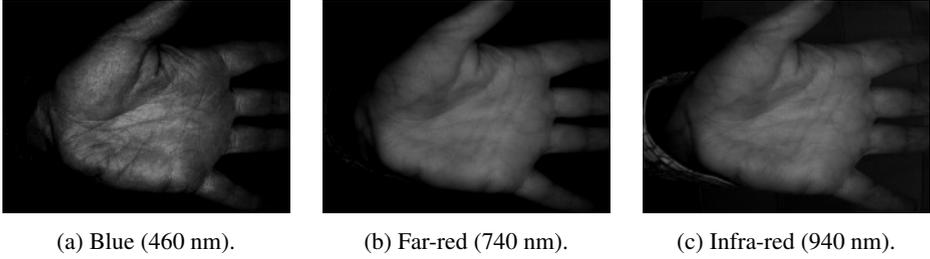


Figure 4: Palm images of the same hand using the multi-spectral acquisition system.

### 3.4 Finger Vein Prototype

When using the reflection method, the contrast is usually weaker on finger vein images than on palm vein images. This is due to the fact that finger veins are located deeper than palm veins. The finger vein prototype is therefore optimized to deliver a very homogeneous illumination to guarantee quality finger vein images. The prototype was also developed for mobile applications. Indeed, the small color image sensor is similar to those found in smart phones, the control software is running on a smart phone (Android) and interacts with the prototype through USB communication. The dimensions of the prototype are 52 mm x 114 mm x 47 mm. Figure 5b shows the finger vein prototype.

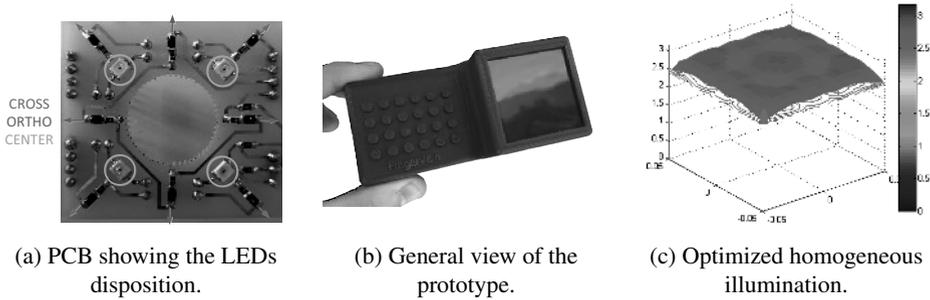


Figure 5: Reflection-based finger vein imaging prototype.

**Illumination** The illumination is based on the same principles described in Paragraph 3.3 but with additional LED orientation optimization. Due to the relatively small dimensions of the prototype, the illuminated area is larger than the PCB surface. To avoid saturation in the central spot and insufficient illumination on the edges, some LEDs are not mounted flat on the PCB so that not all illumination beams are perpendicular to the PCB plane.

The 850 nm LEDs are organized in three groups of four LEDs, as shown in Figure 5a. The central group provides global illumination using wide angle 120° VSMG3700 LEDs. The two other groups, one positioned orthogonally (+) and one in cross (×), compensate to provide an optimized homogeneous illumination using 20° SFH4059 LEDs. These

last eight LEDs are orientated away from the center and mounted on the PCB with a 20° angle with respect to the plane of the PCB. The resulting optimized illumination is shown in Figure 5c. This optimized LED disposition and orientation provides by itself a very homogeneous illumination. Therefore, the use of a diffusing sheet as for the palm vein prototypes previously presented is unnecessary.

The power of each LED group is adjustable by software allowing fine tuning of the illumination. The optimal illumination is reached with a relative power of 1.0 for the central LEDs and of 0.3 for the orthogonal and cross LEDs. The central LEDs have a total radiant flux of 160 mW and a total power of 600 mW whereas the orthogonal and cross LEDs have a total radiant flux of 320 mW and a total power of 900 mW.

**Camera** The imaging device used for the finger vein prototype is a low-cost OV7670 Color 640x480 pixel CMOS sensor coupled with a wide angle 2.1 mm lens allowing image acquisition from a minimal distance of 10 cm. An infra-red long-pass filter is used to filter out visible light. A relatively low cut-off wavelength of 740 nm has been chosen to allow further modification of the illumination wavelength.

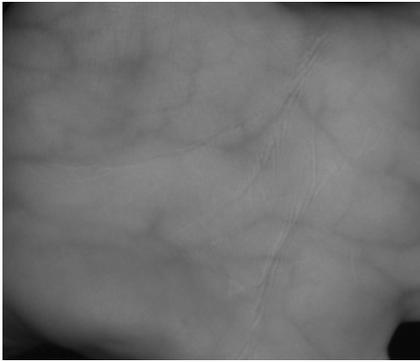
This finger vein prototype shows that the Bayer pattern used in most single-chip digital image sensor to create color images, does not significantly alter infra-red vein imaging. This means that for a future integration of a similar miniaturized system on a smart-phone or a tablet, the embedded camera can be used for normal photography as well as for finger vein authentication. Smart phones are currently equipped with an infra-red filter to prevent inaccurate colors in images caused by NIR light naturally present in ambient light. It is possible to get around this problem by using a dual band-pass filter passing all visible light as well as a specific narrow NIR band.

## 4 Databases

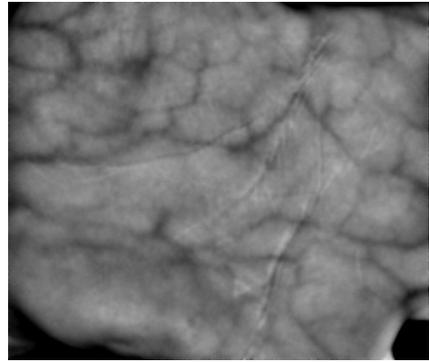
A publicly available finger or palm vein image database has been captured with each of the three prototypes presented in this paper. This section presents the details of these three databases. The second (multi-spectral palm vein) database and the third (finger vein) database were acquired simultaneously. Further research combining and/or comparing palm vein and finger vein is therefore possible using these two databases.

### 4.1 Palm Vein Database

The palm vein database for palm vein recognition consists of 2200 images from 110 subjects. This database was acquired in our premises using the first prototype described in Section 3.2. Each of the 110 subjects presented their both hands and five palm images were recorded per hand. Each subject participate in two recording sessions. The database consists therefore of 2200 palm vein images. (110 subjects x 2 hands x 2 sessions x 5



(a) Raw image.



(b) Processed image.

Figure 6: Raw (left) and processed (unsharp mask, right) palm vein images (after extraction of the region of interest) of the same hand using the first palm vein prototype (940 nm).

images = 2200 images). The database is composed of 40 women and 70 men whose ages are between 18 and 60 with an average of 33. The format of the images is PNG and the image resolution is 640 x 480.

Figure 6a shows a typical raw palm vein image after extraction of the region of interest (ROI). Figure 6b shows the same image after applying a simple unsharp mask to enhance the contrast and therefore better show the vein pattern. The ROI is extracted according to the following procedure. First a threshold (value 50) is applied to the 8-bit raw images. This basically turns the hand to white and the background to black. Then the contours of the hand are extracted using standard edge detection. Only the surface with the maximum area (the hand) is kept, the other possibly existing surfaces are discarded. The next step consists of a closing of the image (about 80 iterations) to split the palm and the fingers. A rectangular boundary box is then applied to the remaining palm surface and is accepted as ROI if it contains more than 20'000 pixels and less than 240'000 pixels.

## 4.2 Multi-spectral Palm Vein Database

The second palm vein database for palm vein recognition consists of 2520 images from 84 subjects. The database was acquired in our premises using the multi-spectral prototype described in Section 3.3.

Each of the 84 subjects presented both their hands and five images were recorder per hand at each of the three wavelengths. Each subject participated in a single recording session. The database consists therefore of 2520 palm vein images (84 subjects x 2 hands x 3 wavelengths x 1 session x 5 images = 2520 images). The database is composed of 24 women and 60 men whose ages are between 15 and 61 with an average of 32. The format of the images is PNG and the image resolution is 656 x 490.

Figure 7 shows a typical set of images of this second database after extraction of the region of interest. The top row images (a, b and c) show the raw images acquired at 460 nm, 740 nm and 940 nm, respectively. The bottom row shows the same images after applying the same unsharp mask mentioned in Section 4.1. The region of interest (ROI) is extracted following the technique described in Section 4.1.

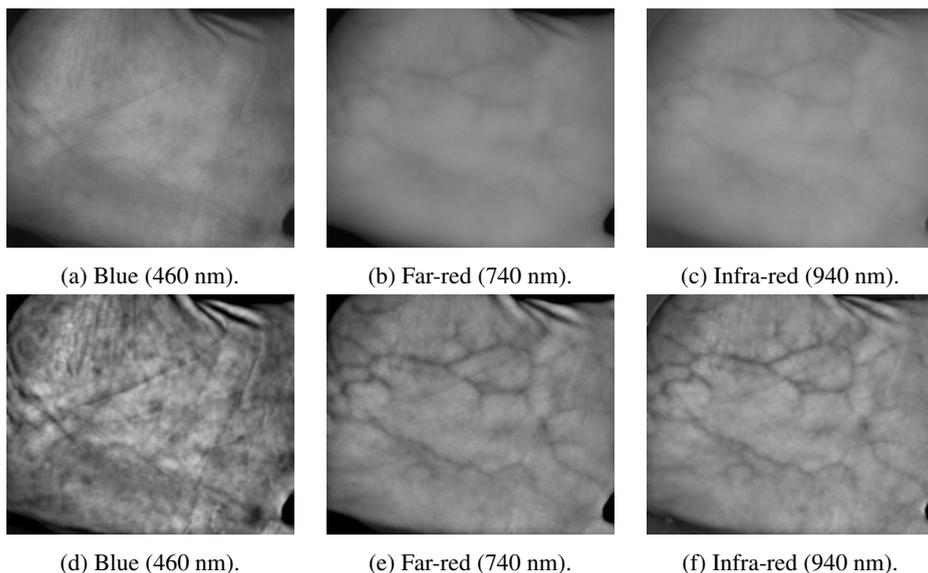


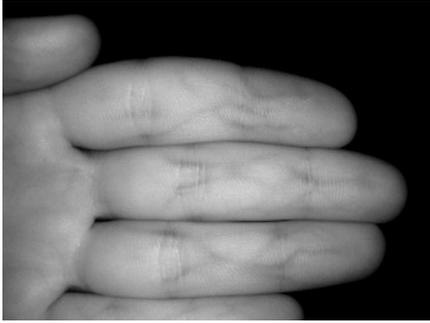
Figure 7: Raw (top row) and processed (bottom row) palm vein images (after extraction of the region of interest) of the same hand using the multi-spectral acquisition system.

### 4.3 FingerVein Database

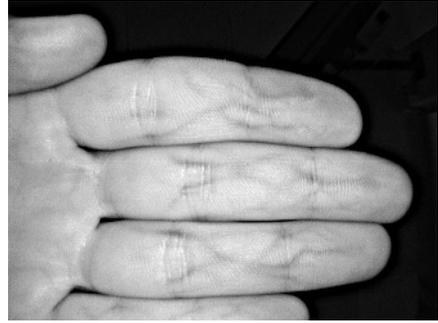
The finger vein database for finger vein recognition consists of 840 images from 84 subjects. The database was acquired in our premises using the finger vein prototype described in Section 3.4.

Each of the 84 subjects presented both their hands and five images were recorded per hand. Each subject participated in a single recording session. The database consists therefore of 840 palm vein images (84 subjects x 2 hands x 1 session x 5 images = 840 images). The gender and age statistics of this database is identical to those of the multi-spectral palm vein database as it was acquired with the same subjects. The format of the images is PNG and the image resolution is 640 x 480.

Figure 8a shows a typical raw finger vein image after extraction of the region of interest (ROI). Figure 8b shows the same images after applying the same unsharp mask mentioned in Section 4.1.



(a) Raw image.



(b) Processed image.

Figure 8: Raw (left) and processed (unsharp mask, right) finger vein images of the same hand using the finger vein prototype (850 nm).

## 5 Conclusion

This paper presents three different palm or finger vein imaging prototypes developed in our institution in the framework of previous projects. The vein image databases acquired with each of these three prototypes are also described in details in this paper. These databases are also publicly available. The qualitative quality of the acquired palm vein images is similar to the one of images in scientific publications or in publicly available databases. However, they were collected with European subjects, contrary to most other publicly available databases that were collected with Asian subjects. The difference in skin color and thickness of these two general ethnic groups could have an impact on recognition algorithms. Furthermore, the presented databases show a wide variety of skin surface and thickness due to the age of the subjects (15 to 60) and their job (e.g. office, mechanical workshop). Finally, the quality of the acquired finger vein images is difficult to assess, as no other publicly available database is based on reflected light.

### 5.1 Future Work

Several research trails could be followed in the future. These prototypes can always be improved, especially in terms of illumination quality and homogeneity as well as robustness to environmental conditions to provide the best finger or palm vein images possible. In parallel to the improvement of the hardware, our research team will soon start deep investigations on the software development. This includes algorithm development for raw image processing, feature selection, feature extraction and finally classification to turn the image acquisition prototypes into fully operational finger or palm vein authentication systems. The presented multi-spectral palm vein and finger vein databases, acquired simultaneously, will be extended to reach at least one hundred subjects.

Another future topic will consist in investigating the feasibility of a reliable multi-spectral

low-cost mobile finger vein authentication system that could be embedded or plugged to a smart phone or a tablet. Several challenges will have to be addressed during the development of such a system. For example, for volume reduction, the system will work with reflected light rather than transmitted light. The system will still have to provide good quality images and perform well in uncontrolled environments to be accepted as an alternative authentication method while consuming a minimum of energy for a minimum impact on the battery of a mobile device.

## References

- [Bad06] A.M. Badawi. Hand Vein Biometric Verification Prototype: A Testing Performance and Patterns Similarity. In *Proceedings of the 2006 International Conference on Image Processing, Computer Vision, & Pattern Recognition*, pages 3–9, 2006.
- [CAS] CASIA Multi-Spectral Palmprint Database. <http://biometrics.idealtest.org/dbDetailForUser.do?id=6>.
- [CIE] Vein Dataset. <http://biometrics.put.poznan.pl/vein-dataset>.
- [DJM<sup>+</sup>11] M. Distler, S.H.N. Jensen, N.G. Myrtue, C. Petitimberty, K. Nasrollahi, and T.B. Moeslund. Low-Cost Hand Vein Pattern Recognition. <http://vbn.aau.dk/files/63418244/CSIP.pdf>, 2011.
- [Fuj] PalmSecure. <http://www.fujitsu.com/us/services/biometrics/palm-vein/>.
- [FV-] FV-USM database. [http://blog.eng.usm.my/fendi/?page\\_id=262](http://blog.eng.usm.my/fendi/?page_id=262).
- [Hit] Finger Vein Authentication Technology. <http://www.hitachi.eu/veinid/>.
- [HKP] HKPU Finger Image Database. <http://www4.comp.polyu.edu.hk/~csajaykr/fvdatabase.htm>.
- [KK11] R. Kabacinski and M. Kowalski. Vein pattern database and benchmark results. *Electronics Letters*, (20):1127–1128, 2011.
- [Lee12] J.-C. Lee. A novel biometric system based on palm vein image. *Pattern Recognition Letters*, 33(12):1520–1528, 2012.
- [MMC] Finger Vein Database: MNCBNU 6000. [http://multilab.jbnu.ac.kr/MNCBNU\\_6000/](http://multilab.jbnu.ac.kr/MNCBNU_6000/).
- [MNM02] N. Miura, A. Nagasaka, and T. Miyatake. Automatic Feature Extraction from non-uniform Finger Vein Image and its Application to Personal Identification. In *Proceedings of the IAPR Workshop on Machine Vision Applications*, pages 253–256, 2002.
- [MSP] Multispectral Palmprint Database. <http://www4.comp.polyu.edu.hk/~biometrics/MultispectralPalmprint/MSP.htm>.
- [RRSB14] R. Raghavendra, K.B. Raja, J. Surbiryala, and C. Busch. A low-cost multimodal biometric sensor to capture finger vein and fingerprint. In *Biometrics (IJCB), 2014 IEEE International Joint Conference on*, pages 1–7, 2014.
- [SDU] SDUMLA-HMT Database. <http://mla.sdu.edu.cn/sdumla-hmt.htm>.

- [TM15] P. Tome and S. Marcel. On the Vulnerability of Palm Vein Recognition to Spoofing Attacks. In *The 8th IAPR International Conference on Biometrics (ICB)*, 2015.
- [TV13] B.T. Ton and R.N.J. Veldhuis. A high quality finger vascular pattern dataset collected using a custom designed capturing device. In *International Conference on Biometrics*, pages 1–5, 2013.
- [VER] VERA Fingervein Database. <https://www.idiap.ch/dataset/vera-fingervein>.
- [VTESM14] M. Vanoni, P. Tome, L. El Shafey, and S. Marcel. Cross-database evaluation using an open finger vein sensor. In *Biometric Measurements and Systems for Security and Medical Applications (BIOMS) Proceedings, 2014 IEEE Workshop on*, pages 30–35, 2014.
- [WESS05] M. Watanabe, T. Endoh, M. Shiohara, and S. Sasaki. Palm vein authentication technology and its application. In *Proceedings of The Biometric Consortium Conference*, 2005.
- [ZKYW03] D. Zhang, W.-K. Kong, J. You, and M. Wong. Online Palmprint Identification. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 25(9):1041–1050, 2003.
- [ZLYB07] Y.-B. Zhang, Q. Li, J. You, and P. Bhattacharya. Palm vein extraction and matching for personnal authentication. *Lecture Notes in Computer Science*, pages 154–164, 2007.

# A Fingerprint Indexing Scheme with Robustness against Sample Translation and Rotation

Guoqiang Li, Bian Yang, Christoph Busch

Norwegian Biometric Laboratory, Gjøvik University College, Norway  
guoqiang.li@hig.no, bian.yang@hig.no, christoph.busch@hig.no

**Abstract:** Automatic fingerprint identification systems (AFIS) are getting prevalent around the world, and the size of fingerprint databases involved in AFIS is continuously growing. Thus, studying fingerprint indexing algorithms is desirable in order to facilitate the search process in a large-scale database. In this paper, we firstly propose a feature extraction method to generate a binary template based on minutia information. A fingerprint indexing is designed by combining this binary template and Locality Sensitive Hashing indexing algorithm developed in a state-of-the-art fingerprint indexing method (minutia cylinder-code based indexing method). Experiments have been conducted on several public databases with different settings. The results show that the proposed approach achieves competitive performance or even better performance when benchmarked to the state-of-the-art fingerprint indexing methods.

## 1 Introduction

Fingerprint recognition system has been increasingly gaining attention around the world. Many systems have been deployed such as FBI's Integrated Automated Fingerprint Identification System (IAFIS), the European Visa Information System (VIS) and many other systems are under construction. Depending on the application context, there are two types of fingerprint recognition systems [MMJP09]

**Verification system:** this system carries out a one-to-one comparison to verify whether the identity is the person who he/she claims. A typical scenario is to compare the fingerprint data stored in a European passport with the fingerprint data captured from the subject, who holds the passport.

**Identification system:** this system identifies a person by searching the whole database, which results in a one-to-N comparison process. A typical scenario is to check whether a criminal suspect has been recorded in FBI's IAFIS by using his/her probe fingerprint samples for the query.

According to the information published on FBI's website [FBI], the FBI IAFIS contains enrolled fingerprint from more than 100 million subjects. Thus it is almost impossible

---

This work is funded by the EU 7th Framework Program under grant agreement  $n^{\circ}$  284862 for the large-scale integrated project FIDELITY.

to conduct a one-by-N comparison in such a large-scale database. Therefore, studying fingerprint indexing techniques is desirable in order to reduce the number of candidate identities, which will be further considered by a verification algorithm [LBA07].

A variety of fingerprint indexing approaches have been presented in the literature. Extracting appropriate features for building indexing tables is the core of a fingerprint indexing approach. The approaches in the literature can be grouped into three categories based on their feature extraction methods: local feature based – primarily focusing on using minutia [RM07, CFM11] or local ridge information [LBA07, BRAC08]; global feature based – using orientation field and singular points as a global reference points [WHP07]; other feature based – such as using symmetric filters [LYW06] or scale invariant feature transformation (SIFT) [SZH08].

Instead of exploring global features, we are focusing on only using minutia location and direction to extract a compact feature vector, since minutia information has been recognized as most reliable and basic feature representing fingerprints, and a standardized definition of this feature vector is given by ISO/IEC 19794-2:2011 [ISOa]. In addition, the majority of existing fingerprint indexing approaches generate the features with real values, which might lead to more computational complexity comparing to binary features. Thus we explore to extract a binary feature vector in this paper, and further build the indexing tables by using Locality Sensitive Hashing (LSH) indexing method which was developed by Cappelli et al. [CFM11, CFM15] and has been proven to be suitable for binary feature vectors.

The remainder of this paper is structured as follows: Section 2 introduces the proposed feature extraction method; the details of creating indexing tables and candidates retrieval are described in Section 3; Section 4 reports the experimental results under different settings; the conclusion is drawn in Section 5.

## **2 Feature Extraction method for Fingerprint Indexing**

The feature extraction method is the critical component in a fingerprint indexing scheme due to the fingerprint sample variations caused at acquisition stage. The proposed feature generation method generates a set of fixed-length binary vectors for a fingerprint template. The number of binary vectors for a fingerprint template depends on the minutiae' number in this template. The proposed feature extraction method consists of three stages: local alignment, training and binary vectors generation. The details will be discussed in the following subsections.

### **2.1 Local Alignment and Quantization**

Instead of detecting a singular point or considering the ridge information surrounding the minutia, we focus on using a local alignment concept to extract a binary vector, which can represent this local area. The basic idea of the local alignment is considering each

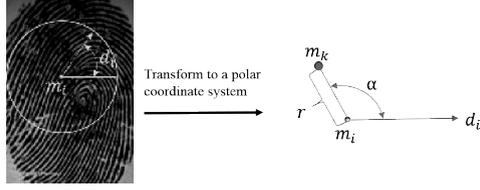


Figure 1: Local alignment: all minutiae included in the yellow circle are aligned with the central minutia in a polar coordinate system whose reference point is  $m_i$  and reference angle is the direction (denoted by  $d_i$ ) of  $m_i$ . This central minutia and another minutiae included in the circle are named as a minutiae-disk.

minutia as a reference point, and then nearby minutiae are aligned with respect to this reference point (called central minutia). As illustrated in Figure 1, all minutiae included in the yellow circle are aligned with the central minutia in a polar coordinate system. This central minutia and another minutiae included in the yellow circle are defined as a minutiae-disk.

We assume a fingerprint template  $T$  including  $n$  minutiae  $\{m_1, m_2, \dots, m_n\}$ , and each minutia comprises three properties:  $m_i(x, y, d)$ , where  $x$  and  $y$  are the minutia location and  $d$  is the minutia direction. A minutiae-disk ( $MD_i$ ) can be formed for each minutia  $m_i$ . A polar coordinate system is defined by using  $m_i$  as reference point and the direction (denoted by  $d_i$ ) of  $m_i$  as reference angle. Then each minutia  $m_k$  included in the minutiae-disk will have a new coordinate  $m'_k(r', \alpha')$  denoted in Equation (1) and (2).

$$r' = DIS(m_k, m_i) \quad (1)$$

where  $DIS$  is Euclidean distance between the two minutiae.

$$\alpha' = \frac{(atan2(m_k(y) - m_i(y), m_k(x) - m_i(x)) + 2\pi) * 180}{\pi} \quad (2)$$

where function  $atan2$  is 'Four-quadrant inverse tangent' defined in [ata].

In addition, the minutiae angle difference  $\theta'$  between  $m_i$  and  $m_k$  is denoted by the following equation:

$$\theta' = |m_k(d) - m_i(d)| \quad (3)$$

In order to further tolerate the sample variation, three attributes ( $r'$ ,  $\alpha'$ ,  $\theta'$ ) are quantified by using Equation (4)~(6).

$$r = floor(r'/5) \quad (4)$$

$$\alpha = floor(\alpha'/5) \quad (5)$$

$$\theta = floor(\theta'/5) \quad (6)$$

where, function  $\text{floor}(X)$  returns the nearest integer less than the variable  $X$ .

Eventually, an aligned minutia  $m'_k$  with three attributes  $(r, \alpha, \theta)$  is created. Since the proposed feature generation method applies local alignment for each minutia, this indicates that each minutiae-disk will be associated with a minutia which is called central minutia. The radius of the minutiae-disk is denoted as  $R$ .

## 2.2 Training and Binary Vectors Generation

A training step is required in the proposed feature extraction method prior to the binary vector generation. The unsupervised learning scheme  $K - \text{means}$  is chosen for this training step, since it has been proven to be appropriate for fingerprint indexing by other researchers [RM07, LYB14]. The input of  $K - \text{means}$  is a set of  $(r, \alpha, \theta)$  vectors generated from the training samples.  $K - \text{means}$  classifies these  $(r, a, \theta)$  vectors into  $K$  clusters, and each cluster is represented by its centroid  $\{C_1(r, \alpha, \theta), C_2(r, \alpha, \theta), \dots, C_K(r, \alpha, \theta)\}$ .

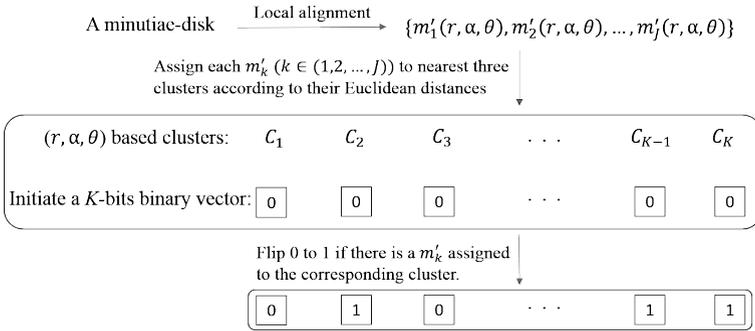


Figure 2: Procedures of generating the binary vector for a minutiae-disk.

The proposed feature extraction method generates a fixed-length binary vector for each minutiae disk. Since one minutia will form one minutiae disk, the number of binary vectors for a template is equal to the number of minutiae in this template. Figure 2 illustrates the procedures of generating the binary vector for a minutiae-disk. There are four steps involved in this process:

**Step 1:** Apply local alignment on this minutiae-disk to generate alignment minutiae:  $m'_1(r, \alpha, \theta), m'_2(r, \alpha, \theta), \dots, m'_j(r, \alpha, \theta)$ .

**Step 2:** Initiate a  $K$  bits binary vector with all components set to 0.

**Step 3:** Assign each  $m'_k(r, \alpha, \theta), k \in (1, 2, \dots, j)$  to nearest three clusters (closest cluster, second closest cluster and third closest cluster) according to their Euclidean distances.

**Step 4:** Flip 0 to 1 if there is a  $m'_k$  assigned to the corresponding cluster. Eventually, a binary vector is generated to represent this minutiae-disk.

Note that only the first change will take effect even if multiple  $m'_k$  have been assigned to the same cluster. The reason of choosing nearest three clusters is to tolerate sample intra-class variations.

### 3 The Indexing Algorithm

Cappelli et al. [CFM11] have proofed that Locality Sensitive Hashing (LSH) is suitable to index the binary vector. We follow their techniques proposed in paper [CFM11] to build the indexing tables and retrieve the candidates by using our newly generated binary vectors. The following subsections give the details of indexing tables creation and candidates retrieval.

#### 3.1 Creating Indexing Tables

---

**Algorithm 1 .** Indexing tables creation

---

**Require:** Minutiae templates of enrolled subjects:  $\{T_1, T_2, \dots, T_E\}$ ;

Hash functions:  $\{f_{H_1}, f_{H_2}, \dots, f_{H_\Lambda}\}$  ( $\Lambda$  is the number of hash functions);

**Ensure:** Indexing tables:  $H_1, H_2, \dots, H_\Lambda$

```

1: for each template  $T_i (i \in 1, 2, \dots, E)$  do
2:   Generate binary template from minutia template by using proposed feature extraction method:  $\{T(i, 1), T(i, 2), \dots, T(i, J)\}$  ( $J$  is the number of binary vector generated from minutiae temple  $T_i$ )
3:   for each binary vector  $T(i, j) (j \in 1, 2, \dots, J)$  do
4:     for each hash function  $f_{H_\lambda}$  do
5:        $b = f_{H_\lambda}(T(i, j))$ 
6:       if  $CountOneBits(b) \geq min_{OneBits}$  then
7:         record  $(i, j)$  in  $b - th$  bucket of indexing table  $H_\lambda$ .
8:       end if
9:     end for
10:  end for
11: end for

```

---

Before describing the algorithm of creating LSH-based indexing tables for fingerprint templates, it is necessary to introduce the techniques of LSH indexing method. Figure 4 gives an example of creating indexing tables by using a set of hash functions  $\{f_{H_1}, f_{H_2}, \dots, f_{H_\Lambda}\}$ , where the number of hash functions is  $\Lambda = 3$ , and the number of bits selected by each hash function is  $\eta = 3$ . Let assume there is a binary  $(T_1, V_1)$  which denotes the first binary vector of the first fingerprint template. Each hash function will randomly select 3 bits from  $(T_1, V_1)$ , then calculate the decimal value based on selected bits and store the pair  $(1, 1)$  in a corresponding bucket. For instance, the decimal value calculated from  $f_{H_2}$  is 3, then  $(1, 1)$  will be stored in the third bucket in hash table  $H_2$ . The number of hash tables is equal

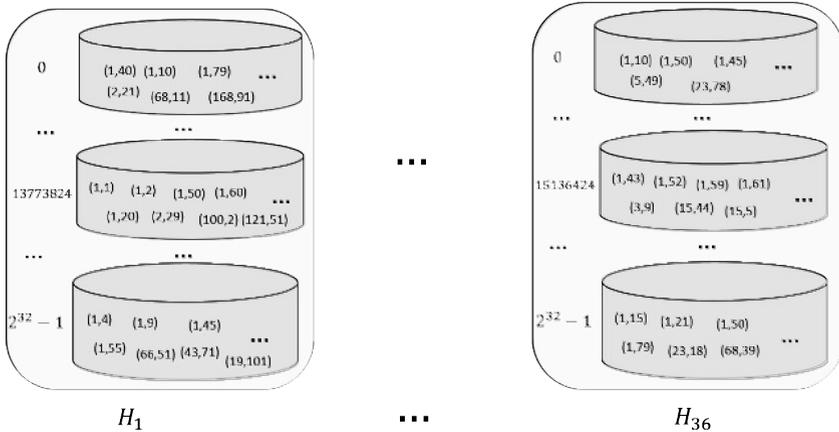


Figure 3: An example of created indexing tables, where the pair  $(i, j)$  indicates the  $j$ -th binary vector of  $i$ -th fingerprint template.

to the number of hash functions, and the number of buckets in each hash table is  $2^n$ .

During indexing tables creation stage, we apply the similar procedures illustrated in Figure 4 to enrol fingerprint minutia template. **Algorithm 1** gives the details of creating indexing tables for a set of fingerprint minutia templates:  $\{T_1, T_2, \dots, T_E\}$ . The first step of enrolling these minutia template is to generate the binary template by using proposed feature extraction method. The function  $CountOneBits(b)$  is to count the number of 1 bits in selected bits, for instance  $CountOneBits(1010001) = 3$ . The pair  $(i, j)$  will be recorded only when  $CountOneBits(b)$  is not less than a parameter  $minOneBits$ . Figure 3 gives an example of the indexing tables after completing enrolment. In addition, the original minutiae templates need to be stored somewhere else (minutiae template can be indexed by their template ID), since they will be used during candidate retrieval stage.

### 3.2 Candidates Retrieval

**Algorithm 2** lists the procedures of retrieving candidates for a probe sample  $P$ . The same hash functions used in enrolment are used as input for candidates retrieval. Another inputs are: indexing tables  $\{H_1, H_2, \dots, H_\Lambda\}$ , enrolled minutiae templates  $\{T_1, T_2, \dots, T_E\}$  as well as the minutia template of the probe sample  $P$ . The function ‘ $Mated(m_\omega, m(i, j))$ ’ is to measure whether minutia  $m_\omega$  from probe sample and minutia  $m(i, j)$  from the reference sample meet a pre-defined geometric constraint. If they satisfy the geometric constraint, the similarity score between this probe sample and the reference sample will increase 1. ‘ $Mated(m_\omega, m(i, j))$ ’ is defined in Equation (7). In order to reduce the computational complexity, we don’t normalize the similarity score which is different to the candidates retrieval method used in paper [CFM11].

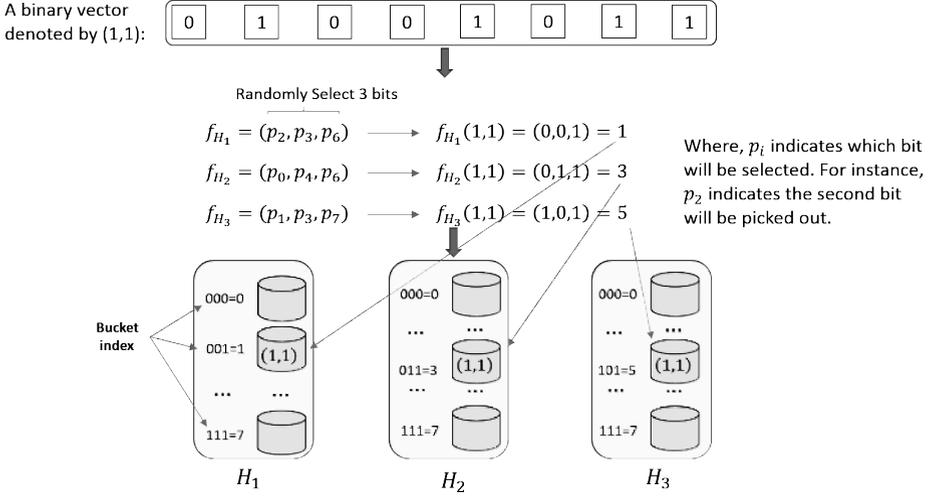


Figure 4: An example of Locality Sensitive Hashing (LSH) indexing algorithm.

$$Mated(m_\omega, m(i, j)) = \begin{cases} true & \text{if } DIS((m_\omega, m(i, j)) \leq \rho \text{ and } |m_\omega(d) - m(i, j)(d)| \leq \sigma \\ false & \text{otherwise.} \end{cases} \quad (7)$$

where,  $|m_\omega(d) - m(i, j)(d)|$  is the direction difference between two minutiae.

## 4 Experimental Settings and Results

In order to evaluate the performance of proposed indexing approach, a couple of experiments have been conducted on several public databases. In accordance with ISO/IEC 19795-1 [ISO], the performance of fingerprint indexing algorithm is reported by two criteria: penetration rate and pre-selection error rate. Penetration rate is a proportion of enrolled references in a database where the identification system has to search. A pre-selection error occurs when the enrolled reference corresponding to the probe sample is not included in the pre-selected candidates. Generally speaking, the better fingerprint indexing approach will achieve lower pre-selection error rate at the same penetration rate comparing to other approaches. The minutia cylinder-code based indexing method (shortly called MCC-Index) [CFM11] was used as benchmark under same protocol in our experiments.

---

**Algorithm 2 .** Candidates retrieval

---

**Require:** Indexing tables:  $H_1, H_2, \dots, H_\Lambda$ ;

Hash functions:  $\{f_{H_1}, f_{H_2}, \dots, f_{H_\Lambda}\}$ ;

Minutiae template of enrolled subjects:  $\{T_1, T_2, \dots, T_E\}$ ;

Minutiae template of probe sample:  $P$ .

**Ensure:** Candidate entities.

- 1: Generate the binary template for probe sample:  $V_1, V_2, \dots, V_\Omega$  ( $\Omega$  is the number of binary vectors);
  - 2: Initiate an array to store similarity score:  $S[E]$ ;
  - 3: **for** each binary vector  $V_\omega$  **do**
  - 4:   Assume  $m_\omega$  is the central minutia associated with binary vector  $V_\omega$
  - 5:   **for** each hash function  $f_{H_\lambda}$  **do**
  - 6:      $b = f_{H_\lambda}(V_\omega)$
  - 7:     **if**  $CountOneBits(b) \geq minOneBits$  **then**
  - 8:       **for** each pair  $(i, j)$  in  $b$ -th bucket of indexing table  $H_\lambda$  **do**
  - 9:         Assume  $m(i, j)$  is the central minutia associated with the pair  $(i, j)$ ;
  - 10:        **if**  $Mated(m_\omega, m(i, j)) == true$  **then**
  - 11:          $S[i] = S[i] + 1$ ;
  - 12:        **end if**
  - 13:       **end for**
  - 14:     **end if**
  - 15:    **end for**
  - 16: **end for**
  - 17: Sort  $S[E]$  by descending order, and select the top- $N$  as candidate entities.
- 

#### 4.1 Databases Preparation and Common Settings for All Experiments

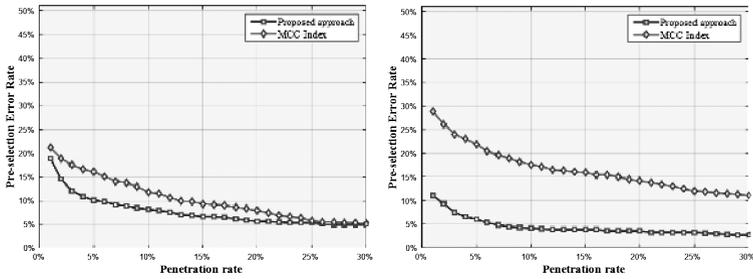
Several *FVC* databases are selected for the experiments: *FVC2002* [MMJP09], *FVC2004* and *FVC2006* [CFFM07]. The details for the respective database will be described in the following sections. The minutia templates were extracted by a commercial product ‘NeuroTechnology Verifinger extractor 6.0’ [Neu]. The experimental results of MCC-Indexing method were generated by MCC sdk v1.4 [CFM10, CFM11, FMC12]. And Table 1 lists the settings of some parameters used for all experiments.

#### 4.2 Experiments on *FVC2002*

We run the experiments on *FVC2002\_DB1* and *FVC2002\_DB2* respectively. There are two subsets in *FVC2002\_DB1* as well as in *FVC2002\_DB2*. We used *FVC2002\_DB1\_B* consisting of 80 samples as a training set for the test set *FVC2002\_DB1\_A* which comprises 800 sample from 100 fingers. The first sample of each finger was enrolled in the indexing tables and the rest of samples were used for searching, since the quality of first sample is relatively better than the rest of samples in *FVC2002*. The similar settings

Parameter	value	Remark
$R$	300 pixels	the radius of the minutiae-disk
$K$	1024	the length of binary vector
$\Lambda$	48	the number of hash functions
$\eta$	32	the number of bits selected by hash function
$\rho$	256	minutia distance threshold
$\sigma$	45	minutia direction difference threshold
$min_{OneBits}$	2	the number of '1' bits in a binary vector

Table 1: Parameters setting for all experiments.



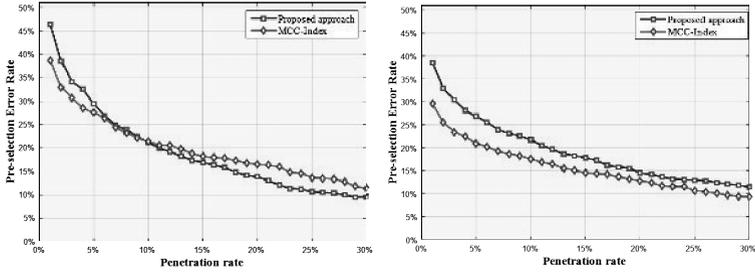
(a) Experiment on *FVC2002\_DB1\_A* (b) Experiment on *FVC2002\_DB2\_A*

Figure 5: Performance evaluation on *FVC2002\_DB1\_A* and *FVC2002\_DB2\_A*

were applied on *FVC2002\_DB2*: *FVC2002\_DB2.B* was used for the training set; *FVC2002\_DB2.A* was used for the test set; the first sample was used for enrolment, and the rest of samples were used for probe samples. Figure 5 demonstrates the performance running experiment on *FVC2002\_DB1* a *FVC2002\_DB2*. The figures show the significant improvements of proposed approach on these databases.

### 4.3 Experiments on *FVC2004*

In order to establish similar settings as for the experiments on *FVC2002*, we used the *FVC2004\_DB1.B* as a training set for the test set *FVC2004\_DB1.A*, and used the *FVC2004\_DB2.B* as a training set for the test set *FVC2004\_DB2.A*. Again we enrolled the first sample of each finger to the indexing tables as we did for *FVC2002*. Figure 6 shows that the MCC-Index method outperformed our proposed approach. Then we observed the sample images of *FVC2004*. We found that the first sample of each finger doesn't have higher quality, and even it can be seen as partial fingerprint comparing other sample from the same finger as seen in Figure 7. This 'partial sample' trait might have more impact on the proposed approach than MCC-Index method, since the radius of



(a) Experiment on *FVC2004\_DB1\_A* (b) Experiment on *FVC2004\_DB2\_A*

Figure 6: Performance evaluation on *FVC2004\_DB1\_A* and *FVC2004\_DB2\_A*: the **first sample** of each subject was enrolled in indexing tables.

minutiae-disk is 300 pixels in proposed approach and MCC-Index method used 70 pixels. In order to investigate the impact of ‘partial sample’, we enrolled the forth sample of each finger and used the rest of sample as probes. Figure 8 depicts the results of using forth sample as enrolled template. The performance of proposed approach are both improved, especially on *FVC2004\_DB1\_A*.



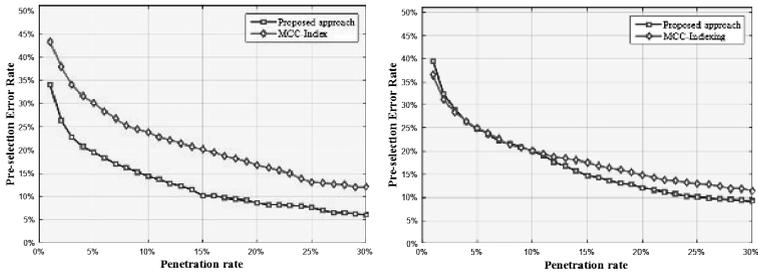
Figure 7: Fingerprint samples selected from *FVC2004\_DB1\_A*.

#### 4.4 Experiments on *FVC2006*

*FVC\_2006\_DB2* was selected to evaluate the performance. The training set is *FVC\_2006\_DB2\_B* which consists of 120 samples, and the test set is *FVC\_2006\_DB2\_A* consisting of 1680 samples which were captured from 140 fingers (12 sample per finger). The first sample of each finger was used for enrolment. Another 11 samples were chosen as probe samples for searching. In total, there are 1540 probe samples. Figure 9 shows the improvement of the proposed approach. The improvement is relatively low, since the performance of MCC-Index method is already a good baseline.

## 5 Conclusion

In this paper, a fingerprint indexing algorithm is designed by only using minutia location and direction information. It is invariant to sample translation and rotation, since the



(a) Experiment on *FVC2004\_DB1\_A* (b) Experiment on *FVC2004\_DB2\_A*

Figure 8: Performance evaluation on *FVC2004\_DB1\_A* and *FVC2004\_DB2\_A*: the **forth** sample of each subject was enrolled in indexing tables.

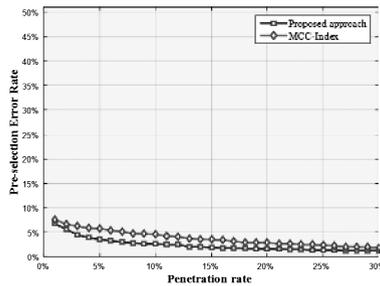


Figure 9: Performance evaluation on *FVC2006\_DB2\_A*.

proposed approach applies the local alignment on each minutia to generate a binary vector rather than using a global reference point. Based on the binary vectors for the template, an indexing approach is designed by combining LSH indexing algorithm developed in MCC-Index method. The experiments on several public database have demonstrated that the proposed approach achieved comparative performance or even better performance than the state-of-the-art fingerprint indexing method. Our future work will extend the experiment to larger-sized databases as well as investigate the impact of the radius of minutiae-disk in order to make the proposed approach more robust to a partial fingerprint sample.

## References

- [ata] Four-quadrant inverse tangent. <http://se.mathworks.com/help/matlab/ref/atan2.html#buact8h0-4>. Accessed: 2015-01-30.
- [BRAC08] Soma Biswas, Nalini K Ratha, Gaurav Aggarwal, and Jonathan Connell. Exploring ridge curvature for fingerprint indexing. In *Biometrics: Theory, Applications and Sys-*

- tems, 2008. *BTAS 2008. 2nd IEEE International Conference on*, pages 1–6. IEEE, 2008.
- [CFFM07] Raffaele Cappelli, Matteo Ferrara, Annalisa Franco, and Davide Maltoni. Fingerprint verification competition 2006. *Biometric Technology Today*, 15(7):7–9, 2007.
- [CFM10] Raffaele Cappelli, Matteo Ferrara, and Davide Maltoni. Minutia cylinder-code: A new representation and matching technique for fingerprint recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(12):2128–2141, 2010.
- [CFM11] Raffaele Cappelli, Matteo Ferrara, and Davide Maltoni. Fingerprint indexing based on minutia cylinder-code. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 33(5):1051–1057, 2011.
- [CFM15] Raffaele Cappelli, Matteo Ferrara, and Davide Maltoni. Large-scale fingerprint identification on GPU. *Information Sciences*, 306:1–20, 2015.
- [FBI] FBI IAFIS. [http://www.fbi.gov/about-us/cjis/fingerprints\\_biometrics/iafis/iafis](http://www.fbi.gov/about-us/cjis/fingerprints_biometrics/iafis/iafis). Accessed: 2015-01-30.
- [FMC12] Matteo Ferrara, Davide Maltoni, and Raffaele Cappelli. Non-invertible Minutia Cylinder-Code Representation. 2012.
- [ISOa] ISO/IEC 19794-2:2011. Information technology – Biometric data interchange Formats – Part 2: Finger minutiae data.
- [ISOb] ISO/IEC 19795-1:2006. Information technology – Biometric performance testing and reporting – Part 1: Principles and framework.
- [LBA07] Xuefeng Liang, Arijit Bishnu, and Tetsuo Asano. A robust fingerprint indexing scheme using minutia neighborhood structure and low-order delaunay triangles. *Information Forensics and Security, IEEE Transactions on*, 2(4):721–733, 2007.
- [LYB14] Guoqiang Li, Bian Yang, and Christoph Busch. A score-level fusion fingerprint indexing approach based on minutiae vicinity and minutia cylinder-code. In *2014 International Workshop on Biometrics and Forensics (IWBF)*, pages 1–6. IEEE, 2014.
- [LYW06] Jun Li, Wei-Yun Yau, and Han Wang. Fingerprint indexing based on symmetrical measurement. In *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, volume 1, pages 1038–1041. IEEE, 2006.
- [MMJP09] Davide Maltoni, Dario Maio, Anil K Jain, and Salil Prabhakar. *Handbook of fingerprint recognition*. springer, 2009.
- [Neu] Verifinger. <http://www.neurotechnology.com/verifying-er.html>. Accessed: 2015-01-30.
- [RM07] Arun Ross and Rajiv Mukherjee. Augmenting ridge curves with minutiae triplets for fingerprint indexing. In *Proceedings of SPIE Conference on Biometric Technology for Human Identification IV*, volume 6539, page 65390C, 2007.
- [SZH08] Xin Shuai, Chao Zhang, and Pengwei Hao. Fingerprint indexing based on composite set of reduced SIFT features. In *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*. IEEE, 2008.
- [WHP07] Yi Wang, Jiankun Hu, and Damien Phillips. A fingerprint orientation model based on 2d fourier expansion (fomfe) and its application to singular-point detection and fingerprint indexing. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 29(4):573–585, 2007.

# Evaluating the Change in Fingerprint Directional Patterns under Variation of Rotation and Number of Regions

Kribashnee Dorasamy<sup>1,2</sup>, Leandra Webb<sup>1</sup> and Jules Tapamo<sup>2</sup>

<sup>1</sup>CSIR, Modelling and Digital Science,

P.O. Box 395, Building 17B, Pretoria, 0001, South Africa

<sup>2</sup>School of Engineering, Howard College, University of KwaZulu-Natal,

King George V Avenue, Durban, 4041, South Africa

KDorasamy@csir.co.za, LWebb@csir.co.za, tapamoj@ukzn.ac.za

**Abstract:** Directional patterns (DPs), which are formed by grouping regions of orientation fields falling within a specific range, vary under rotation and the number of regions. For fingerprint classification schemes, this can result in misclassification due to inconsistency of patterns. Knowing the optimal angle by which to rotate the image and the optimal number of orientation regions to divide it into can be beneficial in analysing specific properties of a class. Furthermore, the number of regions directly impacts singular point (SP) detection, therefore using the optimal number of regions prevents loss of SPs. However, no previous work justifies the use of a specific number of regions or angle of rotation. More so, no explicit studies have been conducted to establish the optimal number of regions or angle of rotation that result in gaining the most information from a pattern. Therefore, this research investigates the change in DPs under the variation of rotation and number of regions to determine which condition provides the best representation of the fingerprint that is less prone to noise and minimizes interclass variability issues with fewer possible patterns for each class. This can serve as a baseline for future works using DPs. The experiments were conducted on the Fingerprint Verification Competition (FVC) 2002 database (DB) 1a. It was found that using a small number of regions produces the most accurate SPs detection and increasing the region number to more than 6 regions drastically depletes the accuracy of SP detection. Furthermore, aligning the SPs of a fingerprint containing a single loop and delta, highlights the essential properties of a class better, with fewer layouts for each class.

## 1 Introduction

Exclusive classification is one of the processes applied in fingerprint recognition systems to decrease the ratio of penetration, in which comparisons are only made against fingerprints which have a common pattern. Galton defined these patterns as classes, namely Whorl (W), Right Loop (RL), Left Loop (LL), Tented Arch (TA) and Plain Arch (PA) [Gal91]. Recently, directional pattern (DP) techniques, stemming from the popular structural approach, have been commonly used in fingerprint classification [LHH08, DWTK15]. In this technique, the arrangement of regions composed of a specific range of orientation fields forms a pattern, known as a DP. Each fingerprint class is represented by a unique DP, which can be used for classification. Globally analysing the arrangement of the regions that constitute a DP of a specific class provides more information about the structural properties of that class, than the traditional structural features. These attributes have

proven to be advantageous in solving common challenges, inclusive of noise, by moving away from the uncertainty of local orientation fields [DWTK15]. In addition, this technique has the potential to classify fingerprints that are not fully captured resulting in loss of singular points (SPs), a very critical issue that is caused from incorrect finger placement on scanners [DWTK15].

To achieve the maximum potential of this technique, DPs have to be consistent for each class. This can be achieved by using the optimal number of regions at a rotation that best emphasises the characteristics of a class. However, classification techniques that have used this fingerprint feature have not undertaken any explicit study examining the ideal conditions of DPs and researchers have not provided reasoning behind the choice of a specific number of regions and rotation [WBS12, LHH08, BKAM13].

Generally many practitioners use three [LHH08, DWTK15] and four [WBS12, MM96, CMM02] region partitions for DPs. Wang *et al.* stated that 4 regions were used as it provides sufficient information to locate SPs and sizes of the regions make it easier to validate a true SP [WBS12]. The authors also claim that eliminating excess noise becomes difficult as the number of regions increases to more than eight partitions [WBS12]. However, no experiments were conducted to justify his statement.

Liu *et al.* and Dorasamy *et al.* both use a DP technique for classification, however the way in which the fingerprints are rotated differ [LHH08, DWTK15]. For each method, unique DPs were produced, showing that they vary under rotation. This inconsistency of patterns can lead to excess amounts of computation. Finding an optimal rotation that has less layouts for each class and distinct patterns that are less prone to noise and interclass variability issues could reduce misclassification and computational cost.

In this paper, the impact of varying the number of regions and rotation of a DP is analysed to establish the optimal number of regions and rotation for classification. The ideal way to test this would be to implement a number of classification schemes for each number of regions and rotation. However, most automated schemes are specific to a single number of regions, and class patterns vary for different numbers of regions and segmentation strategies. Therefore visual comparisons are made of changes in patterns instead.

The following sections present the details of the experiments and the analysis on the impact of the variation of number of regions and rotation of DPs. The description of the methodology is discussed in Section 2. Section 3 provides the discussion of the findings obtained from the change in DPs, influenced by fingerprint rotation and varying the number of regions. Based on the findings, Section 4 states the recommendations on the number of regions and amount of rotation that can be used as a baseline. The conclusions are summarised in Section 5.

## 2 Experimental Set-up

In order to assess the impact of varying the number of regions and varying rotation, two stages of experiments were carried out. The first assessed the impact of varying the number of regions and the second varying the rotation. The details of the two stages are covered in Section 2.1 and 2.2, respectively.

## 2.1 Experimental set-up for the variation of number of regions

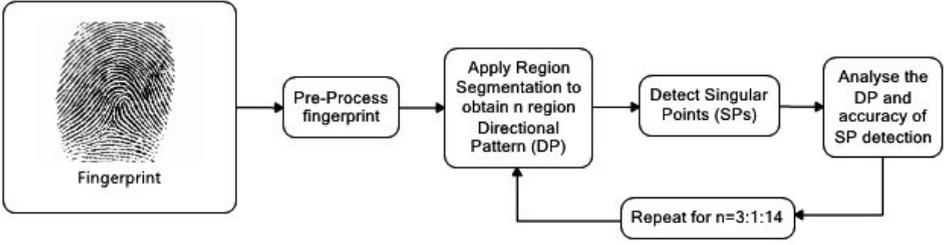


Figure 1: Overview of the experimental set-up for varying the number of regions

Figure 1 shows the overview of the process undertaken to conduct the analysis of the variation of the number of regions. An input fingerprint is initially pre-processed to eliminate the background in the fingerprint image. The orientation fields were computed in the region segmentation stage, where they were grouped into  $n$  homogeneous regions to form a DP. The uniqueness of the pattern and accuracy of the SPs were then observed. The region segmentation, SP detection and observation were repeated with  $n$  being varied from 3 to 14. The pre-processing, region segmentation, SP detection and the observation are further discussed in Section 2.1.1, Section 2.1.2, Section 2.1.3, and Section 2.1.4, respectively.

### 2.1.1 Pre-Processing

In this step, the foreground which includes the ridges and valleys of a fingerprint are separated from the background. The segmentation is performed using the method in [WBG10].

### 2.1.2 Region segmentation

The orientation fields within a specific range  $range_i$  are grouped and assigned a particular region number to form a DP, as shown in Figure 2. The orientation map is represented by a matrix  $Orient(r, c)$ , where  $r$  is the row and  $c$  is the column of the fingerprint image. The computation of the orientation map is covered in [HWJ98], in which the axis advances clockwise from  $0^\circ$  to  $180^\circ$ . To reduce local orientation uncertainty, smoothing is applied using a Gaussian filter. The interval and ranges of each region used to form a DP, are computed using the formulae in Equations 1 to 3 [DWTK15]. Equation 1 is used to obtain the step  $\Delta\phi$  in which the region is discretized into, where  $n$  is the number of regions.

$$\Delta\phi = \pi/n \quad (1)$$

Regions are formed by grouping the orientation fields within a specific range. This range is computed using Equation 2, where  $i = 1 \dots n$ .

$$range_i = [(i - 1) * \Delta\phi] : [i * \Delta\phi] \quad (2)$$

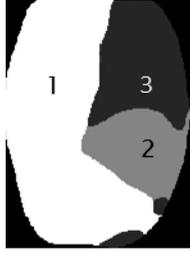


Figure 2: The DP of a particular fingerprint, using three orientation ranges [DWTK15]

Figure 2 depicts region numbers on a DP, at a specific location  $(r, c)$  that is obtained using Equation 3, where  $Orient$  is the orientation value at location  $(r, c)$ .

$$region_{num}(r, c) = \lceil Orient(r, c) / \Delta \phi \rceil \quad (3)$$

### 2.1.3 Singular point detection

The point of intersection between  $n$  regions on a DP, indicates SPs [DWTK15]. The intersection may not always occur at a single point, so a neighbour of 24 or 48 pixels is searched instead, referred to here as 24 ND and 48 ND. Figure 3 (a) shows an example used to indicate an SP by detecting 3 different pixel values, within a 5x5 matrix (24 ND) and around location  $region_{num}(r, c)$ . Loops are identified by regions advancing in a counter-clockwise direction and deltas are represented by regions flowing in a clockwise direction, as illustrated in Figure 3. The coordinates of the loop and delta were compared to a ground truth of manually located SPs.

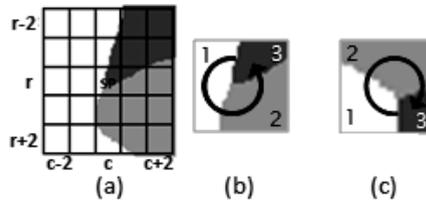


Figure 3: Points of a three region intersection representing (a) SP using a 24 ND, (b) loop and (c) delta [DWTK15]

### 2.1.4 Observation made on the DP

To assess the effect of varying the number of regions, the accuracy of the SP detection is observed as the number of regions are varied. In addition, the pattern of a noisy fingerprint is analysed, as the number of regions are increased.

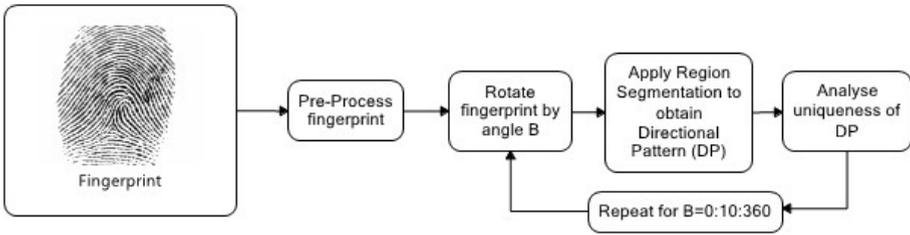


Figure 4: Overview of the experimental set-up for varying angle of rotation

## 2.2 Experimental set-up for the variation of rotation

Figure 4 shows the overview of the process carried out to perform the study on variation of DPs under rotation. Fingerprints are pre-processed to remove the excess background of the image and the fingerprint is rotated by a specified amount. Once the image is rotated, the region segmentation is applied. The change in DP is then observed. The pre-processing and region segmentation algorithm was covered in Section 2.1.1 and Section 2.1.2, respectively. Section 2.2.1 provides the details of obtaining the amount which the image is rotated by. The method of observation is concluded in Section 2.2.2.

### 2.2.1 Increments for rotation

Fingerprints are rotated by an angle  $B$ , where  $B$  increments every  $10^\circ$ , from  $0^\circ$  to  $360^\circ$ . The increment of  $10^\circ$  was selected since it is a small rotational value allowing the observation of the subtle change of a DP.

### 2.2.2 Observation made on the DP

Rotation changes the layout of a DP. To assess the effect of varying rotation, the uniqueness of the pattern is manually observed as the fingerprint is rotated.

## 2.3 Testing Samples

To conduct the experiment 1 as illustrated in Figure 1, fingerprint images from the Fingerprint Verification Competition (FVC) 2002 database (DB)1a were selected [MMC<sup>+</sup>02]. Of the 100 finger impressions contained in FVC 2002 DB1a, 63 upright complete fingerprints were chosen, due to the manual analysis required for experiment 2 depicted in Figure 4.

## 3 Discussion

### 3.1 Variation of the number of regions

The effects of varying the number of regions is observed in terms of SP accuracy and pattern consistency.

### 3.1.1 Impact of the variation of the number of regions on SP accuracy

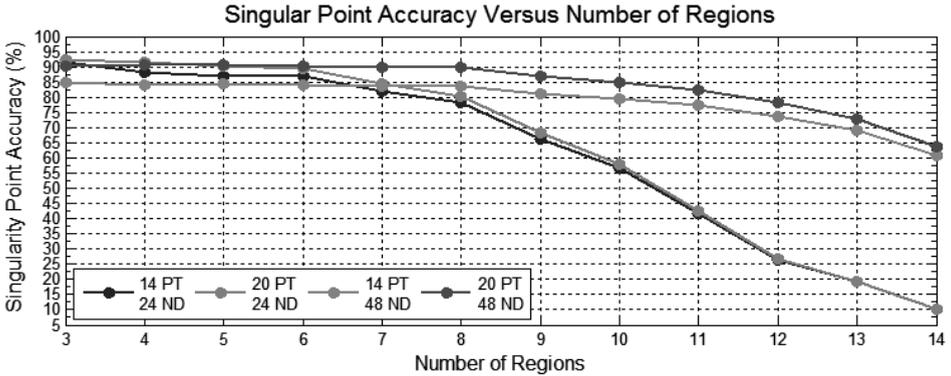


Figure 5: Accuracy of singular points (SPs) versus number of regions

Figure 5 shows the accuracy of SPs detected on a fingerprint for a specific number of regions. The SP detection accuracy is seen to decrease as the number of regions increases from 3 to 14. Four different tests were conducted by varying the pixel tolerance (PT) and ND values. The PT value is the maximum number of pixels that the automatically detected SP can be from the true SP location to still be considered correct. Two different PT values were tested. A PT of 14 is applied to accommodate human error when obtaining the ground truth values. A larger PT of 20 was also selected to evaluate the impact of pixel distortion, resulting in a larger deviation from the location of the ground truth results. Two NDs were observed, a 24 ND since it was the smallest ND that can detect SPs of the largest number of regions under observation; and a 48 ND was used to analysis the impact of pixel distortion on the final accuracy. If any false SPs occurred or any true SPs were not detected, it is considered to be an incorrect detection. From the graph, it is observed that when the number of regions is greater than 6, all lines experience a drop in accuracy, therefore a smaller number of regions is better.

An example of a W fingerprint directional pattern with 15 regions is shown in Figure 6, zoomed in to illustrate the reason for the rapid decrease in SP detection for higher number of regions. From observation all regions don't converge at a single point. Increasing the ND from 24 to 48 can elevate this problem which can be shown by the higher accuracy for 6 to 14 regions. However, using the larger ND range can result in SPs being further from the true SP, therefore a higher PT of 20 shows an increase in the overall accuracy. This indicates that the located SP is 20 pixels off from the true SP. When observing Figure 6 (b), if the ND is too small, all regions may not be detected for larger number of regions, hence the sudden drop from 6 to 14 regions when using a 24 ND. However, even with a high ND and PT, a small number of regions consistently produces better accuracy for all four cases, since there is less distortion and therefore the detected location is generally closer to the true location. A 3 region DP with 14 PT using 24 ND, having a high overall accuracy of 91.38% with a minimum PT, was the optimal output.

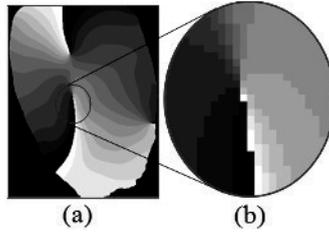


Figure 6: Illustration for the reason for decreased SP detection with a high region number, (a) shows a W with 15 regions and (b) shows the intersection of a 15 region W, which is not a single point

### 3.1.2 Impact of the variation of the number of regions on noisy images

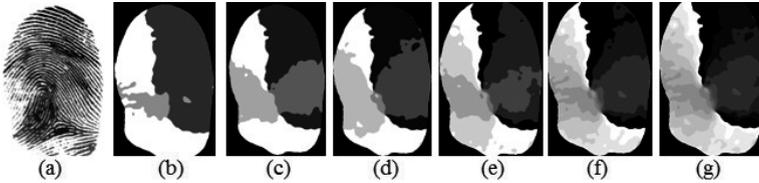


Figure 7: Noisy fingerprint image with different number of regions, (b) 3, (c) 4, (d) 5, (e) 7, (f) 12 and (g) 15

For this case, the smoothing of orientation fields was removed in the pre-processing stage in order to observe how the number of regions affects noise in the DP. Figure 7 (a) depicts a LL containing smudges and missing ridges. The number of regions was varied to observe the behaviour of noise. When the amount of regions increased, the amount of noise increased. For three region segmentation, the least amount of noise is detected and the pattern is more distinct. Therefore, a smaller number of regions is advantageous to both SP detection and noise reduction, as it is likely to encounter false SPs or loss of SPs.

## 3.2 Variation of the amount of rotation

To observe the effect of varying rotation, the pattern changes of each class were observed as the images were rotated. Only the angles at which the fingerprint DP changes the most will be depicted in the figures used in this section.

### 3.2.1 Observation of the PA DP under rotation

Figure 8 (b) - Figure 8 (g) shows the DPs of a PA as the angle is varied. The PA fingerprint has unique structural properties containing no SPs, resulting in the regions never intersecting at any rotation. Therefore, for the PA class, there is no optimal rotation as the pattern can easily be identified for any rotation.

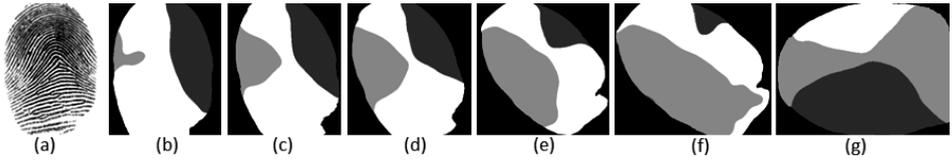


Figure 8: The different patterns of a PA under rotation where the fingerprint is rotated at angles of, (b)  $0^\circ$ , (c)  $20^\circ$ , (d)  $30^\circ$ , (e)  $40^\circ$ , (f)  $60^\circ$  and (g)  $90^\circ$

### 3.2.2 Observation of the W DP under rotation

Figure 9 shows a W fingerprint rotated at angles, (b)  $0^\circ$ , (c)  $10^\circ$ , (d)  $30^\circ$ , (e)  $50^\circ$  and (f)  $90^\circ$ . The DP of a W produces the most consistent pattern compared to RL, LL and TA, due to its symmetrical structure between the top and bottom loop. This forms a single common region between all SPs. Only the region number of the common region changes under rotation. Hence, any rotation is suitable. However, it was found that region loss occurs at the bottom loop, when the fingerprint was at an angle other than  $0^\circ$ ,  $90^\circ$ ,  $180^\circ$ ,  $270^\circ$  and  $360^\circ$ . Therefore, the most reliable patterns are produced at approximate angles of  $0^\circ$ ,  $90^\circ$ ,  $180^\circ$ ,  $270^\circ$  and  $360^\circ$ .

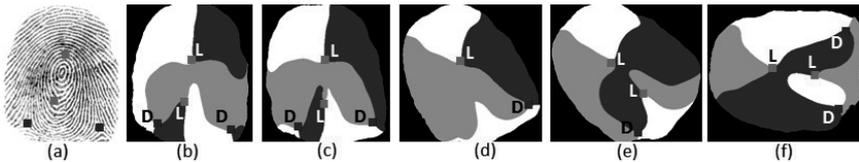


Figure 9: The DPs of a W under rotation, where the fingerprint is rotated at angles of, (b)  $0^\circ$ , (c)  $10^\circ$ , (d)  $30^\circ$ , (e)  $50^\circ$  and (f)  $90^\circ$

### 3.2.3 Observation of the RL, LL and TA DPs under rotation

Since RLs, LLs and TAs have one loop and one delta, these fingerprints were analysed simultaneously. Analysing all the RL, LL and TA DPs produced for angles  $0^\circ$  to  $360^\circ$ , revealed that different layouts were formed for different angles. These layouts differ in the number of regions that link the loop and delta which will be referred to as Common Regions (CRs).

Both RLs and LLs produce three different layouts based on the number of CRs, as shown in Figure 10 and 11. The three layouts of a RL and LL are 3 CR, 2 CR and 1 CR. For a TA only two layouts, 3 CRs and 2 CRs can be produced, as seen in Figure 12. In this section, the changes of the DP of each type of layout of the various classes are analysed to determine the optimal rotation. Since a small rotational change of approximately  $5^\circ$  can result in a DP of a 3 CR TA converting to a 2 CR TA, it was only necessary to analyse one of these two layouts. The 3 CRs TA was used for analysis.

Unlike W and PA fingerprints, patterns for a RL, LL and TL are not consistent. Figures

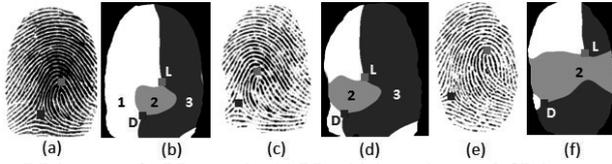


Figure 10: Three DP layouts of a RL namely, (a) RL which produces a 3 CR layout, (b) 3 CR layout, (c) RL which produces a 2 CR layout, (d) 2 CR layout, (e) RL which produces a 1 CR layout and (f) 1 CR layout

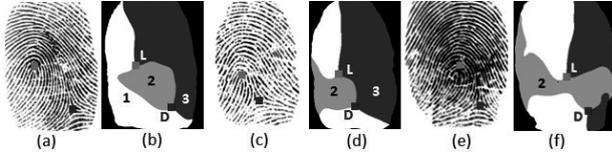


Figure 11: Three DP layouts of a LL namely, (a) LL which produces a 3 CR layout, (b) 3 CR layout, (c) LL which produces a 2 CR layout, (d) 2 CR layout, (e) LL which produces a 1 CR layout and (f) 1 CR layout

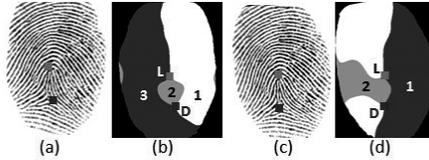


Figure 12: Two DP layouts of a TA namely, (a) TA which produces a 3 CR layout, (b) 3 CR layout, (c) TA which produces a 2 CR layout and (d) 2 CR layout

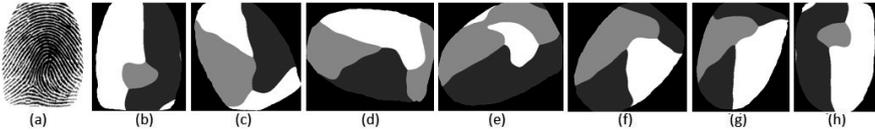


Figure 13: The different patterns produced when a 3 CR RL undergoes rotation at angles of (b)  $0^\circ$ , (c)  $30^\circ$ , (d)  $80^\circ$ , (e)  $120^\circ$ , (f)  $140^\circ$ , (g)  $160^\circ$  and (h)  $180^\circ$

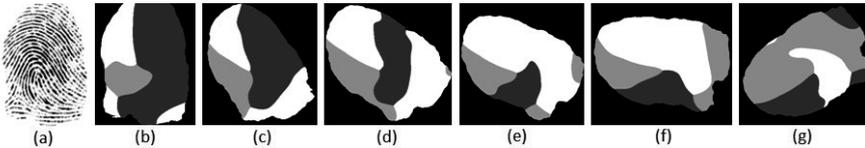


Figure 14: The DPs produced when a 2 CR RL undergoes rotation at angles of (b)  $0^\circ$ , (c)  $20^\circ$ , (d)  $50^\circ$ , (e)  $60^\circ$ , (f)  $80^\circ$  and (g)  $120^\circ$

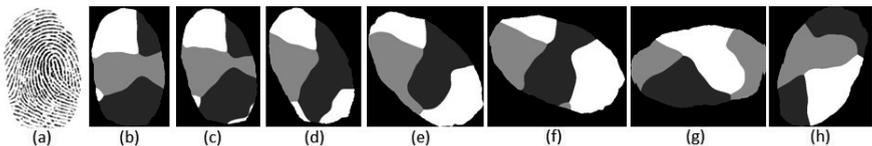


Figure 15: The different patterns produced when a 1 CR RL undergoes rotation at angles of (b)  $0^\circ$ , (c)  $10^\circ$ , (d)  $20^\circ$ , (e)  $30^\circ$ , (f)  $60^\circ$ , (g)  $90^\circ$  and (h)  $150^\circ$

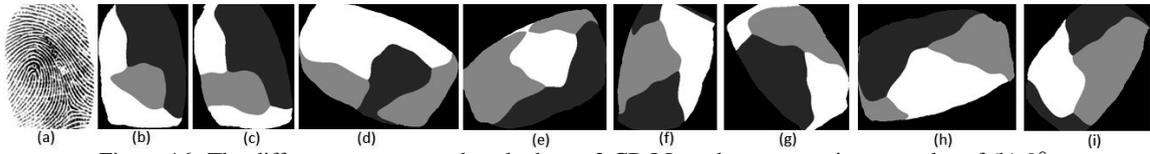


Figure 16: The different patterns produced when a 3 CR LL undergoes rotation at angles of (b)  $0^\circ$ , (c)  $10^\circ$ , (d)  $60^\circ$ , (e)  $110^\circ$ , (f)  $160^\circ$ , (g)  $210^\circ$ , (h)  $280^\circ$  and (i)  $320^\circ$

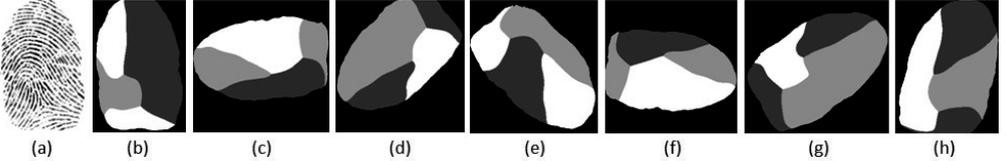


Figure 17: The different patterns produced when a 2 CR LL undergoes rotation at angles of (b)  $0^\circ$ , (c)  $90^\circ$ , (d)  $130^\circ$ , (e)  $220^\circ$ , (f)  $260^\circ$ , (g)  $300^\circ$  and (h)  $340^\circ$

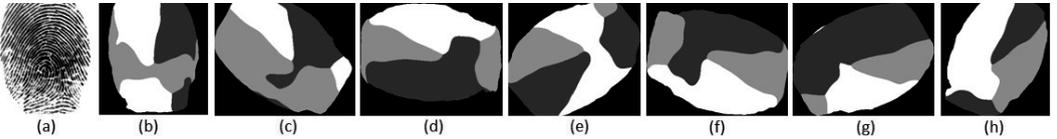


Figure 18: The DPs produced when a 1 CR LL undergoes rotation at angles of (b)  $10^\circ$ , (c)  $60^\circ$ , (d)  $90^\circ$ , (e)  $130^\circ$ , (f)  $260^\circ$ , (g)  $300^\circ$  and (h)  $340^\circ$

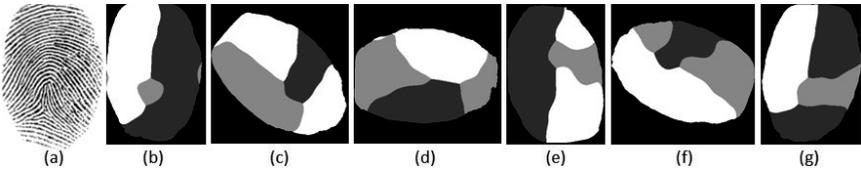


Figure 19: The different patterns produced when a 3 CR TA undergoes rotation at angles of (b)  $10^\circ$ , (c)  $50^\circ$ , (d)  $90^\circ$ , (e)  $190^\circ$ , (f)  $250^\circ$  and (g)  $350^\circ$

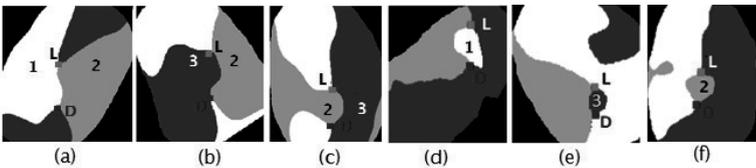


Figure 20: Fingerprint directional patterns produced when aligning SPs for a 2 CR layout (a) LL, (b) RL and (c) TA; and a 3 CR layout (d) LL, (e) RL and (f) TA

13 to 19 show a great variation of patterns under rotation. The number of CRs vary, for example a 2 CR pattern can change to a 3 CR pattern and as it varies the region numbers of the CRs change. Furthermore, RLs, LLs and TAs have highly similar patterns, since any of these classes with the same number of CRs, produce the same region number for the CRs, at each angle. This makes it difficult to differentiate between classes. Since it increases interclass variability among classes, classification accuracy is affected. The only visual difference between these classes, is the location of the SPs on the pattern. However, the location of SPs can only be used to differentiate the classes if fingerprints are all at the same rotation which is unlikely to occur. Therefore, in order to classify a RL, LL and TA consistently from the DPs, a particular method of rotation must be found that produces a unique pattern for each class. In addition, the method of rotation must not be reliant on the fingerprint being initially upright, as this is often not the case. Hence, the method of rotation cannot simply be a global angle to rotate by.

After, further analysing the DP under rotation, it was observed that when the SPs are aligned either vertically or horizontally, both of these requirements are achieved, i.e, a unique pattern is produced for each class and the rotation can be achieved consistently since it is based on internal landmarks rather than a global angle. This is shown by Figure 20, in which SPs are vertically aligned. Figure 20 (a) - Figure 20 (c) show a 2 CR pattern where a LL, RL and TA can be identified by unique region numbers: 1 and 2; 3 and 2; and 2 and 3, respectively. Figure 20 (d)- Figure 20 (f) shows a 3 CR pattern where a LL, RL and TA can be detected by the unique region number of the smallest CR being: 1; 2 and 3, respectively. Since it is independent of using the position of the SP to identify a class, it makes it more reliable by overcoming interclass variability issues and rotation. Therefore, a proposed solution to achieve consistency is to align the SPs.

## 4 Recommendation

Based on the findings in Section 3, the recommendation on the number of regions and amount of rotation for each class are as follows:-

1. A three region partition is recommended, since it produces distinct patterns with the highest SP detection accuracy compared to all other region quantities. Furthermore, the size of each region makes it easier to detect accurate SPs, to eliminate false SPs and noise.
2. For PAs and Ws, consistent patterns can be established at any rotation.
3. For RL, LL and TA fingerprints, vertically aligning the SPs is suggested. This produces consistent unique patterns consisting of only two types of layout, 2 CR and 3 CR.

## 5 Conclusion

This investigation focused on analysing the different DP layouts produced as the number of regions and rotation of the fingerprint was varied and which patterns represent the essential properties of a class better. It was found that a smaller number of regions produced higher accuracy of SP detection, since there are less regions at the point of intersection and they intersect at a single point. Furthermore, it was found that vertically aligning the SPs of a single loop and delta image produces consistent results with fewer possible layouts. The

region between the loop and delta is represented by a unique region number for a particular class type. Therefore, it does not rely only on the positioning of the SPs, making it less prone to interclass variability issues.

## References

- [BKAM13] G.A. Bahgat, A.H. Khalil, N.S. Abdel Kader, and S. Mashali. Fast and accurate algorithm for core point detection in fingerprint images. *Egypt. Informatics J.*, 14(1):15–25, March 2013.
- [CMM02] R. Cappelli, D. Maio, and D. Maltoni. Multi-Classifer Approach to Fingerprint Classification. *Pattern Anal. Appl.*, 5(2):136–144, 2002.
- [DWTk15] K. Dorasamy, L. Webb, J. Tapamo, and N. Khanyile. Fingerprint Classification using a Simplified Rule-Set based on Directional Patterns and Singularity Features. In *8th IAPR Int. Conf. Biometrics*, Phuket, Thailand, 2015. IEEE.
- [Gal91] F. Galton. The Patterns in Thumb and Finger Marks. On their Arrangement into naturally distinct Classes, the Permanence of the Papillary Ridges that make them, and the Resemblance of their Classes to Ordinary Genera. *Philos. Trans. R. Soc. London.B*, (January):1–25, 1891.
- [HWJ98] L. Hong, Y. Wan, and A. Jain. Fingerprint Image Enhancement : Algorithm and Performance Evaluation. 20(8):777–789, 1998.
- [LHH08] L. Liu, C. Huang, and D. C. D. Hung. Directional Approach to Fingerprint Classification. *Int. J. Pattern Recognit. Artif. Intell.*, 22(02):347–365, March 2008.
- [MM96] D. Maio and D. Maltoni. A Structural Approach to Fingerprint Classification. In *Proc. 13th Int. Conf. Pattern Recognit.*, pages 578–585, Italy, 1996. IEEE.
- [MMC<sup>+</sup>02] D. Maio, D. Maltoni, R. Cappelli, J. Wayman, and A. Jain. FVC2002: Second Fingerprint Verification Competition, 2002.
- [WBGs10] L. Wang, N. Bhattacharjee, G. Gupta, and B. Srinivasan. Adaptive approach to fingerprint image enhancement. In *Proc. 8th Int. Conf. Adv. Mob. Comput. Multimed.*, pages 42–49, 2010.
- [WBS12] L. Wang, N. Bhattacharjee, and B. Srinivasan. Fingerprint Reference Point Detection Based on Local Ridge Orientation Patterns of Fingerprints. In *WCCI 2012 IEEE World Congr. Comput. Intell.*, pages 10–15, Brisbans, Australia, 2012. IEEE.

**BIOSIG 2015**

**Further Conference Contributions**



# Robustness Evaluation of Hand Vein Recognition Systems

Christof Kauba, and Andreas Uhl

University of Salzburg, Department of Computer Sciences  
Jakob-Haringer-Str. 2, 5020 Salzburg, AUSTRIA  
{ckauba, uhl}@cosy.sbg.ac.at

**Abstract:** Hand vein recognition systems are more robust against external influences which degrade the image quality like dust or dirt on the sensor or skin surface conditions than fingerprint ones. We investigate the robustness of several hand vein feature extraction and matching schemes against different types of image distortions, related to conditions occurring during the acquisition of hand vein images. These distortions correspond to sensor defects, bad system design and problems in the use of the sensor. The impact on the recognition accuracy is quantified in terms of the EER and compared across different schemes and different types of distortions.

## 1 Introduction

Hand vein recognition systems gain more and more attention nowadays as they provide several advantages over the well established fingerprint ones. Hand vein recognition is more robust against skin surface conditions like dust, dirt, cuts and moisture than fingerprint recognition and can thus be used in scenarios where fingerprint systems cannot because of environment or finger surface conditions.

However other issues might affect the image quality and therefore the recognition accuracy of hand vein systems. These include misplacement of the hand, compression, noise, transmission errors, blurring and sensor ageing related pixel defects. Different strategies have been proposed to assess the robustness of fingerprints, e.g. benchmarking tests like the fingerprint verification contests (FVC [MMJP09a]) and the BioSecure evaluation framework [PDCD09]. An alternative approach is to generate synthetic fingerprints (SFinGe [MMJP09b]) or to artificially degrade real fingerprint images (in [HUPU13] StirMark is used). However for hand and finger vein recognition systems there are neither benchmark data sets nor robustness evaluation results available, except our previous work on the impact of sensor ageing on the recognition performance of finger vein recognition [KU15].

The main goal of this work is to evaluate the robustness of hand vein recognition systems (different feature extraction and matching schemes) against certain kinds of image degradations related to capturing conditions occurring in practice. We use the same methodology as proposed in [HUPU13], i.e. generating the degraded data sets based on a data set captured at the University of Salzburg. StirMark is used to apply image distortions to hand vein images where appropriate and generate degraded data sets. Additionally we generate several “aged” data sets using our image sensor ageing simulation algorithm [KU15].

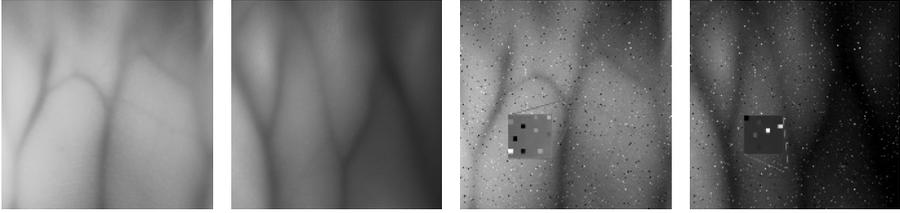


Figure 1: Sample aged images, left: original images, right: aged images containing 10000 hot and 10000 stuck pixels generated by the ageing simulation algorithm

Utilizing StirMark and the image sensor ageing simulation algorithm to generate these data sets has several advantages. First of all the tests are reproducible if their parameters are known and the test data set is available. Further it becomes feasible to isolate specific external influences from others if there is the need to investigate the impact of a specific type of influence. Moreover it is possible to systematically simulate different strengths of distortions corresponding to different levels of external influence, which may not only be a tedious and time-consuming work but also hardly possible to achieve using real data.

Section 2 briefly describes image sensor ageing related pixel defects and presents the StirMark toolkit's image manipulations we utilized. Section 3 gives a short review of the evaluated preprocessing, feature extraction and matching schemes and explains the experimental setup. It continues with the experimental results on the degraded data sets and a short discussion. Section 4 concludes this paper and gives an outlook on future work.

## 2 Image Degradations

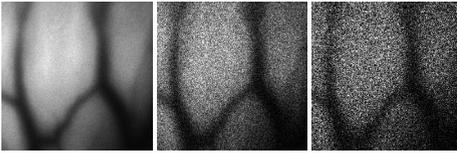
### 2.1 Image Sensor Ageing

In principle a hand vein scanner consists of an infrared light source and an image sensor. An image sensor is an electronic device, containing an array of photosensitive cells, also called pixels, which captures the incoming light and transforms it into an electric signal. The pixels may become defective due to ageing effects. Defective pixels appear as spiky shot noise in the output images. Pixel defects are permanent, their number increases linearly with time, they are randomly distributed over the sensor area and they do not appear in clusters [LCKK09].

**Defective Pixel Types:** there are two main types of in-field pixel defects, hot and stuck pixels [Fri13, LCKK09]. Both are showing different characteristics than at manufacturing time. Example hand vein images containing defective pixels can be seen in figure 1.

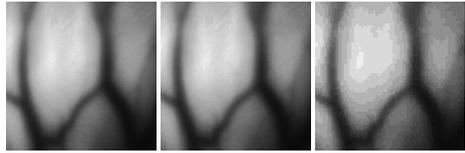
A **stuck pixel** has always the same arbitrary but fixed output value independent of the incoming illumination and exposure settings.

A **hot pixel** adds a light independent offset to the pixel's output which increases linearly with exposure time and might lead to saturation of that pixel.



(a) Level 1 (b) Level 7 (c) Level 15

Figure 2: Additive noise



(a) Level 80 (b) Level 50 (c) Level 10

Figure 3: JPEG compression

## 2.2 The StirMark Toolkit

Fabien A. P. Petitcolas et al. [PAK98] developed a benchmark test in the context of robustness evaluation for digital image watermarking methods, called StirMark (Currently version 4.0 of the toolkit is available at <http://www.petitcolas.net/fabien/watermarking/stirmark/>). It provides specific types of perturbations which are pre-defined and their intensity can be adjusted via a given set of parameters for each type.

In the following we describe the StirMark image manipulations which are chosen to be appropriate for hand vein images and used during the experiments. Not all manipulations provided by StirMark are suitable to simulate natural acquisition conditions. For each manipulation the relation to realistic hand vein capturing scenarios which could be modelled thereby is outlined. The example images shown have been generated by applying the respective StirMark manipulations. The different kinds of manipulations have different meanings in the context of a biometric system and can be grouped into several classes: Sensor ageing related pixel defects and remove lines and columns correspond to a defective sensor. JPEG compression influences result from a bad system design. Median cut filtering and additive noise are due to defects regarding the use of the sensor.

**Median Cut Filtering** results in non directional blur, additionally corrupting the clarity of the vein structure. This sums up small hand movements during the image acquisition and in general blurry vein structures due to the interaction of the infrared light with different types of tissue inside the finger. The size of the filter mask can be set from 1 to 15.

**Additive Noise** simulates noise that might naturally appear in hand vein images due to dust, graining caused by the acquisition equipment itself (e.g. thermal sensor noise), shot noise due to high ISO setting or other errors introduced during processing, storage and transmission of the acquired images. This noise is added to the input image. Its amount can be adjusted by a single parameter ranging from 0 to 100 where 0 means “none” and 100 means “completely random image”. Some example images can be seen in figure 2.

**Remove Lines and Columns** corresponds to errors in hand vein images resulting either from transmission/processing or errors of the biometric sensor while reading the hand vein image (might not be able to read the whole hand and miss or skip some lines). This could be caused by a defective image sensor suffering from dead lines/columns. This manipulation removes lines and columns from the input image. The amount can be adjusted by a single parameter  $k$  which corresponds to the frequency of removing lines, where  $k$  means “remove 1 line in every  $k$  lines”. The dimensions of the output image are reduced.

**JPEG Compression** is applied to the hand vein images in order to save storage space. It is lossy and leads to a general loss in sharpness, reduced edge clarity, loss of colour detail and introduces compression artefacts (blocking and ringing artefacts). This leads to a reduced visibility and breaking of the vein lines. The higher the compression the more severe the artefacts become. The quality level can be set from 0 to 100 where lower numbers indicate higher compression. Figure 3 shows some example JPEG compressed images.

### 3 Experiments

At first a brief overview of the evaluated preprocessing, feature extraction and matching methods is given. Then the hand vein data set and the test protocol are outlined. Subsequently, our experimental results with respect to the different schemes and types of image degradations are presented and discussed.

To improve the visibility of the vein pattern we use **High Frequency Emphasis Filtering** (HFE), **Circular Gabor Filter** (CGF) and simple **CLAHE** (local histogram equalisation) as **preprocessing**.

Different binarisation type **feature extraction** and one key point based technique are used. **Repeated Line Tracking** (RLT), **Maximum Curvature** (MC) and **Wide Line Detector** (WLD) aim to extract the vein pattern from the background resulting in a binary image, followed by a comparison of these binary images. For RLT, MC and WLD the MATLAB implementation by B.T. Ton (publicly available on MATLAB Central: <http://www.mathworks.nl/matlabcentral/fileexchange/authors/57311>) is used. In addition **Local Binary Patterns** (LBP) and a simple **Adaptive Binarisation** (AB) are evaluated as representatives of binarisation type feature extraction methods. **Matching** the binary feature images is done using a correlation measure, calculated between the input images and in x- and y-direction shifted and rotated versions of the reference image. Moreover a **SIFT** based technique with additional key-point filtering is used. AB, LBP and the preprocessing techniques as well as the SIFT (based on VLFeat SIFT: <http://www.vlfeat.org/>) approach are custom implementations. For more details on the preprocessing, feature extraction and matching methods please refer to [KRU14].

**Hand Vein Data Set:** A custom subset of the hand vein data set collected at the University of Salzburg [GU15] is used. Our custom data set contains only images captured using transillumination and includes images of 100 hands, 3 images per hand. This is a relatively low number of images to derive profound statements. Thus we plan to extend the whole data set in the future to include more subjects and also more images per subject/hand.

**Image Sensor Ageing** related pixel defects are simulated using our algorithm proposed in [KU15]. Although in practice only very few defective pixels occur under normal conditions, we use a defect rate of 1000 hot and 1000 stuck pixels per year over a period of 10 years during our experiments to account for environments with higher radiation or other external stress imposed to the sensor.

**EER Determination** is done according to the FVC2004's [MMC<sup>+</sup>04] test procedure, resulting in in 300 genuine matches and 4950 impostor matches ( $3 \times 100$  images).

EER	MC	SIFT	WLD	RLT	LBP	AB
Baseline	0.013	0.02	0.073	0.044	0.297	0.157
5000 Hot	0.015	0.031	0.101	0.063	0.323	0.163
5000 Stuck	0.015	0.027	0.143	0.09	0.337	0.166
10000 Hot + Stuck	0.016	0.029	0.161	0.103	0.35	0.174
Noise level 1	0.023	0.036	0.24	0.143	0.37	0.223
Noise level 3	0.127	0.159	0.477	0.276	0.443	0.374
Noise level 15	0.457	0.294	0.527	0.49	0.457	0.5
RML 1 in 100	0.014	0.023	0.079	0.047	0.317	0.157
RML 1 in 30	0.02	0.017	0.107	0.103	0.36	0.18
RML 1 in 10	0.077	0.03	0.227	0.263	0.433	0.243
Median Filter 3	0.013	0.027	0.076	0.033	0.287	0.15
Median Filter 9	0.013	0.02	0.163	0.033	0.287	0.15
Median Filter 15	0.013	0.03	0.27	0.043	0.337	0.166
JPEG 90	0.013	0.026	0.07	0.043	0.3	0.157
JPEG 50	0.02	0.02	0.123	0.07	0.354	0.167
JPEG 15	0.066	0.118	0.363	0.21	0.45	0.28

Table 1: EER for baseline performance and degraded images

### 3.1 Experimental Results

From table 1 it can be clearly seen that MC and SIFT achieve the best baseline EER and show the highest robustness against all tested distortions.

Figure 4 shows the results for an increasing number of hot (left), stuck (right) and combined hot and stuck (bottom) pixels. In general hot pixels have less influence than stuck pixels. The influence on SIFT and MC is almost negligible up to 20000 defective pixels. AB is influenced starting from 6000 defects. WLD, RLT and LBP are affected starting from several hundred defects, especially for stuck and hot and stuck pixels combined. But in practice more than several hundred defects are very unlikely to occur and most of the schemes are robust against such a number of defects. Thus hand vein recognition systems are robust against a realistic number of pixel defects occurring in practical applications.

If additive noise is applied to the images, each of the tested schemes suffers significantly, which can be seen in the left of figure 5. It has the most severe impact among all of the tested image manipulations. At a noise level of 5 except for SIFT and MC there is no meaningful recognition possible any more.

The difference between image sensor ageing related pixel defects and additive noise is that the defective pixels caused by ageing always have the same fixed locations and characteristics in all output images. Random noise varies from image to image in both, its location

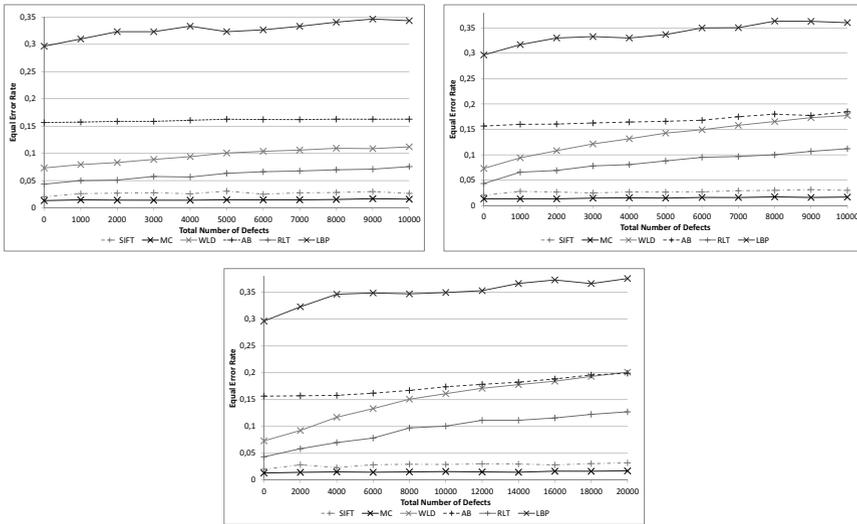


Figure 4: EER for hot, stuck and both, hot and stuck pixels combined

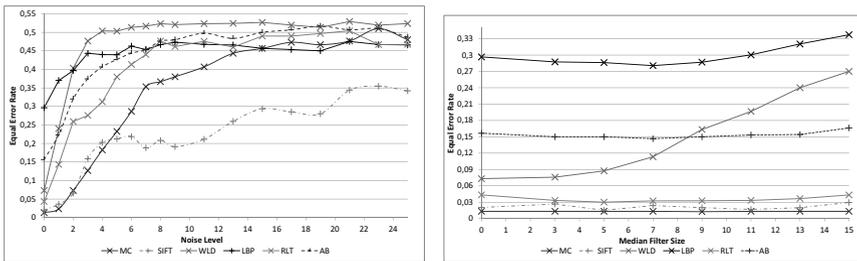


Figure 5: EER for additive noise test / median cut filtering test

and characteristics. The results clearly show that additive noise has a much more severe impact than sensor ageing. This might be due to the type of noise content introduced to the images but more likely due to the different amount of noise that is added. The average PSNR of images with 10000 defects is 24.92. The average PSNR of all images with noise level 1 is 28.28 and for noise level 7 it is 15.3.

Median cut filtering corresponds to blur. As figure 5 right shows, MC is completely insensitive to this type of distortion. SIFT shows quite a good robustness too, except some variations. Actually all schemes except WLD show an improvement in their EER if slight median cut filtering is applied.

Figure 6 left shows the excellent robustness of SIFT against the removal of lines even up to every 5th line is removed. MC is robust against the removal of lines up to 1 in 20 lines are removed. WLD, RLT and AB are affected more and LBP is affected most. Removing only

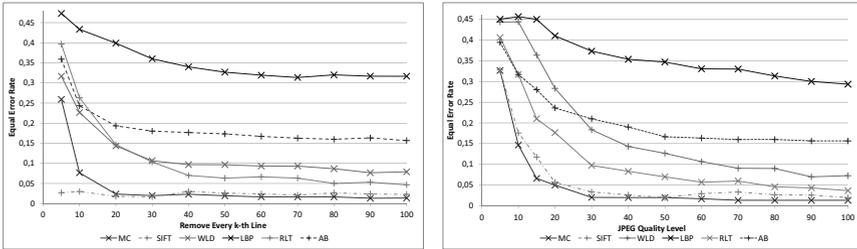


Figure 6: EER for remove lines / JPEG compression

a few lines does not “destroy” the vein lines, i.e. it is unlikely to break them. It shortens them and makes them thinner which has only a minor impact on the actual vein structure. If more lines are removed, vein lines may get broken or disappear completely.

Figure 6 right shows the very high stability of MC against JPEG compression down to a quality level of 30. AB is only slightly affected down to a quality level of 50, from there its EER increases rapidly. The performance of SIFT decreases at first but at a quality level of 50 it is equal to its baseline performance. WLD, RLT and LBP are not robust against JPEG compression at all. All schemes are severely affected below a quality level of 20.

## 4 Conclusion

We assessed the robustness of hand vein recognition systems against several image distortions related to real acquisition conditions. Therefore we generated several test data sets using different StirMark image manipulations and a sensor ageing simulation algorithm. Our experimental results clearly show a large variability in the robustness of the different schemes against the tested types of image distortions. The performance on unperturbed data and even the performance on lower strength levels of the perturbations cannot predict general robustness properties. This necessitates the need for a standardised test tool or common test data sets for the evaluation of hand vein recognition systems like they are available in fingerprint recognition.

Our experiments are a first step towards a systematic robustness evaluation for hand vein recognition. These first results are only theoretical but they provide a basis for further investigations. In practice not only a single kind of distortion will occur but several conditions distorting the images. Our first goal was to have a look at the single distortions and their influence on the recognition performance. Future work will include tests with combined distortions and also more specific image manipulations to be able to exactly model different acquisition conditions occurring in real applications. E.g. the influence of background illumination and constricted or dilated placement of the hand has to be investigated. In addition we will perform further tests on other public available data sets and

using more feature extraction and matching schemes.

## Acknowledgements

This work has been partially supported by the Austrian Science Fund FWF, project no. P26630.

## References

- [Fri13] Jessica Fridrich. Sensor defects in digital image forensic. In *Digital Image Forensics*, pages 179–218. Springer, 2013.
- [GU15] Alexander Gruschina and Andreas Uhl. VeinPLUS: A Transillumination and Reflection-based Hand Vein Database. In *Proceedings of the 39th OAGM/AAPR Workshop*, pages 1–8, Salzburg, Austria, 29-30 May 2015. Austrian Association for Pattern Recognition.
- [HUPU13] J. Hämmerle-Uhl, M. Pober, and A. Uhl. Towards Standardised Fingerprint Matching Robustness Assessment: The StirMark Toolkit – Cross-Database Comparisons with Minutiae-based Matching. In *Proceedings of the 1st ACM Workshop on Information Hiding and Multimedia Security (IH&MMSec’13)*, pages 111–116, Montpellier, France, June 2013.
- [KRU14] Christof Kauba, Jakob Reissig, and Andreas Uhl. Pre-processing cascades and fusion in finger vein recognition. In *Proceedings of the International Conference of the Biometrics Special Interest Group (BIOSIG’14)*, Darmstadt, Germany, September 2014.
- [KU15] Christof Kauba and Andreas Uhl. Sensor Ageing Impact on Finger-Vein Recognition. In *Proceedings of the 8th IAPR/IEEE International Conference on Biometrics (ICB’15)*, pages 1–8, Phuket, Thailand, May 2015.
- [LCKK09] Jenny Leung, Glenn H Chapman, Zahava Koren, and Israel Koren. Statistical identification and analysis of defect development in digital imagers. In *IS&T/SPIE Electronic Imaging*, pages 1–12. International Society for Optics and Photonics, 2009.
- [MMC<sup>+</sup>04] Dario Maio, Davide Maltoni, Raffaele Cappelli, Jim L Wayman, and Anil K Jain. FVC2004: Third fingerprint verification competition. In *Biometric Authentication*, pages 1–7. Springer, 2004.
- [MMJP09a] Davide Maltoni, Dario Maio, Anil K Jain, and Salil Prabhakar. *Handbook of fingerprint recognition*. Springer Science & Business Media, 2009.
- [MMJP09b] Davide Maltoni, Dario Maio, Anil K Jain, and Salil Prabhakar. Synthetic fingerprint generation. *Handbook of fingerprint recognition*, pages 271–302, 2009.
- [PAK98] Fabien AP Petitcolas, Ross J Anderson, and Markus G Kuhn. Attacks on copyright marking systems. In *Information Hiding*, pages 218–238. Springer, 1998.
- [PDCD09] Dijana Petrovska-Delacrétaz, Gérard Chollet, and Bernadette Dorizzi. *Guide to biometric reference systems and performance evaluation*. Springer, 2009.

# Finite Context Modeling of Keystroke Dynamics in Free Text

Nahuel González, Enrique P. Calot

Laboratorio de Sistemas de Información Avanzados  
Facultad de Ingeniería, Universidad de Buenos Aires  
Ciudad Autónoma de Buenos Aires, Argentina  
ngonzalez@lsia.fi.uba.ar, ecalot@lsia.fi.uba.ar

**Abstract:** Keystroke dynamics analysis has been applied successfully to password or fixed short texts verification as a means to reduce their inherent security limitations, because their length and the fact of being typed often makes their characteristic timings fairly stable. On the other hand, free text analysis has been neglected until recent years due to the inherent difficulties of dealing with short term behavioral noise and long term effects over the typing rhythm. In this paper we examine finite context modeling of keystroke dynamics in free text and report promising results for user verification over an extensive data set collected from a real world environment outside the laboratory setting that we make publicly available.

## 1 Introduction

Keystroke dynamics modeling has been applied to password verification as a means to reduce their inherent security limitations. Though its classification performance might not reach those of other biometric schemes making it unsuitable for international security standards, some reported results are outstanding considering how noisy the source data can be and how little extra effort is required for deployment.

Passwords and fixed short texts fit well the general framework of keystroke dynamics; being short sequences typed in a row and repeated often, their characteristic timings tend to be fairly stable in a broad sense. Free text does not enjoy the same privileges. Typing errors, corrections, misspellings, interruptions, pauses to think and attention lapses which are impossible to predict poison the source timing data. What is more, short term variations due to daily tiredness, stress or emotional shifts and long term effects of health and typing skills reeducation are strictly unavoidable. Due to the mentioned difficulties, the problem of free text analysis of keystroke dynamics has only began to be attacked recently [GP05] and in rather controlled conditions. In this paper<sup>1</sup> we examine finite context modeling of keystroke dynamics in free text and report promising results. The general framework we used admits many implementation parameters and selection strategies; their effects on classification performance are studied.

---

<sup>1</sup>The full length version of this paper can be found at <http://lsia.fi.uba.ar/papers/gonzalez15.pdf>

## 2 Theoretical considerations and common techniques

The most commonly used characteristic parameters are hold time -latency between key press and release-, wait time -latency between key release and next key press- and flight time -latency between successive key press events-; to a lesser extent, average typing speed and probability of error or usage frequency of backspace and delete. Usually, timing between consecutive keys -called digraphs- are used but occasionally latencies of bigger groups are chosen [BGP02].

Verification of static passwords, usually in the range of eight to twenty characters long, starts by forming a characteristic vector containing the sequence of values for one or more of the aforementioned parameters, measured during the password entry. The characteristic vector is compared with a pattern vector which is the result of a training process where the same password is entered multiple times. The pattern vector generally contains information on the sample mean, variance and eventually the shape of the adjustment function for all the parameters measured at each keystroke of the password.

Almost every technique for classification and machine learning has been tried for the analysis of keystroke dynamics, ranging from simple metric spaces and k-NN to state of the art classifiers like SVM or random forests; artificial neural networks, fuzzy logic and genetic algorithms have been tried too, with results not as promising. Killourhy and Maxion's review [KM09] has become a classic; an updated one can be found at [KAK11].

Two simple metrics are generally used to grade the quality of biometric systems, including keystroke dynamics modeling: *false acceptance rate* and *false rejection rate*. When a single metric is required for comparison, the *equal error rate* is preferred. Different reporting methodologies in the literature have hindered the cross comparison of results, a problem which is worsened by the diversity of acquisition protocols, data set size, depth or time span, and implementation parameters like minimum required observations for training or classifier threshold. See [GEAHR11] for a detailed review and methodological critique of past studies. We emphasize the lack of public real world data for evaluation, as opposed to that captured in a laboratory setting that artificially diminishes expected variations in the typing rhythm.

Though some isolated attempts were made at tackling its problems, keystroke dynamics analysis of free text has been elusive until recent years [MMCA11]. An original distance metric [GP05] using the degree of disorder of a latencies vector for  $n$ -graphs has shown promising classification performance and high tolerance to variations in typing rhythms.

## 3 Finite context modeling

The motivation behind the usage of finite contexts for modeling keystroke dynamics is based on the fact that prediction by partial matching [CW84] is one the best performing schemes -and asymptotically optimal- for natural language compression. Given that continuous use of a terminal involves extended usage of natural or artificial languages, a similar approach can be expected to be suitable to predict characteristic parameters of keystroke dynamics.

### 3.1 General modeling framework

A partition  $P$  is a sequence  $k_0 \dots k_m$  of key identifiers, codified in a hardware and software independent way to avoid variations due to differences in regional configuration if the sessions are captured in different terminals; a session or input text  $T$  is a sequence of partitions and a training text is a sequence of sessions. The  $n$ -order finite context  $C_j^n$  of the  $j$ -th key  $k_j$  -called leading- in a partition  $P$  is the sequence  $k_{j-n}k_{j-n+1} \dots k_{j-2}k_{j-1}$  of  $n$  preceding key identifiers. For each characteristic parameter  $p$  -hold time, wait time, flight time, applied force or possibly others-, each context  $C$  that appears in the training set of a user  $u$  and its leading key  $k$ , a keystroke model  $M_u(p, C, k)$  is created. The set  $\mathcal{U}$  of all models  $M_u$  is called the user model. The structure of specific keystroke models is open and implementation dependent, being able to provide to the classifier meaningful statistical information about the sampled distribution of timing data for the parameter, context and key. Actually we used models that provided sampled mean and standard deviation, as well as being able to detect outliers using tail probabilities without assuming an underlying normal distribution. However, testing a different set of classifiers or refining techniques might require additional capabilities from the models.

### 3.2 Partitioning and filtering

As the full session content is not generally typed continuously but includes arbitrary pauses, it is split in partitions before being fed to the trainer or classifier with the purpose of restarting the contexts to zero length. A combined strategy was used to partition the text whenever the flight time exceeded a certain fixed threshold or three times the simple moving average for the previous flight times in the partition. Manual inspection showed the criteria matched partitions to humanly perceived pauses in typing; testing more complex strategies did not improve classification metrics. It was found that considering timing data of certain special keys -including modifiers (shift, ctrl, alt), navigation (arrows, page up, page down) and correctors (backspace and delete)- worsened the classifier's performance even though their consideration in the contexts of the next keystrokes improved it. Consequently, their timings but not the keystrokes themselves are removed from the characteristic vector of the session being classified together with the digraph delays of the next keystroke.

### 3.3 Pattern vector reconstruction and classification

It can be assumed in general that not enough repetitions of the input text -free and thus unpredictable- will be found in the training set, if it can be found at all. Thus, to feed a pattern vector to the classifier so it can be compared with the characteristic timing vector of the input text, the former must be reconstructed from the keystroke models in  $\mathcal{U}$ . For every keystroke  $k_i$  in the input text and every characteristic parameter  $p$ , one of the models

$M_u(p, C, k_j)$  is chosen for  $C$  in  $C_j^0$  to  $C_j^j$ . The best strategy is not necessarily the obvious one of picking the longest available context; the order of the optimal predicting context can vary dynamically and, surprisingly, can be much shorter than expected. Also, not all the parameters  $p$  might be meaningful at all times and some can be excluded from the recreated pattern vector (i.e. the first key in every partition can only have an empty context and a single meaningful parameter, hold time). Using an expected vector pattern rebuilt from fragments of training texts does not introduce explicit constraints to classifiers like distance metrics and outlier counting. Unluckily, adapting state of the art classifiers like random forests or SVM to the task does not seem trivial. For this experiment we tested only the aforementioned; a clever scheme using the latter ones can probably improve over our results.

### 3.4 Experimental setup

The data set for the evaluation of the proposed method contains 17158 sessions from 146 users spanning a five months period; the sessions contain up to 10013 keystrokes and an average of 743 keystrokes. Intervals between typing sessions of a certain user can range from hours to days. Sessions were not collected in a special purpose or laboratory setting but were live recorded from daily work with written consent of those involved and their employers; thus, noise factors affecting typing cadences such as interruptions, attention lapses and short and long term emotional, mood or health variations of the individuals are not excluded. Different keyboards with different regional configurations were used, even by the same individuals. The collecting application whose original purpose was to create written reports of professional activities was modified to record keystroke timing; the content does not include personal data or activity trails of the participants. Users' positions are not computer related except for the mentioned purpose of report writing and their typing skills vary from mediocre to excellent, without single finger typists. For each session the timing of key down and key up events were recorded with a precision of 1 millisecond and the presumed user identity, previously authenticated with a passphrase, is included.

In order to encourage collaboration, independent verification of the presented results and further research of keystroke dynamics in realistic scenarios, we make the data set publicly available at the laboratory website and promise to update it as it grows. While most public data sets focus on repetitions of similar strings by the same or different users, we are not aware at the moment of writing this article of any other with this size consisting purely of extended free text without repetitions, except [MMCA11] which is about one third the size of ours but spans a longer period of twelve months.

## 4 Implementation and results

All users with at least 200 sessions were selected for evaluation while the rest were kept only as impostors; 18 users had enough sessions to be evaluated. For every legitimate

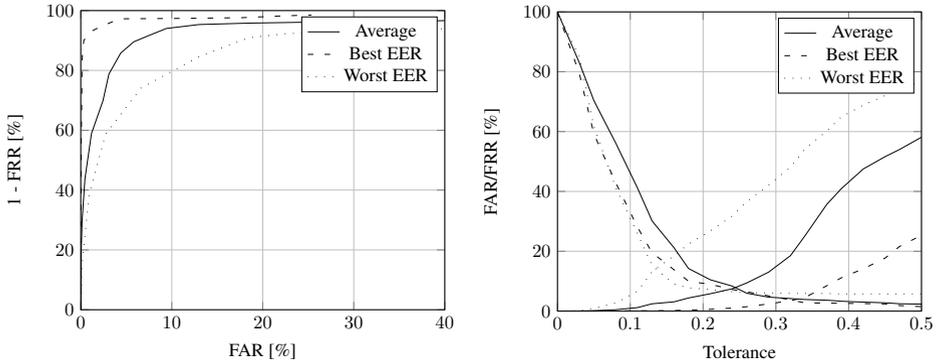


Figure 1: Detailed classifier performance. *Left*, left side of the ROC curve. *Right*, FAR and FRR versus tolerance. Best, worst and average user results for the best strategies combination.

user under evaluation, their first 150 sessions were used as the initial training set to build their keystroke dynamics profile; the remaining sessions were used as challenges, both legitimate and fake, in the same chronological order in which they were captured. After recording the classification result from a legitimate session, its content was used to update the user model.

Due to computational power constraints, a maximum context order of 6 was used; in spite of this limitation, the optimum classification performance seems to be reached around an order of 4. Using less than 100 sessions for the initial training degrades the performance of the method rather fast so a reasonable margin of 150 was chosen; identically, a minimum of 50 legitimate evaluation sessions was used to avoid meaningless variations in the false rejection rate if just a few were misclassified. The observed parameters were hold time, wait time and flight time.

Following [Kil12] we do not consider a certain method to have an EER pointwise defined but randomly distributed; thus, results are reported as average EER together with their standard deviation, maximum and minimum values for the considered users. The best results, with an average EER of 7.56% -minimum of 3.49%, maximum of 13.58%- and a standard deviation of 3.58 were obtained with the combination of choosing the longest available context as model selection strategy, updating models with an exponential moving average, requiring only 10 observations before considering a model valid and estimating the best classifier after 100 evaluation sessions. A detailed plot of FAR and FRR versus tolerance (calculated as fraction of maximum allowed distance or outlier count percentage threshold before rejection) and ROC curves for the best, worst and average user results is shown in figure 1. The results for other combinations of strategies mentioned in this section are shown in table 1.

Three averaging strategies were tested to assess the impact on classification metrics of model updating in addition to a base case where models were not updated after initial training: averaging over all past observations, simple ( $n = 50$ ) and exponential ( $\alpha = 0.98$ ) moving averages. EERs by maximum context order are shown in figure 2. As expected, updating strategies beat the static base case and their performance improves with the importance given to recent observations, making the exponential moving average the best.

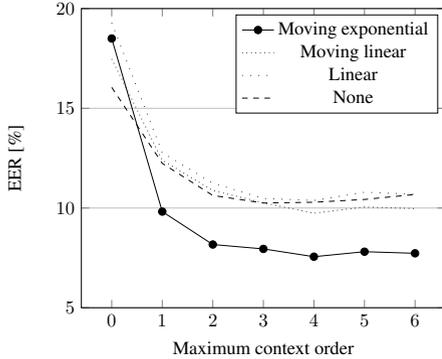


Figure 2: Comparison of model updating and averaging strategies.

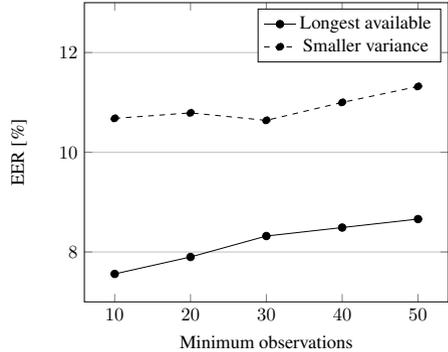


Figure 3: Effect of minimum required model observations on EER for both update strategies.

The minimum required observations to consider a certain model valid is an implementation parameter with influence over the classification performance not only due to the assumed convergence speed of the models but also because of the early availability of higher order models in the classification process, and their updating speed. It is generally claimed that a minimum of 20 observations is required because of the normality assumptions of the underlying variables; however, figure 3 shows that as little as 10 observations are enough to reach optimum or almost optimum EER values with the context selection strategies evaluated next and that increasing their minimum number is slightly but steadily harmful.

A comparison of two context selection strategies shows that the simplest one, selecting for every key  $k_j$  the longest available context, outperforms for every order the strategy of choosing from  $C_j^0$  to  $C_j^{MAX}$  the one that gives the model with the smallest variance.

As expected, the choice of classifier has a noticeable effect on performance. The average results by maximum context order are shown in figure 4 for euclidean distance, outlier count and Manhattan distance, which are named in ascending order of achievement; longest available context and 10 minimum required observations per model were used. It has been shown [Kil12] that the classifier which consistently gives best results is user dependent and not necessarily the one with the best average. Surprisingly, this phenomenon can be exploited with ease to improve classification metrics by splitting the evaluation set and using some sessions -ignoring their results not to bias the reports- for a second training phase that estimates which classifier will outperform the others for that user. We used 50 legitimate sessions and 50 impostor ones for that purpose; even though the best classifier for a certain user was not always chosen, selecting the classifier on a per user basis outperformed every single one on average.

Looking for predictors of classifier performance, it was found that the average of the sample standard deviation for all models is highly correlated to the first. In figure 5 the EER of each user for the best performing scheme is plotted against the predictor, together with a simple regression line. Different approaches like averaging only over the standard deviation of zero order models and over the mean hold time, flight time or both, also show some correlation but not as clearly meaningful as the initial approach.

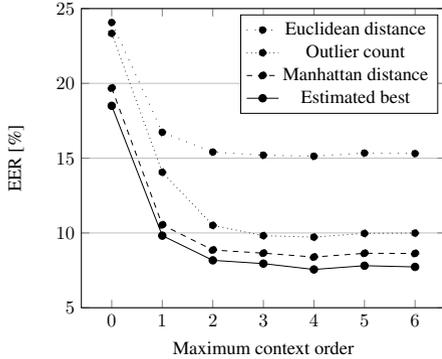


Figure 4: Comparison of basic classifiers and estimated best on a per user basis.

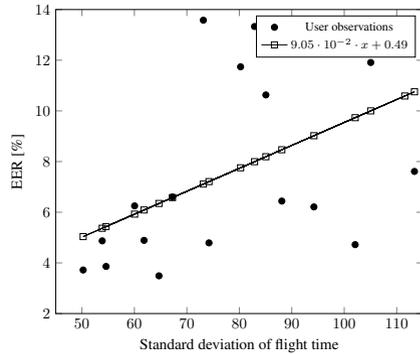


Figure 5: Regression of flight time standard deviation as a predictor of classification performance.

Observations	Order	Context	Classifier	Average EER	St. Dev.	Min	Max
$\geq 10$	$\leq 4$	Longest	Estimated best	7.56%	3.58	3.49%	13.58%
$\geq 10$	$\leq 6$	Longest	Estimated best	7.73%	3.89	3.26%	14.11%
$\geq 10$	$\leq 6$	Longest	Manhattan	8.63%	3.65	3.58%	14.29%
$\geq 50$	$\leq 6$	Longest	Estimated best	8.66%	4.79	2.93%	16.17%
$\geq 10$	$\leq 6$	Longest	Outlier count	9.99%	6.25	3.26%	23.76%
$\geq 10$	$\leq 6$	Min. var.	Estimated best	10.64%	5.7	3.36%	20.40%
$\geq 10$	$\leq 6$	Longest	Euclidean	15.31%	6.47	7.05%	26.14%

Table 1: Ranking of results

The general framework presented in this article admits extensions that were not evaluated and some questions remain unanswered. Can a smarter parameter weighting scheme improve classification performance? Are there better context selection or weighting strategies than those tested? Obviously, state of the art classifiers can beat distance metrics, but how to adapt them to the current scheme is not evident. Particularly, the unexpected results of outlier counts, which surpass euclidean distance -an excellent ranker for static passwords-, points to the fact that specific features of a reduced set of keystrokes might be more important than average deviation from a trained pattern vector. Feature extraction has already proved useful for static passwords before [YC03] and could scale well for free text. Consideration of additional biometrics characteristics which can be successfully extracted from the data with more confidence than identity (i.e. handedness or special key usage patterns) could boost performance of classifiers like random forests.

Finally, as the average of the sample standard deviation for all models is highly correlated with performance, we think the main path to improve the latter is identifying the sources of noise and removing their correlation from the main data. Stress levels, at least, seem to drift key latencies predictably [VZS09]; decorrelating the short term variation might reduce classification errors.

## 5 Conclusion

In this study we have shown the feasibility of user identity verification through keystroke dynamics analysis of free texts captured in a noisy real world environment; the data set was made public to encourage further research in the topic. The performance cannot compete with other biometric systems but has the advantage of being completely transparent and not requiring additional hardware or user actions; however, considering the additional difficulties in comparison with static password verification, our method seems promising and still has room for further improvements. Adapting models to short and long term variations in typing rhythms has proven critical to improve performance, as much as considering individual users as particular sources the classifiers need to adapt to. On the other hand, uncertainties in individual typing rhythms have been shown to be excellent predictors of performance and, as such, the ultimate limit to improvements. We consider this fact proves that keystroke dynamics, not only in fixed short texts but also during unconstrained typing, are rather unique for each user and that finite context modeling captures its essential features. Thus, our future research lines will focus on identifying noise sources of the timing characteristics to mitigate their negative effects on classification.

## References

- [BGP02] F. Bergadano, D. Gunetti, and C. Picardi. User authentication through keystroke dynamics. *ACM Transactions on Information and System Security (TISSEC)*, 5(4):367–397, 2002.
- [CW84] John G. Cleary and Ian Witten. Data compression using adaptive coding and partial string matching. *Communications, IEEE Transactions on*, 32(4):396–402, 1984.
- [GEAHR11] R. Giot, M. El-Abed, B. Hemery, and C. Rosenberger. Unconstrained keystroke dynamics authentication with shared secret. *Computers & security*, 30(6):427–445, 2011.
- [GP05] Daniele Gunetti and Claudia Picardi. Keystroke analysis of free text. *ACM Transactions on Information and System Security (TISSEC)*, 8(3):312–347, 2005.
- [KAK11] M. Karnan, M. Akila, and N. Krishnaraj. Biometric personal authentication using keystroke dynamics: A review. *Applied Soft Computing*, 11(2):1565–1573, 2011.
- [Kil12] Kevin S Killourhy. A scientific understanding of keystroke dynamics. Technical report, DTIC Document, 2012.
- [KM09] Kevin S Killourhy and Roy A Maxion. Comparing anomaly-detection algorithms for keystroke dynamics. In *Dependable Systems & Networks, 2009. DSN'09. IEEE/IFIP International Conference on*, pages 125–134. IEEE, 2009.
- [MMCA11] Arik Messerman, Tarik Mustafaic, Seyit Ahmet Camtepe, and Sahin Albayrak. Continuous and non-intrusive identity verification in real-time environments based on free-text keystroke dynamics. In *Biometrics (IJCB), 2011 International Joint Conference on*, pages 1–8. IEEE, 2011.
- [VZS09] Lisa M. Vizer, Lina Zhou, and Andrew Sears. Automated stress detection using keystroke and linguistic features: An exploratory study. *International Journal of Human-Computer Studies*, 67(10):870–886, 2009.
- [YC03] Enzhe Yu and Sungzoon Cho. GA-SVM wrapper approach for feature subset selection in keystroke dynamics identity verification. In *Neural Networks, 2003. Proceedings of the International Joint Conference on*, volume 3, pages 2253–2257. IEEE, 2003.

# Investigation of Better Portable Graphics Compression for Iris Biometric Recognition

H. Hofbauer\*, C. Rathgeb<sup>†</sup>, J. Wagner<sup>†</sup>, A. Uhl\* and C. Busch<sup>†</sup>

\*Multimedia Signal Processing and Security Lab, University of Salzburg, Austria  
{hhofbaue, uhl}@cosy.sbg.ac.at

<sup>†</sup>da/sec - Biometrics and Internet Security Research Group, Hochschule Darmstadt, Germany  
{christian.rathgeb, johannes.wagner, christoph.busch}@cased.de

**Abstract:** In this paper we present the very first study on the effectiveness of the recently proposed Better Portable Graphics (BPG) image compression algorithm in the context of iris recognition. Original and pre-processed iris images of the IITDv1 iris database are compressed at various reasonable bitrates and the impact of BPG on recognition accuracy is estimated in a bilateral and unilateral compression scenario.

In experiments we found that, compared to well-established image compression standards recommended for biometric data interchange, JPEG and JPEG 2000, BPG generally reveals the least impact on the recognition accuracy of two conventional feature extraction techniques. In addition, we observe that iris segmentation is least affected when employing BPG compression. Consequentially, we identify BPG as an adequate choice for image compression in iris recognition.

## 1 Introduction

Biometrics in particular, technologies of iris recognition [Dau04] represent a rapidly evolving field of research where large-scale biometric systems are already deployed, *e.g.* the Indian Aadhaar project [Uni]. Within such deployments efficient storage and rapid transmission of biometric records are a driving implementation factor, especially for biometric recognition in environments with low-powered mobile sensors or smart-cards. In order to retain vendor neutrality, the International Organization for Standardization (ISO/IEC) specifies iris biometric data to be recorded and stored in (raw) image form rather than in extracted templates, *i.e.* iris-codes [fS11]. Existing studies [RM07, DD08, Gro09] confirm the applicability of lossy image compression in iris biometric systems, recommending the JPEG 2000 standard for iris biometric image compression, which generally outperforms JPEG in terms of PSNR rate-distortion behaviour [fS11]. However, recently it has been shown that rate-distortion performance represents a poor predictor for biometric performance (recognition accuracy). In particular, for conventional iris segmentation techniques the use of JPEG compression, which maintains clear edges that assist iris texture boundary localization, has been found to reveal superior results compared to JPEG 2000 [RUW14].

More recently, the H.265/HEVC-based image compression algorithm BPG [Bel] has been proposed showing promising compression results at minimal human visual perceptible image quality degradation. In this work we examine the usefulness of BPG in the context of

iris image compression. In two different scenarios, compression of pre-processed iris textures and compression of original (cropped) iris images, BPG is compared against JPEG and JPEG 2000 at rates ranging from 1.0 to 0.3 bits per pixel (bpp). In experiments on the uncompressed IITDv1 iris database, the impact of compression algorithms on the performance of two conventional iris recognition systems is evaluated. Considering both cases, bilateral as well as unilateral compression of iris images, we identify BPG as a suitable candidate for iris image compression generally impacts the recognition ratio the least in both scenarios, especially at low compression rates.

This paper is organized as follows: section 2 briefly summarizes related works. The experimental setup of this study is described in section 3 and obtained results are presented in section 4. Finally, conclusions are drawn in section 5.

## 2 Related Work

Focusing on standardization of iris image formats, the ISO/IEC IS 19794-6 [fS11] represents the most relevant standard. Supported by studies conducted in the NIST Iris Exchange program [Nat], JPEG 2000 is recommended exclusively for lossy compression in iris data exchange (IREX records). Apart from standardization numerous studies dealing with image compression in iris recognition have been conducted: the very first study is provided by Rakshit *et al.* [RM07]. They show, that moderate compression of up to 0.5 bpp (bits/pixel) using the JPEG 2000 codec improves recognition accuracy. Their 2156 images dataset, however, refers to CASIA Version 1 data, which does not allow for any conclusions on segmentation impact. Matschitsch *et al.* [MTU07] compare a variety of different compression algorithms (JPEG, JPEG 2000, zero-tree based SPIHT, vector quantization PRVQ, and fractal compression FRAC) resulting in JPEG 2000, SPIHT and PRVQ being almost equally well suited for iris compression. Daugman and Downing [DD08] report, that for a file size of 2000 bytes (1:150 compression ratio) bit flips are caused for only 2-3% of bits in extracted templates, while recognition accuracy is maintained using 1425 images of the ICE database. In their evaluation they did not only employ compression on cropped original images, but also segmentation-assisted cropped and masked (IREX K7) images. Grother [Gro09] surveys existing approaches and compares JPEG and JPEG 2000 to give a quantitative support to the revision of the ISO/IEC IS 19794-6 [fS11] including the cropped format (IREX K3), masked and cropped image format (IREX K7), and unsegmented polar format (IREX K16). The author examines the effect of iris radius, limits of cropping, horizontal and vertical margins, eye lashes, and algorithmic resistance to compression. Ives *et al.* [IBDB10] observe, that a compression of normalized textures has no significant impact on recognition accuracy (compression ratio until 1:100 is feasible). The authors argue, that compression processes may add unique patterns assisting the recognition process.

Conventional image compression algorithms are either optimised with respect to human perception (*e.g.* the JPEG default quantisation (Q-)table) or with respect to rate-distortion criteria (*e.g.* Tier-2 coding in JPEG 2000). For applications in pattern recognition, optimisation with respect to these criteria is not necessarily the optimal solution. In [CZK04] the JPEG Q-table is tuned for application in the pattern recognition context by emphasising middle and high frequencies and discarding low frequencies, which has already been con-

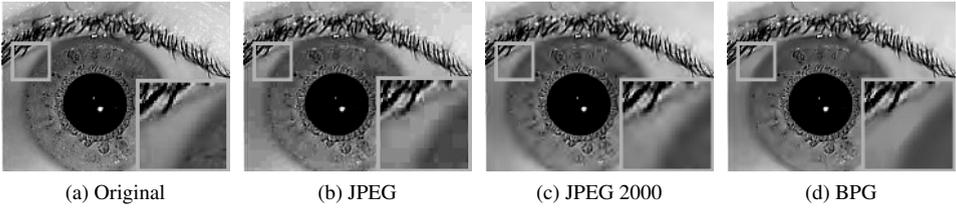


Figure 1: Original version and compressions at 0.3 bpp for 001-01 .bmp of IITD-v1.

considered in face recognition [JKAA06], leading to improved recognition performance. Focusing on iris biometrics, optimisation of JPEG 2000 Part 2 wavelet packet decomposition structures with respect to optimising iris recognition accuracy which provides better results compared to rate-distortion optimised wavelet packet structures [HUKU13]. Further, in [RUW14] it has been shown, that while JPEG 2000 is recommended for compressing iris images [fS11, DD08], the use of JPEG compression yields better segmentation results of cropped iris images (IREX K3), maintaining clearer boundaries between the iris texture and pupil and sclera, respectively, compared to JPEG 2000 and JPEG XR compression.

The BPG format [Bel] is based on the H.265/HEVC video format and a valid HEVC bitstream can be reconstructed from the BPG in case a non modifiable hardware decoder is present. The H.265/HEVC video format is rather complex and there is no single reason for the improvement over the H.264/AVC video codec but rather a large collection of small improvements. The H.265/HEVC format is standardised in [ITU13] and an overview over the various techniques used and improvements over H.264/AVC is given in [BFB<sup>+</sup>13].

The H.265/HEVC subset that comprises BPG was chosen to support a wide variety of features, e.g. animation support, as well support for all features which are present in JPEG, e.g. colorspace, as well as the extended JPEG standard (JPEG XT) [Ric13], e.g. higher dynamic range and lossless compression. The main improvement over the JPEG standard in terms of coding efficiency can be reduced to the smaller block size, combined with an adaptive decomposition quadtree, and intra frame prediction. Focusing on computational complexity, BPG encoding requires slightly more time compared to JPEG, while it still operates in real time (fraction of a second) for iris images considered in this work.

### 3 Experimental Setup

#### 3.1 Database, Segmentation and Recognition

Experiments are carried out using the IITD Iris Database version 1.0 which comprises 2,240 NIR iris images from 224 different subjects. For each subject the first five iris images were acquired from the left eye while the remaining five images were acquired from the right eye, yielding a total number of 448 classes. Original images are sized  $320 \times 240$  pixels and take 77.11kB in uncompressed form, a sample image is depicted in figure 1a. Unrolled and normalized iris texture images are sized  $512 \times 64$  pixels and take 34.58kB in uncompressed form, a sample image is depicted in figure 2a.

We employ (custom) implementations of one segmentation algorithm and two feature extraction techniques publicly available in the USIT software [RUW13]: the Daugman-like

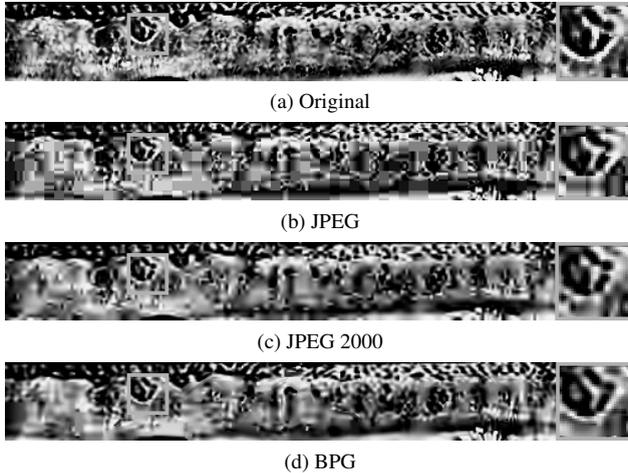


Figure 2: Original version and compressions at 0.3 bpp for texture image of figure 1a.

format	Bits per pixel									
	0.3		0.5		0.7		0.9		1.0	
JPEG	25.8	18.0	28.1	20.1	29.8	21.8	31.2	23.1	31.8	23.7
J2K	27.0	18.7	29.7	21.0	31.9	22.7	33.9	24.3	34.8	25.0
BPG	29.0	20.2	32.0	22.5	35.0	24.3	37.8	25.9	39.2	26.6

Table 1: PSNR[dB] comparison of JPEG, JPEG2000 and BPG for eye and texture images. Entries are given as  $PSNR_{eye} | PSNR_{texture}$  for a compression method (row) and fixed Bits per pixel (column).

[Dau04] Contrast-Adaptive Hough Transform (CAHT) [HUPA09] for segmentation, the feature extraction of Ma *et al.* [MTWZ04], based on dyadic wavelet transforms, and the feature extraction of Masek [MK03] based on 1D Log-Gabor filters. For further details on these implementations the reader is referred to [RUW13].

For a comparison of compression performance between JPEG, JPEG 2000 (as J2K) and BPG see table 1. The comparison is performed on the original cropped eye images and iris texture images (as extracted by the CAHT [RUW13] algorithm). We do not consider segmentation-assisted cropped and masked images as suggested in [DD08] since those already require a segmentation of the iris, *i.e.* we directly employ pre-processed texture images. For JPEG and BPG compression we iteratively configure quality parameters in order to obtain desired bitrates using the *convert* tool and the BPG encoder available at [Bel], for JPEG 2000 we use the *JJ 2000* encoder [Eng]. In [HSU11] effects of JPEG-XR compression on iris recognition are examined revealing similar results to JPEG 2000, *i.e.* we do not consider JPEG-XR in our study. It can be observed that the BPG compression performance in terms of visual quality outperforms the JPEG and JPEG 2000 formats.

### 3.2 Compression Scenarios

In experiments we consider two scenarios: (1) compression of original iris images and (2) compression of iris texture images, depicted in figure 3. In the first scenario, which represents the most relevant case, the iris image is directly compressed after acquisition. If

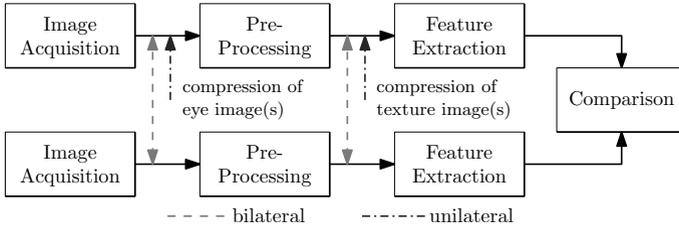


Figure 3: Overview of considered evaluation framework.

Masek; baseline 1.8097				Ma; baseline 1.8300				Masek; baseline 1.8097				Ma; baseline 1.8300			
bpp	BPG	J2K	JPG	bpp	BPG	J2K	JPG	bpp	BPG	J2K	JPG	bpp	BPG	J2K	JPG
0.3	1.79	1.81	1.83	0.3	1.81	1.82	1.96	0.3	1.81	1.47	1.47	0.3	1.77	1.46	1.59
0.5	1.79	1.81	1.83	0.5	1.75	1.79	1.78	0.5	1.78	1.45	1.46	0.5	1.80	1.48	1.48
0.7	1.76	1.79	1.83	0.7	1.77	1.86	1.74	0.7	1.78	1.44	1.45	0.7	1.83	1.49	1.48
0.9	1.79	1.79	1.79	0.9	1.76	1.81	1.73	0.9	1.79	1.44	1.45	0.9	1.81	1.46	1.43
1.0	1.80	1.79	1.81	1.0	1.76	1.79	1.82	1.0	1.80	1.45	1.45	1.0	1.83	1.49	1.46

(a) Bilateral comp. of Masek (left) and Ma (right) (b) Unilateral comp. of Masek (left) and Ma (right)

Table 2: EERs of compressed iris texture images.

biometric sensors do not have the ability to conduct pre-processing it may be necessary to apply lossy image compression to the original eye image for volume reduction of data to be transmitted. Sample images for both scenarios are shown in figure 1b-1d and figure 2b-2d. Further, we evaluate both scenarios using bilateral and unilateral compression, which means that both or only one of the images to be compared are compressed, respectively.

## 4 Experimental Evaluation

### 4.1 Compression of Iris Textures

In a first test, which is given in table 2a and figure 4a, the bilateral case is evaluated. The baseline is the result of the uncompressed comparison. The results show that the performance for all compression types is quite good and comparable. The errors (in terms of EER difference to the baseline) are minor, less than 0.1% for Masek and less than 0.15% for Ma. We restrict to report EERs as single point of measurement since the large number of compression scenarios hinders a comparisons in terms of ROC curves.

However, the compression of the newly acquired texture requires additional computation and might impact the responsiveness of a biometric system. Optimally the unilateral comparison shows the same performance as the bilateral comparison, in which case the additional compression step can be skipped. Table 2b and figure 4b show the results of the bilateral comparison. Only BPG shows a similar behaviour in the bilateral case. The JPEG and JPEG 2000 compression influences the results and lead to a different behaviour. Interestingly, the compression with JPEG and JPEG 2000 increases the performance for the tested feature extraction methods. However, that the compression results in an improvement, can not be guaranteed for other feature extraction methods, and thus should be take as a sign of change to the baseline rather than an improvement. Especially since the compression of the stored textures is final and can not be reversed in case it would impact a feature extraction method aversely.

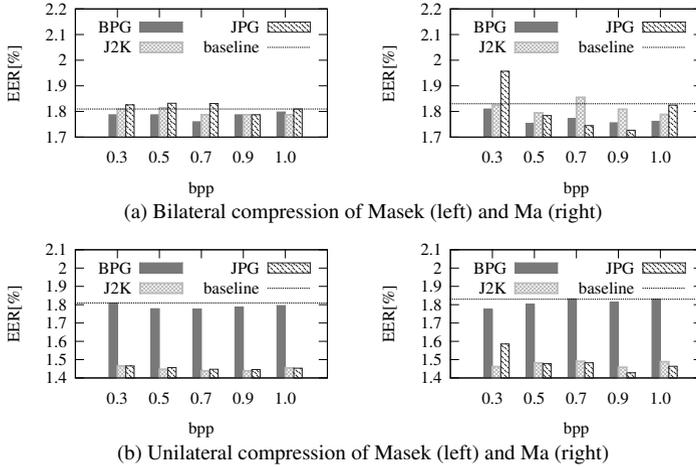


Figure 4: EERs of compressed iris texture images.

Masek: baseline 1.8097				Ma: baseline 1.8300				Masek; baseline 1.8097				Ma; baseline 1.8300							
bpp	BPG	J2K	JPG	baseline	bpp	BPG	J2K	JPG	baseline	bpp	BPG	J2K	JPG	baseline	bpp	BPG	J2K	JPG	baseline
0.3	2.63	3.79	2.90	1.8097	0.3	2.47	3.73	2.79	1.8300	0.3	2.52	3.35	2.52	1.8097	0.3	2.41	3.25	2.32	1.8300
0.5	2.25	2.97	2.61	1.8097	0.5	2.25	2.82	2.54	1.8300	0.5	2.11	2.77	2.36	1.8097	0.5	2.03	2.62	2.41	1.8300
0.7	2.42	2.50	1.94	1.8097	0.7	2.46	2.46	1.99	1.8300	0.7	2.29	2.10	2.14	1.8097	0.7	2.26	2.09	2.10	1.8300
0.9	2.32	2.41	1.85	1.8097	0.9	2.34	2.31	1.81	1.8300	0.9	2.25	2.34	1.91	1.8097	0.9	2.23	2.19	1.85	1.8300
1.0	2.10	2.59	2.01	1.8097	1.0	2.03	2.44	1.99	1.8300	1.0	1.91	2.23	1.90	1.8097	1.0	1.85	2.15	1.81	1.8300

(a) Bilateral comp. of Masek (left) and Ma (right)

(b) Unilateral comp. of Masek (left) and Ma (right)

Table 3: EERs of compressed original eye images.

## 4.2 Compression of Original Images

Similarly to the texture case the question then is in what way the unilateral or bilateral compression influences the result of the biometric system. Compression is even more important in this case since the original eye images are larger and take up more space than the texture images. Also note that the effects that were seen in the texture compression case will also influence the original image case, since the image compression will also result in a compression of the iris texture.

Evaluation results are shown in table 3b and figure 5b for the unilateral case and in table 3a and figure 5a for the bilateral case. The first fact to notice is that the overall performance impact of compression is higher than with texture image compression since the segmentation is now also performed on the compressed images and can introduce errors. JPEG and JPEG 2000 show a similar behaviour as with the textured images, in that the unilateral case is different than the bilateral case, and that again it seems to improve the EER. As was the case with texture images the BPG compression shows a more stable behaviour, being roughly equal in performance for both unilateral and bilateral compression. In this scenario BPG and JPEG outperform JPEG 2000 and while JPEG shows better performance in the low compression tests, BPG performs better for higher compression.

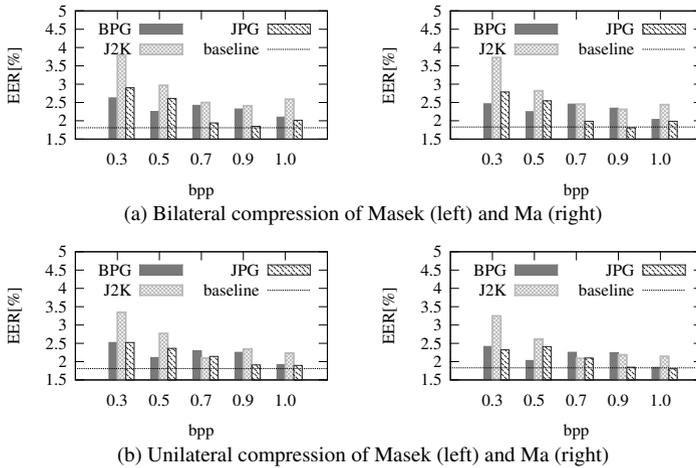


Figure 5: EERs of compressed original eye images.

## 5 Conclusion

We presented the very first studies on the usefulness of BPG compression in iris biometrics. Comparing the BPG format to JPEG and JPEG 2000 in different compression scenarios, including bilateral and unilateral compression, we conclude that BPG is suitable candidate for image compression in iris recognition, especially for low bitrates and for compressing original images rather than texture images, where the former case has more practical relevance. Further, we identify the fact that the BPG encoder/ decoder is released under the BSD license a cutting-edge prerequisite in order to potentially establish BPG compression in the area of biometrics.

## Acknowledgements

This work was partially supported by the European FP7 FIDELITY project (SEC-2011-284862), the Center for Advanced Security Research Darmstadt (CASED) and the Austrian Science Fund, project no. P26630.

## References

- [Bel] F. Bellard. BPG Image format. <http://http://bellard.org/bpg>, retrieved Jan., 2015.
- [BFB<sup>+</sup>13] M. Budagavi, A. Fuldseth, G. Bjontegaard, V. Sze, and M. Sadafale. Core Transform Design for the High Efficiency Video Coding (HEVC) Standard. *IEEE Journal of Selected Topics in Signal Processing*, 2013.
- [CZK04] M. Chen, S. Zhang, and M.A. Karim. Modification of standard image compression methods for correlation-based pattern recognition. *Optical Engineering*, 43(8):1723–1730, 2004.
- [Dau04] J. Daugman. How iris recognition works. *IEEE Trans. Circ. and Syst. for Video Techn.*, 14(1):21–30, 2004.

- [DD08] J. Daugman and C. Downing. Effect of Severe Image Compression on Iris Recognition Performance. *IEEE Trans. Inf. Forensics and Sec.*, 3:52–61, 2008.
- [Eng] D. Engel. JJ 2000. <http://www.wavelab.at/sources/>, retrieved Jan., 2015.
- [fS11] International Organization for Standardization. *ISO/IEC 19794-6:2011. Information technology – Biometric data interchange formats – Part 6: Iris image data*, 2011.
- [Gro09] P. Grother. Quantitative Standardization of Iris Image Formats. In *Proc. of the Biometrics and Electronic Signatures (BIOSIG 2009)*, LNCS, pages 143–154, 2009.
- [HSU11] K. Horvath, H. Stögner, and A. Uhl. Effects of JPEG XR Compression Settings on Iris Recognition Systems. In *Proc. of the 14th Int'l Conf. on Comp. Anal. of Images and Patt.*, LNCS, 6855, pages 73–80, 2011.
- [HUKU13] J. Hämmerle-Uhl, M. Karnutsch, and A. Uhl. Evolutionary Optimisation of JPEG2000 Part 2 Wavelet Packet Structures for Polar Iris Image Compression. In *Proc. of the 18th Iberoamerican Con. on Pattern Recognition (CIARP'13)*, pages 391–398, 2013.
- [HUPA09] J. Hämmerle-Uhl, E. Pschernig, and A. Uhl. Cancelable Iris Biometrics Using Block Re-mapping and Image Warping. In *Proc. of the Information Security Conf. 2009 (ISC'09)* LNCS: 5735, pages 135–142, 2009.
- [IBDB10] R. W. Ives, D. A. Bishop, Y. Du, and C. Belcher. Iris Recognition: The Consequences of Image Compression. *EURASIP J. on Adv. in Signal Processing*, 2010:9 pages, 2010.
- [ITU13] ITU-T H.265. High efficiency video coding, April 2013. <http://www.itu.int/rec/T-REC-H.265-201304-I>, retrieved Jan., 2015.
- [JKAA06] G.-M. Jeong, C. Kim, H.-S. Ahn, and B.-J. Ahn. JPEG Quantization Table Design for Face Images and Its Application to Face Recognition. *IEICE Trans. on Fundamentals of Electronics, Comm. and Computer Science*, E69-A(11):2990 – 2993, 2006.
- [MK03] Libor Masek and Peter Kovesi. MATLAB Source Code for a Biometric Identification System Based on Iris Patterns, 2003. retrieved May, 2012.
- [MTU07] S. Matschitsch, M. Tschinder, and A. Uhl. Comparison of Compression Algorithms' Impact on Iris Recognition Accuracy. In *Proc. of the 2nd Int'l Conf. on Biometrics (ICB 2007)*, LNCS, 4642, pages 232–241, 2007.
- [MTWZ04] L. Ma, T. Tan, Y. Wang, and D. Zhang. Efficient iris recognition by characterizing key local variations. *IEEE Trans. Image Proc.*, 13(6):739–750, 2004.
- [Nat] National Institute of Standards and Technology. NIST Iris Exchange program. <http://iris.nist.gov/irex/>, retrieved Jan., 2015.
- [Ric13] T. Richter. On the standardization of the JPEG XT image compression. In *Picture Coding Symposium (PCS), 2013*, pages 37–40, Dec 2013.
- [RM07] S. Rakshit and D. M. Monro. An Evaluation of Image Sampling and Compression for Human Iris Recognition. *IEEE Trans. Inf. Forensics and Sec.*, 2:605–612, 2007.
- [RUW13] C. Rathgeb, A. Uhl, and P. Wild. *Iris Recognition: From Segmentation to Template Security*, volume 59 of *Advances in Information Security*. Springer Verlag, 2013.
- [RUW14] C. Rathgeb, A. Uhl, and Peter Wild. Effects of Severe Image Compression on Iris Segmentation Performance. In *Proc. of the Int'l Joint Conf. on Biometrics (IJC'14)*, pages 1–6, 2014.
- [Uni] Unique Identification Authority of India. Aadhaar project. <http://uidai.gov.in/>, retrieved Jan., 2015.

# Sparsity-based Iris Classification using Iris Fiber Structures

N. Pattabhi Ramaiah, N. Srilatha\*, C. Krishna Mohan  
Department of Computer Science and Engineering  
Indian Institute of Technology Hyderabad, India, Pin: 502205  
BnPRs Research Lab\*, Andhrapradesh, India, Pin: 533222  
ramaiah.iith@gmail.com, srilatha@bnprs.in, ckm@iith.ac.in

**Abstract:** As there is a growing demand for biometrics usage in e-Society, the biometric recognition system faces the scalability issue as the number of people to be enrolled into the system runs into billions. In this paper, we propose an approach for iris classification using three different iris classes based on iris fiber structures, namely, stream, flower, jewel and shaker for faster retrieval of identities in large scale biometric system. A sparsity based on-line dictionary learning (ODL) algorithm is used in the proposed classification approach where dictionaries are developed for each class using log-Gabor wavelet features. Also, a method for iris adjudication process is illustrated using the iris classification to reduce the search space. The efficacy of the proposed classification approach is demonstrated on the standard UPOL iris database.

## 1 Introduction

Among all the biometrics, fingerprints and iris give more accurate results in uniquely identifying the people based on minutia features. However, the biometric system allows few errors in identification with a threshold at equal error rate. In order to reduce the errors, fingerprint experts look for possible fingerprint matches and enhance the fingerprints to compare the minutia features manually using fingerprint adjudication process. There are scalability issues with the large scale biometric systems where a classification approach is required to reduce the search space. The complex iris texture provides the uniqueness for iris images. Daugman proposed an iris recognition system by using gabor filters and iris codes [Dau93]. Several other researches including Wildes [Wil97], Boles and Boashash [BB98] proposed different iris recognition algorithms by representing the iris texture with Laplacian pyramid construction and 1D wavelet transform, respectively. Few researchers already explored iris classification techniques using hierarchical visual codebook [SZTW13], block-wise texture analysis [RS10] and color information [ZSTW12, PCL13]. So far, there is no classification approach based on the pre-defined iris classes.

Sparse representation has received a lot of attention from researchers in signal and image processing. Sparse coding involves the representation of an image as a linear combination of some atoms in a dictionary [RSS10]. Several algorithms like on-line dictionary learning (ODL) [MBPS09],  $K$ -SVD [AEB06] and method of optimal directions (MOD) [EAHH99] have been developed to process training data. Sparse representation is used to

match the input query image with the appropriate class. Etemand and Chellappa [EC98] proposed a feature extraction method for classification using wavelet packets. In [SS10], a method presented for the learning of dictionaries simultaneously. Recently, similar algorithms for simultaneous sparse signal representation have also been proposed [RS08], [HA06]. The on-line dictionary learning algorithm alternates between sparse coding and dictionary update steps. Several efficient pursuit algorithms have been proposed in the literature for sparse coding [EAHH99],[MZ93]. The simplest one is the  $l_1$ -lasso algorithm [LBRN07]. Main advantage with ODL algorithm is its computational speed as it uses  $l_1$ -lasso algorithm for sparse representation.

The rest of the paper is organized as follows: In section 2, the proposed iris classification approach and the details of on-line dictionary learning are presented.. Experimental results of the proposed classification and adjudication framework are given in section 3. Conclusions are explained in section 4.

## 2 Proposed Iris Classification and Adjudication Framework

The proposed iris classification approach uses three different classes of iris images [Fou09] namely, stream, flower, and jewel-shaker as illustrated in Figure 1. The iris structure can be determined by the arrangement of white fibers radiating from the pupil. In stream iris structure, these fibers are arranged in regular and uniform fashion. The arrangement of fibers is irregular in the flower iris structure. In jewel iris structure, the fibers have some dots. The shaker iris structure have both the characteristics of flower and jewel iris structures. The jewel and shaker classes are merged due to rare occurrence and to make the classification proportional among all the pre-defined classes. The arrangement of fibers are illustrated in Figure 5.

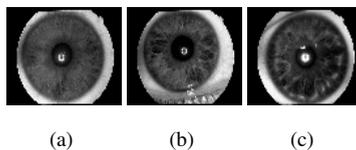


Figure 1: Iris classes: (a) stream, (b) flower and (c) jewel-shaker structures.

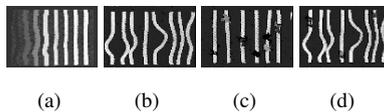


Figure 2: Iris fibers: (a) stream, (b) flower, (c) jewel and (d) shaker fibers.

The following are the steps involved in the proposed iris classification and adjudication

framework:

**Step 1.** *Iris segmentation and normalization* : The pupillary and limbic boundaries [M<sup>+</sup>03] of an iris image are approximated as circles using three parameters: the radius  $r$ , and the coordinates of the center of the circle,  $x_0$  and  $y_0$ . The integrodifferential operator [Dau93] used for iris segmentation is:

$$\max_{r, x_0, y_0} (r, x_0, y_0) G_\sigma(r) * \frac{\partial}{\partial r} \int \frac{I(x, y)}{2\pi r} ds, \quad (1)$$

where  $G_\sigma(r)$  is a smoothing function and  $I(x, y)$  is the image of the eye.

After applying the operator, the resultant segmented iris image is as shown in Figure 3(a). The segmented iris is then converted to a dimensionless polar system based on the Daugman Rubber Sheet model [Dau93] as shown in Figure 3(b).

**Step 2.** *Feature extraction* [M<sup>+</sup>03]: The log-Gabor wavelet feature vector of size  $240 \times 20$  is extracted from the normalized iris image of size  $120 \times 20$ . The resultant feature vector is converted to a single column vector by column major ordering. From each class, some of the iris images are selected to express as a linear weighted sum of the feature vectors in a dictionary belonging to three different classes of iris.

**Step 3.** *Iris classification using ODL*: An on-line dictionary learning (ODL) algorithm is used to classify the iris data into three different classes to reduce the search space. The weights associated with feature vectors in the dictionary are evaluated using ODL algorithm, which is a solution to  $l_1$  optimization for over-determined system of equations. The feature vectors which belong to a particular iris class carry significant weights which are non-zero maximum values.

The class  $C = [C_1, \dots, C_N]$  consists of training samples collected directly from the image of interest. In the proposed sparsity model, images belonging to the same class are assumed to lie approximately in a low dimensional subspace. Given  $N$  training classes, the  $p^{th}$  class has  $K_p$  training images  $\{\mathbf{y}_i^N\}$   $i=1, \dots, K_p$ . Let  $b$  be an image belonging to the  $p^{th}$  class, and it is represented as a linear combination of these training samples:

$$b = \mathbf{D}^p \Phi^p, \quad (2)$$

where  $\mathbf{D}^p$  is a dictionary of size  $m \times K_p$ , whose columns are the training samples in the  $p^{th}$  class and  $\Phi^p$  is a sparse vector.

The following are the steps involved in the proposed classification method:

1. *Dictionary Construction*: Construct the dictionary for each class of training images using on-line dictionary learning algorithm [MBPS09]. Then, the dictionaries  $\mathbf{D} = [D_1, \dots, D_N]$  are computed using the equation:

$$(\hat{\mathbf{D}}_i, \hat{\Phi}_i) = \arg \min_{\mathbf{D}_i, \Phi_i} \frac{1}{N} \sum_{i=1}^N \frac{1}{2} \|\mathbf{C}_i - \mathbf{D}_i \Phi_i\|_2^2 + \lambda \|\Phi_i\|_1, \quad (3)$$

satisfying  $\mathbf{C}_i = \hat{\mathbf{D}}_i \hat{\Phi}_i$ ,  $i = 1, 2, \dots, N$ .

2. *Classification*: In this classification process, the sparse vector  $\Phi$  for given test image is found in the test dataset  $B = [b_1, \dots, b_l]$ . Using the dictionaries of training samples  $D = [D_1, \dots, D_N]$ , the sparse representation  $\Phi$  satisfying  $D\Phi=B$  is obtained by solving the following optimization problem:

$$\Phi^j = \arg \min_{\Phi} \frac{1}{2} \|\mathbf{b}_j - \mathbf{D}\Phi_j\|_2^2 \quad ; \quad (4)$$

subject to  $\|\Phi_j\|_1 \leq T_1$ , and  $\hat{i} = \arg \min_i \|\mathbf{b}_j - \mathbf{D}\delta_i(\Phi^j)\|_2^2$ ,  $j = 1, \dots, t$ .

where  $\delta_i$  is a characteristic function that selects the coefficients. Then  $b_j$  is assigned to  $C_i$  associated with the  $i^{th}$  dictionary. It means, finding the sparsest dictionary for a given test data using  $l_1$ -lasso algorithm. Then, test data is assigned to the class associated with this sparsest dictionary.

- Step 4. *Iris Adjudication*: The matched iris pairs are compared using the adjudication process to illustrate the match-ability of iris images based on the similarity of iris regions marked with three different colors, namely, green, yellow and red. The green, yellow and red colors indicate good, poor and bad match, respectively. The normalized iris image is divided into different regions and the confidence-level of matching for each region is verified and assigned a color code using the dissimilarity measurement.

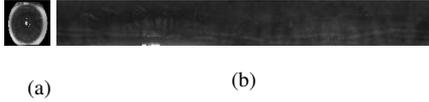


Figure 3: Iris fibers: (a) Iris image segmentation and (b) Normalized Iris Image

### 3 Experimental Results

The experiments were conducted using the iris images taken from the standard UPOL iris database [DMS<sup>+</sup>06], [DMTP04], [DM04]. The iris data is collected from 64 subjects, with three samples of left and right eyes from each subject resulting in a total of 384 iris images. Each iris image is of 24 bit RGB color space with a high resolution image size,  $768 \times 576$ . The images were captured using the optical device (TOPCON TRC50IA) which is connected to a Sony DXC-950p 3CCD camera. In the proposed iris classification approach, three classes are manually identified using the iris patterns stream, flower and jewel-shaker as shown in Table 1. These classes are categorized based on the iris fiber structures (texture information), so the images were converted to gray-scale images for

further processing. The manual identification of the predefined classes is not required for all the data in large-scale applications, but at least those classes should be identified for the training samples.

Table 1: Iris classes defined based on the iris fibers stream, flower and Jewel-Shaker

Class	# of Images (%)	Subject Ids
Class-1 (Stream)	192 (50%)	001,006,007,008,011,013,014,016,018,019,020,021,023,024,026,027,028,033,041,042,044,045,050,051,052,053,058,059,060,061,062,064
Class-2 (Flower)	102 (26.56%)	002,009,010,015,017,022,031,036,037,040,043,047,048,049,054,056,063
Class-3 (Jewel-Shaker)	90 (23.44%)	003,004,005,012,025,029,030,032,034,035,038,039,046,055,057

In order to evaluate the accuracy of proposed classification approach using on-line dictionary learning, the database is split into three sets: training set, testing set and validation set. The distribution of all the three sets are taken in such a way that the 2 samples of each iris image is allotted to the training set and validation set, and the remaining iris sample is given to the test set. The training set consists of 224 images where 112 images are from Class-1 (Stream), 60 images are from Class-2 (Flower) and 52 images are from Class-3 (Jewel-Shaker). The number of test images selected from Class-1, Class-2 and Class-3 are 64, 34 and 30, respectively. A set of 32 iris images is assigned to validation set where 16 images belong to Class-1, 8 images belong to Class-2 and 8 images belong to Class-3.

The experiments were conducted in three different ways of choosing test sets (systematically selecting first, second or third samples of each iris) where the accuracy is almost similar.

In Table 2, the classification accuracy for the validation data set is given. It is observed that 100% classification accuracy is achieved for the dictionary sizes 90 and 120 with residual error value 0.05 as shown in Figure 4. The confusion matrices for both test data and validation data sets are shown in Table 3.

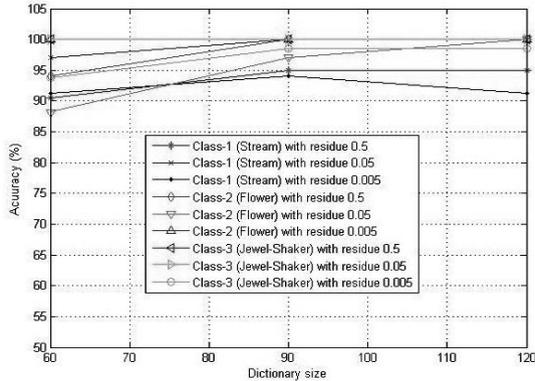


Figure 4: Classification accuracy for three different dictionary sizes 60, 90 and 120

Table 2: Classification accuracy on validation data set

Class	Dictionary Sizes		
	60	90	120
Class-1 (Stream)	91.66	100	100
Class-2 (Flower)	100	100	100
Class-3 (Jewel-Shaker)	100	100	100

Table 3: Confusion matrix for test and validation data

Class	Testing set			Validation set		
	C1	C2	C3	C1	C2	C3
C1	64	0	0	16	0	0
C2	0	34	0	0	8	0
C3	0	0	30	0	0	8

The adjudication results for genuine iris matches are illustrated in Figure 5(a) and for the impostor iris matches are given in Figure 5(b). The normalized images shown on these figures are taken from CASIA database for better illustration of adjudication process.

## 4 Conclusions and Future Work

In this paper, a new methodology for iris classification is proposed to classify the iris images into three different classes namely stream, flower and jewel-shaker. The proposed classification approach achieved 100% classification accuracy with dictionary size 90 and residual error 0.05. Finally the adjudication results are illustrated to avoid the identification errors. The proposed method addressed the scalability issue in large scale iris biometric recognition system for faster retrieval of identities. The proposed approach can be applied

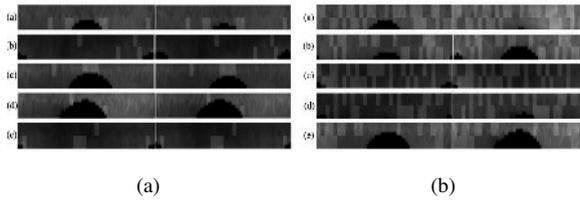


Figure 5: Iris adjudication: LeftSide-(a) genuine iris matches with hamming distances (a) 0.21, (b) 0.19, (c) 0.16, (d) 0.15, (e) 0.19 and RightSide-(b) impostor iris matches with hamming distances (a) 0.48, (b) 0.46, (c) 0.43, (d) 0.51, (e) 0.37

in large scale biometric system in order to reduce the search space and faster retrieval of identities. The manual identification of the predefined classes is not required for all the data in large-scale applications, but at least those classes should be identified for the training samples. The data used for iris classification was collected under visible illumination. Most of the iris recognition systems use the data acquired at near infra-red (NIR) wavelengths. These systems are more accurate among all the existing biometric recognition systems. It is very hard to label the iris classes in the available standard near infra-red databases. The same experimental setup should be executed for the near infra-red iris database which have more texture information to distinguish the iris labels.

## References

- [AEB06] Michal Aharon, Michael Elad, and Alfred Bruckstein. The k-svd: An algorithm for designing overcomplete dictionaries for sparse representation. *Signal Processing, IEEE Transactions on*, 54(11):4311–4322, 2006.
- [BB98] Wageeh W Boles and Boualem Boashash. A human identification technique using images of the iris and wavelet transform. *Signal Processing, IEEE Transactions on*, 46(4):1185–1188, 1998.
- [Dau93] John G Daugman. High confidence visual recognition of persons by a test of statistical independence. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 15(11):1148–1161, 1993.
- [DM04] Michal Dobe and Libor Machala. UPOL Iris Database. <http://www.inf.upol.cz/iris/>, 2004.
- [DMS<sup>+</sup>06] M Dobeš, J Martinek, D Skoupil, Z Dobešová, and J Pospíšil. Human eye localization using the modified Hough transform. *Optik-International Journal for Light and Electron Optics, 2006 Elsevier journal on*, 117(10):468–473, 2006.
- [DMTP04] M Dobeš, L Machala, P Tichavský, and J Pospíšil. Human eye iris recognition using the mutual information. *Optik-International Journal for Light and Electron Optics, 2004 Elsevier journal on*, 115(9):399–404, 2004.

- [EAHH99] Kjersti Engan, Sven Ole Aase, and J Hakon Husoy. Method of optimal directions for frame design. *Acoustics, Speech, and Signal Processing, 1999. Proceedings., 1999 IEEE International Conference on*, 5:2443–2446, 1999.
- [EC98] Kamran Etemad and Rama Chellappa. Separability-based multiscale basis selection and feature extraction for signal and image classification. *Image Processing, IEEE Transactions on*, 7(10):1453–1465, 1998.
- [Fou09] Unitree Foundation. The Rayid model of iris interpretation. <http://rayid.com/main/structures.asp>, 2009.
- [HA06] Ke Huang and Selin Aviyente. Sparse representation for signal classification. *NIPS*, pages 609–616, 2006.
- [LBRN07] Honglak Lee, Alexis Battle, Rajat Raina, and Andrew Y Ng. Efficient sparse coding algorithms. *Advances in neural information processing systems, 2007 MIT Transactions on*, 19:801, 2007.
- [M<sup>+</sup>03] Libor Masek et al. Recognition of human iris patterns for biometric identification. *Bachelor’s thesis, University of Western Australia*, 2003.
- [MBPS09] Julien Mairal, Francis Bach, Jean Ponce, and Guillermo Sapiro. Online dictionary learning for sparse coding. *Machine Learning, 2009 ACM Conference on*, pages 689–696, 2009.
- [MZ93] Stéphane G Mallat and Zhifeng Zhang. Matching pursuits with time-frequency dictionaries. *Signal Processing, IEEE Transactions on*, 41(12):3397–3415, 1993.
- [PCL13] Ioan Pavaloi, Amelia Ciobanu, and Mihaela Luca. Iris classification using WinICC and LAB color features. In *E-Health and Bioengineering Conference (EHB), 2013*, pages 1–4. IEEE, 2013.
- [RS08] Fernando Rodriguez and Guillermo Sapiro. Sparse representations for image classification: Learning discriminative and reconstructive non-parametric dictionaries. 2008.
- [RS10] Arun Ross and Manisha Sam Sunder. Block based texture analysis for iris classification and matching. *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*, pages 30–37, 2010.
- [RSS10] Ignacio Ramirez, Pablo Sprechmann, and Guillermo Sapiro. Classification and clustering via dictionary learning with structured incoherence and shared features. *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 3501–3508, 2010.
- [SS10] Pablo Sprechmann and Guillermo Sapiro. Dictionary learning and sparse coding for unsupervised clustering. *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, pages 2042–2045, 2010.
- [SZTW13] Zhenan Sun, Hui Zhang, Tieniu Tan, and Jianyu Wang. Iris Image Classification Based on Hierarchical Visual Codebook. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 2013.
- [Wil97] Richard P Wildes. Iris recognition: an emerging biometric technology. *Proceedings of the IEEE*, 85(9):1348–1363, 1997.
- [ZSTW12] Hui Zhang, Zhenan Sun, Tieniu Tan, and Jianyu Wang. Iris image classification based on color information. *Pattern Recognition (ICPR), 2012 21st International Conference on*, pages 3427–3430, 2012.

# Palm Vein Database and Experimental Framework for Reproducible Research

Pedro Tome and Sébastien Marcel

Idiap Research Institute

Centre du Parc, Rue Marconi 19, CH-1920 Martigny, Switzerland

{pedro.tome, sebastien.marcel}@idiap.ch

## Abstract:

A palm vein database acquired by a contactless sensor together with an experimental framework freely available for fair reproducible research purposes are described. The palm vein recognition system uses automatic palm region segmentation and circular Gabor filter approach to enhance the veins in the preprocessing, LBP features and histogram intersection as matching. Results are presented comparing two automatic segmentation using the ROI-1 region proportioned by the acquisition sensor and the ROI-2 region generated by the recognition software developed. Complete benchmark results using popular methods and the source code are attached to the database as a reference for other researchers.

## 1 Introduction

Automatic palm vein recognition has emerged as a reliable technology to provide greater level of security to personal authentication system [WESS05]. Among the various human hand biometric characteristics that can be used to recognize a person, such as geometry, fingerprint, palm print or knuckle print, the palm veins are perhaps the most successful form with highest recognition rates achieved between the different characteristics [MCT12] as palm vein patterns are considered stable and reliable. This means that once a person has reached adulthood, the hand structure, veins and configuration remain relatively stable throughout the person's life [YDS06]. In addition, they can be acquired without contact and require the presence of blood in the veins to be registered, which makes more robust these systems against the liveness problem and the spoofing attacks. The palm vein imaging acquisition requires infrared (IR) illumination (generally, NearIR) and standard cameras with a simple CCD or CMOS sensor. Therefore, palm vein images are grayscale images in which dark grey to black veins appear on the grey background.

Because the scarce number of palm vein databases and the different unclear and complex protocols provided by the databases in the literature, no fair reproducible and comparable research can be carried out. For these reasons, the VERA Palm vein database and the experimental framework are introduced and described in this paper freely available for research purposes at [www.idiap.ch/dataset](http://www.idiap.ch/dataset) and [www.idiap.ch/scientific-research/resources](http://www.idiap.ch/scientific-research/resources). Baseline experimental results obtained by the authors using popularly used approaches are also presented.

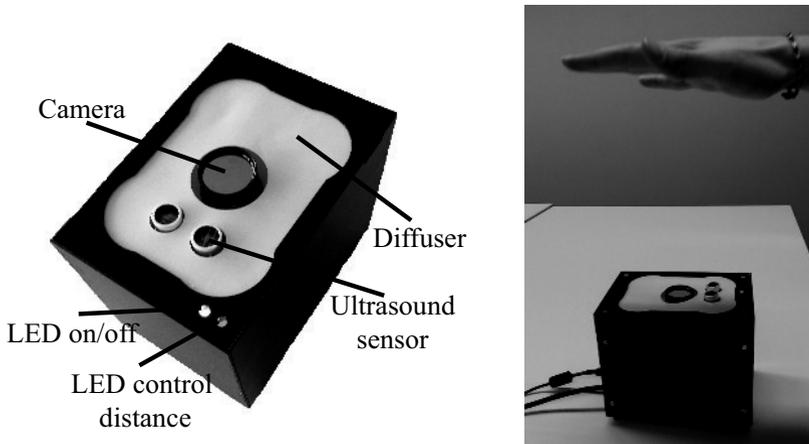


Figure 1: Palm vein prototype sensor description and palm vein image acquisition example.

## 2 State-of-the-art

The most complete research in palm vein pattern recognition was conducted by Fujitsu in Japan, supported by a patent and described in [Wat08]. The database is comprised of 150,000 palm vein images from 75,000 subjects on a different rank of ages. This database was collected for commercial purposes, therefore no details are available and reproduction of the study is impossible. On the other hand, from a non-commercial point of view, there are a scarce number of free available palm vein databases in the literature [HSTR08, KK11]. The most relevant one is the CASIA Multi-spectral [HSTR08], a contactless acquisition from 100 subjects using six different wavelengths (visible, 460, 630, 700, 850 and 940 nm) of the illumination. On the other hand, the PUT database [KK11] is a small database comprises of 50 subjects acquired on a contact sensor using just one wavelength of 880 nm for illumination. It is also important to highlight other database collections mentioned in the literature that are not publicly available such as [MCT12, Lee12]. In this context, researchers working on palm vein recognition built their own acquisition devices to acquire vein pattern images. This resulted in many different proposals for the choice of region of interest (*ROI*), different positioning equipment, various image parameters such as resolution, and different image collection processes. For those reasons, all these works present different protocols and performance results, which in such different conditions are thus difficult to compare. To the best of our knowledge, there are no works in the literature providing any kind of experimental framework which allows the fair comparison of the performance results similar to the new one that we present here.



Figure 2: Image examples from the VERA Palm Vein database. First row shows the *RAW* images acquired and second row shows the ROI-1 images generated by the sensor during the acquisition process. First two columns are male examples while the last two are female examples.

### 3 Database collection and organization

The database introduced in this paper (called VERA Palm vein) consists of 2, 200 images depicting human palm vein patterns. Fig. 2 shows some image examples from the dataset. Palm vein images were acquired by the contactless palm vein prototype sensor developed by University of Applied Sciences Western Switzerland (HES-SO) and the Idiap research institute comprised of a ImagingSource camera, a Sony ICX618 sensor and an infrared illumination of LEDs using a wavelength of 940 nm. The distance between the user hand and the camera lens is measured by a HC-SR04 ultrasound sensor and a led signal that indicates the user the correct position of the hand for the acquisition. This method of contactless acquisition seems to be natural and feasible. Fig. 1 (right) shows an example of the acquisition process and how the user positioning the hand.

Palm vein images were acquired from 110 volunteers for both left and right hands. For each subject, images were obtained in two sessions of five pictures each per hand. Both sessions were separated by an interval of at least 5 minutes. Images of the left and the right hand of the same person in each session were taken alternately, first the left hand and after the right hand. The palm vein images captured by the sensor are saved as bitmap image using a png format with a resolution of  $480 \times 680$ . The database is divided in two datasets: *RAW* and ROI-1 data. The *raw* folder corresponds to the full palm vein image and *roi* folder contains the region of interest (palm vein region) obtained automatically by the sensor during the acquisition process (see Fig. 2). Every dataset contains folders for every person whose id includes the gender of the user (*M*: Male or *F*: Female). User folders are divided into two sessions: 01 and 02, which contain ten images, five from the left hand and five the right hand. Image file names specify all those items of an information

Protocol	World set		Development set			Evaluation set		
	Clients	# Files	Clients	Enrolment	Probe	Client	Enrolment	Probe
nom L&R	20	400	30	120	480	60	240	960
nom L	20	200	30	60	240	60	120	480
nom R	20	200	30	60	240	60	120	480

Table 1: Database detailed description based on number of images for for the three protocols defined and the different sets.

exactly using the next format: “*UUU\_H\_X\_Y.png*”, where *UUU* defines the user id, *H* the hand (*L*: left or *R*: right), *X* the session, and finally, *Y* the number of the acquisition. For example, the image named “021\_L\_1\_2.png” is the second image in the first session of the left palm of the 21<sup>th</sup> user and has the path: “../021-M/01/021\_L\_1\_2.png”.

## 4 Experimental framework

This work presents an open source and extensible experimental palm vein framework called PalmveinRecLib: bob.palmvein<sup>1</sup>, which allows fair and reproducible benchmarks on palm vein recognition. This framework includes a complete module for scores analysis and allows to run a complete palm vein recognition experiment, from the preprocessing of *RAW* images (including segmentation) to the computation of biometric scores and their evaluation. This framework is totally open source and modular, which means that all algorithm parameters are fixed, available and each block can be replaced or improved by new algorithms and approaches. The system implements several baseline methods from the state-of-the-art and is divided on three stages: *i*) segmentation and normalization, *ii*) feature extraction, and *iii*) matching.

In the segmentation process the hand contour is localised by a binarization from grayscale palm vein images. Then the hand landmarks (peaks and valleys) are extracted using the radial distance function (RDF) between the reference point (generally the starting of the wrist) and the contour points extracted [KW14]. The palm region is extracted as a square region based on the located hand landmarks and a scaling and rotation normalization on the extracted palm vein region is performed. Finally, the palm veins are enhanced by using the Circular Gabor Filter (CGF) approach [ZY09]. Once the palm vein region is extracted and normalised, local binary patterns (LBP) are computed to serve as features [MD14] and the histogram intersection metric [SB91] is adopted as a similarity measure to compute the scores.

## 5 Experimental protocol and baseline results

The VERA Palm vein database is presented with three different protocols: *i*) nom L&R - normal operation mode, where left and right hand of the same subject are considered

<sup>1</sup>Freely available at <https://pypi.python.org/pypi/bob.palmvein>

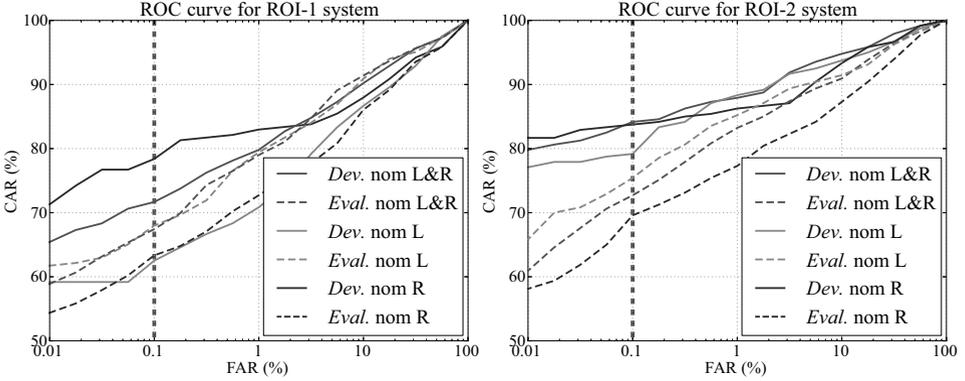


Figure 3: ROC curve of the development set (*Dev.*) and HTER on evaluation set (*Eval.*) for the three protocols defined on the ROI-2 dataset provided by the database and the two ROIs analysed (ROI-1 generated by the sensor during the acquisition and ROI-2 extracted automatically from *RAW* images by the automatic system).

Protocol	ROI-1 data		ROI-2 data	
	Dev. set EER(%)	Eval. set HTER(%)	Dev. set EER(%)	Eval. set HTER(%)
nom L&R	6.66	7.12	3.75	6.80
nom L	8.42	7.73	4.58	5.90
nom R	7.08	8.93	4.57	9.37

Table 2: Benchmark results of EER on the development set (*Dev. set*) and HTER on evaluation set (*Eval. set*) for the three protocols defined on the two datasets (ROI-1 generated by the sensor during the acquisition and ROI-2 extracted automatically from *RAW* images by the automatic system).

different subjects. *ii*) nom L - normal operation mode using just the left hand of the subjects. And *iii*) nom R - normal operation mode using just the right hand of the subjects. Therefore, nom L&R protocol considers a total of 220 subjects and nom L and nom R a total of 110 subjects. For all the protocols, the enrolment is carried out by using the first two images in the first session and the remaining three images plus the five from the second session comprise the probe.

In each protocol, the database is divided on three different sets: *world/training* (subjects 1-20), *development* (subjects 21-50) and *evaluation* (subjects 51-110) as is described in Table 1. Only the images in *world/training set* should be used to train system components such as world/background models, PCA/LDA subspaces, etc., or to otherwise use as background data, for example for score normalisation, etc. The *development set* only should be used to train system hyper-parameters to minimise the chosen error rate metric. The equal error rate (EER) has been used for this purpose. Finally, the *evaluation set* should be used to evaluate palm vein verification accuracy. The decision threshold was determined by tuning on the *development set* (by using the EER), and then applied to palm vein verification scores produced on the *evaluation set*. The half total error rate (HTER), which is the average of false acceptance and false rejection rates after applying the threshold, has been used to measure that accuracy.

Table 2 and Fig. 3 show the benchmark results for the three protocols on the different sets defined. As we can see the ROI-2 images produce better results than the ROI-1 regions, this means that the automatic segmentation implemented align better the palm vein region. Focusing our attention of ROI-2 results, the system achieved a rate of 3.75% of EER on the development set and 6.80% of HTER on the evaluation set on the nom L&R protocol. Results on both hands achieved similar recognition rates of EER on the development set, but however, left hand obtained a rate of 5.90% of HTER on the evaluation set in comparison to the 9.37% of HTER of the right hand. This difference can be explained based on the enrolment images. On the evaluation set of the left hand there are no effects of blurring images, while on the right hand, there are several subjects that experiment this problem on their enrolment images, and therefore, the HTER rate increases.

## 6 Conclusion

This paper presents a new palm vein database acquired by a contactless sensor together with an open source experimental framework freely available for reproducible research purposes. The scarce number of databases and the unclear protocols proposed so far in the literature of this field make this database a valuable reference for the improvement of palm vein recognition systems. The results obtained so far demonstrate the utility of the database and open the opportunity to research on new approaches in the palm vein pattern recognition field. Therefore, the collected database will be useful for the research community as a reference database that provides replicable and clear analysis protocols and a free experimental framework for the fair reproducible research on the palm vein recognition field.

## Acknowledgements

This work has been partially supported by the EU FP7 BEAT (284989) project and the Swiss Centre for Biometrics Research and Testing for support. The authors would like to thank the University of Applied Sciences Western Switzerland (HES-SO) for developing the palm vein sensor.

## References

- [HSTR08] Ying Hao, Zhenan Sun, Tieniu Tan, and Chao Ren. Multispectral palm image fusion for accurate contact-free palmprint recognition. In *Proc. on IEEE International Conference on Image Processing (ICIP)*, pages 281–284, 2008.
- [KK11] R. Kabaciski and M. Kowalski. Vein pattern database and benchmark results. *Electronics Letters*, 47:1127–1128(1), September 2011.

- [KW14] W. Kang and Q. Wu. Contactless Palm Vein Recognition Using a Mutual Foreground-Based Local Binary Pattern. *IEEE Transactions on Information Forensics and Security*, 9(11):1974–1985, Nov 2014.
- [Lee12] Jen-Chun Lee. A novel biometric system based on palm vein image. *Pattern Recognition Letters*, 33(12):1520 – 1528, 2012.
- [MCT12] Goh Kah Ong Michael, Tee Connie, and Andrew Beng Jin Teoh. A contactless biometric system using multiple hand features. *Journal of Visual Communication and Image Representation*, 23(7):1068 – 1084, 2012.
- [MD14] Leila Mirmohamadsadeghi and Andrzej Drygajlo. Palm vein recognition with local texture patterns. *IET Biometrics*, pages 1–9, January 2014.
- [SB91] MichaelJ. Swain and DanaH. Ballard. Color indexing. *International Journal of Computer Vision*, 7(1):11–32, 1991.
- [Wat08] Masaki Watanabe. Palm Vein Authentication. In NaliniK. Ratha and Venu Govindaraju, editors, *Advances in Biometrics*, pages 75–88. Springer London, 2008.
- [WESS05] M. Watanabe, T. Endoh, M. Shiohara, and S. Sasaki. Palm vein authentication technology and its applications. In *Proc. on Biometrics Symposium*, pages 37–38, 2005.
- [YDS06] Erdem Yörük, Helin Dutağaci, and Bülent Sankur. Hand Biometrics. *Image Vision Computing*, 24(5):483–497, May 2006.
- [ZY09] Jing Zhang and Jinfeng Yang. Finger-Vein Image Enhancement Based on Combination of Gray-Level Grouping and Circular Gabor Filter. In *International Conference on Information Engineering and Computer Science (ICIECS)*, pages 1–4, Dec 2009.



# Exploring Gender Prediction From Iris Biometrics

Michael Fairhurst<sup>1</sup>, Meryem Erbilek<sup>2</sup>, Marjory Da Costa-Abreu<sup>3</sup>

<sup>1</sup>EDA, University of Kent, Canterbury, Kent CT2 7NT, UK.

<sup>2</sup>CED, Girne American University, Kyrenia, Cyprus

<sup>3</sup>DIMAp, UFRN, Natal, RN 59078-970, Brasil.

e-mail: M.C.Fairhurst@kent.ac.uk, marjory@dimap.ufrn.br

**Abstract:** Prediction of gender characteristics from iris images has been investigated and some successful results have been reported in the literature, but without considering performance for different iris features and classifiers. This paper investigates for the first time an approach to gender prediction from iris images using different types of features (including a small number of very simple geometric features, texture features and a combination of geometric and texture features) and a more versatile and intelligent classifier structure. Our proposed approaches can achieve gender prediction accuracies of up to 90% in the BioSecure Database.

## 1 INTRODUCTION

The estimation of soft-biometric characteristics of individuals based on extractable features of conventional biometric data has become a very important research topic. Biometric-based estimation of characteristics such as gender, age, and ethnicity is performed by using physical and/or behavioural characteristics embedded in an individual's biometric data. This can be particularly useful in many practical scenarios (checking entitlement claims, for example) including, obviously, forensic investigations. In this paper, our focus is gender prediction from iris biometrics. The literature shows that face biometrics have received the greatest attention in relation to gender prediction [FD12]. This is perhaps not surprising since it is particularly natural and easy to obtain face images for applications such as criminal investigations or profiling from CCTV cameras. However, considerable effort has also been invested in estimating gender from other biometric modalities such as voice [Met07] and text [PDVV11] characteristics. On the other hand, if we consider the predictive properties of the iris in relation to gender characteristics of individuals, only two relevant reported studies [Tho07, LB11] can be found. Indeed, this is a potentially very challenging task, since gender information is not evident from direct human visual inspection of iris images.

In [Tho07], gender prediction is carried out using both geometric and texture features of iris images, and using bagging with the C4.5 decision tree classifier. This proposed gender prediction method was able to achieve 75% and 80% accuracy when tested respectively on the whole dataset and on a subset of this dataset corresponding only to Caucasian subjects. By contrast, in [LB11], gender prediction is carried out using only texture features of

iris images, but adopting a different type, and a larger number of texture features than in [Tho07] while using a support vector machine classifier. When tested on the whole dataset and on a subset corresponding only to single ethnicity subjects, this method was able to achieve an accuracy of around 62% in both cases. Possible reasons for this reduction in the attainable accuracy have been set out and explained in [LB11], summarised as follows:

- Differences in the dataset sizes: experiments in [Tho07] used over 28,000 images whereas in [LB11] 600 images were used, with a factor of around 50 difference in the training set size.
- Differences in the feature vectors: the results in [Tho07] are obtained with combined features computed on the log-Gabor filtered version of the iris image and geometric features, whereas in [LB11] features based on simple spot, line and Laws texture measures were used, without geometric features.
- Differences in the classification structure: the results in [Tho07] were obtained using a multiclassifier configuration (bagging 100 C4.5 decision tree [Qui93]), whereas results in [LB11] were obtained with a single classifier (support vector machine).

A proposed technique for gender prediction from iris samples was presented in [TPB15]. In this paper, once again, the authors use only iris texture and they claim up to 91% accuracy using a variation of fusion of uniform local binary patterns.

An analysis of ageing issues in iris biometrics [FE11] shows that physical ageing effects in iris samples are primarily the result of the physiology of pupil dilation mechanisms, with pupil dilation responsiveness decreasing with age. Hence, pupil dilation is very likely to be related to the geometric appearance of the pupil and the iris, where these findings suggest that geometric features of the iris may also provide useful information for the gender-based biometric prediction task.

Therefore, in this paper, we will investigate and explore the gender prediction task with respect to three different approaches which respectively use (a) only geometric features, (b) only texture features and (c) both geometric and texture features extracted from iris images, and we will use more versatile and intelligence-rich classification structures. We will compare achievable error-rate performance and execution times for each approach.

## 2 GENDER ESTIMATION USING IRIS IMAGES

The basic processing of biometric data in our iris-based gender prediction approach adopts a process based on the following: An eye image is captured in the *Acquisition* step. The *Segmentation* step localises the iris region from the acquired eye image. This step involves detection of the sclera/iris and pupil/iris boundaries. The *Feature extraction* step extracts geometric, texture or both geometric and texture features of the iris according to the configuration required. The *Prediction* step uses the data generated at the output of the previous step and performs the gender classification task itself.

The Data Set 2 (DS2) of the BioSecure Multimodal Database (BMDB) [OG10] is used in this study. The samples were collected as part of an extensive (and commercially available) multimodal database by 11 European institutions participating in the BioSecure Network of Excellence. The eye images were acquired in a standard "office" environment managed by a supervisor and using the LG Iris Access EOU3000 set-up. During the acquisition, spectacles were not allowed to be worn by subjects, although contact lenses were allowed. Four eye images (two left and two right) were acquired in two different sessions with a resolution of 640\*480 pixels, for 210 subjects in total. However, the iris samples of 10 subjects were found to be incorrectly labelled in this database (some of the left eye samples labelled as right or vice versa), and were thus discarded. Hence, this decreased the available number of subjects to 200 (a total of 1600 images).

Using the defined iris dataset, each eye sample is first segmented using the automatic segmentation algorithm as described in [FE11, EF11]. In the event of segmentation failure (this occurred for only 1.87% of images), we segment the irises manually and make sure that all eye images are correctly segmented in order to guarantee the reliability of the further analysis. Subsequently, the obtained iris and pupil parameters from the segmentation process are stored for each eye, to be used in the further processing stages. A full description of these features can be found in [FE11, EF11].

### 2.0.1 Approach 1: Geometric feature extraction and correlation

By using the iris and the pupil parameters saved during the segmentation stage, several features which are related to the geometric characteristics of the iris are extracted. Here, it is important to note that the extraction of these features is computationally simple and fast, since none of them requires the extraction of texture information relating to the iris patterning.

The parameters which were obtained at the segmentation stage are:  $p_x$  (which is the  $x$ -coordinate of the centre of the pupil),  $i_x$  (which is the  $x$ -coordinate of the centre of the iris),  $p_y$  (which is the  $y$ -coordinate of the centre of the pupil),  $i_y$  (which is the  $y$ -coordinate of the centre of the iris),  $i_r$  (which is the iris radius), and  $p_r$  (which is the pupil radius). By using the pupil and iris parameters defined above, 12 (GF1-GF12) geometric features are extracted for our experimental study. Features GF1-GF7 were similarly defined and adopted as in [Tho07], while the remaining five features are specific to this study and adopted from [EFDCA13]. A brief description of these features (specified at the pixel level) is shown in Table 1.

Feature No.	Feature Calculation	Feature No.	Feature Calculation
GF1	$ p_x - i_x $ (distance in $x$ )	GF7	$GF4/GF5$ (area ratio)
GF2	$ p_y - i_y $ (distance in $y$ )	GF8	$i_r/p_r$ (dilation ratio)
GF3	$ GF1 - GF2 $ (distance from centres)	GF9	$p_i * 2 * i_r$ (iris circumference)
GF4	$\pi * i_r^2$ (area iris)	GF10	$p_i * 2 * p_r$ (pupil circumference)
GF5	$\pi * p_r^2$ (area pupil)	GF11	$GF9/GF10$ (circumference ratio)
GF6	$GF4 - GF5$ (true area iris)	GF12	$GF9 - GF10$ (circumference diff)

Table 1: Geometric features

Then, a correlation evaluation across the features is carried out as in [EFDCA13]. By

removing the highly correlated features, efficiency is increased by adopting only the more distinguishing and non-redundant features. The inter-feature correlations were evaluated by using Spearman's rank correlation (a nonparametric-based estimate of correlation).

## 2.0.2 Approach 2: Normalisation and texture feature extraction

After the segmentation stage, this approach performs a normalisation step. This step transforms the iris region into a fixed rectangular block, so that the iris region extracted from the overall eye image is presented at the fixed size necessary for comparisons between samples. A technique [Mas03] based on Daugman's rubber sheet model is employed, which produces a 2D array with horizontal dimensions of angular resolution and vertical dimensions of radial resolution. This produces an unwrapped image of size 20\*240 pixels.

Following the normalisation, 1D Log-Gabor wavelets are used to encode features [Mas03]. Each row of the 2D normalised iris pattern corresponds to a circular ring on the iris region. These rows are divided into a number of 1D signals and convolved with 1D Log-Gabor wavelets which outputs a template of size 20\*480 with both real and imaginary components. As in [Tho07], we only use the real components (which correspond to the array of complex numbers of size 20\*240 of the template) to extract texture features, which are defined in Table ???. Features  $TF1$ ,  $TF2$ ,  $TF6$  were similarly defined and adopted in [Tho07], while the remaining three features are specific to this study and adopted from [EFDCA14].

## 2.0.3 Approach 3: Combining geometric and texture features

This approach simply adopts the combination of approach 1 and approach 2. Hence, geometric and texture features obtained from approach 1 and approach 2 respectively, are combined simply by concatenating them.

## 2.1 Prediction

The gender prediction task involves the specification of how to form the training and testing sets as well as the classification method to be applied. In order reliably to evaluate the performance of the gender classification task, we divide the available samples into person-disjoint testing and training sets. Thus, samples from approximately 72% of the male and the female subjects are used as a training set and the remaining subjects' samples are used as a testing set.

One of the more difficult aspects of designing any classification task is making the best choice of classifier or, in the case of a multiclassifier approach, choosing the set of base classifiers for the fusion method. A guarantee of high diversity among the individual components is essential in the latter context. In order to achieve diversity, we have selected a pool of well known classifiers that have fundamentally different base structures for this experimental study, named: Multi-Layer Perceptron (MLP) [Hay99], Support Vector

Machine (SVM) [FAE08], Optimised IREP (Incremental Reduced Error Pruning) (JRip) [FW94], Decision Tree (DT) [Qui93], K-Nearest Neighbour (KNN) [Ary98].

In order to analyse the full potential of using geometrical, texture and both geometrical and texture features, we have also considered a range of traditional fusion techniques and more intelligent combination techniques, named: Sum-based fusion (Sum) [KA03], Majority Voting (Vote) [Kun04] and Bagging [BB96]. We are especially interested in the use of intelligent agent-based architectures, which we have shown to be well suited to processing biometric data (see, for example, [DCAF11, AF09]). In this paper, we have chosen to analyse the performance of two different techniques, named: The Sensitivity-based Negotiation Method (Sens) and The Game Theory-based Negotiation Method (GT).

### 3 EXPERIMENTAL RESULTS AND DISCUSSION

In this section, we will present experimental results for the three approaches defined above, which are based on geometric (approach 1), texture (approach 2) and both geometric and texture (approach 3) features. We will analyse the proposed gender prediction approaches with respect to both the accuracy achieved and the execution time incurred at the classification stage, after the features were extracted and selected, using a Pentium IV computer with 2.40 GHz processor and 2048 MB RAM. The classifiers were implemented in Java.

For approach 1, all iris samples in the dataset are processed to form the biometric templates, passing through the steps of segmentation, geometric feature extraction and correlation as described in Section 2. Here, highly correlated features are designated as those with a correlation value greater than 0.4 ( $-0.4 \leq \rho \leq 0.4$ ) as in [EFDCA13]. These features are discarded. The remaining uncorrelated 5 features (*GF1 - GF4*, and *GF8*) are used to form a feature vector for each iris sample in the dataset (with size of  $1 * 5$ ). For approach 2, all iris samples in the dataset are processed to form the biometric templates, passing through the steps of segmentation, normalisation and texture feature extraction as described in Section 2. Six texture-related features are used to form a feature vector for each iris sample in the dataset (with size of  $1 * 780$ ). For approach 3, the geometric and texture features from approach 1 and approach 2 are combined to form a feature vector for each iris sample in the dataset (with size of  $1 * 785$ ).

An initial experiment is performed to test the accuracy achieved and the execution time incurred at the classification stage of the proposed prediction approaches by using the defined feature vectors. The results are shown in Table 2.

Approach	Results	SVM	MLP	Jrip	KNN	DT
1	ACC (%)	55.68	57.86	56.64	49.61	56.81
	ET (sec)	0.39	0.98	0.29	0.37	0.51
2	ACC (%)	65.68	67.86	56.03	59.61	66.81
	ET (sec)	1.47	0.49	0.41	0.74	1.47
3	ACC (%)	81.43	76.64	64.51	73.72	81.43
	ET (sec)	1.97	0.89	1.27	1.31	1.97

Table 2: Accuracy (ACC) and execution time (ET) of individual classifiers

The results obtained show that approach 2 (texture features) achieves a better prediction accuracy rate than approach 1 (geometric features) with all classifiers (except the Jrip classifier) while approach 1 completes the classification stage with lower execution time than approach 2 with all classifiers. This suggests that texture features provide more useful information for the gender prediction task. The results also show that approach 3 achieves the best error-rate performance with all classifiers, but with the highest execution time. Of course, this result is not surprising, since approach 3 is the combination of approach 1 and approach 2 (i.e. adopts both geometric and texture features).

Considering these results further from the classification perspective, it is unsurprising to note that different classifiers return the best performance for different approaches, since they perform solution space search in different ways. However, it is very encouraging to see that these initial results for the process of gender prediction from iris images show that our approaches can outperform the systems previously described in the literature, where peak accuracy currently reported is typically around 75-80% [Tho07].

Hence, following these observations, and in order better to exploit the full potential of using the chosen geometrical and texture features, a second experiment is performed to investigate the attainable accuracy and execution time of the proposed gender prediction approaches when using the defined feature vectors with the combination-based classifiers presented in Section 2, with respect to the adopted dataset. The results obtained are shown in Table 3.

Approach	Results	GT	Sens	Sum	Vote	Bagging
1	ACC (%)	70.89	72.46	69.23	59.18	59.72
	ET (sec)	1.83	1.96	0.86	1.31	0.54
2	ACC (%)	72.46	75.96	70.86	70.30	68.00
	ET (sec)	2.05	2.37	1.42	1.47	1.09
3	ACC (%)	87.31	89.74	85.39	85.03	71.24
	ET (sec)	2.84	2.59	1.99	1.84	1.58

Table 3: Accuracy (ACC) and execution time (ET) of combined based classifiers

Thomas et al. [Tho07], reported around 80% accuracy by using a multiclassifier bagging with the C4.5 approach. In the work presented here, the proposed iris based gender prediction approach 1 uses only five simple geometric features of iris images and can reach accuracies close to 73% within approximately 2 seconds for classification (with the multiagent system using negotiation). Also our approach 3, which adopts both geometric and texture features as in [Tho07], is able to reach accuracies close to 90% within approximately 3 seconds using also the multiagent system.

## 4 CONCLUSION

In this paper we have investigated experimentally three approaches to gender prediction from iris images which use a combination of a small number of very simple (and therefore easily and efficiently computable) geometric features (ignoring texture-based information), or which uses texture features alone, or which uses both geometric and texture

features. By also adopting an intelligent classification structure, which we have previously found to be especially well suited to more conventional identity prediction from biometric data, we have developed a particularly effective gender prediction approach. Thus, our study has investigated how performance is influenced by the choice of the types of features used, and we have shown how implementing a more flexible and "intelligent" classification technique can support more efficient prediction using smaller number of features.

The performance we have been able to achieve - assigning each tested subject to one of two gender groups (corresponding to male and female categories) in relation to prediction accuracy, even with a small and limited feature set, is seen to be comparable to that reported elsewhere for the prediction of a gender determination problem, but which used a much larger and more diverse feature set. This comparative study based on different feature sets (i.e. geometric, texture and both geometric and texture features) and different classification approaches, provides valuable information to inform and guide the choice of feature and classification approaches in relation to particular application requirements.

This is a very positive outcome in a task domain which has been relatively little investigated to date. Although further work can still be carried out to improve and enhance the levels of achievable performance, our reported results show real promise in relation to the suitability of our basic techniques for application to a number of practical scenarios of importance and considerable current interest.

## References

- [AF09] M.C.C. Abreu and M.C. Fairhurst. Analysing the Benefits of a Novel Multiagent Approach in a Multimodal Biometrics Identification Task. *IEEE Systems Journal*, 3(4):410–417, 2009.
- [Ary98] A. Arya. An optimal algorithm for approximate nearest neighbors searching fixed dimensions. *Journal of ACM*, 45(6):891–923, 1998.
- [BB96] L. Breiman and L. Breiman. Bagging predictors. In *Machine Learning*, pages 123–140, 1996.
- [DCAF11] M.C. Da Costa-Abreu and M.C. Fairhurst. Combining multiagent negotiation and an interacting verification process to enhance biometric-based identification. In *The COST 2101 European conference on Biometrics and ID management*, pages 95–105, 2011.
- [EF11] M. Erbilek and M.C. Fairhurst. Evaluating iris segmentation for scenario optimisation. In *The 4th International Conference on Imaging for Crime Detection and Prevention*, pages 1–6, 2011.
- [EFDCA13] M. Erbilek, M. Fairhurst, and M. Da Costa-Abreu. Age Prediction from Iris Biometrics. *IET Conference Proceedings*, 2013.
- [EFDCA14] M. Erbilek, M.C. Fairhurst, and M. Da Costa-Abreu. Analysis of physical ageing effects in iris biometrics. In *IEEE International Conference of the Biometrics Special Interest Group (BIOSIG)*, 2014.
- [FAE08] M.S. Fahmy, A.F. Atyia, and R.S. Elfouly. Biometric Fusion Using Enhanced SVM Classification. In *The International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, pages 1043–1048, 2008.

- [FD12] G. Farinella and J. Dugelay. Demographic classification: Do gender and ethnicity affect each other? In *International Conference on Informatics, Electronics Vision*, pages 383–390, 2012.
- [FE11] M. Fairhurst and M. Erbilek. Analysis of Physical Ageing Effects in Iris Biometrics. *IET Computer Vision*, 5(6):358–366, 2011. Special issue on Future Trends in Biometric Processing.
- [FW94] J. Furnkranz and G. Widmer. Incremental Reduced Error Pruning. In *Proceedings the 11st International Conference on Machine Learning*, pages 70–77, 1994.
- [Hay99] S. Haykin. *Neural networks: a comprehensive foundation*, volume 13. Cambridge University Press, 1999.
- [KA03] J. Kittler and F. M. Alkoot. Sum versus vote fusion in multiple classifier systems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(1):110–115, 2003.
- [Kun04] L.I. Kuncheva. *Combining Pattern Classifiers: Methods and Algorithms*. Wiley-Interscience, 2004.
- [LB11] S. Lagree and K.W. Bowyer. Predicting ethnicity and gender from iris texture. In *IEEE International Conference on Technologies for Homeland Security*, pages 440–445, 2011.
- [Mas03] L. Masek. Recognition of Human Iris Patterns for Biometric Identification. Bachelor of engineering degree of the school of computer science and software engineering, The University of Western Australia, 2003.
- [Met07] F. et. al. Metze. Comparison of Four Approaches to Age and Gender Recognition for Telephone Applications. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, volume 4 of *ICASSP 2007*, pages 1089–1092, 2007.
- [OG10] J. et al. Ortega-Garcia. The Multiscenario Multienvironment BioSecure Multimodal Database (BMDB). *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32:1097–1111, 2010.
- [PDVV11] C. Peersman, W. Daelemans, and L. Van-Vaerenbergh. Predicting Age and Gender in Online Social Networks. In *The 3rd International Workshop on Search and Mining User-generated Contents*, pages 37–44, 2011.
- [Qui93] J.R. Quinlan. *C4.5: programs for machine learning*. Morgan Kaufmann Publishers Inc., 1993.
- [Tho07] V. et al. Thomas. Learning to predict gender from iris images. In *The 1st IEEE International Conference on Biometrics: Theory, Applications, and Systems*, pages 1–5, 2007.
- [TPB15] J.E. Tapia, C.A. Perez, and K.W. Bowyer. Gender Classification from Iris Images Using Fusion of Uniform Local Binary Patterns. In *Computer Vision - ECCV 2014 Workshops*, volume 8926 of *Lecture Notes in Computer Science*, pages 751–763. 2015.

# A Comparative Study on Image Hashing for Document Authentication \*

Dominik Klein<sup>†</sup>

Bundesamt für Sicherheit in der Informationstechnik, 53133 Bonn  
dominik.klein@bsi.bund.de

Jan Kruse<sup>‡</sup>

Hochschule Emden/Leer, 26723 Emden Stadt  
kruse.jan.g@gmail.com

**Abstract:** A digital seal is a cryptographically signed 2D barcode printed on a document to verify the document’s authenticity and integrity. In order to secure a printed image against tampering, a compact *image hash* can be stored inside the barcode. We investigate and experimentally evaluate various methods from the areas of image hashing and face verification w.r.t. its appropriateness for such an image hash that is resistant to a print-scan transformation. Exhaustive experiments with genuine security paper show that compressed local binary patterns (LBPs) and feature extraction by DCT-II perform best in this setting.

## 1 Introduction

We are concerned with a very specific and practical scenario of image hashing to verify the authenticity of a document. Consider a paper document which has important information printed on it and is usually protected against forgery by physical properties of the paper. Such documents are used in various domains, here we focus on visa stickers as an example. A simple, yet highly effective way to ensure integrity and authenticity of the printed information is the use of a *digital seal*<sup>1</sup>. A cryptographic signature of the printed information is created using a private key, and this cryptographic signature is encoded into a two-dimensional barcode and printed on the visa itself. The corresponding public key is distributed, and any interested party can verify the printed information on the visa sticker by reading the barcode and verifying the signature with the public key.

To prevent impersonation attacks (person B uses a visa issued to person A), a facial image of the visa holder is usually printed on the sticker. Barcodes can store up to approximately

---

\*Portions of the research in this paper use the FERET database of facial images collected under the FERET program, sponsored by the DOD Counterdrug Technology Development Program Office

<sup>†</sup>Corresponding author

<sup>‡</sup>The research presented here is based on the second author’s Bachelor thesis *Authentifizierung von Lichtbildern durch Image Hashing*, 2015.

<sup>1</sup>BSI TR-03137: Optically Verifiable Cryptographic Protection of non-electronic Documents (Digital Seal)

3000 bytes, but the usable capacity depends on the available physical size, and practically it is impossible to store a photo inside. One solution to detect tampering of the printed image is to create a *perceptual image hash* of the image. Such a hash is much smaller in size compared to the original image and can thus be stored inside the cryptographically signed barcode. To verify, one can scan or photograph the image from the document, compute its image hash, and compare that image hash against the digitally signed image hash of the enrollment image that is stored inside the barcode. Note that this scenario is both *offline* and *local*, and hence distinct from any client-server based method for tamper detection, where information about, or the document itself, is stored in a central database.

The desired property of an image hash function is to always map perceptually similar images to similar hash values, whereas perceptually different images should be always mapped to different hash values. Two hash values are then compared w.r.t. some measure, and considered to be the same, if their distance is below a threshold. Our notion of perceptual difference is inspired by the above scenario: We consider two images to be perceptually similar, if both are the same, or one image is the result of printing and then re-scanning the other one. Two images are considered perceptually different, otherwise. To construct image hashing functions in practice, image invariant features are extracted from the image, and a difference measure compares these extracted features.

Typical tampering is cutting out and replacing the image from a genuine visa. To prevent that, the hash ideally has second preimage resistance and collision resistance. Even if this is practically infeasible, by just ensuring that the number of collisions is low and the image can no longer be replaced by another *arbitrary* one, typical fraud scenarios are prevented.

The contribution of this paper is: 1) We survey different feature extraction methods and distance measures and identify those suitable for the considered scenario, and 2) we perform exhaustive experiments to compare and evaluate their performance under practical conditions. Our implementation is freely available<sup>2</sup>. This paper is structured as follows: In Section 2, related work and foundations for our experiments are identified. The experimental setup, including feature extraction methods and distance measures, is described in Section 3, and results are provided in Section 4. We conclude the discussion in Section 5.

## 2 Foundations and Related Work

Two main areas of research are related to our setting: The first area is (*perceptual*) *image hashing*. The goal is to construct a perceptual hash function that takes an image as input, and maps it to a compact hash value such that perceptually similar images always map to similar values, and to different values if the images are perceptually different. To construct such a function, usually different image invariant features are extracted, for example by mapping to the frequency domain, by using the Radon Transform [DRDVL05, WZN09], or by using Zernike Moments [RLB09]. It is open whether these approaches are suitable here, since robustness of hashes is often measured only w.r.t. moderate digital modifications, whereas in our setting artifacts are quite severe (cf. Figure 1). Often, key-based hash functions (akin to MACs) are considered, where the input of the hash function is not only an image, but also a secret key. However, secure distribution of a symmetric secret key to

---

<sup>2</sup><http://www.github.com/d-klein>

all interested third parties for verification is practically impossible in our scenario.

The second area is face recognition and face verification [HRBLM07]. Here, one usually tries to extract image invariant features from images with a specific focus on capturing the essential features of faces. In most methods, the size of the extracted features by far exceed the capacity of a barcode, and it is non-trivial to shrink the size of extracted features without severely affecting performance. In our setting we do not strive for face recognition, but only consider tamper-detection of documents; hence lighting, pose and expression will be the same in both the enrollment image and the scan; only the print-scan transformation introduces a measurable difference. Procedures that classify two different poses of the same person as different are no problem in our setting, and in fact desirable as this implies the ability to detect even slightest tampering with the document.

### 3 Experimental Setup

For our setup, we prepared sets of facial images (“enrollment images”) and compute their hashes by a feature extraction method. Next, we print and scan the images, optionally pre-process the scans to enhance the quality, and compute hash values of the scans. Next, we compare hash values of the enrollment images and the scans by some distance measure.

#### 3.1 Image Sets and Processing of Scans

We took two image sets of front images from the FEI Face Database<sup>3</sup> and the Color FERET database [PWHR98, PMRR00]. Both image sets do not satisfy requirements set by ICAO for biometric face images – but in practice, these requirements are often not adhered to anyway. Thus from the FEI Face Database the subset frontal images, manually aligned, was used, and from FERET subset *fa* and subset *fb* those frontal images were selected that meet best the above mentioned requirements, especially w.r.t. lighting. Image contrast was then enhanced using IrfanView’s<sup>4</sup> auto-enhancement, the images were printed on genuine Schengen-visa stickers using an off-the-shelf inkjet printer, and then re-scanned at 600dpi using a Kyocera TASKalfa 420i office scanner. After sorting out misprints, we were left with 368 images from FEI from 184 persons (one smiling, one with a neutral expression), and 768 images from FERET from 768 persons (one image per person).



**Figure 1:** FEI, Original and Scan

	Accuracy	Size
$DCT(8x8)_{raw}^{mh}$	0.922	500
$DCT(11x11)_{raw}^{mh}$	0.921	1000
$DCT(blocks)_{raw}^{mh}$	0.885	2000
$LBP_{raw}^{fb,x^2}$	0.961	1350
$LBP_{raw}^{binary,x^2}$	0.906	100

**Table 1:** Accuracy vs. Size (~ Bytes)

<sup>3</sup><http://fei.edu.br/~cet/facedatabase.html>

<sup>4</sup><http://www.irfanview.com>

An example of FEI is displayed in Figure 1: Printing and scanning introduces excessive noise, changes the aspect-ratio (due to a processing step in the visa-printing software) and adds a different background. Also, random guilloche lines reach over the image for security.

Having scanned the images, our experimental setup consists of the following steps: First, the face from the print-scan of the whole visa-sticker is extracted using a Haar cascade classifier, and automatically cropped to resemble the enrollment image. Optionally, some preprocessing of the images is applied to remove artifacts. Next, a feature vector is extracted from each image. Then, given one feature vector from an enrollment image, and one feature vector from a print-scanned image, the two feature vectors are compared by some metric to yield a score value.

Depending on the classification task, the score value is used either to classify the scanned image as being a print-scan of the enrollment image, or as a different image (image verification), or the score value is used to identify the corresponding print-scanned image among a set of scans (image recognition).

Feature extraction methods have different resistance to artifacts. Given a scan, we consider the following additional steps before extracting features of scanned images: a) No preprocessing, b) a slight Gaussian blur (as this effectively removes visible guilloche lines without losing too much sharpness), and c) the complete image pre-processing pipeline that has been proposed in [TT07]. After that, the image is resized to dimension  $N \times N$  and converted to grayscale. The result is the input of the feature extraction algorithm.

### 3.2 Feature Extraction and Hash Comparison

The following feature extraction methods were used in the experiment. None depends on a symmetric key, and all are suitable to create very compact hash values.

**Discrete Cosine Transform** Switching from the spatial domain to the frequency domain is a technique frequently applied in various areas of image processing, and the DCT-II has been used for image hashing in the open-source tool `pHash`<sup>5</sup>. Similar, we take as the hash value the first  $(8 * 8) - 1$  values (lower frequencies) of the DCT-II of a given image.

**Radon Transform: RASH** The radial variance based hash (RASH) was introduced by De Roover et al. in [DRDVLM05]. We directly apply their method, but do not threshold at the end and instead take the resulting real values as the hash.

**Radon Transform: Method by Wu et al.** For RASH, an approximation for the projection line is used to compute a hash inspired by the Radon transform. The method by Wu et al. [WZN09] uses the radon transform more directly. Some implementation details are left out or are ambiguous in their description of the algorithm. In our implementation we use  $40 * 10$  blocks and employ the radon transform approximation of `skimage`<sup>6</sup>.

**Zernike Moments (ZM)** Zernike Moments are image-invariant features that are especially robust to rotation, which makes them appealing in our setting. Their use was explored for

---

<sup>5</sup><http://www.phash.org>

<sup>6</sup><http://scikit-image.org>

image recognition in [TC88, KH90]. **Pseudo Zernike Moments (PZM)** are adapted from Zernike Moments and less sensitive to noise [TC88]. In both cases, the absolute values of moments upto some chosen boundary are taken as the hash value.

**Local Binary Patterns** The use of local binary patterns (LBPs) for face recognition has been pioneered in [AHP04]. Since taking raw LBPs as hash values is not suitable due to their large size, we adapt them here by reducing dimensionality with: (1) Histograms with fewer bins: We interpret each local binary pattern as an unsigned integer value. Then, in a somewhat ad-hoc approach, the number of bins in each local LBP histogram is reduced. (2) PCA: We take a training set of face images, compute the LBP hash for each image to get a high dimension feature vector, and perform principal component analysis (PCA) to compute a *static* PCA basis for LBPs of face images. For an arbitrary image (not related to the training set) the hash is computed by mapping its LBPs into PCA space.

**Eigenfaces (EF)** For LBPs, we mentioned how PCA can be used to reduce the dimensionality of a feature vector. The original Eigenface approach [TP91] used the pixel values directly as features to perform PCA. Whereas Eigenfaces are nowadays not competitive in arbitrary face verification [HRBLM07], the appealing point of this approach is the low-dimensionality, and hence low storage requirements of the hash. Also, the weaknesses w.r.t. arbitrary face recognition (different facial expressions, different lightning etc.) become less important in our setting, since those parameters do not change for the print-scan compared to the enrollment image. To handle *both* image verification and image recognition we adapt PCA by selecting a small set of images to learn an overall *fixed* PCA basis for faces. These training images are not part of the set of enrollment images and scans. A hash for an image is then generated by projecting a query image into PCA space.

**Distance Measures for Hash Comparison** It is sometimes suggested (e.g. [Zau10]) to further compact the hash by thresholding over the mean and comparing by Hamming distance. This however resulted in poor performance in our setting. Instead we consider the following distance measures to compare hashes: manhattan distance, euclidian distance, and peaks of cross correlation (PCC, c.f. [Zau10]). These measures are not suitable for LBPs, since those consist of concatenated histograms. Hence, we use the  $\chi^2$  distance (c.f. [AHP04]) and a weight of 1 for each region.

## 4 Experimental Results

First we consider the *image recognition problem* to get a notion on which feature extraction method works best in our context, which distance measure is optimal, and which pre-processing steps are best in reducing the artifacts introduced by printing and scanning. Given the hash of an enrollment image, we query the set of scans for the corresponding image. If correctly identified, we record a hit, otherwise a miss. The results are depicted in Table 2. Here experiments were run for the FERET dataset, for the FEC dataset, and both combined. For eigenfaces and PCA-based LBPs, we used 200 images from FEC as training data for FERET, 200 images from FERET as training data for FEC, and we split the combined set of 1136 images into 200 for training, and 936 for recognition.

	FERET	FEC	combined		FERET	FEC	combined
DCT <sub>raw</sub> <sup>mh</sup>	0.935	0.899 (0.967)	0.922 (0.944)	RASH <sub>raw</sub> <sup>mh</sup>	0.641	0.609 (0.698)	0.620 (0.643)
DCT <sub>raw</sub> <sup>pcc</sup>	0.951	0.891 (0.957)	0.929 (0.951)	RASH <sub>raw</sub> <sup>pcc</sup>	0.704	0.802 (0.883)	0.715 (0.739)
DCT <sub>raw</sub> <sup>auc</sup>	0.927	0.878 (0.957)	0.909 (0.936)	RASH <sub>raw</sub> <sup>auc</sup>	0.482	0.242 (0.342)	0.388 (0.438)
DCT <sub>gauss</sub> <sup>pcc</sup>	0.951	0.891 (0.957)	0.932 (0.949)	RASH <sub>gauss</sub> <sup>pcc</sup>	0.721	0.813 (0.889)	0.732 (0.757)
DCT <sub>tt</sub> <sup>pcc</sup>	0.928	0.832 (0.910)	0.893 (0.918)	RASH <sub>tt</sub> <sup>pcc</sup>	0.831	0.633 (0.701)	0.750 (0.768)
WU <sub>raw</sub> <sup>mh</sup>	0.747	0.840 (0.899)	0.767 (0.786)	ZM <sub>raw</sub> <sup>mh</sup>	0.673	0.829 (0.867)	0.717 (0.729)
WU <sub>raw</sub> <sup>pcc</sup>	0.876	0.883 (0.938)	0.872 (0.890)	ZM <sub>raw</sub> <sup>pcc</sup>	0.746	0.774 (0.807)	0.742 (0.752)
WU <sub>raw</sub> <sup>auc</sup>	0.724	0.829 (0.883)	0.745 (0.761)	ZM <sub>raw</sub> <sup>auc</sup>	0.408	0.785 (0.837)	0.521 (0.537)
WU <sub>gauss</sub> <sup>pcc</sup>	0.891	0.899 (0.946)	0.886 (0.901)	ZM <sub>gauss</sub> <sup>pcc</sup>	0.755	0.764 (0.796)	0.748 (0.759)
WU <sub>tt</sub> <sup>pcc</sup>	0.888	0.894 (0.954)	0.884 (0.902)	ZM <sub>tt</sub> <sup>pcc</sup>	0.693	0.582 (0.628)	0.639 (0.654)
PZM <sub>raw</sub> <sup>mh</sup>	0.531	0.772 (0.807)	0.596 (0.607)	EF <sub>raw</sub> <sup>mh</sup>	0.711	0.867 (0.948)	0.709 (0.753)
PZM <sub>raw</sub> <sup>pcc</sup>	0.697	0.750 (0.791)	0.688 (0.701)	EF <sub>raw</sub> <sup>pcc</sup>	0.538	0.840 (0.946)	0.586 (0.616)
PZM <sub>raw</sub> <sup>auc</sup>	0.247	0.736 (0.791)	0.391 (0.408)	EF <sub>raw</sub> <sup>auc</sup>	0.464	0.818 (0.935)	0.549 (0.584)
PZM <sub>gauss</sub> <sup>pcc</sup>	0.712	0.742 (0.780)	0.703 (0.715)	EF <sub>gauss</sub> <sup>mh</sup>	0.837	0.867 (0.943)	0.780 (0.816)
PZM <sub>tt</sub> <sup>pcc</sup>	0.634	0.595 (0.639)	0.592 (0.605)	EF <sub>tt</sub> <sup>mh</sup>	0.784	0.736 (0.867)	0.776 (0.811)
LBP <sub>raw</sub> <sup>fb,χ<sup>2</sup></sup>	<b>0.975</b>	0.938 (0.967)	<b>0.961 (0.970)</b>	LBP <sub>raw</sub> <sup>pca,mh</sup>	0.811	0.791 (0.829)	0.792 (0.808)
LBP <sub>raw</sub> <sup>pca,pcc</sup>	0.652	0.742 (0.772)	0.618 (0.624)	LBP <sub>raw</sub> <sup>pca,auc</sup>	0.658	0.679 (0.712)	0.667 (0.674)
LBP <sub>gauss</sub> <sup>pca,χ<sup>2</sup></sup>	0.962	<b>0.962 (0.986)</b>	0.961 (0.969)	LBP <sub>tt</sub> <sup>fb,χ<sup>2</sup></sup>	0.958	0.913 (0.962)	0.944 (0.960)

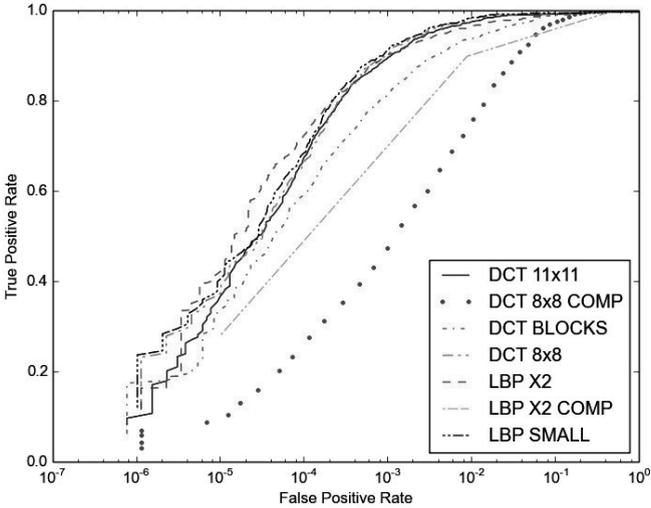
**Table 2:** Evaluation for Image Recognition (Recognition Rate).

Results in brackets denote the recognition score when one does not distinguish between images of the same person. For example if the hash originates from an image with a person smiling, and the lowest distance is to an image of that same person not smiling, we still count that as a hit. For preprocessing, *raw* denotes none, *gb* denotes a slight gaussian blur, and *tt* the pipeline of [TT07]. For distances, *mh* is manhattan distance, *auc* euclidian distance, and *pcc* peaks of cross correlation. For LBPs, *fb* means fewer bins, and *pca* size reduction by PCA. To keep the number of combinations feasible, we first identify the best distance measure for each method without preprocessing. Then taking the best performing measure, we test w.r.t. the best performing preprocessing. Abbreviations for feature extractions are defined in Subsection 3.2.

Clearly, local binary patterns outperform any other feature extraction method. Somewhat surprising is that the very simple and straight-forward transformation into frequency domain by DCT-II is the second best feature. The method by Wu et al. performs adequate if an appropriate distance measure (PCC) is chosen. However their method based on the radon transform, as well all other methods, were apparently designed to cope with rather minor distortions compared to our setting, and perform much worse than LBPs and DCT.

Each of these feature extraction methods results in a different hash size. It would be desirable to measure their performance for equal hash sizes, but for a given method it is non-trivial to arbitrarily decrease the resulting size, and trivially increasing the size needs not to result in better performance. We investigate accuracy vs. size for the best performing methods LBP and DCT-II: Increasing the compact DCT-II is done (a) by taking  $11 \times 11 - 1 = 120$  DCT-II coefficients instead 63, and (b), akin to JPEG compression, by

dividing images into 4x4 blocks, computing 63 DCT-II coefficients for each block, and concatenating the result. To generate a smaller hash from LBPs, we thresholded each resulting local histogram by its mean to generate a binary stream, and used hamming distance to compare two hashes. Results are depicted in Table 1. Considering higher frequency information by taking more DCT-II information apparently picks up more noise and distortions, and does not lead to more accuracy. On the other hand, LBPs have much discriminatory power and there seems to be much potential to create a more compact hash.



**Figure 2:** ROC for Image Verification

For all methods from Table 1, receiver operating characteristics (ROC) are depicted in Figure 2. It is not easy to turn the comparatively good results for image recognition into a good performance for *image verification*, i.e. distance values are such that it is possible to identify the corresponding image among a set by lowest distance, but difficult to derive a numerical global threshold for the *image verification* problem (given an arbitrary pair, decide whether they belong together or not). A powerful method is *One Shot Similarity* [WHT09]: Given two arbitrary images  $x$  and  $y$ , just compare them w.r.t. to a set of images  $B$  for which it is known that each element of  $B$  shows a different person than both  $x$  and  $y$ , and train two classifiers based on this idea. Instead of training classifiers, we adopted this idea in a somewhat ad-hoc manner here (marked COMP in Figure 2), by splitting the combined test set into training and recognition data, and measuring distance between an enrollment image and a scan by first computing the minimal distance between the enrollment image and all training samples. The final distance is then the difference between the distance of scan and image minus the computed minimal distance to the training images. Unfortunately, this ad-hoc approach does not result in a better TPR/FPR ratio.

## 5 Conclusion and Future Work

We identified several potentially suitable hash methods for tamper detection. Security implications were only briefly touched by considering a set with smiling and non-smiling faces, and hence thorough evaluation of the security of the hashes is future work. We anticipate that OSS [WHT09] can be adapted for better performance for image verification.

## References

- [AHP04] T. Ahonen, A. Hadid, and M. Pietikäinen. Face Recognition with Local Binary Patterns. In *Proc. ECCV*, volume 3021 of *LNCS*, pages 469–481, 2004.
- [DRDVLM05] C. De Roover, C. De Vleeschouwer, F. Lefebvre, and B. Macq. Robust image hashing based on radial variance of pixels. In *Proc. ICIP*, pages 3,77–80, 2005.
- [HRBLM07] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller. Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments. Technical Report 07-49, Univ. of Massachusetts, 2007.
- [KH90] A. Khotanzad and Yaw H. H. Invariant image recognition by Zernike moments. *Pattern Analysis and Machine Intelligence*, 12(5):489–497, 1990.
- [PMRR00] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss. The FERET Evaluation Methodology for Face-Recognition Algorithms. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(10):1090–1104, 2000.
- [PWHR98] P. J. Phillips, H. Wechsler, J. Huang, and P. J. Rauss. The FERET database and evaluation procedure for face-recognition algorithms. *Image Vision Comput.*, 16(5):295–306, 1998.
- [RLB09] J. Revaud, G. Lavoué, and A. Baskurt. Improving Zernike moments comparison for optimal similarity and rotation angle retrieval. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31(4):627–636, 2009.
- [TC88] C.-H. Teh and R.T. Chin. On image analysis by the methods of moments. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 10(4):496–513, Jul 1988.
- [TP91] M. Turk and A. Pentland. Eigenfaces for Recognition. *J. Cognitive Neuroscience*, 3(1):71–86, January 1991.
- [TT07] X. Tan and B. Triggs. Enhanced Local Texture Feature Sets for Face Recognition under Difficult Lighting Conditions. In *Analysis and Modelling of Faces and Gestures*, volume 4778 of *LNCS*, pages 168–182, oct 2007.
- [WHT09] L. Wolf, T. Hassner, and Y. Taigman. The One-Shot similarity kernel. In *Proc. ICCV*, pages 897–902, 2009.
- [WZN09] D. Wu, X. Zhou, and X. Niu. A novel image hash algorithm resistant to print-scan. *Signal Processing*, 89(12):2415–2424, 2009.
- [Zau10] C. Zauner. Implementation and Benchmarking of Perceptual Image Hash Functions, 2010. Diploma Thesis.

# On Accuracy of Keystroke Authentications Based on Commonly Used English Words

Alaa Darabseh and Akbar Siami Namin

Department of Computer Science  
Texas Tech University  
Lubbock, TX, USA  
alaa.darabseh@ttu.edu  
akbar.namin@ttu.edu

**Abstract:** The aim of this research is to advance the user active authentication using keystroke dynamics. Through this research, we assess the performance and influence of various keystroke features on keystroke dynamics authentication systems. In particular, we investigate the performance of keystroke features on a subset of most frequently used English words. The performance of four features such as i) key duration, ii) flight time latency, iii) digraph time latency, and iv) word total time duration are analyzed. Experiments are performed to measure the performance of each feature individually as well as the results from the different subsets of these features. Four machine learning techniques are employed for assessing keystroke authentications. The selected classification methods are two-class support vector machine (TC) SVM, one-class support vector machine (OC) SVM,  $k$ -nearest neighbor classifier (K-NN), and Naive Bayes classifier (NB). The logged experimental data are captured for 28 users. The experimental results show that key duration time offers the best performance result among all four keystroke features, followed by word total time. Furthermore, our results show that TC SVM and KNN perform the best among the four classifiers.

## 1 Introduction

Biometric authentication technique concerns the use of human characteristics that make each individual unique. It involves any personal characteristics that can be used to uniquely verify a person's identity [MR00]. Biometrics are mainly classified as physiological biometrics features like fingerprint, face, iris, or behavioral biometrics features such as gait, handwritten signature, keystroke dynamics, etc.

Keystroke dynamics are defined as “*a behavioral biometric characteristic which involves analyzing a computer users' habitual typing pattern when interacting with a computer keyboard*”[MR00]. There are several benefits of using keystroke dynamics: First, keystroke dynamics are practical and feasible, since every computer user types on a keyboard; Second, it is inexpensive due to the fact that it does not require any additional or special tools nor components; Thirdly, typing rhythms can be still available even after the authentication phase has passed [GP05].

In building a keystroke-based authentication system, a number of features or measures pertinent to individuals typing styles are used. These features are one of the most important factors that may influence the performance and error rates of keystroke-dynamic detectors [KM10]. Hence, the proper selection of features plays an important role in enhancing the performance of such authentication system when adapting keystroke dynamic detectors.

This paper focuses on studying the influence of various keystroke features on the keystroke dynamics authentication system performance. The major contribution of this paper is the utilization of most frequently used English words in deciding about authenticating users when typing. Our keystroke authentication scheme captures necessary features such as latencies and duration times to determine which timing feature performs better in keystroke dynamics. The promising results demonstrate the performance accuracy of the proposed authentication approach.

The remainder of the paper is organized as follows: Section 2 highlights the motivation and the contributions of this paper. Section 3 describes our experiment procedure, having components for data capture, feature extraction, and classification. Section 4 describes the experiments and presents the experimental results. Section 5 presents the conclusions and future work.

## 2 Motivation and Contributions

When a person types on a keyboard there are two main timing events that occur: 1) the key down event when the person presses a key, and 2) the key up event when a person releases a key. Timestamps of each event are recorded to keep track of pressing and releasing a key. A variety of timing features can then be extracted from this timing information. Two of the most used features are 1) *duration of the key*, which is the time in which the key is being held down, and 2) *keystroke latency*, which is the time between two successive keystrokes.

Many different methods can be used to calculate latency. The most commonly used methods are: *press-to-press* (PP) latency, which is the time interval between consecutive key presses, PP is also called *digraph time*, *release-to-press* (RP) latency, which is the time interval between releasing the key and pressing the next one, RP is also called *flight time*, and *release-to-release* (RR) latency which is the time interval between releases of two consecutive keys.

It is also possible to capture other keystroke dynamic information such as the time it takes to write a word, two letters (digraph) and three letters (tri-graph).

Keystroke features are one of the most important factors that may influence the error rates of keystroke-dynamic detectors [KM10]. The process of feature selection plays a critical role in improving the performance when designing keystroke dynamic detectors. Revett et al [RGG<sup>+</sup>07] states that the classification accuracy is substantially influenced by the feature selection process and to a lesser extent on the authentication algorithm employed. A recent survey of keystroke dynamics perceives that certain features have a tendency to be more helpful than others [BW12].

Existing works in the literature of keystroke dynamics demonstrate conflicting results regarding which feature is the most effective timing feature in keystroke dynamics domain. The existing works indicate that duration times are more important than latencies times in reducing false positive error rates [TYT12, RLCM98]. It is also observed that using tri-graph time offers better classification results than using digraphs or higher order n-graphs [BGP02]. Revett et al [RGG<sup>+</sup>07] also reported that the digraph and tri-graph times were more effective compared to duration time and flight time.

Another observation we can conclude from the existing literature of keystroke dynamics is the lack of adequate studies on less common features such as speed and frequency of typing errors which occur during typing. These less common features may embrace effective timing information that might improve the performance of keystroke dynamics systems.

This paper focuses on studying the influence of various keystroke features on the keystroke dynamics authentication system performance. In particular, we investigate the performance of keystroke features on a subset of most 20 frequently used English alphabet letters, most 20 frequent appearing pairs of English alphabet letters, and most 20 frequent appearing of English words. Four features including key duration, flight time latency, digraph time latency, and word total time duration were analyzed using four machine learning techniques namely two different classes of support vector machines (SVM),  $k$ -nearest, and Naive Bayes. This work poses the following research questions and addresses them throughout the paper:

1. *What feature items contribute more to the accuracy of the feature?* The feature items are defined as the exact instances of letters in each feature, e.g. “a”, “Th”, etc.
2. *What timing feature (e.g. flight time) performs better in keystroke dynamics?*
3. *How the accuracy of the features is impacted from one prediction technique (e.g. SVM) to another?*
4. *How comparable is the total time to flight, digraph, and duration time when typing an entire word?*

### 3 Experimental Setup

This section describes the experimental procedure, data collection, and the keystroke features used in this experimental study.

#### 3.1 Data Collection

A VB.NET form application was developed to capture raw keystroke data samples. For each keystroke, the time (in milliseconds) when a key was pressed and the time when a

key was released, were captured. The participants had the choice of using all keys on the keyboard including special characters such as `Shift Key` and `Caps Lock`, as well as combination of keys. More importantly, they were also able to use the `Backspace` key to correct or modify their typing.

Our experiment of keystroke timing dynamic involved 28 participants. The participants were graduate students majoring in Computer Science. All participants were asked to type the same prepared English passage (5000 characters) one time. Participants had the choice to either use our laptop computer or download the application for collecting data on their own devices. The layout of the application was split into two sections. The top section displayed text that was required to be typed, i.e., fixed text. The bottom section had a space to allow the participant type the text seeing on the top. The texts used in our experiment were a collection of English sentences and passages drawn from a randomly selected pool of English passages.

### 3.2 Extracted Keystroke Features

The collected raw data from both data sets were used to extract the following features from the experimental data:

1. *Duration* (F1) of the key presses for the most 20 frequent appearing English alphabet letters (e, a, r, i, o, t, n, s, h, d, l, c, u, m, w, f, g, y, p, b) [Gai89].
2. *Flight Time Latency* (F2) for the most 20 frequent appearing pairs of English alphabet letters (in, th, ti, on, an, he, at, er, re, nd, ha, en, to, it, ou, ea, hi, is, or, te) [Gai89].
3. *Digraph Time Latency* (F3) for the most 20 frequent appearing pairs of English alphabet letters (in, th, ti, on, an, he, at, er, re, nd, ha, en, to, it, ou, ea, hi, is, or, te) [Gai89].
4. *Word Total Duration* (F4) for the most 20 frequent appearing English words (for, and, the, is, it, you, have, of, be, to, that, he, she, this, they, will, i, all, a, him) [FKF00].

Every feature (F1, F2, F3, and F4) consisted of 20 items data where feature items are the possible instance values that the underlying feature may have. For instance, F1 contained 20 English alphabet letters where each letter represented an item. Figure 1 illustrates an example of four keystroke features extracted for the word “THE.”

### 3.3 Classification Algorithms

Four machine learning techniques are adapted to classify the data and address the posed research questions. The selected classification methods are two-class support vector ma-

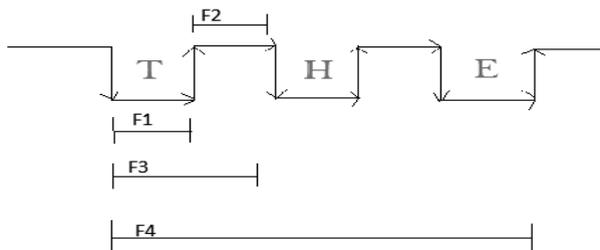


Figure 1: An example of four keystroke features extracted for the word “THE.”

chine (TC) SVM, one-class support vector machine (OC) SVM, the  $k$ -nearest neighbors classifier (K-NN), and Naive Bayes (NB). Due to the space limitations, we do not provide any discussion about mathematical and fundamental background of these algorithms.

## 4 Results

### 4.1 Experiments

Authentication system performance was evaluated by assigning each one of the 28 users in the training data set the role of a genuine user. Then, each of the remaining users played the role of the impostor. Classifiers were trained and tested to measure the accuracy of each feature’s item individually in terms of their ability to classify users. On each repetition, we kept track the number of positive and negative testing cases were correctly classified by the classifier between two selected users.

The accuracy of each feature item was then measured as the percentage of positive and negative testing cases correctly classified by the classifier of all users. In other words, accuracy of the features was measured as the percentage of patterns that were correctly mapped to their true users, and was evaluated by the following formula:

$$Accuracy = \frac{T}{N} \quad (1)$$

where  $T$  is the number of cases correctly classified and  $N$  is the total number of sample cases.

For each user, the first 25 typed timing repetitions of all feature’s items were selected to build a user’s profile. The first 20 repetitions of each feature’s item (of the 25 samples) for each user were chosen for the training phase, and the remaining 5 repetitions of each feature’s item were used for the testing phase. The results of this experiment had two major benefits: Firstly, it measured the accuracy of each feature’s item individually, so we could determine which items contributed more for the precision of the features. Secondly, we were able to compare the features’ performances by measuring the accuracy of each feature individually.

## 4.2 Results

This section reports the accuracy measures of each feature's item individually. As stated earlier, the purpose is to determine which items are more salient for the accuracy of each feature. Tables 1 - 4 illustrate the accuracy values for all feature's items in each experiment based on using TC SCM, OC SVM, KNN, and NB, respectively. We can observe that some items are performing better than the others. For instance, the accuracy value measured using TC SVM for letter "B" was 90% while the accuracy value measured for letter "R" was 73%. The high accuracy values for the feature's items suggest that these items might be good indicators for distinguishing users.

Moreover, as we can observe from Tables 1 - 4, some feature's items appear more than once in the topmost of each classifier results. For instance: for F1 feature, B, H, T, I, Y items appear in the 10 topmost-appearing items of F1 feature item. For F2 feature, IN, OU, ED, EA, TH items appear in the 10 topmost-appearing items of F2 feature items. For F3 feature, IN, OU, ED, HI, EA items appear in the 10 topmost-appearing items of F3 feature items. For F4 feature, WILL, ALL, A, AND, HAVE items appear in the 10 topmost-appearing items of F4 feature items. Similarly, These feature items that appear in the topmost suggest that these items are good indicators for distinguishing users.

Another interesting finding we can point out, by observing Tables 1 - 4, is that F4 items are performing better than F2 and F3 items, which shows that the total time of typing the most frequent used English words might contain valuable timing information that could be utilized in order to differentiate users.

Apart from the accuracy values of feature's items, Tables 1 - 4 report the overall accuracy of each feature individually by calculating the average of its items accuracy values. We use these averages to compare between the four different types of keystroke features. By observing Tables 1 - 4, we can note that by using F1, we are able to obtain a better result compared to the other keystroke features (F2, F3, and F4). Three classifiers (OC SVM, TC SCM, KNN) out of the four used classifiers agree that F1 is performing better than other features followed by F4. For instance, the average accuracy value measured using TC SVM for F1 was 84% followed by F4 with the accuracy value was 81%. Thus, we can conclude that duration times are more effective timing feature than latencies times. This result is consistent with existing works [RLCM98, TYT12].

Finally, by observing Tables 1 - 4 we can note that TC SVM and KNN perform the best among the four classifiers. Both classifiers achieved good recognition accuracy. The other two classifiers, OC SVM and NB, did not perform well.

## 5 CONCLUSION AND FUTURE WORKS

This paper reports the results of a series of experiments to study the influence of four keystroke features when using four major classifiers algorithms. The experimental results show some feature's items are contributing more to the accuracy of active authentication. Also, our experimental results show that duration time (F1) feature offers the best per-

Table 1: Items performances using TC SVM (ACC: Accuracy).

F1		F2		F3		F4	
Item	Acc	Item	Acc	Item	Acc	Item	Acc
B	0.90	IN	0.85	IN	0.86	WILL	0.87
H	0.88	OU	0.80	HI	0.82	ALL	0.86
P	0.88	ED	0.79	ED	0.81	A	0.85
T	0.88	HI	0.79	OU	0.81	AND	0.84
Y	0.88	EA	0.78	EA	0.80	HAVE	0.84
I	0.86	TH	0.77	HE	0.80	THE	0.84
N	0.86	ER	0.76	TH	0.80	TO	0.84
U	0.86	RE	0.73	TO	0.80	IS	0.83
W	0.86	TO	0.73	EN	0.79	OF	0.83
G	0.85	AT	0.70	IS	0.79	BE	0.83
L	0.85	HE	0.70	ER	0.78	FOR	0.81
M	0.84	AN	0.68	HA	0.78	IT	0.81
O	0.82	EN	0.68	OR	0.76	SHE	0.81
S	0.82	IS	0.67	RE	0.76	THIS	0.80
C	0.81	IT	0.67	AN	0.75	HE	0.79
D	0.81	ND	0.67	IT	0.75	I	0.79
F	0.80	HA	0.66	AT	0.74	YOU	0.78
A	0.79	OR	0.66	ND	0.72	THEY	0.77
E	0.76	ON	0.64	ON	0.70	THAT	0.74
R	0.73	TE	0.62	TE	0.70	HIM	0.73
AVG	0.84	AVG	0.72	AVG	0.78	AVG	0.81

Table 2: Items performances using OC SVM (ACC: Accuracy).

F1		F2		F3		F4	
Item	Acc	Item	Acc	Item	Acc	Item	Acc
Y	0.80	OU	0.73	ED	0.72	TO	0.79
L	0.79	EA	0.71	IN	0.71	HAVE	0.77
M	0.79	IN	0.70	EA	0.69	OF	0.77
T	0.79	HI	0.68	OU	0.69	THE	0.76
G	0.78	TH	0.68	HA	0.65	A	0.76
H	0.77	RE	0.67	OR	0.65	ALL	0.75
I	0.77	ER	0.65	TH	0.65	AND	0.74
N	0.77	ED	0.63	TO	0.65	YOU	0.73
O	0.77	AN	0.60	HE	0.64	WILL	0.73
U	0.76	HA	0.60	HI	0.64	FOR	0.72
W	0.75	ND	0.60	IS	0.64	IS	0.72
S	0.74	TO	0.60	IT	0.64	IT	0.72
B	0.73	IS	0.59	ND	0.64	BE	0.72
C	0.73	IT	0.59	TE	0.62	HE	0.72
F	0.73	AT	0.58	EN	0.61	THIS	0.72
D	0.72	EN	0.57	AN	0.60	I	0.72
A	0.71	TE	0.57	RE	0.59	SHE	0.71
R	0.71	HE	0.54	AT	0.58	THEY	0.69
P	0.70	ON	0.53	ER	0.58	HIM	0.69
E	0.69	OR	0.53	ON	0.54	THAT	0.64
AVG	0.75	AVG	0.62	AVG	0.64	AVG	0.73

Table 3: Items performances using KNN (ACC: Accuracy).

F1		F2		F3		F4	
Item	Acc	Item	Acc	Item	Acc	Item	Acc
B	0.86	IN	0.86	IN	0.86	A	0.91
M	0.86	OU	0.85	IS	0.84	WILL	0.90
H	0.85	EA	0.81	TH	0.83	ALL	0.88
L	0.85	ED	0.81	ED	0.82	BE	0.87
O	0.85	TH	0.81	HE	0.82	TO	0.87
A	0.84	HI	0.80	HI	0.82	I	0.87
I	0.84	RE	0.80	OU	0.82	AND	0.86
S	0.84	ER	0.79	EN	0.81	OF	0.86
T	0.84	TO	0.74	TO	0.81	THE	0.86
Y	0.84	AT	0.73	EA	0.80	HAVE	0.84
C	0.83	AN	0.72	AN	0.78	IS	0.83
D	0.83	ND	0.71	ER	0.78	IT	0.83
F	0.83	EN	0.70	RE	0.78	FOR	0.82
G	0.83	HE	0.70	AT	0.77	YOU	0.82
N	0.83	IT	0.70	HA	0.77	HE	0.82
P	0.83	HA	0.69	IT	0.77	SHE	0.82
R	0.83	IS	0.69	OR	0.76	THIS	0.82
E	0.82	OR	0.69	ND	0.74	THEY	0.82
U	0.82	ON	0.65	TE	0.71	THAT	0.74
W	0.82	TE	0.65	ON	0.70	HIM	0.74
AVG	0.84	AVG	0.75	AVG	0.79	AVG	0.84

Table 4: Items performances using NB (ACC: Accuracy).

F1		F2		F3		F4	
Item	Acc	Item	Acc	Item	Acc	Item	Acc
F	0.71	IN	0.82	IN	0.83	BE	0.79
B	0.70	TH	0.74	ED	0.78	TO	0.79
C	0.70	ED	0.73	TH	0.78	WILL	0.79
P	0.70	ER	0.73	HE	0.77	ALL	0.79
H	0.69	HI	0.73	HI	0.77	HAVE	0.78
O	0.69	EA	0.70	TO	0.76	THE	0.78
T	0.69	HE	0.70	EA	0.75	OF	0.77
I	0.68	OU	0.70	OU	0.75	A	0.77
G	0.67	RE	0.68	HA	0.74	AND	0.76
R	0.67	TO	0.67	EN	0.73	IS	0.76
S	0.67	HA	0.65	ER	0.73	THIS	0.75
U	0.66	IS	0.65	IS	0.73	HIM	0.74
L	0.65	AN	0.63	AN	0.71	FOR	0.73
W	0.65	IT	0.63	RE	0.71	IT	0.73
Y	0.65	OR	0.63	IT	0.71	YOU	0.73
D	0.64	AT	0.62	OR	0.69	SHE	0.73
M	0.64	EN	0.61	RE	0.69	I	0.71
A	0.63	ND	0.61	ND	0.67	THAT	0.70
E	0.63	ON	0.59	ON	0.66	HE	0.70
N	0.61	TE	0.59	TE	0.65	THEY	0.66
AVG	0.66	AVG	0.67	AVG	0.73	AVG	0.74

formance result among all four keystroke features, followed by word duration time (F4). Lastly, the paper introduced new features' items that could be effectively used in keystroke dynamics domain. Future work will involve more classifiers and trying to train the classifiers using less samples sizes and test the impact of that on the performance accuracy.

## References

- [BGP02] Francesco Bergadano, Daniele Gunetti, and Claudia Picardi. User Authentication Through Keystroke Dynamics. *ACM Trans. Inf. Syst. Secur.*, 5(4):367–397, November 2002.
- [BW12] Salil Banerjee and Damon Woodard. Biometric authentication and identification using keystroke dynamics: A survey. *Journal of Pattern Recognition Research*, 7(1):116–139, 2012.
- [FKF00] E.B. Fry, J.E. Kress, and D.L. Fountoukidis. *The Reading Teachers Book of Lists*. Jossey-Bass, 3rd edition, 2000.
- [Gai89] H.F. Gaines. *Cryptanalysis: A Study of Ciphers and Their Solution*. Dover Publications, Dover, New York, 1989.
- [GP05] D. Gunetti and C. Picardi. Keystroke analysis of free text. *ACM Trans. Inf. Syst. Secur.*, 8(3):312–347, 2005.
- [KM10] Kevin Killourhy and Roy Maxion. Why Did My Detector Do That?: Predicting Keystroke-dynamics Error Rates. In *Proceedings of the 13th International Conference on Recent Advances in Intrusion Detection (RAID'10)*, pages 256–276. Springer-Verlag, 2010.
- [MR00] Fabian Monrose and Aviel D. Rubin. Keystroke dynamics as a biometric for authentication. *Future Generation Comp. Syst.*, 16(4):351–359, 2000.
- [RGG<sup>+</sup>07] K. Revett, F. Gorunescu, M. Gorunescu, M. Ene, S. Magahaes, and H. Santos. A machine learning approach to keystroke dynamics based user authentication. *International Journal of Electronic Security and Digital Forensics*, 1(1):55–70, 2007.
- [RLCM98] J. A. Robinson, V. M. Liang, J. A. M. Chambers, and C. L. MacKenzie. Computer User Verification Using Login String Keystroke Dynamics. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, 28(2):236–241, March 1998.
- [TYT12] Pin S. Tech, Shigang Yue, and Andrew B.J. Teoh. Feature Fusion Approach on Keystroke Dynamics Efficiency Enhancement. *International Journal of Cyber-Security and Digital Forensics (IJCSDF)*, 1(1), 2012.

# Towards Touchless Palm and Finger Detection for Fingerprint Extraction with Mobile Devices\*

Christof Jonietz<sup>1</sup>, Eduardo Monari<sup>1</sup>, Chengchao Qu<sup>2,1</sup>

<sup>1</sup>Fraunhofer IOSB

<sup>2</sup>Vision and Fusion Laboratory (IES), Karlsruhe Institute of Technology (KIT)  
{firstname.lastname}@iosb.fraunhofer.de

**Abstract:** In this paper, contactless palm and finger detection for biometric fingerprint verification/identification process with mobile devices is considered. In order to speed up the border checking verification process, we focus on capturing the whole palm in order to extract each fingertip instead of successively capturing each fingertip. The workflow comprises palm detection in order to detect the skin region within the image prior to detection of fingertips. A machine learning based algorithm with Aggregated Channel Features (ACFs) adopted for palm detection is considered. Furthermore, a geometric shape based approach for fingertip detection has been designed to reconstruct long lines along fingers. Results demonstrate the performance of both algorithms.

## 1 Introduction

Automatic Border Control (ABC) will allow border control authorities to check travelers in a comfortable, fast and secure way. Here, a mobile scenario where border guards are using a device as a means of the travelers identity check is considered. In the European Union's Seventh Framework Program the project *MobilePass* focuses on research and development towards technologically advanced mobile equipment at land border crossing points. This will allow border control authorities to check European, visa-holding and frequent third country travelers in a comfortable, fast and secure way. Here, contactless palm and finger detection for biometric fingerprint verification/identification process with mobile devices is considered.

Several approaches for finger detection with different sensors are discussed in the literature. In [JHB13, PK14], finger detection is based on the top-hat transform, which uses morphological operators on a hand blob by subtracting the results from the original hand blob. However, a reliable segmentation of the hand palm/fingers in the RGB image is a prerequisite of many algorithms that analyze a binary hand blob image and the performance of these algorithms crucially depends on the segmentation result. In [RYMZ13], a shape representation based on time-series curve is used for fingertip detection. In [BLL07], a

---

\*This project has received funding from the European Union's Seventh Framework program for research, technological development and demonstration under grant agreement No. 608016.

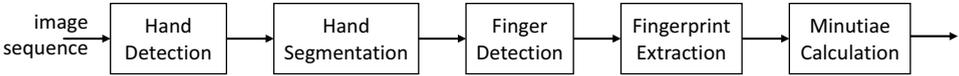


Figure 1: Processing chain.

skeleton-based approach is proposed for fingertip detection.

In order to speed up the border checking verification process, we focus on capturing the whole palm, which is presented by the person to be checked, in order to extract the fingertips, instead of successively capturing each fingertip. As a requirement for the capturing device, the resolution of the sensor should be such, that each fingerprint image can be cropped with a resolution of at least 500 dots per inch (DPI), which is the minimum resolution in a fingerprint verification/identification process. The workflow comprises palm detection in order to detect the skin region within the image prior to the detection of fingertips. A machine learning based algorithm with ACFs adopted for palm detection and finger detection, respectively, are considered. Furthermore, a fast geometric shape based approach for fingertip detection has been designed to reconstruct long lines along fingers.

The considered processing workflow is depicted in Fig. 1. In the first steps hand detection and hand segmentation are carried out, cf. Section 2. Based on the segmentation result, the fingers are detected in order to extract the fingerprint images. cf. Section 2.3.

## 2 Palm Detection, Segmentation, and Fingertip Extraction

### 2.1 Palm Detection using Aggregated Channel Feature

Object detection is one of the fundamental topics of research in the pattern recognition community. The Histograms of Oriented Gradients (HOG) introduced by Dalal and Triggs [DT05], although not state-of-the-art anymore, presented a classic approach for dealing with these tasks by combining rich feature descriptors and effective learning methods. In the past few years, a new family of features, namely the Integral Channel Features (ICF) [DTPB09] has attracted increasing interests, which integrate diversity of information into the computational efficiency using integral images. Recently, Dollár et al. proposed the Aggregated Channel Feature (ACF) in [DABP14]. The random rectangular blocks for calculating integral images in ICF [DTPB09, DBP10] are substituted by the sum of small squares. By exploiting the same learning framework, further performance and efficiency enhancements are reached.

Palm or hand detection in the context of *MobilePass* resembles that of pedestrian detection in the aforementioned work. With the constrained recording condition, the subjects are guided to show upright palm in front of the camera, analog to the upper body of pedestrians. However, the fingers can be spread to a variable extent, which is like the human legs. Considering the similarity between the two detection tasks, we explore the possibility of employing ACF for detection palms.



Figure 2: Illustration of the computed ACF from an example palm image.

**Features.** A bunch of feature combinations with regarding to different color space, intensity and gradient were experimented in [DTPB09], which yields a clear advantage of taking into account color information than the pure grayscale and gradient feature. LUV stands out among the color spaces. Moreover, the gradient magnitude and 6 gradient orientation bins are also included, resulting in totally 10 channels. Before and after dividing the channels into  $4 \times 4$  blocks for summing up the pixels, pre-smoothing and post-smoothing with a  $\frac{[1,2,1]}{4}$  filter are conducted. An example of the computed ACFs from a palm image is illustrated in Fig. 2. It is in particular obvious that the aggregated gradient magnitude and histograms can characterize the hand shape in a compact representation.

**Multiscale ACF Approximation.** The standard routine for object detection is based on sliding windows over the multiscale image pyramids. As a fine-scale pyramid is essential for successive detection, more than 50 scale steps are needed to process an image of  $640 \times 480$  pixels. However, computing ACFs dozens of times on the image pyramid turns out to be costly. To circumvent the drawback of the conventional pipeline, the rich image features are ideally computed as few times as possible. Regarding to ACF, since the color channels can be directly resized to match different scale spaces, Dollár et al. [DBP10, DABP14] proposed a fast resampling scheme for the HOG features in ACF. Based on the observation that the statistics of natural images conform to the power law w.r.t. the ratio of scales, they proved that for the shift-invariant ACFs, a similar approximation also holds  $\frac{\mathbf{f}_\Omega(\mathbf{I}_{s_1})}{\mathbf{f}_\Omega(\mathbf{I}_{s_0})} = \left(\frac{s_1}{s_0}\right)^{-\lambda_\Omega}$ , where  $\mathbf{f}_\Omega$ , given the input image  $\mathbf{I}$  of scale  $s_0$ , computes the channel feature  $\Omega$ , which has its own corresponding scale factor  $\lambda_\Omega$ . Accordingly, assuming  $\lambda_\Omega$  is obtained by least squares fitting of the training images, ACF channels  $\mathbf{f}_\Omega(\mathbf{I}_{s_1})$  of an arbitrary scale  $s_1$  are simply a recalculation of that in the original scale space  $s_0$ , yielding

$$\mathbf{f}_\Omega(\mathbf{I}_{s_1}) = \mathbf{f}_\Omega(\mathbf{I}_{s_0}) \cdot \left(\frac{s_1}{s_0}\right)^{-\lambda_\Omega}. \quad (1)$$

It is suggested that computing  $\mathbf{f}_\Omega$  only once per octave (resizing with doubled or halved size), and approximating the intermediate scales by Eq. (1) suffices to find an ideal trade-off between accuracy and speed. Moreover, by virtue of the aggregation of the features, evaluating ACF approximation  $\mathbf{f}_\Omega(\mathbf{I}_s)$  is even faster than resizing the image  $\mathbf{I}_s$  itself.

**Training.** Despite of the sparsity of hand and palm datasets, we were able to collect 345 images from 3 publicly available hand gesture datasets, i.e. the Sébastien Marcel Static Hand Posture Database [Mar99], the Database for Hand Gesture Recognition (HGR) [NGK14], and the hand gesture dataset acquired with Leap Motion and the Kinect devices [MDZ14]. The hand gesture “five” in these datasets resembles the upright open palm scenario in *MobilePass*. Selected samples of the combined dataset are shown in Fig. 3, demonstrating a large variation in gesture shape, image quality and resolution, illumina-

tion and background clutter, etc. All these nuisance factors are extremely challenging for training a powerful ACF detector.



Figure 3: Example images of our dataset for training the ACF palm detector.

We manually annotated the training images with square bounding boxes and resized the hand crops to  $50 \times 50$  pixels. Boosted tree with soft cascades [FHT00] is leveraged to train and combine 2048 depth-two trees over all candidate ACF channel lookups. Bootstrapped learning employed by the Viola-Jones detector [VJ01] is also exploited. For sampling negative samples, the INRIA pedestrian dataset [DT05] is used.

## 2.2 Palm Segmentation

Given the bounding box of the detected palm, provided by the ACF detector as described in Sec. 2.1, the next step is to determine the subset of pixels which belong to the hand. This segmentation step allows for sub-sequential shape analysis of the hand, and as a consequence to search for finger tips.

One straight forward approach for segmentation of body parts is using skin color. In literature, several approaches exist which mainly try to detect pixels with human skin tones, without any prior knowledge on images content [LP10, FAKK02]. However, due to the applied palm detection step as pre-processing, in our case skin color estimation and segmentation can be designed in a more robust way.

Our palm segmentation approach is sub-divided in the following 4 steps:

**Pixel-based Skin Tone Detection.** For pixel-based skin color detection, we use the algorithms proposed in [CCCM09]. While most alternative approaches basically try to remove illumination component from images to obtain an illumination invariant color classifier, the authors in [CCCM09] claims, that illumination is also an important feature for pixel-wise skin tone classification.

The algorithm first determines a grayscale map of the RGB color image given an usual transformation matrix  $\mathbf{a} = [0.298936, 0.587043, 0.140209]^T$  by standard product operation  $\mathbf{I}'(\mathbf{x}) = [r(\mathbf{x}), g(\mathbf{x}), b(\mathbf{x})] \otimes \mathbf{a}$ . The resulting 1D image  $\mathbf{I}'$  is considered as the grayscale map of the original image, taking into account all color channels. Additionally to  $\mathbf{I}'$  a second illumination map  $\hat{\mathbf{I}}'$  is determined, considering green and blue channel components only. Red channel is discarded from this grayscale map, since it is the most contributing one in skin pixels,  $\hat{\mathbf{I}}'(\mathbf{x}) = \max(g(\mathbf{x}), b(\mathbf{x}))$ . Now, a skin color probability map is generated by pixel wise signal error calculation as  $\mathbf{E}(\mathbf{x}) = \mathbf{I}'(\mathbf{x}) - \hat{\mathbf{I}}'(\mathbf{x})$ .

Finally, given a large training dataset containing regions with skin from persons of a range of races/cultures with extreme variation of lighting effect, the interval of  $\mathbf{E}(\mathbf{x})$  for the

majority of skin pixels has been estimated. Given a lower and upper bound, skin pixels are classified as follows:

$$M_{skin}(\mathbf{x}) = \begin{cases} 1 & \text{if } 0.02511 \leq \mathbf{E}(\mathbf{x}) \leq 0.1177 \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

One important reason for choosing this skin tone detection method was also its efficiency and real-time capability. No complex color space transform is needed and reduction of color space dimensionality from 3D (RGB) to 1D (normalized grayscale) allows for very low computational load and as a consequence efficient pixel-wise classification.



(a) Original image with ACF detector ROI. (b) Image after skin color segmentation. (c) Clutter removal, blob filling and morphology.

Figure 4: Hand segmentation by skin color.

**Connected Component Analysis for Clutter Removal / De-noising.** Given the skin pixel segmentation image as shown in Fig. 4b, in a second step a connected component analysis is performed for blob analysis and clutter removal. All small segments with an unreasonable too small area related to palm region-of-interest (ROI) are classified as clutter. In our examples, we choose 5% of ROI area as threshold.

Also segments outside palm detector ROI are discarded in further processing. Finally, the remaining blob with the largest size is selected as hand segment candidate for further processing. The result is a segmentation image, containing blobs of skin colored areas of significant size, only.

**Post-Processing / Morphology (hole filling).** Given the hand segment candidate as described above, in a final post-processing morphological filters are applied for hole filling. Hereby, a background flood fill approach, followed by a binary image inversion is applied. The resulting segmentation result is shown in Fig. 4c.

### 2.3 Fingertip Extraction

The main idea of the finger detection algorithm is to associate the “left” and “right” edges to fingers (edge-pairing) and to extract them by their respective tips and angles. The edge-pairing algorithm is subdivided into the following steps.

**Hand palm segmentation.** The input RGB image (Fig. 5a) is segmented in order to extract the finger edges. Hand segmentation is based on the algorithm in Section 2.2.

**Converting the Binary Blob Image into a Contour Line.** In order to process and examine the blobs contained in the binary input image (Fig. 5b) efficiently, the blobs are

converted into contour lines, cf. Fig. 5c. A contour is a sequence of pixels located along the boundaries of the blob.

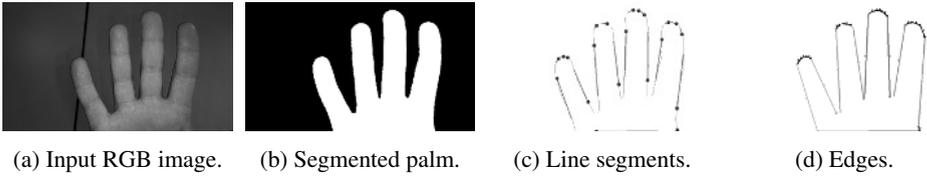


Figure 5: Finger detection by edge pairing.

**Line Segmentation of the Contour Line.** The most important recognition features of fingers are approximately long lines along them. In order to detect such lines and deduce the presence of fingers, it is necessary to isolate them from the contour line of a blob by isolating line segments within a contour. Line segments are detected by analyzing the variation of the tangent angle on the contour. Within a line segment the variation is low, i.e. the tangent angle remains almost constant.

**Forming fingers by matching edges.** The formation of fingers based on the previously computed line segments is provided by edge pairing. A pair of edges is consequently interpreted as the “left” and “right” edge of the fingers, respectively. The method is illustrated in Fig. 5d, where the line segments used for reconstruction of paired edges are plotted in the same color. An edge can only be combined with another edge. Unpaired edges are discarded and are not recognized as a finger. In order to find the best match for each edge, a match quality metric for a hypothetical pair of edges is calculated, which depends on the orientation of edges, the distance between center points, the maximum allowed pairing angle, and the length of edges.

### 3 Results

Results are provided for the ACF palm detector and the finger extraction algorithm based on assigning associated edges through pairing in the following.

Results of the ACF approach for palm detection are presented in Fig. 6. Obviously, the trained detector localize the palms independent of subjects and shape. Robustness against closed and spread fingers is demonstrated. By virtue of the relatively simple background, no false positives are seen in the test cases, although the confidence threshold is set to a very low level.

Results of the proposed finger extraction algorithm based on assigning associated edges through pairing are presented. In Fig. 7, the finger axes and the ROIs are depicted. The finger roots and tips are labeled as red and blue circles, respectively. The ROIs containing that part of the finger are used for a later biometric verification or identification process. After computing the finger axis and the fingertip, cf. Section 2.3, the size of the ROI is determined dynamically in order to take into account individual finger sizes. The width of the ROI corresponds to the width of the finger and is determined by examining the

binary blob-image on a virtual line perpendicular to the finger-axis. The height of the ROI depends on its width and is determined by multiplying the finger width with a constant factor (here: 1.7).

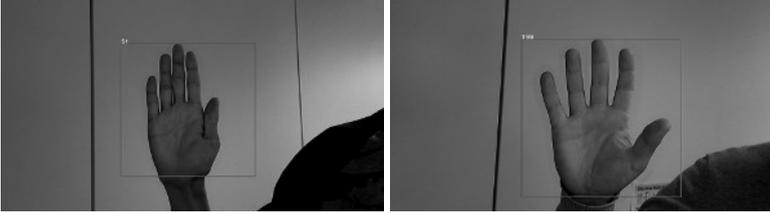


Figure 6: Example results of palm detection by the ACF detector.

Here, different results for palms and fingers are presented. The most important recognition feature of fingers are approximately long lines along them. Since this algorithm has been designed to reconstruct long lines along fingers, only well separated fingers can be reconstructed, i.e., results with spread fingers are shown here. Since extraction of edges is based on a reliably segmented palm in order to deduce the presence of fingers, the segmentation of the palm is crucial. If these two conditions are fulfilled, the dynamic ROI can be determined reliably, as depicted in Fig. 7.

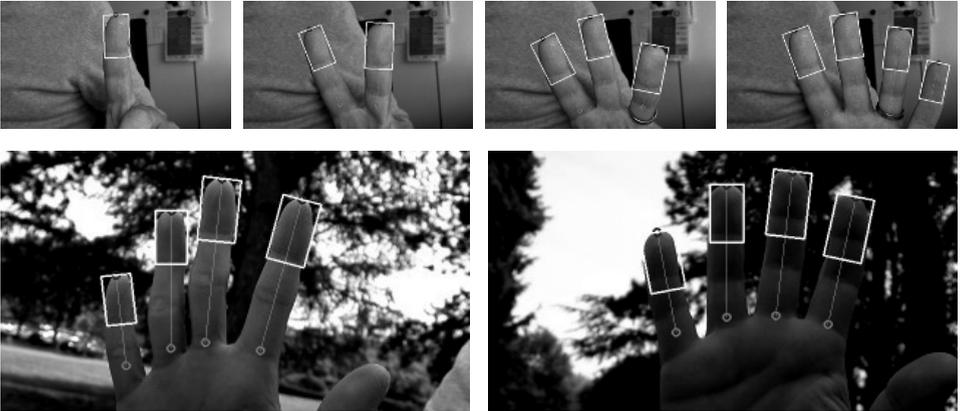


Figure 7: Results of the proposed fingertip detection algorithm in different scenarios.

## 4 Conclusions

In this paper, contactless palm and finger detection for the biometric recognition process for mobile devices has been considered. ACF hand detection and finger detection based on edge-pairing are proposed. Evaluation based on a measurement campaign in different indoor and outdoor scenarios demonstrate the suitability of the geometric approach by edge pairing in extracting fingertips for a biometric identification/verification process.

## References

- [BLL07] X. Bai, L. J. Latecki, and W.-Y. Liu. Skeleton Pruning by Contour Partitioning with Discrete Curve Evolution. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 29(3):449–462, March 2007.
- [CCCM09] A. Cheddad, J. Condell, K. Curran, and P. McKeivitt. A new colour space for skin tone detection. In *Image Processing (ICIP), 2009 16th IEEE International Conference on*, pages 497–500, Nov 2009.
- [DABP14] P. Dollár, R. Appel, S. Belongie, and P. Perona. Fast Feature Pyramids for Object Detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 36(8):1532–1545, Aug 2014.
- [DBP10] P. Dollár, S. Belongie, and P. Perona. The Fastest Pedestrian Detector in the West. In *Proceedings of the British Machine Vision Conference*, pages 68.1–68.11. BMVA Press, 2010.
- [DT05] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 886–893, June 2005.
- [DTPB09] P. Dollár, Z. Tu, P. Perona, and S. Belongie. Integral Channel Features. In *Proceedings of the British Machine Vision Conference*, pages 91.1–91.11. BMVA Press, 2009.
- [FAKK02] A. Farrukh, A. Ahmad, M. I. Khan, and N. Khan. Automated segmentation of skin-tone regions in video sequences. In *Students Conference, 2002. ISCON '02. Proceedings. IEEE*, volume 1, pages 122–128, Aug 2002.
- [FHT00] J. Friedman, T. Hastie, and R. Tibshirani. Additive logistic regression: a statistical view of boosting. *The Annals of Statistics*, 28(2):337–407, 04 2000.
- [JHB13] S. B. Jemaa, M. Hammami, and H. Ben-Abdallah. Data-mining process: application for hand detection in contact free settings. *Image Processing*, 7(8):742–750, November 2013.
- [LP10] H. C. V. Lakshmi and S. PatilKulakarni. Segmentation Algorithm for Multiple Face Detection for Color Images with Skin Tone Regions. In *Signal Acquisition and Processing, 2010. ICSAP '10. International Conference on*, pages 162–166, Feb 2010.
- [Mar99] S. Marcel. Hand Posture Recognition in a Body-face Centered Space. In *Proceedings of the Conference on Human Factors in Computer Systems (CHI)*, CHI EA '99, pages 302–303, New York, NY, USA, 1999. ACM.
- [MDZ14] G. Marin, F. Dominio, and P. Zanuttigh. Hand gesture recognition with leap motion and kinect devices. In *Image Processing (ICIP), 2014 IEEE International Conference on*, pages 1565–1569, Oct 2014.
- [NGK14] J. Nalepa, T. Grzejszczak, and M. Kawulok. Wrist Localization in Color Images for Hand Gesture Recognition. In Dr. A. Gruca, T. Czachórski, and S. Kozielski, editors, *Man-Machine Interactions 3*, volume 242 of *Advances in Intelligent Systems and Computing*, pages 79–86. Springer International Publishing, 2014.
- [PK14] P. Prasertsakul and T. Kondo. A fingertip detection method based on the top-hat transform. In *11th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*, pages 1–5, May 2014.
- [RYMZ13] Z. Ren, J. Yuan, J. Meng, and Z. Zhang. Robust Part-Based Hand Gesture Recognition Using Kinect Sensor. *Multimedia, IEEE Transactions on*, 15(5):1110–1120, Aug 2013.
- [VJ01] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, volume 1, pages 511–518, 2001.

# Weighted Integration of Neighbors Distance Ratio in Multi-biometric Fusion

Naser Damer, Alexander Nouak

Competence Center Identification and Biometrics  
Fraunhofer Institute for Computer Graphics Research (IGD)  
Darmstadt, Germany  
naser.damer@igd.fraunhofer.de  
alexander.nouak@igd.fraunhofer.de

**Abstract:** This work presents an approach to integrate biometric source weighting in the calculation of neighbors distance ratios to be used within a classification-based multi-biometric fusion process. The neighbors distance ratio represents the elevation of the top ranked identification match to the following ranks. Using biometric source weighing can help achieve more accurate initial identity ranking necessary for neighbors distance ratios. It also influences the effect of each biometric source on the ratios values. The proposed approach is developed and evaluated using the Biometric Scores Set BSSR1 database. The results are presented in the verification scenario as receiver operating curves (ROC). The achieved performance is compared to a number of baseline solutions and a satisfying and stable performance was achieved with a clear benefit of integrating the biometric source weights.

## 1 Introduction

Biometrics technology aims at identifying or verifying the identity of individuals based on their physical or behavior characteristics. Combining more than one biometric source is often performed to increase the accuracy, robustness and usability of biometrics [DOS13]. The different biometric sources can be based on different characteristics, captures, algorithms, sensors, or instances. Putting together the information provided by those sources and creating a unified biometric decision is referred to as multi-biometric fusion.

The fusion process can be applied on different levels such as the data, feature, score, or rank level. Higher levels such as score and rank provide a more flexible and integrable solution. Data and feature fusion levels provide more information but affect the integrability and may be hard to achieve in certain multi-biometric combinations. In this work, the score-level fusion will be considered as it provides a fair trade-off between performance and integrability.

Score-level biometric fusion techniques can be categorized into two main groups, combination-based and classification-based fusion. Combination-based fusion consists of simple operations performed on the normalized scores of different biometric sources. Those operations

produce a combined score that is used to build a biometric decision. One of the most used combination rules is the weighted-sum rule, where each biometric source is assigned a relative weight that optimizes the source effect on the final fused decision. The weights are related to the performance metrics of the biometric sources, a comparative study of biometric source weighting is presented by Chia et al. [CSN10] and extended later by Damer et al. [DON14a][DON14b].

Classification-based fusion views the biometric scores of a certain comparison as a feature vector. A classifier is trained to classify those vectors optimally into genuine or imposter comparisons. Different types of classifiers were used to perform multi-biometric fusion, some of those are support vector machines (SVM) [SVN07][GV00][DO14], neural networks [Als10], and the likelihood ratio methods [NCDJ08].

A biometric system usually operates under one of two scenarios, verification or identification. Verification is the authentication of a claimed identity based on the captured biometric characteristics. Identification is assigning an identity to an unknown individual based on their biometric characteristics. Identification can operate as a closed-set identification where the user is known to be included in the biometric references set, or as an open-set identification where the user is not definitely included in the references set. In open-set identification, a verification final step is required to verify that the top ranked identification match is certainly the same captured subject and not an unenrolled subject.

Keeping the open-set identification scenario in mind, previous work by Damer et al. [DO14] tried to use the information provided by the ranked set of comparisons to perform more accurate verification of the top rank. The assumption was that a genuine top rank comparison has a lower distance ratio to its rank neighbors than that of an imposter comparison, this distance ratio was referred to as the neighbor distance ratio (NDR). Those information were integrated into a classification-based fusion approach using SVM.

This work proposes introducing information about the performance of the different biometric sources into the NDR calculation process. This will help in producing more accurate initial ranking and more informative NDR values as discussed later in Section 2. The performance information were included as biometric source weights calculated based on the Overlap Deviation Weighting (OLDW) [DON14a].

The proposed fusion technique is evaluated over the Biometric Scores Set BSSR1 - multimodal database [BSS]. A number of previously proposed base-line fusion approaches were evaluated including state-of-the-art combination rules and the use of SVM with and without consideration of the neighbor distance ratio. The proposed technique proved to outperform the base-line solution and the results are presented as Receiver Operating Characteristics (ROC) curves.

In Section 2 the proposed solution is discussed along with the evaluated base-line solutions. The experiment setup and the achieved results are then presented in Section 3. Finally, in Section 4, a conclusion of the work is drawn.

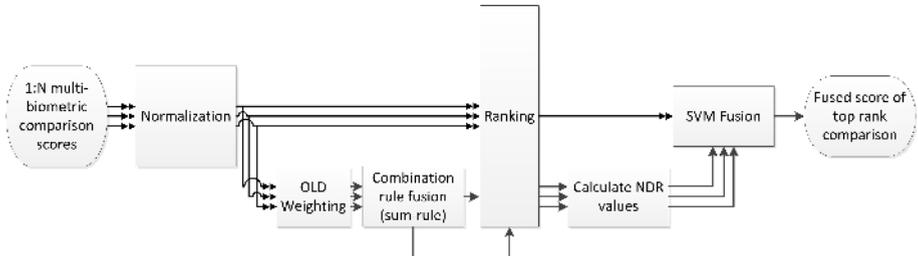


Figure 1: An overview of the proposed solution. The input scores of an 1:N comparison is weighted (OLDW) and fused by simple sum combination rule then ranked based on the resulted scores. The NDR values based on this ranking is concatenated with the original scores of the comparison to be verified. The concatenated vector is fed into the SVM to create a final fused score.

## 2 Methodology

### 2.1 Proposed solution

The assumption that builds the basis of the proposed solution in this work is anchored on the Neighbor Distance Ratio (NDR). Given a rank set of comparison scores that represents an 1:N comparison, NDR is defined as the ratio between one score in this set and a score of a higher rank (neighbor distance). NDR was previously used in the literature to match interest key point descriptors in images [MS05]. Looking into the NDR from the biometric prospective, the inverse ratio between a genuine similarity score and the next highest score (within a ranked 1:N comparison) is assumed to be lower than this ratio between an imposter score and the next highest score.

The contribution of this work is based on providing more accurate initial ranking to calculate NDR values. This is achieved by using OLD weighting [DON14a] approach to assign relative weights to different biometric sources to influence their effect on the overall initial ranking, and thus the accuracy of the NDR values. The weighted biometric scores also effects the values of the NDR as the initially fused scores are also fused by a weighted sum rule.

The proposed solution in this work aims at considering both, the scores absolute values and the relative distances to higher ranks in order to perform more accurate biometric verification. To achieve that, a classification-based fusion approach based on support vector machines (SVM) was used. In classification-based multi-biometric fusion, the fusion process is viewed as a binary classification problem that aims to separate between two classes, genuine and impostor.

Support vector machines [Vap95] is a statistical learning technique often used to learn binary classifiers, i.e. to learn how to separate two classes using information gained from known examples (training data). Classical learning techniques, such as Neural Networks

(NN), focused on minimizing the empirical error (error on the training set). This approach is commonly referred to as Empirical Risk Minimization (ERM). However, the SVM follows the Structural Risk Minimization (SRM) instead of the ERM approach. The SRM insures a high generalization performance as it tries to minimize the upper bound of the generalization error. In simple words, SVM tries to build a class-separation surface in the feature space that is optimized in a manner which considers generalized unknown data.

In order to map the input data space into a feature space where the data is linearly separable, SVM uses kernel functions. In general, those functions help in enhancing the discrimination power. In this work, the Radial Basis Function (RBF) is used as a kernel function as it proved to outperform linear kernels when dealing with low dimensional space [SZL<sup>+</sup>11], such as the problem dealt with in this work.

The feature vector considered for the SVM fusion process consisted of two concatenated Parts, the initial comparison scores of different sources  $N$  and the NDR values based on the initial weighted fusion. Here, three NDR values were considered, 2nd-rank-to-1st-rank, 3rd-rank-to-1st-rank, and the 3rd-rank-to-2nd-rank. This will result in a feature vector of size  $N + 3$ . The SVM classifies the input feature vector and the resulting decision function value (the signed distance to the margin) is considered as the final fused score. An overall look on the proposed method is presented in Figure 1

## 2.2 Baseline solution

A number of baseline solutions are presented here to build a reference for the performance evaluation presented in the next Section 3. The first baseline solution aims at direct comparison by using the SVM based solution that integrates NDR values without weighted ranking [DO14]. The second baseline solution will be SVM based approach without using NDR information. Two other solutions utilized the widely used weighted-sum approach are also discussed, one utilizes the EER as a source performance measure while the second uses the Non-Confidence Width (NCW).

The baseline SVM based approach that integrates NDR values is similar to the proposed solution here, however it does not use weighted scores for initial ranking and NDR calculation. Instead it uses simple equal weight sum rule fusion. This approach will be referred to as SVM-NDR.

The conventional SVM baseline approach takes the biometric comparison scores of different sources  $\{S_1, \dots, S_n\}$  as a feature vector. The SVM is trained to classify this feature vector into genuine or impostor classes and reports the resulted decision function value as the fused score. The SVM used here also uses similar configuration to the proposed approach with RBF as a kernel function.

The two other baseline approaches are based on the weighted-sum combination rule that assigns each score value  $S_k$  with the weight of its source  $w_k$  to produce the fused score. The weights  $w_k$  are calculated from the training data of each biometric source. The fused score  $F$  by the weighted sum rule for  $N$  score sources is given as:

$$F = \sum_{k=1}^N w_k S_k, k = \{1, \dots, N\} \quad (1)$$

The weights used here are based on either EER (equal error rate) or NCW (Non-Confidence Width Weight) values. The EER weighting (EERW) is based on the EER value which is the common value of the false acceptance rate ( $FAR$ ) and the false rejection rate ( $FRR$ ) at the operational point where both  $FAR$  and  $FRR$  are equal. EER weighting was used to linearly combine biometric scores in the work of Jain et al. [JNR05]. The EER is inversely proportional to the performance of the biometric source. Therefore, for a multi-biometric system that combines  $N$  biometric source, the EER weight for a biometric source  $k$  is given by:

$$w_k = \frac{\frac{1}{EER_k}}{\sum_{k=1}^N \frac{1}{EER_k}} \quad (2)$$

The Non-Confidence Width Weight (NCWW) was proposed by Chia et al. [CSN10] to weight biometric sources for score-level multi-biometric fusion. NCW corresponds to the width of the overlap area between the genuine and impostor scores distributions. Given that  $Max_k^I$  is the maximum impostor score and  $Min_k^G$  is the minimum genuine score, NCW is given by:

$$NCW_k = Max_k^I - Min_k^G \quad (3)$$

as the NCW is inversely proportional to the biometric source performance, the weights based on the NCW is given as:

$$w_k = \frac{\frac{1}{NCW_k}}{\sum_{k=1}^N \frac{1}{NCW_k}} \quad (4)$$

### 3 Experiments and Results

The database used to develop and evaluate the proposed solution is the Biometric Scores Set BSSR1 - multimodal database [BSS]. The database contains comparison scores for left and right fingerprints (Fli and Fri) and two face matchers (Fc and Fg). BSSR1 - multimodal database contains 517 genuine and 266, 772 impostor scores. The experiments here considered all possible pairs between finger and face matchers. To evaluate the statistical performance of the proposed solutions, the database was split into three equal-sized partitions. Experiments were performed on all possible fold combinations where one partition is used as an evaluation set and the other two are used as a development set. All the reported results are the averaged results of the three evaluation/development combinations.

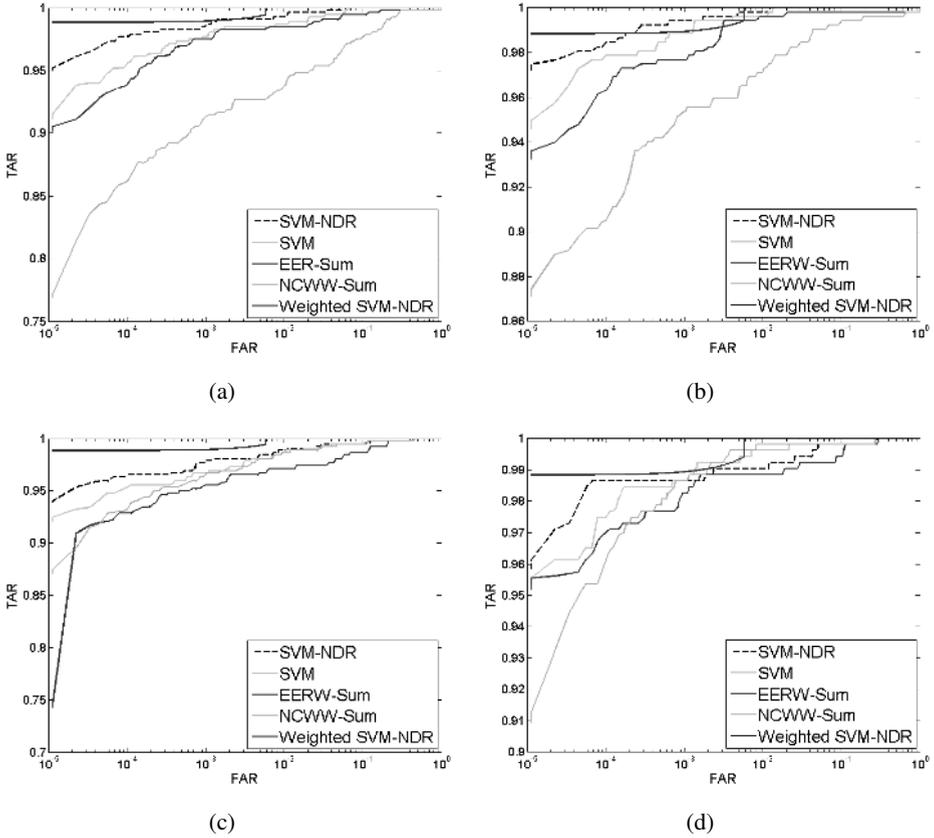


Figure 2: ROC curves (a) achieved on the BSSR1 database: The rates shown here are for bi-modal combinations of face matchers (Fc and Fg) and finger matchers (Fli and Flr) in the BSSR1 database. a) Fc and Fli, b) Fg and Fli, c) Fc and Flr, d) Fg and Flr.

Min-max normalization was used to bring comparison scores produced by different biometric sources to a comparable range. Min-max normalized score is given as:

$$S' = \frac{S - \min\{S_k\}}{\max\{S_k\} - \min\{S_k\}} \quad (5)$$

Where  $\min\{S_k\}$  and  $\max\{S_k\}$  are the minimum and maximum value of scores existing in the training data of the corresponding biometric source. And  $S'$  is the normalized score.

To train and test the proposed approach, every possible open-set identification scenario that can occur in the database was simulated. To do that, the comparisons in the database were split into separated 1:N comparison sets. Each comparison of those sets were fused using the OLD weighted sum-rule fusion, then ranked according to the resulting fused

scores. The three considered NDR values were calculated for each entry in the ranked comparison sets, except the last two ranks, as the second and third rank to those entries does not exist and thus the NDR values cannot be calculated. The resulted NDR values of each comparison are concatenated with the original scores of the comparison to create the final feature vector for that comparison. The resulted feature vectors are passed along with their genuine/imposter labels to train the SVM classifier in the training mode.

For evaluation, similar concatenated feature vectors are created from the testing data. Those features are evaluated by the trained SVM classifier to produce a final fused score from each comparison. The performance achieved by the proposed solution and the baseline approaches is presented as ROC curves in Figure 2. Performance was presented for all possible bi-modal (two face and two fingerprint matchers). ROC curves plots the false acceptance rate (FAR) and the true acceptance rate (TAR) at different operational points (thresholds) and presents the tradeoff performance between the two error rates. Generally, for high secure biometric systems, the area to the left of the curve (low values of FAR) is of main interest. The results shown in Figure 2 shows the high performance of the proposed approach compared to the baseline solutions. This is clearer at lower FAR values.

## 4 Conclusion

This work focused on the process of multi-biometric score-level fusion. The fusion approach is based on including the neighbor distance ratios in a classification-based fusion framework. The proposed solution aimed at including biometric source weighting information in the NDR calculation process, this helped creating a more accurate initial ranking and influence the biometric sources effect on the NDR value. The evaluation was performed on the BSSR1 database and proved the superiority of the proposed solution compared to a number of baseline methods. The results clearly show the benefit of the initial biometric source weighting within an NDR-based biometric verification.

## Acknowledgment

The work leading to these results has received funding from the European Community's Framework Programme (FP7/2007-2013) FIDELITY project under grant agreement n° 284862.

## References

- [Als10] Fawaz Alsaade. A Study of Neural Network and its Properties of Training and Adaptability in Enhancing Accuracy in a Multimodal Biometrics Scenario. *Information Technology Journal*, 2010.

- [BSS] *National Institute of Standards and Technology: NIST Biometric Scores Set.*
- [CSN10] Chaw Chia, N. Sherkat, and L. Nolle. Towards a Best Linear Combination for Multi-modal Biometric Fusion. In *Pattern Recognition (ICPR), 2010 20th International Conference on*, pages 1176–1179, 2010.
- [DO14] Naser Damer and Alexander Opel. Multi-biometric Score-Level Fusion and the Integration of the Neighbors Distance Ratio. In Aurélio J. C. Campilho and Mohamed S. Kamel, editors, *Image Analysis and Recognition - 11th International Conference, ICIAR 2014, Vilamoura, Portugal, October 22-24, 2014, Proceedings, Part II*, volume 8815 of *Lecture Notes in Computer Science*, pages 85–93. Springer, 2014.
- [DON14a] Naser Damer, Alexander Opel, and Alexander Nouak. Biometric source weighting in multi-biometric fusion: Towards a generalized and robust solution. In *22nd European Signal Processing Conference, EUSIPCO 2014, Lisbon, Portugal, September 1-5, 2014*, pages 1382–1386. IEEE, 2014.
- [DON14b] Naser Damer, Alexander Opel, and Alexander Nouak. CMC curve properties and biometric source weighting in multi-biometric score-level fusion. In *17th International Conference on Information Fusion, FUSION 2014, Salamanca, Spain, July 7-10, 2014*, pages 1–6. IEEE, 2014.
- [DOS13] Naser Damer, Alexander Opel, and Andreas Shahverdyan. An Overview on Multi-biometric Score-level Fusion - Verification and Identification. In Maria De Marsico and Ana L. N. Fred, editors, *ICPRAM*, pages 647–653. SciTePress, 2013.
- [GV00] B. Gutschoven and P. Verlinde. Multi-modal identity verification using support vector machines (SVM). In *Information Fusion, 2000. FUSION 2000. Proceedings of the Third International Conference on*, volume 2, pages THB3/3–THB3/8 vol.2, July 2000.
- [JNR05] Anil Jain, Karthik Nandakumar, and Arun Ross. Score normalization in multimodal biometric systems. *Pattern Recognition*, 38(12):2270 – 2285, 2005.
- [MS05] Krystian Mikolajczyk and Cordelia Schmid. A Performance Evaluation of Local Descriptors. *IEEE Trans. Pattern Anal. Mach. Intell.*, 27(10):1615–1630, October 2005.
- [NCDJ08] Karthik Nandakumar, Yi Chen, Sarat C. Dass, and Anil Jain. Likelihood Ratio-Based Biometric Score Fusion. *IEEE Trans. Pattern Anal. Mach. Intell.*, 30(2):342–347, February 2008.
- [SVN07] Richa Singh, Mayank Vatsa, and Afzel Noore. Intelligent Biometric Information Fusion using Support Vector Machine. In Mike Nachttegael, Dietrich Van der Weken, EtienneE. Kerre, and Wilfried Philips, editors, *Soft Computing in Image Processing*, volume 210 of *Studies in Fuzziness and Soft Computing*, pages 325–349. Springer Berlin Heidelberg, 2007.
- [SZL<sup>+</sup>11] Sutaog Song, Zhichao Zhan, Zhiying Long, Jiakai Zhang, and Li Yao. Comparative Study of SVM Methods Combined with Voxel Selection for Object Category Classification on fMRI Data. *PLoS ONE*, 6(2):e17191, 02 2011.
- [Vap95] Vladimir N. Vapnik. *The Nature of Statistical Learning Theory*. Springer-Verlag New York, Inc., New York, NY, USA, 1995.

# Does Context matter for the Performance of Continuous Authentication Biometric Systems? An Empirical Study on Mobile Devices

Soumik Mondal and Patrick Bours  
Norwegian Information Security Laboratory (NISLab)  
Gjvik University College  
firstname.lastname@hig.no

**Abstract:** In this paper we will show that context has an influence on the performance of a continuous authentication system. When context is considered we notice that the performance of the system improves by a factor of approximately 3. Even when testing and training are not based on exactly the same task, but on a similar task, we see an improvement of the performance over a system where the context is not included. In fact, we prove that the performance of the system depends on which particular kind of task is used for the training.

## 1 Introduction

Access control on a mobile device (*i.e.* smart phone or tablet) is generally implemented as a one-time proof of identity during the initial log on procedure [CSW<sup>+</sup>13]. The validity of the user is assumed to be the same during the full session. Unfortunately, when a device is left unlocked, any person can have access to the same sources as the genuine user. This type of access control is referred to as static authentication. On the other hand we have *Continuous Authentication* (sometimes also called *Active Authentication*), where the genuineness of a user is continuously verified based on the activity of the current user operating the device. When doubt arises about the genuineness of the user, the system can lock, and the user has to revert to the static authentication access control mechanism to continue working.

Due to its novelty, little research was done in this area [FBM<sup>+</sup>13, RHM14]. *Continuous Authentication (CA)* by analysing the user's behaviour profile on mobile input devices is challenging due to the limited amount of information and the large intra-class variations. Most of the previous research was actually done as periodic authentication, where the analysis was based on a fixed number of actions or fixed time period.

In our research we address a fundamental question, whether context really matters for the performance of a biometric CA system. We look at how much the performance changes if a user's behaviour profile is created by performing a particular task in a specific applications on the mobile device compared to test data coming from various application with different tasks. We use a CA biometric system proposed by [MB15], which checks the genuineness

of the user during the full session.

## 2 Background Knowledge

### Classifier(s)

We used two classifiers in a *Multi Classifier Fusion (MCF)* architecture to achieve an acceptable system performance [KHDM98]. Due to the nature of the data we found that one prediction model and one regression model gave us a better learning accuracy. We used *Support Vector Machine (SVM)* as a prediction model and *Counter-propagation Artificial Neural Network (CPANN)* as a regression model.

### Trust Model

In our analysis we look at every single action performed by the user and we have used the *Trust Model* [Bou12] in our analysis. The basic idea of the *Trust Model* is that the trust of the system in the genuineness of the current user depends on the deviations from the way this user performs various actions on the system. If a specific action is performed in accordance with how the genuine user would perform the task (*i.e.* as stored in the template), then the systems trust in the genuineness of this user will increase. We call this a *Reward*. If there is a large deviation between the behaviour of the genuine user and the current user, then the trust of the system in that user will decrease, which is called a *Penalty*. If the trust of the system in the genuineness of the user is too low, then the user will be locked out of the system. In particular if the trust drops below a pre-defined threshold  $T_{lockout}$  (global or user specific), then the system locks itself and will require static authentication of the user to continue working. In our research, we use the Dynamic Trust Model [MB15] where, the penalty/reward is calculated based on the resultant classifier score  $sc_i$  according to

$$\Delta_{Trust}(sc_i) = \min\left\{-D + D \times \left(\frac{1 + \frac{1}{C}}{\frac{1}{C} + \exp\left(-\frac{sc_i - A}{B}\right)}\right), C\right\}. \quad (1)$$

Let the trust value after  $i$  actions be denoted by  $Trust_i$ , and let the  $i^{th}$  action have a classification score  $sc_i$ . Then we have the following relation between  $Trust_{i-1}$  and  $Trust_i$ :

$$Trust_i = \min\{\max\{Trust_{i-1} + \Delta_{Trust}(sc_i), 0\}, 100\} \quad (2)$$

The trust level never exceeds 100 to assure that an impostor cannot profit from a longer period of time where the genuine user behaved according to his own profile.

### Performance Measure

In this research, we focus on actual CA that reacts on every single action from a user. Therefore, we use the *Average Number of Genuine Actions (ANGA)* and *Average Number of Impostor Actions (ANIA)* as a performance evaluation metric [MB15]. A detailed explanation of the performed actions is given in Section 3.

Our goal is obviously to have ANGA as high as possible, while at the same time the ANIA value must be as low as possible. The last is to assure that an impostor user can do as

little harm as possible, hence he/she must be detected as quick as possible. In our analysis, whenever a user is locked out, we reset the trust value to 100 to simulate a new session starting after the set of actions that lead to this lockout.

### 3 Data Description and Feature Extraction

In our research, we used a publicly available continuous mobile touch gesture dataset [FBM<sup>+</sup>13]. To the best of our knowledge is this the only dataset publicly available and the structure is suitable for our analysis. A brief description of the dataset is given below.

#### Data Description

During the data collection process a custom application was deployed on 5 different Android mobile devices and touch gestures from 41 volunteers with 5 to 7 session per participants was collected [FBM<sup>+</sup>13]. Data was collected in 7 different tasks, *i.e.* 4 different Wikipedia Reading articles and 3 different Image Comparison Games.

We noticed in the data that all 41 users completed tasks 1-4 (*i.e.* 3 different reading tasks and one image comparison task). Task 5 (image comparison) was completed by 40 users, while tasks 6 and 7 (reading article and image comparison) were completed by only 14 users. These last 14 users were a subset of the 40 users that completed task 5.

#### Feature Extraction and Selection

In our analysis, we divided the sequence of consecutive tiny movement data into actions (*i.e.* strokes). From the raw data 31 different features were calculated. [FBM<sup>+</sup>13] shows the details of these feature extraction process.

Before building the classifier models we applied the feature selection technique proposed by [VK09]. We also analysed another feature selection technique [LTM<sup>+</sup>12], but found that the learning accuracy of the SVM dropped from 97.7% to 52.4% and the CPANN learning accuracy dropped from 97.5% to 89.9%. We have also observed that in some cases the SVM models of some users were biased, meaning that the models always classified one particular class.

### 4 Methodology

#### Verification Process

In our research, we used three verification processes. We split the data of each of the users into a part for training and a part for testing. In all cases the classifiers are trained with genuine and impostor (training) data. The amount of training data of the genuine user is 50% of the total amount of data of that user. The training data of the impostor users is taken such that the total amount of impostor data equals the amount of training data of the genuine user. This is done to avoid bias towards either the genuine or the impostor class. The three verification processes described below might be seen to correspond with

an "Internal System" (*VP-1*) where all the impostor users are known to the system, an "External System" (*VP-3*) where impostor users during testing are not known to the system and a combination of these (*VP-2*) where 50% of the impostor users are known to the system.

### Classifier Fusion

Below we describe the two fusion techniques that we applied in this research. The score vector we use for further analysis is  $(f^1, f^2) = (Score_{svm}, Score_{cpann})$ . We have applied *Score Fusion (SF)* and *Penalty-Reward Fusion (PRF)*.

In SF a single score is calculated from the 2 classifier scores and this score is used in Equation 1 (see Section 2) to calculate the penalty-reward. The final score is the weighed sum of the two scores, where  $f^1$  gets weight  $w$  and  $f^2$  gets weight  $1 - w$ . The calculated score is used in Equation 2 to calculate the updated system trust level. The weights are optimized using linear search.

In case of PRF we individually calculate the penalty-reward from Equation 1 for the 2 classifier scores. Let  $\Delta_{Trust}(sc_i^1)$  represents the penalty-reward from Classifier-1 (*i.e.* SVM) and  $\Delta_{Trust}(sc_i^2)$  stands for the penalty-reward from Classifier-2 (*i.e.* CPANN). We combine these with weighted fusion to calculate the system trust from Equation 2 in the same way as above for SF. As before are the weights optimized using linear search.

### System Architecture

The system is divided into two basic phases, the training phase and the testing phase. In the training phase, the training data is used to build the classifier models and the models are stored in a database for use during the testing phase. Each genuine user has his/her own classifier models and training features.

In the testing phase, we use the test data which was separated from the training data for comparison. In the comparison, we use the models and training features stored in the database and obtain the classifier score (probability) of each sample of the test data according to the performed action. This score will then be used to update the trust value  $Trust$  in the trust model (see Section 2). Finally, the trust value  $Trust$  is used in the decision module, to determine if the user will be locked out or can continue to use the device. This decision is made based on the current trust value and the lockout threshold ( $T_{lockout}$ ).

## 5 Result Analysis

In this section, we analyse the results that we obtained from our performance analysis. We divide our analysis into three major parts based on the verification process (see Section 4). We use one-hold-out cross validation testing. The total number of data sets of genuine users is 41 for tasks 1 to 4 and hence, the total number of data sets of impostor users for these tasks is 1640 ( $41 \times 40$ ). For task 5 these numbers are 40 resp. 1560 ( $40 \times 39$ ), while for tasks 6 and 7 these numbers are 14 and 182 ( $14 \times 13$ ). We report the results for user specific lockout thresholds ( $T_{us}$ ), where each threshold satisfies  $50 \leq T_{us} < \min(Trust_{genuine})$  and is optimized using linear search.

**Interpretation of the tables:** The results from our analysis are divided into 4 possible categories. The categories are divided based on genuine user lockout (+ if the genuine user is not locked out and – in case of lock out) and non-detection of impostor users (+ if an impostor user is locked out and – otherwise). Each of the genuine users can thus be classified of 4 categories: (+/+), (+/-), (-/+), and (-/-), where the first sign is related to the genuine user and the second to the impostor users. In our analysis are genuine users never locked out, so in Table 1 only the categories +/+ and +/- are shown.

The column *# Users* shows how many users fall within each of the 4 categories (*i.e.* the values sum up to 41). In the column ANGA a value will indicate the Average Number of Genuine Actions in case indeed genuine users are locked out by the system. If the genuine users are not locked out, then we actually cannot calculate ANGA, which is indicated by  $\infty$ . The column ANIA will display the Average Number of Impostor Actions, and is based on all impostors that are detected. The actions of the impostors that are not detected are not used in this calculation, but the number of impostors that is not detected is given in the column *# Imp. ND*. This number should be seen in relation to the number of users in that particular category. For example, in the VP-3, PRF '+/-' category described in Table 1, we see that *# Users* equals 3, *i.e.* there are  $3 \times 40 = 120$  impostor test sets, and 4 impostors within these 120 impostors are not detected by the system as being an impostor. In this particular case for the full system, 4 out of 1640 of the impostors are not detected, hence we have a 0.24% Impostor Non-Detected Rate (INDR).

## 5.1 Result Analysis Without Context

We first considered the CA system performance irrespective of the context. Table 1, shows the result we have obtained for this analysis. We clearly observe from the table that *Score Fusion (SF)* performs better than *Penalty-Reward Fusion (PRF)* for each of the verification processes. For this reason we only used *Score Fusion (SF)* for our further analysis. Note that the summary for each verification process shows the system ANGA and ANIA values, as well as the INPR value.

Table 1: Results obtained from our analysis for the context independent evaluation.

Categories	SF				PRF				
	# User	ANGA	ANIA	# Imp. ND	# User	ANGA	ANIA	# Imp. ND	
VP-1	+/+	41	$\infty$	11	40	$\infty$	13		
	+/-				1	$\infty$	435	14	
	Summary		$\infty$	11	0%		24	0.85%	
VP-2	+/+	40	$\infty$	17	38	$\infty$	23		
	+/-	1	$\infty$	88	1	3	$\infty$	275	39
	Summary		$\infty$	19	0.06%		47	2.38%	
VP-3	+/+	40	$\infty$	22	38	$\infty$	19		
	+/-	1	$\infty$	286	1	3	$\infty$	233	4
	Summary		$\infty$	29	0.06%		35	0.24%	

## Comparison with Previous Research

We compared our results with previous results based on the same dataset in [FBM<sup>+</sup>13]. Table 2, shows the previous research results in terms of ANIA/ANGA by using the conversion technique described in [MB15]. We see that our methods outperform the previous research for VP-1, while for VP-2 and VP-3 our ANIA values are higher. Note that the other researches that are based on the same dataset have done the analysis in the same manner as we have done for VP-1.

Table 2: Comparison with Previous Research.

Reference	# Users	FNMR	FMR/INDR	Blocksize	ANGA	ANIA	<i>P-value</i>
[FBM <sup>+</sup> 13]	41	3%	3%	12	400	12	
[RHM14] - Horizontal	41	1.75%	1.75%	11	629	11	
[RHM14] - Vertical	41	2.8%	2.8%	11	393	11	
Our (VP-1 with SF)	41	0%	0%	NA	$\infty$	11( $\pm$ 9)	0.79
Our (VP-2 with SF)	41	0%	0.06%	NA	$\infty$	19( $\pm$ 13)	0.98
Our (VP-3 with SF)	41	0%	0.06%	NA	$\infty$	27( $\pm$ 15)	0.93

## 5.2 Result Analysis With Context

The main objective of this paper is not to find a better CA method, but to determine if including the context in the analysis has an impact on the performance. In this section we present the results we obtained from our analysis by considering this context. We measure the CA system performance by training the system using the data obtained from a particular task and then testing the system with the data from the various tasks performed by the users. Tables 3, 4, and 5 show the results we obtained from the VP-1, VP-2 and VP-3 verification processes respectively, using the *Score Fusion (SF)* technique. In all cases we noted that genuine users were never locked out, hence we found that ANGA= $\infty$ , so these values are not reported. The tables only contain the ANIA values, as well as the INPR values between brackets. The values on the diagonal are displayed in bold to signify that training and testing is done using the same task, while for all off-diagonal values the training and testing is done with 2 different tasks.

Table 3: Results obtained from our analysis for context dependent evaluation for VP-1.

Train \ Test	Task-1	Task-2	Task-3	Task-4	Task-5	Task-6	Task-7
Task-1	<b>3 (0.1%)</b>	3 (0.1%)	16 (4.1%)	30 (3.1%)	30 (2.8%)	11 (4.9%)	42 (1.1%)
Task-2	7 (0.6%)	<b>3 (0.1%)</b>	15 (3.6%)	28 (2.7%)	29 (2.4%)	12 (4.9%)	33 (0.5%)
Task-3	14 (2.8%)	19 (4.4%)	<b>4 (0.2%)</b>	30 (2.6%)	32 (2.8%)	13 (7.7%)	53 (3.8%)
Task-4	22 (4.5%)	20 (3.6%)	20 (4.4%)	<b>3 (0.1%)</b>	6 (0.1%)	15 (4.4%)	15 (0.5%)
Task-5	16 (2.3%)	27 (4.9%)	19 (4.4%)	7 (0.6%)	<b>4 (0.0%)</b>	17 (9.3%)	11 (0.5%)
Task-6	33 (19.2%)	30 (12.6%)	22 (9.3%)	54 (17.0%)	43 (13.7%)	<b>4 (0.0%)</b>	42 (1.6%)
Task-7	32 (19.2%)	47 (25.8%)	48 (33.0%)	13 (3.8%)	12 (2.7%)	15 (8.8%)	<b>4 (0.0%)</b>

Recall that tasks 1 to 4 had 41 participants, while task 5 had only 40 participants and the last 2 tasks had only 14 participants. When training with task  $i$  and testing with task  $j$  we only considered those participants that participated in both tasks. Also recall that tasks 1,

2, 3, and 6 are article reading tasks and tasks 4, 5, and 7 are image comparison games, which will help to understand the results and its impact based on the type of the tasks.

Table 4: Results obtained from our analysis for context dependent evaluation for VP-2.

Train \ Test	Task-1	Task-2	Task-3	Task-4	Task-5	Task-6	Task-7
Task-1	<b>6 (0.2%)</b>	21 (4.0%)	17 (3.8%)	31 (3.3%)	29 (1.9%)	20 (12.1%)	51 (2.7%)
Task-2	11 (1.7%)	<b>5 (0.2%)</b>	18 (3.2%)	31 (3.5%)	28 (2.5%)	18 (11.5%)	46 (2.7%)
Task-3	26 (7.7%)	24 (6.7%)	<b>10 (1.6%)</b>	35 (4.7%)	35 (3.7%)	20 (13.2%)	48 (6.0%)
Task-4	27 (6.8%)	30 (4.4%)	33 (8.3%)	<b>10 (1.1%)</b>	15 (1.5%)	19 (10.4%)	14 (1.6%)
Task-5	22 (4.2%)	27 (4.6%)	24 (4.6%)	8 (0.7%)	<b>6 (0.3%)</b>	18 (6.6%)	18 (2.2%)
Task-6	16 (4.4%)	16 (1.6%)	21 (2.7%)	34 (0.0%)	38 (0.5%)	<b>7 (2.2%)</b>	39 (2.2%)
Task-7	22 (0.4%)	26 (1.2%)	30 (2.7%)	18 (0.0%)	17 (1.1%)	14 (4.4%)	<b>7 (0.5%)</b>

Table 5: Results obtained from our analysis for context dependent evaluation for VP-3.

Train \ Test	Task-1	Task-2	Task-3	Task-4	Task-5	Task-6	Task-7
Task-1	<b>9 (1.3%)</b>	18 (3.9%)	20 (5.2%)	28 (2.3%)	23 (2.1%)	19 (10.4%)	36 (4.4%)
Task-2	11 (1.6%)	<b>7 (0.9%)</b>	19 (5.0%)	30 (3.5%)	27 (1.9%)	15 (10.4%)	40 (2.2%)
Task-3	17 (4.5%)	20 (0.9%)	<b>7 (0.9%)</b>	24 (2.7%)	22 (1.3%)	14 (5.5%)	33 (2.2%)
Task-4	22 (4.2%)	21 (2.1%)	20 (5.1%)	<b>6 (0.5%)</b>	8 (0.7%)	15 (7.1%)	21 (1.1%)
Task-5	17 (2.4%)	24 (4.5%)	19 (4.1%)	9 (0.9%)	<b>8 (0.6%)</b>	12 (1.6%)	7 (0.0%)
Task-6	14 (4.9%)	13 (1.1%)	32 (8.8%)	39 (1.6%)	38 (0.5%)	<b>11 (4.4%)</b>	30 (1.1%)
Task-7	23 (8.2%)	33 (12.6%)	30 (7.1%)	16 (0.5%)	19 (1.6%)	15 (7.6%)	<b>6 (0.0%)</b>

On average we find that ANIA equals 3.6/7.3/7.7 for VP-1/VP-2/VP-3 when training and testing with the same task, while it equals on average 23.9/25.1/21.8 when using different tasks for training and testing. The next values are split according to the type of task (reading or image comparison) and we found that if the task is different, but the type of task is the same, then the ANIA equals 14.4/17.7/16.3, while if also the type of tasks differs, then the ANIA values are much higher: 31.0/30.7/25.9. Finally, if we consider a reading task for training then the system ANIA for any of the other tasks is 26.7/28.0/24.3 and for an image comparison tasks these values are 20.1/21.2/18.4.

### 5.3 Discussion

We have to stress that the amount of samples in the dataset allowed for Linear Search optimization of the parameters in the algorithm. The amount of data was not sufficient to be used for more advanced optimization methods like Genetic Algorithm.

The context dependent analysis performs better than context independent analysis for all the verification processes (3.6/7.3/7.7 vs. 11/19/29 for VP-1/VP-2/VP-3). This finding is in line with the findings from [KH14]. We do note that the INDR goes up slightly when including context in the analysis, but the average values stay low when the same task is considered. Training with an image comparison task gave slightly better results when testing with an arbitrary other task.

## 6 Conclusion

In this research we looked at the performance of a system when including context in the analysis. We saw that the performance improved (i.e. a lower ANIA value) when context was considered. The optimal performance was obtained by training and testing with the data of the exact same task. Even when testing is done with a similar kind of task is the performance still better than when no context is used, but the differences are not as significant.

Our results show that we cannot train and test a system with a specific task and then assume that the full system will perform according to the results that are found. Our future work will include training with multiple sessions of similar and different kind of tasks and see how this affects the performance of the system.

## References

- [Bou12] P. Bours. Continuous keystroke dynamics: A different perspective towards biometric evaluation. *Information Security Technical Report*, 17:36–43, 2012.
- [CSW<sup>+</sup>13] Zhongmin Cai, Chao Shen, Miao Wang, Yunpeng Song, and Jialin Wang. Mobile Authentication through Touch-Behavior Features. In *Biometric Recognition*, volume 8232 of *Lecture Notes in Computer Science*, pages 386–393. Springer, 2013.
- [FBM<sup>+</sup>13] M. Frank, R. Biedert, E. Ma, I. Martinovic, and D. Song. Touchalytics: On the Applicability of Touchscreen Input as a Behavioral Biometric for Continuous Authentication. *IEEE Trans. on Information Forensics and Security*, 8(1):136–148, 2013.
- [KH14] Hassan Khan and Urs Hengartner. Towards Application-centric Implicit Authentication on Smartphones. In *15th Workshop on Mobile Computing Systems and Applications*, pages 10:1–10:6. ACM, 2014.
- [KHDM98] J. Kittler, M. Hatef, R. P W Duin, and J. Matas. On combining classifiers. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 20(3):226–239, 1998.
- [LTM<sup>+</sup>12] Cosmin Lazar, Jonatan Taminau, Stijn Meganck, David Steenhoff, Alain Coletta, Colin Molter, Virginie de Schaetzen, Robin Duque, Hugues Bersini, and Ann Nowe. A Survey on Filter Techniques for Feature Selection in Gene Expression Microarray Analysis. *IEEE/ACM Trans. on Computational Biology and Bioinformatics*, 9(4):1106–1119, 2012.
- [MB15] Soumik Mondal and Patrick Bours. A computational approach to the continuous authentication biometric system. *Information Sciences*, 304:28 – 53, 2015.
- [RHM14] A. Roy, T. Halevi, and N. Memon. An HMM-based behavior modeling approach for continuous mobile authentication. In *2014 IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, pages 3789–3793, 2014.
- [VK09] D. Ververidis and C. Kotropoulos. Information Loss of the Mahalanobis Distance in High Dimensions: Application to Feature Selection. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 31(12):2275–2281, 2009.

# Fingerprint Image Enhancement with easy to use algorithms

Thomas Klir  
Technische Universität Darmstadt  
thomas.klir@gmx.de

**Abstract:** This paper looks at measures to enhance image quality with the intention of improving recognition rates of finger recognition systems. There are thus different algorithms when compared to each other. The methods cover sharpness enhancement in general; wavelet sharpness, photocopy filters, EAW filters, DoG filters and Cartoon filters. The FVC2000 Db1a is used as reference dataset. The scores for the comparison are based on the NFIQ quality scores. The approach of the paper is to use a few single algorithms of the GIMP program and evaluate which one improves the reference database the most. GIMP is open-source software so everybody can use it to improve fingerprint images without having advanced programming skills.

## 1 Introduction

The most frequently used biometric systems are systems with fingerprint recognition [GAKD00]. For example police agencies use them to identify people on a crime scene. The ridge characteristics are one of the last recognisable features after person's death [TB06]. Another example is smartphones, some of which have fingerprint sensors in their default configurations. With those sensors, the user is able to unlock their phone without entering the passcode [Inc15]. Minutiae extraction is an important process for recognising fingerprints. The most frequent minutiae used are ridge ending and ridge bifurcation [HWJ98]. The ridge ending is the point where the ridge terminates, whilst the bifurcation is the point where the ridge splits (forks) into two ridges. A good-quality fingerprint has between 40 and 100 minutiae [HWJ98]. Poor image quality might result if the finger is too dry, wet, worn-out or dirty at the scanning process [Kri06]. Another influence can be to high or low pressure or scratches. The problem is that many incorrect minutiae are extracted and a large percentage of correct minutiae may be ignored [HWJ98]. The goal of enhancement algorithm is to improve the ridge structure of the input fingerprint images and remove the unrecoverable regions [HWJ98].

The paper is organised as follows. The next section introduces the related work to give an overview of the already published papers. Section 3 introduces the reference dataset. In Section 4 the method sharpness (general), wavelet sharpness, photocopy filter, EAW filter, DoG filter and Cartoon filter will be described. In Section 5 the results of the methods will be shown. Finally, in the last section some conclusions are drawn.

## 2 Related work

In the article [DPP13] GIMP is used as one part of the pre-processing chain to alter the colours of fingerprint images. In [FAF07] they use various quality measures to compare fingerprint images. The article [KAY14] is about the comparison of fingerprint enhancement algorithms for poor-quality fingerprint images. They also use NFIQ as a comparison score. The article [KN14] also discusses the enhancement used for low-quality images, using the FVC2000 database as a reference. This paper gives a closer look at the algorithms which are available in the image manipulation program GIMP and where these can improve fingerprint images with FVC2000 as the reference database.

## 3 Dataset

The experimental results have been evaluated on the public fingerprint database FVC2000 Db1a<sup>1</sup>. The database is 100 fingers wide and 8 impressions per finger deep. The resolution of each fingerprint image is 300 x 300 pixels. In the modified framework (see section 4) there is a resolution of 500 x 500 pixels needed as a minimum. For this requirement the terminal program *composite*<sup>2</sup> with an empty image (blanko.tif) of resolution 500 x 500 pixels is used. Subsequently, you can find an example command:

```
Example command: composite -gravity center 1_1.tif blanko.tif 1_1.tif
```

For scoring the fingerprint image quality of the reference dataset, the NFIQ scores are used. The score is standardised in NIST IR 7151. In the standard are five quality measures: *EXCELLENT*, *VERY\_GOOD*, *GOOD*, *FAIR*, *POOR*. In total the database Db1a has 800 images of a quality 94.375 % good (*EXCELLENT*, *VERY\_GOOD*, *GOOD*), 0.875 % fair (*FAIR*) and 4.750 % poor (*POOR*). The quality is measured by the fingerprint feature extractor and comparator Verifinger SDK<sup>3</sup> and used as baseline.

## 4 Fingerprint enhancement (experiment)

### 4.1 Minutiae extraction

The process of fingerprint image enhancement involves getting a fingerprint image as an input, applying a set of intermediate steps to the image, and outputting the enhanced image for further feature extraction [HWJ98].

For the first analysis the VeriFinger 7.1/MegaMatcher 5.1 Identification Technology Algorithm Demo from Neurotechnology were used. With this software it is possible to *enrol* a set of fingerprints and *identify* whether a given fingerprint image is in the enrolled set. For the experiment we enrol all images from database *Db1a*. With an improvement of the input images there will be more and better minutiae extracted. Thus with an improvement of the NFIQ score there will be better input images.

<sup>1</sup><http://bias.csr.unibo.it/fvc2000/default.asp>

<sup>2</sup><http://www.imagemagick.org/script/composite.php>

<sup>3</sup>[http://www.neurotechnology.com/download.html#verifinger\\_sdk\\_trial](http://www.neurotechnology.com/download.html#verifinger_sdk_trial)

## 4.2 Image improvements with PILLOW

As we have heard in the sections above, for a good comparison score the quality of the minutiae extraction is important. The main point of this extraction is that the image has a high quality. For this purpose, we use an easy sharpening algorithm which was implemented in the Python Imaging Library *pillow*<sup>4</sup>. The algorithm is based on an article of P. Haeberli and D. Voorhies [HV94]. In the algorithm the image is manipulated with linear interpolation and extrapolation and a factor  $\alpha$ , which acts as a kernel scale factor. For degenerate images a blurred image may be used; interpolation reduces the high frequencies and extrapolation increases them. The image is sharpened by the unsharpening mask. Independent of the size of the kernel, a series of convolutions can be easily done [HV94].

## 4.3 Image improvements with GIMP

GIMP<sup>5</sup>, the GNU Image Manipulation Program is a widespread image manipulation program. In our experiment we have used the Mac OS X version GIMP 2.8. Some of the image manipulation algorithms looked very useful for improving the quality of the fingerprint images. A selection of some of the algorithms is shown in the following subsections. These algorithms looked very promising in the algorithm preview (GIMP).

### 4.3.1 Wavelet sharpness

The wavelet-sharpening filter<sup>6</sup> enhances the original image by increasing the contrast in high frequency space. By adjusting the sharpening radius, the amount of unsharpness can be taken into account. You can also adjust the amount of sharpening separately. The algorithm was designed by Marco Rossini in the year 2008.

### 4.3.2 Photocopy filter

The idea of the filter is to give the image the look of a photocopy, such as that made by a copier, with the relative darkness of the values of the neighbouring particular regions. This is done by darkening areas of the image which are darker than the neighbourhood average and setting other pixels to white. In this manner, enough large shifts are darkened to black. The degree of darkening is set by the *pct\_black* value. The *mask\_radius* parameter manages the size of the pixel neighbourhood whose average is computed. Large values for *mask\_radius* result in very thick black areas and less detail. Smaller values result in less toner overall and more details. Small *pct\_black* values make the mixture of white regions to black border lines smoother and the toner regions themselves thinner and less perceptible. The algorithm was designed by Spencer Kimball and Peter Mattis in the year 1995. In Figure 1 you can see the original Image 27\_3 and the image with the photocopy manipulation.

---

<sup>4</sup><http://python-pillow.github.io>

<sup>5</sup><http://www.gimp.org>

<sup>6</sup><https://github.com/gimp-plugins-justice/wavelet-sharpen>



Figure 1: Original image 27\_3 (left) and photocopy filtered image (right)

### 4.3.3 EAW filter

The filter *edge avoiding wavelets* (EAW) is based on an algorithm from Rannan Fattal from the year 2009 [Fat09]. The algorithm sharpens an image without halos and can be used for local contrast boosting. The algorithm uses a robust data-prediction lifting process. The scheme works on the edge content and avoids taking pixels from both sides of an edge. The multi-resolution analysis shows a better decorrelation compared to common linear translations-invariants. The wavelets encode - in shape and smoothness - information at every scale. That can be used to derive a new edge-aware interpolation scheme. The effectiveness of the wavelet can be used for various computational photography applications such as multi-scale dynamic range compression, edge-preserving smoothing and detail enhancement and image colorisation.

### 4.3.4 DoG filter

The DoG filter is used as edge detection the *Difference of Gaussians* (DoG) method. This method is based on a 2004 algorithm from William Skaggs. The edges are found by doing two Gaussian blurs with different radii and subtracting the results. There are efficient methods for Gaussian blurs so the algorithm is very fast. The important parameter is the blurring radius for the Gaussian blurs. Decreasing the radius leads to an increase in the *threshold* for recognising edges.

### 4.3.5 Cartoon

The Cartoon filter is based on a 1995 algorithm from Spencer Kimball and Peter Mattis. The filter spreads dark values in an image based on each pixel's darkness relative to a neighbouring average. The filter also modifies the active layer so that it looks like a cartoon draft. The draft is comparable with a black felt pen drawing which is then shaded with colour. This effect is created by darkening areas that are already noticeably darker than those of their neighbourhood. The parameters in the implementation are *mask\_radius* and *percent\_black*. The first parameter controls the size of the working area, with large values yielding very thick black areas and less detail. Small values yield subtle pen strokes. The parameter *percent\_black* determines the amount of black colour added to the image. Larger values create thicker, sharper and darker lines.

## 5 Results

When the quality of the input fingerprint image is poor, the performance of the extraction methods degrades rapidly. Therefore, fingerprint image enhancement is one of the key steps in an automated fingerprint identification system (AFIS). In the following subsections the results of the different methods will be described in detail.

### 5.1 Sharpness

We can see in Figure 2 that the total quality increases from sharpness 10 to sharpness 25. When the sharpness is over 25 the quality falls again. The best quality value is at sharpness 25, with 85.25 % of images *excellent*.

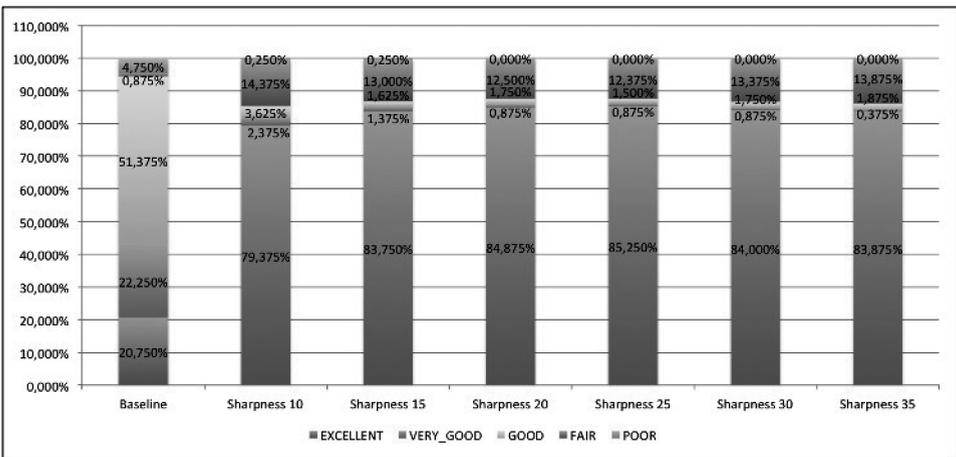


Figure 2: Different sharpness configurations

### 5.2 Wavelet sharpening

In Figure 3 the results of the wavelet sharpening are given in comparison to the baseline (data without enhancement). In Figure 3 you can see that

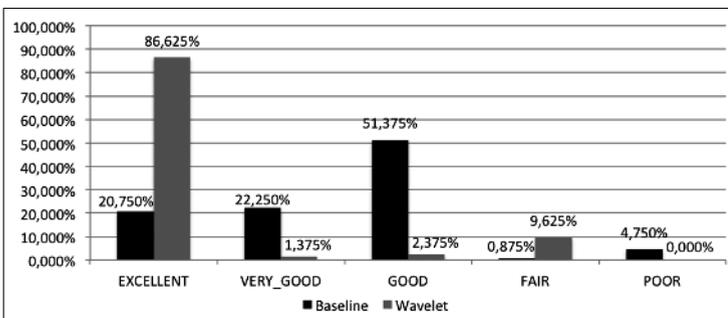


Figure 3: Wavelet sharpening

the score of the *excellent* value has increased to 86.625 %. The parameters for this experiment are radius=10.0, amount=0.5 and luminance=0, these being the standard parameters.

### 5.3 Photocopy filter

The parameters for the experiment are *mask-radius*=8.0, *sharpness*=0.8, *pct-black*=0.2 and *pct-white*=0.2 (standard values). The *excellent* value has increased to 73.875 %, but the *poor* value has also increased to 10.750 %. This suggests that some images overfitted with these values. It is obvious that nearly 3 % of the *excellent* values have decreased.

### 5.4 EAW filter

The *poor* value has also increased to 10.750 %. This is a clue that some images overfitted with these values. In this experiment the EAW filter is used with the standard parameter *alpha*=1.5, *maxband*=10, *inband*=5 and *mode*=0. The result is a minimal total increase (see Figure 4). It is clear that this filter with these parameters is not well suited to the application of fingerprint image enhancement. The highest loss is in *VERY\_GOOD* and *EXCELLENT*. This can also be a sign of overfitting.

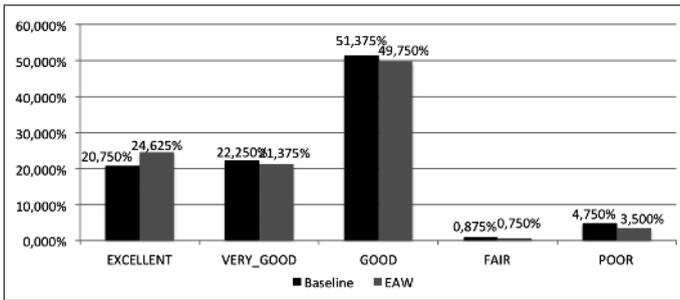


Figure 4: EAW filter

### 5.5 Difference to the normal distribution (DoG)

In this experiment the standard parameters *inner*=3.0, *outer*=1.0, *normalize*=True and *invert*=True are used. Figure 5 shows that the *EXCELLENT* value has a very high increase (90.625 %). The values in general are very good. None of the *EXCELLENT* values are decreasing. In this case the overfitting is very much limited.

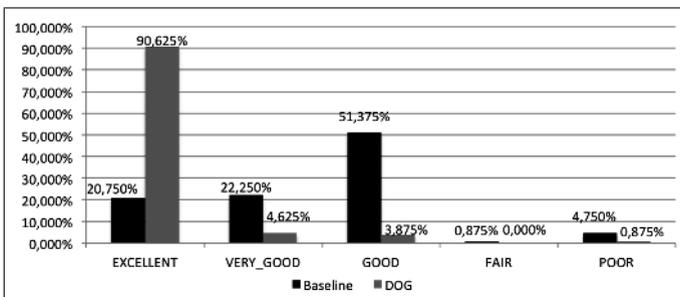


Figure 5: Difference to the normal distribution (find edges)

## 5.6 Cartoon filter

The last method has the standard parameters set at *mask-radius*=7.0 and *pct-black*=0.2. This method also shows good improvement when compared to the baseline (*EXCELLENT* 80.5 %). The *FAIR* value sees a drastic increase from 0.875 % to 12.875 %. This is a sign of overfitting. The *poor* values of *FAIR* come mainly from the decreasing of the *GOOD* values.

## 6 Conclusion

Finally, Figure 6 shows the comparison of all methods. In this figure you can see that the highest impact on the *EXCELLENT* level is with the DoG filter (90.625 %).

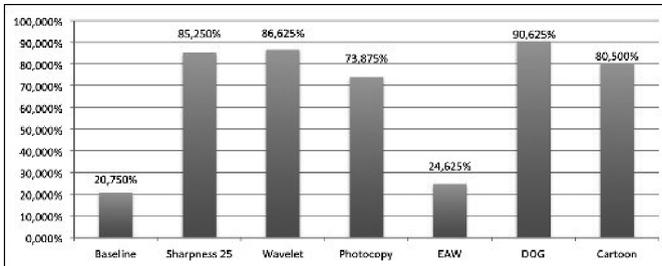


Figure 6: Comparison of the different quality scores (*EXCELLENT*)

Most of the filters are in the range of 80 %  $\pm$  7. The smallest impact is with the EAW filter. In general every presented method has a positive impact on the basic dataset (baseline).

Figure 7 shows the relationship between the FMR and the FNMR values. In Figure 6 we have seen that the quality score of the image shows a large increase, but if you look at

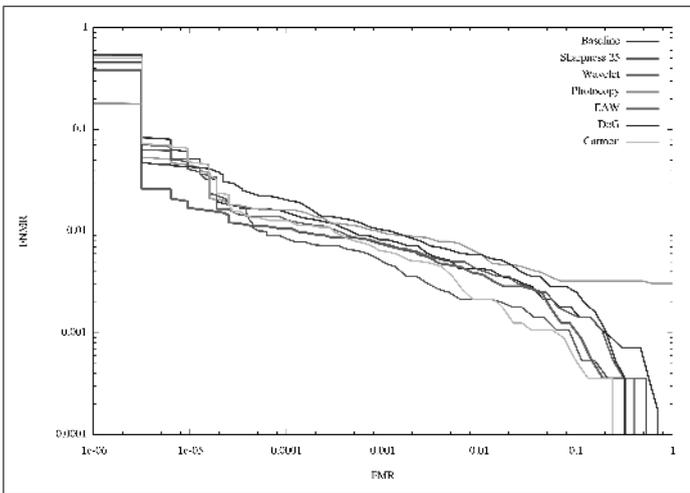


Figure 7: DET-Curve

Figure 7 there is only a significant improvement with the DoG algorithm. The photocopy and the cartoon filter are both on the lower side and EAW is closer to the baseline. Wavelet is also under the baseline and Photocopy is above the DoG filter in the last segment.

## References

- [DPP13] Stephanie F. Williams Robert C. Shaler-Akhlesh Lakhtaki Drew P. Pulsifer, Sarah A. Muhlberger. An objective fingerprint quality-grading system. *Forensic Science International*, 231(1):204–207, 2013.
- [FAF07] Gian Luca Marcialis Julian Fierrez Javier Ortega-Garcia Fernando Alonso-Fernandez, Fabio Roli. Comparison of fingerprint quality measures using an optical and a capacitive sensor. *Biometrics: Theory, Applications, and Systems, 2007. BTAS 2007. First IEEE International Conference on*, pages 1 – 6, 2007.
- [Fat09] Rannan Fattal. Edge-avoiding wavelets and their applications. *ACM Trans. Graph.*, 28(3):1–10, 2009.
- [GAKD00] S. Greenberg, M. Aladjem, D. Kogan, and I. Dimitrov. Fingerprint image enhancement using filtering techniques. *15th International Conference on Pattern Recognition*, 3:322–325, 2000.
- [HV94] P. Haeberli and D. Voorhies. Image Processing by Linear Interpolation and Extrapolation. *IRIS Universe Magazine*, pages 8–9, 1994.
- [HWJ98] L. Hong, Y. Wan, and A. Jain. Fingerprint image enhancement algorithm and performance evaluation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(8):777–789, 1998.
- [Inc15] Apple Inc. Use Touch ID on iPhone and iPad. <https://support.apple.com/en-us/HT201371>, March 2015.
- [KAY14] James Purnama Maulahikmah Galinium Kevin Arighi Yusharyahya, Anto Satriyo Nugroho. A Comparison of Fingerprint Enhancement Algorithms for Poor Quality Fingerprint Images. *International Conference of Advanced Informatics: Concept, Theory and Application (ICAICTA)*, 2014.
- [KN14] Rahul Joshi Kanpariya Nilam. Adaptive Fingerprint Image Enhancement for Low-Quality of Images by Learning From the Images and Features Extraction. *International Journal of Software and Hardware Resarch in Engineering*, 2, 2014.
- [Kri06] Dr. Rama Krishnan. *Fingerprint Capture Challenges and Opportunities*. IDENT - Biometrics Quality Lead, 2006.
- [TB06] Tim Thompson and Sue Black, editors. *Forensic Human Identification: An Introduction*. CRC Press, 2006.

# Exploring How User Routine Affects the Recognition Performance of a Lock Pattern

Lisa de Wilde, Luuk Spreeuwiers, Raymond Veldhuis

Faculty of Electrical Engineering, Mathematics and Computer Science  
University of Twente  
P.O. Box 217, 7500 AE Enschede, The Netherlands  
l.dewilde@student.utwente.nl  
l.j.spreeuwiers@utwente.nl  
r.n.j.veldhuis@utwente.nl

**Abstract:** To protect an Android smartphone against attackers, a lock pattern can be used. Nevertheless, shoulder-surfing and smudge attacks can be used to get access despite of this protection. To combat these attacks, biometric recognition can be added to the lock pattern, such that the lock-pattern application keeps track of the way users draw the pattern. This research explores how users change the way they draw lock patterns over time and its effect on the recognition performance of the pattern. A lock-pattern dataset has been collected and a classifier is proposed. In this research the best result was obtained using the x- and y-coordinate as the user's biometrics. Unfortunately, in this paper it is shown that adding biometrics to a lock pattern is only an additional security that provides no guarantee for a secure lock pattern. It is just a small improvement over using a lock pattern without biometric identification.

## 1 Introduction

Nowadays all Android smartphones can be unlocked by drawing a pattern in a grid of 3 x 3 points. A pattern is valid if it obeys three rules: the pattern connects at least four points, each point is used only once and when two points are connected by a straight line there is no unused point between them [AGM<sup>+</sup>10].

Unfortunately, it is easy for an attacker to trace the pattern of a user and unlock the smartphone by shoulder-surfing or smudge-attacks. With a shoulder-surfing attack an attacker records the user's pattern. A smudge-attack occurs when an attacker extracts information from the smudges left on the screen [SLS13]. The success chances of these attacks can be reduced by adding the user's biometrics as an authentication factor [AW12, JRP06].

During enrolment the user draws the pattern while the lock pattern application records the user's biometrics, here the location, pressure and contact area of the finger as functions of time. During verification, the user draws the pattern to unlock the screen. The application checks if the pattern is correct and compares the biometrics to the enrolled data.

In this research it is explored how the biometrics of the user evolve over time where the

user repeatedly enters the same pattern. This is done by a data analysis on a dataset collected in a user study. In this study the application *Touch Signature* was used to keep track of the way a user draws a lock pattern. It is also explored how the evolution of the way the user draws the lock pattern affects the recognition performance of the lock pattern. This is important in order to find out if the user can still access his smartphone when his drawing behaviour naturally changes. Next to that, imposters that know the owner's lock pattern should not be granted access to the smartphone.

This leads to the following research question: *How does user routine affect the recognition performance of a lock pattern?* and more specifically: *How do users change the way they draw lock patterns over time?* and *How will this affect the recognition performance of the lock pattern?*

The remainder of this paper is as follows. Section 2 discusses the methods used to answer the research questions. Section 3 presents the results of the analysis of the data that has been collected. Section 4 discusses the results and presents conclusions.

## 2 Methods

### 2.1 User study and dataset

People were asked to install the application *Touch Signature* on their Android smartphone<sup>1</sup>. The application asks to draw the same pattern (see Figure 1) ten times every day for eight days. The pattern starts at the last row's central point. While drawing, the application keeps track of where and when the finger touches the screen, the pressure from the finger on the screen and the area that is touched by the finger.

In total 144 individuals participated in the user study. Their gender, age and handedness are in Table 1. In addition the application collected information about the device: manufacturer, model, screen dpi and screen resolution. In this research the screen dpi and resolution are not taken into account.

The collected dataset is used to explore how users draw their lock pattern over a course of time. At first, the data is validated: the results of individuals who participated more than one time to the experiment were discarded from the dataset. As stated before a total of 144 participants have participated in this study. In the best case there should be eight days with ten measurements per participant. But only 43,06% of the participants did measurements for eight days (see Table 2). Besides that 4.76% of the participants did not do exactly ten measurements per day.

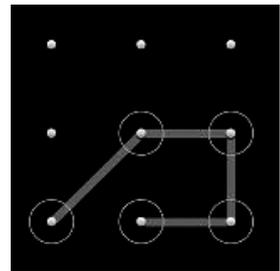


Figure 1: The lock pattern used in the user study.

---

<sup>1</sup>This application was developed by the students Stijn van Winsen, Joep Peeters, and Ties de Kock who are kindly acknowledged for this work.

Table 1: Partition of gender, age and handedness among participants.

	Amount	Percentage
<b>Gender</b>		
Female	25	17,36%
Male	119	82,64%
<b>Age</b>		
0 - 19	42	29,17%
20 - 39	96	66,67%
40 - 59	4	2,78%
60 - 79	1	0,69%
80 - 100	1	0,69%
<b>Handedness</b>		
Left	12	8,33%
Right	125	86,81%
Not sure	7	4,86%

Table 2: Partitioning of number of measurement days among participants.

Days	1	2	3	4	5	6	7	8	9	12
Number	43	5	5	3	2	7	11	62	5	1
Percentage	30%	4%	4%	2%	1%	5%	8%	43%	4%	1%

To get a first view of the measurement four 2d-plots were created (see Figure 2). Each plot contains the measurements of day one (blue lines) and day eight (dashed red lines) from one specific individual, whereby both days consists of ten measurements. The plots visualizes the x-coordinate versus time, y-coordinate versus time, pressure versus time and area touched by the finger versus time. The measurements of this individual in these plots represents the most of the measurements of other individuals.

Ideally, different measurements of one individual should be (almost) the same, so the deviation between the different lines in the plots should be small. Unfortunately, it is clearly visible that there is a large deviation of the pressure and the area touched by the finger at different measurements. On the other hand the x-coordinate and y-coordinate seem more consistent and have a small deviation. For that reason, only the x- and y-coordinate and time are used to analyse the measurements and to answer the research questions. Figure 3 shows the x- and y-coordinate over time of two individuals. Each line represents one measurement. The lines are far apart from each other, which means that person B has drawn the pattern faster than person A. This suggests that the measurements of different persons are different and that these biometrics could be used to distinguish different individuals.

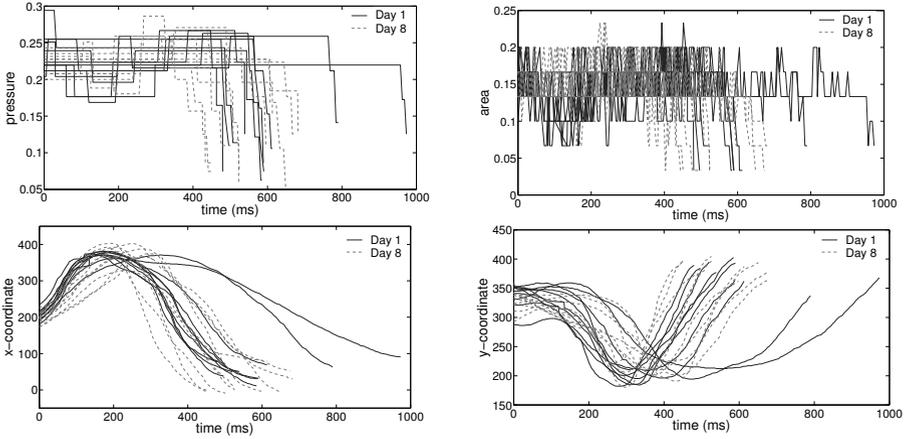


Figure 2: Top left: pressure versus time; top right: area touched by the finger versus time; bottom left: x-coordinate versus time; bottom right y-coordinate versus time.

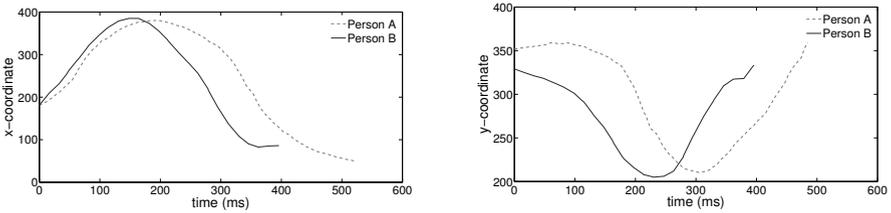


Figure 3: Left: x-coordinate versus time of two persons; right: y-coordinate versus time of two persons.

## 2.2 Classification

To compare the different measurements per individual a likelihood-ratio based classifier [SVSD14] is used. This classifier compares the biometric features of two lock patterns and produces a similarity score. The score increases with the similarity of the patterns. The score is compared with a threshold. If it is higher than the threshold the classifier decides that the lock patterns are from the same individual. Otherwise it decides that the lock patterns are from different individuals. In order to use this classifier every measurement must have the same amount of measuring points. This is achieved by inter- and extrapolating the measuring points of the dataset to 93 equidistant points, 93 being the median of the measuring points of the dataset. The classifier is based on transforming the data using two transformations. The first is called principle component analysis (PCA) and has the purpose of reducing the dimensionality of the data such that noisy components are removed. The second is called linear discriminant analysis (LDA) and retains only discriminative components. Hence, the classifier has two parameters that need to be tuned to the data: the number of PCA coefficients and the number of LDA coefficients. Given these param-

Table 3: Results of the comparison of all measurements.

Dataset	TMR@FMR=0,1	TMR@FMR=0,01	EER
x- and y-coordinate and time	53,1%	8,6%	19,0%
x- and y-coordinate	58,0%	6,5%	16,9%
x- and y-coordinate and time/time <sub>max</sub>	54,8%	12,2%	18,2%

Table 4: Results of the comparison of the second and last day.

Dataset	TMR@FMR=0,1	TMR@FMR=0,01	EER
x- and y-coordinate and time	36,3%	3,9%	23,7%
x- and y-coordinate	52,6%	6,4%	18,7%
x- and y-coordinate and time/time <sub>max</sub>	45,6%	11,1%	20,3%

eters, the transformations are learned from a a part of the data that is set aside as a training set. There is no overlap between individuals in this training set and the remaining test set. The dataset is split based on the characteristics of the participants and their smartphones. The training set consists of data of 96 individuals and the test set contains data of 48 individuals.

The classifier calculates a score matrix using the test set, containing the comparison scores of all pairs of individuals in the test set. The score matrix consists of two types of scores, namely the so-called genuine scores resulting from comparisons of individuals with themselves and the impostor scores resulting from comparing individuals with other individuals. These scores are used to plot a receiver operating characteristic (ROC)-curve which plots the true match rate (TMR) as a function of the false match rate (FMR). The closer the curve is to the upper left corner, the higher the overall accuracy of the classifier [Rev08]. To get closer to the upper left corner the numbers of PCA and LDA coefficients are optimised.

In addition to the dataset with the x-coordinate, y-coordinate and time, the analysis above is also done with two other datasets. One dataset only contains the x-coordinate and y-coordinate. The other dataset contains also the x-coordinate, y-coordinate and normalised time, which is time divided by the maximum time of that measurement (time/time<sub>max</sub>).

To get a better view of how biometrics of the user's lock pattern changes with, ROC-curves were created using only the second and last measurement day. The first day is not used, because that was the first day the participants have drawn the pattern and could contain some large differences in the way of drawing. The last day should be day eight, but not every participant completed exactly eight measurement days. For that reason the last day in the ROC-curve is the last day the participant has drawn the pattern, this day varies from day 5 to day 9. In the test set, there were 26 participants who did measurements on the second day and day 5 or higher (2x day 9, 15x day 8, 4x day 7, 4x day 6 and 1x day 5). To create new ROC-curves containing only measurements of the second and last day, a new score matrix is created. This is done by extracting the scores of these days from the original score matrix.

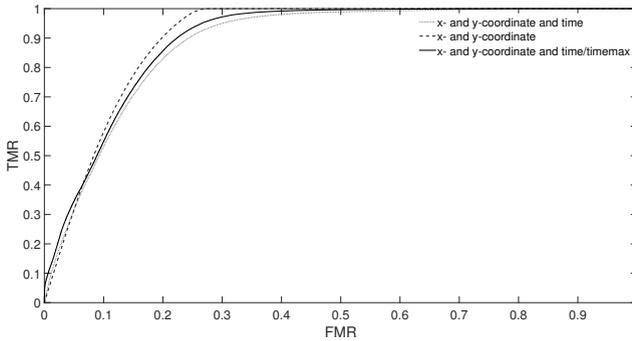


Figure 4: ROC-curves for different features. Dashed blue: x- and y-coordinate (PCA 50 and LDA 49); solid red: x- and y-coordinate and time (PCA 50 and LDA 49, solid black: x- and y-coordinate and time/time<sub>max</sub> (PCA 7 and LDA 4).

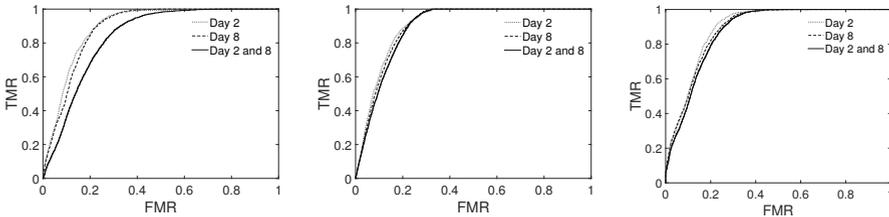


Figure 5: ROC-curves for different features across time. Left: x- and y-coordinate and time; centre: x- and y-coordinate; right: x- and y-coordinate and time/time<sub>max</sub>.

### 3 Results

ROC-curves were created with the dataset containing the x-coordinate, y-coordinate and time. The deviation between different numbers of PCA and LDA coefficients was small. The best result was achieved using 50 for the PCA and 49 for the LDA (see Fig 4, red curve). For the ROC-curve with the dataset with the x-coordinate and y-coordinate the same PCA and LDA coefficients were the best choice (see Figure 4, dashed blue curve). Finally, the dataset with x-coordinate, y-coordinate and time/time<sub>max</sub> was used to create ROC-curves. In this case PCA 7 and LDA 4 yielded the best result (see Figure 4, black curve). These curves show the true match rate (TMR) as a function of the false match rate (FMR). It is assumed that the owner of the smartphone and the imposter who impersonate the user's biometrics always draw the correct pattern. Table 3 shows the chances that the owner of the smartphone can get access to his smartphone when the chances that an imposter can get access to the smartphone is 10% or 1%. The higher TMR, the more secure the lock pattern and the better the dataset. Next to that, the best description of the error rate, the equal error rate (EER) is given (see Table 3). The biometrics of the dataset with the lowest EER contains the least errors and are the most secure to use.

The three plots with the comparison of the second and last day (see Figure 5) are made to explore how users change the way they draw lock patterns over a course of time. The smaller the distance between the different lines, the smaller the changes in the way of drawing the lock pattern. The leftmost plot is created using the x-coordinate and y-coordinate and time and has a small deviation between the second and the last day and the curve of comparison of the second and last day has a larger deviation and is slightly more to the right. The centre plot is created using only the x-coordinate and y-coordinate has all three ROC-curves close to each other and they are more to the upper left corner than the curves of the first plot. Finally, the ROC-curves in the rightmost plot using the x-coordinate and y-coordinate and time/time<sub>max</sub> are also close to each other, but is less close to the upper left corner than the centre plot.

For the ROC-curves of the comparison of the second and last day the TMRs are given as well for a FMR of 10% and 1% for the three different datasets (see Table 4). Besides that, the EER is calculated for all three datasets (see Table 4). The biometrics used in the dataset with the highest TMR and lowest EER are the most secure over a course of time.

## 4 Discussion and conclusions

To collect a dataset a user study has been done with use of the application "Touch Signature". This application kept track of five of the user's biometrics: the location of the finger (x- and y-coordinate), the time to draw the pattern (in milliseconds), the pressure of the finger and the area touched by the finger. During this research it was concluded that the pressure and the area touched by the finger deviated too much to use for this research. The other three biometrics (x- and y-coordinate and time) were used in three different datasets to classify the measurements. The first dataset contains all three biometrics and the second only the x- and y-coordinate. Finally, the third dataset contains next, to the x-coordinate and y-coordinate, the normalized time which is: time/time<sub>max</sub>.

The best dataset is determined by the height of the TMR when the FMR is 10% or 1% and the EER. The higher the TMR and the lower the EER, the more secure the lock pattern is. When a FMR of 10% is desirable the dataset using the x- and y-coordinate has the highest TMR, thus is the best dataset (see Table 3 and 4). The TMR for this dataset is 58,0% for all measurements together and 52,6% for the comparison of the second and last day. Thereafter the dataset using the x- and y-coordinate and time/time<sub>max</sub> is the best and at last the dataset using the x- and y-coordinate and time. This order is also the best for the EER, whereby the highest EER of 16,9% for all measurements together and 18,7% for the comparison of the second and last day using the x- and y-coordinate (see Table 3 and 4).

According to a FMR of 1%, the dataset using the x- and y-coordinate and time/time<sub>max</sub> is the best dataset with a FMR of 12,2% for all measurements and 11,1% for the comparison of the second and last day. In this case the dataset using the x- and y-coordinate is the second best dataset to use (see Tables 3 and 4).

In conclusion, the way users draw their pattern over a course of time changes the most using the dataset with the x- and y-coordinate and time. In other words the time in which

a user draws a pattern differs relatively more per drawing than the location of the finger. The normalized time gives a better result than just the time to draw a pattern. Still, the way users draw their lock pattern over a course of time changes the least using a lock pattern with the location of the finger as user's biometric. Thus, the lock pattern using the x- and y-coordinates can be used best. However, when a FMR of 1% is desired, the biometric time/time<sub>max</sub> should be added to the lock pattern application.

Finally, the less the user's biometrics changes, the higher the TMR and the lower the EER, with the consequence that the recognition performance of the lock pattern will be higher. Unfortunately, the TMR is still too low and the EER is still too high to give reliable results. In other words, an imposter can still access the smartphone easily when he knows the pattern, however it is made more difficult due the addition of the user's biometrics. Concluding, the lock pattern using user's biometrics can be used as an additional security, but provides no guarantee for a secure lock pattern.

## References

- [AGM<sup>+</sup>10] Adam J Aviv, Katherine Gibson, Evan Mossop, Matt Blaze, and Jonathan M Smith. Smudge Attacks on Smartphone Touch Screens. *Proceedings of the 4th USENIX Conference on Offensive Technologies, WOOT'10*, 10:2, 2010.
- [AW12] Julio Angulo and Erik Wästlund. Exploring Touch-Screen Biometrics for User Identification on Smart Phones. In *Privacy and Identity Management for Life*, volume 375 of *IFIP Advances in Information and Communication Technology*, page 139. 2012.
- [JRP06] AK. Jain, A Ross, and S. Pankanti. Biometrics: a tool for information security. *Information Forensics and Security, IEEE Transactions on*, page 125, 2006.
- [Rev08] Kenneth Revett. *Behavioral biometrics: a remote access approach*. John Wiley & Sons, 2008.
- [SLS13] Muhammad Shahzad, Alex X. Liu, and Arjmand Samuel. Secure Unlocking of Mobile Touch Screen Devices by Simple Gestures: You Can See It but You Can Not Do It. In *Proceedings of the 19th Annual International Conference on Mobile Computing & Networking, MobiCom '13*, page 39, 2013.
- [SVSD14] L. J. Spreeuwens, R. N. J. Veldhuis, S. Sultanali, and J. Diephuis. Fixed FAR Vote Fusion of regional Facial Classifiers. In C. Busch and A. Brömme, editors, *BIOSIG 2013: Proceedings of the 13th International Conference of the Biometrics Special Interest Group, Darmstadt, Germany, Darmstadt, September 2014*. Gesellschaft für Informatik.

# JPEG Optimisation for Fingerprint Recognition: Generalisation Potential of an Evolutionary Approach

Thomas Herzog, Andreas Uhl  
{therzog, uhl}@cosy.sbg.ac.at

Department of Computer Sciences, University of Salzburg  
Jakob-Haringer-Straße 2, Salzburg, Austria

**Abstract:** For fingerprint-based biometric systems, JPEG has traditionally been one of the primary storage formats. In this paper, we investigate methods of optimising JPEG compression for increased matching scores at constant compression rates. To achieve this, an evolutionary approach is employed to optimise the quantisation matrix used by JPEG. While finding matrices that are better suited for different sets of fingerprints, different fingerprint sensors, and different recognition algorithms as they have been optimised for remains elusive, we find improvements in matching performance for dataset fitting scenarios.

## 1 Introduction

Due to its inclusion in a former version of the ISO/IEC 19794 standard on Biometric Data Interchange Formats, biometric systems relying on fingerprint data have traditionally used JPEG [PM93] as their image storage format, besides other specialised formats such as WSQ [BBH93] or, more recently, JPEG2000, the current ISO/IEC 19794 recommendation. The optimisation of existing standardised compression algorithms to meet the specific properties of the biometric data to be compressed is a natural strategy. For example, JPEG quantisation matrix (QM) optimisation has been done to optimise results in face recognition [JKAA06] and iris recognition [KSU09]. JPEG2000 has been optimised w.r.t. possible wavelet packet decomposition structures for fingerprint recognition [MSU10] and iris recognition [HUKU13]. Even JPEG XR has been optimised for the iris recognition context [HSU12]. Optimisation of JPEG for fingerprint image compression is unexplored up to now.

Thus, in this paper, we will suggest to optimise the JPEG QM for optimal usage in fingerprint recognition by employing evolutionary optimisation. In particular, we will focus on the potential generalisation of evolved QMs to different datasets as well as different feature extraction and matching schemes. In the remainder of this section, we will motivate our approach. Section 2 explains the methodology used in this study while detailed results are presented in Section 3. Section 4 concludes the paper.

Because our goal is to optimise the QM, it makes sense to first look at the distribution of DCT coefficients. To this end, we took two generic images (photographs), and visualised



Figure 1: Average DCT coefficients of generic and fingerprint images (lighter areas indicate higher values)

the average of DCT coefficients over all image blocks, and did the same for two fingerprint images. DCT coefficients are laid out in a matrix: horizontal frequency increases as the  $x$  coordinate (left to right) increases, vertical frequency increases as the  $y$  coordinate (top to bottom) increases. Thus, the top left corner contains the lowest frequency and the lower right corner the highest frequency. As can be seen in Figure 1, more energy is contained in the lower frequency coefficients for fingerprints and increasingly high frequencies seem to be less relevant for fingerprint images than for generic images, at least for the four chosen example images.

The default QM in the luminance channel (as shown in Figure 3a) was obtained from a series of psychovisual experiments, and has “been known to offer satisfactory performance, on the average, over a wide variety of applications and viewing conditions” [Bov09]. As is stated in the JPEG specification itself: “These tables are provided as examples only and are not necessarily suitable for any particular application.”

Our hypothesis was that, since the default QM was designed for the general case, it should be possible to find better candidates for specific use cases. To test this hypothesis, we created a manually crafted QM, based on the default QM but with most higher frequency components set to 255 (the maximum value for baseline JPEG) due to the observations in Fig. 1. This QM (which we will call “manual QM” from now on) can be seen in Figure 3b.

In this paper, we employ genetic algorithms to optimise the QM for fingerprint images in general, fingerprint images from a specific scanning device, and a specific set of fingerprint images (dataset fitting). Lastly, we investigate whether the results obtained with a specific fingerprint recognition algorithm do generalise to different algorithms.

## 2 Methodology

### 2.1 Evolution

Our main approach for finding better adapted QMs was to use genetic algorithms. We used *Watchmaker* [Dye10], an open source framework for implementing genetic algorithms

Database	Description
FVC2002 DB3	capacitive sensor "100 SC" (Precise Biometrics)
FVC2004 DB1	optical sensor "V300" (CrossMatch)
FVC2004 DB2	optical sensor "U.are.U 4000" (Digital Persona)
FVC2004 DB3	thermal sweeping sensor "FingerChip FCD4B14CB" (Atmel)

Table 1: Fingerprint Databases

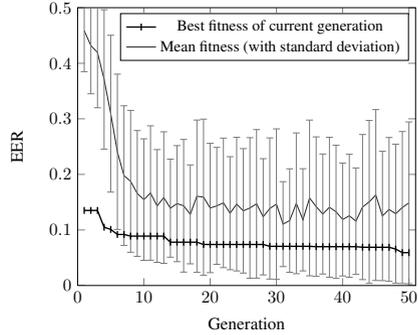


Figure 2: Example of fitness increase during an evolutionary run. (Data for this graph was collected by evolving a QM for 10 fingers (80 impressions) from FVC2002 DB3 over 50 generations.)

in Java. As source data, fingerprints from the databases of the Fingerprint Verification Competitions (FVC) of 2002 and 2004 were used, as shown in Table 1. Each database contains 8 impressions of 110 fingers.<sup>1</sup> We restricted our experiments to subsets of the first 100 fingers per database.

The basic idea of a genetic algorithm is as follows: There exists a population of individuals, and a fitness function that assigns each individual a specific fitness (a numeric value). The goal is to either maximise or minimise this fitness. An evolutionary run consists of multiple generations. In each generation, depending on the selection scheme, a subset of individuals are chosen to produce offspring (mostly via cross-over and mutation) for the next generation. The offspring constitute a new population and the algorithm starts over. This continues until some termination criteria is reached (e.g. a maximum number of generations, a target fitness threshold, fitness stagnation, ...).

Here, the individuals are QMs, represented by an array of  $8 \times 8 = 64$  integer values (the quantisation coefficients). To get from a matrix to a one-dimensional array, the QM values are scanned in zigzag order [PM93]. As fitness function for a QM, the equal error rate (EER) is computed following the official FVC protocol [MMJP05]. Before that, the entire dataset subject to optimisation is compressed with JPEG as close as possible to the specified compression ratio (CR) using this QM.

Note that for evolutionary optimisation of the QMs, we use the freely available biometric software *NBIS* [NIS13] for template extraction (MINDTCT is used for minutiae extraction) and matching (using BOZORTH3).

For evolutionary cross-over, we use the existing class `IntArrayCrossover`, configured to use a single cross-over point. Given two parent individuals QM  $a$  and  $b$ , an index

<sup>1</sup>In our copy of the data, images from FVC2004 DB3, fingerprint sets 5 and 54-60 were corrupt. Those sets have, for our experiments, been replaced with sets 102-109.

(point)  $x \in \{0, \dots, 63\}$  is selected at random to generate the child  $c$ , where

$$c_i = \begin{cases} a_i & \text{if } i \leq x \\ b_i & \text{if } i > x \end{cases}, 0 \leq i < 64$$

For evolutionary mutation, we use an instance of the custom class `IntArrayMutation`, which (with a probability of  $p = 0.02$  per array element) mutates an element  $x$  by a random value

$$m \in \mathbb{N}, -42 < m < 42$$

so that the final value is

$$y = \max(\min(x + m, 255), 0)$$

## 3 Experiments

### 3.1 Setup

Each evolution run in our experiments has the same basic steps and parameters: An instance of `QuantisationMatrixEvaluator` is created and parameterised with the employed fingerprint database, the number of individual fingers used from this database (10 to 100, in steps of 10), and the compression rate to aim for (which is set to 30 for our experiments, since this value represents a good compromise of file size reduction and still sensible matching results). A custom `IntArrayFactory` is created, which can create arrays of  $8 \times 8 = 64$  randomly generated integer values of the form

$$x \in \mathbb{N}, 0 \leq x < 256$$

which are used as the initial population of QMs. A *Watchmaker* evolution engine instance is created and configured to use a roulette wheel selection algorithm, and a Mersenne Twister pseudorandom number generator. In addition to the initial random population, the engine is seeded with two predefined QMs: The default luminance QM from Figure 3a, and the manually crafted QM from Figure 3b. The evolution engine is started to run with a population of 100 individuals and an elite count<sup>2</sup> of 5 for 50 generations.

Figure 2 shows example results of an evolutionary run that took 12 hours and 3 minutes to complete on a Windows 7 (64 bit) computer with an Intel<sup>®</sup> Core<sup>™</sup> i5-2300 CPU and 8 GB of RAM.

### 3.2 Experimental Results

A visualization of two evolved QMs is shown in Figure 3, next to the default luminance (Figure 3a) and manual (Figure 3b) QMs.

---

<sup>2</sup>i.e. the number of individuals with highest fitness that are copied over to the next generation unchanged

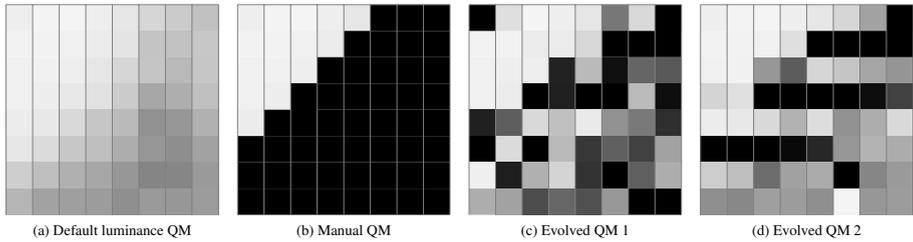


Figure 3: Visualization of select QMs (darker areas correspond to higher values)

The QM shown in Figure 3c was evolved using a set of 10 fingers (i.e. 80 impressions) from FVC2002 DB3 over 50 generations. Figure 3d shows a QM that was evolved on the same DB, but using a set of 100 fingers (i.e. 800 impressions). This suggests that it is possible, using an evolutionary search approach, to find non-obvious QM optimisations for a given set of images and a given application (i.e. biometric matching).

Figure 4 shows the results for dataset fitting ( $\text{---}\triangle\text{---}$ ), i.e. a QM that is evolved for a certain set of fingers is evaluated over the same set of fingers of the identical database. The results are compared with default luminance ( $\text{---}\times\text{---}$ ) and manual ( $\text{---}\text{+}\text{---}$ ) QM performance.

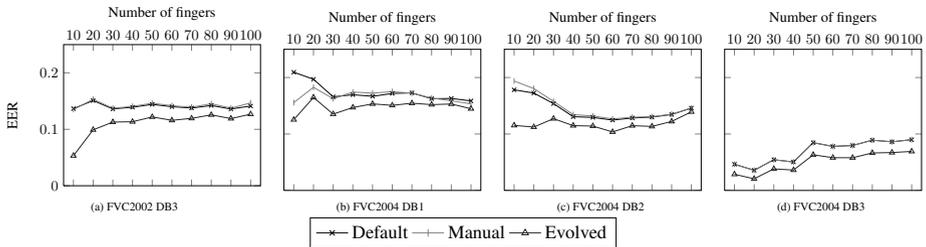


Figure 4: Dataset fitting EER for default, manual and evolved QMs using *NBIS*

We clearly notice that the EER is improved (i.e. decreased) for all four datasets in case of the evolved QM. Moreover, it turns out that, except for FVC2004 DB3, a low number of fingers involved in the optimisation process leads to higher gains in EER. This is to be expected since the adaptation process can really adapt to a low number of fingers almost individually.

To examine intra-database generalisation behaviour, we evolved QMs on a set of  $n = 10k, k \in \{1, \dots, 9\}$  fingers, and evaluated the resulting QM on the remaining  $100 - n$  fingers of the same database. The results are shown in Figure 5 ( $\text{---}\circ\text{---}$ ), plotted against performance of the default luminance ( $\text{---}\times\text{---}$ ) and manual ( $\text{---}\text{+}\text{---}$ ) QMs.

Thus we can conclude that even for fingerprint images from the same sensor, matching performance gains from QM optimisation do not readily generalise across different sets of fingerprint impressions.

Figure 6 shows the EER for each of the four databases, when using QMs evolved for each

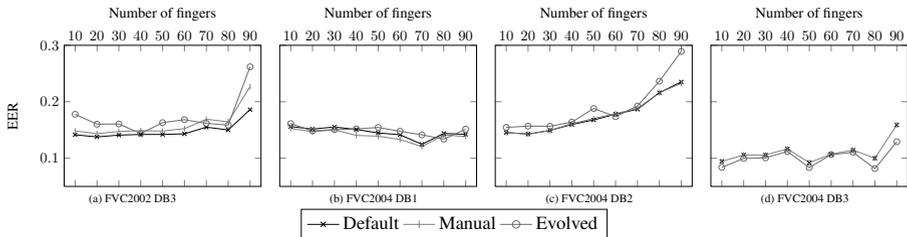


Figure 5: Intra-database EER for default, manual and evolved QMs. (Note that the number of fingers  $n$  indicates the number used for evolving the QM.)

database separately (on 100 fingers each, i.e. dataset fitting), compared with results for the default luminance and manual QMs (inter-database generalisation).

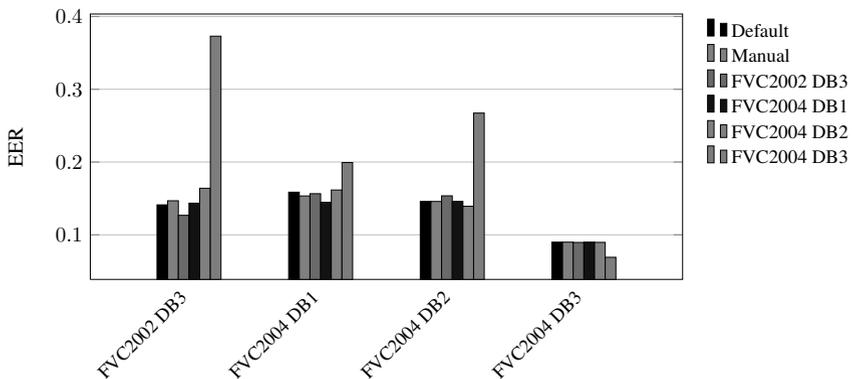


Figure 6: EER for default, manual and evolved QMs.

The EER when using a QM evolved for the same DB (on the same set of 100 fingers) clearly improves (decreases) as observed before. In contrast, when using “foreign” QMs (excluding the default and manual QMs), the average EER *increases* significantly. Using the manual QM leads to a slight average performance decrease. These results show that matching performance gains resulting from QM optimisation for a specific set of fingerprint images do not readily generalise to other sources of fingerprint image data.

Finally, we examine matching performance of different (non-*NBIS*) template extraction and matching engines when using QMs evolved using *NBIS* – cross-algorithm generalisation: *VeriFinger* and *Phase Only Correlation Matcher* (as re-implemented in [HUPU13]) are used.

Figures 7 and 8 show detailed results for cross-engine evaluation, with performance data for the default luminance ( $-x-$ ), manual ( $-+-$ ) and evolved (*VeriFinger* ( $-□-$ ) and *Phase Only Correlation* ( $-◊-$ ) QMs plotted against each other.

These results indicate that, while small performance increases appear for *Phase Only Cor-*

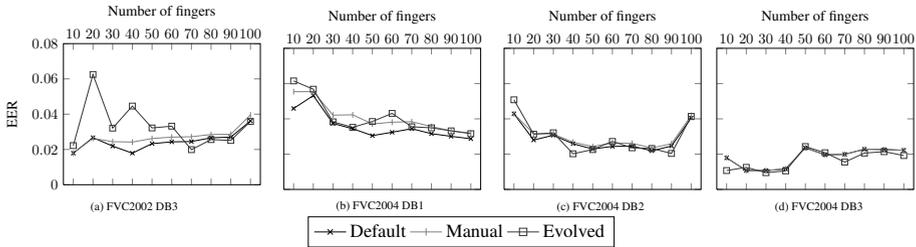


Figure 7: Dataset fitting EER for default, manual and evolved QMs using *VeriFinger*

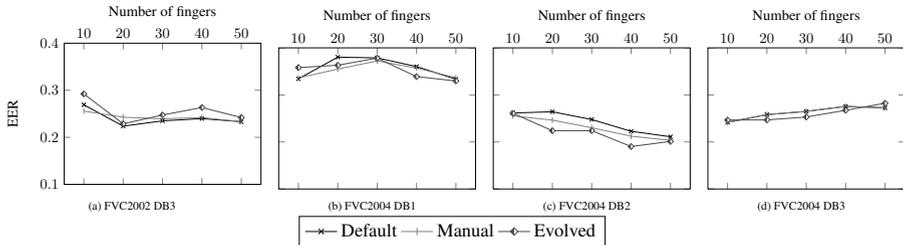


Figure 8: Dataset fitting EER for default, manual and evolved QMs using *Phase Only Correlation*

*relation*, significant performance gains (like those evidenced in the case of dataset fitting using *NBIS*) do not generalise to other template extraction and matching algorithms. In the case of *VeriFinger*, the manual as well as the evolved QMs lead to significantly worse performance.

## 4 Conclusion

Our experiments show that, while it is possible to tailor-fit JPEG for a specific set of fingerprint images by finding appropriate quantisation matrices via genetic algorithms, the resulting matching performance improvements do not, on average, generalise to other sets of fingerprint images, be it from the same database (and thus the same type of sensor) or from different ones. In addition, the evolved quantisation matrices do not, on average, lead to improved performance when employing template extraction and matching algorithms different than those used for evolution. The most promising result is an increase of on average more than 20% in matching performance for dataset fitting scenarios. Similar results and trends of poor generalisation behaviour have been reported for optimising JPEG2000 Part2 wavelet packet subband structures for iris recognition [HUKU13].

## References

- [BBH93] Jonathan N. Bradley, C. M. Brislawn, and T. Hopper. The FBI Wavelet/Scalar Quantization Standard for Gray-scale Fingerprint Image Compression. In *SPIE Proceedings, Visual Information Processing II*, volume 1961, pages 293–304, Orlando, FL, USA, April 1993.
- [Bov09] A.C. Bovik. *The Essential Guide to Image Processing*. Elsevier Press, 2009.
- [Dye10] Daniel W. Dyer. Watchmaker Framework for Evolutionary Computation. Online: <http://watchmaker.uncommons.org/>, January 2010.
- [HSU12] Kurt Horvath, Herbert Stögner, and Andreas Uhl. Optimisation of JPEG XR quantisation settings in iris recognition systems. In P. Davies and D. Newell, editors, *Proceedings of the 4th International Conference on Advances in Multimedia (MMEDIA 2012)*, pages 88–93. IARIA, 2012.
- [HUKU13] J. Hämmerle-Uhl, M. Karnutsch, and A. Uhl. Evolutionary Optimisation of JPEG2000 Part 2 Wavelet Packet Structures for Polar Iris Image Compression. In *Proceedings of the 18th Iberoamerican Congress on Pattern Recognition (CIARP'13)*, volume 8258 of *Springer LNCS*, pages 391–398, Havana, Cuba, 2013.
- [HUPU13] J. Hämmerle-Uhl, M. Pober, and A. Uhl. Towards Standardised Fingerprint Matching Robustness Assessment: The StirMark Toolkit – Cross-Database Comparisons with Minutiae-based Matching. In *Proceedings of the 1st ACM Workshop on Information Hiding and Multimedia Security (IH&MMSec'13)*, pages 111–116, Montpellier, France, June 2013.
- [JKAA06] G.-M. Jeong, C. Kim, H.-S. Ahn, and B.-J. Ahn. JPEG Quantization Table Design for Face Images and Its Application to Face Recognition. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Science*, E69-A(11):2990 – 2993, 2006.
- [KSU09] Mario Konrad, Herbert Stögner, and Andreas Uhl. Custom Design of JPEG quantization tables for compressing iris polar images to improve recognition accuracy. In M. Tistarelli and M.S. Nixon, editors, *Proceedings of the 3rd International Conference on Biometrics 2009 (ICB'09)*, volume 5558 of *LNCS*, pages 1091–1101. Springer Verlag, 2009.
- [MMJP05] Davide Maltoni, Dario Maio, Anil K Jain, and Salil Prabhakar. *Handbook of Fingerprint Recognition*. Springer Science+Business Media, Inc., 2005.
- [MSU10] B. Mühlbacher, T. Stütz, and A. Uhl. JPEG2000 Part 2 wavelet packet subband structures in fingerprint recognition. In P. Frossard, H. Li, F. Wu, B. Girod, S. Li, and G. Wei, editors, *Visual Communications and Image Processing 2010 (VCIP'10)*, number 7744 in *Proceedings of SPIE*, pages 77442C–1 – 77442C–10, Huang Shan, China, July 2010. SPIE.
- [NIS13] NIST. NIST Biometric Image Software, Version 4.2.0. Online: <http://www.nist.gov/itl/iad/ig/nbis.cfm>, October 2013.
- [PM93] W.B. Pennebaker and J.L. Mitchell. *JPEG – Still image compression standard*. Van Nostrand Reinhold, New York, 1993.

# Corneal Topography: An Emerging Biometric System for Person Authentication

Nassima Kihal<sup>1,2</sup>, Arnaud Polette<sup>2,3</sup>, Salim Chitroub<sup>1</sup>, Isabelle Brunette<sup>4</sup>  
and Jean Meunier<sup>2</sup>

<sup>1</sup>Signal and Image Processing Laboratory Electronics and Computer Science Faculty, USTHB, Algiers, Algeria.

<sup>2</sup>Department of Computer Science and Operations Research (DIRO), University of Montreal, Canada.

<sup>3</sup>Aix-Marseille University, CNRS, LSIS UMR 7296, France.

<sup>4</sup>Maisonneuve-Rosemont Hospital, Department of Ophthalmology, University of Montreal, Canada.

Corresponding author e-mail adress: meunier@iro.umontreal.ca

**Abstract:** Corneal topography is a non-invasive medical imaging technique to assess the shape of the cornea in ophthalmology. In this paper we demonstrate that in addition to its health care use, corneal topography could provide valuable biometric measurements for person authentication. To extract a feature vector from these images (topographies), we propose to fit the geometry of the corneal surface with Zernike polynomials, followed by a linear discriminant analysis (LDA) of the Zernike coefficients to select the most discriminating features. The results show that the proposed method reduced the typical d-dimensional Zernike feature vector ( $d=36$ ) into a much lower  $r$ -dimensional feature vector ( $r=3$ ), and improved the Equal Error Rate from 2.88% to 0.96%, with the added benefit of faster computation time.

## 1 Introduction

Biometrics refers to identity recognition of persons according to their physical or behavioral characteristics [CDJ05] [LBS11] [D03] [ZK11]. Many physical body parts and personal features have been used for biometric systems: fingers, hands, feet, faces, irises, retinas, ears, teeth, veins, voices, signatures, typing styles, gaits, odors, and DNA. Person recognition based on biometric features has attracted more attention in designing security system. In this paper we present a new biometric system based on corneal topography. Corneal topography is a non-invasive medical imaging technique to assess the shape of the cornea in ophthalmology. Figure 1 shows typical corneal topographies (images) of the anterior surface elevation from 2 different subjects. These images (a.k.a. elevation maps) show the measured height with respect to a reference (best-fit) sphere with pseudo-colors where warm colors depict points higher than the sphere and cool colors correspond to lower points. One can easily see that these maps are different from one individual to the other (uniqueness). The idea of using this physical characteristic for biometrics also comes from its stability during the life of the person

(permanence) [BCI01]. However with age, the shape of the anterior and the posterior corneal surface might change slightly [DSV06], but this is a slow process that would only necessitate occasional update (e.g. every 5 years). Also corneal topographies are more practical to manipulate (measurability) compared with other data of biometric modalities that require pretreatment such as filtering, extraction of the region of interests etc. Actually, the corneal shape is certainly suitable for biometrics because it satisfies the following requirements: *Universality, Distinctiveness, Permanence and Collectability* [RNJ06].

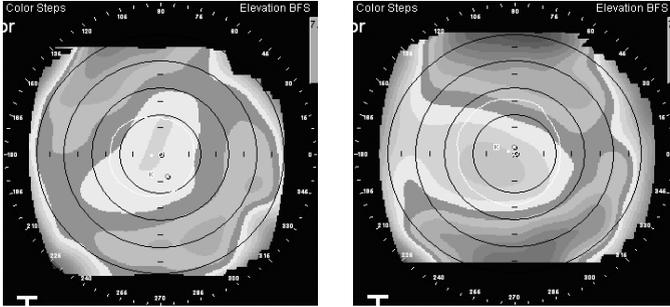


Figure 1: Typical topographies of 2 individuals (anterior surface elevation maps).

The proposed method describes the prototype of a biometric recognition system based on cornea where the corneal surface is modeled by using a Zernike polynomial decomposition [JGJ95][W92] limited to the first 36 coefficients ( $C_0^0$  to  $C_7^7$ ) and compared to evaluate their potential as biometric indicators. Our work extends the work of N.D. Lewis [L11] and shows that the corneal shape can really be a good biometric alternative for individual recognition by selecting the most discriminating features of its geometry. For this reason we propose to apply Zernike polynomial decomposition, and then, LDA (*linear discriminant analysis*), [SML10] [NT10], to find better shape features. This new biometric system based on corneal topography is described as a block-diagram in Figure 2.

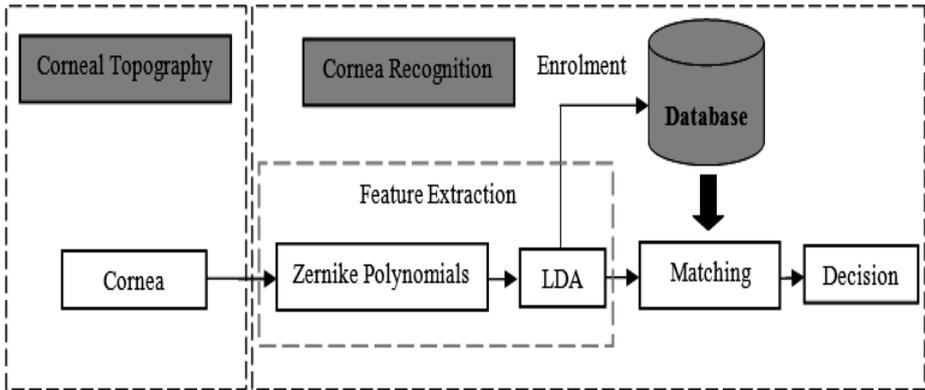


Figure 2: Block-diagram of the proposed cornea authentication system

## 2. Description of the cornea database

### 2.1 Cornea

The cornea is the outer transparent part of the eye, and covers nearly a fifth of the eyeball surface, with an average diameter of 11 mm. It is the main lens of the eye, responsible for two-thirds of the dioptric power (the remaining third is the eye lens), that transmits and focuses light into the eye with a refractive index of 1.377. The curvature radius of the anterior surface varies between 7 and 9 mm and is approximately 6.5 mm for the posterior surface. In this paper only the anterior surface geometry is considered, but the methodology could be applied to the posterior surface as well, or a combination of both.

### 2.2 Data and capture device

Corneal topography is a medical imaging method for the examination of the corneal shape. It is fast and easy (within a few seconds) and could be adapted and simplified for biometric applications in the future. Currently a corneal topographer is relatively expensive (\$20K and up) but this price tag could diminish with its wider use for biometrics. In this paper elevation maps are considered because they provide the full shape of the cornea while curvature maps are useful but limited to localized variations of shape. However curvature will be investigated in the future since they have had some success in matching 3D biometric data in the past [KK13] [KKZ11]. The database was done by using the Orbscan II topographer (Bausch & Lomb). Utilizing a scanning slit of light, it gives anterior (and posterior) surface elevation data with an error margin of 1 micron. The corneal shape was recorded as a uniformly spaced  $101 \times 101$  grid (image) of raw elevations ( $Z$ ), spaced by 0.1 mm along the  $X$  and  $Y$  axis. The cornea database is composed of 104 subjects, each has two (within-class) measures to assess repeatability leading to a total of 208 images (elevation topographies).

## 3. Feature extraction

In order to extract the features of the corneal shape, we present a methodology for analysing Orbscan II data. The technique involves decomposing the corneal height data in terms of the orthonormal set of Zernike polynomials [AHB94] [W92]. Then, a linear discriminant analysis (LDA) is used to select the features (combination of coefficients) which are the most effective to produce optimal cluster separability and consequently accurate recognition results.

### 3.1 Zernike polynomials

The Zernike polynomials are a set of functions  $\{Z_n^{\pm m}(\rho, \theta)\}$  that are orthonormal over the continuous unit circle. They have been used extensively for phase contrast microscopy, optical aberration theory, and interferometric testing to fit wave-front data. These functions are characterized by a polynomial variation in the radial direction  $\rho$  (for  $0 \leq \rho \leq 1$ ) and a sinusoidal variation in the azimuthal direction  $\theta$ . The polynomials are defined mathematically by

$$Z_n^{\mp m} = \begin{cases} \sqrt{2(n+1)}R_n^m(\rho) \cos m\theta & \text{for } +m \\ \sqrt{2(n+1)}R_n^m(\rho) \sin m\theta & \text{for } -m \\ \sqrt{(n+1)}R_n^m(\rho) & \text{for } m = 0 \end{cases} \quad (1)$$

Where

$$R_n^m(\rho) = \sum_{s=0}^{\frac{n-m}{2}} \frac{(-1)^s (n-s)!}{s! \left(\frac{n+m}{2} - s\right)! \left(\frac{n-m}{2} - s\right)!} \rho^{n-2s} \quad (2)$$

$n$  is the order of the polynomial in the radial direction  $\rho$ , and  $m$  is the frequency in the azimuthal direction  $\theta$ . Since the Zernike polynomials are orthogonal over the continuous unit circle and the lower-order terms represent familiar corneal shapes. They appear to be an ideal set of functions for decomposing and analyzing corneal surface height. The reader is referred to [JGJ95] for more details on the use of Zernike polynomials for 3D surface shape encoding.

### 3.2 Preliminary tests

The corneal height data were decomposed into a linear combination of the Zernike functions, we took the first 36 Zernike coefficients as a feature vector for one cornea ( $d=36$ ). For each individual we therefore have two feature vectors (two measures) of size 36. To show that the corneal topography can be a good biometric alternative, two sets of comparison were processed, 104 matching comparisons (with two different acquisitions from the same subject) and 5356 non-matching-comparisons (with two different acquisitions from two different subjects), by computing the absolute difference (AD) between all coefficients. Figure 3 shows the mean AD for each coefficient for the two tests. The more the difference between green and red bars for a particular

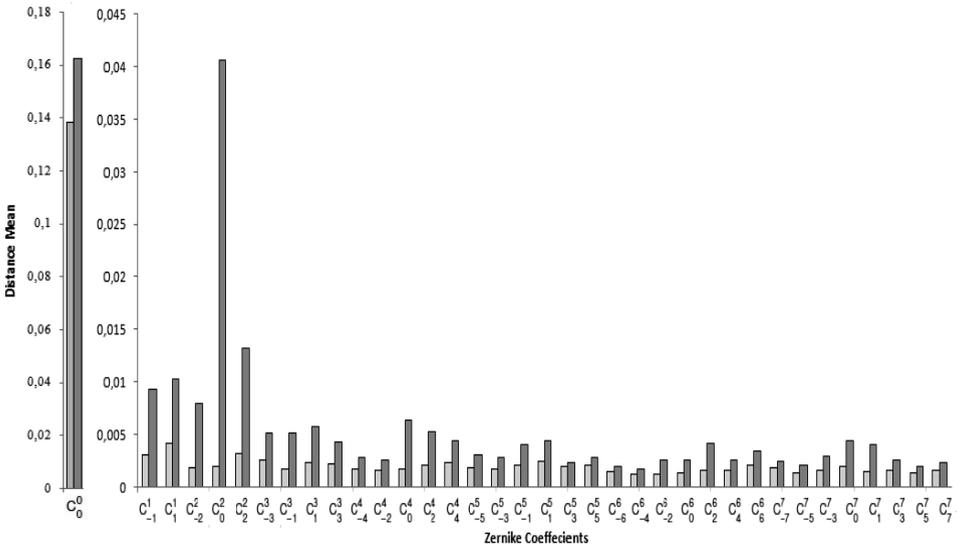


Figure 3: Mean difference for each Zernike coefficients within (green) and between (red) classes.

coefficient, the more this coefficient is selective for a biometric application. For this reason, we propose to select a combination of the most informative coefficients with LDA in the next section.

### 3.3 LDA for feature extraction

Linear discriminant analysis is a powerful method for pattern recognition yielding an effective representation that linearly transforms the original data space into a lower dimensional feature space where the data is as well separated as possible. We briefly describe it. Suppose that there are  $c$  classes ( $c=104$ ) and each class has  $n$  training feature vector samples ( $n=2$ ). The between-class and total scatter matrices of LDA are calculated using Eq (4) and Eq (5):

$$S_b = \sum_{i=1}^c (a_i^j - a_i)(a_i^j - a_i)^T \quad (4)$$

$$S_t = \sum_{i=1}^c \sum_{j=1}^n \left( a_i^j - \bar{a} \right) \left( a_i^j - \bar{a} \right)^T \quad (5)$$

where  $a_i^j$  denotes the  $j$ th training sample of the  $i$ th class,  $a_i$  stands for the mean of the  $i$ th class and  $\bar{a}$  represents the mean of all the training samples. The eigen-equation of LDA is as follows:

$$S_b v = \lambda S_t v \quad (6)$$

If all the eigenvalues of Eq (12) are ordered  $\lambda_1 \geq \lambda_2 \geq \dots$  and the corresponding eigenvectors are,  $v_1, v_2, \dots$ , LDA exploits the eigenvectors corresponding to the first largest eigenvalues to transform the original  $d$ -dimensional vector of each sample ( $d=36$ ) into a  $r$  dimensional vector. Let  $x = [x_1 \dots x_d]^T$  denote a sample, the LDA-based feature extraction result for  $x$  with regard to the first eigenvector is:

$$y_1 = x^T v_1 \quad (7)$$

and so on for the other eigenvectors. Table1 shows an example of the Zernike coefficients and the results of LDA for one individual, where  $x=A_1$  and  $A_2$  represent the first and the second acquisition respectively. The  $\lambda_i$  represent eigenvalues in descending order.  $v_1$  is the first eigenvector corresponding to the largest eigenvalue  $\lambda_1$ . From the eigenvalues in Table1 we see that the original information is mostly kept in the first eigenvectors. This can be interpreted as some Zernike coefficients (and their appropriate combinations) have a more powerful weight for discrimination of corneal topography. For instance,  $C_0^2$  is discriminative (see Figure 3) because the red bar is much higher than the green bar (95% difference), this corresponds to a much higher (absolute) value in the vector  $v_1$ , conversely  $C_0^0$  is not discriminative (9% difference) and the corresponding values in much smaller as expected. It is interesting to notice that Lewis [L11] removed this coefficient (with 3 others) in his analysis due to its high variance.

Table 1: Zernike coefficients and the results of LDA for one individual ( $x=A_1$  and  $A_2$ )

	Zernike Coefficients		LDA Results	
	$A_1$	$A_2$	$\lambda_i \times 10^3$	$v_1$
$C_0^0$	2.1833	2.3125	1.8856	<b>-0.0005</b>
$C_{-1}^1$	-0.0106	-0.0136	0.1438	-0.0309
$C_1^1$	0.0023	0.0064	0.0718	-0.2686
$C_{-2}^2$	-0.0012	0.0007	0.0505	0.1110
$C_0^2$	-0.8383	-0.8389	0.0424	<b>-0.4379</b>
$C_2^2$	0.0017	-0.0027	0.0335	-0.0406
$C_{-3}^3$	-0.0024	-0.0004	0.0289	-0.0984
$C_{-1}^3$	-0.0021	-0.0044	0.0190	0.0911
$C_1^3$	0.0031	0.0034	0.0150	0.5013
$C_3^3$	-0.0045	-0.0015	0.0131	0.0290
...	...	...	...	...

#### 4. Experimental result

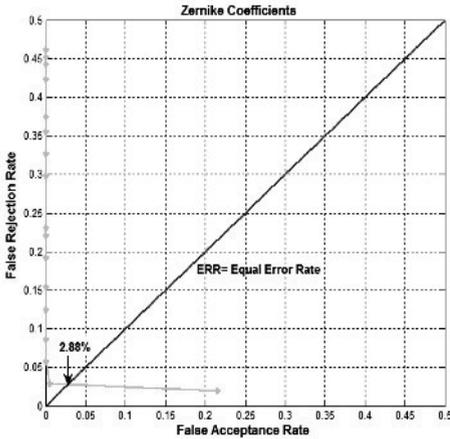
In order to analyze the performance of the proposed corneal biometric algorithm, we compared the feature vectors  $x$  and  $x'$  by computing this mean distance:

$$D(x, x') = \frac{1}{N} \sum_{i=1}^N |x_i - x'_i| \quad (8)$$

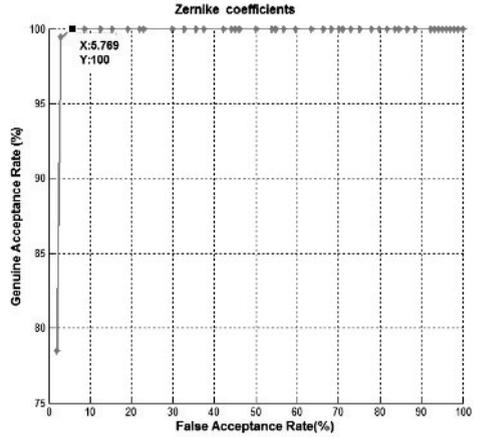
For the first experiment, we used the  $N=36$  Zernike coefficients as cornea matcher 1, then a combination of Zernike coefficients with LDA as cornea matcher 2. To evaluate the performance of the system, the Equal Error Rate (EER) criterion was employed. The system threshold value was obtained using the EER criteria when False Acceptance Rate (FAR) equals False Reject Rate (FRR). This was determined from the Receiver Operating Characteristic (ROC) Curve. The lower the EER, the better is the system performance. Another performance measure is the Genuine Acceptance Rate:  $GAR=1-FRR$ . The lowest FAR that yields a GAR of 100% was selected from the ROC curve. Table 2 shows these results for matcher1 and matcher 2. In the latter case, different numbers ( $r$ ) of features were tested, the best choice was  $r = 3$  features and is used in the following. Fig. 4 and Fig. 5 show the ROC curves for the two matchers. With all Zernike coefficients the ERR was 2.88% and the GAR was improved to 100% with a FAR of 5.77% (See Figure 4 (a) and Figure 4 (b)). These results are similar to those of Lewis [L11] who reported EER of less than 4 percents with a similar approach and another dataset. With LDA and  $r=3$  we achieved a 0.96% EER and a FAR of 0.96% (See Figure 5 (a) and Figure 5 (b)). This value 0.96% corresponds to only one false acceptance out of the 104 identification attempts. This confirms the efficiency of cornea as biometrics and LDA for the selection of the most representative features from the combination of Zernike coefficients.

Table 2 Results for the two cornea matchers

Methodology		FAR(%)	EER(%)
Zernike	$r$	5.77	2.88
Zernike + LDA	1	20.2	7.82
	2	4.8	4.5
	3	0.96	0.96
	4	1.92	0.96
	5	0.96	0.96

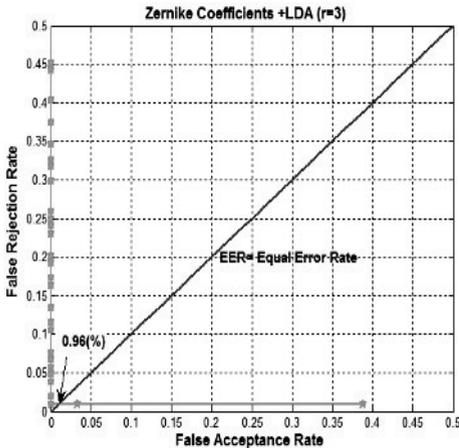


(a)

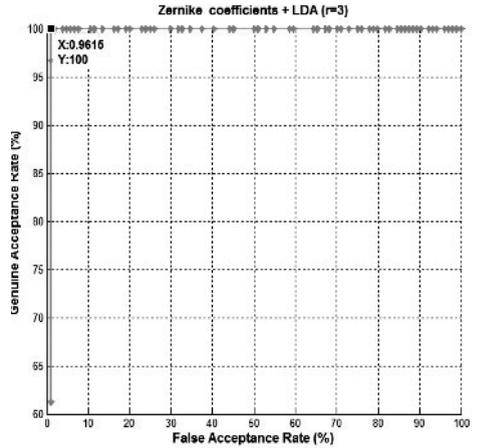


(b)

Figure: 4 ROC curves for all Zernike Coefficients ( $d = 36$ )



(a)



(b)

Fig 5 ROC curves for Zernike Coefficients with LDA ( $r = 3$ )

## 5. Conclusion and future work

The objective of this work was to investigate *corneal topography* as an accurate biometric modality using shape discriminating features. Our proposed method by using LDA, convert a  $d$ -dimensional Zernike feature vector ( $d=36$ ) into a smaller  $r$ -dimensional feature vector ( $r=3$ ) allowing to keep the relevant 88% of information of the initial feature vector. The results obtained (EER less than 1%) confirm that corneal topography could be an effective biometric method. Moreover, we expect that the fusion of corneal features with other biometric modalities could achieve higher performance. In the future we plan to study : (1) other corneal shape descriptors such as curvature, (2) include the posterior surface in the biometric assessment, (3) realize a new biometric database with more within-class comparisons and (4) test other topographers such as the Pentacam (Oculus).

## References

- [CDJ05] Chen Y., Dass S. C., and Jain. A. K. "Fingerprint Quality Indices for Predicting Authentication Performance". In Fifth AVBPA, pages 160–170, Rye Brook, July 2005.
- [LBS11] R. P. Lemes, O. R. P. Bellon, L. Silva, and Anil K. Jain, "Biometric Recognition of Newborns: Identification using Palmprints", IJCB, Washington, DC, Oct. 11-13, 2011.
- [D03] J. Daugman, "The importance of being random: statistical principles of iris recognition", *Patterns Recognition*, 36 (2003), 279-291.
- [ZK11] Y. Zhou and A. Kumar, "Human identification using palm-vein images", *IEEE Trans. Information Forensics & Security*, vol. 6, pp. 1159-1274, Dec. 2011.
- [DSV06] M. Dubbelman, V.A.D.P. Sicam, G.L. Van der Heijde. "The shape of the anterior and posterior surface of the aging human cornea", *Vision Research* 46 (2006) 993–1001.
- [BCI01] T. Buehren, M. J. Collins, D. R. Iskander, B. Davis, and B. Lingelbach, "The stability of corneal topography in the post-blink interval", *Cornea*, vol. 20, no. 8, 2001.
- [RNJ06] A. Ross, K. Nandakumar and A.K. Jain, "Handbook of Multibiometrics", Springer Verlag, page19,2006.
- [JGJ95] John E. Greivenkamp and Joseph M. Miller, "Representation of videokeratoscopic height data with Zernike polynomials", *J. Opt. Soc. Am. A*/Vol. 12, No. 10/October 1995.
- [W92] R. H. Webb, "Zernike polynomial description of ophthalmic surfaces", in *Ophthalmic and Visual Optics*, Vol. 3 of 1992 OSA Technical Digest Series (Optical Society of America, Washington, D.C., 1992), pp. 38–41.
- [L11] N. D. Lewis, "Corneal topography measurements for biometric applications", Ph.D. dissertation, University of Arizona, 2011.
- [SML10] F. Song, D. Mei, and H. Li, "Feature selection based on linear discriminant analysis", in *Proc. 2010 Int. Conf. Intelligent System Design and Engineering Application*, Changsha, China, Nov./Dec. 2010, vol. 1, pp. 746-749.
- [NT10] M. H. Nguyen, F. d. Torre. "Optimal feature selection for support vector machines", *Pattern Recognition*, 43 (3): March 2010.
- [G11] D. Gatinel, "Topographie cornéenne", 2011, Elsevier Masson SAS. ISBN:978-2-294-71134-3
- [KK13] A. Kumar and C. Kwong, "Towards contactless, low-cost and accurate 3D fingerprint identification", *Proc. CVPR 2013*, pp. 3438-3443 June 2013.
- [KKZ11] V. Kanhangad, A. Kumar and D. Zhang, "A unified frame work for contactless hand verification", *IEEE Trans. Info. Security Forensics*, vol. 20, pp. 1415-1424, May 2011.

# EEG Biometrics for User Recognition using Visually Evoked Potentials

Rig Das<sup>1</sup>, Emanuele Maiorana<sup>2</sup>, Daria La Rocca<sup>3</sup>, Patrizio Campisi<sup>4</sup>

Section of Applied Electronics, Department of Engineering  
Roma Tre University

Via Vito Volterra 62, 00146 Roma, Italy,

{rig.das<sup>1</sup>, emanuele.maiorana<sup>2</sup>, daria.larocca<sup>3</sup>, patrizio.campisi<sup>4</sup>}@uniroma3.it

**Abstract:** Electroencephalographic signals (EEG) have been long supposed to contain features characteristic of each individual, yet a substantial interest for exploiting them as a potential biometrics for people recognition has only recently grown. The biggest advantages of EEG-based biometrics lie in its universality and security, while its major concerns are related to the acquisition protocol that can be inconvenient and time consuming. This paper investigates the use of EEG signals, elicited using visual stimuli, for the purpose of biometric recognition, and evaluates the performance obtained considering various frequency bands, different number of visual stimuli, and various subsets of time intervals after the stimuli presentation. An exhaustive set of experimental tests has been performed by employing EEG data of 50 different healthy subjects acquired in two different sessions, separated by one week time.

## 1 Introduction

Biometrics-based recognition is an active area of research which has brought to the deployment of automatic recognition systems using mainly fingerprints and face for real life application [JRN11]. In recent years, a growing interest emerged for alternative biometric identifiers like vein patterns, electrocardiographic (ECG) signals, electrodermal response, or electroencephalographic (EEG) signals, to cite a few. Specifically, brain signals, acquired through electroencephalography, have been investigated mainly in the medical arena. However, despite having large interest in medical applications, EEG signal's use as a biometric identifier is relatively new [CLR14]. The advantage of using EEG biometrics relies mainly in its security being brain signals not acquirable at a distance which makes difficult their synthetic replica [CLR14]. However, one disadvantage of using EEG signals for people recognition is the difficulty for setting up the subject for EEG acquisition and creating an ideal environment for it. The acquisition process in fact requires placing multiple electrodes over the subject's scalp. These electrodes sense the electrical field generated by the brain during resting states or while performing specific tasks, such as receiving audio-visual stimuli, performing imagined or real body movements, speech, etc. The resting state condition or protocol has been for instance considered in [RCS13], where the repeatability of discriminative characteristics of EEG signal over time has been partially addressed, which is essential for biometric recognition. Among the different brain responses that can be acquired as the result of a brain stimulation, in this paper we rely on

the visually evoked potentials (VEP), a kind of Event-Related potentials (ERPs) that refer to the electrical potential modification due to brief visual stimuli and recorded from the scalp over the visual cortex [GPP12]. Within this regard, the focus of this paper is the performance evaluation of EEG-based biometric verification based on VEP responses. Specifically, EEG data collected from 50 healthy subjects during two different sessions, acquired at time  $T_0$  and  $T_0 + 1$  week, are employed to test the effectiveness of VEP as a potential biometrics. This large and multiple-sessions database allows us having a significant number of comparisons for a stable and practical result. It is in fact worth remarking that, although several techniques have been recently proposed for EEG-based biometric recognition, most of them have been tested either on small databases with more than a single session, or on datasets comprising recordings from a single session for performance evaluation. In more detail, this paper evaluates the performance achievable when considering various EEG frequency bands. The performed experiments investigate which is the best performing subband among different combinations of different bands in  $[0.5; 14] Hz$ , the most relevant for our analysis [Bas99, CLR14, RCS13]. Moreover, experiments are performed for finding the minimum number of visual stimuli that are required for generating a single ERP, and the best time interval after producing the visual stimuli that can be considered as a EEG signal latency.

## 2 Related Work on VEP Based Biometric Recognition

In this section some of the earlier works which had evaluated VEP as a potential biometrics are briefly reviewed (Table 1). In [Tou09] Touyama has investigated the possibility of person identification by extracting the P300 evoked potentials from the generated EEG signal, during a target and non-target photo retrieval task. The authors have used Principal Component Analysis (PCA) on the time sequences along with Linear Discriminant Analysis (LDA) for classification, and examined the identification performance. Five male subjects have been considered, with every subject's EEG response acquired for five sessions on a same day upon producing target and non-target stimuli. Each session contained 20 trials with 9 images. For performance evaluation the  $0.5 - 30$  Hz sub-band has been considered, while only the  $C_z$  Channel (according to 10-20 international standard) has been used. A leave-one-out approach has been employed by mixing the EEG signals from different sessions for training purposes. Both the target and non-target stimuli are considered together and a performance accuracy of 97.6% achieved. In [GPP12] Gupta et al. have considered EEG signals recorded as responses to three variations of the oddball paradigm: standard oddball, spatially varying oddball and Rapid Serial Visual Paradigm (RSVP), which is nothing but the stimuli on the same spatial position which minimizes the influence of irrelevant stimuli. Eight subjects (4 males and 4 females) have been employed for testing purposes, with acquisitions from a single sessions. Authors achieved a maximum Correct Recognition Rate (CRR) of about 97% when exploiting the RSVP paradigm, with the  $1 - 12$  Hz bandpass frequency. In [YSL13] Yeom et al. have evaluated the differences of the averaged EEG signals generated in response to self-face and non-self-face images. Tests have been performed over 10 different subjects with signals captured in two sessions on different days, where each session included two runs, and each run further composed of 50 trials. For each trial, a total of 20 face images were presented (10 of self-face and

Table 1: Overview of state-of-the-art contributions using VEP from EEG signals as a biometrics.

Paper	DB	Ch.s	Features	Classifier	Performance	Sessions
[DZGE09]	20	20	LDA	KNN	CRR=94%	1
[PM07]	102	61	MUSIC spectrogram	Elman NN	GAR=98.12%	1
[Pal04]	20	61	spectral power ratio	BP NN	CRR=99.15%	1
[GPP12]	8	8	P300	LDA	CRR=97%	1
[Tou09]	5	1 (Cz)	PCA	LDA	CRR=97.6%	5, same day
[YSL13]	10	8	Adapt. discriminative feat.	Non-Linear SVM	CRR=86.1%	2, diff. days

10 of non-self-face images), i.e. 100 trial for each session and 200 trials overall. Cross-validations are performed with random selection of 180 trials for training and remaining 20 trials for test. The proposed method has achieved an overall CRR of about 86.1% with both false acceptance rate (FAR) and false rejection rate (FRR) of 13.9%.

In [DZGE09] Das et al. have used VEP data for person identification. 20 different subject's EEG signal have been collected, using a visual perceptual task in which filtered noise was added with the visual stimuli. Face and car images have been used as a visual stimuli and each of them appeared for 40ms, after that the subjects had to identify whether a stimuli is car or face. They have also shown that a period of 120 – 200ms after the stimulus is the most informative with respect to the individual discrimination. Authors have obtained a classification accuracy around 75% to 94% for the best performing post-stimulus set. This shows that the VEP signals are crucial for person recognition. In [Pal04] R. Palaniappan has investigated VEP data recorded from 20 subjects by producing a single stimuli, which consists of pictures of common objects represented by black and white line. A classification accuracy of 99.6% has been achieved by ANOVA tests on each of the 61 channels. Also in [PM07] Palaniappan et al. used similar protocol for 300ms VEP stimuli to collect the EEG signals. A total of 102 subjects were used for collection of EEG data from 61 channels for a total of 3560 VEP and the acquired signals were filtered through a 25 – 56 Hz bandpass filter to retain the  $\gamma$  waves. The authors have also used CAR filter for reduction of intra-class variance and they have achieved 98.12% of accuracy using all channels. As already remarked in Section 1, most of already proposed works have evaluated recognition performance achievable with EEG signals when using acquisitions from a single session, or from multiple session yet while considering only few subjects. The present paper is the first one properly investigating the discriminative capabilities of VEP responses, by comparing the signals captured from a large number of users during distinct recording sessions, as described in Section 4.

### 3 Proposed Biometric Recognition System using VEP

#### 3.1 Employed VEP Acquisition Protocol

In this framework we use eight different geometric shapes, as described in Figure 1(c), to generate VEPs. Among these shapes, the circle is considered as target stimulus, with every subject asked to concentrate on it as it appears in the screen. An ERP response is therefore expected to be noted when the target shape appears on the display. All the other

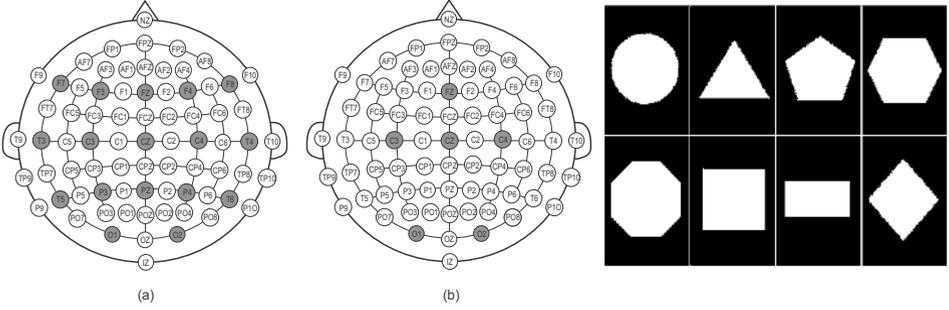


Figure 1: Selected electrode positions. (a) 17 Selected electrodes or Channels, (b) 6 Selected electrodes or Channels  $[F_z, C_3, C_z, C_4, O_1, O_2]$ , (c) Shapes employed for the Geometric Protocol.

shapes are considered as a non-target stimulus. Each of these shapes is shown for 250 ms, and repeated for 60 times. The EEG data are acquired from 19 different electrodes that are positioned on brain scalp according to the 10-20 international standards. For experimental purposes we consider both kind of responses, generated after the presentation of target and non-target shapes. Specifically, for verification purpose each and every users' acquired EEG signals are compared using the following three schemes, these are: **[Target vs. Target]**: where the recorded EEG signal has been generated due to the target stimulus; **[Non-Target vs. Non-Target]**: EEG signal that has been generated due to the non-target stimulus; **[(Target - Non-Target) vs. (Target - Non-Target)]**: generated by subtracting the EEG signal generated by target and non-target stimuli

### 3.2 EEG Data Analysis

A preprocessing step is first carried out on the recorded EEG data to increase their signal-to-noise ratio. Specifically, for each subject a Common Average Referencing (CAR) filter is used to calculate the mean of all channels, and subtract this value from all the output channels [CLR14]. The CAR filtered data are then down-sampled up to 128 Hz from existing 256 Hz using a proper anti-aliasing bandpass filter, and then spectral filtered to 0.5 – 40 Hz to retain all the relevant information which are required for feature extraction. The data are then normalized using Z-score transformation to have a zero mean and unitary standard deviation. Finally the data is detrended, by subtracting the mean or a best fit line from the data. After the described preprocessing, performed in all the considered experimental tests, a specific subband is isolated for extracting discriminative information from the available signals. As discussed in Section 4, we in fact perform an analysis of the best performing EEG subband or subband combinations in the range  $[0.5; 14]$  Hz, conducted when considering the [Target vs. Target] scenario. We also analyze which is the best performing time interval after the stimulus presentation, with a maximum interval of 0 – 700ms. Specifically, all the possible time intervals  $\Delta_T = [t_b; t_e]$ , where  $t_b$  and  $t_e$  values can vary in between  $\{0; 100; 200; 300; 400; 500; 600; 700\}$  ms, can be considered for this aim. As reported in Section 4, also for this analysis a large set of tests are performed for user verification by considering [Target vs. Target] scenario. Once the recorded signals have been filtered, and a specific time interval after the presentation of the stimuli has been determined, the EEG signal corresponding to the target and non-target stimulus

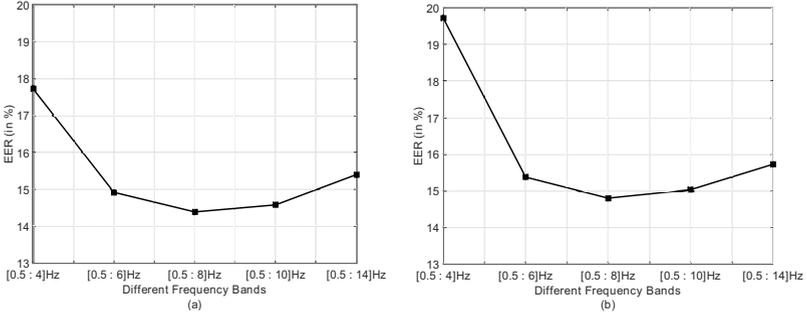


Figure 2: EER vs frequency range for (a) 17 and (b) 6 channels fusion

can be processed in order to extract a VEP waveform from them. This task is performed by averaging the responses taken from  $N$  events. Since we have 60 events available for each considered shape, as described in Section 4 we select  $N = \{20; 30; 40; 45; 50\}$  in the performed tests, to investigate the performance dependence on the number of considered events. The comparison between two VEP waveforms, extracted from two different recording sessions, is performed by resorting to the cosine or to the Euclidean distance. Specifically, the responses extracted from each channel are first compared between them, and the  $M$  computed distances are then fused into a single score by taking their average as the output of the matching process. In more detail, in our experiments we consider the responses collected from  $M = 17$  channels, selected from the available 19 ones by excluding the two frontal ones, i.e.  $F_{p1}$  and  $F_{p2}$ , since VEP responses are known to be mostly present in central and occipital regions. In order to minimize the number of employed electrodes, thus reducing the user inconvenience, we also consider a configuration using  $M = 6$  channels:  $[F_z, C_3, C_z, C_4, O_1, O_2]$ . This latter configuration is in fact often employed for BCI applications based on VEP protocols [WW12]. Both the employed montages are shown in Figure 1(a),(b).

## 4 Experimental Results and Discussion

As already remarked, the experimental validation is performed over a database of EEG signals collected from 50 healthy subjects during 2 different sessions, acquired at one week distance each other. We first focus on the Target vs. Target scenario, where the VEPs generated in correspondence of the selected trigger are employed for biometric recognition. Within this scenario, we evaluate the best performing frequency range, time interval and number of observed events in the following subsection. In more detail, all these evaluations are performed through a cross-validation test by selecting, for 10 different times, a total of  $N$  events from the EEG signals during the first recording session of each user, to build the associated enrollment dataset. For each time, intra-class comparisons are performed by randomly selecting, for 10 different times,  $N$  events from the second recording session of the tested user as authentication probes. The dimensionality of the number of intra-class comparisons considered for performance evaluation is therefore  $[50 \times 10 \times 10]$ . Inter-class comparisons are instead obtained by randomly selecting, for each enrolled user, other 20 subject acting as impostors, and selecting  $N$  random events from their recordings

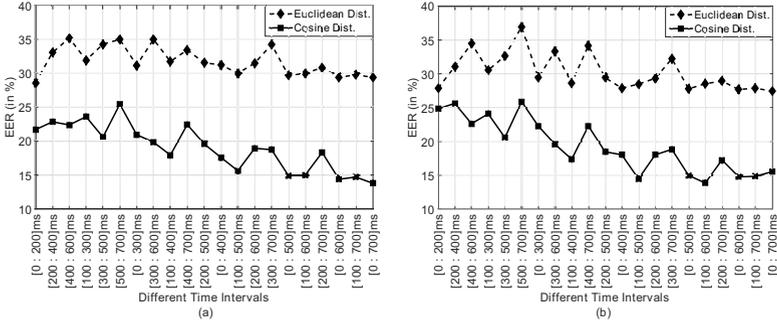


Figure 3: EER vs time intervals  $\Delta_T$  (after visual stimuli) for (a) 17 and (b) 6 channels fusion

from session-2. The dimensionality of the number of inter-class comparisons is therefore  $[50 \times 10 \times 20]$ . This way, we can evaluate the performance in terms of FRR, FAR and equal error rate (EER) through a very large sets of genuine and impostors comparison scores.

#### 4.1 Results of Frequency Range Selection

For the selection of the best performing frequency range we have set the  $\Delta_T = [0; 600]$ ms time interval, and evaluated the performance of frequency ranges in between  $[0.5; 4]Hz$ ,  $[0.5; 6]Hz$ ,  $[0.5; 8]Hz$ ,  $[0.5; 10]Hz$ ,  $[0.5; 14]Hz$  when considering the cosine distance for generating matching scores. The performance evaluation tests are done for both the  $M = 17$  channels and the  $M = 6$  channels configuration, as detailed in Section 3.2.  $N = 45$  events are considered for generating ERP responses from the considered target stimuli. Figure 2 shows the performance of all the above mentioned subbands, and it can be clearly seen that  $[0.5; 8]Hz$  sub-band is the better performing one for both the 17-channel and the 6-channel combinations. Similar experiments are also performed for other frequency ranges such as  $[4; 8]Hz$ ,  $[4; 14]Hz$  and  $[8; 14]Hz$ . Nonetheless, the EER increases substantially for all these latter cases.

#### 4.2 Results of Time Interval Selection

Given the  $[0.5; 8]Hz$  frequency band, several time intervals  $\Delta_T$ , starting at the visual stimuli time presentation, are then considered. Figure 3 shows the performance obtained for various  $\Delta_T$  values, for both the 17-channel and the 6-channel configurations and considering VEP responses obtained with  $N = 45$  events. We omit the results obtained for time intervals lasting 100ms from Figure 3, since in these cases the EER is always greater than 25%. To find the best performing distance metric, we also consider the Euclidean (L2) and manhattan (L1) distances along with the cosine one, finding out that the L1 distance performs almost as similar as L2 distance, whose associated performance is reported in Figure 3. It can be clearly seen that matching performed using the cosine distance outperforms the Euclidean distance calculation as cosine distance measures the similarity of vectors by considering the direction of the signal with respect to the origin while Euclidean distance measures the distance between particular points of interest along the vector; e.g. magnitude. So, by considering the entire EEG signal and its direction produces better result than only considering the magnitudes at some particular points; as the magnitude may be same for different EEG signals coming from different sources. Better results are

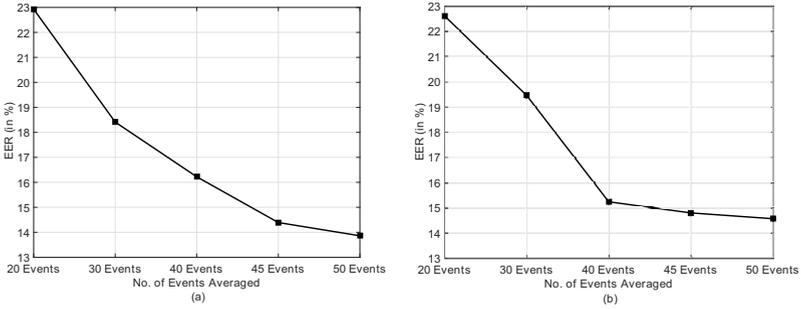


Figure 4: EER vs different no of events  $\Delta_N$  for (a) 17 and (b) 6 channels fusion

typically obtained when considering a large time interval for the considered VEPs, not focusing only on the temporal interval characteristic of ERP responses.

### 4.3 Results of Required Minimum Number of Events

Having set the  $[0.5; 8] Hz$  frequency range and the  $\Delta_T = [0; 600] ms$  time interval, we also evaluate the performance dependency on the number of events considered for generating VEP responses in Figure 4. The reported results are obtained having set the  $[0.5; 8] Hz$  frequency range and the  $\Delta_T = [0; 600] ms$  time interval, while changing the number of events  $N$  as explained in Subsection 3.2. It can be clearly seen that the EER is inversely proportional to the number of events, and VEPs generated by averaging  $N = 50$  events produce the lowest EER.

### 4.4 Performance Evaluation of Proposed Recognition System

By considering  $[0.5; 8] Hz$  frequency range,  $\Delta_T = [0; 600] ms$  time interval and  $\Delta_N = 50$  events as selected system parameter, further experiments are carried out to compare the performance achievable when considering the [Target vs. Target], [Non-Target vs. Non-Target] and [Target – Non-Target vs. Target – Non-Target] scenarios. Specifically, in order to present results with a high statistically significance, in this case we carry out a cross-validation process by selecting, for 20 different runs, 40 subjects out of the available 50 for estimating the achievable recognition performance. The rates reported in the following are obtained as the average of the results obtained during each run. Specifically, for each run we evaluate intra- and inter-class scores as described at the beginning of this section, performing several divisions of the data in sessions 1 and 2 to respectively generate the enrollment and testing datasets. Therefore, now the dimensionality of the number of intra-class comparisons is now  $[20 \times 40 \times 10 \times 10]$ , while the number of inter-class comparisons employed to evaluate the FAR is now  $[20 \times 40 \times 10 \times 30]$ . Table 2 shows the calculated EER after performing the above mentioned comprehensive testing. It can be seen that by using either using the 17-channel or the 6-channel configuration, EER of around 14.5% can be achieved for the [Target vs. Target] scenario. However, for [Non-Target vs. Non-Target], the considered configurations achieve around 14% and 13% EER respectively. Worse results are obtained when performing comparisons between waveforms obtained as difference between the target and non-target responses. Therefore, it can be concluded from the above results that the [Non-Target vs. Non-Target] scheme is

Table 2: Performance Evaluation of Target and Non-Target stimuli by EER calculation.

Testing Schemes	EER (in %) for 17 Ch. Fusion	EER (in %) for 6 Ch. Fusion
[Target vs Target]	14.82	14.45
[Non-Target vs Non-Target]	13.55	14.01
[(Target – Non-Target) vs (Target – Non-Target)]	23.50	19.66

more stable than the other two, for both the considered channel configurations. This result leads us to consider [Non-Target vs. Non-Target] scheme and the 6-channel scheme for further research on achieving performance stability for EEG-based biometrics recognition.

## 5 Conclusions

In this paper we have investigated about the use of EEG for the purpose of automatic people recognition. Specifically, visually evoked potentials have been employed in our approach. An extensive dataset of 50 healthy people acquired in two sessions one week time apart has been employed. Different frequency subbands, time intervals and channel configurations have been tested. In summary our analysis can be considered as a preliminary step towards the assumption that, EEG signals generated as a result of VEP, are stable enough for its consideration as a biometric identifier.

## References

- [Bas99] Erol Basar. *Brain Function and Oscillations: Integrative Brain Function. Neurophysiology and Cognitive Processes*. Springer Series in Synergetics. Berlin: Springer, 1999.
- [CLR14] P. Campisi and D. La Rocca. Brain waves for automatic biometric-based user recognition. *IEEE TIFS*, 9(5):782–800, May 2014.
- [DZGE09] K. Das, S. Zhang, B. Giesbrecht, and M.P. Eckstein. Using rapid visually evoked EEG activity for person identification. In *IEEE EMBC*, pages 2490–2493, Sept 2009.
- [GPP12] C.N. Gupta, R. Palaniappan, and R. Paramesran. Exploiting the P300 paradigm for cognitive biometrics. *Int. Journal of Cognitive Biometrics*, 1(1):26–38, May 2012.
- [JRN11] A. Jain, A. Ross, and K. Nandakumar. *Introduction to Biometrics*. Springer US, 2011.
- [Pal04] R. Palaniappan. Method of identifying individuals using VEP signals and neural network. *Science, Measurement and Technology, IEE Proceedings*, 151(1):16–20, 2004.
- [PM07] R. Palaniappan and D.P. Mandic. Biometrics from Brain Electrical Activity: A Machine Learning Approach. *IEEE TPAMI*, 29(4):738–742, April 2007.
- [RCS13] D. La Rocca, P. Campisi, and G. Scarano. On the Repeatability of EEG Features in a Biometric Recognition Framework using a Resting State Protocol. In *IEEE BIOSIGNALS*, '13.
- [Tou09] H. Touyama. EEG-Based Personal Identification. *Biomedical Engineering, InTech Journal on*, (22):415–424, October 2009.
- [WW12] J.R. Wolpaw and E.W. Wolpaw. *Brain-Computer Interfaces: Principles and Practice*. Oxford University Press, NY, USA, 2012.
- [YSL13] S.K. Yeom, H.II Suk, and S.W. Lee. Person Authentication from Neural Activity of Face-specific Visual Self-representation. *Pattern Recogn.*, 46(4):1159–1169, April '13.

# Liveness Detection in Biometrics

Maximilian Krieg and Nils Rogmann

Hochschule Darmstadt – Fachbereich Informatik  
Schöffnerstr. 8b, 64295 Darmstadt, Germany  
{Maximilian.Krieg; Nils.Rogmann}@stud.h-da.de

**Abstract:** The use of biometrics as an alternative for PIN and password-based authentication systems becomes increasingly attractive in this day and age. Biometric systems perform an authentication by personal biological or behavioral characteristics. A negative side effect of the increasing use of biometric systems is the progressing development of sophisticated attacks, in particular presentation attacks. Liveness detection has the aim to identify a living and during the biometric authentication process present individual as such and to repel spoofing attacks at the data capture subsystem. In this paper different current attack scenarios are described. Based on these scenarios, several liveness detection techniques are elaborated and investigated as possible countermeasures.

**Keywords:** Liveness Detection, Presentation Attacks, Spoofing, Presentation Attack Detection

## 1 Introduction

The use of biometric systems has experienced an enormous growth of interest in recent years. While the recognition of fingerprints has been strongly influenced by the forensic application it becomes more and more accepted in occupational and personal life. The stigmatization as a criminological tool is still present, but is increasingly losing its significance [Ev15], [KS13]. Essential factors for the increasing proliferation are technical advancements which provide convenience and mobility as well as affordable sensors for the user. The increasing digitization has led to a rapid growth of systems that are protected primarily through passwords against unauthorized access. A password-based authentication is flawed with a variety of problems, because passwords can be guessed, forgotten or passed. Biometrics remedies these problems by complementing authentication as a second factor or even substitute passwords completely [Wi15].

The research and advisory company Gartner Inc. predicts that 30 percent of organizations will use biometric authentication for their mobile devices by 2016. This represents a growth of roughly 25 percent within two years [Ga14]. To give an example, the introduction of the electronic passport (ePass) in 2005 is an indicator of the growing importance of biometric systems at national and international level [Ev15], [BI15].

The probability of a successful attack should not be underestimated. Biometric samples such as latent fingerprints are left unconsciously and often unavoidable in many places in everyday life. A latent fingerprint can be collected, for instance, from the surface of a drinking glass or the touch-screen of a smartphone, artificially copied and used for the unauthorized authentication in a biometric system [KS13]. These replicas are referred to

as artifacts. Attacks which rely on the presentation of such artifacts are called presentation attacks and are the main focus of this paper. The unauthorized possession of a high-quality biometric sample poses a risk to the entire system and the user, since the biometric characteristics as opposed to a password cannot be changed. They are persistent and keep their validity typically for a lifetime [Wi15]. Thus, the responsibility lies with the biometric system and its ability to detect presentation attacks.

In addition, the increased use of biometric recognition increases the likelihood of a compromise. A potential risk of inadequately protected data storage subsystems is the theft of stored biometric references. The responsible use and therefore the decision for which system a biometric recognition is required and even suited has to be performed individually [Wi15]. The application of biometric template protection-techniques at this point ensures that a stolen biometric reference does not lead to the uselessness of the complete source. The protection of biometric references is however not part of liveness detection and will therefore not be addressed in this paper.

In this paper current and practice-oriented attack scenarios are investigated and possible countermeasures are worked out. The emphasis is on the assessment of facial, fingerprint and vein recognition. At last, based on the presented knowledge, an overall assessment of recent attack scenarios against biometric systems is conducted.

## 2 Attack Scenarios for Presentation Attacks

A recent example of the threat posed by these attacks was presented at the 31<sup>st</sup> Chaos Communication Congress (31C3) in December 2014. Using a digital camera, thumb photo of federal minister of defense Ursula of the Leyen was taken out of a distance of approximately three meters and was processed digitally. A biometric sample can already be generated from a single image (see. Fig. 1). In the next step the biometric features can be extracted from the processed sample [CC14].



Fig. 1: Reconstruction of a fingerprint through photography [CC14].

Furthermore, it was shown how a latent fingerprint can be taken from a smartphone touch-screen to produce an artifact with capacitive properties. For this the touchscreen was scanned with a high-resolution scanner (2400 DPI). The image was preprocessed digitally to print it on a transparent foil. This was used to expose a printed circuit board (PCB), which later serves as a template for the artifact. The PCB was sprayed with graphite spray and coated with wooden glue. An artifact produced in this way can be used to unlock the smartphone. The middle illustration in figure 2 shows how such an artifact may look like [CC14].

Beside fingerprint sensors, other systems are also affected by spoofing attacks. For instance, facial recognition systems can be deceived by printed facial images. This kind

of presentation attack, however, can be identified by many systems today. Mask and video attacks are considered more modern and effective approaches. A video attack can relatively easily be manufactured. For this purpose, a high-resolution video of a person is recorded and presented to the sensor in the next step using a tablet computer. Other sources for such recordings already exist through the internet and television, if it concerns a person of the public life. Corresponding videos can be obtained from sites like YouTube. The generation of 3D masks (see. Fig. 2) is considered more complex. To begin with, at least two images of the head are required, one in frontal and one in profile view. The masks must be shaped and modeled in the next step. The peculiarity of these attacks lies in the ability to collect biometric characteristics at a distance and without direct interaction [EM13], [An14].

Another technique is vein recognition which is considered comparatively reliable. This paper focuses only on finger vein recognition. An essential advantage of this technique is that the biometric characteristic cannot be collected from objects, like with fingerprints, and requires an additional illumination in near infrared area to make the veins visible for a sensor. If an attacker obtains the vein image of a biometric sample, the possibility for the production of an artifact exists. The image is preprocessed using histogram equalization for contrast enhancement and a Gaussian filter for noise reduction. In addition, it is scaled on a size corresponding to the finger, is bordered with black pixels and mirrored to reverse the internal reflection of the sensor. The resulting image is printed on high-quality paper, contoured with a marker if necessary and finally presented to the sensor. This attack achieved a false acceptance rate between 76 and 86 percent. Figure 2 shows on the right side how the sensor “sees” the printed artifact [To14].

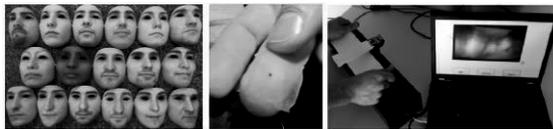


Fig. 2: Examples of Presentation Attack Instruments according to [EM13], [To14], [CC14].

### 3 Liveness Detection

The detection of presentation attacks can be accomplished by many different *presentation attack detection* (PAD) methods and techniques. Basically, capturing these attacks either takes place through a whole system-monitoring approach or through additional features being implemented into the data capture subsystem that is integrated into a biometric system [Am15]. One of the most frequently used PAD techniques is called *liveness detection*. The relevant aspects of this technique will be preliminarily elaborated and explained within this chapter. Subsequently, the obtained knowledge will be used to acquire and compare several different liveness detection techniques to recognize the previously described attack scenarios against biometric systems.

The general task of liveness detection is to detect whether a biometric probe (e.g. a fingerprint) belongs to a living subject that is present at the point of biometric capture [Am15]. Using liveness detection techniques, a reliable recognition of dead fingers or

photographed faces can be established, for example. Consequently, the risk of successful presentation attacks is significantly reduced. Thus, in addition to the regular biometric recognition, liveness detection is an important procedure aiming at an increased reliability of biometric systems.

In the global market, several different methods verifying the liveness of biometric features are already established. Among others these methods include an evaluation of anatomical characteristics, physiological processes of the human body and involuntary reactions to stimuli as well as various predictable human behaviors [Am15].

### **3.1 Hardware and Software-based Approaches**

Typically, liveness detection methods are divided into hardware and software-based approaches. Giving an example, special medical hardware can be used to perform an electrocardiogram or pulse oximetry to detect living subjects. For this purpose, the acquisition of additional sensors such as devices for measuring body temperature or pulse rate ([LJ09b], pp.924-925) that needs to be combined with the regular biometric test system is required. As a consequence, account should be taken of additional costs for acquisition as well as for routine maintenance.

Software-based techniques, by contrast, make use of biometric data already being captured for biometric recognition of individuals. In general, these solutions are implemented as supplementary algorithms being integrated into a consisting biometric system. To give an example, these algorithms are then applied to the extracted biometric fingerprint probe in order to detect the deformation of a living finger that is pressed on the sensor ([LJ09b], p. 925).

### **3.2 Passive and active Techniques**

Besides the distinction of hardware and software-based techniques another common attempt consists of a separation between passive (non-stimulating) and active (stimulating) automated liveness detection methods [Am15]. In general, passive detection techniques make use of biometric probes which were recorded through a biometric sensor. According to this, further interactions with the data subject are not necessary. For this, a typical example would be a temperature or pulse measurement taking place while the biometric probe is collected [MA14].

Active detection techniques normally require additional interaction of the biometric data subject with the biometric system. These further interactions should be requested using challenge response procedures. The different challenge response approaches can be read in *ISO/IEC DIS 30107-1* as they cannot be discussed in this paper [Am15].

### **3.3 Defense against Presentation Attacks**

There are several different liveness detection techniques that have already been evaluated on the market and successfully used for presentation attack detection. Various

techniques that behave as possible countermeasures against previously described attack scenarios will be elaborated first. Based on the main results of this paper, an overall evaluation of these scenarios will be carried out in chapter 3.4 subsequently. The table below contains liveness detection techniques that could be used as counter-measures for detecting various presentation attacks:

<b>Biometric Sensor</b>	<b>Presentation Attack</b>	<b>Liveness Detection Technique</b>	<b>Remarks</b>
Fingerprint scanner	2D print, dead finger, artificial finger, capacitive finger	<b>Passive:</b> pulse measurement* [Am15], temperature measurement** [MA14], sweat detection [Am15], skin resistance detection ** [KS13]  <b>Active</b> (challenge response): Request of different fingers in random order [Am15]	* not working against capacitive finger artifacts  ** depends on the consistency of the artificial finger
Vein scanner	2D print, dead finger, artificial finger,	<b>Passive:</b> pulse measurement, temperature measurement, (sweat detection), skin resistance detection  <b>Active</b> (challenge response): Request of different fingers in random order	
Face scanner	2D print, 3D face mask, video attack	<b>Passive:**</b> natural eye blinking * [Am15], natural muscle movements while speaking [MA14]  <b>Active</b> (challenge response): eye closing request [Am15], voice usage request** [MA14], head turning request** [Am15]	* Possibly not working against face masks  ** No protection against video attacks
Fingerprint, vein & face scanner	2D print, 3D face mask, dead body parts, artificial or digital fingers, veins & faces	<b>Passive:</b> Infrared & ultraviolet light, thermal scans* [MA14], medical techniques like ECG, pulse oximetry or blood pressure reading [KS13]	* Possible in combination with all optical sensors

Tab. 1: Liveness detection techniques against presentation attacks.

As shown in table 1, there are several techniques that can be used against different presentation attacks. However, an individual application of one method such as a pulse

or temperature measurement might get tricked with an acceptable effort and without specific knowledge [MA14]. Thus, a combination of multiple liveness detection techniques is reasonable as the security of the whole biometric system will be increased.

### 3.4 Evaluation of Presentation Attacks

The newly acquired knowledge about liveness detection techniques can be used for a final evaluation of the previously described presentation attack scenarios. For this purpose, first the most important findings will be shown in table 2 and described in detail afterwards. Here it must be considered that the term paper does not focus an assessment of the effectiveness of different PAD techniques. Consequently, this evaluation just gives estimation about the efficiency of these techniques.

Attack scenario	Production cost	Technical complexity	Availability and advantage of counter-measures
High resolution 2D fingerprint image	Low	Low	High
Artificial capacitive finger	Medium	Medium	High
2D face image	Low	Low	High
Video attack	Low	Low	Low
3D face mask	High	High	Medium
2D finger vein image	High	High	Medium to high

Tab. 2: Comparative summary of attack scenarios.

Table 2 contains different attack scenarios and discusses them in terms of production cost as well as technical complexity. Moreover, availabilities and advantages of counter-measures are considered as they provide an important factor in regard to the general risk of a specific presentation attack. Based on these features, each attack scenario will be discussed in detail:

1. The technical capabilities to extract a fingerprint from a high-resolution 2D image have been increased consistently in the past years. Nowadays, digital processing and post-processing no longer present difficult technical challenges while creating a fake print. Finally, several methods reduce the possibility of a successful presentation attack execution, e.g. pulse and temperature measurement or sweat detection as well as several active liveness detection techniques mentioned in table 1.
2. In general, the production of artificial capacitive finger artifacts requires a higher technical knowledge as well as an access to a high-resolution scanner. Furthermore, the usage of optical sensors would reduce the risk of these attacks completely, since only capacitive sensors are vulnerable. However, due to their cost-effective integration into smartphones and tablets, capacitive sensors are widespread. In this case, software-based liveness detection approaches such as sweat detection algorithms can be used to protect the authenticity of these devices [Ne14].
3. In order to create printed 2D face images, no specific technical knowledge or high amount of time is required. Nevertheless, due to their low complexity these attacks

are relatively easy and fast to detect. Giving an example, the detection of natural eye blinking of a living subject provides easy protection against this kind of attacks [MA14]. In addition, challenge response mechanisms such as requests for head rotation or eye closing offer conceivable protection. Taking these circumstances into account, this attack scenario is not dangerous for modern biometric systems.

4. In contrast, the detection of video attacks is more complex. In some cases the turning of a head or the closing of an eye might be recorded and used for a replay attack against biometric systems afterwards. As part of this attack the misuse of voice recordings is also conceivable. As a consequence, biometric face recognition systems should be protected with additional hardware such as infrared, ultraviolet or thermal scanners. The use of medical hardware (e.g. an electrocardiogram) would also be feasible. However, it must be mentioned that due to high costs resulting from these additional systems, some cost-benefit analyses should be done first.
5. The production of a 3D face mask requires multiple high-resolution images of a head as well as the knowledge and capability to create such a mask. Consequently, compared to the video attack scenario the technical and temporal complexity is much higher. However, challenge response mechanisms (e.g. a voice request and comparison) can provide protection against this scenario. Depending on the underlying material of the mask, infrared, ultraviolet or thermal scanners might also detect presentation attacks using 3D masks.
6. The complex procedure for creating spoofing finger vein images from real finger vein samples was already described in detail in the paper “On the Vulnerability of Finger Vein Recognition to Spoofing” [To14]. In general, this scenario represents a promising approach to attack finger vein scanners. However, since the vein pattern that is extracted by the biometric system is embodied inside the finger it cannot be captured or collected like a fingerprint. If an attacker still succeeded to make a high quality copy, liveness detection techniques such as pulse or temperature measurement could be used to protect the biometric system. As the biometric sensor implements a thermal scan to extract vein samples, pulse detection can be realized easily by capturing a video over a short period of time. In this case, no additional hardware would be necessary. Finally, the texture quality of vein images can be compared to fake images using *Fourier Transformation* or an extraction of *Binarized Statistical Image Features*. Since these techniques cannot be discussed in detail within this paper, further information relating this topic can be found in “The 1st Competition on Counter Measures to Finger Vein Spoofing Attacks” [To15].

## 4 Conclusion

Since the use of biometric systems for authentication purposes has experienced an enormous growth of interest, the amount and the complexity of attacks has increased dramatically, too. This particularly includes presentation attacks. However, the threat originated from these attacks can be reduced by using liveness detection techniques.

As shown in this paper, there are several different methods and techniques working

against current presentation attack scenarios efficiently. Here it must be mentioned that none of these techniques provide an entire protection to biometric systems. Especially, the detection of video attacks is a particular challenge (see chapter 3.4). As a consequence, a combination of different liveness detection techniques is strongly recommended. Moreover, as already mentioned in chapter 3, there are several other detection techniques that should be used for detecting presentation attacks and protecting against manipulations of biometric systems to increase the overall security [Am15].

Finally, the results of this term paper can be used for future research tasks regarding liveness detection. Here, an evaluation of the techniques performance and reliability would be of peculiar interest.

## References

- [Am15] American National Standards Institute: ISO/IEC DIS 30107-1 - Part 1: Framework, pp. 2-7, 2015.
- [An14] Anjos, André et.al: Handbook of Biometric Anti-Spoofing: Face Anti-spoofing: Visual Approach, pp. 65-82, Springer London, 2014.
- [BI15] Federal Ministry of the Interior, <http://www.bmi.bund.de/DE/Themen/Moderne-Verwaltung/Ausweise-Paesse/Reisepass/reisepass.html>, last visited: 20.05.2015.
- [CC14] Chaos Computer Club e.V., [https://media.ccc.de/browse/congress/2014/31c3\\_-\\_6450\\_-\\_de\\_-\\_saal\\_1\\_-\\_201412272030\\_-\\_ich\\_sehe\\_also\\_bin\\_ich\\_du\\_-\\_starbug.html#video](https://media.ccc.de/browse/congress/2014/31c3_-_6450_-_de_-_saal_1_-_201412272030_-_ich_sehe_also_bin_ich_du_-_starbug.html#video), last visited: 20.05.2015.
- [EM13] Erdogmus, Nesli; Marcel, Sebastien: Spoofing in 2D Face Recognition with 3D Masks and Anti-spoofing with Kinect, Idiap Research Institute, 2013.
- [Ev15] Evans, Nicholas et.al.: Guest Editorial Special Issue on Biometric Spoofing and Countermeasures. IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, VOL. 10, NO. 4, APRIL 2015, pp. 699-702, 2015.
- [Ga14] Gartner, Inc., <http://www.gartner.com/newsroom/id/2661115>, last visited: 22.05.2015.
- [KS13] Kalla, Christian; Schuch, Patrick: Sicherheit in der Fingerabdruck-Identifikation. Datenschutz und Datensicherheit - DuD Volume 37, Issue 6, 352-357, 2013.
- [LJ09b] Li, Stan Z.; Jain, Anil: Encyclopedia of Biometrics, pp. 883-952, Springer US, 2009.
- [MA14] Matthew, Peter; Anderson Mark: Novel Approaches to Developing Multimodal Biometric Systems with Autonomic Liveness Detection Characteristics, 2014.
- [Ne14] NextID Biometrics: Liveness Detection for the Mobile Biometrics Market, <http://nexidbiometrics.com/wp-content/uploads/2014/01/NexID-White-Paper-Mobile-Biometrics.pdf>, 2014, last visited: 22.05.2015.
- [To14] Tome, Pedro; Vanoni, Matthias; Marcel, Sebastien: On the Vulnerability of Finger Vein Recognition to Spoofing, Idiap Research Institute, 2014.
- [To15] Tome, Pedro et.al.: The 1st Competition on Counter Measures to Finger Vein Spoofing Attacks, [http://publications.idiap.ch/downloads/papers/2015/Tome\\_ICB-2015\\_2015.pdf](http://publications.idiap.ch/downloads/papers/2015/Tome_ICB-2015_2015.pdf), last visited: 05.06.2015
- [Wi15] Wikibooks: Biometrie, <http://de.wikibooks.org/wiki/Biometrie>, last visited: 21.05.2015.

# Discarding low quality Minutia Cylinder-Code pairs for improved fingerprint comparison

M. Hamed Izadi, Andrzej Drygajlo

Swiss Federal Institute of Technology (EPFL)  
CH-1015 Lausanne, Switzerland

{hamed.izadi, andrzej.drygajlo}@epfl.ch

**Abstract:** Local minutiae descriptors such as Minutia Cylinder-Code (MCC) are becoming increasingly popular in modern fingerprint verification systems. The verification performance depends on the fingerprint image quality in global and local levels. Discarding part of the lowest quality samples based on quality measures is a universal approach being widely used for improving the performance of biometric recognition systems. In this work, we evaluate several different discarding methods to filter out low quality pairs of MCC descriptors using minutiae qualities, with the final aim of improving global comparison accuracy. Moreover, we propose an efficient MCC based fingerprint comparison method based on discarding the low quality elements from local similarity matrix. Our extensive experiments on three different databases (FVC2002\_DB2, FVC2002\_DB3 and FVC2004\_DB3) show that 1) the proper discarding of low quality MCC pairs from local similarity matrix either independently or using pairwise measures can improve the MCC based comparison performance, 2) for the proposed discarding method, the quality of central minutiae is more efficient as cylinder quality measure than the average minutiae qualities in each descriptor.

## 1 Introduction

Fingerprint verification systems are widely used every day for security purposes. The most common method in these systems is minutiae based fingerprint comparison, whose performance depends a lot on the fingerprint image quality. Low quality regions in fingerprint images may harm the verification systems by the extraction of false minutiae. One common solution to this problem is to filter out the false minutiae using minutiae quality [MMJP09, CCM07]. Apart from the false minutiae removal, minutiae quality have been already utilized via other approaches such as quality-based weighting [CDJ05].

Minutia quality is usually computed either using local quality assessment of underlying fingerprint in the minutia neighborhood or based on the correlation with a set of previously selected high quality minutia images [CCM07]. Several methods have been investigated in [CCM07] for embedding minutia quality scores in fingerprint comparison. However, since the reliability of the existing minutiae quality assessment algorithms in discriminating genuine and false minutiae is far from optimal, only certain carefully designed combination of minutiae quality and fingerprint comparison strategies could achieve improvement in

verification performance [CCM07].

Discarding a portion of low quality samples based on quality measures is a universal approach being widely used for improving the performance of many biometric recognition systems. For fingerprint comparison, such a discarding approach have been used in global setting, e.g., in the NIST Fingerprint Image Quality (NFIQ) developments [TWW04, OTMB13], as baseline algorithm for comparing global quality measures [AFFOG<sup>+</sup>07].

Modern fingerprint comparison algorithms are more and more exploiting local minutiae descriptors [CFM12]. Local minutiae descriptors generally encode the relationships between each minutia and its neighboring minutiae within the fingerprint image in terms of some invariant measures. Minutia Cylinder-Code (MCC) [CFM10] is one of the most efficient local minutiae descriptors, which is known for being rotation and translation invariant, robust to skin distortions, and computationally fast. Moreover, it has shown a high performance comparing to other minutiae descriptors [PGT<sup>+</sup>15].

Similar to minutiae qualities, cylinder quality measures have been introduced in [IMD12] as local quality measures for minutiae descriptors, together with some methods to embed such quality measures into the MCC based comparison. These methods are generally based on: 1) quality based weighting scheme [IMD12] or 2) quality based modification of local similarity scores using a training data set of synthetic fingerprints [ID15].

In this paper, we focus mainly on another widely recognized approach, which is discarding of the low quality elements, for improved MCC based fingerprint verification. We evaluate several discarding scenarios for MCC based fingerprint comparison using minutiae qualities. Then, we propose an efficient method based on discarding a portion of low quality elements from local similarity matrix using only the central minutia quality in each cylinder.

The rest of this paper is organized as follows: In Section 2, we briefly introduce the MCC based fingerprint comparison procedure, then in Section 3, we discuss several discarding scenarios for MCC based comparison and propose a novel and efficient discarding method for it. In Section 4, we present the results of our evaluations on three different FVC databases followed by a short discussion. Finally, in Section 5, conclusions are summarized together with some directions for future work.

## 2 MCC based fingerprint comparison

The MCC is a fixed-length descriptor computed for each minutia, encoding its relationships with the neighboring minutiae in a fixed-radius circular area around it. In addition to distance, the angular difference is taken into account using an additional dimension, finally creating a discrete 3D cylinder-shaped structure for each minutia, whose base and height are related to the spatial and directional information, respectively. This 3D structure is then linearized into a vector, whose entries can be binarized into bits by simply setting a threshold. For an MCC pair, a local similarity score can then be computed rather fast using simple bit-based operations, comparing the underlying binary vectors.

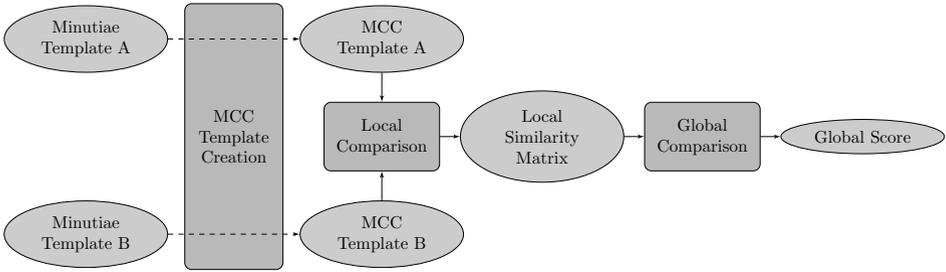


Figure 1: MCC based fingerprint comparison.

Given two MCC templates, say  $A = \{a_1, a_2, \dots, a_{n_A}\}$  and  $B = \{b_1, b_2, \dots, b_{n_B}\}$ , we assume that  $\Gamma(a_r, b_c)$  is the local similarity score between two cylinders  $a_r$  and  $b_c$  from the templates  $A$  and  $B$  respectively.  $r$  and  $c$  denote the cylinder indices in the templates  $A$  and  $B$  respectively ( $1 \leq r \leq n_A$ ,  $1 \leq c \leq n_B$ ). Hence, there are  $n_A \times n_B$  MCC pairs in total. Local similarity scores can be also represented in the form of a matrix  $\Gamma$ , called local similarity matrix, with  $n_A$  rows and  $n_B$  columns. In the next step, usually a set of candidate pairs is pre-selected from the  $n_A \times n_B$  pairs available in  $\Gamma$ . The local similarity score of these pre-selected pairs can be relaxed in an iterative process using second-order compatibility measures, taking into account their global relationship with other pairs. Finally, a small number of pairs (usually between 3 to 12) is selected depending on the minimum number of minutiae in the two templates. The global score is then computed by averaging the similarity scores of the final pairs. These steps are illustrated in Figure 1.

It is worth mentioning that if no relaxation procedure is used for comparison, the final selection will be merged into the pre-selection step, meaning the final pairs will be selected from the matrix  $\Gamma$  to contribute directly into the global score. This is usually the case when there is no information about position and direction of minutiae available in the templates, such as in Noninvertible P-MCC templates [FMC12].

### 3 Discarding approach for MCC based fingerprint comparison

In this work, we focus on discarding approach as a widely recognized methodology for embedding quality measures into biometric recognition systems. This approach is based on the fact that a part of biometric data (samples, regions, ...) with lowest quality can be considered as unreliable data, and discarding them may improve the comparison accuracy in general. There are thresholds needed to be set for discarding criteria. It can be for example an absolute threshold on the value of corresponding quality measures, or the percentage of low quality data to be discarded. Difficulty in setting universal discarding thresholds makes this approach challenging as an embedding technique. Another important application of this approach is to evaluate and compare different quality measures in terms of their ability in discriminating between unreliable and reliable data.

Quality based rejection is usually applied early during the minutiae extraction process in any minutiae based comparison technique. But in this work, we focus on the rejection approach after minutiae extraction process within the MCC based comparison framework. Almost at any stage of MCC based comparison, shown in Figure 1, a discarding method can be applied. For example:

1. Some low-quality or unreliable minutiae can be discarded from minutiae templates.
2. Some unreliable cells can be considered as invalid inside each descriptor.
3. Some low-quality or unreliable descriptors can be discarded from MCC templates before local comparison.
4. Some low-quality or unreliable MCC pairs can be discarded from local similarity matrix before global comparison.

Other discarding scenarios can be considered depending on the global comparison method being used. From the possible approaches listed above, the first one is usually performed during minutiae extraction. The second and third ones are already considered in original MCC algorithm by introducing some cell and cylinder validity criteria for descriptors. Here in this paper, we consider mainly the fourth approach, where low-quality MCC pairs are discarded from local similarity matrix based on quality measures. Other than performance improvement, we also aim at designing a baseline algorithm for evaluating local quality measures within the framework of MCC based comparison.

Assuming a local similarity matrix  $\Gamma$  of size  $n_A \times n_B$ , and a given percentage ( $100 \cdot \alpha$ ) of the MCC pairs to be discarded from  $\Gamma$ , we can consider the following discarding scenarios:

1. *Discarding independently:* We discard  $\text{round}(n_A \times \sqrt{\alpha})$  rows and  $\text{round}(n_B \times \sqrt{\alpha})$  columns entirely from the matrix  $\Gamma$ . These rows and columns are corresponding to the descriptors having the lowest quality in each template independent of the other one.  $\text{round}(x)$  is a rounding operator which returns the nearest integer to  $x$ .
2. *Discarding based on pairwise quality measures:* We discard the  $\text{round}(n_A \times n_B \times \alpha)$  elements from the matrix  $\Gamma$ , corresponding to those MCC pairs having the lowest pairwise quality based on some pairwise function such as square root or minimum.

Discarding elements from local similarity matrix means to replace them with zero. In other words, the local similarity matrix will be multiplied element-wise with a  $n_A \times n_B$  binary mask which is zero where the elements are going to be discarded, and one elsewhere.

## 4 Experiments and Results

### 4.1 Experimental setting

**Databases:** For our evaluations, we have chosen three FVC databases which are captured by different types of sensors: FVC2002.DB2 (optical sensor), FVC2002.DB3 (capacitive

sensor) and FVC2004\_DB3 (thermal sweeping sensor). Each database contains 800 fingerprint images, including 100 different fingers and 8 samples for each finger. Each sample is compared against the remaining samples of the same finger, creating 2800 genuine pairs, and the first sample of each finger is compared to the first sample of the remaining fingers, providing 4950 impostor pairs for each database.

**Minutiae extraction:** The open source minutiae extractor FingerJetFX OSE is used to extract minutiae for all fingerprints. This extractor also provides a quality value for each minutia using a correlation-based method and keeps by default only those minutiae having quality above 40 (out of 100) up to maximum 68 minutiae for each fingerprint.

**MCC parameters:** All parameters for MCC template creation and comparison have been set according to the last published version in [CFMT10].

**MCC template creation and comparison:** The publicly available MCC SDK Version 1.4 has been used to create the bit-based MCC descriptors (MCC16b). The Local Greedy Similarity (LGS) method [CFMT10] is applied for global comparison using the SDK in all cases. Therefore, the global score is directly computed from the local similarity matrix, without any iterative relaxation in between.

#### 4.2 Cylinder quality: average vs. central minutiae quality

In [IMD12], the cylinder quality measures have been proposed based on a weighted average of minutiae qualities inside cylinders, with much bigger weights given to the minutiae close to the center. Here we consider two extreme cases of such cylinder quality measures: 1) a simple average of minutiae qualities inside each descriptor, 2) only the quality of central minutia in each descriptor. The Equal Error Rate (EER) has been evaluated on all three databases for different percentages of MCC pairs discarded independently. The results are shown in Figure 2. One can interpret from this figure that the central minutia quality is usually more efficient than the average minutiae quality to be used in the proposed approach, especially for higher discarding percentages. This could be due to the overlap between the cylinder areas within the fingerprint image.

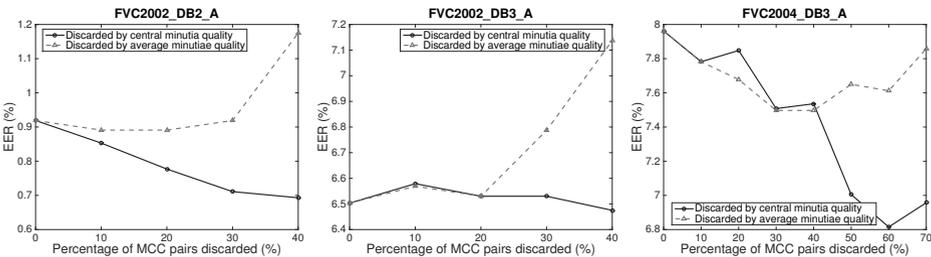


Figure 2: EER vs. percentage of MCC pairs discarded based on central minutia quality (solid line) and based on average minutiae quality (dashed line).

### 4.3 Evaluation of the proposed methods

Following the results presented in Section 4.2, central minutia quality is assumed here to be the cylinder quality measure. Given two MCC descriptors with cylinder quality measures  $Q_a$  and  $Q_b$ , we consider two common pairwise measures for our experiments:  $\sqrt{Q_a \times Q_b}$  and  $\min(Q_a, Q_b)$ . The Equal Error Rate (EER) has been evaluated on each database for different percentages of MCC pairs discarded from local similarity matrix via the methods proposed in Section 3, i.e., (1) discarding MCC pairs independently (independent discarding of rows and columns from local similarity matrix), (2) discarding of MCC pairs based on the pairwise quality  $\sqrt{Q_a \times Q_b}$ , and (3) discarding of MCC pairs based on the pairwise quality  $\min(Q_a, Q_b)$ . The results given in Figure 3 show that all the methods improve the global verification performance to some extent after discarding a proper portion of low-quality MCC pairs. The performance improvement differs for different methods depending on the database and the percentage of discarding. The independent discarding of MCC pairs performs equally well or even better in some cases than the pairwise methods, with  $\min$  function outperforming the  $\sqrt{\phantom{x}}$  function most of the times.

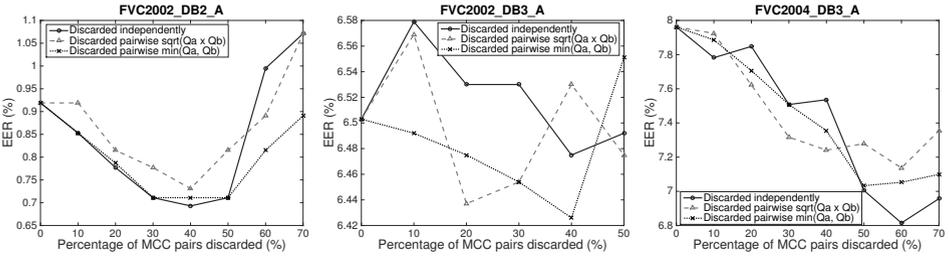


Figure 3: EER vs. percentage of MCC pairs discarded independently (solid line), discarded using pairwise quality-square root (dashed line) and discarded using pairwise quality-minimum (dotted line).

### 4.4 Discussion

In Figure 3, one can see that the proposed discarding method performs much better for FVC2004\_DB3 than FVC2002\_DB3 for example. This could be due to the fact that the average number of minutiae extracted for each fingerprint in FVC2004\_DB3 is much higher (almost double), as seen in Table 1. On the other hand, there are several fingerprints in FVC2002\_DB3 with only a few minutiae, while there is no fingerprint with less than 19 minutiae in the FVC 2004\_DB3. Another interesting difference among these databases is the distribution of minutiae qualities, shown in Figure 4. The distribution looks closer to normal for the FVC2004\_DB3, with only a small percentage of minutiae having very low quality. On the other hand, the minutiae qualities have a rather different distribution in the FVC2002\_DB3 with relatively a high percentage of low quality minutiae.

Table 1: Some statistics on the number of extracted minutiae per fingerprint.

Database	Mean	Min	Max	Std
FVC2002_DB2_A	50.8	9	68	14.0
FVC2002_DB3_A	31.2	6	68	11.5
FVC2004_DB3_A	64.1	19	68	8.7

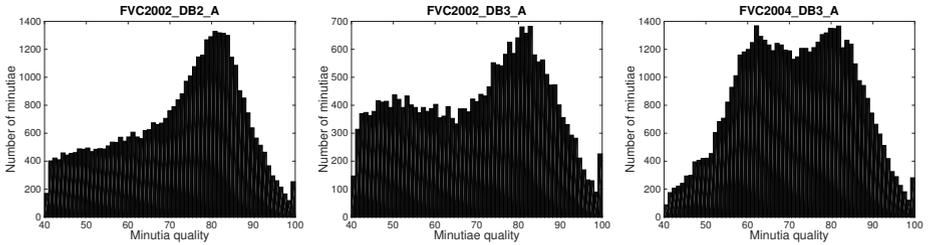


Figure 4: Distribution of minutiae qualities extracted by FingerJetFX.

## 5 Conclusions and future works

In this paper, we focused on the quality based discarding approach for improving the MCC based fingerprint comparison. We evaluated several different discarding scenarios in this context, and proposed an efficient discarding method based on discarding the low quality elements from the local similarity matrix. These elements could be discarded independently or by using some pairwise quality measures. Our experiments on three different FVC databases show that proper discarding of low quality MCC pairs either independently or pairwise can improve the comparison performance. On the other hand, the quality of central minutiae was shown to be more efficient for this discarding method than the average minutiae qualities in each descriptor. As a future step, we aim at using a similar discarding approach to evaluate different local quality measures in the context of minutiae based fingerprint comparison. The adaptation to other comparison methods involving relaxation and other applications such as palm print comparison will be considered as well. Setting a threshold on the minimum number of minutiae per fingerprint might be also helpful to be combined with this approach.

## 6 Acknowledgment

This work was supported in part by the Swiss National Science Foundation (SNSF) through the grant 200020-146826.

## References

- [AFFOG<sup>+</sup>07] F. Alonso-Fernandez, J. Fierrez, J. Ortega-Garcia, J. Gonzalez-Rodriguez, H. Fronthaler, K. Kollreider, and J. Bigun. A Comparative Study of Fingerprint Image-Quality Estimation Methods. *IEEE Transactions on Information Forensics and Security*, 2(4):734–743, December 2007.
- [CCM07] J. Chen, F. Chan, and Y.-S. Moon. Fingerprint Matching with Minutiae Quality Score. In Seong-Whan Lee and Stan Li, editors, *Advances in Biometrics*, volume 4642, pages 663–672. Springer Berlin Heidelberg, 2007.
- [CDJ05] Y. Chen, S. C. Dass, and A. K. Jain. Fingerprint quality indices for predicting authentication performance. In *International Conference on Audio- and Video-Based Biometric Person Authentication (AVBPA)*, pages 160–170, July 2005.
- [CFM10] R. Cappelli, M. Ferrara, and D. Maltoni. Minutia Cylinder-Code: A New Representation and Matching Technique for Fingerprint Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(12):2128–2141, December 2010.
- [CFM12] R. Cappelli, M. Ferrara, and D. Maltoni. Minutiae-Based Fingerprint Matching. In C. Liu and V. Mago, editors, *Cross Disciplinary Biometric Systems*, volume 37, pages 117–150. Springer Berlin Heidelberg, 2012.
- [CFMT10] R. Cappelli, M. Ferrara, D. Maltoni, and M. Tistarelli. MCC: a Baseline Algorithm for Fingerprint Verification in FVC-onGoing. In *International Conference on Control Automation Robotics Vision (ICARCV)*, pages 19–23, December 2010.
- [FMC12] M. Ferrara, D. Maltoni, and R. Cappelli. Noninvertible Minutia Cylinder-Code Representation. *IEEE Transactions on Information Forensics and Security*, 7(6):1727–1737, Dec 2012.
- [ID15] M. H. Izadi and A. Drygajlo. How synthetic fingerprints can improve pre-selection of MCC pairs using local quality measures. In *3rd International Workshop on Biometrics and Forensics (IWBF)*, March 2015.
- [IMD12] M. H. Izadi, L. Mirmohamadsadeghi, and A. Drygajlo. Introduction of Cylinder Quality Measure into Minutia Cylinder-Code based Fingerprint Matching. In *IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, September 2012.
- [MMJP09] D. Maltoni, D. Maio, A. K. Jain, and S. Prabhakar. *Handbook of Fingerprint Recognition*. Springer, 2nd edition, 2009.
- [OTMB13] M. A. Olsen, E. Tabassi, A. Makarov, and C. Busch. Self-Organizing Maps for Fingerprint Image Quality Assessment. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 138–145, June 2013.
- [PGT<sup>+</sup>15] Daniel Peralta, Mikel Galar, Isaac Triguero, Daniel Paternain, Salvador Garca, Edurne Barrenechea, Jos M. Bentez, Humberto Bustince, and Francisco Herrera. A survey on fingerprint minutiae-based local matching for verification and identification: Taxonomy and experimental evaluation. *Information Sciences*, 315:67 – 87, 2015.
- [TWW04] E. Tabassi, C. L. Wilson, and C. I. Watson. Fingerprint Image Quality. *NISTIR 7151*, 2004.

# A Signature Complexity Measure to select Reference Signatures for Online Signature Verification

Christian Kahindo, Sonia Garcia-Salicetti, Nesma Houmani

Institut Mines-Telecom  
Télécom SudParis, CNRS UMR 5157 SAMOVAR, CEA Saclay Nano-Innov  
PC176 Bât. 861  
91191 Gif sur Yvette Cedex, France  
christian.kahindo@telecom-sudparis.eu  
sonia.garcia@telecom-sudparis.eu  
int.nesma@gmail.com

**Abstract:** This paper presents an original procedure for selecting the reference online signature instances of a writer, an important issue for any effective signature verifier. To this end, for each signature instance, we propose a novel complexity measure, by exploiting a global description of signatures in the frequency domain as well as a global statistical modelling of each signature instance. To select the reference signatures, we propose a method based on the distribution of complexity values for all the available genuine signatures. The 2500 genuine samples of MCYT-100 online database are used in this study. Experimental results show the effectiveness of the method and of the here proposed complexity measure for this specific task.

## 1 Introduction

One of the most widespread means to verify the identity of a person in our society is handwritten signature, and that since a long time, for example as a mean of guaranteeing the validity of a document in the legal field or in banking transactions [HG14]. Nowadays, signatures can either be acquired online as a temporal signal on a digitizer or a smartphone [HG14, IP08] or offline as a static image [IP08, SPL92]. Our study is carried out in the online framework.

The implementation of an automatic signature verification system consists of two phases: enrolment and verification. The enrolment consists in the acquisition of signatures that will be stored as references or be used to build a writer-model. During verification, the writer claims an identity and captures his/her probe signature, then given as input to the verification system; its outcome is the acceptance or rejection of the writer's claimed identity. Thus, enrolment is the first step of any verifier, and is a crucial phase for improving the reliability of the verification system [BP89, Di99].

It is well-known that signature is a behavioral biometric modality with high intra-class variability. Such variability is the main obstacle for accurate signature verification. For

this reason, it is essential to have an effective criterion for selecting the pertinent signatures of the reference set. Most previous works on handwritten signature are focused on the verification step, while only very few tackle the selection of reference signatures based on signature stability and signature complexity criteria.

Brault and Plamondon proposed in [BP89] a measure of signature complexity, the “difficulty coefficient”, which is a function of the rate of geometric modifications such as length, direction of strokes and curvature per unit of time. This coefficient was used to accept or reject a signature at the enrolment step, by accepting a signature only if it is complex enough. The authors also proposed a “dissimilarity index” based on elastic matching between two signatures, for measuring the intra-class variability within the genuine signatures of a writer. They conclude that signers with low intra-class variability have a low rate of false acceptance and propose to select as references those signatures showing a low dissimilarity index. Di Lecce et al. [Di99] proposed a method to select reference signatures based on the analysis of stability in handwritten dynamic signatures. They compute signature stability as a sum of local stability indices. Elastic matching techniques are used to compute the correlation between different signatures of a writer, and a subset of signatures with the highest correlation is selected as reference set [Di02]. More recently, Guest and Fairhurst [GF06] carried out a sample signature selection at the enrolment step based on the assessment of the “Coefficient of Variance” (COV) for each global feature across all samples for a particular subject. The triplet of signatures with lowest COV value, namely with lowest variance, are selected as references. All such works of the literature point out the impact of complexity and stability criteria on improving the performance of signature verification systems.

The aim of this work is to propose an original approach for selecting the reference signatures of a writer based on a new complexity measure. Such measure is constructed by exploiting a global description of signatures in the frequency domain as well as a global statistical modelling of each signature instance. The hypothesis of this work is that the proposed complexity measure is fine enough for reflecting the variations of a writer’s signature from one instance to the next. For this reason, in order to select reference signatures, we exploit the distribution of complexity values of all the available genuine signatures. Experimental results on the widely used MCYT-100 database validate our hypothesis: the complexity measure characterizes well each genuine signature and can thus be used successfully for building a criterion to select reference signatures.

The organization of the paper is the following: Section 2 presents the complexity measure of a signature instance, Section 3 describes the experimental setup and the analysis of results. Finally, we conclude on the scope of this study in Section 4.

## 2 The novel complexity measure for selecting reference signatures

### 2.1 Quantifying complexity on the raw description of a signature instance

Online handwritten signatures are acquired on a digitizer, and according to this sensor's properties, different time functions are available (pen coordinates, pen pressure, pen inclination through time) [HG14]. In this study, we consider a signature as a raw sequence of pen coordinates  $(x(t), y(t))$  since this description of signatures is common to digitizers, tablets and smartphones. If such sequence of points representing an online signature is considered as being the outcome of a random variable, the concept of entropy can be used for estimating the degree of disorder associated to this random variable. The entropy of this variable depends on its associated probability density function [CT06]. To this end, an accurate estimation of the probability density associated to each signature instance must be achieved. We exploit for this purpose a Gaussian Mixture Model (GMM) [Re95], since this model has proven its efficiency in modeling signatures [MM08]. A GMM [Re95] is a weighted sum of  $M$  component Gaussian densities as given by the equation:

$$p(x|\lambda) = \sum_{i=1}^M w_i g(x|\mu_i, \Sigma_i) \quad (1.1)$$

where  $x$  is a  $D$ -dimensional continuous-valued data vector (i.e. feature vector),  $w_i$  for  $i = 1, \dots, M$ , are the mixture weights, and  $g(x|\mu_i, \Sigma_i)$ ,  $i = 1, \dots, M$ , are the component Gaussian densities. Each component density is a  $D$ -variate Gaussian function of the form,

$$g(x|\mu_i, \Sigma_i) = \frac{1}{(2\pi)^{D/2} |\Sigma_i|^{1/2}} \exp \left\{ -\frac{1}{2} (x - \mu_i)' \Sigma_i^{-1} (x - \mu_i) \right\} \quad (1.2)$$

with mean vector  $\mu_i$  and covariance matrix  $\Sigma_i$ .

The statistical complexity measure here proposed is based on the concept of differential entropy of information theory. For a given random variable  $X$  with a probability distribution  $f$ , the differential entropy  $h(x)$  is defined as follows:

$$h(x) = -\int_x f(x) \ln f(x) dx \quad (1.3)$$

For the multidimensional Gaussian distribution defined in Equation 1.2, such entropy has the following simplified form:

$$H(t) = \frac{1}{2} \ln \{ (2\pi e)^N \det(\Sigma) \} \quad (1.4)$$

For each signature of a given writer, we compute its complexity index as follows: the Gaussian component that gives the highest probability (the maximum value of the expression in Equation 1.2) is assigned to each point  $(x(t), y(t))$ . Then, we assign to the current point its corresponding differential entropy using Equation 1.4. For a signature sample of length  $N$ , the complexity index is defined as follows:

$$C = \frac{\sum_{t=1}^N H(t)}{N} \quad (1.5)$$

## 2.2 Quantifying complexity on the frequency domain of a signature instance

Fourier descriptors of a signature have already been used in the literature [KY08]. Fourier transform gives a global description of what happens in the temporal domain, by breaking down the signal into constituent sinusoids of different frequencies. The Fourier Transform coefficients of a given signal  $y(t)$  of length  $N$  are defined as follows:

$$C_k = \frac{1}{N} \sum_{t=0}^{N-1} y(t) e^{-j2\pi tk/N} \quad k=0,1 \dots N-1. \quad (1.6)$$

For a given signature, Fourier analysis is carried out separately on  $x(t)$  and  $y(t)$ ,  $N$  being the number of points in the signature, and  $C_k$  the  $k$ -th Fourier coefficient  $C_k = a_k + jb_k$ . We exploit the magnitude of such coefficient, namely  $\sqrt{a_k^2 + b_k^2}$ , which measures the energy of the signal for the  $k$ -th harmonic. The resulting energy spectrum on  $x$  and  $y$  is then given as input to a GMM, this way using the same approach described in the previous section (2.1). Indeed, we aim at comparing a *global* description of signatures in the frequency domain, with its raw description in the time domain.

The next section presents our proposal of exploiting this complexity index computed on a signature instance for selecting the reference signatures of a given writer.

## 2.3 Selection of reference signatures based on the complexity index

Based on the complexity index above defined, we perform a Hierarchical Clustering in order to study the behavior of such measure on *all* genuine signature samples available.

Our study is carried out on the freely available and the widely used MCYT-100 subset of 100 persons [Or03]. We chose this database because it contains Western signatures of different styles, varying from simple flourish signatures to very complex flourish ones (rather close to cursive handwriting). Indeed, this allows assessing whether the complexity measure quantifies the existing gaps in complexity between different writers.

We determined the optimal number of clusters by computing different validity indices of the literature, namely Krzanowski-Laï index [DF02], Davies-Bouldin index [DB79], silhouette [Ro87], and Weighted intrer-intra index [St02]. The optimal number of clusters is 3, namely 3 categories of signatures according to their complexity, respectively displayed in Figure 1(a), 1(b) and 1(c).

The same validity indices are used to assess the optimal number of Gaussian components for the statistical model (GMM). We obtained that 24 mixture components is the optimal configuration because it optimizes the 4 validity indices, ensuring the best clustering.

In the same way, we assess the quality of the clustering on both the raw description of signatures and on their global description in the frequency domain. These 4 indices point out that the clustering on complexity values obtained after performing Fourier analysis on signatures is by far better than that obtained with the raw description of signatures.

### 2.3.1 Signatures' categories obtained by Hierarchical Clustering

As mentioned in the previous section, we retrieve 3 categories of signatures on the 2500 genuine signatures available. Figure 1 shows that each of such categories has a different degree of complexity and Table 1 gives the average complexity and its standard deviation per category. We clearly obtain a low complexity category (Figure 1(a)), a medium one (Figure 1(b)) and a high complexity category (Figure 1(c)). This result shows that our complexity index behaves well. Moreover, Figure 1(d) displays the values of the complexity index for all signatures per category, revealing that *its variance differs significantly between categories*. Indeed, this variance lowers with complexity; this can be seen in the upper part of Figure 1(d) (category in red of lowest variance), then in the medium complexity category with a higher variance (see complexity values in blue), and finally in the lowest complexity category with the highest variance (see values in green). This result confirms that complexity and stability are correlated in signatures as previously shown in the literature [BP89, GHD09].

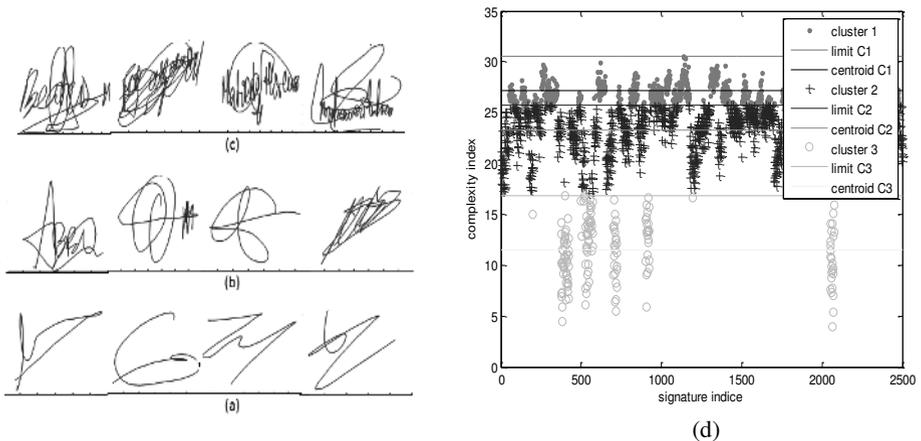


Figure 1: Examples of signatures in 3 complexity categories obtained by Hierarchical Clustering, (a) low, (b) medium, (c) high complexity. Such signatures were already published [Or03].

Complexity index-based categories	Percentage of signatures	Mean value	Std value
Low complexity	6.96%	11.50	2.9186
Medium complexity	51.16%	23.27	1.9494
High complexity	41.88%	27.15	0.9618

Table 1: Distribution of signatures of the MCYT-100 database in each complexity-based category; mean and standard deviation (Std) values of complexity per category.

### 2.3.2 The proposed method for selecting reference signatures

To select the best reference signatures, we analyze the distribution of complexity values on *all* genuine signature instances of a writer. The five nearest signatures to the median (indicated in red inside the boxplot of Figure (2b)) that is found between the first quartile ( $Q1=25\%$  of values) and the third quartile ( $Q3=75\%$  of values) are selected. Figure 2(a)

illustrates this method on the 25 genuine signatures from the first person in MCYT-100. This person belongs to the medium complexity category; note that complexity values of his/her signatures are spread in a quite large interval (17 to 24). This shows that the *intra-class variation is well reflected by our novel complexity measure* since it is *sensitive to differences in signature instances of a same writer*. This fact has also an impact on the standard deviation of complexity values reported in Table 1.

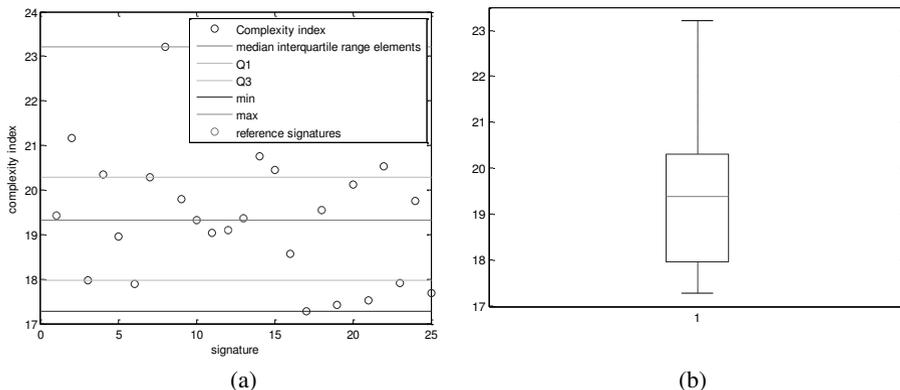


Figure 2: (a) The statistical distribution of the complexity index for all signatures and (b) the boxplot of the first person in MCYT-100 database.

### 3. Experiments

In the following, we evaluate the impact of the proposed method for selecting reference signatures in performance of a signature verification system. We compare the proposed method to a random selection of reference signatures. The 25 genuine signatures available per writer are used. The signature verification approach exploited for this evaluation is Dynamic Time Warping (DTW), proven to be one of the best approaches for signature verification [Ye04]. Concerning our method, reference signatures are selected in two ways: the 5 nearest to the median between quartiles Q1 and Q3 as explained above, and the 5 nearest to the mean value in the same interval (boxplot).

The random selection consists in sampling 5 reference signatures among the 25 genuine signatures of a given writer and consider his/her remaining 20 genuine signatures and the available 25 forgeries for verification purposes. We repeat the process 5 times. We compare in Figure 3 classifier performance with the 3 methods for reference signatures' selection: the 5 nearest to the median between quartiles Q1 and Q3, the 5 nearest to the mean value in the same interval (boxplot), and the random selection. Results are also displayed in Table 2 in terms of the Minimum Half Total Error Rate (minHTER). Our method for selecting signatures results in a significant relative improvement of 18% compared to the random selection. This result points out the pertinence of our approach that is based on an accurate complexity measure.

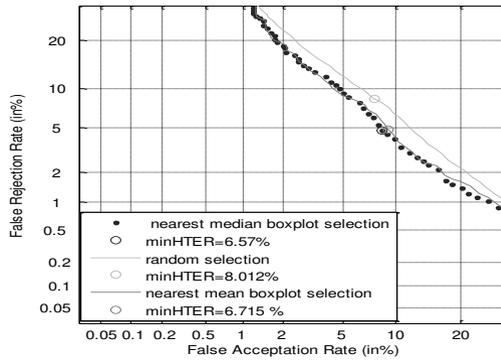


Figure 3: Detection Error Trade-off (DET)-Curves for selection of reference signatures based on complexity index.

Selection method	minHTER	Relative improvement compared to random selection
5 nearest to the median between Q1 & Q3	6.57%	18%
5 nearest to the mean between Q1 & Q3	6.715%	16%
5 random selected	8.012 %	--

Table 2: minHTER and relative improvement compared to random selection

## 4. Conclusions and future work

In this paper, a novel method for selecting reference signatures of a given writer is proposed. It is based on an original complexity measure that exploits a statistical global approach and a global description of signatures in the frequency domain. Experimental results reveal the effectiveness of the method, by generating a significant relative improvement of verification performance compared to a random selection of reference signatures. This proves that our complexity measure is not only able to reflect the gap in complexity between different writers and categories of writers, but even able to reflect *the variations in signature instances of a same writer*. In other words, it is *sensitive to intraclass variation* and thus an accurate tool for selecting references. Future work will be focused on studying how dynamic parameters have an influence on complexity, aiming at improving our selection method.

## Acknowledgments

This work was partially funded by Fondation MAIF through project “Biométrie et santé sur tablette” ([http://www.fondation-maif.fr/notre-action.php?rub=1&sous\\_rub=3&id=269](http://www.fondation-maif.fr/notre-action.php?rub=1&sous_rub=3&id=269)).

## References

- [BP89] Brault, J., Plamondon, R.: How to detect problematic signers for automatic signature verification. In *Proc. Int. Carnahan Conf. On Security Technology*, 1989; pp. 127-132.

- [CT06] Cover, T.M., Thomas, J. A.: Elements of Information Theory, Second Edition, John Wiley & Sons, (2006).
- [DB79] Davies, D. L., Bouldin D. W.: "A Cluster Separation Measure." *IEEE Transactions on Pattern Analysis and Machine Intelligence*. Vol. PAMI-1, No. 2, 1979, pp. 224–227.
- [Di99] Di Lecce, V., Di Mauro, G., Guerriero, A., Impedovo, S., Pirlo, G., Salzo, A., Sarcinella, L.: Selection of Reference Signatures for Automatic Signature Verification. In: Int Conf. on Document Analysis and Recognition (ICDAR'99), Bangalore, India, 1999.; pp. 597-600.
- [DF02] Dudoit, S., Fridlyand., J.: A prediction-based resampling method for estimating the number of clusters in a dataset, *Genome Biology*, 3(7): Research0036.1–0036.21, 2002.
- [Di02] Dimauro, G., Impedovo, S., Modugno, R., Pirlo, G., & Sarcinella, L.: Analysis of stability in hand-written dynamic signatures: IEEE Computer Society. In : *Ninth Int. Workshop on Frontiers in Handwriting Recognition.*, 2002 pp. 259-259;.
- [GF06] Guest, R. et Fairhurst, M., Sample selection for optimising signature enrolment. In : *Tenth International Workshop on Frontiers in Handwriting Recognition*. Suvisoft, 2006.
- [GHD09] Garcia-Salicetti, S., Houmani, N., Dorizzi, B.: A Novel Criterion for Writer Enrolment based on a Time- Normalized Signature Sample Entropy Measure, *EURASIP Journal on Advances in SignalProcessing* 2009, doi:10.1155/2009/964746.
- [HG14] N. Houmani and S. Garcia-Salicetti, "Digitizing Tablet", in "Encyclopedia of Biometrics", ISBN 978-3-642-27733-7(Online), DOI 10.1007/978-3-642-27733-7\_19-3, Eds: Stan Li, Anil K. Jain, Springer Science + Business Media New York, 2014.
- [IP08] Impedovo, D., Pirlo, G.: Automatic Signature Verification: The State of the Art, *IEEE Transactions on Systems, Man, and Cybernetics-Part C: Applications and Reviews*, 38(5): 2008.
- [KY08] Kholmatov, A. et Yanikoglu, B.: An individuality model for online signatures using global Fourier descriptors. In : *SPIE Defense and Security Symposium*. Int.Society for Optics and Photonics, 2008. p. 694407-694407-12.
- [MM08] Miguel-Hurtado, O., Mengibar-Pozo, L. Pacut, A.: A new algorithm for signature verification system based on DTW and GMM, 42<sup>nd</sup> Annual IEEE int. Conf., Carnahan, 2008; pp 206-213.
- [Or03] Ortega-Garcia, J., Fierrez-Aguilar, J., Simon, D., Gonzalez, J., Faundez-Zanuy, M., Espinosa, V., Satue, A., Hernaez, I., Igarza, J.-J., Vivaracho, C., Escudero, D., Moro, Q.-I.: MCYT Baseline Corpus: A Bimodal Biometric Database, *IEE Proc. Vision, Image and Signal Processing, Special Issue on Biometrics on the Internet*, 150(6): 395-401, 2003.
- [Re95] Douglas A. Reynolds.: Robust Text-Independent Speaker identification Using Gaussian Mixture Speaker Models. *IEEE transaction on speech and audio processing*, vol 3, 1995.
- [Ro87] Rouseeuw, P. J. "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis." *Journal of Computational and Applied Mathematics*. Vol. 20, No. 1, 1987, pp. 53–65.
- [SPL92] Sabourin, R., Plamondon, R., Lorette, G.: Off-line Identification with Handwritten Signature Images: Survey and Perspectives, in *Structured Document Image Analysis*, Eds: Baird, H. S., Bunke, H., Yamamoto, K., Publisher Springer Berlin Heidelberg, pp. 219-234, 1992.
- [St02] A. Strehl, Relationship-Based Clustering and Cluster Ensembles for High-dimensional Data Mining, Ph.D Thesis, University of Texas at Austin, May 2002.
- [Ye04] Yeung, D., Chang, H., Xiong, Y., George, S., Kashi, R., Matsumoto, T., Rigoll, G.: SVC2004: Proc. First Int. Signature Verification Competition. In: the Int. Conf. on Biometric Authentication (ICBA), LNCS 3072, Springer, , Hong Kong, China, 2004; pp. 16 - 22.

## *GI-Edition Lecture Notes in Informatics*

- P-1 Gregor Engels, Andreas Oberweis, Albert Zündorf (Hrsg.): Modellierung 2001.
- P-2 Mikhail Godlevsky, Heinrich C. Mayr (Hrsg.): Information Systems Technology and its Applications, ISTA'2001.
- P-3 Ana M. Moreno, Reind P. van de Riet (Hrsg.): Applications of Natural Language to Information Systems, NLDB'2001.
- P-4 H. Wörn, J. Mühling, C. Vahl, H.-P. Meinzer (Hrsg.): Rechner- und sensor-gestützte Chirurgie; Workshop des SFB 414.
- P-5 Andy Schürr (Hg.): OMER – Object-Oriented Modeling of Embedded Real-Time Systems.
- P-6 Hans-Jürgen Appelrath, Rolf Beyer, Uwe Marquardt, Heinrich C. Mayr, Claudia Steinberger (Hrsg.): Unternehmen Hochschule, UH'2001.
- P-7 Andy Evans, Robert France, Ana Moreira, Bernhard Rumpe (Hrsg.): Practical UML-Based Rigorous Development Methods – Countering or Integrating the extremists, pUML'2001.
- P-8 Reinhard Keil-Slawik, Johannes Magenheim (Hrsg.): Informatikunterricht und Medienbildung, INFOS'2001.
- P-9 Jan von Knop, Wilhelm Haverkamp (Hrsg.): Innovative Anwendungen in Kommunikationsnetzen, 15. DFN Arbeitstagung.
- P-10 Mirjam Minor, Steffen Staab (Hrsg.): 1st German Workshop on Experience Management: Sharing Experiences about the Sharing Experience.
- P-11 Michael Weber, Frank Kargl (Hrsg.): Mobile Ad-Hoc Netzwerke, WMAN 2002.
- P-12 Martin Glinz, Günther Müller-Luschnat (Hrsg.): Modellierung 2002.
- P-13 Jan von Knop, Peter Schirmbacher and Viljan Mahni\_ (Hrsg.): The Changing Universities – The Role of Technology.
- P-14 Robert Tolksdorf, Rainer Eckstein (Hrsg.): XML-Technologien für das Semantic Web – XSW 2002.
- P-15 Hans-Bernd Bludau, Andreas Koop (Hrsg.): Mobile Computing in Medicine.
- P-16 J. Felix Hampe, Gerhard Schwabe (Hrsg.): Mobile and Collaborative Business 2002.
- P-17 Jan von Knop, Wilhelm Haverkamp (Hrsg.): Zukunft der Netze –Die Verletzbarkeit meistern, 16. DFN Arbeitstagung.
- P-18 Elmar J. Sinz, Markus Plaha (Hrsg.): Modellierung betrieblicher Informationssysteme – MobIS 2002.
- P-19 Sigrid Schubert, Bernd Reusch, Norbert Jesse (Hrsg.): Informatik bewegt – Informatik 2002 – 32. Jahrestagung der Gesellschaft für Informatik e.V. (GI) 30.Sept.-3. Okt. 2002 in Dortmund.
- P-20 Sigrid Schubert, Bernd Reusch, Norbert Jesse (Hrsg.): Informatik bewegt – Informatik 2002 – 32. Jahrestagung der Gesellschaft für Informatik e.V. (GI) 30.Sept.-3. Okt. 2002 in Dortmund (Ergänzungsband).
- P-21 Jörg Desel, Mathias Weske (Hrsg.): Promise 2002: Prozessorientierte Methoden und Werkzeuge für die Entwicklung von Informationssystemen.
- P-22 Sigrid Schubert, Johannes Magenheim, Peter Hubwieser, Torsten Brinda (Hrsg.): Forschungsbeiträge zur "Didaktik der Informatik" – Theorie, Praxis, Evaluation.
- P-23 Thorsten Spitta, Jens Borchers, Harry M. Sneed (Hrsg.): Software Management 2002 – Fortschritt durch Beständigkeit
- P-24 Rainer Eckstein, Robert Tolksdorf (Hrsg.): XMIDX 2003 – XML-Technologien für Middleware – Middleware für XML-Anwendungen
- P-25 Key Pousttchi, Klaus Turowski (Hrsg.): Mobile Commerce – Anwendungen und Perspektiven – 3. Workshop Mobile Commerce, Universität Augsburg, 04.02.2003
- P-26 Gerhard Weikum, Harald Schöning, Erhard Rahm (Hrsg.): BTW 2003: Datenbanksysteme für Business, Technologie und Web
- P-27 Michael Kroll, Hans-Gerd Lipinski, Kay Melzer (Hrsg.): Mobiles Computing in der Medizin
- P-28 Ulrich Reimer, Andreas Abecker, Steffen Staab, Gerd Stumme (Hrsg.): WM 2003: Professionelles Wissensmanagement – Erfahrungen und Visionen
- P-29 Antje Düsterhöft, Bernhard Thalheim (Eds.): NLDB'2003: Natural Language Processing and Information Systems
- P-30 Mikhail Godlevsky, Stephen Liddle, Heinrich C. Mayr (Eds.): Information Systems Technology and its Applications
- P-31 Arslan Brömme, Christoph Busch (Eds.): BIOSIG 2003: Biometrics and Electronic Signatures

- P-32 Peter Hubwieser (Hrsg.): Informatische Fachkonzepte im Unterricht – INFOS 2003
- P-33 Andreas Geyer-Schulz, Alfred Taudes (Hrsg.): Informationswirtschaft: Ein Sektor mit Zukunft
- P-34 Klaus Dittrich, Wolfgang König, Andreas Oberweis, Kai Rannenber, Wolfgang Wahlster (Hrsg.): Informatik 2003 – Innovative Informatikanwendungen (Band 1)
- P-35 Klaus Dittrich, Wolfgang König, Andreas Oberweis, Kai Rannenber, Wolfgang Wahlster (Hrsg.): Informatik 2003 – Innovative Informatikanwendungen (Band 2)
- P-36 Rüdiger Grimm, Hubert B. Keller, Kai Rannenber (Hrsg.): Informatik 2003 – Mit Sicherheit Informatik
- P-37 Arndt Bode, Jörg Desel, Sabine Rathmayer, Martin Wessner (Hrsg.): DeLFI 2003: e-Learning Fachtagung Informatik
- P-38 E.J. Sinz, M. Plaha, P. Neckel (Hrsg.): Modellierung betrieblicher Informationssysteme – MobIS 2003
- P-39 Jens Nedon, Sandra Frings, Oliver Göbel (Hrsg.): IT-Incident Management & IT-Forensics – IMF 2003
- P-40 Michael Rebstock (Hrsg.): Modellierung betrieblicher Informationssysteme – MobIS 2004
- P-41 Uwe Brinkschulte, Jürgen Becker, Dietmar Fey, Karl-Erwin Großpietsch, Christian Hochberger, Erik Maehle, Thomas Runkler (Edts.): ARCS 2004 – Organic and Pervasive Computing
- P-42 Key Pousttchi, Klaus Turowski (Hrsg.): Mobile Economy – Transaktionen und Prozesse, Anwendungen und Dienste
- P-43 Birgitta König-Ries, Michael Klein, Philipp Obreiter (Hrsg.): Persistence, Scalability, Transactions – Database Mechanisms for Mobile Applications
- P-44 Jan von Knop, Wilhelm Haverkamp, Eike Jessen (Hrsg.): Security, E-Learning, E-Services
- P-45 Bernhard Rumpe, Wolfgang Hesse (Hrsg.): Modellierung 2004
- P-46 Ulrich Flegel, Michael Meier (Hrsg.): Detection of Intrusions of Malware & Vulnerability Assessment
- P-47 Alexander Prosser, Robert Krimmer (Hrsg.): Electronic Voting in Europe – Technology, Law, Politics and Society
- P-48 Anatoly Doroshenko, Terry Halpin, Stephen W. Liddle, Heinrich C. Mayr (Hrsg.): Information Systems Technology and its Applications
- P-49 G. Schiefer, P. Wagner, M. Morgenstern, U. Rickert (Hrsg.): Integration und Datensicherheit – Anforderungen, Konflikte und Perspektiven
- P-50 Peter Dadam, Manfred Reichert (Hrsg.): INFORMATIK 2004 – Informatik verbindet (Band 1) Beiträge der 34. Jahrestagung der Gesellschaft für Informatik e.V. (GI), 20.-24. September 2004 in Ulm
- P-51 Peter Dadam, Manfred Reichert (Hrsg.): INFORMATIK 2004 – Informatik verbindet (Band 2) Beiträge der 34. Jahrestagung der Gesellschaft für Informatik e.V. (GI), 20.-24. September 2004 in Ulm
- P-52 Gregor Engels, Silke Seehusen (Hrsg.): DELFI 2004 – Tagungsband der 2. e-Learning Fachtagung Informatik
- P-53 Robert Giegerich, Jens Stoye (Hrsg.): German Conference on Bioinformatics – GCB 2004
- P-54 Jens Borchers, Ralf Kneuper (Hrsg.): Softwaremanagement 2004 – Outsourcing und Integration
- P-55 Jan von Knop, Wilhelm Haverkamp, Eike Jessen (Hrsg.): E-Science und Grid Ad-hoc-Netze Medienintegration
- P-56 Fernand Feltz, Andreas Oberweis, Benoit Otjacques (Hrsg.): EMISA 2004 – Informationssysteme im E-Business und E-Government
- P-57 Klaus Turowski (Hrsg.): Architekturen, Komponenten, Anwendungen
- P-58 Sami Beydeda, Volker Gruhn, Johannes Mayer, Ralf Reussner, Franz Schweiggert (Hrsg.): Testing of Component-Based Systems and Software Quality
- P-59 J. Felix Hampe, Franz Lehner, Key Pousttchi, Kai Rannenber, Klaus Turowski (Hrsg.): Mobile Business – Processes, Platforms, Payments
- P-60 Steffen Friedrich (Hrsg.): Unterrichtskonzepte für informatische Bildung
- P-61 Paul Müller, Reinhard Gotzhein, Jens B. Schmitt (Hrsg.): Kommunikation in verteilten Systemen
- P-62 Federrath, Hannes (Hrsg.): „Sicherheit 2005“ – Sicherheit – Schutz und Zuverlässigkeit
- P-63 Roland Kaschek, Heinrich C. Mayr, Stephen Liddle (Hrsg.): Information Systems – Technology and its Applications

- P-64 Peter Liggesmeyer, Klaus Pohl, Michael Goedicke (Hrsg.): Software Engineering 2005
- P-65 Gottfried Vossen, Frank Leymann, Peter Lockemann, Wolffried Stucky (Hrsg.): Datenbanksysteme in Business, Technologie und Web
- P-66 Jörg M. Haake, Ulrike Lucke, Djamshid Tavangarian (Hrsg.): DeLFI 2005: 3. deutsche e-Learning Fachtagung Informatik
- P-67 Armin B. Cremers, Rainer Manthey, Peter Martini, Volker Steinhage (Hrsg.): INFORMATIK 2005 – Informatik LIVE (Band 1)
- P-68 Armin B. Cremers, Rainer Manthey, Peter Martini, Volker Steinhage (Hrsg.): INFORMATIK 2005 – Informatik LIVE (Band 2)
- P-69 Robert Hirschfeld, Ryszard Kowalczyk, Andreas Polze, Matthias Weske (Hrsg.): NODe 2005, GSEM 2005
- P-70 Klaus Turowski, Johannes-Maria Zaha (Hrsg.): Component-oriented Enterprise Application (COAE 2005)
- P-71 Andrew Torda, Stefan Kurz, Matthias Rarey (Hrsg.): German Conference on Bioinformatics 2005
- P-72 Klaus P. Jantke, Klaus-Peter Fähnrich, Wolfgang S. Wittig (Hrsg.): Marktplatz Internet: Von e-Learning bis e-Payment
- P-73 Jan von Knop, Wilhelm Haverkamp, Eike Jessen (Hrsg.): "Heute schon das Morgen sehen"
- P-74 Christopher Wolf, Stefan Lucks, Po-Wah Yau (Hrsg.): WEWoRC 2005 – Western European Workshop on Research in Cryptology
- P-75 Jörg Desel, Ulrich Frank (Hrsg.): Enterprise Modelling and Information Systems Architecture
- P-76 Thomas Kirste, Birgitta König-Riess, Key Pousttchi, Klaus Turowski (Hrsg.): Mobile Informationssysteme – Potentiale, Hindernisse, Einsatz
- P-77 Jana Dittmann (Hrsg.): SICHERHEIT 2006
- P-78 K.-O. Wenkel, P. Wagner, M. Morgens-tern, K. Luzi, P. Eisermann (Hrsg.): Land- und Ernährungswirtschaft im Wandel
- P-79 Bettina Biel, Matthias Book, Volker Gruhn (Hrsg.): Softwareengineering 2006
- P-80 Mareike Schoop, Christian Huemer, Michael Rebstock, Martin Bichler (Hrsg.): Service-Oriented Electronic Commerce
- P-81 Wolfgang Karl, Jürgen Becker, Karl-Erwin Großpietsch, Christian Hochberger, Erik Machle (Hrsg.): ARCS'06
- P-82 Heinrich C. Mayr, Ruth Breu (Hrsg.): Modellierung 2006
- P-83 Daniel Huson, Oliver Kohlbacher, Andrei Lupas, Kay Nieselt and Andreas Zell (eds.): German Conference on Bioinformatics
- P-84 Dimitris Karagiannis, Heinrich C. Mayr, (Hrsg.): Information Systems Technology and its Applications
- P-85 Witold Abramowicz, Heinrich C. Mayr, (Hrsg.): Business Information Systems
- P-86 Robert Krimmer (Ed.): Electronic Voting 2006
- P-87 Max Mühlhäuser, Guido Rößling, Ralf Steinmetz (Hrsg.): DELFI 2006: 4. e-Learning Fachtagung Informatik
- P-88 Robert Hirschfeld, Andreas Polze, Ryszard Kowalczyk (Hrsg.): NODe 2006, GSEM 2006
- P-90 Joachim Schelp, Robert Winter, Ulrich Frank, Bodo Rieger, Klaus Turowski (Hrsg.): Integration, Informationslogistik und Architektur
- P-91 Henrik Stormer, Andreas Meier, Michael Schumacher (Eds.): European Conference on eHealth 2006
- P-92 Fernand Feltz, Benoît Otjacques, Andreas Oberweis, Nicolas Poussing (Eds.): AIM 2006
- P-93 Christian Hochberger, Rüdiger Liskowsky (Eds.): INFORMATIK 2006 – Informatik für Menschen, Band 1
- P-94 Christian Hochberger, Rüdiger Liskowsky (Eds.): INFORMATIK 2006 – Informatik für Menschen, Band 2
- P-95 Matthias Weske, Markus Nüttgens (Eds.): EMISA 2005: Methoden, Konzepte und Technologien für die Entwicklung von dienstbasierten Informationssystemen
- P-96 Saartje Brockmans, Jürgen Jung, York Sure (Eds.): Meta-Modelling and Ontologies
- P-97 Oliver Göbel, Dirk Schadt, Sandra Frings, Hardo Hase, Detlef Günther, Jens Nedon (Eds.): IT-Incident Mangament & IT-Forensics – IMF 2006

- P-98 Hans Brandt-Pook, Werner Simonsmeier und Thorsten Spitta (Hrsg.): Beratung in der Softwareentwicklung – Modelle, Methoden, Best Practices
- P-99 Andreas Schwill, Carsten Schulte, Marco Thomas (Hrsg.): Didaktik der Informatik
- P-100 Peter Forbrig, Günter Siegel, Markus Schneider (Hrsg.): HDI 2006: Hochschuldidaktik der Informatik
- P-101 Stefan Böttinger, Ludwig Theuvsen, Susanne Rank, Marlies Morgenstern (Hrsg.): Agrarinformatik im Spannungsfeld zwischen Regionalisierung und globalen Wertschöpfungsketten
- P-102 Otto Spaniol (Eds.): Mobile Services and Personalized Environments
- P-103 Alfons Kemper, Harald Schöning, Thomas Rose, Matthias Jarke, Thomas Seidl, Christoph Quix, Christoph Brochhaus (Hrsg.): Datenbanksysteme in Business, Technologie und Web (BTW 2007)
- P-104 Birgitta König-Ries, Franz Lehner, Rainer Malaka, Can Türker (Hrsg.) MMS 2007: Mobilität und mobile Informationssysteme
- P-105 Wolf-Gideon Bleek, Jörg Raasch, Heinz Züllighoven (Hrsg.) Software Engineering 2007
- P-106 Wolf-Gideon Bleek, Henning Schwentner, Heinz Züllighoven (Hrsg.) Software Engineering 2007 – Beiträge zu den Workshops
- P-107 Heinrich C. Mayr, Dimitris Karagiannis (eds.) Information Systems Technology and its Applications
- P-108 Arslan Brömme, Christoph Busch, Detlef Hühnlein (eds.) BIOSIG 2007: Biometrics and Electronic Signatures
- P-109 Rainer Koschke, Otthein Herzog, Karl-Heinz Rödiger, Marc Ronthaler (Hrsg.) INFORMATIK 2007 Informatik trifft Logistik Band 1
- P-110 Rainer Koschke, Otthein Herzog, Karl-Heinz Rödiger, Marc Ronthaler (Hrsg.) INFORMATIK 2007 Informatik trifft Logistik Band 2
- P-111 Christian Eibl, Johannes Magenheimer, Sigrid Schubert, Martin Wessner (Hrsg.) DeLFI 2007: 5. e-Learning Fachtagung Informatik
- P-112 Sigrid Schubert (Hrsg.) Didaktik der Informatik in Theorie und Praxis
- P-113 Sören Auer, Christian Bizer, Claudia Müller, Anna V. Zhdanova (Eds.) The Social Semantic Web 2007 Proceedings of the 1<sup>st</sup> Conference on Social Semantic Web (CSSW)
- P-114 Sandra Frings, Oliver Göbel, Detlef Günther, Hardo G. Hase, Jens Nedon, Dirk Schadt, Arslan Brömme (Eds.) IMF2007 IT-incident management & IT-forensics Proceedings of the 3<sup>rd</sup> International Conference on IT-Incident Management & IT-Forensics
- P-115 Claudia Falter, Alexander Schliep, Joachim Selbig, Martin Vingron and Dirk Walther (Eds.) German conference on bioinformatics GCB 2007
- P-116 Witold Abramowicz, Leszek Maciszek (Eds.) Business Process and Services Computing 1<sup>st</sup> International Working Conference on Business Process and Services Computing BPSC 2007
- P-117 Ryszard Kowalczyk (Ed.) Grid service engineering and management The 4<sup>th</sup> International Conference on Grid Service Engineering and Management GSEM 2007
- P-118 Andreas Hein, Wilfried Thoben, Hans-Jürgen Appelrath, Peter Jensch (Eds.) European Conference on ehealth 2007
- P-119 Manfred Reichert, Stefan Strecker, Klaus Turowski (Eds.) Enterprise Modelling and Information Systems Architectures Concepts and Applications
- P-120 Adam Pawlak, Kurt Sandkuhl, Wojciech Cholewa, Leandro Soares Indrusiak (Eds.) Coordination of Collaborative Engineering - State of the Art and Future Challenges
- P-121 Korbinian Herrmann, Bernd Bruegge (Hrsg.) Software Engineering 2008 Fachtagung des GI-Fachbereichs Softwaretechnik
- P-122 Walid Maalej, Bernd Bruegge (Hrsg.) Software Engineering 2008 - Workshopband Fachtagung des GI-Fachbereichs Softwaretechnik

- P-123 Michael H. Breitner, Martin Breunig, Elgar Fleisch, Ley Pousttchi, Klaus Turowski (Hrsg.)  
Mobile und Ubiquitäre Informationssysteme – Technologien, Prozesse, Marktfähigkeit  
Proceedings zur 3. Konferenz Mobile und Ubiquitäre Informationssysteme (MMS 2008)
- P-124 Wolfgang E. Nagel, Rolf Hoffmann, Andreas Koch (Eds.)  
9<sup>th</sup> Workshop on Parallel Systems and Algorithms (PASA)  
Workshop of the GI/ITG Special Interest Groups PARS and PARVA
- P-125 Rolf A.E. Müller, Hans-H. Sundermeier, Ludwig Theuvsen, Stephanie Schütze, Marlies Morgenstern (Hrsg.)  
Unternehmens-IT:  
Führungsinstrument oder Verwaltungsbürde  
Referate der 28. GIL Jahrestagung
- P-126 Rainer Gimnich, Uwe Kaiser, Jochen Quante, Andreas Winter (Hrsg.)  
10<sup>th</sup> Workshop Software Reengineering (WSR 2008)
- P-127 Thomas Kühne, Wolfgang Reisig, Friedrich Steimann (Hrsg.)  
Modellierung 2008
- P-128 Ammar Alkassar, Jörg Siekmann (Hrsg.)  
Sicherheit 2008  
Sicherheit, Schutz und Zuverlässigkeit  
Beiträge der 4. Jahrestagung des Fachbereichs Sicherheit der Gesellschaft für Informatik e.V. (GI)  
2.-4. April 2008  
Saarbrücken, Germany
- P-129 Wolfgang Hesse, Andreas Oberweis (Eds.)  
Sigsand-Europe 2008  
Proceedings of the Third AIS SIGSAND European Symposium on Analysis, Design, Use and Societal Impact of Information Systems
- P-130 Paul Müller, Bernhard Neumair, Gabi Dreö Rodosek (Hrsg.)  
1. DFN-Forum Kommunikationstechnologien Beiträge der Fachtagung
- P-131 Robert Krimmer, Rüdiger Grimm (Eds.)  
3<sup>rd</sup> International Conference on Electronic Voting 2008  
Co-organized by Council of Europe, Gesellschaft für Informatik und E-Voting.  
CC
- P-132 Silke Seehusen, Ulrike Lucke, Stefan Fischer (Hrsg.)  
DeLFI 2008:  
Die 6. e-Learning Fachtagung Informatik
- P-133 Heinz-Gerd Hegering, Axel Lehmann, Hans Jürgen Ohlbach, Christian Scheideler (Hrsg.)  
INFORMATIK 2008  
Beherrschbare Systeme – dank Informatik Band 1
- P-134 Heinz-Gerd Hegering, Axel Lehmann, Hans Jürgen Ohlbach, Christian Scheideler (Hrsg.)  
INFORMATIK 2008  
Beherrschbare Systeme – dank Informatik Band 2
- P-135 Torsten Brinda, Michael Fothe, Peter Hubwieser, Kirsten Schlüter (Hrsg.)  
Didaktik der Informatik –  
Aktuelle Forschungsergebnisse
- P-136 Andreas Beyer, Michael Schroeder (Eds.)  
German Conference on Bioinformatics GCB 2008
- P-137 Arslan Brömme, Christoph Busch, Detlef Hühnlein (Eds.)  
BIOSIG 2008: Biometrics and Electronic Signatures
- P-138 Barbara Dinter, Robert Winter, Peter Chamoni, Norbert Gronau, Klaus Turowski (Hrsg.)  
Synergien durch Integration und Informationslogistik  
Proceedings zur DW2008
- P-139 Georg Herzwurm, Martin Mikusz (Hrsg.)  
Industrialisierung des Software-Managements  
Fachtagung des GI-Fachausschusses Management der Anwendungsentwicklung und -wartung im Fachbereich Wirtschaftsinformatik
- P-140 Oliver Göbel, Sandra Frings, Detlef Günther, Jens Nedon, Dirk Schadt (Eds.)  
IMF 2008 - IT Incident Management & IT Forensics
- P-141 Peter Loos, Markus Nüttgens, Klaus Turowski, Dirk Werth (Hrsg.)  
Modellierung betrieblicher Informationssysteme (MobIS 2008)  
Modellierung zwischen SOA und Compliance Management
- P-142 R. Bill, P. Korduan, L. Theuvsen, M. Morgenstern (Hrsg.)  
Anforderungen an die Agrarinformatik durch Globalisierung und Klimaveränderung
- P-143 Peter Liggesmeyer, Gregor Engels, Jürgen Münch, Jörg Dörr, Norman Riegel (Hrsg.)  
Software Engineering 2009  
Fachtagung des GI-Fachbereichs Softwaretechnik

- P-144 Johann-Christoph Freytag, Thomas Ruf, Wolfgang Lehner, Gottfried Vossen (Hrsg.)  
Datenbanksysteme in Business, Technologie und Web (BTW)
- P-145 Knut Hinkelmann, Holger Wache (Eds.)  
WM2009: 5th Conference on Professional Knowledge Management
- P-146 Markus Bick, Martin Breunig, Hagen Höpfner (Hrsg.)  
Mobile und Ubiquitäre Informationssysteme – Entwicklung, Implementierung und Anwendung  
4. Konferenz Mobile und Ubiquitäre Informationssysteme (MMS 2009)
- P-147 Witold Abramowicz, Leszek Maciaszek, Ryszard Kowalczyk, Andreas Speck (Eds.)  
Business Process, Services Computing and Intelligent Service Management  
BPSC 2009 · ISM 2009 · YRW-MBP 2009
- P-148 Christian Erfurth, Gerald Eichler, Volkmar Schau (Eds.)  
9<sup>th</sup> International Conference on Innovative Internet Community Systems  
I<sup>2</sup>CS 2009
- P-149 Paul Müller, Bernhard Neumair, Gabi Dreo Rodosek (Hrsg.)  
2. DFN-Forum  
Kommunikationstechnologien  
Beiträge der Fachtagung
- P-150 Jürgen Münch, Peter Liggesmeyer (Hrsg.)  
Software Engineering  
2009 - Workshopband
- P-151 Armin Heinzl, Peter Dadam, Stefan Kirn, Peter Lockemann (Eds.)  
PRIMIUM  
Process Innovation for Enterprise Software
- P-152 Jan Mendling, Stefanie Rinderle-Ma, Werner Esswein (Eds.)  
Enterprise Modelling and Information Systems Architectures  
Proceedings of the 3<sup>rd</sup> Int'l Workshop EMISA 2009
- P-153 Andreas Schwill, Nicolas Apostolopoulos (Hrsg.)  
Lernen im Digitalen Zeitalter  
DeLFI 2009 – Die 7. E-Learning Fachtagung Informatik
- P-154 Stefan Fischer, Erik Maehle, Rüdiger Reischuk (Hrsg.)  
INFORMATIK 2009  
Im Focus das Leben
- P-155 Arslan Brömme, Christoph Busch, Detlef Hühnlein (Eds.)  
BIOSIG 2009:  
Biometrics and Electronic Signatures  
Proceedings of the Special Interest Group on Biometrics and Electronic Signatures
- P-156 Bernhard Koerber (Hrsg.)  
Zukunft braucht Herkunft  
25 Jahre »INFOS – Informatik und Schule«
- P-157 Ivo Grosse, Steffen Neumann, Stefan Posch, Falk Schreiber, Peter Stadler (Eds.)  
German Conference on Bioinformatics 2009
- P-158 W. Claupein, L. Theuvsen, A. Kämpf, M. Morgenstern (Hrsg.)  
Precision Agriculture  
Reloaded – Informationsgestützte Landwirtschaft
- P-159 Gregor Engels, Markus Luckey, Wilhelm Schäfer (Hrsg.)  
Software Engineering 2010
- P-160 Gregor Engels, Markus Luckey, Alexander Pretschner, Ralf Reussner (Hrsg.)  
Software Engineering 2010 –  
Workshopband  
(inkl. Doktorandensymposium)
- P-161 Gregor Engels, Dimitris Karagiannis, Heinrich C. Mayr (Hrsg.)  
Modellierung 2010
- P-162 Maria A. Wimmer, Uwe Brinkhoff, Siegfried Kaiser, Dagmar Lück-Schneider, Erich Schweighofer, Andreas Wiebe (Hrsg.)  
Vernetzte IT für einen effektiven Staat  
Gemeinsame Fachtagung  
Verwaltungsinformatik (FTVI) und  
Fachtagung Rechtsinformatik (FTRI) 2010
- P-163 Markus Bick, Stefan Eulgem, Elgar Fleisch, J. Felix Hampe, Birgitta König-Ries, Franz Lehner, Key Pousttchi, Kai Rannenber (Hrsg.)  
Mobile und Ubiquitäre Informationssysteme  
Technologien, Anwendungen und Dienste zur Unterstützung von mobiler Kollaboration
- P-164 Arslan Brömme, Christoph Busch (Eds.)  
BIOSIG 2010: Biometrics and Electronic Signatures  
Proceedings of the Special Interest Group on Biometrics and Electronic Signatures

- P-165 Gerald Eichler, Peter Kropf, Ulrike Lechner, Phayung Meesad, Herwig Unger (Eds.)  
10<sup>th</sup> International Conference on Innovative Internet Community Systems (I<sup>2</sup>CS) – Jubilee Edition 2010 –
- P-166 Paul Müller, Bernhard Neumair, Gabi Dreo Rodosek (Hrsg.)  
3. DFN-Forum Kommunikationstechnologien Beiträge der Fachtagung
- P-167 Robert Krimmer, Rüdiger Grimm (Eds.)  
4<sup>th</sup> International Conference on Electronic Voting 2010  
co-organized by the Council of Europe, Gesellschaft für Informatik and E-Voting.CC
- P-168 Ira Diethelm, Christina Dörge, Claudia Hildebrandt, Carsten Schulte (Hrsg.)  
Didaktik der Informatik  
Möglichkeiten empirischer Forschungsmethoden und Perspektiven der Fachdidaktik
- P-169 Michael Keres, Nadine Ojstersek, Ulrik Schroeder, Ulrich Hoppe (Hrsg.)  
DeLFI 2010 - 8. Tagung der Fachgruppe E-Learning der Gesellschaft für Informatik e.V.
- P-170 Felix C. Freiling (Hrsg.)  
Sicherheit 2010  
Sicherheit, Schutz und Zuverlässigkeit
- P-171 Werner Esswein, Klaus Turowski, Martin Juhrisch (Hrsg.)  
Modellierung betrieblicher Informationssysteme (MobIS 2010)  
Modellgestütztes Management
- P-172 Stefan Klink, Agnes Koschmider, Marco Mevius, Andreas Oberweis (Hrsg.)  
EMISA 2010  
Einflussfaktoren auf die Entwicklung flexibler, integrierter Informationssysteme  
Beiträge des Workshops der GI-Fachgruppe EMISA (Entwicklungsmethoden für Informationssysteme und deren Anwendung)
- P-173 Dietmar Schomburg, Andreas Grote (Eds.)  
German Conference on Bioinformatics 2010
- P-174 Arslan Brömme, Torsten Eymann, Detlef Hühnlein, Heiko Roßnagel, Paul Schmücker (Hrsg.)  
perspeGktive 2010  
Workshop „Innovative und sichere Informationstechnologie für das Gesundheitswesen von morgen“
- P-175 Klaus-Peter Fähnrich, Bogdan Franczyk (Hrsg.)  
INFORMATIK 2010  
Service Science – Neue Perspektiven für die Informatik  
Band 1
- P-176 Klaus-Peter Fähnrich, Bogdan Franczyk (Hrsg.)  
INFORMATIK 2010  
Service Science – Neue Perspektiven für die Informatik  
Band 2
- P-177 Witold Abramowicz, Rainer Alt, Klaus-Peter Fähnrich, Bogdan Franczyk, Leszek A. Maciaszek (Eds.)  
INFORMATIK 2010  
Business Process and Service Science – Proceedings of ISSS and BPSC
- P-178 Wolfram Pietsch, Benedikt Krams (Hrsg.)  
Vom Projekt zum Produkt  
Fachtagung des GI-Fachausschusses Management der Anwendungsentwicklung und -wartung im Fachbereich Wirtschaftsinformatik (WI-MAW), Aachen, 2010
- P-179 Stefan Gruner, Bernhard Rumpe (Eds.)  
FM+AM'2010  
Second International Workshop on Formal Methods and Agile Methods
- P-180 Theo Härder, Wolfgang Lehner, Bernhard Mitschang, Harald Schöning, Holger Schwarz (Hrsg.)  
Datenbanksysteme für Business, Technologie und Web (BTW)  
14. Fachtagung des GI-Fachbereichs „Datenbanken und Informationssysteme“ (DBIS)
- P-181 Michael Clasen, Otto Schätzel, Brigitte Theuvsen (Hrsg.)  
Qualität und Effizienz durch informationsgestützte Landwirtschaft, Fokus: Moderne Weinwirtschaft
- P-182 Ronald Maier (Hrsg.)  
6<sup>th</sup> Conference on Professional Knowledge Management  
From Knowledge to Action
- P-183 Ralf Reussner, Matthias Grund, Andreas Oberweis, Walter Tichy (Hrsg.)  
Software Engineering 2011  
Fachtagung des GI-Fachbereichs Softwaretechnik
- P-184 Ralf Reussner, Alexander Pretschner, Stefan Jähnichen (Hrsg.)  
Software Engineering 2011  
Workshopband  
(inkl. Doktorandensymposium)

- P-185 Hagen Höpfner, Günther Specht, Thomas Ritz, Christian Bunse (Hrsg.)  
MMS 2011: Mobile und ubiquitäre Informationssysteme Proceedings zur 6. Konferenz Mobile und Ubiquitäre Informationssysteme (MMS 2011)
- P-186 Gerald Eichler, Axel Küpper, Volkmar Schau, Hacène Fouchal, Herwig Unger (Eds.)  
11<sup>th</sup> International Conference on Innovative Internet Community Systems (I<sup>2</sup>CS)
- P-187 Paul Müller, Bernhard Neumair, Gabi Dreo Rodosek (Hrsg.)  
4. DFN-Forum Kommunikationstechnologien, Beiträge der Fachtagung 20. Juni bis 21. Juni 2011 Bonn
- P-188 Holger Rohland, Andrea Kienle, Steffen Friedrich (Hrsg.)  
DeLFI 2011 – Die 9. e-Learning Fachtagung Informatik der Gesellschaft für Informatik e.V. 5.–8. September 2011, Dresden
- P-189 Thomas, Marco (Hrsg.)  
Informatik in Bildung und Beruf INFOS 2011  
14. GI-Fachtagung Informatik und Schule
- P-190 Markus Nüttgens, Oliver Thomas, Barbara Weber (Eds.)  
Enterprise Modelling and Information Systems Architectures (EMISA 2011)
- P-191 Arslan Brömme, Christoph Busch (Eds.)  
BIOSIG 2011  
International Conference of the Biometrics Special Interest Group
- P-192 Hans-Ulrich Heiß, Peter Pepper, Holger Schlingloff, Jörg Schneider (Hrsg.)  
INFORMATIK 2011  
Informatik schafft Communities
- P-193 Wolfgang Lehner, Gunther Piller (Hrsg.)  
IMDM 2011
- P-194 M. Clasen, G. Fröhlich, H. Bernhardt, K. Hildebrand, B. Theuvsen (Hrsg.)  
Informationstechnologie für eine nachhaltige Landbewirtschaftung  
Fokus Forstwirtschaft
- P-195 Neeraj Suri, Michael Waidner (Hrsg.)  
Sicherheit 2012  
Sicherheit, Schutz und Zuverlässigkeit  
Beiträge der 6. Jahrestagung des Fachbereichs Sicherheit der Gesellschaft für Informatik e.V. (GI)
- P-196 Arslan Brömme, Christoph Busch (Eds.)  
BIOSIG 2012  
Proceedings of the 11<sup>th</sup> International Conference of the Biometrics Special Interest Group
- P-197 Jörn von Lucke, Christian P. Geiger, Siegfried Kaiser, Erich Schweighofer, Maria A. Wimmer (Hrsg.)  
Auf dem Weg zu einer offenen, smarten und vernetzten Verwaltungskultur  
Gemeinsame Fachtagung Verwaltungsinformatik (FTVI) und Fachtagung Rechtsinformatik (FTRI) 2012
- P-198 Stefan Jähnichen, Axel Küpper, Sahin Albayrak (Hrsg.)  
Software Engineering 2012  
Fachtagung des GI-Fachbereichs Softwaretechnik
- P-199 Stefan Jähnichen, Bernhard Rumpe, Holger Schlingloff (Hrsg.)  
Software Engineering 2012  
Workshopband
- P-200 Gero Mühl, Jan Richling, Andreas Herkersdorf (Hrsg.)  
ARCS 2012 Workshops
- P-201 Elmar J. Sinz Andy Schür (Hrsg.)  
Modellierung 2012
- P-202 Andrea Back, Markus Bick, Martin Breunig, Key Pousttchi, Frédéric Thiesse (Hrsg.)  
MMS 2012: Mobile und Ubiquitäre Informationssysteme
- P-203 Paul Müller, Bernhard Neumair, Helmut Reiser, Gabi Dreo Rodosek (Hrsg.)  
5. DFN-Forum Kommunikationstechnologien  
Beiträge der Fachtagung
- P-204 Gerald Eichler, Leendert W. M. Wienhofen, Anders Kofod-Petersen, Herwig Unger (Eds.)  
12<sup>th</sup> International Conference on Innovative Internet Community Systems (I2CS 2012)
- P-205 Manuel J. Kripp, Melanie Volkamer, Rüdiger Grimm (Eds.)  
5<sup>th</sup> International Conference on Electronic Voting 2012 (EVOTE2012)  
Co-organized by the Council of Europe, Gesellschaft für Informatik und E-Voting.CC
- P-206 Stefanie Rinderle-Ma, Mathias Weske (Hrsg.)  
EMISA 2012  
Der Mensch im Zentrum der Modellierung
- P-207 Jörg Desel, Jörg M. Haake, Christian Spannagel (Hrsg.)  
DeLFI 2012: Die 10. e-Learning Fachtagung Informatik der Gesellschaft für Informatik e.V.  
24.–26. September 2012

- P-208 Ursula Goltz, Marcus Magnor, Hans-Jürgen Appelrath, Herbert Matthies, Wolf-Tilo Balke, Lars Wolf (Hrsg.)  
INFORMATIK 2012
- P-209 Hans Brandt-Pook, André Fleer, Thorsten Spitta, Malte Wattenberg (Hrsg.)  
Nachhaltiges Software Management
- P-210 Erhard Plödereder, Peter Dencker, Herbert Klenk, Hubert B. Keller, Silke Spitzer (Hrsg.)  
Automotive – Safety & Security 2012  
Sicherheit und Zuverlässigkeit für automobile Informationstechnik
- P-211 M. Clasen, K. C. Kersebaum, A. Meyer-Aurich, B. Theuvsen (Hrsg.)  
Massendatenmanagement in der Agrar- und Ernährungswirtschaft  
Erhebung - Verarbeitung - Nutzung  
Referate der 33. GIL-Jahrestagung  
20. – 21. Februar 2013, Potsdam
- P-212 Arslan Brömme, Christoph Busch (Eds.)  
BIOSIG 2013  
Proceedings of the 12th International Conference of the Biometrics Special Interest Group  
04.–06. September 2013  
Darmstadt, Germany
- P-213 Stefan Kowalewski, Bernhard Rumpel (Hrsg.)  
Software Engineering 2013  
Fachtagung des GI-Fachbereichs Softwaretechnik
- P-214 Volker Markl, Gunter Saake, Kai-Uwe Sattler, Gregor Hackenbroich, Bernhard Mitschang, Theo Härder, Veit Köppen (Hrsg.)  
Datenbanksysteme für Business, Technologie und Web (BTW) 2013  
13. – 15. März 2013, Magdeburg
- P-215 Stefan Wagner, Horst Lichter (Hrsg.)  
Software Engineering 2013  
Workshopband  
(inkl. Doktorandensymposium)  
26. Februar – 1. März 2013, Aachen
- P-216 Gunter Saake, Andreas Henrich, Wolfgang Lehner, Thomas Neumann, Veit Köppen (Hrsg.)  
Datenbanksysteme für Business, Technologie und Web (BTW) 2013 – Workshopband  
11. – 12. März 2013, Magdeburg
- P-217 Paul Müller, Bernhard Neumair, Helmut Reiser, Gabi Dreö Rodosek (Hrsg.)  
6. DFN-Forum Kommunikationstechnologien  
Beiträge der Fachtagung  
03.–04. Juni 2013, Erlangen
- P-218 Andreas Breiter, Christoph Rensing (Hrsg.)  
DeLFI 2013: Die 11 e-Learning Fachtagung Informatik der Gesellschaft für Informatik e.V. (GI)  
8. – 11. September 2013, Bremen
- P-219 Norbert Breier, Peer Stechert, Thomas Wilke (Hrsg.)  
Informatik erweitert Horizonte  
INFOS 2013  
15. GI-Fachtagung Informatik und Schule  
26. – 28. September 2013
- P-220 Matthias Horbach (Hrsg.)  
INFORMATIK 2013  
Informatik angepasst an Mensch, Organisation und Umwelt  
16. – 20. September 2013, Koblenz
- P-221 Maria A. Wimmer, Marijn Janssen, Ann Macintosh, Hans Jochen Scholl, Efthimos Tambouris (Eds.)  
Electronic Government and Electronic Participation  
Joint Proceedings of Ongoing Research of IFIP EGOV and IFIP ePart 2013  
16. – 19. September 2013, Koblenz
- P-222 Reinhard Jung, Manfred Reichert (Eds.)  
Enterprise Modelling and Information Systems Architectures (EMISA 2013)  
St. Gallen, Switzerland  
September 5. – 6. 2013
- P-223 Detlef Hühnlein, Heiko Roßnagel (Hrsg.)  
Open Identity Summit 2013  
10. – 11. September 2013  
Kloster Banz, Germany
- P-224 Eckhart Hanser, Martin Mikusz, Masud Fazal-Baqaie (Hrsg.)  
Vorgehensmodelle 2013  
Vorgehensmodelle – Anspruch und Wirklichkeit  
20. Tagung der Fachgruppe Vorgehensmodelle im Fachgebiet Wirtschaftsinformatik (WI-VM) der Gesellschaft für Informatik e.V.  
Lörrach, 2013
- P-225 Hans-Georg Fill, Dimitris Karagiannis, Ulrich Reimer (Hrsg.)  
Modellierung 2014  
19. – 21. März 2014, Wien
- P-226 M. Clasen, M. Hamer, S. Lehnert, B. Petersen, B. Theuvsen (Hrsg.)  
IT-Standards in der Agrar- und Ernährungswirtschaft Fokus: Risiko- und Krisenmanagement  
Referate der 34. GIL-Jahrestagung  
24. – 25. Februar 2014, Bonn

- P-227 Wilhelm Hasselbring,  
Nils Christian Ehmke (Hrsg.)  
Software Engineering 2014  
Fachtagung des GI-Fachbereichs  
Softwaretechnik  
25. – 28. Februar 2014  
Kiel, Deutschland
- P-228 Stefan Katzenbeisser, Volkmar Lotz,  
Edgar Weippl (Hrsg.)  
Sicherheit 2014  
Sicherheit, Schutz und Zuverlässigkeit  
Beiträge der 7. Jahrestagung des  
Fachbereichs Sicherheit der  
Gesellschaft für Informatik e.V. (GI)  
19. – 21. März 2014, Wien
- P-230 Arslan Brömme, Christoph Busch (Eds.)  
BIOSIG 2014  
Proceedings of the 13<sup>th</sup> International  
Conference of the Biometrics Special  
Interest Group  
10. – 12. September 2014 in  
Darmstadt, Germany
- P-231 Paul Müller, Bernhard Neumair,  
Helmut Reiser, Gabi Dreo Rodosek  
(Hrsg.)  
7. DFN-Forum  
Kommunikationstechnologien  
16. – 17. Juni 2014  
Fulda
- P-232 E. Plödereder, L. Grunske, E. Schneider,  
D. Ull (Hrsg.)  
INFORMATIK 2014  
Big Data – Komplexität meistern  
22. – 26. September 2014  
Stuttgart
- P-233 Stephan Trahasch, Rolf Plötzner, Gerhard  
Schneider, Claudia Gayer, Daniel Sassiati,  
Nicole Wöhrle (Hrsg.)  
DeLFI 2014 – Die 12. e-Learning  
Fachtagung Informatik  
der Gesellschaft für Informatik e.V.  
15. – 17. September 2014  
Freiburg
- P-234 Fernand Feltz, Bela Mutschler, Benoît  
Otjacques (Eds.)  
Enterprise Modelling and Information  
Systems Architectures  
(EMISA 2014)  
Luxembourg, September 25-26, 2014
- P-235 Robert Giegerich,  
Ralf Hofestädt,  
Tim W. Nattkemper (Eds.)  
German Conference on  
Bioinformatics 2014  
September 28 – October 1  
Bielefeld, Germany
- P-236 Martin Engstler, Eckhart Hanser,  
Martin Mikusz, Georg Herzwurm (Hrsg.)  
Projektmanagement und  
Vorgehensmodelle 2014  
Soziale Aspekte und Standardisierung  
Gemeinsame Tagung der Fachgruppen  
Projektmanagement (WI-PM) und  
Vorgehensmodelle (WI-VM) im  
Fachgebiet Wirtschaftsinformatik der  
Gesellschaft für Informatik e.V., Stuttgart  
2014
- P-237 Detlef Hühnlein, Heiko Roßnagel (Hrsg.)  
Open Identity Summit 2014  
4.–6. November 2014  
Stuttgart, Germany
- P-238 Arno Ruckelshausen, Hans-Peter  
Schwarz, Brigitte Theuvsen (Hrsg.)  
Informatik in der Land-, Forst- und  
Ernährungswirtschaft  
Referate der 35. GIL-Jahrestagung  
23. – 24. Februar 2015, Geisenheim
- P-239 Uwe Aßmann, Birgit Demuth, Thorsten  
Spitta, Georg Püschel, Ronny Kaiser  
(Hrsg.)  
Software Engineering & Management  
2015  
17.-20. März 2015, Dresden
- P-240 Herbert Klenk, Hubert B. Keller, Erhard  
Plödereder, Peter Dencker (Hrsg.)  
Automotive – Safety & Security 2015  
Sicherheit und Zuverlässigkeit für  
automobile Informationstechnik  
21.–22. April 2015, Stuttgart
- P-241 Thomas Seidl, Norbert Ritter,  
Harald Schöning, Kai-Uwe Sattler,  
Theo Härder, Steffen Friedrich,  
Wolfram Wingerath (Hrsg.)  
Datenbanksysteme für Business,  
Technologie und Web (BTW 2015)  
04. – 06. März 2015, Hamburg
- P-242 Norbert Ritter, Andreas Henrich,  
Wolfgang Lehner, Andreas Thor,  
Steffen Friedrich, Wolfram Wingerath  
(Hrsg.)  
Datenbanksysteme für Business,  
Technologie und Web (BTW 2015) –  
Workshopband  
02. – 03. März 2015, Hamburg
- P-243 Paul Müller, Bernhard Neumair, Helmut  
Reiser, Gabi Dreo Rodosek (Hrsg.)  
8. DFN-Forum  
Kommunikationstechnologien  
06.–09. Juni 2015, Lübeck

- P-244 Alfred Zimmermann,  
Alexander Rossmann (Eds.)  
Digital Enterprise Computing  
(DEC 2015)  
Böblingen, Germany June 25-26, 2015
- P-245 Arslan Brömme, Christoph Busch ,  
Christian Rathgeb, Andreas Uhl (Eds.)  
BIOSIG 2015  
Proceedings of the 14th International  
Conference of the Biometrics Special  
Interest Group  
09.–11. September 2015  
Darmstadt, Germany
- P-246 Douglas W. Cunningham, Petra Hofstedt,  
Klaus Meer, Ingo Schmitt (Hrsg.)  
INFORMATIK 2015  
28.9.-2.10. 2015, Cottbus
- P-247 Hans Pongratz, Reinhard Keil (Hrsg.)  
DeLFI 2015 – Die 13. E-Learning  
Fachtagung Informatik der  
Gesellschaft für Informatik e.V. (GI)  
1.–4. September 2015  
München
- P-248 Jens Kolb, Henrik Leopold, Jan  
Mendling (Eds.)  
Enterprise Modelling and  
Information Systems Architectures  
Proceedings of the 6th Int. Workshop  
on Enterprise Modelling and  
Information Systems Architectures,  
Innsbruck, Austria  
September 3-4, 2015
- P-249 Jens Gallenbacher (Hrsg.)  
Informatik  
allgemeinbildend begreifen  
INFOS 2015 16. GI-Fachtagung  
Informatik und Schule  
20.–23. September 2015

The titles can be purchased at:

**Köllen Druck + Verlag GmbH**

Ernst-Robert-Curtius-Str. 14 · D-53117 Bonn

Fax: +49 (0)228/9898222

E-Mail: [druckverlag@koellen.de](mailto:druckverlag@koellen.de)

