Challenges for the Implementation and Revision of International Biometric Standards Demonstrated by the Example of Face Image Data

Peter Ebinger and Margarida Castro Neves
Security Technology Dept., Fraunhofer Institute for Computer Graphics Research IGD
Fraunhoferstr. 5, 64283 Darmstadt, Germany
{peter.ebinger|margarida.castro-neves}@igd.fraunhofer.de

René Salamon and Oliver Bausinger
Federal Office for Information Security (BSI)
Godesberger Allee 185-189, 53175 Bonn, Germany
{rene.salamon|oliver.bausinger}@bsi.bund.de

Abstract: Travel documents such as the electronic passport (ePass) ensure that each person can be uniquely identified by a single document. The development of new ePass security chip technologies allows for the inclusion of biometric properties in the data carrier of the ePass. The International Civil Aviation Organization (ICAO) has determined a personal photograph as being the interoperable feature for all global travel documents; ICAO [Gro04] regulations reference quality requirements for facial images as defined in ISO standard ISO/IEC 19794-5 [Intb]. Project FIReBIRDs goal is to prepare an international facial image database for conformity tests based on ISO/IEC 19794-5 [Intb], to analyze the requirements in the regulating documents, and to develop suggestions for adaptations and extensions of these standards.

The FIReBIRD database shall provide a well-defined ground truth for level 3 conformance testing. For this purpose the specifications in the standard were thoroughly analyzed and in some parts refined to allow for a precise definition of ground truth. We show with two examples that there might be a defined common-sense definition for some parameters, but they are not measurable and their specification is not scientifically founded: the definition of full frontal view and the definition of eye and hair colors. Our results show that specifications and requirements should always be checked for necessity, practicability and usability and that a continued review and revision of biometric standards is necessary.

1 Standards and Guidelines: History and Development

An identity or travel document such as the electronic passport (ePass) ensures that a person can be clearly tied to the document and be verified (in the traditional sense) by sight check. Supplementing conventional optical security features with new digital security chips has allowed the inclusion of biometric information that more effectively binds each document with an individual, thus introducing an additional layer of identity verification.

The power of biometric verification with the gross capacity of available storage raises several questions: which biometric features should be included in passports, which features should be mandatory, which optional, how will interoperability be ensured? In which way should the biometric feature be stored in country A so that it can be used for verification at border control of country B? These questions illustrate the obvious need for international standardization.

For quite some time the international body responsible for passport standards has been the International Civil Aviation Organization (ICAO), a sub-organization of the United Nations. They specified the photograph and thus the biometric properties of the face as the globally interoperable feature to be stored on a chip, referring the activities of the working group for the international standardization in the field of biometrics[1]. For this aim the standard ISO/IEC 19794-5 Biometric data interchange formats - Part 5: Face image data [Intb] was produced and is now referenced by the ICAO [Gro04].

Many countries have adopted these specifications. In Germany for example, biometrically enhanced passports (called "ePass") based on these standards have been issued since 2005. Quality demands on facial images corresponding to [Gro04] or [Intb] are implemented by German authorities in [Bun, table 6]. Biometrically enhanced travel documents have become increasingly common in their use throughout the world and the reliability of the biometric mechanisms is no longer generally questioned. One question remains however: Are the defined standards and specifications applicable?

The following parts of the paper are organized as follows: Section 2 provides a description of standard implementation and conformance testing followed by a definition of ground truth. The objectives and the realization of the project FIReBIRD are outlined in section 3. In section 4 the general revision process of standards and subsequently related findings of FIReBIRD are presented. Lessons learned finalize the paper in section 5.

2 Standard Implementation and Conformance Testing

In this section we describe experiences with the implementation of the standards mentioned above and how the conformance of biometric systems that are based on them can be tested.

2.1 Standard Implementation

The implementation of these standards and their application revealed some issues that were not foreseen when originally released. Problems in everyday life (e.g. rejection of photographs by passport offices) have led to some workarounds. Photographers want to raise the acceptance rate of face images they produce in order to satisfy an increased

[1] This is the Subcommittee "Biometrics" of the Joint Technical Committee of ISO and IEC (ISO/IEC JTC1/SC37).

number of their customers. On ePass applications image properties are primarily checked by means of a sample photo table. Template and quality assurance (QA) software checks additional photographic image properties such as pose, head/image size, width-to-height ratio and photographic image properties. Experiences with the QA software show that a flawless assessment of facial images based on the current standard is not in every case possible.

Practical experiences with the use of biometric systems for face recognition have shown that certain image characteristics and scene properties such as pose variations have a considerable impact on the recognition performance. Other properties however are far less important for identification than previously suspected. This has shown the need for a scientifically based approach to address these issues.

2.2 Conformance Testing

Data produced by one biometric system or component should be able to be processed by systems or components from other vendors. ISO/IEC 19794-1 Information Technology – Biometric data interchange formats – Part 1: Framework [Inta] defines biometric data interchange formats to ensure vendor independency. Verification mechanisms are needed to check conformance claims of vendors regarding their biometric products. ISO/IEC 29109-1 – Conformance Testing Methodology for Biometric Interchange Records Format Part 1: Generalized conformance testing methodology [Intc] defines a methodology for testing the conformance for various parts of ISO/IEC 19794.

ISO/IEC 29109-1 defines three level of conformance testing:

1. *Level 1 Data Format Conformance* Field by field and byte by byte conformance checking with the specification, both in terms of fields included and the ranges of the values in those fields.

2. *Level 2 Internal Consistency Checking* Testing the internal consistency of the biometric data, relating values from one part or field of the data to values from other parts.

3. *Level 3 Content Checking* Testing that biometric data produced by a system is a faithful representation of the subject.

ISO/IEC 29109-1 [Intc] summarizes conformance requirements for each modality (in our example for face image data) in a requirement table based on ISO/IEC 19794. Products or implementations can comply with a subset of these requirements as some of them are defined as mandatory and others as optional. It is important that all requirements are precisely defined so that they are unambiguous and can be correctly implemented and followed. In particular for face images in travel documents it is important that standard conformance is achieved. Passports are used worldwide and there is a variety of face recognition products and vendors to choose from. Therefore, only clearly specified requirements and a sound standard enable suppliers of facial image processing software to implement standard conformant and interoperable systems.

Standards have no effect as long as they are not applied and used. Applying a standard means to implement systems that are conformant to the standard. Being conformant to a standard means that input data, output data and processes of the system are valid in respect to data format and content as defined by the standard.

The reliability of conformance tests depends on the comprehensibility of the testing scheme and the test data used. To generate conformance testing data, a measure for the classification of valid and invalid values regarding the standard is needed. This measure should allow to determine the degree of validity of a testing data record. For every property the standard deals with the following has to be clearly defined:

1. whether the testing data record is valid or invalid,
2. if the data record is valid: where inside the boundaries of validity the data record lies,
3. if the data record is invalid: where outside the boundaries of validity the data record lies.

Regarding test data it is absolutely necessary to know which part of it is valid and what part is invalid in respect to which part of the standard.

2.3 Ground Truth

The term ground truth was originally used in the analysis of aerial photographs and satellite imagery in which data are gathered at a distance with the objective to relate image data to real features and materials on the ground. In this context ground truth refers to reliable information that is collected "on location", in contrast to the information that is captured by remote sensing which has to be interpreted and categorized afterwards. Ground-truth data enables the calibration, training and evaluation of systems and algorithms for remote-sensing and interpretation and analysis collected data.

Although the term ground truth is commonly used in the field of biometrics (including related ISO standards) there is no exact definition by ISO yet. There have been some discussions to include the term ground truth into the Working Draft for the Standing Document 2, Harmonized Biometric Vocabulary [Inth]. The discussed definition pointed out that ground truth data should be captured by other means than the normally used measuring instrument. This way the measuring instrument can be validated with sufficient accuracy.

ISO/IEC 19795-2, Information technology – Biometric performance testing and reporting – Part 2: Testing methodologies for technology and scenario evaluation [Intg] distinguishes between ground truth for technology evaluations and for scenario evaluations. In the context of technology evaluations ground truth refers to known associations between data samples and source of samples whereas for scenario evaluations it is described as associations between system decisions and independently recorded sources of presented samples.

Definition In this context we define *ground truth* as *reliable biometric data* captured within a *defined setup and known parameters* which are *well documented* and available as metadata.

The provision of ground truth test data based on the international standards is of basic importance for performance tests and conformity tests of facial image data. To meet these requirements a pool of facial images containing both kinds of images – images that are valid and those invalid according to ISO/IEC 19794-5 [Intb] – is collected by the FIReBIRD project as described in the following section.

3 FIReBIRD: Project Aim and Realization

In 2008 the Facial Image Recognition Benchmark including Realistic Disturbances (FIReBIRD) [ENSS08] project was started by the Federal Office for Information Security BSI jointly with Fraunhofer IGD.

3.1 Project Aim

FIReBIRD aims at assuring and improving the quality of systems for processing facial images based on a scientifically grounded implementation of the standards mentioned in the above section (see Fig. 3.1). For this purpose a facial image database shall be created based on the requirements defined in the standards that can be used for conformity and performance tests of systems for processing facial images (e.g. systems for an automated face recognition and/or systems for quality assessment of facial images based on ISO/IEC 19794-5 [Intb]). Furthermore the requirements in the regulating documents shall be analyzed and suggestions be developed for adaptations and extensions of the standard.

A concept for the development of an internationally composed facial image database has been prepared. This concept has been internationally coordinated in cooperation with the US-American National Institute of Standards and Technology (NIST) and the British National Physical Laboratory (NPL).

3.2 Project Realization

The requirements defined in ISO/IEC 19794-5 [Intb] concerning photographic and phototechnical image properties are subject to a revision. Based on an analysis of the relevant

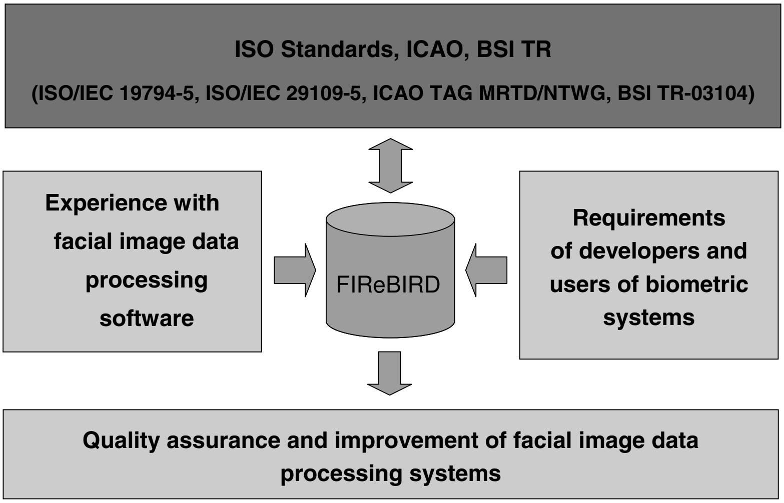

Figure 1: Interaction of FIReBIRD with biometric standards, developers and users of biometric systems

image properties the qualitative size of the image database is determined. A ground truth is defined by specifying and describing a precise and universally applicable reference value or a reference point for each property respectively.

Morphologic face features are also addressed and recorded as morphologic peculiarities (such as certain face or nose shapes, the mouth line, the special manifestation of the eye area, eye, skin and hair color and hair type) may have an impact on the performance of face recognition systems. The "semantics" of the face based on state-of-the-art anthropological and forensic findings are captured. Schematic representations are generated allowing a classification of faces. These facial properties shall be stored in the database as meta data for each facial image for each identity respectively.

4 Standard Revision

Most ISO standards require periodic revision because of technological evolution, new methods and materials and/or new quality and safety requirements. ISO has therefore established the general rule that all ISO standards should be reviewed at intervals of not more than five years.

Accordingly also the international standard ISO/IEC 19794-5 [Intb] is periodically revised and has undergone several extensions and modifications represented by the following documents:

- a first (informative) addendum [Intd] providing more detailed descriptions of photographic scenarios for taking suitable photos,
- a first technical corrigendum [Inte] correcting typos,
- a second technical corrigendum [Intf] considerably softening the tolerances for the criteria mentioned above.

One of the outcomes of FIReBIRD is to give some feedback on the revision process of related standards. For this purpose suggestions shall be made based on the analysis of the demands on facial images and the experiences with the verification of sovereign documents for inclusion in subsequent releases of ISO/IEC 19794-5 [Intb] and [Bun] (or related amendments to these documents respectively).

To allow an automated but reliable quality assurance of the presented facial images no image properties should be enforced that are not automatically measurable and/or calculable.

The following aspects were particularly affected:

- head size relative to the image size,
- width of the head relative to the image width,
- horizontal centering of the image, and
- roll angle (rotation about the horizontal back to front (z) axis).

FIReBIRD demonstrated that the true reason for these problems in the specification of requirements for facial images was the definition and the measurement of ground truth data and reference points (e.g. full frontal positioning of the head) for all relevant image properties. In particular "soft" properties – such as morphologic features, head positioning and color representation – are difficult to measure and therefore it is not easy to define quality requirements for them. The findings of FIReBIRD show that it is necessary to considerably extend tolerances where necessary and to limit the standards to measurable properties.

4.1 Refining the Definition of Ground Truth

Standards often deal with properties that are clearly defined, but also with properties that are hard to measure.

On the one side, for example, [Intb] specifies the relation of face height to width, or the range of valid widths and heights of images. These properties can be easily measured using a ruler or counting pixels: There is a well-defined reference/zero point, as well as a defined optimum value and there are well-defined algorithms to determine the values. For these properties there are generally accepted standards or established definitions and methodologies that define measurement and processing.

On the other side, the standard also defines properties that are not as easy to determine, e.g. the range of valid pitch of the head. In this case there is a defined common-sense optimum value, but there is no well-defined reference point and scale. However, without them it is impossible to exactly measure a property.

To make bad things worse the character of a test data record often cannot be completely described using properties specified in the standard. A facial image, for example, can be described more exactly if some basic information about the person shown on the image is given, e.g. age, hairdo, eye, hair or skin colors. However, most of these additional properties (as pitch of head mentioned above) do not have well-defined reference points and they are therefore hard to measure. However, for some of them this might not be obvious.

So how can the problem of measurability be solved? A practical approach: If there is no standardized measure, define and build one.

The following requirements should hold for a ground truth measure:

1. The measure has to be close to reality.
2. The measure has to be internationally reproducible.
3. The measure has to be applicable.
4. The reference points have to be clearly separated from each other.

Within FIReBIRD we applied the requirements described above to determine measures that can be applied to collect ground truth data for facial images.

4.2 Definition of Ground Truth for Eye, Hair and Skin Color

One of the challenges within FIReBIRD is to define measures for eye, hair and skin colors. Tables for eye and hair color referring to reality and scientific knowledge about the typical appearance of human beings reference are defined as described in the following.

ISO/IEC 19794-5 [Intb] specifies seven *eye colors*: “black”, “blue”, “brown”, “gray”, “green”, “multi-colored” and “pink” where “pink” probably refers to the eye color of people who have albinism. These seven classes may be sufficient if we have to describe eye colors for traveling purposes (in identity documents). In case of traveling or migration the border control staff is first of all interested in reliably verifying and distinguishing eye colors. When building a database such as the FIReBIRD database – for performance or conformance testing it is necessary to describe the classification and measuring equipment as precisely as possible. For this purposes we use a more comprehensive eye color classification scheme as it is defined in the ISO standard.

Our goal is a reliable and unambiguous classification of eye colors during the acquisition of facial images. Therefore eye colors are categorized into five color classes with three color depths each. To provide a realistic representation of eye colors to allow an easy and

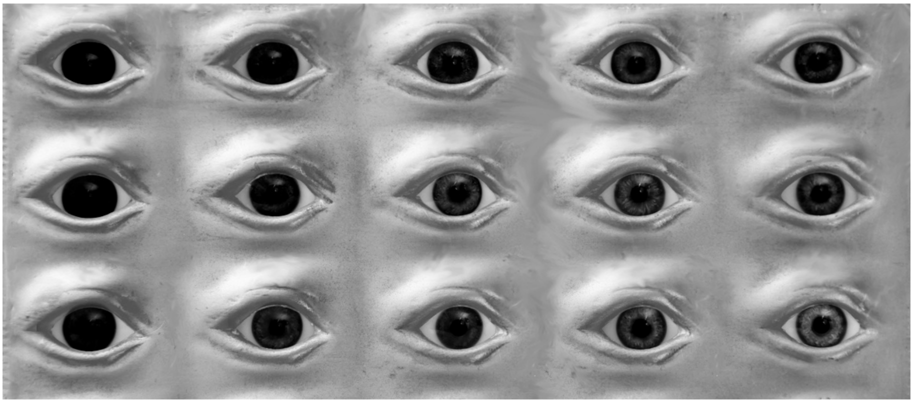

Figure 2: Box with glass eyes to categorize eye colors into five color classes (from left to right: brown, grey-brown, multiple, grey-blue and blue) with three color depths each (from top to bottom: dark, medium, bright)

precise classification by human operators these 15 colors are represented by glass eyes in a framed box covered with a plate of eyelids (see figure 4.2).

Using our extended eye color table we want to reliably classify the eye colors for more than 99% of the subjects. We hope that we can also demonstrate that an extended eye color table does not necessarily lead to an unreliable, ambiguous and/or slow classification process and that our color table is closer to reality than the ISO eye color categorization. If our experiments show that our classification scheme works as expected we will recommend to extend and redefine the eye color table in the ISO standard based on our eye color table. If the classification process does not work as expected or if a redefinition of the standard is not be possible we could try to map our extended eye color table to ISO eye colors by rearranging and grouping several eye colors of our table and assign them to one of the ISO eye colors.

Six *hair colors* are specified by ISO/IEC 19794-5 [Intb]: "black", "blonde","brown", "gray", "white", and "red". An additional category covers the case of no hair and therefore no hair color ("bald"). This is a very simplistic and rough classification scheme, but probably sufficient for traveling purposes. The same considerations and expectations apply to hair colors analogously to those described above for eye colors.

For a more detailed capturing of hair colors we propose to use 10 color depths (from black to brown to blond) with another three natural red colors (light red, medium red, dark red). 5 stages of grey (0%, 25%, 50%, 75% and 100% grey) and 5 additional colors (blue, green, yellow, clear red) were added to cover the overall range of natural and artificial hair colors. These color categories are – in analogy to glass eyes for eye colors – represented by a ring of artificial hair.

Skin color classification using artificial color tables poses some problems since available color tables are based on clear colors which makes a comparison with multi-pigmented skin very difficult. This may lead to false classification or even the lack of a suitable reference class for some skin types. Therefore, it is investigated if a spectrophotometer may be used for an objective skin-color categorization.

First experiences with the classification scheme for eye, hair and skin colors defined above are under development.

4.3 Definition of Ground Truth for Full Frontal View

In the current development of the revision process of 19794-5 (latest version 2nd CD as of January 2009), the following differences in comparison to the first revision are already visible.

The problem of defining a zero degree reference point in pitch and yaw (see figure 4.3) is at least acknowledged, stating that the definition of zero pitch and yaw is "not obvious". As a partial aid and based on the inclusion of 3D image representations in the base standard, the Frankfurt horizontal (defined by a line through the tragion and the lowest point of the right eye socket) is at least included, although not used as a normative definition.

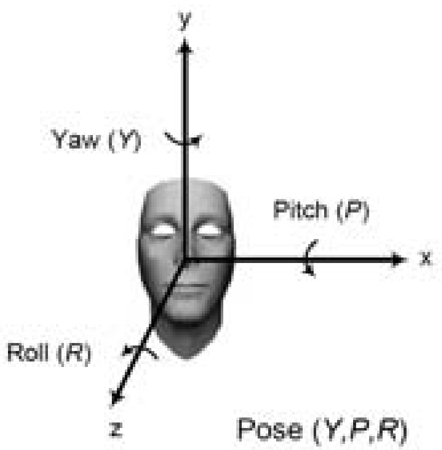

Figure 3: The definition of pose angles with respect to the frontal view of the subject according to ISO/IEC 19794-5 [Intb]

In general, the standard now tries to be more precise in its definition, e.g. it has been given an upper bound for an acceptable thickness of problematical heavy glasses frames (not more than 5% of the images inter-eye distance), or, as an other example, the definition of equal subject lighting has been enhanced by a formula for automatic computation.

The best-practice guide (in Annex A) has been refined and extended by several definitions for recommended image properties, e.g. a technical definition of hair covering the eyes and rims of glasses covering the eyes was added.

The general trend in the revision process that can be deduced from these observed changes:

- Much care has been taken to be more precise on the specification of several technical requirements where possible.
- The limitations of this approach have been recognized, although a sensible definition of pose error is still not given (because of the impossibility of establishing a useful ground truth), the requirement has still been kept in the revision of the standard. This was mainly based on the feeling of many participants that a normative requirement on pose was and is still necessary.

As a side note: Based on the introduction of the 3D image types in the base standard and the possibility of having both 2D and 3D representations of the same person in the record, in some settings a better way of determining pose deviations based on the 3D representation might be possible. Therefore, it is valuable to specify these pose properties in the standard since there are application domains of the standard beyond electronic passports or similar use cases.

5 Standards and Guidelines: Lessons Learned

In particular the second technical corrigendum [Intf] shows the particularities of the standard ISO/IEC 19794-5 [Intb]. Experiences with the application of the standard show a need for corrections in fundamental elements of related standards. Improved requirements are now available for applications that deal with the quality assurance of images as well as the assignment and production of electronic passports.

In ISO standardization processes a standard is usually reviewed every three to five years resulting in a continuance of stipulations, revisions or withdrawals of specifications. Accordingly ISO/IEC 19794-5 [Intb] is also currently under revision. Now the time pressure for the international group of experts to (re-)define requirements for facial image properties is lower than when the first version of the standard was released. The second version of the standard is planned for 2011.

The following conclusions can be drawn for the implementation and revision of standards:

- Specifications and requirements should always be checked for necessity, practicability and usability.
- Application improves the standard.
- Time pressure is an enemy of quality.
- Continued review and revision is always necessary.

These tenets should be taken into account in the development of future standards so that revision and extension efforts are kept to a minimum ensuring an efficient and target-oriented revision process.

References

[Bun] Bundesamt für Sicherheit in der Informationstechnik. BSI TR-03104 Annex 1 (QS-Gesicht) – Technische Richtlinie zur Produktionsdatenerfassung, -qualitätsprüfung und -übermittlung für Pässe, Version 2.1.1 – Qualitätsanforderungen bei der Erfassung und Übertragung der Lichtbilder als biometrische Merkmale für elektronische Pässe.

[ENSS08] Peter Ebinger, Margarida Castro Neves, René Salamon, and Helmut Seibert. International Database of Facial Images for Performance and ISO/IEC 19794-5 Conformance Tests. In *BIOSIG 2008*, pages 165–174, 2008.

[Gro04] International Civil Aviation Organization Technical Advisory Group 15 Machine Readable Travel Documents/New Technologies Working Group. Biometrics Deployment of Machine Readable Travel Documents, Version 2.0, May 2004.

[Inta] International Organization for Standardization. ISO/IEC CD 19794-1 Information Technology – Biometric data interchange formats – Part 1: Framework.

[Intb] International Organization for Standardization. ISO/IEC CD 19794-5 Information Technology – Biometric data interchange formats – Part 5: Face Image Data.

[Intc] International Organization for Standardization. ISO/IEC FDIS 29109-1 – Conformance Testing Methodology for Biometric Interchange Records Format Part 1: Generalized conformance testing methodology.

[Intd] International Organization for Standardization. ISO/IEC IS 19794-5:2005, Information Technology – Biometric data interchange formats – Part 5: Face Image Data – Amendment 1 – Conditions for Taking Photographs for Face Image Data.

[Inte] International Organization for Standardization. ISO/IEC IS 19794-5:2005, Information Technology – Biometric data interchange formats – Part 5: Face Image Data – Technical Corrigendum 1.

[Intf] International Organization for Standardization. ISO/IEC IS 19794-5:2005, Information Technology – Biometric data interchange formats – Part 5: Face Image Data – Technical Corrigendum 2.

[Intg] International Organization for Standardization. ISO/IEC IS 19795-2:2007, Information technology – Biometric performance testing and reporting – Part 2: Testing methodologies for technology and scenario evaluation.

[Inth] International Organization for Standardization. JTC 1/SC 37 Biometrics – WD Standing Document 2 Version 11 (SD 2), Harmonized Biometric Vocabulary.