3D Face Recognition on Low-Cost Depth Sensors

Štěpán Mráček[*], Martin Drahanský[*×], Radim Dvořák[*], Ivo Provazník[+], and Jan Váňa[*]
[*]Faculty of Information Technology
[+]Faculty of Electrical Engineering and Communication
Brno University of Technology, Czech Republic

[×]Department of Computer Science
Graduate School of Information Science and Engineering
Tokyo Institute of Technology, Japan

{imracek,drahan,idvorak,ivanajan}@fit.vutbr.cz, provaznik@feec.vutbr.cz

Abstract: This paper deals with the biometric recognition of 3D faces with the emphasis on the low-cost depth sensors; such are Microsoft Kinect and SoftKinetic DS325. The presented approach is based on the score-level fusion of individual recognition units. Each unit processes the input face mesh and produces a curvature, depth, or texture representation. This image representation is further processed by specific Gabor or Gauss-Laguerre complex filter. The absolute response is then projected to lower-dimension representations and the feature vector is thus extracted. Comparison scores of individual recognition units are combined using transformation-based, classifier-based, or density-based score-level fusion. The results suggest that even poor quality low-resolution scans containing holes and noise might be successfully used for recognition in relatively small databases.

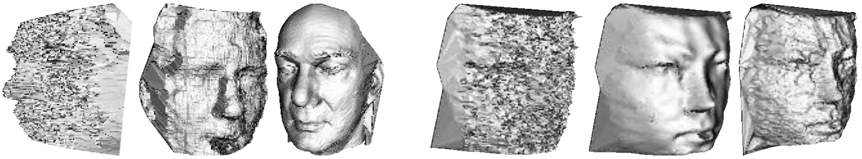
1 Introduction

The face is one of the most used biometric modalities. Although there has been a rapid development in recent years [ANRS07] and the facial biometric is also accepted in the industry, there are still some challenges that should be considered when one is designing a face recognition system. The classical approach utilizing 2D photographs has to deal with illumination and pose variation. This can be solved when the 3D face recognition is used, however, the biggest disadvantage of this approach are much higher acquisition costs.[1]
The expansion of personal depth sensors related with the new ways of the human-computer interaction in recent years markedly lowered the price of 3D acquiring devices for personal use. This paper describes the face recognition pipeline utilizing such low-cost devices, i.e., Microsoft Kinect 360[2] and SoftKinetic DS325[3] sensors.
The biggest challenge of the face recognition based on the low-cost depth sensors is the quality of acquired scans. While, for example, the Minolta Vivid or Artec 3D M scanners provide a highly precise geometry with outstanding resolution and level of detail, the scans retrieved from the Kinect or DS325 sensors are noisy, have low resolution and sometimes contain holes (see Figure 1(a)).

[1] Full-length version of this paper can be found at: http://www.fit.vutbr.cz/ imracek/pubs.php?id=10679

[2] http://www.xbox.com/kinect/

[3] http://www.softkinetic.com/products/depthsensecameras.aspx



(a) Scans from SoftKinetic (left), Kinect (middle), and Minolta Vivid (right) sensors.

(b) Application of feature preserving mesh denoising – before (left) and after (middle). Basic Gaussian smoothing is on the right side of the figure.

Figure 1: Input scans and mesh denoising algorithm.

1.1 Related Work

Our approach represents a combination of holistic and feature-based methods. We are using the holistic feature extraction method - *Principal Component Analysis* (PCA) [DGG05] performed on the image representation of the face surface. However, we process the image with the bank of filters first. E.g. the Gabor filter offers localization of specific properties of the image in spatial as well as frequency domain. Thus our approach may be also considered as the feature based recognition.
The similar approach, where the holistic and local features are combined, is presented in [Ard12]. Their method is based on a set of facial depth maps extracted by multiscale extended *Local Binary Patterns* (eLBP). The following SIFT-based matching strategy combines local and holistic analysis. In [KED12] a block based face analysis approach is proposed which provides the advantage of robustness to nose plastic surgery alterations. The method utilizes local description. PCA, *Linear Discriminant Analysis* (LDA) and *Circular Local Binary Pattern* (CLBP) are applied over image blocks to extract block features.
The utilization of Kinect sensor for face recognition was proposed in [LMLK13] where *Sparse Representation Classifier* (SRC) is applied on the range images as well as on the texture. Moreover, the RGB channels of the texure are transformed using *Tensor Discriminant Color Space* (TDCS).
The application of the Gabor and Gauss-Laguerre filters for thermal face recognition has been previously proposed in our work [VMD^{+}12]. We have shown that the score-level fusion of individual face recognition classifiers based on PCA and ICA applied on images processed by Gabor and Gauss-Laguerre filter banks significantly outperforms individual face classifiers.
We also investigated the utilization of image filters and score-level fusion in our previous work that deals with 3D face recognition [MVL^{+}14]. In this paper, the recognition pipeline is generalized in order to deal with poor-quality scans.

2 Pre-processing

The pose invariation of our recognition algorithm is solved using the *Iterative Closest Point* (ICP) algorithm. The input face mesh is aligned to the reference face template, such that the sum of distances between corresponding points of template and input mesh are minimal. *Fast Library for Approximate Nearest Neighbors* (FLANN) is used in order to achieve a fast calculation of corresponding points.

The scans acquired using the SoftKinetic sensor suffer from high noise and peak presence. Although one can use stronger Gaussian smooth filter, our experiments show that much better, in terms of recognition performance, is the application of the feature-preserving mesh denoising algorithm [SRML07]. The example of application of such filter is in Figure 1(b).

We estimate the principal curvatures $k_{1_\mathbf{P}}$ and $k_{2_\mathbf{P}}$ at each point $\mathbf{P}$ from the range image representation of the aligned mesh [MBDD11]. Several important surface image representations can be directly deduced from the principal curvature values. The mean curvature $H_\mathbf{P}$, Gaussian curvature $K_\mathbf{P}$, and the shape index $S_\mathbf{P}$:

$$H_\mathbf{P} = \frac{1}{2}\left(k_{1_\mathbf{P}} + k_{2_\mathbf{P}}\right), \quad K_\mathbf{P} = k_{1_\mathbf{P}} k_{2_\mathbf{P}}, \quad S_\mathbf{P} = \frac{1}{2} - \frac{1}{\pi}\operatorname{atan}\left(\frac{k_{1_\mathbf{P}} + k_{2_\mathbf{P}}}{k_{2_\mathbf{P}} - k_{1_\mathbf{P}}}\right) \tag{1}$$

Another image curvature representation is the *eigencurvature* [Rus09] that is computed from the image point $\mathbf{P} = (p_x, p_y, p_z)^T$ and its 8 surroundings $(\mathbf{P}_1, \mathbf{P}_2, \ldots, \mathbf{P}_8)$. It is based on the PCA of the matrix $\mathbf{M} = \begin{pmatrix}\mathbf{P} & \mathbf{P}_1 & \cdots & \mathbf{P}_8\end{pmatrix}$. The PCA reveals 3 eigenvectors and their corresponding eigenvalues l_0, l_1, and l_2 $(l_0 > l_1 > l_2)$. The *eigencurvature* $E_\mathbf{P}$ is then $E_\mathbf{P} = \frac{l_2}{l_0+l_1+l_2}$. The examples of various texture, depth, and curvature representations are in Figure 2.

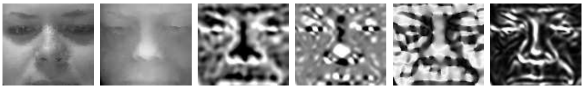


Figure 2: From left to right: texture, range image, mean curvature, Gaussian curvature, shape index, and eigencurvature.

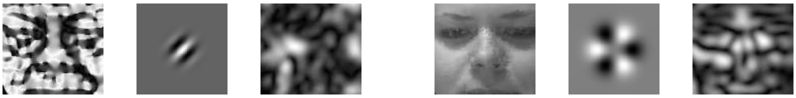
3 Feature extraction

3.1 Filter Banks

The image filter banks are set of m 2D kernels that are convolved with the input image. This convolution provides m new images that are further used for the feature extraction and comparison. We utilize the Gabor filter bank and Gauss-Laguerre filter bank.

The complex Gabor filter [Lee96] is defined as the product of a Gaussian kernel and a complex sinusoid. Our Gabor filter bank consists of 56 filters with the varying orientation $o \in (0, 1, \ldots, 7)$ and frequency $f \in (1, 2, \ldots, 7)$.

The Gauss-Laguerre wavelets [AP07] are polar-separable functions with harmonic angular shape. They are steerable in any desired direction by simple multiplication with a complex steering factor and as such they are referred to self-steerable wavelets. Our Gauss-Laguerre filter bank consists of 35 filters that were created with parameters $n \in (1, 2, 3, 4, 5)$, $k = 0$, $j = 0$ with sizes 16×16, 24×24, 32×32, 48×48, 64×64, 72×72, and 96×96 pixels. The examples of the application of Gabor and Gauss-Laguerre filters are in Figures 3(a) and 3(b) respectively.



(a) Gabor filter applied on the shape index image. (b) Gauss-Laguerre filter applied on the texture image.

Figure 3: Example of application of Gabor and Gauss-Laguerre image filters. From left in each sub-figure: input image, real part of the kernel, and absolute response (modulus).

3.2 Modified PCA

Probably the most crucial part of every biometric system is the selection of the feature extraction algorithm and the subsequent comparison metric. In the area of face recognition, a well established feature methods are PCA, LDA, and ICA [DGG05]. We have compared PCA, LDA as well as ICA in our experiments and selected modified PCA as the feature extraction method that best suits our needs.

In plain PCA, the components of the projected vector are proportional to the variability that is expressed as the corresponding eigenvalue. This unbalance of individual feature vector components leads to neglect of those feature vector components that may have positive impact on the recognition performance, however their associated eigenvalue is too small. In order to avoid that, individual feature vector components are normalized after PCA projection using z-score normalization. That is, an arbitrary feature vector $X = (x_1, x_2, \ldots, x_m)$ is modified such $x_i \leftarrow \frac{x_i - \bar{x}_i}{\sigma_i}$, where $\bar{x}_i$ is the mean value of the component i and σ_i is corresponding standard deviation.

Usually, the basic Euclidean distance is used in order to compare two feature vectors. We have tried other metric functions as well and the correlation metric achieved the best results in our experiments.

4 Score-level Fusion

According to [NCDJ08], score fusion techniques can be divided into the following three categories: *Transformation-based* – the scores are first normalized (transformed) to a common domain and then combined. *Classifier-based* – scores from multiple matchers are treated as a feature vector and a classifier is constructed to discriminate genuine and impostor scores. *Density-based* score – this approach is based on the likelihood ratio test and it requires explicit estimation of genuine and impostor comparison score densities.

We use a weighted sum as a representative of transformation-based fusion. The classifier-based fusion is provided by the SVM classifier with linear kernel. The density-based fusion is represented by the *Gaussian Mixture Model* (GMM) [NCDJ08].

When the fusion of scores from individual classifiers is involved, the emphasis should be put on the selection of classifiers in order to avoid degradation of recognition performance caused by score correlation and performance bias [PB05]. Our face pre-processing produces 6 representations of the face texture, shape, and curvature. Moreover, each representation is optionally convoluted with one of 56 Gabor filters or 35 Gauss-Laguerre filters. That yields to $6 \cdot (1 + 56 + 35) = 552$ possible score-level fusion inputs (units). The exhaustive search of all potential combinations of input classifiers ($2^{552} - 1$) is therefore impossible.

We employ a greedy hill-climbing wrapper selection mechanism. The optimization criterion is the achieved EER of the fusion on the training set. The selection wrapper selects the best units in the first iteration. In subsequent iterations, the unit that best improves the fusion is added to the selected units set. The selection is ended when there is no further unit to add or if there is no improvement.

5 Evaluation

Our databases were acquired using Microsoft Kinect and SoftKinetic DS325 depth sensors. We developed a simple enrollment application, where users had to position their head to the specific distance from the sensor. The process of capturing was fully automatic – once the face was detected, users were notified not to move and their 3D face model was acquired. The SoftKinetic database consists of 320 scans divided to 3 portions – training set (13 subjects, 94 scans), validation set (12 subjects, 60 scans), and evaluation set (26 subjects, 166 scans). The Kinect database consists of 110 scans divided to 2 equally sized portions – training and evaluation sets, both with 55 scans and 9 subjects.

5.1 SoftKinetic

Table 1 brings the detailed overview of the unit selection process using the wrapper. The individual units as well as the SVM-based fusion were trained on the training portion of the SoftKinetic database. Values in the table show that even if the particular unit has EER 26% it can contribute to the overall recognition performance.
The Gabor(f, o) in Table 1 stands for the application of Gabor filter with frequency f and orientation o. The G-L(s, n) stands for the application of Gauss-Laguerre filter with size s and appropriate parameter n.

Table 1: Wrapper unit selection training - SVM fusion on the SoftKinetic database.

Iteration	Selected unit		Unit EER	Fusion EER
	Image data	Applied Filter		
1	Depth	Gabor(7,2)	0.0867	0.0867
2	Eigen	Gabor(4,5)	0.1404	0.0657
⋮	⋮	⋮	⋮	⋮
14	Gauss	G-L(16,1)	0.2580	**0.0262**

We have evaluated all three major score-level fusion techniques on the SoftKinetic database. The individual weights of the weighted sum fusion are proportional to the achieved EERs on the training portion of the database. The weight w_i of unit i is:

$$w_i = \frac{0.5 - eer_i}{\sum_{i=j}^{n}(0.5 - eer_j)} \tag{2}$$

where eer_i is the achieved EER for unit i and n is the number of units. The transformation-based fusion requires a normalization of the input scores prior to the fusion itself. We are using a simple normalization of input score s:

$$s \leftarrow \frac{s - gen_i}{imp_i - gen_i} \quad (3)$$

where gen_i is the mean genuine score for unit i and imp_i is correponding mean impostor score.

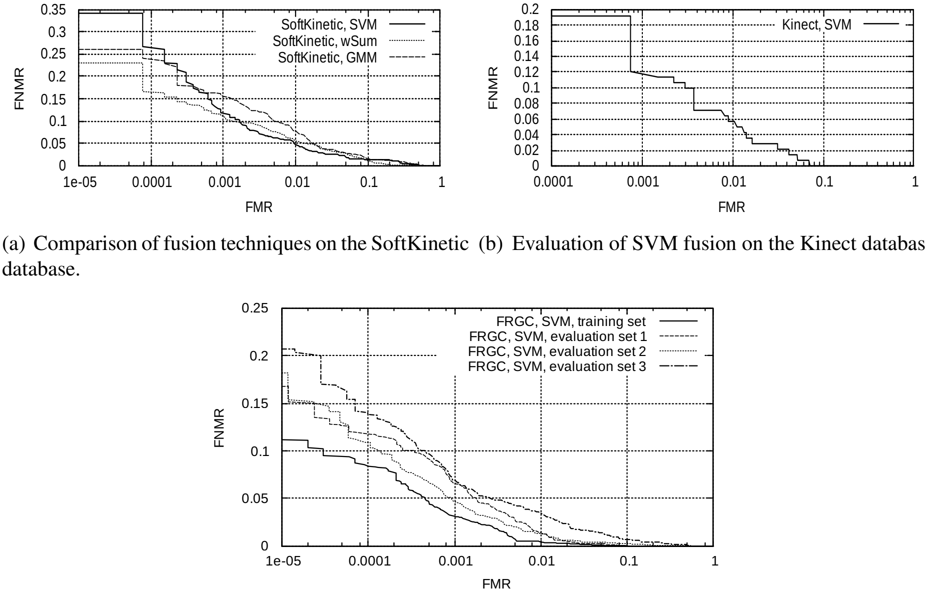
The SVM classifier is using a simple linear kernel. Although there should be no need for prior score normalization, we are using the same normalization technique as in weighted sum fusion. Our experiments shown that this has positive impact on the recognition performance.

The GMM-based fusion is trained using the expectation-maximization algorithm. Both genuine and impostor distributions are modeled using 5 Gaussian mixtures with diagonal covariance matrices.

The results are shown in Table 2. It has emerged that there is not a significant difference between individual fusion techniques. For example, the lowest FNMR for a given FMR = 0.001 is achieved with SVM-based fusion, but the best at FMR = 0.0001 is the weighted sum. Figure 4(a) shows DET curves of evaluated techniques.

Table 2: Fusion techniques on the SoftKinetic database.

Fusion type	EER on the evaluation set	FNMR at FMR = 0.001
Weighted sum	0.0321	0.1097
SVM	0.0259	0.1172
GMM	0.0350	0.1556



(a) Comparison of fusion techniques on the SoftKinetic database.

(b) Evaluation of SVM fusion on the Kinect database.

(c) Evaluation on the FRGC v2.0 database.

Figure 4: DET curves.

5.2 Kinect

In this subsection, we present performance of our face recognition algorithm on the Kinect database with SVM fusion. As it was shown in previous subsection, the results with weighted sum or GMM-based fusion are similar. The only difference between SoftKinetic and Kinect input face mesh pre-processing is the absence of the feature-preserving denoising. Scans acquired with Kinect are less noisy and thus they need no special denoising treatment. On the other hand, they have lower resolution. This is because the Kinect is able to capture depth data from greater distance than SoftKinetic sensor. The DET curve of our recognition algorithm evaluated on the Kinect database is in Figure 4(b).

5.3 FRGC

In order to allow a direct comparison of our recognition algorithm with others, we also made evaluations on the FRGC v2.0 database. We used the „spring2004“ part of the database divided into 5 isolated non-overlapping portions. First portion (416 scans) was used for training of individual PCA projections, the second portion (451 scans) was used for the selection of suitable fusion units and training of the SVM classifier. The last three portions (414, 417, and 308 scans) were reserved for evaluation. Each subject was present just in one portion. The achieved EERs as well as FNMR values at specific FMRs are summarized in Table 3. Corresponding DET curves are in Figure 4(c).

Table 3: Evaluation on the FRGC database.

Set	EER	FNMR at FMR = 0.001	FNMR at FMR = 0.0001
Training set	0.0053	0.0314	0.0837
Evaluation #1	0.0117	0.0659	0.1176
Evaluation #2	0.0116	0.0466	0.1087
Evaluation #3	0.0214	0.0688	0.1381

6 Conclusion

The presented 3D face recognition algorithm is robust enough in order to deal with poor quality scans acquired with the Kinect or SoftKinectic DS325 sensors. We have also made evaluations on the FRGC v2.0 database. Our experiments show that the real-world application of the face recognition employing a low-cost device may be limited by desired security and the expected size of the database. The verification or identification within the database consisting of 26 persons employing SoftKinectic DS325 sensor is convenient for users even when the desired security of the system is set to FMR = 0.001. Further robustness of the recognition may be achieved using more than one reference template.

Acknowledgment

This research has been realized under the support of the following grants: “The IT4Innovations Centre of Excellence” ED1.1.00/02.0070, “Verification and Optimization of Computer Systems” FIT-S-12-1, “Reliability and Security in IT” FIT-S-14-2486, “Tools and Methods for Video and Image Processing for the Fight against Terrorism” VG20102015006, and the fellowship research activity at the Tokyo Institute of Technology supported by the Matsumae International Foundation (Japan).

References

[ANRS07] A. F. Abate, M. Nappi, D. Riccio, and G. Sabatino. 2D and 3D face recognition: A survey. *Pattern Recognition Letters*, 28(14):1885–1906, October 2007.

[AP07] H. Ahmadi and A. Pousaberi. An Efficient Iris Coding Based on Gauss-Laguerre Wavelets. In *Advances in Biometrics*, pages 917–926. 2007.

[Ard12] M. Ardabilian. 3-D Face Recognition Using eLBP-Based Facial Description and Local Feature Hybrid Matching. *IEEE Transactions on Information Forensics and Security*, 7(5):1551–1565, October 2012.

[DGG05] K. Delac, M. Grgic, and S. Grgic. Independent Comparative Study of PCA, ICA, and LDA on the FERET Data Set. *International Journal of Imaging Systems and Technology*, 15(5):252–260, 2005.

[KED12] N. Kose, N. Erdogmus, and J. L. Dugelay. Block based face recognition approach robust to nose alterations. In *2012 IEEE Fifth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, pages 121–126. IEEE, September 2012.

[Lee96] T. Lee. Image representation using 2D Gabor wavelets. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(10):959–971, 1996.

[LMLK13] B. Y. L. Li, A. S. Mian, W. Liu, and A. Krishna. Using Kinect for face recognition under varying poses, expressions, illumination and disguise. In *IEEE Workshop on Applications of Computer Vision (WACV)*, pages 186–192. IEEE, January 2013.

[MBDD11] Š. Mráček, C. Busch, R. Dvořák, and M. Drahanský. Inspired by Bertillon - Recognition Based on Anatomical Features from 3D Face Scans. In *Proceedings of the 3rd International Workshop on Security and Communication Networks*, pages 53–58, 2011.

[MVL$^+$14] Š. Mráček, J. Váňa, K. Lankašová, M. Drahanský, and M. Doležel. 3D face recognition based on the hierarchical score-level fusion classifiers. In *Biometric and Surveillance Technology for Human and Activity Identification XI*, page 12, 2014.

[NCDJ08] K. Nandakumar, Y. Chen, S. C. Dass, and A. K. Jain. Likelihood ratio-based biometric score fusion. *IEEE transactions on pattern analysis and machine intelligence*, 30(2):342–347, February 2008.

[PB05] N. Poh and S. Bengio. How Do Correlation and Variance of Base-Experts Affect Fusion in Biometric Authentication Tasks? *IEEE Transacations on Signal Processing*, 53(11):4384–4396, 2005.

[Rus09] R. B. Rusu. *Semantic 3D Object Maps for Everyday Manipulation in Human Living Environments*. PhD thesis, August 2009.

[SRML07] X. Sun, P. Rosin, R. Martin, and F. Langbein. Fast and effective feature-preserving mesh denoising. *IEEE transactions on visualization and computer graphics*, 13(5):925–938, 2007.

[VMD$^+$12] J. Váňa, Š. Mráček, M. Drahanský, A. Poursaberi, and S. Yanushkevich. Applying Fusion in Thermal Face Recognition. In *International Conference of the Biometrics Special Interest Group (BIOSIG)*, pages 1–7, 2012.