An Automatic 2D, 2.5D & 3D Score-Based Fusion Face Verification System

Cristina Conde, Ángel Serrano, Licesio J. Rodríguez-Aragón and Enrique Cabello
Face Recognition & Artificial Vision Group (FRAV)
Universidad Rey Juan Carlos, c/ Tulipán, s/n, Móstoles (Madrid) 28933-E, Spain
{cristina.conde, angel.serrano, licesio.rodriguez.aragon, enrique.cabello}@urjc.es

Abstract

A score-based fusion for face verification is presented from FRAV3D Face Database (2D, 2.5D and 3D face images). In the case of 2.5D and 3D data, an automatic correction of pose has been carried out by detecting the nose tip and the eyes. For each kind of image a different feature extraction has been applied (Principal Component Analysis and Support Vector Machine for 2D and 2.5D, and Iterative Closest Point algorithm for 3D). A fusion at score level has been performed two by two, after a minimum-maximum normalization (MM) and a Z-score standardization (ZS). We have found an optimal combination that reduces (or at least does not worsen) the Equal Error Rate of the classifiers applied independently. In the most optimal situation, the improvement of the EER is higher than 80% for the fusion of 2D and 2.5D data, as well as for 2.5D and 3D data.

1. Introduction

Due to a great interest in security applications, the automatic use of human faces to recognize or verify the identity of people has received increasing attention, as it is user-friendly and privacy-respectful. Since the recent emergence of cheap 3D digitizers, face-based biometrics has made use of a variety of detectors and types of data, from the usual 2D colour images, to the 3D model meshes, including the intermediate range images, which combine 2D and 3D information (therefore known as 2.5D images).

The fusion of data from different biometrics or from different detectors has proven to yield better results than considering the data independently [1, 2, 3]. Several fusion scenarios can be taken into account [2]. Data can be fused directly after their acquisition, or after the main features have been extracted. The scores from different classifiers can be combined, or even a voting process can be carried out by separate classifiers. The method used here is score-based.

The remainder of this paper is organized as follows. In section 2, a literature review about fusion of classifiers applied to face recognition can be read. In section 3, we describe the face database we have used for this work. In section 4, the feature localization and face normalization in pose for 2.5D and 3D images is depicted. In section 5 we describe the classifiers module for the different kinds of images, while in section 6 the fusion of classifiers is presented. In section 7 we discuss our results. Finally conclusions are to be found in section 8.

2. Previous work

Several works about data fusion have been published in the recent years. On the one hand, Chang et al. [4] used range images and texture images, which were analysed with the eigenfaces method [5] and the Mahalanobis distance. They obtained better results when the results provided by both kind of images were combined by a weighted sum (92.8% success rate).

Chang et al. [6] also studied the improvement in classification for 2D images and range images. They tried to find out the influence on the recognition rate by the fusion of data or the amount images used. They obtained a multimodal Equal Error Rate (EER) of 0.019 and a value of 0.045 for only range images.

In [7] a comparison of two systems working with 2D, 3D and their combination is presented. They used the FaceIt system, based on Local Feature Analysis (LFA), and the Fisherfaces method, based on Principal Component Analysis (PCA). The level of fusion was different for each system: detector-based (Fisherfaces) or classifier-based (FaceIt). They obtained a lower EER for the fused data for both methods. For FaceIt, the EER fell to 0 for the fusion (0.027 for 3D, 0.15 for 2D), while for Fisherfaces, the combined EER was 0.96 (1.82 for 3D, 1.55 for 2D).

1-4244-0686-2/06/$20.00 C2006 IEEE

3. FRAV3D: Our Face Database

We used our FRAV3D Face Database [8], which contains 2D colour images, as well as 3D mesh models. It was acquired using a Minolta 700 VIVID red laser under indoor controlled lighting. It contains 105 subjects, mainly young adults, Caucasian (one woman every three men), with the person sitting opposite the scanner and in front of a dark plain background. No hats, scarves or glasses were allowed.

There are 16 captures per person, which include 4 frontal views, 2 turns to the right (25º), 2 turns to the left (5º), 2 turns in the Z direction, 2 views looking up and down, 2 views with gestures and 2 frontal images with lighting changes. Only one parameter was changed in every new capture.

The scanner provided two kinds of images: a 2D colour picture and a 3D mesh model. We also computed the corresponding range images from the 3D model, as can be read in the following section.

Hair, blue areas in the scene and occlusions constrain the effectiveness of the verification process, although they do not impede it as a whole.

4. Feature location and face normalization

We used an automatic algorithm to locate the face features in order to correct changes in pose in 3D meshes. This normalization is very necessary in order to take advantage of the three-dimensional properties of human face for verification purposes. In order to guarantee a good normalization, it is essential to detect and localize face features accurately, such as nose tip and inside corners of the eyes. This step is carried out by means of the so-called spin images [9, 10, 11].

4.1. Spin images

A spin image is a representation of a 3D mesh with respect to an "oriented" point $\vec{p}$ and the unitary vector $\vec{n}$ containing it, which is perpendicular to the 3D mesh in this point. In this reference frame, for each point $\vec{x}$=(x, y, z), there is a coordinate change to a 2D system (α, β), which defines the distance from the point $\vec{x}$ to the oriented point $\vec{p}$ as follows:

$$S_0 : \Re^3 \to \Re^2$$
$$S_0(\vec{x}) \to (\alpha, \beta) = \left(\sqrt{\|\vec{x}-\vec{p}\|^2 - (\vec{n}\cdot(\vec{x}-\vec{p}))^2}, \vec{n}\cdot(\vec{x}-\vec{p})\right) \quad (1)$$

where α is the distance from a point $\vec{x}$ to the straight line L parallel to $\vec{n}$ and containing $\vec{p}$, while β is the distance from $\vec{x}$ to the plane P (Figure 1).

A spin image can be computed considering any point in a 3D mesh as a reference. It is easy to understand that spin images for a certain facial feature are similar for different 3D face models, as all human faces are alike in shape (two eyes, a nose and a mouth) and size. Therefore spin images can be used to detect face features.

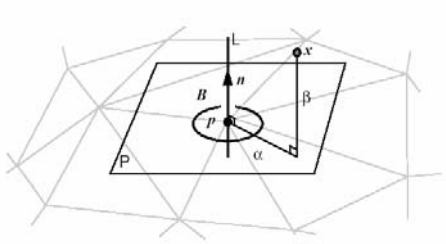

Figure 1: Parameters of a spin image [9].

4.2. Localization of the nose tip

We trained an SVM classifier [12, 13] to detect the nose tip in a 3D face image using 90 face meshes, where the nose location had been found manually. The spin image computed for each nose tip was converted to a 1D vector and fed to an SVM as a positive case. Spin images for face features other than nose tips were used as negative cases.

In the test stage, for an unknown mesh, the nose was searched by looking for the most jutting point in the 3D model and providing its spin image to the SVM. This process was repeated for the next most jutting point until a positive SVM score was found. That point was considered the nose tip.

4.3. Localization of the eyes

Once the nose has been localized, an approximate region for both eyes is extracted, in order to accelerate their finding process. This allows to remove controversial features, such as the hair or the chin.

Within the candidate area, the discrete mean curvature was computed at each point with information on the angles between adjacent polygons [14]. Using clustering techniques based on Euclidean distance, three groups were selected around those points with the highest curvature: both eyes and nose.

For the eye clusters, a second SVM classifier was trained to detect each inside corner using the corresponding spin images. Those points with the deepest Z coordinate in the 3D model were preselected as candidates for an eye. The spin images for those points were fed to the SVM. This process was iteratively repeated until two points produced a positive SVM score, one for each eye.

4.4. Face normalization

With the coordinates of both eyes and nose tip, we computed the adequate turn so that the 3D face model seemed to look directly forward. This correction of pose can be divided in three steps, each one regarding the three directions X, Y, Z. The background wall is considered as the XY plane, while the direction of the laser beam of the scanner is the Z axis.

With the Z coordinates of both eyes, the face pose is turned around Y axis, so that the face is not in profile any more. By taking into account the points belonging to the nose bridge, from the middle of the brows to the nose tip, a turn around Z axis can be applied so that the face seems to be at an upright position. This correction is better performed by considering only the points of the face inside an ellipse centred at the nose, in order to avoid noise from the external parts such as the hair. Finally, the face can be corrected for X axis, by computing by least squares the best straight line that fits the projection of the face onto the YZ plane.

4.5. Range data computation

After pose normalization, we computed the corresponding 200×200 range image, where the value of each pixel is related to the distance to the scanner (Z coordinate). 2.5D images represent 3D information in a 2D style. Following the method proposed by Conde et al. [15], the exact value for each pixel was computed by applying an exponential equalization, which allows an optimization in the range values used and produces better verification results.

5. Verification process

For our verification process, we used the standard 2D colour images, for which no correction for face orientation is possible, as well as the 2.5D range images and the 3D mesh models, which can be normalized in pose as seen in the previous section. The feature extraction techniques and the used classifiers will be different for each case (Figure 2).

5.1. 2D and 2.5D data

For 2D images as well as 2.5D range images, which in practice can be also considered as 2D images, we applied a usual Principal Component Analysis algorithm (PCA) as a dimension reduction method [5]. After finding the face in the picture with a template, every image was cropped to a 130×140 pixels size and converted to grey scale. Repeating this process for all the subjects in the database, the usual covariance matrix and the eigenfaces were computed, considering only the 150 most dominant eigenvalues. This is done twice, once for the 2D pictures and another time for the 2.5D range images.

All these images were projected onto the eigenfaces framework. The 150 projection coefficients were used to train a set of SVM classifiers [12, 13], each one for every subject in the database, considering this subject as a genuine case and the others as impostors. For the decision module, the SVM associated with that person is used. We performed 15 experiments using different kinds of images per person to train and test these SVMs (Table 1), in order to compute the influence of turns, gestures and lighting conditions in the score fusion.

The score provided by the classifier will be used to make the decision. This value can be positive or negative. The more positive it is, the more confident is the verification.

5.2. 3D data

For the 3D mesh models, we used the usual Iterative Closest Point (ICP) algorithm [16, 17]. In this situation, two 3D clouds of points have to be fitted. One of them is the model and is fixed. The other one is the scene.

By means of an iterative process, the scene is rotated and moved until the fitting square error is lower than a certain threshold. The value of this error will be used in the subsequent verification process. Obviously the error cannot be negative. Therefore, the lower it is, the better coincidence between the model and the scene and the higher the probability that they belong to the same person.

Unlike 2.5D range images, which have been interpolated in order to remove holes due to occlusions or hair, 3D meshes have not been interpolated.

6. Fusion of classifiers

6.1. Score normalization

It is important to take into account the different nature of both classifiers before the fusion. On the one hand, for SVM the higher and more positive the value, the more confident is the verification. This results can be positive or negative. On the other hand, for ICP, the more similar to 0, the better fit between the model and the scene. In this case, the quadratic error is always positive or zero.

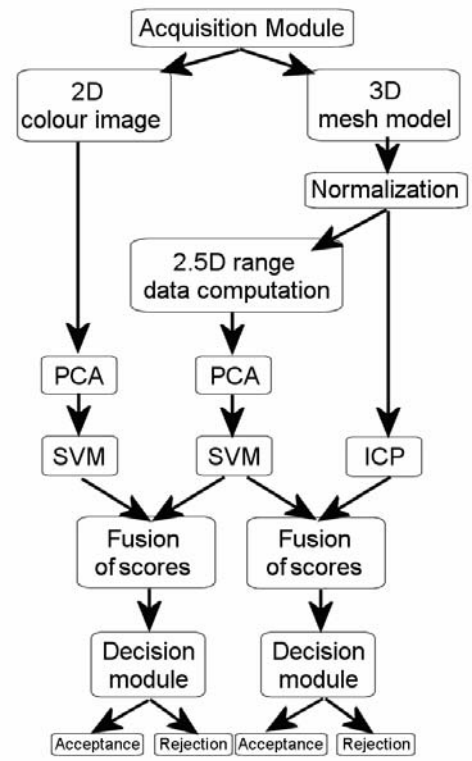

Figure 2: Scheme of our verification system.

After multiplying ICP score by –1 so that the acceptance verification is obtained when the score exceeds the corresponding threshold, two kinds of methods were applied before the data fusion [18]:

- Minimum-Maximum normalization (MM): Scores are normalized to range from 0 to 1.
- Z-score standardization (ZS): Scores are normalized to a Gaussian distribution with zero mean and standard deviation equal to 1.

6.2. Rule of fusion

The fusion of data was done by considering the normalized scores of two classifiers with a weight λ, such as $0 \leq \lambda \leq 1$:

$$score_{fusion} = \lambda \cdot score_1 + (1-\lambda) \cdot score_2 \qquad (2)$$

The fusion was performed two by two, i.e. 2D and 2.5D images on the one hand, 2.5D and 3D on the other one. The 2.5D score always corresponds to $score_1$. Therefore, when $\lambda = 1$, the 2.5D SVM output is the only one that contributes to the total score. On the contrary, when $\lambda = 0$, it is the 2D SVM output or the 3D ICP score the one considered for the fusion.

In the following section we present our results for both MM normalized scores and ZS standardized ones.

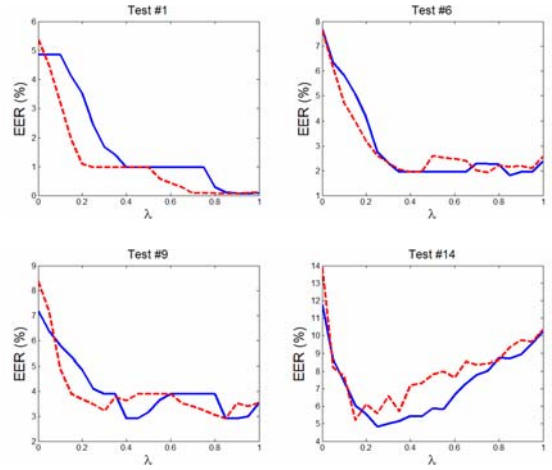

Figure 3: EER for different values of λ for certain tests for the fusion of 2.5D+3D ZS scores (continuous) and MM scores (dashed).

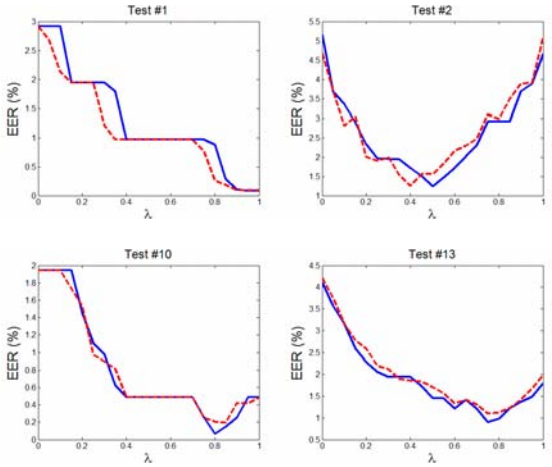

Figure 4: EER for different values of λ for certain tests for the fusion of 2D+2.5D ZS scores (continuous) and MM scores (dashed).

7. Results

For each classifier, before and after the fusion, we computed the so-called Receiver Operating Characteristic curve (ROC), which plots the False Acceptance Rate (FAR) versus the False Rejection Rate (FRR). As both rates indicate how well the distinguishing properties of each feature of each classifier work, but in an inverse way (the higher FAR, the lower FRR, and vice versa), the classifier can be better described with its Equal Error Rate (EER), which is the rate at which FAR equals FRR.

In order to study the goodness of the fusion, we computed the EER for $0 \leq \lambda \leq 1$ for the 15 experiments considered in Table 1.

Table 1: Images per person (training & tests).

Experiment	No. of training images (type)	No. of testing images (type)
1	3 (frontal)	1 (frontal)
2	4 (frontal)	1 (smiling)
3	4 (frontal)	1 (open mouth)
4	4 (frontal)	2 (illumination)
5	4 (frontal)	2 (5º Y-turn)
6	4 (frontal)	1 (Z-turn)
7	4 (frontal)	2 (X-turn)
8	4 (frontal)	2 (25º Y-turn)
9	4 (frontal)	1 (Z-turn)
10	4 (frontal) 2 (illumination)	2 (frontal)
11	3 (frontal) 1 (5º Y-turn)	1 (frontal) 1 (5º Y-turn)
12	3 (frontal) 1 (5º Y-turn) 1 (illumination)	1 (frontal) 1 (5º Y-turn) 1 (illumination)
13	3 (frontal) 1 (5º Y-turn) 1 (illumination)	1 (frontal) 1 (5º Y-turn) 1 (illumination) 1 (gesture)
14	4 (frontal)	2 (gestures)
15	2 (frontal)	2 (frontal)

Table 2: EER (%) for the 2.5D+3D fusion.

Exper-iment	2.5D + 3D (ZS)	λ_{ZS}	2.5D + 3D (MM)	λ_{MM}
1	0.07	0.95	0.07	0.85
2	3.5	0.30	3.2	0.2
3	5.9	0.25	6.5	0.15
4	0.2	0.45	0.3	0.35
5	1.0	0.80	0.9	0.55
6	1.8	0.85	1.9	0.75
7	1.3	0.65	1.4	0.6
8	1.0	0.65	1.1	0.4
9	2.9	0.40	2.9	0.85
10	0.5	1.00	0.5	0.95
11	0.5	1.00	0.5	0.65
12	0.6	0.85	0.6	0.8
13	1.4	0.90	1.5	0.8
14	4.8	0.25	5.2	0.15
15	0.5	0.85	0.5	0.7

7.1. Fusion of 2.5D+3D data

Figure 3 shows there is always an optimum value of λ for which the total EER is reduced (or at least, not worsened). For example, for experiment #9 (with turns around Z axis), the 3D classifier produces an EER of 7.37%, while the 2.5D classifier has a 3.71% error. The fusion of both classifiers for ZS scores yield an improved error of 2.9% for λ=0.40.

In all cases the EER obtained for the fusion of ZS scores is lower than that for MM scores.

Table 3: EER (%) for the 2D+2.5D fusion.

Exper-iment	2D + 2.5D (ZS)	λ_{ZS}	2D + 2.5D (MM)	λ_{MM}
1	0.09	0.95	0.09	0.9
2	1.2	0.50	1.3	0.4
3	6.4	0.30	6.8	0.25
4	0.6	0.85	0.8	0.9
5	1.0	0.80	1.0	0.9
6	1.9	0.75	2.1	0.85
7	1.1	0.70	1.2	0.7
8	3.8	0.85	3.9	0.85
9	3.7	1.00	3.5	1
10	0.06	0.80	0.2	0.85
11	0.5	1.00	0.5	0.8
12	0.4	0.75	0.5	0.85
13	0.9	0.75	1.1	0.75
14	4.7	0.35	4.5	0.2
15	0.4	0.85	0.5	0.85

7.2. Fusion of 2.5D+2D data

We have also computed the fusion of data for 2.5D and 2D images, as can be seen in Figure 4. In this case the total error is also lower than the individual errors. For example, in experiment #2 the EER using only 2D ZS score is 5.15%, while the 2.5D error is 4.67%. The fusion achieves an error of only 1.2% for λ=0.50.

In all cases the EER obtained for the fusion of ZS scores is also lower than that for MM scores.

7.3. Discussion

As can be seen in Tables 2 and 3, we have found a value of λ that reduces or equals the EER computed independently for each of the 15 experiments proposed here (Table 1). Moreover, the fusion of data gets an improvement of about 87% in experiment 10 (training with 4 frontal images and 2 images with illumination change, test with 2 frontal images), when 2D and 2.5D captures are fused (Table 4).

This improvement is of 83% when fusing 2D and 3D images for experiment 4 (training with 4 frontal images and test with 2 Z-axis turned images).

Tables 2 and 3 show the optimal values for λ for the data fusion are quite variable, depending on the type of images and the fusion selected. Except for the cases considering images with gestures (experiments #2, #3, #14), we can consider an approximate optimal range for λ between 0.6 and 0.9, i.e., the SVM score is more important than the ICP one. It can be seen that, in certain conditions, no improvement is obtained at all when fusing data of different kinds (experiments #9, #10 and #11). However no worsening in EER is obtained in these cases either.

Table 4: Percentage of improvement.

Exper-iment	2D+2.5D (ZS)	2D+2.5D (MM)	2.5D+3D (ZS)	2.5D+3D (MM)
1	8.2	0.95	29.3	29.32
2	73.3	73.02	24.7	31.68
3	49.5	46.19	53.7	47.96
4	36.9	17.20	83.5	64.21
5	3.9	0.00	3.9	11.14
6	11.3	4.59	17.1	11.46
7	43.4	38.29	33.2	26.87
8	21.0	20.00	80.0	77.20
9	0.0	4.53	21.5	21.52
10	86.7	59.46	0.0	5.66
11	0.0	0.00	0.0	0.00
12	41.1	17.32	6.6	2.25
13	45.9	33.23	12.9	7.38
14	48.5	51.11	52.9	49.19
15	20.9	16.70	7.3	14.08

8. Conclusions

A multimodal face verification system, using 2D, 2.5D and 3D images, has been presented. All the data have been acquired with the same detector. The processing has been different for each kind of images.

On the one hand, 2D and 2.5D images have been used for a PCA algorithm, whose projections onto the corresponding eigenfaces framework have been used as inputs for an SVM classifier. On the other hand, 3D meshes have been used for an ICP algorithm to fit a 3D model and an 3D scene. While the SVM score can be positive or negative (the more positive, the more confident the verification), the ICP score can only be positive or zero.

The fusion exposed here has been applied after an MM or a ZS score normalization. This fusion has only been applied for two classifiers at a time, 2D and 2.5D, 2.5D and 3D. A linear combination of the classifiers scores with a λ parameter has been considered for the fusion. In every case, an ROC curve and the corresponding EER has been computed with the fused data. In all cases, there exists a certain value of λ for which the EER for the fusion is always lower (or at least equal) compared to the verification with independent classifiers. In fact, in certain conditions, the improvement rate can reach up to 80% for both 2D+2.5D and 2.5D+3D data. In all cases, the ZS score normalization achieves lower error rates than MM.

Acknowledgments

This paper has been supported by grants from Universidad Rey Juan Carlos. The authors would like to thank Jorge Pérez for his enthusiastic work.

References

[1] R. Duin and D. Tax. Classifier conditional posterior probabilities. *Advance in Pattern Recognition*, LNCS 1451. Springer, 1998.

[2] A. K. Jain, R. P. W. Duin, and J. Mao. Statistical pattern recognition: a review. *IEEE Trans. on Pattern Analysis and Machine* intelligence, 22(1), 2000.

[3] J. Kittler, M. Hatef, R. P. W. Duin, and J. Matas. On combining classifiers. *IEEE Trans. on Pattern Recognition and Machine Intelligence*, 20(3):226–239, 1998.

[4] K.I. Chang, K.W. Bowyer, P.J. Flynn. Multi-modal 2D and 3D Biometrics for Face Recognition. *Proceedings of the IEEE Workshop on Analysis and Modeling of Faces and Gestures (AMFG)*, 187:194, 2003.

[5] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71-86, 1991.

[6] K.I. Chang, K.W. Bowyer, P.J. Flynn. An evaluation of multimodal 2D+3D face biometrics. *IEEE Transactions on PAMI*, 27(4):619-624, 2005.

[7] C. BenAbdelkader and P.A. Griffin. Comparing and combining depth and texture cues for face recognition. *Image and Vision Computing*, 23(3):339-352, 2005.

[8] http://frav.escet.urjc.es/databases/FRAV3D/

[9] A. E. Johnson. Spin-images: A representation for 3-D surface matching. PhD. Thesis, Robotics Institute, Carnegie Mellon University, 1997.

[10] A. E. Johnson and M. Hebert. Surface matching for object recognition in complex three-dimensional spaces. *Image Vision Computing*, 16, 635-651, 1998.

[11] A. E. Johnson and M. Hebert. Using spin-images for efficient object recognition in cluttered 3D scenes. *IEEE Trans. PAMI*, 21(5), 433-449, 1999.

[12] C. Cortes and V. Vapnik. Support vector network. *Machine Learning*, 20, 273-297, 1995.

[13] T. Joachims. Making large-scale SVM learning practical. *Advances in Kernel Methods: Support Vector Learning*, MIT Press, Cambridge (USA), 1999.

[14] T. Lyche and L. L. Schumaker (eds.). *Mathematical Methods for Curves and Surfaces.* Vanderbilt University Press, Oslo, 135-146, 2000.

[15] C. Conde, Á. Serrano, L. J. Rodríguez-Aragón, E. Cabello. 3D Facial Normalization with Spin Images and Influence of Range Data Calculation over Face Verification. *IEEE CVPR*, San Diego, 2005.

[16] P. J. Besl and N. D. McKay. A Method for Registration of 3-D Shapes. *IEEE PAMI*, 14(2):239-256, 1992.

[17] Z. Zhang. Iterative Point Matching for Registration of Free-Form Curves and Surfaces. *International Journal of Computer Vision*, 13(2):119-152, 1994.

[18] R. Snelick, U. Uludag, A. Mink, M. Indovina and A. Jain. Large-scale evaluation of multimodal biometric authentication using state-of-the-art systems. *IEEE Transactions on PAMI*, 27(3):450-455, 2005.