MULTIMODAL 2D, 2.5D & 3D FACE VERIFICATION

Cristina Conde, Ángel Serrano and Enrique Cabello

Face Recognition and Artificial Vision Group, Universidad Rey Juan Carlos
C/ Tulipán, s/n, Móstoles E-28933 Madrid (Spain)
{cristina.conde, angel.serrano, enrique.cabello}@urjc.es, http://frav.escet.urjc.es/

ABSTRACT

A multimodal face verification process is presented for standard 2D color images, 2.5D range images and 3D meshes. A normalization in orientation and position is essential for 2.5D and 3D images to obtain a corrected frontal image. This is achieved using the spin images of the nose tip and both eyes, which feed an SVM classifier. First, a traditional Principal Component Analysis followed by an SVM classifier are applied to both 2D and 2.5D images. Second, an Iterative Closest Point algorithm is used to match 3D meshes. In all cases, the equal error rate is computed for different kinds of images in the training and test phases. In general, 2.5D range images show the best results (0.1% EER for frontal images). A special improvement in success rate for turned faces has been obtained for normalized 2.5D and 3D images compared to standard 2D images.

Index Terms— *Biometrics, Pattern Recognition, Image processing.*

1. INTRODUCTION

During the last years face biometrics for 2D images has experienced an important emergence, in part due to a demand increase of applications on security and police purposes. There exists a long tradition of research of 2D face recognition [1].

Not until recently have these algorithms been applied to 3D data, in part due to the high cost of 3D digitizers and also due to the high load of work needed to handle these data. The independence from the lighting conditions, as well as the study of a 3D object such as the human face by means of 3D techniques are some of the advantages of these methods respect of 2D cases [2, 3, 4].

The remainder of this paper is organized as follows. In section 2 a short literature review is presented. In section 3 we describe our multimodal face database. In section 4 the face normalization process is explained. In section 5 the verification process for 2D, 2.5D and 3D data can be found. In section 6 we give our results. Finally conclusions are to be found in section 7.

2. PREVIOUS WORK

Several kinds of methods have been applied for 3D face recognition. On the one hand, methods based on local characteristics take advantage of geometrical features of the 3D mesh, where they measure principal curvatures, saddle points and valley lines [5].

On the other hand, methods based on global characteristics have also been used, from properties of the 3D surface, such as intersection with diverse planes and the study of profiles [6], to the face representation with Gaussian images [7]. Some others have compared PCA [8] with other methods: in [9] with ICA, in [10] with Haussdorff distance within surfaces. Mavridis et al. [11] computed 3D eigenfaces, while Chang et al. [12] compared eigenfaces for range images and texture images for face recognition.

Some methods for 3D face recognition are based on the comparison with a template. For example, in [13] ICP is used, as in this work. Others have proposed 3D deformable models [14] or texture information with ASM models [15].

3. FRAV3D: A MULTIMODAL FACE DATABASE

We have used a multimodal database, the so-called FRAV3D. It has been acquired during ten months with 105 volunteers. The totality of the subjects are young adults (18 – 35 years old), Caucasian, with a certain bias towards men (81 males/24 women). Some parts of the database are available from its web page [16].

A scanner MINOLTA VIVID-700 red laser light-stripe triangulation range finder was used under controlled indoor conditions. As a result, both a 3D mesh with up to 4000 points and 7500 triangles and a classical 2D color image were produced. The subjects were asked to sit opposite the scanner, with a dark plain background behind them. No hats, scarves or glasses were allowed. For security reasons, all the participants kept their eyes closed during the acquisition. All scans were acquired using a strict protocol for standardizing reasons. Each shot differed from the previous one in only one acquisition parameter, which included turns, presence or absence of gestures and changes in illumination. In Figure 1 several examples of the FRAV3D face database are shown.

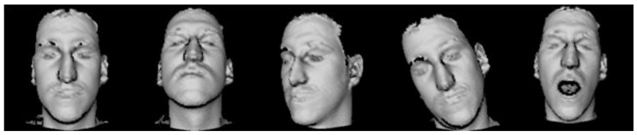

Figure 1 – Sample of 3D meshes from FRAV3D database. From left to right: frontal face, X-axis turn, Y-axis turn, Z-axis turn, face with gesture.

4. AUTOMATIC FEATURE LOCATION AND NORMALIZATION

4.1. Introduction to Spin Images

To find automatically face features in our 3D meshes, such as nose and eyes, a global registration technique called spin images has been used [17], which can be understood as spatial histograms with respect to an origin "oriented" point **p**. This refers to a reference frame defined as the tangent plane P containing **p** and the unitary vector **n** perpendicular to this plane through **p**.

There is a dimension reduction from the 3D coordinates (x, y, z) to a 2D system (α, β) which represent the relative distance between the oriented point **p** and the other points. The spin image transformation S_0 is defined as follows:

$$S_0 : R^3 \to R^2$$

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

where α is the distance from a point **x** to the straight line L parallel to **n** and containing **p**, while β is the distance from **x** to the plane P (Figure 2).

4.2. Localization of nose tip

Spin images are dependent on the point selected as origin. As all human faces look alike in shape and size, this implies that all spin images computed for example for the tip nose will also be similar, even for different people.

We have trained an SVM classifier [18] to detect nose tips. In an iterative process, the most jutting point of the 3D mesh model is identified and classified by means of its spin image until the genuine nose tip is localized.

This method is very accurate and could be also used to search the eyes or other face features. However, due to the associated computational effort, a more intelligent algorithm has to be applied to lighten the load of work. A process divided in two steps is necessary to assure an efficient and fast localization of both eyes.

4.3. Preprocess for Eyes Localization

Once the nose tip has been located, a coarse search of the eyes is performed first only in a reduced area around the nose. In this candidate region, the discrete mean curvature is computed at each point [19]. The eyes and the nose are located at the positions with highest curvature.

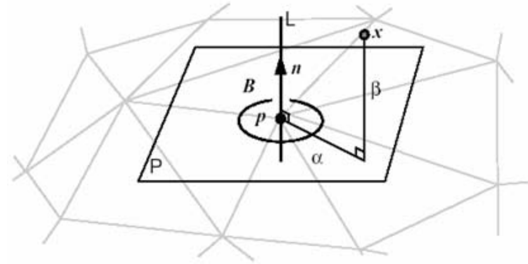

Figure 2 – Parameters of Johnson's spin image [17].

Using a clustering technique based on Euclidean distance, the selected points in the previous step are allocated to three groups or clusters: left eye, right eye and nose.

4.4. Detection of the eyes

A SVM classifier has been trained in order to detect each eye inside corner. For each cluster the deepest points are selected (those with lowest Z coordinate). The corresponding spin image is computed and fed into the classifier. If the output of SVM is negative, a new candidate point is selected. The process is repeated until the lachrymal contained in that cluster has been found.

4.5. Normalization

The face has to be normalized in orientation, this is, the effects produced by any turn with respect to the X, Y or Z axes have to be corrected, in order to obtain a frontal view. The better the normalization, the better verification rates are obtained later. This process can be divided in several steps:

First, the face orientation is corrected for Y axis, by computing the adequate angle to level the depth of both eyes, so that the face does not seem to be in profile.

Second, the face orientation is corrected for Z axis. This is harder to do, because the eye detection algorithm is sometimes not accurate enough. A linear fit for the points between the nose tip and the middle point between the brows has been computed. When this correction is applied, the nose will seem to be vertical.

Finally, the face orientation is corrected for X axis. This is done by computing the straight line that best fits the face in the YZ plane. When this correction is applied, the face will not look upward nor downward, but directly forward.

5. VERIFICATION PROCESS

We have performed a face verification process using three types of data: standard 2D images, range or depth images (also known as 2.5D) and 3D mesh models.

Table 1 – Images used for training and tests per person

No. of test	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)

5.1. Verification using 2D and 2.5D Data

Range data have been computed using the method proposed in [20]. For each pixel a gray level is assigned according to the distance to the laser. In order to optimize this range, an exponential equalization has been applied. Holes because lost points are interpolated in the image.

The verification process used here, both for 2D and 2.5D images, makes use of a Principal Component Analysis (PCA) [8], followed by an SVM classifier [18]. First of all, the face contained in range image is searched by means of a template. It is then cropped in order to reduce its size to 130×140 pixels. This procedure was applied for all the subjects in the database, so a traditional covariance matrix is constructed and the eigenfaces are computed.

We only retain the 150 most significant eigenvalues. With this dimension reduction, every image is projected into the eigenfaces framework. With these projections, an SVM classifier is trained for each person. The SVM output will be used to verify the identity of each subject.

5.2. Verification using 3D Data

For the 3D mesh models, the usual Iterative Closest Point (ICP) algorithm has been used [21, 22]. 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 and has to be rotated and moved adequately so that it adapts to the model with minimal error.

Table 2 – EER computed for all the tests

No. of test	2D	3D	2.5D
1	2.9	4.9	0.1
2	5.1	10.3	4.6
3	12.6	12.6	12.6
4	8.7	3.4	1.0
5	14.6	4.4	1.0
6	33.9	7.4	2.2
7	13.3	5.3	1.9
8	27.0	5.8	4.9
9	41.1	7.4	3.7
10	1.9	3.7	0.5
11	4.9	5.3	0.5
12	4.2	4.2	0.6
13	4.1	6.7	1.7
14	9.2	11.8	10.3
15	1.9	3.7	0.6

By means of an iterative process, the scene is rotated and moved until the fitting error is lower than a certain threshold. The value of this error will be used in the subsequent verification process. The lower the error is, the better coincidence between the model and the scene and the higher the probability that they belong to the same person.

6. RESULTS AND DISCUSSION

Fifteen tests have been carried out. For 2D and 2.5D an SVM classifier has been trained with a set of images (Table 1) and tested with a disjoint set of images. For 3D only frontal meshes have been considered for training and tests.

To verify the reliability of this process, a Receiver Operating Characteristic (ROC) has been computed. The error at which the False Acceptance Rate equals the False Rejection Rate is called Equal Error Rate (EER).

In Table 2 we show the EER for the 15 types of tests for 2D, 2.5D and 3D data. In general 2.5D data produce better results compared with only 2D data. The only exceptions can be obtained in the tests 3 and 14, which make use of images with gestures. It is specially remarkable that tests 5, 6, 7, 8 and 9, which include images with turns, show an important decrease for the EER from around 40% to barely 5% in test 9. This improvement is due to the correction of orientation for 2.5D data, which allows to remove the influence of turns of the face in any direction. Obviously in standard 2D images this correction cannot be made.

Comparing the results obtained for 3D data with ICP, no improvement is obtained. In general a lower EER is obtained for 3D data with respect to 2D, but greater than those of 2.5D. Again for tests 5, 6, 7, 8 and 9, where normalized faces are used, the 3D algorithm yields much better results than in 2D.

For frontal images (test 1), the three methods show similar good results, in this order: 2.5D (0.1% EER), 2D (3% EER) and finally 3D (5% EER).

7. CONCLUSIONS

A study of face recognition for three kinds of images (2D color images, 2.5D range data and 3D meshes) has been carried out for a 105 people multimodal face database (FRAV3D). A traditional PCA followed by an SVM classifier has been applied for both 2D and 2.5D images. As well, an ICP algorithm has been used to compare 3D meshes. A normalization in face orientation has been applied for both 2.5D and 3D images, so that in these cases, the face looks frontal after the correction.

For frontal images, the best performance is obtained for range images (99.9% success rate). Standard 2D color images yield a 97.1% rate and 3D images hit only 95.1%.

Faces showing a smile have better results than faces with the mouth open, as in this case the geometry of the face is not affected so much. 2.5D images have the better results.

Changes of illumination affect mainly 2D data. The performance of 3D images remains the same, and 2.5D also behaves well, with a lower success rate than for frontal images. Turned faces decrease enormously the performance for 2D images. Normalized faces that for 2.5D and 3D images show similar results to frontal faces.

For tests 10 to 14, where several kinds of images have been used in the training phase, an improvement in performance is achieved for 2D and 2.5D images, as the classifier is prepared for more different conditions. As ICP algorithm has no training phase, we obtain similar results to other cases.

In general, better results are obtained for 2.5D range images, where the normalization in position and orientation is essential. As well, we have to point out that our 2D classifier yields better results compared to 3D, although the geometrical information contained in 2.5D images shows to be more robust and powerful than 2D texture only.

ACKNOWLEDGEMENTS

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

8. REFERENCES

[1] W. Zhao, R. Chellappa, P.J. Phillips and A. Rosenfeld, “Face Recognition: A Literature Survey”, *ACM Computing*, Surveys, Vol. 35, Issue 4, December 2003, pp. 399 – 458.

[2] K.W. Bowyer, K. Chang and P.J. Flynn, “A survey of approaches to three-dimensional face recognition”, *International Conference of Pattern Recognition*, Vol. 1, 2004 pp. 358–361.

[3] A. Scheenstra, A. Ruifrok and R. C. Veltkamp, “A Survey of 3D Face Recognition Methods”, *Lecture Notes in Computer Science*, Vol. 3546, 2005, pp. 891 – 899.

[4] J. Kittler, A. Hilton, M. Hamouz and J. Illingworth, “3D Assisted Face Recognition: A Survey of 3D Imaging, Modelling and Recognition Approaches”, *Proc. IEEE Workshop on Advanced 3D Imaging for Safety and Security A3DISS*, San Diego, CA, June 25, 2005.

[5] G. Gordon, “Face recognition based on depth maps and surface curvature”, *SPIE Geometric Methods in Computer Vision*, San Diego, CA, Vol. 1570, 1991.

[6] T. Nagamine, T. Uemura and I. Masuda, “3D facial image analysis for human identification”, *International Conference on Pattern Recognition*, 1992, pp. 324-327.

[7] H. Tanaka, M. Ikeda and H. Chiaki, “Curvature-based face surface recognition using spherical correlation - principal directions for curved object recognition”, *3rd International Conference on Automatic Face and Gesture Recognition*, 1998, pp. 372-377.

[8] M. Turk and A. Pentland, “Eigenfaces for recognition”, *Journal of Cognitive Neuroscience*, Vol. 3, No. 1, 1991, pp. 71-86.

[9] C. Hesher, A. Srivastava and G. Erlebacher, “A novel technique for face recognition using range imaging”, *7th International Symposium on Signal Processing and its Applications*, 2003, Vol 2, pp. 201-204.

[10] G. Pan, Z. Wu and Y. Pan, “Automated 3D face verification from range data”, *International Conference on Acoustics, Speech and Signal Processing*, 2003, pp. 192-196.

[11] N. Mavridis, F. Tsalakanidou, D. Pantazis, S. Malasiotis and M. Strintzis, “The HISCORE face recognition application: Affordable desktop face recognition based on a novel 3D camera”, *Proceedings of the International Conference on Augmented Virtual Environments and 3D Images*, 2001, http://uranus.ee.auth.gr/hiscore

[12] K.I. Chang, K.W. Bowyer and P.J. Flynn, “Multi-Modal 2D and 3D Biometrics for Face Recognition”, Proceedings of the IEEE International Workshop on Analysis and Modeling of Faces and Gestures, 2003, pp. 187.

[13] X. Lu, D. Colbry and A. Jain, “Three-dimensional model based face recognition”, *International Conference on Pattern Recognition*, 2004, pp. 362-366.

[14] V. Blanz and T. Vetter, “Face Recognition Based on Fitting a 3D Morphable Model”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 25, 2003.

[15] A. Ansari and M. Abdel-Mottaleb, “3D Face modeling using two views and a generic face model with application to 3D face recognition”, *Proceedings of the IEEE Conference on Advanced Video and Signal Based Surveillance*, 2003, pp. 37–44.

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

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

[18] C. Cortes and V. Vapnik, “Support vector network”, *Machine Learning*, Vol. 20, 1995, pp. 273-297.

[19] T. Lyche and L. L. Schumaker (eds.), *Mathematical Methods for Curves and Surfaces*, Vanderbilt University Press, Oslo, 2000, pp. 135-146, ISBN: 0-8265-1378-6.

[20] 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”, *Proc. IEEE Workshop on Advanced 3D Imaging for Safety and Security A3DISS, San Diego, CA, June 25, 2005.*

[21] P. J. Besl and N. D. McKay, “A Method for Registration of 3-D Shapes”, *IEEE PAMI*, Vol. 14, No. 2, 1992, pp. 239-256.

[22] Z. Zhang, “Iterative Point Matching for Registration of Free-Form Curves and Surfaces”, *International Journal of Computer Vision*, Vol. 13, No. 2, 1994, pp. 119-152.