3D Facial Normalization with Spin Images and Influence of Range Data Calculation over Face Verification

Cristina Conde, Ángel Serrano,
Licesio J. Rodríguez-Aragón, Enrique Cabello
Universidad Rey Juan Carlos (URJC)
Escuela Superior de Ciencias Experimentales y Tecnología (ESCET)
Face Recognition & Artificial Vision Group (FRAV)
c/ Tulipán, s/n, Móstoles, E-28933, Madrid, Spain
{cristina.conde, angel.serrano, licesio.rodriguez.aragon, enrique.cabello}@urjc.es
http://frav.escet.urjc.es/

Abstract

In this paper face verification techniques have been performed over 3D data acquired by a Laser Scanner. Advantages of 3D face models have been used to perform a normalization task over each face. First, an original method to detect local features points, based on Spin Images, has been developed. Once local features, as the nose tip or eyes corners, have been detected, a normalization process is carried out. After face normalization, different depth maps are calculated using several transform functions to equalize the images. The adequacy of each equalization to face verification has been measured to determine which one emphasizes most the feature discrimination. Face verification has been performed through a Principal Component Analysis and a Support Vector Machine. Final results show the importance of a careful 3D normalization and an optimal election of the depth map towards a improvement in the verification method.

1. Introduction

In the previous years in which prices of 3D digitizer devices have dropped considerably, biometrics using 3D images, in particular 3D face models have reached the great public and are also accessible to many more researchers. Until now 2D face recognition has received attention for long time due to its cheaper acquisition systems and to a long tradition of researching, as can be read in Zhao et al.'s literature survey [1]. Very well known constraints are the great influence of illumination, because shades on the face decrease considerably the performance in face recognition systems, and changes of pose, which provoke different views over the same face.

However, nowadays 3D techniques for face verification can improve our performances, as they use geometrical information on human faces, which really are 3D objects. Removing the influence at changes on pose (no 2D projections are considered) as well as the invariability under diverse lighting conditions, make 3D face recognition very promising. Bowyer et al. have written a review on the current status of these techniques [2].

In this paper we present the preliminary results of a 3D face verification system. Our 3D digitizer provides us with the 3D mesh representations, which we turn into the equivalent 2D depth map or distance to the acquisition system.

When working with 3D images, it is of great importance to achieve high quality face normalization. In order to do so, the feature points have to be found correctly, as they serve as control data for the normalization.

The method proposed here finds the face feature points in real-time and is intended to be integrated into a 3D face verification system. Despite the great computational efforts that are needed to handle 3D data, we have found that our algorithm is fast enough to fulfil our requirements.

The remainder of the paper is organized as follows. Section 2 describes the data acquisition set-up and the database used in this paper. The feature location and normalization method is presented in section 3. Section 4 explains the different depth map calculation techniques. The Face Verification system is described in section 5, and the last section shows the experimental results concluding the paper.

Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)
1063-6919/05 $20.00 © 2005 **IEEE**

2. Data Acquisition Set-up

For our experiments, we have acquired our own 3D face database, the so-called FRAV3D Database. A Minolta VI-700 3D laser light-stripe triangulation rangefinder has been used, which provides a polygonal 3D mesh model [3].

We have scanned 51 human faces, but an enlargement up to 100 subjects is currently in process. Every face was scanned several times. Frontal views were preferred, although little turns were allowed in the acquisition process. Due to these changes in the face pose, normalization has to be done, as we shall explain in the following section.

All the individuals were asked not to wear glasses during the acquisition process, because the laser light beam from our scanner is absorbed by the glass and therefore it did not work properly.

In all cases the images were obtained under controlled laboratory conditions. A regulating stool and a uniform dark background were used in order to obtain clearly homogeneous images and meshes. The distance from the detector to the person was fixed. In fact, the distance to the background was 2 meters long and every subject had to hold its neck against the wall during the acquisition process. Scanner resolution at that distance is at least 1.4 mm in X-Y axis and 2.8 mm in depth (Z axis).

Our scanner provided us in each scan with a VRML file corresponding to a triangular mesh with about 15000 points.

At least part of this database will be available for research purposes in our website [4].

3. Feature Location and Normalization

With our 3D face mesh models, we have computed the so-called Spin Images. These consist on a global registration technique developed by Johnson [5] and Herbert [6, 7]. In this representation, each point belonging to a 3D surface is linked to an oriented point on the surface working as origin. There is a dimension reduction, as from three spatial coordinates (x, y, and z), we obtain a 2D system (α, β), which represents the relative distance between the oriented point p and the other points p_i in the surface. This is similar to a distance histogram respecting to a certain point.

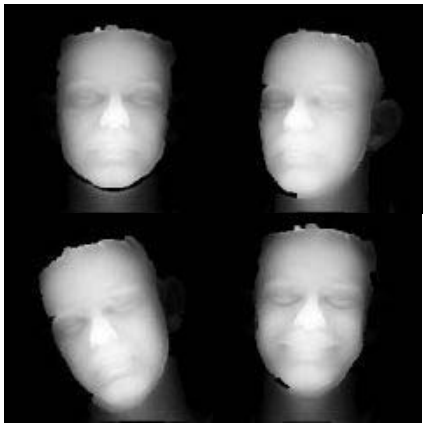

Figure 1 – Samples from our FRAV3D Database, where the depth maps corresponding to a single person are shown for frontal views with no or little turns. These raw images have not been normalized, as we shall explain in section 3.

The Spin-map S_0 components can be computed as follows:

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

Encoding the density of points in the Spin-map, the 2D array representation of a Spin Image can be produced.

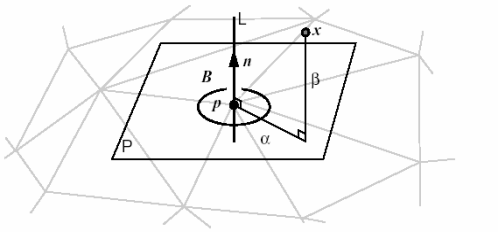

Figure 2 – Parameters of Johnson's geometrical Spin Image [5].

As the Spin images depend on the origin point, different facial points generate particular different images (see Figure 3). We have considered Spin images for the nose-tip and the eyes corners, which provide us with similar images, even for different persons. The main reason for this is that all faces are similar to each other, and therefore, the distribution of relative distances between points does not vary too much.

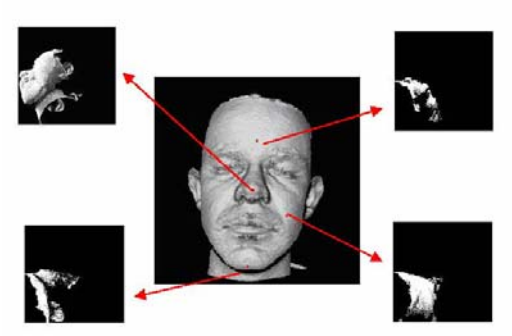

Figure 3 – Examples of Spin Images calculated with respect to different points of the face.

By comparing the Spin images for different facial points, we can select points with a similar geometry. This is a straightforward method to find feature points in a face from a 3D mesh model. In particular, three feature points for each face have been searched: nose-tip, left and right eye inside corners.

In order to do so, an SVM classifier has been trained [8], which allows us to identify these three control points. With them, the size and position of the face can be estimated and later normalized in order to obtain a frontal view of the face, so turns of the face with respect to any axis are corrected.

Despite its accuracy, this method requires a great computational effort. That is the reason why an intelligent point selection must be carried out before computing any Spin Images. In the following subsections, we describe the process in two stages.

3.1. Preprocess: Candidate Areas Selection

First of all, the candidate areas to contain facial points of interest are identified. In our case, three areas are considered, one regarding each feature.

This stage is divided in two steps:

- We compute the discrete mean curvature at each point [9]. The areas of interest are supposed to have a higher curvature, as they contain facial features.
- Using clustering techniques [10] in relation to Euclidean distance, three different clusters are identified, each one containing a feature point (Figure 5).

3.2. Feature Points Selection with Spin Images

Once the candidate areas have been found, using an *apriori* knowledge of the face, one candidate point is selected in each area.

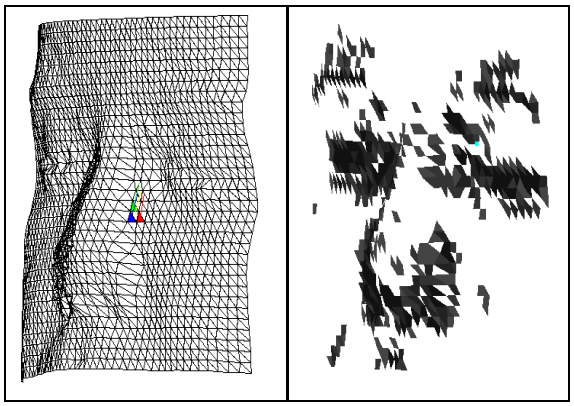

Figure 4 – Areas with a higher mean discrete curvature in the face.

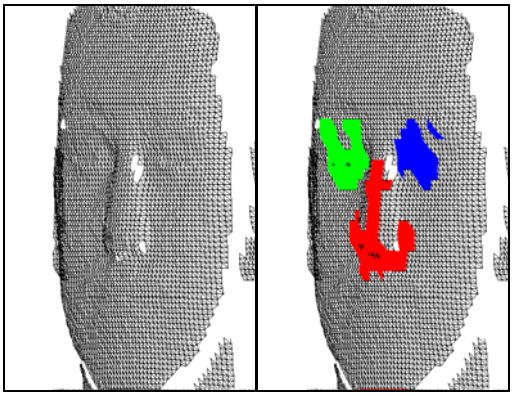

Figure 5 – The three candidate areas containing the searched feature points.

As we have said above, each point produces different Spin images. In order to compare these images for each point, a Support Vector Machine classifier has been used, as it has proven to be very robust even for faces with small displacements. This is the situation of Spin Images calculated at points belonging to different laser captures.

A SVM model for our three feature points has been calculated. It is very interesting to remind that this model can be used for different subjects because it has information which is shared by all the feature points in all faces (all noses have a similar shape, and so on). Therefore it is not necessary to train a new SVM model every time a new subject is added to the database.

On the basis of the classifier output, the candidate point is accepted as facial feature point, or it is rejected and the process is repeated in an iterative way.

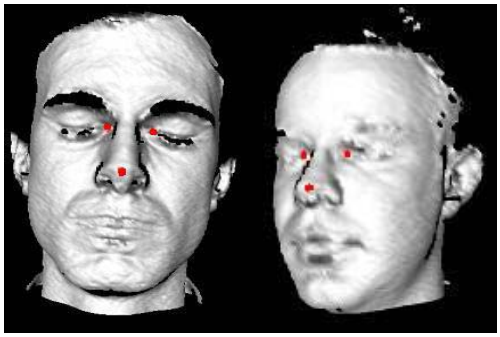

Figure 6 – Location Results (feature points are brought out) in a front capture (left) and rotated capture (right).

After applying the location method above exposed, three feature points are located in a satisfactory way, independently from acquisition conditions (Figure 6). These three points allow us to normalize the face in pose and size.

4. Depth Map Calculation

Once the face is normalized, the depth map is calculated. It is a projection of the 3D data onto a plane that would be the image plane. A gray level is assigned to each pixel based on the depth coordinate of the 3D point that is projected on that pixel.

Despite the high resolution of the laser scanner, in the image plane there could be pixels than do not correspond directly to a 3D point acquired. In this case, it must be calculated the equation of the plane that corresponds to the facet of the mesh where the projection of the pixel is held. The gray level of the pixel will be given by the depth coordinate calculated by this equation. In Figure 7 both cases are shown: directly projection of the 3D point (point A) and depth calculation (point B).

The assignment of a gray level corresponding to a depth distance is relevant for the outcome quality from the depth map. Most approaches apply a linear relation between them [11, 12] but basing on the face geometry, the most discriminate features are around the nose, in a small depth range, so the gray variation should be greater in this range.

In our approach, first a linear relation is applied (equation 2) to calculate the gray level corresponding a z coordinate. The z range between z maximum and z minimum is linearly corresponding to the 255 gray level scale.

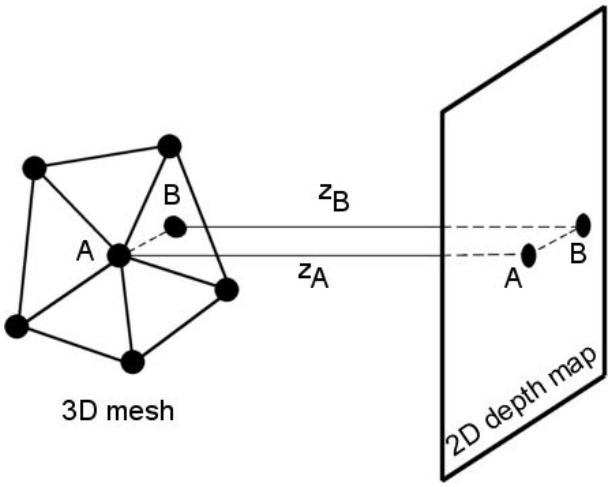

Figure 7 – From a 3D mesh a 2D depth map has to be computed. For a point located at a node of the mesh, the z-coordinate or depth can be computed directly (point A). For a point not located at any node, the equation of the plane of the facet which contains it has to be found before computing the corresponding z-coordinate (point B).

$$Gray(z) = Int\left\{\frac{Int(z - z_{\min}) \cdot 255}{(z_{\max} - z_{\min})}\right\} \qquad (2)$$

Secondly, several equalizations are made to emphasize different depth ranges with a wider or a narrower gray range. Four equalization functions have been applied, with several parameters which can be fitted by the user:

- Linear: image obtained directly from equation 2.
- Exponential centred in z=0 (the tip of the nose), where *a* is the basis and *b* is a bias constant.

$$Gray_{new}(gray_{old}) = a^{gray_{old}} + \mathrm{b} \qquad (3)$$

- Gaussian centred in z=0 (the tip of the nose), with standard deviation σ and mean valueμ.

$$Gray_{new}(gray_{old}) = 255 \cdot \exp\left\{\frac{(gray_{old} - \mu)^2}{2\sigma^2}\right\} \qquad (4)$$

- Linear by sections, where m_1, m_2 and m_3 are the slopes and c_1, c_2 and c_3 are the independent terms.

$$Gray_{new}(gray_{old}) = \begin{cases} m_1 \cdot gray_{old} + \mathrm{c}_1 & gray_{old} < 120 \\ m_2 \cdot gray_{old} + \mathrm{c}_2 & 120 < gray_{old} < 200 \\ m_3 \cdot gray_{old} + \mathrm{c}_3 & gray_{old} > 200 \end{cases} \qquad (5)$$

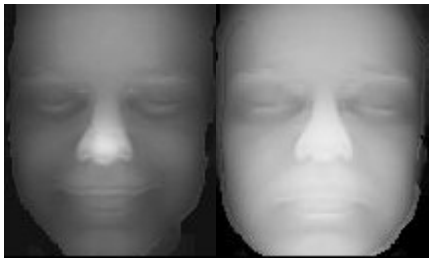

Figure 8 – Depth Maps of the same person with different equalizations: exponential (left) and Gaussian (right).

Figure 9 – Depth Maps of the same person with different equalizations: linear (left) and linear by sections (right).

Figures 8 and 9 show the final depth map calculated applying these different equalizations. The features are differently stood out in each case. Particularly the Gaussian equalization provides a depth map with more discriminated features.

5. Face Verification

5.1. Dimensional Reduction

With the depth maps calculated previously, we perform a traditional Principal Component Analysis (PCA), in the Turk and Pentland's style [13]. Considering N depth maps of size h×w, the d highest eigenvalues and their associated eigenvectors are obtained.

Such method is used because it provides with an important dimensional reduction. Instead of considering all the possible eigenvectors, only the d highest eigenvalues and their associated eigenvectors are selected, where d is lower than N and is typically 150.

After computing the PCA transformation matrix, we calculate the projection of each depth map into the eigenvector subspace. These d components are now fed to a classifier in order to carry out the verification.

5.2. Classification with SVM

We have used a Support Vector Machine classifier (SVM) for our verification process [14], as SVM has proven to be one of the most efficient and robust binary classifiers [15]. Given a set of feature points, SVM computes the optimal hyperplane that best separates them into two classes.

Each SVM classifier requires two stages: training and test. The first one, which is a supervised phase, makes use of the d components from the projections of depth maps into the eigenvector subspace for several individuals. In the test phase, a disjoint set of images are considered. As we know *apriori* whose image they are, we can estimate the performance obtained by our SVM classifiers, so an optimal threshold value for the classification decision can be computed.

For training stage three images per subject are selected. Test is performed with 1 image per subject different to the training set.

6. Results and Conclusions

An original 3D face feature location system based on Spin Images is presented and the influence of the depth map method calculation is measured by a whole face verification system.

We represent our results provided by the face verification system for different depth maps in a Receiver Operating Characteristic curve (ROC) (Figure 10), which plots false negative versus false positive cases.

Table 1 shows the Equal Error Rate (EER) using each set of depth maps calculating by the different conditions explained above. The EER is the value for which false positive rate is equal to false negative rate.

	EER %
Gaussian	2.59
Exponential	4.44
Linear	3.40
Linear by sections	3.21

Table 1 – EER, values for which the rate of incorrect classification of positive verifications is equal to the rate of incorrect classification of negative verifications, for each depth map calculation method.

From the EER values presented for each method, better results are achieved for the Gaussian equalization. Assigning a wide gray range to the depth

area where the features of the face are, improves the verification result in nearly 1%.

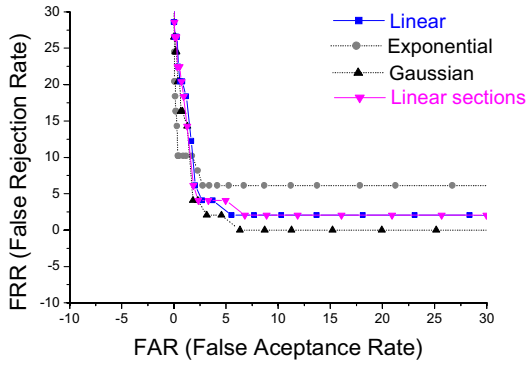

Figure 10 – ROC Curve representing the results for different equalizations.

7. Acknowledgments

Authors would like to thank Jorge Pérez López for his enthusiastic work. Also thanks must be given to every one that offered his help to join FRAV3D database. This work has been partially supported by URJC grant GVC-2004 - 04.

8. References

[1] Face Recognition: A Literature Survey *W. Zhao, R. Chellappa, P.J. Phillips, and A. Rosenfeld. ACM Computing Surveys, Volume 35, Issue 4, December 2003*

[2] A Survey of 3D and Multi-Modal 3D+2D Face Recognition. Kevin W. Bowyer, Kyong Chang, and Patrick Flynn International Conference on Pattern Recognition, August 2004.

[3] "Minolta VIVID 700 3D laser scanner" http://kmpi.konicaminolta.us/vivid/

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

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

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

[7] A. E. Johnson, M.Hebert. Using Spin Images for e.cient object recognition in cluttered 3D scenes.*IEEE Trans. PAMI*.1999, 21(5): 433-449

[8] T. Joachims,"Making large-Scale SVM Learning Practical". *Advances in Kernel Methods Support Vector Learning*, B. Schölkopf and C. Burges and A. Smola (ed.), MIT-Press, 1999.

[9] Mathematical Methods for Curves and Surfaces. Oslo 2000. Tom Lyche and Larry L. Schumaker (eds.), pp. 135–146.Copyright oc 2001 by Vanderbilt University Press, Nashville, TN. ISBN 0-8265-1378-6.

[10] Pattern Recognition. Sergios Theodoridis and Konstantinos Koutroumbas. Academic Press. 1999. Chapter 11.

[11] Kyong I. Chang, Kevin W. Bowyer and Patrick J. Flynn. Face Recognition Using 2D and 3D Facial Data. Workshop in Multimodal User Authentication, pp 25-32, Santa Bárbara, California, December 2003.

[12] Chenghua Xu, Yunhong , Tieniu Tan d Long Quan. Depth vs. Intensity : Wich is More Important for Face Recognition?. Proceedings of the 17th International Conference on Pattern Recognition.

[13] M. Turk & A. Pentland. "Eigenfaces for Recognition". Journal of Cognitive Neuroscience. 1991, Vol. 3, No. 1, p. 71.

[14] Cortes, C. and Vapnik, V.: Support vector network. Machine Learning. 20 (1995) 273–297.

[15] Fortuna, J. and Capson, D.: Improved support vector classification using PCA and ICA feature space modification. Pattern Recognition 37 (2004) 1117–1129