MVA2005 IAPR Conference on Machine VIsion Applications, May 16-18, 2005 Tsukuba Science City, Japan
11-3

3D FACIAL FEATURE LOCATION WITH SPIN IMAGES

Cristina Conde[1], *Roberto Cipolla*[2],
Licesio J. Rodríguez-Aragón[1], *Ángel Serrano*[1] *and Enrique Cabello*[1]

[1]Universidad Rey Juan Carlos (URJC)
Escuela Superior de Ciencias Experimentales y Tecnología (ESCET)
Face Recognition and Artificial Vision Group (FRAV)
c/ Tulipán, s/n, Móstoles, Madrid E-28933, Spain
{cristina.conde, ljrodriguez, aserranos, ecabello}@ urjc.es
http://frav.escet.urjc.es/

[2]University of Cambridge
Department of Engineering, Cambridge, CB2 1PZ
cipolla@eng.cam.ac.uk

ABSTRACT

This paper shows an original 3D facial feature localization method. The localization of feature points in the face is a relevant step that allows the normalization of the data in size and orientation. This stage is essential previous to a range data calculation or a verification process. The localization procedure proposed is based on both, clustering techniques over discrete curvatures calculated, and Spin Images as a global registration method. This method has been tested with a 3D Face Database acquired by a laser scanner in several conditions of position and gesture. The aim is to find three specific feature points: nose-tip and left eye and right eye inside corners. Results show a success rate of 100% and a low computation time allowing the system to work in real time.

1. INTRODUCTION

During the last years, the interest on biometrics has increased largely. These systems can store physical information such us face and iris structure, fingerprint, voice or gait, in order to be used for verification tasks at airports, hospitals, and controlled access areas, among other places where security is a concern.

Of all these types of biometric techniques, face recognition has shown to be one of the most reliable even when collaborative behaviour is not present. Recently, many 2D face recognition applications have been carried out with optimal results obtained for images acquired in good conditions [1]. The main constraints of these techniques are: firstly, the influence of illumination, as the shaded parts of the face may mislead the verification process, and secondly, the changes of pose.

However, the human face is three-dimensional, so projecting it as a 2D object provokes information loss. With the development and improvement of 3D data acquisition devices, especially with the price reduction of laser scanners, 3D face recognition techniques have received more interest. Nowadays, this is one of the most vigorous research areas within biometrics.

Due to the novelty of 3D face recognition techniques, there are not many published results [2]. Two kind of aspects can be already pointed out: firstly the use of range data, translating the 3D information into a 2D depth map or distance to the acquisition system, and secondly, the use of 3D mesh face representation themselves.

As these methods are based mainly on geometric information, in 3D techniques it is even more important than in 2D to carry out a good normalization of the face. In order to normalize the face, the first step is to find the feature points, which will serve as control points for the normalization process.

The method here proposed is based on the localization of facial candidate areas to contain feature points, and its location inside each of these areas.

The system has been designed to perform a real time localization of this feature points. One characteristic of 3D data process is its high computational cost, making even impossible their use under real time conditions. As we shall show, our method fulfils our speed requirements.

2. DATA ACQUISITION SETUP

In order to develop the tests here exposed, a set of 3D facial data have been acquired. Different orientations and gestures have been considered to build a set of models that would help in the development of methods devoted to facial verification. A MINOLTA VI-700 laser light-stripe triangulation rangefinder has been used for the acquisition process obtaining a set of three-dimensional points organized in a triangular mesh for each subject [3].

Captures have been acquired under laboratory controlled conditions; scanner and subject positions were fixed and controlled during all the process; a regulating stool and a uniform background helped in the process of obtaining homogeneous images and meshes. Distance from the scanner to the background was set to 2m and every subject held one's head against the background during the scanning course. Scanner resolution at that distance is at least 5.5mm in X-Y axis and 7.23mm in depth (Z axis). A correlative order in the position and gestures of the subject has been established to require a change in the facial expression between two consecutive acquisitions.

For each scanning process the data provided from the scanner have been processed obtaining a VRML file with about 15000 points organised in a triangular mesh.

During two months, a set of 51 subjects were scanned following the same instructions for each scanning. For each subject 14 captures were performed. The acquired captures were as followed: 6 frontal, 2 with a 5º right rotation and 2 images with a 15º left rotation in the horizontal axis (X axis), 2 captures with a vertical rotation (Y axis) one of them 5º over the scanner and the other under the scanner, and finally 2 captures with gesture expressions (mouth open and smiling). The total amount of captures is 714.

The objective was the construction of a facial database with 3D and 2D information, called FRAV3D. This database is not yet completed as more data are being acquired for new subjects, and, in order to evaluate changes in the facial expression through time, new sets of images are being taken for already scanned subjects. At least a part of this database will be available for research purposes and will be accessible from our website [4].

Some of the problems that appeared within the acquisition period were:

- There are lost points in the acquired captures because of the loss of the laser signal in different areas with hair, or occlusions,.

- The scanner produced some noise and perturbations in horizontal borders as nose and lower chin, even though a noise filter provided by the scanner was active.

3. ALGORITHM DESCRIPTION

Spin Images method is a global registration technique initially introduced by Johnson [5] and further developed in his joint work with Hebert [6,7]. It is based on a local characterization of the surface by images associated with each oriented point on the surface. This image represents the relative distance between the oriented point and the other points in the data. Therefore, it would be similar to a space histogram.

Given an oriented point ***p***, each other point ***p**$_i$* can be defined, with reference to ***p***, by two parameters: α and β (Figure 1). This parameters are represented in the Spin-map S_0:

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

$$S_0(\mathbf{x}) \rightarrow (\alpha, \beta) = (\sqrt{\|\mathbf{x}-\mathbf{p}\|^2 - (\mathbf{n}\cdot(\mathbf{x}-\mathbf{p}))^2}, \mathbf{n}\cdot(\mathbf{x}-\mathbf{p}))$$

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

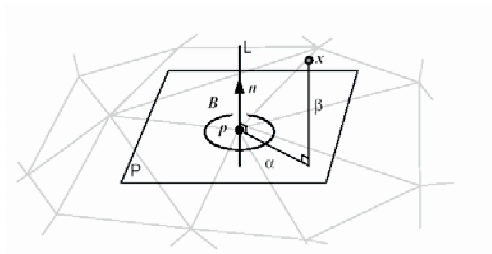

Figure 1– Parameters of Johnson's geometrical Spin Image [5]

As shown in Figure 2, different facial points generate different Spin Images. Points with a specific geometry generate very particular Spin Images. This is the case of the feature points in all faces (nose-tip, eyes corners, etc.), all of them with a very special geometry, even for different persons.

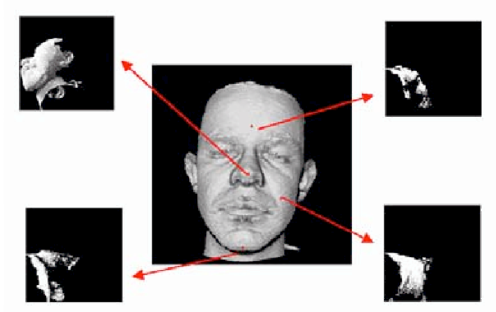

Figure 2– Spin Images examples calculated in different points of the face

This technique was selected on account of its simplicity and suitableness to find points with a specific geometry like facial feature points.

Through the comparison of the Spin Images calculated at different points of several faces, points with a similar geometry can be selected. As result, the facial feature points are located. Three feature points for each face have been searched: nose-tip, and left and right eye inside corners.

The comparison of different Spin Images has been performed by a SVM classifier [8] . The localization of these three points allows us to estimate size and position of the face, and to carry out the normalization for a further recognition stage.

Spin Images have proven to be an accurate method in this task, but it has appeared computationally expensive, requiring an intelligent point selection before proceeding to compute Spin Images. A preprocess step has been introduced to select candidate areas that contain the feature points. This is a relevant stage in the final goal of designing a 3D feature points location method that could work in real time.

Two stages can be identified: preprocess and feature points selection.

3.1. Preprocess: Candidate Areas Selection

At this stage a selection of the candidate areas to contain facial feature points has been done. As three feature points have been considered, three areas are found after this stage, one regarding each feature. Two steps have been involved in this task.

- First, the discrete mean curvature at each point is calculated [9] and a selection of the areas with a higher curvature is made. Areas with a higher curvature in a face correspond to the feature areas (Figure 3).

- Second, clustering techniques [10] in relation to euclidean distance are applied to the points selected in the previous step. The data are then organized in three different clusters, each one containing a feature point (Figure 4).

3.2. Feature Points Selection with Spin Images

Once the candidate areas have been found, using an a-priori knowledge of the face, one candidate point is selected in each area. This point is characterized by the corresponding Spin image that describes the position of the rest of the points with reference to this candidate point.

Different methods have been considered to compare Spin Images that correspond to different points. After several tests, a Support Vector Machine classifier has been selected. This binary classifier is robust and has the capability to recognize the global structure of an image.

Even images that represent the same object with different details like small displacements are classified as equivalent. This is the situation of Spin Images calculated at points belonging to different laser captures.

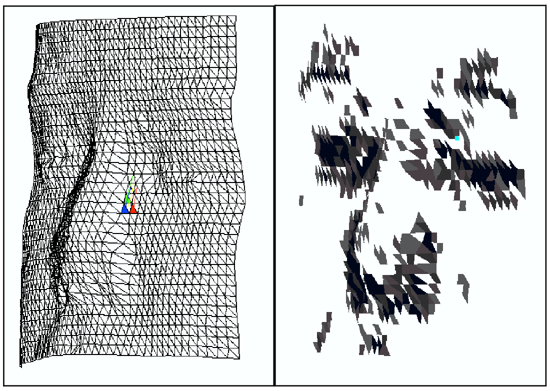

Figure 3– Areas with a higher mean discrete curvature in the face

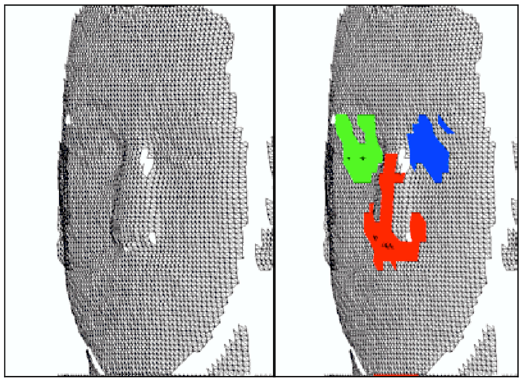

Figure 4 – The three candidate areas containing the searched feature points

An SVM model for each feature point has been generated through a supervised training. It is important to remark that this model is valid for different subjects, because it contains the geometrical information of the feature point, which is common to all the faces. For instance, all noses have a similar shape, consequently 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.

4. RESULTS

The method above exposed has been applied to the database described in section 2.

The database has been split in two different sets: one made up of 2 front captures and one rotated around X axis per person, and a second set with the rest of captures. The training step was performed using the first group while the testing step used the whole database.

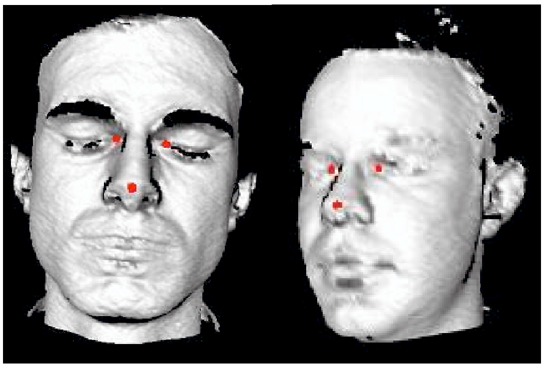

Figure 5 – Location Results (feature points are brought out) in a front capture (left) and rotated capture (right)

After applying the localization method above exposed, the results show a global 98.65% success rate over the 714 captures. Table 1 shows the number of incorrect locations for each acquisition condition of the database in percentage over the whole captures in the same conditions. In case of correct location, three feature points are located in a satisfactory way, independent from acquisition conditions (Figure 5).

	Incorrect Location (%)
Front	0.33
Light rotation	2.61
Strong rotation	3.92
Gesture	0

Table 1 – Incorrect Location rate over the whole captures in the same acquisition conditions.

5. CONCLUSIONS AND FUTURE WORK

An original 3D feature location method has been proposed. It has been tested in real conditions with a 3D database (FRAV3D), at present composed by 51 people (714 total captures) in different pose conditions. The results show an accuracy rate of 98.65% in the location of three facial feature point (nose-tip, left and right eye inside corners).

Our method is robust under real environmental conditions: rotations and gestures, small lost points and noise due to laser scanner acquisition process. Rotation or gesture conditions do not strong affect the localization process as it is based on the local characterization of points through the Spin Images method.

The SVM classifier makes the system independent from small loss of points, such as the ones caused by laser scanner limitations at hairy surfaces or occlusions in rotated faces. The classifier is strong enough to recognize whether a Spin Image belongs to a feature point, even if a part of the image is lost .

Incorrect location cases are presented when there is an important loss of points during acquisition process and the feature points are not acquired by the scanner.

The location method presented is not dependent from the mesh resolution if the quality is equal during the train and test step of the classifier.

It is important to remark the system capability to work under real time conditions. The processing time was shorter than ten seconds even for the worst situation met during test step.

We consider this research as the first step towards the designing a 3D face verification system. Currently the process acquisition of the FRAV3D is being held to enlarge the database for future researches.

6. REFERENCES

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

[2] Kevin W. Bowyer, Kyong Chang, and Patrick Flynn "A Survey of 3D and Multi-Modal 3D+2D Face Recognition" *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 efficient 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] Tom Lyche and Larry L. Schumaker (eds.), "Mathematical Methods for Curves and Surfaces." *Oslo 2000. pp. 135–146.Copyright 2001 by Vanderbilt University Press, Nashville, TN. ISBN 0-8265-1378-6.*

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