ACCURACY OF 3D FACE RECOGNITION FRAMEWORKS

V. Bevilacqua [a, b], M. Caprioli [c], M. Cortellino [b], M. Giannini [a, b], G. Mastronardi [a, b], V. Santarcangelo [b]

[a] DEE (Dipartimento di Elettrotecnica ed Elettronica) Polytechnic of Bari
Via Orabona, 4 – 70125 Bari, Italy

[b] e.B.I.S. s.r.l. (electronic Business In Security) Spin-Off of Polytechnic of Bari
Via Pavoncelli, 139 – 70125 Bari, Italy

[c] DVT (Dipartimento di Vie e Trasporti) Polytechnic of Bari
Via Orabona, 4 – 70125 Bari, Italy
m.caprioli@poliba.it

KEY WORDS: CAD, 3D Geometric Modelling, Image processing.

ABSTRACT

This paper represents a survey of the state of art reached in 3D Face Recognition frameworks and show some different approaches developed and tested by its authors. We have designed a strong algorithm that is based on genetic algorithms, Principal Component Analysis (PCA) and face geometry assumptions, for head pose normalization of 3D scanned face models. Experiments conducted on the GavaDB database show a 100% success rate in correctly those models that “stare” at the camera (with a perfect alignment in 83% of the cases). A previously developed algorithm for nose-tip detection based on an adapted Khoshelham GHT has been used to create an automatic repere’s points detection system with the purpose of obtaining a biometric system for AFR (Automatic Face Recognition) using 3D Face templates. Subsequently two different methodologies, based respectively on an unsupervised self-organizing neural network (SOM) and upon a graph matching, have been implemented to validate the performance of the new 3D facial feature identification and localization algorithm. Experiments have been performed on a dataset of 23 3D faces acquired by a 3D laser camera at eBIS lab with pose and expression variations. Then an optimization of the search of the points ALS and ALD of the nose and a new graph approach for the recognition base on several new points has been implemented. Experiments have been performed on a dataset of 44 faces, acquired by a 3D laser camera at eBIS lab with pose and expression variations, scoring a result of 100% of recognition.

1. INTRODUCTION

In Computer Vision object or shape detection in 2D/3D images is very hard to solve because shapes can be subject to translations, can change by color, can be subject to scale and orientation, can endure occlusions and moreover data acquisition can introduce high levels of noise. Always more and more researchers work on 3D face processing including modelling and recognition. Usually, a 3D face is a group of high dimensional vectors of the x, y and z positions of the vertexes of a face surface. Face recognition based on 3D has the potential to overcome the challenging problems caused by the expression, illumination variations. However, many 3D face recognition approaches, especially the feature based ones require a robust and accurate facial feature localization. To the best of our knowledge, most of the methods do not use benchmark datasets to evaluate their results. Romero et al. presented the first work on benchmark datasets based on FRGC database. They manually marked landmarks of eleven facial features including nose tip and eye corners. With those marked feature locations, the results of automatic feature identification can be measured and evaluated. This paper focuses on a different task and proposes two different steps where the former concerns the automatic identification and localization of a 3D Nose facial features, and the latter is the validation of the correct data obtained by means two independent 3D face recognition tasks. Experiments was performed by using an ASE format dataset of 23 images acquired in by a 3D laser scanner based on structured light with a resolution of 640 by 480. The applied 3D reconstruction method is based on analysis of two images, obtained illuminating sequentially the target surface, with two sinusoidal fringe patterns shifted in phase by 180° and the obtained data are a 3D clouds of point in ASE format. Due to the short integration time, images may be acquired without room darkening (normal ambient illumination). Optical methods used for measurement of 3D geometry are based on two general principles, one related to the almost direct measurement/analysis of the light flight time (lidars, interferometers and holography) and the second related to geometrical triangulation. In the last group the viewing angles of a point of the object can be calculated from its position in the registered image/images or from the applied system settings (scanning devices). Single image acquisition with a projected fringe pattern of high spatial frequency is another kind of compromise towards one-shot entire scene 3D photography. This type of acquisition is usually related to the Fourier analysis of the image. As the main drawback of this approach is its sensitivity to the quality of the registered image (fringe pattern), which directly influences the precision of the phase unwrapping procedures as well as the final accuracy. Another weak point of this method is its intrinsic low resolution and artefacts caused by the extensive filtering applied in the spatial frequency domain, and in particular, in zones close to the analyzed lobe borders. As nose tip is the most prominent feature of the face, most of the previous work perform nose tip detection and uses the nose tip as the foundation to detect other features. However, many previous facial feature identification algorithms use an assumption that the nose is the closest point to the camera or device which acquires the 3D data. Although this supposition is true in most cases, there is no 100% guarantee due to the noise. Various pose rotations and the complex situation of hair and clothes could make some places closer than the nose. Some algorithms make use of the corresponding 2D texture information to detect the face area first then localize the nose tip within the selected 3D face crop. This technique requires 2D texture and 3D shape to correspond correctly. However, for example, in some face datasets such as the Spring2003 subset of FRGC, the 2D texture channel is not always perfectly matched with the 3D shape channel. In (Bevilacqua, V., Casorio, P., Mastronardi, G., 2008) the authors addressed automatic nose tip localisation adapting Khoshelham GHT and describe an automatic repere point detection system with the purpose of obtaining a biometric system for AFR (Automatic Face Recognition) using 3DFace templates. That research was lead on a database of 3D-faces in ASE format, the GavaDB, given by the GAVAB research group of computer science department at the University of King Juan Carlos in Madrid and the authors show their results and claim successful localization rate of nose tip by means several correspondences in terms of very close results obtained by different algorithms but using a limited dataset without benchmark evaluation. Then starting from the 3D nose tip co-ordinates, obtained running the same previous code developed to automatically localize nose tip in ASE format cloud of points, in (Bevilacqua, V., Mastronardi, G., Santarcangelo, V., Scaramuzzi, R., 2010) (Bevilacqua, V., Mastronardi, G., Piarulli, R., Santarcangelo, V., Scaramuzzi, R., Zaccaglino, P., 2009) the authors, propose firstly a new algorithm to localize other four 3D nose features, and two other different approaches to perform a 3D face recognition by using all the five nose points. In (Fazl-Ersi, E., Zelek, J. S., Tsotsos, J. K., 2007) a novel 2D face recognition method is proposed, in which face images are represented by a set of local labelled graphs, each containing information about the appearance and geometry of a 3-tuple of face feature points, extracted using Local Feature Analysis (LFA) technique. That method automatically learns a model set and builds a graph space for each individual, then proposes a two-stage method for optimal matching between the graphs extracted from a probe image and the trained model graphs is proposed and achieves perfect result on the ORL face set and an accuracy rate of 98.4% on the FERET face set.

2. PREVIOUS WORKS AND BACKGROUND

2.1 Stereo Matching

In the first work (Bevilacqua, V., Mastronardi, G., Melonascina, F., Nitti, D., 2006) has been proposed a passive intensity based stereo-matching algorithm using a constraint handling GA to search matched points. Approach used search correspondences on corresponding epipolar lines (not on the whole image), then, selected N points on the epipolar line of the first image, N points on the corresponding epipolar line in the second image are researched, the research is carried out for each couple of epipolar lines. Two stereo-matching algorithms have been proposed: for generic scenes using images from parallel cameras (or rectified images), for 3D face reconstruction using images from parallel or non-parallel cameras (camera calibration is required to compute epipolar lines). In the last case 3D reconstruction process has been implemented using these steps: Stereo-matching, calculation of 3D coordinates from matched points using triangulation, generation and visualization of a 3D mesh.

Figure 1. 3D Reconstruction through Stereo Matching.

2.2 Hough Transform

In previous work (Bevilacqua, V., Casorio, P., Mastronardi, G., 2008) we adapted Khoshelham GHT in an attempt to apply it on 3D-Face shaded model for nose-tip detection. In that way it was possible to create an automatic repere's points detection system with the purpose of obtaining a biometric system for AFR (Automatic Face Recognition) using 3DFace templates. The research was lead on a database of 3D-faces in ASE format, the GavaDB, given by the GAVAB research group of computer science department at the
University of King Juan Carlos in Madrid. One of the more effective solutions for shape detection is the Hough Transform. Formulated for the first time in early '60s, it originally was able to recognize shapes that had analytical description such as straight lines, circles and ellipses in 2D intensity images. In 1981 Ballard (Ballard, D. H., 1981) proposed an extension defined Generalized Hough Transform (GHT) for generic shape detection by using the R-Table, a table that describes the shape to search respect to a reference point that could represent the center of the . Many efforts have been done in order to try to extend the GHT in three- pattern dimensional images. Khoshelham (Khoshelham, K., 2007) proposed an extension of Ballard GHT for three-dimensional images constituted by point clouds obtained by means of laser- scanner acquisitions, for generic applications.

2.3 3D Head Pose Normalization

In (Bevilacqua, V., Andriani, F., Mastronardi, G., 2009) we have first proposed a model reconstruction scheme for human head's point cloud, as it's directly returned from scanning hardware, consisting of a simple pipelining of existing, fairly renowned algorithms. Then we have designed a two-step process for model alignment, based on two different error measures and their corresponding minimization techniques. In particular we presented a software system for fully automatic alignment of the 3D model of a human face. Starting with the point cloud of a human head, previously segmented from the background, pose normalization is attended with a novel, purely geometric approach. In order to solve the 6 degrees of freedom of this problem, we exploit natural mirror symmetry of human faces, then analyse frontal profile shape and finally align model's bounding box with nose tip position.
Normalizing the position of a human face implies achievement of a perfect (or quasi-perfect) alignment of the head 3D model

wrt some kind of reference, so that subject's transformed model appears to be staring at the camera. This problem exposes the following six degrees of freedom: three angles of rotation (referred to as pitch, yaw and roll) and three measures of bounding box translation (xc, yc, zc), respectively relative to the three main axes (X, Y and Z), as shown in Figure 2:

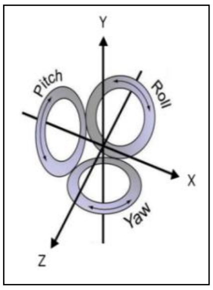

Figure 2. Main axes and corresponding rotation angles representation.

The whole algorithm is conceived as a two-step optimization process. In the first phase we exploit natural vertical symmetry of human face, acquiring knowledge about the yaw and roll angles, along with the xc translation parameter. We designed a simple measure of the asymmetry between the two halves of a human head defined by intersection of its 3D model with a hypothetical symmetry plane, and called it Mean Symmetry Distance (MSD). It is obtained by projecting a bundle of parallel lines, perpendicular to the given plane, and intersecting it with the triangle mesh of the head. Then, for each line in the bundle, intersecting points closest to the plane are selected by each side (two total points) and the absolute difference between the distances of these points from the plane itself is computed, if applicable (that is, if the projected line actually intersects each part of the head at least once). Finally, the mean of these absolute differences will be assumed as a symmetry error value (MSD) for the considered plane. The density of the line bundle may vary according to desired accuracy. In the second phase we lean on natural vertical development of the central profile shape of human faces, gaining clues about the pitch angle. First, a sampling of the profile is extracted, resulting in a quasi-continuous vector of points, then the parameters of a first order equation are adapted for the best possible approximation of this vector. In other words, we try to find the line that best fits the shape of the profile, describing it by its explicit slope-intercept parameters. The error measure between the profile and its estimate, named Mean Profile Distance (MPD), is straight-forwardly computed as the mean of point-line distances between the given linear equation and the sampling vector.

Figure 3. Extracted frontal profile shape and possible approximating lines, with the right one intuitively having a heavily lower mpd value.

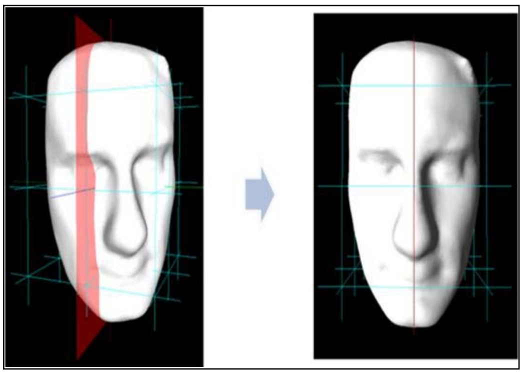

Figure 4. Head 3D model before (left) and after (right) MSD minimization process.

3. 3D FACE ENROLLMENT AND RECOGNITION

For the acquisitions of ".ase" files has been used the MainAxis's "3D CAM SURFACER", a 3D laser scanner based on structured light with a resolution of 640 by 480, which is a good resolution for 3D processing. 3D acquisitions can be affected by random noise that consists of spikes and holes (Bowyer, K., Chang, K., Flynn, P., 2006), so has been necessary a pre-processing phase. The presented technique of 3D vectorial photography is a potentially low cost and has solid-state robustness. It can be used as an alternative for other scanning or multi-image techniques. For distances ranging from1 to10miters the same order precision with the high acquisition rate. Furthermore, it has no moving parts(excluding light source cool fan),and uses white light illumination. In spite of two-acquisition based principles it permits registration, with a normal speed camera, of 3D images of objects moving with the velocity up to1m/sec.in darkened laboratory and up to20mm/sec.in normal ambient illumination conditions.

3.1 Individuation of nasal pyramid points

We have implemented the algorithms to individuate the point of the nasal pyramid.

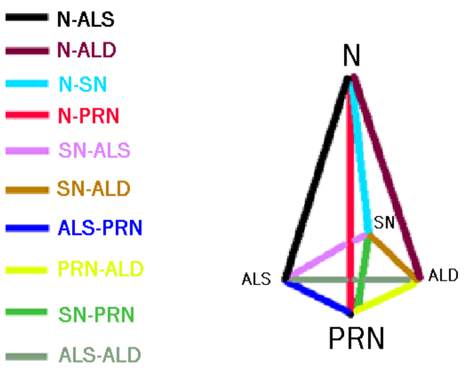

Figure 5. Nasal Pyramid

Connecting those five points we have been able to determine nose graph characterized by ten lengths. We have used a graph matching method (Fazl-Ersi, E., Zelek, J. S., Tsotsos, J. K., 2007) for the recognition. This is a geometric identification method and starts with the consideration of nose graph (G), full graph K_5 that has as vertexes V(G) = {P, PRN, SN, ALS, ALD} and as edges E(G) = {a, b, c, d, e, f, g, h, i, l}.
This graph has been considered as weighed graph where weighs associated are the distances. Then, we have considered the dissimilarity between the geometry of the graphs as (Bevilacqua, V., Andriani, F., Mastronardi, G., 2009):

$$Z(G_i, G_j) = \frac{1}{n} \sum_{k=1}^{n} \frac{(e_{ik} - e_{jk})^2}{(e_{jk})^2}$$

where $G_i = \{(e_{i1}, e_{i2}, \ldots, e_{in}\}$ (MODEL GRAPH) and

$G_j = \{(e_{j1}, e_{j2}, \ldots, e_{jn}\}$ (TEST GRAPH).

This method works considering G_j as the graph to identify, so Z-coefficients are calculated for all the graphs of the database. At the end the graph G_i that minimizes better Z is associated to G_j.

3.2 PRN Matching

Differently from previous approach we have implemented other algorithms for the search of points of repere, considering also the points of the eyebrow arched. An unstructured organization of points as that obtained from ASE file is complicated to manage, because it is impossible to move easily in the cloud. In this context has been useful to have the points organized in a YY matrix where each position has the relative Y value. This matrix has been obtained through the creation of a polygonal mesh that interpolates the terns of the point cloud. This approach permits a more easier scansion of the surface of the face using the easy management of the data in the matrix structure. The search of points of repere has been done by the scansion of the face made by the use of a “sliding vector” on the polygonal mesh, in order to determine the geometric-statistic features typical of the face. The “sliding vector” is an observation window that contains some elements of the YY matrix that at every step “scrolls” along the particular direction of movement.
This methodology allows the individuation of the following points:

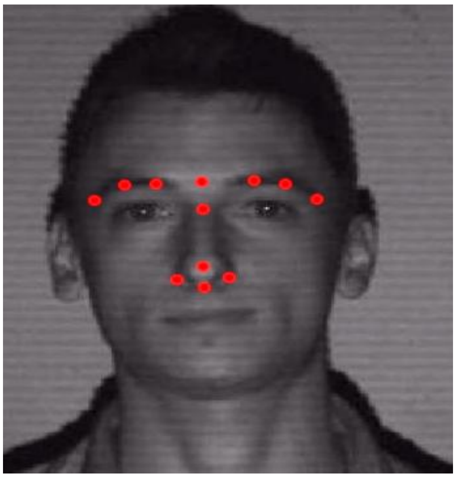

Figure 6. Map of points of repere detected

For the recognition has been implemented PRN Graph Matching. This approach is based on the matching of the distances calculated regarding a point of reference (PRN point). In this technique are considered 9 points of reference and PRN for a total of 9 distances to match.

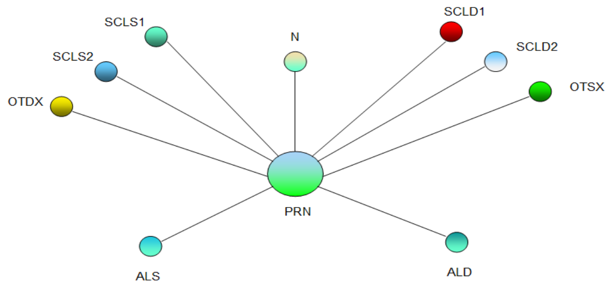

Figure. 7. PRN Graph

The technique as that of K5 Graph Matching (Bevilacqua, V., Mastronardi, G., Piarulli, R., Santarcangelo, V., Scaramuzzi, R., Zaccaglino, P., 2009) is based on the use of the dissimilarity coefficient:

$$Z(G_i, G_j) = \frac{1}{n} \sum_{k=1}^{n} \frac{(e_{ik} - e_{jk})^2}{(e_{jk})^2}$$

where $G_i = \{(e_{i1}, e_{i2}, \ldots, e_{in}\}$ (MODEL GRAPH)

and $G_j = \{(e_{j1}, e_{j2}, \ldots, e_{jn}\}$ (TEST GRAPH),

where e_{ij} are the edges.

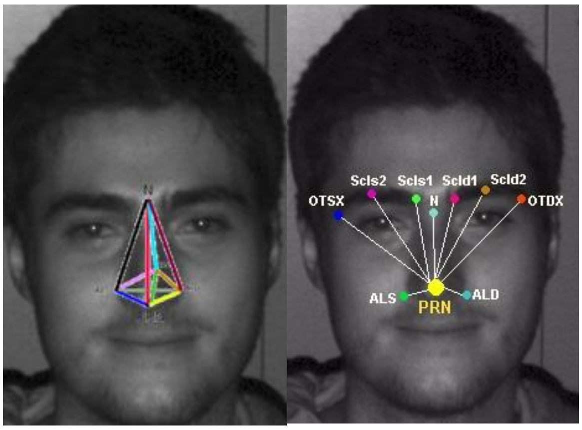

Figure 8. K5 Graph Matching and PRN Graph Matching

Figure 9. 2D images of the two twin brothers

Besides to consider only the distances between points and PRN, another difference regarding K5 Graph Matching is not considering the SN point because, being the SN-PRN distance little important, the performance get worse considering it.

3.3 Twins benchmark

During the acquisition phase has been possible acquire the faces of two twin brothers. It’s very difficult for a person that doesn’t know them to distinguish these two boys (in Fig.9) , but acquiring their 3D faces and obtaining the nose distances with our algorithms has been possible to notice that the two twins are different for a nose distance (Fig. 10).

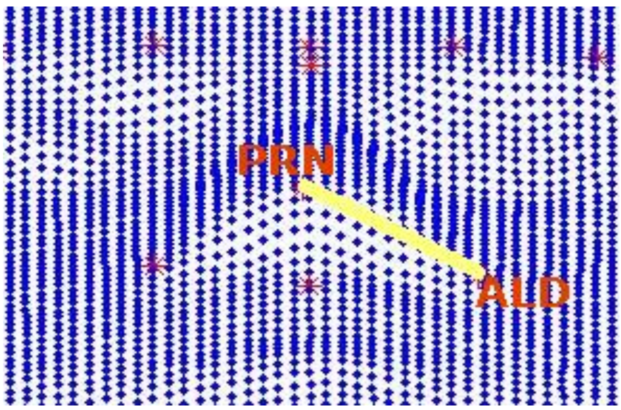

Figure 10. The PRN-ALD distance

Considering a dataset characterized by 11 acquisitions for each brother the number of misclassified acquisition has been of only two items.

4. CONCLUSIONS AND EXPERIMENTAL RESULTS

We have considered a dataset of 5 people A, B, C, D, E, and precisely 11 acquisitions of A, 6 acquisitions of B, 10 acquisitions of C, 9 acquisitions of D and then 8 acquisitions of E for a total of 44 acquisitions. Using K5 Graph Matching the performances are of 77,2%, while considering PRN Graph Matching the performances of correct recognition are of 100% showing interesting results that anyway need some further investigations.

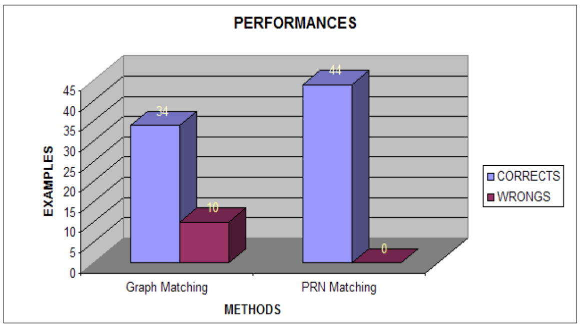

Figure 11. Performances

5. REFERENCES

Ballard, D. H., 1981: Generalizing the Hough Transform to Detect Arbitrary Shapes, Pattern Recognition, Vol. 13, No. 2, 1981, pp. 111-122

Bevilacqua, V., Mastronardi, G., Melonascina, F., Nitti, D., 2006: Stereo-Matching Techniques Optimisation Using Evolutionary Algorithms, D.-S. Huang, K. Li, and G.W. Irwin (Eds.): ICIC 2006, LNCS 4113 © Springer-Verlag Berlin Heidelberg 2006, pp. 612-621

Bevilacqua, V., Andriani, F., Mastronardi, G., 2009: 3D Head Pose Normalization with Face Geometry Analysis, Genetic Algorithms and PCA, Journal of Circuits, Systems and Computers, December 2009, Vol. 18, No. 8, pp. 1425-1439

Bevilacqua, V., Casorio, P., Mastronardi, G., 2008: Extending Hough Transform to a Points’ Cloud for 3D-Face Nose-Tip Detection, D.-S. Huang et al. (Eds.): ICIC 2008, LNAI 5227 - © Springer-Verlag Berlin Heidelberg 2008, pp. 1200-1209

Bevilacqua, V., Mastronardi, G., Santarcangelo, V., Scaramuzzi, R., 2010: 3D Nose Feature Identification and Localisation Through Self-Organizing Map and Graph Matching, Journal of Circuits, Systems and Computers, 2010, Vol. 19, N. 1, pp. 1-12

Bevilacqua, V., Mastronardi, G., Piarulli, R., Santarcangelo, V., Scaramuzzi, R., Zaccaglino, P., 2009: Experimental Comparison among 3D innovative face recognition frameworks D.-S. Huang et al. (Eds.): ICIC 2009, LNAI 5755 - © Springer-Verlag Berlin Heidelberg 2009, pp. 1096-1105

Bowyer, K., Chang, K., Flynn, P., 2006: A Survey of Approaches and Challenges in 3D and Multi-Model 3D+2D Face Recognition, Computer Vision and Image Understanding, January 2006, Vol. 101, No. 1, pp. 1-15

Fazl-Ersi, E., Zelek, J. S., Tsotsos, J. K., 2007: Robust Face Recognition through local graph matching. Journal of Multimedia, Vol. 2, No. 5, September 2007, pp. 31-33

Khoshelham, K., 2007: Extending Generalized Hough Transform to Detect 3D Objects in Laser Range Data. Optical and Laser Remote Sensing Research Group, Delft University of Technology, Kluyverweg 1, 2629 HS Delft, The Netherlands; ISPRS Workshop on Laser Scanning 2007 and SilviLaser 2007, Espoo, September 12-14, Finland (2007)