Face Recognition Using 3D Local Geometrical Features: PCA vs. SVM

Ana Belén Moreno, Ángel Sánchez, José Fco. Vélez and Fco. Javier Díaz
Escuela Superior de CC. Experimentales e Ingeniería, Universidad Rey Juan Carlos, Campus de Móstoles, C/ Tulipán s/n; 28933 Móstoles; Spain
{belen.moreno, angel.sanchez, jose.velez, francisco.diaz}@urjc.es

Abstract

Thirty local geometrical features extracted from 3D human face surfaces have been used to model the face for face recognition. They are the most discriminating ones selected from a set of 86. We have experimented with 420 3D-facial meshes (without texture) of 60 individuals. There are 7 images per subject including views presenting light rotations and facial expressions. The HK algorithm, based in the signs of the Mean and Gaussian curvatures, has been used for region segmentation. Experiments under controlled and non-controlled acquisition conditions, considering pose variations and facial expressions, have been achieved to analyze the robustness of the selected characteristics. Success recognition results of 82.0% and 90.16% were obtained when the images are frontal views with neutral expression using PCA and SVM, respectively. The recognition rates only decrease to 76.2% and 77.9% using PCA and SVM matching schemes respectively, under gesture and light face rotation.

1. Introduction

The need of security and fraud control applications to establish personal authentication, has caused the increase of research in biometric systems [1]. Automatic face recognition is one of the less intrusive biometric modality, which has increased the interest [2]. It has many applications in areas like: personal identification, security applications, law enforcements and smart environments, among others.

Traditionally, research in face recognition was focused on 2D intensity images. Under this context, the recognition accuracy is sensitive to the lighting conditions (position and kind of the light source), expressions, head orientation (rotation), and/or a variety of elements such as hair, moustache, glasses, etc. Automatic Face Recognition using 2D images provides excellent results when the image acquisition conditions (illumination, pose and face variations) are controlled [3]. Recent efforts are oriented to reduce image acquisition restrictions [4]. Some methods have been proposed to tackle non-controlled variations of pose and illumination, but they do not work well in arbitrary conditions [5].

Working with 3D face images has some advantages over 2D face images: (1) more geometrical information can be obtained from the 3D data than from 2D images because 2D images loose depth information as they are formed through projections of 3D objects, (2) the measured features from real 3D data are not affected by the scale and rotation and (3) if the 3D face recognition system does not consider texture information, the recognition is immune to the effect of illumination variations. The advances in computational processing capabilities as well as the reduction of both cost and size of the 3D digitizers, have contributed to the development 3D face recognition systems.

Recently, the interest in model-based 3D automatic face recognition systems has increased [6][7]. One common technique on 3D object recognition is based on the matching among 2D image points and the corresponding ones of a generic head model as a previous stage to infer information. Another technique of representation of 3D range face images using local information is the in computing the point signature over determinate 3D points in order to obtain descriptors. The point signature has been used in the scope of 3D face recognition [8]. Differential geometry has been used for feature extraction in the context of free-form three-dimensional object recognition and also used by some authors for facial feature extraction [9]. The local curvature of a 3D facial surface, as well as angle among surface normal vectors evaluated in a 3D point have been proposed as 3D free-form object descriptors in the scope of 3D face recognition. An antecedent of this work is [9], where a set of twelve 3D features extracted from curvature based segmented face regions, were used for face recognition over a reduced set of individuals.

Some recent approaches [10][11] consider principal component analysis (PCA) to obtain a low dimensional representation of 3D images consisting in depth maps of the complete face. The objective is to evaluate the influence of color, depth and the combination of both in face recognition. These works compared the use of 2D images and their corresponding 3D depth maps, both independently and combined, concluding that the combination of 2D and 3D information provided better performance. These systems have not considered images presenting facial expressions. Pose variations have been limited [11] or not considered [10]. Depth and texture play complementary roles in the coding of faces [12].

Two aspects that characterize a face recognition system are: (1) representation or face modeling and (2) recognition (or matching) technique. Face modeling transforms the information of the facial image into a set of characteristics that represents or models the original data. Matching scheme involves the used method to select the best match from the set of identities of the face database. While the matching scheme can be efficiently implemented using standard machine learning techniques like Neural Networks or Support Vector Machines, 3D Face Modeling is still an open problem [12]. This paper continues the work presented in [13], where a practical representation of 3D face images based on segmentation using curvature information, feature extraction and analysis of the discriminating power of obtained local descriptors, were presented. In this paper, we analyze the robustness of proposed 3D face model and improve the recognition rate by face recognition experiments using a controlled and a non-controlled environment for two pattern classification methods: (1) Support Vector Machines (SVM) and (2) Principal Components Analysis (PCA) in combination with a Euclidean distance classifier. Also, due to the lack of representative 3D face databases that present a high degree of variability among the images of each individual, specially related to facial expressions, we have created our 3D face database named *GavabDB*. This database was described in Moreno et al [14], and its 3D images can be found (without texture) in http://gavab.escet.urjc.es.

The paper is organized as follows. Section 2 resumes the *GavabDB* database. Section 3 describes the proposed face recognition system. Section 4 explains the segmented regions and lines, shows the employed segmentation method and offers the obtained results in this stage. Section 5 describes the extracted facial features from the segmented regions and lines. Section 6 analyzes the discriminating capacity of the features providing the most discriminating subset, used for our experiments. Section 7 reports the face recognition experiments. Finally, section 8 covers the conclusions and future work.

2. Database description

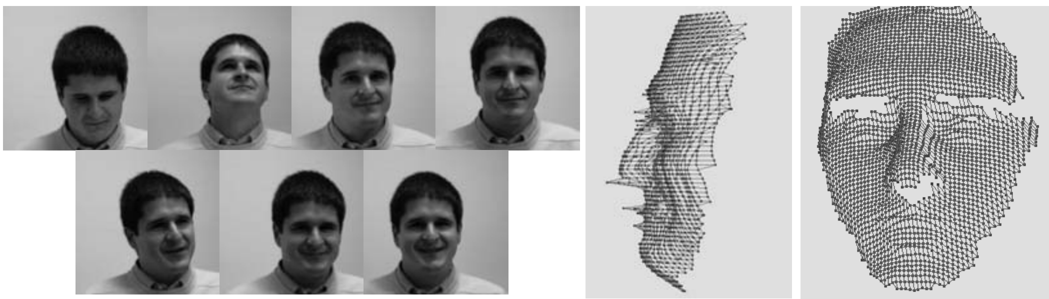
A set of 420 3D facial surfaces corresponding to the 60 individuals of the *GavabDB* database (having 7 different captures per individual) has been used for the experiments. Each set of the seven images of an individual contains (Figure 1(a)): one facial image in which individual is looking down (+35º *x*-rotation approximately), one facial image in which individual is looking up (-35º *x*-rotation approximately), and five frontal views from which three of them (last ones in Figure 1(a)) present facial expressions (a random gesture, smile and laugh, respectively). Facial surfaces are represented by meshes provided by the 3D digitizer VI 700 of Konica-Minolta. Cells of each mesh have four non-coplanar nodes, and occasionally three (in the contour). The grid provides an easy way of establishing two “orthogonal” directions (horizontal and vertical) along which it is possible going from one node to its neighbour in the mesh. The average points per face mesh in our database is 2,186 (at 1/4 of the scanning resolution). The captured images have three main problems: (1) noise (Figure 1(b)), (2) holes (Figure 1(c)) in dark or occluded parts of the face which are not sampled and (3) uncompleted contour caused by the auto-occlusion of the face when it is (unconsciously or not) rotated. Only one shot per captured image was performed. Neck, ears and hair were removed manually when they were present. A pre-processing stage for removing noise and smoothing were carried out over all the 3D facial meshes using Median and Gaussian filters, respectively.

3. Face recognition system description

The proposed recognition system performs the following classical stages: segmentation of regions and lines of interest, feature extraction from the segmented regions and lines, and classification of the feature vectors that model the faces. This last stage is performed by the two matching schemes: (1) PCA in combination with a Euclidean classifier and (2) SVM. The more discriminating set of extracted features were selected to model the face in the recognition system. Following sections explain these stages in detail.

4. Segmentation stage

The HK algorithm [15] can be used to classify 3D points of facial meshes assigning to each point of the mesh a label describing the local shape of the surface using dictionary of shape classes. The algorithm labels each point of 3D facial meshes as concave elliptical, convex elliptical, hyperbolical, concave cylindrical,



(a) (b) (c)

Figure 1. (a) Views of an individual whose corresponding 3D facial surface meshes belong to the database, (b) noisy mesh and (c) mesh with presence of holes.

convex cylindrical or planar. The classification is obtained using the signs of the median (H) and Gaussian (K) curvatures as criteria. The previous classification is qualitative in the sense that the label assigned to each point depends only on the sign of the main curvatures and not on its absolute value. This increases the robustness of the classification, because the sign can be correctly classified even though some noise may have affected the original data.

The first step is to compute both median and Gaussian curvatures in each point, using the following formulae:

$$H = \frac{(1+I_x^2)I_{yy} - 2I_x I_y I_{xy} + (1+I_y^2)I_{xx}}{2(1+I_x^2+I_y^2)^{3/2}} \quad (1)$$

$$K = \frac{I_{xx}I_{yy} - I_{xy}^2}{(1+I_x^2+I_y^2)^2} \quad (2)$$

where I_x, I_y, are the first derivative of the surface along x and y directions respectively, and I_{xy}, I_{xx}, I_{yy} represent the corresponding second derivatives. Signs of the H and K curvatures of each point for an example face are represented in the Figures 2(a) and 2(b) respectively. Brighter points are those in which the corresponding curvature value is positive, and darker ones are those with negative curvature values. Curvatures can not be evaluated in points of the surface contour, where it is no possible computing the second derivative. Figure 2(c) shows the resulting point classification based on sign of curvatures. Three kinds of points are obtained: darkest ones are hyperbolic points (which held K<0), intermediate grey level ones are elliptical convex points (which held H<0 and K>0), and brightest ones are elliptic concave points (H>0 and K>0). Cylindrical points have K=0 (and H>0 when they are concave and H<0 when they are convex, respectively). Planar points have H=0 and K=0. Cylindrical and planar points do not appear since it would be difficult to obtain exactly zero curvature values when working with real numbers.

In order to isolate regions of high curvature and to obtain a greater variety of classification points, different curvature thresholds were experimentally tested. The goal of these thresholds is to isolate regions of interest that can be used to identify a face. The considered regions should comply a set of criteria: (1) they should not change with different facial expressions, (2) these regions should be present in most of the images, (3) the regions should be easily detected and (4) the regions should be outside of areas susceptible of containing hear (moustache, beard, etc.). The first three requisites basically define facial regions with high curvature values, in which the skull is near the skin, so they are not affected by facial expression and can be easily identified. The fourth requisite restricts the face areas to those ones outside the mandible and the mouth. In order to isolate such regions of high curvature, different curvature thresholds were experimentally tested. After a threshold stage, H and K curvatures values that satisfy: -0.05<H<0.05 and -0.0005<K<0.0005 were set to zero. In this way, the considered regions appear isolated in most of the faces as shown in Figure 2(d). From all the regions that appear, taking into account the previous curvature criteria, seven regions have been selected. They are shown in Figure 3 and their location method is resumed next.

Region 1 is easily distinguishable because is the region which has the highest number of convex elliptic nodes in the threshold image. Region 2 is formed by the convex cylindrical points and it is a neighbouring region 1. Points of Region 3 are hyperbolic ones and located in a neighbouring region of the Region 2 which is up from it. Further relative position restrictions with respect to the eye regions can be used if more than one candidate region were obtained. Regions 4 and 5 are formed by concave elliptic points. There are few candidate regions to them, and the selection is performed according to their relative position with respect to the nose regions. Regions 6 and 7 are formed by convex elliptic points in the non-threshold image, after eliminating candidate regions that did not followed certain restrictions. Such restrictions were: their relative positions with respect to regions 3, 4 and 5 (for example, their points had to be found in a band containing these three regions), height (1.7 mm. maximum) and width of the candidate regions. Lines 1 and 2 are looked for from the external border of Region 2 at left and right directions, respectively, until a sign change of the median curvature is found.

A region or line location result can be labeled as: a) failure, b) non-existence and c) correct location. Failure means that a searched region or line was found wrongly and then, a false positive in the region location was found as the chosen region. Non-existence means that there was not any region resulting for the searched region. Correct location occurs when the segmented region was correctly found. Failure, non-existence and correct percentage results for the different regions in the whole database are shown in Table 1. More detailed

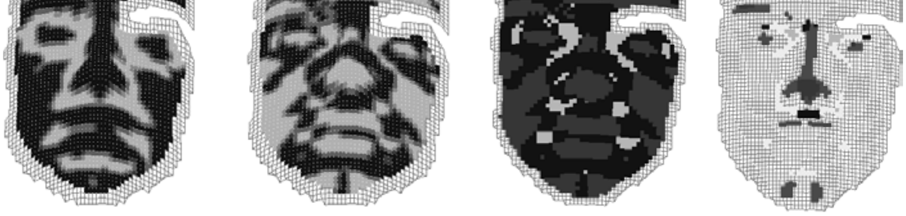


Figure 2. (a) Sign of H and (b) sign of K of the points. (c) Classification of the points before the threshold stage and (d) classification of the points after the threshold stage.

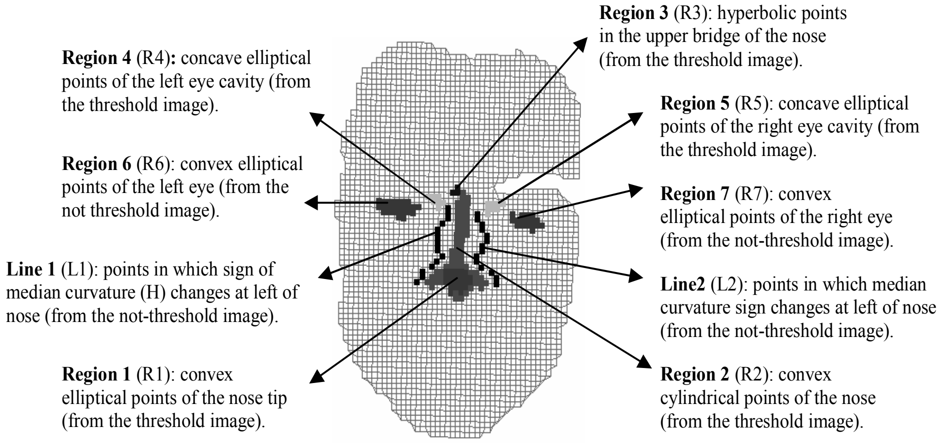


Figure 3. Segmented regions and lines of a face from which facial features are extracted. Regions number 6 and 7 are from non-threshold image shown in Figure 2 (c). Lines 1 and 2 can be observed in the non-threshold image shown in Figure 2 (a).

results can be found in [13].

The poorest segmentation region was the eye region when the individual was looking down. The main reason for that is caused by the fact that occluded eye regions were automatically reconstructed by the software of the 3D digitizer using its own interpolation algorithm. This produced a pernicious effect in the surface, that was smoothed losing information of the real curvature value in that region. This phenomenon also occurred in Region 3, when individuals were looking up. Nose point region (Region 1) offers better results than the other regions. While a failure on a region searched produces a recognition error, a non existence of some region makes possible the face recognition stage.

5. Feature extraction

The goal of this section is to extract a relevant set of features from the automatically segmented regions in order to characterize and model each 3D mesh. Using the automatically segmented areas, a total of eighty six non-independent features were defined. These features are classified in thirteen categories: (1) areas of the regions, (2) area relations, (3) mean of areas, (4) distance between mass centers of regions, (5) relations of distances between mass centers of regions, (6) mean of distances, (7) angles among mass centers of regions, (8) mean of angles, (9) mean of H of the points belonging to a region, (10) mean of K of the points belonging to a region, (11) variance of H of the points belonging to a region, (12) variance of K evaluated in the points belonging to a region, (13) lines 1 and 2 based features. They are detailed in [9].

The discrimination power of each feature Φ has been computed using the Fisher coefficient [9], which represents the ratio of between-class variance to within-class variance, according to the formula:

$$CF_{\Phi} = \frac{\sum_{i=1}^{c}(m_i - m)^2}{\sum_{i=1}^{c}\frac{1}{n_i}\sum_{x\in\Phi_i}(x-m_i)^2} \quad (3)$$

where CF_{Φ} is the Fisher coefficient of the Φ descriptor, c is the number of classes (individuals in our case), Φ_i is the set of feature values for the class i, n_i is the size of Φ_i, m_i is the mean of Φ_i, and m is the global mean of the feature over all classes, and x is each value of descriptor Φ in the class i.

There are 60 classes corresponding to the number of distinct individuals in the database. Although there are seven images per person, when computing the Fisher coefficients, only the 3D facial images having the whole regions and consequently the whole set of features correctly extracted have been employed (a total of 310 images).

Table 2 includes the mathematical definition of each one of the 30 more discriminating features. Each feature has an associated identification number and a ranking position. This position corresponds to the ordered discriminating power of the feature (a smaller value in the ranking corresponds to a higher discrimination rate).

When a feature can not be computed because of the non-existence of a region from which it is derived, it was zero valued, except when a symmetric feature exists. In this case, the non-existent feature is valued like its symmetric one. Extracted features have been normalized to values in the interval [0, 1].

6. Feature selection

The goal of this section is to identify the minimum subset of features that maximize the recognition rate of

Table 1. Percentage results of the region location stage: failure, non-existence and correct location percentages grouped per searched regions and lines over the 420 database images.

	Region 1	Region 2	Region 3	Region 4	Region 5	Region 6	Region 7	Line 1	Line 2
Failure	0,2%	0,2%	0,9%	0,2%	0,4%	2,8%	1,9%	1,1%	0,7%
Non-existence	0%	0,2%	7,8%	2,8%	6,1%	13,3%	10,2%	0,2%	0,4%
Correct location	99,8%	99,6%	91,3%	97%	93,5%	83,9%	87,9%	98,7%	98,9%

a generic face recognition system. In order to select an optimal subset of descriptors, eighty six feature subsets were tested. Each subset was composed by the first *n* features of the ordered features list according to their Fisher coefficients, where $1 \leq n \leq 86$.

The training set consisted of 60 frontal images (one per individual) which, after segmentation, contained all the regions and lines necessary to obtain the feature vector. Six test image sets were used : (1) the rest of frontal view images, which was composed of 47 images, (2) smiling images, which contained 48 images, (3) laughing images, which contained 45 images, (4) random gesture, containing 47 images, (5) looking down images, with 19 images and (6) looking up images, with 41 images. Because the goal is to obtain the minimum set of features that produces the highest recognition rate, these test sets were constructed exclusively with the images of the individuals that contained all the regions and lines necessary to calculate all features.

The conclusion obtained after the experiments is that, in these six test sets, the best recognition results were obtained when the number of features used to represent the individuals was between the first 30 and 37 features (those of lower positions in the ordered list according to Fisher coefficients).

7. Face recognition experiments

The robustness of the model is measured by how the changes in pose and gesture of the faces influence over the recognition rate of a face recognition system. The model has been tested by a face recognition system that uses as matching schemes Principal Component Analysis (PCA) with a Euclidean classifier and Support Vector Machines (SVM). Both PCA and SVM have produced very good results in 2D recognition problems. The comparison of the recognition rates obtained by the recognition system in both controlled and non-controlled environment will provide us an indication of how robust the designed model is.

PCA and SVM have been used with good performance for high-dimensional pattern recognition problems [16] in general, and in face recognition problems in particular [17]. PCA algorithm is used in pattern recognition problems for reducing the dimensionality of the feature vectors, while retaining as much information as possible. The new compressed representation is used to implement a pattern classifier, typically a distance-based classifier, neural networks or Bayesian classifiers.

The problem that SVM try to solve is to find an optimal hyperplane that correctly classifies data points and separates the points of two classes as much as possible.

The matching schemes have been implemented using a free PCA implementation [18] and SVMTorch [19]. The PCA implementation used Euclidean distance to identify a given face. The kernel to run all the SVM experiments was a Gaussian kernel. In each case the classification system has been constructed by using a vector with the 30 more discriminating features presented in section 5.

Each matching scheme has been tested with a controlled and a non-controlled environment. In this context, a controlled environment means that the pose and gestures are controlled. In the non-controlled experiments, from the seven images of each individual, five have been randomly chosen to be in the training set, an the other two in the testing set, this produces a training set with 305 images (5 per each one of the 61 individuals) and 122 test images. In the controlled environment, six images are used for training, which include one of the frontal images, and the other frontal image is used for testing. This produces a training set with 366 images and a testing set with 61 images. This is an indication of the robustness of the model against variations of the pose and gesture. Figure 4 presents the results of the experiments. The correct recognition rate indicates when the classification provided by the matching scheme identified the individual correctly.

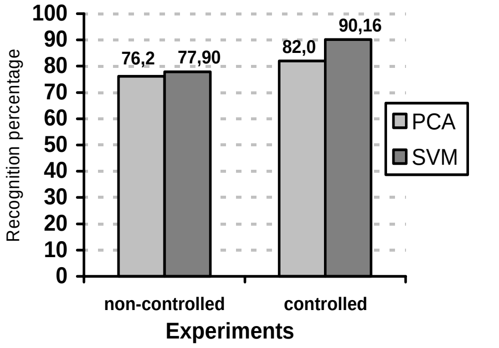
In general, SVM produces better recognition rates than PCA as shown in Figure 4. The small difference in correct recognition rate between controlled and non-controlled environments implies that the proposed 3D

Table 2. 30 more discriminating features and their position in an ordered list according to their discriminating power. The used acronyms are: *Ri* = region *i*; *Li* = line *i*; *A_Ri* = area of region *i*; *C_Ri* = centroid of region *i*; *d*(*P1*, *P2*) = Euclidean distance between 3D points *P1* and *P2*; ang(*P1*,*P2*,*P3*) = angle defined by the 3D points *P1*,*P2* and *P3*, being *P2* the intermediate vertex; *H_Ri* and *K_Ri* are the respective averages of the Mean and Gaussian curvatures, evaluated in points belonging to the region *i*; *VH_Ri* and *VK_Ri* are the respective variances of the Mean and Gaussian curvatures evaluated in points belonging to the region *i*

Rank pos	Feature description		Rank pos	Feature description
1	ang(C_R4, C_R3, C_R5)		16	H_R3
2	ang(C_R4, C_R3, C_R1)		17	H_R4
3	1/2[ang(C_R4,C_R3,C_R1)+ang(C_R5, C_R3,C_R1)]		18	d(C_R4, C_R5)
4	ang(C_R5, C_R3, C_R1)		19	d(C_R5, C_R1)
5	1/2[d(C_R4, C_R3)+d(C_R5,C_R3)]		20	K_R4
6	d(C_R1, C_R3)		21	H_R5
7	ang(upper point of L1,C_R3, upper end of L2)		22	K_R5
8	d(C_R4, C_R3)		23	ang(C_R6, C_R3, C_R1)
9	d(C_R5, C_R3)		24	H_R2
10	d(C_R4, C_R5)/ d(C_R1, C_R3)		25	1/2[ang(C_R6,C_R3,C_R1) +ang(C_R7, C_R3,C_R1)]
11	K_R3		26	ang(C_R7, C_R3, C_R1)
12	H_(R4 U R5)		27	ang(C_R6, C_R3, C_R7)
13	K_(R4 U R5)		28	H_R1
14	1/2[d(C_R4, C_R1)+d(C_R5,C_R1)]		29	1/2[d(C_R6, C_R3)+d(C_R7,C_R3)]
15	d(C_R4, C_R1)		30	A_R2

face model is robust enough to capture many variations that real face recognition systems have.

Figure 4. Recognition rates of PCA and SVM matching methods for controlled and non-controlled environments.



8. Conclusions

This paper presented a 3D face modeling system based on a HK segmentation algorithm. Six regions and two lines were automatically obtained from each 3D facial mesh. These are then used to obtain the most relevant facial features according to Fisher's coefficient. The robustness of our model has been tested by implementing two face recognition systems based on PCA and SVM as matching schemes, under controlled and non-controlled environments. Experimental results show that the correct recognition rates are enough to implement real face recognition applications and that the proposed model is robust. Using SVM in a non-controlled environment produces 77.9% correct recognition rate, while a 90.16% is obtained in a controlled environment. We can conclude our system has a recognition rate in the interval [78%, 90%] for the considered variations using our database.

The *GavabDB* database contains 420 3D facial images of 60 different individuals, with seven images per subject: two frontal views without gesture, one rotated looking up, one rotated looking down, one smiling, one laughing and one with a random gesture.

Future work will consider other features different from those belonging to the eye regions when faces are rotated looking up and looking down because their corresponding regions appear occluded in many patterns. Complete set of images have been used independently of the segmentation results in the recognition experiments. A validation procedure is needed to improve the recognition to reject false positives in the region location stage is needed to improve the recognition results.

9. References

[1] The Biometric Consortium, http://www.biometrics.org.

[2] A. Pentland and T. Choudhury, "Face Recognition for Smart Environments", *Computer IEEE*, February 2000,50-55.

[3] W. Zhao, "Face Recognition: A Literature Survey", UMD-CFAR, *Technical Report* CAR-TR-948, 2000.

[4] W. Zhao, R. Chellapa, P. J. Phillips and A. Rosenfeld, "Face Recognition: A Literature Survey", *ACM Computing Surveys* 35 (4), December 2003, 399-458.

[5] W. Zhao, "Improving the Robustness of Face Recognition", Proc. *Intl. Conf. on Audio and Video Based Person Authentication*, 1999, 78-83.

[6] M. W. Lee and S. Ranganath, "Pose-invariant face recognition using a 3D deformable model", *Pattern Recognition* 36, 2003, 1835-1846.

[7] J. Huang, T. Heisele and V. Blanz, "Component-based Face Recognition with 3D Morphable Models", 4th *Conf. on Audio and Video-based Person Authentication*, 2003.

[8] C.-S. Chua, F. Han and Y.-K. Ho, "3D Human Face Recognition Using Point Signature*", Fourth IEEE Intl. Conf. on Automatic Face and Gesture Recognition*, 2000.

[9] P. Hallinan, G. Gordon, A. L. Yuille, P. Giblin and D. Mumford, "Two-and Three-dimensional patterns of the face", Ed. A. K. Peters, 1999.

[10] K.I. Chang, K.W. Bowyer y P.J.Flynn,"Face Recognition Using 2D and 3D Facial Data", *Workshop on Multimodal User Authentication* (MMUA), Santa Bárbara CA, 2003.

[11] F. Tsalakanidou, D. Tzovaras and M. G. Strintzis, "Use of depth and colour eigenfaces for face recognition", *Pattern Recognition Letters* 24, 2003, 1427-1435.

[12] C. Ben Abdelkader, P.A. Griffin, "Comparing and combining depth and texture cues for face recognition", *Image and Vision Computing* 23, 2005, 339-352.

[13] A. B. Moreno, Á. Sánchez and J. F. Vélez, "Robust Normalization of 3D Facial Meshes using Automatically Located Feature Points", *Proc. 12th Portuguese Conference on Pattern Recognition,* June 2002.

[14] A. B. Moreno and A. Sánchez, "GavabDB: a 3D Face Database", *Workshop on Biometrics on the Internet* COST275, Vigo, March 25-26, 2004, 77-85.

[15] E. Trucco and A. Verri, "Introductory Techniques for 3-D Computer Vision", *Prentice-Hall,* 1998.

[16] H. Bynm, "A survey on pattern recognition applications of Support Vector Machines", Intl. Journal of Pattern Recognition and Artificial Intelligence (17) 3, 2003, 459-486.

[17] P. J. Phillips, "Support Vector Machines applied to face recognition", *Advanced Neural Information. Processing Systems* 11, 1998, 803-809.

[18] S. Romdhani,"Face Recognition Using PCA", http://www.vision.im.usp.br/~teo/pca/, 1996.

[19] R. Collobert and S.Bengio, "SVMTorch: Support Vector Machines for Large-Scale Regression Problems", *Journal of Machine Learning Research*, vol. 1, pp. 143-160, 2002.