GavabDB: a 3D Face Database

A. B. Moreno, A. Sánchez

Departamento de Informática, Estadística y Telemática, Escuela Superior de Ciencias Experimentales y Tecnología (Universidad Rey Juan Carlos), c/Tulipán s/n, 28933 Móstoles (Madrid), Spain

{a.b.moreno, an.sanchez}@escet.urjc.es

Abstract

A 3D face image database (*GavabDB*) has been built for automatic face recognition experiments and other possible image applications such as pose correction and 3D face model registration. This paper presents the database features and also explains the process of building the database in detail. Additionally, our motivation for creating this database, the image acquisition process, the image properties and a discussion about them, as well as some conclusions extracted from the database creation are presented. *GavabDB* contains 427 three-dimensional facial surface images corresponding to 61 individuals (45 male and 16 female), and there are 7 different images per each person. The whole set of individuals are Caucasian and most of them are aged between 18 y 40. Each image consists in a three-dimensional mesh representing a face surface. There are systematic variations over the pose and facial expression of each person. In particular, there are 2 frontal and 4 rotated images without any facial expressions, and 3 frontal images in which the subject presents different and accentuated facial expressions. The purpose of this publication is to present this database such that other researchers in this field can use it for related facial experiments.

1. Introduction

Automatic face recognition is actually a research area in auge. Face Recognition techniques mainly come from Computer Vision, Pattern Recognition and Artificial Intelligence [1]. This biometric feature has many applications in areas like: personal identification, security applications, law enforcements and forensic medicine, among others. Face and voice are the ones less intrusive biometric features. In order to improve robustness of the biometric systems, the research in the multimodal biometric identification (which simultaneously employs more than one biometric features), is actually growing [2][3].

The development of automatic face recognition systems needs from face image databases, such that comparative evaluations of the systems can be possible. These systems must have a certain variability degree (over the pose, illumination, facial expressions, etc.) among the different views of the same individual.

There exist some published face image databases orientated to different experimentation purposes: automatic face recognition, facial expression analysis, pose estimation, face segmentation from a background plenty of objects, etc. Our face image database has been created to be used in face recognition experiments which need from 3D face images.

Next, some databases of face images in the literature are described.

- *FERET* database [4]: it was built for the development of automatic face recognition to assist the law enforcement. It was sponsored by the *Conterdrug Technology Development Program through the Defense Advanced Research Product Agency*. It contains 14.051 2D face images, 8-bits grey scale, including pose, facial expression and illumination variations. There are some individuals wearing glasses, different hair length, and both. It has many systematic pose variations. It is one of the most popular databases.

- *Yale DB* database [5]: created by the *Center for Computational Vision and Control* of the Yale University. It contains 2D grey level face images corresponding to 15 individuals. There are 11 different images per each one of them; in particular, normal images, wearing glasses or not, presenting illumination variations (varying the position of the source of light: at the centre, at left and at right) and face images presenting different facial expressions such as happiness, sad, surprise, etc.).

- *AR* database [6]: it was created by the *Centro de Visión por Computador* of the *Universidad Autónoma* of *Barcelona* (Spain). It contains colour 2D face images corresponding to 126 individuals (70 male and 56 female). They are frontal views presenting four different facial expressions, illumination conditions and occlusions (sun-glasses, etc). Images were captured under strictly controlled conditions. There were not restrictions about the presence of glasses, make-up, style of hearing, etc., of the individuals. Each person collaborated in two sessions, separated by two weeks. The same images were captured in both sessions.

- *MIT* database [7]: it was built by the *MIT Media Laboratory*. It contains 2D face images corresponding to 16 male individuals. It includes images corresponding to 3 different orientations (up at right, right and left), 3 illumination variations and 3 scale variations (varying the camera zoom). Images at different resolution (those corresponding to 6 levels of the Gaussian pyramid whose sizes are 480x512, 240x256, 120x128, 60x64, 30x32 y 15x16, respectively), are present.

- *ORL* database [8]: it was created by The *AT&T* laboratories (Cambridge). It contains the 2D face images corresponding to 40 persons, in which there are 10 different images per person. For some subjects, the images were captured at different times, varying the illumination, the facial expressions (*open eyes*/*closed eyes*, *laugh* / *no laugh*) and other details (*glasses* / *no glasses*). The background was dark and homogeneous. Each subject was captured looking up at right and in a frontal position tolerant with respect to some face movement.

- *CVL* database [9]: it was created by the *Computer Vision Laboratory* (CVL) belonging to the University *of Ljubljana* (Slovenia). It contains 7 2D images corresponding to 114 different individuals.

- *PF01* database (*Postech Faces '01*) [10]: it contains face images of Asian individuals, most of them from Corea. It includes variations with respect to illumination, pose and facial expression. This database offers quite enough variations among the images of each individual. It is appropriate for the evaluation of automatic face recognition systems over Asian individuals.

- *XM2VTS* database [11][12][13]: it is a large multimodal database, oriented to test multimodal verification. It was created by the *Center for Vision, Speech and Signal Processing* of the Surrey University (United Kingdom). It contains 4 capture sessions of 295 persons, over intervals of 1 month. Every session consisted on an individual capture while the subject is speaking and rotating his/her head. As a result, high quality colour images, sound files and video sequences were digitized. A 3D model of each individual was later obtained in the third session. This model was generated using an active stereo system and was converted to VRML format. This is a commercial database, and the different datasets are separately sold.

- *3D_RMA* database [14]: created in the *Signal and Image Center (SIC)* of Brussels (Belgium). The 3D acquisition system was based on structured-light, being constructed using a camera and a projector, and generating the 3D coordinates of the surface points with a high precision. Glasses and dark parts of faces could not be captured. It contains 120 individuals captured in two different sessions, separated by 2 months. Digitisations consisted in 3 shots grabbing different and limited orientations of the heads (*frontal*, *at left or right* and *at up or down*). The whole individuals belonged to the same Caucasian race.

- 3D face database of York University [15]: 3D face databases above mentioned contained a reduced set of face images per subject. This database has images corresponding to 97 individuals. It contains 10 captures per individual including different poses. However, only 2 of these views of each individual present light facial expressions (happiness and frown), and one presents face occlusion.

As it can be observed, there are not many varied changes in the image modifications among the different images of each individual in the databases. Some of them offer variations related to some aspects but not to others. However when some database has a certain richness of systematic changes, the range of variation is very limited. Most of the reported databases only offer 2D face images. Among the databases above related, only the last three ones contain 3D face images, and only the last one includes some facial expression but not pronounced.

Our *GavabDB* database contains 3D face images consisting in facial surfaces that represent the faces by three-dimensional meshes. As a new feature, this database offers three views per individual in which there are facial expressions (two of them very pronounced). It also includes many variations with respect to the pose of each individual. Such facial expressions are: smile, laugh and a random gesture chosen by the individual. In this last case, occlusions of the face by the hand or by the tongue are permitted. In the other hand, the texture (colour) has been omitted in the published images. Then, they consist in 3D meshes exclusively. GavabDB database images can be obtained from the following URL: `http://gavab.escet.urjc.es`.

The rest of the paper is organized as follows. Section 2 describes acquisition process of the face images. Section 3 describes in detail the set of the acquired image data. Section 4 analyzes some aspects related to the generated images and some found problems during the acquisition process. Section 5 presents two example applications for the described database. Finally, Section 6 outlines the conclusions.

2. 3D image acquisition process

In this Section, the designed protocol for the acquisition of the 3D face image database is explained. The objective was to produce a database quite enough numerous with respect to the individuals and simultaneously rich in accentuated variations for the same person. These goals have allowed us to evaluate the degree of invariance with respect to those changes of the experimented face recognition systems. Variations offered by a face images can be classified as intrinsic or extrinsic [16]. They are respectively related to the individual face, or to the position/orientation of his face in the scene. Thus, examples of intrinsic variations are the facial expression, the wear, the face occlusion or the presence of hair. Examples of extrinsic variations are the background, the illumination, the camera settings, the scale, etc. Our database includes three intrinsic accentuated systematic variations and four extrinsic variations related to the pose.

2.1 Subject variations

A facial database should include enough variations in the images of each individual such that different face recognition techniques could be tested, and these techniques should be robust under the considered variations. There are 9 different images (views) captured per each individual. Their names and variations with respect to pose and expression are shown in the Table 1.

View number	File name	Head orientation	Facial expression
1	cara*i*_frontal1	Frontal	Neutral
2	cara*i*_frontal2	Frontal	Neutral
3	cara*i*_derecha	Right Profile: ≈ +90º of rotation around *y* axis.	Neutral
4	cara*i*_izquierda	Left profile: ≈ -90º of rotation around *y* axis.	Neutral
5	cara*i*_arriba	Looking up: ≈ +35º of rotation around *x* axis.	Neutral
6	cara*i*_abajo	Looking down: ≈ -35º of rotation around *x* axis.	Neutral
7	cara*i*_sonrisa	Frontal	Smile
8	cara*i*_risa	Frontal	Accentuated laugh
9	cara*i*_gesto	Frontal	Random gesture chosen by the subject.

Table 1: Set of images of the i-th individual of the database and their variation with respect to pose and facial expression.

Figure 1 shows the 2D images corresponding to the captured views of the individual of the database. These views are captured by the scanner from its viewpoint in the same moment in which the corresponding 3D meshes are acquired. The first two images correspond to frontal views of the face with a neutral expression. The third and fourth images correspond to both right and left profile views of the individual, respectively. In such views the subject presents his/her head rotated approximately 90º around the vertical axis in both directions. The fifth and sixth images correspond to head rotations around vertical direction ±35º approximately, looking up and at down, respectively. The three last captured images correspond to frontal views with the facial expressions of smile, pronounced laugh and a random gesture chosen by the person. We have not considered the variations with respect to presence of glasses nor the illumination, but there exist individuals in the database with bear, moustache, etc.

Figure 1:2D images corresponding to the different variations of one individual in the database.

Figure 2 shows the nine 3D captured images of the same individual of Figure 1. This figure shows the three-dimensional mesh: (a) with texture mapped and (b) without texture (both are showed like they are viewed from the point of view of the scanner). It can be appreciated that the scanner projects rays corresponding to the vertex of a regular grid in the *x-y* plane. Meshes are formed by those connected vertices, forming cells of four nodes. Figure 2 (c) shows the same rotated mesh to provide a visualization in which the depth changes in the *z*-axis can be appreciated.

2.2 Nature of the individuals

We found 42 volunteers among the students through announcements placed in many points of the Campus, aged between 18 and 22 approximately. The remaining 19 individuals were teachers and researchers of our Center (Escuela Superior de CC. Experimentales y Tecnología, ESCET – Universidad Rey Juan Carlos, Madrid). Most of them aged between 23 and 40. The total amount of the individuals corresponds to 45 male and 16 female people of a homogeneous human race (Caucasian).

2.3 Acquisition environment

Next, we describe the 3D digitizer employed to construct the database. It is the Minolta VI-700 digitizer [17], a laser sensor which captures in less than a second a range image of the scene with colour information. It generates a three-dimensional surface mesh of the visible face surface in the scene (not occluded from the scanner viewpoint) if it is not too dark and if it reflects the light emitted by the scanner. The captured images have been directly grabbed by the scanner, making only one shot for each one. It means that, they have not been built using a registering process of several captured images from different scanner views.

No effort was done to control the illumination conditions during the capture process. Sunlight was predominating, although the room was provided with typical artificial light in office environments. The individuals were placed near of large windows without special focus (see Figure 3).

The faces were placed at 1.5 ± 0.5 m. of distance from the scanner. Although the chair (with wheels) was at 1.5 m. of distance, the individuals had flexibility to move their head normally. This fact produces some resolution changes from some images to others. The height of the chair could be changed in order to make the head always visible to the scanner. In order to capture the different head orientations, individuals looked at certain marks (without forcing his/her head to a fix position, maintaining a precise angle).

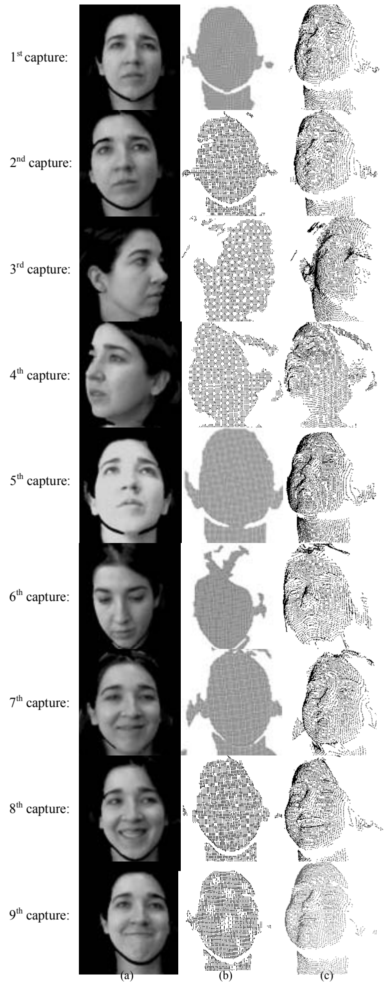


Figure 2: 3D views of an individual: (a) 3D images with texture mapped (b) the same images without texture at 1/4 of resolution (both from scanner's point of view) and (c) the same face rotated images.

Each capture shot was performed in less than one second. Ten minutes approximately were employed to capture the nine images of an individual (including the previous time for explaining the subject pose and gesture changes, the indication before each shot and its preparation). Figure 3 shows the acquisition system used and one individual positioned during the capture process.

Figure 3: Acquisition system of the 3D images and position of an individual with respect to it.

3. Captured 3D images

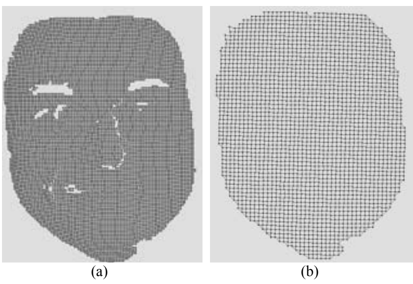
The scanner is sold together with a software tool named VIVID, which contains among others, functions to export the scanned images to files with different formats (those typical in 3D object modelling, such as VRL, DXF, etc.). Different information of the captured images can be selected to be exported, for example:

- The whole set of captured points (resolution 1/1) or the remaining ones after reducing regularly the resolution to 1/4, 1/9 or 1/16 (1 point of each 4, 1 of each 9, and so on). Such reductions decrease the size of data files in a considerable way.

- The colour information from the mesh. Texture elimination when not going to be used is another way considerably to reduce the size of facial images.

- Filled holes of the mesh that could exist in not sampled zones. The scanner software offers the possibility of filling little holes using the option *Fill holes*.

Figure 4 shows the facial mesh corresponding to an individual (a) selecting the whole sampled points without applying the option *Fill holes* and (b) after filling holes and choosing a resolution reduction of 1/4.



(a) (b)

Figure 4: (b) 3D Facial mesh at resolution 1/1 without fill holes and (b) the same mesh at resolution 1/4 after filling holes.

Every mesh (captured image) is composed of points of the facial surface and their connections, forming cells of four non-coplanar nodes each one. The coordinates (*x, y, z*) of their 3D points are referred to a coordinate origin placed in the scanner during the capture time. Each mesh has been exported to a VRML format file, of 1/1 resolution without colour information and after a filling-holes process. By eliminating the colour information the file sizes are reduced to 0.7 Mbytes approximately from their original size of 2.7 Mbytes.

4. Analysis of the acquisition stage

Both analysis and discussion about some important properties of the generated three-dimensional meshes, as well as some aspects related to the capture process, are presented in this Section.

Some different sources of error and deterioration in the generated surfaces are the following: presence of holes, occlusions, uncompleted face contour, presence of noise, etc. Although many restrictions could be introduced in the capture stage in order to reduce some mentioned errors, it has not been done in order to obtain the more general and real images in which can appear those phenomena, letting us their use in systems independent of them. Next, these problems are analyzed:

- Presence of holes: the 3D digitizer is a range laser sensor which uses an active technique for the depth capture, based in the projection of controlled light to establish the geometry of sampling. As mentioned above, dark parts of the surface do not reflect the projected light and they can not be sampled producing holes in the mesh. Also, occluded parts from the scanner viewpoint can not be sampled (for example, part of neck placed after the chin). The *Fill holes* function is supplied by the scanner software using some interpolation method, and consequently introducing some error in the generated surfaces. Figure 5 shows in the first row some images in which holes exists; and in the second, how the number of holes is reduced when the interpolation function *Fill holes* is applied.

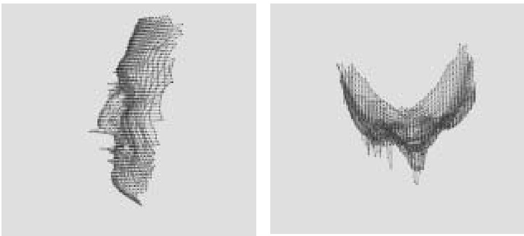
(a) Face 8: gesture view | (b) Face 13: smile view | (c) Face 2: frontal 2 view | (d) Face 30: looking down

Figure 5: (a) Presence of holes in four face examples and (b) the same images after a Fill holes stage.

- Presence of occlusions caused by beard, moustache, goatee, etc.: there exist some individuals in the database who have some of these facial features, which occlude their real face surface. These occlusions are a source of noise or holes around these regions. Figure 5 (d) shows an example.

- Occlusion of face contour: as each image was captured by making a single shot without applying a registering stage among several images from different angles. Only the visible surface from the scanner view point was sampled. Consequently, the contour is uncompleted cause of this occlusion. A light face rotation around the *y* axis produces a large occlusion of one of the face sides, which causes significant differences among two rotated images of the same individual.

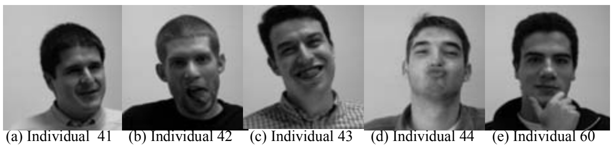
- Presence of noise: the images have a lot of noisy points that must be carefully eliminated by including some pre-processing stage in the systems that use them. Figure 6 represents the mesh of Figure 4, from two different viewpoints (profile and from above, respectively) showing the noisy points.



(a) (b)

Figure 6: Face of the Figure 4 rotated to show the existing noise of the image.

Next, some aspects related to the acquisition process are emphasized. First, we have to point out the existing richness of pronounced variations of each individual's face surface, not only with respect to the pose but also related to the facial shape (through pronounced facial expressions). In particular, it is a salient database feature that both the laugh and the random gesture views are usually quite pronounced. The Figure 7 shows some examples of gestures corresponding to captured meshes of that view. Notice that a user can partially occlude his face with his tongue (b) or his hand (e).



(a) Individual 41 (b) Individual 42 (c) Individual 43 (d) Individual 44 (e) Individual 60

Figure 7: Same examples of random gesture images.

It is very common in images with gestures, the presence of head rotations as observed in the examples of the Figures 7 (a) and (c). This phenomena also happens in the others frontal views. In the other hand, sometimes the individuals have an expression caused by their happiness or nerves, while the neutral expression images are captured. Such images have been kept instead of replacing them by new captures.

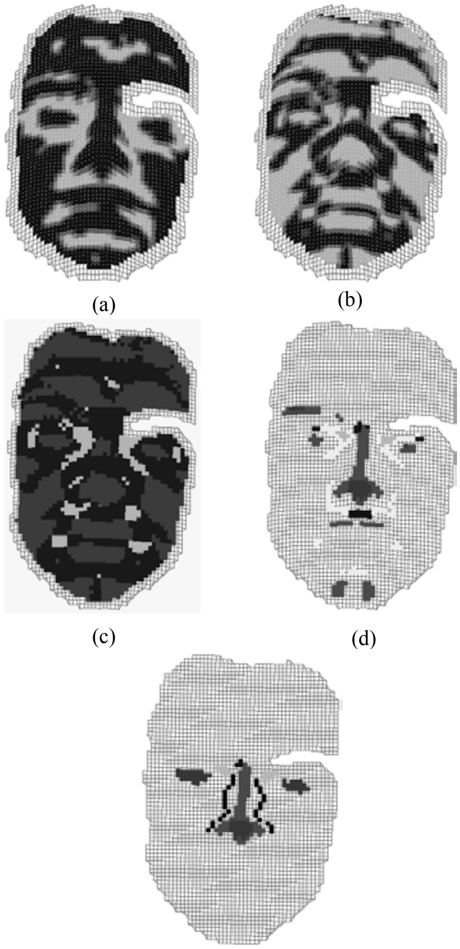
Finally, it is important to comment the difficulties to find volunteers. Many people did not have much interest and others did not want to be grabbed.

5. Possible Database Applications

This section resumes two different database applications, although new possible application examples related to facial recognition, facial expression analysis, pose estimation and face registration could be suggested.

- The discriminating power of three dimensional (3D) descriptors extracted from 3D human face surfaces has been analyzed and presented in [18]. An automatic face recognition system using different subsets of the descriptor set has been implemented and tested. An HK segmentation (based in the signs of the mean and Gaussian curvatures) for isolating regions of pronounced curvature has been performed, whose results are showed in Figure 8. Eighty six descriptors and ratios have been obtained from the segmented regions selected. Some of them are showed in Figure 8 (e). The thirty five most discriminating ones in frontal views provided 78% of recognition success rate when best match is selected, and a 92% of recognition success was obtained when the five best matches are selected, using a Euclidean distance based classifier. In order to reduce the volume of data and to increase the efficiency, we used images in this application of 1/4 of resolution.

- Feature point extraction from 2D images is another work in which *GavabDB* has been used. The method was initially presented in [19]. The objective was to locate nine 3D feature points in the facial meshes without considering



(a) (b) (c) (d) (e)

Figure 8: Sign of (a) median and (b) Gaussian curvatures; (c) HK point classification, (d) HK classification after curvature thresholding, and (e) regions and lines of an automatically-segmented face from which facial features have been extracted for face recognition.

colour information. Figure 9 shows some candidate points for the extracted feature points in (a) and final extracted feature points in (b). In this approach we only use local geometrical properties of the mesh surface in the neighbourhood of points. Captured points were used for the task of pose normalisation in a database of facial surfaces presenting facial expressions and light rotations for recognition purposes. Those points were used to normalize the database. Normalization experiments were presented in [20]. Pronasal (98%), nasion (93%) and entocanthions (92%) have the highest location rate. This rate will increase if only frontal images with light rotation are used. Frontal subset (presenting or not facial expressions) of the database images provided a 94,2% of correct location rate over all feature points. Frontal subset (presenting or not facial expressions) of the database images provided a 94,2% of correct location rate over all of the feature points.

- Possible applications of this 3D facial database are to perform a) pose correction of a 2D image for recognition purposes and b) face encoding for transmission of compressed images. A 2D rotated input image of a face is matched to a 3D face model defined by the corresponding facial meshes. The parameters obtained can be used to construct a different 2D view (for example a frontal view) of the same face. This task can be also easily adapted to the Internet. In the other hand, the parameters encode a face that can be transmitted from an emission place of a communication (i.e. videoconference) instead to transmit the complete face image. These parameters together with the 3D model can reconstruct the original face image in the reception point.

6. Conclusions

We presented the 3D face database Gavab3D. Our database contains 3D face images of 61 people including 9 images for each person. A number of 45 male and 15 female are included. 3D images in the database include systematic variations for pose and facial expressions. The richness of images with pronounced expression variations is one of the most relevant aspects of our database. We hope that *GavabDB* can help other researchers in their related work.

7. References

[1] R. Chellapa, C. L. Wilson and Sirohey, “Human and Machine Recognition of Faces: A Survey”, Proceedings of the IEEE. Vol. 83, nº5, May 1995, pp.704-740.

[2] “Face Recognition: a new dimension”, Biometric Technology Today, November-December 2002.

[3] R. de Luis-García, C. Alberola-López, O. Aghzout and J. Ruiz-Alzola, “Biometric identification systems”, Signal Processing 83, pp. 2539-2557, 2003.

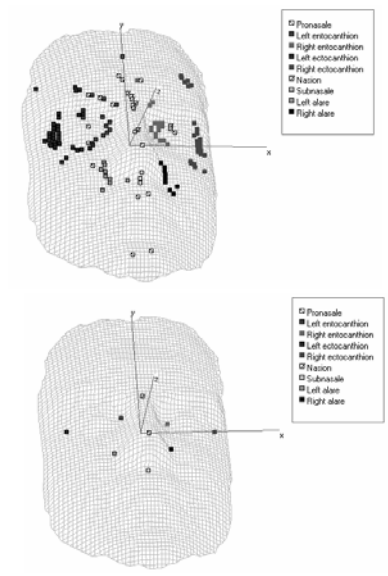
[4] P. J. Phillips, H. Moon, P. J. Rauss and S. Rizvi, “The FERET evaluation methodology for face recognition algorithms”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 22, No. 10, 2000.

[5] P. Belhumeur, J. Hespanha, D. Kriegman, “Eigenfaces vs. Fisherfaces: class specific linear projection”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 19, No. 7, pp. 711-720, 1997.

[6] A. M. Martínez and R. Benavente, “The AR Face Database”. CVC Technical Report #24, June 1998.

[7] M. Turk and A. Pentland, “Eigenfaces for recognition”, J. Cognitive Neuroscience, Vol. 3 (1), 1991.

(a)



(b)

[8] F. Samaria, A. Harter, “Parameterisation of a Stochastic Model for Human Face Identification”, Proc.of 2nd IEEE Workshop on Applications of Computer Vision, Sarasota, FL, 1994.

[9] http://lrv.fri.uni-lj.si/facedb.htm.

[10] Intelligent Multimedia Lab., “Asian Face Image Database PF01”, Technical Report (Department of Computer Science and Engineering, Pohang University of Science and Technology, Korea), 2001.

[11] K. Messer, J. Matas, J. Kittler, J. Luettin and G. Maitre. “Xm2vtsdb: The extended m2vts database”, Second International Conference on Audio and Video-based Biometric Person Authentication, March 1999.

[12] http://www.ee.surrey.ac.uk/Research/VSSP/xm2vtsdb/

[13] J. Matas, M. Hamouz, K.Jonsson, J. Kittler, Y. Li, C. Kotropoulos, A. Tefas, I. Pitas, T. Tan, H. Yan, F. Smeraldi, J. Bigun, N. Capdevielle, W. Gerstner, S. Ben-Yacoub, Y. Abdeljaoued and E. Mayoraz, “Comparison of face verification results on the XM2VTS database”, Proceedings of the 15th International Conference on Pattern Recognition, Barcelona (Spain), vol. 4, September, pp. 858-863, 2000.

[14] http://www.sic.rma.ac.be/~beumier/DB/3d_rma.htm

[15] http://www-users.cs.york.ac.uk/~tomh/3DFaceDatabase.htm.

[16] S. Gong, S. J. McKenna, A. Psarrou, Dynamic Vision: From Images to Face Recognition, ImperCooege Press, 2000.

[17] http://konicaminolta.com

[18] A. B. Moreno, Á. Sánchez, José Fco. Vélez y Fco. Javier Díaz, “Face recognition using 3D surface-extracted descriptors”, The Irish Machine Vision & Image Processing Conference (IMVIP'03), September 2003.

[19] A. B. Moreno, Á. Sánchez and J. F. Vélez, “Automatic Location of 3D Feature Points in Facial Meshes”, Proc. of the IASTED Intl. Conf. on Applied Informatics, v. II: Artificial Intelligence & Applications, pp. 571-576, February 2001.

[20] 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.