State of The Art

In

3D Face Recognition

Index

Index of Figures

INDEX OF TABLES

1 From 2D to 3D

Evaluations such FERET Tests and Face Recognition Vendor Test 2002 [6], have underlined that the current state of the art in 2D face recognition is not yet sufficient for the biometric applications. Indeed, the performance of the 2D face recognition algorithms are still heavily affected by the pose and illumination conditions of the subject. A lot of efforts have been done in order to overcome these drawbacks, but obtained results, in terms of recognition rate confirmed that this is a very hard problem. Nowadays 3D capturing process is becoming cheaper and faster, so that recent works attempt to solve the problem directly on a 3D model of the face. Few databases of 3D face model are publicly available until today and in the most cases the number of subjects is limited as well as the quality is often very low. The constant progress of the 3D capturing technologies influenced also the kind of the recognition algorithms. Indeed, the first algorithms applied directly on a clouds of points [8], after a suitable triangulation, while the more recent ones work directly on a mesh, considering in some cases the information provided both by the 3D shape and texture. The 3D_RMA is an example of a database of 3D face models represented by clouds of points. For long time it has been the only publicly available database, even if its quality is rather low. On the contrary 3D meshes (wrl, 3ds, …) are available today from the newer technologies, but in most cases they are just proprietary databases.

2 The most interesting algorithms

The vast majority of face recognition research, and all of the major commercial face recognition systems, use normal intensity images of the face. All these methods are referred as 2D. In opposite, the 3D considers the entire shape of the head. Furthermore, as well as done for the 2D, the word *identification* is used to refer a one-to-many matching, so that a best matching can be find, while with *verification* is meant a one-to-one matching to verify a claimed identity. At last *multi-modal* refers to a strategy, that combines 2D and 3D information in order to achieve the recognition task. Recently, 3D Face Recognition is attracting the attention of the researcher because it is often though that the use of 3D has the potential for greater recognition accuracy than the use of 2D face images. The greatest advantages are the invariants with respect to the pose and illumination conditions. Furthermore, allowing the use of such curvature descriptors the 3D models have the potential for higher accuracy in describing the facial surfaces, particularly for areas such the cheeks, forehead and chin. However also drawbacks are present. Indeed, time spent in computations on 3D models are often so long, even minutes. On the other hand, the acquisition process is still too much delicate process, making the 3D approach not yet well applicable in practice. Although early work on 3D face recognition was done over ten years ago, the number of published papers is relatively small. A brief description of the most interesting algorithms in the literature is given in the following.

2.1 The algorithms

2.1.1 Face Recognition Using Range Images [1]

Some of the first approaches to the 3D face recognition worked on range data directly obtained by range sensors, due to the low costs of this hardware, respect to the laser scanners used for example by Gordon in [5]. In fact, in [1] the range images are acquired by means of a structured light approach. The most important disadvantage of this choice are the missing data due to the occlusions or improperly reflected regions. This problem is then avoided using two sensors rather than one, and applying a merging step for integrating the obtained 3D data. The initial step, consists in calibrating the sensors, so that such parameters as projection matrix, camera direction, … are computed. Then merged images are computed, that is for every original 3D data point, the coordinates in the merged range image are calculated based on the parameters of the virtual sensor. If 3D points of two different surfaces have to be mapped onto the same pixel, a sort of z-buffering is applied to disambiguate the mapping. The template images obtained from this acquisition process are then used as training and testing set for two different approaches. The first one are the eigenfaces. The dimension of the space of face templates is reduces applying the Principal Component Analysis both for training and testing, so that for each testing image the nearest one in terms of euclidean distance is searched. Another method is also tested on the template images, the HMMs (Hidden Markov Models). As this technique is only applicable on one-dimensional signals, the template images are first transformed in a mono-dimensional signal by means of a slide window, that move on the image from the top to the bottom and from the left to the right. The HMM has five states. For every person in the database, the parameters of the HMM are calculated in a training phase. When a test image is presented, the probability of producing this image is computed by means of the Viterbi algorithm. All images in the database have a size of 75×150 pixels. Since both the method require a training phase, the dataset has been partitioned in two subsets of 120 training and 120 test images. Results are shown in Table 1. For the experiments reported under the category *smoothing*, no rotation was done, but an additional smoothing step was applied, with $\sigma=0.5$ and $\sigma=1.5$. On the contrary, the rotation are controlled, so that the rotation around the *y* axis is constantly 30° and the rotation around the *x* axis is 20°.

Preprocessing	Eigenfaces	HMM
No preprocessing	97.50 %	90.83 %
Smoothing s =0.5	98.33 %	90.00 %
Smoothing s =1.5	98.33 %	76.67 %
Rotation	100.00 %	89.17 %

Table 1 Results of the range based method in terms of recognition rate.

2.1.2 Face Recognition Based on Depth and Curvature Features [5]

In past 2D approaches, the features used in describing faces have been limited to eyes, nose, mouth and face boundaries, neglecting the additional information provided by low contrast area, such as jaw boundary, cheeks and forehead. Then is clear that an approach based on range and curvature data has several advantages over intensity image approaches by virtue of the more available information. Furthermore curvature has the valuable characteristic of being *viewpoint invariant*. This method defines a set of high level features, which are eyes, nose and head, and includes the following features: *Nose bridge* (nasion), *Nose base* (base of septum), *Nose ridge*, *Eye corner cavities* (inner and outer), *Convex center of the eye* (eyeball), *Eye socket boundary*, *Boundary surrounding nose*, *Opposing positions on the cheeks*. Each of these regions on the face image is described in terms of a set of relationships of depth and curvature values. Since several region can respect a set of constraints, this set is designed in order to reduce the search to a single definition of the feature. The set of constraints is given by:

- sign of Gaussian and mean curvature,
- absolute extent of a region on the surface,
- distance from the symmetry plane,
- proximity to a target on the surface,
- protrusion from the surrounding surface,
- local conFig.tion of curvature extrema.

The high level features and regions are used to compute a set of low level features, where the most basic scalar features correspond to measurements of distances. The set of low level descriptors is given by: *left and right eye width*, *eye separation*, *total span of eyes*, *nose height/width/depth*, *head width*, *maximum Gaussian curvature on the nose ridge*, *average minimum curvature on the nose ridge*, *Gaussian curvature at the nose bridge and base*. For each face image, this set of features is computed, placing it in the space of all possible faces, while the Euclidean distance is used as measure in the scaled feature space. Two different training sets are used for feature detection and recognition. The former consists of 26 subjects, while the latter includes 8 faces with 3 views each for a total of 24 faces. Two different characteristics of the method are assessed in the experimentation. The first is a classification of the low level features on the basis of two main properties: *robustness in detection* (their measurements have to be consistent for the same face also when pose and/or expression change), *discriminating power* (their values must vary distinctly over the range of different individuals). The second is the recognition rate obtained for different sets of low level features. The results are shown in Table 2. For each of the targets there are two faces with the same identity remaining in the database. Table 2 shows for each feature set the percentage of targets for which the best match was correct (top), and the percentage of the targets for which the second best match was also correct (bottom). The basic set denoted (I), includes the best 4 features head width, nose height, depth and width, while the other three set include increasing numbers of features added according to their discriminating power.

Feature set considered	Recognition Rates
I	75.0 % 70.8 %

II	91.7 % 83.3 %
IV	95.8 % 79.2 %
III	100.00 % 79.2 %

Table 2 Results of the depth and Curvature based face recognition method.

2.1.3 A New Attempt To Face Recognition Using 3D Eigenfaces [8]

This method apply on the face models of the 3D_RMA database, in which models are represented by scattered point clouds. So the first problem to be addressed consists in building the mesh from the clouds of points. This is done by means of an iterative algorithm. At first the nose tip is localised as the most prominent point in the point cloud. Then a basic mesh is aligned with the point cloud, subdivided and tuned step by step as shown in Fig. 1. Four step are considered enough for the refinement process. Point clouds have different orientations, and resulting meshes preserve this orientation, so an average model is computed and all the meshes are aligned with it, tuning six parameters for the rotations and six for the translations. Due to the noise, some built mesh models cannot describe the geometric shape of the individual. These mesh models are called non face models. Each mesh contains 545 nodes and it is used as a bi-dimensional intensity image, in which the intensity of the pixel is the Z-coordinate of the corresponding node. The eigenfaces technique is applied to these intensity images. A subset of the computed images is used for the training. Let M_1, M_2, …, M_n the mesh images in the training set and let be M_{aver} the average mesh image and Fi= M_1 - M_{aver}. Then the covariance matrix is computed as:

$$C = \frac{1}{n}\sum_{i=1}^{n} \Phi_i \Phi_i^T = AA^T$$

The eigenvalues and the eigenvectors of the matrix C are computed, keeping only the $e<n$ orthogonal eigenvectors u_1, u_2, …, u_e, which correspond to the $e<n$ largest eigenvalues. Both training and testing mesh images are projected on this space, while 20 most dominant 3D eigenfaces are considered, so that a vector $V \in R^{1\times 20}$ is associated to each mesh model. The database used for the experiments is the 3D_RMA, which consists of subjects acquired in two different sessions. From these sessions two databases are built: automatic DB (120 persons) and manual DB (30 persons). After computing the similarity differences between test samples and the training data, the nearest neighbor classifier (NN) and the k-nearest neighbor classifier (KNN) are used for recognition.

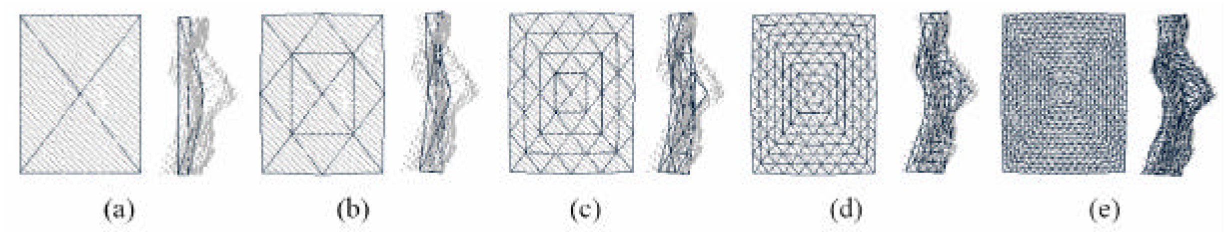

Fig. 1 The regulated mesh models in different levels. (a) Basic mesh. (b) Level one. (c) Level two. (d) Level three. (e) Level four.

The identification accuracy is evaluated on different subsets of the 3D_RMA database. Because of the limited quantity of samples, the Leave-one-out Cross Validation method is used (i.e. each time a mesh image is left out as a test sample and the training is done on the remainder), while the CMS (Cumulative Match Score) is used to evaluate the identification performances. Results are reported in the following tables which show the CCR (Correct Classification Rate) in manual DB (using NN and KNN), the CCR in the first 30 persons of the automatic DB (with and without non-face meshes, using NN and KNN) and the CCR in the automatic DB (with and without non-face meshes, using NN and KNN). Table 3, Table 4 and Table 5

Database	NN	KNN
Manual DB, session 1 (3 instances for each)	92.2 %	92.2 %
Manual DB, session 2 (3 instances for each)	84.4 %	84.4 %
Manual DB, sessions 1 and 2 (6 instances for each)	93.9 %	93.9 %

Table 3 CCR in manual DB (using NN and KNN).

Database	**First 30 Models**		**Non-face meshes removed (22 persons)**	
	NN	**KNN**	**NN**	**KNN**
Automatic DB, session 1 (3 instances for each)	71.1 %	73.3 %	83.3 %	83.3 %
Automatic DB, session 2 (3 instances for each)	80.0 %	80.0 %	89.4 %	89.4 %
Automatic DB, session 2 (3 instances for each)	80.6 %	82.2 %	92.4 %	92.4 %

Table 4 CCR in the first 30 persons of the automatic DB (with and without non-face meshes, using NN and KNN).

Database	All Models (120 persons)		Non-face meshes removed (91 persons)	
	NN	KNN	NN	KNN
Automatic DB, session 1 (3 instances for each)	59.2 %	60.3 %	71.1 %	71.8 %
Automatic DB, session 2 (3 instances for each)	59.2 %	61.1 %	67.4 %	68.5 %
Automatic DB, session 2 (3 instances for each)	69.4 %	71.1 %	79.3 %	80.2 %

Table 5 CCR in the automatic DB (with and without non-face meshes, using NN and KNN).

2.1.4 Face Recognition Based on Fitting a 3D Morphable Model [3]

This face recognition system combines deformable 3D models with a computer graphics simulation of projection and illumination. Given a single image of a person the algorithm automatically estimates 3D shape, texture and all relevant 3D scene parameters. The morphable face model is based on a vector space representation of faces. This space is constructed, such that any convex combination of the examples S_i and T_i belonging to the space, describes a human face:

$$S = \sum_{i=1}^{m} a_i S_i \ \ and\ T = \sum_{i=1}^{m} a_i T_i$$

In order to assure that continues changes on a_i and b_i represent a transition from one face to another, avoiding artifacts a dense point-to-point correspondence constraint has to be guaranteed. This is done by means of a generalization of the optical flow technique on gray-level images to the three-dimensional surfaces. Vectors S and T are directly extracted from the 3D model, where S is the concatenation of the Cartesian coordinates (x, y, z) of the 3D points and T is the concatenation of the corresponding texture information (R, G, B). Furthermore the PCA is applied to the vectors S_i and T_i of the example faces i=1, 2, …, m, while the Phong’s model is used to describe the diffuse and specular reflection of the surface. In this way an average morphable model is derived from scans (obtained with a Cyberware™ 3030PS laser scanner) of 100 males and 100 females, from 18 to 45 years old. Then, by means of a cost function, the fitting algorithm optimizes a set of shape coefficients and texture coefficients along with 22 rendering parameters concatenated in a feature vector ρ, such as pose angles, 3D translations, focal length, … . Two paradigms have been used in order to test the method. In the first one all gallery images are analyzed by the fitting algorithm, and the shape and texture coefficients are stored. In the same way, for a probe image all the coefficients are computed and compared with all gallery data, in order to find the best match, a graphical representation is given in Fig. 2. In the second one, the three-dimension face reconstruction is used in order to generate synthetic views of the subjects in a 2D face database, which are then transferred to a second viewpoint-dependent recognition system.

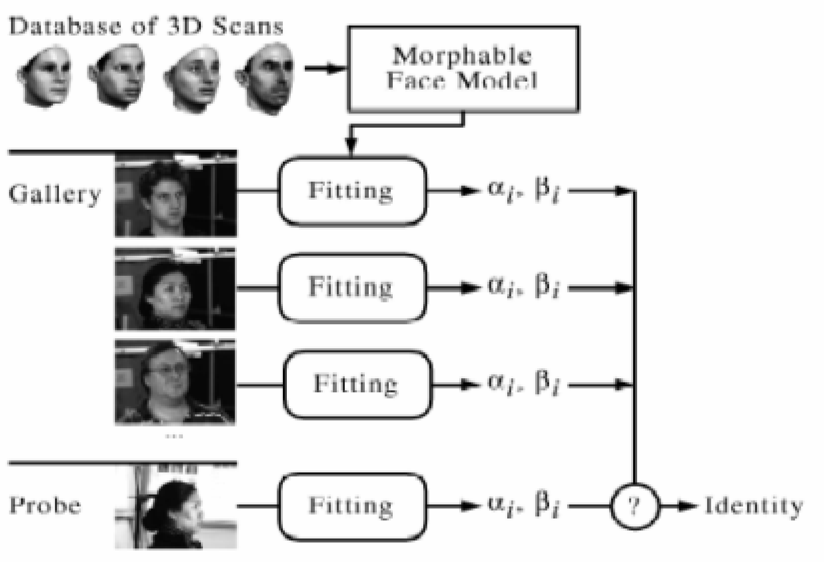

Fig. 2 (source: article [2]): Derived from a database of 200 laser scans, the 3d morphable model is used to encode gallery and probe images. For identification, the model coefficients a and ß of the probe image are compared with the stored coefficients of all gallery images.

Model fitting and identification have been tested on two publicly available databases of images: CMU-PIE et FERET. The individuals in these databases are not contained in the set of 3D scans that form the morphable face model. The reconstruction algorithm is run on all 4,488 PIE and 1,940 FERET images. For all images, the starting condition is the average face at a front view and with frontal illumination. On each image, between six and eight feature points are manually defined, while the following distance function has been used to compare two different faces c_i:

$d_w = \frac{\langle c_1, c_2 \rangle_W}{\|c_1\|_W \cdot \|c_2\|_W}$ with $\langle c_1, c_2 \rangle_W = \langle c_1, C_W^{-1} c_2 \rangle$ and where C_W is the covariance matrix.

Table 6 and Table 7 show some detailed results on the PIE and FERET databases:

Probe view	Gallery view		
	Front	Side	Profile
Front	99.8 %	99.5 %	83.0 %
Side	97.8 %	99.9 %	86.2 %
Profile	79.5 %	85.7 %	98.3 %
Total	92.3 %	95.0 %	89.0 %

Table 6 Mean average identification rate on the CMU-PIE data set, averaged over all lightening conditions for front, side and profile view galleries.

Probe view	Correct identification
Ba	**(gallery)**
Bb	94.8%
Bc	95.4%
Bd	96.9%
Be	99.5%
Bf	97.4%
Bg	96.4%
Bh	95.4%
Bi	90.7%
Bk	96.9%
Total	95.9%

Table 7 Gallery images are frontal views extracted from Ba. Only condition Bk has different illumination condition than the others.

2.1.5 3D Face Modeling Using Two Orthogonal Views and a Generic Face Model [2]

This method uses the 3D coordinates of a set of facial feature points, calculated from two images of a subject, in order to deform a generic 3D face model. Images are grabbed by two cameras with perpendicular optical axes. The 3D generic model is centered and aligned by means of the procrustes analysis, which models the global deformation, while local deformation are described by means of the 3D spline curves. The front and profile view of a subject are shown in Fig. 3. They are used in order to locate facial features, such as eyes and mouth, by means of a probabilistic approaches. An example of the obtained 3D model is given in Fig. 4. Twenty-nine vertices are kept on this model, divided in two subsets, 15 principal vertices and 14 additional vertices.

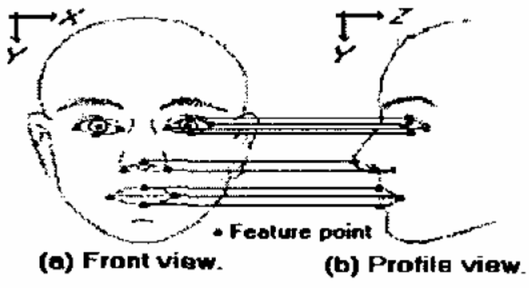

Fig. 3 Front and profile view of a subject.

The algorithm has been applied to 26 subjects, giving a gallery of 26 models, each one characterized by a feature vector of 29 vertices coordinates. Then given the two views of a test face, the coordinates of the 29 feature points are computed and compared with all the models in the gallery, in order to find the best match. It results that 25 people of the 26 subjects are classified correctly, with a recognition rate of 96.2%.

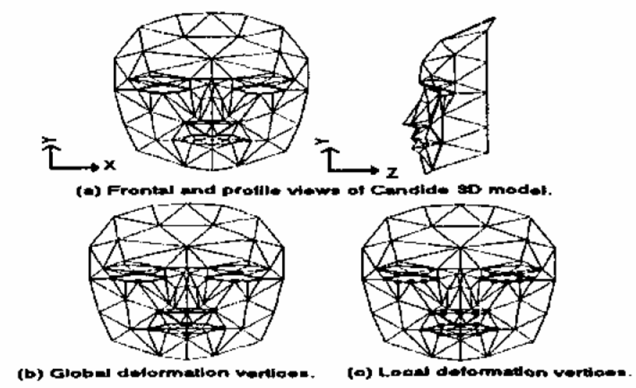

Fig. 4 Generic 3D model

2.1.6 2D+3D Multi-modal Systems

Multi-Modal systems have the aim to improve the recognition rate of the existing 2D approaches, combining results obtained from both a 3D and a 2D recognition system. Two interesting approaches among the recent works are:

Wang [9]: the method apply on both range data and texture. In the 3D domain, the Point Signature is used in order to describe the shape of the face, while the Gabor filters are applied on the texture in order to localize and characterize ten control points (corners of the eyes, nose tip, corner of the mouth, …). The PCA analysis is then applied separately to the obtained 3D and 2D feature vectors, and then the resulting vectors are integrated to form an augmented vector which is used to represent each facial images. For a given test facial image, the best match in the galley is identified according to similarity function or SVM (Support Vector Machine). The experiments are conducted on a database of 50 subjects, with different poses.

Chang [4]: this work is a report on PCA-based recognition experiments performed using shape data and texture data from 200 persons. A total of 278 subject have been acquired with a Minolta Vivid 900 range scanner. The probe set consists of 166 individuals, while the training set contain 278 subjects, including which are acquired for testing. The scanned images are normalized, in order to fill holes and remove spikes. The 3D and 2D data are treaty separately and then a different fusion technique are tested for integrating the data, while Euclidean and Mahalanobis distance are tested. The experiments shown in Fig. 5. Confirm that the Multi-modal overcomes both the 2D and 3D, when considered singularly.

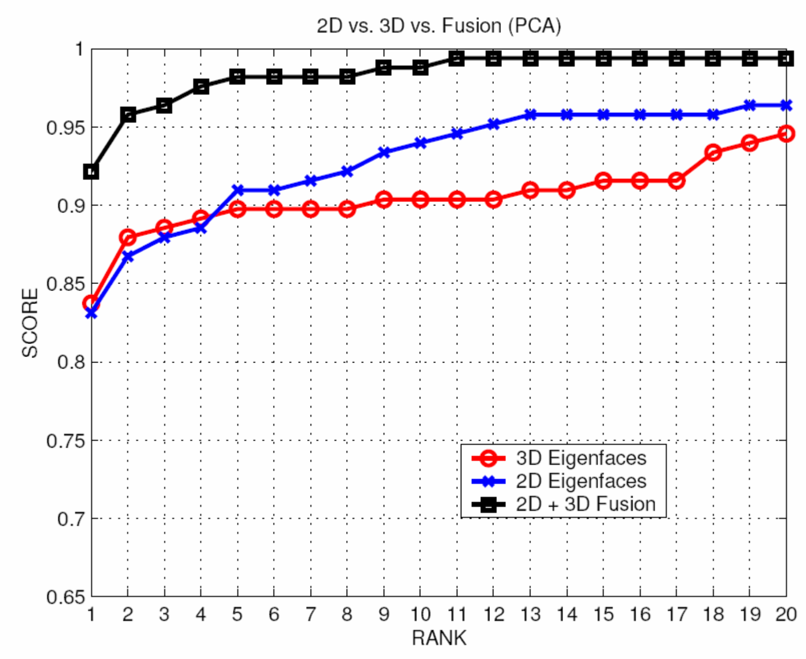

Fig. 5 Single- versus multi-modal biometrics.

3 Discussion

One limitation to some existing approaches to 3D face recognition involves sensitivity to size variation. Approaches that use a purely curvature based representation, such as extended Gaussian images, are not able to distinguish between two faces of similar shape but different size. On the contrary, approaches based on PCA or ICP (Iterative closest Point), avoid this problem but their performances throw down when changes in expression are present between gallery and probe images. Also for multi-modal recognition system there is the need of more sophisticated combination. Indeed in most cases, the score results are combined separately for 2D and 3D, and then combined together. It is at least potentially more powerful to exploit possible synergies between the two modalities. Another interesting problem is the absence of an appropriate standard dataset, with a large number and demographic variety of people, which images have been taken at repeated intervals of time, with meaningful changes in expression, pose and lighting as exists in 2D.

References

[1] Achermann B, Jiang X., Bunke H., "*Face Recognition Using Range Images*", Proc. Of the International Conference on the Virtual Systems and Multimedia (VSMM97), pp: 129-136, September 1997.

[2] Ansari A., Abdel-Mottaleb M.,"*3D Face Modeling Using Two Orthogonal Views and a Generic Face Model*", Proc. Of the International Conference on Multimedia and Expo (ICME '03), Vol. 3, pp.: 289-292, July 2003.

[3] Blanz V. and Vetter T., "*Face Recognition Based on Fitting a 3D Morphable Model*", IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 25, no. 9, pp.: 1063-1074, September 2003.

[4] Chang K. I., Bowyer K. W. and Flynn P. J., "*Multi-Modal 2D and 3D Biometrics for Face recognition*", Proc. of the IEEE International Workshop on Analysis and Modeling of Faces and Gestures (AMFG'03), pp. 187–194, October 2003.

[5] Gordon G. Gaile. "*Face Recognition Based on Depth and Curvature Features*", Proc. of IEEE Computer Society Conference on Computer Vision & Pattern recognition, pp: 808-810, June 1992.

[6] Phillips P. J., Grother P., Michaels R. J., Blackburn D. M., Tabassi E. and Bone J., "*FRVT 2002: overview and summary*". Available at www.frvt.org, March 2003.

[7] Phillips P. J., Wechsler H., Huang J. and Rauss P., „The FERET Database and Evaluation Procedure for Face Recognition Algorithms", Image and Vision Computing J., vol. 16, no. 5, pp.: 295-306, 1998.

[8] Xu C., Wang Y., Tan T., Long Q., "*A new Attempt to Face Recognition Using 3D Eigenfaces*", the 6th Asian Conference on Computer Vision (ACCV), Vol. 2, pp.884-889, 2004.

[9] Wang Y., Chua C. and Ho Y., "*Facial Feature Detection and Face Recognition from 2D and 3D images*", Pattern Recognition Letters, 23:1191-1202, 2002