A face recognition algorithm based on multiple individual discriminative models

Jens Fagertun, David Delgado Gomez, Bjarne K. Ersbøll, Rasmus Larsen

***Abstract*—In this paper, a novel algorithm for facial recognition is proposed. The technique combines the color texture and geometrical configuration provided by face images. Landmarks and pixel intensities are used by Principal Component Analysis and Fisher Linear Discriminant Analysis to associate a one dimensional projection to each person belonging to a reference data set. Each of these projections discriminates the associated person with respect to all other people in the data set. These projections combined with a proposed classification algorithm are able to statistically deciding if a new facial image corresponds to a person in the database. Each projection is also able to visualizing the most discriminative facial features of the person associated to the projection. The performance of the proposed method is tested in two experiments. Results point out the proposed technique as an accurate and robust tool for facial identification and unknown detection.**

***Index Terms*—Face recognition, Principal Component Analysis, Fisher Linear Discriminant Analysis, Biometrics, Multi-Subspace Method.**

I. INTRODUCTION

Regrettable events which happened during the last years (New York, Madrid) have revealed flaws in the existing security systems. The vulnerability of most of the current security and personal identification system is frequently shown. Falsification of identity cards or intrusion into physical and virtual areas by cracking alphanumerical passwords appear frequently in the media. These facts have triggered a real necessity for reliable, user-friendly and widely acceptable control mechanisms for person identification and verification.

Biometrics, which bases the person authentication on the intrinsic aspects of a human being, appears as a viable alternative to more traditional approaches (such as PIN codes or passwords). Among the oldest biometrics techniques, fingerprint recognition can be found. It is known that this technique was used in China around 700 AD to officially certify contracts. Afterwards, in Europe, it was used as person identification in the middle of the 19^{th} century. A more recent biometric technique used for people identification is iris recognition [8]. It has been calculated that the chance of finding two randomly formed identical irises is one in 10^{78} (The population of the earth is below 10^{10}) [7]. This technique has started to be used as and alternative to passport in some airports in United Kingdom, Canada and Netherlands. It is also used as employee control access to restricted areas in Canadian airports and in the New York JFK airport. The inconvenient of these techniques is the necessity of interaction with the individual who wants to be identified or authenticated. This fact has caused that face recognition, a non-intrusive technique, has increasedly received the interest from the scientific community in recent years.

The first developed techniques that aimed at identifying people from facial images were based on geometrical information. Relative distances between key points, such as mouth corners or eyes, were calculated and used to characterize faces [17]. Therefore, most of the developed techniques during the first stages of facial recognition focused on the automatic detection of individual facial features. However, facial feature detection and measurements techniques developed to date are not reliable enough for the geometric feature based recognition, and such geometric properties alone are inadequate for face recognition because rich information contained in the facial texture or appearance is discarded [6], [13]. This fact produced that gradually most of the geometrical approaches were abandoned for color based techniques, which provided better results. These methods aligned the different faces to obtain a correspondence between pixels intensities. A nearest neighbor classifier used these aligned values to classify the different faces. This coarse method was notably enhanced with the appearance of the Eigenfaces technique [15]. Instead of directly comparing the pixel intensities of the different face images, the dimension of these input intensities were first reduced by a principal component analysis (PCA). This technique settled the basis of many of the current image based facial recognition schemes. Among these current techniques, Fisherfaces can be found. This technique, widely used and referred [2], [4], combines the Eigenfaces with Fisher linear discriminant analysis (FLDA) to obtain a better separation of the individuals. In Fisherfaces, the dimension of the input intensity vectors is reduced by PCA and then FLDA is applied to obtain a good separation of the different persons.

After Fisherfaces, many related techniques have been proposed. These new techniques aim at providing a projection that attain a good person discrimination and also are robust at differences in illumination or image pose. Kernel Fisherfaces [16], Laplacianfaces [10] or discriminative common vectors [3] can be found among these new approaches. Typically, these techniques have been tested assuming that the image to be classified corresponds to one of the people in the database. In these approaches, the image is usually classified to the person with the smallest Euclidean distance.

However, some inconveniences appear when the person to be analyzed may not belong to the data set. In this case, a criterium to decide if the person belongs to the data set has to be chosen. E.g. only people with an euclidian distance less than a given threshold are considered as belonging to the data set. However, this threshold has not to be necessarily the same for all the classes (different persons) and different thresholds would need to be found. The estimation of these thresholds is not straightforward and additional data might be needed.

In this work, a new technique that addresses the different inconveniences is proposed. The proposed techniques takes advantage of two novelties in order to deal with these inconveniences. First, not only the texture intensities are taken into account but also the geometrical information. Second, the data are projected into n one-dimensional spaces instead of a $(n-1)$-dimensional space, where n is the number of people in the data set.

Each of these individual models aims at characterizing a given person uniquely. This means that every person in the data set is represented by one model. These multi one-dimensional models allow to statistically interpret the "degree of membership" of a person to the data set and to detect unknowns. Furthermore, these two facts have several advantages in interpretability, characterization, accuracy and easiness to update the model.

II. Algorithm Description

The proposed algorithm is made up of two steps. In the first step, an individual model is built for each person in the database using the color and geometrical information provided by the available images. Each model characterizes a given person and discriminates it from the other people in the database. The second step carries out the identification. A classifier, related with the standard Gaussian distribution, decides if a face image belongs to one person in the database or not. In this section, the two parts of the algorithm are described in detail. A diagram of the algorithm is displayed in Fig. 1. This diagram will be referred during the description of the algorithm to obtain an easier understanding.

A. Creating the individual models

1) Obtaining the geometry of the face: The geometrical characterization of a given face is obtained by means of the theory of statistical shape analysis [1]. In this theory, objects (faces) are represented by shapes. According to Kendall [11], a shape is all the geometrical information that remains when location, scale and rotational effects are filtered out from an object. In order to describe a shape, a set of landmarks or points of correspondence that matches between and within populations are placed on each face. As an example, Fig. 2A displays a set of 22 landmarks. These landmarks indicate the position of the eyebrows, eyes, nose, mouth, jaw and size of a given face.

To obtain a shape representation according to the definition, the obtained landmarks are aligned in order to remove the location, rotational and scaling effects. To achieve this goal, the 2D-full Procrustes analysis is carried out. Briefly, let:

$$\mathbf{X} = \{\mathbf{x}_i\} = \{x_i + i \cdot y_i\}, \quad i = 1, \ldots, n$$

be a set of n landmarks expressed in complex notation. In order to apply full Procrustes analysis, the shapes are initially

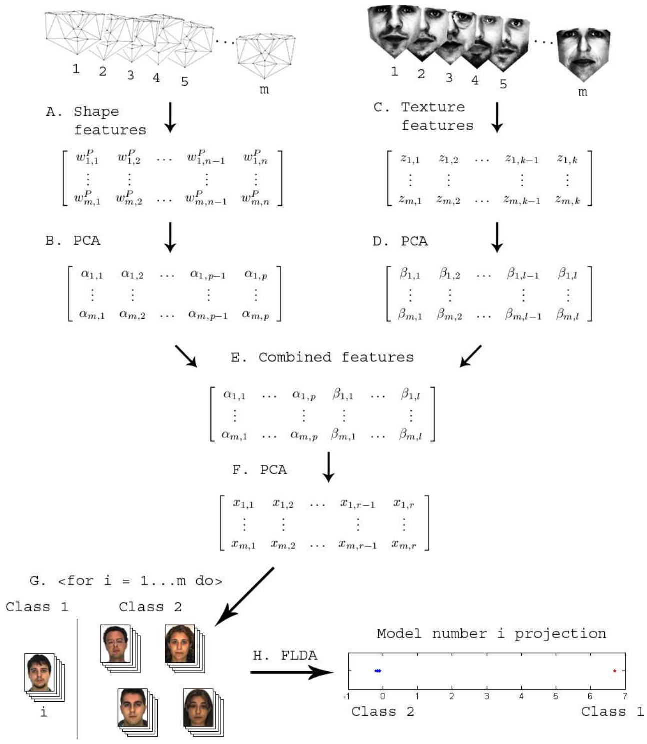
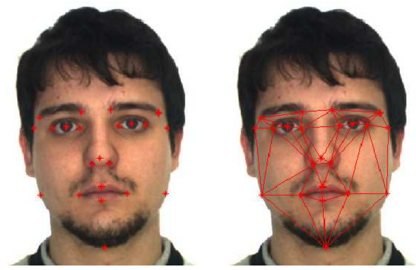


Fig. 1. Algorithm overview. A: Landmarks alignment using full Procrustes analysis. B: PCA on aligned landmarks to remove redundancy. C: Texture normalization using global histogram equalization. D: PCA on normalized texture to remove redundancy. E: Combining shape and texture features. F: PCA on combined features to remove redundancy. G & H :In turn build the individual model using FLDA.



(A) (B)

Fig. 2. (A) Set of 22 landmarks placed on a face image. (B) The Delaunay triangulation of the 22 landmarks.

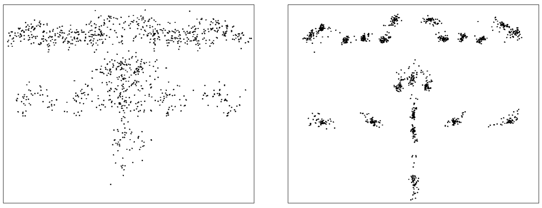
centered. To center the different shapes, the mean of the shape, $\bar{\mathbf{x}}$, is subtracted from each landmark:

$$\mathbf{w}_i = \mathbf{x}_i - \bar{\mathbf{x}}, \quad i = 1, \ldots, n$$

.

The full Procrustes mean shape [12], $\hat{\mu}$, is found as the eigenvector corresponding to the largest eigenvalue of the complex sum of squares and products matrix

$$\sum_{i=1}^{n} \mathbf{w}_i \mathbf{w}_i^* / (\mathbf{w}_i^* \mathbf{w}_i)$$



(A) (B)

Fig. 3. (A) Superimposition of the sets of 22 landmarks obtained over 49 different face images. (B) Alignment of the landmarks.

where $\mathbf{w}_i^*$ denotes the transpose of the complex conjugate of $\mathbf{w}_i$. Using this Procrustes mean shape, the full Procrustes coordinates of $\mathbf{w}_1, \ldots, \mathbf{w}_n$ (Fig. 1) A) are obtained by

$$\mathbf{w}_i^P = \mathbf{w}_i^* \hat{\mu} \mathbf{w}_i / (\mathbf{w}_i^* \mathbf{w}_i) \quad i = 1, \ldots, n$$

Fig. 3A displays the superimposition of the set of 22 landmarks described in Fig. 2, obtained on 49 different face images. The result obtained after applying the full Procustres alignment on theses landmarks can be observed in Fig. 3B. In order to remove redundancy in the data, a Principal Component Analysis is applied to the aligned landmarks (Fig. 1 B).

2) Texture formulation: To form a complete model of the face appearance, the algorithm also captures the texture information provided by the pixels. In order to collect this texture representation, the Delaunay triangulation of every shape is obtained. The Delanuay triangulation connects the aligned landmark set of each image by a mesh of triangles, so no triangle has any of the other points of the set inside its circumcircle. The Delaunay triangulation obtained for each image is warped onto the Delaunay triangulation of the mean shape. The Delanuay triangulation of the 22 landmarks is displayed in Fig. 2B.

Formally, let I be a given image and M the mean shape previously obtained. Let $\mathbf{u}_1 = [x_1, y_1], \mathbf{u}_2 = [x_2, y_2]$ and $\mathbf{u}_3 = [x_3, y_3]$ denote the vertices of a triangle T in I, and let $\mathbf{v}_1, \mathbf{v}_2$ and $\mathbf{v}_3$ be the associated vertices of the corresponding triangle in M. Given any internal point $\hat{\mathbf{u}} = [x, y]$ in the triangle T, the corresponding point in the associated triangle in the mean shape can be written as $\hat{\mathbf{v}} = \alpha\mathbf{v}_1 + \beta\mathbf{v}_2 + \gamma\mathbf{v}_3$ where:

$$\begin{aligned}
\alpha &= 1 - (\beta + \gamma) \\
\beta &= \frac{yx_3 - x_1y - x_3y_1 - y_3x + x_1y_3 + xy_1}{-x_2y_3 + x_2y_1 + x_1y_3 + x_3y_2 - x_3y_1 - x_1y_2} \\
\gamma &= \frac{xy_2 - xy_1 - x_1y_2 - x_2y + x_2y_1 + x_1y}{-x_2y_3 + x_2y_1 + x_1y_3 + x_3y_2 - x_3y_1 - x_1y_2}
\end{aligned}$$

This transformation extracts the texture of a given face image. A histogram equalization is applied to the collected texture to reduce the effects of differences in illumination [9]. This histogram equalization is performed independently in each of the three color channels. Afterwards, the three color channels are converted into gray scale to obtain a more compact representation (Fig. 1 C).

Similarly to the shape analysis, a PCA is conducted in the texture data to reduce dimensionality and data redundancy (Fig. 1 D). However, notice that the large dimension of the texture vectors will produce memory problems because of the huge dimension of the covariance matrix. In order to avoid this difficulty, the Eckart-Young theorem is used [5]. Formally, let $\mathbf{D}$ represents the texture data matrix composed by s n-dimensional texture vectors after the mean of the texture vectors has been subtracted from each one of them ($s << n$). Then the $n \times n$ dimensional covariance matrix can be written as:

$$\mathbf{\Sigma_D} = \frac{1}{s}\mathbf{D}\mathbf{D}^T$$

Let $\mathbf{\Sigma_S}$ be the smaller $s \times s$ dimensional matrix defined by

$$\mathbf{\Sigma_S} = \frac{1}{s}\mathbf{D}^T\mathbf{D}$$

Then the non-zero eigenvalues of the matrices $\mathbf{\Sigma_S}$ and $\mathbf{\Sigma_D}$ are equal. Moreover, the columns of:

$$\mathbf{\Phi_D} = \mathbf{D} \cdot \mathbf{\Phi_S}$$

where the columns of $\mathbf{\Phi_S}$ contain the eigenvectors of $\mathbf{\Sigma_S}$, correspond with the the eigenvectors associated to the non-zero eigenvalues of $\mathbf{\Sigma_D}$ in the sense they have the same direction. Therefore, if the columns of $\mathbf{\Phi_D}$ are normalized, then $\mathbf{\Phi_D}$ holds the normalized eigenvectors of $\mathbf{\Sigma_D}$ that has eigenvalues bigger than zero. This not only avoid problems with the memory but also it gives a substantial speed up of the calculations.

3) Combining color and geometry: The shape and texture features are concatenated in a matrix (Fig. 1 E). In order to remove correlation between shape and texture and also to make the data representation more compact, a third PCA is performed on the concatenated shape and texture matrix (Fig. 1 F).

4) Building an individual model: Once the geometry and texture of the face have been captured, the proposed algorithm builds an individual model for each person in the data set. Each model is built using Fisher linear discriminant analysis. Formally, let $\mathbf{X}$ be the data obtained after combining the shape and texture and applying the PCA. Let n_1 be the number of data elements corresponding to the person for whom the model is being created (class 1) and let n_2 be the number of elements corresponding to the other people (class 2), (Fig. 1 G). Let $\bar{\mathbf{x}}_1$ and $\bar{\mathbf{x}}_2$ be the class mean vectors, $\bar{\mathbf{x}}$ be the total mean vector and $\mathbf{x}_{i,j}$ be the jth sample in the ith class. Then the between matrix is defined by:

$$\mathbf{B} = n_1(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}})(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}})^T + n_2(\bar{\mathbf{x}}_2 - \bar{\mathbf{x}})(\bar{\mathbf{x}}_2 - \bar{\mathbf{x}})^T$$

and the within matrix is defined by:

$$\mathbf{W} = \sum_{i=1}^{2}\sum_{j=1}^{n_i}(\mathbf{x}_{i,j} - \bar{\mathbf{x}}_i)(\mathbf{x}_{i,j} - \bar{\mathbf{x}}_i)^T$$

The projection that best discriminates the two populations is given by the direction of the eigenvector associated to the maximum eigenvalue of $\mathbf{W}^{-1}\mathbf{B}$ (Fig. 1 H). To ensure that the within matrix $\mathbf{W}$ is not singular, only the f first data variables are taken into account, where f is the number of non-zero eigenvalues of the within matrix $\mathbf{W}$.

B. classification

In order to obtain a method to classify a given image, the different individual models are firstly standardized so they can be compared. The standardization of model $i = 1, \ldots, m$ is based on two assumptions. First, the number of observations for person i is much smaller than the number of the observations of all other people. The second assumption is that the projection of the other people follows a Gaussian distribution. These two assumptions imply that the distribution of all the projected facial images on a particular discriminative individual model can be assumed as a Gaussian distribution with outliers. The standardization of model i is then achieved by transforming the projections into a standard Gaussian distribution, keeping the projections of the person i positive. Formally, let $\bar{x}_i$ be the mean of the projections on model i, σ_i the standard deviation, and let $x_{i,j}$ be the projection of image j in model i. These projections are standardized by:

$$\hat{x}_{i,j} = (x_{i,j} - \bar{x}_i)/\sigma_i$$

If the standardized projection for the images corresponding to person i are negative, then $\hat{x}_{i,j}$ are replaced by $-\hat{x}_{i,j}$ for all projections. This causes the projection of the images corresponding to person i to be positive and far from the mean of the gaussian.

Once that the model i is standardized, the probability of a projected image of belonging to the person i is given by the value of the standard normal cumulative function in the projected value. This fact is used to classify a given image. If it is assumed that the image belongs to a person from the data set, the image is projected by all the models and classified as belonging to the model that gives the largest probability. Moreover, it is also statistically possible to decide if a given person belongs to the data set or it is unknown. This can be achieved by comparing the largest projection obtained in all the models with a probabilistic threshold. E.g, if a 99.9% of probability is required, a given image will only be considered as belonging to the database if the projection in one of the individual models is higher than 3.1 standard deviations.

III. Experimental Results

Two experiments are conducted in order to evaluate the performance of the proposed method. The objective of the first experiment is to evaluate the recognition ability in terms of correct classification rates. This first experiment also aims at ranking the importance of shape and texture. The second experiment aims at analyzing if the proposed method can be incorporated into a biometrical facial recognition scheme. The robustness of the proposed method to the presence of unknowns is considered in this second experiment.

A. Experiment one

The first experiment aims at comparing the performance of the proposed method with respect to the Fisherfaces method in terms of correct classification rates. In order to be consistent with a previously published work [15], unknown people are not taken into account.

To achieve this first goal the AR face database [14] is used. The database is composed of two independent sessions, recorded 14 days apart. At both sessions, each person was recorded 13 times, under various facial poses (all frontal), lighting conditions and occlusions. The size of the images in the database is 768 $\times$ 576 pixels, represented in 24 bits RGB color format.

In this study, a subset of 50 persons (25 male and 25 female) from the database was randomly selected. Seven images per person without occlusions are used from each session. Therefore, the experiment data set is composed of 700 images, with 14 images per person. An example of the selected images for one person is displayed in Fig. 4.

Fig. 4. The AR data set: (Top row) The seven images without occlusions from first session, (Bottom row) The seven images without occlusions from the second session.

All the images were manually annotated with the 22 landmarks previously mentioned.

The data set was divided into two sets. The images of the first session were used to build the individual discriminative models, and images from the second session were subsequently used to test the performance.

The landmarks corresponding to the images in the training set were aligned using full Procrustes analysis. The 44 (x,y)-coordinates were obtained to represent the geometrical configuration of each face. In order to obtain the texture of each face in the training set, the different images were warped with respect to the mean shape. Each of the textures received a histogram equalization in each color band to reduce the differences in global illumination. The textures were converted to gray scale and represented by 41337 pixels. The geometrical and color representation of each face was combined, reduced and the individual models were built as described in Section II.

The test set was used to evaluate and compare the proposed method with respect to the Fisherface technique. In order to evaluate the importance of the geometrical information, the Fisherface technique was modified replacing the texture data with the shape data and also combining the shape with the texture. These two modified techniques will be referred to as Fishershape and Fishercombined from now on. The Euclidean Nearest-Neighbor algorithm was used as classifier algorithm

Method	Input features	Correct Classification Rate[1]
Proposed method	Shape	86.4% (95)
Proposed method	Texture	99.6% (3)
Proposed method	Texture and Shape	99.9% (1)
Fishershape	Shape	85% (105)
Fisherface	Texture	98.9% (8)
Fishercombined	Texture and Shape	99.7% (2)

TABLE I

AVERAGE CORRECT CLASSIFICATION RATES.

in the Fisher methods. The proposed method classified the images as the person associated to the model that yields the highest probability.

The test was repeated a second time changing the roles of the training and test sets. So session two was used as training data and session one as test data. The average correct classification rates for the different techniques are shown in Table I.

From Table I, it is observed that the proposed method has a slightly better performance than the Fisher methods. Moreover, it is also noticed that using the texture data one obtains a higher accuracy than when the shape is used. This implies that the information contained in the texture is more significant than that included in the shape. However, the information contained in the shape data is not insignificant. The highest correct classification rate in both techniques is attained when both shape and texture are considered.

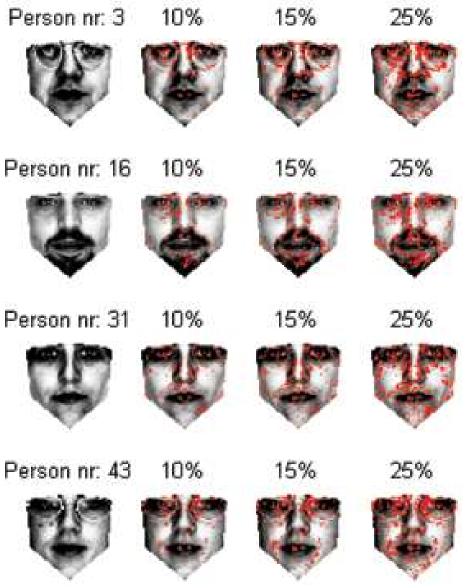


Fig. 5. The 10, 15 and 25% most important pixels (shown in red) for discriminating between the 50 test persons.

An interesting property of the proposed algorithm are that it is possible to determine which are the most discriminative features of a given person. In order to illustrate this fact, four models were built using only the texture. The pixels of the faces corresponding to these models which received the 10, 15 and 25% highest weights in the model are displayed (in red) in Fig. 5. It is clear that important discriminating features include eyes, noses, glasses, moles and beards. Notice that the algorithm detects the glasses and the two moles of person 43 as discriminate features.

[1]Number of misclassified images reported in parentheses.

B. Experiment two

The objective of this second experiment is to test the possibility of incorporating the proposed technique into a biometrical facial recognition scheme. This conveys the identification of people in a data set and also the detection of unknown people. The good performance of the proposed technique in person identification was shown in the previous experiment. Therefore, this second experiment aims at evaluating the performance of the technique in detection of unknown people.

To achieve this goal, the data set used in the previous experiment is selected. In order to evaluate the performance of the technique, a 25-fold crossvalidation was conducted. The seven face images from one male and other seven face images from one female were left out in each iteration. These two people are considered as not belonging to the data set and therefore unknowns. The images of the remaining 48 people were used to train the algorithm.

The average False Acceptance Rate (FAR) and average False Rejection Rate (FRR) graph, can be observed in Fig. 6. The corresponding average Receiver Operating Characteristic curve (ROC) is displayed in Fig. 7.

Both graphs show that the known and unknown populations have a good separability. The best separation happens at the Equal Error Rate (3.1 standard deviations), giving a FAR and FRR of 2%. Moreover, notice that, if the algorithm belongs to a security scheme, the degree of accessibility can be established by increasing or diminishing the standard deviation threshold. E.g., if in the test a false rejection rate of 5.5% is allowed, then a 0% false acceptance rate is obtained. This accommodates biometrical security systems that requires a high level of control access.

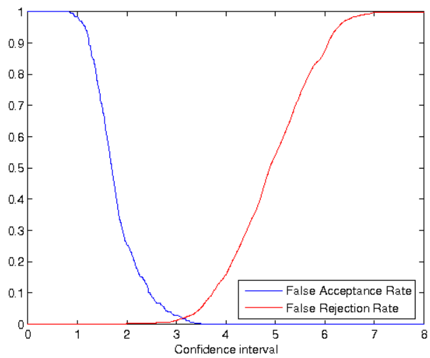


Fig. 6. Average False Acceptance Rate/False Rejections Rate graph obtained by the 25-fold crossvalidation.

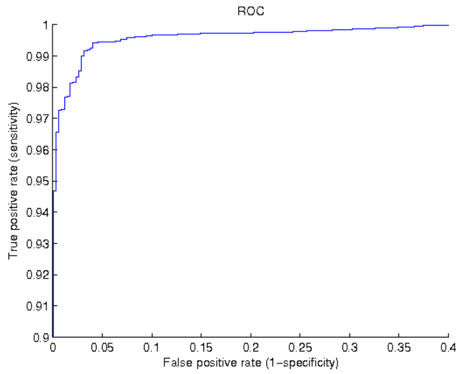


Fig. 7. Average Receiver Operating Characteristic (ROC) curve obtained by the 25-fold crossvalidation. Notice that only the top left part of the ROC curve is displayed here.

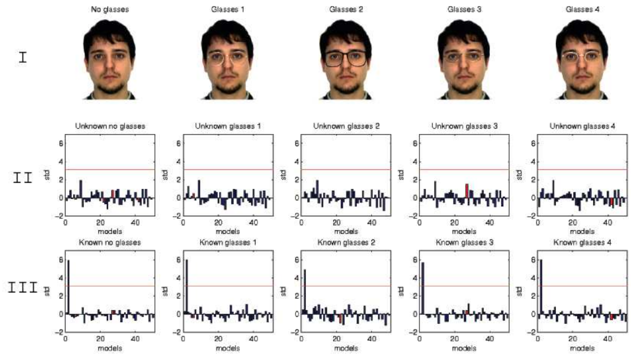


Fig. 8. Impact of changing glasses. (I) Person without glasses and syntectic fitted with 4 glasses form the data set. (II) The corresponding projections in the models as unknown. (III) The corresponding projections in the models as known. Red columns is the model corresponding to the superimposed glasses.

A second test is conducted in order to assess the robustness of the proposed method. This test also aims at showing that the method not only discriminates on removable features, such as glasses. To achieve this goal, eight people (four male and four female) are synthetically fitted with four different glasses taken from people belonging to the data set, giving 32 synthetic images.

This second test consists of two steps. First, these eight people are not used to built the individual models. The goal is to examine if these eight people who do not belong to the data set are considered as one of the person in the data set. Results show that none of the 32 images is misclassified when a threshold of 3.1 standard deviations is considered (probability of correct classification of 99.9%). This fact can be noticed in Fig. 8 II, where the projections of one of the eight unknown people on the different models are displayed. It is observed that, when the person is considered unknown, his projections onto the individual models belonging to the data set are under the selected threshold. This means that the proposed method does not classify any of the unknown people as belonging to the data set.

In the second step, the eight people (without glasses) are also used to build the individuals models. In this case the goal is to analyze if the method can still recognize people belonging to the data set who has slight changes (same people with glasses). In this second step, the 32 images are also classified correctly by the method. In Fig. 8 III, it is observed that the projections onto the individual model associated with this person clearly surpass the threshold. It is also observed that the projections into the individual models associated to the glasses's owners do not increase significantly. Similar graphs are obtained for the other seven people. These results show the suitability of the proposed technique in being incorporated into a biometrical security system.

IV. Discussion and Conclusion

In this paper, a novel method to identify people from face images has been proposed. The developed technique aims at being a precise and robust algorithm that can be incorporated into biometrical security systems. The technique has been tested on face images, but it can also be used in other biometrical data, such as speech. Experimental results have proved that the method can attain better classifications rates than an other widely used technique. Moreover, the final one-dimensional projection allows for a simple interpretation of the results. If a given face image is projected onto the different individual models, it is visually possible to determine if this person belongs to one of the models. Moreover, it is also statistically possible to observe the degree of belonging to that model.

Another of the attracting characteristics of the proposed method is its ability to deal with unknowns. The degree of belonging to the data set can be determined statistically. A decision threshold can be determined in relation to a standard Gaussian distribution. This threshold value is used to set the degree of security of the system. The higher this value is set, the smaller the probability of a person being considered as belonging to the data set.

The robustness of the algorithm has been tested using both known and unknown people. The algorithm has been shown to be robust to the inclusion of artifacts such as glasses. On one hand, unknown people using glasses belonging to people from the data set are still classified as unknown. This fact implies that unknown people would not get access to a security system when they use simple removable features belonging to people from the data set. On the other hand, known people using glasses, belonging to other people from the data set, are still recognized as themselves. This means if someone gets glasses, the associated model does not need to be recalculated. Moreover, this fact suggests that the database should be composed of facial images without glasses. This was also shown by observing that the individual model projections do not change significantly when the glasses were placed.

Another interesting property of the proposed method is its easiness to be maintained and updated. If a large data set is available, it is not needed to recalculate all the existing individual models when a new person has to be registered. Simply, a new individual model for the new person is created.

Similarly, if a person has to be removed from the database, it is only needed to remove the corresponding individual model. In conclusion, an accurate, robust and easily adaptable technique to be used for facial recognition has been developed and demonstrated.

REFERENCES

[1] *Statistical Shape Analysis*. Wiley series in probability and statistics, 1998.

[2] P.N. Belhumeur, J.P. Hespanha, and D.J. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 19(7):711–720, July 1997.

[3] Hakan Cevikalp, Marian Neamtu, Mitch Wilkes, and Atalay Barkana. Discriminative common vectors for face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(1):4–13, 2005.

[4] Songcan Chen and Daohong Li. Modified linear discriminant analysis. *Pattern Recognition*, (38):441 443, 2005.

[5] T. F. Cootes and C. J. Taylor. Statistical models of appearance for computer vision. Technical report, Imaging Science and Biomedical Engineering. University of Manchester, March 2004.

[6] I. J. Cox, J. Ghosn, and P. N. Yianilos. Feature-based face recognition using mixture-distance. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 209–216, June 1996.

[7] J. Daugman. High confidence visual recognition of persons by a test of statistical independence. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(11):1148–1161, 1993.

[8] J. Daugman. How iris recognition works. *Proceedings of 2002 International Conf. on Image Processing*, 1, 2002.

[9] G. Finlayson, S. Hordley, G. Schaefer, and G. Y. Tian. Illuminant and device invariant colour using histogram equalisation. *Pattern Recognition*, 38, 2005.

[10] Xiaofei He, Shuicheng Yan, Yuxiao Hu, Partha Niyogi, and Hong-Jiang Zhang. Face recognition using laplacianfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(3):328–340, 2005.

[11] D.G. Kendall. The diffusion of shape. *Advances in Applied Probabilities*, (9):428–430, 1977.

[12] J. T. Kent. The complex bingham distribution and shape analysis. *Proceedings in current Issues in Statistical Shape Analysis*, pages 167–175.

[13] S. Z. Li and A. K. Jain. *Handbook of face recognition*. Springer, 2005.

[14] A.M. Martinez and R. Benavente. The ar face database. Technical Report 24, Computer Vision Center Purdue University, June 1998.

[15] M. Turk and A.P. Pentland. Face recognition using eigenfaces. *IEEE Conf. Computer Vision and Pattern Recognition*, 1991.

[16] Ming-Hsuan Yang. Kernel eigenfaces vs. kernel fisherfaces: Face recognition using kernel methods. *Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition*, pages 205–211, 2002.

[17] A. Yuille, P. Hallinan, and D. Cohen. Feature extraction from faces using deformable templates. *International Journal of computer Vision*, 1.