A multi-linear discriminant analysis of 2D frontal face images

Carlos Eduardo Thomaz and Vagner do Amaral
Department of Electrical Engineering, FEI, São Paulo, Brazil

Gilson Antonio Giraldi
Department of Computer Science, LNCC, Rio de Janeiro, Brazil

Edson Caoru Kitani
Department of Electrical Engineering, USP, São Paulo, Brazil

João Ricardo Sato
Center of Mathematics, Computation and Cognition, UFABC, São Paulo, Brazil

Duncan Fyfe Gillies
Department of Computing, Imperial College, London, UK

Abstract

We have designed and implemented a multi-linear discriminant method of constructing and quantifying statistically significant changes on human identity photographs. The method is based on a general multivariate two-stage linear framework that addresses the small sample size problem in high-dimensional spaces. Starting with a 2D face data set of well framed images, we determine a most characteristic direction of change by organizing the data according to the features of interest. Our goal here is to use all the facial image features simultaneously rather than separate models for texture and shape information. Our experiments show that the method does produce plausible unseen views for gender, facial expression and ageing changes. We believe that this method could be widely applied for normalization in face recognition and in identifying subjects after a lapse of time.

1. Introduction

Multivariate statistical approaches have played an important role of recognizing face images and characterizing their differences. The importance of using multivariate techniques to analyze face images is related to the well-known fact that face images are highly redundant not only owing to the evidence that the image intensities of adjacent pixels are often correlated but also because every individual has some common facial features such as mouth, nose, and eyes. As a consequence, an input image with n pixels can be projected onto a lower dimensional space without significant loss of information.

The most straightforward and successful statistical models for visual interpretation of well-framed face images have been based on Principal Component Analysis (PCA). Since the pioneering works of Sirovich and Kirby [16], and Turk and Pentland [20], published approximately 20 years ago, several subsequent works have projected face images on a Principal Component Analysis (PCA) feature space to not only reduce the dimensionality of the original samples for further classification and analysis but also to interpret and reconstruct the most expressive components [17] described by all the training images. Impressive results on this latter goal have been achieved by the well-known Active Appearance Model (AAM) proposed by Cootes et al. [3, 1, 4, 2]. Unfortunately, since the AAMs rely on PCA directions ranked by the principle of maximum variance, the first principal components with the largest eigenvalues do not necessarily represent important discriminant directions to separate sample groups.

In this paper, we have designed and implemented a multi-linear discriminant method of constructing and quantifying statistically significant unseen views of human identity photographs. Given a single photograph of an unseen subject it is possible to construct new images with, for example, a range of different expressions or with different gender characteristics. The method could be widely applied for normalization in face recognition and in identifying subjects after a lapse of time. It is based on the use of a two-stage separating hyper-plane called Statistical Discriminant Model (SDM)[10]. Starting with a data set of well framed images, we determine a most characteristic direction of change by organizing the data according to the features of interest. For example, we may identify one group where all the subjects are smiling, and a second group where all the subjects have a neutral expression. If we now find the best separating hyperplane of these two groups, for example by using a linear discriminant regularized method called Maximum uncertainty Linear Discriminant Analysis (MLDA) [19, 13, 12], we can use its normal vector to define the most characteristic direction of change. Given a new subject image we can adjust it by moving parallel to this direction in the image space. So, for instance, we can transform a face image with a neutral expression into a smiling one or vice versa. The constructed images represent the maximum likelihood estimate of the appearance of the subject given the data set that we start with. The method uses all the facial image features simultaneously rather than as separate texture and shape models. Our experiments on two dimensional image sets show that the method does produce visually plausible results for gender, facial expression and ageing facial changes.

The remainder of this paper is divided as follows. In section 2, we briefly review PCA and highlight its importance as a multivariate technique on reducing the high dimensionality of face images but without losing variance information. Section 3 describes the standard linear discriminant analysis (LDA) and states the reasons for using a maximum uncertainty version of this approach to perform the face experiments required. The estimation of the separating hyperplane and the implementation of the Statistical Discriminant Model are described in section 4. In section 5, we describe all the experiments carried out in this work for the gender, facial expression and ageing analyzes. In section 6, we discuss a potential application for the SDM approach focused on the age-progression using familiar features. Finally, in section 7, we conclude the paper, summarizing its main contributions and describing possible future work.

2. Principal Components Analysis (PCA)

Although PCA is a well-known statistical technique [6, 9] that has been used in several image recognition problems, especially for dimensionality reduction, we provide next a brief description of PCA for the sake of completeness.

Let us consider the face recognition problem as an example to illustrate the main idea of the PCA. In any image analysis, and particularly in face image analysis, an input image with n pixels can be treated as a point in an n-dimensional space called the image space. The coordinates of this point represent the values of each pixel of the image and form a vector $x^T = [x_1, x_2, \ldots, x_n]$ obtained by concatenating the rows (or columns) of the image matrix. It is well-known that well-framed face images are highly redundant. As a consequence, an input image with n pixels can be projected onto a lower dimensional space without significant loss of information.

Thus, let an $N \times n$ training set matrix X be composed of N input face images with n pixels. This means that each column of matrix X represents the values of a particular pixel observed all over the N images. Let this data matrix X have covariance matrix S with respectively P and Λ eigenvector and eigenvalue matrices, that is,

$$P^T S P = \Lambda. \tag{1}$$

It is a proven result that the set of m ($m \leq n$) eigenvectors of S, which corresponds to the m largest eigenvalues, minimizes the mean square reconstruction error over all choices of m orthonormal basis vectors [6]. Such a set of eigenvectors that defines a new uncorrelated coordinate system for the training set matrix X is known as the principal components. In the context of face recognition, those $P_{pca} = [p_1, p_2, ..., p_m]$ components are frequently called eigenfaces [20]. The m principal components can then replace the initial n variables and the original data set, consisting of N measurements on n variables, is reduced to a data set consisting of N measurements on m principal components.

3. Maximum Uncertainty LDA (MLDA)

The primary purpose of the Linear Discriminant Analysis, or simply LDA, is to separate samples of distinct groups by maximizing their between-class separability while minimizing their within-class variability. LDA assumes implicitly that the true covariance matrices of each class are equal because the same within-class scatter matrix is used for all the classes considered.

Let the between-class scatter matrix S_b be defined as

$$S_b = \sum_{i=1}^{g} N_i(\overline{x}_i - \overline{x})(\overline{x}_i - \overline{x})^T \tag{2}$$

and the within-class scatter matrix S_w be defined as

$$S_w = \sum_{i=1}^{g}(N_i - 1)S_i = \sum_{i=1}^{g}\sum_{j=1}^{N_i}(x_{i,j} - \overline{x}_i)(x_{i,j} - \overline{x}_i)^T \tag{3}$$

where $x_{i,j}$ is the n-dimensional pattern (or sample) j from class π_i, N_i is the number of training patterns from class π_i, and g is the total number of classes or groups. The vector $\overline{x}_i$ and matrix S_i are respectively the unbiased sample mean and sample covariance matrix of class π_i [6]. The grand mean vector $\overline{x}$ is given by

$$\overline{x} = \frac{1}{N}\sum_{i=1}^{g} N_i \overline{x}_i = \frac{1}{N}\sum_{i=1}^{g}\sum_{j=1}^{N_i} x_{i,j}, \tag{4}$$

where N is, as described earlier, the total number of samples, that is, $N = N_1 + N_2 + \ldots + N_g$.

The main objective of LDA is to find a projection matrix W_{lda} that maximizes the ratio of the determinant of the between-class scatter matrix to the determinant of the within-class scatter matrix (Fisher's criterion), that is,

$$W_{lda} = \arg\max_{W} \frac{\left|W^T S_b W\right|}{\left|W^T S_w W\right|}. \tag{5}$$

The Fisher's criterion described in equation (5) is maximized when the projection matrix W_{lda} is composed of the eigenvectors of $S_w^{-1}S_b$ with at most $(g-1)$ nonzero corresponding eigenvalues [6, 5]. In the case of a two-class problem, the LDA projection matrix is in fact the leading eigenvector w_{lda} of $S_w^{-1}S_b$, assuming that S_w is invertible.

However, in limited sample and high dimensional problems, such as in face images analysis, S_w is either singular or mathematically unstable and the standard LDA cannot be used to perform the separating task. To avoid both critical issues, we have calculated w_{lda} by using a maximum uncertainty LDA-based approach (MLDA) that considers the issue of stabilizing the S_w estimate with a multiple of the identity matrix [18, 19]. In a study [19] with application to the face recognition problem, Thomaz et. al showed that the MLDA approach improved the LDA classification performance with or without a PCA intermediate step and using less linear discriminant features. The w_{mlda} is calculated by replacing S_w in the Fisher's criterion formula described in equation (5) with its regularization version. This regularization is based on the maximum entropy idea that in limited sample size and high dimensional problems where the within-class scatter matrix is singular or poorly estimated, the Fisher's linear basis found by minimizing a more difficult but appropriate *inflated* within-class scatter matrix would also minimize a less reliable *shrivelled* within-class estimate [19].

4. Statistical Discriminant Model (SDM)

The Statistical Discriminant Model is a two-stage PCA+MLDA separating hyperplane that reduces the dimensionality of the original images and extracts discriminant information from images [10].

In order to estimate the SDM separating hyper-plane, we use training examples and their corresponding labels to construct the classifier. First a training set is selected and the average image vector of all the training images is calculated and subtracted from each n-dimensional vector. Then the training matrix composed of zero mean image vectors is used as input to compute the P_{pca} transformation matrix. The columns of this n x m transformation matrix are eigenvectors, not necessarily in eigenvalues descending order. We have retained all the PCA eigenvectors with nonzero eigenvalues, that is, $m = N - 1$, to reproduce the total variability of the samples with no loss of information. It is important to emphasize that this PCA intermediate step has been applied here because $n \gg N$, allowing the MLDA scatter matrices to be calculable in computers with a normal memory size. In situations where $N \gg n$, the SDM approach does not need such PCA step for dimensionality reduction.

Thus, the zero mean image vectors are projected on the principal components and reduced to m-dimensional vectors representing the most expressive features of each one of the n-dimensional image vector. Afterwards, this N x m data matrix is used as input to calculate the W_{mlda} discriminant transformation matrix, as described in the previous section. Since in this work we have limited ourselves to two-group classification problems, there is only one w_{mlda} discriminant eigenvector. The most discriminant feature of each one of the m-dimensional vectors is obtained by multiplying the N x m most expressive features matrix by the m x 1 MLDA linear discriminant eigenvector. Hence, the initial training set of face images consisting of N measurements on n variables, is reduced to a data set consisting of N measurements on only 1 most discriminant feature.

Once the two-stage SDM classifier has been constructed, we can move along its corresponding projection vector and extract the discriminant differences captured by the classifier. Any point on the discriminant feature space can be converted to its corresponding n-dimensional image vector by simply: (1) multiplying that particular point by the w_{mlda} linear discriminant eigenvector previously computed; (2) multiplying its m most expressive features by the P_{pca} transformation matrix; and (3) adding the average image calculated in the training stage to the n-dimensional image vector. Therefore, assuming that the spreads of the classes follow a Gaussian distribution and applying limits to the variance of each group, such as $\pm 3\sigma_i$, where σ_i is the standard deviation of each group $i \in \{1, 2\}$, we can move along the SDM most discriminant features and map the results back into the image domain.

Additionally, any face image $x_{i,j}$ that followed the same acquisition and spatial normalization protocols can incorporate the discriminant information captured by the two-stage linear classifier. More specifically, this procedure of transferring the most discriminant feature can be generated through the following expression [19, 7, 13, 12]:

$$y_{i,j} = x_{i,j} + j\sigma_i \cdot P_{pca} w_{mlda}, \tag{6}$$

where $j \in \{-3, -2, -1, 0, 1, 2, 3\}$. This operation is useful not only to transfer the most discriminant feature to any point on the original image space, but also, and most importantly, predict how the discriminant information can affect a particular sample that does not necessarily belong to the training set.

5. Experimental Results

We have used frontal images of a face database publicly available[1] to carry out the experiments. This face database contains a subset of 400 well framed frontal 2D face images, with 2 images (one with a neutral or non-smiling expression and the other with a smiling facial expression) for each of 200 individuals (100 men and 100 women). All images are taken against a white homogenous background in an upright frontal position and scale might vary about 10%. The original size of each image is 640x480 pixels.

To minimize image variations that are not necessarily related to differences between the faces, we automatically aligned all the frontal face images using the directions of the eyes as a measure of reference so that the pixel-wise features extracted from the images correspond roughly to the same location across all subjects. For implementation convenience, all the frontal images were then cropped to the size of 300x250 pixels and converted to 8-bit grey scale. Most images represent subjects between 19 and 40 years old with distinct appearance, hairstyle, and adornments.

We have performed the following multi-linear discriminant analyzes: male versus female (gender), non-smiling versus smiling (facial expression) and young versus old (ageing) experiments. For the gender experiment we have composed a training set of 200 frontal male images, i.e. a mixture of non-smiling and smiling male images, and 200 analogous frontal female images. For the expression experiments, we have used the 200 frontal non-smiling images, i.e. a mixture of male and female images, and their respective frontal smiling images. For the ageing experiments, we have composed the young training set of 354 images (a mixture of non-smiling and smiling face images of 177 subjects under 30 years of age) and the old training set of 46 images (a mixture of non-smiling and smiling face images of 23 subjects over 30 years of age).

5.1. Interpretation and Reconstruction of the PCA Most Expressive Components

As the average face image is an n-dimensional point ($n = 300 \times 250 = 75000$) that retains all common features from the training sets, we could use this point to understand what happens visually when we move along the principal components and reconstruct the respective coordinates on the image space. Analogously to the works by Cootes et al. [1, 2, 3, 4], we have reconstructed the new average face images by changing each principal component separately using the limits of $\pm 3\sqrt{\lambda_i}$, where λ_i are the corresponding largest eigenvalues.

Figure 1. Interpretation and reconstruction of the first eight, from top to bottom, PCA most expressive components. From left to right: $[-3\sqrt{\lambda_i}, -2\sqrt{\lambda_i}, -1\sqrt{\lambda_i}, \overline{x}, +1\sqrt{\lambda_i}, +2\sqrt{\lambda_i}, +3\sqrt{\lambda_i}]$, where $i = \{1, 2, \ldots, 8\}$.

Figure 1 illustrates the transformations on the first eight PCA most expressive components, that is, the first principal components that describe at least 2% of the total variance of all the 2D frontal face images. Since we have used the same training images for gender, facial expression and ageing experiments, the PCA most expressive components are equal in all the experiments.

Looking at the Figure 1, it is important to note that because changes in either facial expression or ageing are much less significant, PCA is unable to capture such minor variations in its first most expressive components. These results are expected because PCA tends to capture features

1 http://www.fei.edu.br/∼cet/facedatabase.html

that have a considerable variation between all training samples, like changes in illumination, gender, and shape of the head in our experiments. Therefore, if we need to identify specific changes such as the variation in facial expression, PCA has not proved to be a useful solution for this problem. In fact, when we consider a whole intensity-level model to perform the PCA analysis, there is no guarantee that a single principal component will capture a specific variation alone, no matter how discriminant that variation might be.

5.2. Interpretation and Reconstruction of the Most Discriminant Hyper-planes

To perform the multi-linear discriminant analysis on the face images, we have used the training sets previously selected and their respective labels to construct the linear classifiers corresponding to the gender, facial expression and ageing separation tasks. Since in these experiments we have limited ourselves to two-group classification problems, there is only one SDM discriminant eigenvector per separation task.

Figure 2 presents, from top to bottom, the SDM most discriminant features for the gender, facial expression and ageing experiments. It displays the image regions captured by the SDM approach that change when we move from the left side (group 1 of male, non-smiling and young labeled samples) of the dividing hyper-plane to the right one (corresponding group 2 of female, smiling and old labeled samples), following limits to the standard deviation and mean of each sample group.

As can be seen, the SDM hyper-plane effectively extracts the group differences, showing separately the gender, facial expression and ageing features that mainly distinguish the sample groups, without enhancing other image artifacts. For instance, in the first row of Figure 2, from top to bottom, there are some gender variations that are more significant and consequently predominant in the most discriminant direction selected, such as the shape of the face or head, presence or absence of beard or moustache, flatness and length of the nose, and thickness of the eyebrows and shape of the eyes. Analogously to the gender experiments, it is possible to see that the SDM hyper-plane has been able to capture the subtle facial expression changes and nothing else, showing exactly what we should expect intuitively from a face image when someone changes their expression from non-smiling to smiling. In fact, it is possible to note that the SDM most discriminant direction has predicted a facial expression not necessarily present in all the face images, that is, the ”definitely non-smiling” or may be ”anger” status and the ”definitely smiling” or may be ”happy” status described respectively by the images $-3\sigma_1$ and $+3\sigma_2$ in the second row, from top to bottom, of Figure 2. The third row of Figure 2, from top to bottom, displays the image regions captured by the SDM hyper-plane that change when we move from the group 1 of labeled samples under 30 years of age (left side) of the dividing hyper-plane to the group 2 of labeled samples over 30 years of age (right side). Despite the very different sample sizes of these groups, it is possible to see that the SDM hyper-plane captures a number of plausible changes owing to ageing, such as thickness of the eyelids and lips, growth of the nose, and a general reduction of the skin elasticity with the appearance of facial wrinkles.

Figure 2. Interpretation and reconstruction, from top to bottom, of the most discriminant features captured by the SDM hyper-planes for the gender, facial expression and ageing experiments. From left (group 1 of either male, non-smiling or young labeled samples) to right (group 2 of either female, smiling or old labeled samples): $[-3\sigma_1, \overline{x}_1, +1\sigma_1, boundary, -1\sigma_2, \overline{x}_2, +3\sigma_2]$.

5.3. Effect Size of the Multi-Linear Discriminant Differences Found

In the previous sub-section, the detection of the differences have been based only on visual inspection of the most discriminant features. In this sub-section, we investigate the effectiveness of the separating hyper-planes on recognizing the group samples and the statistical significance of the discriminant changes found for all the gender, facial expression and ageing experiments.

We have adopted the leave-one-out method to evaluate the classification performance of the multi-linear classifiers. Throughout all the classification experiments, we have assumed that the prior probabilities and misclassification costs are equal for both groups of the two-group experiments. On the PCA+MLDA subspace, the mean of each class has been calculated from the corresponding training samples and the Euclidean distance from each class mean has been used to assign a test observation to either the male

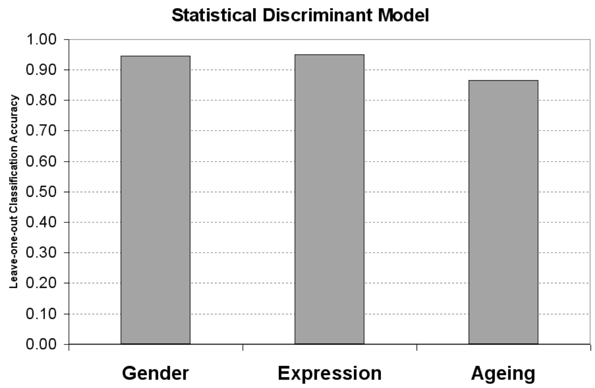

Figure 3. Leave-one-face-out recognition rates of the multi-linear classifiers.

or female groups in the gender experiment, to either the non-smiling or smiling groups in the facial expression experiment, and to either the young or old groups in the ageing experiment.

Figure 3 shows the leave-one-out recognition rates of the multi-linear classifiers. As can be seen, in all discriminant experiments, the SDM approach achieves high recognition rates, that is, 94.5%, 95% and 86.5% for the gender, facial expression and ageing experiments respectively. These classification results indicate that all the discriminant changes found using the 2D frontal pre-aligned face images can be sufficiently extracted by the linear SDM classifiers.

In order to determine and rank the statistical significance of the SDM changes and avoid the use of raw units that are quite arbitrary or lack meaning outside the investigation, we have calculated the effect size of the differences, that is,

$$e = \frac{x_1^* - x_2^*}{\sigma_p} \tag{7}$$

where x_1^* and x_2^* denote, as shown in Figure 4, the transformed images for the statistical extremes points calculated at 3 standard deviations from each corresponding sample group on the SDM separating hyper-plane and projected back on the original image space. The parameter σ_p corresponds to the pooled standard deviation of the sample images and is given by

$$\sigma_p = \sqrt{\frac{(N_1 - 1)\sigma_1^2 + (N_2 - 1)\sigma_2^2}{(N_1 + N_2 - 2)}} \tag{8}$$

where σ_1^2 and σ_2^2 are the variances of each sample group and, as a reminder, N_1 and N_2 are the number of training samples of group 1 and group 2. We have used the pooled variance rather than the variance of each sample group in the e-values because the number of samples is limited and N_1 and N_2 might be very different from each other.

Figure 4. SDM statistical extremes: (top) male and female models; (middle) non-smiling and smiling models; (bottom) young and old models.

Figure 5 illustrates the spatial distribution of the intensity e-changes superimposed on the average face image. We consider a SDM difference important if its e-value exceeds 1 pooled standard deviation, that is, $\|e\| \geq 1$. In both pictures the colour-scales red-yellow and blue-green shows relative intensity change as a range of the effect size. In red-yellow the tissues contained within the lines are brighter ($e \geq 1$) in the male, non-smiling and young groups compared to the female, smiling and old groups, respectively. Analogously, the areas in blue-green show discriminant regions of relative tissue darkness ($e \leq -1$) in the male, non-smiling and young groups compared to the female, smiling and old ones. Face regions contained within the lines and closer to the spectrum of yellow and green show areas of relatively larger statistical significance.

We can see clearly that by exploring the separating hyper-planes found by the multi-linear discriminant analysis and quantifying its most statistically significant changes with the e-values we are able to identify and highlight facial features that are most discriminant between the group samples, such as: forehead, eyebrow, eyes, nose, upper lip, chin and neck for the gender experiments; eyes, shadow, cheek, lips and mouth for the facial expression experiments; and eyebrow, eyes, nose, and skin elasticity around the lips for

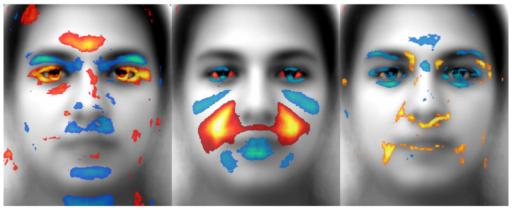

Figure 5. Statistical significance of the SDM differences: (left) effect size of the intensity changes described by the male and female SDM models; (middle) effect size of the intensity changes described by the non-smiling and smiling SDM models; (right) effect size of the intensity changes described by the young and old SDM models.

the ageing experiments.

6. Discussion

We believe that the importance of the SDM results for face interpretation and reconstruction, particularly on the ageing experiments, is beyond the scope of capturing and extracting discriminant information. For instance, the proposed method can be seen as an automatic framework to predict the age-progression in frontal face images using familiar features acquired under controlled conditions. Recently, a number of researchers [11, 8, 15, 14] have modeled facial changes with age using either PCA on different models for texture and shape information [11, 15, 14] or 3D facial meshes in shape space only [8]. The main difference of our approach over these works is based on the fact that we use a linear classifier and all the intensity features simultaneously to estimate the ageing effects on the original image space.

In order to evaluate the SDM approach on the problem of estimating the age-progression in frontal face images using familiar features, we carried out some further experiments by selecting two sets of frontal images (with at least one neutral and one smiling facial expression) of two distinct families composed respectively of a very small number of 6 and 3 subjects over 37 years of age. Figure 6 shows the frontal images used in this experiment, where the first 26 images, from top to bottom and left to right, are from the family 1 and the others 22 from family 2.

Figure 7 illustrates the transformed images for the family 1 and family 2 statistical extremes points calculated at 3 standard deviations from each corresponding sample group on the SDM separating hyper-plane and the spatial distribution of these e-changes superimposed on a reference image. Despite the small sizes of both sample groups, it is possible to see that the SDM hyper-plane captures a number of changes inherent to the families considered and never quantified, such as the relative size of the head and neck, thickness of the eyebrows, shape of the nose, and a light difference on the colour of the facial skin.

Figure 6. Families face samples.

Figure 8 illustrates these most discriminant family features transferred to an example image, not included on the training set, when we move it to the statistical extremes us-

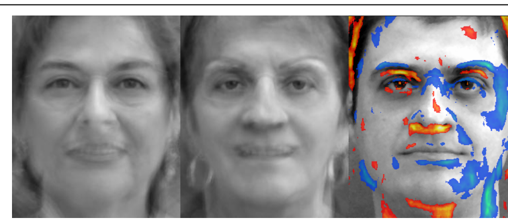

Figure 7. SDM statistical extremes and corresponding significance of the family e-changes superimposed on a reference image: (left) family 1 model; (middle) family 2 model; (right) statistical significance of the SDM differences.

Figure 8. SDM reconstruction when we move an example image (middle) in parallel to the statistical extremes of the family hyper-plane using equation (6): (left) to family 1 model; (right) to family 2 model.

ing equation (6) and the direction defined by the SDM separating hyper-plane. Despite the appearance of some artifacts especially around the head, such as the earrings, due the lack of an exclusive facial cropping on the pre-processing step, we can note the main differences between the families captured by the SDM hyperplane and translated to the example image, like the relative size, position and shape of the eyes, eyebrows, nose, and head.

7. Conclusion

In this work, we designed and implemented a multi-linear method of constructing plausible and statistical significant unseen views of human identity photographs. Since the multi-linear statistical discriminant approach is not restricted to any particular set of samples and involves the same operations irrespective of the complexity of the experiment, straightforward improvements can be made to this approach by using larger training sets with non-rigid spatially normalized images. We believe that the detailed description provided by the multi-linear analysis can facilitate, for instance, forensic specialists on the task of recognizing missing children and adults, particularly in situations where ethnic, gender, and parental face image samples are available. Further work is being undertaken to investigate this possibility in practice.

Acknowledgements

The authors would like to thank the support provided by FAPESP (2005/02899-4), CNPq (472386/2007-7, 301858/2007-1, 303976/2008-0) and CAPES (094/2007).

References

[1] T. Cootes, G. Edwards, and C. Taylor. Active appearance models. *Proceedings of ECCV*, 2:484–498, 1998.

[2] T. Cootes and A. Lanitis. Statistical models of appearance for computer vision. *Technical Report*, 2004.

[3] T. Cootes, C. Taylor, D. Cooper, and J. Graham. Active shape models - their training and application. *Computer Vision and Image Understanding*, 61(1):38–59, 1995.

[4] T. Cootes, K. Walker, and C. Taylor. View-based active appearance models. *4th International Conference on Automatic Face and Gesture Recognition*, pages 227–232, 2000.

[5] P. Devijver and J. Kittler. *Pattern Classification: A Statistical Approach*. Prentice-Hall, 1982.

[6] K. Fukunaga. *Introduction to Statistical Pattern Recognition*. Academic Press, New York, 1990.

[7] G. A. Giraldi, P. S. Rodrigues, E. C. Kitani, J. R. Sato, and C. E. Thomaz. Statistical learning approaches for discriminant features selection. *Journal of the Brazilian Computer Society*, 14(2):7–22, 2008.

[8] T. Hutton, B. Buxton, P. Hammond, and H. Potts. Estimating average growth trajectories in shape-space using kernel smoothing. *IEEE Trans. Med. Imag.*, 22(6):747–753, 2003.

[9] R. Johnson and D. Wichern. *Applied Multivariate Statistical Analysis*. New Jersey: Prentice Hall, 1998.

[10] E. Kitani, C. Thomaz, and D. Gillies. A statistical discriminant model for face interpretation and reconstruction. In *Proceedings of SIBGRAPI'06*, pages 247–254. IEEE CS Press, October 2006.

[11] A. Lanitis, C. Taylor, and T. Cootes. Toward automatic simulation of ageing effects on face images. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(4):442–455, 2002.

[12] J. R. Sato, A. Fujita, C. E. Thomaz, M. G. Morais-Martin, J. Mourao-Miranda, M. J. Brammer, and E. A. Junior. Evaluating svm and mlda in the extraction of discriminant regions for mental state prediction. *NeuroImage*, 46(1):105–114, 2009.

[13] J. R. Sato, C. E. Thomaz, E. F. Cardoso, A. Fujita, M. G. Morais-Martin, and E. A. Junior. Hyperplane navigation: a method to set individual scores in fmri group datasets. *NeuroImage*, 42(4):1473–1480, 2008.

[14] C. Scandrett, C. Solomon, and S. Gibson. A person-specific, rigorous aging model of the human face. *Pattern Recognition Letters*, 27:1776–1787, 2006.

[15] C. Scandrett, C. Solomon, and S. Gibson. Towards a semi-automatic method for the statistically rigorous ageing of the human face. *IEE Proc.-Vis. Image Signal Process.*, 153(5):639–649, 2006.

[16] L. Sirovich and M. Kirby. Low-dimensional procedure for the characterization of human faces. *Journal of Optical Society of America*, 4:519–524, 1987.

[17] D. Swets and J. Weng. Using discriminants eigenfeatures for image retrieval. *IEEE Trans. Patterns Anal. Mach Intell.*, 18(8):831–836, 1996.

[18] C. Thomaz, D. Gillies, and R. Feitosa. A new covariance estimate for bayesian classifiers in biometric recognition. *IEEE Transactions on Circuits and Systems for Video Technology*, 14(2):214–223, 2004.

[19] C. Thomaz, E. Kitani, and D. Gillies. A maximum uncertainty lda-based approach for limited sample size problems - with application to face recognition. *Journal of the Brazilian Computer Society*, 12(2):7–18, 2006.

[20] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3:71–86, 1991.