Regression Techniques versus Discriminative Methods for Face Recognition

Vitomir Štruc, France Mihelič, Rok Gajšek and Nikola Pavešić

*Abstract***— In the field of face recognition it is generally believed that "state of the art" recognition rates can only be achieved when discriminative (e.g., linear or generalized discriminant analysis) rather than expressive (e.g., principal or kernel principal component analysis) methods are used for facial feature extraction. However, while being superior in terms of the recognition rates, the discriminative techniques still exhibit some shortcomings when compared to the expressive approaches. More specifically, they suffer from the so-called *small sample size* (SSS) problem which is regularly encountered in the field of face recognition and occurs when the sample dimensionality is larger than the number of available training samples per subject. In this type of problems, the discriminative techniques need modifications in order to be feasible, but even in their most elaborate forms require at least two training samples per subject. The expressive approaches, on the other hand, are not susceptible to the SSS problem and are thus applicable even in the most extreme case of the small sample size problem, i.e., when only one training sample per subject is available. Nevertheless, in this paper we will show that the recognition performance of the expressive methods can match (or in some cases surpass) that of the discriminative techniques if the expressive feature extraction approaches are used as multivariate regression techniques with a pre-designed response matrix that encodes the class-membership of the training samples. The effectiveness of the regression techniques for face recognition is demonstrated in a series of experiments performed on the ORL database. Additionally a comparative assessment of the regression techniques and popular discriminative approaches is presented.**

I. INTRODUCTION

Over the past decades, automatic face recognition has become a highly active research area, mainly due to the countless application possibilities in both the private as well as the public sector [1]. Automated face recognition systems offer a possible way of improving security in various domains ranging from access control, e-commerce, e-banking, e-government and health monitoring applications to automated user-authentications at ATMs, borders and airports.

Face recognition, being a sub-discipline of biometrics[1], has several advantages when compared to the classically employed knowledge- (e.g., passwords, PINs) or token-based (e.g., ID cards) security schemes [2]. Passwords and PINs can be forgotten, ID cards can be lost or stolen. The human face, on the other hand, cannot be stolen nor forgotten and is, furthermore, unique for each individual. Clearly, it holds a great potential in serving as means for authentication and/or identification of people.

Security schemes, however, are not the only application domain of face recognition systems. They are often found in conjunction with ambient intelligence (and smart house/home) applications where they are used for profile managing. For example, when a person enters a room or house, the face recognition system identifies the person and adjusts the environment (e.g., lighting conditions, music, etc.) in accordance with his/her personal profile.

As we have seen, automated face recognition system are suitable for various applications, however, a number of shortcomings have to be sorted out to improve their performance. One of the major issues with face recognition systems is their performance under the lack of training data. While the existing face recognition techniques work well when a sufficient number of facial images is available for training, the majority of them suffers with their recognition performance when only a small number of images is at hand for training. However, as this is the case with many real-life applications (due to limited memory or processing resources as, for example, in many mobile devices) researchers have directed a considerable research effort towards developing algorithms that require only a small number of training images and still achieve high recognition rates.

If we confine ourselves to the dominant face recognition techniques, i.e., appearance-based methods, two main research trends in respect to the lack of training data can be identified: *(i)* researchers try to improve the performance of the expressive approaches, such as principal component analysis (PCA)[3] or kernel principal component analysis (KPCA)[4], which are feasible regardless of the number of available training images but usually result in an insufficient recognition performance, and *(ii)* researchers try to modify the discriminative approaches, such as linear discriminant analysis (LDA)[5] or generalized discriminant analysis (GDA)[6] which commonly ensure high recognition rates, but require a sufficient number of training images to be applicable. The problem where the sample dimensionality is larger than the number of available training samples per subject is usually referred to as the small sample size (SSS) problem.

Several techniques were presented in the literature to cope with the lack of training data. Wu and Zhou [7], for example, tried to improve the performance of the PCA-based Eigenface technique and introduced a technique called

V. Štruc, R. Gajšek, F. Mihelič and N. Pavešić are all with the Faculty of Electrical Engineering, University of Ljubljana, Tržaška 25, SI-1000 Ljubljana, Slovenia (vitomir.struc, rok.gajsek, france.mihelic,nikola.pavesic)@fe.uni-lj.si

[1]The term biometrics refers to a scientific discipline which involves methods of automatically recognizing (verifying or identifying) people by their physical and/or behavioral characteristics.

$(PC)^2A$. In $(PC)^2A$ the face images were combined with their first-order vertical and horizontal projection images before being projected into the PCA sub-space with the goal of emphasizing the discriminative information contained in the facial images. The authors reported that the proposed method outperformed the traditionally employed PCA technique. Wang et al. [8] reported that good recognition performance under the lack of client-specific training data can be achieved when appearance-based methods are trained using a generic database. The authors evaluated the performance of several established appearance-based methods within this framework and achieved promising results.

Belhumeuer et al. [5] presented a LDA-based technique called the Fisherfaces, where LDA is performed in the PCA sub-space rather than the original pixel space, hence, the technique avoids the SSS problem with a sub-space projection preceding the discriminant analysis. Chen et al. [9] described a modification of the LDA approach tailored towards the SSS problem. The authors proposed to divide each face image from the training set into multiple non-overlapping images-blocks and then to employ these newly produced samples for training of the classical LDA technique. With this approach the training set is artificially enlarged, hence, LDA is applicable. Chen et al. reported that their recognition method outperformed the $(PC)^2A$ method in their experiments.

In this paper we propose a novel approach to handle the SSS problem. Rather than using discriminative techniques to achieve high recognition rates, we propose to use sub-space projection based regression techniques (which use expressive approaches and do not suffer from the SSS problem - they are always applicable) with an appropriately designed response matrix. These techniques are often used for classification purposes in the field of chemometrics, but have not yet been considered for the purposes of face recognition. As we will show in Section VI, where the experiments and their results are presented, regression techniques successfully cope with the SSS problem and simultaneously achieve recognition rates comparable to those achieved by established discriminative techniques.

The rest of the paper is organized as follows. In Section II the main concepts of two linear and two non-linear regression techniques are briefly reviewed. Section III presents the classification approach based on regression techniques, while the final classification rule is introduced in Section IV. In Section V the database and experimental protocol employed in our experiments are described. The experimental results are given and commented on in Section VI. The paper concludes with some final remarks in Section VII.

II. REGRESSION TECHNIQUES

This section presents the basic concepts of four sub-space projection based regression techniques, namely, principal component regression, partial least squares regression, kernel principal component regression and kernel partial least squares regression. While the former two techniques represent linear methods, the latter two represent non-linear or kernel regression approaches.

A. Linear regression techniques

Linear sub-space projection based regression techniques are comprised of two basic steps. In the first step the training data is projected into a linear sub-space, while in the second step multivariate regression is used to build the regression model. We will use the following notation throughout this paper:

- $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n]$ - a matrix containing in its columns the centered training pattern vectors, i.e., training facial images, from N classes,
- n - the number of pattern vectors in the training set,
- d - dimensionality of the pattern vectors,
- $\mathbf{Z}$ - the matrix of sub-space projection coefficients of the training pattern vectors in $\mathbf{X}$,
- $\mathbf{W}$ - the transformation matrix of the sub-space projection technique which defines the basis of the given sub-space,
- d' - dimensionality of the given sub-space and
- $\mathbf{Y}$ - the matrix containing the response variables (the response matrix).

Principal component regression (PCR) is a linear regression technique based on the popular sub-space projection (and/or feature extraction) approach called principal component analysis (PCA) and the classical multivariate regression technique. Similar to PCA, PCR first constructs the PCA transformation matrix $\mathbf{W}$ by means of the leading eigenvectors of the covariance matrix of the training data, i.e., $\Sigma = XX^T$, and uses the computed transformation matrix to project the training data in $\mathbf{X}$ into the principal component sub-space, i.e., $\mathbf{Z} = \mathbf{W}^T\mathbf{X}$. Here, the dimensionality of the sub-space is defined with the selected number d' of eigenvectors in $\mathbf{W}$, where $d' < n$. Next, PCR uses the principal component sub-space projections in $\mathbf{Z}$ to define a linear regression model, i.e., $\mathbf{Y} = \mathbf{ZB}$ with $\mathbf{B}$ being the response coefficient matrix computed as $\mathbf{B} = (\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{ZY}$.

Partial least squares regression (PLSR) computes the sub-space representation of the training data in $\mathbf{X}$ in form of latent vectors which account for as much of the covariance between the training pattern vectors in $\mathbf{X}$ and the responses in $\mathbf{Y}$ as possible. Thus, unlike PCR which considers only the variance present in the training data to construct a sub-space, PLSR computes latent vectors (which define the sub-space) from $\mathbf{X}$ which are relevant for predicting $\mathbf{Y}$. The regression model is commonly determined with the help of the non-linear iterative partial least squares (NIPALS) algorithm, which can be described as follows:

1. randomly initialize vector $\mathbf{u}$,
2. $\mathbf{w} = \mathbf{X}^T\mathbf{u}$,
3. $\mathbf{z} = \mathbf{Xw}, \mathbf{z} \leftarrow \mathbf{z}/\|\mathbf{z}\|$,
4. $\mathbf{c} = \mathbf{Y}^T\mathbf{z}$
5. $\mathbf{u} = \mathbf{Yc}, \mathbf{u} \leftarrow \mathbf{u}/\|\mathbf{u}\|$,
6. repeat steps 2.-5. until convergence

7. $\mathbf{X} \leftarrow \mathbf{X} - \mathbf{z}\mathbf{z}^T\mathbf{X}, \mathbf{Y} \leftarrow \mathbf{Y} - \mathbf{z}\mathbf{z}^T\mathbf{Y}$,
8. continue with step 2. using the new matrices $\mathbf{X}$ and $\mathbf{Y}$.

Here $\mathbf{w}$ and $\mathbf{u}$ denote the latent vectors that represent columns of the matrices $\mathbf{W}$ and $\mathbf{U}$, $\mathbf{c}$ and $\mathbf{z}$ stand for weight vectors comprising the matrices $\mathbf{C}$ and $\mathbf{Z}$, and the regression model $\mathbf{Y} = \mathbf{XB}$ is defined by the the regression coefficient matrix $\mathbf{B} = \mathbf{X}^T\mathbf{U}(\mathbf{Z}^T\mathbf{X}\mathbf{X}^T\mathbf{U})^{-1}\mathbf{Z}^T\mathbf{Y}$. Note that the NIPALS algorithm is repeated until the desired (or the maximum) number of latent vectors is computed. A more detailed description of the technique can be found in [10].

B. Non-linear regression techniques

Similar to their linear counterparts the non-linear regression techniques also use a two stage approach for building regression models. They construct a linear regression model (similar to the one presented in the paragraph on PCR) in a high-dimensional feature space $\mathcal{F}$ to which the d-dimensional input pattern vectors $\mathbf{x}_i$ were non-linearly mapped, i.e., $\Phi : \mathbf{x} \in \mathbb{R}^d \rightarrow \Phi(\mathbf{x}) \in \mathcal{F}$, which corresponds to a non-linear regression model in the original input space. Kernel methods commonly avoid direct calculation of the computationally expensive non-linear mapping Φ, but rather make use of the so-called *kernel-trick* which uses kernel matrices of the training data to achieve non-linear regression. Again we introduce the notation for the kernel regression techniques used in this paper:

- $\mathbf{K} = \left[\Phi(\mathbf{x}_i)\Phi(\mathbf{x}_j)^T\right] = \left[K(\mathbf{x}_i, \mathbf{x}_j)\right]$; $\forall i, j$ - the kernel matrix of the training data,
- $K(\mathbf{x}_i, \mathbf{x}_j)$ - a kernel function, e.g., $K(\mathbf{x}_i, \mathbf{x}_j) = e^{(\|\mathbf{x}_i - \mathbf{x}_j\|^2/2\sigma^2)}$ or $K(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i^T\mathbf{x}_j)^p$,
- $\mathbf{K}_c$ - the centered kernel matrix of the training data computed as: $\mathbf{K}_c = (\mathbf{I} - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T)\mathbf{K}(\mathbf{I} - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T)$, where $\mathbf{I}$ represents an n-dimensional identity matrix and $\mathbf{1}_n$ stands for a vector of all ones, with length n.
- $\boldsymbol{\alpha}$ - the transformation matrix of the given kernel method.

Kernel principal component regression (KPCR) is a non-linear regression technique based on the non-linear form of PCA called kernel principal component analysis (KPCA)[4] and the classical multivariate regression technique. The authors of [4] show that the projection matrix of KPCA $\boldsymbol{\alpha}$ can be computed by solving the following eigenproblem:

$$\mathbf{K}_c\boldsymbol{\alpha} = \boldsymbol{\Lambda}\boldsymbol{\alpha}, \tag{1}$$

where $\boldsymbol{\alpha} \in \mathbb{R}^{n\times n}$ is an orthogonal eigenvector matrix and $\boldsymbol{\Lambda} \in \mathbb{R}^{n\times n}$ is a diagonal eigenvalue matrix. Once the transformation matrix $\boldsymbol{\alpha}$ is computed it used to project the data in the kernel matrix into the non-linear sub-space, i.e., $\mathbf{Z} = \boldsymbol{\alpha}^T\mathbf{K}_c$. The projection coefficients in $\mathbf{Z}$ are then used to construct the regression model $\mathbf{Y} = \mathbf{ZB}$ with $\mathbf{B}$ being the response coefficient matrix computed as $\mathbf{B} = (\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{ZY}$.

Kernel partial least squares regression (KPLSR) is a non-linear extension of the PLSR technique. Similar to the linear case, KPLSR again searches for latent vectors which account for most of the covariance between the predictors and the responses, however, it does so in the high-dimensional feature space $\mathcal{F}$ with the help of kernel matrices. The KPLSR technique is based on the following modification of the NIPALS algorithm:

1. randomly initialize the vector $\mathbf{u}$
2. $\mathbf{z} = \mathbf{K}_c\mathbf{K}_c^T\mathbf{u}$, $\mathbf{z} \leftarrow \mathbf{z}/\|\mathbf{z}\|$
3. $\mathbf{u} = \mathbf{K}_c\mathbf{K}_c^T\mathbf{z}$, $\mathbf{u} \leftarrow \mathbf{u}/\|\mathbf{u}\|$
4. repeat steps 2 - 3 until convergence
5. $\mathbf{K}_c = \mathbf{K}_c - \mathbf{t}\mathbf{t}^T\mathbf{K}_c$, $\mathbf{Y} = \mathbf{Y} - \mathbf{t}\mathbf{t}^T\mathbf{Y}$
6. continue with step 2. using the new matrices $\mathbf{K}_c$ and $\mathbf{Y}$.

Again, the NIPALS algorithm is repeated until an appropriate number of latent vectors is found. The final regression model is then constructed as $\mathbf{Y} = \mathbf{K}_c\mathbf{A}$, where the matrix of regression coefficients $\mathbf{A}$ equals $\mathbf{A} = \mathbf{K}_c^T\mathbf{U}(\mathbf{Z}^T\mathbf{K}_c\mathbf{K}_c^T\mathbf{U})^{-1}\mathbf{Z}^T\mathbf{Y}$.

III. REGRESSION TECHNIQUES AND CLASSIFICATION

As the name suggests, regression techniques are usually employed for building regression models which relate a number of predictor variables to a number of response variables. However, they can also be used for classification purposes if the matrix $\mathbf{Y}$ containing the response variables is constructed in such a way that it encodes the class-membership of the predictors. Thus, the following response matrix must be employed for construction of the regression model:

$$\mathbf{Y} = \begin{bmatrix} \mathbf{1}_{m_1} & \mathbf{0}_{m_1} & \cdots & \mathbf{0}_{m_1} \\ \mathbf{0}_{m_2} & \mathbf{1}_{m_2} & \cdots & \mathbf{0}_{m_2} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{m_N} & \mathbf{0}_{m_N} & \cdots & \mathbf{1}_{m_N} \end{bmatrix}, \tag{2}$$

where N represents the number of classes in the set of n d-dimensional inputs (matrix $\mathbf{X}$), m_i represents the number of inputs in class I_i, $\mathbf{1}_{m_i}$ $(i = 1, 2, ..., N)$ denotes a $m_i \times 1$ vector of all ones and $\mathbf{0}_{m_i}$ $(i = 1, 2, ..., N)$ is a $m_i \times 1$ vector of all zeros. Clearly, when using regression techniques for classification, the user-template $\bar{\mathbf{y}}_i$ for the identity I_i is constructed by averaging the responses corresponding to the predictors, i.e., training images, of the i-th identity.

IV. THE CLASSIFICATION RULE

The recognition performance of the regression techniques will be assessed within a face recognition system operating in identification mode. The problem statement for such systems can according to [2] be defined as follows: given an input feature vector $\mathbf{y}$ extracted from the biometric data of the person currently presented to the system, determine the identity I_k, $k \in \{1, 2, ..., N, N+1\}$. Here $I_1, I_2, ..., I_N$ are the identities enrolled in the system and I_{N+1} indicates the reject case where no suitable identity can be determined for the person presented to the system [1]. Hence

$$\mathbf{y} \in \begin{cases} I_k, & \text{if } \max_i\{\delta(\mathbf{y}, \bar{\mathbf{y}}_i)\} \geq t,\ k = 1, 2, ..., N \\ I_{N+1}, & \text{otherwise} \end{cases}.$$

Here δ denotes a function that measures the similarity between the feature vector $\mathbf{y}$ and the user-template $\overline{\mathbf{y}}_i$ which corresponds to identity I_k, t stands for a pre-defined threshold. We can see that the success of the identification procedure heavily depends on the employed similarity measure δ. In this paper the commonly used cosine similarity measure is used as the scoring function, i.e.,

$$\delta_{cos}(\mathbf{y}, \overline{\mathbf{y}}_i) = \frac{\mathbf{y} \cdot \overline{\mathbf{y}}_i}{|\mathbf{y}||\overline{\mathbf{y}}_i|}, \quad (3)$$

where $\cdot$ denotes the dot product.

Using the function δ_{cos} the system searches for the user-template that results in the highest matching score among all user-templates and consequently assigns the identity corresponding to that user-template to the feature vector $\mathbf{y}$.

V. THE DATABASE AND EXPERIMENTAL PROTOCOL

The effectiveness of regression techniques for face recognition was assessed on the publicly available ORL database acquired at the Olliveti Research Laboratory in Cambridge, U.K. [11]. The database contains grey-scale images of 40 distinct subjects with each subject being represented with 10 facial images. Thus, a total of 400 images is available for training and testing of a given recognition technique. The images of the database are stored at a resolution of 112×92 and archived in the portable grey map format.

To ensure that the face recognition performance of the tested techniques is not affected by factors not related to the face, e.g., hair style or background, a pre-processing procedure was applied to all the images from the database. The procedure first aligned, i.e., rotated and scaled, the images in such a way that the eye-centers were located at predefined position[2], then cropped the face-region to a fixed size of 64×64 pixels and finally used a photometric normalization technique to reduce the impact of the lighting conditions present during the image-acquisition stage on the appearance of the images. Some examples of the pre-processed images form the ORL database are shown in Fig. 1.

Fig. 1. Examples of the pre-processed images from the ORL database

For experimental purposes the 10 images of each subject in the database were randomly partitioned into two non-overlapping image-groups: the first group was used for training, while the second group was employed for testing.

[2]Note that the coordinates of the eye-centers were located manually.

For assessing the performance of the recognition techniques five sets of experiments were performed. In the first set one images of each subject was chosen to serve as the training image, while the remaining images were used for testing. In the second set, the number of randomly selected training images was increased to two, in the third to three, in the fourth to four and in the last set the number of training images was set to five. Of course, the left over images, i.e., not used for training, from each set of experiments were employed for the performance assessment. All the experiments were repeated five times (with five different partitions of the facial images into the training and test groups), hence, the recognition results presented in the following section are presented in terms of the average rank-one recognition rates. Here, the rank-one recognition rate (RORR) denotes the percentage of correctly recognized images, while the average rank-one recognition rate represents the mean value of the RORR over several repetitions of the experiments.

VI. EXPERIMENTS

As already indicated in the previous section, our experiments aimed at assessing the recognition performance of the (linear and non-linear) regression techniques PCR, PLSR, KPCR and KPLSR and compare their performance to that of some established discriminative recognition techniques. To that end, two discriminative approaches were implemented, trained (using the protocol described in Section V) and tested on the ORL database, namely, the linear Fisherface (or linear discriminant analysis - LDA) approach [5] and the non-linear Generalized discriminant analysis (GDA) approach [6]. The results of the experiments are presented in Fig. 2 and Table I - here the symbol N/A denotes that the method is not applicable considering the available number of training images.

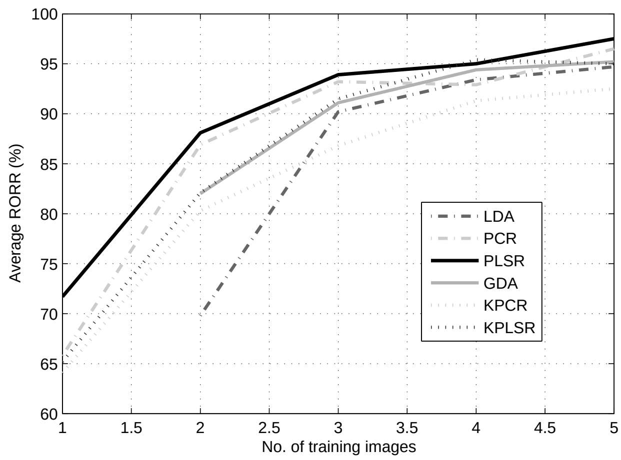

Fig. 2. Average RORRs in respect to different numbers of training images

Amongst the linear methods the PLSR approach performed the best for all numbers of training images, followed in order by the PCR and LDA techniques. Among the non-linear methods the best recognition rates were still achieved

TABLE I

AVERAGE RANK-ONE RECOGNITION RATES IN % FOR THE IDENTIFICATION EXPERIMENTS PERFORMED ON THE ORL DATABASE

No. of training samples	LDA	PCR	PSLR	GDA	KPCR	KPSLR
1	N/A	65.8	71.7	N/A	64.1	65.2
2	69.8	86.9	88.1	82.0	80.3	82.1
3	90.2	93.2	93.9	91.1	86.8	91.5
4	93.4	92.9	95.0	94.4	91.3	95.4
5	94.7	96.5	97.5	95.2	92.5	95.0

by the non-linear form of the PLSR technique, however, the difference to the GDA and KPCR methods has decreased. Interestingly, while the GDA performed better than its linear counterpart, the non-linear versions of the regression techniques performed worse than the linear ones. Thus, the best recognition rate overall was observed with the PLSR approach.

The results clearly show the potential of the regression techniques for face recognition. Not only that they are applicable in the most severe case of the SSS, i.e., when only one face image per subject is available for training, they also achieve recognition rates comparable to those of the discriminative techniques.

VII. CONCLUSION

In this paper four sub-space projection-based regression techniques were introduced to the field of face recognition and tested for their face-recognition accuracy. Their performance was assessed on the ORL database and subsequently compared to that of two established discriminative feature-extraction techniques, i.e., linear and generalized discriminant analysis. Our experimental results show that regression technique successfully cope with the small sample size problem of face recognition while simultaneously achieving recognition rates comparable to those of the more established discriminative feature-extraction approaches.

VIII. ACKNOWLEDGMENTS

This work was supported in part by the European Commission under contract FP7-217762 HIDE - Homeland security, biometric Identification and personal Detection Ethics, the national research program P2-0250(C) Metrology and Biometric Systems, the bilateral project with the Bulgarian Academy of Sciences - Face and Signature Biometrics, the bilateral project with the People's Republic of China Bi-CN/07-09-019 and the national project AvID M2-0210.

REFERENCES

[1] V. Štruc and N. Pavešić, "The Corrected Normalized Correlation Coefficient: A Novel Way Of Matching Score Calculation for LDA-Based Face Verification", *To appear in the Proceedings of the 5th International Conference on Fuzzy Systems and Knowledge Discovery*, Jinan, China, 2008.

[2] A.K. Jain, A. Ross and S. Prabhakar, An Introduction to Biometric Recognition, *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 1, 2004, pp. 4–20.

[3] M. Turk and A. Pentland, Eigenfaces for Recognition. *Journal of Cognitive Neurosicence*, vol. 3, no. 1, 1991, pp. 71–86.

[4] B. Schölkopf, A.J. Smola and K.R. Müller, Nonlinear component analysis as a kernel eigenvalue problem, *Neural Computation*, vol. 10, 1998, pp. 1299–1319.

[5] P.N. Belhumeur, J.P. Hespanha and D.J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection", *Proceedings of the 4th ECCV*, Cambridge, UK, 1996, pp.45–58.

[6] G. Baudat and F.E. Anouar, Generalized Discriminant Analysis Using a Kernel Approach, *Neural Copmutation*, vol. 12, no. 10, 2000, pp. 2385–2404.

[7] J. Wu and Z.H. Zhou, Face recognition with one training image per person. *Pattern Recogntion Letters*, vol. 23, no. 14, 2002, pp. 1711–1719.

[8] J. Wang, K.N. Plataniotis, J. Lu and A.N. Venetsanopoulos, On solving the FR problem with one training sample per subject. *Patteern Recognition*, vol. 39, no. 9, 2006, pp. 1746–1762.

[9] S. Chen, J. Liu and Z.H. Zhou, Making FLDA apllicable to face recognition with one sample per person. *Pattern recognition*, vol. 37, no. 7, 2004, pp. 1553–1555.

[10] R. Rosipal and L.J. Trejo, Kernel partial least squares regression in reproducing kernel Hilbert space. *Journal of Machine Learning Research*, vol. 2, 2001, pp. 97–123.

[11] F. Samaria and A. Harter, "Parameterisation of a Stochastic Model for Human Face Identification", *2nd IEEE Workshop on Applications of Computer Vision*, Sarasota (Florida), 1994, pp. 138–142.