A NOVEL KERNEL DISCRIMINANT ANALYSIS FOR FACE VERIFICATION

Georgios Goudelis, Stefanos Zafeiriou, Anastasios Tefas and Ioannis Pitas

Department of Informatics, Aristotle university of Thessaloniki, Greece

ABSTRACT

In this paper a novel non-linear subspace method for face verification is proposed. The problem of face verification is considered as a two-class problem (genuine versus impostor class). The typical *Fisher's Linear Discriminant Analysis* (FLDA) gives only one or two projections in a two-class problem. This is a very strict limitation to the search of discriminant dimensions. As for the FLDA for N class problems (N is greater than two) the transformation is not person specific. In order to remedy these limitations of FLDA, exploit the individuality of human faces and take into consideration the fact that the distribution of facial images, under different viewpoints, illumination variations and facial expression is highly complex and non-linear, novel kernel discriminant algorithms are proposed. The new methods are tested in the face verification problem using the XM2VTS database where it is verified that they outperform other commonly used kernel approaches.

Index Terms— Kernel techniques, Face Verification, Fisher's Linear Discriminant Analysis

1. INTRODUCTION

In this paper, face verification is modelled as a two-class problem. The motivations of such a modelling are supported by various methods that take into account the individuality of facial features [6, 1, 2]. The use of person-specific graphs with nodes placed at discriminant facial landmarks greatly improves the performance of elastic graph matching in frontal face verification [6]. In [1] it has been shown that discriminant non-negative matrix factorization methods with class-specific bases perform better than other approaches with common bases. Additional details about modeling face verification as a two-class problem are given in [2] introducing the class-specific Fisherfaces.

The methods proposed in this paper exploit the individuality of the human face in order to find a nonlinear subspace representation with enhanced discriminant power. In detail, in this paper we propose a novel class-specific discriminant criterion, which when optimized, it leads to a discriminant low dimensional representation of faces. Furthermore, in order to represent better the face in various poses, we combine the proposed criterion with kernel techniques and we present a technique for optimizing the criterion in arbitrary dimensional Hilbert spaces leading to a novel Kernel Discriminant Analysis. However, the main contribution of the proposed discriminant analysis, is that it tries to remedy some of the limitations of the kernel methods based on the Fisher's discriminant criterion that provide very limited number of features in two-class problems. For example the so-called Complete Kernel Fisher Discriminant Analysis (CKFDA) [3] only two discriminant dimensions are found in two-class problems. This space of very limited number of dimensions may be proved to be insufficient for correctly representing facial images. The proposed approach discovers a low dimensional space with the number of dimensions to be proportional to the number of images available for training. Experiments conducted in the XM2VTS database using facial images at various poses demonstrate the potential of the proposed methods.

This work has been funded by the network of excellence BioSecure IST-2002-507634 (Biometrics for Secure Authentication, http://www.biosecure.info), under Information Society Technologies (IST) priority of the 6th Framework Programme of the European Community

2. DISCRIMINANT CRITERION

Before we develop the new optimization problem, we will introduce some notation that is used throughout this paper. Let r be the reference person that will be used for defining the person specific algorithms. Let $\mathcal{U}_r$ be the class of genuine vectors and $\mathcal{I}_r$ be the class of impostor vector. Let L_G, and L_I be the numbers of genuine and impostor images in the training set for the person r, respectively. Usually, the number of genuine images is much smaller than the number of impostor images for a reference person r. Thus, in the following analysis we will work under the assumption that $L_I > L_G$. Let $L = L_G + L_I$ be the total number of images in the training database.

The genuine vectors $\mathbf{z}_i$ of the person r will be denoted as $\boldsymbol{\rho}_i = \mathbf{z}_i$ ($\mathbf{z}_i \in \mathcal{U}_r$), while the impostor images $\mathbf{z}_i$ of the person r will be denoted as $\boldsymbol{\kappa}_i = \mathbf{z}_i$ ($\mathbf{z}_i \in \mathcal{I}_r$). Let also $\bar{\boldsymbol{\rho}} = \frac{1}{L_G}\sum_{i=1}^{L_G} \phi(\boldsymbol{\rho}_i)$, $\bar{\boldsymbol{\kappa}} = \frac{1}{L_I}\sum_{i=1}^{L_I} \phi(\boldsymbol{\kappa}_i)$ and $\bar{\mathbf{m}} = \frac{1}{L}\sum_{i=1}^{L} \phi(\mathbf{z}_i)$ be the mean vectors of the genuine class, the impostor class and total mean of the facial vectors in the Hilbert space $\mathcal{F}$. Any function k satisfying the Mercer's condition can be used as a kernel. The dot product of $\phi(\mathbf{z}_i)$ and $\phi(\mathbf{z}_j)$ in the Hilbert space can be calculated without having to evaluate explicitly the mapping $\phi(\cdot)$ as $k(\mathbf{z}_i, \mathbf{z}_j) = \phi(\mathbf{z}_i)^T \phi(\mathbf{z}_j)$

(this is also known as the kernel trick [4]. The kernels that have been used in our experiments have been the polynomial kernels $k(\mathbf{x},\mathbf{y}) = \phi(\mathbf{x})^T\phi(\mathbf{y}) = (\mathbf{x}^T\mathbf{y}+1)^d$ where d is the degree of the polynomial.

2.1. The Novel Kernel Criterion

The criterion that is used in this paper, will be formed using a simple similarity measure in the Hilbert space $\mathcal{F}$. This measure quantifies the similarity of a given feature vector $\mathbf{z}$ to the reference facial class r in the subspace spanned by the columns of the matrix $\boldsymbol{\Psi} = [\boldsymbol{\psi}_1 \dots \boldsymbol{\psi}_K]$, with $\boldsymbol{\psi}_i \in \mathcal{F}$. The L_2 norm in the reduced space spanned by the columns of $\boldsymbol{\Psi}$, is used as similarity measure:

$$\begin{aligned} d_r(\mathbf{z}) &= ||\boldsymbol{\Psi}^T(\phi(\mathbf{z}) - \bar{\boldsymbol{\rho}})||^2 \\ &= \boldsymbol{\Psi}^T(\phi(\mathbf{z}) - \bar{\boldsymbol{\rho}})(\phi(\mathbf{z}) - \bar{\boldsymbol{\rho}})^T\boldsymbol{\Psi} \\ &= \sum_{i=1}^{K} \boldsymbol{\psi}_i^T(\phi(\mathbf{z}) - \bar{\boldsymbol{\rho}})(\phi(\mathbf{z}) - \bar{\boldsymbol{\rho}})^T\boldsymbol{\psi}_i \end{aligned} \quad (1)$$

which is actually the Euclidean distance of a projected sample to the projected mean of the reference class and is one of most usually employed measures in pattern recognition applications (i.e, distance from the center of the class). This distance should be low for the samples of the genuine class and should be high for the samples of the impostor class.

Now, in order to find a discriminant linear transformation in $\mathcal{F}$ we demand that the sum of the similarity measures $d_r(\mathbf{z})$ for all $\mathbf{z} \in \mathcal{I}_r$ (impostor similarity measures) to be maximized while minimizing the sum of the similarity measures $d_r(\mathbf{z})$ for all $\mathbf{z} \in \mathcal{U}_r$ (client similarity measures). Thus, the discriminant projections $\boldsymbol{\psi}_i \in \mathcal{F}$ are found in the training set as the ones that maximize the ratio:

$$\begin{aligned} D^{\Phi}(\boldsymbol{\Psi}) &= \frac{\sum_{\mathbf{z}\in\mathcal{I}_r} d_r(\mathbf{z})}{\sum_{\mathbf{z}\in\mathcal{U}_r} d_r(\mathbf{z})} \\ &= \frac{\sum_{\mathbf{z}\in\mathcal{I}_r}\sum_{i=1}^{K} \boldsymbol{\psi}_i^T(\phi(\mathbf{z})-\bar{\boldsymbol{\rho}})(\phi(\mathbf{z})-\bar{\boldsymbol{\rho}})^T\boldsymbol{\psi}_i}{\sum_{\mathbf{z}\in\mathcal{U}_r}\sum_{i=1}^{K} \boldsymbol{\psi}_i^T(\phi(\mathbf{z})-\bar{\boldsymbol{\rho}})(\phi(\mathbf{z})-\bar{\boldsymbol{\rho}})^T\boldsymbol{\psi}_i} \\ &= \frac{\mathrm{tr}[\boldsymbol{\Psi}^T\mathbf{W}^{\Phi}\boldsymbol{\Psi}]}{\mathrm{tr}[\boldsymbol{\Psi}^T\mathbf{B}^{\Phi}\boldsymbol{\Psi}]}. \end{aligned} \quad (2)$$

where $\mathbf{W}^{\Phi} = \sum_{\mathbf{z}\in\mathcal{I}_r}(\phi(\mathbf{z}) - \bar{\boldsymbol{\rho}})(\phi(\mathbf{z}) - \bar{\boldsymbol{\rho}})^T$, $\mathbf{B}^{\Phi} = \sum_{\mathbf{z}\in\mathcal{U}_r}(\phi(\mathbf{z}) - \bar{\boldsymbol{\rho}})(\phi(\mathbf{z}) - \bar{\boldsymbol{\rho}})^T$ and tr[$\mathbf{M}$] is the trace of matrix $\mathbf{M}$.

2.2. Two Step Optimization method for the Discriminant Criterion

In the Hilbert space $\mathcal{F}$ it is almost impossible to make $\mathbf{B}^{\Phi}$ invertible (the matrix $\mathbf{B}^{\Phi}$ is invertible if the dimension of the feature vectors is smaller than the number of the genuine images). Thus, vectors $\boldsymbol{\psi}_i$ such that $\boldsymbol{\psi}_i^T\mathbf{B}^{\Phi}\boldsymbol{\psi}_i = 0$ always exist. These vectors are very effective for discrimination if they satisfy $\boldsymbol{\psi}_i^T\mathbf{W}^{\Phi}\boldsymbol{\psi}_i > 0$ at the same time, since for these vectors it is valid that $D^{\Phi}(\boldsymbol{\Psi}) \to +\infty$. In such a case, the criterion (2) degenerates into the following:

$$D_b^{\Phi}(\boldsymbol{\Psi}) = \mathrm{tr}[\boldsymbol{\Psi}^T\mathbf{W}^{\Phi}\boldsymbol{\Psi}] \ (\boldsymbol{\Psi} = [\dots\boldsymbol{\psi}_i\dots],\ ||\boldsymbol{\psi}_i|| = 1). \quad (3)$$

Using the criteria D_b^{Φ} and D^{Φ} two kind of discriminant features can be calculated. We will call the discriminant projections of the criterion D^{Φ} as regular while the ones of the criterion D_b^{Φ} will be called irregular.

2.2.1. Reducing $\mathcal{F}$

The first step is to reduce the Hilbert space $\mathcal{F}$ by using a linear mapping without discarding any discriminant information. This mapping is comprised of the non-null eigenvectors of $\mathbf{S}^{\Phi} = \mathbf{W}^{\Phi} + \mathbf{B}^{\Phi}$. The non-null eigenvectors of $\mathbf{S}^{\Phi}$ can be calculated using the kernel matrices:

$$\begin{aligned} [\mathbf{K}_1]_{i,j} &= \phi(\boldsymbol{\rho}_i)^T\phi(\boldsymbol{\rho}_j) = k(\boldsymbol{\rho}_i, \boldsymbol{\rho}_j) \\ [\mathbf{K}_2]_{i,j} &= \phi(\boldsymbol{\kappa}_i)^T\phi(\boldsymbol{\rho}_j) = k(\boldsymbol{\kappa}_i, \boldsymbol{\rho}_j) \\ [\mathbf{K}_3]_{i,j} &= \phi(\boldsymbol{\rho}_i)^T\phi(\boldsymbol{\kappa}_j) = k(\boldsymbol{\rho}_i, \boldsymbol{\kappa}_j) = \mathbf{K}_2^T \\ [\mathbf{K}_4]_{i,j} &= \phi(\boldsymbol{\kappa}_i)^T\phi(\boldsymbol{\kappa}_j) = k(\boldsymbol{\kappa}_i, \boldsymbol{\kappa}_j). \end{aligned} \quad (4)$$

The kernel matrix $\mathbf{K}$ is the total kernel function defined as:

$$\mathbf{K} = \begin{bmatrix} \mathbf{K}_4 & \mathbf{K}_2 \\ \mathbf{K}_3 & \mathbf{K}_1 \end{bmatrix} \quad (5)$$

and $\mathbf{E}$ is defined as:

$$\mathbf{E} = \begin{bmatrix} \mathbf{K}_2 \\ \mathbf{K}_3 \end{bmatrix}. \quad (6)$$

First $\mathbf{S}^{\Phi}$ can be written as:

$$\begin{aligned} \mathbf{S}^{\Phi} &= \mathbf{W}^{\Phi} + \mathbf{B}^{\Phi} \\ &= \sum_{\mathbf{z}\in\mathcal{I}_r}(\phi(\mathbf{z}) - \bar{\boldsymbol{\rho}})(\phi(\mathbf{z}) - \bar{\boldsymbol{\rho}})^T \\ &+ \sum_{\mathbf{z}\in\mathcal{U}_r}(\phi(\mathbf{z}) - \bar{\boldsymbol{\rho}})(\phi(\mathbf{z}) - \bar{\boldsymbol{\rho}})^T \\ &= \sum_{\mathbf{z}\in\mathcal{U}}(\phi(\mathbf{z}) - \bar{\boldsymbol{\rho}})(\phi(\mathbf{z}) - \bar{\boldsymbol{\rho}})^T = \sum_{i=1}^{L} \tilde{\boldsymbol{\mu}}_i\tilde{\boldsymbol{\mu}}_j^T \\ &= \boldsymbol{\Phi}_s\boldsymbol{\Phi}_s^T \end{aligned} \quad (7)$$

where $\tilde{\boldsymbol{\mu}}_i = \phi(\mathbf{z}_i) - \bar{\boldsymbol{\rho}}$ and $\boldsymbol{\Phi}_s = [\tilde{\boldsymbol{\mu}}_1 \dots \tilde{\boldsymbol{\mu}}_L]$. Only the first n (with $n \leq L-1$) positive eigenvalues of $\mathbf{S}^{\Phi}$ are of interest to us. These eigenvectors can be indirectly derived from the eigenvectors of the matrix $\boldsymbol{\Phi}_s^T\boldsymbol{\Phi}_s$ ($L \times L$).

The $\boldsymbol{\Phi}_s^T\boldsymbol{\Phi}_s$ can be expanded as:

$$\boldsymbol{\Phi}_s^T\boldsymbol{\Phi}_s = \mathbf{K} - \frac{1}{L_G}\mathbf{E}\mathbf{1}_{L_G L} - \frac{1}{L_G}\mathbf{1}_{L L_G}\mathbf{E} + \frac{1}{N_G^2}\mathbf{1}_{L L_G}\mathbf{K}_1\mathbf{1}_{L_G L}. \quad (8)$$

where $\mathbf{1}_{LL_G}$ is a $L \times L_G$ matrix with elements all equal to one. Let λ_i^s and $\mathbf{c}_i (i = 1 \dots L_I)$ be the i-th eigenvalue and the corresponding eigenvector of $\boldsymbol{\Phi}_s^T\boldsymbol{\Phi}_s$, sorted in ascending order of eigenvalues. It's true that $(\boldsymbol{\Phi}_s\boldsymbol{\Phi}_s^T)(\boldsymbol{\Phi}_s\boldsymbol{\varpi}_i) = \lambda_i^s(\boldsymbol{\Phi}_s\mathbf{c}_i)$. Thus, $\boldsymbol{\varpi}_i = \boldsymbol{\Phi}_s\mathbf{c}_i$ are the eigenvectors of $\mathbf{S}^{\Phi}$. In order to remove the null space of $\mathbf{S}^{\Phi}$, the first $n \leq L_I - 1$ eigenvectors (given in the matrix $\boldsymbol{\Pi} = [\boldsymbol{\varpi}_1 \dots \boldsymbol{\varpi}_n] = \boldsymbol{\Phi}_s\mathbf{C}$, where $\mathbf{C} = [\mathbf{c}_1 \dots \mathbf{c}_n]$), whose corresponding eigenvalues are non zero, should be calculated. Thus, $\boldsymbol{\Pi}^T\mathbf{S}^{\Phi}\boldsymbol{\Pi} = \boldsymbol{\Lambda}_s$, with $\boldsymbol{\Lambda}_s = \mathrm{diag}[{\lambda_1^s}^2 \dots {\lambda_n^s}^2]$, a $n \times n$ diagonal matrix. The orthonormal eigenvectors of $\mathbf{S}^{\Phi}$ are the columns of the matrix:

$$\boldsymbol{\Pi}_1 = \boldsymbol{\Phi}_s\boldsymbol{\Pi}\boldsymbol{\Lambda}_{\mathbf{s}}^{-1/2}. \quad (9)$$

It can be easily proven that $\mathbf{S}^\Phi$ is compact and self-adjoint and thus the columns of the matrix $\mathbf{\Pi}_1$ form an orthonormal basis in $\mathcal{F}$. We define the two orthogonal complementary subspaces $\mathcal{O}$ and $\mathcal{O}^\perp$ of $\mathcal{F}$ ($\mathcal{F} = \mathcal{O} \oplus \mathcal{O}^\perp$). $\mathcal{O}$ is spanned by the column vectors of $\mathbf{\Pi}_1$. Its orthogonal $\mathcal{O}^\perp$ is the one that corresponds to the null space of $\mathbf{S}^\Phi$. We can now easily prove that there is no discriminant information in $\mathcal{O}^\perp$ in respect to the criterions $\mathbf{D}^\Phi$ and $\mathbf{D}_b^\Phi$, since for the vectors $\boldsymbol{\zeta} \in \mathcal{O}^\perp$ it is valid that $\boldsymbol{\zeta}^T \mathbf{B}^\Phi \boldsymbol{\zeta} = 0$ and $\boldsymbol{\zeta}^T \mathbf{W}^\Phi \boldsymbol{\zeta} = 0$ at the same time. Thus, all the discriminant information lie inside $\mathcal{O}$.

Now, based on the previous remarks, the two alternative discriminant criterions can be defined as:

$$D(\mathbf{H}) = \frac{\text{tr}[\mathbf{H}^T \mathbf{W} \mathbf{H}]}{\text{tr}[\mathbf{H}^T \mathbf{B} \mathbf{H}]} \tag{10}$$

and

$$D_b(\mathbf{H}) = \text{tr}[\mathbf{H}^T \mathbf{W} \mathbf{H}] \;\; (||\boldsymbol{\eta}_i|| = 1 \text{ and } \boldsymbol{\eta}_i^T \mathbf{B} \boldsymbol{\eta}_i = 0) \tag{11}$$

where $\mathbf{W} = \mathbf{\Pi}_1^T \mathbf{W}^\Phi \mathbf{\Pi}_1$, $\mathbf{B} = \mathbf{\Pi}_1^T \mathbf{B}^\Phi \mathbf{\Pi}_1$ and $\mathbf{H} = [\ldots \ \boldsymbol{\eta}_i \ldots]$ with $\boldsymbol{\eta}_i \in \Re^n$.

2.2.2. Feature Extraction

Let $\mathbf{\Xi} = [\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_n]$ be all the eigenvectors of $\mathbf{B}$. The first $q = L_G - 1$ eigenvectors correspond to the nonzero eigenvalues (range space). The two orthogonal complementary subspaces of $\mathbf{B}$ are defined as $\mathcal{O}_B = \text{span}\{\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_q\}$ and $\mathcal{O}_B^\perp = \text{span}\{\boldsymbol{\xi}_{q+1}, \ldots, \boldsymbol{\xi}_n\}$. Thus, $\Re^n = \mathcal{O}_B^\perp \oplus \mathcal{O}_B$. In the space $\mathcal{O}_B$ we seek for the regular discriminant projections, while in the space $\mathcal{O}_B^\perp$ we seek for the irregular discriminant projections. We can now summarize the previous procedure for learning the class-specific discriminant transform. For each client r, the following steps should be applied:

Step 1 . Calculate the eigenvalues and the eigenvectors of $\mathbf{\Phi}_s^T \mathbf{\Phi}_s$ and project each facial vector $\mathbf{z}_i \in \mathcal{U}$ as:

$$\begin{aligned}
\mathbf{\Pi}_1^T \phi(\mathbf{z}_i) &= (\mathbf{\Pi} \mathbf{\Lambda_s}^{-1/2})^T \mathbf{\Phi}_s^T \phi(\mathbf{z}_i) \\
&= (\mathbf{\Pi} \mathbf{\Lambda_s}^{-1/2})^T [\tilde{\boldsymbol{\mu}}_1 \ldots \tilde{\boldsymbol{\mu}}_{L_I}]^T \phi(\mathbf{z}_i) \\
&= (\mathbf{\Pi} \mathbf{\Lambda_s}^{-1/2})^T ([\phi(\mathbf{z}_1) \ldots \phi(\mathbf{z}_L)]^T \phi(\mathbf{z}_i) \\
&\quad -[\bar{\boldsymbol{\rho}} \ldots \bar{\boldsymbol{\rho}}]^T \phi(\mathbf{z}_i)) \\
&= (\mathbf{\Pi} \mathbf{\Lambda_s}^{-1/2})^T ([\phi(\mathbf{z}_1) \ldots \phi(\mathbf{z}_L)]^T \phi(\mathbf{z}_i) \\
&\quad -\tfrac{1}{L_G} \mathbf{1}_{L L_G} [\phi(\boldsymbol{\rho}_1) \ldots \phi(\boldsymbol{\rho}_{L_G})]^T \phi(\mathbf{z}_i)) \\
&= (\mathbf{\Pi} \mathbf{\Lambda_s}^{-1/2})^T ([k(\mathbf{z}_1, \mathbf{z}_i) \ldots k(\mathbf{z}_L, \mathbf{z}_i)]^T \\
&\quad -\tfrac{1}{L_G} \mathbf{1}_{L L_G} [k(\boldsymbol{\rho}_1, \mathbf{z}_i) \ldots k(\boldsymbol{\rho}_{L_G}, \mathbf{z}_i)]).
\end{aligned}$$

Step 2 . In the new space calculate $\mathbf{W}$ and $\mathbf{B}$. Perform eigenanalysis to $\mathbf{B}$ and obtain a set of orthonormal eigenvectors. Create the two matrices $\mathbf{\Xi}_1 = [\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_q]$ and $\mathbf{\Xi}_2 = [\boldsymbol{\xi}_{q+1}, \ldots, \boldsymbol{\xi}_n]$ where $q = \text{rank}(\mathbf{B})$ that correspond to non-zero and zero eigenvalues, respectively.

Step 3 . Calculate $\tilde{\mathbf{W}} = \mathbf{\Xi}_1^T \mathbf{W} \mathbf{\Xi}_1$, $\tilde{\mathbf{B}} = \mathbf{\Xi}_1^T \mathbf{B} \mathbf{\Xi}_1$ and find the regular discriminant features using the matrix $\tilde{\mathbf{J}} = [\tilde{\boldsymbol{\zeta}}_1 \ldots \tilde{\boldsymbol{\zeta}}_q]$ whose columns are the eigenvectors of $\tilde{\mathbf{B}}^{-1} \tilde{\mathbf{W}}$ in descending order of the eigenvalues.

Step 4 . Calculate $\hat{\mathbf{W}} = \mathbf{\Xi}_2^T \mathbf{W} \mathbf{\Xi}_2$ and find the irregular discriminant projections using the matrix $\tilde{\mathbf{T}} = [\tilde{\boldsymbol{\tau}}_1 \ldots \tilde{\boldsymbol{\tau}}_q]$ whose columns are the orthonormal eigenvectors of $\hat{\mathbf{W}}$.

After following these steps the regular discriminant projection for a test facial vector $\mathbf{y}$ are given by:

$$\begin{aligned}
\acute{\mathbf{y}}_1 &= (\mathbf{\Pi} \mathbf{\Lambda_s}^{-1/2} \tilde{\mathbf{J}} \mathbf{\Xi}_1)^T ([k(\mathbf{z}_1, \mathbf{y}) \ldots k(\mathbf{z}_L, \mathbf{y})]^T - \\
&\quad -\tfrac{1}{L_G} \mathbf{1}_{L L_G} [k(\boldsymbol{\rho}_1, \mathbf{y}) \ldots k(\boldsymbol{\rho}_{L_G}, \mathbf{y})]).
\end{aligned} \tag{12}$$

The number of dimensions of the regular discriminant vectors is less or equal to $L_G - 1$. The irregular discriminant projection for the facial vector $\mathbf{y}$ is given by:

$$\begin{aligned}
\acute{\mathbf{y}}_2 &= (\mathbf{\Pi} \mathbf{\Lambda_s}^{-1/2} \tilde{\mathbf{T}} \mathbf{\Xi}_2)^T ([k(\mathbf{z}_1, \mathbf{y}) \ldots k(\mathbf{z}_L, \mathbf{y})]^T - \\
&\quad -\tfrac{1}{L_G} \mathbf{1}_{L L_G} [k(\boldsymbol{\rho}_1, \mathbf{y}) \ldots k(\boldsymbol{\rho}_{L_G}, \mathbf{y})]).
\end{aligned} \tag{13}$$

The number of dimensions of the feature vector $\acute{\mathbf{y}}_2$ is less or equal to $L_I - 1$. Two distinct similarity measures can be defined. The first corresponds to the regular discriminant information:

$$d_r(\acute{\mathbf{y}}_1) = ||\acute{\mathbf{y}}_1 - \bar{\boldsymbol{\rho}}_1||^2 \tag{14}$$

where $\bar{\boldsymbol{\rho}}_1$ is the regular discriminant vector of $\bar{\boldsymbol{\rho}}$. The second similarity measure corresponds to the irregular discriminant information:

$$d_r(\mathbf{y}) = ||\acute{\mathbf{y}}_2 - \bar{\boldsymbol{\rho}}_2||^2 \tag{15}$$

where $\bar{\boldsymbol{\rho}}_2$ is the irregular discriminant vector of $\bar{\boldsymbol{\rho}}$. The two similarity measures can be used in an independent fashion or can be fused using empirical or discriminant fusion rules [1, 3].

3. EXPERIMENTAL RESULTS AND DISCUSSION

The performance of a verification system is often quoted by a particular operating point of the *Receiver Operating Characteristic* (ROC) where *False Rejection Rate* (FAR)=*False Acceptance Rate* (FRR). This operating point is called *Equal Error Rate* (EER). The EER will be used to quantify the performance of the tested methods. The specific database contains four recordings of 295 subjects taken over a period of four months. Each recording contains a speaking head shot and a rotating head shot. In the specific procedure only the rotation shots have been used.

The testing database comprises of 120 subjects, 4 recording sessions and one shot of moving head per recording session. We should note here, that each session in the XM2VTS [5], as well as in the video XM2VTS database, has been captured with one month time intervals between each other. The database was randomly divided into 60 clients and 60 impostors. 2 sessions out of 4 of the clients' class where used for training the system, while 1 session was used for evaluation

Fig. 1. Data samples used for the experimental procedure. Each row represents the images taken from one session to consist one person's class.

and 1 for testing. For the impostors, 2 sessions were used for evaluation and 2 for testing. The number of images taken from each session for one person was 10. So, for the training set 1200 images were used. The number of images that were used was 1200 for the evaluation and the test set respectively. Thus, we have a total of 600 client claims and 36000 impostor claims for both, the evaluation and the test sets. A similarity measure $d_r(\mathbf{y})$ between faces is found in all the tested methods. In the proposed approaches the similarity measures were the ones defined in (14) and (15). In order to reject or accept an identity claim, a threshold should be used on this similarity measure. The methods in [6, 1] have been used for class-specific threshold selection. We have tested Kernel Principal Component Analysis (KPCA), multiclass CKFDA and class-specific CKFDA. The multiclass CKFDA gives 59 regular features and 59 irregular features using common bases for all the classes. The class specific CKFDA produces 2 features, one for the regular discriminant direction and one for the irregular one.

In Figure 2, the EERs for the test set are plotted for various polynomial kernel parameters for the multiclass KPCA [3], multiclass and class-specific CKFDA [3] approaches (regular and irregular information) and the proposed kernel discriminant analysis for polynomial kernels of power from 1 to 6. The best EER achieved for these methods has been measured

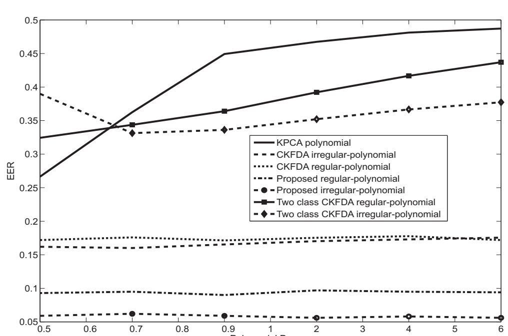

Fig. 2. ERR for KPCA, multiclass and two-class CKFDA methods (regular and irregular space) and the proposed technique with polynomial kernels.

at about 15% for the multiclass CKFDA while for the class specific CKFDA has been measured more than 30%. As can be seen the performance of the two-class variants of CKFDA is worse than the multiclass CKFDA. This is attributed to the very limited feature space that is provided by the two-class CKFDA. The best EER that has been achieved by our method was measured at about 5.5% which is a very good performance considering that the database contains faces at various poses.

4. CONCLUSION

Novel kernel based methods for discriminant feature extractions in two-class problems has been defined. The novel methods overcome the problems of the typical kernel-FLDA and of other approaches (like CKFDA) that give a very limited discriminant subspace spanned by one or two discriminant directions for two-class problems. The proposed approaches have been tested in face verification using facial images under various poses, where they show to outperform many other popular kernel methods.

5. REFERENCES

[1] S. Zafeiriou, A. Tefas, and I. Pitas, "Learning discriminant person-specific facial models using expandable graphs," *IEEE Transactions on Information Forensics and Security*, vol. 2, no. 1, pp. 55 – 69, 2007.

[2] S. Zafeiriou, A. Tefas, I. Buciu, and I. Pitas, "Exploiting discriminant information in nonnegative matrix factorization with application to frontal face verification," *IEEE Transactions on Neural Networks*, vol. 17, no. 3, pp. 683 – 695, 2006.

[3] Y.P. Kittler, J Li and J. Matas, "Face verification using client specific Fisher faces," *In J T Kent and R G Aykroyd, editors, The Statistics of Directions, Shapes and Images*, pp. 63–66, 2000.

[4] J. Yang, A.F. Frangi, J. Yang, D. Zhang, and Z. Jin, "KPCA plus LDA: A complete kernel Fisher discriminant framework for feature extraction and recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 2, pp. 230–244, 2005.

[5] K.-R. Muller, S. Mika, G. Ratsch, K. Tsuda, and B. Scholkopf, "An introduction to kernel-based learning algorithms," *IEEE Transactions on Neural Networks*, vol. 12, no. 2, pp. 181–201, 2001.

[6] K. Messer, J. Matas, J.V. Kittler, J. Luettin, and G. Maitre, "XM2VTSDB: The extended M2VTS database," in *AVBPA'99*, 1999, pp. 72–77.