Weighted Piecewise LDA for Solving the Small Sample Size Problem in Face Verification

Marios Kyperountas, Anastasios Tefas, and Ioannis Pitas

Aristotle University of Thessaloniki

Department of Informatics

Artificial Intelligence and Information Analysis Laboratory

Box 451, 54006 Thessaloniki, Greece

Address for correspondence:

Professor Ioannis Pitas

Aristotle University of Thessaloniki

Department of Informatics

Artificial Intelligence and Information Analysis Laboratory

Box 451, 54006 Thessaloniki, Greece

Tel. ++ 30 231 099 6304

Fax ++ 30 231 099 6304

Email: pitas@zeus.csd.auth.gr

Abstract

A novel algorithm that can be used to boost the performance of face verification methods that utilize Fisher's criterion is presented and evaluated. The algorithm is applied to similarity, or matching error, data and provides a general solution for overcoming the "*small sample size*" (*SSS*) problem, where the lack of sufficient training samples causes improper estimation of a linear separation hyper-plane between the classes. Two independent phases constitute the proposed method. Initially, a set of weighted piecewise discriminant hyper-planes are used in order to provide a more accurate discriminant decision than the one produced by the traditional linear discriminant analysis (LDA) methodology. The expected classification ability of this method is investigated throughout a series of simulations. The second phase defines proper combinations for person-specific similarity scores and describes an outlier removal process that further enhances the classification ability. The proposed technique has been tested on the M2VTS and XM2VTS frontal face databases. Experimental results indicate that the proposed framework greatly improves the face verification performance.

Index Terms: Face verification, linear discriminant analysis, small sample size problem

I. INTRODUCTION

Linear discriminant analysis is an important statistical tool for pattern recognition, verification, and, in general, classification applications. It has been shown that LDA can be effective in face recognition or verification problems [1, 2, 3]. In face recognition systems, the N closest faces, from a set of reference faces, to a test face are found. In face verification systems, a test face is compared against a reference face and a decision is made whether the test face is identical to the reference face (meaning the test face corresponds to a client) or not (meaning the test face corresponds to an impostor). The aforementioned problems are conceptually different. On one hand, a face recognition system usually assists a human face-recognition expert to determine the identity of the test face by computing all similarity scores between the test face and each human face stored in the system database and by ranking them. On the other hand, a face verification system should decide itself if the test face is a client or an impostor [4].

The evaluation criteria for face recognition systems are different from those applied to face verification systems. The performance of face recognition systems is quantified in terms of the percentage of correctly identified faces within the N best matches. By varying the rank N of the match, the curve of cumulative match score versus rank is obtained [5]. The performance of face verification systems is measured in terms of the *false rejection rate* (*FRR*) achieved at a fixed *false acceptance rate* (*FAR*) or vice versa. By varying FAR, the Receiver Operating Characteristic (ROC) curve is obtained. For a face verification system, there is a trade-off between the FAR and the FRR. The choice of the performance metric, i.e., FAR or FRR, that should be low depends on the nature of the application [6]. If a scalar figure of merit is used to judge the performance of a verification algorithm, it is usually the operating point where the FAR and FRR are equal, the so called *Equal Error Rate* (*EER*). A third difference is in the requirements needed when face recognition/verification systems are trained. Face recognition systems are usually trained on sets having one frontal image per person. For example, in face recognition experiments conducted on FERET database [7], the *fa* (regular facial expression) frontal images are used to train the system, while the *fb* (alternative facial expression) frontal images are used to test the system. Face verification systems usually need more images per individual for training to capture intra-class variability (i.e., to model the variations of the face images corresponding to the same individual). The requirements in the number of images increase dramatically when linear discriminant analysis is employed to accomplish feature selection [8].

In many cases, the available facial images are insufficient for carrying out the LDA process in a statistically proper manner. In this type of problems, Fisher's linear discriminant [9] is not expected to be able to discriminate well between face pattern distributions that in many cases cannot be separated linearly, unless a sufficiently large training set is available. More specifically, in face recognition or verification systems LDA-based approaches often suffer from the *SSS* problem, where the sample dimensionality is larger than the number of available training samples per subject [10]. In fact, when this problem becomes severe, traditional LDA shows poor generalization ability and degrades the classification performance.

In recent years, an increasing interest has developed in the research community in order to improve LDA-based methods and provide solutions for the *SSS* problem. The traditional solution to this problem is to apply LDA in a lower-dimensional PCA subspace, so as to discard the null space (i.e., the subspace defined by the eigenvectors that correspond to zero eigenvalues) of the within-class scatter matrix of the training data set [1]. However, it has been shown [11] that significant discriminant information is contained in the discarded space and alternative solutions have been sought. Specifically, in [12] a direct-LDA algorithm is presented that discards the null space of the between-class scatter matrix, which is claimed to contain no useful information, rather than discard the null space of the within-class scatter matrix. This approach was also used in [13], where a subspace of the null space of the within-class scatter matrix is used to solve the small sample size problem. First the common null space of the between-class scatter matrix and the within-class scatter matrix is removed, since it is useless for discrimination. Then, the null space of the resulting within-class scatter matrix is calculated in the lower-dimensional projected space. This null space, combined with the previous projection, represents a subspace which is useful for discrimination. The optimal discriminant vectors of LDA are derived from it.

The key to the approach in [14] is to use the direct-LDA techniques for dimensionality reduction and meanwhile utilize a modified Fisher criterion that is more closely related to the classification error. To obtain this modified criterion, weighted schemes should be introduced into the traditional Fisher criterion to penalize the classes that are close and can lead to potential misclassifications in the output space. In [14], however, simple weighted schemes are introduced into the reconstruction of the between-class scatter matrix in the dimensionality reduced subspace, such that the optimization can be carried out by solving a generalized eigenvalue problem without having to resort to complex iterative optimization schemes. The method in [15] utilizes a variant of direct-LDA to safely remove the null space of the between-class scatter matrix and applies a fractional step LDA scheme to enhance the discriminatory power of the obtained direct-LDA feature space. More recently, the authors in [10] formed a mixture of LDA models that can be used to address the high nonlinearity in face pattern distributions, a problem that is commonly encountered in complex face recognition tasks. They present a machine-learning technique that is able to boost an ensemble of weak learners, operating slightly better than random guessing, to a more accurate learner.

In [16], a linear feature extraction method which is capable of deriving discriminatory information of the LDA criterion in singular cases is used. This is a two-stage method, where PCA is first used to reduce the dimensionality of the original space and then a Fisher-based linear algorithm, called Optimal Fisher Linear Discriminant, finds the best linear discriminant features on the PCA subspace. One of the major disadvantages of using the Fisher criterion is that the number of its discriminating vectors capable to be found is equal to the number of classes minus one. Recently, it was shown [17] that alternative LDA schemes that give more than one discriminative dimensions, in a two class problem, have better classification performance than those that give one projection. This is done by only replacing the original between scatter with a new scatter measure.

In another attempt to address the SSS problem, the authors in [18] present the regularized LDA method (RLDA) that employs a regularized Fisher's separability criterion. The purpose of regularization is to reduce the high variance related to the eigenvalue estimates of the within-class scatter matrix, at the expense of potentially increased bias. By adjusting the regularization parameter R, a set of LDA variants are obtained, such as the *direct*-LDA of [12] for $R=0$, and the DLDA of [15] for $R=1$. The trade-off between the variance and the bias, depending on the severity of the SSS problem, is controlled by the strength of regularization. The determination of the optimal value for R is computationally demanding as it is based on exhaustive search [18].

Similarly, in [19] a new Quadratic Discriminant Analysis (QDA)-like method that effectively addresses the SSS problem using a regularization technique is presented. The *direct*-LDA technique is utilized to map the original face patterns to a low-dimensional discriminant feature space, where a regularized QDA is then readily applied. The regularization strategy used provides a balance between the variance and the bias in sample-based estimates and this significantly relieves the SSS problem.

In [20] a kernel machine-based discriminant analysis method, which deals with the nonlinearity of the face patterns' distribution is proposed which also attempts to solve the SSS problem. Initially, the original input space is non-linearly mapped to an implicit high-dimensional feature space, where the distribution of face patterns is hoped to be linearized and simplified. Then, a new variant of the *direct*-LDA method is introduced to effectively solve the SSS problem and derive a set of optimal discriminant basis vectors in the feature space. Unlike the original *direct*-LDA method of [12], zero eigenvalues of the within-class scatter matrix are never used as divisors in the proposed one. In this way, the optimal discriminant features can be exactly extracted from both inside and outside of the within-class scatter matrix's null space. In [21] the kernel trick is applied to transform the linear-domain Foley-Sammon optimal – w.r.t. orthogonality constraints– discriminant vectors, resulting in a new nonlinear feature extraction method. The Foley-Sammon method can obtain more discriminant vectors than LDA, however, it does not show good performance when having to deal with nonlinear patterns, such as face patterns. Thus, the kernel trick is employed to provide a nonlinear solution. In addition, this method handles the SSS problem effectively by ensuring that most of its discriminant solutions lie in the null space of the within-class matrix. In [22] a kernel optimization method is presented that maximizes a measure of class separability in the empirical feature space. The empirical feature space is a Euclidean space in which the training data are embedded in such a way that their geometrical structure –such as pair-wise distance and angle– in this feature space is preserved. This leads to a data-dependent kernel optimization capability where the optimized kernel can improve classification performance.

The *feature selection via linear programming* (FSLP) method [23] incorporates a feature selection process based on margin size, where margin is defined as the minimum distance between two bounding hyper-planes. The FSLP method can select features by maximizing the margin, thus circumventing the 'curse of dimensionality' problem in the small sample case, when the number of features is large. In addition, pair-wise feature selection is employed to choose the most relevant features for each pair of classes rather than select a fixed subset of features for each class to discriminate it from all other classes. The FSLP technique determines the number of features to select for each pair of classes. In [24] a probabilistic model is used to generalize LDA in finding components that are informative of or relevant for data classes, thus removing the restrictive assumption of normal distribution with equal covariance matrices in each class. The discriminative components maximize the predictability of the class distribution which is asymptotically equivalent to maximizing mutual information within the classes and finding principal components in the so-called learning or Fisher metrics. In [25] verification performance was increased by employing a divide-and-conquer technique. Specifically, a support vector machine (SVM) tree is used, in which the size of the class and the members in the class can be changed dynamically. Initially, a recursive data partition is realized by membership-based locally linear embedding data clustering. Then SVM classification is carried out in each partitioned feature subset. Thus, the authors attempt to solve the classification problem by forming multiple easier sub-problems.

This paper presents a framework of two independent and general solutions that aim to improve the performance of LDA-based approaches. This methodology is not restricted to face verification, but is able to deal with any problem that fits into the same formalism. In the first step, the dimensionality of the samples is reduced by breaking them down, creating subsets of feature vectors with smaller dimensionality, and applying discriminant analysis on each subset. The resulting discriminant weight sets are themselves weighted under a normalization criterion, thus making the piecewise discriminant functions continuous in this sense, so as to provide the overall discriminant solution. This process gives direct improvements to the two aforementioned problems as the non-linearity between the data pattern distributions is now restricted, whereas the reduced dimensionality also helps mend the *SSS* problem. A series of simulations that aim to formulate the face verification problem illustrate the cases for which this method outperforms traditional LDA. Various statistical observations are made about the discriminant coefficients that are generated. Remaining strong nonlinearities between corresponding subsets lead to a bad estimation of a number of discriminant coefficients due to the small training set used. These coefficients are identified and re-estimated in an iterative fashion, if needed. In the second stage, the set of similarity scores, that correspond to the reference images of each person, is used in a second discriminant analysis step. In addition, this step is complemented by an outlier removal process in order to produce the final verification decision that is a weighted version of the sorted similarity scores.

The outline of this paper is as follows: Section II describes the discriminant problem at hand in order to illustrate how the proposed framework contributes to tackling a standard face verification problem. Section III presents the two aforementioned stages that comprise the novel discriminant solution that is proposed in this paper. Section IV describes the structure of a series of simulations that can be used to provide indications on the expected performance of the algorithm. Section V describes the implementation of these simulations and provides the corresponding experimental results. Moreover, in the same section, the proposed methodology is tested on two well-established frontal face databases, the M2VTS and XM2VTS, in order to assess its performance on standard data sets. The *Brussels* protocol, which is used and described in [26], was applied to the M2VTS database and *Configuration I* of the *Lausanne* protocol [27] to the XM2VTS database training and testing procedures.

II. Problem Statement

A widely known face verification algorithm is elastic graph matching [28]. The method is based on the analysis of a facial image region and its representation by a set of local descriptors (i.e. feature vectors) extracted at the nodes of a sparse grid:

$$\mathbf{j}(\mathbf{x}) = \left(\hat{f}_1(\mathbf{x}), \ldots, \hat{f}_M(\mathbf{x}) \right) \tag{1}$$

where $\hat{f}_i(\mathbf{x})$ denotes the output of a local operator applied to image f at the $i-\mathrm{th}$ scale or the $i-\mathrm{th}$ pair (scale, orientation), $\mathbf{x}$ defines the pixel coordinates and M denotes the dimensionality of the feature vector. The grid nodes are either evenly distributed over a rectangular image region or placed on certain facial features (e.g., nose, eyes, etc.), called fiducial points. The basic form of the image analysis algorithm that was used to collect the feature vectors $\mathbf{j}$ from each face is based on multiscale morphological dilation and erosion and is described in [26]. All the feature vectors $\mathbf{j}$ that have been produced are normalized in order to have zero mean and unit magnitude. Let the superscripts r and t denote a reference and a test person (or grid) respectively. Then, the L_2 norm between the feature vectors at the $l-\mathrm{th}$ grid node is used as a (signal) similarity measure:

$$C_l = \left\| \mathbf{j}(\mathbf{x}_l^t) - \mathbf{j}(\mathbf{x}_l^r) \right\|. \tag{2}$$

Let $\mathbf{c}_t$ be a column vector comprised by the similarity values between a test and a reference person at all L grid nodes, i.e.:

$$\mathbf{c}_t = \left[C_1, \ldots, C_L \right]^{\mathrm{T}}, \tag{3}$$

In order to make a decision of whether a test vector corresponds to a client or an impostor, the following simple distance measure can be used, where $\mathbf{i}$ is an $L \times 1$ vector of ones:

$$D(t, r) = \mathbf{i}^{\mathrm{T}} \mathbf{c}_t . \tag{4}$$

The first phase of the algorithm proposed in this paper introduces a general LDA-based technique that is carried out in the training stage and finds weights for each similarity vector $\mathbf{c}_t$ in order to enhance the discriminatory ability of the distance measure.

As is the case in most face verification applications, both the M2VTS and XM2VTS databases, and the protocols they were evaluated under, allow for the final decision, of whether a test facial image corresponds to a client or an impostor, to be made by processing T different images of the reference face. That is, the test face is compared against all the images of the reference person contained in the training set. As a result, we end up with T similarity, or matching error, scores; traditionally, the final classification decision is based solely on the lowest error value. The second phase of the proposed algorithm provides an alternative score weighting method that improves the final classification rate significantly. The two methods are independent from one another and are proposed as general solutions for classification problems of analogous form.

III. Boosting Linear Discriminant Analysis

Let $\mathbf{m}_{r,C}$ and $\mathbf{m}_{r,I}$ denote the sample mean of the class of similarity vectors $\mathbf{c}_t$ that corresponds to client claims relating to the reference person r (intra-class mean) and those corresponding to impostor claims relating to person r (inter-class mean), respectively. In addition, let N_C and N_I be the corresponding numbers of similarity vectors that belong to these two classes and N be their sum, i.e., the total number of similarity vectors. Let $\mathbf{S}_W$ and $\mathbf{S}_B$ be the within-class and between-class scatter matrices, respectively [29]. Suppose that we would like to transform linearly the similarity vectors:

$$D'(t,r) = \mathbf{w}_r^{\mathrm{T}} \mathbf{c}_t \,. \tag{5}$$

The most known and plausible criterion is to find a projection, or, equivalently, choose $\mathbf{w}_r$ that maximizes the ratio of the between-class scatter against the within-class scatter (Fisher's criterion):

$$J(\mathbf{w}_r) = \frac{\mathbf{w}_r^{\mathrm{T}} \mathbf{S}_B \mathbf{w}_r}{\mathbf{w}_r^{\mathrm{T}} \mathbf{S}_W \mathbf{w}_r} \,. \tag{6}$$

For the two-class problem, as is the case of face verification, Fisher's linear discriminant provides the vector that maximizes (6) and is given by:

$$\mathbf{w}_{r,0} = \mathbf{S}_W^{-1} (\mathbf{m}_{r,I} - \mathbf{m}_{r,C}) \,. \tag{7}$$

A. Weighted Piecewise Linear Discriminant Analysis (WPLDA) Model

Our experiments, which are discussed in Section IV, revealed that the traditional Fisher's linear discriminant process not only performs poorly, but, in certain cases, degrades the classification capability of the face verification algorithm, particularly when training data from the M2VTS database was used as can be seen in Table III. That is, the distance measure in (4) provided a much better solution than (5) after traditional LDA was used to determine the values of $\mathbf{w}_r$. This malady can be attributed to the insufficient number N_C of client similarity vectors, with respect to the dimensionality L of each vector $\mathbf{c}_t$. This is the case for most face verification problems and for the *Brussels* [26] and *Lausanne* protocols [27] as well.

The first thing that is done is to provide better estimation to Fisher's linear discriminant function. The main problem is that the class of client claims is very small in relation to the impostor class. This fact may affect the training [30]. As a result, a modified Fisher's linear discriminant is used by redefining (7) to:

$$\mathbf{w}_{r,0} = \mathbf{S}_W^{-1}\left(\mathbf{m}_{r,I}\frac{N_I}{N} - \mathbf{m}_{r,C}\frac{N_C}{N}\right), \tag{8}$$

so as to accommodate the prior probabilities of how well the mean of each class is estimated. Secondly, and for claims related to each reference person r, grid nodes that do not possess any discriminatory power are discarded. At an average 4 nodes, out of 64, are discarded for a 8×8 grid. Simply, each of the L' remaining nodes in $\mathbf{c}_t^{'}$ must satisfy:

$$\mathrm{m}_{r,I}(l) \geq \mathrm{m}_{r,C}(l), \quad l = 1,\ldots,L'. \tag{9}$$

The novelty of our approach is that in order to give remedy to the *SSS* problem, each similarity vector $\mathbf{c}_t^{'}$ with dimensionality L' is broken down to P smaller dimensionality vectors, $\mathbf{c}_{t,i}^{'}$, $i = 1,\ldots P$, each one of length M, where $M \leq (N_C - 1)$, thus forming P subsets. The more statistically independent the vectors $\mathbf{c}_{t,i}^{'}$ are to each other, the better the discriminant analysis is expected to be. As a result, P separate Fisher linear discriminant processes are carried out and each of the weight vectors produced is normalized, so that the within group variance equals to one, by applying:

$$\mathbf{w}_{r,0,i}^{'} = \mathbf{w}_{r,0,i}(\mathbf{w}_{r,0,i}{}^{\mathrm{T}}\mathbf{S}_{W,i}\mathbf{w}_{r,0,i})^{-\frac{1}{2}}, \tag{10}$$

where $i = 1,\ldots,P$ is the index of $M-$dimensionality vector $\mathbf{c}_{t,i}^{'}$ corresponding to a subset of similarity vector coordinates. This normalization step enables the proper merging of all weight vectors to a single column weight vector, $\mathbf{w}_{r,0}^{'}$, as such:

$$\mathbf{w}'_{r,0} = \left[\mathbf{w}'^{\mathrm{T}}_{r,0,1}, \ldots, \mathbf{w}'^{\mathrm{T}}_{r,0,P}\right]^{\mathrm{T}}. \tag{11}$$

B. Re-estimating the Defective Discriminant Coefficients

By meeting condition (9), all discriminant coefficients that correspond to the remaining grid nodes should indicate a constructive contribution to the overall discriminatory process. Since the matching error is always positive and impostor matching errors should be larger than the client errors, $\mathbf{w}'_{r,0}$ should be a vector of L' positive weights only. The exception to this is the possibility to have zero-valued weights that would indicate that certain grid nodes do not contribute to the classification process. In spite of this, when the set of client similarity data presents overlap with the set of impostor similarity data, such that no single linear hyper-plane can separate the two classes, it is likely, depending on the amount of overlap and/or how severe the SSS problem is, that a number of the discriminant coefficients in $\mathbf{w}_{r,0}$ may be found to be negative by the discriminant algorithm during the training phase. The WPLDA model that is introduced is less susceptible to these occurrences, as it settles the SSS problem. Any negative discriminant coefficients that remain in $\mathbf{w}'_{r,0}$ are caused by large nonlinearities of the separation surface between the distribution patterns of corresponding subsets and/or the lack of a sufficiently large number of training samples.

By having the a-priory knowledge that negative discriminant coefficients are the direct result of a faulty estimation process and assuming that L_r is the number of negative weights found in $\mathbf{w}'_{r,0}$, the following two cases are considered:

Case 1: $L_r > 0.5 \cdot M$

In this case, all the grid node training data that correspond to the negative coefficients in $\mathbf{w}'_{r,0}$ are collected and re-distributed into P' $M-$dimensionality vectors where, ideally, each subset again holds M similarity values. Now, an additional discriminant process can be applied on the data contained in one of the P' subsets in order to produce M new discriminant weights. Indeed, P' separate Fisher linear discriminant operations are carried out by using (8) and each of the P' weight vectors of length M that is produced is normalized by using (10).

Successively, all positive weights from all P' vectors are collected and used as the final multipliers of $\mathbf{c}_t$ (discriminant coefficients). On the other hand, all negative weights are collected and once again tested against cases 1 and 2. This process is carried out in as many iterations as are required for *Case 2* to apply. That is, the total number of negative discriminant weights should ideally be zero, or at least smaller or equal to M. The number of iterations in this procedure should drop if training is carried out on the low (M) dimensionality similarity vectors, as opposed to having been applied on the full L' – dimensional vectors.

Case 2: $L_r \leq 0.5 \cdot M$

All negative weights are set equal to zero and no further processing is required. The factor 0.5 is used to indicate that, if the number of similarity values, in the final M – dimensionality vector that holds similarity values corresponding to negative discriminant coefficients, is not equal to more than half of its full capacity M, the corresponding linear discriminant equation depends on too few variables and is likely to give large inaccuracies to the overall discriminant solution. In order to avoid building the overall discriminant solution by also including these large inaccuracies, which will essentially translate to defective discriminant coefficient values, zero weights are assigned to indicate that these coefficients no longer have any discriminant significance.

C. Weighting the Multiple Classification Scores

Most, if not all, face verification applications allow for a test person to be classified as an impostor or a client by using numerous images of the reference face. Thus, numerous verification tests are carried out whenever a claim is considered. As a result, multiple (T) classification scores $D'(t, r_d)$, where $d = 1, \ldots, T$, are available for each claim of an identity r, by a test person t. Traditionally, the test person t is classified as a client if the minimum value out of the total T scores, $D'(t, r_{\min}) \overset{\triangle}{=} \min_{i=1,\ldots T} \{D'(t, r_d),\ d = 1, \ldots, T\}$, is below a predefined threshold, and as an impostor, if it is above this threshold. In this work, training data are used once again to derive person specific weights to be used for the combination of the T scores. The motivation behind this process is that, ideally, all T scores should contribute to the final classification decision, as, in certain cases, the impostor image that corresponds to a minimum score may have accidentally - e.g., due to a particular facial expression or due to similar eyeglasses- had close similarity to a certain reference image. In such a case, the remaining reference images can be used in an effort to repair the false classification decision. Now the problem becomes:

$$D''(t, r) = \sum_{d=1}^{T} v_{r,d} D'(t, r_d) \tag{12}$$

Again, Fisher's modified linear discriminant (8) is applied to determine the vector $\mathbf{v}_r$ which contains the T weights $v_{r,d}$, $d=1,\ldots,T$, of the classification scores.

A much larger number of impostor, rather than client, similarity scores is usually available in the training set of a face verification database. This increases the probability that some impostor images may randomly give a close match to a reference photo, even closer than some of the client images give. Whenever this happens, the process of estimating a separation between the two classes degrades significantly because of the small number of client training similarity scores, which equals to the number of training samples in section III-A. Thus, an outlier removal process is incorporated, where the minimum impostor similarity scores in the training set of each reference person, i.e. all $D'(t,r_1)$ scores that correspond to impostor matches, are ordered and the smallest Q% of these values are discarded. As a result, the linear discriminant process gives a more accurate separation that helps increase the classification performance.

IV. Simulated and Experimental Results

In this section, the efficiency of the proposed discriminant solution is evaluated using both simulated and real data sets. The simulated data sets are used in order to deduce experimental evidence on the performance of WPLDA, whereas the real data that are taken from the M2VTS and XM2VTS databases are used to test the classification ability of the overall discriminant algorithm that is presented in this paper.

A. *Classification Performance on Simulated Data*

In order to provide relevant background on the expected performance of the proposed WPLDA algorithm in face verification, simulations that tackle the 2-class problem are carried out. We intent to investigate the cases where one can expect the WPLDA algorithm to outperform the traditional LDA algorithm, with respect to the size of the impostor and client classes. For each verification experiment, two classes of matching vectors, one that corresponds to the clients and the other to the impostors, are created. Each class contains N sample vectors of dimensionality L. Each of these sample vectors contains entries drawn from a normal (Gaussian) distribution. The L random entries to each sample vector of class X_j, which is the $j-\mathrm{th}$ client or impostor class, are generated by

$$\mathrm{N}_{X_j}\left(x_i^j : \mu_i^j, \sigma_i^j\right) = \frac{1}{\sigma_i^j \sqrt{2\pi}} e^{-\frac{\left(x_i^j - \mu_i^j\right)^2}{2\left(\sigma_i^j\right)^2}}, i = 1 \ldots L, \tag{13}$$

where $\mu_i^j = G_j + \alpha r_i$ and $\sigma_i^j = K_j + \beta r_i$. G_j is the expected mean value and K_j the expected standard deviation for the $i-\text{th}$ random entry of the $j-\text{th}$ class and r_i is a random number, chosen from a normal distribution with zero mean and unit variance. The scalars α and β affect the uniformity among the vectors of each class.

The dimensionality of the sample vectors is set to $L = 64$ in order to be identical to the dimensionality of the feature vectors - or to the number of grid nodes - of the real face verification problem that we are trying to solve in section IV-C. Each class contains $N = 2000$ sample vectors. Let I be the impostor class and C_1 and C_2 be two client classes. Let the random entries to each sample vector of the impostor class I and the client classes C_1 and C_2 be generated based on the following normal distributions, respectively:

$$\mathrm{N}_I\left(x_i : \mu_i = 100 + 5r_i, \sigma_i = 25 + 5r_i\right), \quad i = 1 \ldots 64. \tag{14}$$

$$\mathrm{N}_{C_1}\left(x_i : \mu_i = 87 + 5r_i, \sigma_i = 35 + 5r_i\right), \quad i = 1 \ldots 64. \tag{15}$$

$$\mathrm{N}_{C_2}\left(x_i : \mu_i = 85 + 5r_i, \sigma_i = 35 + 5r_i\right), \quad i = 1 \ldots 64. \tag{16}$$

It is clear that the mean of the random entries of C_2 is expected to deviate more, w.r.t. the mean of the entries of C_1, from the mean of the entries of I.

When the elastic graph matching algorithm is applied to face verification tasks, it is expected that certain nodes should provide more discriminant information than others. This is also true for most feature-based verification methods. For example, in general a node that lies at the location of the nose will be more useful than a node that lies at a location on the forehead. In order to simulate a similar situation, we create a subset of L_B nodes (out of the total L), that is expected to be more discriminant than the remaining nodes. We name this set of L_B nodes as 'most discriminant coefficients'. Let a client class C_3 be created, such that the entries at the L_B nodes are taken from the C_2 client class (since the entries from C_2 are more separated from the entries in I than the entries of C_1 are) and the rest of the node entries from the C_1 class. For this first set of experiments we let $L_B = 5$ and the positions of the 5 most discriminant coefficients are selected so as to be evenly spaced from one another, e.g. their coefficient index is given by $\{11,22,33,44,55\}$ for $L = 64$.

The data that were created are used to compare the discrimination ability of traditional LDA and the proposed WPLDA for various numbers of training sample vectors for the impostor and client class. For each 2-class problem that is formulated, one training and one test set are created. The training set of LDA and WPLDA is formed based on the random selection out of the complete set of N sample vectors of each class. The remaining sample vectors of each class, obtained by excluding the training set of LDA and WPLDA, form the test set that is used to evaluate the classification performance.

In order to approximate the ideal linear discriminant solution, a third method that will be referred to as Ideal LDA (ILDA) will always apply the traditional LDA algorithm making use of the complete sets of N client sample vectors and N impostor sample vectors, during the training phase. We consider this number of samples to be large enough for the traditional LDA algorithm to produce a statistically correct discriminant solution. The test set where the performance of ILDA will be evaluated on, is identical to the test set of LDA and WPLDA. Thus, the test set is always included in the training set of ILDA, so as to best approximate the ideal linear discriminant solution and provide ground-truth results. In addition, and again for comparison purposes, the classification performance of a fourth method will be considered, where this method simply computes the mean of the sample vectors (MSV) and produces a non-weighted result which can be used to indicate how difficult the 2-class classification problem is.

In order to evaluate the performance of the four aforementioned methods the equal error rate (*EER*) is employed. Each of the *EER* values reported has been averaged over 20 independent runs of an identical experiment for more accurate results. The simulation data are used in various discriminant processes that aim to separate out the client and impostor classes. The 2-class problem that is studied next uses data from I and C_3. Figures 1-3 show the *EER* when the number of client sample vectors varies from 2 to 100. It is noted that logarithmic scales are used for the *y*-axis. Figure 1 shows the EER results when the number of impostor sample vectors is 10. For the LDA algorithm, the *SSS* is expected to have the most severe effects on the EER when the client class has less than $(L+1) = 65$ samples. In theory, in this case neither the client class nor the impostor class can be properly modelled by traditional LDA and, as a result, an appropriate separation between the two classes cannot be found. On the other hand, WPLDA is not affected by the *SSS* problem as can be seen in Figure 1. The small variations in the *EER* of ILDA indicate the amount of randomness in our results since only the *y*-axis showing *EER* is significant for the ILDA results. Figure 2 and 3 show the *EER* rates for 100 and 1000 impostor sample vectors respectively. It is clearly seen in these Figures that, unless a relatively large number of client and impostor sample vectors are available, WPLDA outperforms LDA.

Figure 2 shows that, when 100 impostor and 83 client sample vectors are available, the performance of LDA becomes better than that of MSV. Figure 3 shows that when the number of impostor sample vectors becomes 1000, 20 client sample vectors are required for LDA to outperform WPLDA. For most current biometric databases, having 20, or more, client samples per person is quite uncommon. Figure 3 also shows that when the client and impostor class sizes are large enough such that traditional LDA can find a proper estimation of a linear separation hyper-plane between the classes, traditional LDA presents a stronger classification performance since the proper higher-dimensionality solution is more general than the lower-dimensionality solutions offered by WPLDA. For reference, it is stated that in simulations we run where the client class consisted only sample vectors from either C_1 or C_2 the average drop in the *EER* rate of LDA when 1000 impostor sample vectors are used instead of 10 is 25.50%, whereas for WPLDA 0.37%. Of course, it is expected that a quite large number of impostor sample vectors are required for the LDA algorithm to outperform WPLDA when the number of client sample vectors is only 6, as is in the face verification problem that we are trying to solve in section IV-C. As a result, we expect that WPLDA should provide better verification performance.

B. Discriminant Characteristics under the SSS Problem

The second set of experiments using simulated data involves investigating the statistical behaviour of the discriminant coefficients of the LDA and WPLDA processes with reference to ILDA. Moreover, *EER* rates are reported for different numbers of ‘most discriminant coefficients’ contained in each class, that is, for various values of L_B. The L_B most discriminant coefficients are evenly spread out, as much as possible, in the L-dimensional space. In order to determine how efficient each discriminant method is in recognizing the importance of the most discriminant coefficients, a separation criterion between the most discriminant and the remaining coefficients is defined as:

$$H = \frac{|m_B - m_R|}{s_B + s_R}, \tag{17}$$

where m_B and s_B are scalars representing the average mean and the average standard deviation of the set of most discriminant coefficients and m_R and s_R those of the remaining coefficients. If $H \geq 1$, the separation criterion is satisfied, since then the values of the most discriminant coefficients vary significantly from those of the remaining coefficients.

Based on the practical considerations of the face verification problem at hand, this set of simulations is modelled under the *SSS* problem, where the client class has less sample vectors than the dimensionality of the similarity vectors. The M2VTS and XM2VTS face verification test protocols specify for the client class to avail 6 training samples and for the impostor class to avail 210 or 1791 training samples for the M2VTS and XM2VTS databases respectively. Therefore, in order to correlate the simulation results with the expected performance of the MSV, LDA and WPLDA algorithms in these protocols, we randomly select 6 sample vectors from the C_3 client class and 1000 sample vectors from the I impostor class to train LDA and WPLDA. The coefficients of ILDA are once again generated by a training set of 2000 client and 2000 impostor sample vectors. To observe the statistical behaviour of the discriminant coefficients, 1000 independent runs were carried out. The entries at the position of the L_B elements are expected to have a larger distance from the corresponding element entries of class I, than the rest. As a result, the discriminant process should give larger weights for the element entries at these L_B specific positions, since they are expected to be the most useful in producing a meaningful separation between the impostor and the client class.

Figures 4,5 and 6 show the boxplots [31] that provide statistical information about the calculation of the 64 discriminant coefficients, $\mathbf{w}_i^{'}$, $i = 1 \ldots 64$, throughout the 1000 independent runs, by ILDA, LDA and WPLDA respectively. These three methods processed the C_3, (with $L_B = 5$), and I training data. The boxes have lines at the lower quartile, median, and upper quartile values. The whiskers are lines extending from each end of the boxes to show the extent of the rest of the data, specified as 1.5 times the inter-quartile range. Outliers are data with values beyond the ends of the whiskers and are indicated using '+'. It is clear that WPLDA, unlike LDA, provides a complete separation to all the most discriminant coefficients from the remaining coefficients, in terms of assigning largest weights, while ILDA almost does the same. Results such as the ones shown in Figures 4 through 6 were produced for various values of L_B and the corresponding *EER* and H values (17) were calculated. These results are summarized in Table I. Once again, the case where half the coefficients are the most discriminant provides an exception to our results since, in all other cases, WPLDA is related with the largest H value. In addition, WPLDA provides the *EER* rate that is closest to the corresponding ILDA rate. The *EER* of WPLDA and MSV is identical in the cases where no 'most discriminant coefficients' exist, i.e. when as many as half the coefficients are most discriminant ones. As expected, WPLDA always shows a better classification performance than traditional LDA, under the *SSS* problem.

C. Performance Evaluation on the M2VTS and XM2VTS Databases.

In this section, experimental tests are carried out by applying the testing protocols of the M2VTS and XM2VTS databases. The M2VTS database contains video data of 37 persons. Four recordings/shots of the 37 persons have been collected, each containing a frontal-face pose. The *Brussels* protocol, which is used in [26, 29] requires the implementation of four experimental sessions by employing the "leave-one-out" and "rotation" estimates. In each session, one shot is left out to be used as the test set. In order to implement test impostor claims, rotations over the 37 person identities are carried out by considering the frontal face image of each person in the test set as impostor. By excluding any frontal face image of the test impostor from the remaining three shots, a training set that consisted of 36 clients is built. The test impostor pretends to be one of the 36 clients and this attempt is repeated for all client identities. As a result, 36 impostor claims are produced. In a similar manner, 36 test client claims are tested by employing the client frontal faces from the shot that is left out, and those of the training set. The training procedure is analogous to the test procedure that was just described. It is applied to the training set of the 36 clients. Three frontal face images are available for each client. By considering all permutations of the three frontal images of the same person, taken two at a time, 6 training client claims can be implemented. Moreover, 210 training impostor claims, when each of the other 35 persons attempt to access the system with the identity of the person under consideration, are implemented. That is, another 6 raw similarity vectors corresponding to all pair-wise comparisons between the frontal images of any two different persons taken from different shots are produced. For a more detailed description of the *Brussels* protocol the reader is referred to [26].

The XM2VTS database contains four recordings of 295 subjects taken over a period of four months. The *Lausanne* protocol described in [27] splits randomly all subjects into client and impostor groups. The client group contains 200 subjects, the impostor group is divided into 25 evaluation impostors and 70 test impostors. Eight images from 4 sessions are used. From these sets consisting of face images, a training set, an evaluation set and a test set are built. There exist two configurations that differ in the selection of particular shots of people into the training, evaluation and test sets. The training set of the *Configuration I* contains 200 persons with 3 images per person. The evaluation set contains 3 images per client for genuine claims and 25 evaluation impostors with 8 images per impostor. The test set has 2 images per client and 70 impostors with 8 images per impostor. The training set is used to construct client models. The evaluation set is selected to produce client and impostor access scores, which are used to find a threshold that determines if a person is accepted or not as a client (it can be a client-specific threshold or global threshold). According to the *Lausanne* protocol the threshold is set to satisfy certain performance levels (error rates) on the evaluation set. Finally the test set is selected to simulate realistic authentication tests where the impostor identity is unknown to the system [32]. For a more detailed description of the *Lausanne* protocol the reader is referred to [27].

The proposed methodology is now evaluated using the *Brussels* and *Lausanne* standard protocols described above that, as is the case with most face verification applications, suffer from the SSS problem. Specifically, the number of client similarity vectors N_C for each individual that were available in the training set was only 6, whereas the 8×8 grid that was used set the dimensionality of the similarity vector to $L = 64$. The value of N_I was set to 210 and 1791, when training the algorithm using M2VTS and XM2VTS data respectively. For the XM2VTS database and *Configuration I* of the *Lausanne* protocol, a total of 600 (3 client shots x 200 clients) client claim tests and 40,000 (25 impostors x 8 shots x 200 clients) impostor claim tests were carried out for the evaluation set and 400 (2 client shots x 200 clients) client claims and 112,000 (70 impostors x 8 shots x 200 clients) impostor claims for the test set. For the M2VTS database and the *Brussels* protocol, a total of 5,328 client claim tests and 5,328 impostor claim tests (1 client or impostor x 36 rotations x 4 shots x 37 individuals) were carried out. Face verification decision thresholds from the corresponding training process of each database were collected and used to evaluate the verification results, except for the evaluation of the XM2VTS test set, where thresholds from the evaluation process were used, as [27] suggests.

The discriminant coefficient vectors $\mathbf{w}'$ derived by the processes described in sub-sections III-A and III-B have been used to weigh the normalized similarity vectors $\mathbf{c}$ that are provided by the Morphological Elastic Graph Matching procedure applied to frontal face verification, based on the algorithm described in [26]. Our tests revealed that the optimum value for M is 4. Moreover, it was observed that, on the average, 36.54% of the discriminant coefficients in $\mathbf{w}_{r,0}$ and 6.27% of the discriminant coefficients in $\mathbf{w}^{'}_{r,0}$ were found to be negative for the M2VTS training set. Additionally, 24.39% of the discriminant coefficients in $\mathbf{w}_{r,0}$ and 0.76% of the discriminant coefficients in $\mathbf{w}^{'}_{r,0}$ were found to be negative, when the larger XM2VTS training set was used. In addition, during the training stages of the M2VTS database, 3 to 5 iterations described in sub-section III-B, are usually required when $M = L'$, whereas no more than 2 iterations are required when M is set to 4. For the latter value of M, one, at the most, iteration is needed when processing XM2VTS data.

The procedure described in Section III-C is used to calculate a more accurate similarity score for each tested individual. The testing protocols specify that a test person can be classified to be an impostor or a client by using three different images of the reference person $(T = 3)$. Thus, three tests are carried out and three similarity scores are available for each individual. Unfortunately, the training data which we can work with to derive these weights only provide two combinations, since a total of 6 training client combinations are available for the 3 different images of each person. Thus, we are forced to set the largest similarity score to zero and set $T = 2$ in (12). The two weights are found using (8). For the outlier removal process, Q is set to 4, that is, 4% of the minimum impostor similarity scores is discarded.

Let us denote the combination of the morphological elastic graph matching, (MEGM), and the weighting approach that makes up for the first phase of the proposed algorithm, as is described in sub-sections III-A and III-B, by, once again, WPLDA. Moreover, let MS-WPLDA be the second phase of the algorithm that is applied on WPLDA and is described in sub-Section III-C, where ‘MS’ stands for multiple score. In order to evaluate the performance of these methods the *FAR* and *FRR rate* measures are used. Figure 7 shows a critical region of the ROC curves for the raw MEGM data using (4), classical LDA (7) applied on the raw MEGM data, WPLDA and MS-WPLDA evaluated on the M2VTS database. Figure 8 shows the same corresponding ROC curves when the algorithms were evaluated on the XM2VTS evaluation set and Figure 9 the corresponding ones for the XM2VTS test set. Results are presented in logarithmic scales. In addition, Table II shows the *EER* for each algorithm.

When the M2VTS data are used, the traditional LDA algorithm degrades the classification performance significantly, having a poor generalization ability, which stems from the largely inadequate, in terms of size, training set that was available. Traditional LDA underperforms, with respect to WPLDA, at a larger degree on the M2VTS, rather than on the XM2VTS, experiments. This can be attributed to the larger data set that is used in the XM2VTS training process. In addition, the experimental results show that the proposed WPLDA algorithm performs better than either MEGM or LDA, as was previously indicated by the simulation results of section IV. Furthermore, the independent MS-WPLDA process provides additional improvement to the classification ability of WPLDA.

In order to compare the performance of WPLDA with a state-of-the-art method it is important to select an algorithm which is expected to perform well not only under the *SSS* problem, but also when dealing with the 2-class problem. For example, the algorithms in [20] and [23] are designed under the assumption of a multi-class problem. On the contrary, the RLDA algorithm that was recently proposed in [18] is not designed around the multi-class problem. Thus, we apply the *Brussels* and *Lausanne* face verification protocols to evaluate its performance. For salient comparisons, we report results generated by the RLDA algorithm after discarding the nodes that do not possess any discriminatory power, by making use of (9). The EER performance of RLDA is shown in Table II. These results illustrate that WPLDA always gives better classification results that the RLDA algorithm. It is anticipated that the bias introduced by the regularization in order to reduce the high variance related to the eigenvalue estimates of the within-class scatter matrix limits the classification accuracy of RLDA (essentially due to having insufficient samples to represent the client class), whereas WPLDA achieves a better solution since it decomposes the discriminant analysis problem into multiple lower-dimensionality problems.

V. CONCLUSION

A novel methodology is proposed in this paper that provides general solutions for LDA-based algorithms that encounter problems relating to inadequate training and to the *SSS* problem in particular. This methodology was tested on two well-established databases under their standard protocols for evaluating face verification algorithms. Moreover, a set of simulations gave indications on when the proposed weighted piecewise linear discriminant analysis algorithm outperforms traditional LDA. Results indicate that the processes described in this paper boost the performance of the verification algorithm significantly (31.2%, 21.7% and 17.6% drop of the *EER* rate in the three experimental sets). It is anticipated that the performance of other LDA variants may be enhanced by utilizing processes that stem from this framework.

VI. ACKNOWLEDGEMENTS

This work was partially funded by the integrated project BioSec IST-2002-001766 (Biometric Security, http://www.biosec.org), under Information Society Technologies (IST) priority of the 6th Framework Programme of the European Community and is partially 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.

REFERENCES

[1] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. Fisherfaces: recognition using class specific linear projection", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711–720, 1997.

[2] K. Etemad and R. Chellappa, "Discriminant analysis for recognition of human face images", *Journal of the Optical Society of America A: Optics Image Science and Vision*, vol. 14, no 8, pp. 1724-1733, 1997.

[3] W. Zhao, R. Chellappa, and P.J. Phillips, "Subspace Linear Discriminant Analysis for face recognition", *Center for Automation Research*, University of Maryland, College Park, Technical Report, CAR-TR-914, 1999.

[4] C. Kotropoulos, A. Tefas, and I. Pitas, "Frontal face authentication using discriminating grids with morphological feature vectors", *IEEE Trans. on Multimedia*, vol. 2, no. 1, pp. 14-26, March 2000.

[5] P. Jonathon Phillips, "Matching pursuit filters applied to face identification", *IEEE Trans. on Image Processing*, vol. 7, pp. 1150-1164, 1998.

[6] A. Tefas, C. Kotropoulos, and I. Pitas, "Face verification using elastic graph matching based on morphological signal decomposition", *Signal Processing*, vol. 82, no. 6, pp. 833-851, June 2002.

[7] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, "The FERET evaluation methodology for face-recognition algorithms", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 22, no. 10, pp. 1090-1104, Oct. 2000.

[8] C. Kotropoulos, A. Tefas, and I. Pitas, "Frontal face authentication using variants of Dynamic Link Matching based on mathematical morphology", in *Proc. IEEE Int. Conf. on Image Processing*, vol. I, pp. 122-126, 1998.

[9] R.A. Fisher, "The use of multiple measurements in taxonomic problems", in *Annals of Eugenics*, vol. 7, no. 2, pp. 179-188, 1936.

[10] J. Lu, K.N. Plataniotis, and A.N. Venetsanopoulos, "Boosting linear discriminant analysis for face recognition", in *Proc. IEEE Int. Conf. on Image Processing,* vol. 1, pp. I - 657-60, Barcelona, Spain, 14-17 Sep., 2003.

[11] L-F Chen, M. H-Y Liao, J-C Lin, M-T Ko, G-J Yu, "A new LDA-based face recognition system which can solve the small sample size problem", *Pattern Recognition*, vol. 33, pp. 1713–1726, 2000.

[12] H. Yu and J. Yang, "A direct LDA algorithm for high- dimensional data with application to face recognition", *Pattern Recognition*, vol. 34, pp. 2067-2070, 2001.

[13] R. Huang; Q. Liu, H. Lu, and S. Ma**,** "Solving the small sample size problem of LDA", *in Proc. 16th Int. Conf. on Pattern Recognition*, vol. 3 , pp. 29-32, 11-15 Aug., 2002.

[14] D. Zhou and X. Yang: "Face recognition using Improved-LDA", in *Proc. Int. Conf. on Image Analysis and Recognition (ICIAR'04),* vol. 2, pp. 692-699, 2004.

[15] J. Lu, K.N. Plataniotis, and A.N. Venetsanopoulos, "Face recognition using LDA based algorithms", *IEEE Trans. on Neural Networks*, vol. 14, no. 1, pp. 195-200, Jan. 2003.

[16] J. Yang and J. Yang, "Why can LDA be performed in PCA transformed space?", *Pattern Recognition*, vol. 36, pp. 563-566, 2003.

[17] C. Songcan and Y. Xubing, "Alternative linear discriminant classifier", *Pattern Recognition*, vol. 37, no. 7, pp. 1545-47, 2004.

[18] J. Lu, K.N. Plataniotis, A.N. Venetsanopoulos, "Regularization studies of linear discriminant analysis in small sample size scenarios with application to face recognition", *Pattern Recognition Letters*, vol. 26, no. 2, pp. 181-191, 2005.

[19] J. Lu, K.N. Plataniotis, A.N. Venetsanopoulos, "Regularized discriminant analysis for the small sample size problem in face recognition", *Pattern Recognition Letters*, vol. 24, no. 16, pp. 3079-3087, 2003.

[20] J. Lu, K.N. Plataniotis, A.N. Venetsanopoulos, "Face recognition using kernel direct discriminant analysis algorithms", *IEEE Trans. on Neural Networks*, vol. 14, no. 1, pp. 117-126, Jan. 2003.

[21] W. Zheng, L. Zhao, C. Zou, "Foley-Sammon optimal discriminant vectors using kernel approach", *IEEE Trans. on Neural Networks*, vol. 16, no. 1, pp. 1-9, Jan. 2005.

[22] H. Xiong, M.N.S Swamy, M.O. Ahmad, "Optimizing the kernel in the empirical feature space", *IEEE Trans. on Neural Networks*, vol. 16, no. 2, pp. 460-474, March 2005.

[23] G. Guo and C.R. Dyer, "Learning from examples in the small sample case: face expression recognition", *IEEE Trans. on Systems, Man and Cybernetics, Part B*, vol. 35, no. 3, pp. 477-488, June 2005.

[24] J. Peltonen and S. Kaski, "Discriminative components of data", *IEEE Trans. on Neural Networks*, vol. 16, no. 1, pp. 68-83, Jan. 2005.

[25] S. Pang, D. Kim, S.Y. Bang, "Face membership authentication using SVM classification tree generated by membership-based LLE data partition", *IEEE Trans. on Neural Networks*, vol. 16, no. 2, pp. 436-446, March 2005.

[26] C. Kotropoulos, A. Tefas and I. Pitas, "Frontal face authentication using morphological elastic graph matching", *IEEE Trans. on Image Processing*, vol. 9, no. 4, pp. 555-560, April 2000.

[27] J. Luettin, and G. Maitre, "Evaluation protocol for the extended M2VTS database (XM2VTSDB)", in *IDIAP Communication 98-05*, IDIAP, Martigny, Switzerland, 1998.

[28] M. Lades, J. Vorbrggen, J. Buhmann, J. Lange, C. von der Malsburg, R. Wrtz, and W. Konen, "Distortion invariant object recognition in the dynamic link architecture", *IEEE Trans. on Computers*, vol. 42, no. 3, pp. 300-311, March 1993.

[29] A. Tefas, C. Kotropoulos and I. Pitas, "Using support vector machines to enhance the performance of elastic graph matching for frontal face authentication", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 23, no. 7, pp. 735-746, July 2001.

[30] K.-A. Toh, J. Xudong, and W.-Y. Yau, "Exploiting global and local decisions for multimodal biometrics verification", *IEEE Trans. on Signal Processing*, vol. 52, no. 10, pp. 3059-3072, 2004.

[31] R. McGill, J.W. Tukey, and W.A. Larsen, "Variations of boxplots", *The American Statistician*, vol. 32, pp.12-16, 1978.

[32] K. Messer *et. al.*, "Face verification competition on the XM2VTS database", in *Proc. 4th Int. Conf. Audio-and-Video-Based Biometric Person Authentication*, Guildford, UK, June 9-11, 2003.

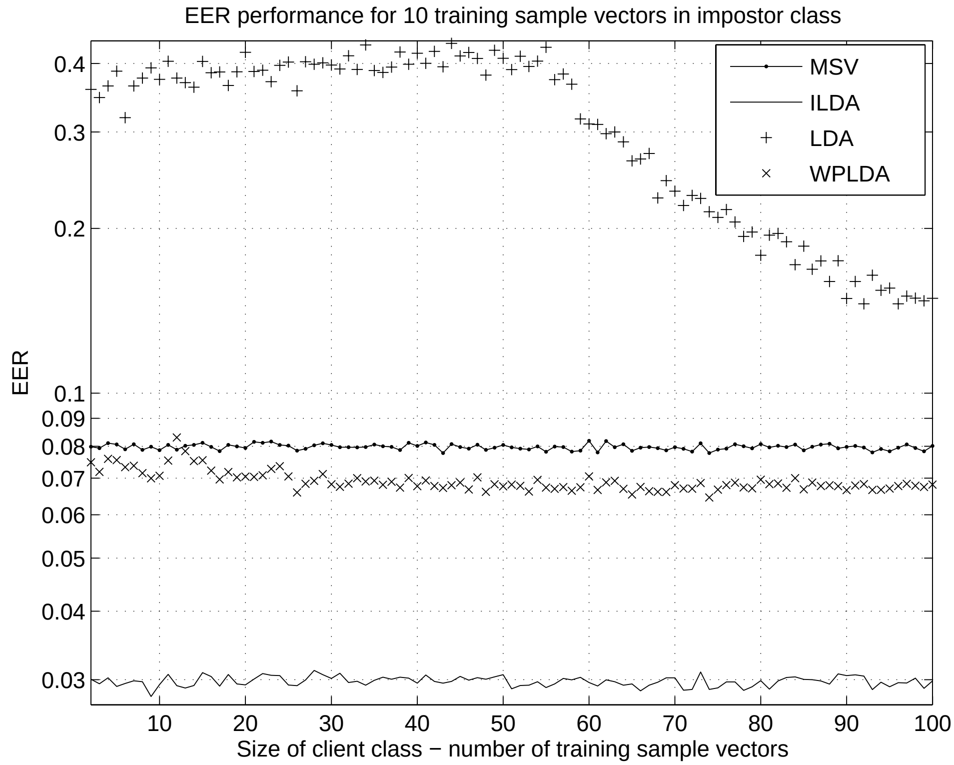

Figure 1: *EER* when varying the number of client sample vectors selected from class C_3 and for 10 impostor training vectors selected from class I. These sample vectors are used to train LDA and WPLDA whereas ILDA indicates the ideal performance and MSV the performance before training is used.

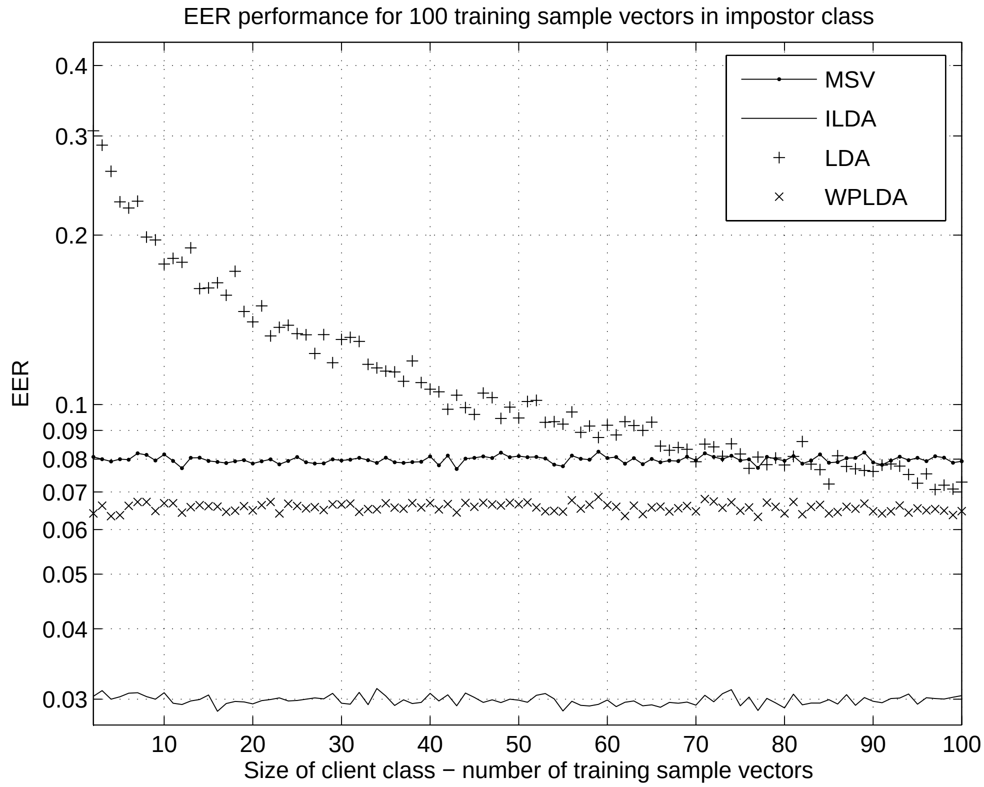

Figure 2: *EER* when varying the number of client sample vectors selected from class C_3 and for 100 impostor training vectors selected from class I. These sample vectors are used to train LDA and WPLDA whereas ILDA indicates the ideal performance and MSV the performance before training is used.

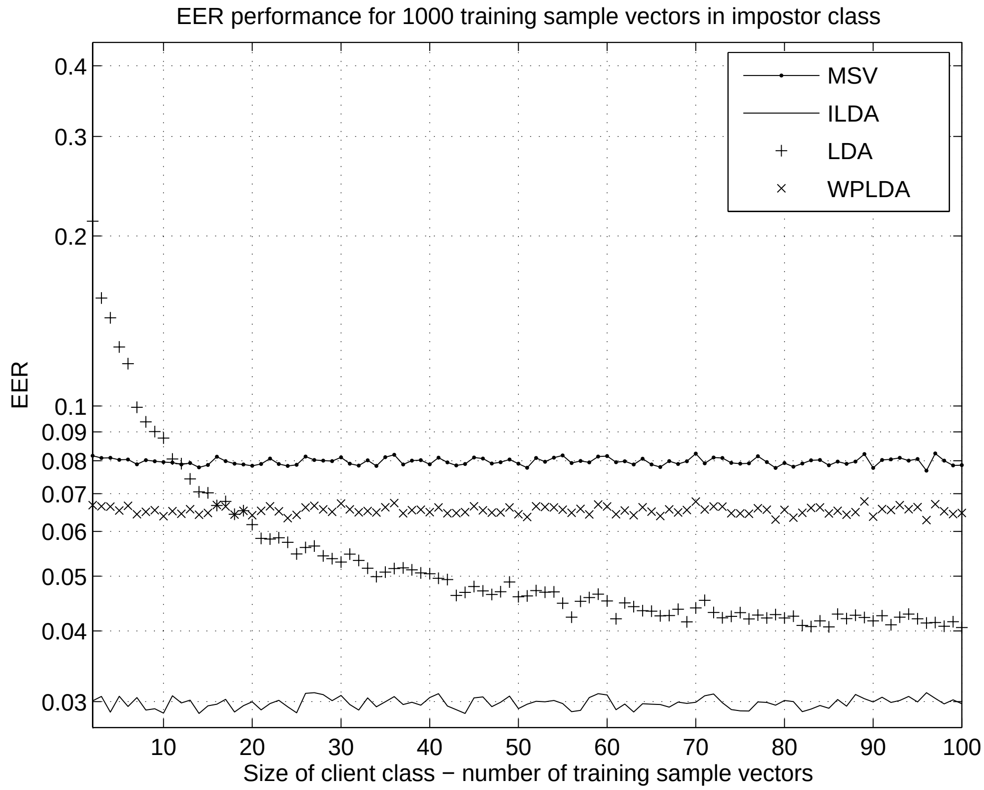

Figure 3: *EER* when varying the number of client sample vectors selected from class C_3 and for 1000 impostor training vectors selected from class I. These sample vectors are used to train LDA and WPLDA whereas ILDA indicates the ideal performance and MSV the performance before training is used.

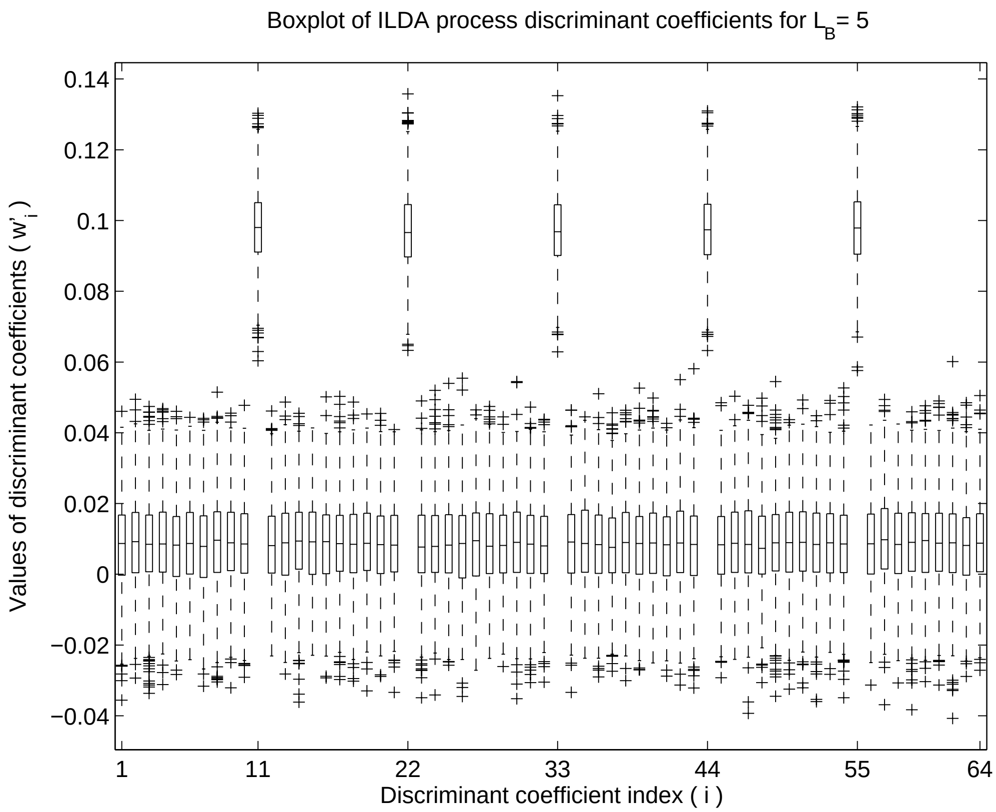

Figure 4: Boxplot of 1000 discriminant coefficient sets of ILDA trained by 1000 impostor (class I) and 6 client (class C_3) vectors. The 'most discriminant' coefficients are indexed at 11, 22, 33, 44 and 55 on the *x*-axis.

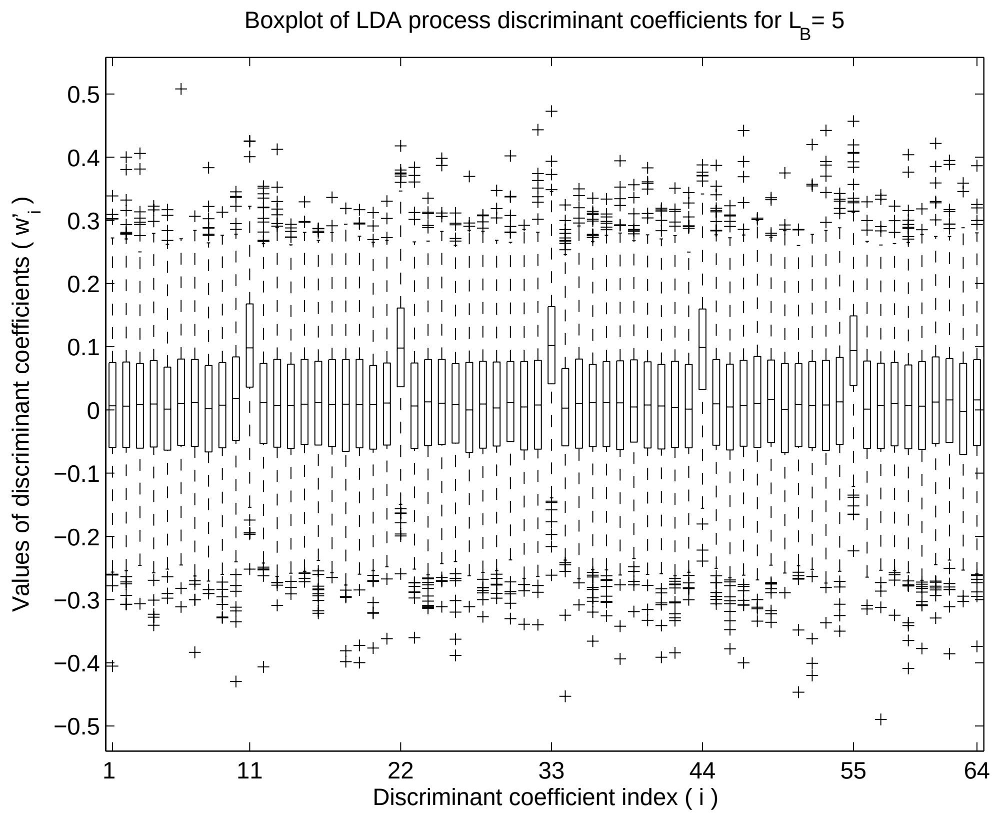

Figure 5: Boxplot of 1000 discriminant coefficient sets of LDA trained by 1000 impostor (class I) and 6 client (class C_3) vectors. The 'most discriminant' coefficients are indexed at 11, 22, 33, 44 and 55 on the *x*-axis.

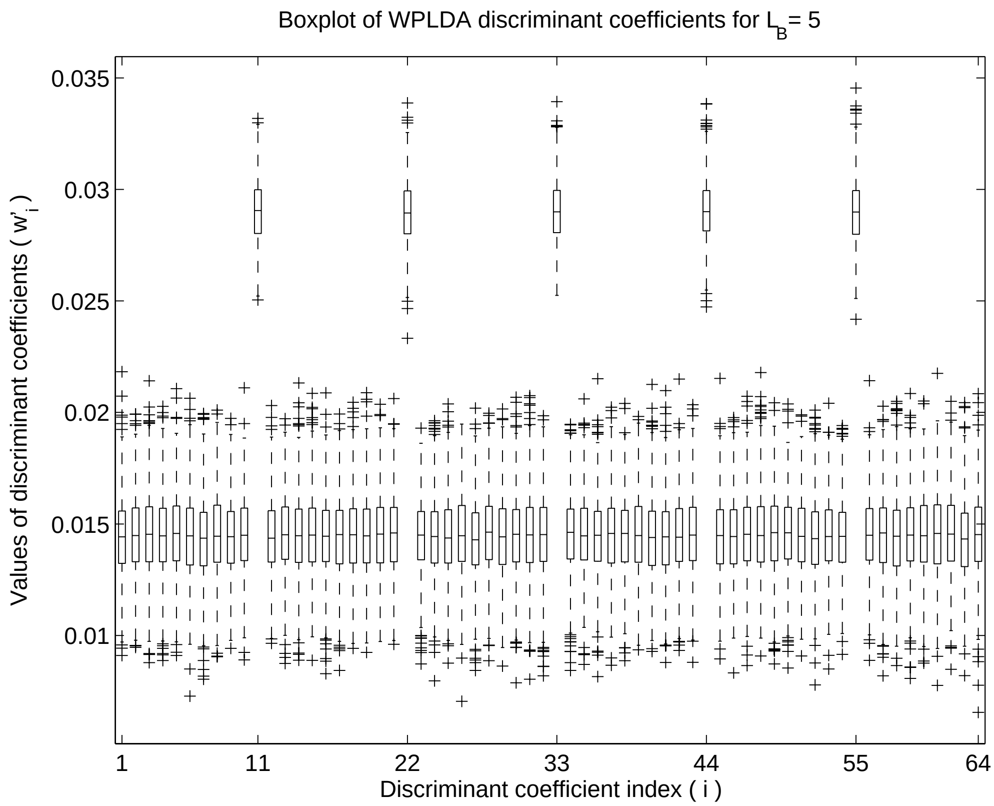

Figure 6: Boxplot of 1000 discriminant coefficient sets of WPLDA trained by 1000 impostor (class I) and 6 client (class C_3) vectors. The 'most discriminant' coefficients are indexed at 11, 22, 33, 44 and 55 on the *x*-axis.

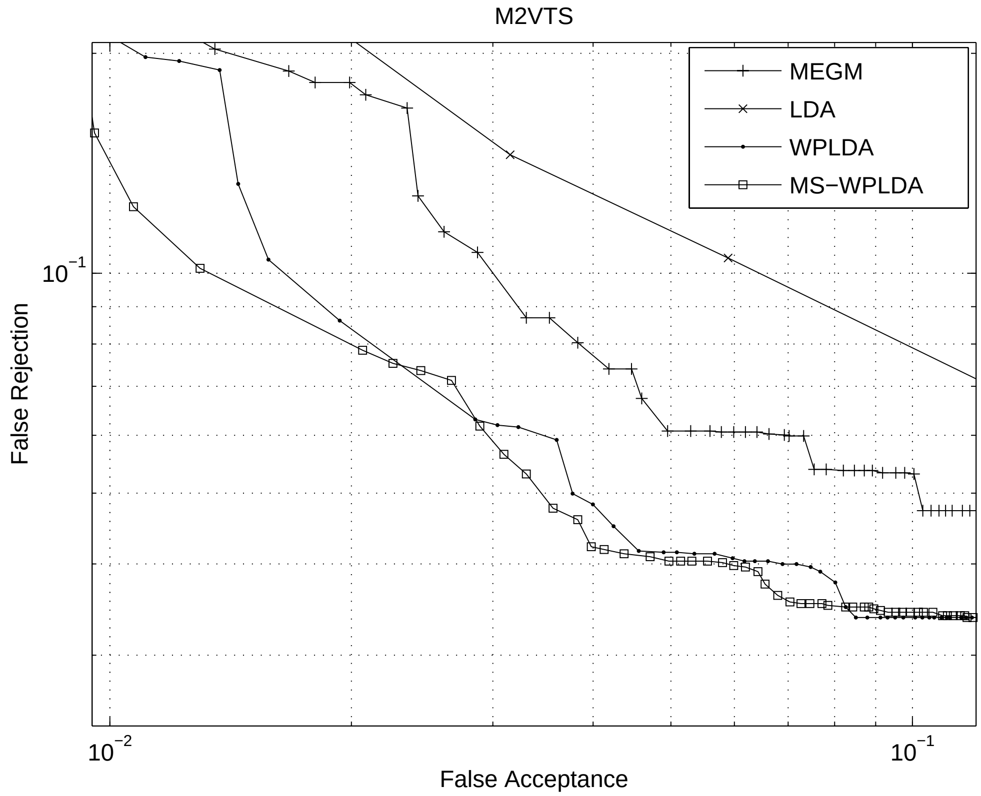

Figure 7: ROC curve of M2VTS experiments that are carried out under the *Brussels* protocol. The relevant training procedure for LDA, WPLDA, and MS-WPLDA uses 6 client similarity vectors and 210 impostor similarity vectors. MEGM indicates the classification performance before training.

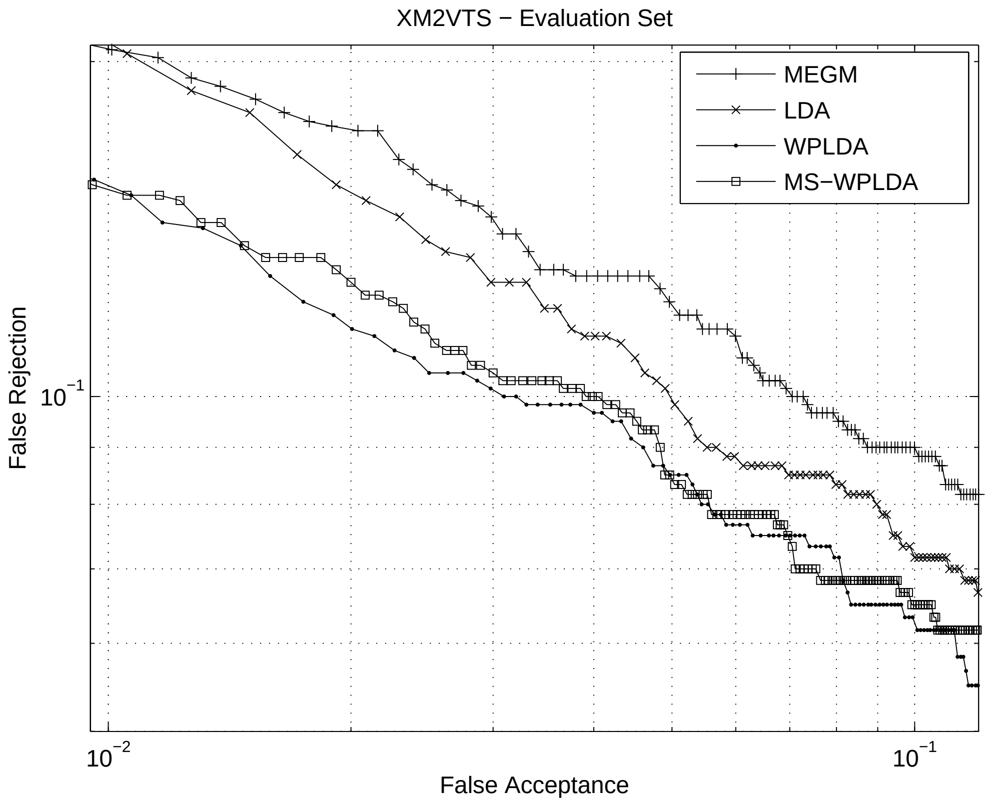

Figure 8: ROC curve of XM2VTS – evaluation set experiments that are carried out under the and *Lausanne* protocol. The relevant training procedure for LDA, WPLDA, and MS-WPLDA uses 6 client similarity vectors and 1791 impostor similarity vectors. MEGM indicates the classification performance before training.

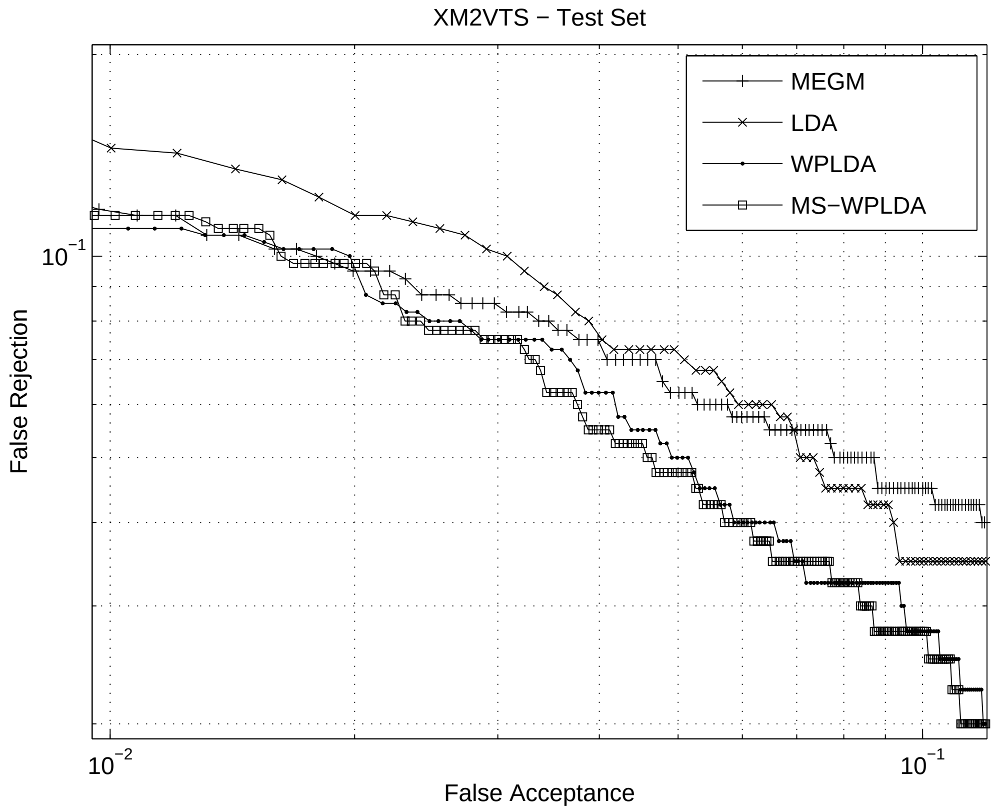

Figure 9: ROC curve of XM2VTS – test set experiments that are carried out under the and *Lausanne* protocol. The relevant training procedure for LDA, WPLDA, and MS-WPLDA uses 6 client similarity vectors and 1791 impostor similarity vectors. MEGM indicates the classification performance before training.

Table I: Mean *EER* and mean H for 1000 independent runs for 6 clients and 1000 impostors.

L_B	*EER*	*EER*	*EER*	*EER*	H	H	H
	ILDA	MSV	LDA	WPLDA	ILDA	LDA	WPLDA
0	0.0961	**0.0989**	0.2578	**0.0989**	-	-	-
5	0.0300	0.0798	0.1142	**0.0652**	3.7908	0.4580	**4.5505**
15	0.0255	0.0486	0.1032	**0.0318**	1.0739	0.1321	**4.7119**
25	0.0247	0.0306	0.1008	**0.0272**	0.3986	0.0489	**1.6225**
32	0.0245	**0.0260**	0.1018	**0.0260**	**0.0959**	0.0123	0.0077

Table II: EER of the various EGM methods on M2VTS and XM2VTS data.

Experiment	*EER* (%)				
	MEGM	LDA	RLDA	WPLDA	MS-WPLDA
M2VTS	6.06	8.94	6.86	4.37	**4.17**
XM2VTS Evaluation Set	9.01	8.22	7.50	7.37	**7.06**
XM2VTS Test Set	5.75	5.96	8.25	4.99	**4.74**