METHODS FOR IMPROVING DISCRIMINANT ANALYSIS FOR FACE AUTHENTICATION

Marios Kyperountas, Anastasios Tefas, Ioannis Pitas

Aristotle University of Thessaloniki, Department of Informatics
Box 451, GR-54124 Thessaloniki, Greece, email: pitas@zeus.csd.auth.gr

ABSTRACT

A novel algorithm that can be used to boost the performance of face authentication methods that utilize Fisher's criterion is presented. The algorithm is applied to 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 locally linear discriminant models is used in order to calculate discriminant weights in a more accurate way than the traditional linear discriminant analysis (LDA) methodology. Additionally, defective discriminant coefficients are identified and re-estimated. The second phase defines proper combinations for person-specific matching scores and describes an outlier removal process that enhances the classification ability. Our technique was tested on the M2VTS and XM2VTS frontal face databases. Experimental results indicate that the proposed framework greatly improves the authentication algorithm's performance.

1. INTRODUCTION

Linear discriminant analysis is an important statistical tool for recognition, verification, and in general classification applications. In many cases, however, and in particular when face data is used, there is insufficient data available so as to carry out the LDA process in a statistically proper manner. In face authentication 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 is a client) or not (meaning the test face is an impostor). In this type of problems, Fisher's linear discriminant is not expected to be able to discriminate well between face pattern distributions that are in many cases highly nonlinear (i.e. they cannot be separated linearly), unless a sufficiently large training set is available. More specifically, in face recognition or authentication systems LDA-based approaches often suffer from the SSS problem where the dimensionality of the samples is larger than the number of available training samples [1]. In fact, when this problem becomes severe, traditional LDA actually degrades the classification performance and shows poor generalization ability.

This work is funded by the research project BioSec IST-2002-001766 (Biometrics and Security), under IST priority of the 6th Framework Programme of the European Community.

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 through LDA applied in a lower-dimensional PCA subspace so as to discard the null space of the within-class scatter matrix of the training data set [2]. However, it has been shown [3] that significant discriminatory information is contained in the discarded space and alternative solutions have been sought. Specifically, in [4] a direct-LDA algorithm is presented, which 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. More recently, the authors in [1] form 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 slightly better than random guessing to a more accurate learner.

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 authentication, 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 and creating subsets of feature vectors with small dimensionality, and applying discriminant analysis on each subset. The resulting discriminant weights are normalized so as to provide the overall discriminatory solution. This process gives direct improvements to the two aforementioned problems as the non-linearity between the data pattern distributions is now restricted while the reduced dimensionality also helps mend the SSS problem. Remaining high nonlinearities between corresponding subsets lead to a number of discriminant coefficients being badly estimated due to the small training set. These coefficients are identified and re-estimated in an iterative fashion, if needed. In the second stage the set of matching scores that correspond to each person's reference photos 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 matching scores.

The proposed methodology was tested on two well-established frontal face databases, M2VTS and XM2VTS. The experimental results are presented and analyzed in section 4 in order to assess the performance of the proposed methodology.

2. PROBLEM STATEMENT

A widely known face verification algorithm is elastic graph matching [5]. 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:

$$\mathbf{j}(\mathbf{x}) = \left(\hat{f}_1(\mathbf{x}), \ldots, \hat{f}_{\mathrm{M}}(\mathbf{x}) \right) \quad (1)$$

where $\hat{f}_i(\mathbf{x})$ denotes the output of a local operator applied to image f at the i^{th} scale or the i^{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 described in [6]. Let the superscripts r and t denote a test and a reference person (or grid) respectively. The L_2 norm between the feature vectors at the l^{th} grid node is used as a (signal) similarity measure:

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

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

$$\mathbf{c}_t = [C_1, \ldots, C_L]^{\mathrm{T}}, \quad (3)$$

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:

$$D(t,r) = \mathbf{I}^{\mathrm{T}} \mathbf{c}_t, \quad (4)$$

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

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 database. As a result, we end up with T matching error values, or 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.

3. BOOSTING LINEAR DISCRIMINANT ANALYSIS

Let $\mathbf{m}_C$ and $\mathbf{m}_I$ denote the sample mean of the class of matching 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 matching vectors that belong to these two classes and N be their sum, or the total number of matching vectors. Let $\mathbf{S}_W$ and $\mathbf{S}_B$ be the within-class and between-class scatter matrices, respectively [7]. Suppose that we would like to linearly transform the matching vectors as such:

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

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):

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

For the two-class problem, as is the case of face authentication, Fisher's linear discriminant, $\mathbf{w}_{r,0}^{\mathrm{T}} \mathbf{c}_t$, which is essentially a specific choice of direction of the data down to one dimension, provides the vector that maximizes (6) and is given by:

$$\mathbf{w}_{r,0} = \mathbf{S}_W^{-1} (\mathbf{m}_I - \mathbf{m}_C). \quad (7)$$

3.1. Locally Linear Discriminant Analysis Model

Our experiments revealed that the traditional Fisher's linear discriminant process not only performed poorly, but also degraded the classification capability of the face authentication algorithm when training data from the M2VTS database was used. That is, (4) provided a much better solution than (5) after traditional LDA was used to determine the values of $\mathbf{w}_r$. This statistical malady can be attributed to the matching error vectors not being linearly separable and to the insufficient availability of mostly client matching error vectors, with respect to the dimensionality of each vector. Moreover, using nonlinear separating surfaces can lead to overtraining and thus to lower performance. Specifically, the number of client matching error 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 (L), or number of grid nodes, at 64. The value of N_I while training the algorithm using M2VTS data was set at 210 and using XM2VTS data at 1791. All these numbers are compatible with the training protocols of each database for authentication purposes – the *Brussels* protocol, which is used and described in [6], was applied to the M2VTS database and *Configuration I* of the *Lausanne* protocol [8] to the XM2VTS database training and testing procedures. The two aforementioned problems are related since a larger training data set can help deal more efficiently with the nonlinearity problem.

The first thing that is done is to provide better estimation to Fisher's linear discriminant by redefining (7) to:

$$\mathbf{w}_{r,0} = \mathbf{S}_W^{-1} \left(\mathbf{m}_I \frac{\mathrm{N}_I}{\mathrm{N}} - \mathbf{m}_C \frac{\mathrm{N}_C}{\mathrm{N}} \right), \quad (8)$$

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 are discarded. Simply, each of the L' remaining nodes must satisfy:

$$\mathbf{m}_I(r,l) \geq \mathbf{m}_C(r,l). \quad (9)$$

In order to give remedy to the SSS problem each matching vector with dimensionality L' is broken down to P smaller dimensionality vectors, each one of length M, where $\mathrm{M} \leq (\mathrm{N}_C - 1)$, thus forming P subsets. The more statistically independent the subsets are among one another, the better the discriminant analysis is expected to be. Our tests revealed that the optimum value for M is 4. 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:

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

where $p = 1,\ldots,P$ is the subsets' index. This normalization step enables the proper merging of all weight vectors to a single column weight vector, $\mathbf{w}'_{r,0}$, as such:

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

3.2. Re-estimating the Negative 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. Thus, and since matching, or error, data are always positive, $\mathbf{w}'_{r,0}$ 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, it was observed that on an 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 when the M2VTS training set was used. 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. The locally linear discriminant analysis model that was introduced in 3.1 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 the combination of large nonlinearities between the distribution patterns of corresponding subsets and the lack of a sufficiently large training sample space.

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

Case 1: $Q_p \le 1.5 \cdot \mathrm{M}$

All negative weights are set to zero and no further processing is required. The factor 1.5 is used to indicate that if the number of values in the final subset 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.

Case 2: $Q_p > 1.5 \cdot \mathrm{M}$

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' subsets where each subset again holds M discriminant variables. In turn, P' separate Fisher linear discriminant operations are carried out by following (8) and each of the weight vectors produced is normalized by following (10).

Successively, all positive weights from all P' subsets are collected and set as the final multipliers of $\mathbf{c}_t$, or 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 1* to apply. Indicatively, it is stated that during the training stages of the M2VTS database 3 to 5 iterations are usually required when $\mathrm{M} = L'$ 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.

3.3. Weighting the Classification Scores

The protocols that the authentication algorithm was tested under specified that a test person could be classified to be an impostor or a client by using three, $T = 3$, different photos of the reference face; thus, three tests are carried out. As a result, three classification scores are available for each individual, i.e. $v_{r,1}$, $v_{r,2}$ and $v_{r,3}$. Traditionally, the test person is classified as a client if the minimum value out of the three, i.e. $v_{r,1}$, is below a set threshold, and as an impostor if it is above that threshold. In this work, training data are used once again to derive person specific weights for the T scores. The motive behind this process is that ideally all three scores should contribute to the final classification decision as in certain cases the impostor's photo that corresponds to a minimum score may have accidentally - e.g. due to a particular facial expression - had close similarity to a certain reference photo. In such a case, the remaining two reference images can be used in an effort to repair the false classification decision. Now the problem becomes:

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

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, are forced to set $v_{r,3}$ to zero, where this would be the weight that corresponds to the largest matching error score, and set $T = 2$ in (12). Fisher's modified linear discriminant (8) is applied and the two weights are found.

A much larger number of impostor, rather than client, matching scores is available in the training set of each database which 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 matching scores, which is as many as the number of training samples in 3.1. Thus, an outlier removal process is incorporated where the minimum impostor matching scores in the training set of each reference person, i.e. all $v_{r,1}$ scores that correspond to impostor matches, are ordered and the smallest 4% of these values is discarded. As a result, the linear discriminant process gives a more accurate separation that helps increase the classification performance.

4. EXPERIMENTAL RESULTS

The discriminant coefficient vectors $\mathbf{w}'$ derived by the processes described in 3.1 and 3.2 have been used to weigh the raw matching error vectors $\mathbf{c}$ that are provided by the morphological elastic graph matching applied to frontal face authentication, based on the algorithm described in [6]. Moreover, the procedure in 3.3 was used to calculate a more accurate matching score for each tested individual. The training, evaluation and test sets of the XM2VTS database were processed under 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 and 112,000 (70 impostors x 8 shots x 200 clients) impostor claims

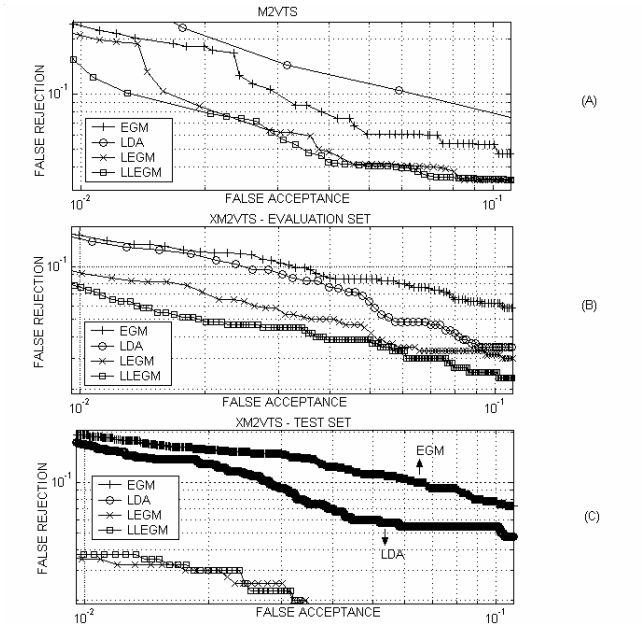

Figure 1: ROC curves. **A**: M2VTS; **B**: XM2VTS – evaluation set; **C**: XM2VTS – test set.

Method	*EER* (%)		
	M2VTS	XM2VTS – Evaluation Set	XM2VTS – Test Set
EGM	6.06	7.33	8.51
LDA	8.94	5.32	5.75
LEGM	4.37	4.67	2.51
LLEGM	4.17	3.90	2.46

Table 1: Evaluation results for each process.

for the test set. For the M2VTS database the *Brussels* protocol was implemented that employs the 'leave-one-out' and 'rotation' estimates, and 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. The M2VTS data were normalized so that the feature vectors would have zero mean and unit variance. Thresholds from the training process of each database were used to evaluate the authentication results, except for the evaluation of the XM2VTS test set, where thresholds from the evaluation process were used, as [8] suggests.

Let us call the combination of the morphological elastic graph matching, EGM, and the weighting approach that makes up for the first phase of the proposed algorithm, as is described in Subsections 3.1 and 3.2, as LEGM. Moreover, let LLEGM be the second phase of the algorithm that is applied on LEGM and is described in Subsection 3.3. In order to evaluate the performance of these methods the *False Acceptance* (*FAR*) and *False Rejection* (*FRR*) *rate* measures are used. Figure 1-A shows a critical region of the ROC curves for the raw EGM data using (4), classical LDA (7) applied on the raw EGM data, LEGM and LLEGM evaluated on the M2VTS database. Figure 1-B shows the same corresponding ROC curves when the algorithms were evaluated on the XM2VTS evaluation set and Figure 1-C the corresponding ones for the XM2VTS test set. Results are presented in logarithmic scales. In addition, Table 1 shows the equal error rates (*EER*) for each algorithm, a common face authentication evaluation measure that is specified as the point where *FAR* and *FRR* are identical.

When M2VTS data is used, the traditional LDA algorithm degrades the classification rate, having a poor generalization ability which stems from the largely inadequate, in terms of size, training set that was available. The proposed algorithm provides the most dramatic improvement to the XM2VTS test set experiments – the outlier removal process was bypassed on this specific set as it slightly weakened the performance. Furthermore, the evaluation tests on the two databases show that for both *FAR* and *FRR*, LEGM is indisputably a better performer than either EGM or LDA while LLEGM almost always provides additional improvement to the classification ability of LEGM.

5. CONCLUSION

A novel methodology is proposed in this paper that provides general solutions for LDA-based algorithms that encounter problems relating to high nonlinearity between the data pattern distributions, small training sets and to the SSS problem in particular. This methodology was tested on two well-established databases under their standard protocols for evaluating face authentication algorithms. Results indicate that the processes described in this paper boost the performance of the authentication algorithm significantly (31.2%, 46.8% and 71.1% 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.

6. REFERENCES

[1] J. Lu, K.N. Plataniotis, and A.N. Venetsanopoulos, "Boosting linear discriminant analysis for face recognition", in *Proc. IEEE Int. Conf. on Image Processing (ICIP'03),* vol. 1, pp. I - 657-60, Barcelona, Spain, 14-17 Sep., 2003.

[2] 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.

[3] L-F Chen, *et al*., "A new LDA-based face recognition system which can solve the small sample size problem", *Pattern Recognition*, vol. 33, pp. 1713–1726, 2000.

[4] 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.

[5] M. Lades, *et al*., "Distortion invariant object recognition in the dynamic link architecture", *IEEE Trans. on Computers*, vol. 42, no. 3, pp. 300–311, March 1993.

[6] 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.

[7] 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.

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