USING SUPPORT VECTOR MACHINES FOR FACE AUTHENTICATION BASED ON ELASTIC GRAPH MATCHING

Anastasios Tefas, Constantine Kotropoulos and Ioannis Pitas

Department of Informatics, Aristotle University of Thessaloniki
Box 451, Thessaloniki 540 06, GREECE, {tefas,costas,pitas}@zeus.csd.auth.gr

ABSTRACT

In this paper, a novel method for enhancing the performance of elastic graph matching in face authentication is proposed. Our objective is to weigh the local matching errors at the nodes of an elastic graph according to their discriminatory power. We propose a novel approach to discriminant analysis that re-formulates Fisher's Linear Discriminant ratio to a quadratic optimization problem subject to inequality constraints by combining statistical pattern recognition and support vector machines. The method is applied to frontal face authentication on the M2VTS database.

1. INTRODUCTION

Many face recognition techniques have been developed for more than two decades whose principles span several disciplines, such as image processing, pattern recognition, computer vision and neural networks [1]. The increasing interest in face recognition is mainly driven by application demands, such as nonintrusive verification for credit cards and automatic teller machine transactions, nonintrusive access-control to buildings, identification for law enforcement, etc.

A well-known approach to face recognition and authentication is the so-called dynamic link architecture (DLA), a general object recognition technique, that represents an object by projecting its image onto a rectangular elastic grid where a Gabor wavelet bank response is measured at each node [2]. Recently, a variant of dynamic link architecture based on multiscale dilation-erosion, the so-called *morphological dynamic link architecture* (MDLA), has been proposed and tested for face authentication [3].

This paper addresses the derivation of optimal coefficients that weigh the local matching errors determined by the elastic graph matching procedure at each grid node. We propose to weigh the local matching errors by a novel approach that combines statistical pattern recognition (i.e., discriminant analysis) [4] and Support Vector Machines [5]. Our approach re-formulates Fisher's Linear Discriminant ratio to a quadratic optimization problem subject to inequality constraints. Linear and nonlinear Support Vector Machines are then constructed to yield the optimal separating hyperplanes.

2. PROBLEM STATEMENT

A widely known face recognition algorithm is the elastic graph matching [2]. The method is based on the analysis of a facial image region and its representation by a set of M local descriptors (i.e., a feature vector) 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-th scale or at the i-th pair of scale and orientation and $\mathbf{x}$ defines the pixel coordinates. The grid nodes are either distributed evenly over a rectangular image region or they are placed on certain facial features (e.g., nose, eyes, etc.) called fiducial points. In both cases, a face/facial feature detection algorithm is needed.

Let the superscripts t and r 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, i.e., $C_v(\mathbf{j}(\mathbf{x}_l^t), \mathbf{j}(\mathbf{x}_l^r)) = \|\mathbf{j}(\mathbf{x}_l^t) - \mathbf{j}(\mathbf{x}_l^r)\|$. The objective in elastic graph matching is to find the set of test grid node coordinates $\{\mathbf{x}_l^t,\ l \in \mathcal{V}\}$ that minimizes the cost function:

$$D(t,r) = \sum_{l \in \mathcal{V}} C_v(\mathbf{j}(\mathbf{x}_l^t), \mathbf{j}(\mathbf{x}_l^r))$$
$$\text{subject to } \mathbf{x}_l^t = \mathbf{x}_l^r + \mathbf{s} + \boldsymbol{\delta}_l, \quad \|\boldsymbol{\delta}_l\| \leq \delta_{\max} \tag{2}$$

where $\mathbf{s}$ denotes a global translation of the graph, $\boldsymbol{\delta}_l$ is a local perturbation and $\delta_{\max}$ controls the rigidity/plasticity of the graph.

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

$$\mathbf{c}_t = \begin{bmatrix} C_v(\mathbf{j}(\mathbf{x}_1^t), \mathbf{j}(\mathbf{x}_1^r)) \\ C_v(\mathbf{j}(\mathbf{x}_2^t), \mathbf{j}(\mathbf{x}_2^r)) \\ \vdots \\ C_v(\mathbf{j}(\mathbf{x}_L^t), \mathbf{j}(\mathbf{x}_L^r)) \end{bmatrix} \tag{3}$$

where L is the cardinality of $\mathcal{V}$. Hereafter, $\mathbf{c}_t$ is referred as the matching vector between the test person t and the reference person r. Using matrix notation, (2) is rewritten as $D(t,r) = \mathbf{1}^T \mathbf{c}_t$ where $\mathbf{1}$ is an $L \times 1$ vector of ones. That is, the classical elastic graph matching treats uniformly all local matching errors $C_v(\mathbf{j}(\mathbf{x}_l^t), \mathbf{j}(\mathbf{x}_l^r))$. We would like to weigh the local matching errors, i.e., to compute a weighted distance measure:

$$D'(t,r) = \mathbf{w}_r^T \mathbf{c}_t \tag{4}$$

where $\mathbf{w}_r$ is an appropriate vector of weights. Let us denote by $\mathcal{S}_r$ the class of matching vectors that belong to the reference person. Let also $\mathcal{S}$ denote the set of matching errors of the training set. Throughout the paper we study a two-class problem, namely, to separate efficiently all matching vectors that are attributed to a client (i.e., the reference person r) from the matching vectors that belong to anybody else (i.e., the class of $\mathbf{c}_t \in (\mathcal{S} - \mathcal{S}_r)$, which constitutes the set of impostors for client r). The most known criterion is to choose $\mathbf{w}_r$ so that the ratio of the trace of the between-class scatter matrix over the trace of the within-class scatter matrix of the transformed matching vectors is maximized. Since, in our case, the transformed matching vector is merely the scalar $\mathbf{w}_r^T \mathbf{c}_t$, the optimization criterion is simplified to the ratio of between-class and within-class variance, i.e.:

$$J(\mathbf{w}_r) = \frac{\mathbf{w}_r^T \mathbf{S}_B \mathbf{w}_r}{\mathbf{w}_r^T \mathbf{S}_W \mathbf{w}_r}. \tag{5}$$

This is the so-called *Fisher's discriminant ratio*. It can easily be verified that the criterion:

$$\begin{aligned} &\text{minimize} && \mathbf{w}_r^T \mathbf{S}_W \mathbf{w}_r && (6)\\ &\text{subject to} && \mathbf{w}_r^T (\mathbf{m}_I - \mathbf{m}_C) \geq \mathbf{1}^T (\mathbf{m}_I - \mathbf{m}_C) && (7) \end{aligned}$$

has an interpretation that agrees with that of Fisher's linear discriminant ratio (5). Indeed, (6) minimizes the within-class variance, while the constraint inequality (7) has the following interpretation: The difference between the class centers (i.e., the average distance measure over client claims $E\{D'(t,r) \mid \mathbf{c}_t \in \mathcal{S}_r\}$ and the average distance measure over impostor claims $E\{D'(t,r) \mid \mathbf{c}_t \in (\mathcal{S} - \mathcal{S}_r)\}$) is not reduced after linear weighting. In the following, we shall elaborate the criterion (6) and (7). More specifically, we shall impose more than one inequality constraints demanding that the distance measures $D'(t,r)$ related to impostor claims to be linearly separable from the distance measures $D'(t,r)$ related to client claims on the training set.

3. SUPPORT VECTOR MACHINE SOLUTION

Support Vector Machines (SVMs) is a state-of-the-art pattern recognition technique whose foundations are stemming from statistical learning theory [5]. SVM is a learning machine capable of implementing a set of functions that approximate best the supervisor's response with an expected risk bounded by the sum of the empirical risk and the Vapnik-Chervonenkis (VC) confidence, a bound on the generalization ability of the learning machine, that depends on the so-called VC dimension of the set of functions implemented by the machine. Motivated by the fact that SVM training algorithm consists of a quadratic programming problem, we reformulate the criterion of minimizing the within-class variance, which appears in Fisher's linear discriminant ratio, so that it can be solved by constructing the optimal separating hyperplane (linear SVM).

3.1. The Separable Case

Suppose the training data:

$$(\mathbf{c}_1, y_1), \ldots, (\mathbf{c}_N, y_N), \quad \mathbf{c}_t \in \mathbb{R}^L, \tag{8}$$
$$y_t = \begin{cases} 1 & \text{if } \mathbf{c}_t \in (\mathcal{S} - \mathcal{S}_r) \\ -1 & \text{if } \mathbf{c}_t \in \mathcal{S}_r \end{cases}$$

can be separated by a hyperplane:

$$g_{\mathbf{w}_r, b}(\mathbf{c}_t) = \mathbf{w}_r^T \ \mathbf{c}_t - b = 0 \tag{9}$$

with the property:

$$y_t(\mathbf{w}_r^T \mathbf{c}_t - b) - 1 \geq 0 \quad t = 1, \ldots, N \tag{10}$$

where b is a bias term. Let us define the distance $v(\mathbf{w}_r, b; \mathbf{c}_t)$ of a matching vector $\mathbf{c}_t$ from the hyperplane (9) as:

$$v(\mathbf{w}_r, b; \mathbf{c}_t) = \frac{\mid \mathbf{w}_r^T \ \mathbf{c}_t - b \mid}{\|\mathbf{w}_r\|_{\mathbf{S}_W}} = \frac{\mid \mathbf{w}_r^T \ \mathbf{c}_t - b \mid}{(\mathbf{w}_r^T \mathbf{S}_W \mathbf{w}_r)^{1/2}} \tag{11}$$

where the norm of the coefficient vector $\mathbf{w}_r$ is measured with respect to the within-scatter matrix $\mathbf{S}_W$. In our case, the optimal hyperplane is given by maximizing the margin:

$$\begin{aligned} \varrho(\mathbf{w}_r) = &\min_{\mathbf{c}_t \in (\mathcal{S} - \mathcal{S}_r)} v(\mathbf{w}_r, b; \mathbf{c}_t) + \\ &+ \min_{\mathbf{c}_t \in \mathcal{S}_r} v(\mathbf{w}_r, b; \mathbf{c}_t) = \frac{2}{(\mathbf{w}_r^T \mathbf{S}_W \mathbf{w}_r)^{1/2}}. \end{aligned} \tag{12}$$

Equivalently, the optimal hyperplane separates the data so that the within-class variance is minimized. The optimization is subject to the constraint functions (10). For completeness, we mention that the standard SVM would solve the problem [5]:

$$\text{minimize } J_{\text{SVM}}(\mathbf{w}_r) = \mathbf{w}_r^T \ \mathbf{w}_r \text{ subject to (10).} \tag{13}$$

The solution of the optimization problem under study is given by the saddle point of the Lagrangian:

$$L(\mathbf{w}_r, b, \boldsymbol{\alpha}) = \mathbf{w}_r^T \mathbf{S}_W \mathbf{w}_r - \sum_{t=1}^{N} \alpha_t \{y_t(\mathbf{w}_r^T \mathbf{c}_t - b) - 1\} \tag{14}$$

where $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_N)^T$ is the vector of Lagrange multipliers. The Lagrangian has to be minimized with respect to $\mathbf{w}_r$ and b and maximized with respect to $\alpha_t > 0$. The Kuhn-Tucker (KT) conditions [6] imply that:

$$\begin{aligned} &\nabla_{\mathbf{w}_r} L(\mathbf{w}_{r,o}, b_o, \boldsymbol{\alpha}_o) = \mathbf{0} \Leftrightarrow \mathbf{w}_{r,o} = \frac{1}{2}\mathbf{S}_W^{-1} \sum_{t=1}^{N} \alpha_{t,o} y_t \mathbf{c}_t \\ &\frac{\partial}{\partial b} L(\mathbf{w}_{r,o}, b_o, \boldsymbol{\alpha}_o) = 0 \Leftrightarrow \sum_{t=1}^{N} \alpha_{t,o} y_t = 0 \\ &y_t \ (\mathbf{w}_{r,o}^T \mathbf{c}_t - b_o) - 1 \geq 0 \quad t = 1, \ldots, N \\ &\alpha_{t,o} \geq 0 \quad t = 1, \ldots, N \\ &\alpha_{t,o} \ \{y_t \ (\mathbf{w}_{r,o}^T \mathbf{c}_t - b_o) - 1\} = 0 \quad t = 1, \ldots, N. \end{aligned} \tag{15}$$

From the conditions (15), one can see that the weighting vector, we search for, is the linear combination of the matching vectors having nonzero Lagrange multipliers α_t. These matching vectors are the *support vectors* [5]. Putting the expression for $\mathbf{w}_{r,o}$ into the Lagrangian (14) and taking into account the KT conditions, we obtain the Wolf dual functional:

$$\mathcal{W}(\boldsymbol{\alpha}) = \sum_{t=1}^{N} \alpha_t - \frac{1}{4} \sum_{t=1}^{N} \sum_{j=1}^{N} \alpha_t \alpha_j \underbrace{y_t y_j \left(\mathbf{c}_t^T \mathbf{S}_W^{-1} \mathbf{c}_j\right)}_{\mathbf{H}_{tj}} \tag{16}$$

where $\mathbf{H}_{tj}$ is the ij-th element of the Hessian matrix $\mathbf{H}$. The maximization of (16) in the non-negative quadrant of α_t, i.e.:

$$\alpha_t \geq 0 \quad t = 1, \ldots, N \tag{17}$$

under the constraint:

$$\sum_{t=1}^{N} \alpha_t y_t = 0 \tag{18}$$

is equivalent to the optimization problem:

$$\text{minimize } \ \frac{1}{4}\boldsymbol{\alpha}^T \ \mathbf{H} \ \boldsymbol{\alpha} - \mathbf{1}^T \ \boldsymbol{\alpha} \ \text{ subject to (17) and (18).} \tag{19}$$

Having found the non-zero Lagrange multipliers $\alpha_{t,o}$, the optimal separating hyperplane is given by:

$$g(\mathbf{c}) = \text{sgn}\left(\frac{1}{2}\sum_{\alpha_{t,o}>0} y_t\alpha_{t,o}(\mathbf{c}_t^T\mathbf{S}_W^{-1}\mathbf{c}) - b_o\right) \quad (20)$$

where $b_o = \frac{1}{2}\mathbf{w}_{r,o}^T(\mathbf{c}_p + \mathbf{c}_q)$ for any pair of support vectors $\mathbf{c}_p$ and $\mathbf{c}_q$, such that $y_p = 1$ and $y_q = -1$. The weighted distance measure is given by (4).

3.2. The Non-Separable Case

When the matching errors are not linearly separable, we would like to relax the constraints (10) by introducing non-negative slack variables $\xi_t, t = 1, \ldots, N$ [5], such that:

$$\mathbf{w}_r^T\mathbf{c}_t \geq b + 1 - \xi_t \qquad \text{if } y_t = 1 \quad (21)$$

$$\mathbf{w}_r^T\mathbf{c}_t \leq b - 1 + \xi_t \qquad \text{if } y_t = -1 \quad (22)$$

$$\xi_t \geq 0, \qquad t = 1, \ldots, N. \quad (23)$$

The above constraints can be given in a compact notation as:

$$y_t(\mathbf{w}_r^T\mathbf{c}_t - b) + \xi_t - 1 \geq 0 \quad t = 1, \ldots, N. \quad (24)$$

The so-called generalized optimal hyperplane is determined by the vector $\mathbf{w}_{r,o}$ that minimizes the functional:

$$J(\mathbf{w}_r, b, \boldsymbol{\xi}) = \mathbf{w}_r^T\mathbf{S}_W\mathbf{w}_r + Q(\sum_{t=1}^{N}\xi_t)^\sigma, \quad \sigma > 0 \quad (25)$$

where Q is a given value, that defines the cost of constraint violations, subject to $\xi_t \geq 0 \quad t = 1, \ldots, N$. The larger the Q is, the higher penalty to the errors is assigned. The minimization of (25) subject to (24) is a convex programming problem for any integer σ. For $\sigma = 1, 2$, it is a quadratic programming problem. Moreover, the choice $\sigma = 1$ has the advantage that neither ξ_t nor their Lagrange multipliers appear in the Wolfe dual problem [7]. The Lagrangian of the optimization problem is given by:

$$L(\mathbf{w}_r, b, \boldsymbol{\alpha}, \boldsymbol{\xi}, \boldsymbol{\mu}) = \mathbf{w}_r^T\mathbf{S}_W\mathbf{w}_r + Q\sum_{t=1}^{N}\xi_t - \sum_{t=1}^{N}\mu_t\xi_t - \sum_{t=1}^{N}\alpha_t\left\{y_t(\mathbf{w}_r^T\mathbf{c}_t - b) + \xi_t - 1\right\} \quad (26)$$

where μ_t are Lagrange multipliers enforcing the positivity of ξ_t. The KT conditions for the primal Lagrangian (26) are the following:

$$\begin{aligned}
&\nabla_{\mathbf{w}_r}L(\mathbf{w}_{r,o}, b_o, \boldsymbol{\alpha}_o, \boldsymbol{\xi}_o, \boldsymbol{\mu}_o) = \mathbf{0} \Leftrightarrow \mathbf{w}_{r,o} = \frac{1}{2}\mathbf{S}_W^{-1}\sum_{t=1}^{N}\alpha_{t,o}y_t\mathbf{c}_t \\
&\frac{\partial}{\partial b}L(\mathbf{w}_{r,o}, b_o, \boldsymbol{\alpha}_o, \boldsymbol{\xi}_o, \boldsymbol{\mu}_o) = 0 \Leftrightarrow \sum_{t=1}^{N}\alpha_t y_t = 0 \\
&\nabla_{\boldsymbol{\xi}}L(\mathbf{w}_{r,o}, b_o, \boldsymbol{\alpha}_o, \boldsymbol{\xi}_o, \boldsymbol{\mu}_o) = \mathbf{0} \Leftrightarrow \alpha_{t,o} + \mu_{t,o} = Q \\
&y_t(\mathbf{w}_{r,o}^T\mathbf{c}_t - b_o) + \xi_{t,o} - 1 \geq 0 \quad t = 1, \ldots, N \qquad (27) \\
&\alpha_{t,o} \geq 0 \quad t = 1, \ldots, N \\
&\xi_{t,o} \geq 0 \quad t = 1, \ldots, N \\
&\mu_{t,o} \geq 0 \quad t = 1, \ldots, N \\
&\alpha_{t,o}\left\{y_t(\mathbf{w}_{r,o}^T\mathbf{c}_t - b) + \xi_{t,o} - 1\right\} = 0 \quad t = 1, \ldots, N \\
&\mu_{t,o}\,\xi_{t,o} = 0 \quad t = 1, \ldots, N.
\end{aligned}$$

To find the coefficients of the generalized optimal hyperplane $\mathbf{w}_{r,o}$ in (27) one has to find the Lagrange multipliers α_t, $t = 1, \ldots, N$ that maximize the Wolfe dual problem

$$\mathcal{W}(\alpha) = \sum_{t=1}^{N}\alpha_t - \frac{1}{4}\sum_{t=1}^{N}\sum_{j=1}^{N}\alpha_t\alpha_j \underbrace{y_t y_j\left(\mathbf{c}_t\mathbf{S}_W^{-1}\mathbf{c}_j\right)}_{\mathbf{H}_{tj}} \quad (28)$$

$$\text{subject to } \sum_{t=1}^{N}\alpha_t y_t = 0 \quad \text{and} \quad 0 \leq \alpha_t \leq Q. \quad (29)$$

By comparing (28)-(29) and (19) reveals that the objective function (28) and the equality constraint (29) remain unchanged, while the Lagrange multipliers are now upper-bounded by Q. As in the separable case, only some of the Lagrange multipliers α_t are non-zero. These multipliers are used to determine the support vectors. Having determined the support vectors, $\mathbf{w}_{r,o}$ is determined by the first equation in (27) and the weighted distance measure is computed by (4). The equations derived for the optimal separating hyperplane and the bias term in the separable case are valid for the non-separable case as well.

4. NONLINEAR SUPPORT VECTOR MACHINES

Thus far, we have described the case of linear decision surfaces. By examining the training procedure (28)-(29), one may notice that the matching vectors $\mathbf{c}$ appear in quadratic forms $\mathbf{c}_t^T\mathbf{S}_W^{-1}\mathbf{c}_j$. The just described quadratic form can be expressed by an inner product of the form $(\mathbf{S}_W^{-1/2}\mathbf{c}_t)^T(\mathbf{S}_W^{-1/2}\mathbf{c}_j)$, because $\mathbf{S}_W^{-1}$ is a positive definite matrix. To allow for a more complex decision surface, the rotated matching vectors $\mathbf{S}_W^{-1/2}\mathbf{c}_t$, $t = 1, \ldots, N$ are nonlinearly transformed into a high-dimensional feature space by a map to a Hilbert space, $\Phi : I\!\!R^L \mapsto \mathcal{H}$, and then linear separation is done in the Hilbert space $\mathcal{H}$. Hilbert space is any linear space, with an inner product defined that is also complete with respect to the corresponding norm (i.e., any Cauchy sequence of points converges to a point to the space) [7]. It is obvious that the training procedure in $\mathcal{H}$ would depend only on inner products of the form $< \Phi(\mathbf{S}_W^{-1/2}\mathbf{c}_t), \Phi(\mathbf{S}_W^{-1/2}\mathbf{c}_j) >$. If the inner product in space $\mathcal{H}$ had an equivalent kernel in the input space $I\!\!R^L$, i.e.:

$$< \Phi(\mathbf{S}_W^{-1/2}\mathbf{c}_t), \Phi(\mathbf{S}_W^{-1/2}\mathbf{c}_j) >= K(\mathbf{S}_W^{-1/2}\mathbf{c}_t, \mathbf{S}_W^{-1/2}\mathbf{c}_j) \quad (30)$$

the inner product would not need to be evaluated in the feature space, thus avoiding the curse of dimensionality problem. In order (30) to hold, the kernel function has to be a positive definite function that satisfies Mercer's condition [5]. The polynomial kernel $\left(\mathbf{c}_t^T\mathbf{S}_W^{-1}\mathbf{c}_j + 1\right)^p$ for $p = 4$ was used in the experiments reported in the next section.

Nonlinear SVMs yield a higher computational cost than linear SVMs during the test phase. Indeed, in nonlinear SVMs the distance between the reference person r and the test person τ is given by:

$$D(\tau, r) = \frac{1}{2}\sum_{t=1}^{N_s}\alpha_t y_t K(\mathbf{S}_W^{-\frac{1}{2}}\mathbf{c}_t, \mathbf{S}_W^{-\frac{1}{2}}\mathbf{c}_\tau), \quad (31)$$

where N_s denotes the number of support vectors extracted in the training phase, instead of the much simpler distance computed by the linear SVM which can be simplified to (4). In the latter case, the inner product between the optimal weighting vector, found in

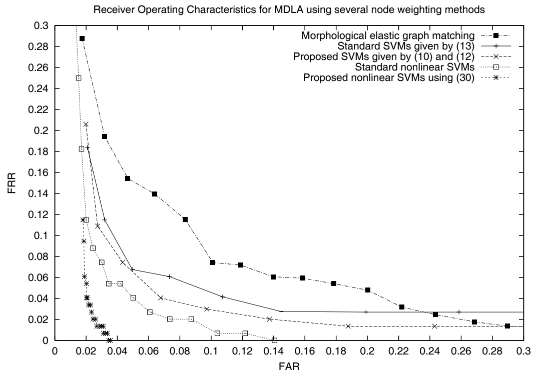

Figure 1: Receiver Operating Characteristics of MDLA for several discriminatory power coefficients.

the training phase, and the test matching vector suffices. This is not the case in (31), where a sum of N_s terms has to be computed. Thus, the test phase of nonlinear SVMs is N_s times slower than that of the linear SVMs.

5. EXPERIMENTAL RESULTS

The optimal coefficient vectors derived by the procedures described in Sections 3 and 4 have been used to weigh the raw matching vectors **c** that are provided by the morphological dynamic link architecture [3], a variant of elastic graph matching, applied to frontal face authentication. Let us denote by weighted MDLA the combination of the SVM weighting approach and the morphological dynamic link architecture. The weighted MDLA has been tested on the M2VTS database. This database contains 37 persons' video data, which include speech consisting of uttering digits and image sequences of rotated heads. Four recordings (i.e., shots) of the 37 persons have been collected. Four experimental sessions have been implemented by employing the "leave-one-out" principle [8]. To apply the proposed methods additional client images are extracted from the database in order to create a large enough set of intra-class distances for each client class. Moreover, additional client images are extracted in order to prevent overfitting during the training caused by the lack of data.

For comparison reasons we have also weighted the raw matching vectors by the coefficient vector determined by the standard SVM algorithm for pattern recognition (13), for both linear and nonlinear separating hyperplanes. By using the coefficient vector derived by the standard SVM to weigh the raw matching vectors an EER equal to 6.4% were obtained. The performance of MDLA was considerably improved by reaching an EER equal to 5.6% when the proposed linear support vector machine that minimizes (14) was applied. The classic nonlinear SVMs resulted an EER equal to 4.5%. The best authentication performance was obtained when the proposed nonlinear SVMs were used reaching an EER of 2.4%. Furthermore, the minimum and the maximum number of support vectors found considering all persons in the training sets that are constructed according to the experimental protocol is given in Table 1 for the standard and the proposed SVMs. It is obvious that the number of SV does not change significantly by using the proposed methods. In any case the number of support vectors is between 10% and 20% of the trained vectors.

The *Receiver Operating Characteristics* (ROC) curves of MDLA for several algorithms are depicted in Figure 1. In the same Figure, the ROC curve for the original MDLA is also plotted for comparison reasons. We can see that the area under the ROC for the proposed methods is much smaller than the initial one. In Table 2, a performance comparison is reported between several face authentication algorithms tested on the same database according to the same protocol. It is clearly seen that the weighted MDLA algorithm attains the best performance.

Table 1: Number of support vectors found.

SVM method	Number of support vectors	
	Minimum	Maximum
standard SVMs	35	43
proposed SVMs	34	40
standard nonlinear SVMs	40	50
proposed nonlinear SVMs	41	51

Table 2: Comparison of equal error rates for several authentication techniques in the M2VTS database.

Authentication Technique	EER (%)
MDLA with discriminating grids	**2.4-5.6**
MDLA	9.2
Gray level frontal face matching [9]	8.5
Discriminant GDLA [10]	6.0-9.2
GDLA [10]	10.8-14.4

6. REFERENCES

[1] R. Chellapa, C.L. Wilson, and S. Sirohey, "Human and machine recognition of faces: A survey," *Proceedings of the IEEE*, vol. 83, no. 5, pp. 705–740, May 1995.

[2] M. Lades, J.C. Vorbrüggen, J. Buhmann, J. Lange, C.v.d. Malsburg, R.P. Würtz, 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.

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

[4] P.A. Devijver and J. Kittler, *Pattern Recognition: A Statistical Approach*, Prentice-Hall International, London, 1982.

[5] V. Vapnik, *The Nature of Statistical Learning Theory*, Springer Verlag, New York, 1995.

[6] R. Fletcher, *Practical Methods of Optimization,* 2nd ed., John Wiley, New York, 1987.

[7] C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining and Knowledge Discovery*, pp. 1–43, 1998.

[8] 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.

[9] S. Pigeon and L. Vandendorpe, "Image-based multimodal face authentication," *Signal Processing*, vol. 69, pp. 59–79, August 1998.

[10] B. Duc, S. Fischer, and J. Bigün, "Face authentication with gabor information on deformable graphs," *IEEE Transactions on Image Processing*, vol. 8, no. 4, pp. 504–516, April 1999.