Learning Support Vectors for Face Authentication: Sensitivity to Mis-Registrations

K. Jonsson[1,2], J. Kittler[1] and J. Matas[1,2]

[1] **CVSSP, University of Surrey**
Guildford, Surrey GU2 5XH, United Kingdom

[2] **CMP, Czech Technical University**
121 35 Prague, Czech Republic

ABSTRACT

The paper studies Support Vector Machines (SVMs) in the context of face authentication. We present an evaluation of the impact of registration errors on the SVM performance. The results obtained by randomly perturbing the groundtruth eye coordinates are compared to those of a fully automatic registration technique based on a robust form of correlation. In the latter method, mis-registrations occur when the optimisation of the similarity function fails and the search terminates in a local maximum. The results of the comparison show that, unless prior knowledge about the problem-specific invariances is incorporated into the SVM training algorithm, the performance is likely to suffer. Nevertheless, we present competitive results obtained on a large face database of 295 subjects and we show that the combined system based on the robust correlation for face registration and the SVMs for verification compares favourably with two other methods based on the dynamic link architecture.

Keywords: support vector machines, face registration, face verification, robust correlation.

1. INTRODUCTION

High-security verification systems based on biometric modalities such as iris, retina and fingerprints have been commercially available for some time. However, one of the most attractive sources of biometric information is the human face since highly discriminative measurements can be acquired without user interaction. The recognition of faces is a well established field of research and a large number of algorithms have been proposed in the literature. Popular approaches include the ones based on Eigenfaces [15], dynamic link matching [9], and active appearance models [3]. These techniques vary in complexity and performance and the choice of algorithm is typically dependent on the specific application. The verification problem, on the other hand, is less explored. Recent examples include [7] in which a robust form of correlation is applied to face authentication.

In [6], we investigated how the SVM performance is affected by different representations and pre-processing techniques. In this paper, we extend this analysis to registration errors and present a fully automatic face registration and verification system. An earlier study of SVMs in face verification has been reported by Phillips [12]. An SVM verification system design was compared with a standard Principal Component Analysis (PCA) face authentication method and the former was found to be significantly better. In this approach the SVM was trained to distinguish between the populations of within-client and between-client difference images respectively, as originally proposed by Moghaddam [11]. This method gives client non-specific support vectors.

In our approach we adopt a client-specific solution which requires learning client-specific support vectors. However, this is not the main distinguishing feature of our work, it only reflects the choice of representation which is different from [12]. In the proposed system, the training images are semi-automatically registered, photometrically normalised and projected into a subspace generated using PCA. The projection coefficients are then used to train one SVM for each client using the other clients as impostors. In the test phase, the images are registered using two different techniques. In the first approach, random displacements are added to the groundtruth eye coordinates generating two-dimensional translations, rotations and scalings. In the second approach, an automatic registration technique based on a robust form of correlation is used. Mis-registrations occur when the optimisation of the similarity function fails and the search terminates in a local maximum.

Since the primary objective of our study was to investigate how the SVM verification performance is affected by registration errors, we did not modify the training procedure. However, recent developments have shown that prior knowledge about the problem-specific invariances (e.g. spatial shifts) can be incorporated into the SVM training process [2]. Their procedure is based on generating virtual support vectors by applying the invariance transformations to the support vectors obtained through standard SVM training. The virtual support vectors are not support vectors in their original sense and they do not correspond to samples in the training set. However, they are located close to the original support vectors and they encode known invariances. These two properties have a positive influence on the classification performance and the re-trained system using the virtual support vectors outperforms standard SVM classifiers.

In our study, the invariance transformations are two-dimensional translations, rotations, and scalings. The sensitivity of the SVMs to noise, spatial shifts and occlusion was investigated in [13]. In this approach, one SVM is trained

for each pair of object classes yielding a total of $\frac{n(n-1)}{2}$ different hyperplanes where n is the number of classes. In the test phase, images are corrupted by adding noise, changing the spatial position of the objects and occluding parts of the images. The degraded objects are then recognised by combining the outputs of the different classifiers using a voting scheme similar to a tennis tournament. Their results show that the SVMs are robust against zero-mean random noise, small displacements and partial occlusions.

Focusing on registration errors and generalising to all rigid transformations, we show that, unless prior knowledge about the problem-specific invariances is incorporated into the SVM training algorithm, the performance is likely to suffer. Nevertheless, we present competitive results obtained on a large face database of 295 subjects and we show that the combined system based on the robust correlation for face registration and the SVMs for verification compares favourably with two other methods based on the dynamic link architecture.

The paper is organised as follows. In the next section, we present the methods used for pre-processing of face images including the different registration techniques and the photometric normalisation. We also describe the representation space into which the normalised images are projected and the decision making tool used for face identity verification. Section 3 introduces the face database used in experimentation and describes the experiments carried out, their objectives and the results obtained. Finally, in Section 4 conclusions are drawn.

2. FACE AUTHENTICATION

Any authentication process involves two basic computational stages. In the first stage a suitable representation is derived with the multiple objective of making the subsequent, decision-making stage computationally feasible, immune to environmental changes during the biometric data acquisition, and effective by providing it only with information which is pertinent to the authentication task. The purpose of the second stage is to accept or reject the identity claim corresponding to a probe biometric measurement. This is basically a two-class pattern recognition problem. In the following subsections we introduce the methods adopted for the design of each of these two stages in the context of the face authentication study pursued in this paper.

2.1. Representation of faces

The first step in the face representation process involves image pre-processing in order to establish correspondence between the face images to be compared. The aim of our study was to evaluate the sensitivity of the SVM approach to registration errors and to meet this objective we implemented two different approaches: a semi-automatic technique used as a baseline method for comparison and a fully automatic approach providing realistic errors. Following the registration, the images are photometrically normalised and projected into a coordinate system which facilitates the decision making process computationally and hopefully emphasises the important attributes for face verification.

Semi-automatic registration: The baseline method for face registration is based on manually localised eye positions. Four parameters computed from the eye coordinates (rotation, scaling and translation in the horizontal and vertical directions) are used to crop the face part from the original image and scale it to any desired resolution. This approach provides the groundtruth for our evaluation by allowing us to separate the issues of localisation and verification (compare with the FERET face recognition test [14]).

Fully automatic registration: Automatic alignment of the face images is achieved using a robust form of correlation [7]. Given an affine transformation $\vec{a}$, the error function expressing the intensity difference between a pixel s in the model image I_m and its projection in the probe image I_p is defined as

$$\epsilon(s, \vec{a}) = I_m(s) - I_p(T_{\vec{a}}(s))$$

where $T_{\vec{a}}$ denotes a geometric transformation function. In the current implementation, $T_{\vec{a}}$ is based on a 6-parameter model incorporating translation, rotation and non-uniform scaling. The score function used to evaluate a match between the transformed model image and the probe image is

$$S(\mathcal{R}, \vec{a}) = \frac{1}{|\mathcal{R}| \cdot \rho_{\max}} \sum_{s \in \mathcal{R}} \rho(\epsilon(s, \vec{a}))$$

where ρ denotes a robust kernel. In words, this function is the average percentage of the maximum kernel response taken over some region R typically obtained by segmenting the model image. Note that this can be done off-line and there is no need for segmenting the probe image. Possible kernel functions are the Huber Minimax and the Hampel (1,1,2) [5].

The optimum of the score function is found using a gradient-based optimisation technique [7]. To meet the real-time requirements of the verification scenario, we apply this search method to each level in a Gaussian pyramid. The estimate obtained on one level is used to initialise the search at the next level. In addition to the speed-up, this multi-resolution search also has the benefit of removing local optima from the search space effectively improving the convergence characteristics of the method.

In the experiments reported in Section 3, we initialise the search at several different spatial positions and we match each probe image with a set of client models. The transformation corresponding to the highest score is then selected and used for the registration of the probe. This procedure reduces the number of mis-registrations and avoids the dependency on single models (typically referred to as 'golden templates') which might be unsuitable for registration.

Photometric normalisation: As the focus of the paper is on the accuracy of the registration and how it affects the SVM performance, we have tried to eliminate the dependency of our experiments on other processes which may lack robustness. For this reason, the cropped images were photometrically normalised. The procedure is based on flattening the distribution of image intensities using histogram equalisation. The reader is referred to [6] for an evaluation of the sensitivity of the SVM approach to different photometric normalisation techniques.

Image projection: Suppose that we have c clients and M training face images $x_i, i = 1, \ldots, M$, $x_i \in R^D$ each belonging to one of the client classes $\{C_1, C_2, \ldots, C_c\}$. Then we can define the following second-order statistics:

- Between-class scatter matrix:

$$S_B = \frac{1}{c}\sum_{k=1}^{c}(\mu_k - \mu)(\mu_k - \mu)^T \quad (1)$$

- Within-class scatter matrix:

$$S_W = \frac{1}{M}\sum_{k=1}^{c}\sum_{i|x_i \in C_k}(x_i - \mu_k)(x_i - \mu_k)^T \quad (2)$$

- Total scatter matrix:

$$S_T = S_W + S_B \quad (3)$$

where μ is the grand mean and μ_k is the mean of class C_k.

The aim of the Principal Component Analysis is to identify the subspace of the image space spanned by the training face image data and to decorrelate the pixel values. This can be achieved by finding the eigenvectors W of matrix S_T associated with nonzero eigenvalues Λ by solving

$$S_T W - W\Lambda = 0 \quad (4)$$

These eigenvectors are referred to as Eigenfaces. The classical representation of a face image is obtained by projecting it into the coordinate system defined by the Eigenfaces. This projection achieves information compression, decorrelation and dimensionality reduction which facilitates the subsequent decision making. If one is also interested in identifying important attributes (features) for face verification, one can adopt a feature extraction mapping. A popular technique is to find the Fisher linear discriminants (Fisherfaces). A comparison between the Eigenface and Fisherface approaches can be found in [6]. In Section 3, a sample face image $\mathbf{y}$ will be represented by its projection into the PCA subspace $\mathbf{x}$ obtained as $\mathbf{x} = W^T\mathbf{y}$.

2.2. Support vector machines

The decision making tool investigated in this paper is the Support Vector Machine. Below we give a brief presentation of the basic theory. The reader is referred to [1] for a more comprehensive introduction. SVMs are based on the principle of structural risk minimisation. The aim is to minimise the upper bound on the expected (or actual) risk defined as[1]

$$R(\alpha) = \int \frac{1}{2}|z - f(\mathbf{x}, \alpha)|dP(\mathbf{x}, z) \quad (5)$$

where α is a set of parameters defining the trained machine, z a class label associated with a training sample $\mathbf{x}$, $f(\mathbf{x}, \alpha)$ a function providing a mapping from training samples to class labels, and $P(\mathbf{x}, z)$ the unknown probability distribution associating a class label with each training sample. Let l denote the number of training samples and choose some η such that $0 \leq \eta \leq 1$. Then, with probability $1 - \eta$, the following bound on the expected risk holds:

$$R(\alpha) \leq R_{emp}(\alpha) + \sqrt{\frac{h(\log(2l/h) + 1) - \log(\eta/4)}{l}} \quad (6)$$

[1]The notation is similar to the one in [1].

where $R_{emp}(\alpha)$ is the empirical risk as measured on the training set and h is the so called Vapnik Chervonenkis (VC) dimension. The second term on the right hand side is called the VC confidence. There are two strategies for minimising the upper bound. The first one is to keep the VC confidence fixed and to minimise the empirical risk and the second one is to fix the empirical risk (to a small value) and minimise the VC confidence. The latter approach is the basis for SVMs and below we will briefly outline this procedure.

First consider the linear separable case. We are looking for the optimal hyperplane in the set of hyperplanes separating the given training samples. This hyperplane minimises the VC confidence and provides the best generalisation capabilities. Giving a geometric interpretation, the optimal hyperplane maximises the sum of the distances to the closest positive and negative training samples. This sum is called the *margin* of the separating hyperplane. It can be shown that the optimal hyperplane $\mathbf{w} \cdot \mathbf{x} + b = 0$ (where $\mathbf{w}$ is normal to the hyperplane) is obtained by minimising $\|\mathbf{w}\|^2$ subject to a set of constraints. This is a quadratic optimisation problem.

These concepts can be extended to the non-separable and non-linear case. The separability problem is solved by adding a term to the expression subject to minimisation. This term is the sum of the deviations of the non-separable training samples from the boundary of the margin. This sum is weighted using a parameter controlling the cost of misclassification. The second problem is how to handle non-linear decision boundaries. This is solved by mapping the training samples to a high-dimensional feature space using kernel functions. In this space the decision boundary is linear and the techniques outlined above can be directly applied. The kernel functions used in the experiments reported in Section 3 are radial basis functions defined as

$$K(\mathbf{x}_i, \mathbf{x}_j) \quad = \quad e^{-\gamma\|\mathbf{x}_i - \mathbf{x}_j\|^2} \quad (7)$$

where $\mathbf{x}_i$ and $\mathbf{x}_j$ denote two samples. The γ-value is a user-controlled parameter. The reader is referred to [6] for an evaluation of the relative performances of different kernels.

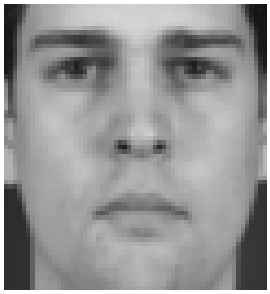
Client-specific thresholding: In verification, the output of the SVM decision function is thresholded to determine whether the identity claim is authentic or not. The results reported in Section 3 were obtained using client-specific thresholds obtained from the distribution of impostor distances (we only have a few samples per client which prevents us from using the client distribution). Given the mean μ_k and the standard deviation σ_k of the impostor distances for client k, the threshold is computed as

$$\tau_k = \mu_k + \tau \cdot \sigma_k \quad (8)$$

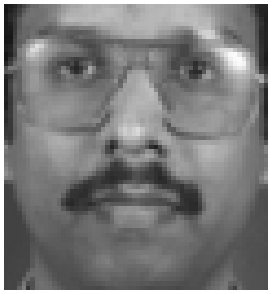
where τ is a global threshold.

3. EXPERIMENTAL RESULTS

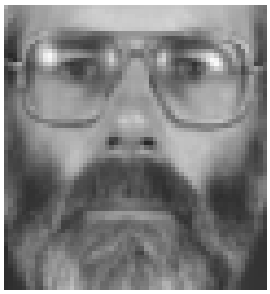
The experiments summarised below were all performed on frontal-face images from the extended M2VTS multimodal database [10]. This publicly available database contains face images and speech recordings of 295 persons. The subjects were recorded in four separate sessions uniformly distributed over a period of 5 months, and within each session a number of shots were taken including both frontal-view and rotation sequences. In the frontal-view sequences



(a) Client 000



(b) Client 002



(c) Client 007

(d) Client 030

Figure 1. Examples of client and impostor images registered using groundtruth eye coordinates.

(a) Client 000

(b) Client 099

(c) Client 000

(d) Client 002

Figure 2. Examples of impostor mis-registrations obtained from (a–b) perturbed groundtruth coordinates (standard deviation 2) and (c–d) failed optimisations of the robust correlation.

the subjects read a specific text (providing synchronised image and speech data), and in the rotation sequences the head was moved vertically and horizontally (providing information useful for 3D surface modelling of the head).

The verification experiments were conducted according to the Laussane evaluation protocol [10]. This protocol provides a framework within which the performance of vision- and speech-based person authentication systems running on the extended M2VTS database can be measured. The protocol specifies a partitioning of the database into disjoint sets used for training and testing. Within the protocol, the verification performance is measured using the false acceptance and the false rejection rates. The operating point where these two error rates equal each other is typically referred to as the equal error rate.

3.1. Authentication results

Verification experiments were carried out for four different sets of eye coordinates and the results are listed in Table 1. The first set of coordinates was obtained by manually localising the eyes in the model and probe images. This set provides the groundtruth and the corresponding verification errors are used as baselines for experimental comparison. Examples of client images registered using the groundtruth coordinates are shown in Figure 1. The second and third sets were obtained synthetically by perturbing the manually localised coordinates in the first set. The horizontal and vertical components of the eye positions were independently perturbed by adding random displacements drawn from a normal distribution (see Figures 2a and 2b for examples). As can be seen in Table 1, the mean and median registration errors (computed from the distribution of Euclidean distances from the groundtruth client coordinates) increase linearly with the standard deviation of the normal distribution. A similar relationship is observed for the equal error rate showing that the SVM verification performance is, as expected, severely affected by large mis-registrations. The fourth set of eye coordinates was obtained fully automatically using the robust correlation described in Section 2. This set provides coordinates which have errors distributed in a way dependent on the image data and the specific method used for registration. As far as the method is concerned, there are two main reasons why the registration sometimes fails. Firstly, the similarity function used in the robust correlation is not necessarily optimal for registration. Secondly, the gradient-based optimisation technique is not guaranteed to reach the global maximum and the search occasionally fails producing erroneous coordinates. Two example mis-registrations are shown in Figures 2c and 2d.

The mean and median registration errors listed in Table 1 give an indication of how the corresponding distributions are positioned. However, it can sometimes be useful to look at the full distributions as represented by the cumulative histograms in Figure 3. In the case of the automatically localised coordinates, we have plotted the distance distributions for both the clients and the impostors. One can see that the eye coordinates of the client images are located with a significantly higher accuracy compared to the impostor images. This is mainly due to the fact that we are using client-specific models in the registration process. For the verifi-

Coords	Std dev	Mean	Median	EER
Man	N/A	0.00	0.00	3.50
Per	1.0	1.25	1.17	4.70
Per	2.0	2.53	2.35	8.52
Aut	N/A	1.78	1.17	6.45

Table 1. Registration errors and verification performances: mean and median Euclidean distances from groundtruth eye coordinates and equal error rates (EER) for the manual (Man), perturbed (Per) and automatic (Aut) sets of eye coordinates. The standard deviations of the normal distributions are also listed.

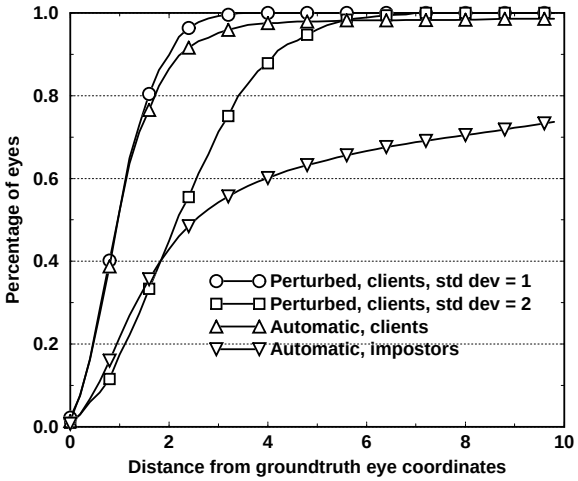
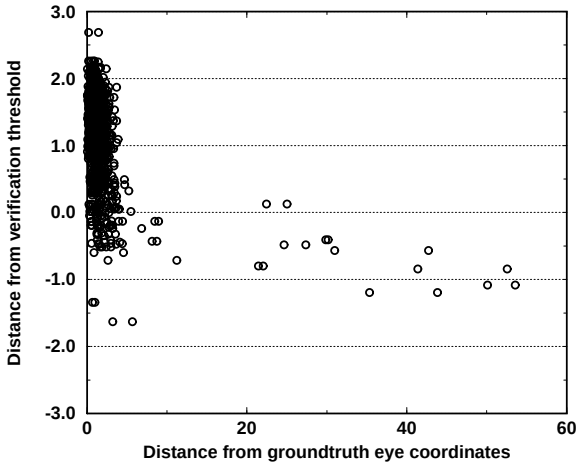
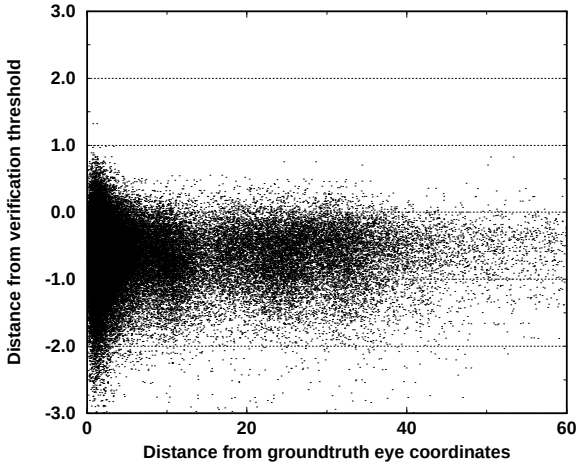


Figure 3. Registration errors: cumulative histograms of Euclidean distances from groundtruth eye coordinates for randomly perturbed (clients only, standard deviations 1 and 2) and automatically located (clients and impostors) eye coordinates.



(a) Clients



(b) Impostors

Figure 4. Distances from the verification threshold versus registration errors: scatter diagrams of (a) clients (600 eye pairs) and (b) impostors (40000 eye pairs).

cation scenario, this is not necessarily a disadvantage since our primary objective is correct alignment of images corresponding to authentic identity claims. However, as can be seen from the scatter diagram in Figure 4b, it is not always the case that a mis-registered image acquired from an impostor is correctly classified (the points to the far right located above the verification threshold). The registration errors can sometimes be quite large and these images do not necessarily contain faces. Since the objective of our study was to investigate how sensitive the SVMs are to registration errors, we only used correctly aligned images in the training process. The SVM response to these non-face patterns is therefore not always the desired one. However, the mis-classified impostor images are dominated by correctly or nearly correctly aligned faces suggesting that these persons are similar and, therefore, occupy neighbouring subspaces in the face space. In the case of the clients, this dominance is not as pronounced and a larger proportion of the mis-classifications are also mis-registrations.

The verification rates listed in Table 1 correspond to single points (the equal error rates) in the receiver operating characteristics. By varying the verification threshold, we obtain a distribution of points showing the trade-off between the false rejection and the false acceptance. This is shown in Figure 5 for the four different sets of eye coordinates.

In Table 2, we list the verification rates obtained by two other partners within the M2VTS project. Both methods are based on the dynamic link architecture as originally proposed in [9]. The methods differ in the choice of features: in the first technique [4], linear disriminants were used and, in the second method [8], the features were derived using mathematical morphology. The error rates shown in Tables 1 and 2 were computed from the same data set and using identical evaluation protocols. We can therefore draw the conclusion that the fully automatic verification system based on the robust correlation and SVMs outperforms the two versions of the dynamic link architecture presented here.

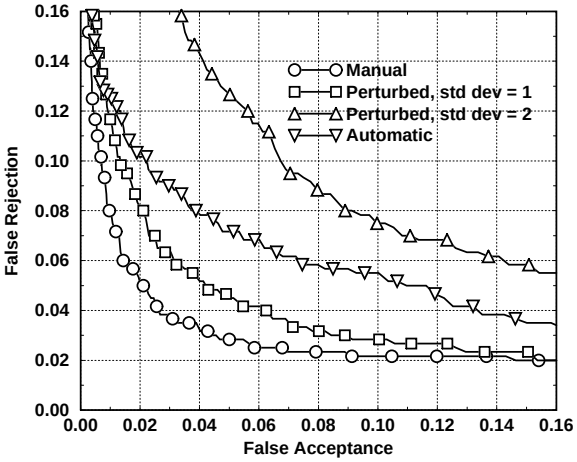


Figure 5. Verification performance: false rejection versus false acceptance for the four different sets of eye coordinates.

Partner	Method	FR	FA	ME	Ref
EPFL	DLA-LD	8.36	8.36	8.36	[4]
AUT	DLA-MM	8.11	8.00	8.06	[8]

Table 2. Verification results obtained by other partners within the M2VTS project: false rejection (FR), false acceptance (FA) and mean error rate (ME) for the dynamic link architecture based on local discriminants (DLA-LD) and mathematical morphology (DLA-MM).

4. CONCLUSIONS

The paper studied SVMs in the context of face authentication. We presented an evaluation of the impact of registration errors on the SVM performance. The results obtained by randomly perturbing the groundtruth eye coordinates were compared to those of a fully automatic registration technique based on a robust form of correlation. In the latter method, mis-registrations occurred when the optimisation of the similarity function failed and the search terminated in a local maximum. The results of the comparison show that, unless prior knowledge about the problem-specific invariances is incorporated into the SVM training algorithm, the performance is likely to suffer. We presented competitive results obtained on a large face database of 295 subjects and we showed that the combined system based on the robust correlation for face registration and the SVMs for verification compares favourably with two other methods based on the dynamic link architecture.

5. ACKNOWLEDGEMENTS

The research reported in this paper was carried out with partial support from EPSRC grant GR/L61095.

References

[1] C. J. C. Burges. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2(2):121–167, 1998.

[2] C. J. C. Burges and B. Schölkopf. Improving the accuracy and speed of support vector machines. In M. Mozer, M. Jordan, and T. Petsche, editors, *Advances in Neural Information Processing Systems 10*, 1997.

[3] T. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. In H. Burkhardt and B. Neumann, editors, *ECCV'98*, pages 484–498, 1998.

[4] Y. A. (editor). ACTS-M2VTS deliverable 351b: Selection, fusion and decision strategies. Technical report, Signal Processing Laboratory, Swiss Federal Institute of Technology, Lausanne, Switzerland, 1998.

[5] F. R. Hampel, E. M. Ronchetti, P. J. Rouseseeuw, and W. A. Stahel. *Robust Statistics: The Approach Based on Influence Functions*. John Wiley, 1986.

[6] K. Jonsson, J. Kittler, Y. P. Li, and J. Matas. Support vector machines for face authentication. In T. Pridmore and D. Elliman, editors, *BMVC'99*, pages 543–553, 1999.

[7] K. Jonsson, J. Matas, and J. Kittler. Learning salient features for real-time face verification. In S. Akunuri and C. Kullman, editors, *AVBPA'99*, pages 60–65, 1999.

[8] C. Kotropoulos, A. Tefas, I. Pitas, C. Fernandez, and F. Fernandez. Performance assessment of morphological dynamic link architecture under optimal and real operating conditions. In *Int Workshop on Nonlinear Signal and Image Processing, Antalya, Turkey*, 1999.

[9] M. Lades, J. C. Vorbrüggen, J. Buhmann, J. Lange, C. von der Malsburg, R. P. Würtz, and W. Konen. Distortion invariant object recognition in the dynamic link architecture. *IEEE Trans on Computers*, 42(3):300–311, Mar 1993.

[10] K. Messer, J. Matas, J. Kittler, J. Luettin, and G. Maitre. XM2VTSDB: The extended M2VTS database. In S. Akunuri and C. Kullman, editors, *AVBPA'99*, pages 72–77, 1999.

[11] B. Moghaddam, W. Wahid, and A. Pentland. Beyond eigenfaces: Probabilistic matching for face recognition. In *FG'98*, pages 30–35, 1998.

[12] P. J. Phillips. Support vector machines applied to face recognition. In M. I. Jordan, M. J. Kearns, and S. A. Solla, editors, *Advances in Neural Information Processing Systems 11*, 1998.

[13] M. Pontil and A. Verri. Support vector machines for 3D object recognition. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 20:637–646, 1998.

[14] S. A. Rizvi, P. J. Phillips, and H. Moon. The FERET verification testing protocol for face recognition algorithms. In *FG'98*, pages 48–53, 1998.

[15] M. A. Turk and A. P. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, 1991.