

Support Vector Machines for Face Authentication

K Jonsson^{1,2}, J Kittler¹, Y P Li¹ and J Matas^{1,2}

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

²CMP, Czech Technical University
121 35 Prague, Czech Republic

Abstract

The paper studies Support Vector Machines (SVMs) in the context of face authentication. Our study supports the hypothesis that the SVM approach is able to extract the relevant discriminatory information from the training data. We believe this is the main reason for its superior performance over benchmark methods. When the representation space already captures and emphasises the discriminatory information content as in the case of Fisherfaces, SVMs lose their superiority. SVMs can also cope with illumination changes, provided these are adequately represented in the training data. However, on data which has been sanitised by feature extraction (Fisherfaces) and/or normalisation, SVMs can get over-trained, resulting in the loss of the ability to generalise. SVMs involve many parameters and can employ different kernels. This makes the optimisation space rather extensive, without the guarantee that it has been fully explored to find the best solution.

1 Introduction

Verification of person identity based on biometric information is important for many security applications. Examples include access control to buildings, surveillance and intrusion detection. Furthermore, there are many emerging fields that would benefit from developments in person verification technology such as advanced human-computer interfaces and tele-services including tele-shopping and tele-banking. Comparing verification to recognition there are several aspects which differ. First, a client – an authorised user of a personal identification system – is assumed to be co-operative and makes an identity claim. Computationally this means that it is not necessary to consult the complete set of database images (denoted *model* images below) in order to verify a claim. An incoming image (referred to as a *probe* image) is thus compared to a small number of model images of the person whose identity is claimed and not, as in the recognition scenario, with every image (or some descriptor of an image) in a potentially large database. Second, an automatic authentication system must operate in near-real time to be acceptable to users. Finally, in recognition experiments, only images of people from the training database are presented to the system, whereas the case of an imposter (most likely a previously unseen person) is of utmost importance for authentication.

The field of face recognition is well established and a large number of algorithms have been proposed in the literature. Popular approaches include the ones based on deformable templates [10], dynamic link matching [4], and Eigenfaces [9]. 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 [3] in which a robust form of correlation is applied to face authentication.

The aim of the paper is to evaluate the effectiveness of the SVM approach to face authentication. An earlier study of SVMs in face verification has been reported by Phillips [8]. 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 [7]. 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 our focus on authentication as opposed to recognition which is of concern in [8]. Our primary motivation for carrying out a similar study was to establish why the performance of the SVM approach is superior. We want to investigate the inherent potential of SVMs to extract the relevant discriminatory information from the training data irrespective of representation and preprocessing. In order to achieve this objective we have designed experiments in which faces are represented in both Principal Component (PC) and Linear Discriminant (LD) subspaces. The latter basis (Fisherfaces) is used as an example of a face representation with focus on discriminatory feature extraction while the former achieves simply data compression. We also study the effect of image photometric normalisation on the performance of the SVM method.

A number of criteria have been considered as a basis for the SVM approach evaluation, using other baseline techniques as a benchmark. We have included as benchmark verification methods not only the classical PC variants with the L^2 norm and correlation coefficient respectively, but also the LD space with the same two decision schemes. As criteria for evaluating SVMs in relation to the benchmark methods we have concentrated on the following: computational complexity (training mode and routine verification mode), robustness (sensitivity to input data conditioning), ability to extract discriminatory information, the merits of nonlinear boundaries. The last three criteria are expressed quantitatively in terms of the false client rejection and impostor acceptance rates.

The findings of our study strongly support the hypothesis that the SVM approach is powerful in the sense of being able to extract the relevant discriminatory information from the training data. This is the main reason for the large difference between the observed performance of the classical Eigenface classification methods and SVMs (factor of almost 3). When the representation space already captures and emphasises the discriminatory information content as in the case of LD bases, SVMs cease to be superior to the simple Euclidean distance or correlation decision rules.

SVMs also show a superior capability to cope with illumination changes, provided these are adequately represented in the training data. However, on data which has been sanitised by feature extraction (Fisherfaces) and/or normalisation, SVMs can get over-trained, resulting in the loss of the ability to generalise. SVMs involve many parameters and can employ different kernels. This makes the optimisation space rather extensive,

without the guarantee that it has been fully explored to find the best solution.

The paper is organised as follows. In the next section we introduce the two face representation spaces used in our study, namely Eigenfaces and Fisherfaces. In Section 2 we overview the SVM approach to face identity verification and summarise the benchmark classification methods. Section 3 introduces the face database used in experimentation and describes the experiments carried out, their objectives and the results obtained. The results are discussed in Section 4 and conclusions are drawn in Section 5.

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 face images to be compared. Once an image is registered, it can further be normalised photometrically. In our study we set out to investigate the resilience of different decision making methods to varying illumination and thus this step was applied only in a subset of experiments. In the final step of processing, the image is projected into a coordinate system which facilitates the decision making process computationally and possibly emphasises the important attributes for face verification.

Geometric Normalisation. As the focus of the paper is on the decision making aspects of face authentication we have tried to eliminate the dependency of our experiments on processes which may lack robustness. For this reason we have performed face registration semi-automatically. The procedure 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.

Photometric Normalisation. When applied, the photometric normalisation consisted of removing the mean of the geometrically normalised image and scaling the pixel values by their standard deviation, estimated over the whole cropped image.

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

BMVC99

- 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_{pca} of matrix S_T associated with nonzero eigenvalues Λ by solving

$$S_T W_{pca} - W_{pca} \Lambda = 0 \quad (4)$$

These eigenvectors are referred to as Eigenfaces. The classical representation of a face image is obtained by projecting it to the coordinate system defined by the Eigenfaces.

The projection of face images into the Principal Component (Eigenface) subspace achieves information compression, decorrelation and dimensionality reduction to facilitate decision making. If one is also interested in identifying important attributes (features) for face authentication, one can adopt a feature extraction mapping. A popular technique is to find the Fisher linear discriminants (Fisherfaces) by solving

$$S_B W_{lda} - S_W W_{lda} \Lambda = 0 \quad (5)$$

The projection of a face image into the system of Fisherfaces associated with nonzero eigenvalues will yield a representation which will emphasise the discriminatory content of the image. The solution of the generalised eigenvalue problem in Equation 5 is known, but due to the high dimensionality many standard methods fail and the choice of a stable numerical algorithm is non-trivial [5]. Figure 1 shows the first few PC and LD basis images.

In Section 3, we perform experiments with different number of basis vectors taken from either the Eigenface or Fisherface systems. In the following, for the sake of notational simplicity, we shall not distinguish between the two different basis systems, nor shall we explicitly denote the dimensionality of the representation space. The actual representation used will be clear from the experiment description. Thus in general, in each experiment we shall work with some transformation matrix W . A sample face image \mathbf{y} will then be represented by a projection \mathbf{x} obtained as $\mathbf{x} = W^T \mathbf{y}$. Similarly, the client model μ_k will be projected into a vector ω_k in the appropriate representation space.

2.2 Classification

Support Vector Machines. The main decision making tool investigated in this paper is the Support Vector Machine. Below we give a brief presentation of the basic theory. The

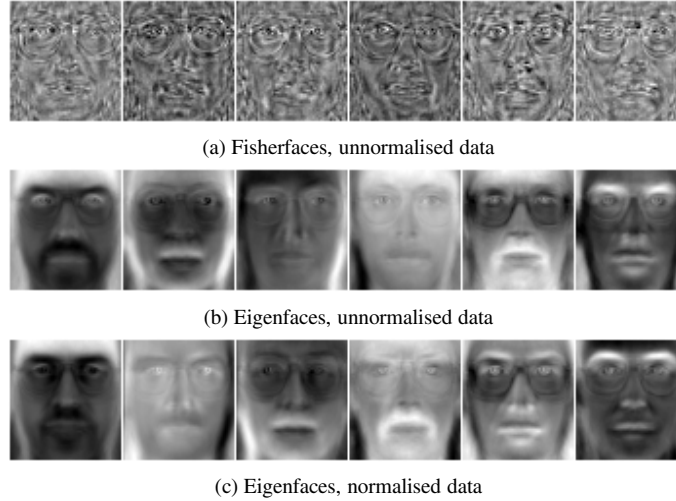


Figure 1: Basis vectors: Subspace computed using (a) LDA (unnormalised data) and (b–c) PCA (unnormalised and normalised data). In all three cases, the first six basis vectors are shown.

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 as¹

$$R(\alpha) = \int \frac{1}{2} |z - f(\mathbf{x}, \alpha)| dP(\mathbf{x}, z) \quad (6)$$

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:

$$R(\alpha) \leq R_{emp}(\alpha) + \sqrt{\frac{h(\log(2l/h) + 1) - \log(\eta/4)}{l}} \quad (7)$$

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 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

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

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 linear, polynomial and radial basis functions (RBFs) defined as

$$K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i \cdot \mathbf{x}_j \quad (8)$$

$$K(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i \cdot \mathbf{x}_j + 1)^d \quad (9)$$

$$K(\mathbf{x}_i, \mathbf{x}_j) = e^{-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2} \quad (10)$$

where \mathbf{x}_i and \mathbf{x}_j denote two samples. The user-controlled parameters are the degree d in the case of the polynomial and the γ -value in the case of the RBF kernel.

In addition to the SVMs with different kernels we have also implemented the following standard classification rules as baselines for experimental comparison.

Euclidean Distance. The most commonly used decision rule is based on the Euclidean distance between the sample projection \mathbf{x} and the projection of the k -th client mean ω_k , i.e.

$$d_E(\mathbf{x}, \omega_k) = \sqrt{(\mathbf{x} - \omega_k)^T (\mathbf{x} - \omega_k)} \quad (11)$$

The claimed client identity is accepted if $d_E(\mathbf{x}, \omega_k)$ is below a threshold τ_{Ek} . Otherwise it is rejected.

Normalised Correlation. Alternatively, the decision can be based on the correlation score

$$d_C(\mathbf{x}, \omega_k) = \frac{|\mathbf{x}^T \omega_k|}{|\mathbf{x}| |\omega_k|} \quad (12)$$

In the case of the correlation measure the claimed identity is accepted if $d_C(\mathbf{x}, \omega_k)$ exceeds a pre-specified threshold τ_{Ck} .

Client-Specific Thresholding. The client-specific threshold τ_k can be determined from the receiver operating characteristic (ROC) computed on an independent evaluation set. The procedure amounts to generating ROC curves parametrised by specific percentiles of the imposter distance distributions. Each percentile defines client-specific offsets. The ROC curves are produced by measuring the false rejection and false acceptance rates for different distance increments measured from these offsets. The ROC curve yielding the minimum equal error rate and the actual increment giving this error jointly define the client specific thresholds.

3 Experimental Results

The experiments summarised below were all performed on frontal-face images from the extended M2VTS multi-modal database [6]. 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 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 experiments were conducted according to the Laussane evaluation protocol [6]. This protocol provides a unified 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 three disjoint sets: a training set (200 clients), an evaluation set (200 clients and 25 impostors) and a test set (200 clients and 70 impostors). The training set is used to build client models, the evaluation set to get distributions of client and impostor scores used to establish verification thresholds, and the test set to obtain a reliable estimate of the verification rate on independent data.

3.1 Results on Face Authentication

Experiments were performed for the two different representations with and without face normalisation, giving four results for each authentication method. The results are summarised in Table 1. Let us first of all look at the baseline methods. One can see that the performance of both the Euclidean distance and the correlation classifiers improves monotonically with the data quality (PCs without normalisation, PCs with normalisation, LD without normalisation, LD with normalisation). The Euclidean distance is particularly sensitive to the deviations from the implicit model underlying the approach, i.e. client clusters being very compact and roughly spherical. The correlation coefficient can cope better with deviations from the sphericity. However, once the data is of that form as in the case of the LD bases with normalised data, the inherent flexibility of this classification method results in a slightly worse performance than that achieved by the Euclidean distance classifier.

The verification performance as a function of subspace dimensionality for the PCA subspaces using SVMs as classification scheme is shown in Table 2. For both subspaces (unnormalised and normalised), the total error rate drops when the number of coefficients is increased. However, when a certain point is reached (about 200 coefficients) the performance saturates and there is no further improvement of the verification rates. This series of experiments shows that the SVMs are robust to changes in the quality of the representation and perform well on both under-represented data (low number of coefficients) and when noise is present (high number of coefficients).

A somewhat surprising result of the experiments is that the SVMs tend to perform better for lower data quality. In fact Table 1 shows that the performance is almost without exception inversely related to data quality. The best results have been obtained for PCs without face normalisation and the worst results for LDs with face normalisation when the authentication problem becomes relatively easy. This suggests that SVMs can be over-

BMVC99

SSP	CLM	NOR	KRN	PAR	Evaluation set		Test set		
					EER	TE	FR	FA	TE
PCA	EUD	×	N/A	N/A	9.83	19.66	5.50	12.46	17.96
	NOC	×	N/A	N/A	7.50	15.00	5.25	6.70	11.95
	SVM	×	LIN	N/A	3.90	7.80	2.75	3.75	6.50
		×	POL	2	3.69	7.38	3.25	3.54	6.79
		×	RBF	0.01	3.83	7.66	2.75	3.68	6.43
	EUD	✓	N/A	N/A	7.00	14.00	4.00	7.79	11.79
	NOC	✓	N/A	N/A	6.83	13.66	3.75	6.51	10.26
	SVM	✓	LIN	N/A	3.66	7.32	3.25	3.52	6.77
		✓	POL	2	3.59	7.18	3.00	3.46	6.46
		✓	RBF	0.01	3.54	7.08	3.25	3.43	6.68
*		✓	RBF	0.01	3.50	7.00	2.50	3.25	5.75

SSP	CLM	NOR	KRN	PAR	Evaluation set		Test set		
					EER	TE	FR	FA	TE
LDA	EUD	×	N/A	N/A	8.59	17.18	9.75	7.94	17.69
	NOC	×	N/A	N/A	3.33	6.66	3.75	3.48	7.23
	SVM	×	LIN	N/A	3.47	6.94	3.75	3.39	7.14
		×	POL	2	3.40	6.80	3.75	3.30	7.05
		×	RBF	0.01	3.45	6.90	3.75	3.39	7.14
	EUD	✓	N/A	N/A	3.33	6.66	2.25	3.43	5.68
	NOC	✓	N/A	N/A	3.83	7.66	3.00	3.95	6.95
	SVM	✓	LIN	N/A	3.87	7.74	3.75	3.82	7.57
		✓	POL	2	3.89	7.78	3.75	3.84	7.59
		✓	RBF	0.01	3.90	7.80	3.75	3.84	7.59

Table 1: Verification performance on the extended M2VTS database: False rejection (FR), false acceptance (FA) and total error rate (TE) as functions of subspace (SSP, upper table PCA, lower table LDA), classification method (CLM), photometric normalisation (NOR) and kernel (KRN, if applicable). The classification methods are Euclidean distance (EUD), normalised correlation (NOC) and support vector machines (SVM). The kernels are linear (LIN), polynomial (POL) and radial basis functions (RBF). The kernel parameters (PAR) are indicated when applicable (degree for polynomial and γ for RBF). The row marked with a \star was obtained using a different subspace (see text).

SSP	NoC	ERG	Evaluation set		Test set		
			EER	TE	FR	FA	TE
UN	131	93	4.36	8.72	2.75	4.13	6.88
	148	94	4.00	8.00	2.75	3.83	6.58
	170	95	4.00	8.00	2.75	3.86	6.61
	198	96	3.83	7.66	2.75	3.68	6.43
NO	142	92	3.81	7.62	3.50	3.65	7.15
	158	93	3.91	7.82	3.25	3.77	7.02
	177	94	3.54	7.08	3.25	3.43	6.68
	200	95	3.78	7.56	3.00	3.63	6.63

Table 2: Verification performance as function of PCA subspace dimensionality: False rejection (FR), false acceptance (FA) and total error rate (TE) as functions of subspace (SSP) and number of projection coefficients (NoC). The two subspaces were obtained from unnormalised (UN) and normalised (NO) data. The corresponding percentage of the energy (ERG) in the subspace is also indicated. All results were obtained using RBF kernel ($\gamma = 0.01$).

BMVC99

trained on relatively easy data and cannot generalise to new data so readily. This finding has been confirmed by another experiment where the PC basis system was computed for normalised face data but without removing the global face mean (row marked with \star). In this case the first Eigenface is completely determined by this global mean face and the successive Eigenfaces are influenced accordingly. For this system of PCs the test set performance of 5.75% is comparable to the best achieved with the Euclidean distance decision rule (5.68%).

The following conclusions can be drawn from the study:

- The SVM approach is able to extract the relevant discriminatory information from the data fully automatically. It can also cope with illumination changes. The major role in this characteristic is played by the SVMs ability to learn non-linear decision boundaries.
- On data which has been sanitised by feature extraction (Fisherfaces) and/or normalisation, SVMs can get over-trained, resulting in the loss of the ability to generalise.
- SVMs involve many parameters and can employ different kernels. This makes the optimisation space rather extensive, without the guarantee that it has been fully explored to find the best solution.
- An SVM takes about 5 seconds to train per client (on a Sun Ultra Enterprise 450). This is about an order of magnitude longer than determining client-specific thresholds for the Euclidean and correlation coefficient classifiers. However, from the practical point of view the difference is insignificant.

4 Discussion

It is interesting to note that the total error rates achieved on the evaluation set by all the SVM methods are very similar. The better results on the test set yielded by some of the techniques suggest that the test set data is easier than the evaluation set data. The differences in the test set performance are perhaps indicative of the different generalisation capabilities of the respective methods. While it is true that the SVMs are designed not to over-train, if the representation space used is excessively tuned to the training data already, the SVM cannot mitigate such an inherent problem. This could explain the better test set performance on a less powerful representation afforded in terms of the PC space.

Note also the difference in the information used by the various decision making schemes which can be gleaned from Figure 2. In this figure we show the original of a probe image, its PC reconstruction in Figure 2b, its LD “reconstruction” in Figure 2c and finally its SVM “reconstruction” in Figure 2d. The latter two reconstructions have been produced in an analogical way to the PC reconstruction approach. The classical Euclidean distance and correlation methods use the standard PCA reconstruction of the probe image shown in Figure 2b. The SVM using the PC coefficients work with a similar source of information but some regions in the image are weighted more heavily. Thus, SVMs seem to be capable of performing client-dependent feature extraction. The Fisherface reconstruction uses the global mean image as a starting point. The bright areas indicate the increased weighting applied to some pixels in the image. This weighting is client dependent and is a function of the probe image projection into the Fisher space.

BMVC99



Figure 2: Reconstruction from subspace representations: (a) original image, reconstruction using (b) PCA (191 coefficients), (c) LDA (199 coefficients) and (d) SVM (37 coefficients) subspaces. All subspaces were obtained from unnormalised data.

5 Conclusions

The paper studied SVMs in the context of face authentication. Our study proved the hypothesis that the SVM approach is able to extract the relevant discriminatory information from the training data. This is the main reason for the large difference between the observed performance of the classical Eigenface classification methods used as a benchmark and SVMs (factor of almost 3). When the representation space already captures and emphasises the discriminatory information content as in the case of LD bases, SVMs cease to be superior to the simple Euclidean distance or correlation decision rules.

SVMs also show a superior capability to cope with illumination changes, provided these are adequately represented in the training data. However, on data which has been sanitised by feature extraction (Fisherfaces) and/or normalisation, SVMs can get over-trained, resulting in the loss of the ability to generalise. SVMs involve many parameters and can employ different kernels. This makes the optimisation space rather extensive, without the guarantee that it has been fully explored to find the best solution.

Acknowledgements

The research reported in this paper was carried out within the framework of the European Union ACTS project M2VTS and ESPRIT RETINA. We would also like to acknowledge Thorsten Joachims at the University of Dortmund for making the SVM Light program [2] available to the research community.

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] T. Joachims. Making large-scale SVM learning practical. In B. Schölkopf, C. J. C. Burges, and A. J. Smola, editors, *Advances in Kernel Methods: Support Vector Learning*, pages 41–56, Cambridge, USA, 1998. MIT Press.

BMVC99

- [3] K. Jonsson, J. Matas, and J. Kittler. Learning salient features for real-time face verification. In S. Akunuri and C. Kullman, editors, *Audio- and Video-Based Biometric Person Authentication, Washington D. C., March 1999*, pages 60–65, 1999.
- [4] 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 Transactions on Computers*, 42(3):300–311, Mar 1993.
- [5] Y P Li, J Kittler, and J Matas. Effective implementation of linear discriminant analysis for face recognition and verification. In A. Leonardis and F. Solina, editors, *Computer Analysis of Images and Patterns, Ljubljana, Slovenia, Sept, 1999*. to appear.
- [6] K. Messer, J. Matas, J. Kittler, J. Luetin, and G. Maitre. XM2VTSDB: The extended M2VTS database. In S. Akunuri and C. Kullman, editors, *Audio- and Video-Based Biometric Person Authentication, Washington D. C., March 1999*, pages 72–77, 1999.
- [7] B. Moghaddam, W. Wahid, and A. Pentland. Beyond eigenfaces: Probabilistic matching for face recognition. In *Automatic Face and Gesture Recognition, Nara, Japan, April 1998*, pages 30–35, 1998.
- [8] P. J. Phillips. Support vector machines applied to face recognition. In M. S. Kearns, S. A. Solla, and D. A. Cohn, editors, *Advances in Neural Information Processing Systems 11*, 1998.
- [9] M. A. Turk and A. P. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, 1991.
- [10] A. L. Yuille. Deformable templates for face recognition. *Journal of Cognitive Neuroscience*, 3(1):59–70, 1991.