Proceedings of the First International Workshop on Pattern Recognition with Support Vector Machines (SVM 2002), Niagra Falls, Canada, pages 249–259, August 2002

Support Vector Features and the Role of Dimensionality in Face Authentication

F. Smeraldi[1] ⋆, J. Bigun[1], and W. Gerstner[2]

[1] Halmstad University, Box 823, S-30118 Halmstad, Sweden
[2] Swiss Federal Institute of Technology (EPFL-DI), 1015 Lausanne, Switzerland

Abstract. A study of the dimensionality of the Face Authentication problem using Principal Component Analysis (PCA) and a novel dimensionality reduction algorithm that we call Support Vector Features (SVFs) is presented. Starting from a Gabor feature space, we show that PCA and SVFs identify distinct subspaces with comparable authentication and generalisation performance. Experiments using KNN classifiers and Support Vector Machines (SVMs) on these reduced feature spaces show that the dimensionality at which saturation of the authentication performance is achieved heavily depends on the choice of the classifier. In particular, SVMs involve directions in feature space that carry little variance and therefore appear to be vulnerable to excessive PCA-based compression.

1 Introduction

Face authentication/recognition can be described as a classification problem in a high-dimensional vector space. Much research effort is devoted to determining a subspace (feature space) that would carry all the relevant information from a facial image in a form that is convenient for classification.

To date, there is no general agreement in the literature concerning the optimal number of features required for authentication. Principal Component Analysis (PCA) based approaches call for a reduction of the dimensionality of the input space by projection onto a set of basis vectors (Eigenfaces) [1,2]. Part of the literature indicates a number of Principal Components of the order of 30 to 50 to be sufficient [2,3], while other studies such as [4] and [5], where PCA is used in conjunction with Linear Discriminant Analysis, raise this number to a few hundred components.

On the other hand, approaches based on multi-resolution decompositions involve mapping (regions of) the image to a higher dimensional space such as the Gabor decomposition space [6], which corresponds to an over-complete representation [7] of the image. Discretization of the local spectrum typically leads to a number of features of the order of one thousand or more [8,9,10]. Because of the demonstrated effectiveness of the Gabor decomposition both in face authentication and in a variety of other computer vision problems, we have chosen

⋆ Part of this research was carried out while F. Smeraldi was employed at ([2])

to base our analysis on the Gabor space rather than on the original grey level space of the images.

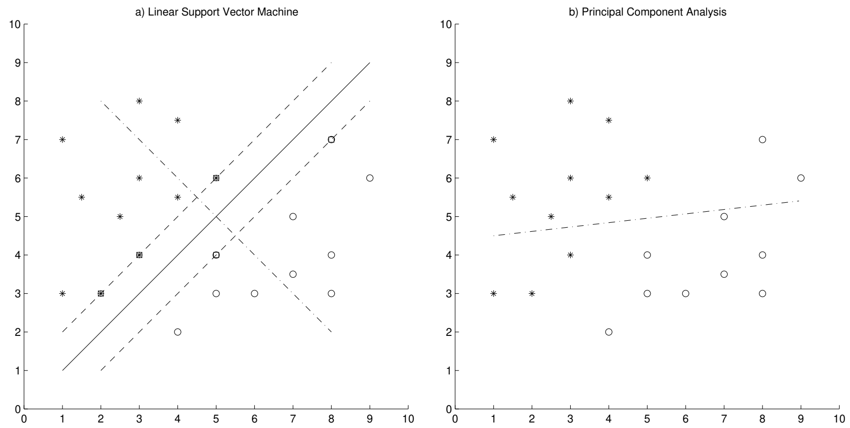
Support Vector Machine (SVM) classifiers also implicitly involve mapping the image onto a high (possibly infinite) dimensional Hilbert space [11]. The joint use of a sparse subset of Gabor features corresponding to the region around the eyes and the mouth of a subject and linear SVMs has been shown to yield very low error rates in authentication experiments [10,12]. Gabor filter responses extracted from images of different subjects turn out to be linearly separable, so that the hyperplane decision functions of linear SVMs are sufficient to discriminate between clients and impostors. Each such hyperplane determines, through its normal vector, a direction in the Gabor space. We show that the directions computed on a set of training subjects can effectively be used to reduce the dimensionality of the feature space used for authenticating a completely different set of subjects, thus achieving a *data–independent compression* of the biometric features. This SVM–based compression technique, to which we refer as Support Vector Features (SVFs), yields a performance comparable to that obtainable by performing PCA in the Gabor space.

Concerning the relation between the dimensionality of the feature space, the type of classifier employed and the resulting performance, we find that the error rate of simple threshold-based classifiers applied after either PCA or SVFs saturates to its minimum value already in a 30 dimensional subspace. In contrast, more complex and reliable classifiers such as SVMs appear to profit from a higher dimensionality. Our results indicate that the classical criterion of discarding the Principal Components associated with the smallest Eigenvalues must be used with utmost care in conjunction with complex classifiers.

2 Gabor Feature Space and Mixture of Experts

We consider a face authentication algorithm based on a Gabor feature space $\mathcal{G}$ constructed following the procedure detailed in [10]. A sparse Log-polar retinotopic sampling grid is centred on the eyes and the mouth of the subjects in three successive and automatic fixations. A set of Gabor filter responses is then extracted at each point of the grid. Each of these sets of responses can be seen as a vector in a separate copy of $\mathcal{G}$. Three machine experts, implemented by classifiers, are then employed to independently authenticate a client based on the Gabor responses obtained from the three fixations. The final decision about the identity claim being true or false is obtained by fusing the opinions of the experts using a simple rule. This procedure is meant to increase robustness; however, each machine expert independently authenticates the client based on a single copy of $\mathcal{G}$.

The set of Gabor filters employed in the experiments reported below consists of 30 filters arranged in 5 resolution (frequency) and 6 orientation channels. Given that the retinotopic grid consists of 50 points, the dimensionality of $\mathcal{G}$ turns out to be $30 \times 50 = 1500$. The mapping between the original grey-level space and $\mathcal{G}$ is effective in separating the response vectors obtained from different



(a) The dash-dot line represents the direction of the normal vector $\boldsymbol{w}$. Support vectors are boxed.

(b) The dash-dot line represents the direction of the first Principal Component.

Fig. 1. Two dimensional "toy" examples of a) Linear SVM, b) Principal Component Analysis.

subjects. This is confirmed by the very low error rates achieved using linear SVM classifiers on $\mathcal{G}$ (see Table 1). Because of this fact, we have decided to base this study on $\mathcal{G}$ rather than on the grey level image space.

In the rest of this paper we employ PCA and the SVFs algorithm to restrict the feature space to subspaces of $\mathcal{G}$.

3 Support Vector Features

Support Vector Machine classifiers are characterised by a decision function of the form

$$\sum_j \alpha_j y_j K(\boldsymbol{s}_j, \boldsymbol{v}) + b \geq 0, \tag{1}$$

where $\boldsymbol{v} \in \mathcal{G}$ is the input to be classified [11]. The Support Vectors $\boldsymbol{s}_j$ constitute a subset of the training examples that is determined through an optimisation process, together with the coefficients α_i. The set of labels y_j identifies the $\boldsymbol{s}_j$ as positive examples ($y_j = +1$) or negative examples ($y_j = -1$).

The kernel function $K(\boldsymbol{u}, \boldsymbol{v})$, is a (generally nonlinear) symmetric scalar function on $\mathcal{G} \times \mathcal{G}$ that satisfies the conditions of Mercer's theorem of Functional Analysis [11]. In the case of linear SVMs, i.e. $K(\boldsymbol{u}, \boldsymbol{v}) = \boldsymbol{u} \cdot \boldsymbol{v}$, the decision surface is a hyperplane in $\mathcal{G}$:

$$\sum_j \alpha_j y_j \boldsymbol{s}_j \cdot \boldsymbol{v} + b = \boldsymbol{w} \cdot \boldsymbol{v} + b \geq 0. \tag{2}$$

The normal vector to the hyperplane is $\boldsymbol{w} = \sum_j \alpha_j y_j \boldsymbol{s}_j$ (Figure 1a).

Given a database of training images for an ordered set $\Gamma = \{1, 2, \ldots, \#\Gamma\}$ of subjects, we consider the set $\{\mathcal{S}_\gamma\}_{\gamma \in \Gamma}$ of the linear Support Vector Machines $\mathcal{S}_\gamma$, each trained using the response vectors belonging to person γ as positive examples and the vectors belonging to the remaining $\#\Gamma - 1$ identities in $\Gamma \setminus \{\gamma\}$ as negative examples (that is, $\mathcal{S}_\gamma$ is trained to distinguish subject γ from all the others). For each classifier $\mathcal{S}_\gamma$, let $\boldsymbol{w}_\gamma \in \mathcal{G}$ denote the normal vector to the separating hyperplane. Note that, whatever the number of support vectors, there is exactly one $\boldsymbol{w}_\gamma$ for each $\mathcal{S}_\gamma$, and therefore one $\boldsymbol{w}_\gamma$ for each identity $\gamma \in \Gamma$.

We now consider the set $\{\boldsymbol{w}_\gamma\}_{\gamma \in \Gamma}$ of all the normal vectors. If $\boldsymbol{v}$ is an arbitrary vector in $\mathcal{G}$, we define the first n Support Vector Features $\boldsymbol{v}|_\Gamma^n$ associated to Γ and $\boldsymbol{v}$ as the ordered list of scalar products

$$\boldsymbol{v}|_\Gamma^n = (\boldsymbol{w}_\gamma \cdot \boldsymbol{v})_{1 \leq \gamma \leq n}. \tag{3}$$

The $\boldsymbol{v}|_\Gamma^n$ can be seen as n-dimensional vectors (SVF vectors) in a compressed feature space $\mathcal{G}|_\Gamma^n$. Classification can therefore be performed in this space much in the same way as one would do after projecting the vectors in $\mathcal{G}$ over the first n Principal Components from a given dataset.

Indeed, Support Vector Features are obtained simply by substituting the normal vectors $\{\boldsymbol{w}_\gamma\}$ for the Principal Components. SVFs therefore compare directly with linear algorithms such as PCA and LDA rather than with their kernel-based nonlinear extensions, Kernel PCA [13] and Kernel LDA [14]. As the toy example reported in Figure 1 suggests, however, normal vectors may determine entirely different directions from the Principal Components, a fact that is analysed in depth in the following Sections.

4 SVFs versus PCA

In this section, we address the question of whether the vectors $\{\boldsymbol{w}_\gamma\}$ obtained from a training set Γ consisting of a (small) number of subjects carry some general information about the structure of the face authentication problem itself, *independently* of the specific identities in Γ.

To this end, a set of 37 normal vectors $\boldsymbol{w}_\gamma$ has been computed starting from the set Γ of 37 identities from the M2VTS database (this procedure has been repeated for left-eye, right-eye and mouth based authentication separately as specified in Section 2; however, such redundancy is exclusively meant for robustness and can be safely ignored for the purposes of this paper).

We have then applied the authentication algorithm described in Section 2 to the entirely different set of 295 subjects from the XM2VTS database (see Appendix A) after replacing the uncompressed Gabor space $\mathcal{G}$ with the space $\mathcal{G}|_\Gamma^n$ of the first n Support Vector Features associated to the set Γ of the identities in the M2VTS DB.

More precisely, the feature space $\mathcal{G}$ has been compressed by computing the dot products between each response vector $\boldsymbol{v} \in \mathcal{G}$ extracted from an XM2VTS image and the first n normal vectors $\boldsymbol{w}_\gamma$ to obtain the first n Support Vector

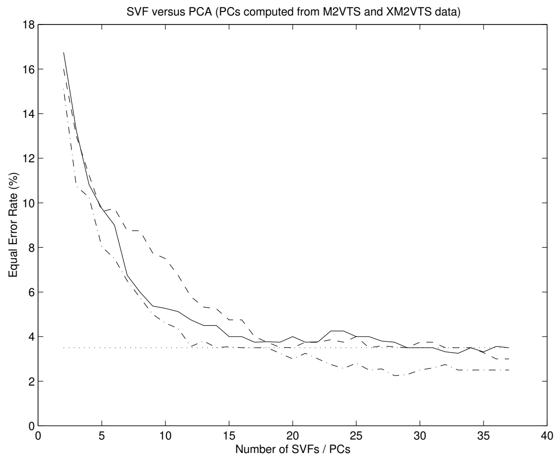


Fig. 2. EER of XM2VTS authentication tests as a function of the number of SVFs (solid line), Principal Components obtained from the M2VTS DB (dashed line) and PCs obtained from the XM2VTS DB (dash-dot line). The dotted line indicates the performance of the threshold-based classifiers in the uncompressed 1500-dimensional Gabor space $\mathcal{G}$.

Features, according to (Equation 3). The machine experts that authenticate the clients have then been restricted to work on the resulting n-dimensional SVF vectors $\boldsymbol{v}|_{\Gamma}^{n}$, with $2 \leq n \leq 37$.

For these experiments, the machine experts have been implemented as simple threshold based classifiers of the type commonly used in conjunction with PCA [2]. An identity claim is accepted if the distance between the corresponding SVF vector $\boldsymbol{v}|_{\Gamma}^{n}$ and the closest training example (SVF vector) $\boldsymbol{x}_{\nu,i}|_{\Gamma}^{n}$ for the claimed identity ν is lower than a fixed threshold τ, otherwise it is rejected. The resulting decision function is

$$\min_{i} \|\boldsymbol{x}_{\nu,i}|_{\Gamma}^{n} - \boldsymbol{v}|_{\Gamma}^{n}\| \leq \tau, \tag{4}$$

where i varies across all the training examples for the claimed identity ν.

Figure 2 (solid line) shows the Equal Error Rate (EER) obtained in XM2VTS authentication tests as a function of the number n of Support Vector Features employed for classification. The EER obtained using the full-length, 1500-dimensional Gabor feature vectors is 3.5%. As can be seen, the SVF technique allows achieving the same result already in a 29-dimensional subspace, with a compression ratio in excess of 50. Even in a 15-dimensional subspace the EER is as low as 4.0%.

Analogous results are obtained by projecting the Gabor vectors onto a variable number of Principal Components determined by performing PCA over the M2VTS DB, that is over the same set Γ of identities that was used in the computation of the SVFs (Figure 2, dashed line; note that no statistical knowledge

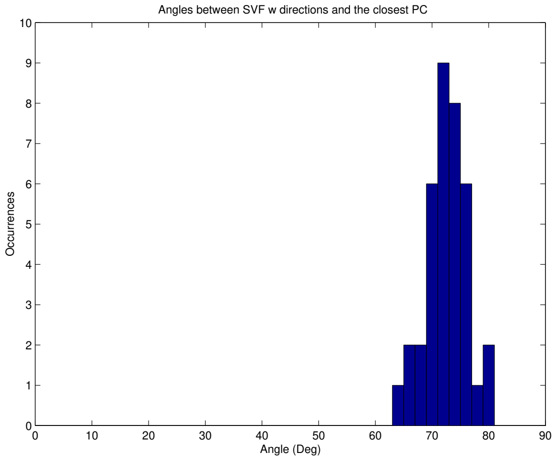


Fig. 3. Distribution of the angle between each normal vector $\boldsymbol{w}_\gamma$ and the closest direction among the 37 most significant Principal Components (median: 73°, M2VTS DB, left eye expert).

about the test database XM2VTS is employed). The dash-dotted line in Figure 2 refers to the EER obtained by extracting the Principal Components from the XM2VTS training data directly. As can be seen, the introduction of statistical knowledge about the very same identities the system is required to authenticate only results in a relatively small improvement in performance, at the expense of making the feature compression scheme data-dependent.

These experiments suggest that a 30-dimensional subspace might be sufficient to perform face authentication, a result that would be in agreement with other studies on PCA using similar classification techniques [2]. Furthermore, such a subspace appears to be characterised by a certain stability, in that the directions $\boldsymbol{w}_\gamma$ or the Principal Components computed on the M2VTS DB can be used to compress the Gabor response vectors from the completely different set of subjects constituting the XM2VTS DB with only a small performance degradation. However, the optimal subspace for authentication need not be unique. The subspace spanned by the Principal Components only accounts, on the average, for 36% of the norm of the $\boldsymbol{w}_\gamma$; therefore, PCA and SVFs essentially map the Gabor vectors to two different subspaces. It follows that a large contribution to the SVFs comes from directions that only carry a small fraction of the variance of M2VTS data, a result that will be confirmed by the experiments reported in Section 5. This is made intuitive in Figure 3, where we show that the angle formed by each normal vector $\boldsymbol{w}_\gamma$ with the closest direction among the first 37 Principal Components of M2VTS data is on the average quite large (median value: 73°).

EER (SVM)	34 SVFs ($\mathcal{G}|_{\Gamma}^{34}$)	34 PCs (M2VTS)	34 PCs (XM2VTS)	$\mathcal{G}$ (1500-dim)
Linear	2.3%	2.1%	1.5%	$\lesssim 0.25\%$
Best	1.6%	1.5%	1.0%	$\lesssim 0.25\%$
Degree	3	2	2	$1-4$

Table 1. Equal Error Rates for XM2VTS authentication experiments. SVM classifiers are trained and tested in four different feature spaces: the first 34 SVFs ($\mathcal{G}|_{\Gamma}^{34}$), the first 34 PCs extracted from the M2VTS DB, the first 34 PCs from the XM2VTS DB and the uncompressed 1500-dimensional Gabor space $\mathcal{G}$. Performance of the linear kernel and the optimal polynomial kernel is shown. The degree of the polynomial is reported in the last line.

5 Dimensionality Reduction and Advanced Classifiers

We might ask whether the apparent convergence of SVF and PCA based authentication to the performance achieved on the unaltered feature space $\mathcal{G}$ indicates that all the available information has been captured, or whether it is rather a consequence of the (simple) classification technique employed.

Substituting linear SVM classifiers for the simple threshold based rule (Equation 4) on the entire Gabor space $\mathcal{G}$ causes the EER achieved on the XM2VTS DB to drop from 3.5% to below the resolution limit of 0.25%, that corresponds to a single falsely rejected identity claim (see Table 1 and Appendix A). In Table 1 we also show the EER obtained by using SVM experts in conjunction with the three choices of subspaces in $\mathcal{G}$ corresponding to 34 Support Vector Features ($\mathcal{G}|_{\Gamma}^{34}$), 34 Principal Components obtained from the M2VTS DB and PCA over the XM2VTS DB. The first two choices both lead to the minimal EER of 3.5% when the simple threshold-based classifiers are used. As can be seen, performance on the compressed Gabor space is considerably improved when SVMs are used for classification. In particular, the dimensionality boost provided by nonlinear kernel functions appears to somewhat compensate for the compression, yielding lower error rates. Performance does not differ significantly on the subspaces determined by SVFs and PCA, with PCA on the XM2VTS database being (as expected) slightly better. However, the best results still fall short of the results achieved using the entire Gabor space $\mathcal{G}$ in conjunction with SVMs. This suggest that SVM classifiers exploit information that is lost in the compression procedure and that the dimensionality of the three subspaces actually is too low.

In order to get a feeling for what the correct dimensionality might be, we revert to the 1500-dimensional Gabor space $\mathcal{G}$ and consider the set Ξ of the 200 clients from the XM2VTS DB. From this we construct the set $\{\mathcal{S}_\xi\}$ of the 200 linear SVMs that distinguish each of these 200 clients from all the others, as has been done in Section 3 for the subjects in Γ. Let $\{\boldsymbol{w}_\xi\}$ denote the corresponding set of normal vectors. On the average, only 28% of the norm of each specific $\boldsymbol{w}_{\bar{\xi}}$ lies in the subspace spanned by the remaining 199 vectors $\{\boldsymbol{w}_\xi\}_{\xi\in\Xi\setminus\{\bar{\xi}\}}$. This indicates that all of the $\boldsymbol{w}_\xi$ identify independent directions, as is confirmed

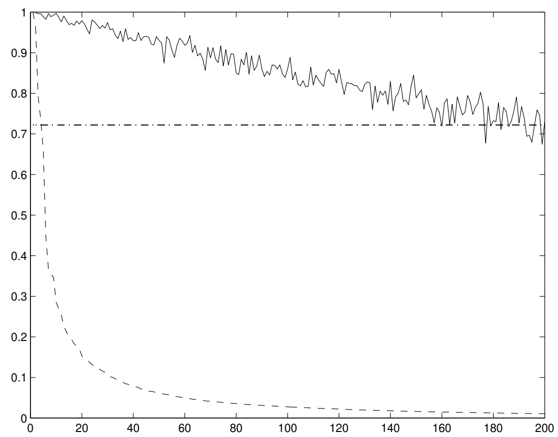


Fig. 4. Solid line: norm of the residues obtained by applying the Gram-Schmidt orthogonalisation procedure to the list of (normalised) vectors $\boldsymbol{w}_\xi$. Dash-dot line: average fraction of the norm of each $\boldsymbol{w}_\xi$ that contributes an independent direction. Comparison with the relative magnitude of the first 200 eigenvalues from PCA (dashed line) shows that most of these directions only carry a small fraction of the variance of the data. All data refer to the set of clients from the XM2VTS Database, left eye expert.

by displaying the norm of the residues obtained by applying the Gram-Schmidt orthogonalisation procedure to the list of (normalised) vectors $\boldsymbol{w}_\xi$ (Figure 4). We conclude that (SVM-based) authentication effectively involves a 200 dimensional subspace of $\mathcal{G}$.

By looking at the distribution of the magnitude of the first 200 PCA Eigenvalues on the same set Ξ of clients, one can see that SVM-based classification involves many directions that carry only a small fraction of the global variance [15].

6 Conclusions

A comparative study of the dimensionality of the face authentication problem is presented. A feature space is constructed using a sparse Log-polar retinotopic grid to sample the Gabor decomposition of facial images around the location of the eyes and the mouth. Two types of compression algorithms are employed to reduce the dimensionality of the feature space, namely PCA and a novel SVM–based technique that we call Support Vector Features (SVFs). We report experimental results using two classification techniques, thresholded distance classifiers and Support Vector Machines. PCA and SVFs are shown to identify two different subspaces in the Gabor space that allow comparably good authentication performance regardless of the classifier employed.

Experiments with a simple threshold-based classifier and SVMs suggest that the saturation of recognition/authentication performance with a low number of

Principal Components or Support Vector Features heavily depends on the choice of the classifier. In particular, authentication of the 200 clients from the XM2VTS DB using linear SVMs is shown to require a 200 dimensional subspace. SVM classifiers appear to rely on directions that are not necessarily associated to a large variance of the training data. Therefore, the classical criterion of discarding Principal Components associated with small Eigenvalues must be used with care in conjunction with SVM classifiers.

Acknowledgements

Thanks are due to B. Schölkopf for an encouraging discussion on the topics of this paper.

A Details on the Experimental Protocol

The experiments described are performed using the disjoint subsets of frontal images from the M2VTS (37 subjects, 551 images) and XM2VTS (295 subjects, 2360 images) databases. All the authentication results refer to the XM2VTS DB and are obtained using the training and test sets defined by the Lausanne Protocol, Configuration II [16] (the images from the 3rd session have been included in the training set). The protocol specifies a total of 400 client tests and 112′000 impostor tests. This allows estimating the EER with a resolution of 0.25%, corresponding to the false rejection of a single image of a client. The position of the eyes and the mouth has been manually detected on the images of the M2VTS DB. Automatic detection has been used on the XM2VTS DB [17]; erroneous detections have been corrected manually. For other published results on the XM2VTS database we refer the reader to [18].

The SVM engine was developed following the ideas in [19] and [20].

References

1. Sirovich, L., Kirby, M.: Application of the Karhunen-Loève procedure for the characterization of human faces. IEEE-PAMI Transactions on Pattern Analysis and Machine Intelligence **12** (1990) 103–108
2. Turk, M., Pentland, A.: Eigenfaces for recognition. J. Cognitive Neuroscience (Winter) **3** (1991) 71–86
3. Liu, C., Wechsler, H.: A unified Bayesian framework for face recognition. In: International Conference on Image Processing, ICIP-98, Chicago, October 4-7. Volume 1. (1998) 151–155
4. Moghaddam, B., Pentland, A.: Beyond linear eigenspaces: Bayesian matching for face recognition. In Wechsler et al., H., ed.: Nato-Asi advanced study on face recogniton. Volume F 163., Springer (1998) 230–243
5. Zhao, W., Krishnaswami, A., Chellappa, R., Swets, D.L., Weng, J.: Discriminant analysis of principal components for face recognition. In Wechsler et al., H., ed.: Nato-Asi advanced study on face recogniton. Volume F 163., Springer (1998) 73–85

6. Porat, M., Zeevi, Y.Y.: The generalized Gabor scheme of image representation in biological and machine vision. IEEE transactions on Pattern Analysis and Machine Intelligence **10** (1988) 452–468
7. Daubechies, I.: Ten lectures in wavelets. Society for industrial and applied mathematics, Philadelphia, USA (1992)
8. Duc, B., Fischer, S., Bigun, J.: Face authentication with Gabor information on deformable graphs. IEEE Transactions on Image Processing **8** (1999) 504–516
9. Lades, M., Vorbruggen, J.C., Buhmann, J., Lange, J., Malsburg, C.v.d., Hurtz, R.P., Konen, W.: Distortion invariant object recognition in the dynamic link architectures. IEEE Trans. on Computers **42** (1993) 300–311
10. Smeraldi, F., Capdevielle, N., Bigun, J.: Face authentication by retinotopic sampling of the Gabor decomposition and Support Vector Machines. In: Audio and Video based Biometric Person Authentication – AVBPA99. (1999) 125–129
11. Vapnik, V.N.: The nature of statistical learning theory. Springer–Verlag (1995)
12. Smeraldi, F., Bigun, J.: Retinal vision applied to facial features detection and face authentication. Pattern Recognition Letters **23** (2002) 463–475
13. Schölkopf, B., Smola, A.J., Müller, K.R.: 20 of Advances in Kernel Methods - Support Vector Learning. In: Kernel Principal Component Analysis. MIT Press (1998) Eds. B. Schölkopf, C. J. C. Burges, A. J. Smola.
14. Mika, S., Weston, G.R.J., Schölkopf, B., Müller, K.R.: Fisher discriminant analysis with kernels. In Hu, Y.H., Larsen, J., Wilson, E., Douglas, S., eds.: Neural Networks for Signal Processing IX, IEEE (1999) 41–48
15. Smeraldi, F.: Attention–driven pattern recognition. PhD thesis, Swiss Federal Institute of Technology – Lausanne, Switzerland (No. 2153 (2000))
16. Messer, K., Matas, J., Kittler, J., Luettin, J., Maitre, G.: XM2VTSDB: the Extended M2VTS database. In: Proceedings of the 2nd international conference on Audio- and Video-based Biometric Person Authentication (AVBPA'99), Washington DC, U.S.A.. (1999) 72–77
17. Smeraldi, F., Capdevielle, N., Bigun, J.: Facial features detection by saccadic exploration of the Gabor decomposition and support vector machines. In: Proceedings of the 11th Scandinavian Conference on Image Analysis – SCIA 99, Kangerlussuaq, Greenland. Volume I. (1999) 39–44
18. Matas, J., Hamouz, M., Jonsson, K., Kittler, J., Li, Y., Kotropoulos, C., Tefas, A., Pitas, I., Tan, T., Yan, H., Smeraldi, F., Bigun, J., Capdevielle, N., Gerstner, W., Ben-Yacoub, S., Abdeljaoued, Y., Mayoraz, E.: Comparison of face verification results on the XM2VTS database. In: Proceedings of the 15th International Conference on Pattern Recognition, Barcelona (Spain), September 2000. Volume 4., IEEE Comp. Soc. Order No. PR00750 (2000) 858–863
19. Joachims, T.: 11 of Advances in Kernel Methods - Support Vector Learning. In: Making Large-scale SVM Learning Practical. MIT Press (1998) Eds. B. Schölkopf, C. J. C. Burges, A. J. Smola.
20. Osuna, E., Freund, R., Girosi, F.: Improved training algorithm for Support Vector Machines. In: Proceedings of IEEE NNSP'97. (1997) 276–285