Face Recognition Using Independent Component Analysis and Support Vector Machines *

O. Déniz †, M. Castrillón, M. Hernández
Universidad de Las Palmas de Gran Canaria
Departamento de Informática y Sistemas
Edificio de Informática y Matemáticas
Campus Universitario de Tafira
35017 Las Palmas - SPAIN
{odeniz,mcastrillon,mhernandez}@dis.ulpgc.es

* Work partially funded by DGUI-Gobierno Autnomo de Canarias *PI1999/153* research project.
† Author supported by graduate grant *D260/54066308-R* of Universidad de Las Palmas de Gran Canaria.

Face Recognition Using Independent Component Analysis and Support Vector Machines *

Abstract

Support Vector Machines (SVM) and Independent Component Analysis (ICA) are two powerful and relatively recent techniques. SVMs are classifiers which have demonstrated high generalization capabilities. ICA is a feature extraction technique which has been mainly used on the problem of blind signal separation. In this paper we combine these two techniques for the face recognition problem. Experiments were made on two different face databases, achieving very high recognition rates. As the results using the combination PCA/SVM were not very far from those obtained with ICA/SVM, our experiments suggest that SVMs are relatively insensitive to the representation space.

Keywords: face recognition, ICA, SVM.

1 Introduction

The face recognition problem has attracted much research effort in the last years. Although it has proven to be a very difficult task even for frontal faces, certain algorithms can perform well under constrained conditions. The most prominent work was [1], which introduced the *eigenfaces* method, widely used as a reference. Recent advances in statistical learning theory, and in particular the introduction of Support Vector Machines [2] have made it possible to obtain very high accuracies for the object recognition problem. Independent Component Analysis [3] is also a relatively recent technique which has been mainly applied to *blind signal separation*, though it has been successfully applied to the face recognition problem too.

* Work partially funded by DGUI-Gobierno Autnomo de Canarias *PI1999/153* research project.

This paper is organized as follows. In the next two sections we give a brief introduction to ICA and SVM. In section 4 we describe the experiments carried out. Section 5 concludes by presenting future directions of research.

2 Independent Component Analysis

Independent Component Analysis is a technique for extracting statistically independent variables from a mixture of them [3]. ICA has been successfully applied to many different problems such as MEG and EEG data analysis, finding hidden factors in financial data and face recognition (see [4] for an introduction and applications).

The ICA technique aims to find a linear transform for the input data using a basis as statistically independent as possible. Thus, ICA can be considered a generalization of Principal Component Analysis (PCA). PCA tries to obtain a representation of the inputs based on uncorrelated variables, whereas ICA provides a representation based on statistically independent variables. Figure 1 shows the difference between PCA and ICA basis images.

Figure 1: Some original (left), PCA (center) and ICA (right) basis images for the Yale Face Database (see section 4).

In the context of face recognition, ICA has been showed to produce better results than those obtained with PCA, see [5] for a comparison. As opposed to PCA, ICA does not provide an intrinsic order for the representation coefficients of the face images. In [6] the best results were obtained with a order based on the ratio of between-class to within-class variance for each coefficient $r = \sigma_{between}/\sigma_{within}$, where $\sigma_{between} = \sum_j (\bar{x}_j - \bar{x})^2$ is the variance of the j class means and $\sigma_{within} = \sum_j \sum_i (x_{ij} - \bar{x})^2$ is the sum of the variances within each class.

3 Support Vector Machines

We only give here a brief presentation of the basic concepts needed. The reader is referred to [7] for a more detailed introduction and to [8] for a list of applications of SVMs. SVMs are based on structural risk minimization, which is the expectation of the test error for the trained machine. This risk is represented as $R(\alpha)$, α being the parameters of the trained machine. Let l be the number of training patterns and $0 \leq \eta \leq 1$. Then, with probability $1 - \eta$ the following bound on the expected risk holds [2]:

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

$R_{emp}(\alpha)$ being the empirical risk, which is the mean error on the training set, and h is the VC dimension. SVMs try to minimize the second term of (1), for a fixed empirical risk.

For the linearly separable case, SVM provides the optimal hyperplane that separates the training patterns. The optimal hyperplane maximizes the sum of the distances to the closest positive and negative training patterns. This sum is called *margin*. In order to weight the cost of missclassification an additional parameter is introduced. For the non-linear case, the training patterns are mapped onto a high-dimensional space using a kernel function. In this space the decision boundary is linear. The most commonly used kernel functions are polynomials, exponential and sigmoidal functions.

4 Experiments

In order to establish the performance of ICA/SVM, in comparison with other schemes, we carried out experiments on two independent face databases, the Yale Face Database [9], and a randomly chosen subset of the AR Face Database [10]. The Yale Face Database contains 165 images (11 per individual), with changes in facial expression, occlusion, and illumination conditions. From the AR Face Database we used 300 face images (12 per individual), with changes in facial expression and illumination conditions, and images taken in two sessions two weeks apart. All the results were obtained using 2-fold (AR) and 5-fold (Yale) cross-validation and varying the number of coefficients used in the range 1-N, N being the number of training images. ICA coefficients were ordered according to the ratio r explained in section 2. All the images were previously converted to 256 gray levels and histogram equalization was applied. The background in the Yale images was manually removed with a rectangle. For the images of the AR database a more plausible normalization was accomplished. Besides histogram equalization, we also performed geometric normalization. The images were firstly cropped with an ellipse, thus removing hair and shoulders. The eyes and mouth were located manually and then the images were shifted both in x and y and warped in order to have the eyes and mouth in the same place for all the images.

The ICA algorithm we used in our experiments was FastICA [11]. FastICA provides rapid convergence and estimates the independent components by maximizing a measure of independence among the estimated original components.

The SVM classifier, as introduced in section 3, is a 2-class classifier. Therefore, we had to adapt it to our multiclass problem. There are two options: using N SVMs (N being the number of classes), separating one class from the rest, or using $N(N-1)/2$ SVMs, one for each pair of classes. As the accuracies between these two approaches are almost the same [12], we chose the first option, which is less complex. The SVM algorithm used in the experiments presented problems of convergence when the input coefficients had a relatively high magnitude. This may be alleviated by dividing the input coefficients by a constant value (Thorsten Joachims, personal communication).

The results obtained appear in table 1. SVM was used only with polynomial (up to degree 3) and gaussian kernels (varying the kernel parameter σ).

		NMC using	SVM			
		euclidean distance	p=1	p=2	p=3	Gaussian
Yale	PCA	92.73 %	98.79 %	98.79 %	98.79 %	99.39 %
	ICA	95.76 %	99.39 %	99.39 %	99.39 %	99.39 %
AR	PCA	48.33 %	92 %	91.67 %	91 %	92.67 %
	ICA	70.33 %	93.33 %	93.33 %	92.67 %	94 %

Table 1: Recognition rates obtained for Yale and AR images, using the Nearest Mean Classifier (NMC) and SVM. For SVM, a value of 1000 was used as missclassification weight. The last column represents the best results obtained varying σ. Note that the combination PCA-NMC corresponds to the *eigenfaces* method.

For the Yale Database, there is no clear difference between ICA/SVM and PCA/SVM. We postulate that this is due to the fact that the classification error is too close to zero, which does not allow for differences to be seen clearly. As for the AR images, although the best absolute results are obtained with ICA and SVM, the performance is not far from that obtained with PCA and SVM. This is consistent with the results reported in [13], which suggested that SVM is relatively insensitive to the representation space.

Figure 2 represents the cumulative rank error and classification error as a function of the number of coefficients for the AR face set.

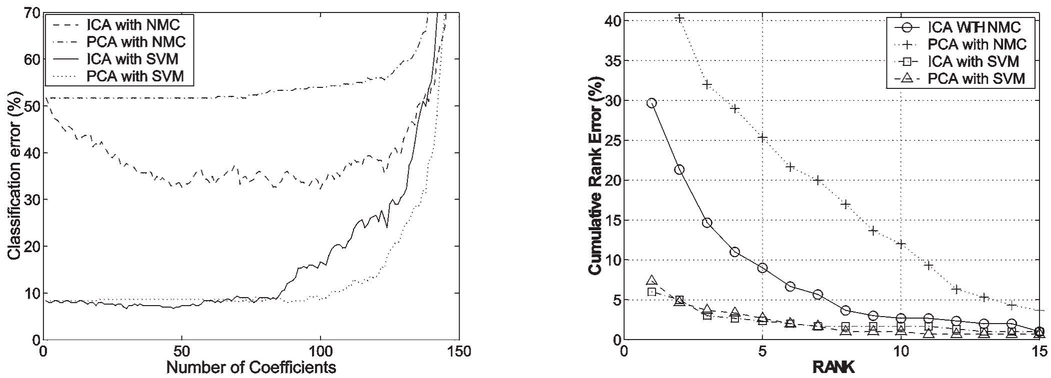

Figure 2: Cumulative rank error and classification error as a function of the number of coefficients for the AR face set (SVM with the best absolute results obtained).

5 Conclusions and Future Work

We obtained experimental results showing that very high recognition rates can be achieved using ICA/SVM, although PCA/SVM also gave good results. Thus, evidence was given for the fact that SVMs are relatively insensitive to the representation space, which is in accordance with the results reported in [13], giving more importance to the trade-off between complexity and performance.

In future work we plan to use techniques such as SVM in the dynamic face recognition problem, the objective being the recognition of face sequences. The robustness of such system can be improved using temporal context.

Acknowledgments

The authors would like to thank Dr. Marian Stewart Bartlett for her interesting comments on ICA. Thanks are also due to Prof. Dr. Robert Duin, David Tax, Thorsten Joachims and Prof. José Javier Lorenzo Navarro for valuable advice during the writing of this paper.

References

[1] M. A. Turk and A.P Pentland. Eigenfaces for Recognition. *Cognitive Neuroscience*, 3(1):71–86, 1991. ftp://whitechapel.media.mit.edu/pub/images/.

[2] V. Vapnik. *The Nature of Statistical Learning Theory.* Springer, N.Y., 1995.

[3] A. Bell and T. Sejnowski. An Information Maximization Approach to Blind Separation and Blind Deconvolution. *Neural Computation 7*, pages 1129–1159, 1995.

[4] Aapo Hyvärinen and Erki Oja. Independent Component Analysis: A Tutorial. http://www.cis.hut.fi/~aapo/papers/IJCNN99_tutorialweb/, 1999.

[5] Chengjun Liu and Harry Wechsler. Comparative Assessment of Independent Component Analysis (ICA) for Face Recognition. Second International Conference on Audio and Video-based Biometric Person Authentication, March 1999.

[6] Marian Stewart Bartlett and Terrence J. Sejnowski. Independent Components of Face Images: a Representation for Face Recognition. In *Procs. of the 4th Annual Joint Symposium on Neural Computation*, Pasadena, CA, May 1997.

[7] C.J.C. Burges. A Tutorial on Support Vector Machines for Pattern Recognition. *Data Mining and Knowledge Discovery*, 2(2):121–167, 1998.

[8] SVM application list. http://www.clopinet.com/isabelle/Projects/SVM/applist.html.

[9] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman. Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):711–720, 1997.

[10] A.M. Martinez and R.Benavente. The AR face database. Technical Report 24, CVC, June 1998.

[11] The FastICA MATLAB package. Available at http://www.cis.hut.fi/projects/ica/fastica/.

[12] Olivier Chapelle, Patrick Haffner, and Vladimir Vapnik. SVMs for Histogram-based Image Classification. *IEEE Trans. on Neural Networks*, 10(5):1055–1065, September 1999.

[13] K. T. Jonsson. *Robust Correlation and Support Vector Machines for Face Identification.* PhD thesis, School of Electronic Engineering, Information Technology and Mathematics, University of Surrey, UK, March 2000.