ESANN'1999 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 21-23 April 1999, D-Facto public., ISBN 2-600049-9-X, pp. 195-200

Face Identification using Support Vector Machines

Rodrigo Fernández, Emmanuel Viennet

LIPN, Institut Galilée-Université Paris 13
Avenue J.B. Clément 93430 Villetaneuse France
rf@lipn.univ-paris13.fr, viennet@lipn.univ-paris13.fr

Abstract. The Support Vector Machine (SVM) is a statistic learning technique proposed by Vapnik and his research group [8]. In this paper, we benchmark SVMs on a face identification problem and propose two approaches incorporating SV classifiers. The first approach maps the images in to a low dimensional features vector via a local Principal Component Analysis (PCA), features vectors are then used as the inputs of a SVM. The second algorithm is a direct SV classifier with invariances. Both approaches are tested on the freely available ORL database. The SV classifier with invariances achieves an error of 1.5%, which is the best result known on ORL database.

1. Introduction

Face identification is an important field of research with many possible applications. On the other hand, SVMs have been applied successfully to pattern recognition problems like OCR[6] and face detection[5][9]. In order to contribute to the development of this promising technique, we think it is interesting to benchmark SVMs with a database widely used in face recognition research. This is the goal of this work. Therefrom, we propose two face identification systems incorporating SVMs: a local feature extraction module which feeds a SVM and a direct SVM with invariances.

The rest of this paper is organized as follows. Section 2 presents the data base ORL and an brief overview of previous work on it. The approaches we propose are detailed in section 3. The final section is devoted to the discussion of the results.

2. Data Base

In this work, we have used the ORL database[1], a set of pictures taken between 1992 and 1994 at Olivetti Research Laboratory. There are images of 40 different persons, 10 images were taken of each person. The series of 10 images presents variations in facial expression, in facial position (there are slightly rotated faces) and in some other details like glasses/no-glasses. All the photos were taken with the persons in a frontal position against a dark background, there are small

[1]Available at `http://www.orl.co.uk/facedatabase.html`

ESANN'1999 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 21-23 April 1999, D-Facto public., ISBN 2-600049-9-X, pp. 195-200

Figure 1: Some images belonging to ORL database. The base is composed by 400 images, there are images of 40 different persons (10 images per person).

Algorithm	error	Algorithm	error
PDBNN	4.0	SOM+CN	3.8
Top-down-HMM	13.0	Pseudo-2d-HMM	5.0
Eigenfaces	10.0	L1-1NN	3.8
n-tuple	14.0	cont-n-tuple	**2.7**

Table 1: Published error rates obtained by several algorithms on ORL database. The reported test error is the average of 5 experiences. The best result is obtained by the Continuous-n-tuple classifier proposed by Lucas.

variations of the background gray level also. The images are 256 gray levels with a resolution of 92x112 pixels. The figure 1 shows a subset of the ORL database.

ORL images have good quality and their size (92x112 pixels) is enough to implement preprocessing modules like local filtering or local feature extraction [1]. However, from the viewpoint of the classification problem, ORL is challenging due to the large number of individuals to identify (40) with respect to the little amount of images per person (10, usually 5 for learning and 5 for test).

Since 1994, ORL has been used to benchmark many face identification systems. We compile in table 1 the performances of different approaches on the 40-classes identification problem. The PDBNN method (*Probabilistic Decision-Based Neural Network*) has been proposed by Lin and al. in [3]; Lawrence and al. in [1], carry a series of experiments with several algorithms, they propose mainly an adaptation of the Hidden Markov Model (HMM) algorithm and a mixture of a non-supervised and a convolutional neural network learning systems (SOM+CN); in the same article, the performance of the well known Eigenfaces algorithm is reported; lastly, we show the results obtained by Lucas' n-tuple and cont-n-tuple classifiers proposed in [4]. Lucas reported also the performance of a Nearest Neighbour classifier based on the L_1 distance (L1-1NN).

Lawrence et al. [1] report the error rate of one of their systems as a functions of the number of subjects considered, with the following results: 1.3% for 10 persons, 4.33% for 20 persons and 5.75% for 40 persons. This result illustrate well the kind of challenge posed by ORL : it requires a classifier able to generalize from few examples.

ESANN'1999 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 21-23 April 1999, D-Facto public., ISBN 2-600049-9-X, pp. 195-200

3. Support Vector Machines

The SVM algorithm, proposed by Vapnik in 1995 uses the Structural Risk Minimization (SRM) Principle to build decision rules with good generalization properties. A good and detailed introduction to SVMs can be found in [8]. SVMs use decision functions of the form:

$$f(x) = sgn\left(\sum_{x_i \in SV} \lambda_i K(x_i, x) + b\right). \tag{1}$$

where $K(x, y)$ is a (possibly non-linear) kernel, and SV is a subset of the training examples, the so-called *support vectors*. The key ideas underlying SMS are: implement SRM by minimizing a worst-case bound on the generalization error; use a non linear kernel K to implicitly map inputs to a high dimensionnal feature space, in which data will be linearly separated. The support vectors x_i and weights λ_i are found by solving a constrained quadratic programming problem.

In our experiments, three types of kernels are used. The first one is the well known polynomial kernel $K(u, v) = \left(\frac{u \cdot v}{m}\right)^d$, here, m is the dimension of the input space. The other two are based on the distance between patterns. We have used L_1 and L_2 distances, which allows us to the following kernels:

- RBF1: $K(u, v) = e^{\frac{\sum_{i=1}^{n} |u_i - v_i|}{\sigma \cdot m}}$,
- RBF2: $K(u, v) = e^{\frac{\sum_{i=1}^{n} (u_i - v_i)^2}{\sigma \cdot m}}$.

4. Classification Systems

This section presents different approaches that use SVMs to identify the faces.

4.1. Local PCA-SV Classifier

This algorithm, based on the one proposed by Lawrence [2] et al.[1], can be split in four modules:

1. **Re-scale**: The images are re-scaled to a size of 43x51 pixels. This is done by a simple linear interpolation algorithm.
2. **Local Sampling**: The 43x51 is transformed in to a list of local windows, each window is 5x5 pixels and the step between neighbors windows is 4 pixels both in horizontal and vertical directions. There are 130 windows to analyze. Let's note w_{ij}^k the window centered on the pixel (i, j) of the k-th image.
3. **Local PCA**: The p first principal components of each window are extracted. PC Analysis is done independently for each position (window) of the image, for example, to compute the principal components of the

[2] The size of the re-scaled image and the size of the mask for local sampling are the same proposed by Lawrence.

ESANN'1999 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 21-23 April 1999, D-Facto public., ISBN 2-600049-9-X, pp. 195-200

window w_{ij}, one use the data set $\{w_{ij}^k : k = 1, \cdots, l\}$. This procedure allows us to carry out 130 different PCAs. After 1,2,3 the original 92x112 are coded in to a vector of $p \times 130$ components.

4. **SV Classifier**: The features vectors obtained in 1,2,3 feed a standard SV classifier.

This processing gives a very fast classifier: to get the principal components of a window, one only has to compute dot products in $\mathbb{R}^{25}$ which has low computing cost.

4.2. Direct SV Classifier

The simplest approach is to plug the original images into a SV classifier. This experience gives an idea of its accuracy without preprocessing the data. Together with the direct SVM we have tested the performances when a little preprocessing is incorporated: re-scale the images in order to reduce its size.

4.3. Invariant SV Classifier

This approach is basically the same as direct SV classifier. The only difference, and important one, is that here, SVMs with invariances are incorporated in order to get a more robust decision surface. Prior knowledge about the problem is incorporated by applying transformations to learning examples[3]. Six transformations are considered, the four standard translations and two zoom deformations (zoom and unzoom). Figure 2 shows an original image and its six transformations. The generation of artificial examples grows the training set from 200 to 1400 examples, which is still quite small for standard SV solvers. The algorithm is summarized as follows.

1. **Synthesis**: An extended training set is generated by applying transformations to the examples.

2. **Re-scale**: The images are re-scaled to a size of 20x25 pixels. This is done by a simple linear interpolation algorithm. The size 20x25 is justified by the results obtained by direct approach.

3. **SV Classifier**: The images obtained in 1,2 feed a SV classifier.

Interested by the excellent performances on OCR obtained by the 1-NN classifier based on the Tangent Distance (TD) (see Simard et al. [2]), we decided to test it on the ORL base. Using 20x25 pixels images, 1-NN-TD reaches an error rate of 4.0%, which is not better than the performance obtained by the faster and simpler 1-NN-L1 classifier (see table 1). We implemented also a SVM classifier provided with a RBF kernel based on the TD. The results are poor: the error rate is about 5% and the classifier is slow. Given some set of invariances to implement, computing TD based kernels is more expensive than use virtual examples. Thus, we decided not to go ahead with TD based SVMs. One possible explanation to the poor results is that TD is a pseudo-distance and, therefore, TD based RBF kernels do not satisfy Mercer's Condition.

[3]This could have been done by applying transformations only to support vectors as Schölkopf et al. suggested in[6], but we decided to apply transformations to the whole learning set for two reasons: ORL is a very little base (200 training examples) and, due to the sparsity of the data, in the original direct SV classifier almost every example is support of at least 2 (of 40) machines.

ESANN'1999 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 21-23 April 1999, D-Facto public., ISBN 2-600049-9-X, pp. 195-200

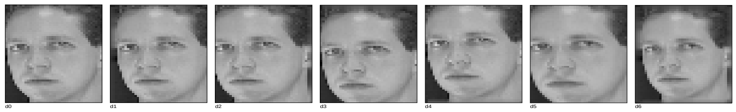

Figure 2: One ORL image and its 6 transformations. The four standard translations(up, down, right, left) and two zoom deformations (zoom and unzoom) are considered.

Direct SVM					
Resolution	14x18	20x25	28x35	40x50	92x112
Error	6.5	**5.5**	6.0	6.5	7.0

Table 2: Influence of the resolution over SVM performance. The optimal resolution is between 20x25 and 28x35 pixels. Note how a reduction of the resolution of the inputs can improve the classification rate. The kernel used is RBF2.

5. Results and Discussion

All experiments were performed with 5 training images and 5 test images per person, without overlapping between training and test sets. In all experiments the 40 individuals of the ORL database were considered. There are 200 original training images and 200 test images. In order to better estimate the accuracy of our classifiers, the results were averaged over 5 random selections of training and test sets.

Direct SV classifiers are tested with two goals: a) to have an idea of how good SVMs perform on the ORL database without preprocessing and b) to determine how the resolution of the images influence the performances and to get the optimal input size for the invariant SV classifiers. Results are shown in Table 2. Note how a reduction of the resolution of the inputs can improve the classification rate.

Using virtual examples improve the performances as the Table 3 shows. In this experiment, two kinds of kernels were used: RBF1 and RBF2, remark how the former, based on the L_1 distance, outperforms the more popular RBF kernel based on the L_2 distance.

The Table 3 shows the results obtained by this approach as a function of the number of principal components p.

The results indicate that SVMs are very suitable for face identification tasks. In fact, the 1.5% error rate obtained by the invariant SVM can be considered as the state of the art performance on ORL. Besides this very accurate system, we have proposed a fast classifier based on local feature extraction which attains a 3.7% error rate, this approach is comparable to the Lawrence's CNNA 3.8% and is significantly faster than the Lucas' Continuous-n-tuple (2.7%).

Some final remarks. The RBF1 kernel outperforms manifestly the Euclidean RBF2 kernel, this result should be kept in mind when choosing kernels for future work. Note also that classification time of the SV classifiers (mainly the one generated from the extended training set) can be reduced using the Reduced Set method proposed by Burges [6]. Another possible issue for future work is to test the Kernel-PCA algorithm [7], instead of the linear PCA to extract local features of the images. SVMs have been proved to be robust to overftting,

ESANN'1999 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 21-23 April 1999, D-Facto public., ISBN 2-600049-9-X, pp. 195-200

Local-PCA SVM	
POLY (d=2)	
p	error
1	3.7
2	3.7

Invariant SVM			
RBF2		RBF1	
σ	error	σ	error
0.1	2.8	0.1	1.7
0.3	2.7	0.3	**1.5**

Table 3: Error rates obtained by Local-PCA SVM and Invariant SVM approaches. The best result is 1.5% obtained by the Invariant SVM with a L_1 RBF kernel. However, the Local-PCA SV classifier is considerably faster and its performance is honorable (3.7%).

the results presented in this paper show that they generalize also well when data are sparse, which confirms the good regularity properties of large margin classifiers.

References

[1] S. Lawrence, C. Giles, A. Tsoi, A. Back. Face recognition: a convolutional neural network approach. *IEEE Transactions on Neural Networks*, vol 8, 98-113, 1997.

[2] P. Simard, Y. Le Cun, J. Denker. Efficient pattern recognition using a new transformation distance. *Advances in Neural Information Processing Systems*, 5, 1993.

[3] S. Lin, S. Kung, L. Lin,. Face recognition/detection by a probabilistic deciaion-based neural network. *IEEE Transactions on Neural Networks*, vol 8, 114-132, 1997.

[4] S.M. Lucas. Face recognition with the continuous n-tuple classifier. *British Machine Vision Conference*, 1997.

[5] E. Osuna, R. Freund, and Girosi F. Training support vector machines:an application to face detection. *Proceedings CVPR, Puerto Rico*, 1997.

[6] B. Schölkopf and C. Burges. Improving the accuracy and speed of support vector machines. *Proceedings Neural Information Processing Systems*, Vol. 9. MIT Press, Cambridge, MA, 1997.

[7] B. Schölkopf, A. Smola, and K-R. Müller. Nonlinear component analysis as a kernel eigenvalue problem. Technical report, Technical Report 44, MPI für biologishe Kybernetik, 1996.

[8] V. Vapnik. *The Nature of Statistical Learning Theory.* Springer, 1995.

[9] Emmanuel Viennet and Francoise Fogelman Soulié. Connectionnists methods for human face processing. In H. Wechsler, J. Phillips, and T.S. Huang, editors, *Face Recognition : From Theory to Applications*, volume 68 of *NATO ASI Series in Computer and Systems Sciences.* Springer Verlag, 1997.