Some Experiments On Face Recognition With Neural Networks

Enrique Cabello[1], Araceli Sánchez[1] and Luis Pastor[2]

[1] Universidad de Salamanca, Facultad de Ciencias, Dpto de Informática y Automática, Plaza de la Merced s/n, 37008 Salamanca, Spain
[2] Universidad Rey Juan Carlos / Universidad Politécnica de Madrid, Fac. de Informática, Dep. de Tec. Fotónica, 28860 Madrid, Spain

Abstract. This paper presents some results on the possibilities offered by neural networks for human face recognition. In particular, two algorithms have been tested: learning vector quantization (LVQ) and multilayer perceptron (MLP). Two different approaches have been taken for each case, using as input data either preprocessed images (gray level or segmented), or geometrical features derived from a set of manually introduced landmarks. The preprocessing steps included resolution reduction and segmentation. For the geometrical features´ case, a Karhunen-Loeve expansion was used to extract features among the different possibilities offered by 14 landmark points.

For the experiments, a database composed of 300 images was used. The pictures correspond to 10 frontal, inclined o rotated views from thirty male persons of similar age and race. If gray level images are used as input data, the experimental results show higher recognition rates for LVQ than for MLP (96.7% versus 83.3%). Applying a previous segmentation stage strongly decreases the recognition rates. For geometrical features, the situation is reversed: MLP yields better results than LVQ (93.3% versus 84.4%).

Keywords. Neural Networks, Face Recognition, Multilayer Perceptron, Learning Vector Quantization

1 Introduction

Computer face recognition is a topic that has been receiving increasing amounts of attention during the last years. In particular, this last decade has witnessed a renewed interest, resulting in large increases both in the number of research centres and personnel involved in this problem and in the methods and techniques proposed to cope with it [3] [9] [12] [5]. Consequently, commercial systems have started to appear.

There are several reasons that explain the present interest in computer processing of human face images. First, there are strong economical reasons: reliable face processing and recognition algorithms will find a myriad of commercial applications. Among them, the following can be pointed out [3]:

- Security. Enabling authorised people to access restricted areas. Detecting particular persons in sensitive areas that are specifically forbidden to them.
- Services (access to information or services). Automatic bank teller machines. Credit card owner identification. Access to computers and networks. Verification of user's identity for accessing medical services, etc.
- Law enforcement. Passport control. Person identification against a face database. Portrait sketching from witnesses' descriptions.

There are other somewhat related fields that can be mentioned, such as forensic medicine (face reconstruction or person identification from skeletal remains), video conferencing (compression of human face video sequences), etc.

There are also academic or technical reasons spurring interest in facial image analysis: unrestricted, robust face recognition is an extremely demanding task that has attracted attention from researchers in disciplines as diverse as cognitive psychology, forensic medicine, computer vision, etc. The problem's difficulty stems both from the complexity of facial patterns as well as from the variability found on face images: a face is intrinsically a three-dimension entity, and therefore, two-dimensional pictures are largely affected by illumination and pose variations. Furthermore, the importance that face aspect and facial expressions play in social life makes us devote a lot of attention to our external appearance. Consequently, face images undergo deep transformations due to variations in hair style, length and colour; presence of eye glasses, beard and/or moustache; presence of make-up, etc. Temporal and/or gradual changes in weight and age additionally affect our appearance, making the problem more difficult yet (even our own vision system can easily misrecognize faces, especially when other cues such as voice or context are not available).

In general, face recognition applications require the use of different approaches either because of the problem itself (matching two known pictures or matching a picture against a - perhaps large - face database, etc.), or because of the availability of supplementary information such as a person's age, race and gender, additional pictures, etc. Furthermore, there can be differences in the way data is acquired (static pictures or live video; from controlled or unrestricted setups; with uniform or cluttered backgrounds; within a limited or a long-term time span; from cooperative or non-cooperative subjects, etc.) [3] [9]. This volume can give an updated overview of approaches and application areas.

Our paper presents some results for human face recognition using different neural networks (NN). For each case, the NN was fed with two different kinds of data: preprocessed gray level images, and feature vectors computed from manually extracted landmarks. The most important criterion used for comparing results is recognition accuracy, although other key aspects that have to be taken into consideration are the algorithms' robustness and flexibility, and the computational aspects involved in the classifier design and operation. The. following sections describe briefly other methods for human face recognition, the experiments performed and results achieved in this work, and the main conclusions drawn from our experimental results.

2 State Of The Art

Many different approaches have been suggested for face recognition, especially during the last years. The techniques proposed in the literature can be classified according to the following scheme [3]:

Geometrical features: computing geometrical features such as angles, indexes or distances on human faces permits the straightforward application of statistical pattern recognition techniques [3] [9]. The main weakness of the geometrical features approach lies in the feature computation stage: current algorithms for the automatic location of landmarks are not consistently accurate.

Eigenfaces: Turk and Pentland [11] presented a face recognition scheme in which face images were projected onto the principal components of the original set of training images. A related technique ("Fisherfaces") has recently been proposed by Belhumeur et al [1].

Template Matching: Brunelli and Poggio [2] and Yullie et al. [13] performed direct correlations of image segments. In general, template matching is effective when the test images have the same scale, orientation and illumination as the training set.

Neural networks: Lawrence [7] proposes a system with a local image sampling, a self organization map (SOM) neural network and a convolutional neural network. Also, Lin [8] considered a probabilistic decision-based neural network for face detection and recognition. A review of connectionist approaches to face analysis can be found in Valentin et al [12].

Many other techniques, such as von der Malsburg´s jets [14] or Nastar et al´s deformable intensity surface models [15], have been developed during the last years. Some of them have been compiled in this volume and in two excellent surveys [3], [9]. Additionally, a recent issue of IEEE Transactions on Pattern Analysus and Machine Intelligence [5] has been devoted to face and gesture recognition.

3 Experiment Description

3.1 Data Set Description

The data base used in this study [4] is composed of 30 subjects, each of which has ten frontal or rotated images (Figure 1). All of the available pictures correspond to men of similar ages. Each person's set of images is composed of two frontal images, two with the face looking up, two looking down, two with the face rotated to the right and two rotated to the left. All of the pictures were taken with a white background and with a controlled lighting. The resolution of each image is 512x342 pixels with 256 gray levels, with the faces covering most of the picture area. In total, 300 images were available.

Fig. 1. Set of images for one subject

3.2 Classification Approaches

The results presented here have been obtained using two basic methods: multilayer perceptron (MLP), and learning vector quantization (LVQ). In both cases, two kinds of data have been fed to the classifiers: reduced resolution images (gray level or segmented), and feature vectors.

The first method used was based on a MLP. Some experiments were performed using a topology of two hidden layers with 674, 75, 15 and 30 neurons in the input, two hidden and output layers, respectively. Some other experiments have used a MLP with only one hidden layer, with a topology of 674, 100 and 30 neurons in the input, hidden and output layers. The second method implemented was LVQ (learning vector quantization, [6]). This method works like a 1-NN classifier differing on how the set of labelled patterns is formed: this set is typically obtained by clustering the training data (to reduce the number of labelled patterns), and then using a supervised learning algorithm to move the cluster centres into positions that minimize the classification error. Usually several codebook vectors are assigned to each class, and each test pattern is assigned to the class with the nearest codebook vector. One, two or three codevectors by class have been used here.

For geometrical characteristics, the same two methods (MLP and LVQ) have been tested. Each of them has been fed with different number of features, provided by a Karhunen-Loeve expansion. For MLP, Table I shows the number of neurons in the input layer (equal to the number of selected features) and the number of neurons in the hidden layer. The number of neurons in the output layer is always 30. For LVQ, one, two or three codevectors by class were considered for 5, 10 and 15 features.

Table 1. Number of neurons considered for MLP using geometrical characteristics

Number of neurons (input layer)	5	6	7	8	9	10	11	12	13	14	15
Number of neurons (hidden layer)	18	18	19	19	20	20	21	21	22	22	23

3.3 Data Preprocessing

The use of gray level images permits the consideration of all the information available in each picture, but the amount of data found in raw images advises the use of preprocessing steps to decrease the computational requirements in the analysis task. In our study this was done by creating reduced resolution versions of each image. The process is conceptually similar to computing a Gaussian pyramid, working with levels which are relatively high up in the pyramid. Additionally, the input images were normalized to decrease the influence of the acquisition conditions. Promising results were obtained by using a resolution level of 32x22 pixels. Higher resolutions imply higher computation times, particularly for the NN training.

Once the reduced resolution images were computed, nine of them were selected for each subject as the training set, while the tenth was used for testing the system performance. For some experiments, the test images were frontal views, whereas for others, rotated views were used to check the recognition system's robustness to changes in the viewing angle. In a different experiment, the reduced resolution images were also segmented before being fed to the recognition system (Figure 2). It should be pointed out that, in order to make results comparable, the same data set was used: in both cases, the same nine reduced resolution (gray level or segmented) pictures were used to train the MLP or to form the codebook vectors, using the tenth image for testing the methods' performance.

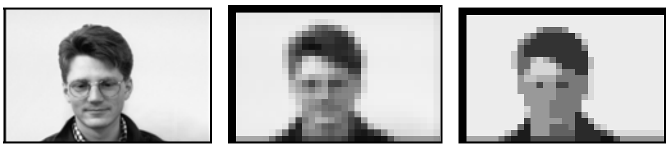

Fig. 2. Original, reduced resolution and segmented images for one subject

Regarding the use of geometrical characteristics for recognition, it is necessary first to define the set of landmarks, and then, to locate their actual position over the face images and to compute the features to be entered to the classification stage. In order to choose the landmarks, a preliminary study was performed over a set of common somatometric points: a group of operators repeatedly introduced the landmarks' positions in different frontal images at different times. The points that showed high placement variability (Figure 3) were discarded, as well as those that were easily occluded by facial hair or small face rotations. Finally, a set of 14 landmarks was selected for the tests. It has to be noted that the manual placement of landmarks permitted the separation of the recognition problem from that of landmark location, a research field in itself.

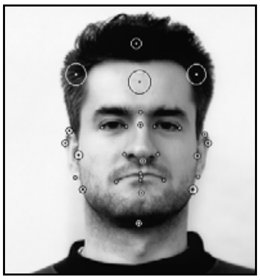

Fig. 3. The initial landmark set for one of the persons in the database (the circles' radii are estimates of placement variability).

With the 14 selected landmarks, 47 different features were computed, including the most commonly used ratios in human identification through cranial measurements (most of these measurements represented distances which had been normalized by the distance between the eyes). To analyze the information contained in these features, a Karhunen-Loeve expansion was performed in order to obtain feature vectors with lower dimensions; these vectors were introduced afterwards to the classifiers. As expected, decreasing the feature vectors' dimension did not have strong effects until certain limit was reached.

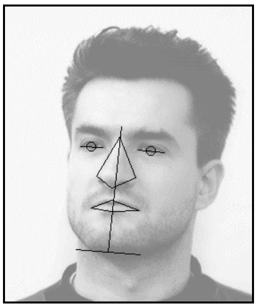

Fig. 4. Selected landmarks for one subject

4 Results And Discussion

The experimental results are summarized in Tables 2 to 4 and in Figures 5 and 6. Table 2 presents the percentages of correct recognition achieved for gray level images and MLP (with 674, 75, 15 and 30 neurons in the input, two hidden and output layers), or LVQ (with one, two or three codevectors per class). The images are reduced resolution versions of the original database pictures, including frontal and rotated views, and they were computed by gaussian averaging and downsampling. In some tests, these coarse images have additionally been normalized or segmented.

Table 2. Percentage of successful recognition for gray level images

Input image	MLP 2 hidden layers	LVQ 1 codevector per class	LVQ 2 codevectors per class	LVQ 3 codevectors per class
Gauss (frontal)	43.3	96.6	96.6	96.6
Gauss (rotated)	26.6	83.3	96.6	96.6
Normalized (frontal)	83.3	93.3	96.6	96.6
Normalized (rotated)	80.0	90.0	93.3	93.3
Gauss & Segment.	None	40.0	40.0	40.0

Figure 5 shows the evolution of the SSE (Sum Square Error) during the training stage of the referred MLP, for normalized and non-normalized input images.

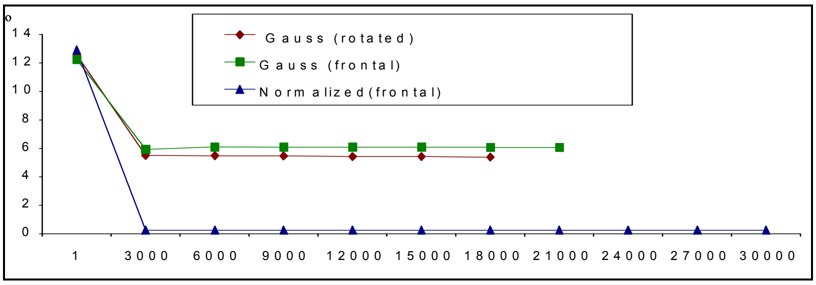

Fig. 5. SSE/o_units for the MLP (with 674, 75, 15 and 30 neurons).

Some tests were also fulfilled to compare the performance of the MLP approach when one or two hidden layers were used. For that purpose, a MLP with 100 neurons in the hidden layer was tested. The results for reduced resolution, non-normalized images are summarized in table 3.

Table 3. Percentage of successful recognition for different MLP topologies

	MLP (one hidden layer)	MLP (two hidden layers)	LVQ (two codevectors per class)
Gauss	73.3%	43.0%	96.0%

Table 4 outlines the main results achieved for the geometrical features case. Like in Table 2, the results have been produced by MLP (with one hidden layer, in this case) and LVQ (with one, two or three codevectors per class). Figure 6 depicts the evolution of success rates when the number of provided features is increased.

Table 4. Percentage of successful recognition for geometrical characteristics

Number of features provided by KL	MLP	LVQ (1 codevector per class)	LVQ (2 codevectors per class)	LVQ (3 codevectors per class)
5	44.4	36.3	45.5	40.0
10	84.8	77.0	78.5	77.7
15	93.3	83.7	81.8	84.4

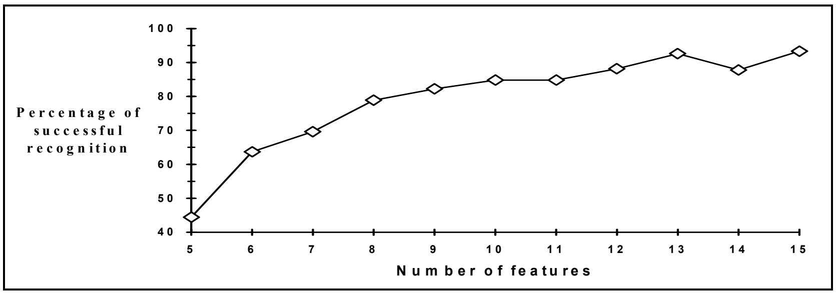

Fig. 6. Percentage of successful recognition versus number of features (number of neurons in the input layer) for MLP

An evaluation of Tables 2 to 4 shows that, for analyzing gray level images, better results were achieved with LVQ than with MLP. This result was consistently obtained for frontal or rotated views, independently of supplementary image preprocessing stages (normalization or segmentation). Additionally, Table 2 indicates that the methods (in particular, LVQ) were quite robust to small or medium head rotations, and that altering the images' gray levels by a segmentation process affected very negatively the methods' results. This can be due to either the gray level information lost in the segmentation process or to the influence that segmentation errors have at such a coarse resolution. For the geometrical features case, the situation was reversed: MLP (with one hidden layer) achieved higher success rates than LVQ. Regarding MLP topology, best results were achieved for only one hidden layer rather than for two (another advantage of using only one hidden layer was reduced training times). In any case, LVQ with gray level images has proven to be the best option tested here, including LVQ with geometrical features. This result is similar to Brunelli's [2], perhaps because gray level images contain more information than geometrical features. It should be noted that automatic or manual introduction of points has lead to similar results (90% in Brunelli [2] and 93.3 and 84.4% in our case).

5 Conclusions

Different NN approaches have been proposed in the literature for dealing with the human face recognition problem. This paper presents some results obtained using LVQ (learning vector quantization) and MLP (multilayer perceptron), fed with gray level images and geometrical features extracted from a set of 14 manually introduced landmarks. When using pictoric face information as input, LVQ behaved better than MLP, showing lower error rates and being more robust against changes introduced during the image preprocessing stage, as can be deducted from the results presented in Table II for segmented or non-normalized images. Furthermore, training times were much shorter for LVQ than for MLP. On the other hand, MLP achieved lower error rates when dealing with geometrical features.

The experimental results also show that, for the approaches considered here, analyzing gray level images produced better results than analyzing geometrical features, either because of the errors introduced during their extraction or because the original images have a richer information content. Increasing the number of geometrical features improved the results, although the process became more time consuming. The poor results achieved for segmented images are also remarkable. In Samaria [10] a similar result was concluded, noting that the process of face segmentation prior to recognition led to unpredictable results. Last, the results presented here indicate that LVQ and MLP are tolerant to small or medium head rotations, such as the ones considered in this work.

6 Acknowledgements

The work presented here was partially supported by the Spanish Commission for Science and Technology (CICYT: TAP94- 0305-C03-02) and Comunidad de Madrid (AE00263/95). Additionally, the help of Juan Alonso on some of the experiments is gratefully acknowledged.

7 References

[1] 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. Vol 19, Nº 7, p 711-720. Jul. 1997.

[2] R. Brunelli and T. Poggio. Face Recognition: Features versus Templates. IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol 15, Nº 10, p.1042-1052. Oct. 1993.
[3] R. Chellappa, C. L. Wilson and S. Sirohey. Human and Machine Recognition of Faces: A Survey. Proceedings of the IEEE. Vol. 83. Nº 5. pp 705-740 .May 1995.
[4] Facial Data Base. University of Bern (iamftp.unibe.ch). Switzerland. 1995.
[5] IEEE Transactions on Pattern Analysis and Machine Intelligence: Special issue on face and gesture recognition (July 1997).
[6] T. Kohonen. Self - Organizing Maps. Springer -Verlag. 1995.
[7] S. Lawrence, C. L. Giles, A. C. Tsoi and A. D. Back. Face Recognition: A Convolutional Neural Network Approach. IEEE Transactions on Neural Networks. Vol 8, Nº 1, p.98-113. Jan. 1997.
[8] S. Lin, S. Y. Kung and L. J. Lin. Face Recognition/Detection by Probabilistic Decision-Based Neural Network. IEEE Transactions on Neural Networks. Vol 8, Nº 1, p.114-132. Jan. 1997.
[9] A. Samal and P. A. Iyengar. Automatic Recognition and Analysis of Human Faces and Facial Expressions: A Survey. Pattern Recognition. Vol 25. Nº 1, pp 65-72.1992.
[10] F. S. Samaria. Face Recognition using Hidden Markov Models. PhD. Thesis.Trinity College. University of Cambridge. Cambridge. UK. 1994.
[11] M. Turk and A. Pentland. Face Recognition Using Eigenfaces. Proceedings Computer Vision and Pattern Recognition. pp 586-591. 1991.
[12] D. Valentin, H. Abdi, A. J. O'Toole and Garrison W. Cottrell. Connectionist models of face processing: a survey. Pattern Recognition, Vol. 27 No. 9, pp.1209-1230, 1994.
[13] A. L. Yulie, D. S. Cohen and P. W. Hallinan. Feature Extraction from Faces Using Deformable templates. Proceedings Computer Vision and Pattern Recognition. 1989.
[14] L. Wiskott, J-M Fellous, N. Krüger, C. von der Malsburg. Face Recognition by Elastic Bunch Graph Matching. IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol 19, Nº 7. p 775-789. Jul. 1997.
[15] C. Nastar, B. Moghaddam and A. Pentland. Flexible images: matching and recognition using learned deformations. Computer Vision and Image Understanding, Vol. 65, No. 2, Feb. 1997, pp. 179-191