Facial Expression Classification Using RBF AND Back-Propagation Neural Networks

R.Q.Feitosa[1,2],
M.M.B.Vellasco[1,2],
D.T.Oliveira[1],
D.V.Andrade[1],
S.A.R.S.Maffra[1]

1 – Catholic University of Rio de Janeiro, Brazil
Department of Electric Engineering

2 – State University of Rio de Janeiro, Brazil
Department of Computer Engineering

e-mail: [raul, marley]@ele.puc -rio.br, tuler@inf.puc-rio.br, [diogo, sam]@tecgraf.puc-rio.br

ABSTRACT

This article presents a facial expressions classification experiment using neural networks. The classification system is based on attributes extracted from human faces images using the principal component analysis (PCA) technique. Well-framed images were used in order to simplify the face detection on the image. Two different models of neural networks have been applied as classifiers: back-propagation and RBF networks. In the experiments for performance evaluation the networks achieved a recognition rate equal to 71.8% and 73.2% respectively for Back Propagation and RBF, which is consistent with the best results reported in the literature for the same data base and testing paradigm. An analysis of the confusion matrix suggest the combination of both networks for a better performance.

Keywords: PCA, facial expression recognition, neural networks

1 INTRODUCTION

The interest in systems for the automatic recognition of facial expressions has recently increased. Such systems are clearly relevant in studies of human behavior, since facial expressions are a manifestation of human emotions. Facial expressions also have an important role in the non-verbal communication among human beings. Studies indicate that the role of facial expressions many times surpasses the one of the actual words [1]. This has awakened the interest in many computer vision researchers, who are trying to develop more effective techniques for computer-human interaction.

Any automatic system for the recognition of facial expressions must deal with three basic problems: detection of the human face in a generic image, extraction of relevant attributes from the facial image; and finally the classification itself.

Locating a face in a generic image is not an easy task, which continues to challenge researchers. Once detected, the image region containing the face is extracted and geometrically normalized, usually maintaining a constant inter-ocular distance. References to detection methods using neural networks and statistical approaches can be found in [2] and [3]. This paper does not tackle the problem of face detection. All of the experiments presented in the next sections used well-framed face images as input.

The second problem concerns with the selection of a set of attributes that could represent appropriately the emotions expressed on the images. Among the proposed approaches for the selection of attributes [4] the Principal Component Analysis algorithm (PCA) has been frequently used [5].

Regarding the third problem, neural networks have been successfully used as classifiers on face recognition systems (as in [6], [7] and [8]). In Rosemblum et al., geometrical characteristics have been extracted from sequences of images and applied to a RBF (Radial Basis Function) neural network, acting as a classifier.

This paper evaluates the performance of two neural network algorithms for the automatic facial expressions recognition: Back-Propagation and RBF neural networks [9]. Unlike [6], the system proposed here utilizes well-framed, static images, obtained by a semi-automatic method. Instead of geometrical attributes, the principal components analysis have been applied to generate the vector of relevant attributes. Many experiments have been carried out in order to evaluate the performance of the system proposed.

The remaining of this paper is organized as follows. Section 2 describes the system proposed, presenting a brief description of the PCA technique and of neural networks. Section 3 describes the experiments that have been performed. The results are then shown in section 4, which is followed by the conclusions in section 5.

2 METHODOLOGY

2.1 System's General Architecture

The automatic system proposed for the recognition of facial expressions is composed of three stages: detection, extraction of attributes and classification, as shown in figure 1.

The first stage is performed by a semi-automatic method. The extraction of attributes is performed using Principal Component Analysis algorithm, as described in section 2.2. On the classification stage, two neural network models have been used: Back-Propagation (BP) e Radial Basis Functions (RBF) [9].

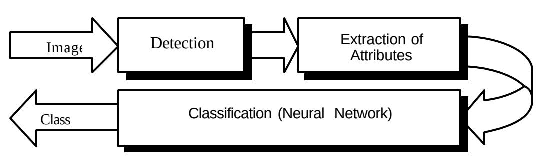

Figure 1: General Architecture of the Facial Expression Recognition System.

2.2 Using PCA for the Extraction of Attributes

The system presented in this work explores the concept of eigenfaces, proposed originally in [10] and extended in [11], [12], [13], [14] and [15]. An image having $n = N?M$ pixels can be seen as a point in an n-dimensional space. (PCA) identifies the orthonormal base of the subspace which concentrates most of the data variance.

An image is represented by a $n ? 1$ row vector obtained by concatenating the rows of each image matrix of size $n = N ? M$ pixels. Given a sample with K grayscale images, a training set $\{\mathbf{x}_i \mid 1 ? i ? K\}$ can be built. The base for the subspace is derived from the solution of the eigenvector/eigenvalue problem:

$$\mathbf{?} \; ? \; \mathbf{F}^{T} \mathbf{S} \mathbf{F} \,, \qquad (1)$$

where $\mathbf{S}$ is the sample covariance matrix, $\mathbf{F}$ is the $n ? n$ matrix of the eigenvectors of $\mathbf{S}$, and $\mathbf{?}$ is the corresponding diagonal matrix of eigenvalues. To reduce the dimensionality, only the p ($p < n$) eigenvectors associated with the p greatest eigenvalues are considered. The representation $\mathbf{y}$ in the p-dimensional subspace of a face $\mathbf{x}$ can be obtained by:

$$\mathbf{y} \; ? \; (\mathbf{x} \; ? \; \bar{\mathbf{x}}) \mathbf{F}_{p} \,, \qquad (2)$$

where $\mathbf{x}$ is the row face vector, $\bar{\mathbf{x}}$ is the corresponding sample mean and $\mathbf{F}_p$ is the $n ? p$ matrix formed by the p eigenvectors with the greatest eigenvalues. It can be proved that this approach leads to the p-dimensional subspace for which the average reconstruction error of the examples in the training set is minimal.

2.3 Neural Networks

Neural computing has re-emerged as an important programming paradigm that attempts to mimic the functionality of the human brain. This area has been developed to solve demanding pattern processing problems, like speech and image processing, which were intractable or extremely cumbersome when implemented using traditional computing [16].

By analogy with the human brain, *neural networks* are massively parallel systems that rely on simple processors and dense arrangements of interconnections [17]. These networks have demonstrated their ability to deliver simple and powerful solutions in areas that for many years have challenged conventional computing approaches.

A neural network is represented by *weighted* interconnections between processing elements (PEs). These weights are the parameters that actually define the non-linear function performed by the neural network. The process of determining such parameters is called training or learning [18], relying on the presentation of many training patterns. Thus, neural networks are inherently *adaptive*, conforming to the imprecise, ambiguous and faulty nature of real-world data.

2.3.1 Back-Propagation Networks

The most widely used neural network is the Back Propagation algorithm. This is due to its relatively simplicity, together with its universal approximation capacity [19].

The back-propagation algorithm defines a systematic way to update the synaptic weights of multi-layer perceptron (MLP) networks. The supervised learning is based on the gradient descent method, minimizing the global error on the output layer.

The learning algorithm is performed in two stages [9]: feed-forward and feed-backward. In the first phase the inputs are propagated through the layers of processing elements, generating an output pattern in response to the input pattern presented. In the second phase, the errors calculated in the output layer are then back propagated to the hidden layers where the synaptic weights are updated to reduce the error.

This learning process is repeated until the output error value, for all patterns in the training set, are below a specified value.

The definition of the network size (the number of hidden layers and of neurons in each layer) is a compromise between generalization and convergence. Convergence is the capacity of the network to learn the patterns on the training set and generalization is the capacity to respond correctly to new patterns. The idea is to implement the smallest network possible, so it is able to learn all patterns and, at the same time, provide good generalization.

The Back Propagation, however, has two major limitations: a very long training process, with problems such as local minima and network paralysis; and the restriction of learning only static input-output mappings [9]. To overcome these restrictions, new algorithms have been devised.

2.3.2 Radial Basis Functions

Radial Basis Functions (RBF) have attracted a great deal of interest due to their rapid training, generality and simplicity [20]. When compared with traditional multi-layer perceptrons, RBF networks present a much faster training, without having to cope with traditional Back Propagation problems, such as network paralysis and the local minima. These improvements have been achieved without compromising the generality of applications. It has been proved that RBF networks, with enough hidden neurons, are also universal approximators *[21]*.

The RBF network is based on the simple idea that an arbitrary function y(x) can be approximated as the linear superposition of a set of localized basis functions ?(x), leading to a structure very similar to the multi-layer perceptron. The RBF is basically composed of three different layers: the input layer, which basically distributes the input data; one hidden layer, with a radially symmetric activation function, hence the network's name; and one output layer, with linear activation function. For most applications, the gaussian format is chosen as the activation function ?(x) of the hidden neurons.

3 EXPERIMENTS

3.1 Image database

The database used in this work has been extracted from ATR Human Information Processing Research Labs. It consists of 213 grayscale images of 9 Japanese women. The database contains pictures with the six basic facial expressions, as defined in the Facial Action Coding System (FACS) [22]: disgust, fear, anger, sadness, surprise and happiness expressions. For each person, there are three or four images of each expression.

The figure below illustrates the six types of expressions used in this work.

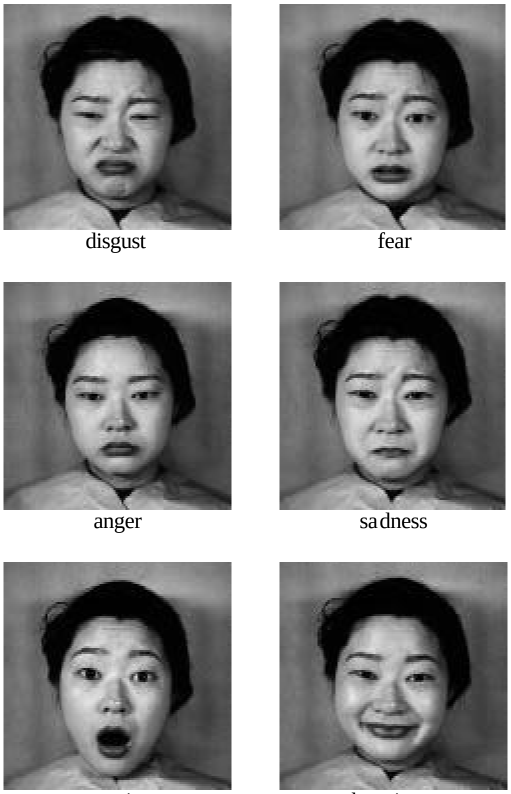

Figure 2: Facial Expressions

A large portion of the images of this database carries no relevant information for the classification procedure. So the image region covered by the face was extracted from the complete image, as shown in Figure 3. This could be easily implemented given that all images were well centered. After extraction, the faces were resized to 64x64 pixels, in order to reduce the computational load.

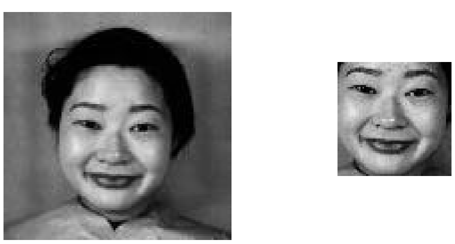

Figure 3: Extractig the image region containing the face

3.2 Dimensionality reduction using PCA

The rows of each image matrix produced by the pre-processing stage are concatenated in order to form line vectors. In this way, vectors with 4096 elements are created from the 64x64 face images.

The image set is divided in two subsets: one used for neural network training and the other one used for validation. An average vector of the training subset is then calculated and subtracted from all vectors, as required by the PCA algorithm (see equation 2).

The covariance matrix is then calculated from the training set. The p eigenvectors, correspondents to the p largest eigenvalues of this matrix, are then used to build the transformation matrix $\mathbf{F}_p$.

By applying the transformation matrix on each face vector, its projection on the attribute space is then obtained, which contains only p elements instead of the initial number of 4096.

3.3 Parameters

This work evaluates the impact of changing the values of some parameters. Several neural networks configurations have been tested.

The first parameter is the number of PCAs, which represents the number of inputs to the network. Values for this parameter ranged from 20 to 90.

Among the 9 persons in the database, the neural network training used 8 of them and the test phase used the remaining person, with the results averaged over all 9 expressers.

3.3.1 Back-Propagation (BP)

The Back-Propagation network used is composed of an input layer, a hidden layer and an output layer. The number of neurons in the hidden layer has been varied from 20 to 50. The number of neurons in the output layer was equal to 6, the number of classes. As mentioned in section 2.3.1, to overcome back-propagation limitations, like local minima, an adaptative learning rate and a momentum constant (0.5) has been used. The stopping criteria for the network training are the Sum of Squared Error (SSE) (1.0) and a maximum number of epochs (100,000).

3.3.2 Radial Basis Functions (RBF)

In RBF network experiments, two parameters were varied: radial basis function spread (1.0, 1.5, 2.0) and minimum squared error (MSE), from 1 to 10. The number of neurons in the hidden layer was chosen based on the MSE by the following rule: neurons are added in the hidden layers until it meets the specified MSE.

4 RESULTS

The objective of the parameter variation presented in the last section was to achieve an optimum network configuration. The results presented in this section were obtained with the best network configuration.

The parameters of the best Back Propagation network were: 30 neurons in the hidden layer and SSE = 0.1. For the RBF network the best configuration was: Radia basis function spread = 2.0 and MSE=2.0.

Figure 4 shows the graph obtained with Back Propagation and RBF varying the number of PCAs. The curves represent the average of the recognition rate to all persons of the database.

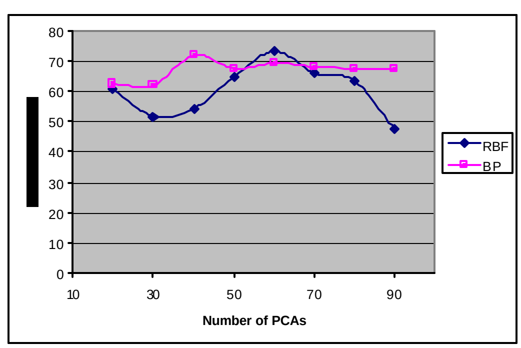

Figure 4: Average Recognition Rate Graphic

This graphic presents two interesting conclusions:

1. RBF seems to be considerably more sensitive in change of number of PCAs than Back Propagation.
2. Both networks presented approximately the same result in their best configuration (71.8% for Back Propagation with 30 PCAs and 73.2% for RBF with 60 PCAs).

Results from a recent work based on a completely different approach [23] used the same database and a similar testing paradigm for its evaluations. The authors report an average recognition rate equal to 75%, therefore slightly superior to our results.

The next two tables present confusion matrices for both networks considered in this work.

	Ang	Disg	Fear	Hap	Sad	Surp
Ang	**81.5**	3.7	3.7	3.7	7.4	0
Disg	3.7	**59.3**	0	11.1	25.9	0
Fear	0	0	**66.7**	7.4	11.1	14.8
Hap	0	0	11.1	**70.4**	3.7	14.8
Sad	7.4	7.4	3.7	0	**74.1**	7.4
Surp	0	0	18.5	3.7	0	**77.8**
Overall:		**71.8%**				

Table 1: Confusion matrix for Back Propagation

	Ang	Disg	Fear	Hap	Sad	Surp
Ang	**92.6**	7.4	0	0	0	0
Disg	50	**22.2**	0	0	27.8	0
Fear	0	3.7	**77.8**	0	11.1	7.4
Hap	0	0	37.0	**29.7**	14.8	18.5
Sad	0	0	0	0	**100**	0
Surp	0	0	0	0	0	**100**

Table 2: Confusion matrix for RBF

These matrixes allow to take some conclusions about the networks behavior:

1. Back Propagation is more stable than RBF among the classes. All classes had 60% (or higher) recognition rate, while in RBF 2 classes had 100%, but 2 other classes had recognition rate lower than 30%. In this sense the average values shown in figure 4 are misleading.
2. RBF is considerably more efficient for some classes. This suggests a future improvement to this system by combining both network responses to obtain a more efficient classifier.

5 CONCLUSIONS

The present work evaluated two automatic facial recognition systems, using PCA for dimensionality reduction followed by two neural networks models: Back Propagation e RBF. Experiments for performance evaluation were carried out on a data base having 213 grayscale images of 9 Japanese representing the six basic facial expressions. Best performance for the Back Propagation and for the RBF network was respectively 71.8% and 73.2%, slightly inferior to the best recognition performance reported in the literature based on the same database and for the same testing procedure. An analysis of the confusion matrices suggests that a combination of both networks would bring an important performance improvement.

Acknowledgment: We are thankful to ATR Human Information Processing Research Labs for making available the face database for our experiments.

6 REFERENCES

[1] EKMAN, P., FRIESEN, W.V., "*The Facial Action Coding System*," Consulting Psychologists Press Inc., 1978.

[2] ROWLEY, H. A., BALUJA, S.; KANADE, T., "*Neural Network-Based Face Detection,*" IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 20, No. 1, Jan 1998, pp. 23-38.

[3] SUNG, K.K, POGGIO , T., "*Example-Based Learning for View-Based Human Face Detection*, "IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 20, No. 1, Jan 1998, pp. 39-50.

[4] CHELLAPA, R., WILSON , C., SIROHEY, S., "*Human and machine recognition of faces: A Survey,*" Proceeding of the IEEE, 83 (5), May 1995, pp. 705-740.

[5] JOPHNSON, R. A.; D., WICHERN, W., *Applied Multivariate Statistical Analysis,* Prentice Hall, 4th Ed., 1998.

[6] ROSENBLUM, M., YACOOB, Y., DAVIS, L.S., "*Human Expression Recognition from Motion Using a Radial Basis Function Network Architecture,*" IEEE Trans. on Neural Networks, 7(5), 1996, 1121-1138.

[7] ZHANG , Z.; LYONS , M.; SCHUSTER, M., AKAMATSU , S., "*Comparison Between geometry-Based and Gabor-Wavelets-Based Facial Expression Recognition Using Multi-Layer Perceptron,*" Proceedings of the 2nd International Conference on Automatic Face and Gesture Recognition, Japan, April, 1998, pp. 545-459.

[8] KAMACHI, M., GYOBA , J., "*Coding Facial Expressions with Gabor Wavelets,*" Proceedings of the 2nd International Conference on Automatic Face and Gesture Recognition, Japan, April, 1998, pp. 200-205.

[9] HAYKIN, S., "*Neural Networks: A Comprehensive Foundation*," Prentice Hall, Upper Saddle River, NJ, 1999.

[10] KIRBY, M., SIROVICH, L., "*Application of the Karhunen-Loeve procedure for the characterization of human faces*," IEEE Transactions on pattern Analysis and Machine Intelligence, Vol. 12, n. 1, 1990, pp. 103-108.

[11] TURK, M., PENTLAND, A., "*Eigenfaces for Recognition*," Journal of Cognitive Neuroscience, Vol. 3, No. 1, 1991.

[12] PENTLAND, A., MOGHADDAN, B., STRAINER, T., "*View-based and modular eigenspaces for face recognition*," Proc. of IEEE Conference on Computer Vision and Pattern Recognition, June, 1994.

[13] PENTLAND, A., et al, "*Experiments with Eigenfaces*," International Joint Conference on Artificial Intelligence, Chamberry, France, August, 1993.

[14] MOGHADDAM, B.; PENTLAND, A., "*Probabilistic Visual Learning for Object Detection*," 5th Int. Conference on Computer Vision, June, 1995.

[15] RAO, R. P. N. , BALLARD, D. H., "*Natural basis functions and topographic memory for the face recognition,*" in International Joint Conference on Artificial Intelligence, Montreal, Canada, pp. 10-17, 1995.

[16] HECHT-NIELSEN, R., *"Neurocomputing,"* (Addison-Wesley, San Diego, 1989).

[17] TRELEAVEN, P., PACHECO M. and VELLASCO, M., *"VLSI Architectures for Neural Networks,"* IEEE Micro, Vol 9, No. 6, pp. 8-27, (ISSN 0272-1732), December, 1989,

[18] BISHOP, C. M., Rev. Sci. Intrum. 65, 1803 (1994).

[19] HORNIK, K., STINCHCOMBE, M. and WHITE, H., *"Neural Networks"*, 2, 359 (1989).

[20] WASSERMAN, P. D. *"Advanced Methods in Neural Computing,"* (Van Nostrand Reinhold, New York, 1993).

[21] GIROSI, F. and POGGIO, T. *"Biological Cybernetics,"* 63, 169 (1990).

[22] EKMAN P.AND FRIESEN, W. V., *"The Facial Action Coding System,"* Consulting Psychologists Press, Inc. San Francisco, CA, 1978.

[23] LYONS, M.J., BUDYNEK, J. AND AKAMATSU S., *"Automatic Classification of Single Facial Images,"* Transactions on Pattern Analysis and Machine Intelligence, Vol. 21, N. 12, December, 1999, pp. 1357-1362.