ESANN'2004 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 28-30 April 2004, d-side publi., ISBN 2-930307-04-8, pp. 567-572

Face Recognition Using Recurrent High-Order Associative Memories

Iulian B. Ciocoiu

Technical University of Iasi, Romania
Faculty of Electronics and Telecommunications
iciocoiu@etc.tuiasi.ro

Abstract.

A novel face recognition approach is proposed, based on the use of compressed discriminative features and recurrent neural classifiers. Low-dimensional feature vectors are extracted through a combined effect of wavelet decomposition and subspace projections. The classifier is implemented as a special gradient-type recurrent analog neural network acting as an associative memory. The system exhibits stable equilibrium points in predefined positions given by the feature vectors of the training set. Experimental results for the Olivetti database are reported, indicating improved performances over standard PCA and LDA-based face recognition approaches.

1. Introduction

Face recognition has represented for more than one decade one of the most active research areas in pattern recognition. A plethora of approaches have been proposed and evaluation standards have been defined, but current state-of-the-art solutions still need to be improved in order to cope with the recognition rates and robustness requirements of commercial products. Most of the approaches may be classified into two categories [2]:
a) geometric feature-based techniques, relying on the identification of specific components of a face such as eyes, nose, mouth, and distances among them
b) holistic template-based techniques, usually based on projecting the original (high-dimensional) images onto lower dimensional subspaces spanned by specific basis vectors. Eigenfaces [13] represent a *de facto* standard for the second approach and, although superior solutions exist, still defines a performance reference against which any new method is compared.
Face recognition systems usually include three modules, *i.e.* the preprocessing stage, feature extraction, and classification. Although the novelty aspect of the present paper is mainly related to the classifier, we present key elements of the other components in the following:

- similar to other approaches [3, 14], we perform a multiresolution decomposition of the original images based on the Discrete Wavelet Transform (DWT) and keep only the low-frequency components (Figure 1). Besides dimensionality reduction this procedure is also known to offer face expression invariance.

ESANN'2004 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 28-30 April 2004, d-side publi., ISBN 2-930307-04-8, pp. 567-572

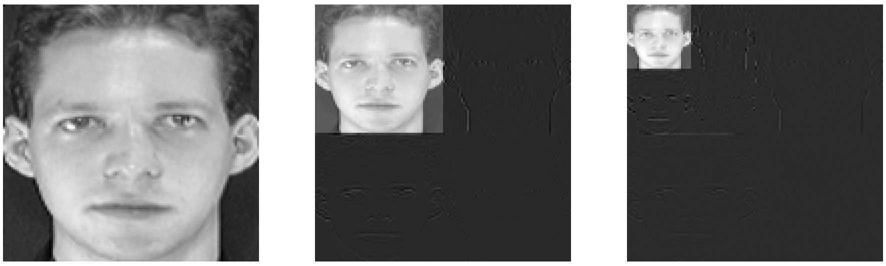

Figure 1: a) original image; b) DWT decomposition at level 1 (upper left corner image represents the level-1 approximation coefficients); c) DWT decomposition at level 2

- we apply 2 standard linear subspace projections of the low-resolution face images, performing Principal Components Analysis (PCA) and Linear Discriminant Analysis (LDA) [1, 12]. PCA is basically a compression procedure based on a linear projection technique on a subspace spanned by the principal eigenvectors (those corresponding to the largest eigenvalues) of the input covariance matrix. When applied to face processing, those basis vectors are called *eigenfaces* and define the directions along which the variance of the original images is maximized). PCA approximates the original images based on the *most expressive* features, which may not be optimal in terms of discriminating power. *Most discriminant* features would be better for classification purposes, and one of the classical choices is related to Fisher's LDA approach that identifies directions in space along which separation of the projections is maximized. *Fisherfaces* are obtained when LDA is applied to the original face images, and *discriminant waveletfaces* if low-resolution DWT filtered images are processed [3]. According to LDA theory, the procedure yields (C-1) projection directions (C is the number of distinct classes) and is sensitive to small training databases made-up of high-dimensional vectors. This is why usually PCA is initially performed on the available images and then LDA acts on the compressed feature vectors. In this respect, performing DWT prior to LDA has the additional benefit of already reducing the dimensionality of the vectors to be processed. While LDA is not always superior to PCA in terms of recognition accuracy, the PCA+LDA approach has been successfully applied in face recognition applications.

2. The Neural Classifier

The neural classifier is implemented as a special recurrent high-order associative memory. Associative memories represent one the most interesting applications of artificial neural networks and many solutions have been reported in the literature. Basically, a set of patterns is to be stored by using a training database and a proper learning procedure. In the testing phase, the system should output correct results even if noisy, incomplete or distorted data is applied as input. Both feedforward and recurrent approaches have been used for designing associative memories. When recurrent networks are used for implementation, desired memories are usually stored

ESANN'2004 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 28-30 April 2004, d-side publi., ISBN 2-930307-04-8, pp. 567-572

as stable states of dynamical systems. When certain conditions are met such systems are *globally stable* and the dynamics evolves from any initial state towards one particular stable equilibrium and no other complex behavior can occur [6]. Such systems should satisfy the following requirements [9]:

- no spurious memories (stable states which do not correspond to the desired ones) should exist
- the number of desired equilibria should be arbitrarily large and the dimension of the corresponding basins of attraction should be controllable
- the addition/elimination of an equilibrium should be performed without redesigning the whole system.

The main drawbacks of existing solutions are related to the presence of many spurious states and limited memory capacity. In order to alleviate these, we use a special gradient-type analog dynamic system defined according to:

$$\frac{dx_i}{dt} = -\frac{\partial V(\mathbf{X})}{\partial x_i}, \quad i = 1 \ldots N \tag{1}$$

where $\mathbf{X}=\{x_i\}$ defines the state-vector, N is the order of the system, and V($\mathbf{X}$) is the associated Lyapunov function. A well-known result states that all isolated minima of V($\mathbf{X}$) are asymptotically stable states of system (3) [6]. The key feature of our approach lies in the special way of constructing the function V($\mathbf{X}$) as a sum of individual functions exhibiting good *space localization* properties, having deep minima in predefined locations and been practically constant in rest [5]:

$$V(\mathbf{X}) = \sum_{m=1}^{M} w_m g_m(\mathbf{X}) \tag{2}$$

where M is the number of memories to be stored, w_m are scalar weights, and $g_m(\mathbf{X})$ is:

$$g_m(\mathbf{X}) = 1 - e^{-\frac{d_p{}^p(\mathbf{X},\mathbf{X}_m)}{2\sigma_m{}^2}} \tag{3}$$

$d_p(\mathbf{X},\mathbf{X}_m)$ is the distance induced by the L_p measure defined on the N-dimensional state vector space. In Figure 2 we present an example of the function $V(\mathbf{X})$ for a system with $N = 2$ and $M = 4$ stable equilibrium points, namely, (–1, –1); (–1, 1); (1, –1); (1, 1). We used a common value for $\sigma_m = 0.25$, and the weights vector was $\{w_m\} = \{1, 1, 1, 1\}$.

ESANN'2004 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 28-30 April 2004, d-side publi., ISBN 2-930307-04-8, pp. 567-572

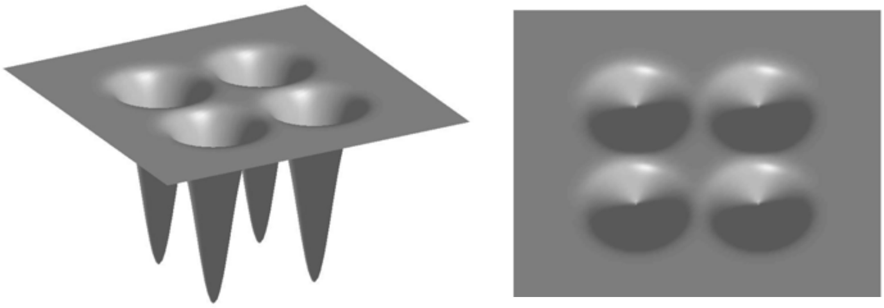

Figure 2: a) Example of a Lyapunov function as in equation (1) (M=4, N=2);
b) upside view

The proposed design procedure has a number of important advantages, including:
- a clear correspondence between the set of memories to be stored and the equations governing the system dynamics
- a transparent interpretation of the effect of the parameters (centers, weights, width) on the time and state-space evolution
- guaranteed convergence based on Lyapunov stability theory
- implementation advantages in terms of limited number of interconnections

Face recognition based on such a recurrent associative memory works as follows: the feature vectors (PCA or PCA+LDA projection coefficients) are extracted from the training set and memorized into the network as stable equilibrium points. When a test face image is presented to the system the corresponding feature vector is first computed and then applied as *initial condition* to the (neural) dynamical system, which will eventually settle down to one of the stable equilibrium points, hopefully to one obtained from a training image of the correct person. According to the positions of the training images, complex basins of attractions are developed around the equilibrium points, which may include besides the available test images many others, *e.g.* ones corresponding to occluded, distorted or noisy versions of the training set. In this respect, it is worth mentioning that proper choice of the individual σ_m parameters offers an additional handle for shaping those basins of attraction. Although a learning algorithm could provide optimized performances, we set their values according to a heuristic rule, namely as a fraction of the distance of vector $\mathbf{X}_m$ from its closest neighbour (neighbours originating from the training images of the same person are excluded). Using distinct σ_m values leads to basins of attraction having unequal widths, which may explain why test vectors closer (in terms of Euclidean distance) to a training image of an erroneous person may still fall into the basin of the correct one).

Moreover, the proposed neural classifiers exhibits implicit modularity, in that storing additional memories doesn't influence the positions of the previously stored ones and, more importantly, the dynamics of the system and thus the final solution is influenced only by a small fraction of the existing stable equilibria (ideally, only by a single stable point whose basin of attraction the test vector falls into).

ESANN'2004 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 28-30 April 2004, d-side publi., ISBN 2-930307-04-8, pp. 567-572

3. Experimental Results

Intensive computer simulations have been performed in order to assess the performances of each preprocessing technique compared to the standard eigenface procedure. The experiments were conducted on the Olivetti database, which comprises 10 distinct images of 40 persons, and includes variations in pose, light conditions, scaling, and expression. Each image has 112x92 pixels. In order to cope with the requirement of the DWT transform that the dimensions are a power of 2, we first interpolate the original images to yield 128x128 pixels resolution, and then apply 2-level DWT decomposition using Daubechies-4 mother wavelet. Only the low-pass approximation coefficients are selected, as 32x32 images. Next, we applied PCA and LDA using a training database of 5 images per person, randomly selected from the available 10, and the rest for the testing phase. The training and test datasets were not overlapping. After performing DWT, several distinct feature extraction procedures were used: a) standard PCA (eigenfaces) using Euclidean measure; b) (PCA+LDA) using Euclidean measure; c) PCA+neural associative memory; d) PCA+LDA+neural associative memory.

The degree of dimensionality reduction obtained by PCA is chosen according to the following criteria: a) small reconstruction error (energy loss is typically less than 10%); b) theory indicates that maximally allowed dimension of the vectors subject to LDA is (No_train-C), where No_train is the number of training images and C is the number of classes. In our case we have No_train=200 images and C=40 classes. We performed separate tests using PCA feature vectors with 150, 100, and 50 dimensions. Recognition performances are given in Table 1, including average error rates after 10 trials. Results obtained by the (discriminant waveletfaces+neural associative memory) combination match some of the top performant solutions reported in the literature.

Table 1: Correct recognition rates (average values over 10 trials, 5 training images per person; L2 – Euclidean distance; NFL – Nearest Feature Line; NFS – Nearest Feature Space)

System	Error rate (%)
Eigenfaces [10]	10
Pseudo-2D HMM [10]	5
Convolutional Neural Network [11]	3.8
Linear SVM [7]	3
Kernel PCA [13]	2.5
Waveletface + L2 [3]	7.5
Discriminant Waveletface + L2 [3]	5.5
Discriminant Waveletface + NFL [3]	5
Discriminant Waveletface + NFS [3]	3.9
ARENA [11]	2.9
Waveletface + PCA + Recurrent Associative Memory	5.4
Discriminant Waveletface + Recurrent Associative Memory	3.1

ESANN'2004 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 28-30 April 2004, d-side publi., ISBN 2-930307-04-8, pp. 567-572

4. Conclusions

The proposed approach yields accurate results, comparable to the best results reported on the Olivetti database. The key aspect is related to the structure of the neural classifier, which is attractive in terms of modularity and memory capacity. Due to the analog nature, high processing speed could be expected when the proposed solution is hardware implemented. Further work will be dedicated to the possibility of inferring invariance to standard transformations (translation, rotation, scale variation) for the proposed system.

References

[1] P.N. Belhumeur, J.P. Hespana, and D.J. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Trans. PAMI*. 19(7):711-720, 1997.
[2] R. Brunelli, and T. Poggio. Face Recognition: Features versus Templates. *IEEE Trans. PAMI*. 15(10):1042-1052, 1993.
[3] J.-T. Chien, and C.-C. Wu. Discriminant Waveletfaces and Nearest Feature Classifiers for Face Recognition. *IEEE Trans. PAMI*. 24(12):1644-1648, 2002.
[4] G. D. Guo, S. Z. Li, and K. L. Chan. Face recognition by support vector machines. In *Proc. ICAFGR*, pages 196–201, 2000.
[5] J.Y. Han, M.R. Sayeh, and J. Zhang. Convergence and limit points of neural networks and its application to pattern recognition. *IEEE Trans. Syst., Man, Cybern*. 15(5):1217-1222, 1989.
[6] M.W. Hirsch, and S. Smale. *Differential Equations, Dynamical Systems, and Linear Algebra*. NY: Academic, 1974.
[7] K.I. Kim, K. Jung, and H.J. Kim. Face Recognition Using Kernel Principal Component Analysis. *IEEE Signal Processing Letters*. 9(2):40-42, 2002.
[8] S. Lawrence, C. L. Giles, A.C. Tsoi, and A.D. Back. Face recognition: A convolutional neural network approach. *IEEE Trans. Neural Networks*. 8:98-113, 1997.
[9] A.N. Michel, and J.A. Farrell. Associative memories via artificial neural networks. *IEEE Control Systems Magazine*. 10:6-17, 1990.
[10] F. S. Samaria. Face recognition using hidden Markov models. *Ph.D Dissertation*. Univ. Cambridge, Cambridge, U.K., 1994.
[11] T. Sim, R. Sukthankar, M. Mullin, and S. Baluja. High-Performance Memory-Based Face Recognition for Visitor Identification. Technical report, JPRC-TR-1999-001-1, 1999.
[12] D.L. Swets, and J. Weng. Using discriminant eigenfeatures for image retrieval. *IEEE Trans. PAMI*. 18(8):831-836, 1996.
[13] M. Turk, and A.P. Pentland. Eigenfaces for recognition. *J. of Cognitive Neuroscience*. 3(1):71-86, 1991.
[14] B.-L. Zhang, H. Zhang, and S.S. Ge. Face Recognition by Applying Wavelet Subband Representation and Kernel Associative Memory. *IEEE Trans. Neural Networks.* 14(6), 2003.