PROC. OF THE INT.CONF. ON SIGNAL PROCESSING, COMMUNICATIONS AND NETWORKING, IEEE ISBN: 1-4244-0997-7/07/$25.00

A Back Propagation Based Face Recognition Model, Using 2D Symmetric Gabor Features

B. Vinay Kumar, Subex Azure Limited, Bangalore, India
B. S. Shreyas, Department of Telecommunication, B.M.S College of Engineering, Bangalore, India
C.N.S. Ganesh Murthy, Daimler Chrysler, Research and Technology, Bangalore, India

***Abstract*– We present a system for recognizing human faces from a database consisting of multiple images per test subject, which spans the normal variations in a human face. The faces are represented based on a Gabor wavelet transform. The features are extracted as a vector of values using a carefully chosen symmetrical Gabor wavelet matrix. This feature extraction is biologically motivated and models systems based on human vision. The extracted features are fed into an Artificial Neural Network, in dual phases. The training and testing phases of the neural network work on the features extracted by the same method. Excellent pattern-recognition-specific neural network like a multi layer perceptron with back propagation provides the necessary classification once the feature extraction is complete.**

I. Introduction

Face recognition has been an active research topic over the past few years due to its scientific challenges and potential applications. In general, there are two main approaches to 2-D face recognition, template-based approach and geometric feature-based approach [1]. In template-based approach statistical methods are used to represent face images as a whole. Template-based approach is characterized by a family of subspace methods originated by "eigenface" [2]. Peter et al. switched from "eigenface" to "fisherface" [3]. Moghaddam et al. proposed to estimate density in high-dimensional spaces using eigenspace decomposition [4] and then derived a probabilistic similarity measure based on Bayesian analysis of image differences [5]. Wang et al. further developed a unified analysis method that uses three subspace dimensions, i.e. intrinsic difference, transform difference and noise, and achieved better recognition performance than the standard subspace methods [6]. On the other hand, feature based methods describe a face through abstraction: An object is represented as a selected collection of abstract features. Simple abstract features are edges, lines, line segments and points. More complex features may be composed from several simple ones. This type of representation leads to an abstraction from the image pixel values. Representative works include Hidden Markov Model (HMM) proposed by Samaria [7], elastic bunch graph matching algorithm proposed by Wiskott et al. [8], and Local Feature Analysis (LFA) proposed by Penev et al. [9]. In most feature-based approaches, the selection of features as well as their description is given a priori through heuristics.

However, although many of the algorithms have demonstrated excellent recognition results, there are still many open problems. A general framework has been proposed for extraction of Gabor features for the purpose of identification or verification of faces [10]. We now focus on the task of face identification. For this purpose we employ the proposed feature extraction method that uses 2D Gabor wavelets and a neural network classifier. Wavelets are a transformation of an image into a frequency representation called a wave. The wave represents how fast things change in certain repetitive patterns. They have advantages over traditional Fourier methods in analyzing physical situations when the signal contains discontinuities and sharp spikes. This is a feature-based approach. The general framework for representing facial images used here is based on topographically ordered, spatially localized feature points to represent pattern in the image. At each feature point a bank of a multi-resolution, multi orientation Gabor wavelet functions are employed. The Gabor wavelet transform of a two-dimensional visual field generates a four-dimensional field: two of the dimensions are spatial, the other two represent spatial frequency and orientation. There are several methods to represent images using Gabor wavelets. At one extreme, images can be represented by placing the wavelets at each pixel. Usually at each point, 60 Gabor wavelets (6 different scales and 10 orientations) are used. If the size of image is 100 x 100, the dimension of the feature vector will be 600,000. Such a high dimensional vector will lead to expensive computation and storage cost. Another approach would be to place a face graph where the nodes of the graph lie on facial features. This approach requires a fine localization of facial feature points. The major disadvantage would be that only the areas which can be reliably located are used for recognition purposes. In between these two approaches, one can use a rectangular grid that is placed over the face region, which is used in this paper.

The Gabor wavelets approach appears to be quite perspective and has several advantages such as invariance to some degree with respect to homogenous illumination changes, small changes in head poise and robustness against facial hair, glasses, image noise. Further advantages include: Saves neighborhood relationships between pixels, easy to update, fast recognition and low computational cost. Experimental results show that the proposed method performs better than traditional approaches in terms of both efficiency and accuracy.

Once the features are extracted from the image, they are fed to an artificial neural network (ANN). ANN is an interconnected group of artificial neurons that uses a mathematical or computational model for information processing based on a connectionist approach to computation.

A comprehensive study of Neural networks [11] and its specific applications in Pattern Recognition [12] help in understanding the intrinsic details. After feature extraction, we train a Multi Layer Perceptron to act as a classifier [13].

The remainder of this paper is organized as follows: In section 2, Design and Implementation of our Face recognition algorithm is presented as the major contribution of this paper. The experimental results are discussed in section 3. Limitations of our algorithm are outlined in section 4 and concluding remarks are made in section 5.

II. Design And Implementation

The entire system has been divided into three modules. These can be represented by three sequential sections which as a whole forms our face recognition algorithm.

- Image acquisition and analysis section
- Feature extraction section
- Neural network classifier

The entire system is shown in Fig. 1. The input consists of the face of the person seeking authentication. The facial image is captured by a digital camera and fed via USB to the central processing unit. The major components attached to the Central Processing Unit are the Display unit and the Memory unit. Let us consider each of the above sections and understand the design details.

A. Image Acquisition and Analysis Section

The input to this section is the facial image of the person seeking recognition. The image is preprocessed if required. This preprocessing involves the use of digital image enhancement techniques. Here we are assuming only variations with respect to the quality of the image and not with respect to the head poise or expressions.

After the preprocessing step has been completed we need to convert the image to the portable gray map (PGM) format. This makes the image a gray scale image. Once the image has been fed to the system, we have to place the wavelets in the "Region of Importance." As our images are faces the region of importance is "The inner face region." A rectangular grid of feature points is placed over "The inner face region." This wavelet net (wavenet) gives us a set of feature points. The number of feature points must be chosen optimally. Increasing the number of feature points gives better representation but at the cost of computational speed. Finding an appropriate trade off between these two factors is of prime importance. The Gabor wavelet expression and feature extraction process is explained in detail in the next section.

B. Feature Extraction Section

This section is the core of our face recognition algorithm. Here the face image is processed with Gabor wavelets and in this process the features are extracted. A set of frequencies and orientations are chosen to represent the image at each feature points of the wavenet.

Now we define "Lower bound" and "Upper bound" frequencies given by

$$f_{LB} = \frac{1}{x_1\sqrt{2}} \quad \text{and} \quad f_{UB} = \frac{1}{x_2\sqrt{2}} \tag{1}$$

The values of x_1 and x_2 are chosen such that $x_1 > x_2$. A set of frequencies to be used at each feature point is obtained by starting at f_{LB} and multiplying by 2 until f_{UB} is reached. The number of frequencies is given by P. For each frequency, a set of orientations is chosen ranging from to $-\pi$ to π. The step size between any two θ is n, where n is chosen appropriately. The number of orientations is given by Q. The feature points are denoted by (c_x, c_y) as per the Cartesian co-ordinates. The number of feature points is given by R.

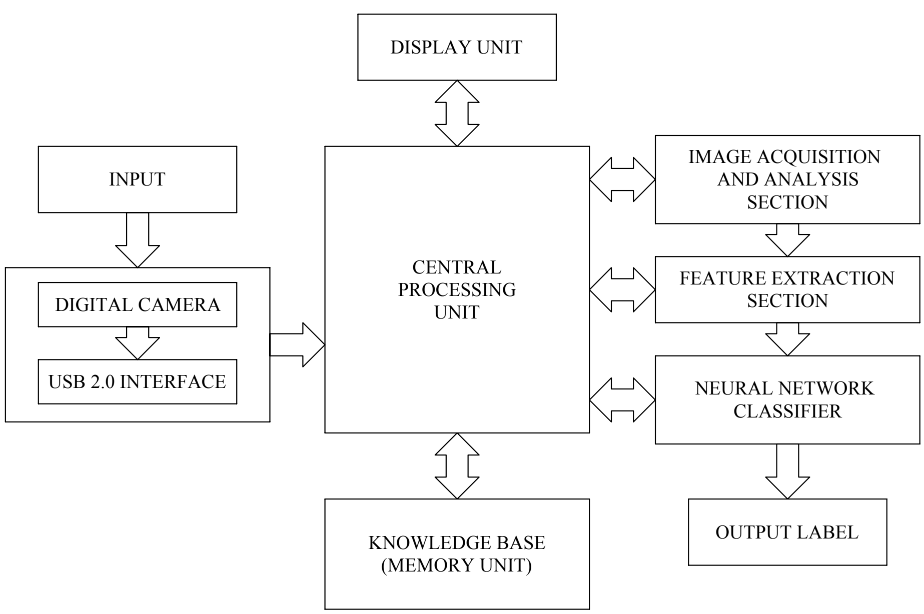

Fig. 1. Block diagram depicting the various modules of the face recognition system.

Now a set of frequencies and orientations are obtained at each feature point (c_x, c_y). The number of wavelets is $N = P*Q*R$. $I(x, y)$ (or just I) represent the input image. 'x' and 'y' represent the coordinates of each pixel in the Cartesian system. Hence 'x' and 'y' ranges from 0 to 'h' and 'w' respectively where 'h' is the height of the image and 'w' is the width of the image.

Now we define a family of N Gabor wavelet functions $\psi = \{\psi_{1,1,1}, \Lambda, \psi_{P,Q,R}\}$ of the form

$$\psi_{i,j,k}(x,y) = \frac{f_i^2}{2\pi} \exp\left\{-0.5 f_i^2 \left[\left(x - c_{x_k}\right)^2 + \left(y - c_{y_k}\right)^2\right]\right\} * \sin\left\{2\pi f_i \left[\left(x - c_{x_k}\right)\cos\theta_j + \left(y - c_{y_k}\right)\sin\theta_j\right]\right\} \quad (2)$$

where f_i denotes the frequency, θ_j denotes orientation and c_{x_k}, c_{y_k} denotes wavelet position.

Now the wavelet function $\psi_{i,j,k}(x,y)$ is normalized by scaling it by the reciprocal of its length $\|\psi_{i,j,k}(x,y)\|$. Thus we obtain the unit vector $\hat{\psi}_{i,j,k}(x, y)$ as shown below

$$\hat{\psi}_{i,j,k} = \frac{\psi_{i,j,k}}{\|\psi_{i,j,k}\|}. \quad (3)$$

The first weight w_1 associated with wavelet Ψ_1 is given by the dot product of I and $(\hat{\psi}_{i,j,k})_1$ as shown below

$$w_1 = I \cdot (\hat{\psi}_{i,j,k})_1. \quad (4)$$

The subsequent weights w_u ($2 \le u \le N$) associated with wavelets Ψ_u are determined by calculating the dot products of I_{diff_u} and $(\hat{\psi}_{i,j,k})_u$ as shown below

$$w_u = I_{diff_u} \cdot (\hat{\psi}_{i,j,k})_u \quad (5)$$

where I_{diff_u} represents the intermediate difference image:

$$I_{diff_u} = I_{diff_{(u-1)}} - \hat{I}_{u-1} \text{ with } I_{diff_1} = I(x,y), \text{ for } (2 \le u \le N) \quad (6)$$

where $\hat{I}_u$ is the intermediate reconstructed image for each wavelet given by

$$\hat{I}_u = w_u \cdot \left(\hat{\psi}_{i,j,k}\right)_u, \text{ for } (2 \le u \le N). \quad (7)$$

The final reconstructed image is given by

$$\hat{I} = \sum_{u=0}^{N-1} \hat{I}_u. \quad (8)$$

The original face image I, wavelet net and the reconstructed image $\hat{I}$ are all shown in Fig. 2.

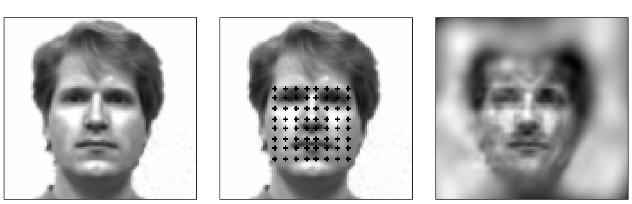

Fig. 2. The left image shows the original face image I, the middle image shows the position of wavelet net and the right image shows the final reconstructed image $\hat{I}$ for $N = 100$.

In the above reconstruction process, the set of weights (projections) are obtained. From this set of weights the 'n' highest weights are selected, where n is chosen suitably. These n weights along with the corresponding set of n frequencies and orientations are the features. For identification purpose we use a neural network classifier. The feature vector supplied to the classifier comprises of n weights along with the corresponding set of n frequencies and n orientations.

C. *Neural Network Classifier*

An artificial neural network (ANN), also called a simulated neural network (SNN) or commonly just neural network (NN) is an interconnected group of artificial neurons that uses a mathematical or computational model for information processing based on a connectionist approach to computation. In most cases an ANN is an adaptive system that changes its structure based on external or internal information that flows through the network. A specific type of neural network "The Multi-layer Perceptron (MLP)" will be dealt with in some detail since we are using a MLP in our face recognition system

This class of networks consists of multiple layers of computational units, usually interconnected in a feed-forward way. Each neuron in one layer has directed connections to the neurons of the subsequent layer. In many applications the units of these networks apply a sigmoid function as an activation function.

The *universal approximation theorem* for neural networks states that every continuous function that maps intervals of real numbers to some output interval of real numbers can be approximated arbitrarily closely by a multi-layer perceptron with just one hidden layer. This result holds only for restricted classes of activation functions, e.g. for the sigmoidal functions.

Multi-layer networks use a variety of learning techniques, the most popular being *back-propagation*. Here the output values are compared with the correct answer to compute the value of some predefined error-function. By various techniques the error is then fed back through the network. Using this information, the algorithm adjusts the weights of each connection in order to reduce the value of the error function by some small amount. After repeating this process for a sufficiently large number of training cycles the network will usually converge to some state where the error of the calculations is small. In this case one says that the network has *learned* a certain target function. To adjust weights properly one applies a general method for non-linear optimization task that is called gradient descent. For this, the derivative of the error function with respect to the network weights is calculated and the weights are then changed such that the error decreases (thus going downhill on the surface of the error function). For this reason back-propagation can only be applied on networks with differentiable activation functions.

In general the problem of teaching a network that performs well, even on samples that were not used as training samples, is a quite subtle issue that requires additional techniques. This is especially important for cases where only very limited number of training samples are available. The danger is that the network overfits the training data and fails to capture the true statistical process generating the data. Computational learning theory is concerned with training classifiers on a limited amount of data. In the context of neural networks a simple heuristic, called early stopping, often ensures that the network will generalize well to examples not in the training set.

Other typical problems of the back-propagation algorithm are the speed of convergence and the possibility to end up in a local minimum of the error function. Today there are practical solutions that make back-propagation in multi-layer perceptrons the solution of choice for many machine learning tasks.

In this section the feature vector obtained in feature extraction section is processed. Neural network classifier is used for this purpose, since a classifier always attempts to find the closest match. In other words, it never rejects any image but simply outputs the closest match or a set of top '*n*' matches, where *n* is chosen suitably.

The type of neural network used is Multi-layer perceptron. It is a feed-forward network that uses back-propagation learning algorithm. The neural network can either be configured in the training mode or recognition mode. Initially the neural network is set to training mode and all the images in the gallery are trained. An optimal output encoding is used to differentiate between individuals. The neural network is trained with respect to all the individuals. Factors such as number of hidden neurons, momentum, learning rate, number of iterations and output encoding are varied for optimal training. While in the recognition mode, the neural network outputs the closest match or a set of top '*n*' matches.

III. EXPERIMENTAL RESULTS

- Wavelet net needs to be placed in the region of importance.
- Placement of wavelets should be consistent with an admissible variation of just few pixels for efficient training and recognition.
- 8 X 8 wavelet matrix gives optimal tradeoff between good representation and computational time required.
- Top 50 projections and corresponding angles and frequencies which form the feature vector is found to be optimal.
- The neural network parameters during training mode were set as follows:
    - The number of hidden neurons was set to 20% of the number of input neurons
    - The learning rate and momentum were both set to 0.5
    - The threshold error level was set to 0.0001
    - Number of iterations was set to 250,001 System is insensitive to homogenous illumination changes to a certain extent.
- System is extremely robust against facial hair, glasses and small changes in head poise if the number of subjects is less than 5.
- System is not robust against extreme changes in facial expression.
- With 5 subjects overall efficiency of close to 100% was achieved.
- System is not very robust against extreme head pose variations and expression if the number of subjects is more than 5.
- With 10 subjects overall efficiency of close to 90% was achieved.
- Overall efficiency of 90% was achieved with tests on the selected images of Yale and Olivetti face databases.
- Efficiency of the neural network depends on the output encoding. One-Hot encoding was found suitable for less than 5 subjects and binary encoding for 5 or more subjects.

Test Images (samples): The images in Fig. 3 show some of the images used for training and the images used for testing. The test images constitute of different head poises, glasses, facial hair and illumination (both gradient changes and sharp step changes).

Trained Images **Test Images**

Fig. 3. The images on the left are the trained images and images on the right are the corresponding tested images for 3 individuals.

Successful recognition was obtained for all of the images below. The total number of subjects was 10. Only 3 have been shown in below.

IV. Limitations

The limitations of our face recognition algorithm are listed below:

- Not so robust against extreme variations in expression.
- Cannot be used for faces with lateral head rotation.
- The computing power that is in general available makes face recognition suitable for offline applications as online applications require higher computing power.
- Face recognition alone as of now, cannot be successfully used for authentication but can be used for identification and classification.
- Identical twins cannot be distinguished using face recognition.

V. Conclusion

In this project we developed Face recognition system using Gabor wavelets. As said earlier, the use of Gabor wavelets is biologically motivated. This approach appears to be quite perspective, insensitive to small changes in head poise and homogenous illumination changes, robust against facial hair, glasses and also generally very robust compared to other methods. However it was found to be sensitive to large facial expression variations. Also, it was found that placement of wavelets should be consistent for efficient recognition. Accurate training of the neural network was found to be overriding factor determining the efficiency of the algorithm. Ad-hoc training depending on the type of images in the gallery was found to give better performance than following a generalized procedure.

Face recognition systems are no longer limited to classification, identity verification and surveillance tasks. Growing numbers of applications are starting to use face recognition as the initial step towards interpreting human actions, intention, and behavior, as a central part of Next-Generation Smart Environments.

Acknowledgment

The images used are derived from the Yale face database and the Olivetti face database.

References

[1] W. Zhao, R. Chellappa, A. Rosenfeld and J. Phillips, "Face recognition: A literature survey," *ACM Computing Surveys*, vol. 12, pp. 399-458, 2003.

[2] M. Turk and A. Pentland, "Face recognition using Eigenfaces," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 586-591, 1991.

[3] N. B. Peter, P. H. Joao and J. K. David, "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711-720, 1997.

[4] B. Moghaddam and A. Pentland, "Probabilistic visual learning for object representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no.7, pp. 696-710, 1997.

[5] B. Moghaddam, B. Jebara and A. Pentland, "Bayesian face recognition," *Pattern Recognition*, vol. 33, pp. 1771-1782, 2000.

[6] X. Wang and X. Tang, "Unified subspace analysis for face recognition," *Proceedings of IEEE International Conference on Computer Vision*, vol. 2, June 2003.

[7] F. S. Samaria, "Face recognition using hidden Markov models," PhD thesis, University of Cambridge, UK, 1994.

[8] L. Wiskott, J. Fellous, N. Kruger and C. Malsburg, "Face recognition by elastic bunch graph matching," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 775-779, 1997.

[9] P. S. Penev and J. J. Atick, "Local feature analysis: A statistical theory for object representation," *Network: Computation in Neural Systems*, vol. 7, pp. 477-500, 1996.

[10] B. Vinay Kumar and B. S. Shreyas, "Face recognition using Gabor wavelets," *Proceedings of the 40th Asilomar Conference on Signals, Systems and Computers*, in press.

[11] S. Haykin, *Neural Networks: A Comprehensive Foundation*, 2nd ed., New York: Prentice Hall, 1999.

[12] Chris Bishop, *Neural Networks for Pattern Recognition*, Oxford, England: Oxford University Press, 1995.

[13] D. W. Ruck, S. K. Rogers, M. Kabrisky, M. E. Oxley and B. W. Suter, "The multilayer perceptron as an approximation to a Bayes optimaldiscriminant function," *IEEE Transactions on Neural Networks*, vol. 1, no. 4, pp. 296-298, Dec 1990.