PROCEEDINGS OF THE IEEE ASILOMAR CONFERENCE ON SIGNALS, SYSTEMS AND COMPUTERS, IEEE ISBN: 1-4244-0785-0/06/$20.00

Face Recognition Using Gabor Wavelets

Vinay Kumar. B
Department of Electronics and Communication Engineering
Global Academy of Technology
Bangalore, India
vinay@ieee.org

Shreyas. B. S
Department of Telecommunication Engineering
B.M.S College of Engineering
Bangalore, India
shreyas.bs@ieee.org

***Abstract*– Face recognition with variant pose, illumination and expression is a challenging problem. Robust face recognition requires the ability to recognize identity despite many variations in appearance the face can have. Today there exist many well known techniques for face recognition, each with its own inherent limitations. In this paper we present a novel approach to face recognition using Gabor wavelets. The Gabor Image representation simulates the function of the human visual system, a design feature which may be important in the field of robotics and computer vision. The Gabor wavelets approach appears to be quite perspective and has several advantages such as invariance to homogenous illumination changes, small changes in head poise and robustness against facial hair, glasses. Experimental results show that the proposed method performs better than traditional approaches in terms of both efficiency and accuracy. It is worth mentioning that Gabor wavelets technique has recently been used not only for face recognition, but also for face tracking and face position estimation.**

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 method describes a face through abstraction: An object is represented as a collection of selected 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 published algorithms have demonstrated excellent recognition results, there are still many open problems. In this paper we introduce 2-D Gabor wavelet-based method face recognition system. Wavelets are mathematical functions that cut up data into different frequency components, and then study each component with a resolution matched to its scale. 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. 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 scales and 10 orientations) are used. If the size of image is 100×100, 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, as suggested in this paper.

The use of 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.

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
- Identification/Verification section

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. The enhancement is 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. 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 at each feature point of the wavenet are chosen to represent the image.

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.

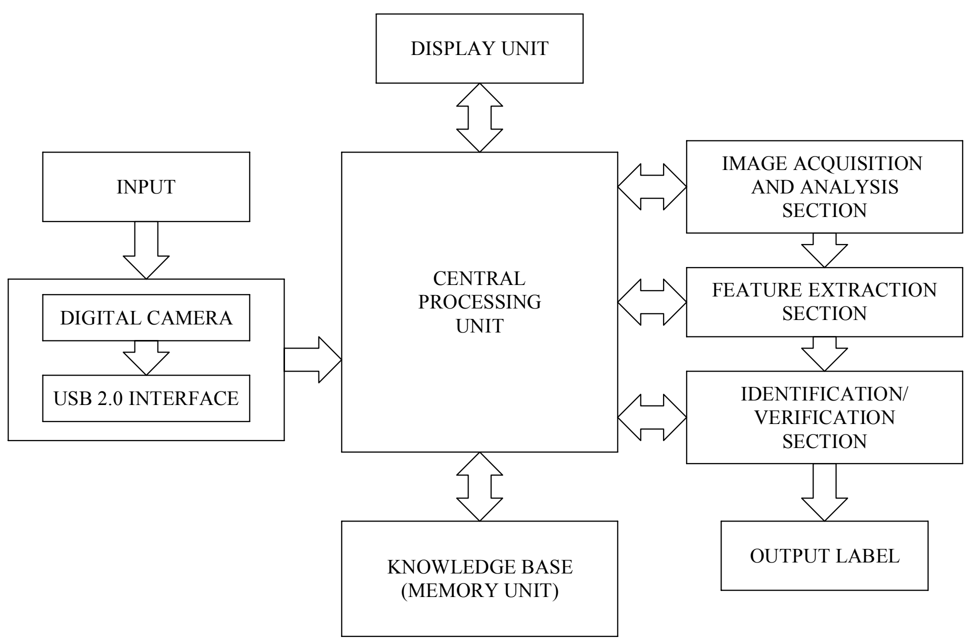

Fig. 1. Block diagram depicting the various modules of the face recognition system.

Now a set of frequencies and orientations are obtained at each feature point (c_x, c_y). The number of wavelets $N = P*Q*R$. $I(x, y)$ (or just I) represents 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 0 to '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 $\psi = \{\psi_{1,1,1}, \ldots, \psi_{P,Q,R}\}$ 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.

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.

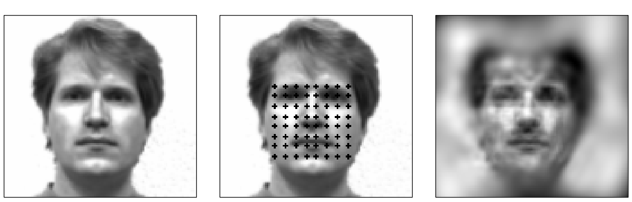

Fig. 2. Image on the left shows the original face image I, image in the middle shows the position of wavelet net and image on the right shows the final reconstructed image $\hat{I}$ for $N = 100$.

In the above reconstruction process, a 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. Now the elements of the feature vector to be used in the next section are determined by the application. 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. For the purpose of verification where distance measure approach is used, feature vector comprises only of 'n' highest weights. Other methods can also be used to choose the feature vectors such as considering the frequency and orientation at each feature point that gives maximum projection, etc.

C. Identification/Verification Section

In this section the feature vector obtained in feature extraction section is processed depending on the task in hand. There are two kinds of face recognition tasks that we shall consider:

1) Identification

Identification is considered as a closed-universe recognition task. The closed-universe recognition task is the task of identifying the individual in the probe image, assuming that the probe is in the gallery. (gallery is a set of image sets. Each image set is associated with a specific individual.) 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 adjusted for optimal training. While in the recognition mode, the neural network outputs the closest match or a set of top 'n' matches.

2) Verification

Verification is an open-universe recognition task. In an open-universe, some probe persons may not be included in the gallery. For verification, a probe image and an identity are given as inputs to the system. The identity is assumed to refer to a gallery face. The system now has to verify whether the identified individual is same as the one in the probe image. The degree of similarity between the gallery image and the probe image is used to decide on recognition or rejection.

The degree of similarity is determined by using simple Euclidean distance measure. A threshold for the similarity measure is usually given in advance by hand, or can be learned during the training phase. While setting the threshold two factors should be considered: False-reject rate and False-alarm rate. A False-reject is an instance where the system rejects a valid identity; a False-alarm is an instance where the system incorrectly accepts an identity. A higher threshold makes the system more secure but at the expense of causing inconvenience to legitimate users. Careful selection of the similarity threshold is important in keeping false-reject rate and false-alarm rate low. Other methods like Mahalanobis distance measure can also be used to determine the degree of similarity.

III. Experimental Results

We selected 8 individuals drawn from the Yale and Olivetti Face Databases. The images were cropped as required and resized to 92 x 112. Each individual has 4 frontal face images from which 1 image was selected as training image and the remaining 3 images were used for testing. Thus our gallery consisted of 8 face images belonging to 8 different individuals. An example of training image is shown in Fig. 3 and the corresponding images used for testing are shown in Fig. 4.

We tested our algorithm for the task of identification, using neural network classifier. The placement of wavelets at feature points was done manually with the help of eye corners and lip corners as references. 64 feature points were distributed evenly over the inner face region in the form of an 8 x 8 wavenet matrix. This was found to be an optimal tradeoff between good representation and computational time required. At each feature point 6 frequencies were chosen and at each frequency 10 orientations were chosen. Image reconstruction was found to deteriorate if the number of frequencies or orientations was either decreased or increased.

Fig. 3. An example of training image. Normal faces with uniform illumination and no facial expressions were chosen for training.

Fig. 4. Image on the left shows normal facial expression, i.e. no particular expression, image in the middle shows illumination from left and image on the right shows illumination from right.

Projections were calculated for all the 64*6*10 wavelets using (4) and (5), out of which highest 50 projections (in terms of magnitude) were selected. These 50 projections along with the corresponding 50 frequencies and 50 orientations formed our feature vector.

The various parameters of the neural network during training mode were set as follows: The number of hidden neurons was set to 20% of the number of input neurons i.e. 30 hidden 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. The error level was checked for every 10,000 iterations. The training was stopped if either the threshold error was reached or the above number of iterations was carried out. One-hot encoding was used for encoding the output neurons.

The experiments were conducted using AMD 1.8 GHz, 64-bit (Athlon(tm) 64) processor with 768 MB of RAM and nVIDIA 128 MB (GeForce4 MX 4000) graphics processing unit. The computation time for testing was 40 seconds. It was found that placement of wavelets should be consistent with an admissible variation of just few pixels for efficient training and recognition. It was found that the system was insensitive to homogenous illumination changes to a certain extent. System was also found to be extremely robust against facial hair, glasses and small changes in head poise if the number of gallery images was less than 5. System was found to be not robust against extreme changes in facial expression. One-hot encoding was found in general, to give better results than binary encoding. But as the number of subjects increase, the number of bits required for one-hot encoding also increases, thus slowing down the training process. With 5 gallery images overall efficiency was close to 100% and with all 8 gallery images, 90% efficiency was achieved.

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 paper 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. It is worth mentioning that Gabor wavelets technique has recently been used not only for face recognition, but also for face tracking and face position estimation.

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, intentions, and behavior, as a central part of Next-Generation Smart Environments.

ACKNOWLEDGMENT

We thank Dr. C. N. S Ganesh Murthy; our guide who has shown immense patience and support, besides providing an incredible amount of guidance. We also extend our heartfelt thanks to all the people who have directly or indirectly helped us in the completion of our project.

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.