REAL TIME FACE RECOGNITION SYSTEM USING AUTOASSOCIATIVE NEURAL NETWORK MODELS

S. Palanivel, B.S. Venkatesh and B. Yegnanarayana

Speech and Vision Laboratory
Department of Computer Science and Engineering
Indian Institute of Technology Madras, Chennai-600 036, India.
Email:{spal, venkatesh, yegna}@cs.iitm.ernet.in

ABSTRACT

This paper proposes a novel method for video-based real time face recognition. The proposed method uses motion information to detect the face region, and the region is processed in YC_rC_b color space to determine the location of the eyes. The system extracts only the gray level features relative to the location of the eyes. Autoassociative Neural Network (AANN) model is used to capture the distribution of the extracted gray level features. Experimental results show that the proposed system gives an average recognition rate of 99% in real time for 25 subjects. The performance of the proposed method is invariant to size, tilt of the face and is also not sensitive to natural lighting conditions.

1. INTRODUCTION

Automatic recognition of human faces is one of the challenging problems in pattern recognition. A comprehensive survey of still and video-based face recognition techniques can be found in [1]. Various methods have been proposed in the literature such as eigenface [2], elastic graph matching [3], neural network [4],[5], line edge map [6] and support vector machines [7]. The performance of a face recognition technique should be:

1. Invariant to size and tilt of the face
2. Invariant to natural lighting conditions
3. Able to produce the result within a reasonable time
4. Invariant to facial expressions, pose and aging of the subject

The first three are considered to be the minimal requirements in order to implement the technique in real time. Most of the methods described in the literature are not able to satisfy at least one of the minimal requirement. The method proposed in this paper satisfies all the minimal requirements; It consists of three modules, namely face detection, feature extraction and face recognition. The face detection and feature extraction techniques are described in Section 2 and 3, respectively. Section 4 describes the AANN model for face recognition. The experimental results are discussed in Section 5.

2. FACE DETECTION

Detecting faces automatically from the intensity or color image is an essential task for many applications like face recognition, face tracking and video indexing [8]. Recent methods for face detection use neural networks [9], skin color segmentation [5],[10] and motion information [11] for tracking faces in the video. We used a simple method to detect the face region using only the motion information in order to implement the system in real time. In our method, the face region is determined from the upper head contour points which are extracted from the thresholded interframe difference image. The RGB image is converted to gray level image (I) and the interframe difference image (D) is obtained by

$$\begin{aligned} D(i,j,k) &= I(i,j,k) - I(i,j,k-1) \\ & \quad 0 \le i < w, 0 \le j < h, k \ge 1 \end{aligned} \tag{1}$$

where k is the frame number in the video, w and h are the width and height of the image respectively.

The thresholded difference image T is obtained by

$$T(i,j,k) = \begin{cases} 1, & if\ D(i,j,k) > \lambda \\ & or\ T(i,j-1,k) = 1 \\ 0, & otherwise \end{cases} \tag{2}$$

where λ is the threshold, which is the smallest integer such that $T(i,j,k) = 0$, $\forall i$ and $\forall j$, whenever there is no moving region in the camera view.

The thresholded difference image is scanned from top to bottom to find out the approximate top pixel (c_x, c_y) of the moving region. The RGB image in the region below the pixel (c_x, c_y) is converted to YC_rC_b space, and then checked for skin pixels as given in [12]. If there are no skin pixels, then the thresholded difference image is scanned until it finds out the pixel (c_x, c_y) or there is no moving region with skin pixels in the image frame. The upper portion of the head contour points are extracted by scanning the thresholded difference image from the pixel (c_x, c_y). The width of the face (w_1) is estimated from the head contour points, and this process is repeated for every two consecutive frames in order to track the face in the video. This method tracks only a single face, and it assumes that there is no motion in the region very nearer to the upper portion of the head in the video. Fig.1(a) shows the thresholded difference image as given by the Eq.(2). Fig.1(b) shows the extracted head contour points and the face region.

(a) Difference image (b) Contour and face region

Fig. 1. Face detection.

3. FEATURE EXTRACTION

Feature extraction is a key step in any pattern recognition task. This paper proposes a new method for extracting features from the face, which are relative to the location of the eyes, and hence the features are invariant to size and tilt of the face. A similar method is proposed in [10] for locating the eyes.

The face region is converted to YC_rC_b space, and the skin pixels are identified as given in [12]. The face region is thresholded to obtain the thresholded face image (R), given by

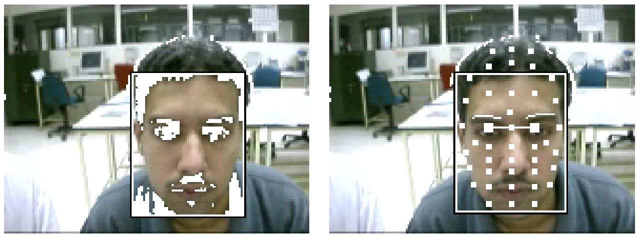
$$R(i,j,k) = \begin{cases} 255, & if \;\; Y(i,j,k) < \lambda_1 \text{ and } \\ & C_r(i,j,k) < \lambda_2 \text{ and } \\ & C_b(i,j,k) > \lambda_3 \\ I(i,j,k), & otherwise \end{cases} \tag{3}$$

where λ_1, λ_2 and λ_3 are the average Y, C_r and C_b values of the detected skin pixels in the face region, respectively. Morphological closing operation is applied to the thresholded face image and the centroid of all the blobs are estimated.

The eyebrow(E) pixels are estimated by

$$E(i,j,k) = \begin{cases} 1, & if \;\; Y(i,j,k) \geq \lambda_1 \text{ and } \\ & Y(i,j+1,k) < \lambda_1 \\ 0, & otherwise \end{cases} \tag{4}$$

The centroid of the blobs which are nearer to the center of the eyebrow pixels are taken as the location of the eyes. Fig.2(a) shows the thresholded face image, and Fig.2(b) shows the eyebrows, location of the eyes and other facial regions. The position of facial regions are estimated relative to the location of the eyes, and each region has 9 pixels. The average gray value in each region is used as an element in the 44 dimensional feature vector $x_1, x_2, ..., x_{44}$. If the distance from the center of the eyes to the pixel (c_x, c_y) is used for estimating the top 8 facial regions in Fig.2(b), then the features are invariant to yaw of the face to certain extent.



(a) Thresholded face (b) Facial regions

Fig. 2. Feature extraction.

4. AUTOASSOCIATIVE NEURAL NETWORK MODEL FOR FACE RECOGNITION

Autoassociative neural network models are feedforward neural networks performing an identity mapping of the input space, and are used to capture the distribution of the input data [13],[14]. AANN models are used to capture the distribution of the feature vectors for each subject.

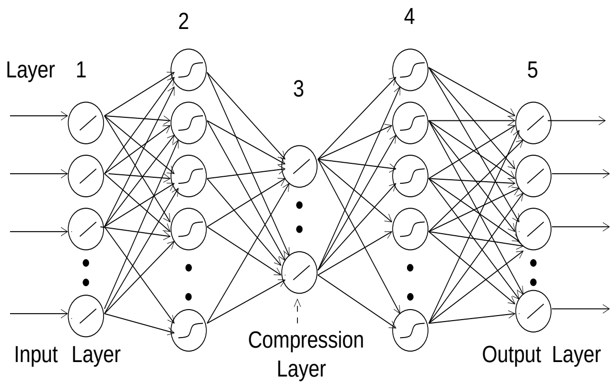


Fig. 3. AANN model used for face recognition.

The five layer Autoassociative neural network model as shown in Fig.3 is used to capture the distribution of the extracted gray level features. The second and fourth layers of the network have more units than the input layer. The third layer has fewer units than the first or fifth. The activation functions at the second, third and fourth layer are nonlinear. The structure of the AANN model used in our study is 44L 60N 20N 60N 44L, where L denotes a linear unit and N denotes a nonlinear unit. The integer value indicates the number of units used in that layer. The nonlinear units use $tanh(s)$ as the the activation function, where s is the activation value of the unit. The standard backpropagation learning algorithm is used to adjust the weights of the network to minimize the mean square error for each feature vector.

The 44 dimensional feature vector $x_1, x_2, ..., x_{44}$, is normalized as follows:

$$y_i = \frac{2(x_i - x_{min})}{(x_{max} - x_{min})} - 1 \tag{5}$$

where x_{max} and x_{min} are the maximum and minimum values in the feature vector. The normalized feature vector $y_1, y_2, ..., y_{44}$, is used as input to the AANN model. Feature vectors are extracted from 400 face images for each subject for training the AANN model. The feature vectors from face images are extracted in two sessions (day time and night time), in order to handle the lighting effects in uncontrolled environment. In each session, 200 face images of each subject with variation in size, tilt of the face, and also variation in yaw and pose to certain extent, are used to form the training set. In this system there is no need to store the face images in the database. The AANN model is trained using the standard backpropagation learning algorithm for 1000 epochs. One such model is created for each subject. A new subject can be easily added to the system without affecting the existing models. The training of a subject with 200 feature vectors for 1000 epochs requires approximately 5 minutes on a Pentium machine at 500 MHz.

For testing, the 44 dimensional feature vector is extracted from the face image of the subject and is given to all the models. The output of each model is compared with the input to compute the square error. The error (ε) is transformed into a confidence value (c) by using the equation $c = exp(-\varepsilon)$. The confidence values from the models are used to decide the identity of the subject.

5. EXPERIMENTAL RESULTS

Performance of the proposed real-time face recognition system is evaluated for 25 subjects using a camera with a resolution of 160x120. The face region is estimated as described in Section 2, and the method is not sensitive to size of the face and lighting conditions. This method requires head movement in order to estimate the face region. The estimated face region is not changed until there is a significant motion of the head in the video. The location of the eyes and other facial regions are identified as described in Section 3, and the feature vector is extracted from the face image only when the following three heuristics are satisfied:

1. The distance between the eyes lies between $0.5 * w_1$ and $0.25 * w_1$, where w_1 is the width of the face.
2. The angle of the line connecting the eyes lies between $+30°$ and $-30°$.
3. All the 44 facial regions are inside the face or head region.

Fig.4 shows sample face images used to extract the feature vector. The results show that the feature extraction technique is invariant to size and tilt of the face.

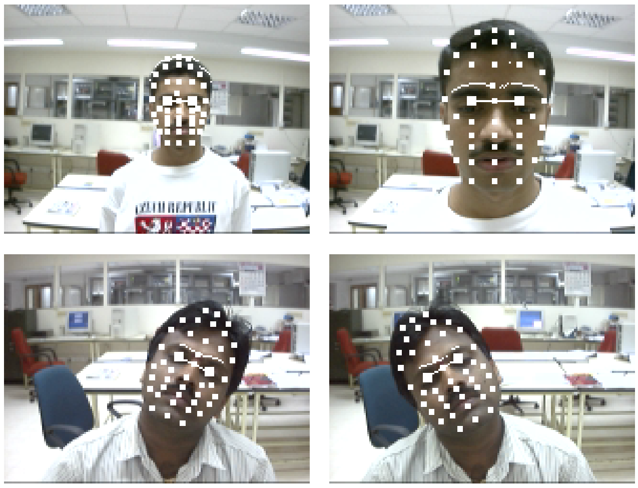


Fig. 4. Sample face images.

The AANN models are trained and tested as described in Section 4 for 25 subjects. For testing, we collected feature vectors from 10 face images for each subject with variation in size and tilt of the face one week after the training. The average confidence value is calculated for each model. The identity of a subject is decided based on the highest confidence value. Fig.5 shows the average confidence values for 10 test subjects of the experiment against 10 corresponding models. High confidence values in the diagonal indicate that the system recognized all the test subjects correctly. The performance of the system is invariant to the size and tilt of the face. The system tests the identity of a subject in real time, and it gives an average recognition rate of 99%. The face detection and feature vector extraction techniques are computationally inexpensive, and testing a feature vector against 25 models requires less than 0.15 sec on a Pentium machine at 500 MHz.

6. CONCLUSION

In this paper, we have proposed a method for real time face recognition using autoassociative neural network models.

Test subjects	Models									
	m1	m2	m3	m4	m5	m6	m7	m8	m9	m10
t1	0.96	0.04	0.67	0.76	0.79	0.88	0.63	0.17	0.82	0.37
t2	0.13	0.92	0.58	0.41	0.09	0.28	0.12	0.76	0.12	0.65
t3	0.87	0.31	0.97	0.90	0.65	0.72	0.79	0.61	0.89	0.72
t4	0.69	0.31	0.74	0.92	0.31	0.58	0.45	0.49	0.71	0.67
t5	0.83	0.11	0.62	0.58	0.93	0.73	0.75	0.40	0.81	0.51
t6	0.87	0.24	0.79	0.79	0.77	0.93	0.78	0.44	0.81	0.58
t7	0.82	0.28	0.81	0.77	0.78	0.84	0.96	0.51	0.92	0.78
t8	0.18	0.76	0.78	0.74	0.16	0.27	0.18	0.97	0.08	0.47
t9	0.84	0.19	0.76	0.87	0.70	0.57	0.73	0.54	0.95	0.73
t10	0.81	0.17	0.76	0.89	0.74	0.83	0.77	0.47	0.84	0.97

Fig. 5. Confidence values for 10 test subjects.

The proposed method extracts features relative to the location of the eyes, and hence it is invariant to size and tilt of the face. Experimental results show that the system gives an average recognition rate of 99% in real time for 25 subjects under natural lighting conditions. The proposed system is efficient in terms of computational power and memory requirement. In this system, a new model can be easily added without changing the existing models. Better discrimination in the confidence values can be achieved by adding additional facial features like appearance of the face and color information. The system can be extended to recognize people in a still image or video database by locating their eyes using the technique described in [10].

7. REFERENCES

[1] W. Zhao, R. Chellappa, A. Rosenfeld, and P.J. Phillips, *Face Recognition: A Literature Survey*, UMD CAFR, Technical Report, CAR-TR-948, October 2000.

[2] M. Turk and A. Pentland, “Eigenfaces for recognition,” *Journal of Cognitive NeuroScience”*, vol. 3, pp. 71–86, 1991.

[3] Constantine L. Kotropoulos, Anastasios Tefas, and Ioannis Pitas, “Frontal face authentication using discriminating grids with morphological feature vectors,” *IEEE Transactions on MultiMedia*, vol. 2, no. 1, pp. 14–26, March 2000.

[4] Dominique Valentin, Herve Abdi, Alice J. O’Toole, and Garrison W.Cottrell, “Connectionist models of face processing: A survey,” *Pattern Recognition*, vol. 27, pp. 1209–1230, 1994.

[5] Lian Hock Koh, Surendra Ranganath, and Y.V. Venkatesh, “An integrated automatic face detection and recognition system,” *Pattern Recognition*, vol. 35, pp. 1259–1273, 2002.

[6] Yongsheng Gao and Maylor K.H.Leung, “Face recognition using line edge map,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 6, pp. 764–779, June 2002.

[7] Constantine Kotropoulos and Ioannis Pitas, “Using support vector machines to enhance the performance of elastic graph matching for frontal face authentication,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 7, pp. 735–746, July 2001.

[8] Ming-Hsuan Yang, David J. Kriegman, and Narendra Ahuja, “Detecting faces in images: A survey,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 1, pp. 34–58, January 2002.

[9] Henry A. Rowley, Shumeet Baluj, and Takeo Kanade, “Neural network-based face detection,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 1, pp. 23–38, January 1998.

[10] Rein-Lien Hsu, Mohamed Abdel-Mottaleb, and Anil K. Jain, “Face detection in color images,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 696–706, May 2002.

[11] Baoxin Li and Rama Chellappa, “A generic approach to simultaneous tracking and verification in video,” *IEEE Transactions on Image Processing*, vol. 11, no. 5, pp. 530–544, May 2002.

[12] Christophe Garcia and Georgios Tziritas, “Face detection using quantized skin colour regions merging and wavelet packet analysis,” *IEEE Transactions on MultiMedia*, vol. 1, no. 3, pp. 264–277, September 1999.

[13] B.Yegnanarayana, S.R. Mahadeva Prasanna, and Suryakanth V. Gangashetty, “Autoassociative neural network models for speaker recognition,” in *Proceedings of International workshop on Embedded Systems*, Hyderabad,India, December 2001.

[14] B.Yegnanarayana, Suryakanth V. Gangashetty, and S. Palanivel, “Autoassociative neural network models for pattern recognition tasks in speech and image,” in *Soft Computing Approach to Pattern Recognition and Image Processing*, World Scientific publishing Co. Pte. Ltd, Singapore, december 2002, pp. 283–305.