Object Recognition, A Particular Reference to Face Recognition

Praseeda Lekshmi.V, Dr.M.Sasikumar, Divya.S.Vidyadharan

Abstract—**Pattern recognition problem rely upon the features inherent in the pattern of images. Face detection and recognition is one of the challenging research areas in the field of computer vision. In this paper, we present a method to identify skin pixels from still and video images using skin color. Face regions are identified from this skin pixel region. Facial features such as eyes, nose and mouth are then located. Faces are recognized from color images using an RBF based neural network. K means clustering algorithm is used to locate different facial elements. Orientation is corrected by using eyes. Parameters like inter eye distance, nose length, mouth position, Discrete Cosine Transform (DCT) coefficients etc. are computed and used for a Radial Basis Function (RBF) based neural network. This approach reliably works for face sequence with orientation in head, expressions etc.**

Index Terms— **Pattern recognition, Face recognition, Radial Basis Function, Discrete Cosine Transform, Skin pixel.**

I. Introduction

The technological advancement in the area of digital processing and imaging has led to the development of different algorithms for various applications such as automated access control, surveillance etc. For automated access control, most common and accepted method is based on face detection and recognition. Face recognition is one of the active research areas with wide range of applications. The problem is to identify facial image/region from a picture/image. Generally pattern recognition problems rely upon the features inherent in the pattern for efficient solution. Though face exhibits distinct features which can be recognized almost instantly by human eyes, it is very difficult to extract and use these features by a computer. Human can identify faces even from a caricature. The challenges associated with face detection and recognition are pose, occlusion, skin color, expression presence or absence of structural components, effects of light, orientation, scale, imaging conditions etc. Most of the currently proposed methods use parameters extracted from facial images. For access control application, the objective is to authenticate a person based on the presence of a recognized face in the database.

In this paper, skin and non-skin pixels are separated and the pixels in the identified skin region are grouped to obtain face region. From the detected face area, the facial features such as eyes, nose and mouth are located.

Rest of the paper is organized as follows. Section 2 gives background and related works. Section 3 discusses the proposed method. Results are given in section 4. Conclusion and future work are given in section 5.

II. BACKGROUND AND RELATED WORK

A lot of research has been going on in the area of human face detection and recognition [3]. Most face detection and recognition methods fall into two categories: Feature based and Holistic. In feature-based method, face recognition relies on localization and detection of facial features such as eyes, nose, mouth and their geometrical relationships. In holistic approach, entire facial image is encoded into a point on high dimensional space. Principal Component Analysis (PCA) and Active Appearance Model (AAM) [9] for recognizing faces are based on holistic approaches. In another approach, fast and accurate face detection is performed by skin color learning by neural network and segmentation technique

Manuscript received March 21, 2008.

F. A. Praseeda Lekshmi.V, Assistant Professor is with the College of Engineering Kidangoor Kottayam dist, Kerala, India, phone: 09447963211; e-mail: vplekshmi@yahoo.com.
S. B. Dr.M.Sasikumar, Professor is with Marian Engineering College, Trivandrum, Kerala, India. e-mail: drmsasikumar@yahoo.com.
T. C. Divya.S.Vidyadharan, Sr.Staff Scientist is with the C-DAC, Trivandrum, Kerala, India. e-mail: divyasvidya@yahoo.co.in

[4]. Independent Component Analysis (ICA) was performed on face images under two different conditions [8]. In one condition, image is treated as a random variable and pixels are treated as outcomes and in the second condition pixels are treated as random variables and image as outcome. Facial expressions are extracted from the detailed analysis of eye region images is given in [2]. Large range of human facial behavior is handled by recognizing facial muscle actions that produce expressions is given in [10]. Video based face recognition is explained in [5].

III. METHOD

The method explains detection of faces from video frames and still images. This is followed by extraction of facial features.

A. Facial region Identification

The first step in face detection problem is to extract facial area from the background. In our approach, both still and video images are used for face detection. Image frames from video are extracted first. The input images contain regions other than face such as hair, hat etc. Hence it is required to identify the face. Each pixel in the image is classified as skin pixel or non skin pixel. Different skin regions are detected from the image. Face regions are identified from the detected skin region as in [7] which addressed the problem of face detection in still images. Some randomly chosen frames with different head pose, far away from the camera, and expressions from video face images are extracted as shown in Fig 1. The difference image at various time instances is shown in Fig 2.

Fig. 1. Some frames from face video database

Fig. 2. Difference image

The portion of the image that is moving is assumed as head. The face detection algorithm used RGB color space for the detection of skin pixels. The pixels corresponding to skin color of the input image is classified according to certain heuristic rules. The skin color is determined from RGB color space as explained in [1] and [6]. A pixel is classified as skin pixel if it satisfies the following conditions.

$$R > 95 \text{ AND } G > 40 \text{ AND } B > 20 \text{ AND } \max\{R, G, B\} - \min\{R, G, B\} > 15 \text{ AND } |R-G| > 15 \text{ AND } R > G \text{ AND } R > B \quad (1)$$

OR

$$R > 220 \text{ AND } G > 210 \text{ AND } B > 170 \text{ AND } |R-G| \leq 15 \text{ AND } R > B \text{ AND } G > B \quad (2)$$

Edge detection is performed in each frame. Edge detection and skin regions identified from the color images of video frames and still images are shown in Fig 3 and Fig 4.

From these skin regions, it is possible to identify whether a pixel belongs to skin region or not. To find the face regions, it is necessary to categorize the skin pixels in to different groups so that it will represent some meaningful groups such as face, hand etc. Connected component labeling is performed to classify the pixels. In the connected component labeling operation, pixels are connected together geometrically. In this, we used 8-connected component labeling so that each pixel is connected to its eight immediate neighbors. At this stage, different regions are identified and we have to classify each region as a face or not.

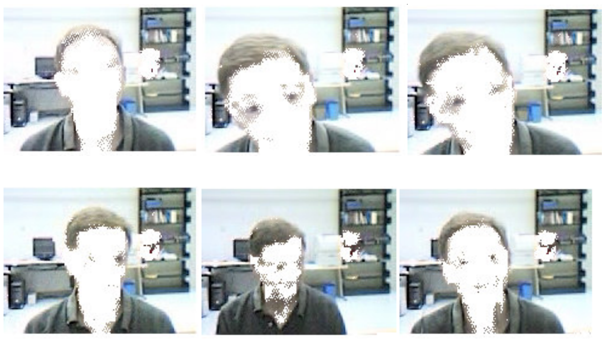

Fig 3. Edge detection and Skin region identification from video frames

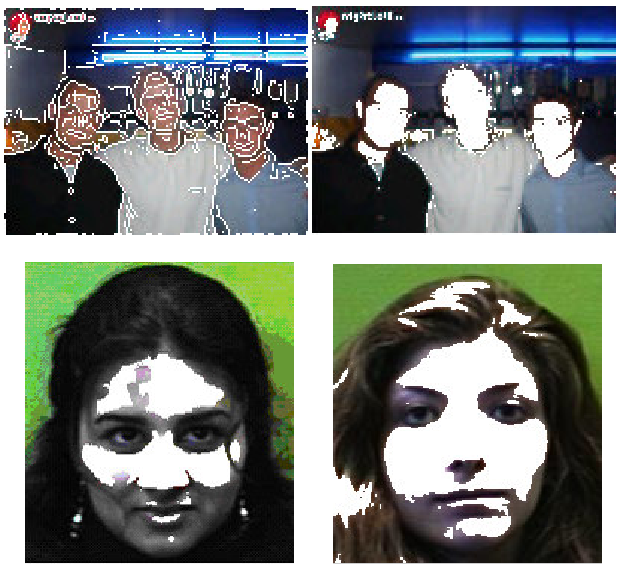

Fig 4. Edge detection and Skin region identification from still images

This is done by finding the skin area of each region. If the height to width ratio of those skin region falls with in the range of golden ratio $((1+\sqrt{5})/2 \pm$ tolerance, then that region is considered as a face region

B. Facial feature Extraction

B1. Segmentation

Image segmentation is a long-standing problem in computer vision. There are different segmentation techniques that divide spatial area with in an image to different meaningful components. Segmentation of images is based on the discontinuity and similarity properties of intensity values. Cluster analysis is a method of grouping objects of similar kind in to respective categories. It is an exploratory data analysis tool which aims at sorting different objects in to groups in a way that degree of association between two objects is maximal if they belong to same group and minimal otherwise. K-Means is one of the unsupervised learning algorithms that solve the clustering problems. The procedure follows a simple and easy way to classify a given data set through a certain number of clusters (assume K clusters) fixed apriori. The idea is to define K centroids, one for each cluster. These centroids should be placed in a cunning way because of different location causes different results. So the better choice is to place them as much as possible far away from each other. The next step is to take each point belonging to a given dataset and associate it to the nearest centroid. When no point is pending, the first step is completed and an early groupage is done. At this point K new centroids need to be re-calculated as barycentres of the clusters resulting from previous step. After getting these K new centroids, a new binding has to be done between the same dataset points and the nearest new centroid. A loop has been generated. As a result of this loop we may notice that the K centroids change their location step by step until no more changes are done. If we know the number of meaningful groups/classes based on the range of pixel intensity, weighted K-means clustering can be used to cluster the spatial intensity values.

In facial images the skin color and useful components can be generally classified as two different classes. But if we use two class based clustering, it may result in components that may be still difficult to identify. So we used three classes and are able to cluster the data in useful manner. Initial cluster centers are calculated using histogram. Then K-means clustering algorithm computes distance between the different pixels and cluster centers and selects a minimum distance cluster for each pixel. This process continues until all pixels are classified properly.

The results of clustering algorithm are shown in Fig 5. Class–I is selected since it is possible to separate the components properly compared to other classes. Then connectivity algorithm is applied to all the components in the clustered face.

Class-I Class-III Class-III

Fig 5.Clustering

ISBN:978-988-17012-3-7

B2. Feature extraction

Eye regions are located in the upper half of skin region and can be extracted using the area information of all the connected components. Using eye centers, orientation is corrected by rotation transformation. After calculating the inter eye distance, nose and mouth are identified as they generally appear along the middle of two eyes in the lower half. The inter eye distance, nose length and mouth area are computed

To keep track of the overall information content in the face area, we used DCT. On applying DCT, most of the energy values can be represented by a few coefficients. First 64 x 64 coefficients are taken as part of the feature set. The face area calculated from the skin region is taken as another parameter.

C. *Radial basis function*

Radial functions are a special class of function. Their characteristic feature is that their response decreases (or increases) monotonically with distance from a central point. The centre, the distance scale, and the precise shape of the radial function are parameters of the model, all fixed if it is linear. A typical radial function is the Gaussian which, in the case of a scalar input, is

$$h(x) = exp(-(x-c)^2/r^2). \quad (3)$$

Its parameters are its centre c and its radius r. Universal approximation theorems show that a feed forward network with a single hidden layer with non linear units can approximate any arbitrary function[Y]. No learning is involved in RBF networks. For pattern classification problems, the numbers of input nodes are equal to the number of elements in the feature vector and the numbers of output nodes are equal to the number of different clusters .

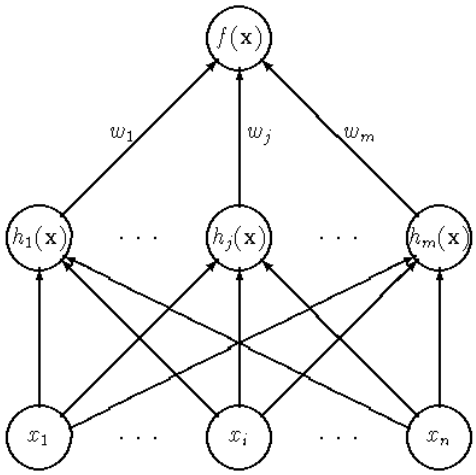

Fig 6. Radial Basis Function Network

IV RESULTS

In this experiment 200 frames of video images with considerable variations in head poses, expressions, camera viewing angle are used. The 'face 94' color image database is used for still images. We selected 100 face images with considerable expression changes and minor variation in head turn, tilt and slant. The performance ratios are 90% for video image sequence and 92% for still images. The results of face detection for both still and video images are shown in figure 7 and figure 8. Figure 9 shows distinct connected components. Rotation transformation is shown in figure 10. Figure 11 shows the located features. Geometric parameters and DCT coefficients were given to a classification network for recognition.

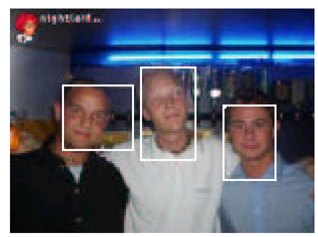

Fig 7. Face detection from still images

Fig. 8.Face detection from video images

Fig 9. Distinct Connected Components

Fig 10.Eye region Image after angle Correction

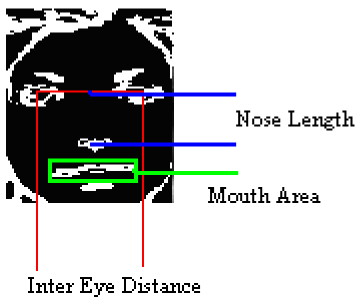

Fig 11. Located Features

V. CONCLUSION

In this paper, faces are detected and facial features are located from video and still images. 'NRC-IIT Facial video database' is used as image sequences and 'face 94 color image database' is used for still images. Skin pixels and non-skin pixels are separated and skin region identification is done by RGB color space. From the extracted skin region, skin pixels are grouped to some meaningful groups to identify the face region. From the face region, facial features are located using segmentation technique. Orientation correction is done by using eyes. Parameters like inter eye distance, nose length, mouth position, and DCT coefficients are computed which is used for a RBF based neural network. In this experiment only one image sequence is used for detection of faces.

Future work in the approach includes recognition of faces from video sequences. It is also proposed to analyze facial expressions from video image sequences.

REFERENCES

[1] P. Peer, F. Solina, "An automatic human face detection method", Proceedings of 4th Computer Vision Winter Workshop (CVWW'99),Rastenfeld, Australia, 1999, pp.122-130

[2] Tsuyoshi Moriyama, Takeo Kanade, Jing Xiao and Jeffrey F. Cohn" Meticulously Detailed Eye region Model and It's Application to Analysis of Facial Images", IEEE Transactions in Pattern Analysis and Machine Intelligence, Vol.28, No.5, May 2006

[3] Chellappa. R, Wilson and Sirohey.S, "Human and machine recognition of faces: A Survey" In Proc.IEEE, 83(5): 1995, 705-740

[4] Hichem Sahbi, Nozha, Boueimma, Tistarelli, J. Bigun, A.K.Jain (Eds), and "Biometric Authentication "LNCS2539, Springer Verlag Berlin Heidelberg,

[5] Dmitry O. Gorodnichy "Video-based framework for face recognition in video", Second Workshop on Face Processing in Video. (FPiV'05) in Proceedings of Second Canadian Conference on Computer and Robot Vision (CRV'05), Victoria, BC, Canada, 9-11 May, 2005. pp. 330-338

[6] Franc Solina,Peter Peer,Borut Batagelj,Samo Juvan, "15 Seconds of Frame-An Interactive Computer Vision Based Art Installation", Proceedings of 7th International Conference on Control ,Automation, Robotics and Vision (ICARCV 2002), Singapore, 2002, pp.198-204,

[7] K.Sandeep and A.N. Rajagopalan,"Human Face Detection in Cluttered Color Images Using Skin Color and Edge Information"

[8] Marian Stewrt Bartlett, Javier R, Movellan and Terrence J Seinowski , "Face Recognition by Independent Component Analysis", IEEE Transactions on Neural Networks, Vol.113, No.6, November 2002

[9] Nathan Faggian, Andrew Paplinski, Tat-Jun Chin," Face Recognition from Video using Active Appearance Model Segmentation"

[10] Yan Tong, Weinhui Lio, Qiang Ji,"Facial Action Unit Recognition by Exploiting Their Dynamic and Semantic Relationships", IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol.29, No.10, October 2007