Fusion of Facial Features using DS Theory for Face Recognition

Dakshina Ranjan Kisku1, Ajita Rattani2, Phalguni Gupta3, and Jamuna Kanta Sing4
1. Asansol Engineering College, Asansol, India
2. University of Cagliari, Cagliari, Italy
3. Indian Institute of Technology Kanpur, Kanpur, India
4. Jadavpur University, Kolkata, India

Abstract—This paper proposes an efficient face recognition system that uses salient facial features extracted from eyes, mouth and nose of a face image. SIFT descriptor is used to determine all facial keypoints from each landmark. These feature points are considered to obtain a relaxation graph. Given two relaxations graphs from a pair of faces, matching scores between two corresponding feature points has been determined with the help of iterative graph relaxation cycles. Dempster-Shafer decision theory is used to fuse all these matching scores obtained from each pair of facial landmarks for taking the final decision. The proposed technique has been tested against the three databases, namely, FERET, ORL and IITK face databases. The experimental results exhibit efficacy of the proposed face recognition system.

Index Terms—Face recognition, SIFT descriptor, Relaxation graph, Feature fusion, Dempster-Shafer Decision theory

I. INTRODUCTION

Face recognition [1] is considered to be a difficult person recognition task in machine vision. It is due to the variable appearance of face images and changes in appearance may occur due to many factors, such as facial attributes compatibility complexity, the motion of face parts, facial expression, pose, illumination and partly occlusion. As a result face recognition problem become ill-posed.

There exist several techniques to similar faces with identical characteristics from a set of faces and to accelerate the face matching strategy. These techniques can broadly classify into three categories: appearance-based, model-based and feature-based techniques. The appearance-based techniques are Principal Component Analysis (PCA) [1], Linear Discriminant Analysis (LDA) [1], Fisher Discriminant Analysis (FDA) [1], and Independent Component Analysis (ICA) [1]. Among the various model based techniques for face recognition AAM [10] is worth to mention. Some local feature based techniques have been proposed in [2], [3]. One of the most well known feature based technique which is known as Elastic Bunch Graph Matching (EBGM) has been discussed in [2]. EBGM has been used to represent faces as graphs where vertices are the fiducial points (e.g., eyes, nose) and edges are the geometric distance between two fiducial points. Each vertex contains a set of 40 complex Gabor wavelet coefficients at different scales and orientations. They are called a Gabor Jet. A graph similarity function has been proposed in [2] and used for identification. The local feature based proposed in [3], is based on graph matching topology drawn on SIFT features [4].

This paper proposes a new local feature based face recognition technique which makes use of mouth, eyes and nose as salient features of face obtained through SIFT operator [4]. To capture the face variations, face characteristics of dynamic (mouth) and static parts (eyes, nose) are further represented by incorporating probabilistic graph relaxations [9] drawn on SIFT features extracted from localized mouth, eyes and nose facial parts. The proposed technique has made an attempt to handle this problem. It first detects and extracts automatically the salient facial parts such as eyes, mouth and nose with the help of the algorithm proposed in [5], [6]. Invariant feature descriptor is used on each of these facial parts to determine SIFT features and finally a relaxation graph is drawn on the SIFT features for each facial part. For Matching between two faces is done by considering the pair of relaxation graphs of each salient facial part (i.e., nose, mouth, left eye and right eye) of two faces. For each point of a relaxation graph corresponding point having maximum score in another relaxation graph of the pair is obtained. These matching scores obtained from various facial parts are fused using Dempster-Shafer theory [7]. The proposed techniques are evaluated against three face databases, namely, FERET [11], ORL (formerly known as AT&T) [8] and IITK face databases.

II. LANDMARK DETECTION AND FEATURE EXTRACTION

The eyes, mouth and nose positions are automatically located by applying the technique proposed in [5], [6]. A circular region of interest (ROI), centered at each extracted facial landmark location is considered to determine the SIFT features of the landmark.

SIFT descriptor [4] is invariant to image rotation, scaling, partly illumination changes and the 3D camera view and it detects feature points efficiently through a staged filtering approach that identifies stable points in the scale-space of the image of pyramid obtained by convolving the image by a set of difference of Gaussian kernels. Each feature point is composed of four types of information – spatial location (x, y), scale (S), orientation (θ) and Keypoint descriptor (K). In this paper, only
keypoint descriptor has been used. The descriptor consists of a vector of 128 elements representing changes in neighborhood intensity of current points. More formally, these keypoint descriptors represent local shape distortions and illumination changes. In Fig. 1, SIFT features are shown for a pair of face images.

![SIFT features of a pair of faces](image)

**III. GRAPH REPRESENTATION**

Deformable objects are generally difficult to characterize with a rigid representation in feature spaces for recognition. Different facial regions, not only convey different and redundant information on the subject’s identity, but also differ from different time variability either due to motion or illumination changes.

In order to interpret the facial landmarks with invariant SIFT points and graph relaxation topology [9], each extracted feature can be thought as a node and the relationship between invariant points can be considered as edge between two nodes. At the level of feature extraction, invariant SIFT feature points are extracted and facial landmarks location with the landmark detection algorithms discussed in [5], [6]. The final similarity measure can be obtained from the feature points extracted from facial landmarks only. Relaxation graphs are then drawn on the features extracted from landmarks. These relaxations are used for matching and generation of match scores.

**IV. FUSION OF FACIAL FEATURES**

The Dempster-Shafer decision theory [7] is applied to combine the matching scores obtained from individual landmark. It is based on combining the evidences obtained from different sources to compute the probability of an event. This is obtained combining three elements: the basic probability assignment function (bpa), the belief function (bf) and the plausibility function (pf).

Let $\Gamma_{\text{left-eye}}$, $\Gamma_{\text{right-eye}}$, $\Gamma_{\text{nose}}$ and $\Gamma_{\text{mouth}}$ be the individual matching scores obtained from the four different matching of salient facial landmarks. Now, in order to obtain the combine matching score determined from the four salient landmarks pairs, Dempster combination rule has been applied. First, we combine the matching scores obtained from the pairs of left-eye and nose landmark features and in the next, we combine the matching scores obtained from the pairs of right-eye and mouth landmark features. Finally, we combine the matching scores determined from the first and second processes. Also, let $m(\Gamma_{\text{left-eye}})$, $m(\Gamma_{\text{right-eye}})$, $m(\Gamma_{\text{nose}})$ and $m(\Gamma_{\text{mouth}})$ be the bpa functions for the Belief measures $\text{Bel}(\Gamma_{\text{left-eye}})$, $\text{Bel}(\Gamma_{\text{right-eye}})$, $\text{Bel}(\Gamma_{\text{nose}})$ and $\text{Bel}(\Gamma_{\text{mouth}})$ for the four classifiers, respectively. Then the Belief probability assignments (bpa) $m(\Gamma_{\text{left-eye}})$, $m(\Gamma_{\text{right-eye}})$, $m(\Gamma_{\text{nose}})$ and $m(\Gamma_{\text{mouth}})$ can be combined together to obtained a Belief committed to a matching score set $C \in \Theta$ according to the following combination rule or orthogonal sum rule

$$m(C_i) = m(\Gamma_{\text{left-eye}}) \oplus m(\Gamma_{\text{nose}})$$

$$\frac{1}{1 - \sum m(\Gamma_{\text{left-eye}}) m(\Gamma_{\text{nose}})}, \ C_i \neq \emptyset.$$  

$$m(C_i) = m(\Gamma_{\text{left-eye}}) \oplus m(\Gamma_{\text{mouth}}) =$$

$$\frac{1}{1 - \sum m(\Gamma_{\text{left-eye}}) m(\Gamma_{\text{mouth}})}, \ C_i \neq \emptyset.$$  

$$m(C) = m(m(C_1)) \oplus m(m(C_2)) =$$

$$\frac{1}{1 - \sum m(m(C_1)) m(m(C_2))}, \ C \neq \emptyset.$$  

The denominator in equations (1), (2) and (3) are the normalizing factors which denotes the art of Belief probability assignments $m(\Gamma_{\text{left-eye}})$, $m(\Gamma_{\text{right-eye}})$, $m(\Gamma_{\text{nose}})$ and $m(\Gamma_{\text{mouth}})$ are conflicting.

Let $m(m(C_1))$ and $m(m(C_2))$ are the two matching score sets obtained from the local and global matching strategies. They can be fused together recursively as:

$$m(\text{FMS}) = m(m(C_1)) \oplus m(m(C_2))$$  

where $\oplus$ denotes the Dempster combination rule. The final decision of user acceptance and rejection can be established by the following equation and by applying the threshold $\Psi$ to the final match $m(\text{FMS})$

$$\text{decision} = \begin{cases} 
\text{accept, if } m(\text{FMS}) \geq \Psi \\
\text{reject, otherwise} 
\end{cases}$$

**V. EXPERIMENTAL RESULTS**

The effectiveness of the proposed facial landmark based face matching strategy has been investigated with the three face databases, namely FERET [11], ORL [8] and IITK face databases.

The FERET face recognition database is a collection of face images acquired by NIST. For this evaluation, 1,396 face images are considered as training dataset out of which...
200 images labeled as bk; that is 1396 face images in database. For query set we have considered 1195 images that are labeled as fafb. All these images have been downsampled to 140x100 from the original size of 150x130. For testing purpose, we take fa labeled dataset of 1,195 and the duplicate 1 dataset of 722 face images as probe set. Prior to processing, the faces are well registered to each other and the background effects are eliminated. Moreover, only the frontal view face images are used, which have natural facial expressions (fa) and the face images which have taken under different lighting conditions. The Receiver operating characteristics (ROC) determined from the FERET dataset given in Fig. 2. The recognition accuracy of the proposed system when tested on FERET dataset found to be 92.34%. Consequently, our proposed result proved to be an appropriate one for changing illumination and facial expression.

The IITK face database consists of 1200 face images with four images per person (300X4). These images are captured under control environment with ±20° changes of head pose and with at most uniform lighting and illumination conditions, and with at most consistent facial expressions. For the face matching, all probe images are matched against all target images. From the ROC curve in Fig. 2 it has been observed that the recognition accuracy is 93.63%, with the false accept rate (FAR) of 5.82%.

![Figure 2. ROC curves Determined for Different Databases](image)

The ORL face database consists of 400 images taken from 40 subjects. Out of these 400 images 200 face images are considered for experiment. It has been observed that there exact changes in orientation in images which lying between -20° and +30°. The face images are found to have the variations in pose and facial expression (smile/not smile, open/closed eyes). The original resolution of the images is 92 x 112 pixels. However, for the experiment the resolution is set to 120x160 pixels. From the ROC curve in Fig. 2 it has been observed that the recognition accuracy for the ORL database is 97.33%, yielding FAR is about 2.14%. The relative accuracy of the proposed matching strategy for ORL database increases of about 3% and 5% over the IITK database and the FERET database respectively.

VI. CONCLUSION

An efficient and robust face recognition technique has proposed in this paper. During the face recognition process, the human faces are characterized on the basis of local salient landmark features. In the proposed face recognition method, local facial landmarks are considered for further processing. The optimal face representation using graph relaxation drawn on local landmarks then allows matching the localized facial features efficiently by searching the correspondence of keypoints using iterative relaxation. The experimental results are found to be encouraging.

REFERENCES