Component based Face Recognition using Hybrid Approach

Srinivasa Perumal R* and Chandra Mouli P.V.S.S.R**
*School of Information Technology and Engineering
r.srinivasaperumal@vit.ac.in
**School of computing Science and Engineering
chandramouli@vit.ac.in

Abstract: This paper presents a hybrid framework for face recognition based on the component features that are extracted by the salient points from the facial components of the face. Face recognition is still challenging under different pose, illumination conditions etc. The Proposed method detects the face from the original image. The components (eyes, nose and mouth) are extracted from the detected face through Viola-Jones algorithm. Extract the component features using Speeded up Robust Features (SURF). Extracted features are concatenated into a single feature vector to represent the face and compute the similarity between the query face and feature vectors stored in the database. The proposed hybrid method is implemented on FERET and ORL datasets and it improves the recognition rate and decreases the system running time.

Keywords: Viola Jones Algorithm, SURF, Face Detection, Hybrid face detection method, component based face detection.

Introduction

Face detection plays a vital role in automated human recognition. The methods are broadly categorized into two categories namely holistic approaches and Component based approaches. The Holistic face approaches utilize global information from face which is represented by features. Local Binary Patterns [7], biologically inspired features [16], Scale Invariant Feature Transform [21] and appearance-based representations [2] are some of the holistic approaches. In component based representations, features are extracted from specific facial components. Using the features from the facial components, the similarity measure is calculated and the decision is chosen based on the similarity score. Component based face recognition is the order of the day.

Schwaninger et al. [14] proved that a human has the ability to identify the faces even with different facial expressions, pose invariant and occluded image through natural intelligence. The current evidence [5] shows that idea supports the face processing to integrate each individual component of the person. The facial components like eyes, nose and mouth are integrated by human and compared with optimal Bayesian integrator based on component Processing. The comparison results show that the Bayesian framework performs slightly better than human performance. In addition, the behavioral characteristics of human components will be taken into account, such as outline of the face, eyes, nose and mouth. The reason behind the selection of these characteristics is that the upper part of the face contains more useful information than the lower part. The existing approaches that fall under this category are Template Matching [1], Elastic Bunch Graph Matching (EBGM) [3], Linear Discriminant Analysis (LDA) [8, 11], Cascaded LDA [10], Local Binary Pattern (LBP) [13], and Appearance based Method [23].

Template matching techniques fall under the component based approaches. R. Brunelli and T. Poggio [1] proposed a template matching based face recognition. In this method, facial components (eyes, eyebrows, nose and mouth) are extracted using geometrical features. From each component, features are extracted based on the intensity value lying in the template and generate set of templates for each face. A set of 35 features are extracted for each facial component and finally face is recognized using Bayes classifier. EBGM [3] computes the fiducial points for each face that is characterized by Gabor wavelet coefficients. The fiducial points and their relative positions are extracted and formed into an image graph. This method recognizes the person from known class and the performance will be decreased if huge variation occurs between the images.

Local Discriminant Analysis (LDA) [8, 11] explores the feature space which maximizes between-class scatter and minimizes within-class scatter. The face is divided into five regions such as lower part and upper part of mouth, nose and left and right eye region. Each region describes the discriminant information. Then it constructs the component bunches in order to discriminate information. It provides higher recognition rate under pose change and lighting conditions. Cascaded LDA [10] method divides the face into four overlapped regions. The features are extracted from overlapped regions of face through Eigen vector and determine the discriminating information of overlapped regions.

Lin and Tang [12] introduced a framework for high resolution face recognition manipulating features in multiple scales. Here the face image is factorized into four layers: global appearance, facial organs, skins, and irregular details. This system applied Multilevel PCA followed by Regularized LDA to model global appearance and facial organs. However, the system is not suitable for representing of irregular details and texture of skin. Lin and Tang address the issue by using SIFT and Multiscale texton features are used to describe the texture of skin and irregular details. To increase performance the systems combine the information conveyed by all layers. K. Pan et al. [13] proposed Local Binary Patterns for extraction of local features. The purpose of LBP is used to extract the texture information and to enhance the feature vectors. LBP feature vectors are adjusted with monotonic transform for higher recognition rate at different illumination conditions. In case of varying pose and facial expressions, the accuracy of LBP fails.

Heisele et al. [15] proposed a component based face detection using 3D models, it determine a set of templates which is specific to faces consisting of the face components and their relations. Component based face detection and verification using online learning based on unsupervised clustering. It provides more accurate results and reduces the execution time. This system is more flexible to use to describe the features in appearance based method as well as geometry based method. Appearance-based [23] used anthropometric metric to extract the components. Active shape model identify the location of facial components such as eyes, eyebrows, mouth and nose. The features are extracted by multi scale local binary pattern. This method improves the recognition rate in pose variation as well as for occluded images.

The above existing methods extract the facial features approximately by predefined position of the face. Instead of using predefined position of the face the hybrid approach identifies the most appropriate method for determining facial component. In component based representation, face images are divided into different facial components and then compute the individual feature vectors for each facial component. The extracted features are finally concatenated. The component based framework demonstrates the robustness in orientation, occlusions and recognition accuracies by learning algorithms. Table.1 summarizes some studies done in component based representation approach for face recognition. The rest of the paper is organized as follows. In Section 2, the proposed method is explained. Experimental results and performance analysis are presented in Section 3 and finally Section 4 concludes the work.

Table 1. Summarization of component based face recognition methods

Description	Component Extraction Method	Component(s) used	Component representation	Database
Template Matching [1]	Rigid	Eyes with Eyebrows, nose and mouth	Pixel representation and Laplacian	Private [6]
EBGM [3]	EBGM	N/A	Gabor wavelet Coefficients	FERET, Bochum
Component based LDA method with component Bunches [8]	Rigid	Left eye, Right eye, nose, left mouth and right mouth	Pixel representation	FERET
Component based Cascade LDA [10]	Rigid	Four overlapped regions from whole face	Pixel representation	FERET
Component based LDA Face Description [11]	Rigid	14 components across whole face	Pixel representation	MPEG-7, XM2VTS, FERET
Framework for High resolution face recognition [12]	Rigid	Left eye, Right eye, nose, mouth, forehead, skin, irregularities	Pixel representation	Private
Part based Face Recognition [13]	Rigid	10 components	LBP	Near Infrared
Component based face identification by 3DModels [15]	Reference points on 3D head models	14 learned components	Histogram equalized gray values	3D face database
Component based representation in automated face recognition [23]	Rigid	Eyebrows, Eyes, nose and mouth	MLBP	LFW, FERET

Proposed System

The proposed system is considered as a hybrid approach as it combines the Viola Jones algorithm and SURF method. In the first step, the face and the facial components are detected by Viola Jones algorithm. In the second step, features are extracted from each component through SURF descriptor. The Key points are described for facial components. The feature vectors of facial components are concatenated into one feature vector. The concatenated feature vector is matched with everyone feature vector that was stored in the database. Finally the decision is taken based on the similarity measure. The architecture of the proposed system is shown in Fig 1. The process in detail is discussed in the subsequent subsections.

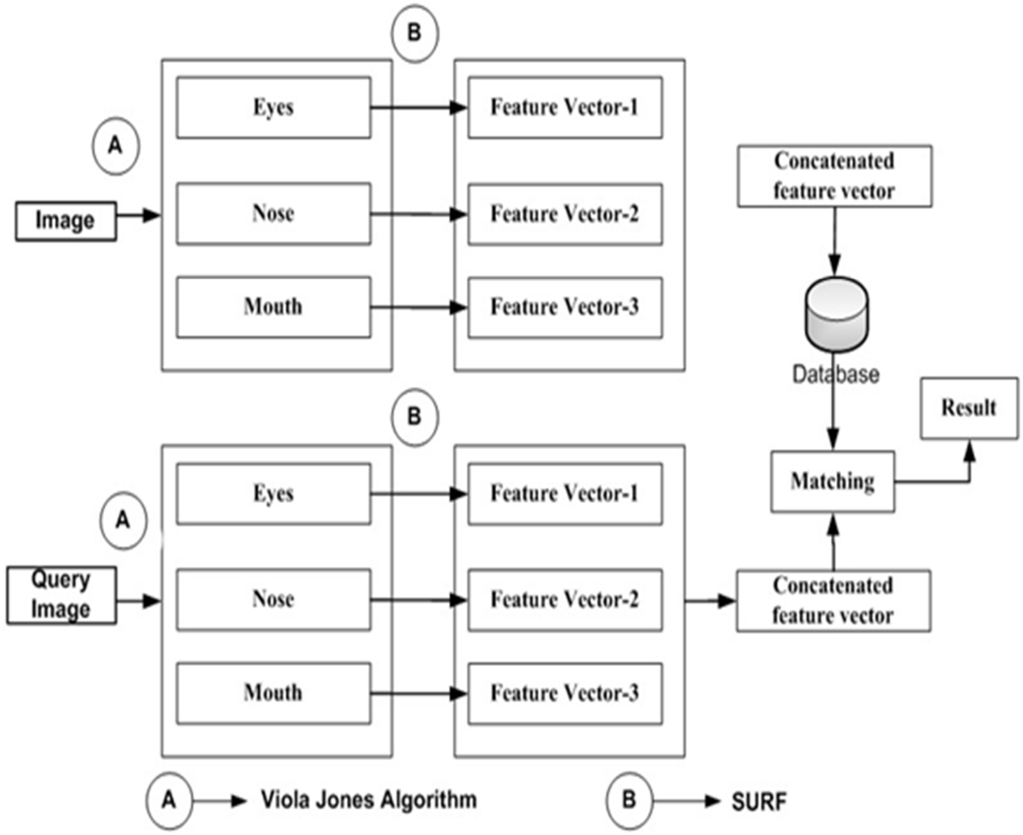

Figure 1. Architecture of the proposed method

The architecture of the proposed system is shown in Fig 1. In Fig 1.A shows the three components that are detected by Viola jones algorithm. In Fig 1.B, SURF descriptor is applied in each component to extract the feature vectors from image. Interest points are extracted from face image such as histogram equalization and normalization. SURF extracts 15 -25 fiducial points per component. Similarly for each component 15-25 fiducial points are extracted. Finally the feature vectors of every component are fused to form a single feature vector for the face image and it is matched with the feature vectors stored in the database to identify the person is recognized person or unknown person.

Component classification by Viola Jones Algorithm

The basic idea of Viola Jones algorithm is to scan a sub-window capable of detecting faces. The traditional image processing could be rescale into different sizes and then run the constant size detector through these images. It is time consuming approach. In Viola Jones algorithm rescale the detector instead of the input image and execute the detector ‘n’ number of times through the image, each time with a different size. Viola Jones algorithm is scale invariant and time consuming approach. Viola Jones algorithm is a real time facial component detection method based on simple features. They used rectangle features for computing the sum and difference of the pairs in the rectangular sub-image. The windows are scaled to detect the components from face [9]. This method is invariant to geometrical operations like rotation, scaling as well as invariant to illumination conditions. The algorithm has three major steps. They are

(i) Scale invariant detection
(ii) Feature selection using Learning Classification Function
(iii) Cascading classifiers to detect the facial component

Scale invariant detector
The first step of the Viola-Jones face detection algorithm is to convert the input image into an integral image. The conversion of input image into integral image is by making each pixel equal to the entire sum of all pixels above and the left of the concerned pixel. It computes the sum of all pixels inside any given rectangle using only four values. These values are the pixels in the integral image that coincide with the corners of the rectangle in the input image. To achieve scale invariant detection, rectangular features are used, with a new image representation defined as integral image. The integral image is defined in equation (1).

$$ii(x,y) = \sum_{x' \leq x, y' \leq y} i(x,y) \qquad (1)$$

where ii(x,y) is the integral image and i(x,y) is the original image. For example

1	1	1
1	1	1
1	1	1

Input Image

1	2	3
2	4	6
3	6	9

Integral Image

Selecting the features with Learning Classification function
The feature set can be classified using Adaboost algorithm. AdaBoost is a machine learning boosting algorithm capable of constructing a strong classifier through a weighted combination of weak classifiers. It is strong and guaranteed classified algorithm. The feature set is classified based on the weighted majority vote. A weak classifier (h(x, f, p, θ)) is defined in equation (2).

$$h(x,f,p,\theta) = \begin{cases} 1 & if\ pf(x) < p\theta \\ 0 & Otherwise \end{cases} \qquad (2)$$

where x is a sub window, f is the feature, θ is threshold and p is polarity indicating the direction of inequality.
Select the best weak classifier with respect to their weight in order to detect the strong classifier because it is time consuming process rather than selecting a strong classifier.

Cascade classifiers to detect the face and facial component
The detector scans the image several times with different size. The cascaded classifier discards the non-face region instead of determining the face. The cascaded classifier is composed of stages each containing a strong classifier. Each stage is to determine whether a given sub-window is definitely not a face or non- face region. When a sub-window is classified to be a non-face by a given stage it is immediately discarded. Conversely a sub-window classified as a non-face is passed on to the next stage in the cascade. It follows that the more stages a given sub-window passes, the higher the chance the sub-window actually contains a face. Similarly the components are also determined. This algorithm will help to reconstruct the classifier which achieves good performance and reduces the time.

Feature Extraction for component by SURF
SURF [18] is a method to extract features. The feature has 64 dimensions and it is invariant to rotation, contrast, brightness and scale. It is based on Haar wavelet and use integral images. It utilizes an integer approximation to the determinant of Hessian matrix, which can be computed extremely fast with the use of integral image. For features, it uses the sum of Haar wavelet responses around the point of interest. Feature vectors can be detected through salient points from the image. The Algorithm contains following steps.

1. Finding interest point from the image.
2. Find most important interest points in scale space by approximate Laplacian of Gaussian.
3. Find Feature description in horizontal and vertical direction
4. Generate feature vector

Interest point from the image

The interest points can be extracted through the SURF for facial component. The original image is convolved with Gaussian kernel to determine the determinant of Hessian Matrix is defined in equation (3).

$$H(x, y, \sigma) = \frac{\mathrm{Gxx}(\mathrm{x}, \mathrm{y}, \sigma).\mathrm{Gyy}(\mathrm{x}, \mathrm{y}, \sigma) - Gxy(x, y, \sigma)^2}{\sigma^2} \tag{3}$$

where Gxx(x,y,σ) is defined in equation (4).

$$Gxx(x, y, \sigma) = \frac{\partial N(0, \sigma)^2}{\partial i . \partial j} * image(x, y) \tag{4}$$

The local maxima filter occurs in facial component with gradient in multiple directions. The integral image location x = (x, y)^T contains the aggregate sum of pixels within the bounded (rectangular) region and opposite corners from the current position of the pixels. The integral image is defined in equation (5).

$$I(x) = \sum_{i=0}^{i \le x} \sum_{j=0}^{j \le y} I(x, y) \tag{5}$$

The interest points are localized in scale space. Once the integral image is computed, it takes three additions to calculate the sum of intensities over rectangular area because the calculation time is independent of its size.

Scale Space Representation

Using scale space value, the feature vectors are generated from the most significant point from the image. These points can be calculated rapidly through the integral image. Due to use of box filter and integral images, the filter can be any size at exactly the speed on the image. Hence the scale space is analyzed by up-scaling the filter size rather than iteratively reducing the image size. The resultant of 9 x 9 filter, is considered as the initial scene layer, to which we will refer as scale s = 1.2 (corresponding to Gaussian derivatives with σ = 1.2). The layers are attained by filtering the image with steadily bigger mask. At larger scales, the step between consecutive filter sizes should also scale accordingly. The motivation for this sampling is computational efficiency.

The scale space is divided into octaves. An octave characterises of filter response maps achieved by convolving the original input image with a filter of increasing size. Each octave is subdivided into a constant number of scale levels because of the discrete nature of integral images; the minimum scale difference between 2 subsequent scales depends on the length l0 of the positive or negative lobes of the partial second order derivative in the direction of derivation (x or y). To localize the most interest point in the original image over scales, non-maximum suppression in a 3 X 3 X 3 neighbourhood is applied. The maxima of the determinant of the Hessian matrix are then interpolated in scale and image space [25].

Feature Description

SURF describes the features by sum up the pixel information of neighbors. First, it determines the orientation of each feature by convolving in its neighborhoods with horizontal and vertical Haar wavelet filter. The filter is shown in Fig 2. These filters help to determine the directional derivatives of the image's intensity. From the edge responses, the descriptor describes the specific orientation of images. The rotation invariant property allows identifying the descriptor accurately than existing methods.

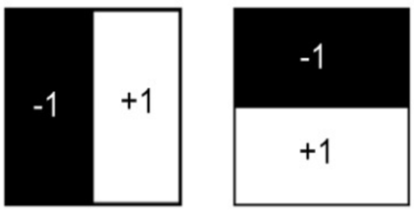

Figure 2. Haar Wavelet Filters

The edge gradient responses are taken from dark versus light, therefore the responses to generate a feature vector representing each component feature and its neighborhood. This property allows improving the recognition rate in illumination conditions. So the neighborhood is divided into smaller regions randomly. From each region, the features are extracted and it is summed up to form single feature vector.

Feature Vector

SURF describes the feature vector of image based on sum of Harr coefficients. Construct a 4 x 4 rectangular region that preserves the information. Apply Haar wavelet transform in scale space to extract the facial component descriptor in both directions (i.e. horizontal and vertical direction). The sum of Haar responses dx and dy for each region are generated and considered as a feature vector. The Sum of pixels for each region feature vector is denoted as |dx| and |dy|. Concatenate the Haar response vector and sum of pixel vector into a single feature vector is defined in equation (6).

$$V = \left(\sum d_x \sum d_y \sum |d_x| \sum |d_y|\right) \tag{6}$$

Feature Matching

Face Recognition determine a database features for each image in the standard database. When given a query image at runtime, the system generates the features of query image and its compare the features within the database using distance metrics. If the features are similar then the system concludes that the query image is present in the database. In component based representation the features are extracted from each component. The extracted features of component are concatenated into a feature vector for face. The features are stored in the database. Given a query image, the features are generated based on the process of extraction. Euclidean distance measure is used for determining the distance between features of query image and features of database images. If the distance between the query image features and the database image features is very close then the person is considered as matched. If there are "n" persons are identified as closer, then it is difficult to identify the exact person. In this case Euclidean distance gives ambiguous results. To overcome such ambiguities, the ratio of the number of features pairs found to the sum of the squared distance is employed for accurate detection. This ratio is defined in equation (7).

$$S_R = \frac{N_i}{\sum_{j=1}^{j=N_i} d_{ij}^2} \tag{7}$$

where Ni is no of feature pairs identified and d^2_{ij} is sum of squared distance.

Experimental Results and Performance Analysis

The experimental results show the effectiveness of component processing in automated face recognition. In this work the face images are detected in two ways. First, the global approach which represents the feature vector from whole face and second, the component approach which represents the separate feature vector for each facial component. The current research is mostly based on the global approach in face recognition. If the face image size increased in global approach then the accuracy is low comparatively with component approach. The proposed system used component based representation for detecting the components. The facial component is considered as a block and the features are extracted from each block. In Fig 3, the first row represents some sample images taken from FERET [4] database. The second row represents the face detected using Viola Jones algorithm. The third, fourth and fifth rows represent the components extracted from each face. The proposed method is implemented on benchmark datasets like FERET and ORL database.

Table 2. Recognition rate on FERET Database

Method	FERET 1a	FERET 1b
PCA	0.83	0.85
LDA	0.89	0.96
SIFT	0.93	0.94
PFD-SIFT	0.97	0.98
RS-LDA	0.99	0.99
Proposed Method	0.99	0.99

Analysis on FERET database

The FERET database [4] consists of 2388 images and the dataset comprised into 1194 persons. Select 2 images from Fa and Fb per person. Faces are cropped by viola jones algorithm and reshape the image size into 100 X100. From the dataset 250 people images are selected randomly for training and rest of the images are used for testing. (FERET 1a) [24] In FERET 1b, 497 images are randomly selected and it is used for training and the rest of the images are used for training. The proposed method applied to the database to estimate the performance. Table 2 and Fig. 4 show the recognition rate of holistic and component based approaches compared with the proposed system. From Table 3, it is observed that RS-LDA and the proposed method have similar recognition rates. However, the proposed method has lower running time than RS-LDA. Fig.5

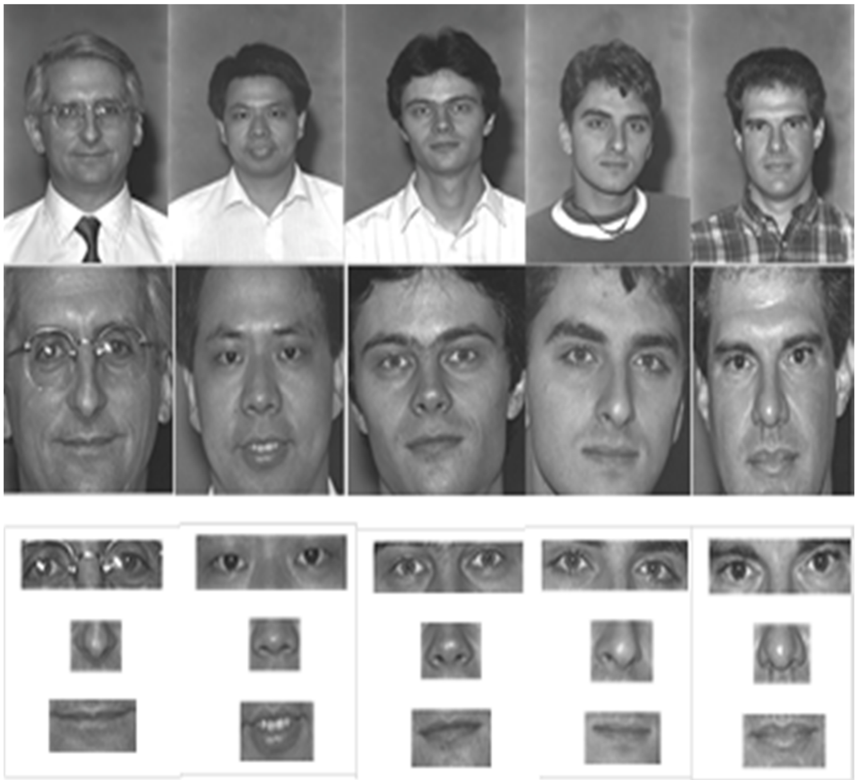

Figure 3. Sample images of FERET database

and Table 3 shows the comparative analysis of the component based method (RS-LDA). In the case of individual components, both the methods have similar recognition rate for nose component and slightly better in rest of the components.

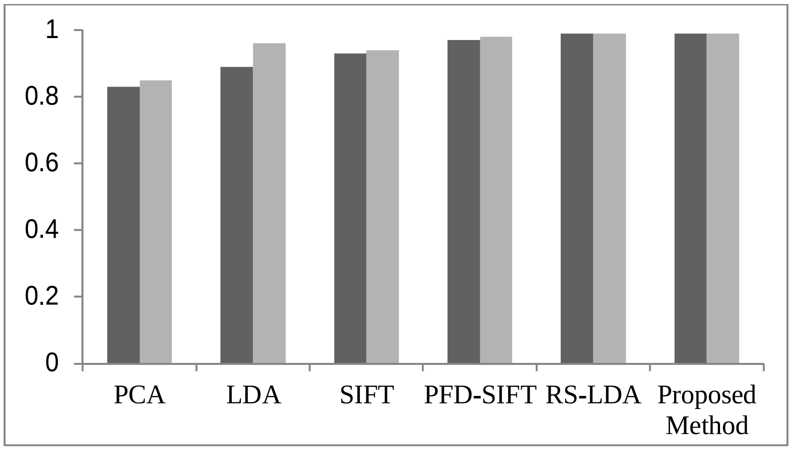

Figure 4. Rank 1 Recognition rate on FERET Database

Table 3. Performance of Each Component on FERET database

Component	RS-LDA	Proposed Method
Eyes	0.86	0.87
Nose	0.99	0.99
Mouth	0.79	0.80

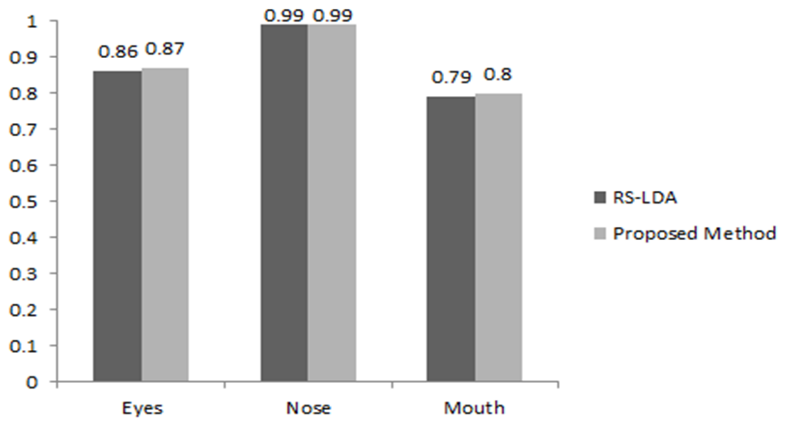

Figure 5. Performance of Each Component on FERET database

Analysis on ORL database

The ORL database contains 40 different persons image. Each subject 10 images with different variation (varying the lighting, facial expressions and facial details). Each subject five images are randomly selected for training and rest 5 images for testing. In ORL database 200 images are used as gallery images and remaining 200 images are used as probe set. The recognition rate of the proposed system is compared with existing holistic approaches and is shown in Fig.6 and Table 4.

Table 4. Percentage of Recognition rate on ORL database

Method	Recognition Rate
PCA	0.85
LDA	0.92
SIFT	0.90
PFDM-SIFT	0.99
Proposed Method	0.99

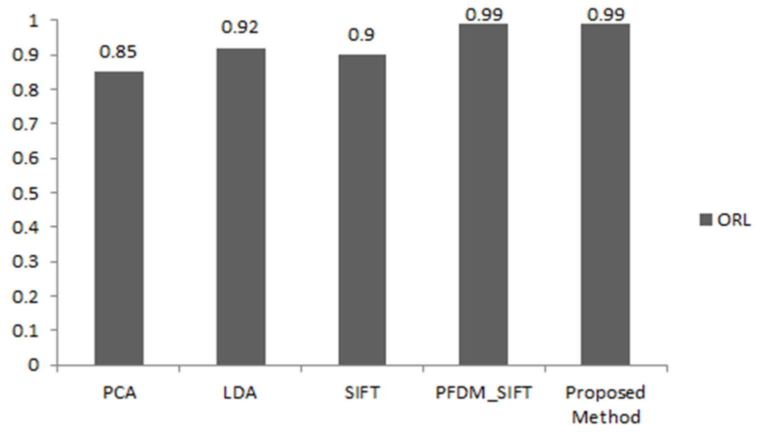

Figure 6. Percentage of Recognition rate on ORL database

Conclusion

This work presents a hybrid approach for automated face recognition based on the facial components. The hybrid approach demonstrates the efficiency in pose variation and illumination condition. The experimental results show that the proposed method provides good results comparatively with holistic approaches. The system slightly improves the recognition rate in component. The proposed method reduces the running time of system.

References

[1] Brunelli and T. Poggio, “Face recognition: Features versus templates,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 15, no. 10, pp. 1042–1053, 1993.
[2] P. Belhumeur, J. Hespanha, and D. Kriegman, “Eigenfaces vs. fisherfaces: Recognition using class specific linear projection,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 19, pp. 711–720, 1997.
[3] L. Wiskott, J.-M. Fellous, N. Kruger, and C. von der Malsburg, “Face recognition by elastic bunch graph matching,” In Proc. IEEE Int. Conference on Image Processing, vol. 1, 1997.
[4] P. J. Phillips, H. Moon, P. J. Rauss, and S. Rizvi, “The feret evaluation methodology for face recognition algorithms,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 10, 2000.
[5] N. Sagiv and S. Bentin, “Structural encoding of human and schematic faces: Holistic and part-based processes,” Journal of Cognitive Neuroscience, vol. 13, no. 7, pp. 937–951, 2001.
[6] H. Jie, H. Yu, and J. Yang, “A direct lda algorithm for high-dimensional data with application to face recognition,” Pattern Recognition, vol. 34, pp. 2067–2070, 2001.
[7] T. Ojala, M. Pietikainen, and T. Maenpaa, “Multi resolution gray-scale and rotation invariant texture classification with local binary patterns,” IEEE Trans. Pattern Analysis & Machine Intelligence, vol. 24, no. 7, pp. 971–987, 2002.
[8] J. Huang, P. Yuen, W.-S. Chen, and J. Lai, “Component-based lda method for face recognition with one training sample,” IEEE International Workshop on Analysis and Modeling of Faces and Gestures, pp.120–126, 2003.
[9] Paul Viola and Michael J.Jones ,” Robust Real Time Face Detection”, International journal of Computer vision.57(2), pp. 137-154,2004.
[10] W. Zhang, S. Shan, W. Gao, Y. Chang, and B. Cao, “Component-based cascade linear discriminant analysis for face recognition,” Advances in Biometric Person Authentication Lecture Notes in Computer Science, vol. 3338, pp. 19–79, 2005.
[11] T.-K. Kim, H. Kim, W. Hwang, and J. Kittler, “Component-based lda face description for image retrieval and mpeg-7 standardisation,” Image and Vision Computing, vol. 23, pp. 631–642, 2005.
[12] D. Lin and X. Tang, “Recognize high resolution faces: From macrocosm to microcosm,” Computer Vision and Pattern Recognition, pp. 1355– 1362, 2006.
[13] K. Pan, S. Liao, S. Zhang, S. Z. Li, and P. Zhang, “Part-based face recognition using near infrared images,” IEEE conference on Computer Vision and Pattern Recognition, 2007.
[14] Schwaninger, S. Schumacher, C. Wallraven, and H. Buelthoff, “Using 3d computer graphics for perception: The role of local and global information in face processing,” Applied Perception in Graphics and Visualization, 2007.
[15] Heisele, T. Serre, and T. Poggio, “A component-based framework for face detection and identification,” International Journal of Computer Vision, vol. 74, no. 2, 2007.
[16] E. Meyers and L. Wolf, “Using biologically inspired features for face processing,” Int. Journal of Computer Vision, vol. 76, no. 1, pp. 93–104, 2008.
[17] X. Jiang, B. Mandal, and A. Kot, “Eigenfeature regularization and extraction in face recognition,” Pattern Analysis and Machine Intelligence, IEE Transactions on, vol. 30, no. 3, pp. 383–394, 2008.
[18] C. Evans, “Notes on the opensurf library,” University of Bristol, Tech. Rep. CSTR-09-001, January 2009. [Online]. Available: http://www.chrisevansdev.com
[19] S. Xie, S. Shan, X. Chen, and J. Chen, “Fusing local patterns of gabor magnitude and phase for face recognition,” IEEE Transactions on Image Processing, vol. 19, 2010.
[20] B. Klare and A. Jain, “On taxonomy of facial features,” in Proc. of IEEE Conference on Biometrics: Theory, Applications and Systems, 2010.
[21] B. Klare, Z. Li, and A. Jain, “Matching forensic sketches to mugshot photos,” IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 33, no. 3, pp. 639–646, 2011.
[22] J. Richler, O. Cheung, and I. Gauthier, “Holistic processing predicts face recognition,” Psychological Science, vol. 22, no. 4, pp. 464–471, 2011.
[23] Kathryn Bonnen, Brendan Klare, and Anil Jain,” Component-Based Representation in Automated Face Recognition,” IEEE Transactions on Information Forensics and Security, 2013.
[24] C. Geng and X. Jiang, “Face recognition based on the multi-scale local image structures”, Pattern Recognition, vol. 44, pp. 2565–2575, 2011.
[25] H. Bay, A. Ess, T. Tuytelaars, L. Van Gool, L. Van Gool, H. Baya, A. Essa, T. Tuytelaarsb, and L. Van Goola, “Speeded-up robust features (SURF),” Comput. Vis. image Underst., vol. 110, no. 3, pp. 346–359, 2008.