Local Feature Based Automatic Face Recognition System

***S.A.Pardeshi and S.N. Talbar*[1]**

Rajarambapu Institute of Technology, Sangli
[1]S.G.G.S.C.O.E & T, Nanded
E-mail : sapardeshi @rediffmail.com ; sntalbar@ yahoo.com

ABSTRACT

We proposes automatic face recognition system based on multi-scale Harris-Laplace detector, robust to the illumination, orientation, scale and translation variations. A face region is separated from given image by face localization algorithm. A separated face region is represented by a set of interest points, detected by Harris-Laplace detector, to achieve scale invariance. Each interest point is described by a local feature vector, based on the Gabor filter, extracted at various scales and orientations to achieve rotation invariance. Similarity between two faces is measured by voting algorithm. Performance of voting algorithm is also checked for various distance metrics. Comparison of results with existing algorithms validates the usefulness of local invariant features for face recognition at less computational cost and less storage requirements. Experimental results show that voting algorithm with cosine distance metric gives best recognition accuracy.

Keywords: *face localization, interest point detection, interest point selection, local feature descriptor, voting algorithm.*

I. INTRODUCTION

In recent years, face recognition has received substantial attention from both research communities and the market, but still remains very challenging problem for real time applications. A large number of face recognition algorithms, along with their modifications, have been developed during the past decades which can be generally classified into two categories: holistic approaches and local feature based approaches.

Holistic approaches uses whole face region as the input to a recognition system while local feature based methods, first locate several facial features, and then classify the faces by combining and comparing the corresponding local statistics. Influential work is given by Wiskott et al. [1], called Elastic Bunch Graph Matching. The elastic bunch graph is a graph-based face model with a set of jets attached to each node of the graph. The algorithm recognizes new faces by first locating a set of facial features to build a graph, which is then used to compute the similarity of both jets and topography. Local binary pattern was originally designed for texture classification and was used for face recognition in [2]. The face area was divided into small windows of size 7 × 7. Each window is represented by local binary pattern. The chi square statistic and the weighted chi square statistic were adopted for comparison and recognition. Zhang *et a*l, [3] proposed local Gabor binary pattern histogram sequence (LGBPHS) by combining Gabor filters and the local binary operator. The face image was first filtered with Gabor filters at five scales and eight orientations. The LBP operator was then applied to all 40 Gabor magnitude pictures to generate the LGBPHS. The recognition was done using histogram intersection and weighted histogram intersection. Luo et al. [4] investigated the use of SIFT (scale invariant feature transform) for face recognition. The SIFT method first detects interest points at different resolutions and uses scale and rotation invariant descriptor to represent the interest points. They also compared SIFT approach with EBGM and Local binary pattern approach for expression variations and illumination variations. We used the same comparison to check the effectiveness of proposed method.

Our approach is similar to SIFT approach. Unlike the SIFT approach, which uses scale-space Difference-Of-Gaussian to detect interest points in images, we used multi-scale Harris-Laplace detector [5] to detect interest points and detected interest points are represented by Gabor descriptor. In general, the following steps are proposed for face recognition:

(a) Face localization used to separate face region from given image.

(b) Detection of interest points on separated face region invariant to scale changes.

(c) Selection of the interest points and extraction of local feature vector.

(d) Reference database formation based on feature vectors extracted from model images.

(e) Voting algorithm based classification of query image using various distance metrics.

Details of the above steps are given in section 2. Comparative results with state-of-art methods are presented in section 3. The concluding remarks are given in section 4.

2. PROPOSED METHOD

2.1 Face localization

Interest points can be considered as pixels in an image which "stand out" from other pixels so that it can be used to describe image content. For face images having inter-class and intra-class variation along with background clutter, interest points are also get detected on background clutter. In order to eliminate background clutter, we perform face localization, to locate and separate face region from input image. We developed skin color based face localization algorithm based on YC_bC_r space. But performance degrades under varying lighting conditions. To investigate the reason, we manually extracted face skin samples from 392 facial images taken under normal and varying lighting conditions and examined the distribution of R, G and B color components. We found that, under normal lighting conditions distribution of R, G and B color components is confined to very narrow ranges as compared to distribution under varying lighting conditions. The distribution of R, G, and B color components is get affected under varying lighting conditions due to non-linear mapping of individual color components and image dependency on lighting geometry and illumination color. Nonlinear mapping of R, G, and B color components is corrected by linear stretching of R, G and B color components. The equations used for linear stretching of R, G and B color components, based on experimental investigation carried out, are as in (1) to (3). R, G and B are color components of original image. $\tilde{R}$, $\tilde{G}$ and $\tilde{B}$ are normalized color components.

$$\tilde{R} = \begin{cases} 75 & \text{if } R \le 45 \\ 150 & \text{if } R = 255 \\ \left(\dfrac{150-75}{255-45}\right) R + 50 & \text{if } 45 < R < 255 \end{cases} \quad \text{................(1)}$$

$$\tilde{G} = \begin{cases} 45 & \text{if } G \le 5 \\ 125 & \text{if } G \ge 175 \\ \left(\dfrac{125-45}{175-5}\right) G + 40 & \text{if } 5 < G < 175 \end{cases} \quad \text{.............(2)}$$

$$\tilde{B} = \begin{cases} 25 & \text{if } B = 0 \\ 100 & \text{if } B \ge 175 \\ \left(\dfrac{100-25}{175}\right) B + 25 & \text{if } 5 < B < 175 \end{cases} \quad \text{................(3)}$$

Image dependency on lighting geometry and illumination color is corrected by method given in [6]. The corrected color image is converted to YC_bC_r space and binary mask is generated based on values of C_b and C_r of every pixel as in equation (4). The values of C_b and C_r are obtained by collecting the skin samples of YC_bC_r images and identifying the ranges of C_b and C_r based on collected samples.

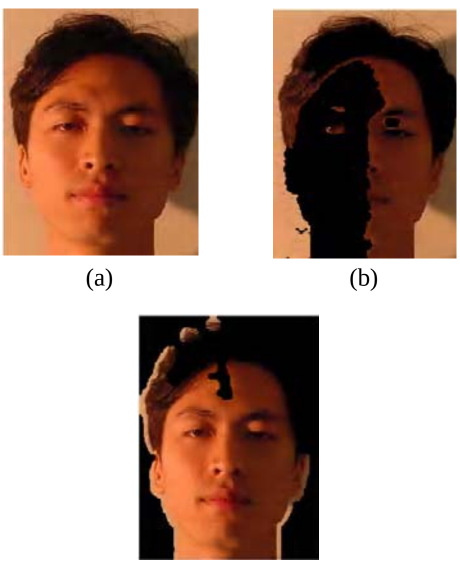
$$map(i,j) = \begin{cases} 1 & \text{if } 110 \le C_b(i,j) \le 140 \text{ and } 110 \le C_r(i,j) \le 140 \\ 0 & \text{otherwise} \end{cases} \quad \text{....(4)}$$

$map(i, j)$ = binary value of pixel with spatial location (i,j)

$c_b(i, j) = c_b$ value of pixel with spatial location (i,j)

$c_r(i, j) = c_r$ value of pixel with spatial location (i,j)

The morphological operations (erosion followed by closing and dilation followed by closing) are performed on generated binary mask. Erosion is used to remove small objects in the background area. Dilation is used to fill small hole in facial areas such as eyes and mouth. The obtained binary mask is used to remove background clutter and localization of face region based on skin color. The result of face localization algorithm is shown in Fig.1 (a) to Fig1 (c), with black pixels indicates non-skin pixels. Figure 1(a) shows original image with illumination direction as left, Fig. 1 (b) shows result of face localization in YC_bC_r space without any normalization and Fig. 1(c) shows result of face localization in YC_bC_r space with proposed method.



(a) (b) (c)

Fig. 1: (a) original image, (b) Face localization without normalization, (c) Face localization with proposed method

2.2 Interest point detection and selection

The evaluation of interest point detector, as presented in [7], demonstrate the excellent performance of Harris detector but it is not invariant to scale changes. Hence we used the Harris-Laplace detector, as proposed by [5], to obtain scale invariant detector. The Harris-Laplace detector uses scale-adapted Harris function, as in equation (5), to localize corner points in scale space.

$$N(X,\sigma_I) = \sigma_D^2\, g(\sigma_I) \otimes \begin{bmatrix} L_x^2(X,\sigma_D) & L_xL_y(X,\sigma_D) \\ L_xL_y(X,\sigma_D) & L_y^2(X,\sigma_D) \end{bmatrix} \quad \text{...........(5)}$$

where σ_I is integration scale, σ_D is differentiation scale and L_a is the derivative computed in the '*a*' direction. The measure of corner response at a point X and scale σ_I is given by equation (6) where λ is constant.

$$R(X,\sigma_I) = \det(N(X,\sigma_I)) - \lambda\, \mathrm{tr}^2(N(X,\sigma_I)) \quad \text{............(6)}$$

The point is selected as a corner point if it satisfies the condition given in equation (7) i.e. corner response of the point must be positive and maximum in the neighborhood of 3 × 3 windows.

$$\left.\begin{array}{l} R(X,\sigma_I)\ \mathrm{f}\ 0 \text{ and} \\ R(X,\sigma_I)\ \mathrm{f}\ R(X_W,\sigma_I)\ \forall X_W \in W \end{array}\right\} \quad \text{...........(7)}$$

w = window of size 3×3

We use gray scale image of original RGB image to build the scale-space representation by using equation (5). The pre-selected scales are used with $\sigma_n = k^n \sigma_o$; σ_o is the initial scale factor set to 1; factor k is scale factor between successive levels (set to 1.4 as proposed in [8], *n* gives number of resolution levels. The matrix $N(X, \sigma_I)$ is computed with $\sigma_I = \sigma_n$ and $\sigma_D = s\sigma_n$, where $s \in [0.7, 0.8, ..., 1.4]$. Large scale change of 1.4 is used to detect initial interest points with k=1.4 and n=1.Small scale changes, specified by s, are used to get better accuracy for the location of *X* as in [5]. The detected interest points are stored with their spatial locations and associated scales i.e. the scales at which the interest point is detected.

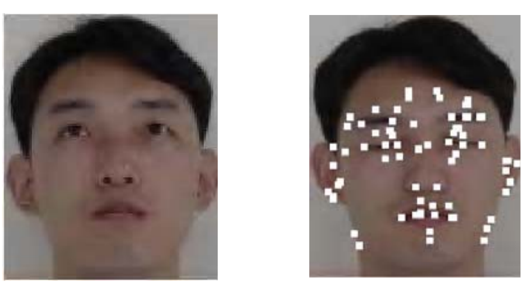
Based on detected interest points, we note the following important facts: 1) Number of interest points decreases as scale increases. 2) Interest point location varies slightly over scales. The higher the scale level, the bigger the possible range of point locations.

The spatial location of the interest point can be detected accurately under image rescaling by verifying whether LoG (Laplacian-of-Gaussian) response of the interest point at certain scale attains a maximum i.e. LoG response is lower for finer and coarser scales. Since the detected interest points are stored with their spatial coordinates and associated scales, we check whether the LoG response of the interest point attains maximum at associated scale. If LoG response attains maximum at associated scale, interest point is retained else rejected. The LoG response is calculated using equation (8).

$$\left|LOG(X,\sigma_n)\right| = \sigma_n^{\ 2}\left|L_{xx}(X,\sigma_n) + L_{yy}(X,\sigma_n)\right| \text{........(8)}$$

where $\sigma_n = set\ of$ scales

The number of interest points detected is very large. We need to select the optimal number of interest points so that computation cost and storage requirement will get reduced without affecting recognition accuracy. We sorted the interest points detected by Harris-Laplace detector in descending order of their corner responses. The optimal number of interest points required is decided by carrying out experiments with different number of interest points and examining resulting recognition accuracy. Fig. 2(a) shows the original image and Fig. 2 (b) shows the 75 selected interest points, based on corner response, superimposed on the original image. In order to highlight the points, 3 × 3 neighborhood of point is used.



(a) (b)

Fig. 2: (a) original image, (b) selected interest points

2.3 Interest point descriptor

Invariance properties of Gabor filter based features were extensively studied in [9]. They proved that Gabor features provides significant degree of robustness to photometric disturbances and natural image variations. Hence we used 2D Gabor filters to detect local structure of the image centered at detected interest point. Gabor filters enhance the low level image features such as the peaks, valleys and ridges so that the eyes, the nose and the mouth, as well as the other salient local features like dimples are get enhanced. These key features can be used for the discrimination of different faces. A family of complex Gabor filters is defined as in equation (9):

$$W(x,y) = 1^{-\frac{x'^2+\gamma^2 y'^2}{2\sigma^2}} \cos\left(2\Pi\frac{x'}{\lambda} + \varphi\right) \text{ .... (9)}$$

where $x' = x\cos\theta + y\sin\theta$ and $y' = -x\sin\theta + y\cos\theta$

We designed Gabor filter bank with 4-scales (λ) and 4-orientations (θ) for feature extraction. This result in 4*4=16 Gabor filters altogether. The values of λ and θ used are given in equation (10).

$$\lambda \in \left(4, 4\sqrt{2}, 8, 8\sqrt{2}\right) \ and \quad \theta \in \left(0, \frac{\Pi}{4}, \frac{\Pi}{2}, \frac{3\Pi}{4}\right) \text{ ....(10)}$$

Feature vectors are generated at interest points as composition of Gabor filter responses. Local image patch of size 25*25, centered at interest point, is extracted. Extracted local image patch is convolved with Gabor filters. Since the phase information of the Gabor filter is time-varying, only magnitude values are used for feature description. K^{th} feature vector of i^{th} reference face is defined as in equation (11).

$$v_{i,k} = \left\{x_k, y_k, R_{i,j}(x_k, y_k);\ j = 1,....,16\right\} \text{..............(11)}$$

where x_k, y_k = spatial co-ordinates of k $^{\text{th}}$ interest point

$R_{i,j}(x_k, y_k)$ = j $^{\text{th}}$ Gabor filter response at (x_k, y_k)

2.4 Matching two faces

To match unknown face, we need reference database. The reference database is used to identify the most similar model image to unknown image. The reference database is developed as follows: several model face images are collected with pose variations, illumination variations and expression variations for each subject. Face localization is performed on model images and interest points are detected by Harris-Laplace detector. The detected interest points are sorted in descending order of corner response and required numbers of interest points are selected. Gabor feature vectors are extracted from these interest points. Extracted Gabor feature vectors, for each of these model images are stored in a reference database, along with a pointer to the model image from which they originate. Thus Reference database consist a set $\{M_k\}$ of models. Each model M_k is defined by set of Gabor feature vectors $\{V_j\}$ extracted from interest points of model images. During storage process, each V_j is added to the reference database with a link to the model k for which it has been computed. So the reference database is table of couples $\{V_j, k\}$.

Matching of unknown face consist of finding the model $M_{\hat{k}}$ which is most similar to unknown image. The most similar model is selected by performing face localization on unknown image followed by detection and selection of interest points by Harris-Laplace detector. Then set of Gabor feature vectors $\{V_l\}$ is extracted from detected interest points. These vectors are compared with each of the vector V_j in reference database for similarity by using three distance metrics i.e. city-block distance i.e. $d_c(x, y)$, squared Euclidean distance i.e. $d_e(x, y)$ and cosine of the angle between feature vectors i.e. $d_\theta(x, y)$. Three distance metrics are as in equations (12 to 14).

$$d_c(x,y) = |x-y| = \sum_{i=1}^{k} |x_i - y_i| \text{.......(12)}$$

$$d_e(x,y) = \|x-y\|^2 = \sum_{i=1}^{k} \left(x_i - y_i\right)^2 \text{..........(13)}$$

$$d_\theta(x,y) = -\frac{x.y}{\|x\|\|y\|} = -\frac{\sum_{i=1}^{k} x_i y_i}{\sum_{i=1}^{k} x_i^2 \sum_{i=1}^{k} y_i^2} \text{........(14)}$$

The most similar V_j for each V_l is identified and vote is given to the corresponding model. We sum number of votes received by each model. This sum is stored in the vector $T(k)$. The model received more votes is considered as best match: the unknown image represents the model $M_{\hat{k}}$ for which $\hat{k} = \arg\max_k T(k)$.

3. EXPERIMENTAL RESULTS

Experiments are carried out on Asian face database. It contains true-color face images of 103 people, 53 men and 50 women, representing 17 various images (1 normal face, 4 illumination variations, 8 pose variations and 4 expression variations) per person. As described in section 2, interest points are extracted, selected and described by magnitudes of Gabor filters.

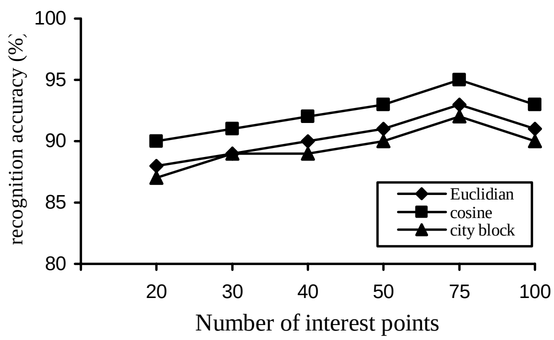
We checked the recognition accuracy of our algorithm; against facial expression variations (EV), illumination variations (IV) and pose variations (PV), as mentioned in [4]. Normal face image and 50% images of each variation are used to create reference database. Before performing the experiments, all images are cropped to 112x92 pixels images but no any preprocessing is performed based on manually selected eye positions as in [4]. TABLE I shows the comparison of percentage of recognition accuracy achieved by proposed method and results reported in [4] for various variations. Reported results are available for EV and IV only. No any results are available for experiment PV. For algorithms marked with '*', results were reported on FERET database.

	EV	IV	PV
EBGM Optimal*	90	42	--
LBP*	97	79	--
SIFT_GRID*	94	35	--
Person specific SIFT*	97	47	--
Our method	100	80	100

Table 1: Comparison of Experimental results

Results displayed in Table 1 shows that proposed method provides 100 % invariance against expression and pose variations. The significant increase in recognition accuracy for illumination variation is also achieved as compared to state-of-art methods. Numbers of interest points used for recognition are also very less i.e. 75 and each interest point is represented by Gabor feature vector of size 1*16 requires very less storage memory. Matching of the faces is performed by simple comparison by Euclidian distance metric results in low computational cost and fast execution speed as well

Comparison of recognition accuracy against number of interest points used along with various distance measures is shown in Fig. 3. Graph shows that recognition accuracy increases linearly with number of interest points up to certain limit but then it starts decreasing. Hence optimal number of interest points required to get better recognition accuracy is 75. Cosine of the angle between feature vectors is better choice for measuring feature vector similarity



as compared to squared Euclidian distance and city block distance measure.

Fig. 3: Comparison of recognition accuracy against number of interest points

4. CONCLUSION

We presented promising capability of invariant local features for face recognition. Algorithm uses multi-scale Harris-Laplace detector to extract a set of scale invariant interest points from face images. Rotation invariance and invariance to photometric disturbances is achieved by extracting local information, centered at interest point, by set of Gabor filters. Similarity of Gabor feature vectors, extracted from query image and reference database, is checked by three different distance metrics. Similarities of Gabor feature vectors are used by voting algorithm to match the model images with query image. Several experiments on the Asian face dataset validate the robustness of our technique against variations in facial expressions, pose and illumination. Comparison with state-of-art methods shows the substantial increase in face recognition accuracy at less computational cost, less storage memory requirement.

In future, we will try to inculcate the scale selected by LoG in the design of Gabor filter so that optimum number of Gabor filters can be used for extraction of interest point local information. It will further reduce the computation cost but resulting recognition accuracy is required to be studied.

REFERENCES

1. Wiskott L, "Face Recognition by Elastic Bunch Graph Matching," *IEEE Trans.PAMI*, Vol. 19(7), pp. 775-779, 1997.
2. Ahonen T, "Face recognition with local binary patterns," *ECCV*, 2004, pp. 469–481.
3. Zhang W, "Local Gabor binary pattern histogram sequence (LGBPHS): A novel non-statistical model for face representation and recognition," *ICCV*, 2005, pp. 786–791.
4. Luo et al," Person-specific sift features for face recognition," *ICASSP*, vol.2, 2007, pp. II-593-II-596.
5. Mikolajczyk K., "Scale & affine invariant interest point detectors", *Computer Vision*, Vol. 60(1), pp. 63–86, 2004.
6. Finlayson G," Comprehensive color image normalization," *ECCV*, vol.1, 1998, pp. 475-490.
7. Schmid, C., "Evaluation of interest point detectors," *Computer Vision*, vol. 37(2), pp.151–172, 2000.
8. Lowe, D.G. "Object recognition from local scale-invariant features," *ICCV*, 1999, pp. 1150–1157.
9. Kamarainen J.K. and Kyrki V., "Invariance Properties of Gabor Filter-Based Features—Overview and Applications," *IEEE transactions on image processing*, VOL. 15(5), pp.1088-1099, 2006.
10. Sang N.,"Local Gabor Fisher classifier for face recognition", *ICIG*, vol. 2, 2007, pp. 620 – 626.