Journal of Computational Intelligence in Bioinformatics.
ISSN: 0973-385X Volume 1 Number 1 (2008) pp. 73–81

http://www.ripublication.com/jcib.htm

Face Recognition Using Local Invariant Features

Sanjay A. Pardeshi

Rajarambapu Institute of Technology,Rajaramnagar
Sangli (M.S.), India
E-mail: sapardeshi@rediffmail.com

S.N. Talbar

S.G.G.S.C.O.E.T, Nanded
Nanded (M.S.), India
E-mail: sntalbar@yahoo.com

Abstract

We propose automatic face recognition system based on multi-scale Harris corner detector, robust to the geometrical distortions of the images and to the changes of lighting conditions. A face region is segmented from given image using face segmentation algorithm. Segmented face region is represented by a set of interest points, detected by multi-scale Harris corner detector to achieve scale invariance. Each interest point is represented by a local feature vector, based on the Gabor filter, extracted at several scales and orientations to achieve rotation invariance. The illumination normalization is achieved by normalization of extracted Gabor feature vectors itself. The subspace principal component analysis method is used for further dimensionality reduction and variance maximization. The similarity between two faces is a measured by 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.

Index terms: face localization, interest point detection, interest point selection, local feature descriptor, face recognition.

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 (EBGM). The elastic bunch graph is a graph-based face model with a set of jets (Gabor wavelet components) attached to each node of the graph. The algorithm recognizes new faces by first locating a set of facial features (graph nodes) to build a graph, which is then used to compute the similarity of both jets and topography.Local binary pattern (LBP) was originally designed for texture classification and was introduced in face recognition in [2]. The face area was divided into small windows of size 7×7. Each window is represented by local binary pattern histogram. 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 [2]. 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.

Our approach is also similar to SIFT approach. Unlike the SIFT approach, which uses scale-space Difference-Of-Gaussian (DOG) to detect interest points in images, we used multi-scale Harris corner detector [5] to detect interest points. Interest points, detected with multi-scale Harris detector and represented by local descriptors, have been used successfully for applications like stereo reconstruction, image indexing and object recognition. However to the best of our knowledge, this is the first attempt to use interest points detected by multi-scale Harris corner detector and represented by Gabor feature vector, for face recognition application.

In general, the following steps are proposed for face recognition:

1. Face segmentation used to separate face region from given image.
2. Detection of interest points on separated face region invariant to scale changes.
3. Selection of the interest points and extraction of local feature vector.
4. Normalization of local feature vector against geometric and photometric transformations.
5. Training based on eigen-space approach.
6. Classification based on various similarity measures.

Detailed description of the above steps is given in sections 2. Experimental results are presented in section 3. The concluding remarks are given in section 4.

Proposed Method

Face segmentation

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 segmentation, to locate and separate face region (face region from forehead to chin and cheek to cheek) from input image, having only one subject in the foreground. We developed skin color based face segmentation algorithm based on YC_bC_r space. The original color images are converted to YC_bC_r space without any pre-processing. The C_b and C_r values of each pixel are compared against the range of C_b and C_r values to check whether the pixel is skin pixel or non-skin pixel. The range of The C_b and C_r values is obtained from a training set of nearly 4000 color skin samples that were manually extracted from set of color images in YC_bC_r space. But performance is very poor under varying lighting conditions.

To investigate the reason, we again manually extracted face skin samples from 392 facial images taken under normal and varying lighting conditions in RGB space and examined the distributions of R, G and B color components. We found that, under normal lighting conditions distributions of R, G and B color components is confined to very narrow ranges as compared to their distributions under varying lighting conditions. The distributions of R, G, and B color components is get stretched 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.

$$\tilde{R} = \begin{cases} 75 & \text{if } R \leq 45 \\ 150 & \text{if } R = 255 \\ \left(\frac{150-75}{255-45}\right) R + 50 & \text{if } 45 \prec R \prec 255 \end{cases} \quad \ldots\ldots\ldots\ldots(1)$$

$$\tilde{G} = \begin{cases} 45 & \text{if } G \leq 5 \\ 125 & \text{if } G \geq 175 \\ \left(\frac{125-45}{175-5}\right) G + 40 & \text{if } 5 \prec G \prec 175 \end{cases} \quad \ldots\ldots\ldots\ldots(2)$$

$$\tilde{B} = \begin{cases} 25 & \text{if } B = 0 \\ 100 & \text{if } B \geq 175 \\ \left(\frac{100-25}{175}\right) B + 25 & \text{if } 5 \prec B \prec 175 \end{cases} \quad \ldots\ldots\ldots\ldots(3)$$

Normalized color components are further processed, as given by [6]. The image obtained by normalized and processed color components is converted to YC_bC_r space and binary mask is generated based on values of C_b and C_r as in equation (4). The ranges of the C_b and C_r used in equation (4) are obtained by manually collecting the color skin samples of corrected face images in YC_bC_r space.

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

$map(i, j)$ = 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 for segmentation of face region based on skin color as given in (5).

If $map(i, j) = 0$ then R,G and B components of original image are set to 0

else R,G and B components of original image are kept as it is.....(5)

The result of face segmentation algorithm is shown in Fig.1 (a) to Fig.1 (c). Fig. 1(a) shows original image with illumination direction as left, Fig.1 (b) shows result of face segmentation in YC_bC_r space without any normalization and Fig. 1(c) shows result of face segmentation in YC_bC_r space with proposed method. Result of Fig.1 (b) shows that left side region of the face area is wrongly detected as non skin area and background is detected as skin area. Result of proposed algorithm, as displayed in Fig. 1(c), shows proper segmentation of the face area from the background.

Fig. 1(a) : Original image

Fig. 1(b) : Face segmentation without normalization

Fig. 1(c) : Face segmentation with proposed method.

Interest Point Detection

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 achieve scale invariance. The Harris-Laplace detector uses scale-adapted Harris function, as in equation (6), to localize corner points in scale space.

$$N(X,\sigma_I) = \sigma_D^2\, g(\sigma_I) \otimes \begin{bmatrix} L_x^2(X,\sigma_D) & L_xL_y(X,\sigma_D) \\ L_xL_y(X,\sigma_D) & L_y^2(X,\sigma_D) \end{bmatrix} \quad \text{.....(6)}$$

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 (7) where λ is constant.

$$R(X,\sigma_I) = \det(N(X,\sigma_I)) - \lambda \operatorname{tr}^2(N(X,\sigma_I)) \quad \text{.....(7)}$$

The point is selected as a corner point if it satisfies the condition given in equation (8) i.e. corner response of the point must be positive and maximum in the neighborhood of 3×3 window.

$$\left.\begin{array}{l} R(X,\sigma_I) \succ 0 \text{ and} \\ R(X,\sigma_I) \succ R(X_w,\sigma_I)\ \forall\ X_w \in W \\ W = \text{window of size } 3\times3 \end{array}\right\} \quad \text{.....(8)}$$

We use gray scale image of original RGB image to build the scale-space representation by using equation (6). 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 $s \in [0.7, 0.8, \ldots\ldots, 1.4]$. Large scale change of 1.4 is used to detect initial interest points with k=1 and then small scale changes, specified by s, are used to get better accuracy for the location of X as in [8, 9].

Interest Point Selection

Scale –space representation of Harris corner detector gives large number of interest points, representing important image contents, at various resolution levels. We selected the interest points, which appeared repeatedly, at each and every level of resolution and then sort them in descending order of Harris corner response, obtained at that point. Experiments are carried out with different number of interest points and recognition accuracy is checked. The objective of these experiments is to determine optimal number of interest points required to get good recognition accuracy. fig (2.a) shows the original image and fig (2.b) shows the selected interest points superimposed on the original image. In order to highlight the points, 3×3 neighborhood of point is highlighted.

Fig (2.a) : Original image

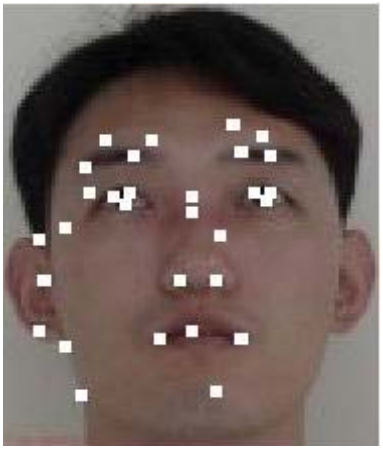

Fig (2.b) : Selected interest points

Interest Point Description

Detected interest points are described by 2D Gabor filter. These 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) and specified in [10]:

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

where $x' = x\cos\theta + y\sin\theta$ and $y' = -x\sin\theta + y\cos\theta$. The values of θ and λ are as in equation (9).Feature extraction is done by placing the Gabor filter mask at each interest point and calculating magnitude value of convolution. Since the phase information of the Gabor filter is time-varying, only magnitude values are used for feature description.

Feature Vector Representation

Gabor filter responses, corresponding to different wavelengths (λ) and orientations (θ), are arranged in feature matrix f_N for each interest point as in equation (10).

$$f_N = \begin{bmatrix} r(x_o,y_o;\lambda_1,\theta_1) & r(x_o,y_o;\lambda_1,\theta_2) & r(x_o,y_o;\lambda_1,\theta_3) & r(x_o,y_o;\lambda_1,\theta_4) \\ r(x_o,y_o;\lambda_2,\theta_1) & r(x_o,y_o;\lambda_2,\theta_2) & r(x_o,y_o;\lambda_2,\theta_3) & r(x_o,y_o;\lambda_2,\theta_4) \\ r(x_o,y_o;\lambda_3,\theta_1) & r(x_o,y_o;\lambda_3,\theta_2) & r(x_o,y_o;\lambda_3,\theta_3) & r(x_o,y_o;\lambda_3,\theta_4) \\ r(x_o,y_o;\lambda_4,\theta_1) & r(x_o,y_o;\lambda_4,\theta_2) & r(x_o,y_o;\lambda_4,\theta_3) & r(x_o,y_o;\lambda_4,\theta_4) \end{bmatrix} \quad \ldots(10)$$

$$where\ \lambda_1 = 4, \lambda_2 = 4\sqrt{2}, \lambda_3 = 8, \lambda_4 = 8\sqrt{2}\ \ and\ \ \theta_1 = 0, \theta_2 = \frac{\Pi}{4}, \theta_3 = \frac{\Pi}{2}, \theta_4 = \frac{3\Pi}{4}$$

For illumination invariance, f_N is normalized as in equation (11)

$$f_N' = \frac{f_N}{\sqrt{\sum_{i,j}(f_{N\,i,j})^2}} \quad \ldots\ldots(11)$$

The normalized feature vector $f_{nor,i}$ is generated by concatenating all rows of $f_N^{'}$ one by one. With' N' interest point for every subject in the database, final feature vector $final^j$ is obtained for each subject in database as in equation (12).

$$final^j = \left[f_{nor,1}^j, f_{nor,2}^j, \ldots\ldots\ldots\ldots f_{nor,i}^j \right] \text{ where i= Number of interest points, j=subject...(12)}$$

Data matrix X is generated by placing $final^j$ obtained for various subjects in database, side-by-side, as in equation (13)

$$X = \left[(final^1)^{'} | (final^2)^{'} | \ldots\ldots\ldots\ldots | (final^j)^{'} \right] \text{ where j= number of subjects.....(13)}$$

Training And Recognition

Training is done by well known eigen-space projection technique based on principal component analysis. Numbers of images used for training are specified in section 3. Data matrix X , generated from training images is mean centered. The mean centered data matrix is subjected to principal component analysis. It gives set of eigen-vectors with associated eigen-values. Eigen-vectors are ordered according to eigen-values from high to low and eigen-vectors associated with non-zero eigen-values are only kept. Each of the training vectors $(final^1)^{'}$ is projected onto eigen-vectors to generate eigen-space.

Similarity measures

For recognition, probe vector is generated by applying same steps i.e. interest point detection, interest point selection, interest point description, on test image to be recognized. Probe vector is projected into eigen-space and projection is compared with the projections of training vectors by similarity measures. Three similarity measures i.e. city-block (L1 norm), squared Euclidean distance (L2 norm), and cosine angle between feature vectors are used as given by equations (14 to 16) respectively.

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

2. $$d(x,y) = \|x-y\|^2 = \sum_{i=1}^{k} (x_i - y_i)^2 \text{ .....(15)}$$

3. $$d(x,y) = -\frac{x.y}{\|x\|\|y\|} = -\frac{\sum_{i=1}^{k} x_i y_i}{\sum_{i=1}^{k} x_i^2 \sum_{i=1}^{k} y_i^2} \text{.....(16)}$$

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. There are three kinds of systematic variations, such as illumination, pose and expression, in the database. 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 (FE), illumination variations (IV) and pose variations (PV), as mentioned in [4]. 50% of images are used for training. Before performing the experiments, all images were cropped to 112x92 pixels images. No any preprocessing is performed based on manually selected eye positions as in [4]. Table.1 shows the percentage of recognition accuracy achieved by proposed method and results reported in [4] for various variations. Reported results are available for expression and illumination variations only. No any results are available for experiment pose variation. For methods marked with '*', results are reported on FERET database.

Results displayed in Table 1 shows that recognition accuracy achieved by our method, for illumination variation, is almost doubled while for expression variation, it is comparable because we processed the raw images as it is without any normalization, as used by [4] and number of interest points is also very less i.e. 30 results in feature vector size of 1*480. Small size feature vector results in less computation cost and less storage requirement.

Table 1 : Comparison of recognition accuracy against number of interest points used for various similarity measures is as shown in fig.(3).

	EV	**IV**	**PV**
EBGM Optimal*	90	42	--
LBP*	97	79	--
SIFT_GRID*	94	35	--
Person specific SIFT*	97	47	--
Our method	95	80	91

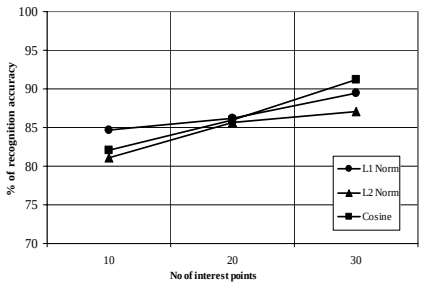

Figure 3

It shows that recognition accuracy increases with number of interest points. We cannot increase the number of interest points beyond 30 because very less number of points satisfies the condition mentioned in section 2.3. Comparison of three similarity measures shows that cosine angle is a good choice for proposed method, as compared L1 and L2 norms.

Conclusion

We presented promising capability of local features for face recognition. Algorithm uses multi-scale Harris corner detector to extract a set of interest points from face images and set of Gabor filters is used to extract local information from selected interest points. Extracted local information is further converted to eigen-space for dimensionality reduction and variance maximization. Recognition is performed using various similarity measures. Several experiments on the Asian face dataset validate the robustness of our technique against variations in facial expressions, pose and illumination. Also the experiments shows that comparable face recognition accuracy, over state-of - art methods can be achieved by using reliable interest points and measuring similarity by cosine angle between feature vectors.

In future, we will try to build multi-modal system, by using different interest point descriptors and combine the recognition result obtained by each descriptor with proper voting algorithm so as to get improved recognition accuracy.

References

[1] Wiskott L, "Face Recognition by Elastic Bunch Graph Matching", *IEEE Trans.PAMI,* Vol. 19(7), 1997, pp. 775-779.

[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] Sluzek A., "Using interest points for robust visual detection and identification of objects in complex scenes", *ICIRS*, 2006, pp. 5321-5326.

[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), 2000, pp.151–172.

[8] Lowe, D.G., "Object recognition from local scale-invariant features", 1999, pp. 1150–1157.

[9] Mikolajczyk K., "Scale & affine invariant interest point detectors", *Computer Vision*, Vol. 60(1), 2004, pp. 63–86.

[10] Weldon T. P., "Efficient Gabor filter design for texture segmentation", *Pattern Recognition*, Vol.29 (12), 1996, pp. 2005-2015.