PERSON-SPECIFIC SIFT FEATURES FOR FACE RECOGNITION

Jun Luo[1], *Yong Ma*[2], *Erina Takikawa*[2], *Shihong Lao*[2], *Masato Kawade*[2], *Bao-Liang Lu*[1]
[1] Department of Computer Science and Engineering
Shanghai Jiao Tong University, Shanghai, China, 200240
{ljfootball, bllu}@sjtu.edu.cn
[2] Sensing & Control Technology Lab., Omron Corporation, Kyoto, Japan, 619-0283
{ma, erinat, lao, kawade}@ari.ncl.omron.co.jp

ABSTRACT

Scale Invariant Feature Transform (SIFT) proposed by Lowe has been widely and successfully applied to object detection and recognition. However, the representation ability of SIFT features in face recognition has rarely been investigated systematically. In this paper, we proposed to use the person-specific SIFT features and a simple non-statistical matching strategy combined with local and global similarity on key-points clusters to solve face recognition problems. Large scale experiments on FERET and CAS-PEAL face databases using only one training sample per person have been carried out to compare it with other non person-specific features such as Gabor wavelet feature and Local Binary Pattern feature. The experimental results demonstrate the robustness of SIFT features to expression, accessory and pose variations.

Index Terms— SIFT, person-specific, face recognition

1. INTRODUCTION

Face recognition has attracted much attention [1] in last decade because of its wide applications. However, face recognition is still an unsolved problem as human face is not rigid object and it can be transformed easily under different situations. Therefore, how to represent the intrinsic attributes of a human face effectively becomes much more important to increase the accuracy of face recognition systems. Various kinds of methods have been proposed for face representation. Subspace methods based on dimension reduction such as Eigenfaces [2] and Fisherfaces [3] are classical paradigms for face recognition. In order to represent the detailed properties of face images, a Gabor wavelet transform method was proposed to compute Gabor-filtered images in different scales and orientations [4] instead of the original gray-scale values and it can make the description of face images more robust to different variations. Recently, a local face texture analysis based method Local Binary Pattern [5] (LBP) has been shown a very successful descriptor for face image as its stability and simplicity. Meanwhile, in LBP method, the representation of a face is very directly derived from facial image without any supervised training set involved and the classification is just a simple non-statistical histogram matching procedure.

Scale Invariant Feature Transform (SIFT) [6] proposed by Lowe becomes one of the research interests for pattern recognition because of its excellent performance on object recognition. The SIFT method first detects the local key-points that are notable and stable for images in different resolutions and uses scale and rotation invariant descriptors to represent the key-points. In this respect, SIFT features are quite similar with LBP features with local histogram patterns representing the whole face image. Although SIFT has very good performance in object recognition, whether it is a good descriptor for face images should be analyzed more. Because object recognition requires only coarse features while face recognition needs much more subtle and refined discriminative features. An investigation of SIFT features on face representation has ever been done as the first attempt to analyze the SIFT approach in face analysis context [7]. In their experiment, the performance of SIFT feature was evaluated on a database of 52 persons with 5 training images and 7 testing images per person. Although the result was promising, only on such a small database the conclusion is not very convincing. It is still well deserved to investigate the performance of SIFT features on a large scale under different conditions. Therefore, we propose to apply SIFT features on face recognition with only one training sample per person and evaluate its performance under various conditions. Meanwhile, as SIFT can detect person-specific features in different images, we use a K-means method instead of overlapping sub-windows in [7] to construct stable effective sub-regions on images and compute the matching similarity of all corresponding region pairs. Moreover, as different sub-regions having different discriminative power we propose a weighting scheme when computing the final average similarity value.

The remaining part of this paper is organized as follows: in section 2, the SIFT method and a new person-specific feature matching strategy are described and in Section 3 experiments on single image per person face recognition are presented followed by some discussion and conclusion in Section 4.

1-4244-0728-1/07/$20.00 ©2007 IEEE

2. FACE RECOGNITION ON MATCHING PERSON-SPECIFIC SIFT FEATURES

In this section, we will introduce face recognition framework based on person-specific SIFT features in three parts: Firstly, each input face image is normalized and extracted with SIFT features; Secondly, a k-means clustering on the locations of features is computed to construct sub-regions in face images; Thirdly, a matching computation is processed between a testing image and all registered images for recognizing face.

2.1. Scale Invariant Feature Transform

Scale Invariant Feature Transform has been proposed for extracting distinctive invariant features from images to perform matching of different views of an object or scene. It consists of two main parts: interest point detector and feature descriptor.

SIFT method uses scale-space Difference-Of-Gaussian (DOG) to detect interest points in images. As for an input image, $I(x,y)$, the scale space is defined as a function, $L(x,y,\sigma)$ produced from the convolution of a variable-scale Gaussian $G(x,y,\sigma)$ with the input image and the difference-of-Gaussian function $D(x,y,\sigma)$ can be computed from the difference of two nearby scales separated by a multiplicative factor *k*:

$$\begin{aligned} D(x,y,\sigma) &= (G(x,y,k\sigma)-G(x,y,\sigma))*I(x,y) \\ &= L(x,y,k\sigma)-L(x,y,\sigma) \end{aligned} \tag{1}$$

Then local maxima and minima of $D(x,y,\sigma)$ are computed based on comparing each sample point to its eight neighbors in current image and nine neighbors in the scale above and below. At this scale, the gradient magnitude, $m(x,y)$, and orientation, $\theta(x,y)$, is computed using pixel differences in Equ.(2). Thereafter, an orientation is determined by building a histogram of gradient orientations weighted by the gradient magnitudes from the key-point's neighborhood and it is assigned to each interest point combined with the scale above and provides a scale and rotation invariant coordinate system for the descriptor.

$$\begin{aligned} m(x,y) &= \sqrt{(L(x+1,y)-L(x-1,y))^2+(L(x,y+1)-L(x,y-1))^2} \\ \theta(x,y) &= \tan^{-1}((L(x,y+1)-L(x,y-1))/(L(x+1,y)-L(x-1,y))) \end{aligned} \tag{2}$$

After detecting the interest points in each image, the SIFT descriptor computes the gradient magnitude and orientation at each image sample point in a region around the key-point location weighted by a Gaussian window. The coordinates of the descriptor and the gradient orientations are rotated relative to the key-point orientation to achieve orientation invariance. Fig.1 shows the SIFT features extracted on sample faces and some corresponding matching points in two face images.

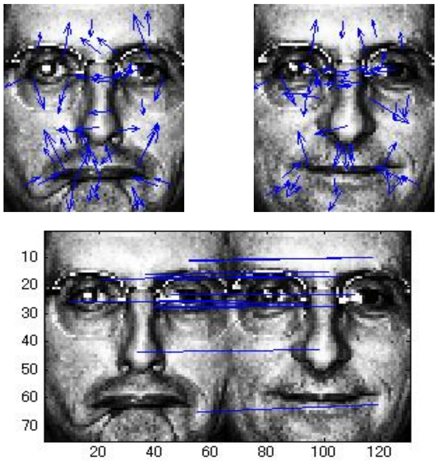


Fig.1 SIFT features on sample images and features matched in faces with expression variation.

2.2. Person-specific feature matching strategy

2.2.1. Sub-region construction for feature matching

In each image the number and the positions of the features selected by SIFT point detector are different, so these features are person-specific. In order to only compare the feature pairs with similar physical meaning between gallery faces and probe faces, same number of sub-regions are constructed in each image to compute the similarity between each pair of sub-regions based on the features inside and at last get the average similarity values. In [7], an overlapping sub-image matching strategy is used for face authentication. However, the definition of the sub-image area is not efficient for the final recognition result. Therefore, we propose to ensemble a K-means clustering scheme to construct the sub-regions automatically based on the locations of features in training samples. The clustering scheme is as follows:

(1) For input registered images, initialize *k* sub-region cluster centers with random values.

(2) Decide the nearest sub-region for each feature point in each image using the Euclidean distance and update the values of each center to reconstruct the sub-regions.

(3) If the new centers remain the same as before recomputed, stop clustering and the remaining *k* sub-regions are the resulting areas for matching.

(4) After constructing the sub-regions on face image, when testing a new image, all the SIFT features extracted from the image are assigned into corresponding sub-regions based on the locations.

The construction of five sub-regions is illustrated in Fig.2 and it can be seen that the centers of regions denoted by crosses just correspond to two eyes, nose and two mouth corners that agree with the opinion of face recognition experience as these areas are the most discriminative parts of face images.

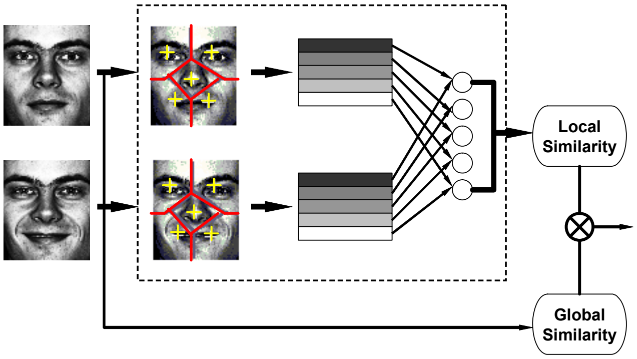


Fig.2 Sub-region construction and similarity computation scheme for the face recognition system

2.2.2. Matching

Based on the constructed sub-regions, a Local-and-Global combined matching strategy is used for face recognition. Assuming a face image I is represented as $\left(m_1, m_2, \cdots m_k\right)$ SIFT feature descriptors scattered in k sub-regions and denoted by:

$$I = \left(f_1^1, \cdots f_1^{m_1}, f_2^1, \cdots f_2^{m_2}, \cdots, f_k^{m_k}\right) \tag{3}$$

where the f_i^j means the jth SIFT descriptor in the ith sub-region. Then the similarity between a testing image and a registered image in training sample I_t, I_r is computed by:

Local Similarity

$$S_L\left(I_t, I_r\right) = \frac{1}{k}\sum_{i=1}^{k}\left(\max_{x\in[1,\cdots m_{ti}], y\in[1,\cdots m_{ri}]}\left(d(f_{ti}^x, f_{ri}^y)\right)\times w_i\right) \tag{4}$$

where $d(f_{ti}^x, f_{ri}^y)$ denotes the similarity between two SIFT feature vectors f_1, f_2:

$$d = \frac{\left\langle f_1, f_2 \right\rangle}{\left\| f_1 \right\| \cdot \left\| f_2 \right\|} \tag{5}$$

and the w_i means the weight for the ith sub-region computed similar as described in [8] according to the recognition rates using each sub-region only based on an additional evaluation set.

Global Similarity

$$S_G\left(I_t, I_r\right) = \frac{match(I_t, I_r)}{\sum_{i=1}^{k} m_{ri}} \tag{6}$$

where the $match(I_t, I_r)$ is the number of validly matched features of two images computed using the same method as Lowe [6] with the distance ratio between the nearest and the second nearest is less than a pre-specified value.

So the final similarity value is:

$$S = S_L \times S_G \tag{7}$$

with the bigger S indicates more similar attribute.

3. EXPERIMENTS

As summarized in [1] that the most difficult recognition problem for face recognition including expression, illumination, multi-pose, accessory and age variation. To investigate the detailed effects of SIFT on face recognition using our approach, we carried out experiments on two large-scale databases FERET [9] and CAS-PEAL [10] and compared the proposed method with other methods in [7, 8, 9]. Using the preprocessing tool from CSU [11], images of FERET and CAS-PEAL are normalized and masked in the way the same as [8] with images from FERET having a size of 150×130 while images from CAS-PEAL database having a size of 75×65. In the following description, we denote LBP_HI and LBP_CHI as LBP combined with *histogram intersection* and *Chi-square* matching strategies, SIFT with grid sub-window matching as SIFT_GRID and our method as SIFT_CLUSTER. As for the proposed method, we constructed the face images into 5 regions.

3.1. Experiment on the FERET face database

The FERET face database is used to evaluate the proposed method according to the standard FERET evaluation protocol with the gallery set including 1196 images of 1196 persons and four kinds of probe sets: fafb (1195 images with expression variations); fafc (194 images with illumination variations); dup.I (722 images taken in less than 18 months); dup.II (234 images taken about 18 months later). For our evaluation set, we choose 200 images from subset of fafc and dup1. The performances are shown in Table.1 including the results of proposed method, SIFT_GRID, and some reported results for one image per person problem such as Elastic Bunch Graph Matching [12] and Uniform LBP [8] and also the results from statistical method like Fisherface [3]. It can be seen that SIFT in our method has almost the same performance as weighted LBP on fafb probe set; however, due to the person-specific condition, SIFT features may fail to detect and describe face images well and the performance becomes not so good.

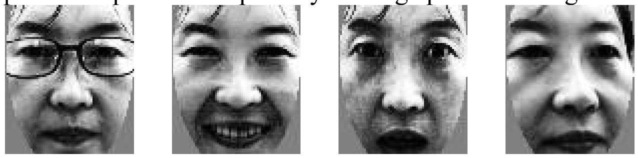
Table.1 The rank-1 recognition rates of different methods on FERET probe data sets

Methods	Fafb	fafc	dup.I	dup.II
Fisherface [3]	0.94	0.73	0.55	0.31
EBGM_Optimal [12]	0.90	0.42	0.46	0.24
LBP results of [8]	0.97	0.79	0.66	0.64
SIFT_GRID [7]	0.94	0.35	0.53	0.36
Our method	0.97	0.47	0.61	0.53

Note: The result of EBGM algorithm is taken from [12] as we use the same normalization method as them.

3.2. Experiment on the CAS-PEAL face database

On the CAS-PEAL face database, similar experiments are carried out on different kinds of face image condition. We enroll 400 persons from the data set as gallery set and each person has one image. Then four different probe sets including 291, 739, 754, and 754 images corresponding to accessory, expression, 15° Pose angle, and 30° Pose angle variations respectively as shown in Fig.3. An evaluation set of 200 images on 15° Pose angle is used to compute the weights of each sub-region for our method. The results are shown in Fig.4. It can be seen that the proposed method performs quite well especially for large pose view angle.



(a) Glass (b) Smile (c) Surprised (d) Pose15°

Fig.3 Sample Normalized Probe images from CAS-PEAL

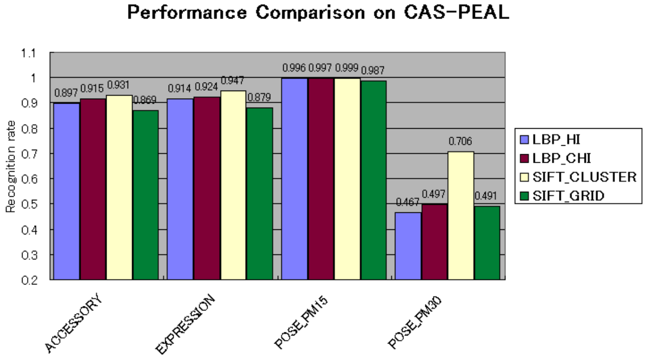


Fig.4 Performance comparison on CAS-PEAL database

4. CONCLUSION

In this paper, we proposed to assemble *Scale Invariant Feature Transform* (SIFT) method with a new matching strategy based on K-means to investigate the robustness of SIFT features to various probe images on face recognition. As the feature number in each image is different, the recognition based on person-specific feature matching is more difficult than general problem. Experiments considering the difficulties of face recognition such as expression, illumination and pose variation were carried out on FERET and CAS-PEAL databases and the results gave the detail information about the performance. As for expression, pose and accessory variations, SIFT features perform quite well and robust even under a single training image. However, the method fails to work under lighting and age variations because of the person-specific features may be more sensitive to these variations.

5. ACKNOWLEDGEMENT

The work of J. Luo and B. L. Lu was partially supported by the National Natural Science Foundation of China via the grants NSFC 60375022 and NSFC 60473040.

6. REFERENCES

[1] W.Y. Zhao, R. Chellappa, P.J. Philips, and A. Rosenfeld, "Face Recognition: A Literature Survey," *ACM Computing Survey*, pp. 399-458, 2003.

[2] M. Turk, and A. Pentland, "Eigenfaces for Recognition," *Journal of Cognitive Neuroscience*, vol.3, no.1, pp.71-86, 1991

[3] P. N. Belhumeur, J.P. Hespanha etc, "Eigenfaces vs Fisherfaces: recognition using class specific linear projection" *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.20, no.7, pp.711-720, 1997

[4] C. Liu and H. Wechsler, "Gabor Feature Based Classification Using the Enhanced Fisher Linear Discriminant Model for Face Recognition," *IEEE Trans. Image Processing,* vol.11, no.4, pp. 467-476, 2002

[5] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns", *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.24, no.7, pp.971-987, 2002

[6] D. Lowe, "Distince image features from scale-invariant keypoints," *Int. Journal of Computer Vision*, vol.60, no.2, pp.91-110, 2004

[7] M. Bicego, A. Lagorio, E. Grosso, and M. Tistarelli, "On the use of SIFT features for face authentication," *Proc. of IEEE Int Workshop on Biometrics, in association with CVPR*, NY, 2006

[8] T. Ahonen, A. Hadid and M. Pietikäinen, "Face recognition with local binary patterns," *Proc.of European Conference on Computer Vision*, , Springer, pp.469-481, 2004

[9] P.J. Phillips, H. Moon, P. Rauss, and S.A. Rizvi, "The FERET Evaluation Methodology for Face Recognition Algorithms," *Proc. of IEEE International conference on Computer Vision and Pattern Recognition*, pp. 137-143, 1997

[10] B. Cao, and S. Shan etc, "Baseline Evaluations On The CAS-PEAL-R1 Face Database", *LNCS 3338, Advances in Biometric Person Authentication*, pp.370-378, 2004

[11] D.S. Bolme, J.R. Beveridge, M. Teixeira and B.A. Draper, "The CSU face identification evaluation system: Its purpose, features and structure," *Proc. of International Conference on Vision Systems*, Springer-Verlag, pp.304-311, 2003

[12] L. Wiskott, R. Fellous, N. Kruger, and C. von Malsburg, "Face recognition by elastic bunch graph matching," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.19, no.7, pp.775-779, 1997