Research on Feature Matching of Multi-pose Face Based on SIFT

Yingjie Xia
School of Information Science and Engineering, University of Jinan, Jinan, China
ise_xiayj@ujn.edu.cn

Yanbin Han
School of Information Science and Engineering, University of Jinan, Jinan, China
ise_hanyb@ujn.edu.cn

Jinping Li
School of Information Science and Engineering, University of Jinan, Jinan, China
ise_lijp@ujn.edu.cn

Rui Chen
The 53RD Research Institute of China North Industries Group Corporation, Jinan, China
shally2004@163.com

Abstract—**Feature matching based on multi-pose faces has become more important in recent years. And it can be used in many fields, such as video monitoring, identity recognition, and so on. In this paper, SIFT algorithm is combined with AdaBoost algorithm, and a method of feature matching based on a multi-pose face is established. Firstly, the face region is extracted from multi-pose face images by AdaBoost. Secondly, SIFT characteristic vectors of the main regions are extracted and matched. The images of the ORL face DB are used in this paper, and some pictures taken in the experiment are used too. The matching results are acceptable and reasonable. Based on multi-pose face, it can be used to research face feature matching, face recognition, video monitoring and 3-D face reconstruction.**

Index Terms—**SIFT, feature matching, AdaBoost, multi-pose face**

I. INTRODUCTION

The number, position, size and pose of face in an input image are confirmed after an analysis, so the face can be recognized. This is called face recognition. This has become important than ever in many fields, such as video monitoring, video compression, identity recognition, and so on. It has been one of the hotspots of pattern recognition, computer vision and human-computer interaction.

Face recognition has developed for over ten years, and there have been many good algorithms. But many of them aimed at obverse and upright faces. Multi-pose face is that face image covers a visual angle range. There are some difficulties in multi-pose face recognition [1]. And there are few effective methods. Compared with frontal-pose face researches, multi-pose face researches are much weaker and further from practical application. The problems are the feature extraction and matching of

multi-pose face. Qu Yanfeng simplified the multi-pose detection by using an eye-analog detection, and it used multi-template matching and rule verification to determine the location of real face [2]. But some of its detection parameters need to be set by a person. A method of multi-pose face recognition based on wavelet transform and LVQ network was proposed in Reference [3]. But the selection of training samples could affect the results. Zhu Changren proposed a multi-pose face recognition algorithm based on a hierarchical model and fusion decision [4]. In his method, traditional face recognition technology based on eigenface was applied in the process of face recognition of every class divided based on their poses. Finally, fusion decision was adopted to get the final result of face recognition. In 2003, a multi-pose face recognition algorithm based on a single view was proposed [5-6]. In Reference [7], a frontal face classifier and a profile face classifier were trained by the AdaBoost algorithm. The two classifiers worked in parallel and their detection results were combined to form candidate face regions. These candidate regions were further verified by a skin color model in YCbCr chrominance space. But it would spend more time on real-time detection of large image. In Reference [8], there is a method for face detection based on the matching of multiple related templates. It used various kinds of templates of various angles to match every possible window of input image. But redundancy matching is so much that detection speed is hard to be improved. Shao Ping put forward a multi-pose knowledge model to acquire the candidate face information from two-value grads image of face organs [9]. According to the sort of pose, the candidate face was affirmed as a true face through template matching with corresponding templates. And it could be used for face detection in either gray image or color image. In Reference [10], a correlative

sub-region mapping method was proposed to recognize the multi-pose face image. But it should be quickened in the search of sub-region, and the results should become more reasonable.

Boosting algorithm was put forward by Freund and Schapire in 1995 [11]. AdaBoost(Adaptive Boosting) algorithm is the promotion of Boosting algorithm [12]. It doesn't need priori knowledge of weak classifiers, and it can be used easily in the practical problems. It is an iterative method, and its main idea is that weak classifiers of different training sets are trained, and they are combined to form a strong classifiers. SIFT (Scale-Invariant Feature Transform) algorithm was proposed by David Lowe in 2004 [13]. And it is an algorithm based on feature matching. The image features extracted by SIFT have such advantages as scale invariability, rotation invariance, affine invariance, which can make the features be matched effectively and accurately, and can reduce mismatching. In this paper, the face region is extracted from multi-pose face images by AdaBoost. Then SIFT characteristic vectors of the face regions are matched. In this paper, the images of the ORL face DB are used, and some pictures taken in the experiment are used too. The matching results are acceptable and effective. It can be used to research face recognition, video monitoring, 3D face reconstruction, and so on.

II. THEORY OF ADABOOST

When AdaBoost algorithm is used for the face detection, it needs more simple features which can distinguish the faces and non-faces to some certain extent.

Haar feature is proposed by Paul Viola [14], and it is a simple rectangle matrix. Fig. 1 enumerates several rectangular characters. All pixels' gray of two or more rectangles of same figure and size are added, and the difference of them is the value of Haar feature. The rectangle features based on pixel gray have a very good performance in edge detection, and it has a good ability of extraction and coding for various kinds of face features.
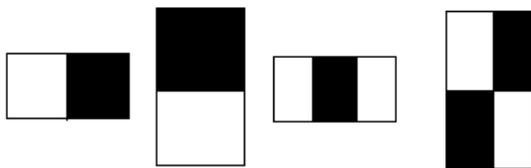

Figure 1. Haar feature

The features who distinguish the faces and non-faces best are attained, then strong classifiers are formed by their corresponding weak classifiers. The detailed steps of this method are as follows:

There is a training set including n training samples $(x_1, y_1), ...,(x_n, y_n)$, where $y_i=\{0,1\}(i=1,2,...,n)$ represents the falseness or trueness of training samples. There are m false samples and a true sample in the training set. The objects which will be classified have k simple features, and they are expressed as $f_j(1 \leqq j \leqq k)$. For the sample x_i, its features are $\{f_1(x_i), f_2(x_i), ..., f_j(x_i), f_k(x_i)\}$, and there is a simple two-value classifier for every input feature f_j.

The weak classifier of the feature j is composed of a threshold θ_j, a feature f_j and an offset p_j (whose value is 1 or -1) which shows the direction of the in equation:

$$h_j = \begin{cases} 1, & if \quad p_j f_j < p_j \theta_j \\ 0, & Otherwise \end{cases} \quad (1)$$

After the analysis of positive and negative samples, the T weak classifiers are selected whose error rates are lowest. Finally, the T weak classifiers are optimized and form a strong classifier. The training steps are following:

1. $w_{t, i}$ is supposed the error weight of the sample i in the circle t. The error weights of the training samples are initialized according to the following formulas: for the sample, $w_{1, i}=1/2m$, if $y_i=0$, or $w_{1,i}=1/2l$, if $y_i=1$.

2. For $t=1, ..., T$

1) The weights are normalized and made according to the following formula:

$$w_{t,i} \leftarrow \frac{w_{t,i}}{\sum_{j=1}^{n} w_{t,j}} \quad (2)$$

2) The weak classifier h_j of every feature j is trained, namely, the threshold θ_j and the offset p_j are confirmed. It makes the error function value least:

$$\varepsilon_j = \sum_{i=1}^{n} w_{t,i} |h_j(x_i) - y_i| \quad (3)$$

3) A weak classifier h_t whose error ε_t is least is attained among the weak classifiers which are confirmed in step 2).

4) The weight of every sample is updated.

$$w_{t+1,i} = w_{t,i} \beta_t^{1-e_i} \quad (4)$$

If x_i is classified accurately, $e_i=0$. $e_i=1$ if not.

$$\beta_t = \frac{\varepsilon_t}{1 - \varepsilon_t}$$

3. The strong classifier formed finally is following:

$$h(x) = \begin{cases} 1, & \sum_{t=1}^{T} a_t h_t(x) \geq \frac{1}{2} \sum_{t=1}^{T} a_t \\ 0, & Otherwise \end{cases} \quad (5)$$

where $a_t = \log \frac{1}{\beta_t}$

A weak classifier whose error rate is at least being searched under the current probability distributing during every iterative process. Then the probability distributing is adjusted, and the probability of the sample which is classified accurately by weak classifiers is reduced. So the samples which are classified inaccurately are given prominence to. Then the next iterative process will aim at the current inaccurate classes, namely, aim at the more difficulties samples, and then the samples classified

inaccurately will be attached importance to. So the weak classifiers formed later will strengthen the training of the samples classified inaccurately.

Fig. 2 shows the results of AdaBoost algorithm. The images of the ORL face DB are used. The left images are the origin images, and the right images are the results of AdaBoost algorithm. Results reveal that the face region can be extracted by AdaBoost, and it has high accuracy and robustness.
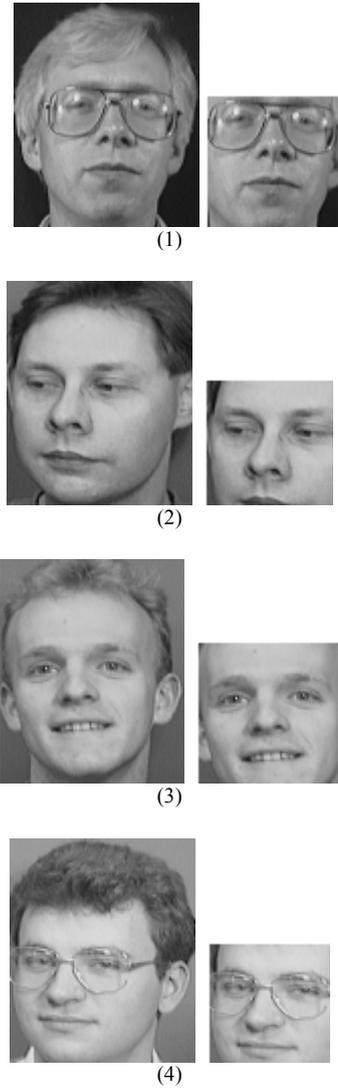

(1)

(2)

(3)

(4)

Figure 2. The results of AdaBoost algorithm

III. SIFT ALGORITHM

Gaussian Scale Space

The concept "scale space" is the complementarily of the famous concept "image pyramid". Lindeberg's research indicated that Gaussian kernel is the only possible linear scale kernel [15]. The scale space of an image is defined a function $L\ (x,\ y,\ \sigma)$. It is the convolution of Gaussian function G and image I, and the scale of G transforms constantly.

$$L(x,y,\sigma) = G(x,y,\sigma) * I(x,y) \tag{6}$$

$$G(x,y,\sigma) = \frac{1}{2\pi\sigma^2} e^{x^2+y^2 / 2\sigma^2} \tag{7}$$

where $(x,\ y)$ is the pixel's coordinate, σ is the scale factor, and L is the scale space. The size of σ determines the smoothness of the image. The features of general picture of corresponding image are shown in big scale, and the detailed features are shown in small scale.

SIFT Algorithm

SIFT algorithm proposed by Lowe involves four steps as following:

(1) Extreme value is detected in scale space. Firstly, the DOG (Difference of Gaussian) pyramid of the image is established. Extreme value is detected in the 26 neighborhoods of scale space. And $D(x,\ y,\ \sigma)$ is the difference of two adjacent scale images, i.e.

$$\begin{aligned} DOG\ (x,y,\sigma) &= (G(x,y,k\sigma) - G(x,y,\sigma)) * I(x,y) \\ &= L(x,y,k\sigma) - L(x,y,\sigma) \end{aligned} \tag{8}$$

A point will be regarded as a feature point of the image of this scale when its value is the biggest or smallest in the 26 neighborhoods of the layer, upper layer and lower layer of the DOG scale space. The schematic diagram is Fig. 3.
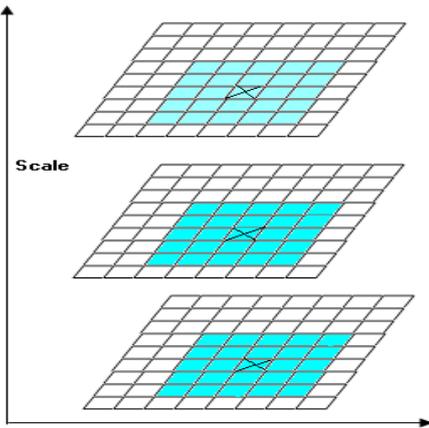

Figure 3. Detection of feature point in Gaussian space

(2) Direction parameters of every key-point are specified by using distributing characteristic of grads direction of the pixels near the key-point. So operators will have advantage of rotation invariance.

$$\begin{aligned} m(x,y) = \sqrt{(L(x+1,y) - L(x-1,y))^2 + (L(x,y+1)} \\ \overline{- L(x,y-1))^2} \end{aligned} \tag{9}$$

$$\theta(x,y) = \arctan\frac{L(x,y+1) - l(x,y-1)}{L(x+1,y) - L(x-1,y)} \tag{10}$$

Module value of the grads at (x, y) will be calculated by (9), and its phase will be calculated by (10). In the formulas, L is the scale of every key-point.

In actual calculation, sampling is executed in neighborhood window whose center is the key point. Then gradient direction of neighboring pixels is counted by histogram. The range of histogram is 0∼360°. And every 10° represents a direction. So there are 36 directions in all. Histogram peak represents the main direction of neighborhood gradient at this feature point, and it is regarded as the direction of this key point. Fig. 4 shows that the main direction of feature point is confirmed by gradient histogram in 7 directions.
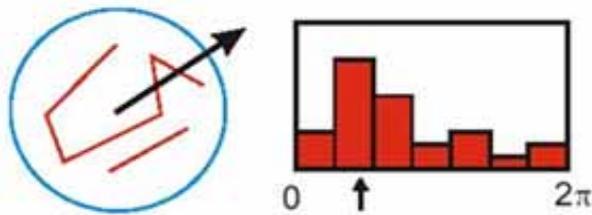
Figure 4. To establish the orientation of feature point

(3) SIFT characteristic vectors are attained. Coordinate axis is rotated to the direction of the characteristic point. So rotation invariance is ensured. Then 8×8 pixel fields whose center is the key-point will be divided into 16 subfields by 4×4. Fig. 5 shows that every small cell represents a pixel near the key point at the scale space. The arrow direction is the gradient direction, and the arrow length is the gradient value. The direction and amplitude in every subfield will be calculated, and then the direction distribution can be obtained. Weighted operators are done by Gaussian window, and a pixel corresponds to a vector. The circle is the weighted range. Every vector is projected to 8 gradient directions in the range of 2×2 child window. And the accumulated value of every gradient direction is drawn. So a 1*8 vector is attained. In practical application, 4×4 window is used, so the 1*128 vector is defined feature description. In order to ensure the illumination invariance, the feature description should be normalized.
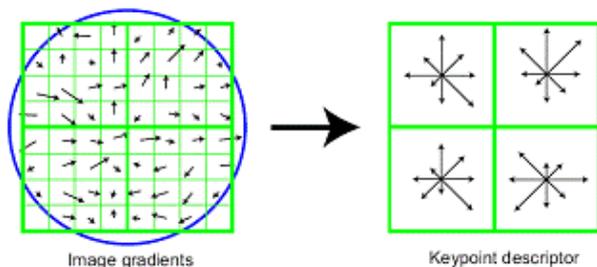

Figure 5. Creating SIFT descriptor from the gradient information of neighbourhood

(4) Features are matched. After SIFT characteristic vectors of two images are attained, the Euclid distance between key-points from two pictures is calculated, and it is regarded as the measurement of the comparability. If

the distance is smaller than the preset threshold, the two images are matched.

From the Fig. 6, it is found that face features can be attained by SIFT. And the distribution area of key points mainly includes eyes, nose and mouth.
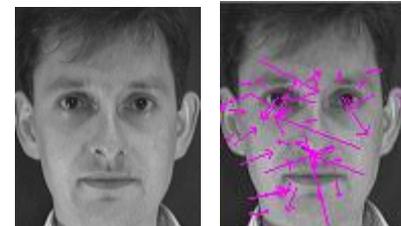
(1)
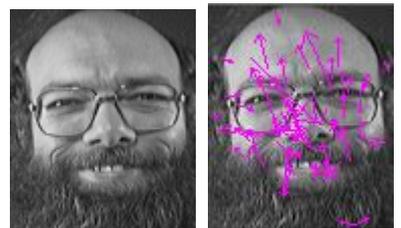
(2)

(3)

(4)

Figure 6. the results of SIFT

IV. EXPERIMENTS AND RESULTS

The images of the ORL face DB are used to be matched in this paper, and some pictures taken in the experiment are used too. The ORL face DB is composed of forty persons' photos, and everyone has ten face photos which have the different lightness, position, expression and so on. The photos of the same people have apparent difference. The size of photo is 92×112.

The distribution area of face's key-points mainly includes eyes, nose and mouth. So face organs area (including eyes, nose and mouth) is attained by AdaBoost algorithm. Neck, hair and background information are ignored. So the features extracted from face region can describe face information more accurately. Then the error will be reduced during the matching.

In the Fig. 7, the first and the second group images are from the ORL face DB, and the third and the fourth group images are shot in the experiment. In every group images, the first two figures are original images, the second two figures are the images of face region got by AdaBoost, the third two figures are the images including face feature vectors attained by SIFT, and the fourth figure is the accurate result of feature matching.

In the Fig. 8, the images are from the ORL face DB. In every group images, the first two figures are original images, the second two figures are the images of face region got by AdaBoost, the third two figures are the images including face feature vectors attained by SIFT, and the fourth figure is the inaccurate result of feature matching.

The experiment shows that different face has different face features, the face features can be reserved in face region, and be extracted by SIFT. They can describe face information adequately.

V. CONCLUSION

A method of feature matching based on multi-pose face is proposed in this paper. Firstly, the face region is extracted from multi-pose face images by AdaBoost. Secondly, SIFT characteristic vectors of the main regions are attained. Lastly, features are matched by them. The images of the ORL face DB are used in this paper, and some pictures taken in the experiment are used too. The matching results are acceptable and reasonable. The experiment shows that the face features extracted by SIFT have a high robustness during the feature matching. It can be used to research face recognition based on multi-pose face.

REFERENCES

[1] Liang Luhong, Ai Haizhou, Xu Gangyou, et al, "A Survey of Human Face Detection," *Chinese Journal of Computers*, vol. 25, pp. 449-458, May 2002.
[2] Qu Yanfeng, Li Weijun, Xu Jian, and Wang Shouju, "Fast Multi-Pose Face Detection in A Complex Background," *Journal Of Computer-Aided Design & Computer Graphics*, vol. 16, pp. 45-50, January 2004.
[3] CHEN Lei, HUANG Xianwu, and SUN Bing, "Pose-varied Face Recognition Based on WT and LVQ Network," *Computer Engineering*, vol. 32, pp. 47-49, November 2006.
[4] Zhu Changren, and Wang Runsheng, "Multi-Pose Face Recognition Based On A Hierarchical Model And Fusion Decision," *Journal Of Electronics And Information Technology*, vol. 24, pp. 1447-1453, November 2002.
[5] Zhu Changren, and Wang Runsheng, "Multi-Pose Face Recognition Based on a Single View," *Chinese Journal Of Computer*, vol. 26, pp. 104-109, January 2003.
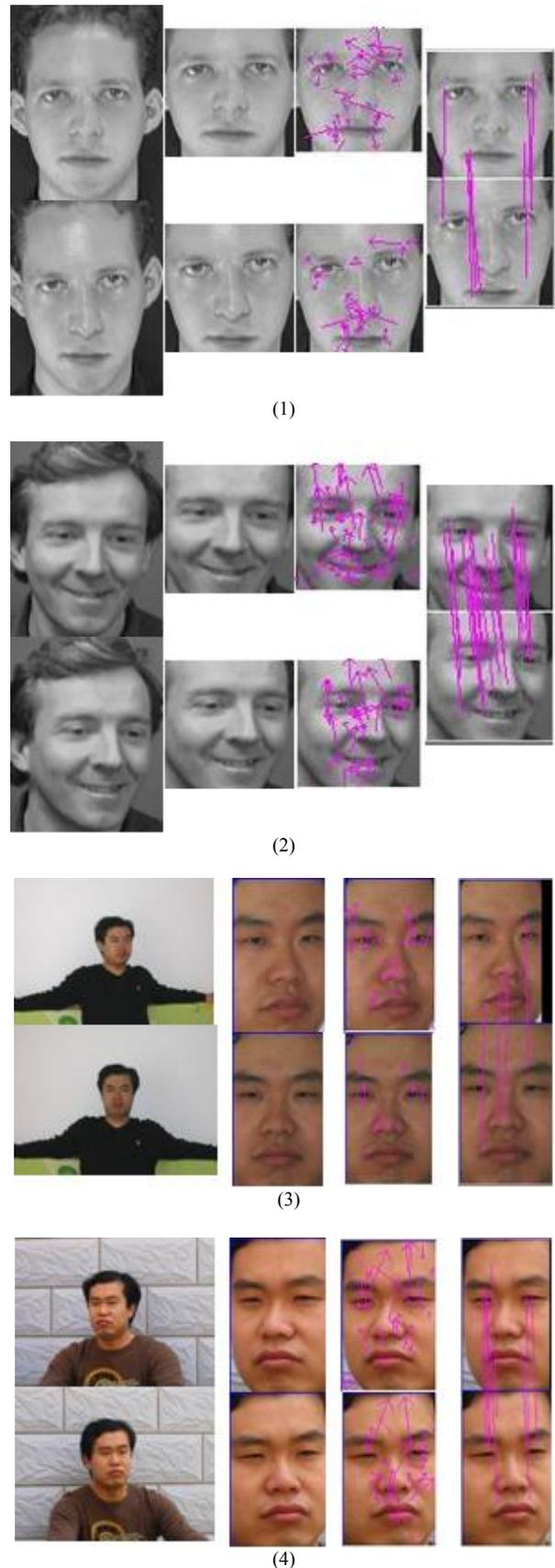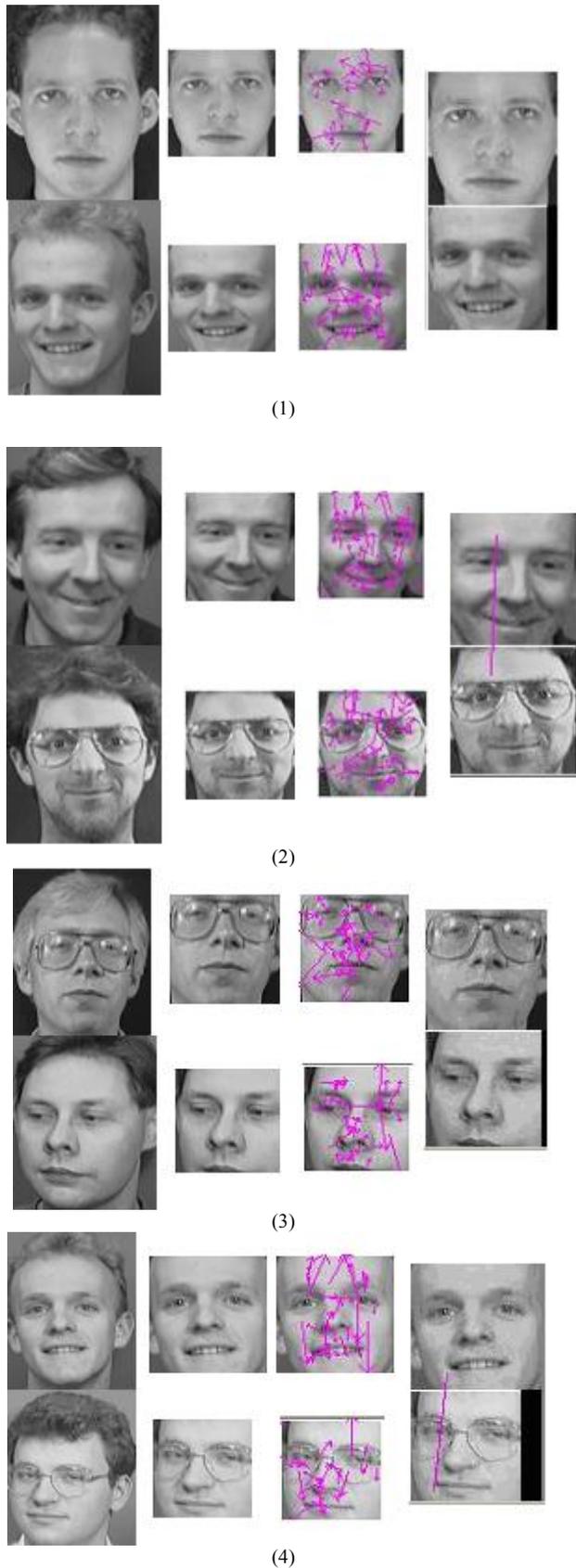

(1)

(2)

(3)

(4)

Figure 7.   Accurate results of feature matching

[6] Zhu Changren, and Wang Runsheng, "Research On Multi-Pose Face Image Synthesis From A Single View," *Journal Of Electronics And Information Technology*, vol. 25, pp. 300-305, March 2003.

[7] BI Ping, ZHAO Heng, and LIANG Ji-min, "Variant Pose Face Detection Based on Multi-classifier Fusion," *Journal of System Simulation*, vol. 21, pp. 6469-6473,6478, October 2009.

[8] Liang Luhong, Ai Haizhou, He Kezhong, el al, "Face Detection Based on The Matching of Multiple Related Templates," *Chinese Journal of Software*, vol. 12, pp. 94-102, January 2001.

[9] SHAO Ping, YANG Lu-ming, HUANG Hai-bin, and ZENG Yao-rong, "Rapid Face Detection with Multi-pose Knowledge Models and Templates," *Journal of Chinese Computer Systems*, vol. 28, pp. 346-350, February 2007.

[10] CHEN Hua-jie, and WEI Wei, "Multi-pose Face Recognition Based on Correlative Sub-region Mapping," *Journal of Image and Graphic*, vol. 12, pp. 1254-1260, July 2007.

[11] Freund Y, "Boosting a weak learning algorithm by majority," *Information and Computation*, vol. 121, pp. 256-285, February 1995.

[12] P Viola , M Jones, "Robust Real-time Object Detection," *Second International Workshop on Statistical and Computational Theories of Vision-Modeling, Learning, Computing and Sampling*, vol. 13, pp. 1-25, 2001.

[13] Lowe D G, "Distinctive image features from Scale-Invariant keypoints," *International Journal of Computer Vision*, vol. 60, pp. 91-110, February 2004.

[14] Paul Viola, Michael Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features," In: Proceeding of IEEE Conference on Computer Vision and Pattern Recognition, Kauai, Hawaii, USA, pp. 905-910, 2001.

[15] Lindeberg T, "Scale-space theory: A basic tool for analyzing structures at different scales," *Journal of Applied Statistics*, vol. 21, pp. 224-270, 1994.

(1)

(2)

(3)

(4)

Figure 8.   Inaccurate results of feature matching
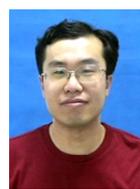
Yingjie Xia, born in Jinan, Shandong Province, China, in November 1978. He received the Bachelor's degree in computer science and technology from University of Jinan in 2001. In the same year, he started working as a teacher at School of Infor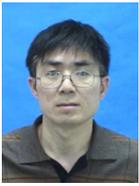mation Science and Engineering, University of Jinan. In 2007, he received Master degree in computer application in School of Information Science and Engineering, University of Jinan.

His major is image processing, pattern recognition and 3-D reconstruction, he has published over 5 papers, and 2 were indexed by EI.

Yanbin Han, born in Weifang, Shandong Province, China, in March 1979. From 1998 to 2005, he studied at School of Information Science and Engineering, University of Jinan, and received the Bachelor and Master of degree at computer application. In the same year, he started working as a teacher at School of Information Science and Engineering, University of Jinan.

His major is image processing and pattern recognition, he has published over 10 papers, and 3 were indexed by EI.

Jinping Li, born in Jiaozuo, Henan Province, China, in August 1968. In 1997, he received Doctor degree in science in China University of Technology. In the same year, started working as a teacher at School of Information Science and Engineering, University of Jinan.

His major is image processing, pattern recognition and optimization. He worked as a visiting scholar at University of Birmingham CERCIA in 2005. He charges Natural Science Foundation of Shandong Province and Doctor Fund of Shandong Province, and he is a member of several 863 Projects and National Natural Science Foundation of China. He has published over 80 papers, and 20 were indexed by SCI/EI.

Rui Chen, born in Baotou, Inner Mongolia Autonomous Region, China, in October 1981. She received the Bachelor's degree in computer science and technology from University of Jinan in 2001. In the same year, started working at the 53RD Research Institute of China North Industries Group Corporation.