Research on Hand Gesture Recognition Based on Inner-distance Contour Point Distribution Features and Histogram Matching

Qiu-yu Zhang
School of Computer and Communication, Lanzhou University of Technology, Lanzhou, China
Email: zhangqylz@163.com

Hui-yi Wei, Mo-yi Zhang, Zhi-gang Xu, Hong-xiang Duan and Lu Lv
School of Computer and Communication, Lanzhou University of Technology, Lanzhou, China
Email: wei.huiyi@163.com

***Abstract*—Aiming at the influence of joints or part structures deformations on the accuracy of gesture recognition in representing, and large amount of calculation with shape matching directly, a method based on inner-distance contour point distribution features (IDCPDF) and histogram matching is proposed in this paper. Firstly, elliptical skin model is used to segment and extract contour. Then IDCPDF of gestures is generated. Finally, histogram matching is used to measure the similarity of IDCPDF and classify. Experimental results show that the method describes distributions of gesture contour points under polar coordinates. It not only reflects significant information of gesture shapes, but also reduces calculations in gesture features extraction and matching on the promise of ensuring gesture recognition accuracy, and achieves better real-time performance. Meanwhile, this method keeps good robustness on joints and part structures deformations of hands.**

***Index Terms*—Hand gesture recognition, inner-distance contour point distribution features, histogram matching, deformable hand gesture**

I. INTRODUCTION

Recent years, with the widely use of computers in the human industries and life, research on Human-Computer Interaction technology becomes extremely active [1-2]. As an important means of communications, gestures are vivid and informative. Therefore, vision-based gesture recognition draws more attention. However, as hand is an articulated deformable object with 27 freedom degrees [3] and the inadaptability of vision itself, hand feature extraction is one of key issues in vision-based gesture recognition. There are many researches on how to extract simple and effective features for gesture recognition, such as hand contour features (Fourier descriptors [4], curvature, etc), regional features (Hu invariant moments [5], density set [6], convexity, area, perimeter, etc), skeleton features [7-8], spatial distribution characteristics [9], shape parameters [10], SIFT features [11-13], multi-feature fusion [14], shape context (SC) [15], inner-distance shape context (IDSC) [16] and so on. Shape context was proposed by Belongie et al. [17], which described the relative position of any pair of points on the contour. Since the extracted statistic histogram will change with positional relationship of joints and part structures, its ability to describe objects with joints is reduced. Ling et al. [18] used the inner-distance instead of the Euclidean distance to measure distance between two points, which can represent non-rigid objects better. It's called inner-distance shape context (IDSC). As a result of the deformations of finger joints and part structures, along with the different opening angles between fingers, contour of the same gesture varies. Thus, the problem of hand shape changes is particularly outstanding. Besides, the method of shape matching needs more computing work and time, which can not meet real-time performance in gesture recognition.

Aiming at the problems mentioned above, a method of gesture recognition based on inner-distance contour point distribution features (IDCPDF) and histogram matching is proposed in this paper. IDCPDF describes global structural information of gesture shapes, which can relieve influence on describing shape structural features caused by deformations, so robustness on deformations of joints and part structures is ensured. Meanwhile, IDCPDF extraction reduces the computation of features extraction and matching and has good real-time performance.

II. PROPOSED METHOD

Gesture recognition process is shown in Fig. 1.

Figure 1. Gesture recognition flowchart

Firstly, gesture area is obtained by segmentation. Then IDCPDF of gestures is extracted. Finally, histogram matching is used to classify and recognize. This experiment is carried out with simple background and natural illumination in laboratory, so elliptical skin model

doi:10.4304/jcp.9.10.2455-2460

proposed by Hsu et al. [19] is used to extract gesture area considering real-time performance.

A. IDCPDF Extraction

Shape is represented by binary image of gestures and can be seen as contour. It is described by location distributions of pixel points on the contour.

In this paper, 8-neighbor chain code tracking algorithm is used to extract contour of binary gesture, then landmark points are sampled uniformly. As a result, a discrete point set is obtained, represented as $P=\{p_1, p_2, \ldots, p_N\}$, where N denotes the number of sample points.

For a given gesture contour, take the centroid as center and the longest inner distance R between the centroid and N sample points on the contour as radius to establish M concentric circles and divide into L equal parts in circumferential direction. The obtained polar histogram bins are shown in Fig. 2. The relative position (inner distance and inner angle) of the shortest path between the centroid and N sample points is simplified the number of points distributed in each bin. Calculate the number of points distributed in each bin separately to obtain a histogram, called inner distance contour point distribution features (IDCPDF). According to the above rules for establishing IDCPDF, translation and scaling invariance is inherent property. In order to ensure rotation invariance, the orientation direction is estimated by computing the scatter direction of gesture images through principal component analysis (PCA) [16], which is taken as the start direction of polar coordinates. The start direction of the shortest path from the centroid to sample points, named as inner-angle, is used to measure the direction of sample points.

Here the inner-distance is calculated as following two steps: (1) Create a graph with N sample points and the centroid. For each pair of sample points x and y on the contour, if the line segment connecting x and y is entirely within the shape, add an edge between x and y with the weight equal to the Euclid distance $\|x - y\|$. (2) Calculate the shortest path between the centroid and sample points with the shortest path algorithm.

As mentioned above, the whole algorithm of building IDCPDF can be summarized as follows:

Step 1: Extract the contour of gestures using the 8-neighbor chain code algorithm.

Step 2: Sample uniformly to obtain a discrete point set P.

Step 3: Calculate the centroid (x_c, y_c) of gesture shape.

Step 4: Create a graph with N sample points and the centroid.

Step 5: Calculate the shortest path and inner angle between the centroid and N sample points using the shortest path algorithm.

Step 6: Estimate the principal direction of gesture area.

Step 7: Calculate the longest inner distance R from the centroid to sample points.

Step 8: Take the centroid (x_c, y_c) as center, principle direction of gesture area as the start direction to build polar coordinates and calculate inner angle in polar coordinates.

Step 9: Divide the longest distance R into M equal parts and circular angles into L equal parts to get polar histogram bins with $M \times L$ sections.

Step 10: Count the number of points on the contour distributed in each bin and construct contour sample point distribution features to finally get the one-dimensional histogram namely IDCPDF.

Fig. 3 shows an articulated object shape with two joints, where O is the centroid of object shape and dashed parts are the contour after joint C and D rotation. Before rotation, the inner distance between O and A is calculated as $d=OC+CA$; the inner angle denoted by θ, is the angle between OC and the start direction of polar coordinates. After rotation, the inner distance between O and A' is calculated as $d'=OC+CA'$. Now d is equal to d', while inner angle still maintains θ. It can be seen that this distance measure way is less influenced by joints and part structures, and insensitive to joints.

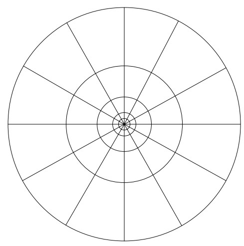


Figure 2. Diagram of polar histogram bins

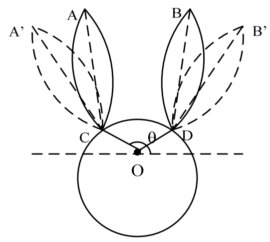


Figure 3. Articulated object example

Fig. 4 shows the IDCPDF of gestures denoted by I and II, which have same semantics and different contours by rotation; III represents another semantic gesture. Here, the number of sample points (N) is 100, the number of distance bins (M) is 8 and the number of angle bins (L) is 12. IDCPDF of gesture I and II are similar though rotation, while IDCPDF of different semantic gestures varies greatly. Therefore, IDCPDF keeps good robustness on deformations caused by finger joints or part structures.

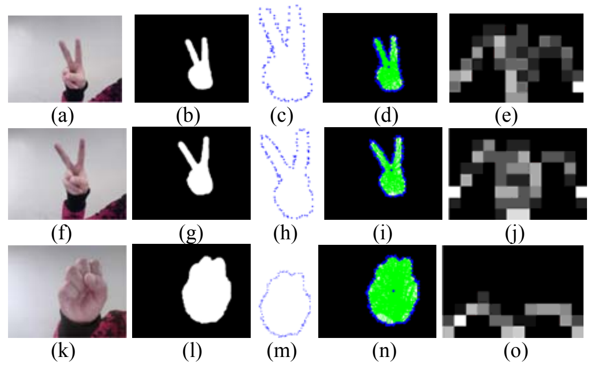


Figure 4(a-o): Illusions of IDSC (a) Gesture I, (b) Image after I segmentation, (c) Sample points of I, (d) Inner-distance of I, (e) IDCPDF of I, (f) Gesture II, (g) Image after II segmentation, (h) Sample points of II, (i) Inner-distance of II, (j) IDCPDF of II, (k) Gesture III, (l) Image after III segmentation, (m) Sample points of III, (n) Inner-distance of III, (o) IDCPDF of III. In the histograms, the vertical axis denotes inner distance bins and horizontal axis denotes orientation bins.

The IDCPDF is a histogram created by statistic distributions of sample points in histogram bins, where the centroid is taken as center and the longest inner distance from the centroid to sample points on the contour is taken as radius; sample points are taken as statistic object; radius and circumferential angles are divided into equal parts. Compared with the SC and IDSC, this method only needs to regard the centroid as reference point to calculate the distributions of sample points relative to the centroid, ie IDCPDF, which can describe a gesture image instead of calculating the SC or IDSC of each sample point. This method greatly reduces computational complexity in gesture features extraction. In addition, IDCPDF uses the inner distance from the centroid to sample points on the contour to describe location distributions of sample points, effectively overcomes the influence of joints and part structures deformations and keeps high robustness.

B. Histogram Matching

Histogram matching method is divided into two categories: bin-to-bin and cross-bin method [20]. The former assumes that the histogram domains are aligned. Considering only the relativity of bins with same index, rather than the relativity of different bins makes the method affected by the number of bins and sensitive to the impacts caused by quantization, deformation and illumination. The latter takes account of the correlation between different bins, not only removing interference of relativity among variables, but also relieving effects of quantization. Histogram matching method will be specifically described as follows.

(1) Chi-Squared [21] is a bin-to-bin histogram distance. It is defined in (1), where P and Q represent the histograms:

$$\chi^2(P,Q)=\frac{1}{2}\sum_i\frac{(P_i-Q_i)^2}{(P_i+Q_i)} \tag{1}$$

(2) Quadratic-Form distance [22] is a cross-bin distance. It is defined in (2), where P and Q represent histograms and A represents the bin-similarity matrix:

$$QF^A(P,Q)=\sqrt{(P-Q)^T A(P-Q)} \tag{2}$$

Quadratic-Form distance is called the Mahalanobis distance when A is the inverse of the covariance matrix.

(3) Earth Mover's Distance (EMD) [20] is a linear programming method used to solve transportation problem.

The method is as follows: m mounds are denoted by $P=\{(p_1,w_{p_1}),\cdots,(p_{m,}w_{p_m})\}$, where w_{p_i} represents the weight of p_i; n pits are denoted by $Q=\{(q_1,w_{q_1}),\cdots,(q_{n,}w_{q_n})\}$, where w_{q_j} represents the capacity of q_j. The distance from i-th mound to j-th pit is denoted by $D=\{d_{ij}:1\le i\le m,1\le j\le n\}$ and the mass of soil transported from i-th mound to j-th pit is denoted by $F=\{f_{ij}:1\le i\le m,1\le j\le n\}$. The goal of EMD model is to seek the minimum work and the corresponding representation is as follows:

$$EMD(P,Q)=\min_{F=\{f_{ij}\}}\frac{\sum_{i=1}^{m}\sum_{j=1}^{n}d_{ij}f_{ij}}{\sum_{i=1}^{m}\sum_{j=1}^{n}f_{ij}} \tag{3}$$

f_{ij} needs to satisfy the following constraints:

$$\begin{aligned}&\sum_{j=1}^{n}f_{ij}\le w_{p_i},1\le i\le m;\\&\sum_{i=1}^{m}f_{ij}\le w_{q_j},1\le j\le n;\\&f_{ij}\ge 0,1\le i\le m,1\le j\le n;\\&\sum_{i=1}^{m}\sum_{j=1}^{n}f_{ij}=\min\left(\sum_{i=1}^{m}w_{p_i},\sum_{j=1}^{n}w_{q_j}\right)\end{aligned} \tag{4}$$

(4) Quadratic-Chi. Ofir Pele et al. [23] combined chi-square distance and quadratic-form distance to form the Quadratic-Chi (QC) distance, defined as follows:

$$QC_m^A(P,Q)=\sqrt{\sum_{ij}\left(\frac{(P_i-Q_i)}{(\sum_c(P_c+Q_c)A_{ci})^m}\right)\left(\frac{(P_j-Q_j)}{(\sum_c(P_c+Q_c)A_{cj})^m}\right)A_{ij}} \tag{5}$$

Here P and Q are histograms and m is a normalized parameter. When I represents the identity matrix and m is equal to 0.5, QC distance is the form of chi-square distance.

$$QC_{0.5}^I(P,Q)=\sqrt{2\chi^2(P,Q)} \tag{6}$$

When m is equal to 0, QC distance is the form of quadratic distance.

$$QC_0^A(P,Q)=QF^A(P,Q) \tag{7}$$

QC distance has Sparseness-Invariance property and Similarity-Matrix-Quantization-Invariance property [23].

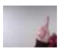
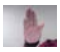
Histogram matching method mentioned above has advantages and disadvantages, we will compare which histogram matching method for IDCPDF in gesture recognition is more effective by experiment, specific matching process is as follows: for IDCPDF of testing gesture images, calculate the matching cost with each kind of gesture image template using above histogram matching method respectively, and the test image will be determined as the category of gesture template with minimum cost.

III. Experimental Results

A. Gesture Image Data Set Collection

Due to the diversity, ambiguity and differences in application backgrounds of gestures, there is no unified gesture image data set in gesture recognition currently, such as the Cambridge Hand Gesture Data Set [16]. Therefore, for universal gesture recognition method, 8 kinds of gestures are defined to verify effectiveness of algorithm proposed in this paper, which are listed in Table I. To avoid the impact caused by perspective on capturing gesture images, as well as the effect of complex background and illumination on gesture segmentation, gesture images are obtained by facing camera in laboratory with natural light and simple background. Each image has a resolution of 640×480. 50 test samples are recoded for each gesture with different scales and rotations (including gesture direction and joint rotations) without any other objects in the background.

TABLE I.
EIGHT KINDS OF GESTURES DEFINED

Label	Images			Label	Images		
1				5			
2				6			
3				7			
4				8			

B. *Gesture Recognition Results*

The experiment is carried out with MATLAB R2012b in Windows7 on a PC with CPU Intel Core i3-2350M 2.30 GHz and 2G RAM. While the parameters N, M and L mentioned above are different, recognition rate of eight gestures and the consumed time with different histogram matching algorithms are not the same. The performance under three different parameters respectively using Chi-Squared [21], Quadratic-Chi [23] and EMD distance [24] is specified in Table II. In this paper improved EMD [24] that simplifies calculation of the distance matrix and has certain robustness and effectiveness, is chosen to avoid higher computational complexity of traditional EMD.

As seen from Table II, when the number of orientation bins is 6 or 8, the recognition rate is relatively good. When the number of orientation bins and the number of angles are constant, consumed time for recognizing 50 images will increase with the number of sample points. In general, recognition rate of cross-bin histogram matching is superior to bin-to-bin method. Chi-Squared method can achieve a higher recognition rate with a smaller number of bins, while recognition rate of QC and EMD is increasing with the number of bins. The consumed time for recognizing 50 images with 3 histogram matching methods has a relationship of EMD>QC>Chi-Squared. Taking into account of real-time and recognition precision, set $N = 100$, $M = 8$ and $L = 12$ as parameters of QC distance. Table Ⅲ lists identification confusion matrix of 50 test samples to eight different gestures.

In order to verify effectiveness of the algorithm, online gesture recognition based on inner distance contour point distribution features and histogram matching is implemented in MATLAB R2012b platform, where gesture templates are same with offline test. Each type of gestures is tested 50 times online and specific results are shown in Table Ⅳ. This method keeps good robustness on deformations caused by joints and part structures, with an average recognition rate of 85.5% to eight kinds of predefined gestures. Meanwhile, processing time of each image is about 0.35s, which can meet real-time requirements. Fig. 5 shows some frames of online recognition.

C. *Experimental Results Analysis*

Ref. [15] illustrated a method for real-time gesture recognition by shape context matching and cost matrix with average recognition rate of 78% in five kinds of gestures. Ref. [25] took the point with larger curvature on the contour as interest point and extract SC features of

TABLE II.
PERFORMANCE UNDER DIFFERENT PARAMETERS WITH DIFFERENT HISTOGRAM MATCHING METHODS

N	M	L	Chi-Squared	QC	EMD	Chi-Square(s)	QC (s)	EMD (s)
100	5	10	79.25%	78.25%	75.75%	15.8503	16.8563	17.1875
150	5	10	84.50%	80.25%	77%	20.0614	20.8725	21.4226
200	5	10	86.50%	82.25%	76%	25.3297	25.9263	26.7294
100	6	10	81.25%	82.25%	77.25%	16.0990	16.8829	17.3417
150	6	10	86.50%	82.75%	79.75%	20.4155	21.2285	21.7910
200	6	10	83.75%	83.75%	78.50%	25.7028	26.5471	27.1312
100	7	10	81%	84%	77.25%	16.2237	17.1728	17.5516
150	7	10	80.50%	81.25%	80%	20.0645	21.4617	21.8675
200	7	10	81%	83.50%	76.75%	25.2122	26.4098	27.5108
100	8	10	79.50%	82.75%	80.50%	16.0826	17.3588	17.7372
150	8	10	81.25%	84.25%	84.50%	20.2938	21.4142	22.3594
200	8	10	83.75%	87%	80.75%	25.9362	27.0173	27.7366
100	5	12	83%	84%	81.5%	16.2247	16.9569	17.2933
150	5	12	82.75%	84%	84%	20.6882	20.9267	21.6559
200	5	12	85%	83%	82.75%	25.6068	26.1665	27.1961
100	6	12	82.75%	87%	83%	16.2430	17.2126	17.6391
150	6	12	85.25%	88%	85%	20.7063	21.5223	21.9882
200	6	12	85.25%	86.50%	84.25%	25.4953	26.9582	27.5310
100	7	12	85.50%	86.25%	83.50%	15.9750	17.3590	17.8518
150	7	12	81.75%	85.25%	84.50%	20.3294	21.6851	22.2607
200	7	12	85.25%	85.75%	85%	25.7722	26.5289	27.6652
100	8	12	83%	87.75%	84.50%	16.0960	17.3931	18.2599
150	8	12	82.75%	83.75%	85%	20.4922	21.6957	22.6814
200	8	12	85.25%	86.25%	86%	25.8260	27.1895	27.8909

TABLE III
GESTURE RECOGNITION CONFUSION MATRIX UNDER SPECIFIC PARAMETERS WITH QC DISTANCE

Gestures Category	1	2	3	4	5	6	7	8
1	47	3	1	0	0	4	0	4
2	0	44	3	0	0	0	0	0
3	0	3	45	13	0	0	0	1
4	0	0	0	35	0	0	0	0
5	0	0	1	1	49	3	0	0
6	0	0	0	0	1	37	1	0
7	0	0	0	0	0	0	49	0
8	3	0	0	1	0	6	0	45
Consumed time of 50 images(*s*)	17.0129	17.6847	17.9015	17.6735	17.4482	17.8735	17.0596	17.1312
accuracy	94%	88%	90%	70%	98%	74%	98%	90%

TABLE Ⅳ:
THE PERFORMANCE ONLINE

Gestures Category	1	2	3	4	5	6	7	8
Accuracy	96%	86%	86%	68%	94%	70%	100%	84%
Consumed time of each image(*s*)	0.3481	0.3563	0.3587	0.3569	0.3547	0.3598	0.3493	0.3526

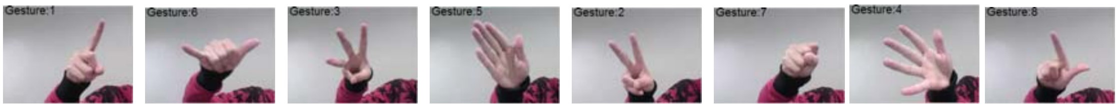


Figure 5. Parts of frames for real-time gesture recognition

interest points to achieve an average recognition rate of 89% in six kinds of gestures. The method in this paper is based on IDCPDF and histogram matching. IDCPDF not only describes the global features of gesture images, but also represents local features by dividing the radius and circumferential angles into equal parts and calculating the relative location distributions between the centroid and sample points. IDCPDF can describe a gesture image with one-dimensional histogram and reduce the computational complexity in features extraction and matching. Ref. [15] and Ref. [25] used distributional features of local contour points which are sample points and interest points respectively, and needed calculate relative distributions between the reference point and remaining sample points. In the process of matching, finding the corresponding points of two gesture shapes not only consumed more time, but also easily caused corresponding errors on the overall gesture shape.

The complexity in SC features extraction of sample points and interest points is O (N^2), and the complexity in matching is O (N^3) in Ref. [15] and Ref. [25]. In the proposed method, the complexity is O (N^3) in calculating inner distance, O (N) in establishing IDCPDF features and O (N) in matching. The proposed method reduces the calculation complexity in gesture feature extraction and matching process, has good robustness on gesture changes with inner-distance.

Ⅳ. CONCLUSIONS

Shape description is a hot topic in image analysis and has many potential applications. In this paper, the method based on inner distance contour point distribution features and histogram matching is proposed for gesture recognition. Firstly, segment and get the contour of gestures, followed by uniformly sample and IDCPDF extraction. Finally, three kinds of histogram matching methods Chi-Squared, Quadratic-Chi and EMD are used for recognition. Experiments on recognizing eight kinds of gestures show that the method has better accuracy and real-time performance in applications and certain robustness on deformations caused by part structures or finger joints.

Since this experiment is carried out in natural illumination and simple background, further research is also needed to aim at the segmentation algorithm about complex background, skin color and illumination changes, as well as gesture-based shape context feature extraction.

ACKNOWLEDGMENT

This work is supported by the National Natural Science Foundation of China (No. 61363078), the Natural Science Foundation of Gansu Province of China (No. 1212RJZA006). The authors would like to thank the anonymous reviewers for their helpful comments and suggestions.

REFERENCES

[1] Zhu H.M, Pun C.M, Lin C. “Robust hand gesture recognition using machine learning with positive and negative samples,” *Journal of Computers*, vol.8, no.7, pp.1831-1835, July 2013.
[2] Zhao F.Q, Li X.Y, Zhang Q.Y, et al. “Polar-radius-invariant-moment based on key-points for hand gesture shape recognition,” *Journal of Software*, vol.7, no.1, pp.88-94, January 2012.
[3] Rehg J.M, Kanade T. “Visual tracking of high dof articulated structures: an application to human hand tracking,” *Computer Vision—ECCV'94*. Springer Berlin Heidelberg, 1994: 35-46.
[4] Qi-jun Chen, Zhen-jiao Zhu, and Shuang Gu, “Hand gesture recognition based on Fourier descriptor and HMM,” *Control Engineering*, vol. 19, no. 4, pp. 634-638, 2012(*in Chinese*).
[5] Xingbao Meng, Jing Lin, and Yingchun Ding, “A robust method based on static hand gesture recognition for human-computer interaction under complex background,” *Proceedings of the Computer, Informatics, Cybernetics and Applications*, Hangzhou, pp. 907-914, 2012.

[6] A. Jinda-apiraksa, W. Pongstiensak, and T. Kondo, “A simple shape-based approach to hand gesture recognition,” *Proceedings of International Conference on Electrical Engineering/Electronics Computer Telecommunications and Information Technology*, Chiang Mai, pp. 851-855, May 2010.
[7] S. M. M. Roomi, R. J. Priya, and H. Jayalakshmi, “Hand gesture recognition for human-computer interaction,” *Journal of Computer Science*, Vol. 6, No. 9, pp.1002-1007, 2010.
[8] Yuan Shi, Qi Wei, Ruijie Liu, and Yuli Ge, “Simplified skeleton model based hand gesture recognition,” *Proceedings of the International Conference on Information Technology and Software Engineering-Software Engineering and Digital Media Technology*, Beijing, pp. 969-977, 2013.
[9] Bo Yang, Xiaona Song, Zhiquan Feng, and Xiaoyan Hao. “Gesture recognition in complex background based on distribution features of hand,” *Journal of Computer-Aided design and Computer Craphics,* vol. 22, no. 10, pp.1841-1848, 2010(*in Chinese*).
[10] M. Panwar. “Hand gesture recognition based on shape parameters,” *Proceedings of International Conference on Computing, Communication and Applications*, Dindigul, pp. 1-6, Feb 2012.
[11] N. Dardas, Q. Chen, N. D. Georganas, E. M. Petriu, “Hand gesture recognition using bag-of-features and multi-class support vector machine,” *Proceedings of IEEE International Symposium on Haptic Audio-Visual Environments and Games*, Phoenix, pp. 1-5, Oct 2010.
[12] Yuelong Chuang, Ling Chen, Gangqiang Zhao, Gencai Chen, “Hand posture recognition and tracking based on bag-of-words for human robot interaction,” *Proceedings of IEEE International Conference on Robotics and Automation*, Shanghai, pp. 538-543, May 2011.
[13] P. Gurjal and K. Kunnur, “Real time hand gesture recognition using SIFT,” *International Journal of Electronics and Electrical Engineering*, vol. 2, no. 3, pp. 19-33, Mar 2012.
[14] Yun Liu, Lifeng Zhang, and Shujun Zhang, “A hand gesture recognition method based on multi-feature fusion and template matching,” *Procedia Engineering*, vol. 29, pp. 1678-1684, 2012.
[15] L. Y. Deng, J. C. Hung, H. C. Keh, K. Y. Lin, Y. J. Liu, and N. C. Huang, “Real-time hand gesture recognition by shape context based matching and cost matrix,” *Journal of Networks*, vol. 6, no. 5, pp. 697-704, May 2011.
[16] S. Khandelwal, “Articulated hand posture recognition system using IDSC features,” *CS676: Computer Vision and Image Processing, Indian Institute of Technology*, Kanpur, 2011.
[17] S. Belongie, J. Malik, and J. Puzicha, “Shape matching and object recognition using shape contexts,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 4, pp. 509-522, Apr 2002.
[18] H. Ling and D. W. Jacobs, “Shape classification using the inner-distance,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 2, pp. 286-299, Feb 2007.
[19] R. L. Hsu, M. Abdel-Mottaleb, and A. K. Jain, “Face detection in color images,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24. no. 5, pp. 696-706, May 2002.
[20] Y. Rubner, C. Tomasi, L. J. Guibas, “The earth mover's distance as a metric for image retrieval,” *International Journal of Computer Vision*, vol. 40, no. 2, pp. 99-121, 2000.
[21] M. Varma and A. Zisserman, “A statistical approach to material classification using image patch exemplars,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 11, pp. 2032-2047, Nov 2009.
[22] J. Hafner, H. S. Sawhney, W. Equitz, M. Flickner, and W. Niblack, “Efficient color histogram indexing for quadratic form distance functions,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, no. 7, pp. 729-736, Jul 1995.
[23] O. Pele and M. Werman, “The quadratic-chi histogram distance family,” *Proceedings of the 11th European Conference on Computer Vision,* Crete, pp. 749-762, Sep 2010.
[24] O. Pele and M. Werman, “Fast and robust earth mover's distances,” *Proceedings of the 12th IEEE International Conference on Computer Vision*, Kyoto, pp. 460-467, 2009.
[25] Shufen Zhang, Liling Ma, and Junzheng Wang, “A framework of static hand shape recognition system,” *Proceedings of the International Conference on Information Technology and Software Engineering-Software Engineering and Digital Media Technology*, Beijing, pp. 577-585, 2013.

Qiu-yu Zhang is currently a researcher and doctoral supervisor at school of computer and communication of Lanzhou University of Technology. His main research interests contain network and information security, information hiding and steganalysis, image understanding and recognition.

Hui-yi Wei received the Bachelor’s degree in 2011. She is currently a postgraduate student at School of Computer and Communication of Lanzhou University of Technology. Her research interests include image understanding and recognition, pattern recognition.

Mo-yi Zhang was born in Gansu Province, China in 1985. She received the Master’s degree in 2010. She is currently a PHD student at School of Computer and Communication of Lanzhou University of Technology. Her research interests include pattern recognition and picture processing.

Zhi-gang Xu is currently a lecturer and PHD at School of Computer and Communication of Lanzhou University of Technology. His research interests contain target tracking, image understanding and recognition.

Hong-xiang Duan received the Master’s degree in 2011. She is currently a lecturer at School of Computer and Communication of Lanzhou University of Technology. Her research interests include target tracking and image processing.

Lu Lv was born in Shanxi Province, China in 1991. She received the Bachelor’s degree in 2012. She is currently a postgraduate student at School of Computer and Communication of Lanzhou University of Technology. Her research interests include pattern recognition and picture processing.