Gesture Detection and Recognition Fused with Multi-feature Based on Kinect

Haifeng Sang and Wei Li(✉)

School of Information Science and Engineering,
Shenyang University of Technology, Shenyang, China
785349306@qq.com

Abstract. Aiming at the high demand of the background environment and the user, this paper designs a real-time human static gesture recognition algorithm based on the depth information of Kinect. The localization of the joint points of the hand is realized by using the Kinect skeleton. The depth image is acquired by the depth sensor, and the joint points of the hand are tracked continuously; After locating the position of the hand, the region of interest is intercepted, and the depth threshold is set up to segment the hand from the depth image; The segmentation image is processed by morphology, and the circular rate, filling rate, perimeter rate, convex hull, convex defect, Hu moments of the hand contour are 9 kinds of features; Six kinds (0 to 5) of gesture are recognized using SVM method. Recognition rate and robustness of gesture recognition experiments are conducted in static and dynamic environment respectively. The experimental results show that the proposed algorithm can achieve better recognition result in a variety of environments.

Keywords: Kinect · Depth data · Multi-feature · SVM

1 Introduction

Gesture recognition is a kind of human-computer interaction technology. Compared with the traditional keyboard and mouse, it's more natural, intuitive and easy to learn.

The limit method is limited by wearing the logo with color or using the background with fixed color, so that we can segment the gesture from the special color. [1] This method reduces the freedom of the gesture. The color detection method based on the color space distributes the color of the image to the corresponding color space for the threshold separation. Using skin detection method can isolated skin color regions directly from the image, but current technology mainly gets following problems: easy to influenced by complex background and light conditions; can not overlap with the face and other position that are similar with skin color, can not wear non skin gloves etc.

Aiming at the high demand of the background environment and the user, the system locates the hand joint by using Kinect skeleton. Segmenting the hand area combined with the depth image. Then extracting the features of hand contour, recognizing the gesture by using SVM classifier. This method overcomes the influence of complex background and light changing efficiently. The user doesn't need to wear any

J. Yang et al. (Eds.): CCBR 2015, LNCS 9428, pp. 597–606, 2015.
DOI: 10.1007/978-3-319-25417-3_70

device to operate system. In this way, the system reflects the convenience of human-computer interaction. The experimental results show that the proposed algorithm can achieve better recognition result in a variety of environments.

2 Human Static Gesture Recognition System

2.1 Whole System

This paper proposes a hand gesture recognition system based on Kinect depth data. It's more efficient to take advantage of depth data for overcoming the influence of complex background and light changing. It's more convenient for user to operate the system because there is no need to wear any device. First the localization of the joint points of the hand is realized by using the Kinect skeleton. The depth image is acquired by the depth sensor, and the joint points of the hand are tracked continuously. After getting the three-dimensional coordinate of hand, the ROI is intercepted and the depth threshold is set up to segment the hand from the depth image; Extracting feature of the segmented gesture and recognizing the gestures by using the SVM classifier. Fig. 1 is the process of gesture recognition system.

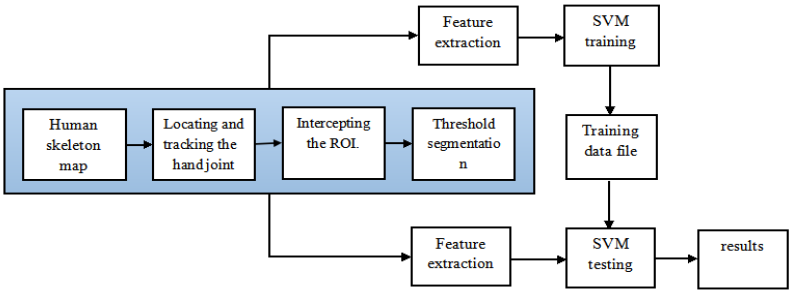

Fig. 1. Process of gesture recognition system.

2.2 Gesture Contour Acquisition Based on Depth Data

In this paper, we obtain the human skeleton map with Kinect. After hand location, we intercept a ROI of 140×140 pixels that contains the hand image. Then double threshold is set to segment the hand. In this way we can remove the influence of foreground and background efficiently. Fig. 2 is process of gesture acquisition.

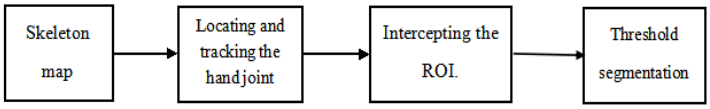

Fig. 2. Process of gesture acquisition.

2.2.1 Human Skeleton Map and Hand Locating and Tracking

Kinect can provide depth image using the depth camera. [2][3] Pixel of the image records the calibration depth. This depth camera can eliminate the background noise and extract the information of the people. Shotton labeled various parts of the characters' body in the picture. [4] Using large, rich and varied training data, he ensures the decision tree classifier assessing the various parts of the body is not misclassified no matter different individuals, clothing and posture. After the division of the body parts, the 3 dimensional position of the body joints can be predicted. [5] In the process of recognition, a depth map is required to identify the 3 dimensional positions of the individual joints, as shown in Fig. 3.

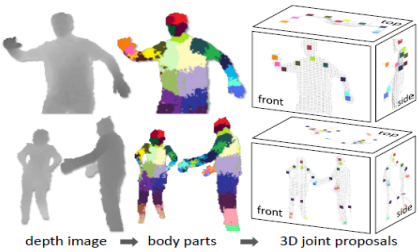

Fig. 3. Process of human body joints recognition.

After the human skeleton recognition, we map the three-dimensional coordinate on the depth image. Then obtaining the real-time 3D gesture information to achieve the gesture tracking.

2.2.2 Intercepting the ROI and Hand Segmentation

1) Intercepting the ROI

According to the coordinate X and Y of the hand, intercepting a 140×140 pixels (according to the experience) ROI which includes the hand image part. In this way we can prevent the influence of the object in the same depth plane.

2) Hand segmentation

① depth theory

The Z coordinates of the hand position represent the relative distance between the hand and the Kinect. The general depth value is represented by 12bit, that is, the maximum is 4095. If the depth data we collected need to be expanded to a gray-scale figure, it should multiplied by a factor 255/4095.

② threshold segmentation

The threshold range is T:

$$[Z \times (255/4095) - 5] < T < [Z \times (255/4095) + 5] \tag{1}$$

The contour of the hand is determined according to the depth information of the hand. Then binaryzation and filtering of the contour are processed. In Fig. 4, A is gesture segmentation sketch map and B is a gesture contour segmentation graph.

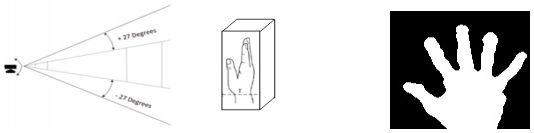

Fig. 4. A) gesture segmentation sketch map B) gesture contour segmentation graph.

2.3 Feature Extraction

After preprocessing, the binary image, edge image and contour matrix of gesture can be obtained. Commonly used feature extraction methods are Hu moment [6], Zernike moment [7], Fourier contour moment [8] and wavelet moment [9] and so on. In view of the characteristic of the image extraction, the features of the translation, rotation and scale invariance are selected. They are circular rate, filling rate, perimeter rate, convex hull, convex defect, Hu moments of the hand contour. These features can effectively reflect the characteristics of gesture images. Fig. 5 is the calculation of rectangle and circle of six kinds of hand contour.

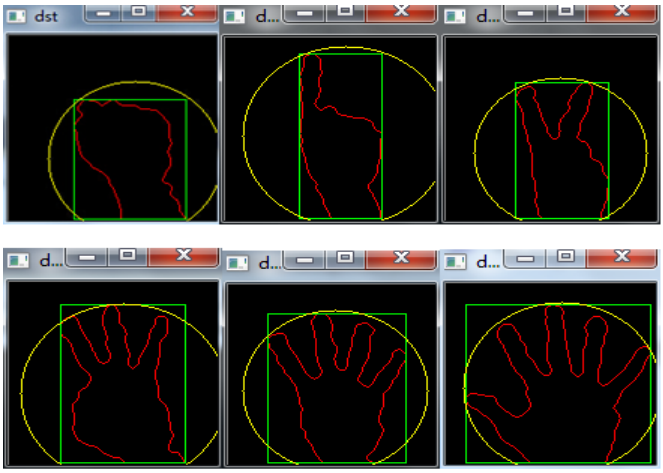

Fig. 5. Rectangle and circle of six kinds of hand contour.

2.3.1 Circular Rate, Filling Rate, Perimeter Rate

In this paper, we take the circular rate, the filling rate, the perimeter ratio of contour as features. Collecting 100 frames of each gesture with the method in 2.2 to analysis the feature. Table 1 is the average value of the circular ratio, filling ratio and perimeter ratio of 6 kinds of gestures.

Table 1. Circular ratio, filling ratio, perimeter ratio of 6 kinds of gestures.

	Zero	One	Two	Three	Four	Five
Circular ratio	1.66	2.75	3.49	4.25	5.33	5.94
Filling ratio	0.70	0.50	0.62	0.52	0.48	0.44
Perimeter ratio	0.73	0.83	1.09	1.17	1.45	1.61

1) Circular ratio

The circular ratio is the degree of the similarity between the contour and the circle. The ratio is close to 1, and the contour is close to the circle. L is the circumference of the contour and A is the area of the contour.

$$circularra\ \ tio = \frac{l^2}{4\pi A} \tag{2}$$

2) Filling ratio

The filling ratio is the ratio of the contour area to the rectangular area of the gesture. The larger the ratio, the more square the contour is. A is the contour area and R is the outer positive rectangular area.

$$fillingratio = \frac{A}{R} \tag{3}$$

3) Perimeter ratio

The perimeter ratio is the ratio of hand contour and perimeter of minimum circumscribed circle. The larger the ratio, the more open the gesture is. L is perimeter of the contour and C is the perimeter of the minimum circumscribed circle.

$$perimeterratio = \frac{l}{C} \tag{4}$$

2.3.2 Convex Hull and Convex Defect

The convex hull and the convex defect of contour are used to describe the shape of the object. The convex hull and the convex defect of the gesture can better reflect the characteristics and status of the the gesture. The main idea of gesture convex hull points and convex defect points extraction is using Douglas Peucker algorithm to simplify the contour. Draw the outline of the simplified contour after the treatment, and then calculate and draw out the convex hull and convex defect points of the contour. Fig. 6 is convex hull points and convex defect points of six kinds of hand contour.

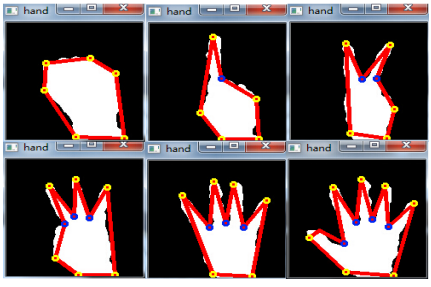

Fig. 6. Convex hull points and convex defect points of hand contour.

2.3.3 Hu Moment

In this paper, the Hu moment is chosen as the feature of gesture. The Hu matrix is a set of moment invariant which is composed of the nonlinear combination of moments. They well resolved the scale changing, image shifting, coordinate conversion and rotation changing during the process of feature matching.

7 invariant moments of Hu are composed of two order and three order central moment. The most useful information of the gesture image is contained in the low order moment. High order moments are complicated and noisy. In order to overcome the effects of noise and reduce the amount of computation, we choose the top four feature vector as the feature parameters of gesture image.

$$\phi_1 = \mu_{20} + \mu_{02} \tag{5}$$

$$\phi_2 = (\mu_{20} - \mu_{02})^2 + 4\mu_{11} \tag{6}$$

Using 2.2 method, gesture images of 100 frames are obtained. The average Hu moment parameter of each gesture is calculated. Table 2 is the average value of the first four invariant moments of the six gestures.

Table 2. The average value of the first four invariant moments.

	Zero	One	Two	Three	Four	Five
ϕ_1	0.1876	0.2683	0.2621	0.2376	0.2297	0.2207
ϕ_2	0.0076	0.0343	0.0256	0.0171	0.0098	0.0022
ϕ_3	0.0003	0.0056	0.0010	0.0007	0.0009	0.0008
ϕ_4	0.0001	0.0038	0.0021	0.0005	0.0002	0.0003

2.4 SVM Classifier

The SVM classifier is based on the theory of VC dimension and the least structural risk of statistical learning. In order to obtain the best generalization ability, SVM Seeks the best compromise between the complexity of the model and the ability to learn according to the limited sample information.

Considering the limitation of the sample and the accuracy of the recognition, this paper uses the SVM classifier to recognize the gesture.

1) In the training phase we extract the gesture contour using 2.2 method, and extract 100 frame images respectively for six different gestures.

2) Filter out the bad sample in the image and save the sample contains the complete gesture contour.

3) Extract features of the sample images and feature vector is sent to SVM training to construct the training parameters file.

4) In the testing phase we obtain real-time gesture contour and extract features by using the method in this paper. SVM classifier is used to test and then returns the recognition results.

Fig. 7 is the process of training and testing of SVM classifier.

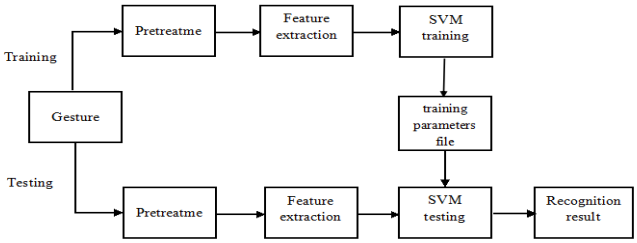

Fig. 7. Training and testing of SVM classifier

3 Experiment and Analysis

3.1 Gesture Recognition in Dynamic Environment

3.1.1 System Robustness Experiment

In addition to the normal environment, the system can still achieve good recognition results in a complex environment. According to the algorithm proposed in this paper, a series of experiments are carried out in the dynamic environment: the condition of complex background; the changing of environmental light conditions; gestures for rotation; the condition of multiple human interference. From Fig. 8 the correspondence of recognition results five and color image gesture, we can be see that the performance of the proposed algorithm is good in dynamic environment, and it shows strong robustness.

Fig. 8. Gesture recognition effect in dynamic environment.

3.1.2 Running Time Experiment

The recognition time of the system in this paper is tested in a complex background. The test set is 100 frames each gesture in SVM training stage. From the gesture detection to recognition, the average time of the system is 0.05s. Compared with other hand gesture recognition system [10][11][12], ours has a large advantage and is more suitable for online real-time gesture recognition applications.

Table 3. Comparison of system recognition time.

System	Paper [10]	Paper [11]	Paper [12]	This paper
Time	0.4s	0.09-0.11s	0.1333s	0.05s

3.2 Gesture Recognition in Static Environment

3.2.1 Fixed Distance Gesture Recognition Experiment

In the static environment, gesture recognition experiment is carried out in a complex background with same light condition. The official recommended distance of Kinect camera is between 1.2 meters to 3.8 meters. Experiments were conducted by 10 operators at a fixed distance of 1.5 m. The operator makes 0-5 six gestures 10 times respectively (guaranteed gestures are within the range) to calculate the recognition rate of gestures by using the proposed algorithm. Table 4 is the average recognition rate for each gesture in a static environment.

Table 4. The average recognition rate for each gesture.

Gesture	Zero	One	Two	Three	Four	Five
Recognition rate	97%	93%	96%	94%	94%	98%

From the table 4, the recognition rate of gesture zero and gesture five is higher and other hand gesture recognition rate is relatively low. The analysis is:

1) Human gesture is a flexible object and there is not a standard for gestures. The recognition error is caused by the randomness of the operator's finger.

2) The features of gesture zero and gesture five are more obvious than other gestures and the recognition rate is higher than other gestures.

3) Gesture recognition is carried out by using depth image. The operator need to locate the palm facing the camera in order to acquire a completed gesture image. When the operator’s gestures have bias, the recognition rate will be affected.

3.2.2 Changing Distance Gesture Recognition Experiment

Gesture recognition experiments are conducted at different distances. From 0.5 meters, 10 operators test 5 times every 0.25 meters for the same gesture. Experimental results are shown in Fig. 9. The results show that when the distance between the camera and operator is 1.2-1.5 meters, recognition rate is the highest. When the distance is less than 0.5 meters, the system cannot obtain complete gesture image to recognize. When the distance is larger than 2 meters. The depth resolution is lower so that some information of the hand is lost and the recognition rate drops.

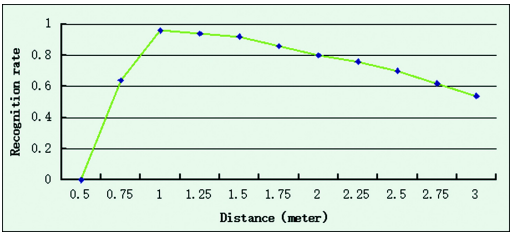

Fig. 9. Recognition results of different distance.

4 Conclusion

This method overcomes the influence of complex background and light changing efficiently. The experimental results show that the proposed algorithm can achieve better recognition result in a variety of environments.

At the same time, this system has some disadvantages, and it is also the research direction in the future.

1) Enrich the static gesture library by defining more gestures to improve the type of static gestures;

2) Try other feature extraction methods and classification. Different features, classification and recognition methods will get different recognition effect. Try new combination to improve the recognition rate;

3) Add gesture control function. Using the static gesture to control picture zoom, the start and stop of program and the game control. Enhance the practicality.

References

1. Yang, B., Sun, X.N., Fang, Z.Q.: Gesture Recognition in Complex Background Based on Distribution Features of Hand. Journal of Computer-Aided Design & Computer Graphics **22**, 1841–1848 (2010)
2. Sungil, K., Annah, R., Hyunki, H.: Using depth and skin color for hand gesture classification. In: IEEE International Conference, Las Vegas, pp. 155–156 (2011)
3. He, J.G., Shao, L., Xiao, D., Jamie, S.: Enhanced Computer Vision with Microsoft Kinect Sensor: A Review. IEEE Transactions on Cybernetics **43**(5) (2013)
4. Jamie, S., Andrew, F., Mat, C.: Real-time human pose recognition in parts from single depth images. In: 2011 IEEE Conference on Computer Vision and Pattern Recognition, Colorado Springs, pp. 1297–1304 (2011)
5. Georg, H., Rod, M., Wolfgang, B.: Lightweight palm and finger tracking for real-time 3D gesture control. In: IEEE Virtual Reality, Singapore, pp. 19–23 (2011)
6. Guo, J.Y., Zhang, Y.W.: Face Recognition Based on Moment Invariants and Neural Networks. Computer Engineering and Applications **7**(2008)
7. Teague, M.R.: Image analysis via the general theory of moments. Optical Society of America. **70**(8), 920–930 (2010)
8. Li, Y., Tong, X.L., Liang, C.Q.: Hand Gesture Recognition Based on Fourier Descriptors with Complex Backgrounds. Computer Simulation **22**(12), 158–161 (2005)

9. Song, C.D., Ali, M.S.: Face recognition using complex wavelet moments. Optics and Laser Technology **47** (2013)
10. Chung, W., Wu, X., Xu, Y.: A real-time hand gesture recognition based on Haar wavelet representation. In: Proc. IEEE Int. Conf. Robot. Biomimetics, pp. 336–341 (2009)
11. Fang, Y., Wang, K., Cheng, J., Lu, H.: A real-time hand gesture recognition method. In: Proc. IEEE Int. Conf. Multimedia Expo, pp. 995–998(2007)
12. Yun, L., Peng, Z.: An automatic hand gesture recognition system based on viola-Jones method and SVMs. In: Proc. 2nd Int. Workshop Comput. Sci. Eng, pp. 72–76(2009)