Hand Gesture Recognition and Its Application in Robot Control

Pei-Guo Wu[1] and Qing-Hu Meng[2]

[1] Information Engineering College, Henan University of Science and Technology, Luoyang, 471023, China

[2] Department of Electronic Engineering, The Chinese University of Hong Kong, Hong Kong, 00852, China

Abstract

In view of the problem that the accuracy and robustness of hand gesture recognition technology based on vision in the process of using gestures to interact with robots were unstable due to its background, illumination and other factors, this paper presented a gesture segmentation and recognition method Combining depth information and color images. First, it used Kinect sensor to obtain depth information and color images, then used depth information to pick the hand part from color images, and then got gesture images through color segmentation method. Second, it calculated HU invariant moments and shape features of the gesture images as feature information. Finally, it used the feature information to train the support vector machine, then implemented hand gesture recognition for static hand gestures. Experimental results show that the method has strong robustness to the influence of background interference, illumination variation, translation, rotation and zoom, and can be applied to control intelligent robot.

Keywords: *Kinect; hand gesture recognition; Hu invariant moments; support vector machine; robot.*

1. Introduction

With the development of the robot technology, robots have entered every aspect of human life. Robots can replace human to work in the factory, can provide entertainment and help to people in daily life. As a result, interactions between humans and robots are becoming more and more frequent. In this case, the traditional human-computer interaction technologies based on mouse and keyboard has already could not satisfy people's needs, people are longing for a more simple, natural and direct way to communicate with the robots. As a common communication way between people, hand gestures can convey rich and complete information, and play an important role in our daily life. The use of hand gestures to control robots can make people manipulate and interact with the robots easily and conveniently. Therefore, hand gesture recognition is an indispensable key technology in the new generation of human-computer interaction technologies [1-3].

Initially, people used data glove to collect the data of hand joints for hand gesture recognition. But the hand gesture recognition Technology based on data glove is not suitable for practical application, because the data glove is expensive and not flexible for user. Now hand gesture recognition Technology based on vision is widely researched and applied, it obtains images with camera, and used image processing methods for processing and analysis, then accomplishes the purpose of hand gesture recognition. Compared with the hand gesture recognition based on data glove, hand gesture recognition based on vision is more natural and convenient, and has a great application value. In the complex environment conditions, images obtained by camera are easily affected by complex background and illumination variation, that makes it difficult for image processing and analysis. Therefore, the accuracy of hand gesture recognition is difficult to improve [4-6]. In this case, this paper presented a hand gesture recognition method based on Kinect, it used Kinect to obtain depth information and color images at the same time. The depth information is only determined by the distance between object and Kinect, it has strong robustness to the influence of complex background. The method combined depth information with color images for image segmentation, then extracted feature information of the hand images after segmentation, and used the feature information to train support vector machine to get classifier, and then used the classifier to recognize hand gestures. Finally, the method was used in robot controlling to achieve natural and convenient human-computer interaction.

2. Hand Gesture Segmentation

Kinect is a 3d sensor which is released by Microsoft Corporation. On the Kinect, there is a color (RGB) camera, an infrared (IR) projector, and an infrared (IR) sensor. With the RGB camera, color images of the scene in front of the Kinect can be obtained immediately. The IR projector can emit a grid of near infrared (NIR) light in

doi:10.20943/IJCSI-201602-1015

front of it. When the NIR light irradiates objects with rough surfaces, according to the distance of the objects, there will be some different patterns on the surface of the objects. The IR sensor is a CMOS camera that can receive the patterns. Then the patterns are decoded in the Kinect to determine the depth information [7-8].

2.1 Foreground Extraction

In this paper, the frame rate of Kinect was set to 30 Hz, and the resolution of the image was set to 640 × 480 pixels. With the Kinect, color images and depth information could be obtained at the same time. It is the value of distance from Kinect to the objects that is stored in each frame of depth information. Then the foreground in the range of 0.5 meters to 1.2 meters was retrieved from color images based on depth information. In this way, the images of hand without any other part of the experimenter's body could be obtained; see Fig. 1.

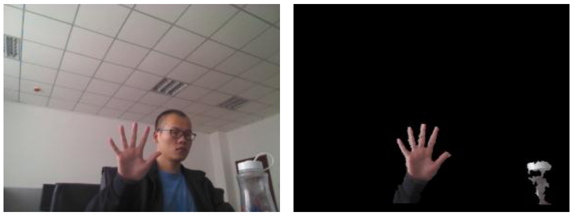
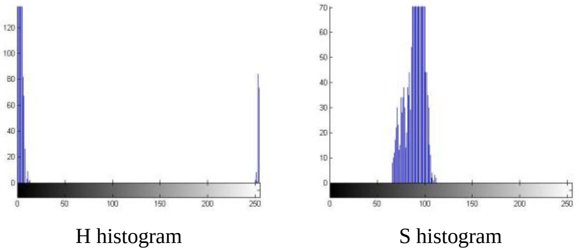


Fig. 1 Foreground extraction.

2.2 Skin Color Segmentation

The hand images which were obtain from Kinect is a RGB (Red, Green, Blue) color image. In RGB color space, the color of each pixel is defined by R, G and B values. The R, G and B values have a high correlation, and are easily affected by illumination variation. As a result, it is not suitable for color segmentation in RGB color space. Therefore, the hand images were transformed from RGB color space to HSV color space. The definition of color in HSV color space accords with the character of human vision perception very much. In HSV (Hue, Saturation, Value) color space, H and S values are used to describe the different colors; V value reflects the light and dark of different colors because of light intensity. Hence, color segmentation based on HSV color space can eliminate the influence of light intensity [9].

Firstly, the hand images in RGB color space were transformed to HSV color space. Then, the color histograms of hand were extracted; see Fig. 2. After analysis of the color histograms, the threshold was determined as follow: $0 \le H \le 15$ or $170 \le H \le 180$, and $50 \le S \le 130$. The values of pixels within the range of threshold were replaced with the value of 1, the other pixels were replaced with the value of 0. Finally, binary images were achieved after the color segmentation of hand images.



H histogram

S histogram

Fig. 2 Hand color histograms.

After skin color segmentation, there was some noise in the binary images. As a result, median filtering was necessary before extracting features. Median filter is a kind of nonlinear spatial filters, it provides excellent noise-reduction capabilities for certain types of random noise, with considerably less blurring than linear smoothing filters of similar size. In order to perform median filtering at a point in an image, the values of the pixel in question and its neighbors must be sorted at first, then determine their median, and assign this value to that pixel. Fig. 3 shows the result of median filtering.

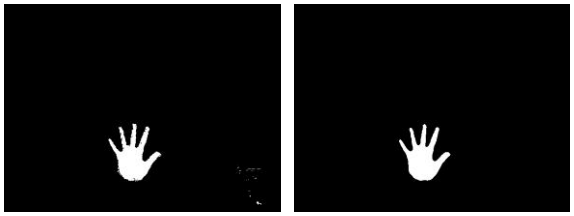


Fig. 3 Median filtering.

3. Feature Extraction

Fig. 4 shows the binary images of hand gestures which are defined in Sebastien Marcel Static Hand Posture Database that were obtained through Kinect. Then the HU invariant moments and compactness of the images were calculated as feature information for hand gesture recognition.

3.1 Hu Invariant Moments

In image processing, moments are usually used to describe geometrical features of images, and are used as the basis of classification. If the resolution of the image is m × n and

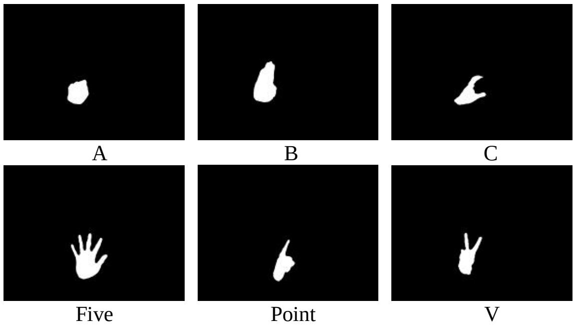


Fig. 4 Binary images of hand gestures.

$F(x, y)$ is the value at point (x, y), then the (p + q) th moments m_{pq} and the (p + q) th central moments μ_{pq} can be calculated as follows:

$$m_{pq} = \sum_{x=1}^{m}\sum_{y=1}^{n} x^p y^q F(x, y) \quad p, q = 0,1,2,... \tag{1}$$

$$\mu_{pq} = \sum_{x=1}^{m}\sum_{y=1}^{n} (x - x_0)^p (y - y_0)^q F(x, y) \tag{2}$$

Where (x_0, y_0) is the center of gravity in image, $x_0 = \frac{m_{10}}{m_{00}}$, $y_0 = \frac{m_{01}}{m_{00}}$.

Moments and central moments can be used as shape feature information for images classification, but they are not invariant of transformation, rotation and scale at the same time. In 1962, HU M.K. put forward the concept of moment invariant, also called Hu invariant moments. Hu invariant moments are invariant of translation, rotation and zoom, therefore they are widely used in image classification and recognition [10]. Hu invariant moments are constructed with normalized second central moments and third central moments, the normalized central moments can be calculated by zero moment as the following equation:

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^{\frac{p+q+2}{2}}} \tag{3}$$

HU invariant moments are calculated as follows:

$$\phi_1 = \eta_{20} + \eta_{02} \tag{4}$$

$$\phi_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11} \tag{5}$$

$$\phi_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \tag{6}$$

$$\phi_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \tag{7}$$

$$\phi_5 = (\eta_{03} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + 3\eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \tag{8}$$

$$\phi_6 = (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \tag{9}$$

$$\phi_7 = (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + (3\eta_{21} - \eta_{30})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \tag{10}$$

3.2 Compactness

The same hand gestures in different images were not identical because of gesture polymorphism and the influence of the camera angle. Therefore, the Hu invariant moments of images changed slightly. As a result, the accuracy of hand gesture recognition was reduced. In order to raise the accuracy, two shape features based on compactness were introduced [11-12], they are calculated as follows:

$$c_1 = \frac{l^2}{4\pi \times S} \tag{11}$$

$$c_2 = \frac{S}{S_{rect}} \tag{12}$$

Where l is the circumference of gesture contour, S is the area of gesture, S_{rect} is the area of minimum enclosing rectangle.

4. Hand Gesture Recognition

SVM was served as classifier for hand gesture recognition. The first three parameters in Hu invariant moments and the two shape features were used as feature information to train SVM, and then the classifier is obtained.

SVM is first put forward by Cortes and Vapnik in 1995, and has become a new method of machine learning. SVM is based on VC dimension theory and structure risk minimization principle, it excels in addressing high-dimension and solving small sample size problem. SVM performs classification and recognition by finding optimal hyperplanes in the feature space. When samples are linear separable, the optimal hyperplanes can be find directly. For the linear non-separable problems, it can be solved by introducing slack variables. In fact, most of the questions are nonlinear, they needs to be transformed into linear problems in a high dimension space by kernel functions to solve [13-14], three common kernel functions are as follows:

1. Polynomial kernel function:

$$K(u,v) = (\gamma u^T v + c)^d \tag{13}$$

2. RBF kernel function:

$$K(u,v) = \exp(-\gamma \| u - v \|^2) \tag{14}$$

3. Sigmoid kernel function:

$$K(u,v) = \tanh(\gamma u^T v + c) \tag{15}$$

RBF kernel function was used in the method, it has good stability and wide convergence domain, and has a wider application range than the others. In the experiment, in addition to the parameters in kernel function, the penalty parameter C is equally important. The value of C determines how important the off-group points data is. The bigger the value of C is, the more important the off-group points data is.

5. Results and Analysis

5.1 Experimental Research

The hardware equipment of hand gesture recognition includes computer and Kinect. The operating system of computer is Windows 7; the driver software of Kinect is Kinect for Windows SDK v1.8. The software developing platform is Visual Studio 2010. Both Open Source Computer Vision Library (OpenCV) and LIBSVM tools are used for software development.

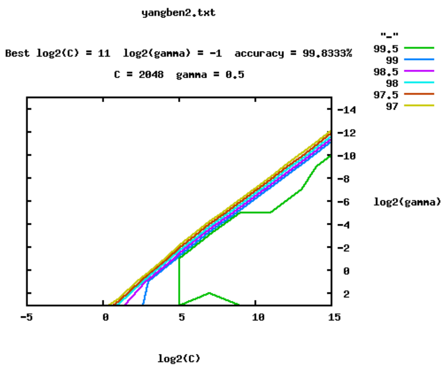


Fig. 5 Parameters optimization.

With the Kinect, one hundred and sixty images were obtained randomly for every hand gesture in Sebastien Marcel Static Hand Posture Database. One hundred of the images were used as training samples, the others were used as test samples. The training samples were used to extract feature information, and the feature information were used for parameters optimization with "grid.by" which is included in the LIBSVM tools. Fig. 5 shows the result of parameters optimization. Then the optimal parameters and feature information were used to train SVM. The test samples were used for hand gesture recognition after feature extracting. Table 1 shows the results of hand gesture recognition, and the suggested method is more accurate than the method which only use Hu invariant moments as feature information.

Table 1: Results of hand gesture recognition

Hand Gesture	***Recognition Rate(%)***	
	Our Method	***Hu Method***
A	100	98.33
B	98.33	96.67
C	96.67	96.67
Five	100	100
Point	100	100
V	100	98.33
Average	99.17	98.33

5.2 Robustness Analysis

In the complex environment conditions, complex background and illumination variation make it difficult for hand gesture segmentation. As a result, the accuracy of hand gesture recognition is reduced. In Fig. 6, a and d are the hand gesture segmentation in dim light, b and e are the hand gesture segmentation in bright light, c and f are the hand gesture segmentation in complex background. From Fig. 6, the hand gesture segmentation method is stable to the influence of illumination change and complex background, and has strong robustness. In the experiment, every hand image which was used to do hand gesture

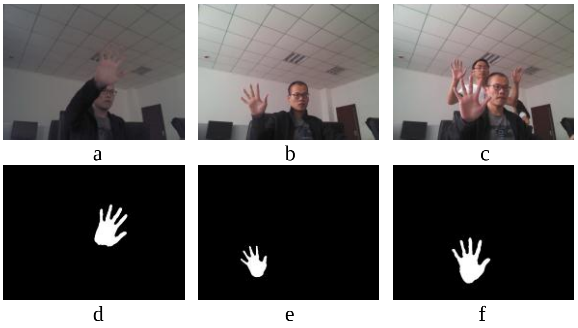


Fig. 6 Hand gesture segmentation in different condition.

recognition is different, they have some changes of translation, rotation and zoom. Combined with the table 1, we can see that the hand gesture recognition method has strong robustness to translation, rotation and zoom.

5.3 Robot Control

NAO was used in this experiment, it is a programmable humanoid robot developed by Aldebaran Robotics, a French robotics company. NAO has 25 degrees of freedom, and has many kind of sensors, it is widely used in scientific research and family service. In the robot control system, hand gestures (a, b, c, five, point, v) were defined as control commands (go, turn left, turn right, sit down, stand up, say hello) to the robot. The system used Kinect to obtain hand images, then do hand gesture recognition. According to the results of hand gesture recognition, the defined control commands were executed. Fig. 7 shows the work flow of robot control system. The results show that the system can be used to control the motion of the robot.

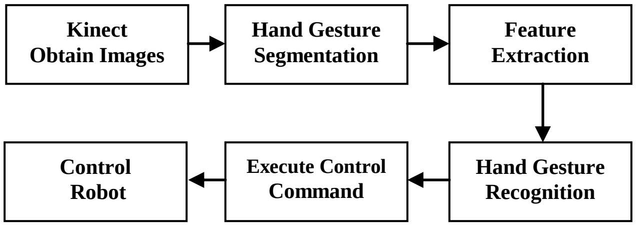


Fig. 7 Work flow of robot control system.

6. Conclusions

In this paper, we present a hand gesture recognition method based on Kinect. The method uses Kinect to obtain hand images, then performs skin color segmentation and feature extraction to the images, and uses SVM as classifier. Experimental results show that the method has strong robustness to the influence of background interference, illumination variation, translation, rotation and zoom, and can be used to control robot. Future work will focus on the research of dynamic gesture recognition and its application.

References

[1] H. Hasan, S. Abdul-Kareem, "Human-computer interaction using vision-based hand gesture recognition systems: a survey", Neural Computing and Applications, Vol. 25, No. 2, 2014, pp. 251-261.

[2] Y. Y. Pang, N. A. Ismail, and P. L. S. Gilbert, "A Real Time Vision-Based Hand Gesture Interaction", in the 4th Asia Modelling Symposium (AMS2010), 2010, pp. 237-242.

[3] M. Takahashi, M. Fujii, M. Naemura, et al., "Human gesture recognition system for TV viewing using time-of-flight camera", Multimedia tools and applications, Vol. 62, No. 3, 2013, pp. 761-783.

[4] REN Hai-bing, ZHU Yuan-xin, and XU Guang-you, et al., "Vision-Based Recognition of Hand Gestures: A Survey", Acta Electronica Sinica, Vol. 28, No. 2, 2000, pp. 118-121.

[5] WU Yu, "Gesture Recognition Simulation in Robot Vision Communication", Computer Simulation, Vol. 32, No. 2, 2015, pp. 405-408.

[6] Y. Han, "A low-cost visual motion data glove as an input device to interpret human hand gestures", IEEE Transactions on Consumer Electronics, Vol. 56, No. 2, 2010, pp. 501-509.

[7] YU tao, Kinect Development and Application in Actual Combat, Beijing: China Machine Press, 2013.

[8] MENG Shang, GAO Chenqiang, and YANG Luyu, "Method of Gesture Recognition Based on Depth Image", Digital Communication, Vol. 41, No. 2, 2014, pp. 22-26.

[9] P. Kakumanu, S. Makrogiannis, and N. Bourbakis, "A survey of skin-color modeling and detection methods", Pattern Recognition, Vol. 40, No. 3, 2007, pp. 1106-1122.

[10] LUO Yuan, XIE Yu, and ZHANG Yi, "Design and Implementation of a Gesture-Driven System for Intelligent Wheelchairs Based on the Kinect Sensor", Robot, Vol. 34, No. 1, 2012, pp. 110-113, 119.

[11] A. Jinda-apiraksa, W. Pongstiensak, and T. Kondo, "A simple shape-based approach to hand gesture recognition", in the 2010 International Conference on Electrical Engineering/Electronics Computer Telecommunications and Information Technology (ECTI-CON), 2010, pp. 851-855.

[12] CHEN tao, "Research of Real-Time Gesture Recognition Technology Based on Depth and Color Information ", M.S. thesis, School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, China, 2013.

[13] F. Bellakhdhar, K. Loukil, and M. Abid, "Face recognition approach using Gabor Wavelets, PCA and SVM", International Journal of Computer Science Issues, Vol. 10, No. 2, 2013, pp. 201-207.

[14] DING Shi-fei, QI Bing-juan, and TAN Hong-yan, "An Overview on Theory and Algorithm of Support Vector Machines", Journal of University of Electronic Science and Technology of China, Vol. 40, No. 1, 2011, pp. 1-10.

Pei-Guo Wu male, born in 1991, is a master candidate in computer application technology at the Information Engineering College, Henan University of Science and Technology, Luoyang, China. His research interests include image processing and recognition.

Qing-Hu Meng male, born in 1962, received the Master's degree from Beijing Institute of Technology, Beijing, China, in 1988, and the Ph.D. degree in electrical and computer engineering from the University of Victoria, BC, Canada, in 1992.
He was a Professor in the Department of Electrical and Computer Engineering at the University of Alberta, Canada, from April 1994 to August 2004. Currently, he is a Professor with the Department of Electronic Engineering, Chinese University of Hong Kong. His research interests are in the areas of biomedical engineering, medical and surgical robotics, active capsule endoscopy, medical

image-based automatic diagnosis, interactive telemedicine and telehealthcare, biosensors and multisensor data fusion, bio-MEMS with medical applications, biomedical devices and robotic assistive technologies and prosthetics, adaptive and intelligent systems, and related medical and industrial applications.

doi:10.20943/IJCSI-201602-1015