Hand Detecting and Positioning Based on Depth Image of Kinect Sensor

Van Bang Le, Anh Tu Nguyen, and Yu Zhu

***Abstract*—This paper presents a new method of detection and identification, called PYTHON programming environment, which can realize the gesture track recognition based on the depth image information get by the Kinect sensor. First, Kinect sensor is used to obtain depth image information. Then it extracts splith and with the official Microsoft SDK. Finally, this paper presents how to calculate the palm center's coordinates based on the moment of hand contour feature. Experiments show that the advantages of using the hand split and gesture recognition of the Kinect's depth image can be very effective to achieve interactive features.**

***Index Terms*—Kinect, depth image, palm center, moment, gesture recognition.**

I. Introduction

As the human hand has unique features, gesture recognition is very conducive to achieve human-computer interaction. Currently, this is a very hot research. The traditional gesture recognition based on color image has gained great achievements, but, it can't describe the depth information of the gesture since it lacks the three dimension parameters. However, the depths chart to make up for this shortcoming has been widely applied for the study of the gesture trajectory tracking, and many scholars have used this method's results. The scholars are the following: Zhou [1] described and identified the palm and fingers relative distance algorithm, Tao [2], with KNN method, used it for fingers recognition, and Wu [3] described the characteristics of gestures using HOG features of the image

In order to improve the recognition rate and its application effect, this study uses the depth information collected by Kinect to achieve precise hand detection and gesture trajectory tracking. Microsoft launched Kinect [4] in 2010. Kinect is part of the Xbox 360 and Windows video game control platform developed by Microsoft. The Infrared range combined with an RGB camera on the device and the voice-activated system could remotely controlled without contact, including human motion gesture detection. At the same time, Microsoft simultaneously launched a software development kit for Kinect support such as C++, Python, C# and other programming environments.

Scholars are really interested to use Kinect, and the research about the acquisition of depth image using the gesture features also shows great results. Zhang [5] tested Kinect with a hidden Markov model (HMM) in order to achieve gesture trajectory recognition and Wu [6] tested the skeleton point cloud.

This study is about the use of image processing library of Pykinect SDK and of OpenCV to process Kinect depth image and get accurate hand segmentation. According to the relationship between the gray level of the depth image and the actual distance, we can detect the hand contour portion, and then calculate the center of gravity to find out the position of the hand. This method of segmentation process is much simpler than the traditional visual inspection methods. In the palm of source it creates a hand coordinate system, and a demonstration of a simple gesture trajectory tracking instance.

II. Depth Image of the Kinect

Fig. 1 shows a Kinect with the cover removed [7], included: infrared projector, infrared camera, RGB camera and microphones.

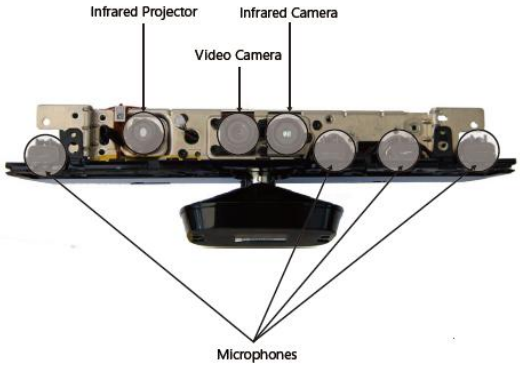

Fig. 1. A Kinect sensor unwrapped.

When using Python, the software bring in Pythoneasy_install function can be conveniently installed device development kit pykinectto collect images. Infrared camera to detect environment space depth output for 640×480 resolution video (low frame rate can be up to 1280×1024 resolution).When an Xbox 360 gaming platform is connected, the available range of 1.2 ~ 3.5 meters, the maximum range reaches six square meters. Kinect sensor field is design with 57 °level angle, 43 °vertical angle, and 27 °of elevation.

Data from the Kinect sensors was shown in Fig. 2. In depth image, the distance between the body and IR camera is proportional to the grayscale. Kinect for windows device recently shows that the distance is 40 cm. Infrared rays on an object's surface reflection to the receiver to form a three-dimensional coordinates of an object point cloud. So

Manuscript received December 14, 2013; revised February 17, 2014. This work was supported in part by the National Science Foundation of China (No. 61370174, No. 61271349) and the Fundamental Research Funds for the Central Universities (No. WH1114030, WH1214015).

The authors are with the East China University of Science and Technology, Shanghai, China (e-mail: banglevan@gmail.com, 378587599@qq.com, zhuyu@ecust.edu.cn; corresponding author: Yu Zhu).

each frame of video has slightly different contour information, and can't solve the problem of shade, and it also, for some special materials or special surface structure, will produce a lot of noise.

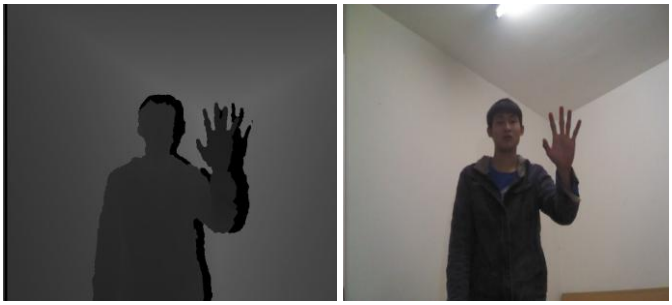

Fig. 2. Sampling data of Kinect. (a) Depth image, (b) RGB image.

Kinect depth image of map formats: an unsigned 16-bit 1 channel (grayscale) image, among them low 12 bits is effectively information. In order to use OpenCV for processing, we will convert them to 8 bits image. Schematic of the hand detection processing of the original image:


Fig. 3. Original image.

The depth of the scenario was described in the original image, in the Fig. 3 the actual distance has been converted into grayscale. Relation of depth map of grayscale and the actual distance [8]:

$$d = K \times \tan\left(d_{gray}/2842.5 + 1.1863\right) - \theta \qquad (1)$$

In equation (1), d as actual distance (unit by met), d_{gray} as grayscale of image, K =0.1236m, $\theta = 0.037$m.

III. Hand Extraction and Palm Positioning

The depth image grayscale distribution is regular, generally uses gestures to realize human-computer interact, and hand is in front of the body. Thus infrared camera outputs image, the hand of gray level is lower than the gray level of the human body's, through the gray image binarization process can be easily extracted the hand from depth image. Hand regional monitoring to extract the implementation process as shown in the figure below:

This paper, using the convenience of Python development environment and OpenCV image process library to achieved human hand segmentation. In the process shown in Fig. 4, for the purpose of hand positioning, through depth image preprocessing to optimize the process of extract fingers and palm center.

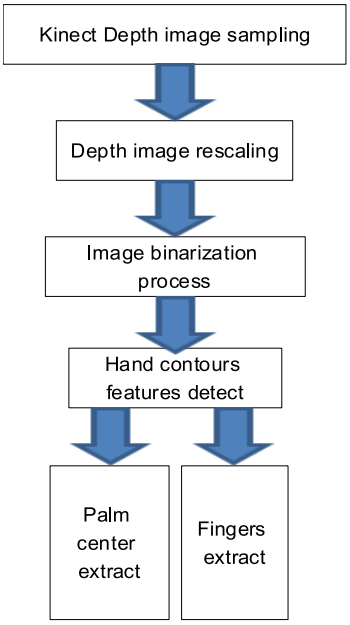

Fig. 4. Hand segmentation extraction process.

A. Hand Detection

According the distance between the hand and the IR camera, use depth image selective display, to achieve transformation equation[9, 10]:

$$\mathrm{dst}(I) = \mathrm{src}(I) + (\mathrm{shift}_0, \mathrm{shift}_1, \ldots) \qquad (2)$$

This study sets that the grayscale display range is 1 ~ 45, the shadow of infrared imaging noise grayscale still is 0, and the original image grayscale greater than 45 points changed to 255.Binarization process math model is:

$$g(x,y) = \begin{cases} 0 & if\ f(x,y) > thresh \\ f(x,y) & others \end{cases} \qquad (3)$$

Using (3) destination image realize binarization process:

$$h(x,y) = \begin{cases} value & if\ g(x,y) > thresh \\ 0 & others \end{cases} \qquad (4)$$

In the (4) equation, value is, *value* is the first set of values.

Fig. 6 represents the hand detection based on depth image extraction process:

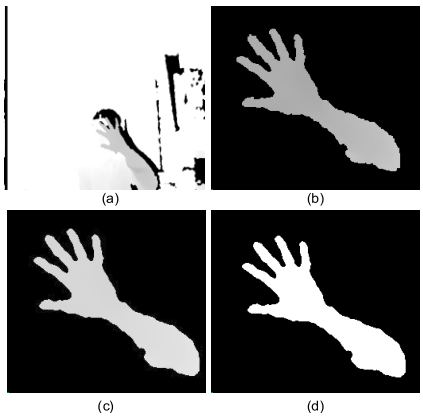

Fig. 5. Hand detect process. (a) Depth image rescaling, (b) Back ground eliminate, (c) Median filter, (d) Image binarization process.

As it can be seen from the Fig. 5, compared with the color image segmentation method to extraction of hand based on depth image effect is much better. Because it has the characteristics of fast segmentation, accurate extraction, handsegmentation based on depth image is a great advantage.

B. Palm Center Positioning

In the process of gesture recognition, palm recognition is a very important step. In this process the palm is a key reference point, whether it is the peak of the fingers or the pits, they are all around to palm center. It can be viewed as "hand coordinate system" source point which can realize the gesture recognition. Implement of hand gesture recognition, Chen [11] used smart circle fitting positioning the palm center, but in this method calculation accuracy is depend on contour defects and its pits. Palm as a hand model coordinate system of the source point so need higher objectivity, to overcome the disadvantages, this paper proposes a palm center detection method based on determination of the contour features moment and positioning the center of gravity of the hand (palm center coordinate).

Spatial moments of image are computed as:

$$m_{ij} = \sum_{x,y}(f(x,y).x^j.y^i) \quad (5)$$

The central moments:

$$mu_{ij} = \sum_{x,y}(f(x,y).(x-\overline{x})^j.(y-\overline{y})^i) \quad (6)$$

where $(\overline{x}, \overline{y})$is the mass center:

$$\overline{x} = m_{10}/m_{00}, \overline{y} = m_{01}/m_{00} \quad (7)$$

This method ensures the objectivity of palm center coordinates, it can also improve accuracy and reduce the calculation time.

IV. Experimental Result

A. Hand Extract

By using the depth image of the Kinect sensor, hand extraction task becomes easily and quickly. Fig. 6 shows the hand contours features detection process.

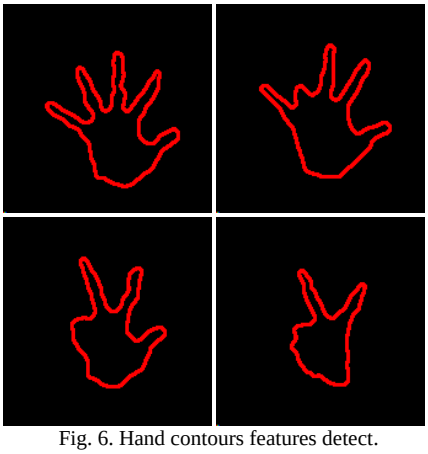

Fig. 6. Hand contours features detect.

The hand has very unique geometrical shapes. Human hands, expressing our intension, are often used for communication [12]. The most important step of gesture trajectory recognition is to calculate the key points of the hand region from contour information's.

B. Palm Center Coordinate Detect

Smart circle fitting method palm coordinates, it has more disadvantages such as a low precision, unstable, and large amount of calculation. This method depends on contour feature pits. When these pits number and position changed, the calculation results will be greatly changed too. Experimental proof, when the hand, wrist - arm cathetus line is overlap, under different hand gestures used contour features moment find the palm center coordinate has very steady positioning results, and little time used in the calculation.

TABLE I: The Comparison of the Processing Time Per Frame (Time Unit by MS)

Circle fitting method	Contour moment method
0.0731791	0.0177147
0.0501991	0.0118246
0.0754548	0.0122709

In the Table I, there is a comparison of the two methods of calculation time under the same frame. For the one used in this study, there are very significant optimization effect; its processing time is quadruple to quintuple of the circle fitting method.

On Fig. 7, shows the palm center position process used gestures of Fig. 6. The right column is results of circle fitting method, and the left is result of contour moment method. Unlike circular intelligent fitting method, the method used in this study can also determine the location of the palm accurately, in the case of a small number of key points. No matter the defects of the size or the number in changes, the contour of the coordinates would not have large fluctuations.

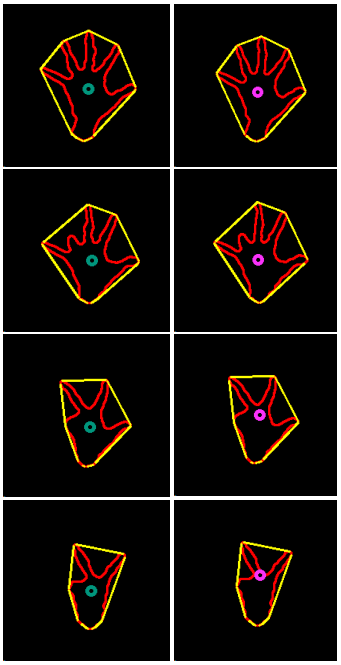

Fig. 7. Comparison of the two methods for the same frames.

The images in the left column show the results of using contours moment detect method proposed in our paper (blue

circles are the calculated palm center coordinates), and the images in the right column are using fit circle method (pink circle are palm center coordinate).

When there are multiple pits, the program has to found three key points, this algorithm is implemented in multiple loops, which leads to the calculation time (As shown in Table I). At this time if using Python-OpenCV programming environment, calculation results closely depend on the number of these key points and its location. When two different gestures appeared in succession, the palm center will be change in the larger. The differences with this paper, it will be calculated according to the outline, so based contour moment method is very accurate, stable, and small amount of calculation.

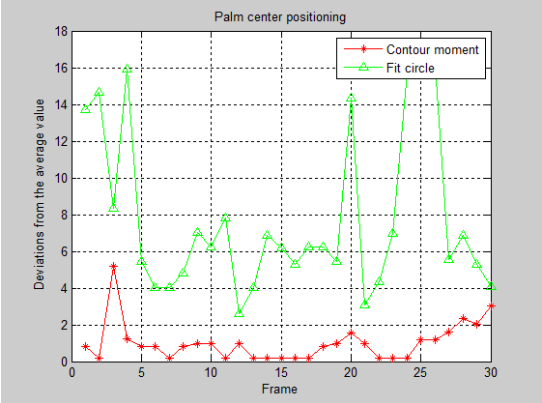

Fig. 8. Compare of the result of the stability under the same gesture.

In Fig. 8 used the first gesture on Fig. 6, in the thirty consecutive frames, if use the circle fitting method, deviation of positioning point to the average coordinates reached 14 pixel, while the method for this paper only reaches 5 pixel. Since it is easy to see, because it is based on the contour moment method detection, the palm center is very reliable.

V. Conclusion

This paper presents a method of hand segmentation based on Kinect sensor depth image. When the Kinect infrared camera output imaging is 1 channel video, the gray level of images and the distance between body-camera have a proportional relationship. With this feature, this article is under OPENCV-PYTHON programming environment. This paper puts forward a palm center detected based on contours moment. The proposed method has many advantages such as: high precision, high stability, small computation time and the influence of the key points is very small.

In the future, with the depth image, we can achieve detection segmentation of the hand accurately. And we can also combine with Kinect RGB images to create a 3D gesture recognition model.

References

[1] Z. Ren and J. S. Yuan, "Robust part-based hand gesture recognition using kinect sensor," *IEEE Transactions On Multimedia*, vol. 15, pp. 1110-1120, May, August 2013.
[2] H. Y. Tao and Y. L. Yu, "Finger tracking and gesture interaction with kinect," in *Proc. IEEE 12th International Conference on Computer and Information Technology*, pp. 214-218, 2012.
[3] X. Y. Wu, C. Yang, Y. W. Wang, H. Li, and S. M. Xu, "An intelligent interactive system based on hand gesture recognition algorithm and kinect," *Fifth International Symposium on Computational Intelligence and Design*, 2012.
[4] Kinect for Windows document. [Online]. Available: http://en.wikipedia.org/wiki/Kinect
[5] Y. Zhang, S. Zhang, Y. Luo, and X. D. Xu, "Gesture trajectory identification and application based Kinect depth image," *Application research of computer*, 2012, vol. 29, no. 9.
[6] S. Wu, F. Jiang, and D. B. Zhao, "Hand gesture recognition based on skeleton of point clouds", in *Proc. 2012 IEEE fifth International Conference on Advanced Computational Intelligence (ICACI)*, pp. 566-569, October 2012.
[7] R. Miles, "Learn the kinect api," *O'Reilly Media, Inc*. 2012.
[8] J. L. Raheja, A. Chaudhary, and K. Singal, "Tracking of fingertips andcentres of palm using Kinect," presented at 2011 Third International Conference on Computational Intelligence, Modelling & Simulation.
[9] Opencv documentation. [Online]. Available: http://docs.opencv.org
[10] G. Bradski and A. Kaehler, "Learning OpenCV," *O'Reilly Media , Inc*. 2008.
[11] W. Z. Chen, "Real-time palm tracking and hand gesture estimation based on fore-arm contour," M. S. Thesis, Dept. Inform. Eng., National Taiwan University of Science and Technology, 2011.
[12] R. Fujiki, D. Arita, and R. I. Taniguchi, "Real-Time 3D hand shape estimation based on inverse kinematics and physical constraints," presented at 13th International Conference, Cagliari, Italy, September 6-8, 2005.

Van Bang Le is currently a master degree student in Signal and Information process at East China University of Science and Technology, Shanghai, China. Before that, he received the B.Eng. in the optical engineering from Nanjing University of Science and Technology, Nanjing, China, in 2012.

His research interests include computer vision, 3D image reconstruction, machine learning and the human-computer interaction.

Anh Tu Nguyen is currently a master degree student in signal and information process at East China University of Science and Technology, Shanghai, China. Before that, he received the B.Eng. in the optical engineering in Nanjing University of Science and Technology, Nanjing, China, in 2013.

His research interests include computer vision, image processing, and software development.

Yu Zhu received the Ph.D. degree in optical engineering from Nanjing University of Science and Technology, China, in 1999. She is currently a professor, Ph.D. supervisor in the department of electrical and communication engineering of East China University of Science and Technology.

Her research interest is image processing, computer vision and multimedia communication. She has published more than 50 papers in journals and conferences.