The Design and Implementation of Natural Human-Robot Interaction System Based on Kinect Sensor

Fengli Ma
College of Electrical Engineering
Zhejiang University
Hangzhou, China
mfl84089770cx@163.com

Hao Wang
Robot Association of Zhejiang University
Zhejiang University
Hangzhou, China
wanghao@zju.edu.cn

Zhifeng Sun
College of Electrical Engineering
Zhejiang University
Hangzhou, China
eeszf@zju.edu.cn

Abstract—In this paper, researchers propose a method to build a natural human-robot interaction system based on Kinect sensor. Taking the advantage of skeleton tracking technology, researchers can easily get the depth data from Kinect sensor and capture body movements, which greatly simplify the recognition algorithms. After establishing the 3D coordinates of human joints, the rotation angles of human waist, shoulder, elbow and wrist can be calculated according to the spatial geometry, and these joints correspond with the joints of the robot arm one to one. Further, researchers need to extract the palm characteristic information to recognize hand gesture based on Hu Moment. Finally, the results of identification will be converted to commands and sent to controller via Bluetooth to manipulate the robot arm. Experiments show that by this method, researchers can effectively control the rotation of the 5-DOF robot arm to grab object. As the relationship between people and robots becomes more and more closely, this kind of human-robot interaction interface will play an important role in the future development of robot.

Keywords- Kinect; Skeleton Tracking; Depth Data; Robot Arm; Hu Moment

I. INTRODUCTION

In recent years, robots in many fields have been able to replace people to accomplish specific jobs, or even do better than human in efficiency, accuracy and stability, which has significantly liberated human and accelerated the development of the society. Actually, robots have moved away from industrial settings and walked into the life, and the question is how to interact with these machines [1]. In the various robot research fields, the topic about interaction between human and robots has gathered great interest [2]. As traditional human-robot interaction system can no longer satisfy people's requirements, researchers need to find other ways to interact with robot more naturally. Fortunately, gesture recognition is a good choice, and it has been an active technology in computer vision and pattern recognition [3].

Traditional gesture recognition technology is mainly based on RGB image which contains a lot of methods and algorithms. For example, Mahmoud et al. had developed a system that could recognize gesture for alphabets from hand motion using Hidden Markov Model (HMM) [4]. Chen et al. proposed a approach that implements the posture recognition with Haar-like features and the AdaBoost learning algorithm [5]. Ghotkar et al. introduced a hand segmentation technique for hand gesture recognition [6].

This paper presents a method to interact with the robot and manipulate it to simulate the movement of human skeleton and hand gesture. Researchers can get the depth data from Kinect sensor and take the advantage of skeleton tracking technology with the help of SimpleOpenNI library. After establishing the 3D coordinates of human joints, the joint rotation angles can be easily calculated according to the spatial geometry. In order to recognize hand gesture, researchers need to extract the palm characteristic information based on Hu Moment to match the real-time images. And the results of identification will be converted to command and sent to controller via Bluetooth to manipulate the robot arm. Researchers build a robot platform with a 5-DOF robot arm, results show it can be controlled effectively by human skeleton movement and hand gesture.

II. OVERVIEW OF SKELETON TRACKING TECHNOLOGY

"You are the controller", it is the declaration of Kinect sensor. As an electronic product of Microsoft, it has caught the eye of human after its launch. Kinect is so powerful because of its advanced image processing algorithms, unlike traditional methods, and it collects depth data through a technology called light coding. In particular, using depth camera offers several advantages over traditional intensity sensors, as they can work in low light levels with color and texture invariant [7], so it is robust to light intensity and complicated background. The architecture of human-robot interaction system is shown in Figure 1. This kind of Human-Robot Interaction interface

allows amateur users or those with disabilities to control robotic systems [8].

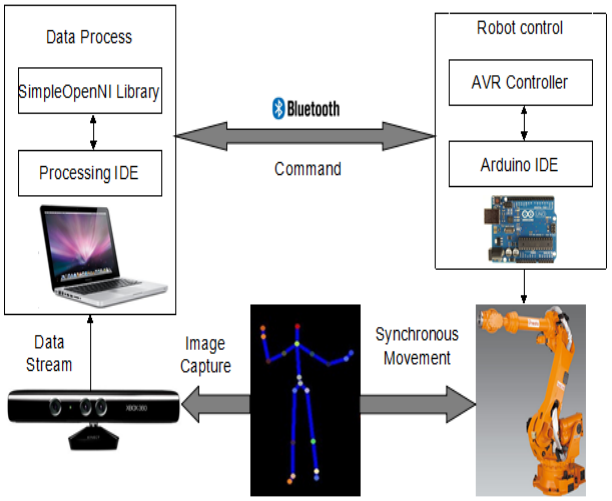

Figure 1. System architecture

Kinect sensor has a strong skeleton tracking ability to give the 3D coordinates of human joints and simplify image processing effectively. Actually, skeleton tracking technology uses machine learning algorithm which requires a large amount of data. So in the early days, people with various body heights, weights and skin colors provided vast amounts of training data. The original system needs to analyze millions of images, extract all kinds of characteristic value and finally establish a decision tree to match different parts of human body. The skeleton tracking model of Kinect is shown Figure 2.

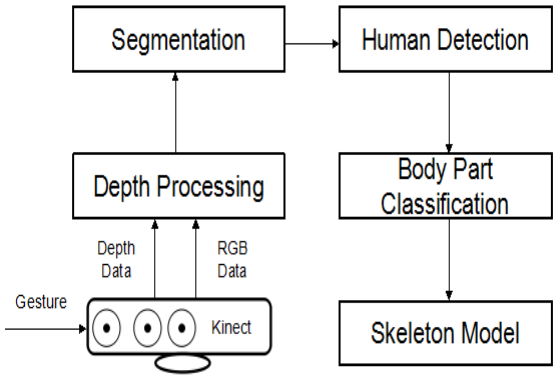

Figure 2. Skeleton tracking model of Kinect

III. CALCULATION OF HUMAN JOINT ANGLE

Kinect skeleton data frames offer 20 joints of human body. Mapping relationship between human skeleton joints and the robot arm in the space coordinates is shown in Figure 3.

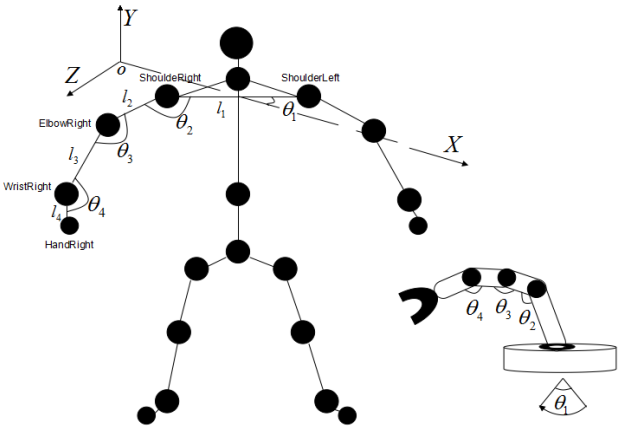

Figure 3. Mapping relationship between human skeleton joints and the robot arm

In this paper, researchers use right hand to manipulate the robot arm, with the aim to calculate the rotation angles of waist, shoulder, elbow and wrist, researchers need to get the 3D coordinates of ShoulderLeft, ShoulderRight, ElbowRight, WristRight and HandRight, and other joints can be used as an auxiliary condition.

A. *Rotation Angle of Waist*

To get the rotation angle of waist, researchers can make use of two human skeleton joints: ShoulderLeft and ShoulderRight. Assuming that the 3D coordinate of ShoulderLeft is (x_1, y_1, z_1), ShoulderRight is (x_2, y_2, z_2). These two points form a straight line l_1 in space. Because Y-axis coordinate remains unchanged when waist is rotating, so just consider the XOZ plane, then the slope of l_1 is K_0, calculated as follows:

$$K_0 = \frac{z_1 - z_2}{x_2 - x_1}, x_1 \neq x_2. \quad (1)$$

The angle between l_1 and X-axis calculated as follows:

$$\theta_1 = \arctan K_0. \quad (2)$$

B. *Rotation Angle of Shoulder*

To get the rotation angle of shoulder, researchers can make use of three human skeleton joints: ShoulderLeft, ShoulderRight and ElbowRight. Assuming that the 3D coordinate of ElbowRight is (x_3, y_3, z_3). ShoulderRight and ElbowRight form a straight line l_2 in space. Because Z-axis coordinate remains unchanged when Shoulder is rotating, so just consider the XOY plane, then the slope of l_1 is K_1, calculated as follows:

$$K_1 = \frac{y_2 - y_1}{x_2 - x_1}, x_1 \neq x_2. \quad (3)$$

The slope of l_2 is K_2, calculated as follows:

$$K_2 = \frac{y_3 - y_2}{x_3 - x_2}, x_2 \neq x_3. \quad (4)$$

The angle between l_1 and l_2, calculated as follows:

$$\theta_2 = \arctan\left|\frac{K_2 - K_1}{1 + K_1 K_2}\right|, K_1 K_2 \neq -1. \quad (5)$$

C. Rotation Angle of Elbow

As described above, researchers can get θ_3, calculated as follows:

$$\theta_3 = \arctan\left|\frac{K_3 - K_2}{1 + K_2 K_3}\right|, K_2 K_3 \neq -1. \quad (6)$$

D. Rotation Angle of Wrist

As described above, researchers can get θ_4, calculated as follows:

$$\theta_4 = \arctan\left|\frac{K_4 - K_3}{1 + K_3 K_4}\right|, K_3 K_4 \neq -1. \quad (7)$$

IV. GESTURE RECOGNITION BASED ON HU MOMENT

A. Image Segmentation Based on Distance

To facilitate the processing of the image, depth data should be converted to 3D point cloud. Assuming the human hand is nearest from the origin point, according to K-nearest neighbors algorithm, researchers need to select k nearest points as feature points. If k is set too small, the mapping will not reflect any global properties. If k is set too high, the mapping will lose its nonlinear character [9]. So the crux in the locally linear embedding algorithm is the selection of the number of nearest neighbors [10]. Through several experiments, researchers propose 2000 as the value of k to get a good segmentation performance.

B. Details of the Recognition Algorithm Based on Hu Moment

There are two main methods in the field of image recognition as follows: Template Matching and Artificial Neural Networks. The former one gives high image recognition accuracy, but when rotating and scaling the image, its recognition efficiency is a little low. The latter one has a strong classification ability to resist noise, and it's widely used in the identification of a static image but performs a great amount of calculation. This article focuses on gesture recognition based on Hu Moment. Researchers capture the hand gesture using Kinect sensor to transmit information of the shape, position and movement of human hand [11].

In 1962, Hu developed the theory of normalized central moment with the characteristic of translation, rotation and scaling invariance. Seven invariant moments are constructed as follows:

$$\varphi_1 = \eta_{20} + \eta_{02}. \quad (8)$$

$$\varphi_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2. \quad (9)$$

$$\varphi_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2. \quad (10)$$

$$\varphi_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2. \quad (11)$$

$$\begin{aligned}\varphi_5 = {} & (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12}) \\ & \cdot\left[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2\right] \\ & + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03}) \\ & \cdot\left[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2\right].\end{aligned} \quad (12)$$

$$\begin{aligned}\varphi_6 = {} & (\eta_{20} - \eta_{02})\left[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2\right] \\ & + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}).\end{aligned} \quad (13)$$

$$\begin{aligned}\varphi_7 = {} & (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12}) \\ & \cdot\left[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2\right] \\ & - (\eta_{30} - 3\eta_{12})(\eta_{21} + \eta_{03}) \\ & \cdot\left[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2\right].\end{aligned} \quad (14)$$

The normalized central moment η_{pq} is defined as:

$$\eta_{pq} = \mu_{pq} / \mu_{00}^r. \quad (15)$$

Where

$$r = (p + q + 2)/2, p, q = 0, 1, 2.... \quad (16)$$

For depth distribution $f(x, y)$ of the image, (x, y) represents pixel location. And $(p+q)$ order central moment is defined as:

$$\mu_{pq} = \sum_{m=1}^{M}\sum_{n=1}^{N}(x - \bar{x})^p (y - \bar{y})^q f(x, y). \quad (17)$$

Where

$$\bar{x} = m_{10} / m_{00}, \bar{y} = m_{01} / m_{00}. \quad (18)$$

$$m_{pq} = \sum_{m=1}^{M}\sum_{n=1}^{N} x^p y^q f(x, y). \quad (19)$$

Extracting gesture features based on Hu moment is an important method for image recognition and matching. Depth feature vectors contain most of the information for specific gesture matching. To make full use of the whole information, researchers consider all of the Hu moments as feature vectors to describe the characteristics of the gesture image that need to be recognized. The feature space is defined as: $(\varphi_1, \varphi_2, \varphi_3, \varphi_4, \varphi_5, \varphi_6, \varphi_7)$. Euclidean distance is always used to measure the similarity of different vectors, here researchers calculate Euclidean distance between the standard template vector and the test image vector to match the predefined gesture.

Subsequent experiments proves that this algorithm is suitable for describing the overall shape of the target and has obvious advantages in the field of pattern recognition and image matching, etc.

V. ANALYSIS AND DISCUSSION ON EXPERIMENTAL RESULTS

A. *Verification of Human Skeleton Recognition Accuracy*

This section presents the results of human skeleton recognition. Researchers use different skeleton rotation angles to demonstrate the validity of the approach proposed in this paper. Figure 4 shows the skeleton tracking results which contain skeleton and depth images. In the experiments, 500 samples with a variety of rotations (from 0° to 90°) are captured and tested. The rotation recognition rate of ShoulderRight in the same environment is shown in Table 1.

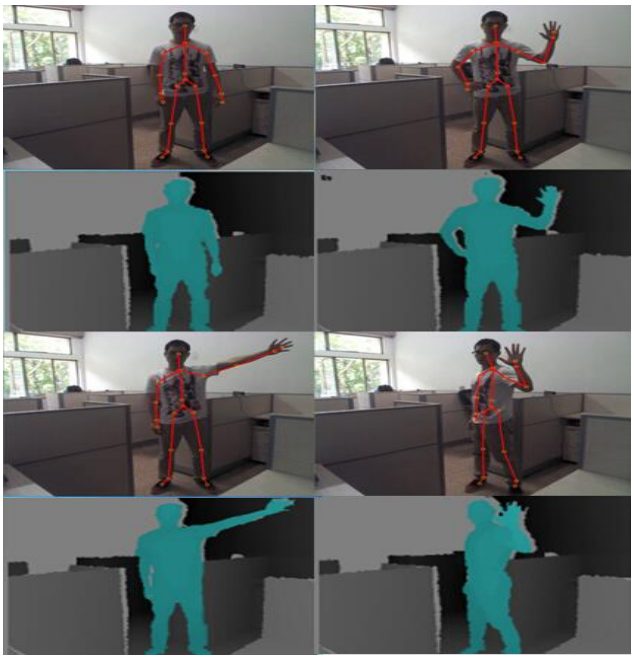

Figure 4. Skeleton tracking results

Experiments indicate that researchers get 451 correct recognition results and the average recognition rate is 90.2%, which is similar to the traditional identification technology with a good robustness. Researchers can also discover that smaller rotation angles of human skeleton easily cause errors in judgment because of the interference of the upper arm and lower arm.

TABLE I. RECOGNITION RATE

Rotation angle	Correct times	Error times	Recognition rate
0°	85	15	85%
30°	88	12	88%
45°	90	10	90%
60°	92	8	92%
90°	96	4	96%

B. *Experiment Based on Hand Gesture Recognition*

We define two kinds of hand gestures: Grab gesture and Release gesture. In the test, 200 sample gestures are captured, 100 of which are treated as training set and another 100 are treated as testing set. Figure 5 shows the gesture recognition and hand segmentation results.

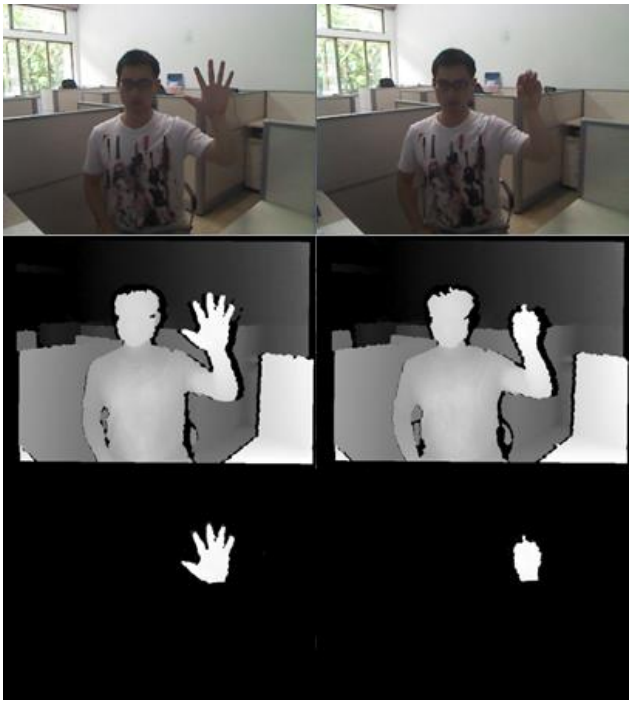

Figure 5. Gesture recognition and hand segmentation results

VI. THE IMPLEMENTATION OF CONTROL SYSTEM

In order to establish the control system, researchers need a Kinect sensor, a computer with Windows 7 operating system, a robot platform with a 5-DOF robot arm, AVR controller and a couple of Bluetooth modules. The system uses Processing and Arduino IDE with SimpleOpenNI library and OpenCV library. After correctly configuring the system, Kinect can be driven successfully and becomes available.

Now consider the working principle as follows. First of all, Kinect sensor captures the human skeleton movement and hand gesture using skeleton tracking technology. Then Processing IDE calculates the human joint angle according to the spatial geometry and extracts the palm characteristic information based on Hu Moment to match the predefined gesture template. And the results of identification will be converted to commands and sent to AVR controller via Bluetooth. After receiving the control instructions, AVR controller outputs the signals to drive the motors and manipulate the robot arm to finish the assigned work. The results researchers obtained demonstrate that researchers can effectively control the rotation of the robot arm to grab object. Figure 6 shows the results of object grab experiment.

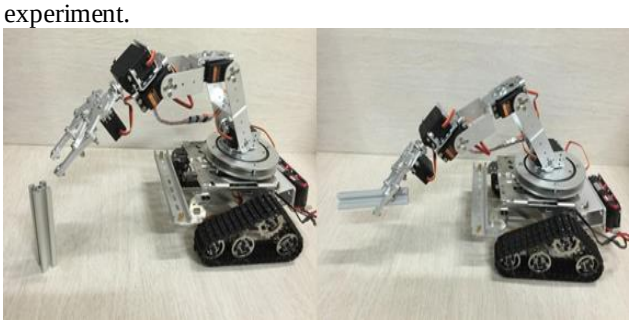

Figure 6. The results of object grab experiment

VII. CONCLUSIONS

This paper proposes a method to build a natural human-robot interaction system. Researchers use Kinect sensor to capture body movements and recognize hand gesture based on Hu Moment. Experiments show that by this method, researchers can effectively control the target with strong robustness.

In the future, researchers hope to improve the control precision and speed of the robot arm and develop the recognition algorithms for more complex applications, researchers will combine voice and face recognition technique to enrich the interactive way with robots.

ACKNOWLEDGMENT

This research for Natural Human-Robot Interaction System project was supported by Intelligent Robot Laboratory, Zhejiang University.

We thank the many skilled engineers in Kinect, particularly Robert Craig, Craig Peeper and Ryan Geiss, who built the Kinect tracking system on top of this research. Researchers also thank my friends for providing their test data.

REFERENCES

[1] Van D B M, Carton D, De Nijs R, "Real-time 3D hand gesture interaction with a robot for understanding directions from humans,"RO-MAN, IEEE, 2011.

[2] Kao M C, Li T S, "Design and implementation of interaction system between Humanoid robot and human hand gesture,"Proceedings of Sice Annual Conference, 2010, pp.1616-1621.

[3] Lin H, Zhao M, "A Fast Algorithm for Hand Gesture Recognition Using Relief,"Fuzzy Systems and Knowledge Discovery, 2009. FSKD "09. Sixth International Conference on, 2009.

[4] Elmezain M, Al-Hamadi A, "Gesture Recognition for Alphabets from Hand Motion Trajectory Using Hidden Markov Models,"Signal Processing and Information Technology, 2007 IEEE International Symposium on, 2008.

[5] Chen Q, Georganas N D, Petriu E M, "Real-time Vision-based Hand Gesture Recognition Using Haar-like Features,"Conference Record - IEEE Instrumentation and Measurement Technology Conference, 2007.

[6] Ghotkar A S, Kharate G K, "Hand Segmentation Techniques to Hand Gesture Recognition for Natural Human Computer Interaction,"International Journal of Human-Computer Interaction, March 2012.

[7] Shotton J, Fitzgibbon A, Cook M, "Real-time human pose recognition in parts from single depth images," in CVPR,March 2011.

[8] Yang H D, Park A Y, Lee S W, "Human-Robot Interaction by Whole Body Gesture Spotting and Recognition,"Proceedings of the 18th International Conference on Pattern Recognition - Volume 04, 2006.

[9] Jolliffe I T, Principal Component Analysis, 2nd ed., J.appl.met, 2002.

[10] Lvarez-Meza A, Valencia-Aguirre J, Daza-Santacoloma G, "Global and local choice of the number of nearest neighbors in locally linear embedding,"Pattern Recognition Letters, 2011, 32(16) , pp.2171-2177.

[11] Kofman J, Wu S V X, "Robot-Manipulator Teleoperation by Markerless Vision-Based HandArm Tracking,"International Journal of Optomechatronics, 2007, 1(3), pp.331-357.