Developing a Gesture Based Remote Human-Robot Interaction System Using Kinect

Kun Qian[1], Jie Niu[2] and Hong Yang[1]

[1]*School of Automation, Southeast University, Nanjing, China*
[2]*School of Electrical and Electronic Engineering,*
Changzhou College of Information Technology, Changzhou, China
kqian@seu.edu.cn

Abstract

Gesture based natural human-robot interface is an important function of robot teleoperation system. Such interface not only enables users to manipulate a remote robot by demonstration, but also ensures user-friendly interaction and software reusability in developing a networked robot system. In this paper, an application of gesture-based remote human-robot interaction is proposed using a Kinect sensor. The gesture recognition method combines depth information with traditional Camshift tracking algorithm by using Kinect and employs HMM in dynamic gesture classification. A Client/Server structured robot teleoperation application system is developed, which provides a favorable function of remotely controlling a dual-arm robot by gestural commands. Experiment results validate the practicability and effectiveness of the application system.

Keywords: *Kinect, Gesture, Robot, Teleoperation, Human-robot Interaction*

1. Introduction

One key long-term goal of developing remote robot control system via local network and Internet requires a human-friendly interface that transfers information of guidance or other types or commands. In the field of gesture recognition and robotics, many studies have been carried out on adapting gesture as an ideal communication interface in the Human-Robot Interaction (HRI) context.

In the last decade, several methods and potential applications in the advanced gesture interfaces for HRI or HCI have been suggested. Yoon [1] develops a hand gesture system in which combination of location, angle and velocity is used for the recognition. Liu [2] develops a system to recognize 26 alphabets by using different HMM topologies. Computational models such as Neural Network [3], Hidden Markov Model (HMM) [4] and Fuzzy Systems have been popular for gesture classification.

Besides from traditional monocular or stereo vision system [5] for building gesture interface, low-cost Kinect sensor [6, 7] has recently become an attractive alternative to expensive laser scanners in robotic applications. The Kinect's depth sensor consists of an infrared projector that emits a constant pattern and an infrared camera that measures the disparity between the observed pattern and a prerecorded image at a known constant depth. The output consists of an image of scaled disparity values in Kinect disparity units.

In this paper, a remote robot control system is described that utilizes Kinect based gesture recognition as human-robot interface. The system achieves synchronized robot arm control that generates human-mimic actions, which is a fundamental function to build a teleoperation application.

2. Gesture Recognition as Human-robot Interface

2.1. Camshift based Hand Tracking

In this paper, we define four basic patterns of gestures which are designed as robot control commands. These gestures are *WaveLeft, WaveRight, RiseUp,* and *PutDown*. An interactive hand tracking strategy combing color data and depth data is designed. Firstly, a person waves his/her hand in front of a Kinect sensor. Then, a motion detector in color camera between successive frames is applied to locate the hand, which is taken as the initial tracking target. After the initialization, Camshift tracking algorithm is employed for real-time hand tracking. Since it is well known that Camshift-based tracking algorithm using only color information may produce tracking error when there is similar color in background. We introduce depth information to facilitate hand tracking according to the fact that hand is usually separated from the background object in depth, and has fixed moving range. Hence threshold segmentation in depth map can accurately distinguish the player from the background. The tracking algorithm extracts trajectory of a single hand movement, denoted as $X = \{X_1, \cdots, X_n\}$, in which $X_t = (x_t, y_t)$ is the two dimensional spatial coordinate at time t. Example of tracking results are shown in Figure 1.

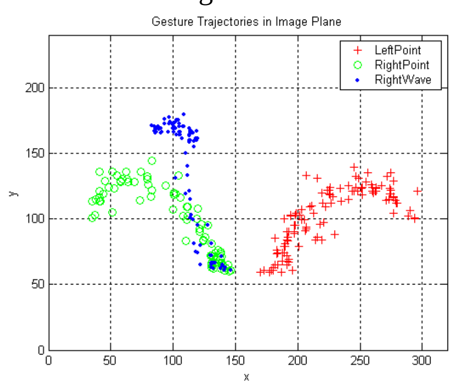


Figure 1. Tracking Trajectories of Three Types of Gestures

2.2. Hand Motion Trajectory Recognition

Hidden Markov models are used to model the underlying processes of gestures and determine the underlying processes behind the individual components of a gesture.

Firstly, the obtained gesture trajectories are converted into 16 direction codes representing direction vectors. Each angle is converted into one of the sixteen direction codes. The angle ranges of the direction codes have different widths. The feature of 2D motion trajectory of hand gesture is comprised of a series of discrete movement points, and it is represented by a series of discrete movement direction value. For the 2D motion plane, we divide the directions into eight discrete values. Therefore, the trajectory of dynamic gesture ban be described by the sequence of discrete direction value $d_1, d_2, ..., d_{16}$. This transforms the original state sequence $X = \{X_1, \cdots, X_n\}$ into an encoded observation sequence $O = \{O_1, \cdots, O_n\}$.

Then HMM is employed to model and classify the gesture played by the user. We use a Left-Right structure HMM and take 3 to 4 different states to model the abovementioned gestures. As an example, Figure 2 shows the basic structure of the HMM model for representing the gestures used in our work.

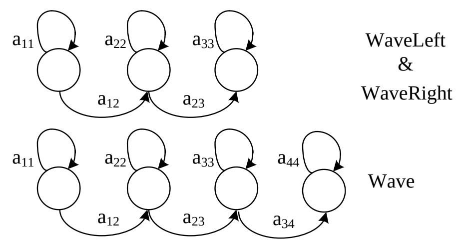


Figure 2. HMM Structure for Gesture Trajectory Recognition

Baum-Welch algorithm is employed to train $\lambda_i(A_i, B_i, \pi_i)$, the computational the Hidden Markov model for each gesture pattern. Regular HMM based recognition algorithm computes the similarity metric $P(O \mid \lambda_i)_{i=1,\cdots,N}$ of observed gesture to each gesture model of $\lambda_1, i = 1,....,N$. The pattern with highest similarity is considered as the classification result. Based on this, we define a probabilistic confidence measure. Denote that the confidence measure of gesture pattern i is C_i, which is computed as follow:

$$C_i = \sum_{k \neq i} \frac{\log(P(O \mid \lambda_k))}{\log(P(O \mid \lambda_i))}$$

$$i = \arg\max_{k \neq i} \{C_k\}$$

We also define and adaptive threshold $\sum_{k \neq i} C_k$ for each i. Only in case that $C_i > \sum_{k \neq i} C_k$ can gesture pattern i be considered as the gesture recognition result. For four candidate gesture patterns $G = \{WaveLeft, WaveRight, RaiseUp, PutDown\}$, an example of being classified as $WaveRight$ gesture is computed as follow:

$$\Pr(G = WR) = \begin{cases} P(O \mid \lambda_{WR}), if\,((C_{WR} > \sum_{k \neq i} C_k) \wedge (C_{WR} = \max_{k \neq i}\{C_k\})) \\ 0 \qquad , otherwise \end{cases}$$

3. Development of Remote Robot Control System with Gestural Interface

The bi-handed Cyton 14D-2G (as shown in Figure 3) includes two arms with 14 degrees of freedom in joint motion plus a pair of actuated two-finger grippers. The bi-handed humanoid manipulators have the property of redundancy and bifurcation which enables placement of multiple hands at a desired position and orientation within its workspace in an unlimited number of ways.

To develop Client/Server structured remote human-robot interaction control software, a lightweight toolkit ActinSE is employed. The software of robot control server contains a control-based calculation API performing inverse-kinematics calculations and other associated commands, and a hardware API to directly control the Cyton arm via codes. Communication interface using TCP/IP protocol is also implemented in the robot controller. Meanwhile, client software contains a gesture recognition engine, which captures human gestures and produces gesture classification results, and similar communication interface, which sends the signals of control commands to the server. The system architecture is shown in Figure 4.

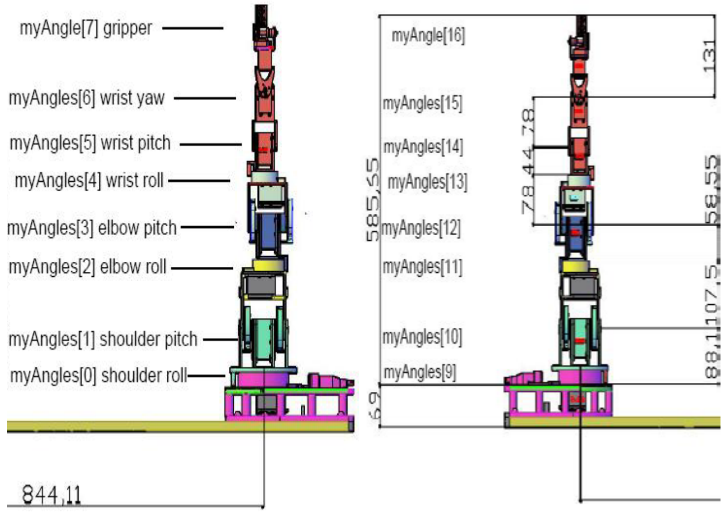


Figure 3. Robot Arm Kinematic Model

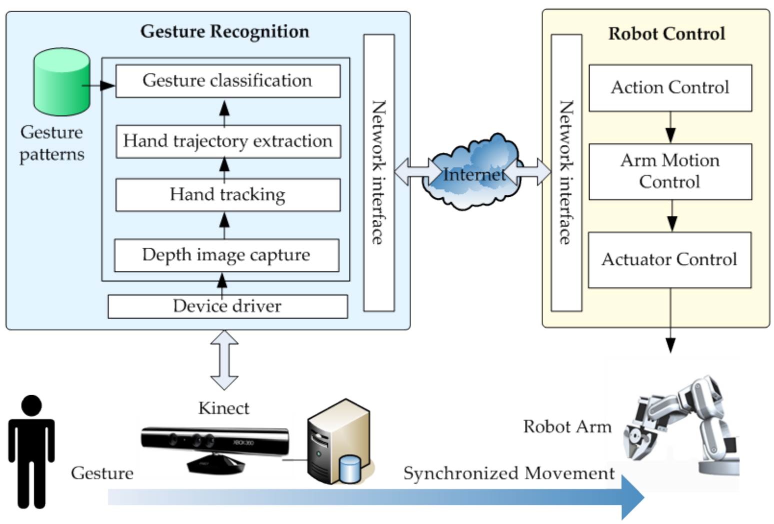


Figure 4. System Architecture

4. Experiments

In the experiment, a gesturer stood 0.3m~0.5m in front of a Kinect sensor connected to a client computer. He made four of the command gestures such as *WaveLeft*, *WaveRight*, *RiseUp*, and *PutDown* to direct robot. Client and server computers are both with Intel Core 2 Duo E8200 2.66GHz CPUs, and the software on both client and server computers was developed using Microsoft Visual Studio 2010.

To test the accuracy of gesture recognition, a set of 50 image sequences containing four of the above and other undefined gestures were collected. These sample sequences of images were performed by different gesturers and backgrounds under normal lighting conditions. Figure 5 shows an example of the confidence estimation of a *WaveRight* gesture. In addition, Figure 6 shows the result that corresponds to the situation that the user was making a sequence of *WaveLeft*, *WaveRight* and *RiseUp* gestures. In general, recognition accuracy of nearly 85% was achieved.

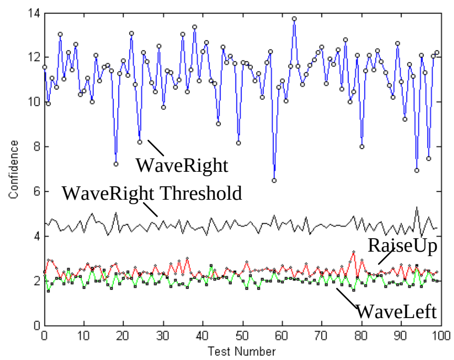


Figure 5. Gesture Confidence

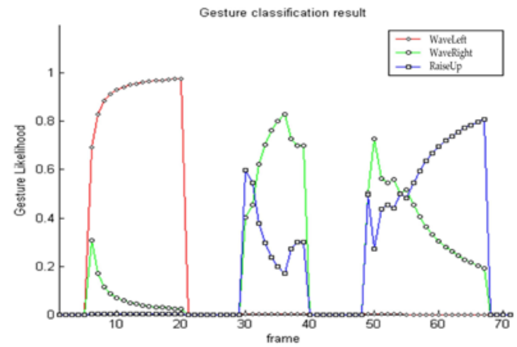


Figure 6. Gesture Classification Estimation Results

The system is developed for a gesture-based remote robot arm control application. The client software made use of OpenNI suite to develop fundamental Kinect sensor data capture and processing modules. TCP/IP protocol based network communication interface was implemented on both client and server sides. Consequently, the application system enabled reliable remote robot arm control. We have implemented four command gestures to guide the robot arm to execute several types of actions, such as rising of putting down one of its arms, as shown in Figure 7.

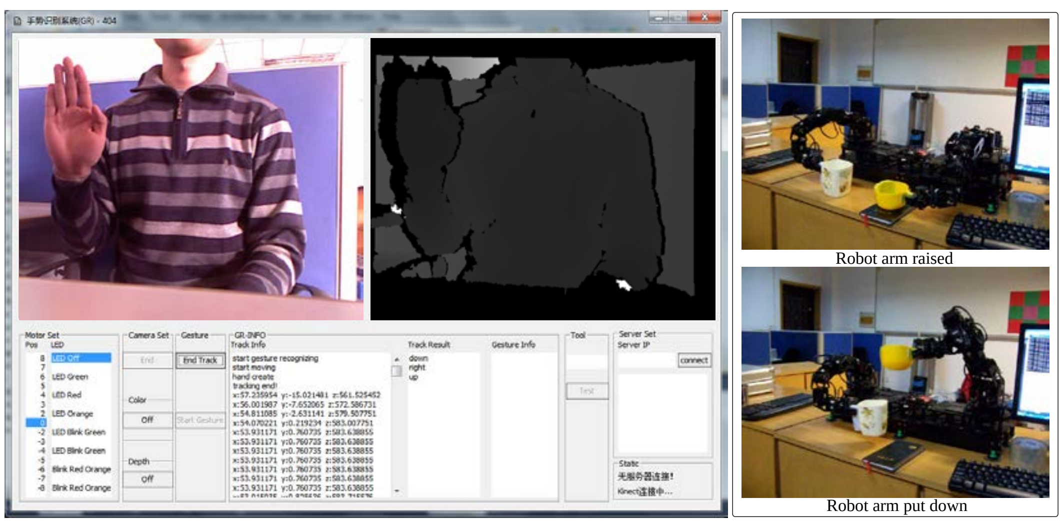


Figure 7. Software GUI

5. Conclusion

In this paper a natural interaction framework is proposed that integrates an intuitive tool for teleoperating a mobile robot through gestures. It provides a comfortable way of interacting with remote robots and offers an alternative input for encouraging non-experts to interact with robots. Application system is developed combing low-cost Kinect device and network technology. For our future work, the gesture recognition will be improved to support mobile robot navigation scenarios. We believe that better gesture recognition ratio will allow the usage of more ergonomic poses and gestures for the user, thus reducing the fatigue associated with the current gesture recognition.

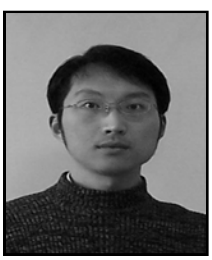
Acknowledgements

This work is supported by the National Natural Science Foundation of China (Grant No. 61105094).

References

[1] H. S. Yoon, J. Soh, Y. J. Bae and H. S. Yang, "Hand gesture recognition using combined features of location, angle and velocity", Pattern Recognition, vol. 34, **(2001)**, pp. 1491-1501.
[2] N. Liu, B. Lovel and P. Kootsookos, "Evaluation of hmm training algorithms for letter hand gesture recognition", Proc. IEEE Int'l Symp. On Signal Processing and Information Technology, Darmstadt, Germany, **(2003)**, pp. 648-651.
[3] X. Deyou, "A Network Approach for Hand Gesture Recognition in Virtual Reality Driving Training System of SPG", ICPR Conference, Hong Kong, China, **(2006)**, pp. 519-522.
[4] M. Elmezain, A. Al-Hamadi and B. Michaelis, "Real-Time Capable System for Hand Gesture Recognition Using Hidden Markov Models in Stereo Color Image Sequences", WSCG Journal, vol. 16, no. 1, **(2008)**, pp. 65-72.
[5] Y. Yu, Y. Ikushi and S. Katsuhiko, "Arm-pointing Gesture Interface Using Surrounded Stereo Cameras System", Proceeding of the 17th International Conference on Pattern Recognition, Cambridge, England, UK, **(2004)** Aug 26-26, pp. 965-970.
[6] X. Liu, R. Xu and Z. Hu, "GPU Based Fast 3D-Object Modeling with Kinect", Acta Automatica Sinica, vol. 38, no. 8, **(2012)**, pp. 1288-1297.
[7] W. Xu and E. J. Lee, "Continuous gesture recognition system using improved HMM algorithm based on 2D and 3D space", International Journal of Multimedia and Ubiquitous Engineering, vol. 7, no. 2, **(2012)**, pp. 335-340.

Authors



Dr. Kun Qian received the Ph.D. in control theory and control engineering from Southeast University in 2010. He is currently working as a lecturer at School of Automation, Southeast University. His research interests are intelligent robotic system and robot control.

Jie Niu received his M.Eng. in control theory and control engineering from Southeast University in 2007. He is currently working as a lecturer at School of Electrical and Electronic Engineering, Changzhou College of Information Technology. His research interests are robot vision.

Hong Yang received his B.Eng. from Shandong University in 2011. He is now a master candidate majored in control theory and control engineering at School of Automation, Southeast University.