Proceedings of the International MultiConference of Engineers and Computer Scientists 2014 Vol I,
IMECS 2014, March 12 - 14, 2014, Hong Kong

Gesture Recognition System for Human-Robot Interaction and Its Application to Robotic Service Task

Tatsuya Fujii, Jae Hoon Lee, *Member, IAENG* and Shingo Okamoto

***Abstract*— This paper presents a gesture recognition system for Human-Robot Interaction. And it has been employed to the development of a service robot that can be operated by the human gesture given as the user's command. Human motion is detected by the Kinect sensor on the robot and recognized as one of the predefined commands by using an algorithm based on Hidden Markov Model (HMM). Its recognition rates about the predefined gestures were verified through several experiments and compared to the region-based recognition method of the previous research. Finally, the developed system has been applied to Human-Robot Interaction for service tasks in office environment.**

***Index Terms*— Service Mobile Robot, Human-Robot Interaction, Gesture Recognition**

I. Introduction

COEXISTENCE and interaction between human and robot has been considered as one of the important issues in robotics research area in these days. It is based on the social expectation that robot's main application area will be changed from the typical industry field to daily environment in the near future. For that, many communication skills for the service robot have been proposed such as voice recognition [1], gesture recognition [2-6], and so on.

There have been proposed the approaches of gesture recognition as follows: HMM (Hidden Markov Model) and Multi-Layer Perceptron with Radial Basis Function (MLP/RBF) have been used as the recognition algorithm in [2, 3]. A method by tracking both hands with camera [4], utilizing a remote controller with inertial sensor [5], and an approach to use multiple inertial sensors that could be attached human body have been researched for recognizing human motion [6]. In addition, the researches to exploit the above mentioned gesture recognition as a tool for communication between human and robot. For example, some researchers have been carried out for the application to operate a service robot and a cleaning robot [7, 8]. Especially in [7], the communication method to use both speech and gesture has been investigated.

A camera has been used as the typical device to detect human motion. However, a new device called 'Kinect' was developed by Microsoft co. several years ago, which can capture not only the image like general camera but also achieve depth image with 3D information. It has been employed in many applications of robotics research recently. The researches [9, 10] of gesture recognition using Kinect sensor have been also reported.

In spite of many researches related to human-robot interaction, there are not so many reports of its successful application to robotic service task. Therefore, this paper aims to develop a service robot with the communication capability based on gesture recognition, and apply it to realistic service tasks to support human in indoor environment. In the previous research of the authors [11], its possibility had been investigated with a robot having region-based gesture recognition system. The recognition system has been redesigned as the algorithm with HMM and confirmed through experiments of real robotic service tasks in this work.

II. System Configuration

Figure 1 shows the concept of the proposed service robot. Firstly, the human gesture which is one of the predefined commands for the service task is given to the robot. Then, it is detected in real time and given as the position information of human arm, i.e., positions of nodes in the skeleton model, by the Kinect sensor installed in the robot. It is translated to the input signal, i.e., symbol sequence, for the recognition engine installed in the robot. After processing to recognize user's command, the robot replies to the human with the display and the audio messages based on the recognition result. At the same time, the robot starts the service task ordered by the user.

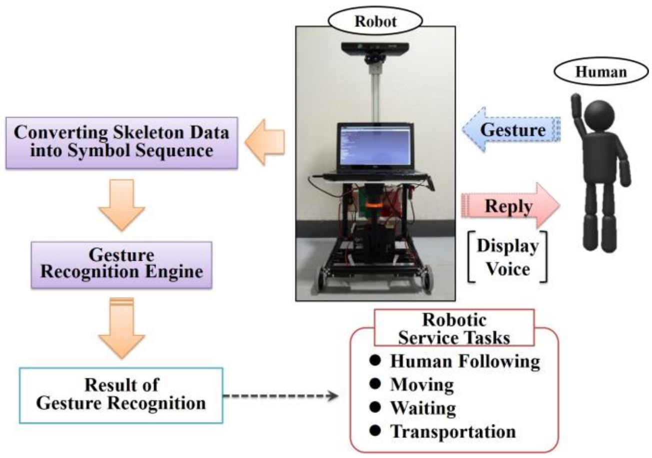


Fig. 1. Concept of the proposed service robot with gesture recognition system.

Manuscript received December 23, 2013; revised January 30, 2014.

Every author is with Mechanical Engineering Course, Graduate School of Science and Engineering, Ehime University, 3 Bunkyo-cho, Matsuyama 790-8577, Japan
E-mail: Tatsuya Fujii <y840021u@mails.cc.ehime-u.ac.jp>,
Jae Hoon Lee <jhlee@ehime-u.ac.jp>,
Shingo Okamoto <okamoto.shingo.mh@ehime-u.ac.jp>

III. Hidden Markov Model for Gesture Recognition

A. *Left-to-Right HMM*

HMM has been widely employed in speech recognition and showed successful performance in real applications [2, 12]. HMM is a kind of stochastic state transition model and makes it possible to deal with uncertain time-series data for recognition. Moreover, HMM is characterized by their learning ability which is achieved by inputting time-sequential data to HMM and automatically optimizing the model with the data.

HMM consists of several states which are connected with the probability of transition from one state to another state. Where, state transitions occur stochastically according to time. The states at any time depend only on the state at the preceding time like Markov models. One symbol is yielded and observed from one of the states according to the probabilities assigned to the states. States are not directly observable, and can be observed only through a sequence of observed symbols. Therefore, it is named "Hidden" Markov model.

The complete parameter set $M(=\{\mathbf{A}, \mathbf{B}, \pi\})$ of the HMM is represented by $\mathbf{A}$, $\mathbf{B}$ and π, and HMM is described as follows. The state transition probability is given as

$$\mathbf{A} = \left\{ a_{s_n} \middle| a_{s_n} = \Pr\left(s_n \text{ at } t+1 \middle| s_n \text{ at } t\right) \right\},$$

where, $a_{s_n s_n}$ is the transition probability from state s_n to state s_n. The observation symbol probability is given as

$$\mathbf{B} = \left\{ b_{s_n}(o_k) \middle| b_{s_n}(o_k) = \Pr\left(o_k \middle| s_n \text{ at } t\right) \right\},$$

where, $b_{s_n}(o_k)$ is the probability of output symbol o_k at state s_n. The initial state probability is given as

$$\pi = \left\{ \pi_{s_n} \middle| \pi_{s_n} = \Pr\left(s_n \text{ at } t = 0\right) \right\},$$

where, π_{s_n} is the probability that initial state is s_n. The set of states are given as

$$S = \{ s_0, s_1, \cdots, s_n \},$$

where, s_n is the state of number n. The observed symbol sequence is given as

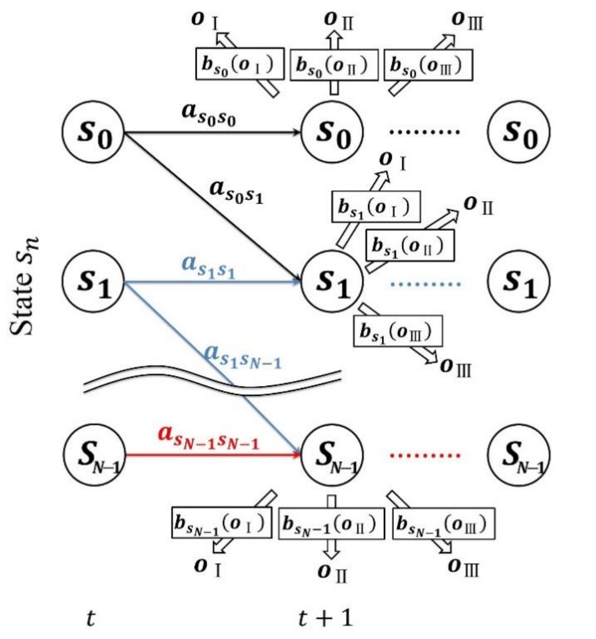
$$O = \{ o_0, o_1, \ldots, o_T \}$$

where, T is length of the observation sequence.

Besides, there are two basic types of model structures in HMM. In a fully connected (Ergodic model) HMM, every state of the model can be reached from every other state of the model. In a Left-to-Right HMM, state transition is not allowed to states whose state of number is lower than the current state. This research utilizes Left-to-Right HMM as a type of HMM to recognize gesture as shown in Fig. 2.

B. *Training Parameters of HMM*

The parameters of HMM should be decided for the recognition of gestures. A recognition model of HMM is defined as $M_m(=\{\mathbf{A}_m, \mathbf{B}_m, \pi_m\})$, where m denotes the type of gesture. In order to estimate parameters of $\mathbf{A}, \mathbf{B}$ and π, training process has been conducted, where Baum-Welch algorithm [2] was employed. A number of example data for each gesture were used for training of HMM. Then, training process has been repeated until the parameters were converged to certain values.



n : Number of state ($n = 0 \sim N-1$) $a_{s_n s_n}$: State transition probability
$b_{s_n}(o_k)$: Observation probability
N : Total number o_k : Observation symbol ($k = \mathrm{I} \sim \mathrm{III}$)
t : Time s_n : State defined in HMM

Fig. 2. Left-to-Right HMM.

C. *Recognizing Observed Symbol Sequence Using HMM*

To recognize observed symbol sequence, calculation is conducted using recognition model M_m. The probability (likelihood) $\Pr(O | M_m)$ is calculated to output symbol sequence $O = o_0, o_1, \ldots, o_T$ using M_m. This probability is calculated recursively by using the forward algorithm [2]. Namely, the probability that sequence was generated by M is calculated by using Eq. (1) and (2).

$$\alpha_0(s_n) = \pi_{s_n} b_{s_n}(o_k) \tag{1}$$

$$\alpha_{t+1}(s_n) = [\sum \alpha_t(s_n) a_{s_n s_n}] b_{s_n}(o_k) \tag{2}$$

Thus, we can calculate the likelihood of each recognition model using above equation and select the most likely HMM as the recognition result.

IV. Gesture Recognition with Kinect Sensor

A. *Detecting Human Arm's Coordinate with Kinect Sensor*

The Kinect sensor can provide the position information of human body as shown in Fig. 3. The information is displayed on the screen of the laptop computer, where the software library called as OpenCV is used in this research. The position data with respect to the sensor coordinate system, (X_s - Y_s - Z_s), is given from the Kinect device as shown in Fig. 4. For making observation symbols, the human coordinate system, (X - Y - Z), is defined, whose origin is located at the center of human right elbow, and its direction is same to the sensor coordinate system.

ISBN: 978-988-19252-5-1
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

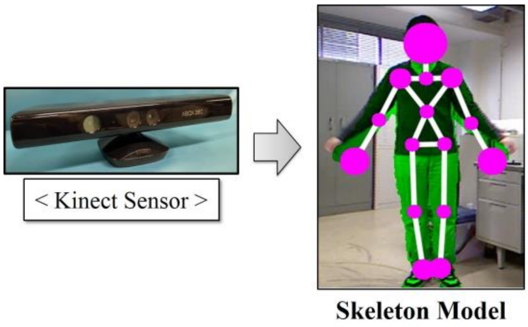


Fig. 3. Human skeleton model detected by Kinect sensor.

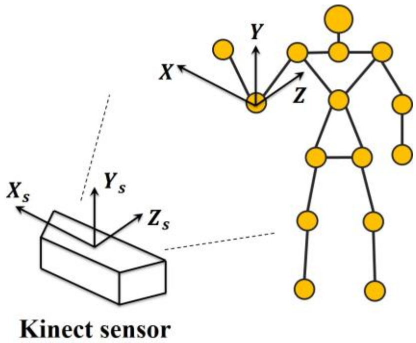


Fig. 4. Sensor and human coordinate systems.

B. Converting Position Data into Symbol of Observation

In this research, for easy and effective symbolization of human motion, we focused on the change of arm direction during human gesture. Namely, when a human moves his arm to make gestures, the direction from the elbow to the hand is changed. The direction is used as the observation for recognizing gestures because the position data cannot be used directly as the symbol of HMM.

However, the motion of human arm in gesture is complicated and spatial actually. Therefore, the symbolization with respect to both front and side view is proposed in this paper. If only one of them is considered, some motions cannot be detected because the human arm moves in a certain hidden plane. For the front view and the side view, the X - Y plane and the Z - Y plane of the human coordinate system are defined as shown in Fig. 5. Besides, both planes are divided into the eight areas as shown in Fig. 5 (1) and (2). The angle θ_{XY} from the X - axis to the arm in the X - Y plane is calculated by using Eq. (3) with the position of human right hand and right elbow.

$$\theta_{XY} = \mathrm{atan}\,2\left(y_1 - y_0 , x_1 - x_0 \right) \tag{3}$$

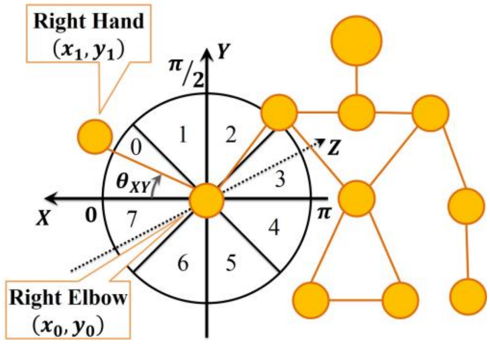
The angle θ_{ZY} from the Z - axis to the arm in the Z - Y plane is calculated by using Eq. (4).

$$\theta_{ZY} = \mathrm{atan}\,2\left(y_1 - y_0 , z_1 - z_0 \right) \tag{4}$$

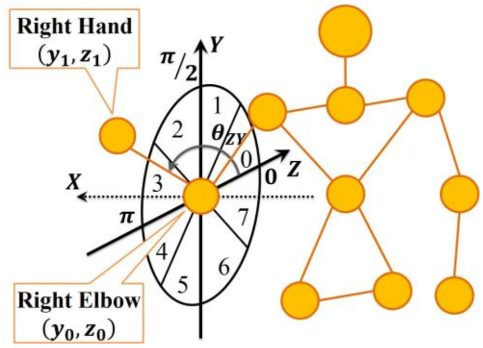
Finally, the angle is converted into the observation symbol of each plane.

C. Gesture Recognition by Using Multiple Recognition Models

Six types of gesture are prepared in this research as shown in Fig. 6. It is assumed that the service robot is operated by the commands of human gestures.



(1) X-Y plane.



(2) Z-Y plane.

Fig. 5. Two planes for converting motion of human arm into directional symbol.

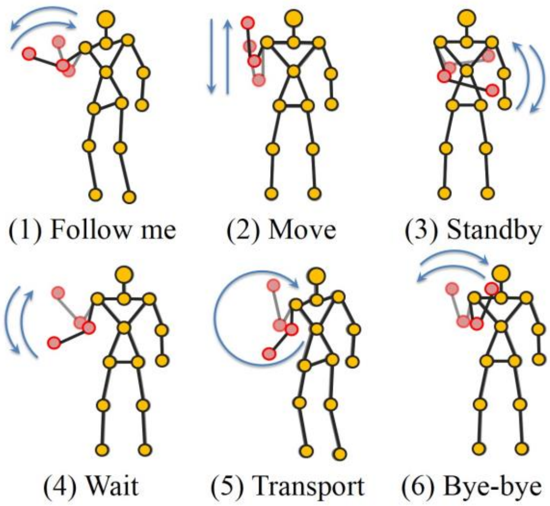


Fig. 6. Types of gesture.

Figure 7 shows the flow of information from user's gesture to robot motion. For that, a gesture recognition engine was developed, where multiple HMMs with trained parameters are employed in it. Parameter estimation is conducted by using the data with respect to both X - Y plane and Z - Y plane. So a gesture is recognized by two models for both planes. The recognition models are defined as M_{lm}. Where, the subscript l denotes the number of recognition plane as $l = 0$: X - Y, $l = 1$: Z - Y. The second subscript m denotes the number of gesture as follows; $m = 0$: 'Follow me', $m = 1$: 'Move', $m = 2$: 'Standby', $m = 3$: 'Wait', $m = 4$: 'Transport', $m = 5$: 'Bye-bye', respectively.

ISBN: 978-988-19252-5-1
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

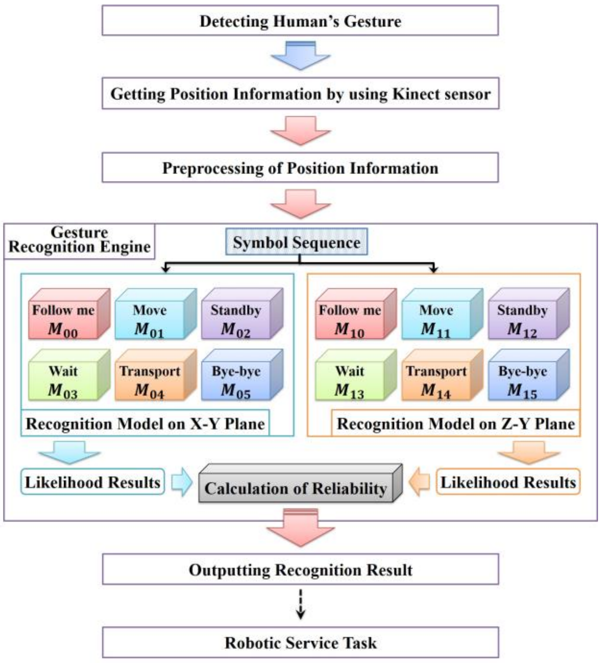


Fig. 7. Flow of information for gesture recognition.

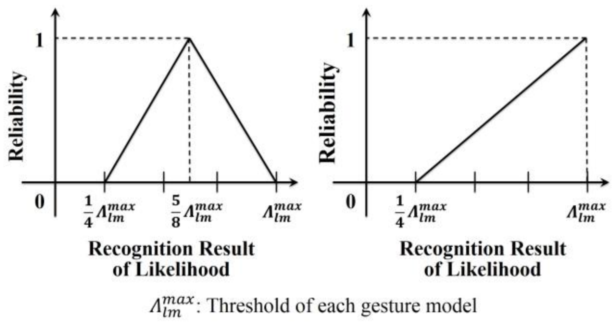


Fig. 8. Concept of reliability calculation.

Besides, the threshold is also prepared for each recognition model. Therefore, the likelihood, i.e., the output of each recognition model, is calculated by using forward algorithm with a symbol sequence as the input. As a result, the likelihood of largest value becomes the output of the recognition engine. However, the correct result cannot be judged because there are multiple results from both recognition models of X - Y and Z - Y planes and their effective range are different with each other. In order to recognize the gesture correctly, the following equations are utilized to extract data in the meaningful region of multiple likelihoods of models in both planes.

$$f(\lambda_{lm}) = \begin{cases} 0 & \left(\lambda_{lm} \le \frac{1}{4}\Lambda_{lm}^{max}\right) \\ \frac{8}{3\Lambda_{lm}^{max}}\lambda_{lm} - \frac{2}{3} & \left(\frac{1}{4}\Lambda_{lm}^{max} < \lambda_{lm} \le \frac{5}{8}\Lambda_{lm}^{max}\right) \\ -\frac{8}{3\Lambda_{lm}^{max}}\lambda_{lm} + \frac{8}{3} & \left(\frac{5}{8}\Lambda_{lm}^{max} < \lambda_{lm} \le \Lambda_{lm}^{max}\right) \\ 0 & \left(\lambda_{lm} > \Lambda_{lm}^{max}\right) \end{cases} \quad (5)$$

$$f(\lambda_{lm}) = \begin{cases} 0 & \left(\lambda_{lm} \le \frac{1}{4}\Lambda_{lm}^{max}\right) \\ \frac{4}{3\Lambda_{lm}^{max}}\lambda_{lm} - \frac{1}{3} & \left(\frac{1}{4}\Lambda_{lm}^{max} < \lambda_{lm} \le \Lambda_{lm}^{max}\right) \\ 0 & \left(\lambda_{lm} > \Lambda_{lm}^{max}\right) \end{cases} \quad (6)$$

Where, λ_{lm} denotes the result of likelihood calculation in recognition model, Λ_{lm}^{max} denotes the threshold value of each recognition model, respectively. The result of likelihood is translated as the reliability according to the shape of function as shown in Fig. 8. Where, the reliable area was decided based on the result of likelihood for all gestures.

Finally, the highest value among reliabilities from the recognition models becomes the resultant output. Then, it is transferred to the robot as the appropriate command connected to the predefined service task.

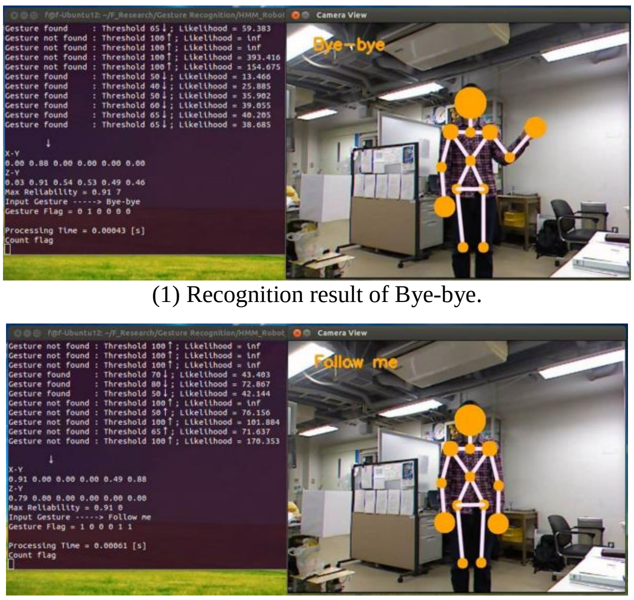
V. Experimental Results

A. Experiment of Gesture Recognition

Experiments to confirm the performance of recognition for each gesture have been conducted. A human subject stands at the position of 1.5[m] in front of the robot with Kinect sensor and conducts all gestures for ten times. Where, three humans, subjects A, B and C, participated in the experiment.

The examples of gesture recognition experiment are shown in Fig. 9 (1) and (2). Their recognition results of three subjects are shown in Fig.10, where the system used the recognition model learned by gestures of subject A. The fourth results were performed by using the region-based algorithm of the previous research [11] for the same human of subject A. Besides, the experimental results with the recognition system including the model learned by gestures of subject B are given in Fig. 11.

It is observed that the recognition rates of subject A are higher than other subjects in Fig. 10, and those of subject B are higher than other subjects in Fig. 11, respectively. It denotes the natural result that the model shows higher performance in the case to recognize the gesture of the human used in training process. In this research and in general case, the communication with owner and robot is fundamental requirement. Therefore, the higher rate for recognizing the owner's gesture is desirable and acceptable for real applications. Thus it is confirmed that the developed recognition system can be used in real field of human-robot communication.



(1) Recognition result of Bye-bye.

(2) Recognition result of Follow me.

Fig. 9. Example of recognition experiment.

ISBN: 978-988-19252-5-1
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

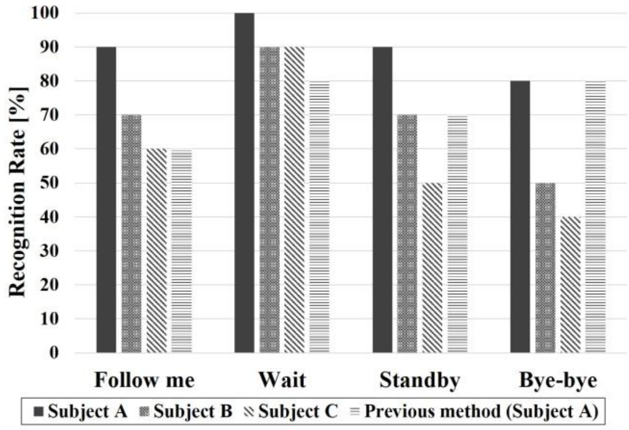


Fig. 10. Experimental results of gesture recognition using learned data of subject A.

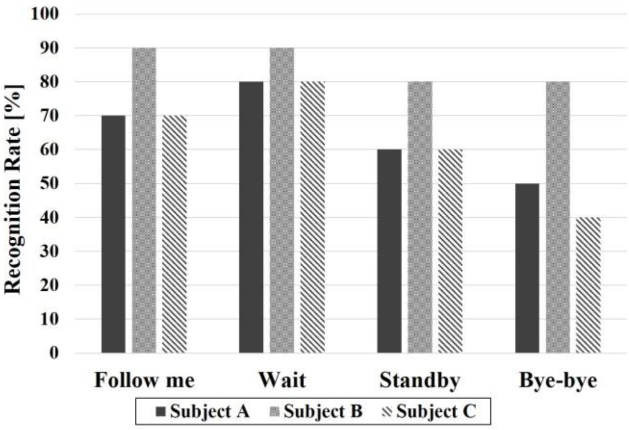


Fig. 11. Experimental results of gesture recognition using learned data of subject B.

B. Scenarios of Robotic Service Task

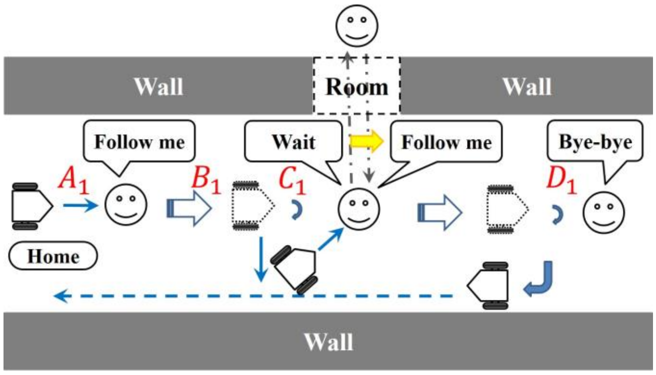
Experiments to check the possibility of the developed system have been carried out in some robotic tasks to support human as follows.

- Scenario 1: human following service.

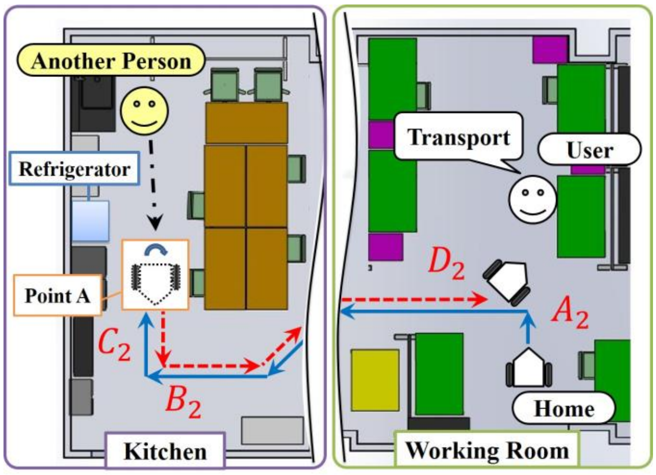
It is assumed that a service robot with a box for transportation has a capability of following human. So the service task starts (A_1), the robot follows the human like as a companion (B_1). The human can make the robot wait at certain position (C_1), and after some time he can restart to follow himself again. After arriving to the goal position, he can give a command to the robot for returning to home position with his baggage (D_1). All commands of the human are given by gestures to the robot, and above the demonstration scenario is shown in Fig. 12 (1).

- Scenario 2: object delivery service.

It is assumed that the user cannot leave from his working place. He needs the object (for example, drink or food) which were located in another room. Then the human gives a command to the robot to bring it to himself (A_2), and the robot moves autonomously to the destination (B_2), (C_2). And then another human who is at the destination puts the object on the robot, and the robot delivers it to the user (D_2). Above the demonstration scenario is shown in Fig. 12 (2).



(1) Scenario 1: human following service.



(2) Scenario 2: object delivery service.

Fig. 12. Scenario of robot's service task.

C. Experiment of Robotic Service Task

Firstly, experiment according to 'scenario 1' was carried out as follows. Its representative scenes are given in Fig.13. A human subject commands to start the task of following himself to the mobile robot in home position (1), the mobile robot follows a subject during (2). Then, he makes the robot wait at (3), after for a while he makes the robot follows again (4). Finally, he makes the robot return to home position at (5), (6).

Secondly, experiment according to 'scenario 2' was carried out as follows. Its representative scenes are given in Fig.14. A human subject commands to start transporting task to the mobile robot in home position (1), the mobile robot moves autonomously to the point A at (2), (3). Then, another human who is at the point A puts an object to the robot at (4). Finally, the robot transports an object to user's position at (5), (6).

The experiments with the developed system has been done successfully according to the scenarios.

VI. Conclusion

In this research, the gesture recognition system using HMM was developed for the communication between human and a service robot. The recognition rates for the predefined gestures familiar to human were verified through several experiments and compared to the region-based method of previous research. Furthermore, the developed system has been installed the robot and applied to Human-Robot Interaction for service tasks in office environment.

The further research is ongoing for the real applications of robotic tasks with various functions.

ISBN: 978-988-19252-5-1
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

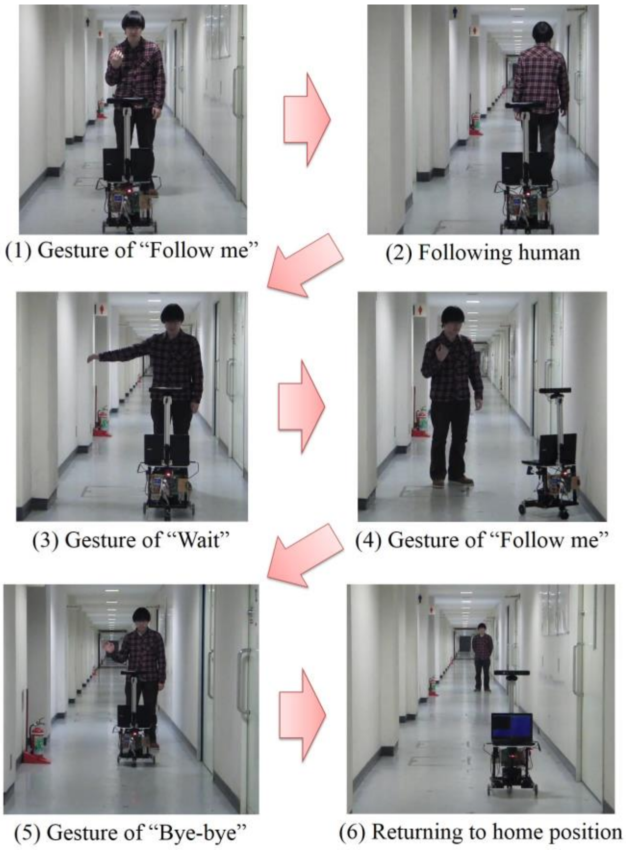


Fig. 13. Example of experimental results. (Scenario 1: human following service)

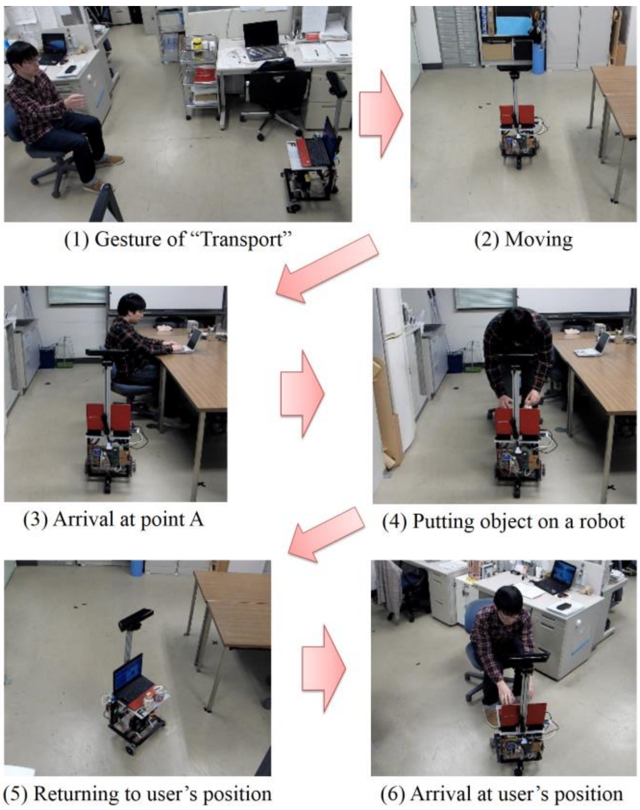


Fig. 14. Example of experimental results. (Scenario 2: object delivery service)

REFERENCES

[1] A. Vijayaraj and N. Velmurugan, "LIMITED SPEECH RECOGNITION FOR CONTROLLING MOVEMENT OF MOBILE ROBOT," *Proceedings of International Journal of Engineering Science and Technology*, Vol. 2. 10, pp. 5275-5279, 2010.

[2] L. R. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," *Proceedings of* the IEEE, Vol. 77, No. 2, pp. 257-286, 1989.

[3] M. Sigalas, H. Baltzakis and P. Trahanias, "Gesture recognition based on arm tracking for human-robot interaction," *Proceedings of* IEEE *International Conference on Intelligent Robots and Systems*, pp. 5424-5429, 2010.

[4] W. Xu and E. Lee, "A New NUI Method for Hand Tracking and Gesture Recognition Based on User Experience," *Proceedings of International Journal of Security and its Applications*, Vol. 7, No. 2, pp. 149-158, 2013.

[5] T. Schlomer, B. Poppinga, N. Henze and S. Boll, "Gesture Recognition with a Wii Controller," *Proceedings of the 2nd international conference on Tangible and embedded interaction*. ACM, pp. 11-14, 2008.

[6] H. D. Yang, A. Y. Park and S. W. Lee, "Gesture Spotting and Recognition for Human-Robot Interaction," *Proceedings of* IEEE *Transactions on Robotics*, Vol. 23, No. 2, pp. 256-270, 2007.

[7] S. Rossi, E. Leone, M. Fiore, A. Finzi and F. Cutugno, "An Extensible Architecture for Robust Multimodal Human-Robot Communication," *Proceedings of IEEE International Conference on Intelligent Robots and Systems*, pp. 2208-2213, 2013.

[8] N. Nguyen-Duc-Thanh, S. Lee and D. Kim, "Two-stage Hidden Markov Model in Gesture Recognition for Human Robot Interaction," *Proceedings of International Journal of Advanced Robotics Systems*, Vol. 9. 39, pp. 1-10, 2012.

[9] O. Patsadu, C. Nukoolkit and B. Watanapa, "Human Gesture Recognition Using Kinect Camera," *Proceedings of Computer Science and Software Engineering (JCSSE), International Joint Conference on.* IEEE. pp. 28-32, 2012.

[10] P. Doliotis, A. Stefan, C. McMurrough, D. Eckhard and V. Athitsos, "Comparing Gesture Recognition Accuracy Using Color and Depth Information," *Proceedings of the 4th International Conference on PErvasive Technologies Related to Assistive Environments*. ACM, 2011.

[11] T. Fujii, J. H. Lee and S. Okamoto, "Communication with a Mobile Robot Based on Recognition of Human Motion by using Kinect sensor," JSME Conference on Robotics and Mechatronics (Domestic). 1A1-002, 2012.

[12] J. Yamato, J. Ohya and K. Ishii, "Recognizing Human Action in Time-Sequential Images using Hidden Markov Model," in *Computer Vision and Pattern Recognition*, 1992. *Proceedings CVPR'92.* IEEE *Computer Society Conference on*, pp. 379-385, 1992.

ISBN: 978-988-19252-5-1
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)