Adaptive Behavior Generation of Social Robots Based on User Behavior Recognition

Woo-Ri Ko, Minsu Jang, Jaeyeon Lee and Jaehong Kim

***Abstract*— Social robots should understand the user's behavior and respond appropriately for natural human-robot interaction. Our work is aimed at recognizing the subtle differences in user behavior and generating behavior appropriately. For the user behavior recognition, we use a Kinect v.2 sensor for skeletal tracking and a deep neural network (DNN) for behavior classification. The weights of the DNN are trained using *AIR-Act2Act*, which is a human-human interaction dataset. For the robot behavior generation, we designed several behavior selection rules by referring to the interaction scenarios of the dataset, and then modify the key pose of the selected behavior taking into account the user's posture, position, and physical characteristics such as height. To demonstrate the effectiveness of the proposed method, we perform experiments using a Pepper robot in the 3D virtual environment. The experimental results show that the proposed method has a 98% accuracy in recognizing the user's behavior and can naturally change the behavior in a situation where the user's intention is confused.**

I. INTRODUCTION

For natural human-robot interaction, social robots should be able to understand a user's behavior and generate the most appropriate responses [1]. In this need, a number of studies have been focused on the behavior selection method of social robots [2], [3], [4]. However, these studies have a limitation that the robot is unable to respond to subtle differences in user behavior because it repeats predefined actions. Our work aims to recognize these subtle differences in user behavior and generate appropriate behavior taking into account the user's posture, position and physical characteristics such as height. For example, when the robot responds to the user's request for a handshake, the robot should stretch its hand to the user's current hand position.

We propose a social behavior generation method that can respond to subtle differences in user behavior, as shown in Fig. 1. First, a Kinect sensor [5] captures the user's 3D joint positions through skeletal tracking. For implementation, we used Kinect for Windows SDK 2.0 provided by Microsoft. Secondly, the user's behavior is classified by a deep neural network (DNN) based on the sequence of captured joint data. Then, the robot's behavior that responds appropriately to the user's behavior is selected according to the predefined rules. Finally, the selected behavior is modified according to the user's posture, position and physical characteristics such as height.

The rest of this paper is organized as follows. Sections II-IV describes the detailed procedures for the user behavior classification, robot behavior selection and adaptation, respectively. Section V presents the experimental results to demonstrate the effectiveness of the proposed method. The concluding remarks follow in Section VI.

Authors are with the Artificial Intelligence Research Laboratory, Electronics and Telecommunications Research Institute (ETRI), Daejeon, KR, {wrko, minsu, leejy, jhkim504}@etri.re.kr

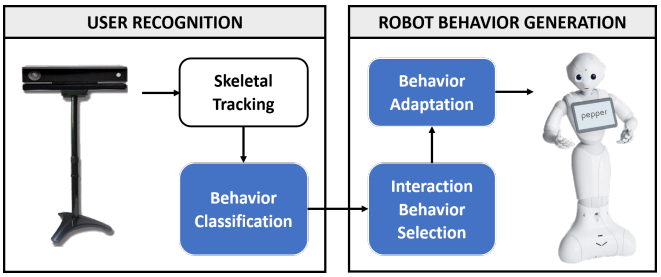


Fig. 1: Overall system.

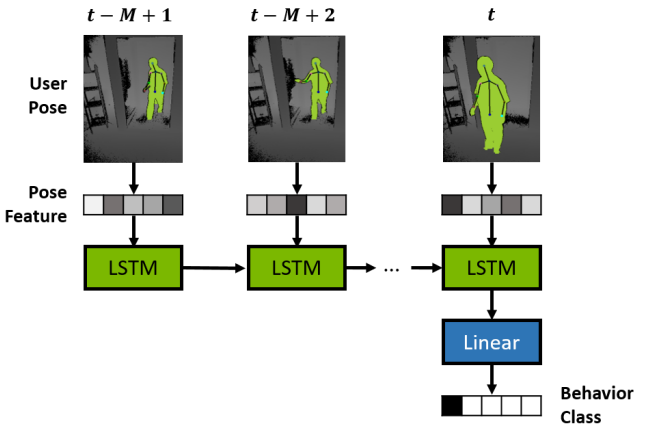


Fig. 2: LSTM-based behavior recognition model.

II. USER BEHAVIOR CLASSIFICATION

A. Deep Neural Network

To recognize the user's behavior, we use a long short-term memory (LSTM) [6] based model, as shown in Fig. 2. It is a popular model in sequential data understanding and makes a great success. In our method, for user behavior recognition, the input of LSTM is set to a sequence of feature vectors of user poses, and the output is set to a one-hot vector of behavior class label. M is the number of user poses inputted to the LSTM at one time, and the dimension of the output is equal to the number of user behavior classes.

Each user pose is represented as

$$\mathbf{P} = [\mathbf{k}_1, \mathbf{k}_2, \ldots, \mathbf{k}_9] \tag{1}$$

where $\mathbf{k}_i = (x_i, y_i, z_i)$ is the 3D position of the i-th keypoint relative to the camera. Figure 3 shows the nine

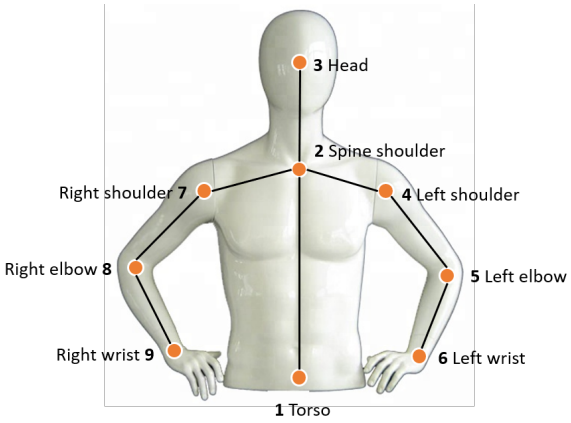


Fig. 3: Nine body keypoints used for behavior classification.

TABLE I: The five interaction scenarios used for training.

	Interaction Scenario (IS)
1	User: enters into the service area through the door. Robot: bows to the elderly person.
2	User: stands still without a purpose. Robot: stares at the elderly person for a command.
3	User: lifts his arm to shake hands. Robot: shakes hands with the elderly person.
4	User: covers his face and cries. Robot: stretches his hands to hug the elderly person.
5	User: threatens to hit the robot. Robot: blocks the face with arms.

body keypoints used for behavior classification.

To promote the convergence of neural network, the feature vector is extracted from the user pose as

$$\bar{\mathbf{P}} = \left[d, \mathbf{k}_1^2, \mathbf{k}_2^3, \mathbf{k}_2^4, \mathbf{k}_4^5, \mathbf{k}_5^6, \mathbf{k}_2^7, \mathbf{k}_7^8, \mathbf{k}_8^9\right], \tag{2}$$

where $d = \|\mathbf{k}_1\|$ is the distance from camera to torso and $\mathbf{k}_m^n = (\mathbf{k}_n - \mathbf{k}_m)/\|(\mathbf{k}_n - \mathbf{k}_m)\|$ is the direction vector from the m-th keypoint to the n-th keypoint, normalized to unity.

B. Training data

We adopted the *AIR-Act2Act* dataset [7] to train the LSTM. It contains 5,000 human-human interaction samples, where each sample captured the interaction between two human participants in an indoor environment. The dataset contains 10 interaction scenarios, and we used only five scenarios that happen frequently to service robots (Table I).

We divide the dataset into two splits: one with 45 subjects for training and the other with 5 subjects for evaluation. In each training data sample, we extracted the poses of the person who initiates the interaction and used as training input. The training output, i.e. user's behavior class, are manually labeled with the help of K-means clustering algorithm. Table II shows the user's behaviors identified in each interaction scenario.

TABLE II: User's behaviors in each interaction scenario.

IS	User's Behavior
1	stand (u_0), open the door (u_1), hand on wall (u_2), not shown (u_3)
2	stand (u_0)
3	stand (u_0), raise right hand (u_4), wave right hand (u_5), lower right hand (u_6)
4	stand (u_0), raise right hand (u_4), lower right hand (u_6), cry with right hand (u_7), raise both hands (u_8), cry with both hands (u_9), lower left hand or both hands (u_{10}), raise or cry with left hand (u_{11})
5	stand (u_0), raise right hand (u_4), lower right hand (u_6), lower left hand or both hands (u_{10}), threaten to hit with right hand (u_{12}), raise or threaten to hit with left hand (u_{13})

TABLE III: Robot's reaction to each user's behavior.

	Robot's Behavior
u_0	stares for a command (r_0)
u_1	bow (r_1)
u_2	-
u_3	-
u_4	shake hand with right hand (r_2)
u_5	-
u_6	stares for a command (r_0)
u_7	stretch hands to hug (r_3)
u_8	-
u_9	stretch hands to hug (r_3)
u_{10}	stares for a command (r_0)
u_{11}	stretch hands to hug (r_3)
u_{12}	block the face with arms (r_4)
u_{13}	block the face with arms (r_4)

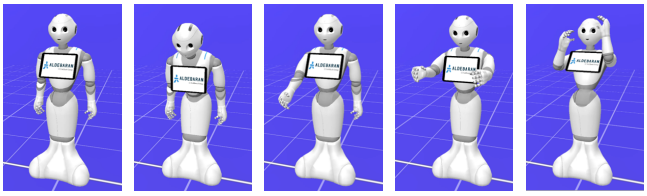
III. ROBOT BEHAVIOR SELECTION

In order to make robots respond appropriately to the user's behavior, we have defined the behavior selection rules listed in Table III. Each robot's behavior was designed by referring to the interaction scenarios of *AIR-Act2Act*.

As depicted in Fig. 4, we also defined key poses for the robot's behaviors. In the experiments, if a robot selects a behavior, it moves from the current pose to the key pose of the selected behavior.

IV. ROBOT BEHAVIOR ADAPTATION

If the robot repeats the same pose for a certain behavior, it may not be able to perform the exact motion the user wants. In order to solve this problem, we modify the key poses of the robot's behaviors according to the user's current posture, position and physical characteristics such as height. Fig. 5 shows adaptation examples of hugging behavior. In the case of hugging behavior, the robot should stretch hands to the user's shoulder. Therefore, according to the y-value (height) of the user's spine shoulder (k_2), a new key pose for hugging behavior is generated by interpolating the original



(a) r_0 (b) r_1 (c) r_2 (d) r_3 (e) r_4

Fig. 4: Key poses of the robot's behaviors.

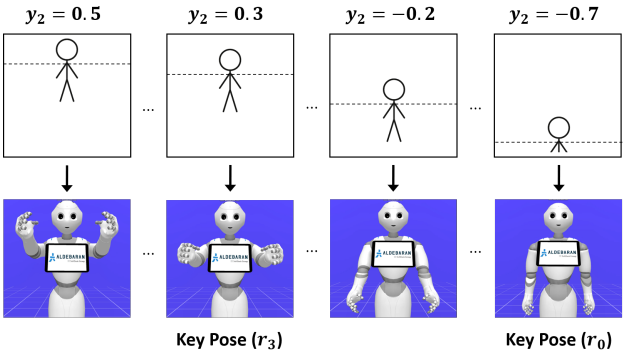


Fig. 5: Adaptation examples of hug behavior.

key poses of hugging behavior (r_3) and standing behavior (r_0). Behavior adaptation is performed in different ways depending on the selected behavior, and explanations for adaptation of behaviors other than hugging behavior are omitted due to space limitations.

V. EXPERIMENTS

To demonstrate the effectiveness of the proposed method, we performed the following two experiments using a Pepper robot in 3D virtual environment. For the neural network, we set M as 15, where the input sequence is 10 fps.

A. User behavior recognition

In this experiment, the accuracy of user behavior recognition is evaluated on the *AIR-Act2Act* dataset. Fig. 6 shows the recognition accuracy of the proposed DNN model with two different feature extraction techniques. One is direction-based feature extraction method, which is presented in (3), and the other is traditional position-based feature extraction method, which extracts a feature vector as

$$\hat{\mathbf{P}} = \left[d, \mathbf{k}_1^2, \mathbf{k}_1^3, \mathbf{k}_1^4, \mathbf{k}_1^5, \mathbf{k}_1^6, \mathbf{k}_1^7, \mathbf{k}_1^8, \mathbf{k}_1^9\right], \tag{3}$$

where $\mathbf{k}_1^n = (\mathbf{k}_n - \mathbf{k}_1)/\|(\mathbf{k}_n - \mathbf{k}_1)\|$ is the position vector of the n-th keypoint with respect to torso, normalized to unity. The recognition accuracy of our LSTM-based behavior recognition model with the proposed feature extraction method was about 98% and that with the traditional feature extraction method was about 93%.

B. Robot behavior generation

The rules for robot behavior selection were defined according to the interaction scenarios of *AIR-Act2Act* dataset, so the robot behavior generation in the test data will perform

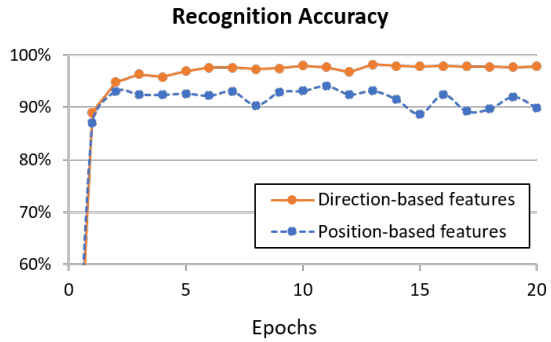


Fig. 6: Recognition accuracy of user's behaviors.

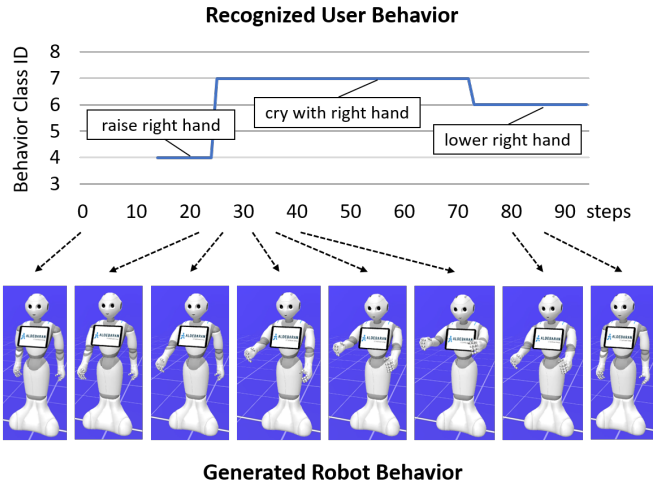


Fig. 7: Generated robot's behavior when the user wipes tears with his right hand.

excellently. Therefore, in this experiment, we show that the robot naturally changes its behavior in a situation where the user's intention is confused. Note that it was not intended in the learning process but it happens frequently in human-human interaction. Fig. 7 shows the generated robot's behavior when the user wipes tears with his right hand. When the user starts to raise his hand to wipe the tears (step 14-24), the robot cannot be sure whether the user is about to cry or to shake hands. So, at first, the robot tries to shake hands (step 18-27), but the moment it notices that the user is about to cry, it naturally changes its behavior to hug (step 28-37).

VI. CONCLUSIONS

In this paper, we proposed a social behavior generation method that can respond to subtle differences in user behavior. The user's behavior was recognized using a Kinect v.2 sensor for skeletal tracking and a long short-term memory (LSTM) based model for behavior classification. The robot's behavior was selected according to the rules established by referring to the interaction scenarios of *AIR-Act2Act* dataset. Then, the selected behavior was modified by taking into account the user's posture, position, and physical charac-

teristics such as height. To show the effectiveness of the proposed method, we performed experiments with a Pepper robot in the 3D virtual environment. The experimental results showed that the user behavior recognition accuracy was 98% and even if the user's intention is confused, the robot could change the behavior naturally.

ACKNOWLEDGMENT

This work was supported by the Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No. 2017-0-00162, Development of Human-care Robot Technology for Aging Society)

REFERENCES

[1] H. Dindo and G. Schillaci, "An adaptive probabilistic approach to goal-level imitation learning," in *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2010, pp. 4452–4457.

[2] C.-M. Huang and B. Mutlu, "Robot behavior toolkit: generating effective social behaviors for robots," in *2012 7th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 2012, pp. 25–32.

[3] A. Zaraki, K. Dautenhahn, L. Wood, O. Novanda, and B. Robins, "Toward autonomous child-robot interaction: development of an interactive architecture for the humanoid kaspar robot," in *3rd Workshop on Child-Robot Interaction (CRI2017) in International Conference on Human Robot Interaction (HRI 2017), Vienna, Austria*, 2017, pp. 6–9.

[4] S. F. Alves, M. Shao, and G. Nejat, "A socially assistive robot to facilitate and assess exercise goals," in *Proc. Int. Conf. Robot. Autom, Montreal, Canada*, 2019.

[5] Microsoft Corp., "Kinect for Windows SDK 2.0 Documentation," 2014.

[6] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[7] W.-R. Ko, "AIR-Act2Act," https://github.com/ai4r/AIR-Act2Act, 2019.