A Method for Driving Humanoid Robot Based on Human Gesture

Kenta Goto [1], Hiroaki Nishino [2], Akihito Yatsuda [1], Hokuto Tsutsumi [1], and Toshiyuki Haramaki [2]
[1] Graduate School of Engineering, Oita University, Oita, Japan
[2] Faculty of Science and Technology, Oita University, Oita, Japan
Email: {v17e3007, hn, v17e3019, v17e3011, haramaki}@oita-u.ac.jp

***Abstract*—When designing human robot communication interface, applying non-verbal interaction modalities such as robot motions is useful. Such modalities should be designed to augment verbal communication contents and set a mood for effective human robot interactions. Designing and implementing a set of useful robot motions, however, is a difficult problem. It requires developers to describe instructions for accurately controlling motors embedded in the robot. In this paper, the authors propose a method for generating a variety of robot motions using human gestures. The proposed method captures human gestures in real time by using motion tracking techniques, converting the acquired data to robot motion instructions, and applying them to a physical robot. We developed a prototype system targeted at supporting multiple motion tracking techniques and physical humanoid robots. We also conducted experiments for controlling the robots by human gestures to verify the effectiveness of the proposed method. It can promote better human robot communication environments with reducing labors in robot motion development.**

***Index Terms*— robot motion design, human motion tracking, human robot interaction**

I. Introduction

Humanoid robots are practically used in various application fields such as education, entertainment, and social welfare. A variety of robots providing useful services have been developed such as robots for performing personal services like watching over the elderly and public services like guiding and navigating at commercial facilities. The authors have been working on several IoT research projects for last few years through the development of some application systems: a senior watching system for monitoring the indoor environments of the elderly, detecting the risk of acquiring indoor heat stroke, and alerting the risk to the elderly [1], and a driver monitoring system for sensing the driver's degree of concentration, scenting the danger of accidents, and warning it to the driver [2].

Fig. 1 shows the organization of the senior watching system. The number of people acquiring indoor heat stroke has rapidly been increased due to global warming, and over half of them are the elderly. The system detects the risk in advance and prevents the elderly from acquiring the indoor heat stroke by alerting the risk. The system constantly monitors the indoor environments such as temperature and humidity, and calculates a value called heat index from the observed data. The system judges the risk level based on the index value. When the value exceeds a threshold, the system alerts to the elderly by robot’s voice messages and gestures and urges them for drinking water. Additionally, the system adjusts indoor temperature by turning on the air conditioner.

Urgent information is usually delivered in the form of images, texts, and voice messages from mobile terminals. Such mechanical push-type information delivery, however, intensifies the psychological burdens of the elderly and becomes an obstacle for effectively helping them [3]. Therefore, we introduce a humanoid robot as a partner for notifying information when detects the risky conditions while closely watching the daily life of the elderly. Implementing an effective interaction scheme between a robot and human is a crucial issue in designing the partner robot. Non-verbal communication modalities such as gestures and facial expressions are especially more useful in the interaction with the elderly than the dialogue by language and speech [4].

In this paper, the authors propose a robot motion design method to provide natural and effective non-verbal communication environments between human and robots. Designers commonly use dedicated software tools for implementing robot motions. They should, however, have expertise and advanced skills for creating natural robot motions in addition to proficiency in the tools. Therefore, we devise a method for capturing human gestures, converting them into natural robot motions, and driving the target robot depending on its allowable movement limits. The proposed method enables to generate and present reasonable robot motions conforming to the dialogue contents by recording and learning human gestures as shown on the left side of Fig. 1. The proposed method senses human gestures with a single camera, converting them to a series of the controlling commands of the target robot, and activating the robot in real time. It enables to construct a general-purpose robot motion design scheme usable for a variety of application systems development.

Manuscript received May 5, 2019; revised December 29, 2019.

doi: 10.18178/ijmerr.9.3.447-452

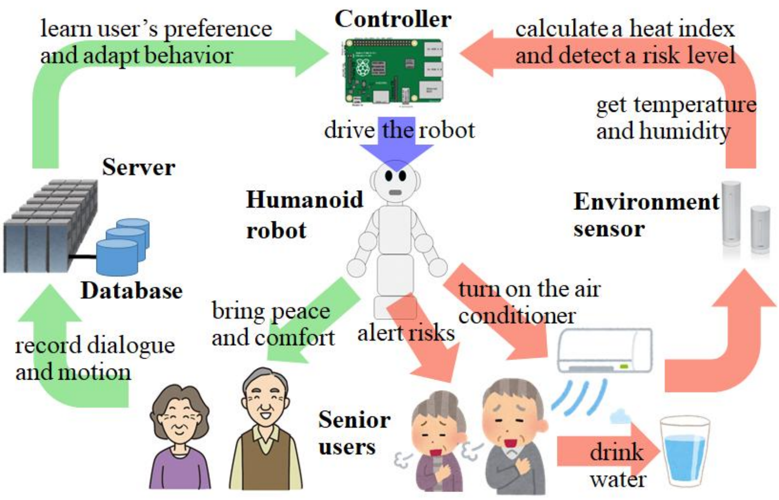


Figure 1. Senior watching system using a humanoid robot.

II. Related Work

Humanoid robots are expected to have abilities for smoothly communicating with human. It is important that the robots look and behave like human [5]. It especially is a crucial issue when the robots are working in tasks to exchange complex and diverse information with human such as watching the elderly and supporting automatic driving, rather than the tasks to be replaced by a computer such as ticketing, reception, and information retrieval [6].

It is a difficult task for appropriately designing and creating motion data so that the robots behave like human. The most promising method is to directly use human behaviors as is and map them to a series of commands for driving the robots. Cicirelli et al. proposed a method for sensing human gestures by using a commercial depth sensor, recognizing them with a neural network, and using the results for controlling a remote robot [7]. Qureshi et al. proposed a method to make a robot directly contacting with humans to acquire human-like behaviors based on a reinforcement learning algorithm [8].

Because recent evolutions of human motion sensing technology produced readily available tools and devices, it becomes easier for capturing human behaviors and converting them into a set of robot motion data. On the other hand, robot motion design tends to fully depend on physical sensors and robots to use in applications. Developing a framework or a scheme enabling even unexperienced users to design robot motions targeted for a specific sensor and robot pair is imperative. Jung et al. devised a guideline for robot motion design from the point of two different contexts depending on behaviors with and without voice conversation [9].

In this paper, we firstly define a robot motion design scheme that encompasses various combinations among different sensing technologies and robots. Then, we elaborate an implementation method for a specific case among the combinations. Our goal is to construct a robot motion design method that is independent of specific robots and sensing technologies as much as possible.

III. System Overview

Fig. 2 shows the overview of the proposed robot motion creation system. It consists of four modules; system controller (SC), human gesture sensing (HGS), gesture to motion conversion (GMC), and robot motion activation (RMA) modules. The SC module provides a GUI interface enabling the user to customize the system for driving a specific robot. Because various kinds of humanoid robots are available, the user needs to define a set of mapping rules for effectively converting human gestures to their corresponding robot motions. It manages the motion database function allowing the user to reuse the generated robot motions between different robots and share the motions with others.

The HGS module drives multiple motion tracking tools using a single camera such as OpenPose [10] and VNect [11]. These tools acquire the user's 3D bone structure of his/her posture held in front of a camera in real time. The HGS module captures the whole human body pose and generates the corresponding robot motion by mimicking the captured body pose.

Commercial robots have embedded servomotors in their joints, and they express various poses by cooperatively activating these motors. Because the motion range and the degrees of freedom of their movements vary, generating the robot motions for directly reflecting the human poses usually is a difficult task. Therefore, the GMC module converts the human poses into a set of servomotor control commands within the maximum motion range of the target robot. Specific definition examples of the conversion are elaborated in the next section.

The RMA module drives the target robot in accordance with the motion data generated by the GMC module. We are developing the system targeted for two kinds of robots; a small desktop communication robot Sota [12], and a humanoid robot Pepper [13] mainly used for public services.

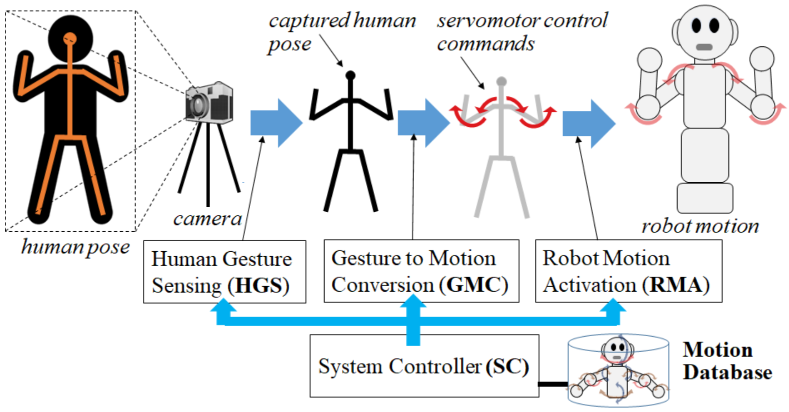
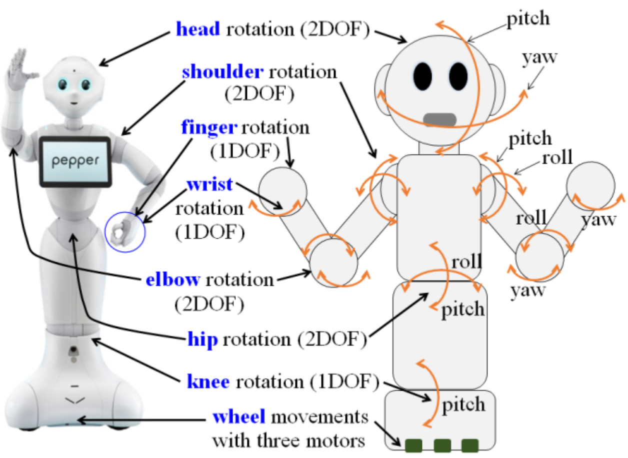


Figure 2. Overview of the proposed robot motion creation system using human gestures.

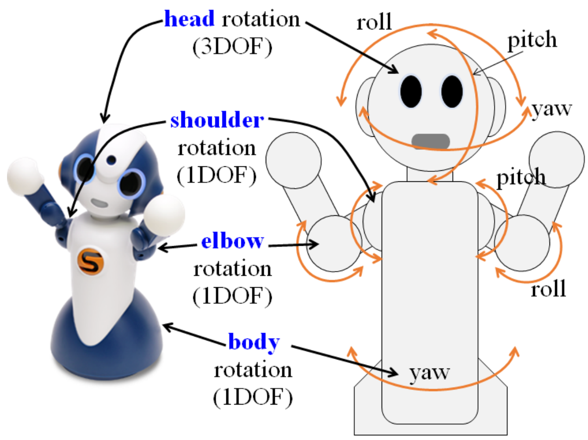
IV. System Implementation

A. Target Robots

We implemented the proposed system targeted at driving two different kinds of humanoid robots. Fig. 3(a) shows a 121cm tall robot called Pepper developed and marketed by Softbank Robotics Corp. It can generate a variety of motions with 20 DOF (degree of freedom) by simultaneously activating twenty built-in motors. Figure 3(b) shows a 28cm tall desktop communication robot called Sota developed and sold by Vstone Co., Ltd. It can generate a set of motions with 8 DOF. While Pepper can generate a three-dimensional complex motion with its higher degree of freedom actuation performance, designing a series of control commands for expressing human-like complex motions is a difficult task. Sota primarily presents planar upper limb motions with its lower degree of freedom performance. However, it can be an intuitive motion generation platform for imitating human movements in real time because of its lightweight processing feature. We elaborate a method for implementing the system using these two robots whose specifications are significantly different in the following subsections.



(a) Appearance and degree of freedom of Pepper



(b) Appearance and degree of freedom of Sota

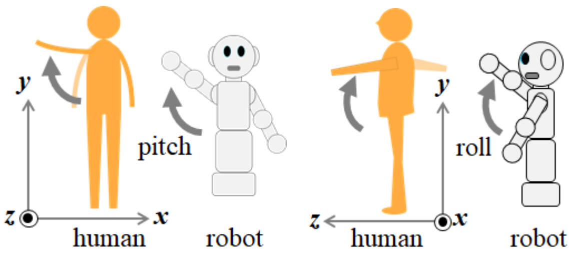
Figure 3. Humanoid robot Pepper and desktop communication robot Sota.

B. Human Gesture Sensing

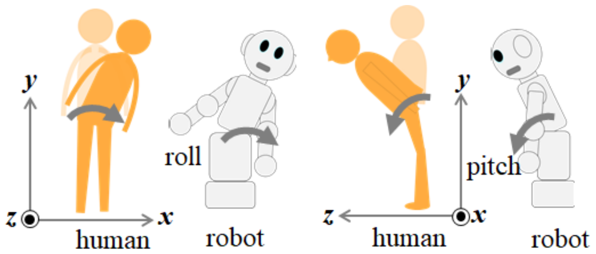
The system senses the human gestures by using a single camera without relying on a dedicated motion tracker such as Kinect or Leap Motion. The user, therefore, can easily set up the system with his/her tablet or notebook with a web camera. We use the VNect motion tracking technology for estimating a human pose with a 3D skeleton to drive Pepper, and the OpenPose method for efficiently detecting a pose with a set of 2D feature points to drive Sota. The VNect tool acquires human 3D skeleton features using a deep learning technique. In order to capture a dynamic gesture in real time, we apply it to a movie input for continuously detecting the 3D human skeleton included in each frame. Because the method requires large computing power for efficiently capturing a series of human movements by using a monocular RGB camera and estimating a 3D pose based on the deep learning, we introduce a GPU-based high-performance computing node, Nvidia's Jetson TX1, for driving the Pepper robot. For driving the Sota robot, we use a normal notebook with a web camera for performing the 2D pose estimation using the OpenPose motion tracker in 30 FPS (frames per second).

C. *Mapping Human Gesture to Robot Motion: Example for Pepper*

Pepper's head, shoulder, elbow, and hip can activate 2DOF motions with two built-in motors as shown in Fig. 3(a). Fig. 4(a) shows example motions where human arm raising and lowering movements are reproduced by combining the shoulder's pitch and roll rotations. Fig. 4(b) presents another example motion for reproducing the body movements from front to back and from side to side by combining the hip's roll and pitch rotations. The system needs to drive the robot with a set of commands to reproduce similar motions from the captured human pose.



(a) Shoulder rotations for arm raising and lowering



(b) Hip rotations for body movements

Figure 4. Gesture to motion mapping examples for Pepper.

Fig. 5 shows a method of calculating a robot motion from the arm raising and lowering human pose. VNect detects human joints as feature points and outputs the human skeletal structure. In this example, 3D coordinates of the right wrist (point A in the figure), the base of the neck (point B), and the left wrist (point C) are obtained. Next, a vector indicating the vertical direction of the human arm is calculated by the following equation.

$$\overrightarrow{BA} = (x_{ba}, y_{ba}, z_{ba}) \\ = (x_b - x_a, y_b - y_a, z_b - z_a) \quad (1)$$

Then, the system calculates the angle between the obtained vector and the x-axis with respect to the xy-plane by the following equation.

$$\theta_{xy} = atan2(y_{ba},\ x_{ba}) \quad (2)$$

This angle coincides with the pitch value of Pepper's arm and is mapped in the range from -1 to 0.

$$PepperRightPitch = \cos(\theta_{xy}) \quad (3)$$

Similarly, the system calculates the angle between the vector and the z-axis with respect to the yz-plane, and then obtains the roll value of Pepper's arm.

$$PepperRightRoll = \cos(atan2(y_{ba},\ z_{ba})) \quad (4)$$

The system can also control the left arm by calculating the pitch and roll values of the left arm using the same procedure. Additionally, the system performs the same procedures on all the joints to get the corresponding robot pose and drive it with the derived datasets.

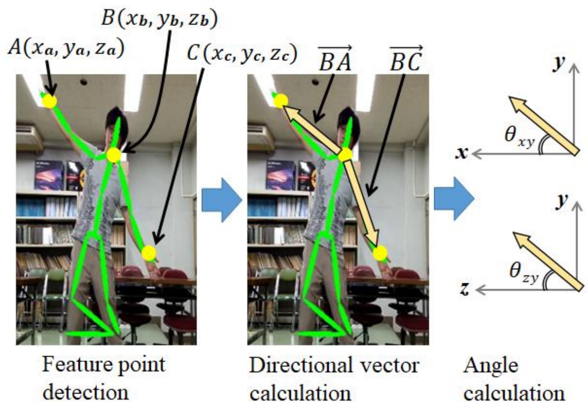
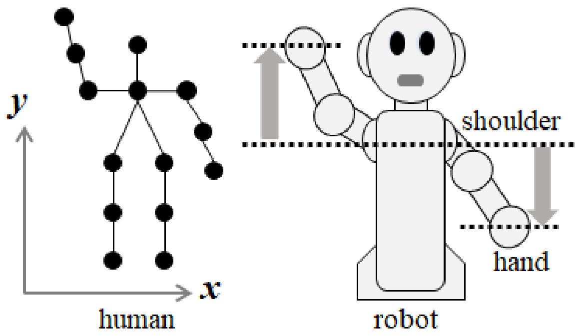


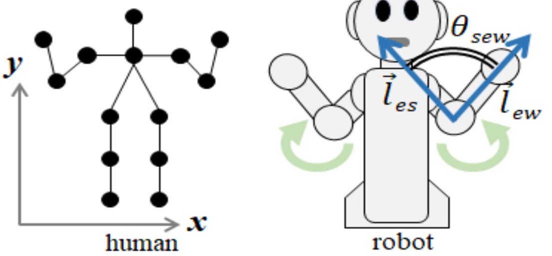
Figure 5. Derivation of a robot motion from an arm raising and lowering human pose.

D. *Mapping Human Gesture to Robot Motion: Example for Sota*

The degree of freedom in the Sota's joints is 1 DOF except for its head as shown in Fig. 3(b), the system calculates the robot motion parameters based on the human's 2D feature points obtained by the OpenPose tracker. Figure 6(a) shows a mapping example of the arm raising and lowering motion. The system calculates a difference between the shoulder's and the wrist's y-coordinate values in the human pose. Then, it raises the robot arm when the difference value is a positive or lowers the arm when negative by driving the motors at the shoulder. Fig. 6(b) shows another example of the arm bending motion. The system derives an angle (θ_{sew}) formed by two vectors, one from the elbow to the wrist ($\overrightarrow{l_{ew}}$) and the other from the elbow to the shoulder ($\overrightarrow{l_{es}}$). Then, it bends the robot arm according to the angle value by driving the motor at the elbow.



(a) Control a motor at shoulder for arm raising and lowering



(b) Control motors at shoulder and elbow for arm bending

Figure 6. Gesture to motion mapping examples for Sota.

V. Preliminary Experiments

The authors made a prototype system based on the design elaborated in Section 4 and conducted preliminary experiments. Fig. 7 shows the system console for activating the human gesture tracking and setting up the gesture to robot motion mapping. We use the types of robot motions as depicted in Fig. 4 for examining the Pepper motions and the motions in Figure 6 for Sota. Fig. 8(a) and 8(b) show the captured images taken in driving Pepper using human gestures, respectively. Fig. 9(a) and 9(b) show the results of Sota. As can be seen in these images, both robots were able to closely reproduce the user's body movements. We also verified the motion reproduction performance of the system. The system response performance that is the delay time between the human movements and the corresponding robot motions is a few second in the case of Pepper and is almost zero in the case of Sota. Although the system currently converts all of the captured human poses to their corresponding robot motions, we can further improve the performance for Pepper by culling some poses when similar ones are continued.

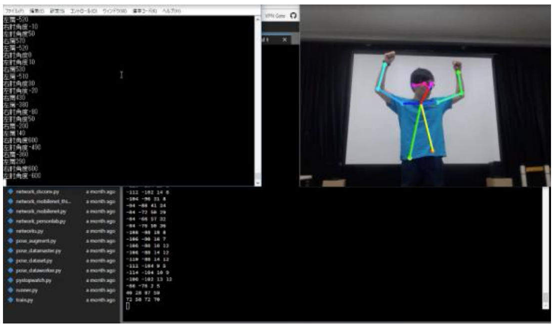
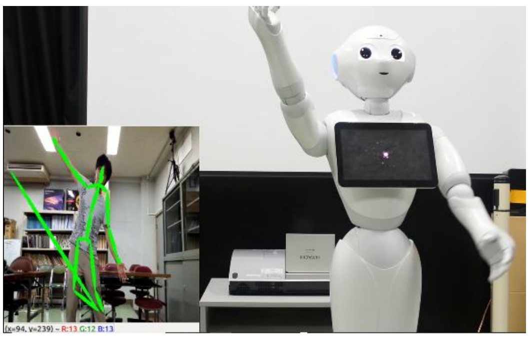
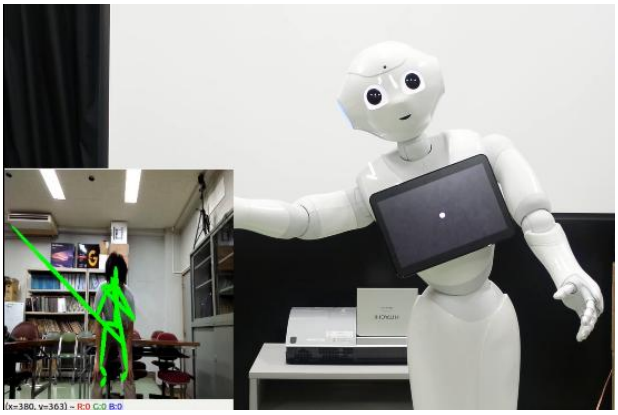


Figure 7. A console screen for controlling the system.

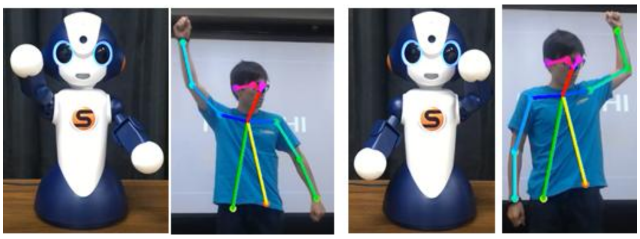


(a) Arm raising and lowering motion as depicted in Figure 4(a)



(b) Hip bending motion as depicted in Figure 4(b)

Figure 8. Robot motion results successfully mimicking human gestures by Pepper.



(a) Arm raising and lowering motions as depicted in Figure 6(a)

(b) Arm bending and stretching motions as depicted in Figure 6(b)

Figure 9. Robot motion results successfully mimicking human gestures by Sota.

VI. Conclusions

Implementing robots to behave like human is a crucial element in human-robot interaction design. To make the design process accessible even for unexperienced users, the authors proposed a method for easily creating robot motions. It captures human gestures and makes robots to imitate the captured movements. We adopted an approach for defining a design scheme to deal with the various combinations of sensors and robots and designing different sets of robot motion data in a unified manner. We configured two systems using different set of sensing technologies and robots and verified that the system could be quickly set up only by modifying some related modules. The system achieved a real-time human gesture to robot motion playback function by quickly generating motion commands from observed gestures.

We would like to further verify other sensor-robot combinations with higher degree of freedom for human gesture sensing and robot motion generation to make more human like robot behaviors. Quantitatively evaluating the flexibility and practicality of the proposed system through sophisticated experiments is another important future work.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] A. Yatsuda, T. Haramaki, H. Nishino "An unsolicited heat stroke alert system for the elderly," in *Proc. IEEE Int'l Conf. on Consumer Electronics - Taiwan*, 2017, pp.347-348.

[2] T. Haramaki, H. Nishino, "An edge computer based driver monitoring system for assisting safety driving," in *Proc. 6th Int'l Conf. on Emerging Internet, Data & Web Technologies*, 2018, pp.639-650.

[3] A. Yatsuda, T. Haramaki, H. Nishino, "Development of an adult care system based on interactions with a communication robot," in *Proc. 12th Int'l Conf. on Broadband and Wireless Computing, Communication and Applications*, 2017, pp.197-205.

[4] A. Yatsuda, T. Haramaki, H. Nishino, "A study on robot motions inducing awareness for elderly care," in *Proc. IEEE Int'l Conf. on Consumer Electronics - Taiwan*, 2018, pp.391-392.

[5] D. Matsui, T. Minato, K. MacDorman, H. Ishiguro, "Generating natural motion in an android by mapping human motion," in *Proc. of IEEE/RSJ Int'l Conf. on Intelligent Robots and Systems*, 2005, pp. 3301-3308.

[6] M. Imai, "Why do we have to use a robot?" *IPSJ Magazine*, vol.59, no.8, pp.692-697. (in Japanese)

[7] G. Cicirelli, C. Attolico, C. Guaragnella, T. D'Orazio, "A kinect-based gesture recognition approach for a natural human robot interface," *International Journal of Advanced Robotic Systems*, vol.12, no.3, pp.1-11.

[8] A. H. Qureshi, Y. Nakamura, Y. Yoshikawa, H. Ishiguro, "Show, attend and interact: perceivable human-robot social interaction through neural attention q-network," in *Proc. IEEE Int'l Conf. on Robotics and Automation*, 2017, pp. 1639-1645.

[9] J. Jung, T. Kanda, M.-S. Kim, "Guidelines for contextual motion design of a humanoid robot," *International Journal of Social Robotics*, vol.5, no.2, pp.153-169.

[10] Z. Cao, T. Simon, S.-E. Wei, Y. Sheikh, "Realtime multi-person 2d pose estimation using part affinity fields," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp.1302-1310.

[11] D. Mehta, S. Sridhar, O. Sotnychenko, H. Rhodin, M. Shafiei, H.-P.Seidel, W. Xu, D. Casas, C. Theobalt, "VNect: Real-time 3D Human Pose Estimation with a Single RGB Camera," *ACM Trans. on Graphics*, vol.36, no.4, pp. 44:1-44:14.

[12] Sota. [Online]. available: https://sota.vstone.co.jp/home/ [Accessed December 2018] (in Japanese)

[13] Pepper. [Online]. available: https://www.softbank.jp/robot/pepper/ [Accessed December 2018] (in Japanese)

Kenta Goto is a master course student in the Graduate School of Engineering, Oita University, Oita, Japan.

Hiroaki Nishino is a full professor in the Faculty of Science and Technology, Oita University, Oita, Japan.

Akihito Yatsuda is a master course student in the Graduate School of Engineering, Oita University, Oita, Japan.

Hokuto Tsutsumi is a master course student in the Graduate School of Engineering, Oita University, Oita, Japan.

Toshiyuki Haramaki is a technical staff in the Faculty of Science and Technology, Oita University, Oita University, Oita, Japan.