25th IEEE International Symposium on
Robot and Human Interactive Communication (RO-MAN)
August 26-31, 2016. Columbia University, NY, USA

# Toward a Robotic Attendant Adaptively Behaving According to Human State

## - Attending position determination based on the target person's state -

Shuji Oishi[1], Yoshiki Kohari[1], and Jun Miura[1]

*Abstract*— **This research aims to develop a robot which can adaptively attend a specific person according to the person's behavior. Transition of the person's state is modeled with finite state machine (FSM), and the robot recognizes events for the state transitions and selects an appropriate attending action for each state. We implemented attending actions for the person's walking and the sitting actions. When a person is walking, the robot takes a following action. When the person is sitting, the robot moves to a waiting position determined by considering the comfort of the person and the others. We carried out attending experiments using a real robot to show the effectiveness of the proposed approach.**

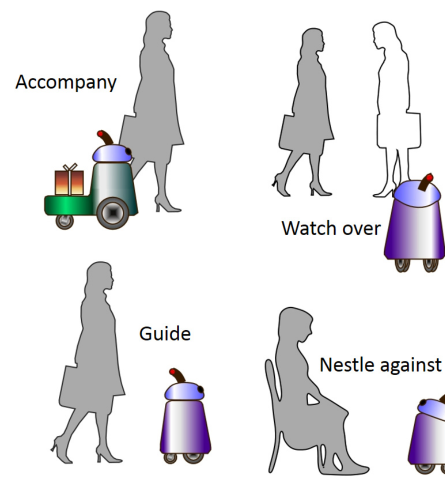



Fig. 1. Adaptive attendance

## I. INTRODUCTION

In recent years, service robots have been attracting much attention for applications such as self-reliance support of the elderly or guiding people in public places, and in a variety of situations, the robots need to provide appropriate service while attending a specific person. For such mobile robots, people detection and tracking are the essential functions and have been well studied in robotics and computer vision fields over the past decades.

Bellotto and Hu [1] develop a multi-sensor fusion technique for people tracking for mobile robots. The method detects human legs in laser scan data considering possible leg postures, and face detection in camera images is also performed to improve the accuracy of the discrimination. Fusing the observations obtained with different sensors, the method tracks people using a sequential unscented Kalman filter, and demonstrated the robustness of the proposed human tracking method in complex indoor environments. As another example, a fast people tracking algorithm for service robots using RGB-D data is proposed by Murano and Menegatti [2]. People in a scene are detected by performing 3D clustering of the point cloud and calculating HOG features from RGB data of each cluster. The algorithm then tracks the detected people considering the clusters' motions, colors, and detection confidences. They tested the proposed algorithm using a variety of datasets and data obtained from a mobile robot in crowded indoor environments, and demonstrated the outstanding performance and robustness.

While a large number of person detection and tracking algorithms have been developed and available for mobile robots, most of them focus on how well to follow the target person and do not pay much attention to the person's state. Leica et al. [3] develop a switched control strategy of a guiding robot in order to allow the robot not only to guide a specific person at a desired distance but also switch the robot's behavior based on the human-robot relative position. The method basically defines several interaction zones around the robot to enable the person to communicate the intention, however, the person explicitly has to enter the defined zone to switch the motion of the robot. When working in social environments, the robot should not only follow the person but also infer the person's state and adaptively change its own behavior for providing better service. For example, when the target person gets tired from walking and sits on a bench to rest, it is nice if the robot could recognize the person's state and move to an appropriate position so as not to give unpleasant feelings by continuing to follow persistently and stay in front of the target person.

In this paper, extending our past work [4], we propose a new attendance methodology for mobile robots. As a prototype, we have developed an attendant robot which recognizes a target person's walking/sitting behavier with laser range finders, and performs the corresponding attending action according to the person's state.

The rest of this paper is organized as follows. In section II, the overview of the proposed method is shown to give the brief idea of this work. In section III, we first introduce our

*This work was supported by JSPS KAKENHI Grant Number 25280093.
[1]S.Oishi, Y.Kohari, and J.Miura are with Department of Computer Science and Engineering, Toyohashi University of Technology, Aichi, Japan oishi@cs.tut.ac.jp, kohari@aisl.cs.tut.ac.jp, jun.miura@tut.jp

robot system, and then describe the method of human state estimation. The proposed attending algorithm based on the estimated human state is also explained. In section IV, we demonstrate how the proposed algorithm works and verify the effectiveness through experiments using a real robot. Finally, in section V, we conclude this paper and discuss the future work.

## II. OVERVIEW OF PROPOSED METHODOLOGY

Although the main task of attendant robots is following the target person for application such as guiding or nursing, appropriate attendance varies depending on the situations as shown in Fig.1. For example, while the robot only has to track the target person avoiding obstacles when the person is walking, the robot should stand by the sitting person at a position where the robot does not obstruct the person. In another example, when the person seems to be getting lost, the attendant robot leads the person and guides to the destination. Besides, when the person is talking with someone, the robot watches over the person at a distance to refrain from interfering with them. As stated above, depending on the state of the target person, the attendant robot should select appropriate behavior to provide desirable service.

This kind of advanced attendance is based on recognition of the person's state. In this study, we model the transition of the person's state with Finite State Machine (FSM). The robot recognizes events for the state transitions from sequential range data obtained by laser range finders, and selects an appropriate attending action for each state. We implemented attending actions for the person's walking and the sitting action as a prototype of the adaptive attending system. In addition, with regards to the sitting action, we determine an appropriate waiting position based on proxemics in Human-Robot interaction. The details are explained in Sec.III-C

## III. PROPOSED METHOD

### A. The attendant robot system

In this research, we use an omni-directional attendant robot (GRACE, KER) shown in Fig.2. It can controle the driving and the steering torque efficiently based on Differential-Drive Steering System (DDSS) [5] achieving a high mobility even in narrow areas. The robot is also equipped with a touch screen monitor as an interface where the robot shows useful information to the user or allows the user to input commands.

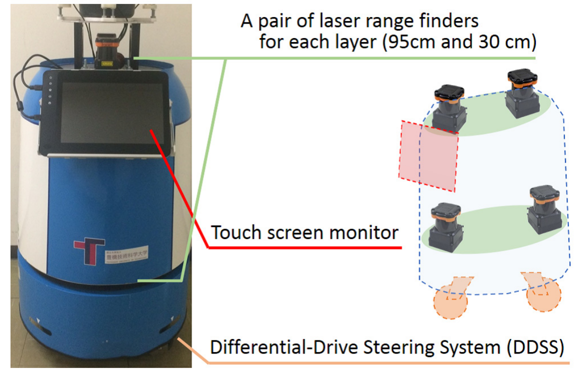
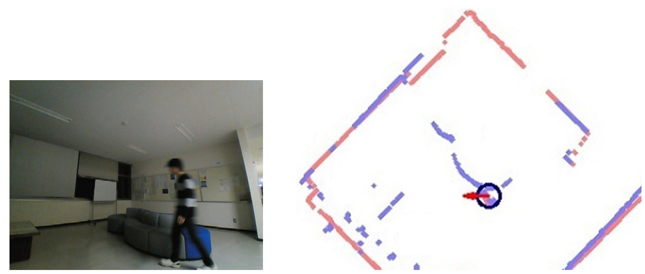



Fig. 2. An omni-directional attendant robot



(a) Scene

(b) Estimated position and orientation of the target person

Fig. 3. Person detection with laser range finders

Two laser range finders (UTM-30LX, HOKUYO) are mounted on the front and rear of waist-high (95cm) and shin-high (30cm) layers respectively in order to measure omni-directional range data at different heights. For person tracking, we first extract leg-like segments in range data focusing on that people in a scene cause local minima in the distance histogram [6]. Next, leg clusters are detected from these segments by calculating clusters's features, such as the length, the mean curvature, and the variance ratio by PCA, and classifying them with Radial Basis Function Support Vector Machine (RBFSVM) as a method of Zainudin et al. [7]. These two steps are applied to each frame, and the robot tracks the target person's legs position using Unscented Kalman Filter (UKF). In order to elaborate the state of the target person, we extend the state variables in UKF using upper range data so that it can estimate not only the position but also the body orientation of the person, as shown in Fig. 3, by comparing the input torso shape data with the model data, 360-degree torso shape data collected in advance.

When attending, the robot follows the person with planning a path using an algorithm proposed by Ardiyanto and Miura [4]. This algorithm calculates a shorter and safer path to the destination in real time utilizing a randomized path search, and enables the robot to move without collision even in dynamic environments.

### B. Modeling and estimating human states

Estimating the target person's state is essential to perform the appropriate attendance. In this research, we adopt Finite State Machine (FSM) to handle the states of the person. FSM is a model of computation that consists of a finite number of states and is capable of managing transition from one state to another. We model the transition of person's state with a FSM where three states are defined: Initial, Walking, and Sitting as shown in Fig.4. Before the recognition, the person's state is defined as "Initial". When the robot recognizes the person walking, the state transits to "Walking" state and the robot starts to follow the person. On the other hand, when the robot detects a sitting behavior of the person, the FSM

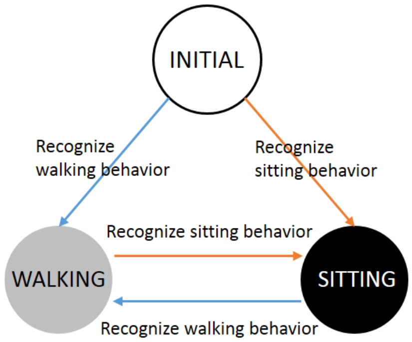



Fig. 4. Finite state machine for adaptive attendance

(a) View from the robot (b) Generated position

Fig. 5. Robot destination when following the walking person

changes its state to "Sitting" and the robot nestle against the person.

In order to recognize the transition of the target person's state, we adopt Hidden Conditional Random Fields (HCRF)[8]. HCRF is an extension of conventional CRFs into the temporal domain with hidden states and has demonstrated outstanding performance in human gesture recognition. In this research, we estimate the state of the person using HCRF based on the target person's behavior observed with laser range finders. The behavior is described by sequential features consisting of the walking speed, the distance and orientation to the nearest chair. Referring to a research of Schindler et al. [9] which has reported that observations of 5 to 7 frames are enough to recognize most human gestures, we have a HCRF learn the best discriminative structure from 5-frame consecutive features, and recognize the current state of the person in real time based on same length of the consecutive features.

### C. Adaptive attendance

*1) Walking state:* The attendant robot follows the target person when the person is walking. Actually, the robot keeps itself 1.0 m behind the person so that the robot can avoid collision even when the person suddenly stops walking. When the person steps into an area within 1.0 m from the attendant robot, it stops following until the person goes out of the area. Fig.5 shows an example of generated destinations at walking state where the robot (orange triangle) has detected a target person walking in front of it (green dot) and moving toward the target position (black cross).

*2) Sitting state:* On the other hand, when the target person is sitting, the attendant robot moves to a appropriate waiting position by considering not only the structure of the environment but also the comfort of the person and the others in the environment. The position is determined based on the following factors.

*a) Collision safety:* Risk of collision increases if the waiting position is near from obstacles in the scene. In other words, the farther from the obstacles the robot is, the more safely the robot can move. Therefore, we define an evaluation function which gives higher scores to the positions that are far from objects in the scene.

$$f_{ob}\left(x_{i,j}\right) = \mathrm{e}^{-s_{ob}(d_{i,j}^{ob})^2}, \tag{1}$$

where $x_{i,j}$ denotes a 2D candidate position, $d_{i,j}^{ob}$ a distance from the nearest obstacle, and $s_{ob}$ a parameter which controles the distribution of the function $f_{ob}$.

*b) Comfortable attendance:* The appropriate distance to the target person depends on what kind of service the robot provides. In this research, the attendant robot is supposed to provide service through a touch screen monitor mounted on itself. We therefore give higher score to the area of around 0.5 m from the sitting person so that the person can reach the monitor.

At the same time, it should be taken into account from which direction the robot attends to the target person for comfortable service. Wood et al. [10] investigated how a service robot should approach and serve the target person by carrying out experiments in real scenes, and revealed that people generally prefer approaches from the front left and the front right. Referring to the work, we define another function that gives a score to the candidate positions considering the relative distance and position to the target person as follows:

$$f_{pos}\left(x_{i,j}\right) = \mathrm{e}^{-s_d\left(d_{i,j}^h - d_g\right)^2} \times \max\left(\mathrm{e}^{-s_{th}\left(th_{x_{i,j}} - th_{FL}\right)^2}, \mathrm{e}^{-s_{th}\left(th_{x_{i,j}} - th_{FR}\right)^2}\right), \tag{2}$$

where $d_{i,j}^h$ is a distance between a candidate position and the target person, $d_g$ the target distance defined as 0.5 meter, $th_{x_{i,j}}$ the relative angle from the front of the person to a candidate position, $th_{FL}$ and $th_{FR}$ angles of diagonally forward left and right from the person, respectively. $s_d$ and $s_{th}$ controle the distribution of the function.

*c) Social distance to the other parson:* According to a proximics study of Hall [11], "Social distance" has been defined as the psychological zone $(1.2m - 3.6m)$ for meeting and interacting with unfamiliar people while "Personal distance" $(0.45m - 1.2m)$ for well-known people. In order not to give unpleasant feelings to the other people in the scene, candidate positions within the social distance to the people are rated lower as follows:

TABLE I

RECOGNITION RESULTS OF SITTING BEHAVIOR

| Behavior Label | Data | Detected | Correct | Precision | Recall | F1 |
|---|---|---|---|---|---|---|
| Pass by the chair | 135 | 134 | 128 | 95.5 | 94.8 | 95.2 |
| Sit on the chair | 61 | 62 | 55 | 88.7 | 90.2 | 89.4 |
| Total | 196 | 196 | 183 | 93.4 | 93.4 | 93.4 |

$$f_{sd}(x_{i,j}) \begin{cases} e^{-s_{sd}\left(d^o_{i,j}-d_{sd}\right)^2} & if\ d^o_{i,j} < d_{sd}, \\ 1 & otherwise \end{cases} \quad (3)$$

where $d^o_{i,j}$ denotes the distance between a candidate position and the other people, $d_{sd}$ the boundary between Social distance and Personal distance ($1.2m$), and $s_{sd}$ a parameter for the distribution.

*d) Determination of the appropriate waiting position:* Integrating the above-mentioned factors, a total score for each candidate position is calculated as follows:

$$f(x_{i,j}) = f_{ob}(x_{i,j}) \times f_{pos}(x_{i,j}) \times f_{sd}(x_{i,j}) \quad (4)$$

The appropriate waiting position is determined by finding a candidate which has the highest score. When there are several candidates which have almost the same score as the best one, the robot prefers to the nearest one for moving efficiency.

Fig. 6 illustrates the procedure of the waiting position calculation. In a scene shown in Fig. 6(a) where the target person is sitting on a chair, the attendant robot calculates scores of candidate positions based on the distance to obstacles (Fig.6(b)), relative position and angles to the target person (Fig.6(c)), and the social distance to the other (Fig.6(d)). Finally, by integrating these scores, the robot finds the most appropriate waiting position as shown in Fig.6(e).

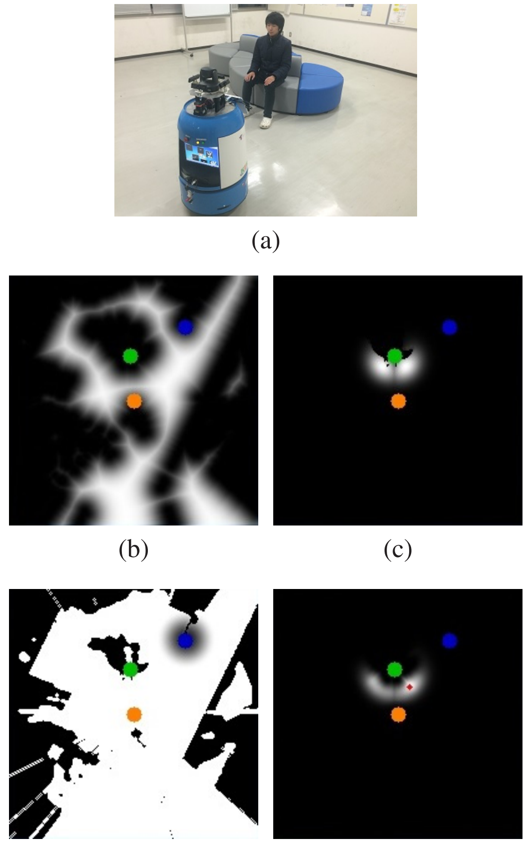

(a) (b) (c) (d) (e)

Fig. 6. **The procedure of waiting position calculation:** In a scene (a), the robot first measure the environment with laser range finders, and calculates a score for each candidate position based on the distance to obstacles (b), the relative position and angle to the target person (c), and the social distance to the other people (d). The most appropriate waiting position is then determined by integrating these scores (e). Note that orange, green, and blue dots denote the attendant robot, the target person, and the other person, respectively.

## IV. EXPERIMENTS

### A. Recognition of sitting behavior

We carried out experiments of sitting behavior recognition to verify the performance of our recognition framework. We first trained a one-class HCRF model using 235 sitting action sequences as positive data and 548 passing by a chair action sequences as negative data. Next, the trained HCRF model was applied to a test dataset which consists of 135 positive and 61 negative sequences. As shown in Table I, the attendant robot successfully recognized the sitting behavior from the simple features derived from range data obtained with laser range finders.

### B. Adaptive attending experiments

We also applied the proposed method to an actual attendant robot described in Sec.III-A. The software is implemented as a set of RT Components (RTCs) [12] and the configuration is shown in Fig. 7. The entire system works as follows: First, the attendant robot detects and starts to track a target person estimating the position and body orientation while it localizes itself using Monte Carlo Localization [13] with the given environment map. Besides, the system recognizes the person's state and determines the appropriate destination considering the state and the surrounding environment. A collision-free path towards the destination is then calculated by a randomized search [4], and the system finally drives the robot to the goal position.

To confirm whether the attendant robot could adaptively select appropriate actions depending on the state of the target person, we carried out experiments according to the following four different scenarios.

- scenario 1: The target person sits on a chair, and there is nobody around the chair.
- scenario 2: The target person sits on a chair, and there is another person standing around the chair.
- scenario 3: The target person sits on a bench, and there is nobody around the bench.
- scenario 4: The target person sits on a bench, and there is another person standing around the bench.

In each scenario, the target person walks along a corridor, approaches to a chair or a bench, and sit on it. The attendant robot follows the target person, and goes

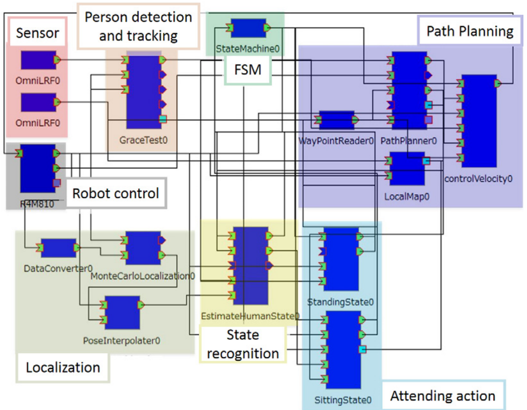
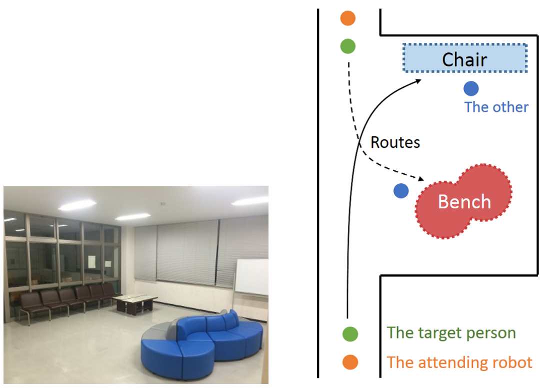



Fig. 7. The entire system implemented using RT-middleware [12]




(a) Scene (b) 2D map of the floor

Fig. 8. Experimental setup

to the most appropriate waiting position considering the surrounding environment when it recognizes sitting behavior of the person. Note that the global 2D map of the floor has been given, and the locations of chair and bench are registered in the map in advance (Fig. 8).

Fig. 9 shows the experimental results for these four scenarios. While the target person is walking, the attendant robot follows the person trying to keep itself 1.0 m behind and avoiding obstacles as well. On the other hand, when the person is sitting, the attendant robot detects the transition of the person's state and determines the appropriate waiting position considering the situation. These results demonstrate that the proposed method successfully enables the robot to attend the target person adaptively according to the state.

### C. Evaluation based on questionnaire

In order to investigate how the robot's behavior affects to the target person's feeling, we collected the subjects' opinions using a questionnaire after each trial. Fifteen male subjects were involved in this study, and we carried out the same four scenarios for each subject with and without applying the proposed adaptive attendance, totally 8 trials. Note that in the 'without' case the robot just follows the subject in the same way as III-C.1 and does not recognize

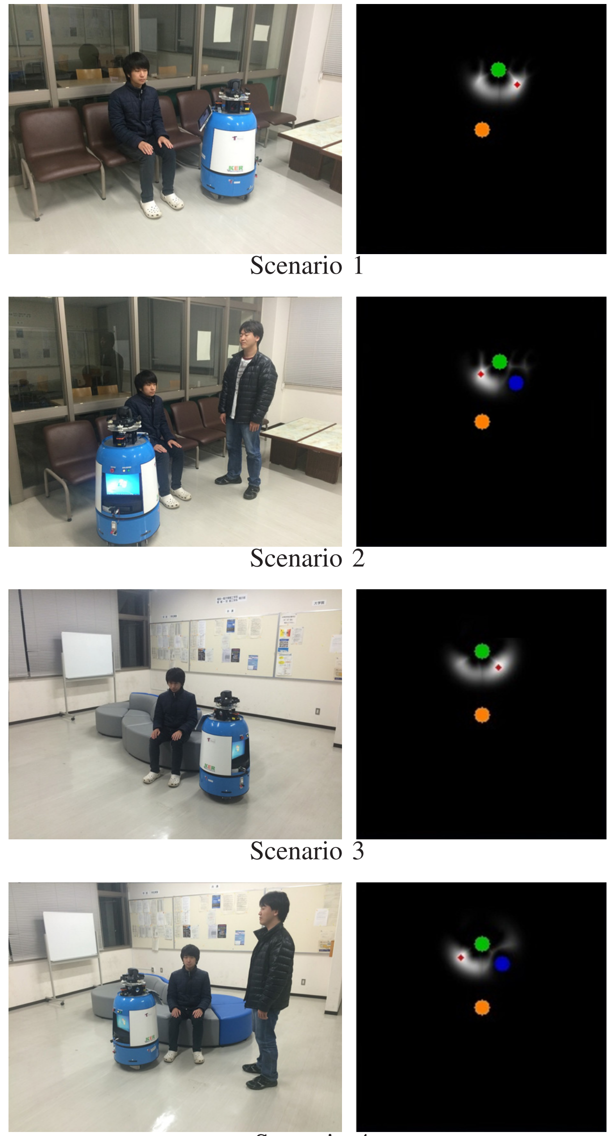



Fig. 9. Waiting position determination based on the proposed scoring. Note that orange, green, blue, and red dots denote the attendant robot, the target person, the other person, and the determined waiting position, respectively.

the state. The questionnaire allows a subject to rate the following items with 5-point Likert scales where 1 denotes highly negative, 2 fairly negative, 3 neutral, 4 fairly positive, and 5 highly positive.

1) It was comfortable to be attended by the robot.
2) The robot was considerate to you.
3) During the attendance, the robot seemed to act considering the surrounding environment such as obstacles and the other people.

We applied paired t-tests to investigate the differences between the samples with and without the proposed method. Table II shows the results of the questionnaires and the corresponding t-scores. Overall, the probabilities $p$ were $p < 0.05$ in most of the cases. In other words, there were significant improvements with regards to the subjects's feelings when

TABLE II

**Questionnaire results:** THE AVERAGES AND VARIANCES OF THE DIFFERENCES BETWEEN SCORES OF TRIALS WITH AND WITHOUT APPLYING THE PROPOSED METHOD ARE CALCULATED. THE TABLE ALSO HAS T-SCORES AND THE CORRESPONDING PROBABILITIES.

| | | Scenario 1 | Scenario 2 | Scenario 3 | Scenario 4 |
|---|---|---|---|---|---|
| Question 1 | Average scores w/ and w/o the proposed method | 2.33 / 3.53 | 2.13 / 2.93 | 3.27 / 3.87 | 2.67 / 4.00 |
| | Average of the score difference | 1.20 | 0.800 | 0.600 | 1.33 |
| | Variance of the score difference | 0.862 | 0.862 | 0.986 | 0.976 |
| | t-value *t(14)* | 5.39 | 3.59 | 2.36 | 5.29 |
| | Probability $p$ | $9.49 \times 10^{-5}$ | $2.93 \times 10^{-3}$ | $3.35 \times 10^{-2}$ | $1.14 \times 10^{-4}$ |
| Question 2 | Average scores w/ and w/o the proposed method | 2.33 / 3.60 | 2.20 / 3.33 | 2.93 / 4.07 | 2.60 / 4.13 |
| | Average of the score difference | 1.27 | 1.13 | 1.13 | 1.53 |
| | Variance of the score difference | 1.16 | 0.915 | 1.13 | 0.743 |
| | t-value *t(14)* | 4.22 | 4.79 | 3.90 | 7.99 |
| | Probability $p$ | $8.59 \times 10^{-4}$ | $2.85 \times 10^{-4}$ | $1.60 \times 10^{-3}$ | $1.39 \times 10^{-6}$ |
| Question 3 | Average scores w/ and w/o the proposed method | 2.40 / 3.13 | 1.73 / 3.33 | 2.87 / 3.40 | 1.67 / 4.33 |
| | Average of the score difference | 0.733 | 1.60 | 0.533 | 2.67 |
| | Variance of the score difference | 0.961 | 1.35 | 1.13 | 0.900 |
| | t-value *t(14)* | 2.95 | 4.58 | 1.84 | 11.5 |
| | Probability $p$ | $1.04 \times 10^{-2}$ | $4.26 \times 10^{-4}$ | $8.78 \times 10^{-2}$ | $1.65 \times 10^{-8}$ |

the proposed adaptive attendance was applied enabling the robot to give pleasant impressions to the target person.

The probability $p$ of the third question in the third scenario was more than 0.05 indicating the proposed method did not significantly improve the subjects' impressions of the robot's awareness or consciousness. This was partly because in the third scenario there was no one except the target person and the robot could get good scores according to the algorithm [4] which generated collision-free paths to deal with the situation.

In conclusion, the results indicate that the proposed method allows the attendant robot to perform more considerate attendance, and also make the target person more comfortable during the attendance.

## V. CONCLUSION

In this paper, we developed an service robot which can attend the target person considering the person's state to provide appropriate service. The transition of person's state is modeled using a Finite State Machine (FSM) to handle "Walking" and "Sitting" states while the state is estimated based on sequential range data captured with laser range finders. We implemented the proposed method in a mobile robot and demonstrated that the robot could select an appropriate action according to the target person's behavior for comfortable attendance. We also evaluated the proposed method based on questionnaires where the robot attended the target person in several situations. The study indicates that the proposed adaptive attendance makes the target person more comfortable and also feel close to the attendant robot.

For future work, we extend the proposed method so that it can recognize objects in the scene such as chairs, desk, doors, and so on for handling a variety of tasks. We also need to discuss the use of FSM. FSM with a number of states may enable the attendant robot to deal with more complicated situations, however, increasing merely the number of states may not necessarily lead to natural interaction. For sophisticated interaction, we should extend the system carefully considering psychological and socialcontextual factors.

## REFERENCES


[1] N. Bellotto and Huosheng Hu, "Multisensor-based human detection and tracking for mobile service robots," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 39, no. 1, pp. 167–181, 2009.

[2] Matteo Munaro and Emanuele Menegatti, "Fast RGB-D people tracking for service robots," *Auton. Robots*, vol. 37, no. 3, pp. 227–242, oct 2014.

[3] Paulo Leica, Juan Marcos Toibero, Flavio Roberti, and Ricardo Carelli, "Switched control to robot-human bilateral interaction for guiding people," *J. Intell. Robotics Syst.*, vol. 77, no. 1, pp. 73–93, jan 2015.

[4] Igi Ardiyanto and Jun Miura, "Real-time navigation using randomized kinodynamic planning with arrival time field," *Robotics and Autonomous Systems*, vol. 60, no. 12, pp. 1579–1591, Dec 2012.

[5] Yuki Ueno, Takashi Ohno, Kazuhiko Terashima, Hideo Kitagawa, Kazuhiro Funato, and Kiyoaki Kakihara, "Novel differential drive steering system with energy saving and normal tire using spur gear for an omni-directional mobile robot," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2010, pp. 3763 – 3768.

[6] Dirk Schulz, Wolfram Burgard, Dieter Fox, and Armin B. Cremers, "People tracking with a mobile robot using sample-based joint probabilistic data association filters," *International Journal of Robotics Research (IJRR)*, vol. 22, no. 2, pp. 99–116, 2003.

[7] Zulkarnain Zainudin, Sarath Kodagoda, and Gamini Dissanayake, "Torso detection and tracking using a 2D laser range finder," in *Australasian Conference on Robotics and Automation*, 2010.

[8] Sy Bor Wang, A. Quattoni, L. P. Morency, D. Demirdjian, and T. Darrell, "Hidden conditional random fields for gesture recognition," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 2006, vol. 2, pp. 1521–1527.

[9] K. Schindler and L. van Gool, "Action snippets: How many frames does human action recognition require?," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2008, pp. 1–8.

[10] S. N. Woods, M. L. Walters, K. L. Koay, and K. Dautenhahn, "Methodological issues in hri: A comparison of live and video-based methods in robot to human approach direction trials," in *The 15th IEEE International Symposium on Robot and Human Interactive Communication (ROMAN)*, Sept 2006, pp. 51–58.

[11] Edward T. (Edward Twitchell) Hall, *The hidden dimension : man's use of space in public and private*, London : Bodley Head, 1966.

[12] Noriaki Ando, Takashi Suehiro, and Tetsuo Kotoku, "A software platform for component based rt-system development: Openrtm-aist," in *The 1st International Conference on Simulation, Modeling, and Programming for Autonomous Robots*, Berlin, Heidelberg, 2008, SIMPAR '08, pp. 87–98, Springer-Verlag.

[13] "MRPT," `http://www.mrpt.org/`.