A Teleoperation Approach for Mobile Social Robots based on Autogazing and 3D Spatial Visualizations

Andres Mora, Dylan F. Glas, Takayuki Kanda, and Norihiro Hagita
Advanced Telecommunications Research Institute International, ATR

***Abstract*—The teleoperation of mobile social robots requires operators to understand facial gestures and other non-verbal communication of the person interacting with the robot. It is also critical for the operator to comprehend the surrounding environment, in order to facilitate an improved human-robot interaction. Allowing the operator to control where the robot looks to obtain visual feedback of the person interacting with the robot can help the operator observe non-verbal communication. However, it can also produce undesirable side effects, such as operators getting disoriented and navigation becoming inaccurate. In order to solve the problems caused by these side effects, the authors developed a graphical user interface which combines an automatic control of the robots gaze and a 3D representation of the surrounding environment, such as location of items and configuration of a shop. A study where a robot plays the role of a shopkeeper was conducted to validate the proposed GUI. It was demonstrated that when providing the operator with the implemented representations of the spatial relationships, the benefits of the proposed automatic gaze control were maintained, the undesirable side effects were reduced, and the quality of the interaction with the customer was improved.**

***Index Terms*—spatial relationships, workload, teleoperation, social mobile robot.**

I. Introduction

MObile social robots are expected to be used in human environments everyday environments such as malls, elderly care centers, museums, etc. While it is ideal that such robots are deployed with full autonomy, we have found that small amount of teleoperation enables such social robots to provide useful service [1], [2] even with immature techniques, such as lack of capability in understanding speech in noisy environments. There are also safety and legal reasons that would require partial supervision from human operators. Teleoperation is also actively used in laboratory studies (using Wizard-of-Oz (WOZ) methods [3]–[5])

The development of social robots which incorporate teleoperation brings together two very different branches of human-robot interaction (HRI). One is social HRI, which focuses on studying psychological aspects of conversational interactions between people (from now on referred to as *customers*) and robots. The other is HRI for teleoperation, which typically focuses on issues like the workload of the *operator* (person remotely controlling the robot), situation awareness and shared autonomy [6] for the remote operation of non-social robots.

A. Mora, D. Glas, T. Kanda and N. Hagita are with the Advanced Telecommunications Research Institute International 2-2-2 Hikaridai, Keihanna Science City, Kyoto, Japan, 619-0288, e-mail: amoravar@asu.edu

Manuscript received ...; revised ...

Little research has explored different techniques for teleoperating mobile social robots, leaving many questions unanswered. What new requirements exist for social robots? What new techniques can aid a teleoperator in controlling social robots effectively?

Keeping track of a person's face is fundamental for social interactions, yet, manually actuating this task requires a large amount of effort by the operator. An automatic gaze control technique of the robot's head was implemented to keep the customer during our study within its field of view and relieve the operator from this routine task.

Although our implemented automatic gaze control has benefits such as reducing the operator's actuation workload and increasing the operator's awareness of the customer's state, including facial expressions and gestures; the authors found that the automatic gaze control has some drawbacks; it reduces the operator's awareness of surrounding environment, and makes an operator less effective in navigation.

Indeed, the field of view of the robot's camera is narrow and at any one time, the video can be showing the customer or the environment (e.g. the area in front of the robot) but not both. Thus, if the proposed automatic gaze control is always engaged, the operator cannot see video of the area in front of the robot. This effectively limits the operator's understanding of the robot's position and surroundings.

To overcome this difficulty, a 3D graphical user interface (GUI) was created to represent the robot's environment which augments the operator's understanding of spatial relationships. In this paper, we establish that a teleoperation system for mobile social robots must provide the operator with an appropriate representation of spatial relationships when automatic gaze control is used.

II. Related Works

A. Teleoperation for navigation tasks

For mobile robots that have to accomplish navigation tasks in order to carry out missions such as search and rescue, military tasks or space exploration, there are two opposite approaches along the ends of a spectrum: being completely teleoperated by humans [7]–[9] or being fully automated [10]. Some of the aspects of research on teleoperation involve increasing and maintaining the level of situational awareness of the operator [11], [12], combining mixed and virtual reality techniques to help the operator improve the navigation of the robot [13], and the design of the Graphical User Interface (GUI) to be used to remotely operate the robot.

Particular to the design of GUIs for navigational robots, a number of studies have been done regarding the way to present information [14], [15]. One notable finding could be summarized as the need to combine different types of information altogether [16], [17]. In specific, how does the navigation of the robot improve with a GUI that integrates a video feed and map data within a 3D environment, in contrast to video-based only or map-based only GUI.

Although existing knowledge in this domain has proven useful, further understanding of the requirements governing the teleoperation of mobile social robots is imperative. The teleoperation of social robots requires observation of new kinds of information (e.g. gestures, facial expressions, tone of voice, relative positioning) as well as to address new problems in actuation that may arise (controlling conversation, gaze direction, and gestures; following someone via locomotion or gaze control). Our approach to solve these issues is presented in later sections of this paper.

B. Teleoperation of social robots

In practice, the WOZ methodology in HRI involves the remote control of a robot system. In that respect, it appears to be similar to teleoperation. However, the system that allows the operator to do so, is seen as a tool and not as a research topic in itself.

In the work carried out by Kuzuoka [18], focus is given to the "ecology" among operators and customers. In Kuzuoka's study the idea of the operator acquiring all the information through a video-only interface is conducted and no map information is provided. It reports the fact that what the operator utilizes (in this case, a three-screen based GUI) is not necessarily a good factor for the interaction with a customer e.g. due to the robot's lack of natural motion.

C. Natural interaction with social robots

In this study, our focus is to enable a "context-sensitive" interaction between a human and social robots, where the robots' interaction go beyond simple question-answer-type or command-receiving-type interaction. In the scope of this paper, the importance is the adaptability of the robot to the customer's context, including a location, surrounding objects, attention, and subtle reaction (see our watch shop scenario in Section V as an example of such interaction). There are a number of studies with social robots conducted for natural interaction. There are many aspects to be studied, such as knowledge on non-verbal behaviors, like natural way of gazing [13], [16], proximity behavior [11], [12], the way of social dialog [3], and social patterns [19]. These studies are certainly useful for future social robots; however, the context of users was often out of focus in this type of studies. Some previous studies in robotics have aimed to recognize users' context, like a way to recognize joint attention behavior [1], attention [20] and engagement [21]. Although new techniques are constantly being developed, the robots' capabilities in context-sensitive interactions have remained highly limited.

III. Design Principles

Previous work on the teleoperation of mobile robots has been mainly focused on navigational robots whereas little is known about the teleoperation of mobile social robots. The basic design of our teleoperation system was created according to this previous knowledge on teleoperation for navigational robots. This section introduces the authors' proposed techniques for the teleoperation of mobile social robots and the guidelines on which these techniques are based on.

A. Guidelines for Navigational Robotics

Research on the teleoperation of mobile robots, using traditional 2D GUIs, has shown that distributing information on different locations of the interface may result in an increased workload and decreased performance of the operator [17]. These results may be caused by poor situation awareness of the operator. Situation awareness can be referred to as the level of understanding of the operator with respect to the environment around the robot that allows the operator to provide accurate instructions to the robot [22].

In [15], a study compares the usefulness of combining map and video information in a navigation task by comparing a side-by-side 2D representation and an integrated 3D representation. This study reports that the integration of map and video information in a 3D-based GUI positively affected the performance of the operator during navigation of the robot. However, the scope of this study is only a navigational task and it does not addresses important issues such as observing facial gestures of a customer and how they would affect the performance of an operator.

From a design perspective, Nielsen et al. [17] summarize that to improve situation awareness in human-robot systems it is recommended to: a) use a map, b) fuse sensor information, c) minimize the use of multiple windows and d) provide more spatial information to the operator. Based on these recommendations, the authors have implemented a GUI that incorporates laser range data, a video feed, a 3D model of the robot used in this research and a 3D representation of the environment where the robot is located.

B. Proposed Techniques

In addition to these guidelines, two fundamental mechanisms for facilitating the teleoperation of a mobile social robot are proposed: automatic gaze control and visualization of spatial relationships. The first one helps relieve the operator from continuously having to direct the camera towards the customer and the second one helps the operator retain the awareness that may be lost by providing the operator with autonomy.

1) Automatic Gaze Control: A critical requirement for the teleoperation of mobile social robots is to allow the operator to observe the facial expressions and gestures of the customer. Typically, this information is provided to the operator as a video feed coming from a camera pointing to the object or location of interest; in this way, the operator can understand the intentions of the customer. However, the actuation required

by the operator to maintain the customer within the field of view of the robot's camera may increase the workload of the operator, especially when the customer may continuously move inside the environment.

Thus, the automation of such task would become useful to reduce the effect of this workload on the performance of the operator. A feature called "automatic gaze control" is proposed to allow the system to automatically control the robot's gaze (i.e. camera direction) to follow a person's location and the person's face. The operator then, is able to observe the facial expressions and gestures of the person interacting with the robot without the tedious responsibility of maintaining the robot's gaze direction manually.

2) Visualization of Spatial Relationships: The proposed automatic gaze control is intended to release the operator from continuously following a customer's face in order to decrease the workload in terms of actuation and allow the operator to easily concentrate on the customer's facial expressions. However, we have observed that this automatic gaze function sometimes caused a "disorienting problem". For instance, we observed that an operator seems to lose track of spatial relationship around the robot after he/she used automatic gaze control for a while. During this moment, a customer moved around the robot, thus the real spatial relationship was changed; however, as the camera direction was controlled by the system, he/she took less attention to this change of the spatial relationship. Thus, when he/she needed to navigate the robot after that, for example because the customer moved to another product, he/she seemed not sure in which direction the robot was oriented, where the robot and the customer were, and where shop products were with respect to the robot. This resulted in a moment of confusion. The operator seemed to take much time in acquiring awareness of spatial relationships, while the robot behave strangely, i.e. spinning around and moving into a direction that does not make sense.

Therefore, it becomes essential to visualize spatial relationships between the robot and both static and dynamic objects in the environment. Using graphical visualization of such spatial relationships in conjunction with a video feed would increase the overall perception of the environment, by releasing the operator from the need to create a mental map of the objects in the environment, since they are represented on the GUI.

Through combination of the design recommendations presented in [15] with our proposed techniques, an enhanced robot control by the operator reflected in an improved human-robot interaction is expected.

IV. System Implementation

Given that our approach incorporates shared autonomy, implementation is necessary on both the robot side and operator side. This section presents how the concepts of visualizing spatial relationships and automatic gaze control are carried out within the proposed teleoperation system.

A. Robot side

The robot testbed used in our research is called "Robovie II". It comprises a mobile base (Pioneer 3) and an upper body that has two arms, each with 4 degrees of freedom (DOF) and a head with 3DOF. The arms can be used to point at the objects of interest as well for other gestures that complement its utterances. The head has a camera, a microphone and a speaker to allow an operator to gather information about the environment and the person the robot is interacting with. Robovie has two laser range sensors attached to its mobile base (about $10[cm]$ from the ground), one in the front and one in the back, in order to cover almost $360[\text{deg}]$ around the robot to detect obstacles.

1) Environmental Human Tracking Sensor System: A tracking system using laser range finders (LRF's) embedded in the environment was used to track the positions of people and localize the robot in the room. Six SICK LMS-200 laser range finders were placed around the perimeter of the room to minimize occlusions. They were set to a detection range of $80[m]$ with precision of $1[cm]$, each scanning an angular area of 180 deg at a resolution of $0.5[\text{deg}]$, providing readings of 361 data points every $26[ms]$.

The LRF's were mounted $85[cm]$ from the ground, a height chosen so the sensors could see above clutter and obstacles such as benches and luggage. Another reason for this placement was that at long range, the scan beams are spaced quite far apart (over $8[cm]$ apart at a range of $10[m]$) and detection of small features like legs is difficult. Detection of larger targets, like a torso, is more robust at these distances.

The sensors were connected directly to a central data acquisition PC in another room, which then streamed all sensor data to the tracking server. The tracking server performed background subtraction on the scan data to remove fixed environment features, then combined the foreground data from all sensors. Particle filters were used to track each entity (human or robot) in the environment according to the algorithm presented in [21], and the system was used to correct the robot's localization according to the method described in [23]. The accuracy of this system varies according to sensor placement but has been measured at +/- $6[cm]$ in field deployments.

2) Automatic Gaze Control System: The proposed automatic gaze control system follows the face and upper body of a person once the subject has been identified by the environmental human tracking sensor system.

The position of the person (in 2D coordinates) is continuously obtained from the environmental human tracking sensor system. The height at which the robot's camera gazes the person is determined by the use of trigonometry and considering the distance separating the robot and the person interacting with it. This relationship is bounded by an angle ranging between $57[\text{deg}]$ and $60[\text{deg}]$ at a minimum distance of $1[m]$ from the person.

The automatic gaze control is enabled through the graphical user interface presented in this study. The operator clicks on the representation of the person of interest in the GUI and the system determines the angle at which the camera should point to, in order to maintain the person's face in focus and allow the operator to observe the person's facial expressions and gestures.

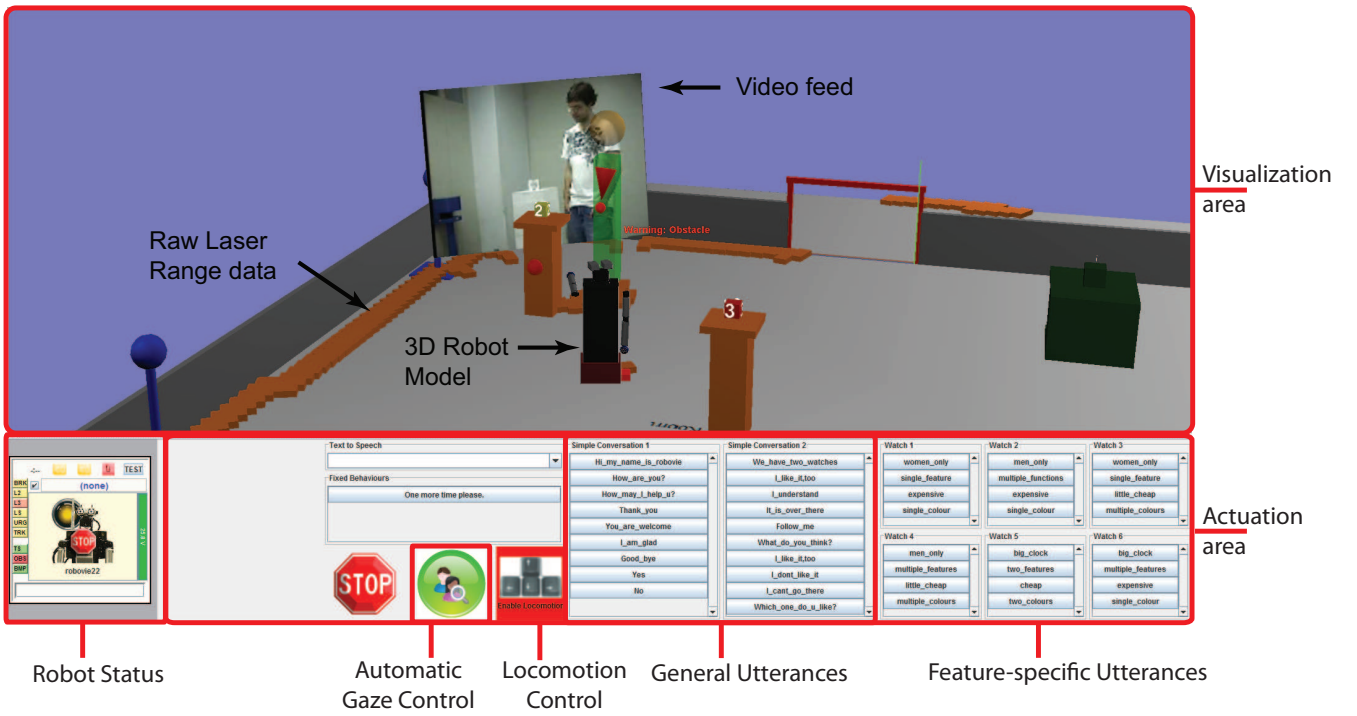


Fig. 1. GUI with the implemented visualization of spatial relationships and automatic gaze control.

B. Operator Side

The data gathered by the sensors onboard of Robovie and by the environmental sensory system (human tracker module) are presented to the operator through a 3D-based interface.

The proposed GUI combines the two factors discussed in Section III-B, and aims to allow the operator to identify and locate a person and objects of interest quickly, as well as to establish social distances accurately. Figure 1 shows an instance of the proposed system's GUI. The interface is divided in two sections: a visualization area (top) where a video feed is combined with a 3D model of the controlled robot and range data from laser sensors, and an actuation control area (bottom).

1) Visualization: The visualization comprises three main elements: map and object representations, video feed and robot representation.

Map and Object Representations

The map representation of the environment was generated using the *a priori* known locations of objects (desks, watch stands, etc.) within the environment. These objects do not move in order to make the environment a static one. 3D computer-generated models of walls, environmental laser sensors, stands and tables represent the different objects of interest in the environment. The laser range data representation is shown as small blocks on the ground.

Video feed

The GUI incorporates a video screen into the 3D environment, the movement of which is synchronized to the movement of the head of the robot. The video screen presents the image of the area at which the robot is looking.

In addition to helping the operator understand the environment in which the robot is located and avoid obstacles, video feedback can help the operator understand the intention of the person interacting directly with the robot.

Robot representation

It is important for the operator to understand the position, orientation and gestures of the teleoperated robot. In order to satisfy this requirement, a 3D model of Robovie II was implemented. This 3D model can represent the different movements of the limbs, head and position and orientation of the robot within the 3D environment. The operator observes the environment from a tethered point of view anchored $3[m]$ behind the head of the 3D model representation of the robot. In addition, the status of the robot and safety warnings are displayed. Information regarding the status of the robot such as battery and identification of the robot are presented in the lower left corner of the GUI as presented in Figure 1. Obstacles are shown spatially on the floor as yellow and red points and they represent the level of danger of navigating the robot in a particular direction. Yellow points represent obstacles that are in the vicinity of the robot but that would not cause any danger to the robot or the customer and red points represent obstacles that would do so. Safety warnings are also shown to bring the operator's attention to possible dangers during the navigation of the robot. These safety warnings are shown on top of the head of the robot's representation and as a drop-down message from the top of the 3D environment visualization. These warnings are intended to help the operator navigate more smoothly and avoid collisions with obstacles or people.

2) Actuation: The three actuation categories the operator can perform are: locomotion and pointing, utterances and gaze control.

Locomotion
The robot is able to move forward and rotate to the left and to the right around its own z-axis in order to reach a desired location. The operator drives the robot using the keyboard's arrow keys.

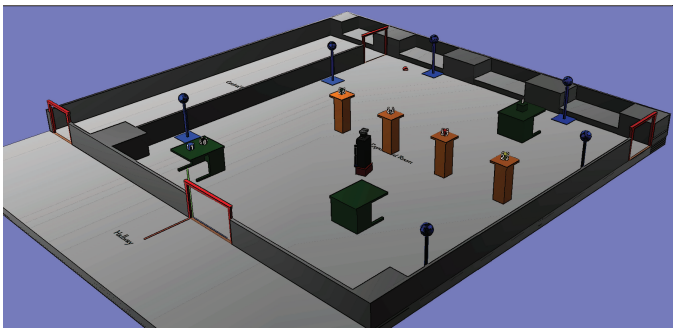


Fig. 2. 3D view of the watch shop.

Pointing
In addition to these translation commands, the operator can also point to a given position or object. The operator right-clicks a location or an object on the 3D environment and selects one out of two utterances the robot can say: "this one" or "that one".

Both of these actions can be performed through the use of the GUI or using a mouse and a keyboard.

Utterances
There are two different sets of utterances given to the operator: general and feature-specific. The general utterances are those utterances designed to help the operator have a smoother interaction with the person, i.e. "would you like to see some other product?". The feature-specific utterances have been designed to allow the operator to give specific information about an object of interest to the person the robot is interacting with, i.e. "this product costs 5,000yen". Both types of utterances are accessed by the operator by clicking on the button having the desired utterance's label. Some of the utterances are accompanied by head and arm gestures to make the robot more expressive.

Gaze
The operator manually controls the gaze of the robot by clicking on the video screen and dragging it to the direction where the operator wants the robot to look. The operator can enable the automatic gaze control by simply pressing a button on the GUI (Figure 1). When the automatic gaze control is used, the robot uses the data obtained from the human tracker module to calculate the location of the robot and person interacting with the robot. These data can be used then to calculate the vector at which the robot's head would look.

V. Experiment

An experiment was conducted to validate the combined effect of the visualization of spatial relationships and the automatic gaze control in the teleoperation of a mobile social robot. While the automatic gaze control is expected to help the operator better understand the facial expressions and gestures of the person interacting with the robot, the visualization of spatial relationships is expected to help the human operator better understand the location of objects in the environment and the robot, and in this manner avoid any possible "disorienting problem" from the automatic gaze control. In addition, the effect of the combination of these two factors on the operator's workload is verified.

A. Scenario

The scenario chosen for the experiment had a Robovie II playing the role of a shopkeeper at a simulated watch shop prepared up in an experiment room INSERT FIGURE HERE!!.

In this scenario, various clocks and watches are located on stands and tables, Figure 2 shows an example of one of the configurations for the location of each of the six watches and clocks. The robot would navigate within the shop showing customers different watches at different locations. A collection of six external laser range finders was used to localize the robot and customers in the environment.

B. Hypothesis and Prediction

The automatic gaze function is designed to help operators look at customers' face. As operators will be released from time-consuming operation when he/she control the robot's gaze to customers' face, we believe that the robot's gaze is more frequently actuated thanks to the automatic gaze function. We further hypothesized that it will reduce the operators' workload, thus operators will be more efficient in operating the robot which result in shorter interaction length. Furthermore, due to either reduced workload and/or more frequent chance of observing customers face and other behavior thanks to the frequent gaze control, better satisfaction from the customer will be obtained.

The visualization of spatial relationships is prepared to support operators to gain awareness to spatial relationships around the robot. Thus, we believe that operators will be more aware of surrounding situation, such as positions of the robot, customers, and watches. We hypothesized that it will reduce the operators' workload, and thus operator will be more efficient in operating the robot and thus interaction will be shorter, which will result in better satisfaction from the customer.

We are also interested in combination of these two factors, thus conducted the experiment having both factors at the same time.

C. Procedure

There were 29 undergraduate students (15 females and 14 males, in average 22 years old) who participated as an operator. In addition, two undergraduate students (1 female and 1 male) constantly participated playing the role of customers, and providing evaluations from the view point of the customer.

The operator participants had an introduction that included an explanation of task during the experiment. They were allowed to ask questions during this practice time to confirm their understanding of the different features of the GUI and

their role in the experiment. The operator participants were located in a separate room from the location where the robot was, and they never directly observed the room until the end of the experiment.

The order of the conditions at each experiment was counter-balanced to avoid a "learning-curve" effect. In addition, to prevent the operators from learning the positions of objects in the room, it was necessary to change the layout of the objects in the room, it was necessary to change the layout of the objects after every trial. Five layouts were created -one for the trainning session and one for each of the four trials. In these layouts, watches were placed an average of 2.3[m] from the center of the room, with a standard deviation of 1.2[m]. In preliminary trials, we had observed that placing watches in close proximity to each other increased the difficulty of the teleoperation task, so we attempted to make each of the layouts used in the experiment similar in difficulty, by placing one pair of watches within 0.8[m] of each other, and spacing the remaining four watches around the room, at least 1.2[m] apart.

1) Operator's Role: The role of the participant working as an operator was to control the robot to behave as a shopkeeper at the simulated watch shop. The operator's tasks included locating a customer who is wandering inside the watch shop, approach the customer and show and talk about the different watches or clocks to the customer based on the customer's non-verbal expressions. Based on a customer's facial expressions, for example, the operator should identify the interest or lack thereof in a given watch or clock and introduce different features of the current watch or guide the customer to another watch that may be of more interest to the customer.

2) Customer's Role: Each of the customer participants behaved as a customer for each session. Thus, for each session, there were two customers visiting the shop. The customer was instructed to walk into the watch shop and wander around until the robot approaches him/her. There is no scripted conversation; instead, the customer is given a situation and a watch that should be the target one. An example of a situation is that the customer will participate in a wedding and is interested in buying a watch. In order to make each interaction equivalent, the customer is also instructed to wait until at least 3 different watches have been presented to make a purchase. If none of the watches that have been presented within those 3 watches is the targeted one, the customer will wait until the robot presents the target one and finally purchase it.

We used two constant participants (customer participants) instead of inviting novice customers, because it is more difficult for novice people to provide an absolute evaluation to an interaction provided by a robot, as they are not sure what is good, what is bad, while they are excited by the novelty of a robot. They were instructed to provide the evaluation with a constant principle across trials.

D. Conditions

A 2×2 within-subjects experimental design was used with the following conditions:

- **Automatic Gaze Control** factor
  - Autogaze; in this condition, there is a button that enables the automatic tracking of the customers. This can be turned off by either pressing the button again, or manually moving the robot's head (via the GUI).
  - No-Autogaze; in this condition, the button is disabled, and the only way to control the robot's gaze (presumably to track and observe the customer) is direct manual control via the GUI.
- **Visualization of Spatial Relationships** factor
  - Spatial-Visualization; this condition adds 3D models of the objects (static, located in the room) and also avatar(s) of the persons (customers, keeping track of their current location; an example is provided in Figure 5).
  - No-Spatial-Visualization; in this condition, only the URG laser sensor raw data are shown, along with a 3D model of the robot, and the video feed coming from one of the robot's cameras (an example is provided in Figure 6).

E. Evaluation

A combination of subjective and objective techniques was employed to measure the performance of the operators in each condition as presented below.

- **Gaze time** was measured as the time the robot's gaze direction was actuated to face toward the customers or anywhere else (e.g. to seek for the location of the watch) either with manual control or automatic control.
- After each condition, subjective evaluations from the operator participants were provided to score on a 7-point Likert scale asking **the operator's awareness to surroundings** measured with an average of 7-point Likert scale items, asking for the awareness of the location of the robot, each customer, and each of the watches.
- **The operator's workload** was evaluated using a NASA-TLX test [24] that the operator had to complete after each condition. The result of this test is in a range between 0 and 100 points. Lower values represent lower workload, whereas higher values represent higher workload. In the context of this study, an operator was typically in a situation where he/she would suffer from overflow of tasks; thus, we consider that lower workload represents a situation where he/she were less suffered from such overflowing situation, and thus had a potential to respond to the customer in a more efficient way.
- As an indicator of how well the performance of the operator was, the authors timed the total **interaction length** of each condition. In our study, if an operator is efficient enough, the interaction length is supposed to be short. Customers were waiting for appropriate information to be provided, while operators were asked to identify the customers' interest and to choose the information contents to be presented by the robot. Clumsy operation and failure in identifying this situation would result in consumption of redundant time. Since the customers were waiting for information, such loss of time in a timely operation would

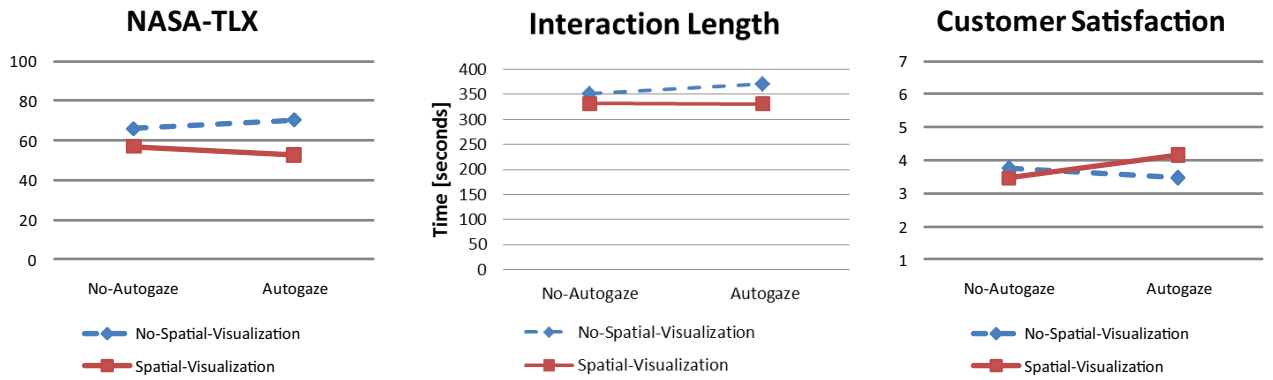


Fig. 3. NASA-TLX (left), Interaction Length (center) and Customer Satisfaction (right).

result in less engaging interactino and would lead to make customers bored.

- After each condition, subjective evaluations from the customer participants were provided to score on a 7-point Likert scale asking "how satisfactory was the robot's service?", i.e. **customer satisfaction**. In this scale, higher values represent higher satisfaction.

F. Hypothesis Testing

The results presented in Figure 3 share the following format: the blue dotted series represent the condition No-Spatial-Visualization, the red continuous series correspond to the Spatial-Visualization condition for the spatial relationships factor. The *x*-axis represents the "No-Autogaze" and "Autogaze" conditions corresponding to the experimental factor "automatic gaze control". A two-way repeated measures Analysis of Variance (ANOVA) was conducted with two within-subject factors, visual relationships and automatic gaze control, for all the results presented in this section.

1) Manipulation check: First, we confirmed that operators gained from the prepared functionality as designed. About automatic gaze, we analyzed "gaze time". There is a significant main effect revealed with automatic gaze control factor ($F(1,21) = 26.384$, $p < .001$, partial $\eta^2 = .557$), but no significant effect in the visualization of spatial relationships factor ($F(1,21) = .182$, $p = .674$, partial $\eta^2 = .009$) and interaction of these factors ($F(1,21) = 1.146$, $p = .297$, partial $\eta^2 = .052$). That is, as expected, gaze was more actuated in automatic gaze condition either with spatial visualization (avg. 233.4 sec, s.d. 196.9) or without spatial visualization (avg. 197.5 sec, s.d. 188.7) than manual gaze conditions either with spatial visualization (avg. 33.0 sec, s.d. 29.9) or without spatial visualization (avg. 46.5 sec, s.d. 40.0).

About spatial visualization, we analyzed "the operator's awareness to surroundings". There is a significant main effect revealed with both visualization of spatial relationships factor ($F(1,21) = 135.746$, $p < .001$, partial $\eta^2 = .866$) and automatic gaze control factor ($F(1,21) = 4.416$, $p = .048$, partial $\eta^2 = .174$). There interaction was not significant ($F(1,21) = 1.004$, $p = .328$, partial $\eta^2 = .046$). As expected, with spatial visualization (in automatic gaze condition: avg. 5.98, s.d. .745, in manual gaze conditions: avg. 5.44, s.d. 1.39) yield better subjective evaluation than without spatial visualization (in automatic gaze condition: avg. 3.38, s.d. 1.03, in manual gaze condition avg. 3.17 sec, s.d. .847).

These results confirmed our prediction. As designed, automatic gaze enabled more frequent actuation of gaze, and spatial visualization provided better awareness of surrounding spatial relationships.

2) NASA-TLX: The results measured by the NASA-TLX test are depicted in Figure 3 (center). A significant main effect was revealed with the visualization of spatial relationships factor ($F(1,21) = 14.693$, $p = .001$, partial $\eta^2 = .412$) but did not show significance with the automatic gaze control factor ($F(1,21) = .006$, $p = .939$ partial $\eta^2 = .000$). Interaction within these factors was significant ($F(1,21) = 4.984$, $p = .037$, partial $\eta^2 = .192$).

Simple main effects in the interaction were further investigated. Regarding visualization of the spatial relationship, there is significant simple main effects in case both with the automatic gaze ($p < .001$) and with manual gaze ($p = .041$). Regarding the simple main effect of automatic gaze control, there is significant trend when there is no spatial visualization ($p = .066$), but the comparison was not significant when there is spatial visualization ($p = .206$).

Overall, these results partially confirm our hypothesis. Spatial visualization affected in a way we hypothesized; however, contrary to what hypothesized, automatic gaze solely does not reduce the workload.

3) Interaction Length: The results measuring the interaction length are shown in Figure 3 (right). A significant main effect was revealed in the visualization of spatial relationships factor ($F(1,21) = 8.747$, $p = .008$, partial $\eta^2 = .294$). The interaction between these two factors did not present a significant effect ($F(1,21) = .798$, $p = .382$, partial $\eta^2 = .037$). No significant effect was shown by the automatic gaze control factor ($F(1,21) = 1.190$, $p = .288$, partial $\eta^2 = .054$). From these results it can be seen that when the operator was provided with the visualization of spatial relationships, interactions were shorter. These results support our hypothesis with respect to the effect from the visualization of spatial relationships. However, our prediction about the effect of automatic gaze was not confirmed.

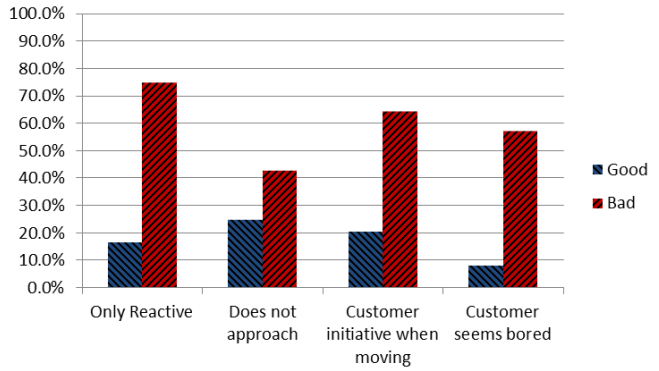


Fig. 4. Comparison between four event categories present in both high and low customer satisfaction cases: a) Robot is only reactive, b) Operator is disoriented, c) Customer ignores robot d) Customer is bored.

4) Customer Satisfaction: Figure 3 (right) shows the results corresponding to the customer satisfaction. No significant main effect was revealed for either the automatic gaze control factor ($F(1,21) = 2.094$, $p = .163$, partial $\eta^2 = .091$) or the visualization of spatial relationships factor ($F(1,21) = 1.817$, $p = .192$, partial $\eta^2 = .080$). The interaction between the visualization of spatial relationships factor and automatic gaze control factor was significant ($F(1,21) = 5.431$, $p = .030$, partial $\eta^2 = .205$). Simple main effects in the interaction were further investigated.

Regarding visualization of spatial relationship, there is the simple main effect was only significant when the conditions were with the automatic gaze ($p = .015$), but not significant when there is with manual gaze ($p = .250$). Regarding the automatic gaze control, the simple main effect was only significant when there is spatial visualization ($p = .003$), but not significant when there is no spatial visualization ($p = .367$).

Overall, these results partially confirm our hypothesis. It did not confirm our hypothesis about neither automatic gaze nor spatial visualization. Only when they are combined, significant effect were observed. The result indicates that both automatic gaze and spatial visualization should be simultaneously used to gain positive effect on customer satisfaction.

G. Qualitative Analysis of Interactions

In order to reveal the specific phenomena that contributed to customer satisfaction scores an analysis was made based on the video data obtained from the simulated environment and the screen of the operator.

The rationale of this analysis is to understand why some interactions could qualified as “good” (higher customer satisfaction score) or “bad” (lower customer satisfaction score) and which specific actions or behaviors led to them. The analysis consisted in categorizing several events the operators and customers incurred in during a single interaction and observing how they affected the quality of the interaction and thus the customer satisfaction.

To observe the causes of the contrast between a good and a bad interaction, the data were organized in four blocks each containing 25% of the entire data. For the scope of this paper, the top and bottom ends of these data were studied.

The next step in this analysis was to select the behavior categories, within this spectrum, where the number of “bad” interaction events registered was larger than the number of “good” interaction events. The authors believe a difference favoring the “bad” interaction events should help understand which actions or lack thereof that may have strongly influenced the customer satisfaction.

Two independent coders analyzed recorded video and audio from these trials to identify whether any of these four behavior patterns occurred. Trials were presented to the coders in an arbitrary order, and the coders were not informed as to whether the trials belonged to the ”good” or ”bad” set. The coders used the following criteria to judge the four behavior patterns:

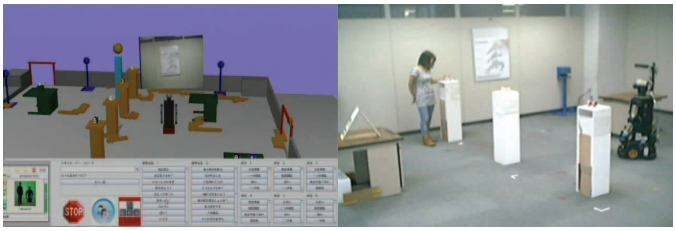
a) **Only reactive**: The robot was judged to be behaving reactively if long periods of silence occurred without the robot taking action, or if the robot uttered very few spontaneous utterances that were not in response to a customer’s question.
b) **Does not approach**: If the coders judged that the robot should have approached the customer but did not, or if the robot’s approach was very late, they considered this behavior to have occurred. If the customer approached the robot immediately and the robot did not need to move to meet the customer, they did not consider this behavior to have occurred.
c) **Customer initiative when moving**: If the customer moved to another watch before the robot mentioned that watch or began moving towards the watch, this behavior was considered to have occurred.
d) **Customer seems bored**: This was judged by the coders subjectively based on the customer’s expressions and actions.

The final results were averaged between the two coders. The raw agreement of their scores was 79.8%, and Cohen’s Kappa was computed to be 0.58. The final results, presented in Figure 4, show that all four of these behavior patterns occurred much more often in the ”bad” set of trials than in the ”good” set.

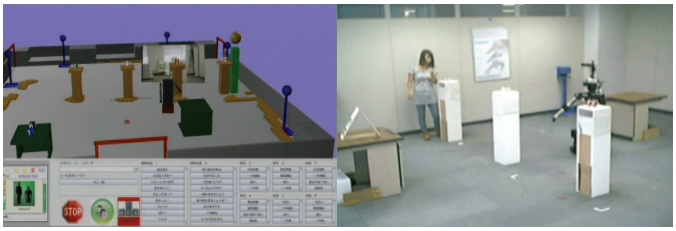
In order to have smooth, rich and dynamic human-robot interactions it is important to enhance the performance of the operator considering these four issues. These observations support the results presented in this paper by showing that an operator not being aware of the environment and the objects surrounding it spends vital time searching for them. In addition the operator may convey the wrong message to the customer by moving the robot in an awkward manner while trying to point to or find an object.

H. Observations

The authors present two interaction examples in this section; one without visualization of spatial relationships or automatic gaze control, and one with both. These examples present an insight on how operators teleoperated the robot under different conditions and the effect of these conditions on their performance. In particular, these two cases represent how the operator benefited from the techniques developed in this research when they were available and how the operator was handicapped when they were absent in the GUI. In

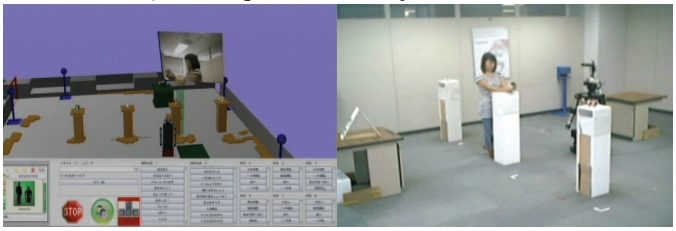


a) Operator finds customer by looking at the interface



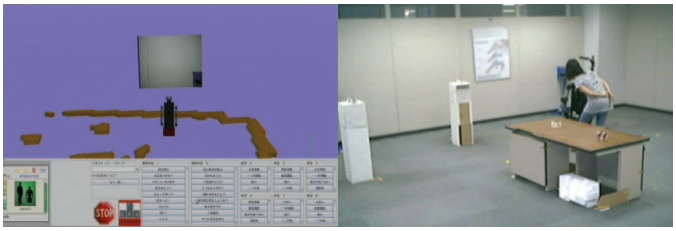
b) Robot points correctly to a watch

c) Operator automatically tracks the customer



d) Operator observes customer's facial expressions

Fig. 5. Natural interaction achieved using visualization and autogaze.



a) Operator looks for the customer

b) Operator finds the customer and interacts

c) Robot spins looking for another watch

d) Robot points incorrectly and loses the customer

Fig. 6. Poor interaction when using no-visualization and no-autogaze.

addition, these examples exhibit the behavior categories listed and analyzed in Section V-G.

1) Case 1: With visualization and with autogaze: This case (Figure 5), serves as an example of a smooth interaction. The operator having an understanding of where objects are in the environment and allowing the robot to track the customer, is able to take an initiative in the interaction by showing the customer a watch in another location. In this particular case it is valuable to notice that the robot does not appear disoriented and furthermore, it has constant body language and spoken interaction with the customer. A transcript of this continuous interaction is given as follows (where, Robot = R and Customer = C):

R: Operator finds the customer just by looking at her 3D representation. "One moment please" (Operator is not disoriented).

R: "Are you looking for something?" (Robot leads interaction) (Figure 5(a)).

Customer: "Yes, I'm looking for a watch to give as a present".

R: Robot moves towards a watch and faces it. "I see."

R: Pointing correctly to the watch, "How about this one?" (Operator is not disoriented) (Figure 5(b))

C: Operator enables automatic gaze control to observe the customer's facial expressions, Customer moves towards the watch. "This one, huh?" (Customer does not ignore the robot and is not bored) (Figure 5(c)).

R: While looking at customer's face. "I would recommend this one" (Robot leads interaction) (Figure 5(d)).

In this example, it can be seen how the operator can differentiate a stand or a watch from a person only by looking at their respective visualizations as 3D models. This case illustrates that the proposed visualization in combination with the automatic gaze control allows the operator to react faster, improving the customer satisfaction and decreasing the operator's workload (as shown in the results presented in Section V-F).

2) Case 2: Without visualization and without autogaze: When the operator has limited understanding of the location of objects in the environment, he/she spends more time looking

for them through the manual use of the robot's camera. This additional actuation task forces the operator to manually control the robot which may produce awkward social behaviors. The interaction case provided in this section shows how the lack of the proposed techniques in this paper result in the presence of the behavior categories identified in Section V-G.

A transcript of this continuous interaction presented in Figure 6 is shown:

R: Operator tries to find the customer spinning the robot (Operator is disoriented) Figure 6(a).
C: Looks at the robot to try to understand its intention.
R: After a long pause the operator finally finds the watch that wants to show. "How about these one?". Figure 6(b).
C: Looks at the watch.
R: "This watch has many functions. It comes in one color".
C: "Really?"
R: Operator tries to look at customer's face (Operator is disoriented). "I like it. Do you want to see another watch?"
C: "Yes".
R: Operator looks around, apparently confused, by spinning the robot. Figure 6(c)
R: Pointing to an incorrect location. "How about that one?".
C: Does not know which watch the robot talks about, long pause (Customer is bored and ignores the robot). After the pause asks for confirmation. "This one?" Figure 6(d).

This example serves to portrait the consequences of lack of a visualization of the objects in the environment. For instance, the operator relies on the video feed in order to understand where static or dynamic objects are. This makes the operator look for these objects by spinning the robot, which can translate into a socially awkward behavior. As presented in Section V-F, this affects negatively the performance of the operator since it increases the operator's workload.

VI. Discussion

A. Summary

The results of our study indicate that when: a) an operator has an understanding of the spatial relationships, and b) the level of actuation the operator has to perform is decreased through automation of necessary and/or routine tasks, the operator can more effectively control the robot in social interactions.

In our setting, the visualization of where the persons and the objects are, combined with automatic gaze control that frees the operator from tracking the person in order to observe them and thus determine their intentions, has resulted in improved customer satisfaction, that could be related to the reduced operator workload.

However, it was observed that the automation of the gaze, by itself, did not enhance the customer satisfaction. The automatic gaze control enabled the operator to effectively observe the facial gestures of the customer while being aware of the surroundings of the environment. An appropriate visualization of the spatial relationships of the environment, as the proposed in this paper, allows the operator to have such intuitive understanding. If this visualization is not available while the automatic gaze control is, the operator may incur in continuous socially awkward movements of the robots head and body which in turn may convey an erroneous message to the customer.

Therefore, the authors would argue towards an approach in teleoperation architecture design that incorporates both the visualization of spatial relationships and the automation of processes that are necessary within an HRI context to aid the operator in improving their understanding of human non-verbal communication and which are crucial for social interactions.

This approach has applications both for teleoperated systems (for improving the operator performance), but also for research towards fully-automated systems, as first steps towards understanding the requirements necessary to implement the social processes to be automated (such as the automatic gaze control in our current work).

B. Limitations

In our current work, the robot can keep track of a single person within its field of view. However, it is conceivable that in a different social context, the robot would have to interact with multiple people at the same location (e.g. guiding a crowd at a museum). In the future, this could be augmented by additional mechanisms that e.g. automatically determine the gaze of the person or any pointing gestures. The visualization of spatial relationships currently relies on *a priori* knowledge of a static environment, as well as the existence of environmental sensors. Both of these limitations may be addressed by using traditional robot navigational and localization techniques and also by relying on on-board sensors. The addressed interaction in the study was rather limited to be simple. This was because of our aim to study the phenomena at the operators' side. Nevertheless, the overall effect to customers' satisfaction when robots will be used in a real field would be affected by various factors, e.g. context role, expectation, design of interaction, thus we consider that the obtained result about customer satisfaction should be carefully interpreted.

VII. Acknowledgments

The authors would like to acknowledge Florent Ferreri, Kyle Sama for their invaluable help on the development of the proposed system and the experiment carried out and presented in this paper. This research was supported by the Ministry of Internal Affairs and Communications of Japan.

References

[1] D. F. Glas, T. Kanda, H. Ishiguro, and N. Hagita, "Simultaneous teleoperation of multiple social robots," in *HRI '08: Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction*. ACM, 2008, pp. 311–318.

[2] K. Zheng, D. F. Glas, T. Kanda, H. Ishiguro, and N. Hagita, "How many social robots can one operator control?" in *Human-Robot Interaction (HRI), 2011 ACM/IEEE 6th Annual Conference on*, March 2011, pp. 379–386.

[3] A. Green, H. Huttenrauch, and K. Eklundh, "Applying the Wizard-of-Oz Framework to Cooperative Service Discovery and Configuration," in *Robot and Human Interactive Communication, 2004. ROMAN 2004. 13th IEEE International Workshop on*, 2004, pp. 575–580.
[4] A. Steinfeld, O. C. Jenkins, and B. Scassellati, "The oz of wizard: Simulating the human for interaction research," in *ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, March 2009.
[5] N. Dahlbäck, A. Jönsson, and L. Ahrenberg, "Wizard of oz studies: why and how," in *Proceedings of the 1st international conference on Intelligent user interfaces*. ACM, 1993, pp. 193–200.
[6] B. Pitzer, M. Styer, C. Bersch, C. DuHadway, and J. Becker, "Towards perceptual shared autonomy for robotic mobile manipulation," in *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, May 2011, pp. 6245 –6251.
[7] J. L. Burke, R. R. Murphy, M. D. Coovert, and D. L. Riddle, "Moonlight in miami: A field study of human-robot interaction in the context of an urban search and rescue disaster response training exercise," *Human Computer Interaction*, vol. 19, pp. 85–116, 2004.
[8] P. Wells and D. Deguire, "Talon: A universal unmanned ground vehicle platform, enabling the mission to be the focus," *Unmanned Ground Vehicle Technology VII*, vol. 5804, no. 1, pp. 747–757, 2005.
[9] B. Yamauchi, "Packbot: A versatile platform for military robotics," in *Proceedings of SPIE 5422*, 2004, pp. 228–237.
[10] M. Buehler, K. Iagnemma, and S. Singh, *The 2005 DARPA Grand Challenge: The Great Robot Race*. Springer Publishing Company, Incorporated, 2007.
[11] J. L. Drury, L. Riek, and N. Rackliffe, "A Decomposition of UAV-related situation awareness," in *HRI '06: Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*. ACM/IEEE, March 2006, pp. 88–94.
[12] J. L. Drury, J. Scholtz, and H. A. Yanco, "Awareness in Human-Robot Interactions," in *Proceedings of the IEEE Conference on Systems, Man and Cybernetics*, October 2003, pp. 111–119.
[13] J. Carff, M. Johnson, E. M. El-Sheikh, and J. E. Pratt, "Human-robot team navigation in visually complex environments," in *Proceedings of 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, October 2009, pp. 3043–3050.
[14] R. Meier, T. Fong, C. Thorpe, and C. Baur, "A sensor fusion based user interface for vehicle teleoperation," in *International conference on field and service robotics (FSR*, 1999, pp. 279–286.
[15] C. W. Nielsen and M. A. Goodrich, "Comparing the usefulness of video and map information in navigation tasks," in *HRI '06: Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*. ACM, 2006, pp. 95–101.
[16] J. L. Drury, B. Keyes, and H. A. Yanco, "LASSOing HRI: analyzing situation awareness in map-centric and video-centric interfaces," in *HRI '07: Proceedings of the ACM/IEEE international conference on Human-robot interaction*. New York, NY, USA: ACM, 2007, pp. 279–286.
[17] C. Nielsen, M. Goodrich, and R. Ricks, "Ecological interfaces for improving mobile robot teleoperation," *Robotics, IEEE Transactions on*, vol. 23, no. 5, pp. 927–941, oct. 2007.
[18] H. Kuzuoka, K. Yamazaki, A. Yamazaki, J. Kosaka, Y. Suga, and C. Heath, "Dual ecologies of robot as communication media: thoughts on coordinating orientations and projectability," in *CHI '04: Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 2004, pp. 183–190.
[19] S. Woods, M. Walters, K. Koay, and K. Dautenhahn, "Comparing Human Robot Interaction Scenarios Using Live and Video Based Methods, Towards a Novel Methodological Approach," in *Advanced Motion Control, 2006. 9th IEEE International Workshop on*, 2006, pp. 750 –755.
[20] D. F. Glas, T. Kanda, H. Ishiguro, and N. Hagita, "Field trial for simultaneous teleoperation of mobile social robots," in *HRI '09: Proceedings of the 4th ACM/IEEE international conference on Human robot interaction*. ACM, 2009, pp. 149–156.
[21] D. F. Glas, T. Miyashita, H. Ishiguro, and N. Hagita, "Laser-based tracking of human position and orientation using parametric shape modeling," *Advanced Robotics*, vol. 23, no. 4, pp. 405–428, 2009.
[22] A. A. Nofi, "Defining and measuring shared situation awareness," in *Center Naval Anal.*, Nov. 2000.
[23] D. F. Glas, T. Miyashita, H. Ishiguro, and N. Hagita, "Automatic position calibration and sensor displacement detection for networks of laser range finders for human tracking," in *Proc. IEEE/RSJ Int'l Conf. Intelligent Robots and Systems (IROS'10)*, 2010, pp. 2938–2945.
[24] S. G. Hart and L. E. Staveland, "Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research," in *Human Mental Workload*, 1988, pp. 139–183.

Andres Mora Andres Mora was born in San Jose, Costa Rica. He received his B.S. in Electronics Engineer from the Universidad Interamericana de Costa Rica and his M.Sc. and Ph.D in Aerospace Engineering (Space Robotics) in 2006 and 2009 respectively at the Space Robotics Laboratory, Tohoku University, Japan. His research interests span path planning, teleoperation and control of mobile robots and the design of graphical user interfaces for mobile robots.

Dylan Glas Dylan F. Glas received S.B. degrees in aerospace engineering and in earth, atmospheric, and planetary science from MIT in 1997, and he received his M.Eng. in aerospace engineering in 2000, also from MIT. He has been a Researcher at the Intelligent Robotics and Communication Laboratories (IRC) at the Advanced Telecommunications Research Institute International (ATR) in Kyoto, Japan since 2005. From 1998-2000 he worked in the Tangible Media Group at the MIT Media Lab. His research interests include networked robot systems, teleoperation for social robots, human-machine interaction, ubiquitous sensing, and artificial intelligence.

Takayuki Kanda Takayuki Kanda (M04) received his B. Eng, M. Eng, and Ph. D. degrees in computer science from Kyoto University, Kyoto, Japan, in 1998, 2000, and 2003, respectively. This author became a Member (M) of IEEE in 2004. From 2000 to 2003, he was an Intern Researcher at ATR Media Information Science Laboratories, and he is currently a Senior Researcher at ATR Intelligent Robotics and Communication Laboratories, Kyoto, Japan. His current research interests include intelligent robotics, human-robot interaction, and vision-based mobile robots.

Norihiro Hagita (M85 SM99) received his Ph.D. degree from Keio University (Japan) in 1986 in electrical engineering and joined Nippon Telegraph and Telephone Public Corporation (NTT) in 1978. He engaged specially in developing handwritten character recognition. He also stayed as a visiting researcher at Prof. Stephen Palmer's lab in University of California, Berkeley (Dep. of Psychology) during 1989-1990. He is currently the director of ATR Intelligent Robotics and Communication Laboratories (IRC) at the Advanced Telecommunications Research Institute International (ATR) in Kyoto, Japan.