A Teleoperation Approach for Mobile Social Robots

Andres Mora Dylan F. Glas Takayuki Kanda Norihito Hagita
{andresmora, dylan, kanda, hagita}@atr.jp

***Abstract*— Mobile social robotics is the field of robotics that deals with interacting with humans in everyday environments such as malls, eldery care centers, museums, etc. Currently, artificial intelligence, speech recognition and other technologies have not reached a level of reliability which would enable robots to be completely autonomous in these situations. In this paper, the authors identify unique requirements for the teleoperation of mobile social robots and developed a system that allowed to examine their effects in a human-robot interaction. In particular, the teleoperation of mobile social robots requires operators to understand facial gestures and other non-verbal communication of the person interacting with the robot, in order to facilitate an improved human-robot interaction. It then becomes critical to provide the operators with tools that increase their comprehension of such communication and of the physical environment in which the robot is involved. A study where a robot plays the role of a shopkeeper was conducted to demonstrate that when the operator is freed from following a customer's face using an automatic control of the robot's gaze, unexpected side effects such as operators getting disoriented and navigation became inaccurate occured. It was demonstrated, however, that when the operator was given a 3D representation of the spatial relationships of the simulated shop in addition to the automatic gaze control, these side effects were solved and the quality of the interaction with the customer improved.**

I. Introduction

The quest for a "natural" interaction with a human is what strongly motivates the development of new methodologies and technologies in the field of social human-robot interaction (HRI).

Due to limitations in today's artificial intelligence and recognition technologies, full automation of social robots will not be possible for some time. However, these limitations can be bypassed by using an person operating the robot (from now on referred to as operator) to perform certain recognition and decision-making tasks.

Using teleoperation to augment partial autonomy is useful not only for laboratory studies (using Wizard-of-Oz (WOZ) methods [12], [18], [4]) but will also be valuable in actual deployments of social robots in commercial applications for safety and legal reasons. The development of social robots which incorporate teleoperation brings together two very different branches of 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 situation awareness, shared autonomy, and operator workload for the remote operation of non-social robots.

As these fields traditionally focus on very different problems, little research has explored the intersection between them, leaving many questions unanswered. How do teleoperation systems for social robots differ from those for traditional robots? 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 of the robot's head was designed to keep the person interacting with the robot within its field of view and relief the operator from this routine task. Although this automatic gaze control has benefits such as reducing the operator's actuation tasks and increasing the operator's awareness of the customer's state, including facial expressions and gestures; it also has some drawbacks.

These drawbacks include an unintended disorienting side effect on the operator (given the operator's reduced awareness of the state of the robot) and thus could not use the video information effectively for navigational purposes. This reduced awareness is due to the large attention the operator gives to the video presenting the customer's face which results in a lower understanding of the location and orientation of the robot within its environment, and the locations of objects and people around the robot.

To overcome this difficulty, a 3D representation of the robot's environment which augments the operator's understanding of spatial relationships was created. The results of an experiment where a robot plays the role of a shopkeeper, demonstrate that when this representation was present, the awareness issues were effectively tackled and in addition, the benefits of our proposed automatic gaze control were also retained.

In this paper, establishes that the teleoperation system must provide the operator with an appropriate representation of spatial relationships when the operator uses our proposed automatic gaze control, and when both factors are available, the operator can improve the quality of the human-robot interaction.

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

A. Mora, D. F. Glas, T. Kanda and N. Hagita are with Advanced Telecommunications Research Institute International, 2-2-2 Hikaridai, Keihanna Science City, Kyoto, Japan, 619-0288

opposite approaches along the ends of a spectrum: being completely teleoperated by humans [2], [19], [21] or being fully automated [1]. Some of the aspects of research on teleoperation involve increasing and maintaining the level of situational awareness of the@operator@[7], [8], combining mixed and virtual reality techniques to help the operator improve the navigation of the robot [3], 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 [15], [17]. One notable finding could be summarized as the need to combine different types of information altogether [6], [16]. 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 in the field of HRI 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 are 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 [14], focus is given to the "ecology" among operators and customers. In this 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 [3], [6], proximity behavior [7], [8], the way of social dialog [12], and social patterns [20]. 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 [9], attention [10] and engagement [11]. Although new techniques are constantly being developed, the robots' capabilities in context-sensitive interactions have remained highly limited.

III. Design Principles

Previous knowledge on the teleoperation of mobile robotics has been mainly focused on navigational robots whereas little is known about the teleoperation of mobile social robotics. The basic design of our teleoperation system was created according to this previous knowledge on teleoperation for navigational robots. Additional techniques for the teleoperation of social robots in particular are also proposed.

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 [16].

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 [17]. This study reports that the integration of map and video information in a 3D-based GUI positively affected the navigation of the robot.

From a design perspective, Nielsen et al. 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, particularly on [17], the authors have implemented a GUI that serves as a baseline to our study. This baseline incorporates laser range data, a video feed and a 3D model of the robot used in this research.

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 relieving the operator from routine tasks and the second one helps the operator retain the awareness that may be lost by providing the operator with autonomy.

1) Automatic Gaze Control: The teleoperation of mobile social robots requires the operator to be capable of navigating the robot smoothly and safely through an environment, to identify people and obstacles and to be able to interact with them accordingly. A critical requirement for such system is to allow the operator to observe the facial expressions and gestures of the person interacting with the robot. Typically, this information is provided to the operator through a video feed; in this way, the operator can understand the intentions of the person interacting with the robot. 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 consequently, 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 of continuously following a customer's face in order to decrease the workload in terms of actuation and increase the concentration of the operator on the customer's facial expressions. However, this level of attention of the operator into the video feed may rise problems in the awareness of the operator, e.g. the location of static and dynamic (people) objects in the areas of the environment not shown by the limited field of view of the video feed. In addition, if the robot is required to interact with multiple objects in the environment, the operator would need to actively observe their locations. Relying solely on the video feed would force the operator to create a "mental map" to remember where several objects are in the environment [5].

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 [17] 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 $360[\deg]$ around the robot to detect obstacles.

The representation of people and their real-time position within the environment is based on the tracking data provided by a human tracker module. This module combines the range data from six SICK©laser sensors located around the interaction environment. The human tracker module detects and tracks people using a technique that is based on the algorithm presented in [11]. The estimation of the positions for a person is done by particle filters and a contour-analysis technique is used to estimate the direction in which a person is facing [10].

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 within the environment, these locations are considered static. 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 (Refer also to Figure 4).

Video feed
The GUI incorporates a video screen into the 3D environment which movement 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 $3[m]$

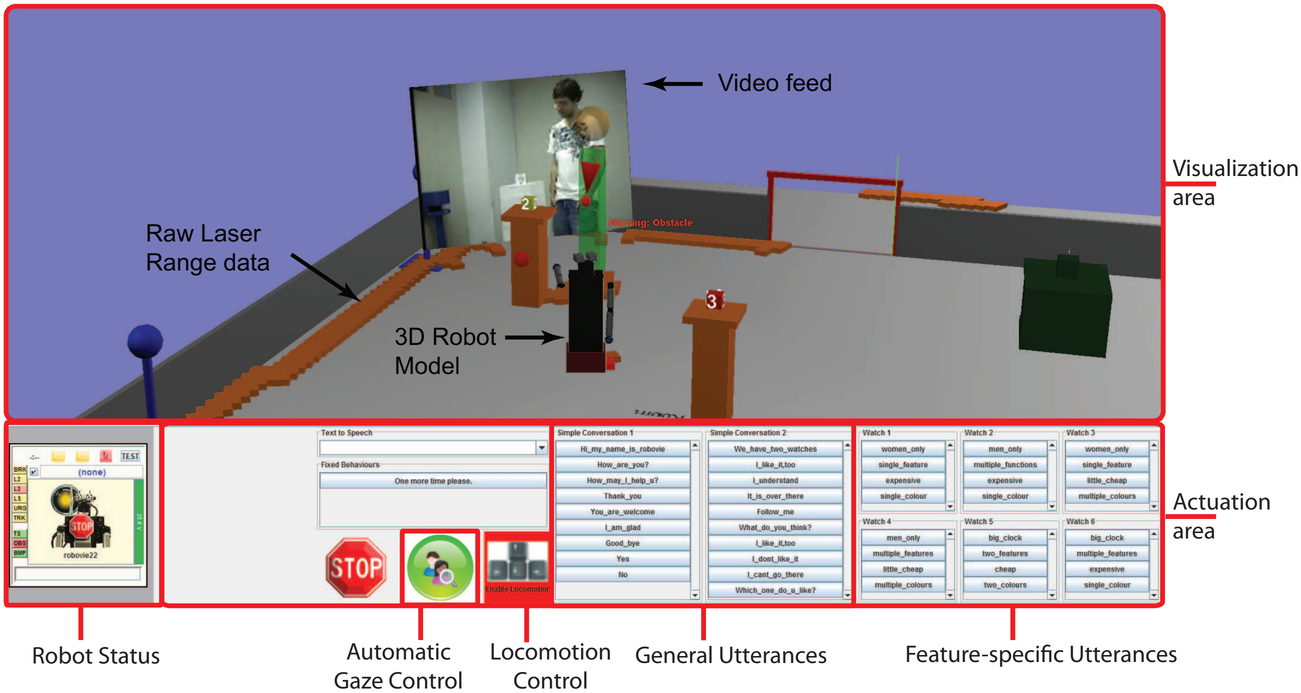


Fig. 1. GUI with the implemented visualization of spatial relationships and automatic gaze control

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. Safety warnings are shown on top of the head of the robot's representation and as a pull-down message from the top of the 3D environment visualization, if objects are within an area considered dangerous to the robot. 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.

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 cost 5,000yen". Both types of utterances are accessed by the operator by click 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 effects 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. In this scenario, various clocks and watches are located on stands and tables.

B. Procedure

The participants of this experiment were 31 undergraduate students both female (16) and male (15), with an age average of 22 years old. There were two type of participants: two constant participants that played the role of customers and participants that worked as operators.

The participants teleoperating the robot had an introduction that included an explanation of the experiment and their role during the experiment was given. The operators then had a 15-minute period of practice during which they could control the robot just as if they would in the real experiment, during this period the location of the watches was randomized. 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 operators 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. After each trial, the position of the stands and tables where the watches and clocks were changed to have different layouts. The layouts were also counter-balanced.

1) Operator's Role: The role of the participant working as an operator was 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 nonverbal expressions. Upon the facial gestures, 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: The customer is 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. 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 this watch and finally purchase it.

C. Conditions

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

- **Visualization of Spatial Relationships** factor
  - 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.
  - 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).
- **Automatic Gaze Control** factor
  - Autogaze; in this condition, a button 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.

D. Hypothesis and Prediction

The proposed automatic gaze control and the 3D representation of spatial relationships are expected to complement each other. The utilization of autonomy to handle the robot's gaze control, should reduce the operator's workload and reduce the possibility that the operator will lose track of the customer's face and not see expressions. The customer satisfaction should increase due to an improved navigation combined with a better positioning of the robot with respect to objects and humans and a faster reaction time during the interaction. The representation of spatial relationships should provide the means by which the operator can transition between observing the customer's face and observing what the customer is referring to in the environment not covered by the video.

Therefore, the authors expect that having automatic gaze control and having the visualization of spatial relationships would increase operator understanding of customer intention by:

- reducing operator workload,
- increasing customer satisfaction, and
- decreasing interaction time.

The combination of the visualization of spatial relations factor and the automatic gaze control factor is expected to result in the highest level of customer satisfaction and present the lowest level of workload. In the other hand, not having either the visualization of spatial relations factor and automatic gaze control factor should result in the lowest level of customer satisfaction and present the highest level of workload.

E. Evaluation

A combination of subjective and objective techniques were employed to measure the performance of the operators in each condition:

- **Customer's satisfaction** was evaluated by asking the customers after each condition to score on a 7-point Likert scale "how satisfactory was the robot's service?".

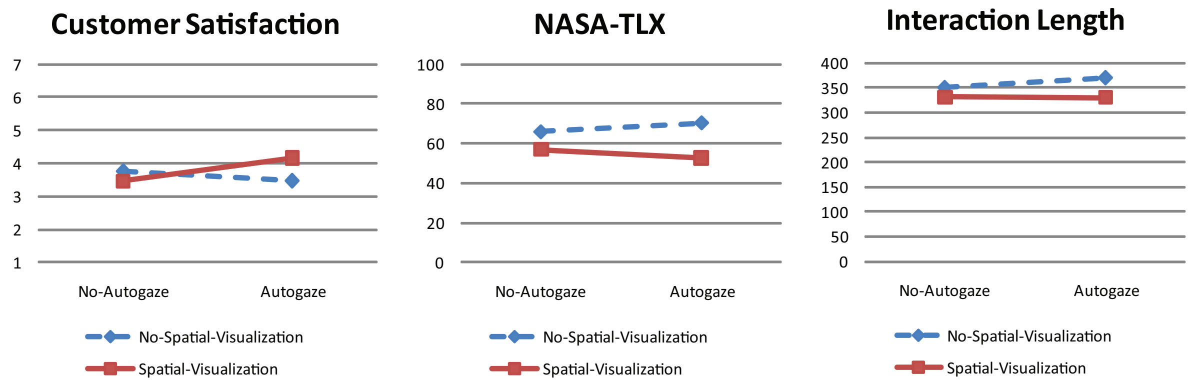


Fig. 2. Customer Satisfaction(left), NASA-TLX(center) and Interaction Length(right)

- **The operator's workload** was evaluated using a NASA-TLX test[13] that the operator had to complete after each condition.
- 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, longer interactions are regarded as inefficient.

VI. Results

A. Verification of Hypothesis

The results obtained from the experiment described above and presented in Figure 2 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.

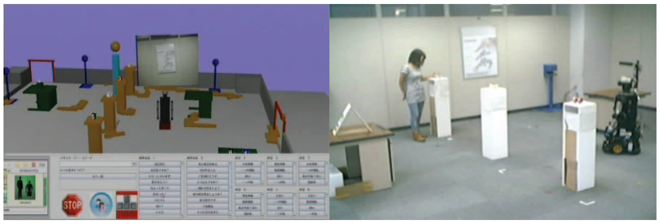
The results presented in Figure 2 (left) support our prediction that the combination of visualization of spatial relationships and automatic gaze control would positively affect the customer's satisfaction. However, the factors by themselves did not contribute to increase the customer's satisfaction. The interaction between the visualization of spatial relationships factor and automatic gaze control factor was significant ($F(1, 21) = 5.431$, $p = .030$, partial $\eta^2 = .205$). The automatic gaze control factor indicated significance for Spatial-Visualization ($p = .003$), however for No-Spatial-Visualization no significant difference was revealed ($p = .367$). The visualization of spatial relationships factor indicated significance for Autogaze ($p = .015$), however for No-Autogaze no significant difference was revealed ($p = .250$). 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 $\eta^2 = .294$).

The results measured by the NASA-TLX test depicted in Figure 2 (center), confirm our prediction that the combination of visualization of spatial relationships and automatic gaze control decrease the workload of the operator. However, the automatic gaze control factor alone does not support this hypothesis. A significant main effect 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 $\eta^2 = .192$). The visualization of spatial relationships factor revealed a significant effect for No-Autogaze ($p = .041$) and it also indicated a significant effect for Autogaze ($p = .000$). The automatic gaze control factor did not reveal a significant difference for either No-Spatial-Visualization or Spatial Visualization.

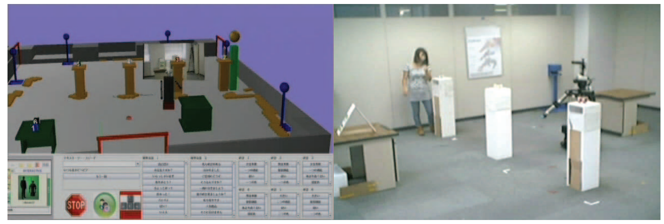
The results depicted in Figure 2 (right), support our hypothesis regarding the visualization of spatial relationships, however, the automatic gaze control does not support our prediction. A significant main effect in the visualization of spatial relationships factor ($F(1, 21) = 8.747$, $p = .008$, partial $\eta^2 = .080$). No significant effect was shown by the automatic gaze control factor ($F(1, 21) = 1.190$, $p = .288$, partial $\eta^2 = .054$). The interaction between these two factors did not present a significant effect ($F(1, 21) = .798$, $p = .382$, partial $\eta^2 = .037$).

These results support our prediction that the representation of spatial relationships and automatic gaze control complement each other and that when combined, they have a positive effect on the customer's satisfaction. It was also demonstrated that the factors by themselves did not contribute to increase the customer's satisfaction.

The fact that the automatic gaze control by itself did not contribute to increase the customer's satisfaction may be due to an absence of an external point of reference from which the operator can understand where the robot is located at any given moment. The operator spends time observing the facial expressions and gestures of the customer and becomes "used" to this view. However, when the customer points to another location or asks about an object in another location, the field of view of the camera is insufficient and the operator must look within the 3D environment for the corresponding



a) Operator finds customer by looking at the interface

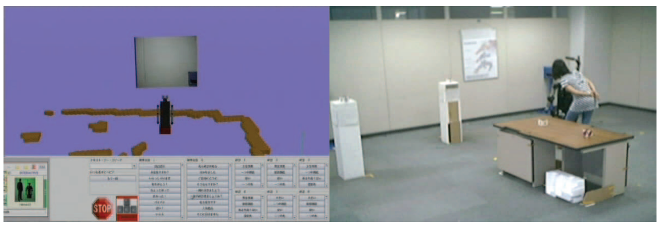


b) Robot points correctly to a watch

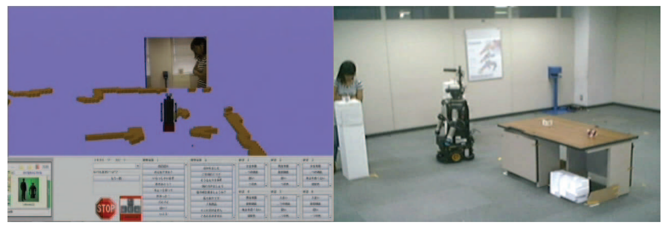
c) Operator automatically tracks the customer

d) Operator observes customer's facial expressions

Fig. 3. 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. 4. Poor interaction when using no-visualization and no-autogaze.

desired location. If the operator is not provided with an appropriate representation of the spatial relationships of the environment, the operator must remember were the desired object's location is, confirm its location by looking at it and finally, while navigating towards the target location, use the robot's camera to avoid collisions. This increased effort translates in longer interactions and "unnatural" interaction with the customer. Therefore, while it is important to alleviate the operator's workload by introducing shared autonomy in the teleoperation system, it is essential to provide, at the same time, the operator with the tools to understand the spatial relationships governing the environment where the robot is located.

B. 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.

1) Case 1: With visualization and with autogaze: This case (Figure 3), 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. 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".
R: "Are you looking for something?" (Figure 3(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?" (Figure 3(b))

C: Operator enables automatic gaze control to observe the customer's facial expressions, Customer moves towards the watch. "This one, huh?" (Figure 3(c))
R: While looking at customer's face. "I would recommend this one". (Figure 3(d))

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 this 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 VI-A.

2) Case 2: Without visualization and without autogaze: When the operator has limited understanding of the location of objects in the environment spends more time looking for them through the use of the robot's camera. This additional actuation task forces the operator to manually control the robot which may produce awkward social behaviors. A transcript of this continuous interaction presented in Figure 4 is shown:

R: Operator tries to find the customer spinning the robot Figure 4(a).
C: Looks at the robot to try to understand its intention.
R: After a long pause the operator finally finds the intended watch. "How about these one?". Figure 4(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. "I like it. Do you want to see another watch?"
C: "Yes".
R: Operator looks around, apparently confused, by spinning the robot. Figure 4(c)
R: Pointing to an incorrect location. "How about that one?".
C: Does not know which watch the robot talks about, long pause. Asks for confirmation. "This one?" Figure 4(d).

When there is lack of a visualization of the objects in the environment, 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 VI-A, this affects negatively the performance of the operator since it increases the operator's workload.

VII. 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 focus more on the overall task. 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. 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.

VIII. 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] M. Buehler, K. Iagnemma, and S. Singh. *The 2005 DARPA Grand Challenge: The Great Robot Race*. Springer Publishing Company, Incorporated, 2007.

[2] 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*, 19:85–116, 2004.

[3] 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*, pages 3043–3050, October 2009.

[4] 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*, pages 193–200. ACM, 1993.

[5] B. DeJong, J. Colgate, and M. Peshkin. Improving teleoperation: reducing mental rotations and translations. In *Robotics and Automation, 2004. Proceedings. ICRA '04. 2004 IEEE International Conference on*, volume 4, pages 3708–3714, apr. 2004.

[6] 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*, pages 279–286, New York, NY, USA, 2007. ACM.

[7] 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*, pages 88–94. ACM/IEEE, March 2006.

[8] 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*, pages 111–119, October 2003.

[9] 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*, pages 311–318. ACM, 2008.

[10] 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*, pages 149–156. ACM, 2009.

[11] D. F. Glas, T. Miyashita, H. Ishiguro, and N. Hagita. Laser-based tracking of human position and orientation using parametric shape modeling. *Advanced Robotics*, 23(4):405–428, 2009.

[12] 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*, pages 575–580, 2004.

[13] S. G. Hart and L. E. Staveland. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Human Mental Workload*, pages 139–183, 1988.

[14] 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*, pages 183–190. ACM, 2004.

[15] 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*, pages 279–286, 1999.

[16] C. Nielsen, M. Goodrich, and R. Ricks. Ecological interfaces for improving mobile robot teleoperation. *Robotics, IEEE Transactions on*, 23(5):927–941, oct. 2007.

[17] 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*, pages 95–101. ACM, 2006.

[18] 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.

[19] P. Wells and D. Deguire. Talon: A universal unmanned ground vehicle platform, enabling the mission to be the focus. *Unmanned Ground Vehicle Technology VII*, 5804(1):747–757, 2005.

[20] 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*, pages 750 –755, 2006.

[21] B. Yamauchi. Packbot: A versatile platform for military robotics. In *Proceedings of SPIE 5422*, pages 228–237, 2004.