Snapshots for Semantic Maps*

Curtis W. Nielsen, Bob Ricks,
Michael A. Goodrich
CS Dept. Brigham Young University
Provo, UT, USA
{curtisn, cyberbob, mike}@cs.byu.edu

David Bruemmer, Doug Few, Miles Walton
Idaho National Engineering and
Environmental Laboratory
Idaho Falls, ID, USA
{bruedj, fewda, mwalton}@inel.gov

Abstract – *A significant area of research in mobile robotics is in the local representation of a remote environment. In order to include a human in a mobile robot task it becomes important to present the remote information efficiently to a human. A relatively new approach to information presentation is semantic maps. Semantic maps provide more detail about an environment than typical maps because they are augmented by icons or symbols that provide meaning for places or objects of interest. In this paper we present snapshot technology as a means to take pictures from the real world and store them in a semantic map. To make the snapshots and semantic map available to an operator, we identify and discuss general attributes for useful displays and present a mixed reality 3D interface that meets the requirements. The interface and snapshot technology are validated through experiments in real and simulated environments.*

Keywords: Snapshots, Semantic Maps, Mixed Reality, 3D Interface, Mobile Robots, Human-Robot Teams

1 Introduction

Human-robot interaction is an area of research that is gaining momentum in mobile robotics. Of importance to the success of a human-robot team is the ability of the human to be aware of the robot's situation. Endsley defined situation awareness as, "the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning and the projection of their status in the near future"[5]. Additionally, Dourish and Bellotti define awareness as, "...an understanding of the activities of others, which provides a context for your own activity"[4]. This implies that a human-robot team's performance is directly related to the human's understanding of the situation around the robot.

Most mobile robots have the ability to communicate some information about their situation back to the operator, so it becomes necessary to present available information in a manner that is useful and intuitive. The presentation of information depends largely on the task at hand. In fact, a display that is incompatible with the task dooms the human-robot team to sub-par performance [15]. Thus, it is important to consider the frame of reference through which the user interacts with the robot as well as the requirements of the task. Wickens and Hollands have identified two types of tasks that are typical to human-robot systems: tasks involving understanding and tasks involving navigation [15]. Typically, tasks involving navigation require an egocentric reference frame while tasks involving understanding require a more exocentric frame of reference [2][14]. The question then arises of what reference frame to use when performing tasks that have both navigation and comprehension requirements. Our solution to this problem is to make the frame of reference adjustable so the operator can obtain a useful perspective for the task at hand.

In addition to creating an interface that supports the user in a variety of changeable tasks, we have developed snapshot technology as a means to store visual information directly in the interface. Snapshot technology allows a user to take pictures of the environment with the camera on a remote robot and store the images at the corresponding position in a local map representation. This technology is aimed at shifting some of the memory requirements of the operator to the interface such that the user is able to focus on other tasks.

This paper will proceed as follows. We first discuss requirements for useful displays and continue with our solution to the requirements. As part of the solution we will describe snapshot technology, semantic maps, and a mixed reality 3D display we have developed. We conclude the paper with experimental results.

2 Requirements for Useful Displays

Situation awareness is a key to successful human-robot teams. The ability of the user to understand the situation around a robot depends largely on the display through which the user interacts with the robot. For a display to be considered useful and effective we re-

*0-7803-8566-7/04/$20.00 © 2004 IEEE.

quire three features. First, it must allow the user to store information in the display. Second, the interface must integrate sensor information into a single coherent display. Finally, it must allow the user to adjust their perspective of the environment to match the task at hand. We discuss each of these requirements in the following sections.

2.1 Information Storage

In tasks where an operator is required to remember where objects are, or what was happening at various places in the environment, it is typically left up to the operator to remember the information. As the complexity of such a task increases or the amount of information that must be remembered increases, it quickly becomes likely that the human will forget some information or their recollection will deteriorate. To reduce the memory requirements on the human operator we require an interface that facilitates information storage.

2.2 Integrate Sensor Information

To accurately store information in the interface that correlates with the actual environment, we require that the information from the robot be integrated into a single display. The information from the robot includes many things including video, laser readings, sonar data, map information, and position data.

In many research environments, the display through which the user interacts with the robot provides a separate window for each of the types of information. A typical interface is shown in Figure 1. As can be seen in the figure, the information from each sensor is shown in its own panel within the display. There is a separate view for the laser data, the sonar data, the image data, and the map data. Such an interface forces the user to mentally combine the different sensory information into an awareness of the situation around the robot. In contrast, an integrated display presents the user with a view of the environment that combines the various sources of data into a single display such that the user is relieved of the mental effort of combining the information [11].

2.3 Adjustable Perspectives

Information storage and integrated displays are important concepts for reducing the mental workload of the user. In addition, research has shown that certain displays are better suited to certain tasks. Specifically, navigation tasks are better performed with egocentric displays and spatial understanding tasks are better performed with exocentric displays [15][2]. Adjustable perspectives enable the user to interact with the robot efficiently regardless of the task at hand. Having an adjustable display enables the human-robot team to embark upon more complex tasks where the requirements of the human-robot team might change throughout the experiment.

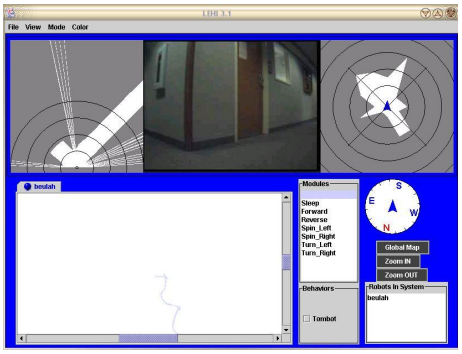

Figure 1: An example of a 2D interface. The various sources of information are displayed in separate panels within the display. In this case, laser data is on the top left, video is on the top center, sonar data is on the top right, map information is on the bottom left and compass information is on the bottom right.

3 Technology for Useful Displays

With the requirements for useful displays set forth, we next present the technologies we developed for useful displays along with the philosophies behind the various technologies.

3.1 Transactive memory

In order to discuss the implementation of information storage within a display we first look at the cognitive science notion of *transactive memory*. Transactive memory is a term that was first introduced by Wegner as the "operation of the memory systems of the individuals and the process of communication that occur within the group" [13]. In Wegner's definition, he is referring to individuals as the storage container for this transactive memory. When someone has expertise in a field, then a good friend of that individual can have access to the information by asking their friend as opposed to remembering everything on their own. Thus, transactive memory is a form of external memory. It is well known that people use external memory for a variety of common memory tasks from appointments to shopping lists to daily events recorded in a journal [6] [9] [13]. Examples of places where information is stored in external memory include such things as a PDA, a calender, or even a scratch piece of paper.

In order to use these forms of external memory, it is important to have a storage device in place that is easy to access. Then the person desiring to find the information does not have to remember the details of the information, just where to find it. This frees the person's mind to focus on other tasks.

Similarly, information available to the operator in a human-robot team can be overwhelming unless the user has a means to store the information in an easily accessible manner. We stated earlier that for an interface to be useful, the operator must have the ability to store

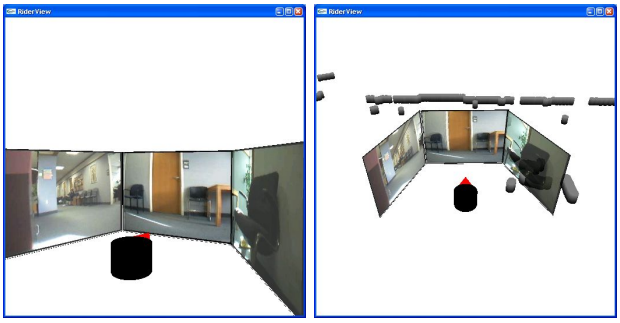

Figure 2: Some snapshots placed within a map to give more information about that place in the environment.

and quickly access the information within the display. One method we have developed for storing information in the interface is via snapshots.

3.2 Snapshot Technology

The idea behind snapshots is that visual images contain a lot of information that is understandable by a human, but not necessarily a computer. In human-robot tasks that involve object recognition and recollection, it is important to aid the user by storing relevant information in the interface, rather than forcing the user to mentally store the information. Consider the case of navigating a robot through an environment looking for objects. Suppose that, at the end of the navigation, the operator is required to tell an administrator, for example, where all of the blue boxes in the environment are located. If the environment is sufficiently large, the operator will likely forget where some of the objects were located.

To aid the user in search and identification tasks, we have created *snapshot technology*. Snapshots are pictures that are taken by the robot and stored at the corresponding location in a map. In Figure 2 we show some snapshots taken from the robot. In the figures, the robot took three pictures from three directions. The pictures are used to show a panoramic view of the visual information around the robot. To take a snapshot, a user indicates the request via a button on the joystick. Upon receiving the user's request, the robot saves the current image along with the position and orientation of the robot when the picture was taken. The position and orientation of the robot are found using Konolige's large-scale map-making algorithm [7]. The snapshot information along with the recorded pose of the robot is then returned to the interface and displayed at the corresponding location and orientation in the operator's view of the map.

In search or identification tasks, the snapshots in the display are an implementation of the aforementioned transactive or external memory. By adding the snapshots to the user's perspective of the map, we make the visual information available to the user whenever they need more information about a corresponding place in the environment.

As an example of the usefulness of snapshot technology consider the following: suppose that part way through a patrolling task a supervisor asks if the operator has seen anything suspicious. If the interface does not support snapshots, the user will have to remember if they observed something, what it was, and where it happened. In contrast, by empowering the user with the ability to record information directly into the display, the necessary information is already correlated with the map of the explored environment. This makes the recollection of a previous experience very accessible to the operator.

In the future, we are interested in allowing the robot to take its own snapshots of the environment when it finds objects or items of interest. In such situations, the operator might simply observe the progress of the experiment and upon arrival of a snapshot, review the findings of the robot and give further directions based on the new information from the robot. Additionally, if the robot becomes disoriented or loses its way, snapshots could be used as a means to alert the operator to the situation of the robot.

Thus, snapshots currently provide a method to store information within a display. In the future, snapshots might also be used to give the user more context of the situation around the robot when the robot needs assistance.

The introduction of snapshot technology leads us to a broader external storage medium, namely semantic maps.

3.3 Semantic Maps

A relatively new approach to information storage is *semantic maps*. Semantic maps can be thought of as a map of an environment that is augmented by information that supports the current task of the operator. Semantics simply gives meaning to something; therefore, a semantic map gives meaning to places on the map. The information that is stored in a semantic map might include snapshots, laser readings, sonar data, map information, and video.

As an example, consider an occupancy grid-based map. The map by itself does a good job of portraying to the operator where the robot can and cannot go. However, with such a map, the user and robot will have difficulty understanding where "Bob's chair" is located, or how to move to "Mike's Door". It is virtually impossible for the robot to learn where Bob's chair is without any user input. By placing semantic information into the map and tying it directly to places in the environment, the human and robot are able to reason about the environment semantically.

Principles of semantic maps have been addressed previously by other researchers. Most notably Kuipers in-

troduced the notion of a spatial semantic hierarchy as a model of large-scale space with both quantitative and qualitative representations. The model is intended to serve as a method for robot exploration and map building and a model for the way humans reason about the structure of an environment [8]. Additionally, Chronis and Skubic have presented a system that allows a user to sketch a map and a route for the robot to follow on a PDA [3]. This map and path are an example of a semantic map where the map made of obstacles is augmented with route information which gives the user an understanding of what the robot will be doing.

3.4 Virtual 3D Display

With the ability to store information inside the interface via snapshots, we next look to the requirement of integrating information into a single display. The challenge with creating a single display is to combine both real and virtual elements representing the remote environment into a single display that is intuitive and supports interaction with the remote environment [12].

When creating a representation of a remote environment, the representation will be primarily virtual because we simply cannot gather and present all the available information from the environment. However, there is some real information available that we want to portray within the virtual environment. For this reason we need a display that supports both virtual and real elements. In the literature there are many terms for displays with virtual and real elements including *virtual environments*, *augmented virtuality*, *augmented reality*, and *mixed reality* [1] [10].

Our display is a mixed reality representation that combines real video with virtual range, map, and robot information into a 3D interface. In a previous version of the integrated display, we combined sonar, laser, and video information into a static display [11]. In the current version of the display we add an occupancy grid-based map and the snapshot technology. Thus, the virtual environment serves as a semantic map because it stores more information than the simple afforded behaviors found in most maps. Figure 3 shows a view of our display. The dark rectangles represent walls or objects identified by the mapping algorithm and the robot model is drawn at its current location with respect to the discovered map. The robot model is also scaled to match the size of the actual environment, thereby enabling the user to comprehend the relative position of the robot in the real environment. A texture mapped plane with the video stream is rendered a small distance in front of the robot, perpendicular to the orientation of the robot [11]. As the robot moves through the environment the visual information displayed by the texture map is updated to match the video stream. Additionally, another texture mapped plane with a snapshot is

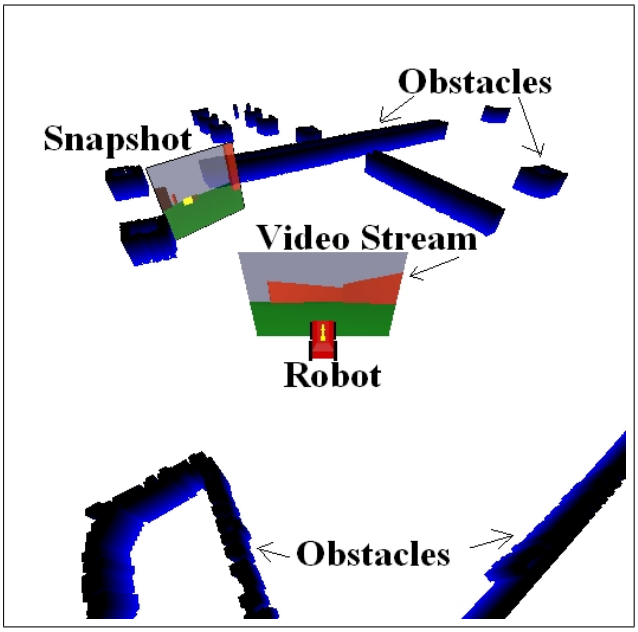

Figure 3: A view of our mixed reality 3D display with video feed, robot position, an occupancy grid-based map, and a snapshot.

shown to the left of the robot. In contrast to the video stream, the snapshot will not change as the robot moves.

In order to display the occupancy-grid based map, we render the occupied squares as three dimensional rectangles and we do not render the unknown or unoccupied squares. As the map is built and neighboring grid cells are identified as occupied, an obstacle begins to form in the display representing the location of the obstacle in the environment. The wall in the display takes on height, width and depth and offers the operator an *in-perspective* view of the map-building process.

In addition to providing an interface that displays the relevant information in a useful way, we desire that the interface be able to support the user in a variety of tasks. We next present an adjustable perspective as a means to supporting the user in complex tasks.

3.5 Adjustable Perspective

In complex tasks it is feasible to imagine that the human-robot team will be required to perform different tasks. Due to the fact that certain displays are better suited to certain tasks [15], we have made the view of our interface adjustable to the needs of the human-robot team. The user adjusts the perspective by manipulating the zoom, pitch, and yaw of the field of view. This is currently implemented in one of two ways: either by clicking on buttons or dragging the mouse. As the display is changed toward a more egocentric perspective, the display gives more support to tasks involving navigation. To support spatial reasoning tasks, the display can be changed to an exocentric perspective. Figure 4 shows a zoomed-in, tethered perspective useful for navigation

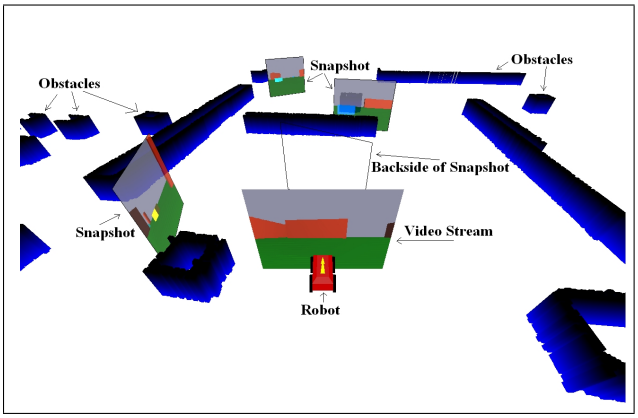

Figure 4: A zoomed-in, tethered perspective of the mixed reality 3D display that supports more egocentric tasks such as navigation and object identification.

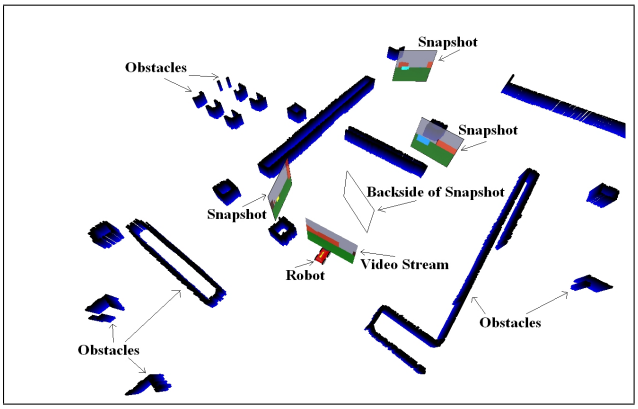

Figure 5: A zoomed-out untethered perspective of the mixed reality 3D display that supports more exocentric tasks such as spatial understanding or global planning.

or identification tasks and Figure 5 shows a zoomed-out untethered perspective of the same environment useful in spatial reasoning tasks. In both figures we see the placement of the robot and the video stream with respect to the map as well as four snapshots. One of the snapshots is displayed as an outline of a rectangle to indicate that we are looking at the back of the snapshot. (We only render the image if we can see the front of the snapshot with the current perspective.)

One observation within the displays is that without the snapshot technology the user would not know what is behind the wall in front of the robot. However, with the snapshot technology and the proper perspective, the user can quickly identify the objects behind the wall.

4 Validation

In a recent experiment at the St. Louis Science Center, 56 volunteers were asked to drive a robot and build a map of a large room containing various obstacles. The volunteers were split into two groups where 23 users were given a video stream and a 2D representation of

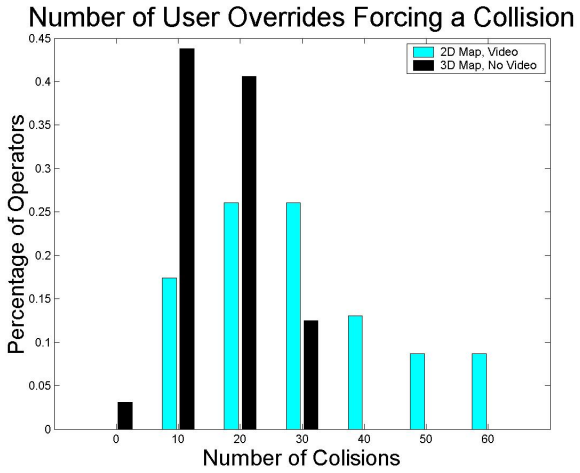

Figure 6: A graph representing the number of collisions experienced by operators when driving with a 2D map and video or with a 3D map and no video.

the map as it was built by the robot, and 33 users were given a prototype version of the mixed reality 3D map that did not have any video stream. The map was built using Konolige's algorithm [7], and is the same underlying map for both the 2D and 3D representations. The only difference was how the map was presented to the user.

The performance of the participants was based primarily on the time to completion. Using the 2D interface, the fastest any user finished was 289 seconds. The average time was 584 seconds with a standard deviation of 233. In contrast, the slowest any user finished when using the 3D interface was 255 seconds with an average time of 191 seconds and a standard deviation of 18. That means that the slowest user in the 3D interface was still 12% faster than the fastest person with the 2D interface. The average time for the 3D interface was 67% faster than the average time for the 2D interface.

Additionally, the number of times that the user directed the robot to move into an obstacle and the robot took initiative to protect itself was recorded. A histogram of the results is shown in Figure 6. Of particular interest is the fact that 87.5% of the 3D display operators had less than 20 collisions with an average of only 11 and a standard devition of 6.56. In contrast, only 43.5% of the 2D display operators had less than 20 collisions with an average of 24 and a standard deviation of 15.65.

Following the experiments, each user was asked to rate their feeling of control over the robot on a scale from 0 to 10, with 0 being 'no control' and 10 being 'complete control'. The distribution of the participants subjective feeling of control is shown in Figure 7.

We find the subjective results remarkable in that such a large percentage (over 65%) said they felt they had no control over the robot with the 2D interface.

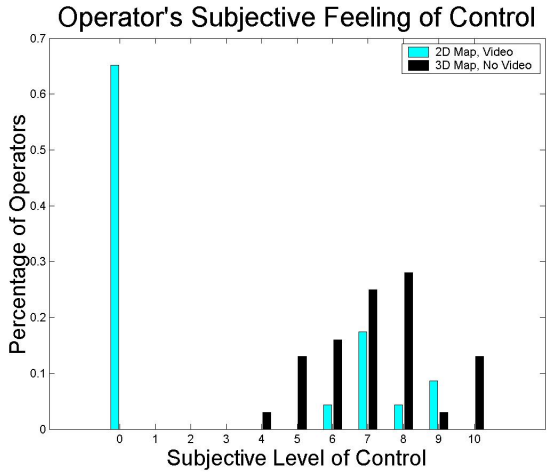

Figure 7: The subjective feeling of control as indicated by the operator when using a 2D map with video and when using a 3D map without video.

In addition to the objective and subjective measurements of performance, we also monitored the number of messages sent between the robot and the interface and joystick bandwidth. These numbers provide an indication of the amount of workload the user is experiencing. The messages to the robot decreased 77% and the messages from the robot decreased 69% with the 3D interface compared to the 2D interface. Additionally, the joystick turning bandwidth decreased 54% and the translation bandwidth decreased 56% with the 3D interface compared to the 2D interface.

This indicates that not only does the 3D interface improve performance, but it actually decreases workload on the operator. The results surprised us and we will continue to perform experiments to further validate our findings.

5 Summary and Future Work

In this paper we presented snapshot technology as a means to take pictures from the real world and store them in a semantic map. The semantic map is based on an occupancy-grid and rendered using a mixed reality 3D interface that presents the information from the robot in an intuitive display.

Experiments have shown that the 3D interface reduces workload and increases performance in comparison to typical 2D interfaces in navigation based tasks. Furthermore, the operators tend to feel that they are in better control of the robot with the 3D display

In the future we plan to continue our user studies to identify the principles that govern the success of the 3D interface. We will also experiment with various icons that can be placed in the map to aid the user in maintaining awareness of the robot. Additionally, we will make the display more interactive so the user can select virtual objects in the interface and label them with an icon or a descriptive phrase such that the human and robot can communicate directly via the interface.

We are also looking into expanding the robot's autonomy such that a human-robot team can handle situations with adjustable levels of human and robot control.

6 Acknowledgments

This work was partially supported by DARPA under grant NBCH1929913.

References

[1] R. T. Azuma. A survey of augmented reality. *Presence: Teleoperators and Virtual Environments*, 6(4):355–385, August 1997.

[2] W. Barfield and C. Rosenberg. Judgments of azimuth and elevation as a function of monoscopic and binocular depth cues using a perspective display. *Human Factors*, 37:173–181, 1995.

[3] G. Chronis and M. Skubic. Sketch-based navigation for mobile robots. In *Proceedings of the IEEE 2003 International Conference on Fuzzy Systems*, St. Louis, MO, May, 2003.

[4] P. Dourish and V. Bellotti. Awareness and coordination in shared workspaces. In *Proceedings of the ACM conference on Computer-supported cooperative work (CSCW-92)*, pages 107–114, Toronto, Ontario, 1992. New York: ACM.

[5] M. Endsley. Design and evaluation for situation awareness enhancement. Paper presented at the Human Factors Society 32nd Annual Meeting, 1988.

[6] J. Harris. External memory aids. In M. Gruneberg, P. Morris, and R. Sykes, editors, *Practical Aspects of Memory*, pages 172–180. London: Academic Press, 1978.

[7] K. Konolige. Large-scale map-making. In *Proceedings of the National Conference on AI(AAAI)*, San Jose, CA, 2004.

[8] B. Kuipers. The spatial semantic hierarchy. *Artificial Intelligence*, 119:191–233, 2004.

[9] J. Meacham and B. Leiman. Remembering to perform future actions. In U. Neisser, editor, *Memory Observed*, pages 327–336. San Francisco: Freeman, 1982.

[10] P. Milgram, H. Takemura, A. Utsumi, and F. Kishino. Augmented reality: A class of displays on the reality-virtuality continuum. In *SPIE: Telemanipulator and Telepresence Technologies*, Boston, MA, 1994.

[11] B. Ricks, C. Nielsen, and M. Goodrich. Ecological displays for robot interaction: A new perspective. to appear IEEE/RSJ International Conference on Intelligent Robots and Systems, 2004.

[12] T. B. Sheridan. *Telerobotics, Automation, and Human Supervisory Control*. MIT Press, Cambridge, MA, 1992.

[13] D. Wegner. Transactive memory: A contemporary analysis of the group mind. In B. Mullen and G. Goethals, editors, *Theories of Group Behavior*, pages 185–208. New York: Springer-Verlag, 1986.

[14] C. Wickens, C.C.Liang, T. Prevett, and O. Olmos. Egocentric and exocentric displays for terminal area navigation. *International journal of aviation psychology*, 6:241–271, 1996.

[15] C. D. Wickens and J. G. Hollands. *Engineering psychology and human performance Third Edition*. Prentice Hall, Upper Saddle River NJ, 2000.