ABSTRACT

BOYCE, CURTIS WESLEY. Augmented Reality as a Perceptual Aid in Robot Teleoperation. (Under the direction of Dr. Robert St. Amant).

Robots have reached a level of sophistication such that it is now productive to include robot-assisted search teams in urban search and rescue scenarios. The perceptual interface to the human operator is frequently a video display that relays images from a camera on the robot. Research into issues surrounding human-robot interaction has suggested that if more intelligence were placed on the robot, it could lessen the load on the human operator and improve performance in the search task. In this thesis we examine the relative performance in a search task where the robot is controlled by a remote human operator under different simulated environmental conditions. Three distinct environmental conditions are represented by: a nominal display from the robot's camera, an artificially degraded display (representing signal interference) of the same information, and an augmented display of the degraded information, in which object recognition algorithms enable the application of a visual overlay to identify a search target. Results show that reduced performance under the degraded condition can be improved to near the nominal level by this augmentation method.

AUGMENTED REALITY AS A PERCEPTUAL AID IN ROBOT TELEOPERATION

by
CURTIS WESLEY BOYCE

A thesis submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Master of Science

COMPUTER SCIENCE

Raleigh

2006

APPROVED BY:

Dr. Christopher G. Healey	Dr. R. Michael Young

Dr. Robert St. Amant
Chair of Advisory Committee

To my wife Sarah whose inspiration, like the normal distribution, is continuous.

Biography

Curt Boyce was born on March 2, 1965. He developed a fondness for programming on his first computer, a Commodore VIC-20 purchased during high school. This interest in computer science lead to undergraduate work at both University of California, San Diego and San Diego State University, culminating in a Bachelor of Science degree at the latter in 1989.

After graduation, he went to work for Unisys Corporation in San Diego for one year prior to transferring to Houston, Texas to work on the space shuttle program at NASA's Johnson Space Center. After a five year stint in Texas and meeting his wife on the job, the call of the Golden State beckoned. Taking a job with Lockheed-Martin developing software for the MILSTAR satellite systems provided the opportunity to return to California. After a year of satellite software he jumped the fence between Lockheed-Martin and Moffett Field to take a position with Sterling Software developing the real-time data system for the wind tunnels of the National Full-Scale Aerodynamic Complex (NFAC) at NASA Ames. After 3 years at NASA Ames, and becoming the Team Lead, family obligations necessitated relocating to his wife's home state of North Carolina.

Curt worked briefly as a contractor for Nortel Networks before leaving to join the startup company ViOS. Upon the exhaustion of venture capital at ViOS he returned to contracting at GlaxoSmithKline (GSK) in 2001. At the completion of the contract at GSK, opportunity and attitude allowed for the return to academic pursuits. He joined the Masters program in Computer Science at North Carolina State University in Spring 2003. He accepted a permanent position at GSK near the end of the Spring 2003 semester, and continued working full-time during the completion of his studies and research.

Acknowledgements

I would like to thank my advisor, Dr. Robert St. Amant for his insight, understanding, and patience with the protracted changing research that begat this thesis. His enthusiasm for the field of human computer interaction inspired me to attempt this degree. The use of his ER1 robot was invaluable as a learning tool, and became the focal point my thesis research.

I would also like to thank Mario Munich, Charles Baker and the team at Evolution Robotics for their gracious assistance with incorporating the Evolution libraries into my Augmented Reality Robot GUI application.

Table of Contents

List of Tables

List of Figures

1. Introduction

The issues associated with the human-computer interface have been well researched by the Human-Computer Interaction (HCI) community and are relatively well understood. For an interface in a Human-Robot Interaction (HRI) scenario however, there are the added complexities associated with the physical environment, robot configuration, communication bandwidth, and system response time. These issues often combine to create a discrepancy between the intended actions of the human operator (HO) and the resultant action of the robot. For this study we are concerned with robot teleoperation; where the HO controls the actions of a robot from a remote location. This teleoperation can be across the solar system or across the hall, either way the same complexities of HRI need to be addressed. For HRI to become practical in real world scenarios the effects of environment, robot hardware, bandwidth limitations and system lag time need to be mitigated.

The physical environment in urban search and rescue scenarios can often times be inhospitable to humans either due to atmospheric conditions or confined space issues. These constraints necessitate the ability for the robot to provide a perceptual interface for the human operator situated outside the search area. The perceptual interface is most frequently a video display presenting images from a camera mounted on the robot [5].

Robot configurations are as variable as the imagination of their human operators. From tethered track vehicles to free roaming dog like quadrupeds, each has its own particular sphere of usefulness. The predominant configuration used in urban robot-assisted search activities utilizes a tether along which communication lines are run to handle the control interface and robot feedback [5][13]. For this experiment it was much more convenient and cost effective to utilize a wireless communication channel; regardless, I believe the concepts investigated translate across platforms.

The communication channel, whether it is wireless, USB, Ethernet or something else, has a limitation to how much information can be sent back and forth in a fixed time frame. The communication bandwidth places constraints on items such as video resolution, the amount and frequency of sensor data returned, and the level of control available to the user. The wireless communication bandwidth directly affected decisions about the perceptual interface adopted for this study, which will be discussed later.

The system response time can be affected by the communication bandwidth (such as for the Mars rovers to use an extreme example [7]), but is also highly dependent on processing time available on-board the robot. Response time is specifically a consideration when image processing is done on the robot, as in the case of the ER1 in this study. When the response time becomes slow enough to affect perceptual feedback from control input it is perceived as system lag. Lag time can become a consideration for standard HCI environments, but in HRI environments it is a significant consideration because damage to robot hardware can result from a loss of control. For this reason, in teleoperation scenarios every effort is made to mitigate lag time to a manageable level by prudent selection of attributes during system configuration.

The specific task of interest in this study is robot-assisted search; a paradigm which is coming into its own in real-world application. Often the primary interface between human and robot is a video display that relays images from a robot mounted camera. It is because robot-assisted search is a perceptual activity that we need to focus on providing the best perceptual information available [13].

If we focus on the perceptual interface to the operator, then we should endeavor to capitalize on known features of human perceptual processing. If we can discover a way to impart intelligence to the systems onboard the robot such that the robot knows for what the HO is looking, then we can give cues to the HO via the display interface when there is an object of interest in view so as to direct their attention to the appropriate region. The distinction between preattentive and attentive perceptual processes was first posited by Neisser. The preattentive perceptual processes are automatic, rapid, and occur in parallel [15]. When an object in a display is easily discerned as being present, without having to actively search for that object, it is said to "pop out". This "pop out" effect seems to be cognitively linked to preattentive processing [2][3][5][14][26].

Exploiting the preattentive processing aspect of visual perception with augmented reality (AR) techniques is not a new concept. Augmented Reality in the form of rectangular image overlays is a common feature in the view finder of the latest digital cameras to highlight regions of interest designated by the built in auto focus assistant. This is AR in practice for a real world application. We almost instinctively understand that this application of AR is helpful, and by extension we would expect that it also would be of use in a robot search task such as that proposed in this study. The HRI issue then becomes a question of how to employ the augmented reality, and how to determine regions of interest.

Object recognition algorithms can be used on a digital image to ascertain regions of interest. There are many "flavors" of digital filters that can be applied to detect edges, and presumably therefore objects. Construction of custom object recognition software was beyond the scope of this study, fortunately the ER1 robot came with object recognition packages. I contend that combining object recognition software with augmented reality techniques results in an enhancement to the visual interface presented to the human operator. The premise of this study is that this enhancement will serve as a perceptual aid in the performance of the search task. The perceptual aid provided by the augmentation is particularly evident when there is signal interference that degrades the image presented from the robot. It is in this scenario where the HO is presented with a degraded visual perception interface that we will gather empirical evidence towards the performance improvement offered by the augmentation of the image.

They key task in the experiment for this thesis is that of positive identification of an object of interest. The HO is searching for two specific objects within a confined arena of operation. Specific criteria are defined by which the HO can positively identify a target object. This positive identification is made difficult when there is interference in the signal providing the video interface to the remote operator. If we can give some intelligence to the systems onboard the robot, then identification assistance can be administered prior to transmission of the image data. A region of interest (ROI) can easily be identified with paired coordinates (upper left & lower right) which requires a minute amount of memory in comparison to the image, so more deliberate error checking can be utilized to ensure the successful communication of the ROI data.

For this study, the signal interference was simulated so that the amount of image degradation would remain constant from test to test across all participants. When there was a recognized target object the ROI information was delivered with every image transmission between the robot and the client side visual interface application.

In the experiment for this study participants were tasked with remote navigation of a robot in a search task. Presented only with an image provided by a camera mounted on the robot each participant must locate and identify two target objects within an enclosed area. The only control the participants had of the image they saw on their interface was through navigational input to the robot via a joystick. Each participant completed the search task 3 times under different simulated environmental conditions. The simulated environmental conditions were achieved utilizing different video modes for the visual image presentation on the operator interface.

There is a nominal video mode which contains a small amount of simulated static. The degraded video mode represents significant signal interference with considerable simulated static, ghost images, and blurring. Lastly, the degraded/augmented video mode is the same as the degraded mode with the addition of a highlight effect which keys off of object recognition software on the robot. The object recognition software runs against the clean image captured by the robot prior to the application of simulated signal interference which is presented to the operator.

The participants encountered the different video modes in random order. This was done in an effort to minimize the effect of practice on the comparative performances. The participants were asked to complete the search task as rapidly as possible, and were informed that time would be kept in order to log the duration of their search task. The expectation was that we would see noticeable improvement in task completion times for the degraded/augmented mode over that of the degraded mode. Anyone who is familiar with the concepts involved in this experiment would think the predicted result intuitively obvious, yet prior to this study there is little empirical evidence to support the assertion in a robot teleoperation paradigm.

2. Related Work

Because the search task for this experiment is inherently perceptual in nature it naturally follows that the area of cognitive psychology that deals with perception, specifically visual perception and how it relates to attention, would be of interest. Augmented reality is employed in this experiment to direct the attention of the operator to an ROI in the image. The use of AR in this manner is designed to capitalize on the cognitive processing benefits of preattentive perception. There has been much study of perception and attention in the field of cognitive psychology; I will just scratch the surface of this research to indicate why it is that AR is a good tool to use as an aid for the perceptual task in this experiment.

This research also builds upon several aspects of HCI and HRI that have been explored in previous studies; aspects such as object recognition, image augmentation, and control interface issues. The findings from these studies can be used to guide current and future work. I have endeavored to utilize the lessons learned in previous research to guide the focus of this study. The following sections discuss some of the previous studies that were particularly relevant to this ER1 experiment.

2.1 Perception and Attention

The subject of perception has sufficient scope to comprise its own subcategory within the broader discipline of cognitive psychology. The research on human visual perception alone is voluminous. Of specific relevance to this study is the matter of preattentive processing in visual perception. The human visual system seems to be able to detect some visual properties rapidly and automatically; this early occurring aspect of the perceptual process is called preattentive processing [15].

There is in cognitive psychology the notion of visual attention being focused like a "spotlight" on an ROI in an image [17][16]. Preattentive processing allows for recognition of target presence prior to focusing the "spotlight" of attention. This feature of the human visual system can be used to our advantage in directing where we want the "spotlight" to be focused when the software identifies the existence of a target in the image presented to the HO.

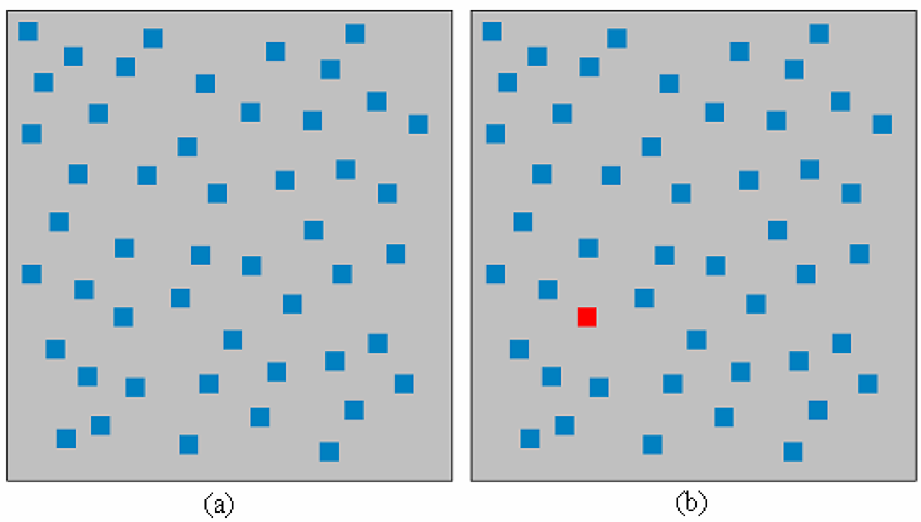

Figure 1. Preattentive Perception and The "Pop Out" Effect

Frame (a) has only distractors, but the target object in frame (b) "pops out" at first glance.

Some of the features that are considered preattentive are size, color or hue, and curvature [26]. A target object placed in a field of distractors that differs in one of these simple properties has a tendency to "pop out" of a display. An observer can readily tell if the target is present in a display with out having to specifically focus attention on the target. A common example of the “pop out” effect can be seen in figure 1, where the red target object is obviously present in frame (b) at first glance.

There has been much research into what properties are preattentive, and several prominent theories have been devised addressing preattentive concepts. Of those theories, perhaps the feature integration theory proposed by Treisman is most recognized [27]. The feature integration theory has served as the jumping off point for many other studies that focus on the role of color in preattentive perception; these are of interest in the discussion of how to use AR as a perceptual aid for this study.

Treisman has stated in the feature integration theory that in the human visual process there exist features that are “registered early, automatically, and in parallel across the visual field”. This sort of early and automatic registration would seem to enable the “pop out” effect discussed earlier. Treisman suggests a model of the low-level human vision system that is composed of feature maps. A number of simple features such as color, size, and orientation are each represented by a mapping in the Treisman model. When simple features in an image match up with the feature map in the low-level visual system they are automatically registered as being present. [27]

Treisman confirmed the “pop out” nature and textural segmentation associated with objects that differ in only a single primary feature such as color, however, it should be noted that when there is greater than a single difference between target and distractors the preattentive “pop out” effect is lessened. This was linked to the feature integration map by suggesting that “pop out” occurs when a target produces activity in a separate feature map that is unaffected by the distractors. [26]

It was later demonstrated in a study by D'Zmura, and confirmed by Bauer, that it was more difficult to detect the presence of a target among a field containing two types of distractors that differed from the target in a single feature but with two separate values[5][2]. One can see in Figure 2 that the orange square in frame (b) does not "pop out" quite to the extent that the red square does in Figure 1 (b). D'Zmura noted that a chromatically narrow band decreases the tendency to "pop out", but that the hard-wired visual mechanisms could still be driven in a "bottom-up" fashion by chromatic visual properties [5]. In other words differences in chromatic properties can be perceived based strictly on the physical stimuli without having to engage higher level cognitive processes.

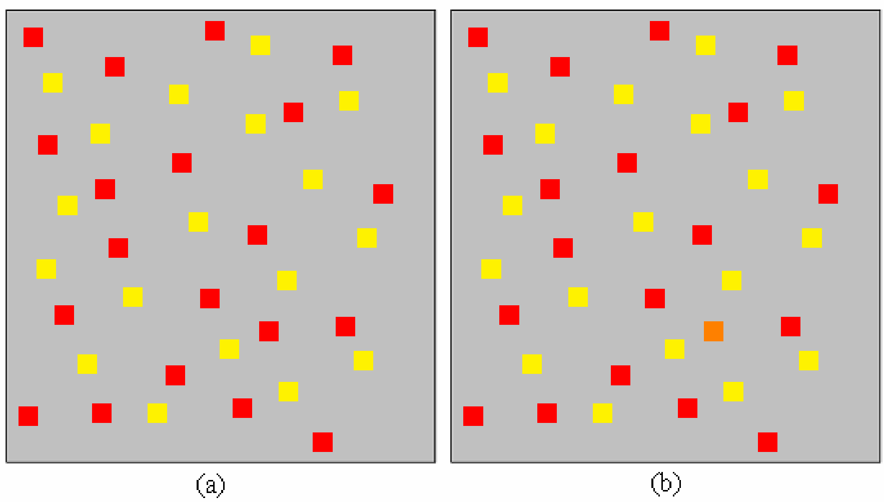

Figure 2. Decreased "Pop Out" With Multi-color Distractors

Frame (a) has red and yellow distractors. The orange target object in frame (b) exhibits less tendency to "pop out" than the previous example in Figure 1.

Based on results of a 1996 study, Bauer et al rejected models based solely on color differences [2]. Somewhat as a corollary to the color mapping suggested by Treisman, Bauer noticed that luminosity variance functioned similar to color differences for preattentive discriminability [2]. To that end, the augmented reality technique used in this study applies a color overlay along with a moderate luminosity increase to the highlighted ROI; if it doesn't "pop out" based on color, perhaps it will "pop out" based on luminosity.

Nagy had some discoveries similar to those of D'Zmura and Bauer with regard to color differentiation between target and distractor. As the target color becomes increasingly similar to that of the distractor, the search time required to locate the target increases [14].

Beck studied textural segmentation and agrees that simple properties like brightness, color, and size may be picked up preattentively [3]. An interesting observation from Beck is that attentional processing may only be necessary for preattentive features when a stimulus is degraded [3]. I might extend this notion to my study scenario to say that AR is only necessary when the stimulus is degraded; this is in order to direct the attention to the ROI for analysis.

The aforementioned studies all agree that an object of a color that contrasts from the background can be perceived preattentively [2][6][14][26]. Evidence also indicates that size and luminosity can also lend to the "pop out" effect [2][26]. Based on that information, it stands to reason that augmented reality could be used to apply an object of contrasting color, size, and luminosity to a display such that it is perceived preattentively and therefore draws the "spotlight" of attention to that ROI in the image. This concept served as the inspiration for this study.

2.2 Object Recognition

Object recognition is a key component of the search task. Cognitive psychologists describe recognition tasks as those requiring selection or identification of an item as being one that was learned previously [24]. Whether performed by the human operator or the robot software, it is the recognition of specific objects that determines the successful completion of the search task for the ER1 experiment.

In addition to the important role recognition plays in the cognitive processes necessary for search tasks, it is also necessary for basic navigation. Recognition of environmental attributes in an image can contribute to location awareness; it can also contribute to establishing navigational goals or triggering of specific behaviors. When working with autonomous robots, recognition of environmental queues is paramount.

Some similarities to the ER1 experiment can be found in the RoboCup legged league with the Sony AIBO robots. The AIBOs utilize a vision system that employs recognition techniques to determine field boundaries, targets, and goals [11]. What is of particular interest in relation to this study is how the vision system enabled the robot to identify objects of interest. Since the RoboCup takes place in a very controlled setting, the AIBOs can take all their queues for positional information and goal location based on colors in the images acquired from their vision system.

The vision system utilized by the AIBO to recognize an ROI in an image is driven off of a color lookup table. Due to the controlled setting in RoboCup soccer this process is functional for the AIBOs, but color recognition is highly susceptible to variation in lighting conditions. It is because lighting conditions contribute to perceived color variance that the AIBO vision system references a range of hues for color recognition. One range of hues indicates field boundaries; another range of hues indicate the field itself; while still another range of hues indicate goals [11].

The color lookup table has been accepted as a standard vision system for the AIBO platform, but there is always interest in improvement where competition is involved. Quinlan et al. investigated how to improve vision of the Sony AIBO robots in the soccer game task. The study identified improvements in performance that could be made with the use of an improved lookup table for color identification [18]. These results suggest the value of careful organization of a database of information that can drive image recognition in HRI tasks.

Another recent application of visual object recognition techniques is the AR_PDA project. The recognition process does not use markers, but only what can be seen in the acquired image. It is a two step process to complete the object registration. First they use image filters to perform edge detection and create a 2D map of the acquired digital image. For the second step "hypotheses about possible combinations of features are generated, verified, and the best matching hypothesis is chosen". The AR_PDA utilizes a server to perform the image processing. This results in a frame rate of 6-10 frames per second for a small PDA screen. [4] This is an attractive methodology, but to maintain the notion of placing intelligence on the robot necessitates the use of a laptop on the ER1 which processes the images. For the resolution desired in this study the heavy processing necessitated by edge detection was not feasible on the laptop which also serves as the control system.

2.3 The HRI Interface

In the world of robot-assisted search Robin Murphy is on the front line of rescue robotics; she worked with the Center for Robot-Assisted Search and Rescue (CRASAR) on the World Trade Center disaster response. Real world experiences such as this give insight into HRI issues and to how things might be improved [5]. Murphy notes that several opportunities exist for artificial intelligence and distributed network systems to improve robot-assisted search and rescue. Particular suggestions hinge on wireless networks and augmentation of digital images from robots [13]. We use both these technologies in an effort to demonstrate that search tasks can be facilitated.

The insights of Robin Murphy reinforced the notion that image augmentation could be a useful tool to aid in visual perception activities of a search task. One of the findings noted from the WTC activities by CRASAR was the lack of image processing capabilities on the robots; which added to the cognitive load of the HO [5]. This is precisely the type of issue this study would like to address. By adding some intelligence on the robot that has “first person” information at the remote location, the HO can be relieved of some effort interpreting the perceptual information.

One of the HRI issues raised by Murphy was that of improving the human to robot ratio in the search task. For the CRASAR search activities the operator is accompanied by a second person (the problem holder) who aids in viewing the video from the robot and is responsible for reporting search results back to the search manager [13]. The inherent communication issues that arise in real-world search situations make it difficult to remove the problem holder from the “hot zone” where the search is being performed, but advances in artificial intelligence and sensor performance could shift some responsibilities off the problem holder to possibly enable them to be located more remotely [13]. This study is a step toward leveraging intelligence on the robot to reduce requirements on the human agents.

Although the robots used by Murphy for CRASAR activities are generally tethered robots, she sees value in the wireless networks for image sharing in non-real time analysis [13]. The ER1 came equipped with wireless communication functionality to enable teleoperation; it made sense to use that which was already available as a control conduit for this research. The affordability and ease of use provided by the wireless teleoperation interface on the ER1 was of singular importance in enabling this research to proceed with minimal funding requirements.

A critical and obvious element of the HRI interface is the display which presents the video from the robot to the HO. In an unpublished experiment, fellow graduate student Christiaan Janssen and I performed a user interface analysis of the ER1 robot. Much of the feedback from participants in that ER1 study indicated the display area was too small for the teleoperation task. This guided the efforts to obtain a larger custom display that was operational within the bandwidth restrictions of our test parameters. The details of the custom ER1 display properties will be covered in a later section.

2.4 Image Augmentation

The ARGOS (Augmented Reality through Graphic Overlays and Stereo-video) system developed by Milgram et al. is designed to improve the HO's comprehension of a remote environment in a teleoperation task. The goal of ARGOS is to provide a set of tools to facilitate the interactive modeling of a remote space. The cornerstone of the ARGOS tool suite is a *virtual tape measure*, which allows the user to use a virtual pointer to select a point on a display and see the 3D coordinates of the point in real space in real world units. The virtual tape measure is implemented with a stereoscopic vision system that uses graphic overlays to augment the video image presented to the remote HO. This graphic overlay method is the style of augmentation selected for my study. [12]

A "tag" concept has been used by multiple researchers to guide the augmenting of images, where a tag might be a black and white bit code [19] or a colored "landmark"

[22]. The general idea is that automated recognition of specific patterns in real world images provides an opportunity to augment that image for human interpretation. The system described by Rekimoto et al. [19] employs a variety 2D tags that feature specific black and white block designs. A digital image containing a tag is binarized to facilitate tag identification. The corners of the tag are located, and then the bit pattern of the design is decoded. Once the bit pattern is decoded, the image can be augmented with any information that is associated with the coding of the tag.

The Mars rover project is probably the most recognizable robot-assisted search exercise ever undertaken. Because of the 20 minute lag time between when a command is sent and when response is received, the Mars rovers cannot operate in an interactive teleoperation mode [7]. So AR is not used as a primary control tool, however, AR is used in planning the actions of the rovers.

The Rover Sequencing and Visualization Program (RSVP) utilizes AR to help plan the navigation of the rovers. A portion of the RSVP system called Hyperdrive is a visualization tool used to "test drive" the commands prior to sending them up to the rovers. 3D terrain maps are generated from stereo images received from the rover cameras. Open GL is used to create the 3D landscape from the terrain maps and 2D images; this Mars-scape is then augmented with a CAD version of the rover. [7]

The RSVP can be used to visualize the rover on location at a task site prior to navigation, and thereby avoid possible pitfalls prior to communication of the planned activity. The RSVP software can also be used to visualize a planned route as in figure 3. Though the rovers don't conform to the teleoperation paradigm in the conventional sense, this is a clear example of AR as a perceptual aid in robot teleoperation. [7]

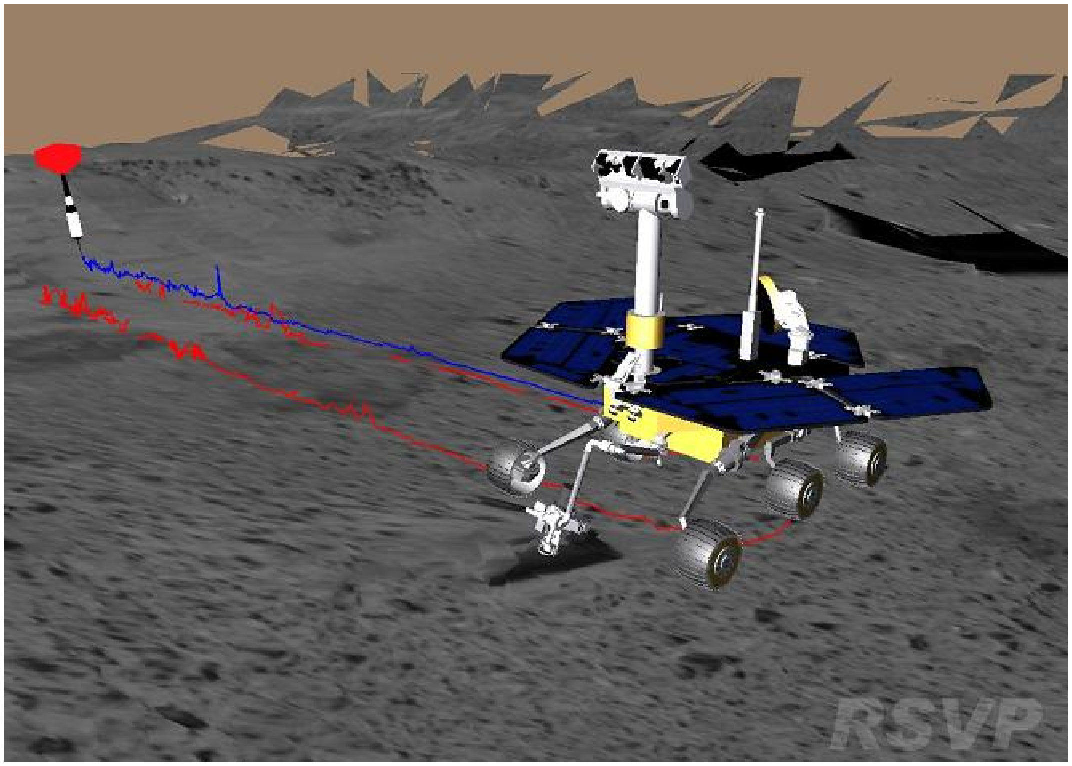

Figure 3. RSVP augmented reality view of *Spirit* and navigation plan (Courtesy NASA/JPL – Caltech)

3. The System

The system used for the teleoperation activity consists of several hardware and software components. Primary among these is the Evolution Robotics (ER) ER1 robot. The ER1 Robot Control Center (ERCC) software that accompanies the ER1 provides key capabilities for our work. The ERCC runs in the background during testing to enable navigational control input from the operator. Some of the libraries from the ERCC are used in a custom client/server software package called the Augmented Reality Robot GUI (ARRG) which is used to perform the image recognition and image manipulation. An 802.11G wireless connection capability is supported via the ERCC; this enables the operator to be in a location separate from the robot.

Figure 4. The ER1 Robot

The ER1 has battery powered servo motors to control locomotion via two drive wheels, but it has no CPU. The CPU for the ER1 is provided by a laptop that rides in a cradle at the front of the robot. The operator can either plug a joystick into the on-board laptop and drive it in a first person scenario, or run the ERCC on another laptop/desktop computer and establish a remote connection for control input over a wireless network. The research for this thesis relied on the latter scenario.

The custom software package developed for this study utilizes an off the shelf camera (Logitech QuickCam) for image capture. The ER libraries were adapted to suit the image acquisition and processing needs of the search task. Once an image buffer is acquired it is passed through the ER image recognition function. The data returned from the image recognition function is held while custom code manipulates the image buffer to selectively degrade the image (depending on the video mode selected for a given test).

Only after the degradation has taken place is the augmentation applied, based on the data returned from the recognition function.

In this section a detailed description of the system architecture and the custom software application is presented. The server side software was developed to run on the laptop which serves as the control unit for the robot. The software on the server side already has access to the raw image data coming in from the camera, and so all the image processing is done here including object recognition. The client side software simply manages the images presented to the HO. The ERCC software is running on both client and server laptops, but it is only utilized as the communication channel for the control input.

Following the architecture discussion, the details of the aspects relating specifically to the image processing are presented. The object recognition process will be examined followed by details of the image resolution, and finally a discussion of the image degradation process. The balance of this section will consist of the details about the setup of the experiment.

Physical characteristics of the test bed will be discussed. First we'll cover the lighting used for the trials; followed by a brief overview of the non-target objects in the arena, the distractor and clutter. The course layout will then be presented.

Lastly, some of the more important variables for this study will be examined. The various video modes used during each test will be detailed, then the exercise attributes encountered by each participant are covered.

The construction of a realistic experiment for a search and rescue type scenario can involve an extraordinary amount of resources that are beyond the reach of this graduate student. So it was decided to construct an experimental arena to test some of the technical aspects of a search and rescue task. A number of technical factors were considered: how to perform image recognition, the plausible effects of signal degradation on captured images, task lighting, resolution and frame rate for image presentation, and the presence of clutter and distractor objects in the test environment.

3.1 The ARRG Architecture

A client/server architecture lends itself perfectly to the robot teleoperation task. The ERCC software that was used to manage navigational control communications is a client/server application, but discussion of that off the shelf software is outside the scope of this study. For the image data however, the custom ARRG software was written to handle the specific processing required to accomplish the desired image variations. This custom software also functions in the client/server paradigm.

Since the robot is providing image data from the remote environment to be viewed by the HO, it only makes sense that the robot is designated as the server side of the equation. To maintain the premise that intelligence can be placed on the robot to aid in search tasks, all the recognition and augmentation operations were performed on the server side prior to image transmission. Additionally, the simulated signal interference that is present in two of the three video modes, to be discussed later, is also performed on the server side.

One issue with a client/server application functioning over a wireless network is that of latency; especially when primary information being transferred is image data. The initial option considered for image acquisition was to allow the ERCC to handle image data as well as control data. This would essentially have made the server side image processing software a slave to the ERCC processing cycle. I did not want to exacerbate the perceived lag in the interface by possibly adding buffer latency to the system. So I designed a software package that utilized the ER imaging libraries to perform image acquisition only when the previous image had been sent to the client. In this way, the image that is processed on the server is the most recent image available from the camera rather than an image from a stored buffer in another application.

Even after these efforts were made to obtain current image data, there was still a perceptible lag in the system. This partially stemmed from the fact that control input and image capture were processed asynchronously. The more likely explanation is that the object recognition processing consumed the lion's share of the server processing cycle. The object recognition process alone averaged greater than 600 milliseconds to complete; this is more than half of our desired 1 second cycle time for image refresh. Many control inputs could be detected, received and processed in the span of 600 milliseconds. Consequently, there was noticeable lag between control input and the perceived results of that action on the display. It should be noted that there also exists display lag when running the ERCC application for imaging and control; it seems to be the nature of the wireless teleoperation for this robot system. Perhaps the cycle time for object recognition could have been reduced with a smaller database of object images, but experimentation with smaller databases resulted in less consistent object recognition than was desired for this study.

Another issue of importance for this study was the viewable area of the images presented to the operator. Based on results of an earlier user interface study of the ER1, it was highly desirable to increase the dimensions of the image displayed to the operator.

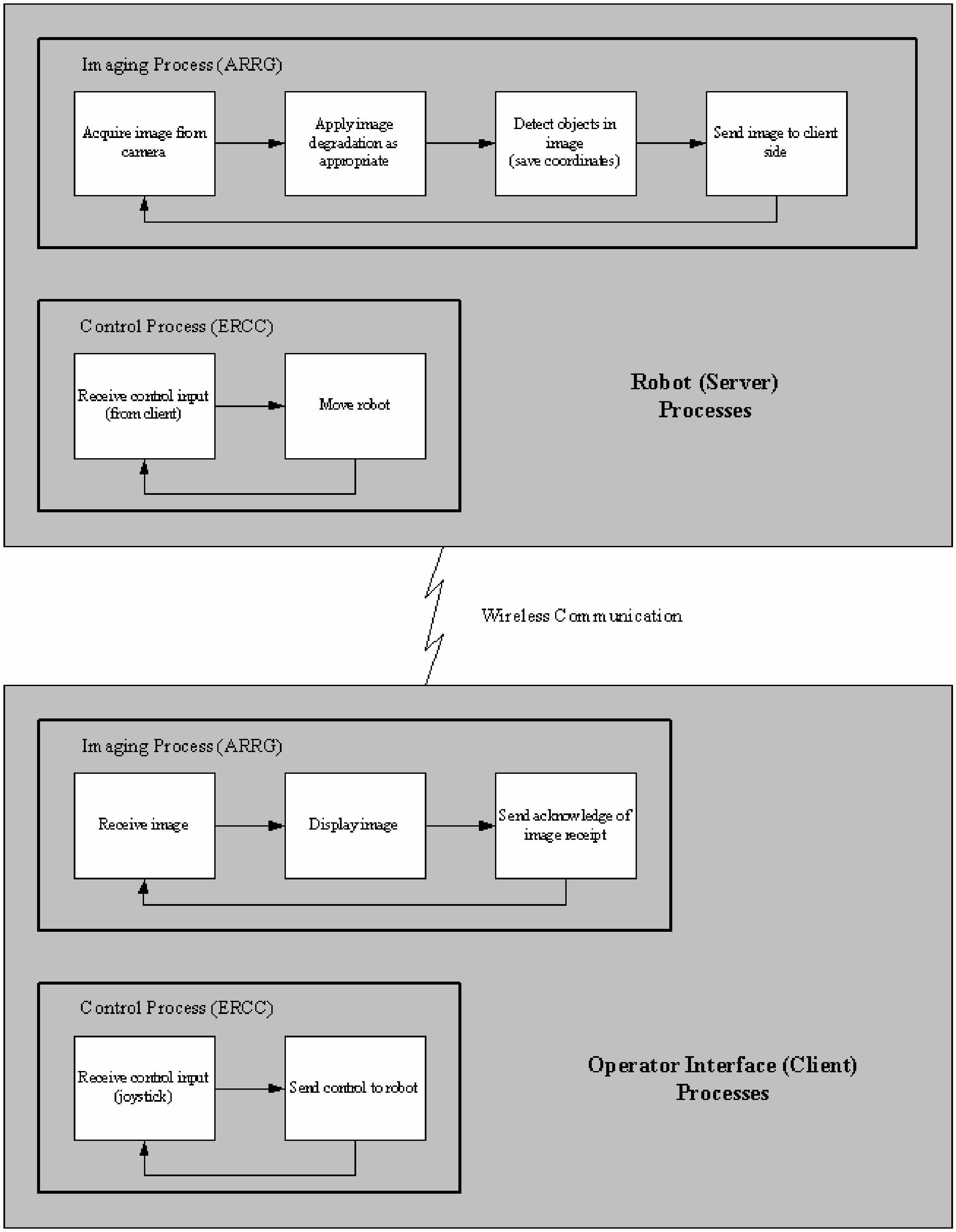

Figure 5. ARRG Process Flow Diagram

The existing display in the ERCC was deemed too small to adequately show environmental features for teleoperation activities. This was supported by both the client and server side of the ARRG application. The server configured the camera on the robot to acquire images at higher resolution, and the client side presented a larger window in which to view the images from the robot. The resolution selected hinged on the cycle time of the server image processing, this subject is addressed further in a later section.

3.2 Object Recognition

The object recognition system provided by Evolution supplies the calling process with an area of interest within a given image. The Evolution object recognition is based on a set of images in an object database. Areas of interest are noted on each image in the database (such as regions of high contrast), and these areas are compared to the current image in the buffer. The recognition system provides an ordinal "match" value indicating the likelihood that the image in the buffer is a match for the image in the database. If that "match" value is above a pre-determined threshold, we consider the image in the buffer to be recognized.

Once an object is recognized, we assess the location in the current image buffer and apply highlighting accordingly.

Figure 6 shows the two target objects for the experiment, the Hofbrau and the Rothenburger. For each of these objects, ten images are stored in the database. Photographs were taken of each object at two different distances from each of 5 different positions. The five positions consisted of directly in front, 45 degrees right, 90 degrees right, 45 degrees left, and 90 degrees left. The distances were chosen to obtain images where the objects appeared approximately the same size (the objects would fill about the same amount of the image frame). This was an effort to create a level playing field from the perspective of the image recognition system, such that each object is of a comparable size in the database for all of their respective images. For the Hofbrau (the smaller object), the two distances used for the photographs were 12" and 34". For the Rothenburger, the distances used were 20" and 44". The close-up series of photos used to populate the recognition database for both nutcrackers is displayed in figure 7.

Figure 6. Target objects

Once the database of images was created, various tests were performed to assess the match values returned from the recognition system. This was simply to serve as calibration of the system prior to testing. A threshold level for the match value was selected which eliminated all false positives.

Figure 7. Close-up series

One of the statistics returned from the recognition system is the coordinates of the approximate center of an object recognized in the current buffer. This coordinate is the sole piece of information required to augment the image. This is important because this piece of information is tiny by comparison to the image data. In a real world scenario greater care and error checking can be employed to ensure that this data makes it from the robot to the user console. Once this coordinate data arrives, the highlighting can be applied to the resultant image regardless of any degradation that may occur due to signal loss or interference.

3.3 Image Resolution

The resolution of the image presented to the operator in these exercises was determined by the refresh rate (frame rate) required to adequately navigate the robot during the search task. For adequate navigation anything less than one frame per second was problematic. Lag time can result in excessive control input and inadvertent collisions. Greater lag time also tends to contribute to disorientation and confusion as to what control input has been received.

In order to achieve the one image per second frame rate, it was necessary to select an image size so that I was able to perform all necessary processing within the one second window. The image dimensions afforded by this time constraint were 320 x 240 pixels.

The image was presented to the operator centered in a blank window with black background that obscured the desktop. This allowed the HO to position the display window at a comfortable viewing location on the screen without concern of revealing distracting background processes such as the ERCC running on the desktop.

3.4 Image Degradation

Based on background knowledge, two primary perceptual effects of signal interference were identified: ghost images and static. A third contributor to perceived image degradation, though not necessarily an effect of signal interference, is blurring. This last item is often caused by environmental factors such as smoke/fog or precipitation/condensation on the camera.

First, ghost images were added, one offset to each side of the original image. The left and right ghost images were overlaid on top of the original image at 40% and 35% of total luminosity respectively. Once the ghost images were applied to the base image frame, simulated static was added. 55% of the pixels in the image were perturbed. Each RGB value for a selected pixel was replaced by a randomly generated chroma value such that the selected pixel would show as a shade of gray.

After the ghosting and static was applied to the image, a blur effect was then applied to impart an impression of clouded vision from the robot’s perspective. This also served to soften the hard edges of the simulated static applied in the second step of the image degradation process. An iterative box blur algorithm was implemented to achieve a simulated Gaussian blur effect as discussed by Jarosz [8]. This gives the impression of blurring the entire image using a Gaussian filter but only requires a fraction of the processing time. By utilizing this method the desired blur effect was achieved and the one frame per second update rate could still be maintained for the user trials.

The speed up suggested by Jarosz was to use a vertical motion blur followed by a horizontal motion blur to achieve the same effect as a box blur. This type of algorithm improves the processing time of the blur function from $O(m^2*n^2)$ to $O(m^2*2n)$. Since the box blur uses a flat weighted kernel the motion blur is a reasonable substitution. The box blur function implemented in this study employs four iterations of the improved box blur algorithm, the end results of which are the perceptual equivalent of having applied a piecewise cubic filter. [8]

3.5 Lighting

The course was lit at the center by an overhead light fixture that cast diffuse light over the entire arena. This lighting was augmented by 40 watt spot lights located opposite each end of the course so as to optimally light the area where our target objects were located. The lights used during the test were the exact same fixtures used when acquiring the images used to populate the database for the object recognition software.

This lighting scheme proved totally adequate for consistent recognition of the Rothenburger. The Hofbrau often proved to have spotty recognition; this is partly attributable to the similarity in color between the Hofbrau nutcracker and the wood used to construct the course. Nevertheless, the recognition software was often able to identify the Hofbrau before the HO.

Figure 8. The Distractor

3.6 Distractor and Clutter

It is important in a search task that there is more in the arena of activity than just the target objects. If this were not so, then anything perceived in the view frame from the robot's camera would be assessed as successful target acquisition. For this experiment, there were an equal number of target and non-target items in the arena.

Figure 9. Clutter Obscuring Augmented Target

There was one nutcracker that was not a target object, shown in Figure 8. This nutcracker (Marvin) is designated the distractor; an object that is similar to the targets and could possibly be confused with a target object.

Some scrap plywood that remained after constructing the course was fastened together forming a triangular shape standing approximately 3 feet in height. The plywood was painted a mottled black and tan color so as to be clearly distinguishable from the wall panels. This plywood structure is designated as clutter. It is not likely to be confused with target objects, but interferes with the field of view by obscuring other objects in the arena as seen in Figure 9.

3.7 Course Layout

The course used for the search task was intentionally constructed in a symmetrical fashion such that it would appear the same regardless of which end was used as the starting point. Each 4' x 4' lobe of the course forms a 240 degree angle at the inside corner where it meets the adjacent lobes. Each participant piloted the robot once in each of the three course configurations. The course is shown in diagrammatic form in Figure 10.

3.8 Video Modes

Three different video modes were used during the exercise. Participants had to perform the search task once under each of the three modes.

Nominal video mode consisted of only the image as captured by the on-board camera with 1% of the pixels perturbed by the static function. No ghosting or blurring was done to the nominal image. All the degradation functions were invoked within the server software in order to maintain a consistent frame rate across all video modes, but for the nominal case the ghost and blur effects were not applied to the final image.

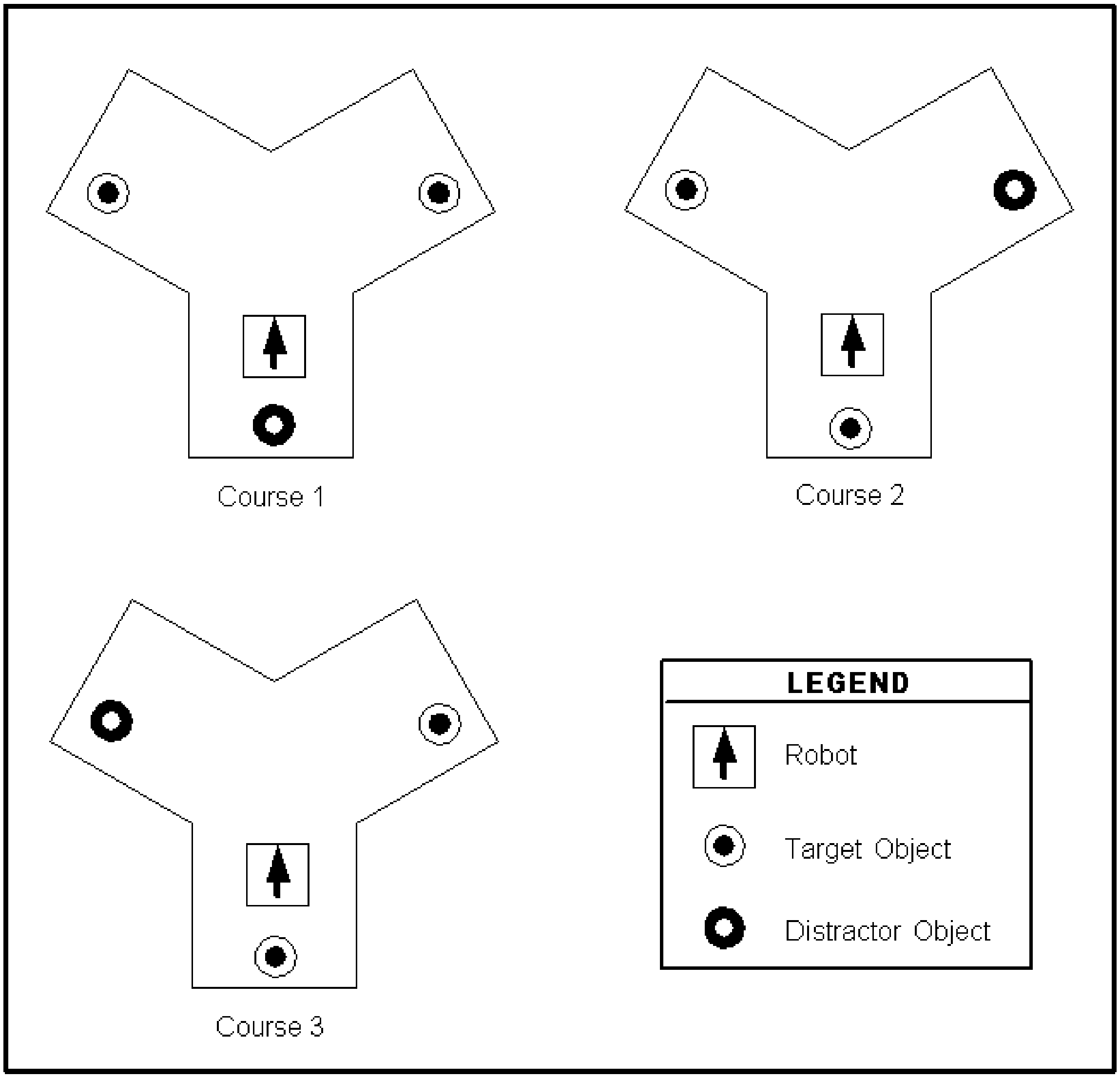

Figure 10. The Course Layout

For degraded video mode the image was modified by applying static perturbation to 55% of the pixels, applying ghost images to the left and right at fractional luminosity values, and applying a simulated Gaussian filter to blur the entire image.

Degraded/augmented video mode consisted of the degraded image discussed above with the addition of a yellow rectangular highlighted region where there was a recognized object in an image.

The three different modes are shown in Figure 11. These images were captured from the client application, and represent just what the human operator would see during a test.

Nominal Mode | Degraded Mode | Degraded/Augmented Mode

Figure 11. Video Modes

3.9 Combinations of Exercise Attributes

It can be seen for the course layout in Figure 10 that the beginning viewpoint appears the same, from the robot's perspective, on each course configuration. To minimize the effect of practice in locating the target objects, each participant drove once on each of the three course configurations. This way the target objects were in a different position relative to the robot start location on each trial.

4. Robot Teleoperation

The experiment conducted for this research involved the teleoperation of an Evolution Robotics ER-1 robot in a simulated search and rescue mission. A symmetrical three pronged course was constructed in which the robot would be navigated during the search task. The symmetrical course was designed so that the view from each "end point" of the course is indistinguishable from the other two "end points". The participants were allowed to view the robot and the empty course prior to the teleoperation exercise simply to get an idea of relative scale of the robot and the course.

Twenty three participants volunteered for the experiment. The participants ranged in age from 8 to 72 years with an average age of 34. There were 11 male and 12 female participants. The order in which the participants encountered the three video modes was randomized. A full set of six permutations (3_P_3 = 3!/0! = 6) of the video modes was used. Over all 23 participants, we cycled completely through the set of six permutations nearly four times.

We wanted the emphasis of this task to be on the search and identification of the objects of interest. Toward that end measures were taken to familiarize participants with their surroundings prior to the teleoperation exercise. Each participant was allowed to view the course in an empty configuration as well as see the robot, nutcrackers, and clutter collected at one open end of the course. The function of this was two-fold: it allowed the participant to understand where they might be driving so as not to be too concerned with trying to discover new passageways rather than searching out their target objects; it also provided the participants with a self-awareness of size when they assumed robot control, so they would be more aware of when they might scrape walls with their robot wheels.

Each participant was also given photos of the two subjects for which they were to search, so that they could identify them when encountered in the course. This made the search task less about cognitive recall, and more about recognition of what is in the image displayed.

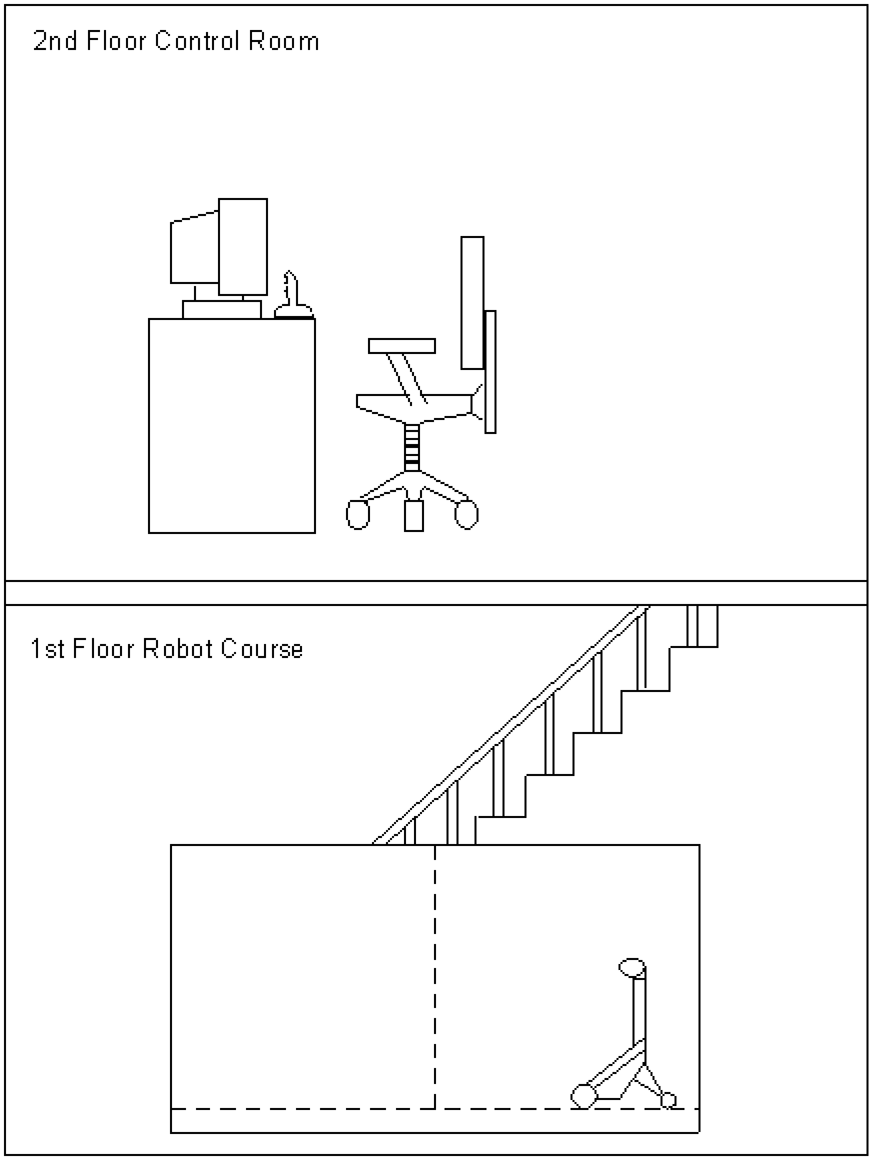

Figure 12. Geographic Separation

Since the experiment is a teleoperation task, it was necessary to separate the operator from the robot geographically. For the trials, the operator was situated in a closed room one floor above where the robot course resided (Figure 12). Though not a great distance, it was sufficient to prevent any direct perception of robot activities by the HO.

4.1 The Search Task

The task given to each participant was to search and locate each of two objects of interest within the robot course. The participants were accompanied by a test moderator who kept the time for the performance of the test. The moderator role was performed by the developer of the system, who could therefore be classified as a subject matter expert and address any atypical performance issues with the system.

This arrangement parallels the real-world scenario discussed by Murphy where the operator is accompanied by the problem holder who monitors and tracks the search activity to then report back to the search coordinator [13]. In this study the moderator, similar to the real-world problem holder, tracks the search activity. The function of reporting search results to the coordinator is replaced by the moderator acknowledging when the HO has successfully identified a target object.

For the participants to successfully locate a target object, we first had to establish ground rules as to what action was sufficient to claim that the target object had been located. Since our goal was to assess the relative performance based on perception, we defined a simple criterion for positive identification of a target object. Each participant was advised by written rules that positive identification would be defined as one of the following:

- Object recognition by the participant (spoken aloud).
- Object recognition by the software.
- Proximity to the object such that it fills greater than 50% of the image frame.

For each of the above positive identifications, the participant was notified by the test moderator with confirmation of a successful identification and was informed to proceed with the next search task.

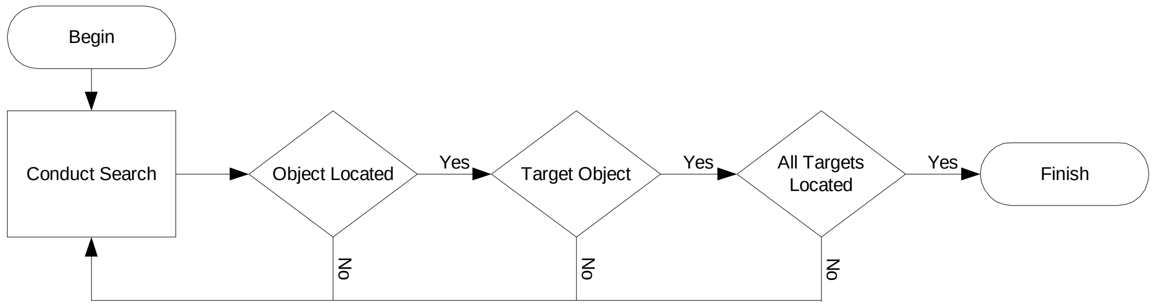

Figure 13. Work Flow

The search task performed by each participant followed the work flow model in Figure 13. Generally speaking the operator searches for a target object of which there are two located somewhere in the course. Once one is located, the search continues until the second of the two target objects is located. The search task is designated as complete when the second of the two target objects is identified.

A single trial by a participant is considered to be the completion of the search and identification task, positively identifying the two target objects which they were shown prior to beginning the teleoperation activity. The video mode and course layout is reset (out of view of the participant) for each subsequent trial. Time is kept separately for each trial. The time is kept by the moderator who is sitting next to the participant during the teleoperation activity. The stopwatch is started with the initiation of navigational control by the participant via the joystick; the stopwatch is stopped upon successful positive identification of the second target object.

The timing data gathered for each participant for their three trials (one for each video mode) was then analyzed to assess the performance under each video mode across all participants.

There were many approaches taken by the various participants. Some approached the navigation task with a deliberate "move-and-wait" technique to carefully manage the time lag associated with the one image per second frame rate. Others made a habit of turning the robot to look back before venturing into the course, while still others would drive backwards in an effort to "broaden" their view.

4.2 The Results

The initial hope was that we would see performance with the degraded/augmented video mode that approached that of the nominal video mode. In hindsight, that would have been somewhat surprising, as the navigation of the robot during the search task is much more than just the recognition of the target objects. The degraded video image made it initially difficult to discern walls from floors, much less target from distractor. Once participants began moving the robot, though, the delta between frames eased the task of making out the course boundaries.

One improvement made to the course during prototype testing was the addition of a strip of blue tape running along the walls 6 inches above floor level. This helped differentiate wall panels from floor panels, and made it easier to perceive the perspective aspects of walls receding in the distance.

Due to the limited size of the course available for these trials, the performance times are in a range that is an order of magnitude less than that of an actual real world search exercise [5], but the task is more relevant to the real world task than would be a complete software simulation with no physical component.

Each participant performed three trials, one under each video mode. The fastest aggregate performance time across all three video modes was just under 4 minutes and the slowest was near twenty minutes. The teleoperation abilities of the participants varied widely. As can be seen in Table 1, the variance between the means is relatively high, yet the relationship between the degraded performance time and the degraded/augmented performance time remains largely consistent from participant to participant. It is this relationship between the degraded and the degraded/augmented video modes that is of greatest interest.

Table 1. Mean trial duration in seconds

	Mean	**Standard deviation**	**Standard error**
Nominal	139.260	99.386	20.723
Degraded	210.608	103.966	21.678
Degraded/Augmented	167.565	82.303	17.161

The mean performance under the degraded video mode was considerably worse than under the other two video modes. As might be expected from the goal of the experiment, it was necessary to establish that performance under the degraded video mode was significantly different from that of the other video modes. This was in fact the case. A one tailed matched-pairs t-test between performance times for the nominal and degraded video modes results in $t(23) = 3.914$, with a p value of 0.0004. The same test between performance times under the degraded and degraded/augmented video modes results in $t(23) = 2.902$, with a p value of 0.0041. There was not a significant difference between the performance times under the nominal and degraded/augmented video modes, resulting in t(23) = 1.484, with a p value of 0.076.

So if we consider the baseline to be established by the performance under the nominal video mode, then the performance under the degraded video mode is greatly reduced. The mean time to complete a single trial with degraded video is nearly 50% greater than under the nominal video mode. Under research conditions, the augmentation of the degraded video improves the performance to a level that is statistically indistinguishable from the baseline performance.

Table 2. Mean trial duration in seconds, by gender

	MALE		
	Mean	**Standard deviation**	**Standard error**
Nominal	147.545	109.008	32.867
Degraded	193.363	105.052	31.674
Degraded/Augmented	171.818	70.643	21.299
	FEMALE		
	Mean	**Standard deviation**	**Standard error**
Nominal	131.666	93.920	27.112
Degraded	226.416	104.954	30.297
Degraded/Augmented	163.666	94.743	27.349

The comparison of the performance characteristics breakdown by gender showed some interesting differences. For the female participants, the difference in the average performance times between the degraded and the degraded/augmented modes was greater than those of the male participants. The augmented performance times for the female participants were approximately 27% improved over the degraded times, while the male participants only showed an 11% improvement for the same conditions. It is unclear what this difference might indicate; whether the degradation affects one gender more prominently than the other, or if the augmentation helps one gender more than the other. Of course there could also be a relationship to differing navigational habits between the genders, but that is a sticky wicket that I will avoid at this time. Perhaps this is a topic that warrants further research to see if there is some statistical significance to this finding.

The differences between the genders not withstanding, the average performance times in both cases showed the improvement trend that echoes our earlier findings in the pair wise analysis. Upon closer inspection, the male performance improvement between the degraded and degraded/augmented video modes does not achieve the level of statistical significance seen by the female performance improvement. A one tailed matched-pairs t-test between male performance times for the degraded and degraded/augmented video modes results in $t(11) = 1.035$, with a p value of 0.163. The same matched-pairs t-test for the female performance times result in $t(12) = 3.105$, with a p value of 0.005

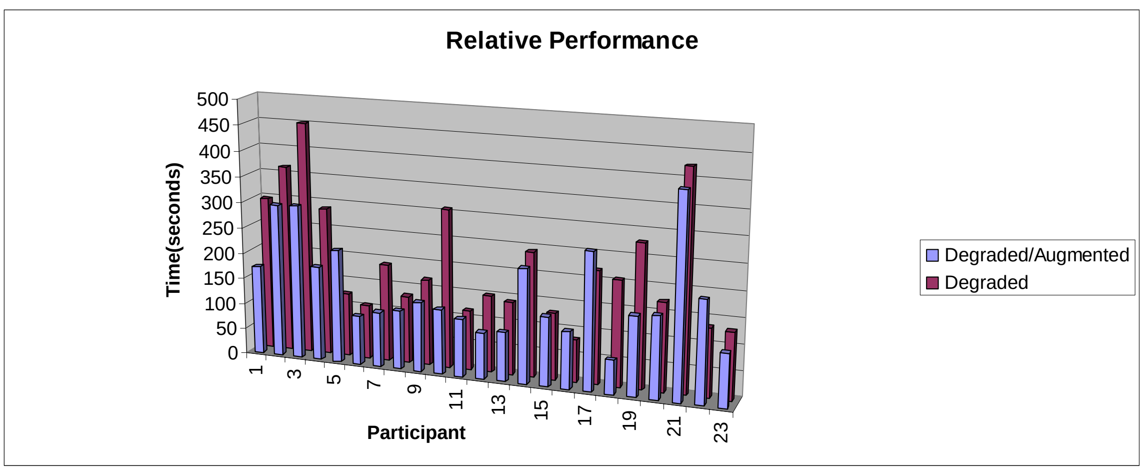

Figure 14. Paired Performance Times

The paired performance times for the degraded and degraded/augmented modes can be seen in figure 14; with the occasional participant performing better under the degraded mode. The prevailing trend demonstrates that the augmentation improves performance.

Another interesting perspective on these statistics is obtained by looking at the breakdown of the performance across the course configuration. If we look at the performance for the differing video modes across the course variations, the relative performance characteristics for those video modes still hold. Some courses seemed to be more difficult than others for any given video mode. Under the nominal mode, course #2 was most difficult with average completion time 11% and 15% greater than courses #1 & #3 respectively. For the degraded mode, course # 1 was most difficult with an average completion time 7% and 26% greater than course #3 and #2 respectively. With the degraded/augmented mode the difficulty level mirrored that of the degraded mode where course #1 was most difficult having average completion time 15% and 16% greater than courses #3 and #2 respectively.

With the small sample size, the averages for the course-by-course breakdown can be unduly influenced by the performance times of outliers on either the high or low end of the spectrum; excluding only one outlier, the trend of this compartmentalized performance data suggests the same conclusion as the matched pairs analysis.

All the statistical analysis of the performance data suggests the same conclusion. Augmented reality, when used as a perceptual aid, can significantly improve performance in a robot-assisted search task.

5. Contributions

Beyond the contribution of the empirical results discussed previously several other insights were gained during the implementation of this study. We have shown that

AR can improve performance in a search task; this result lays a foundation for further study into the benefits of AR as a perceptual aid for teleoperation. Some of the other findings serve as lessons learned or even inspiration for future related studies. The difference in relative performance between the male and female participants is an ancillary finding that was not at all anticipated, but something I think warrants further study. The various system constraints faced in this study such as bandwidth limitations, image processing cycle time, and object recognition limitations will also be faced in future studies. System lag time in a teleoperation scenario can spell disaster; all efforts to mitigate lag time issues are warranted. A combination of approaches worked best for this study. Little could be done about the bandwidth limitations, but measures were taken to improve image processing time, object recognition time, and object recognition success rate.

The measures taken improved the frame rate of the display interface nearly three fold. The frame rate improvement was critical in achieving a level of usability that avoided frequent disorientation or confusion by the human operator observed in early prototypes of the system. Even with the improvement some participants confronted periods of confusion about location and orientation, but these were perhaps less frequent than might have been anticipated with degraded video mode. The addition of the blue tape near the base of the walls on the course proved invaluable for many participants as a means of determining robot location. System prototype testing can reveal simple solutions, like a piece of tape to improve contrast, that pay large dividends in more consistent results during experiment execution.

6. Conclusion

Augmented reality concepts have been around for decades, but there has been limited research into human factors issues [23]. Even though augmented reality has been a popular topic in HCI and many technological improvements have been made over the years, there are few empirical studies about its effectiveness [25]. Where robots are concerned, much of the research is understandably focused on hardware and autonomous control mechanisms. Perhaps as an artifact of the influence of science fiction entertainment in our society, there seems to be much effort towards robots that reflect characteristics of living entities whether they be pets or humans. When robots are utilized as an assistant in the real world the characteristics are much more attuned to functionality than appearance and the opportunity still exists to improve the interface to the human operator with augmented reality. This is supported by the findings of Casper and Murphy at the World Trade Center [5].

An experienced HRI researcher might suggest the results obtained in this experiment were intuitively obvious. That same HRI researcher would also have to concede that all intuitions gained from HCI research do not necessarily transfer to the HRI domain. It is worthwhile to test even the most basic concepts if they have not yet been analyzed in the current domain. During the research process ancillary findings present themselves and give us better insight into the subject of study. Beyond the experimental results of this study, insights have been gained about the constraints of the robot teleoperation paradigm and what methods work in practice.

Part of the contribution to the fields of HCI and HRI from this study is the empirical evidence acquired toward substantiating the hypothesis that AR used as a perceptual aid can improve performance in a teleoperation search task. Additionally the experiment reinforced the importance of improved object recognition algorithms and the elimination of system lag. The value of prototype testing cannot be understated and was pivotal in discerning ways to improve usability prior to bringing in participants for the experiment.

What was believed to be true about AR at the outset of this study turns out to have valid empirical foundation. To that end, I have demonstrated the clear advantage in the application of augmented reality by the performance improvements seen with the degraded image in the search task. Whether it be used directly to aid the search task as in this study, or used indirectly to aid the navigation with graphic representation of sensor data, augmented reality can certainly aid in robot teleoperation tasks.

7. Future Work

For the future we need to determine how augmented reality can best be applied to HRI tasks. Guidance and navigation cues seem a logical application. With the proliferation of handheld GPS units and wireless communication technology positional cues can quite readily be used to augment a display [9][10][20][21]. To complement the visual image from a robot mounted camera, laser range finding can be employed to ascertain distances to objects in a display [1]. There will always be environmental considerations which may limit or eliminate certain solutions, such as smoke or precipitation limiting a visual interface; other sensor configurations can fill the gap. Augmented reality does not have to rely on visual input to drive the visual output. From radio to laser or thermal imaging, a myriad of sensor data can be utilized to augment a visual display. The augmented reality need not be visual either; a user interface could be augmented with auditory cues or even tactile feedback through control interfaces.

The future really lies in finding optimal ways to impart perceived intelligence to the robot. It seems clear that AR is one method to allow a robot's "knowledge" of its environment take some load off of the human operator. As improvements are made in the realms of sensors, detection algorithms, and processing speeds, new directions will become evident as to how best we enable the robot to become a more valuable assistant.

References

1. Arman, F., “Model-Based Object Recognition in Dense-Range Images”, *ACM Computing Surveys*, vol. 25, no. 1, pp. 5-43, Mar. 1993.
2. Bauer, B., Jolicoeur, P., and Cowan, W. B., “Visual search for colour targets that are or are not linearly separable from distractors,” *Vision Research*, vol. 36, no. 10, pp. 1439-1466, May 1996.
3. Beck, J., *Organization and Representation in Perception*, Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1982, pp. 285-318.
4. Beier, D., Billert, R., Bruderlin, B., Stichling, D., and Kleinjohann, B., “Marker-less Vision Based Tracking for Mobile Augmented Reality,” *Proceedings of the Second IEEE and ACM International Symposium on Mixed and Augmented Reality*, pp. 258–259, Oct. 2003.
5. Casper, J., and Murphy, R. R. “Human-Robot Interactions During the Robot-Assisted Urban Search and Rescue Effort at the World Trade Center,” *IEEE Tansactions on Systems Man, and Cybernetics, Part B*, vol. 33, no. 3,pp. 367-385, Jun. 2003.
6. D’Zmura, M. “Color in visual search,” *Vision Research*, vol. 31, no. 6, pp. 951-966, 1991.
7. Hartman, F., and Maxwell, S., “Driving the Mars Rovers,” *Linux Journal*, 2004. http://www.linuxjournal.com/article/7570
8. Jarosz, W. “Fast Image Convolutions.” *ACM SIGGRAPH Workshop @ UUIC 2001*. http://www.acm.uiuc.edu/siggraph/workshops/wjarosz_convolution_2001.pdf.

9. Kangas, K., and Röning, J. "Using code mobility to create ubiquitous and active augmented reality in mobile computing," *Proceedings of the fifth annual ACM/IEEE international conference on Mobile computing and networking*, pp. 48-58, Aug. 1999.

10. Lawson, S., Pretlove, J. R. G., Wheeler, A., and Parker, G. "Augmented Reality As a Tool to Aid the Telerobotic Exploration and Characterization of Remote Environments," *Presence*, vol. 11, no. 4, pp. 352-367, Aug. 2002.

11. Lenser, S., Bruce, J., and Veloso, M. "CMPack: A Complete Software System for Autonomous Legged Soccer Robots," *Proceedings of the Fifth International Conference on Autonomous Agents.* May 2001.

12. Milgram, P., Yin, S., and Grodski, J. J. "An Augmented Reality Based Teleoperation Interface For Unstructured Environments," *Proceedings, American Nuclear Society 7th Topical Meeting on Robotics and Remote Systems*, pp. 966-973, Apr. 1997.

13. Murphy, R. R. "Human-Robot Interaction in Rescue Robotics," *IEEE Transactions on System, Man, and Cybernetics – Part C: Applications and Reviews*, vol. 34, no. 2, May 2004.

14. Nagy, A. L., and Sanchez, R. R. "Critical color differences determined with a visual search task," *Journal of the Optical Society of America*, vol. 7, pp. 1209–1217, Jul. 1990.

15. Neisser, U. *Cognitive Psychology*, Appleton-Century-Crofts, 1967

16. Norman, D., "Toward a Theory of Memory and Attention," *Psychological Review*, vol. 75, pp. 523-536, 1968.

17. Posner, M., Snyder, C., and Davidson, B., “Attention and the Detection of Signals,” *Journal of Experimental Psychology*, vol. 109, no. 2, pp. 160-174, Jun. 1980.

18. Quinlan, M., Chalup, S., and Middleton, R. “Techniques for Improving Vision and Locomotion on the Sony AIBO Robot,” *Proceedings of the 2003 Australasian Conference on Robotics and Automation,* Brisbane, Australia, 2003.

19. Rekimoto, J., and Ayatsuka, Y. “CyberCode: designing augmented reality environments with visual tags,” *Proceedings of DARE 2000 on Designing augmented reality environments*, pp. 1–10, April 2000.

20. Renevier, P., and Nigay, L. “Mobile Collaborative Augmented Reality: The Augmented Stroll,” *EHCI 2001*, pp. 299–316, 2001.

21. Schnädelbach, H., Koleva, B., Paxton, M, Twidale, M., Benford, and Anastasi, R., “The augurscope: refining its design” *Presence: Teleoperators and Virtual Environments*, vol 15, no. 3, pp.278-293, Jun. 2006.

22. State, A., Hirota, G., Chen, D. T., Garrett, W. F., and Livingston, M. A. “Superior augmented reality registration by integrating landmark tracking and magnetic tracking,” *Proceedings of the 23rd annual conference on Computer Graphics and interactive techniques*, pp. 429 – 438, Aug. 1996.

23. Stedmon, A., Kalawsky, R., Hill, K., and Cook, C. “Old theories, new technologies: Cumulative clutter effects using augmented reality,” *Proceeding of the IEEE international conference on information visualization,* pp. 132-137, Jul. 1999.

24. Sternberg, R. J. *Cognitive Psychology*, 3d ed. Belmont, CA: Wadsworth 2003.

25. Tang, A., Owen, C., Biocca, F., and Weimin, M. “New techniques for presenting instructions and transcripts: Comparative effectiveness of augmented reality in object assembly,” *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 73-80, Ft. Lauderdale, FL, Apr. 2003.

26. Treisman, A., “Preattentive processing in vision,” *Computer Vision, Graphics and Image Processing*, vol. 31, no. 2, pp. 156–177, Aug. 1985.

27. Treisman, A., and Gelade, G. “A feature-integration theory of attention,” *Cognitive Psychology*, vol., no. 1, pp. 97-136, Jan. 1980.