A Point-and-Click Interface for the Real World: Laser Designation of Objects for Mobile Manipulation

Charles C. Kemp
Georgia Tech
Atlanta, Georgia 30332
charlie.kemp@bme.gatech.edu

Cressel D. Anderson
Georgia Tech
Atlanta, Georgia 30332
cressel@ieee.org

Hai Nguyen
Georgia Tech
Atlanta, Georgia 30332
haidai@cc.gatech.edu

Alexander J. Trevor
Georgia Tech
Atlanta, Georgia 30332
atrevor@cc.gatech.edu

Zhe Xu
Georgia Tech
Atlanta, Georgia 30332
zhexu@hawaii.edu

ABSTRACT

We present a novel interface for human-robot interaction that enables a human to intuitively and unambiguously select a 3D location in the world and communicate it to a mobile robot. The human points at a location of interest and illuminates it ("clicks it") with an unaltered, off-the-shelf, green laser pointer. The robot detects the resulting laser spot with an omnidirectional, catadioptric camera with a narrow-band green filter. After detection, the robot moves its stereo pan/tilt camera to look at this location and estimates the location's 3D position with respect to the robot's frame of reference.

Unlike previous approaches, this interface for gesture-based pointing requires no instrumentation of the environment, makes use of a non-instrumented everyday pointing device, has low spatial error out to 3 meters, is fully mobile, and is robust enough for use in real-world applications.

We demonstrate that this human-robot interface enables a person to designate a wide variety of everyday objects placed throughout a room. In 99.4% of these tests, the robot successfully looked at the designated object and estimated its 3D position with low average error. We also show that this interface can support object acquisition by a mobile manipulator. For this application, the user selects an object to be picked up from the floor by "clicking" on it with the laser pointer interface. In 90% of these trials, the robot successfully moved to the designated object and picked it up off of the floor.

Categories and Subject Descriptors: I.2.9 [Artificial Intelligence]: Robotics

General Terms: design, human factors.

HRI'08, March 12–15, 2008, Amsterdam, The Netherlands.
Copyright 2008 ACM 978-1-60558-017-3/08/03 ...$5.00.

Keywords: laser pointer interface, mobile manipulation, object fetching, assistive robotics.

1. INTRODUCTION

The everyday hand-held objects commonly found within human environments play an especially important role in people's lives. For robots to be fully integrated into daily life they will need to manipulate these objects, and people will need to be able to direct robots to perform actions on these objects. A robot that finds, grasps, and retrieves a requested everyday object would be an important step on the road to robots that work with us on a daily basis, and people with motor impairments would especially benefit from object fetching robots [18]. For example, a person with ALS could use the robot to fetch an object that he or she has dropped on the floor, or an elderly person who is sitting down could use the robot to fetch an object that is uncomfortably out of reach, Figure 1.

A key issue for this type of application is how the human will designate the object that the robot is to retrieve. If the robot is a personal robot in the sense that it closely follows the person around during the day, then there will be many situations where the human and the robot are observing the same environment and the human wants the robot to retrieve a visible object. In these situations, gesture-based pointing could serve as a natural mode for communication. Furthermore, recent research has demonstrated that autonomous grasping of novel objects is feasible via approaches that take advantage of compliant grippers, tactile sensing, and machine learning [5, 11, 16]. These successes indicate that providing a location of an object to a robot may be sufficient to command the robot to fetch the object.

In this paper, we present a new user interface that enables a human to unambiguously point out a location in the local environment to a robot, and we demonstrate that this is sufficient to command a robot to go to the location and pick up a nearby object. In the next section we review related work. We then describe the design of the laser pointer interface. After this, we present tests showing that people can use the interface to effectively designate locations of objects placed around a room. Finally, we demonstrate that the laser pointer interface can support an object acquisition

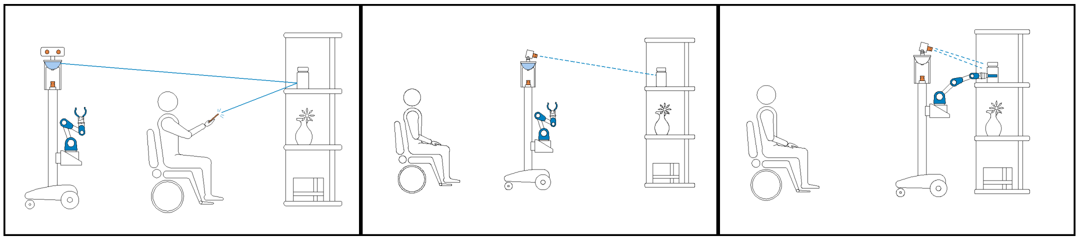

Figure 1: This series of diagrams shows a proposed object-fetching application for wheelchair bound users that makes use of the laser pointer interface we present in this paper. Left: The user points the laser pointer at an object that he or she wants to acquire and illuminates a location on the object ("clicks it"), and the robot detects the resulting laser spot with its omnidirectional camera. Middle: The robot moves its stereo camera to look at the laser spot and makes an estimate of the laser spot's 3D position. Right: The robot grasps the object.

application, where a mobile manipulator navigates to a selected object and picks it up off of the floor.

2. RELATED WORK

Pointing is a fundamental way of communicating. Our system relates to a wide body of work that we summarize here.

2.1 Pointing Through a GUI

GUIs can provide a 2D point-and-click interface with which to control robots. This is common within both teleoperation applications and situations where the robot and human are in the same location. More recently, PDAs have been used to control robots in mobile settings [15, 10]. When controlling a robot from a remote location, direct methods of pointing such as our laser pointer interface are not an option. At best, high-fidelity virtual reality interfaces could give the illusion of directly pointing to a location in the robot's environment.

When in close proximity to one another, direct pointing becomes a viable option with potential advantages. For example, a laser pointer is a compact device that is easy to hold, yet the effective display size for this interface consists of the entire environment visible to both the human and the robot. In contrast, a portable GUI-based interface must contend with potentially conflicting issues, such as display size, device usability, and comfort. A PDA's display size is usually proportional to the PDA's size, and even head-mounted displays introduce issues with comfort and social appropriateness. Pointing through a GUI also requires that the user switch perspectives, which could be frustrating when commanding a mobile manipulator to manipulate an object that is nearby and clearly visible outside of the GUI interface.

2.2 Natural Pointing

Extensive research has been devoted to developing computer interfaces that interpret natural gestures, gaze direction [3], and language so that a user can intuitively select a location or object on a computer screen or in the world [20, 17]. For the great majority of computer applications these approaches have yet to supplant pointers controlled by a physical interface such as a computer mouse.

Communicating a precise location is difficult, especially through language. Human success at this task often relies on shared, abstract interpretations of the local environment. For example, when humans communicate with one another, a coarse gesture combined with the statement, "The cup over there.", may be enough to unambiguously designate a specific object, and by implication, a 3D location or volume.

Given the significant challenges associated with machine perception of objects and people in everyday environments, these approaches seem likely to produce high uncertainty for at least the near term. As an example, recent computer vision research related to the estimation of the configuration of the human arm produces angular errors that would lead to significant positional errors when pointing at an object. When tracking human upper bodies with multiview stereo, the authors of [23] were able to get mean errors of 22.7° for the upper arm and 25.7° for the lower arm. Even if we unrealistically assume that the position of the forearm is perfectly known, and is independent of upper arm angular errors, a forearm-based pointing gesture would have approximately 1.19 meters of error at 3 meters distance. As another example, consider [13] in which the authors looked at using 3D tracking information of a person's head and forearm for gesture classification. It was found that people tend to point along the line connecting the forward direction of the head and the tip of the arm used for pointing. Using this head/arm line with favorable conditions resulted in an average of 14.8 degrees of error (i.e., about 0.77 meters error at 3 meters), and 26% of the test gestures were dropped prior to calculating this error due to tracking failures and other issues. Likewise, [4] describes a robotic system capable of understanding references to objects based on deictic expressions combined with pointing gestures. However, the approach requires preexisting knowledge of the objects to which the user is referring, and has significant uncertainty in detecting arm pose, which forces pointing gestures to be made very close to the objects.

The laser pointer interface results in significantly less error than these methods and offers two additional benefits. First, the green spot provides a feedback signal to the human about how well he or she is pointing at the object. Second, the laser pointer interface directly outputs a 3D position corresponding to the laser spot, while pointing gestures output a ray that must be further interpreted in order to determine the intersection between it and the environment.

2.3 Pointing with Intelligent Devices

A number of intelligent devices have enabled people to select locations in a physical environment by directly pointing at them. The XWand [22] and WorldCursor [21], developed at Microsoft Research are especially relevant to our work.

The XWand is a wand-like device that enables the user to point at an object in the environment and control it using gestures and voice commands. For example, lights can be turned on and off by pointing at the switch and saying ”turn on” or ”turn off”, a media player can be controlled by pointing at it and giving spoken commands such as ”volume up”, ”play”, etc. This work is similar in spirit to ours, since we would eventually like robots to perform such tasks. However, having a robot perform tasks avoids the need for specialized, networked, computer-operated, intelligent devices. Likewise, a robot has the potential to interact with any physical interface or object in addition to electronic interfaces.

As with our interface, the WorldCursor attempts to enable point-and-click operations in the real world. It consists of a laser pointer projected on the environment from a computer-controlled pan/tilt unit mounted to the ceiling. The user controls the WorldCursor by moving the XWand, which is instrumented with orientation sensors that control the WorldCursor differentially. This indirect and arguably awkward method of control was designed to avoid the complexities of estimating the position and orientation of the XWand with respect to the room. For this work, the selected location must be registered with 3D models of the room and the room's contents.

Several other examples of intelligent pointing devices exist, such as Patel and Abowd's 2-way laser assisted selection system [14]. In this work, a cell phone instrumented with a laser communication system selects and communicates with photosensitive tags placed on objects and in the environment. Teller et al. have demonstrated a device that estimates its pose relative to the world frame by using onboard orientation sensors and positions acquired via a form of “indoor GPS” called The Cricket Indoor Location System [19]. This device was designed to enable a user to mark the world by pointing at locations with a laser dot. When pointing the laser dot at something, an integrated laser range finder would estimate the range from the device to the marked location. By combining the pose of the device with this range estimate, the device could produce a 3D position in the world frame to be associated with the location.

All of these systems rely on instrumented devices, instrumented environments, instrumented objects, elaborate models of the world, or some combination thereof. In contrast to these systems, our system does not require any special intelligence in the pointing device, any instrumentation of the environment besides what the robot carries, any instrumentation of the objects to be selected, nor any models of the room or its contents. The user can efficiently and directly select points in the environment using a simple, compact, off-the-shelf device. No alterations to the environment are necessary, so the interface is as mobile as the human and robot. We specifically designed our system to support human-robot interaction for robotic applications. Providing a clear 3D location to a robot enables a robot to direct its sensory resources to the location, move to the location, and physically interact with the location in ways that are not achievable by an inert intelligent environment.

2.4 Laser Designation

Lasers have been used to designate targets for human-human, human-animal, and human-machine interactions. For example, during presentations speakers often direct the audience's attention towards particular points using a laser pointer. Animals, such as cats, can be sensitive to a laser spot. Some quadriplegics use monkeys as helper animals to assist them with everyday tasks. These helper monkeys have served as an inspiration for our research. Helper monkeys have been successfully directed to perform sophisticated tasks, such as bringing a drink and operating light switches, by a mouth-held laser pointer and simple words [2].

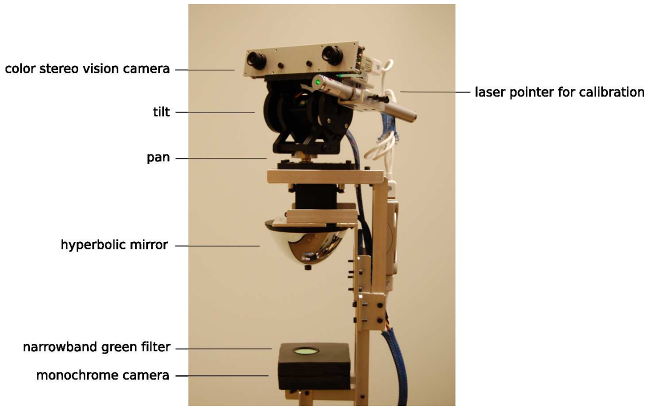

Figure 2: The robot's head and the laser pointer interface.

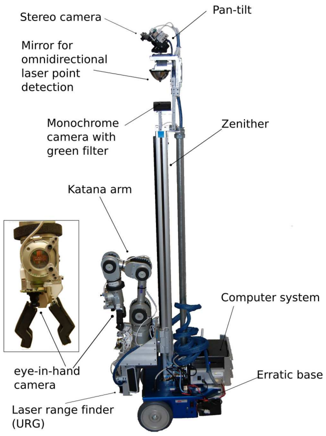

Figure 3: An image of the entire mobile manipulator with the integrated interface system (i.e. the robot's head)

The user interface for the MUSIIC (Multimodal User-Supervised Interface and Intelligent Control) stationary table-top rehabilitation robot from the University of Delaware incorporated a laser pointer with which the user could select some objects over a small area, but the task was highly constrained with just a few well-modeled objects on a tabletop under very controlled conditions [9]. Researchers at MIT have recently attempted to use laser pointers to mark and annotate features in the world for a wearable tour-guide system [8]. More broadly, lasers have been successfully used by the military to mark targets for laser-guided weaponry [1].

3. IMPLEMENTATION

Within this section we present our implementation of the laser pointer interface. This implementation is the result of careful design with the goal of creating a low-complexity device that robustly detects the laser spot over a large area and accurately estimates its 3D position. The interface performs three main steps.

1. Detect the laser spot over a wide field of view.
2. Move a stereo camera to look at the detected laser spot.
3. Estimate the 3D location of the spot relative to the robot's body.

Figure 1 illustrates these steps in the context of a mobile manipulation task.

3.1 Observing the Environment

If the robot could observe the laser spot at all times and over all locations, it would be able to respond to any selection made by the human. Instrumenting the environment could enhance the robot's view of the laser spot, but this has drawbacks, including the complexity of installation, privacy concerns, and a lack of mobility. Even with room-mounted cameras, occlusion and resolution can still be an issue. Other than the laser pointer, our design is completely self-contained. Wherever the robot goes, the system goes with it, so it can be an integral part of its interactions with people. Furthermore, the onboard system is able to directly estimate the 3D location of the spot with respect to the robot's frame of reference, where it is most needed, rather than transforming a position in the room's frame of reference into the robot's frame.

To maximize the area over which the robot can detect the laser pointer, we use a catadioptric [12], omnidirectional camera placed at approximately the height of an average person's neck (1.6m), see Figure 3. This is well-matched to human environments, where objects tend to be placed such that people can see them. As shown in Figure 2, the omnidirectional camera consists of a monochrome camera with a narrow-angle lens (40 degrees horizontal FoV) that looks at a mirror. The resulting camera has a view of approximately 115 degrees in elevation and 280 degrees in azimuth with a blind spot in the rear, see Figure 5. The camera's field of view goes from the floor in front of the robot up to the tops of the walls around the robot.

3.2 Detecting the Laser Spot

By design, green laser pointers are highly visible to people. Laser pointers of this type are typically used to facilitate human-human interactions, so this visibility is a requirement. Unfortunately, this high visibility does not directly translate to high visibility for robots using video cameras. During testing, we found that the high brightness and strong greenness of green laser pointers were often poorly translated into the digitized images used by our robots. When observed by a single-chip, color CCD or CMOS camera (Unibrain Fire-i and PointGrey Firefly respectively), the high-brightness would often saturate red, green, and blue pixels, leading to a white spot. Turning down the gains on these cameras facilitated detection on reflective surfaces (e.g., white walls), but seriously hindered detection on dark surfaces.

We explored the use of several methods to enhance detection under these unfavorable circumstances, including searching for circles and detecting image motion, since people do not hold the laser pointer perfectly steady. With these approaches detections were increased but the overall system robustness was lacking. A high-dynamic range camera or a 3 chip camera might overcome some of these issues.

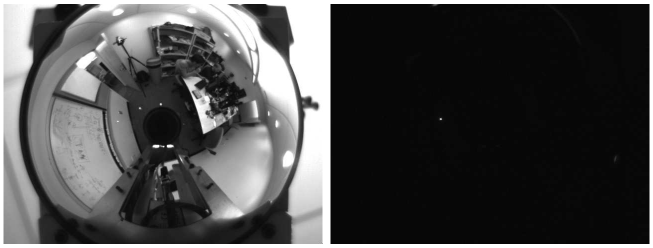

Figure 4: The left image shows the image from the omnidirectional camera without a filter. The right image shows the same environment captured through the narrow-band filter. The spot on the right image corresponds with a laser spot in the world. The bottoms of these images looks towards the back of the robot, and are occluded by the robot's body.

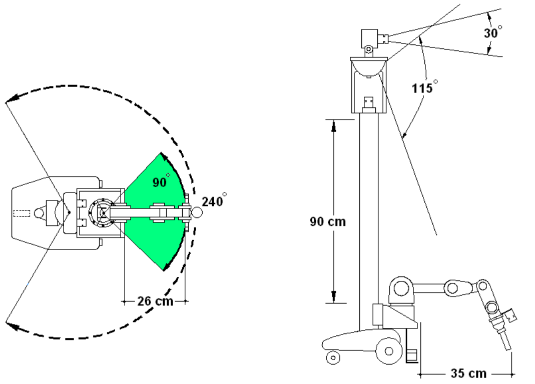

Figure 5: Left, the manipulator workspace shown in green along with the catadioptric camera's field of view. Right, vertical field of view of the stereo and catadioptric camera.

We chose a direct approach to achieving robustness by introducing a monochrome camera with a narrow-band, green filter matched to the frequency of green laser pointers, see Figure 6. As shown in Figure 4, the filter drastically reduces the total incoming light and lets the laser light come through. The monochrome camera effectively increases the pixel resolution, since only green sensitive pixels of a color camera would respond strongly to the incoming green light. Since the filter we use works by way of destructive interference, performance degrades significantly if light comes in at an angle greater than 15 degrees. Consequently, the monochrome camera must have a narrow-angle lens.

The resulting omnidirectional camera supports robust detections, as demonstrated by the tests described in the Results Section. The detector for these tests simply returns the location of the pixel with the maximum value above a fixed threshold. For this paper, a single detection is enough to cause the stereo camera to be moved.

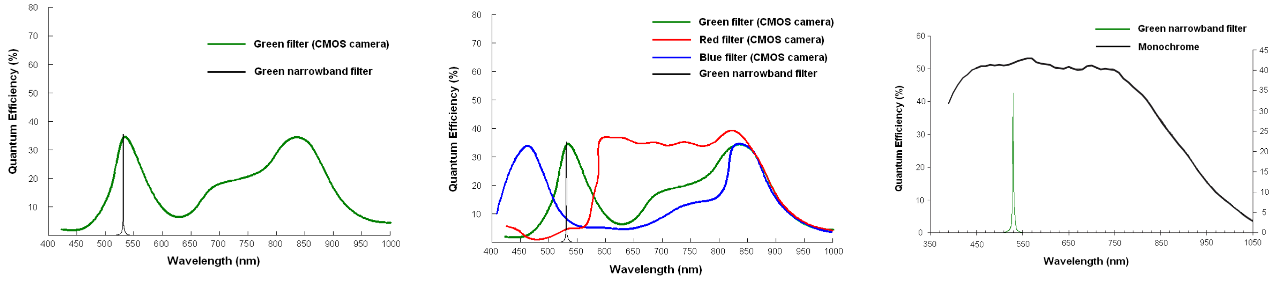

Figure 6: The left image shows the approximate spectral response of our omnidirectional camera with a green narrowband filter compared with the green filter of a common CMOS camera. The middle image shows these response curves in the context of all the color components from a CMOS camera. The right image shows the response curve for the monochrome camera that images the mirror.

3.3 Looking at the Laser Spot

After a laser spot is detected by the omnidirectional camera, the pan/tilt stereo camera with narrow-angle lenses (40 degrees horizontal FoV each) is pointed at the spot. This movement of the stereo camera provides feedback to the human and allocates more perceptual resources to the selected location. After this movement, additional laser spot detections by the omnidirectional camera are ignored while the stereo camera attempts to detect and analyze the laser spot that triggered the move.

The mapping from detection locations on the omnidirectional camera to pan/tilt coordinates is non-trivial, due to the geometry of the mirror, and is further exacerbated by the position of the monochrome camera, which is placed slightly forward in order to achieve a better frontal view. We have developed an easy-to-use, automated calibration method that in our tests has resulted in 99.4% success rate when moving the stereo camera to see a laser spot detected by the omnidirectional camera.

The procedure automatically finds a suitable mapping by using a head-mounted green laser pointer, see Figure 2. During calibration, the pan/tilt head scans around the room with the laser pointer turned on. At each pan/tilt configuration, the omnidirectional camera performs a detection of the laser spot resulting from the head-mounted laser pointer and stores the detection's coordinates in the omnidirectional image paired with the current pan/tilt configuration. This data is then smoothly interpolated using radial basis functions to create a mapping from omnidirectional image coordinates to pan/tilt commands. When a detection is made, an interpolation is performed to produce an appropriate pan/tilt configuration to which the stereo camera is moved.

3.4 Estimating the 3D Location

Once the stereo camera is looking at the laser spot, we turn down its brightness slightly to help distinguish between the laser spot and other bright spots in the environment. We then perform a detection in the left and right cameras and estimate the 3D location that corresponds with the detections. This 3D point is then transformed through the pan/tilt axes into the coordinate system of the robot's mobile base. The system attempts to detect the laser spot and estimate its 3D location eight times. If it succeeds at least once, all of the resulting estimates are averaged and the result is returned as the location of the "click". If it fails to make any detections over these eight attempts, the system goes back to looking for a laser spot with the omnidirectional camera.

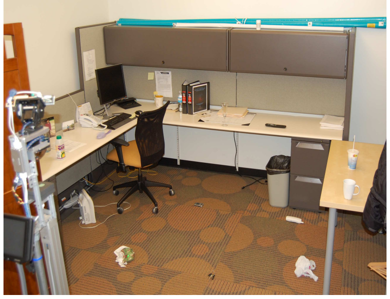

Figure 7: The experimental setup used for the laser-designation tests.

The detector uses a combination of motion, size, and brightness to detect the spot reliably in each camera. First, we look for motion by using image differencing, since people do not hold the laser pointer perfectly steady and small motions of the laser pointer result in observable motions of the laser spot. We threshold this difference image to select for changes in brightness over time that are large enough to have been generated by the motion of the laser spot. Next, we filter out connected binary regions that are unlikely to have been generated by the laser spot due to their large size. We also remove regions whose pixels all have green component values below a threshold. If any regions remain, one of them is labeled as a detection for the associated camera. If a detection is found in each camera simultaneously, we use the pair of image coordinates to generate a coarse stereo estimate. We then use template matching between the two cameras to obtain a subpixel resolution estimate of the 3D position of the laser spot.

The 3D position produced by the stereo camera is transformed to a point in the robot's base reference frame. This transformation depends upon the rotation of the pan/tilt unit. Stereo estimation and the transformation through the pan/tilt unit are dominant sources of error for the system. Stereo error estimates are prone to error in depth, while the pan/tilt RC servos tend to produce errors around the axes of rotation due to their limited angular resolution.

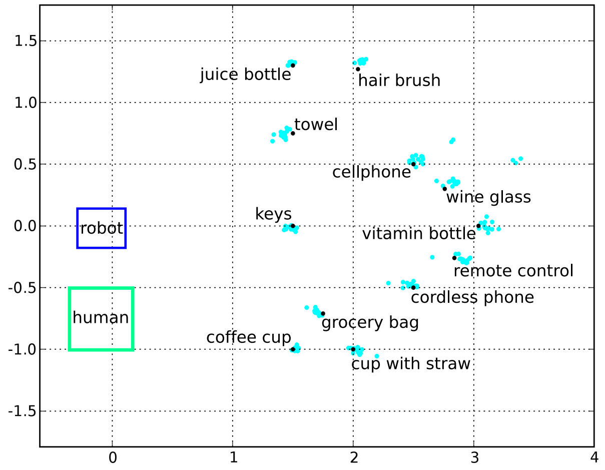

Figure 8: The results of the 178 3D estimates from an overhead view using an orthographic projection. The robot is located at (0,0) and faces to the right. The axes are in meters. The results of a single trial are shown in blue. The black dots are placed at the hand-measured locations associated with the objects.

4. RESULTS

We evaluated the efficacy of the laser pointer interface with two tests. The first test assessed the interface's ability to detect the selection of a location and estimate its 3D position. The second test evaluated the performance of the interface in the context of mobile manipulation.

4.1 Designating Positions of Objects

We conducted a test to determine the robustness of the laser detection and position estimation performed by the interface when using common manipulable objects. The objects that we used can be seen in Figure 9. We chose these objects because of their status as common household items and their distinct appearances when illuminated by the laser pointer. To test the laser pointer's effectiveness as an interface for designating objects, five lab members attempted to "click" 12 objects placed around a room. Each lab member attempted to "click" each object 3 times for a total of 5x12x3 (180) trials of the interface. Each user sat beside the robot to simulate how a person in a wheelchair might use the device. The objects were placed in a variety of locations, and their distances to the robot were measured by hand. The robot was stationary for the duration of the test. The test area was a standard office environment (see Figure 7). Using the festival speech synthesis engine, the robot asked the user to select a particular object. Each user was instructed to wait for the robot's audible "go" before attempting to designate the specified object. After this, the robot would attempt to produce a 3D estimate within 10 seconds by using the procedure described in Section 3. If the time frame was exceeded prior to the generation of a 3D estimate, the trial was deemed a failure.

Table 1: Quantitative results from the laser-designation tests.

Object Location	Object Name	Distance from robot (m)	Average Error (m)
Floor	Keys	1.50	0.05
	Cordless Phone	2.55	0.07
	Cell Phone	1.68	0.14
	Towel	1.68	0.10
	Grocery Bag	1.89	0.07
Desk	Remote Control	2.85	0.12
	Hair Brush	2.40	0.08
	Cup with Straw	2.24	0.07
	Wine Glass	2.78	0.30
	Juice Bottle	1.99	0.05
	Coffee Cup	1.80	0.04
Shelf	Vitamin Bottle	3.04	0.08

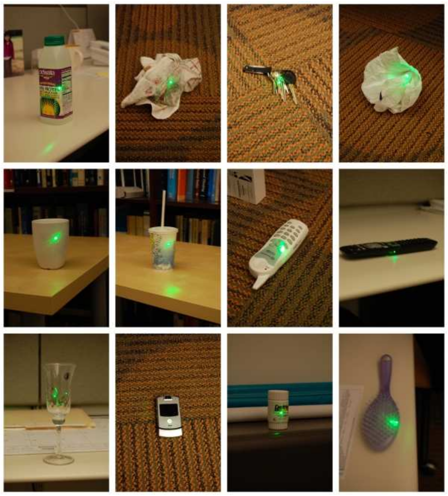

Figure 9: The set of objects used for the laser-designation tests.

This test resulted in 178 successful "clicks" out of 179 attempts (99.4%). Each "click" returned a 3D estimate of the object's location. For these results, we have removed one of the 180 original trials because the user "clicked" the wrong object. Out of the remaining 179 trials, we recorded only a single failure. This single failure appears to have resulted from the user initially missing the object and hitting a far wall, which caused the stereo camera to be grossly misdirected such that the user's subsequent illumination of the object was out of the stereo camera's field of view. Ordinarily, we would expect the system to recover from this situation by using the omnidirectional camera to detect the laser spot again. However, the time limit of 10 seconds was exceeded prior to a successful correction. Except for this failed trial, all trials took less than 3 seconds.

The 3D estimates produced by the 178 trials had an average error of 9.75 cm with respect to hand-measured locations. For these hand-measured locations, we attempted to measure the location of the center of the object's surface

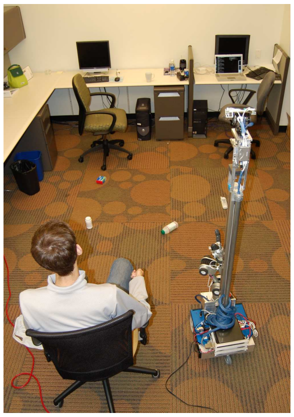

Figure 10: The experimental setup used for the object-acquisition tests.

closest to the user. We did not instruct the users to illuminate this particular location on the objects, which may have resulted in larger errors. Table 1 shows the average error for each object. Figure 8 shows an overhead view of the test area and the resulting 3D estimates.

Except for the remote control, the cell phone, and the wine glass, the average error for each object was no greater than 10cm. The error in the remote control estimates reflect the difficulty of holding the laser on an object with a very small visible profile, 2cm x 22cm. This small target resulted in overshoot. For both the cell phone and the wine glass, performance degraded due to reflection of the laser light. The wine glass resulted in the worst performance due to transmission of the laser light through the glass.

4.2 Mobile Manipulation: Grasping Selected Objects

We performed a second experiment in order to investigate the use of our human-robot interface for mobile manipulation. For this experiment, a single lab member used the interface to select objects for our mobile manipulator to approach, grasp, and lift off of the ground. At any time during the test, three to six objects were strewn about the floor of the office environment. The robot was placed next to a chair in which the lab member was sitting with a laser pointer in hand, see Figure 10.

Except for the designation of the target by the user, this demonstration application was fully autonomous. The robot would detect the laser spot, move its stereo camera to observe it, make a 3D estimate of its location in the coordinate system of its mobile base, navigate towards the selected location while using a laser range finder that scans across the floor, perform final alignment with the object for grasping via the laser range finder, reach out towards the object using the laser range finder, visually segment the object using an eye-in-hand camera, and then attempt to grasp the object from above. Further details of this autonomous operation are beyond the scope of this paper. For this test, the autonomous behavior serves to demonstrate the effectiveness of our human-robot interface in the context of an important robotic application. Specifically, this is the first half of an application that will retrieve designated objects for a person from the floor, which is a robotic capability that has the potential to meet a well-documented need of people with severe motor impairments [18].

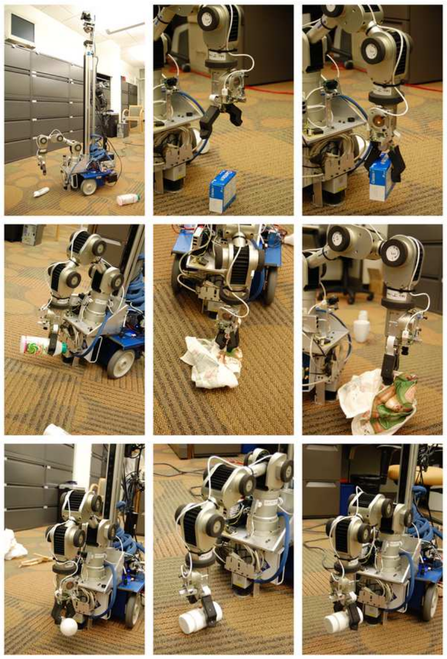

Figure 11: Pictures taken while testing the object-acquisition application.

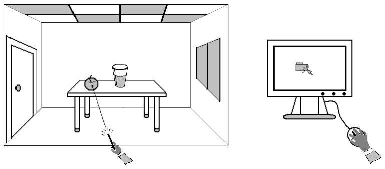

Figure 12: Using the human-robot interface we present in this paper, a person can select a 3D location in the real world with a laser pointer (left). This system is analogous to the selection of 2D locations on a graphical display with a computer mouse (right).

Using this experimental setup, ten trials were performed. The sequence of objects selected was: juice bottle, vitamin bottle, juice bottle, vitamin bottle, allergy medication box, poseable figurine, light bulb, towel, cordless phone, coffee cup. In all ten of these trials, the robot successfully determined which object was selected and navigated to the object such that the object was in manipulable range. The system was also able to successfully grasp nine out of the ten objects, failing only on the cordless phone. Figure 11 shows objects being acquired during the test.

5. DISCUSSION

The majority of personal computers (PCs) today have a graphical user interface (GUI) that allows a user to unambiguously select graphical elements by pointing to them and pressing a button. This point-and-click style of interaction has been extremely successful. Today it enables non-specialist users to efficiently work with a computer to perform sophisticated tasks as diverse as file management, music editing, and web surfing. A critical step towards this form of interaction was the development of the computer mouse at SRI in the 1960's [6], which enabled people to intuitively select a location on a graphic display. This eventually led to the Apple Macintosh in 1984 [7], which popularized the use of a mouse to work with a windowing system.

As with a computer mouse, our interface's operation is straight forward. The human points at a location of interest and illuminates it ("clicks it") with an unaltered, off-the-shelf laser pointer, see Figure 12. The robot detects this location and estimates its 3D position and performs an associated function, thereby forming a point-and-click interface for the real world.

In the long-term, we expect this style of interface to support a diverse array of applications for personal robots, much as point-and-click interfaces support diverse personal computer applications today. By selecting objects, a user should be able to command a robot to perform a variety of tasks. Clearly, additional modalities and autonomy could be complementary to the laser pointer user interface. For example, as object recognition technology improves, it could be integrated into the user interface in order to provide additional context for the robot's actions. If the user points to a light switch and clicks, the robot could recognize it as a light-switch and infer the desired action. Likewise, by clicking on a screwdriver followed by a screw, a robot could recognize the objects and infer that the screw is to be tightened or loosened. Similarly, combining the laser pointer interface with speech could be a powerful approach.

As our results indicate, the current laser pointer interface is robust enough for realistic applications. However, studies with users who are unfamiliar with robotics in the context of full applications will be required to assess the true efficacy of this system and inform future revisions. We believe that this new type of human-robot interface opens up rich areas for investigation of human-robot interaction. Once a human and robot can easily and robustly communicate 3D locations to one another, qualitatively different interactions may be possible.

6. ACKNOWLEDGMENTS

We thank the reviewers for their helpful suggestions. We thank Prof. Julie Jacko for her invaluable advice and support. We thank Dr. Jonathan Glass for his insight and encouragement. This work also benefited from discussions with Prof. Seth Teller at MIT and his student Albert Huang. We graciously thank the Wallace H. Coulter Foundation for funding this work as part of a Translational Research Partnership in Biomedical Engineering Award, "An Assistive Robot to Fetch Everyday Objects for People with Severe Motor Impairments".

7. REFERENCES

[1] Joint laser designation procedures: Training and doctrine command procedures pamphlet 34-3, December 1985.

[2] Helping hands: Monkey helpers for the disabled inc. http://www.helpinghandsmonkeys.org/, dec 2006.

[3] S. Baluja and D. Pomerleau. Non-intrusive gaze tracking using artificial neural networks. Technical report, Carnegie Mellon University, Pittsburgh, PA, USA, 1994.

[4] A. G. Brooks and C. Breazeal. Working with Robots and Objects: Revisiting Deictic Reference for Achieving Spatial Common Ground. 2006.

[5] A. M. Dollar and R. D. Howe. Towards grasping in unstructured environments: Grasper compliance and configuraton optimization. *Advanced Robotics*, 19(5):523–543, 2005.

[6] W. K. English, D. C. Engelbart, and M. L. Berman. Display-selection techniques for text manipulation. *IEEE Transactions on Human Factors in Electronics*, 8(1), March 1967.

[7] F. Guterl. Design case history: Apple's macintosh. *IEEE Spectrum*, 1984.

[8] A. Huang. Finding a laser dot. In *Online progress log for the Ladypack project*, Cambridge, May 2006.

[9] Z. Kazi and R. Foulds. Knowledge driven planning and multimodal control of a telerobot. *Robotica*, 16:509–516, 1998.

[10] C. Lundberg, C. Barck-Holst, J. Folkeson, and H. Christensen. Pda interface for a field robot. *Intelligent Robots and Systems, 2003. (IROS 2003). Proceedings. 2003 IEEE/RSJ International Conference on*, 2003.

[11] L. Natale and E. Torres-Jara. A sensitive approach to grasping. September 2006.

[12] S. K. Nayar. Catadioptric omnidirectional camera. *Computer Vision and Pattern Recognition*, 1997.

[13] K. Nickel and R. Stiefelhagen. Real-time recognition of 3d-pointing gestures for human-machine-interaction. 2003.

[14] S. N. Patel and G. D. Abowd. A 2-way laser-assisted selection scheme for handhelds in a physical environment. *UbiComp 2003: Ubiquitous Computing*, 2003.

[15] D. Perzanowski, A. Schultz, W. Adams, E. Marsh, and M. Bugajska. Building a multimodal human-robot interface. *IEEE Intelligent Systems*, 16, 2001.

[16] A. Saxena, J. Driemeyer, J. Kearns, C. Osondu, and A. Y. Ng. Learning to grasp novel objects using vision. 2006.

[17] B. Scassellati. Mechanisms of shared attention for a humanoid robot. *'Embodied Cognition and Action: Papers from the 1996 AAAI Fall Symposium'*, 1996.

[18] C. A. Stanger, C. Anglin, W. S. Harwin, and D. P. Romilly. Devices for assisting manipulation: a summary of user task priorities. *IEEE Transactions on Rehabilitation Engineering*, 2(4):10, December 1994.

[19] S. Teller, J. Chen, and H. Balakrishnan. Pervasive pose-aware applications and infrastructure. *IEEE CG&A*, pages 14–18, July 2003.

[20] M. Vo and A. Wabel. Multimodal human-computer interaction. *In Proc. Int. Symp. on Spoken Dialogue*, 1993.

[21] A. Wilson and H. Pham. Pointing in intelligent environments with the worldcursor. *Interact*, 2003.

[22] A. Wilson and S. Shafer. Xwand: Ui for intelligent spaces. *Conference on Human Factors in Computing Systems*, 2003.

[23] J. Ziegler, K. Nickel, and R. Stiefelhagen. Tracking of the articulated upper body on multi-view stereo image sequences. 2006.