# El-E: An Assistive Robot that Fetches Objects from Flat Surfaces

Hai Nguyen
Georgia Tech
Atlanta, Georgia 30332
hai@gatech.edu

Cressel Anderson
Georgia Tech
Atlanta, Georgia 30332
cressel@gatech.edu

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

Advait Jain
Georgia Tech
Atlanta, Georgia 30332
advait@cc.gatech.com

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

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

## ABSTRACT

Objects within human environments are usually found on flat surfaces that are orthogonal to gravity, such as floors, tables, and shelves. We first present a new assistive robot that is explicitly designed to take advantage of this common structure in order to retrieve unmodeled, everyday objects for people with motor impairments. This compact, statically stable mobile manipulator has a novel kinematic and sensory configuration that facilitates autonomy and human-robot interaction within indoor human environments. Second, we present a behavior system that enables this robot to fetch objects selected with a laser pointer from the floor and tables. The robot can approach an object selected with the laser pointer interface, detect if the object is on an elevated surface, raise or lower its arm and sensors to this surface, and visually and tacitly grasp the object. Once the object is acquired, the robot can place the object on a laser designated surface above the floor, follow the laser pointer on the floor, or deliver the object to a seated person selected with the laser pointer. Within this paper we present initial results for object acquisition and delivery to a seated, able-bodied individual. For this test, the robot succeeded in 6 out of 7 trials (86%).

## 1. INTRODUCTION

In this paper we present our platform El-E (Elevated - Engagement), a robot designed around two key innovations. First, El-E is equipped with a laser pointer interface that detects when a user illuminates a location with a green laser pointer and estimates the 3D location selected by this "point-and-click". This enables a user to unambiguously communicate a 3D location to the robot with modest effort, which provides a direct way to tell the robot which object to manipulate or where to go. Second, El-E is able translate its manipulator and associated sensors to different heights, which enables it to grasp objects on a variety of surfaces, such as the floor and tables, using the same perception and manipulation strategies. This effectively takes advantage of a common symmetry found within human environments that has likely evolved due to people's desire to place objects in stable configurations. Smooth flat surfaces that are orthogonal to gravity are rarely found outdoors, but are nearly ubiquitous in built-for-human environments.

To use our robot helper, the human points at an item of interest and illuminates it with an unaltered, off-the-shelf, green laser pointer. The robot then moves near the perceived 3D location using the laser range finder. If the location is on the floor, the robot will attempt to grasp the object. If the location is not on the floor, the robot uses the laser range finder to scan for an elevated surface. Once the elevated surface is found the robot docks with the surface and attempts to grasp the object. In both cases grasping is accomplished through a combination of visual segmentation and tactile sensing. After successfully grasping and lifting an object, the robot will await a command from the user. If the user illuminates him or herself with the laser pointer, the robot will deliver the object. If the user illuminates a table, the robot will place the object on the table. If the user illuminates the floor near the robot, the robot will follow the laser pointer.

## 2. MOTIVATION

Autonomous mobile robots with manipulation capabilities offer the potential to dramatically improve the quality of life for people with motor impairments. With over 250,000 people with spinal cord injuries and 3,000,000 stroke survivors in the US alone, the impact of affordable, robust assistive manipulation could be profound [3, 4]. Moreover, as is often noted, the elderly population worldwide is increasing substantially as a percentage of overall population, and there are over 16,000,000 people currently over the age of 75 in the US [16] . This aging population creates a real need for affordable, robust robotic assistance, since 20% of people in the US between 75 and 79 years of age have been shown to require assistance in activities of everyday living,

*The Robotic Helpers Workshop at HRI'08,* March 12, 2008, Amsterdam, The Netherlands.

and this percentage increases precipitously with age, with 50% of people over 85 years of age requiring assistance [1].

Currently, this assistance is most often provided by a human caregiver, such as a spouse or nurse, which reduces privacy and independence, and often places a heavy burden on a loved one or entails high costs. Highly trained animals, such as service dogs or helper monkeys, can also provide physical assistance, but they come with a host of other complications, including high costs ($17000-$35000), difficult training, reliability issues, and their own need for care [2, 10].

As members of the Healthcare Robotics Lab at Georgia Tech, we are working to develop assistive, autonomous mobile manipulators that can meet the needs of people with motor impairments in a manner similar to helper animals and human assistants. We consider object fetching to be an especially important area for research, since people with motor impairments have consistently placed a high priority on the ability to retrieve objects from the floor and shelves [19], and since object fetching is a fundamental capability for autonomous robot manipulation upon which future assistive applications could be built.

Trained helper monkeys for quadriplegics serves as an inspiration for our research. People direct these monkeys using a laser pointer and simple words. This style of interaction helped inspire our design, and helps validate the feasibility of our robot's interface from a usability stand point. With helper monkeys, a quadriplegic can operate the laser pointer with his or her mouth, but other interfaces would also be possible, such as a joystick or sip-and-puff interface to a servoed laser pointer. When discussing his life before he had a helper monkey, Chris Watts, a quadriplegic, stated in an interview with CBS, "If I dropped the phone on the floor, if I dropped my water on the floor, or I dropped my pills or any of those kinds of things, they stayed on the floor until someone came home," [22]. For many people with ALS or spinal cord injuries, the retrieval of dropped objects is a significant concern and a frequent need.

## 3. ROBOT DESIGN

The robot, El-E, is primarily constructed from off-the-shelf components as seen in Figure 1. Its mobile base is an ERRATIC platform from Videre Design, which includes an onboard computer with a Core Duo processor and 1 GB of memory with which the robot performs all its computation. The computer runs Ubuntu GNU/Linux and we have written most of our software with Python and occasionally C++. We also make use of a variety of open source packages including SciPy, Player/Stage and OpenCV.

The ERRATIC base has differential drive steering with two wheels and a caster in the back. Attached to the center of the ERRATIC base is a 1-DoF linear actuator built from a Festo DGE-25-900-SP-KF-GV, which uses a ball screw, and an Animatics SM2315DT servo. This linear actuator, which we refer to as the zenither, raises and lowers an aluminum carriage containing sensors and actuators in order for the robot to interact with objects at various heights. The servo drives the carriage platform using either torque or position commands. Mounted on this carriage is a 5-DoF Neuronics Katana 6M manipulator with a two finger gripper containing

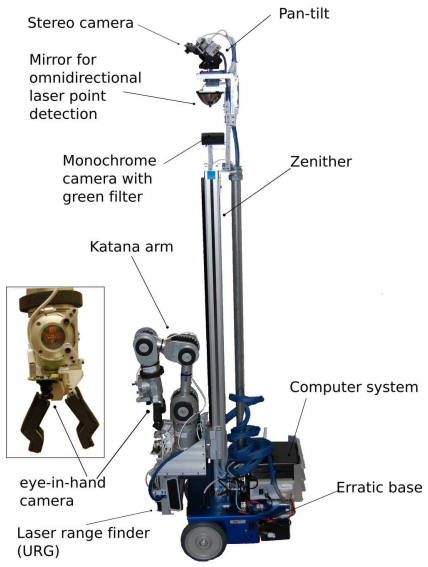


**Figure 1: An image of the entire mobile manipulator with the integrated interface system (i.e. the robot's head)**

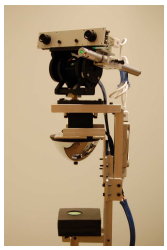

**Figure 2: The laser pointer interface is integrated into the robot's head. It consists of an omnidirectional camera (bottom half) and a pan/tilt stereo camera (top half).**

force and IR sensors.

On a downward extension of the carriage, we mounted a Hokuyo URG laser scanner. The URG scans at a height of 2cm above any flat surface on which the carriage is operating. We mounted a color, eye-in-hand camera, 150° diagonal FoV, at the arm's wrist just behind the gripper. We use the URG and the eye-in-hand camera together for grasping as described in section 4.4. As shown in Figure 2, a monochrome camera, 52° diagonal FoV, a narrow band green filter (532nm, 3 nm bandwidth), and a catadioptric mirror are mounted on top of the zenither. This arrangement is used to detect targets designated by a green laser pointer. Above the catadioptric mirror on a 2-DoF pan/tilt unit we mounted a stereo rig composed of two synchronized color cameras, 52° diagonal field of view (FoV). We used FireFly MV cameras from Point Grey Research for both the monochrome and color cameras.

In total, the robot weighs approximately 38.6 kg and the top of its stereo pair reaches to a height of 1.70 meters above the

ground. During construction, the majority of the robot's weight was intentionally concentrated in its base to assure stability even in the case where the arm is stretched out and its carriage has been raised to the highest possible point. In this configuration the robot can be tilted approximately 10 degrees from upright without tipping over. With the platform raised 3mm from rest–the prefer ed configuration for locomotion–this threshold is increased to 13 degrees. The robot has a total of 11 degrees of freedom and moves at roughly 0.5 meters per second although the maximum speed is 2.0 m/s. The zenither moves at a rate of 0.2 m/s in ascent and 0.46 m/s in decent. Movement of the zenither allow laser scans as low as 2.0 cm and as high as 92.5 cm from the ground to be made.

Currently there are two sources of on-board power: a 12V battery bank and a 24V battery bank. The ERRATIC is capable of supplying 5V for the URG and pan/tilt unit and 19V for the computer. The 24V bank powers the Katana controller and the zenither. The robot is fully untethered and self-powered, and only uses on-board computation.

# 4. IMPLEMENTATION

This section describes several key components we have implemented to give the robot the ability to fetch objects autonomously. We describe how a user can specify a 3D location to the robot with a laser pointer. Then we describe how the robot navigates using the specified 3D location. We show a method for finding flat surfaces such as table tops which allows the robot to reuse a single grasping controller. We describe the grasping controller, and finally explain how the robot delivers a grasped object to the user.

## 4.1 3D Estimation for the Laser Pointer Interface

As seen in Figure 4, having the robot retrieve an object begins with the 3D estimation of a laser point. Our estimation process consists of three stages:

1. Detect the laser spot using the omnidirectional camera.
2. Look at the laser spot using the stereo pair.
3. Estimate the spot's 3D location by detecting corresponding points in the stereo pair.

As shown in Figure 2, the omnidirectional camera consists of a monochrome camera with a 40° medium angle lens that looks at a mirror. As seen in figure 3, the resulting catadioptric camera has a view of approximately 115° in elevation and 240° degrees in azimuth with a blind spot in the rear. The camera views the floor in front of the robot, almost to the tops of the walls around the robot. After a laser spot is detected by the omnidirectional camera, the pan/tilt stereo camera with narrow-angle lenses (40°, 30° vertical) is pointed at the spot. Once the stereo camera looks at the laser spot, it performs a detection in each camera and estimates the 3D location that corresponds to the pair of stereo detections. The details of this interface's operation are described in detail in [11].

We mounted this interface at approximately the height of an average person (1.6m), see Figure 1. This is well-matched to

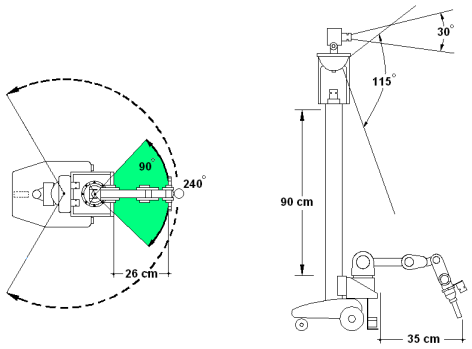


**Figure 3: Left, the manipulator workspace shown in green along with the catadioptric camera's field of view. Right, vertical field of view of the stereo and catadioptric camera.**

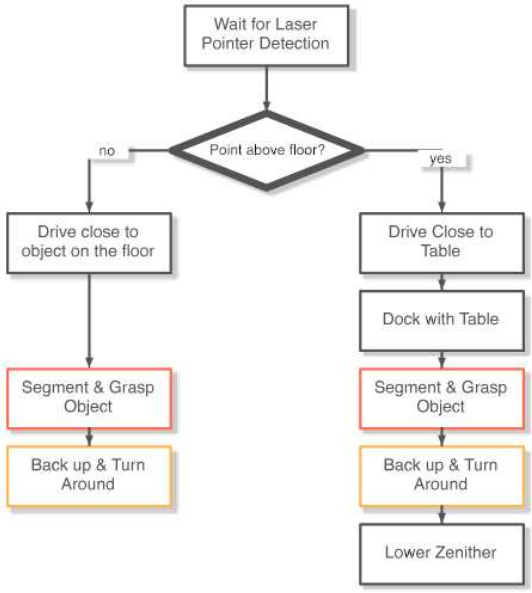


**Figure 4: Object Retrieval Process**

human environments, where objects tend to be placed such that people can see them.

## 4.2 Moving to the Object

For navigation, we use a two step reactive control technique used traditionally for controlling mobile robot bases as in [5]. We use a combination of repulsive potential fields induced by obstacles perceived by the laser range finder and an attractive potential field centered at the laser point to get close to the location indicated by the user. From the 3D laser point estimation process, the robot now possesses a point indicated by the user in the ERRATIC's base frame or $p_l^{base}$. After a point is detected, the robot enters the two step process indicated by the bold diamond in Figure 4. The robot drives toward $p_l^{base}$ until it is within 1.0m. Then the user makes a second selection of the object. This is done to ensure odometric errors do not corrupt the perceived loca-

tion of $p_l^{base}$. The robot selects a behavior to retrieve objects on the floor if the point is less than 0.3 meters in height, as shown in Figure 5. Otherwise the robot switches to its "grasping on elevated surface" behavior, shown in Figure 6 — the process of detecting the height of the flat surface and docking is described in section 4.3.

In both the case of elevated surface and floor manipulation, our system continues onto the second step of driving using its laser range finder and $p_l^{base}$ to enhance estimation of the object's true location. Each laser scan is first processed to find connected components. From this set of connected components we remove components that are either too small or too large, then weight the elements using $N(p_l^{base}, \sigma)$. In the last step, the group's centroid with the highest weight is then selected as the true object location. This selection process is then carried out at every time step to guide an attractive potential field whose influence stops at 15cm away from the object. At this point, our robot executes its grasping controller.

With the above two step driving procedure our system uses information from the laser scanner and natural user interaction to overcome errors from odometry as well errors inherent in the stereo estimation process for laser points that are far away. By basing decisions such as whether to manipulate on the floor or table, and which object to manipulate only on accurate stereo points estimated at relatively close ranges (around 1.0 m) our robot is able to mitigate most sources of errors that would otherwise cause it to fail.

There is significant interest in extracting planar surfaces from 3D data for improved 3D reconstructions of human environments [6, 9, 23, 21, 12, 20]. Unlike datasets used in previous works, 3D scans provided through the combination of our robot's zenither and its laser range finder does not provide readings of surfaces themselves but only of their boundaries. Since existing techniques depend on flat surfaces having many supporting points, we have developed a method for detecting surfaces whose edges are the only observable components. Moreover, we use this approach to detect the task-relevant feature of the edge of the table, which subsequently is directly used as part of a behavior that servos to the table. This is very different from a traditional reconstruction approach to perception.

### 4.2.1 Surface Detection

The system starts out by using the zenither to move the URG laser scanner from a height of 2cm to 92cm. A 3D scan is obtained by registering the 2D URG scans with the corresponding zenither height. After obtaining a 3D scan, we calculate a function whose maximum should give us the height of the planar surface. As the laser scanner updates only at 10 hz, we perform this procedure twice. In the first pass we move the zenither up quickly over its full range of motion. Using this scan the robot calculates an initial estimate of the surface's height, $h_1^*$. In the second scan, the zenither is moved slowly over the interval $[h_1^* - 5 \text{ cm}, h_1^* + 5 \text{ cm}]$ to calculate a more accurate estimate $h_2^*$.

The 3D laser scan is represented by the function $l(\theta, z)$ which yields a range reading for each angle $\theta$ and elevation $z$. To process these full 3D scans, first, for each horizontal scan made by the laser we extract a set of connected components, defined as a set of consecutive points in a scan within 3cm of the neighboring point. Each connected component is treated as a possible planar surface hypothesis $h$ composed of $(\theta_{start}, \theta_{end}, z)$ where $\theta_{start}$ is the angle in the scan at which the connected component starts, $\theta_{end}$ is the angle at which it ends, and $z$ the height at which the scan was made. Collectively, these hypotheses form the set of hypotheses $H$. The best hypothesis, $h^* \in H$, is then given by:

$$h^* = \underset{h \in H}{\operatorname{argmax}}\, r_\mu^2(h) F_z(h) F_\theta(h) f_p(h) \tag{1}$$

The above expression is a product of the three features $F_z(h)$, $F_\theta(h)$, and $f_p(h)$ with the term $r_\mu(h)$ being the average range reading returned by the function $l(\theta, z)$ for all $\theta_k \in [\theta_{start}, \theta_{end}]$ of a given hypothesis $h$. This term corrects the score so that far away hypotheses having fewer laser scan hits than hypotheses closer to the robot are not penalized unfairly. We now give motivations and precise definitions for the remaining terms used in equation (1).

We observed that while scanning upward, most planar surfaces induces a step function in the laser's range readings. Using this, we can define our first feature as:

$$f_z(\theta, z) = \min(\max(\frac{\partial l}{\partial z} l(\theta, z), 0), 0.5 \text{ m}) \tag{2}$$

The term $\frac{\partial l}{\partial z} l(\theta, z)$, gives derivatives with respect to the height variable $z$, which consists of 655 values in all when using the URG laser scanner. The lower bound of 0.0 makes sure that we are only measuring an increase in the amount of free space observed. The upper bound, with an empirically determined value of 0.5 meters, is also important as otherwise the feature would overly prefer surfaces that are either far from a wall or close to the robot. Integrating $f_z(\theta, z)$ over $\theta$, we obtain an expression that measures the degree of increase in freespace over a range of angles:

$$F_z(h = (\theta_{start}, \theta_{end}, z)) = \int_{\theta_{end}}^{\theta_{start}} f_z(\theta, z) d\theta \tag{3}$$

We define for the second feature a measure of the degree of how smoothly the range readings of a hypothesis varies as a function of $\theta$:

$$f_\theta(\theta, z) = 1 - \frac{1}{K} \frac{\partial^2 l}{\partial \theta^2} l(\theta, z) \tag{4}$$

With normalization constant $K$ defined as:

$$K = \max_{\theta, z} \frac{\partial^2 l}{\partial \theta^2} l(\theta, z) \tag{5}$$

Similar to the first case, after integration, (4) results in $F_\theta(h = (\theta_{start}, \theta_{end}, z))$. The third feature, $f_p(h)$, measures the minimum distance between the hypothesis and the user designated laser point. Assuming that the function $L(\theta, r, z)$ takes a range reading $r$ made by the laser scanner into Euclidean space, the third feature we define as:

$$f_p(h) = \max_{\theta_k \in [\theta_{start}, \theta_{end}]} N(L(\theta_k, r_k, z); p_l^{base}, \Sigma) \tag{6}$$

With the function $N(\mu, \Sigma)$ the probability density function of a 3D Gaussian having mean $\mu$ and variance $\Sigma$. In our system, $\Sigma$ is set to be $20I$ cm with $I$ being a 3x3 identity matrix.

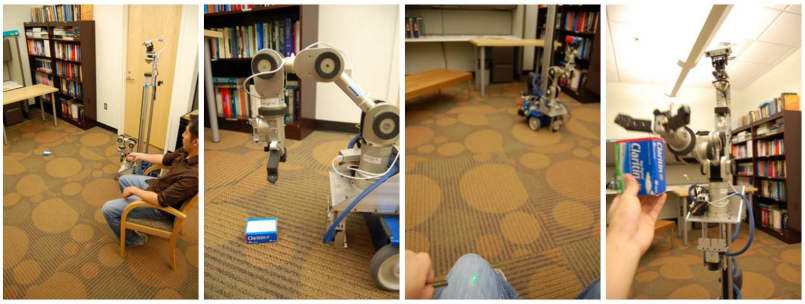

**Figure 5: Sequence to pick an object off the floor.**

**Figure 6: Sequence to pick an object off a table.**

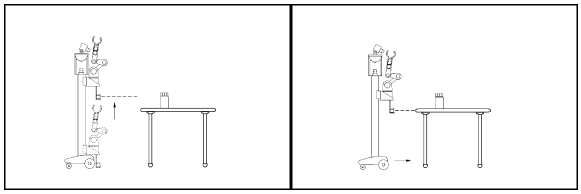

**Figure 7: Using the zenither the robot is able to scan for a table. Once the table with the object is found, the robot servos to the table by aligning itself with the edge.**

Using the above features the algorithm then finds the best hypothesis $h^*$ with (1). To fix this, values of (3) are normalized by dividing values of $F_z(h)$ by its maximum value over $H$ fixing its range to lie in $[0, 1]$.

## 4.3 Finding Flat Surfaces

### 4.3.1 Servoing to Surface's Edge

After having detected the top edge of the surface supporting the object indicated by the user. Our robot drives forward, with its laser range finder set at exactly $h_2^*$, stopping when the surface is within, an empirically determined, 12 cm of the laser scanner. At this point the laser scanner is moved up to clear the surface. Together the 3D scan acquisition and the servoing to the surface's edge compose the docking behavior as seen in Figure 7. After this step, our system then continues on by driving towards the closest connected component to the user designated 3D laser point; however, the robot's behavior is slightly modified compared to the case of manipulation on the floor to prevent it from moving farther forward if the zenither is within a small threshold distance of the laser scan at the height where the second planar hypothesis $h_2*$ was found. Finally, after reaching the object using centroids of connected components sensed by its laser scanner, the robot lowers its sensor carriage so that contact can be made with the surface. At this point the same grasping behavior as was used on the floor is started.

## 4.4 Manipulating on Flat Surfaces

The robot grasps objects by aligning the gripper above the object with the fingers in a suitable orientation and then lowering the gripper onto the object. We describe this method in section 4.4.1. The position and orientation of the object is determined by visual segmentation using the eye-in-hand camera as explained in section 4.4.2

### 4.4.1 Overhead Grasping Behavior

Similar to recent works by [17, 13, 15] our system is able to grasp a variety of unmodeled objects. In contrast to [17], which uses a machine learning algorithm trained on synthetic data generated from object models to locate grasp points on objects in uncluttered environments, our system does not require any training prior to use. Our method also provides an orientation in addition to a grasp position, and makes use of tactile sensing. Work by [13], demonstrates

that haptic feedback on manipulators coupled with simple behavioral controllers can enable successful robot grasping for some objects without the need for elaborate geometric models. Unlike our system, this work uses a stationary platform, but more sophisticated tactile sensors.

Our overhead grasping behavior starts with orienting the robot arm such that the eye-in-hand camera is 30cm above the plane surface at the location of the object to be grasped, as estimated by the laser range finder. The eye-in-hand camera is rigidly mounted to the last link of the robot arm. At this stage, the last link is oriented vertically and the pixels at the center of the image from the eye-in hand camera correspond to rays that are approximately orthogonal to the planar surface over which the robot arm is operating.

Centering the expected object in the camera ensures that the center of the field of view of the camera is overlapping or very near the object to be grasped. This also minimizes 3D projection effects, allowing the system to approximate the problem by treating it as a 2D image processing problem.

After the eye-in-hand camera is positioned vertically above the object to be grasped, we perform visual segmentation on the image captured from the eye-in-hand camera. This is done to refine the estimated position of the object and determine the orientation of the gripper. After visual segmentation, the gripper moves to the appropriate position and orientation and then descends vertically down for an overhead grasp. During the descent, the last link of the robot arm is not constrained to be in a vertical orientation. This increases the workspace of the robot arm and permits variation in relative position of the mobile base and the object (section 4.2). The descent stops when either an IR sensor on the wrist of the gripper (between the two fingers) detects an object, or IR sensors on the finger tips detect a possible collision with the surface or the object. The gripper closes until pressure sensors give high enough readings or the fingers are close to each other. The robot arm then lifts the object by 4cm to test whether the grasp was successful or not. If the pressure sensors in the fingers indicate that no object is present, the system perceives a grasp failure.

### 4.4.2 *Visually Segmenting Objects*

The visual segmentation results in a 2D ellipse that represents the object. The center of this ellipse and the angle of its major axis are used to position the gripper above the object with the gripper oriented to span the object's minimum cross-section.

Our method of visual segmentation closely resembles the approach used by Carson et al's for their work on Blobworld [7]. Flat indoor surfaces tend to have uniform color and texture over large areas. By observing an object from above, the appearance of a surface is more likely to surround and delineate the object. Moreover, the texture will not suffer from perspective effects from this camera angle, so we do not need to use remove perspective effects through the use of a homography. Finally, due to the initial segmentation of the object along the plane by the laser rangefinder, we can expect part of the object to be in the center of the image. We take advantage of this structure by using EM to fit a mixture of Gaussians model with two modes to a set of feature vectors. Each feature vector has six dimensions which consist of a pixel's location in the image, hue, saturation, and three texture features describing a small region around the pixel.

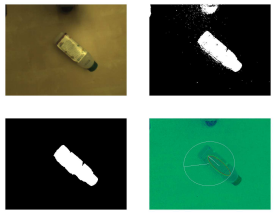

**Figure 8: Different steps of the visual segmentation: image captured by the eye-in-hand camera, classification of each pixel with respect to object and background gaussians, morphological processing to clean up the image and modeling the segmented region with a 2D gaussian (red ellipse).**

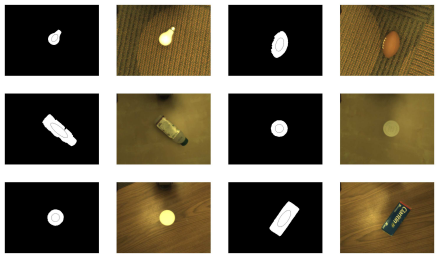

**Figure 9: Collection of sample images and segmentation results.**

We initialize both of the Gaussians to have the same mean such that they are spatially located at the center of the image. We also set one of these Gaussians to have a small mixture weight and low variance, and the other to have a high mixture weight and high variance. This biases these two modes to represent the object and the background, respectively. After using EM, we go back and categorize the entire image with respect to the object and background Gaussians, perform morphological processing to clean up the resulting binary image, and select a connected component that is close to the center of the image and above a threshold size. Finally, we model the spatial distribution of this segmented region with a 2D gaussian, which provides the input to the grasping behavior, as described previously.

## 4.5 Delivery

After a successful grasp has been confirmed, the system will back away from the table 40cm and turn 180°. As in Figure 10, the robot begins waiting for a laser pointer detection. The user's laser command shows the robot where to look and the robot confirms the user's presence. To do this, the robot orients its stereo camera in the direction of the laser detection and uses a Viola-Jones frontal face detector as implemented in OpenCV in a manner similar to [8] to determine whether or not a person is near the laser designated location [14]. As mentioned earlier in Section 4.2, a two step process is used to combat compounding odometric

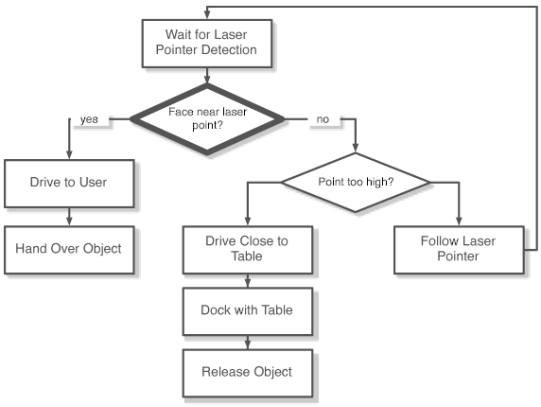


**Figure 10: Object Delivery Process**

error and the absence of a map. If the laser point is farther than 1.0m away, the robot drives toward that location as in section 4.2, detects a second laser indication, and checks for a face; this two step process is indicated the bold diamond in Figure 10. If in the first or second step the robot determines the distance between the face and the laser detection is below 1.0m, the robot uses the same technique as in 4.2 to navigate to 35cm from the user to deliver the object.

If the detected face is not close to the given laser point, the robot will make a decision based on the perceived height of the laser detection. If it is higher than 30cm the robot will initiate the sequence to dock and deliver the object on the table. If it is lower, then the robot assumes that it is to follow the laser pointer. After delivering the object, the robot execute a scripted path that places it next to the person and looking in the same direction, the process then repeats. The advantage of such a system is in the ability for the user to dynamically specify what operations the robot should perform.

By detecting the person's face during delivery, our robot is able to maintain a specified distance from the user. Even though these actions are not planned in advance but arise from the behavioral navigation system, the spirit of maintaining this constraint is similar to work by [18] where robots plan paths that optimize the comfort, safety, and visibility of the action. However, in contrast to this work, our system is implemented on a real robot, and does not require estimation of the full 3D configuration of the human's body.

## 5. RESULTS

To test the effectiveness of our robot in manipulating objects on different surfaces we performed an object retrieval experiment. We chose some common objects to include in the experiment: two vitamin bottles, a football, a medicine box, and a bottle of soy milk. Figure 12 shows the objects that we used for the experiment, and Figure 13 shows the overall setup. We placed two objects each on the floor, a low table, and a higher table. A lab member was seated on the chair and selected an object using a laser pointer. The robot would grasp the selected object and deliver it to the lab member in the chair.

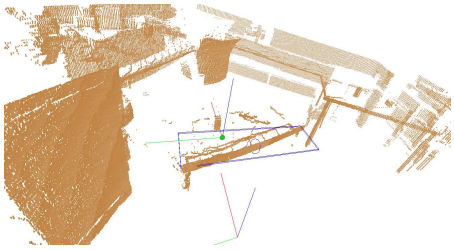

**Figure 11: Plot of 3D scan where the user selected the lower table. The surface detected is outlined in blue; the circular green point is the detected location of the user's laser pointer.**

**Figure 12: Six objects used in the experiment.**

Table 1 shows the results of the experiment. It shows the performance of the robot for the three main components of the object retrieval task: approach, grasping, and delivery. Approach refers to correctly navigating to the selected object, including correct detection of the laser point, deciding whether the selected object is on the floor or on a higher flat surface, estimating the height of the surface, and approaching the object such that it is in the manipulable workspace of the robot arm. The second component, grasping, includes visual segmentation to determine the orientation of the object and then actually picking up the object. Delivery refers to returning the grasped object back to the person sitting in the chair.

The robot was able to successfully navigate itself so that the correct selected object is in its grasp controller's workspace in six out of seven trials (86%). The only failure occurred due to incorrect estimation of the table's height; the rest of the trial was aborted. A successful example of 3D scanning and surface detection can be seen in Figure 11. Grasping was successful in six out of six trials (100%), and delivery of the object to the person in the chair was also successfully completed in six out of six trials.

## 6. CONCLUSION

In this paper we have presented an assistive robot designed to increase quality of life for motor impaired users by helping to retrieve a wide variety of unmodeled objects from the floor or a table and return them to the user. We designed the system to exploit the prevalence of flat planar surfaces such as floors, tables, counter tops, and shelves typically found in human environments. We demonstrated that this ubiquitous structure, to which humans are so accustom, affords a robot useful opportunities for control and perception.

As our results indicate, the current robot system can successfully retrieve objects from the floor and tables. However, further work will be required to robustly extend these capabilities to cluttered environments, novice users, diverse environments, and large sets of everyday objects.

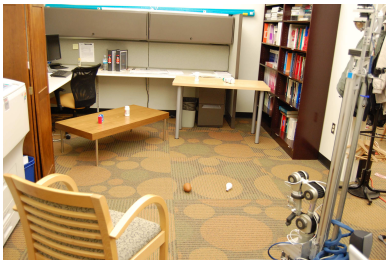

**Figure 13: The experimental setup.**

**Table 1: Retrieving objects from 3 different surfaces**

| Surface | Object | Approach | Grasp | Delivery |
|---|---|---|---|---|
| Floor | Light Bulb | Success | Success | Success |
| | Football | Success | Success | Success |
| Table 1 | Medicine box | Failure | N/A | N/A |
| | Medicine box | Success | Success | Success |
| | Vitamin bottle | Success | Success | Success |
| Table 2 | Soy milk bottle | Success | Success | Success |
| | Vitamin bottle | Success | Success | Success |

## 7. ACKNOWLEDGMENTS

We thank Prof. Julie Jacko, Dr. Jonathan Glass, and Prof. Matt Reynolds for many valuable discussions. 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".

## 8. REFERENCES

[1] Sixty-five plus in the united states. http://www.census.gov, May 1995. Economics and Statistics Administration, U.S. Department of Commerce.

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

[3] National spinal cord injury statistical center (nscisc), supported by the national institute on disability and rehabilitation research. In *http://www.spinalcord.uab.edu/*, 2007.

[4] A. Agranoff, C. Godbout, and J. Johns. Stroke motor impairment. In *http://www.emedicine.com/pmr/topic189.htm*, 2007.

[5] R. Arkin. Motor schema-based mobile robot navigation. *International Journal of Robotics Research*, 8(4):92–112, 1989.

[6] J. Bauer, K. Karner, K. Schindler, A. Klaus, and C. Zach. Segmentation of building models from dense 3d point-clouds. In *Proc. 27th Workshop of the Austrian Association for Pattern Recognition*, pages 253–258, 2003.

[7] C. Carson, S. Belongie, H. Greenspan, and J. Malik. Blobworld: Image segmentation using expectation-maximization and its application to image querying. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(8):1026–1038, 2002.

[8] A. Edsinger and C. C. Kemp. Human-robot interaction for cooperative manipulation: Handing objects to one another. In *Proceedings of the 16th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 2007.

[9] D. Hähnel, W. Burgard, and S. Thrun. Learning compact 3d models of indoor and outdoor environments with a mobile robot. In *Robotics and Autonomous Systems*, volume 44, pages 15–27, 2003.

[10] C. Kemp. Ramona nichols, founder and operator of georgia canines for independence. In *personal communication*, November 2007.

[11] C. C. Kemp, C. Anderson, H. Nguyen, A. Trevor, and Z. Xu. A point-and-click interface for the real world: Laser designation of objects for mobile manipulation. In *3rd ACM/IEEE International Conference on Human-Robot Interaction*, 2008.

[12] Y. Lui, R. Emery, D. Charabarti, W. Burgard, and S. Thrun. Using EM to learn 3D models of indoor environments with mobile robots. In *Proc. 18th International Conf. on Machine Learning*, pages 329–336. Morgan Kaufmann, San Francisco, CA, 2001.

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

[14] Open source computer vision library: Reference manual, 2001.

[15] R. Platt, A. Fagg, and R. Grupen. Reusing schematic grasping policies. dec 2005.

[16] QT-P1. Age groups and sex: 2000 (data set: Census 2000 summary file 1 (sf 1) 100 percent data) for the united states. In *http://factfinder.census.gov/*, 2007.

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

[18] E. A. Sisbot, L. F. Marin, and R. Alami. Spatial reasoning for human robot interaction. In *IEEE Int. Conf. on Intelligent Robots and Systems*, 2007.

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

[20] R. Triebel, W. Burgard, and F. Dellaert. Using hierarchical em to extract planes from 3d range scans. In *Proc. of the IEEE Int. Conf. on Robotics and Automation*, 2005.

[21] Y.-H. Tseng and M. Wang. Automatic plane extraction from lidar data based on octree splitting and merging segmentation. In *Proc. IEEE Int. Geoscience and Remote Sensing Symposium*, 2005.

[22] S. Wahle. Cbs4 news report: Monkey helpers for the disabled, January 2006.

[23] J. W. Weingarten, G. Gruener, and R. Siegwart. Probabilistic plane fitting in 3d and an application to robotic mapping. In *Proc. IEEE Int. Conf. Robotics and Automation*, volume 1, pages 927–932, 2004.