Perception by Proxy: Humans Helping Robots to See in a Manipulation Task

J. Alan Atherton
Computer Science Department
Brigham Young University
3361 TMCB
Provo, UT 84602, USA
j.alan.atherton@byu.edu

Michael A. Goodrich
Computer Science Department
Brigham Young University
3361 TMCB
Provo, UT 84602, USA
mike@cs.byu.edu

ABSTRACT

Robots excel at planning and performing tasks in controlled environments, but poor perception often leads to poor performance in unstructured environments. One typical way of improving robot performance is to give more control to a human operator and then design user interfaces that build the operator's situation awareness. As an alternative, humans can support robot perception to add structure to unstructured environments. We claim that when humans support robot perception, robots can spend more time acting autonomously, which can lead to reduced operator workload and increased overall performance. We present a design process, called *perception by proxy*, and apply it to a simple manipulation task.

Categories and Subject Descriptors

H.5.2 [**Information Systems**]: Information Interfaces and Presentation—*Interaction Styles*

General Terms

Design

Keywords

human-robot interaction, robot perception

1. INTRODUCTION

A robot can be defined as a machine that can sense and act on the world. Although robots can sense the world, they cannot yet perceive the world the same way humans do. Poor perception often leads to poor autonomous performance. One common solution is to give more control to a human operator and design user interfaces to support operator situation awareness. An alternative approach is to have a human augment the robot's perception. With better perception, robots may be able to autonomously perform more complex tasks in unstructured environments than they currently can [1].

With the right human-robot interface, a person can act as a sensor for the robot by giving perceptual cues to the robot. Other researchers in robotics have used human-annotated information to help the robot, but it is usually done to fill in an algorithm that is too time-consuming to implement or as a placeholder for future autonomy [2, 4]. For example, in [4] researchers built a robot system that can autonomously open doors. In order to open the door, the robot must know the approximate location. The researchers manually annotated maps to indicate where the doors are. A general algorithm to find doors on a map could be difficult to develop, but people can be trained quickly to do this perceptual task.

This work constructs a more formal design approach for developing tools that enable a human to augment robot perception. We present a design process for human-robot interaction to help robots overcome roadblocks in perception. Because humans are communicating what they perceive to robots, we call it *perception by proxy*. We apply this process to a simple manipulation task and design a user interface to show how perception by proxy can be useful.

HRI'11, March 6–9, 2011, Lausanne, Switzerland.
ACM 978-1-4503-0561-7/11/03.

2. PERCEPTION BY PROXY DESIGN

To design a system based on perception by proxy, we start by examining the task. Sheridan gives five generic supervisory functions for human-robot interaction [5]. Because our focus is perception instead of supervisory control, we adapt Sheridan's functions to our specific needs. The overall idea is to see where a lack of perception limits the robot, and how a human could help. The following steps guide through the perception by proxy design process.

1. Identify perception requirements for the task.
2. Identify those perception tasks that are outside robot capabilities.
3. Determine robot sensors based on perceptual needs.
4. Choose a way to present sensor data to the human.
5. Establish a way for the human to communicate perception to the robot.

By following these steps we can create a system in which a human supports a robot by augmenting robot perception.

3. EXAMPLE APPLICATION

To illustrate how to design a user interface by this process, we apply the concepts to a manipulation task and implement the design. Our example task is to stack toy alphabet blocks into a specific structure and spell a specific word with a simple robotic manipulator arm. Figure 1a shows our task environment.

Step 1. One perceptual skill required for this task is identifying blocks and their pose (position and orientation)

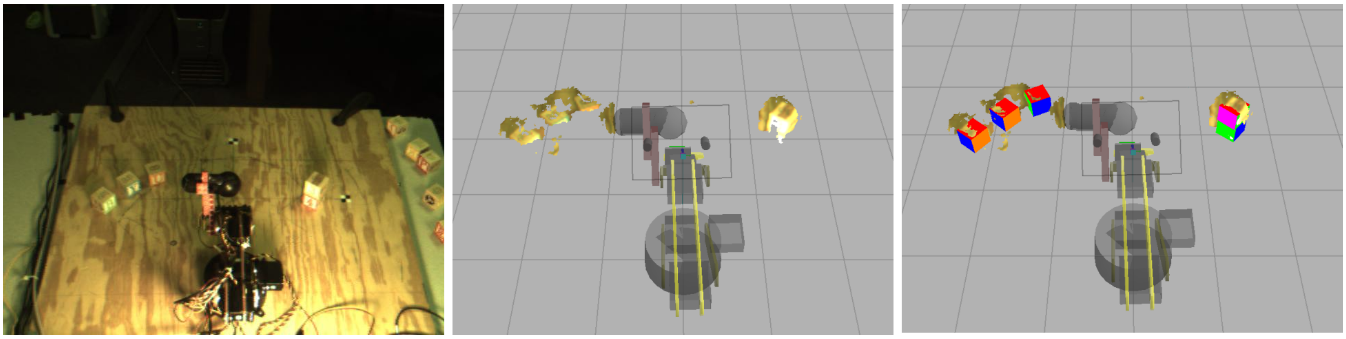

Figure 1: Example task and user interface. (a) Block-stacking task environment from the perspective of the stereo camera. (b) User interface showing virtual robot arm and 3D model data derived from the stereo camera. (c) User interface showing annotations for block positions.

with respect to the robot. The blocks must be stacked a certain way with the correct letter in the correct place, so the task performer needs to know which block goes where.

Step 2. Robot perception is incapable of finding blocks or finding target poses for the blocks in our implementation[1]. As such, the human user must help the robot by giving current block poses and target poses. If the robot was given the current and target block poses, then manipulator motion planning algorithms are sophisticated enough to perform autonomously the remainder of the task.

Step 3. Because the human is remote from the environment, he or she needs to receive information from the robot's sensors to perceive where blocks are and what letters are on the blocks. These needs determine what the robot's sensors must acquire: (a) because the 3D position of each block is important, the visual feedback to the human needs to include some representation of depth; (b) color or intensity information is also important to identify the letters on the blocks. Suitable sensors include stereo cameras, time-of-flight cameras, a color camera combined with a laser rangefinder, etc.

Step 4. A user interface must present sensor information in a way that enables the human to understand the scene. As the user must interpret 3D data and eventually communicate 3D information, it makes sense to use some sort of 3D presentation. Virtual reality, mixed reality, and stereo video all seem to be reasonable methods to present 3D information. Figure 1b illustrates our user interface which uses a type of mixed reality called augmented virtuality (AV). In AV, the display is virtual reality constructed from real-world data (see for example [3]). This AV interface allows the user to adjust the viewpoint of the virtual camera to actively gain accurate spatial understanding of the data.

Step 5. Once the user understands the scene, he or she must communicate the block poses and target poses to the robot. The poses are in 3D space, so the interface must accept input in 3D space. This could be done with special on-screen user interface widgets like those in 3D computer animation software, or by using a 3D input device, for example. We chose a 3D input device for our implementation. The block markers are displayed as multicolored cubes at the same scale as their real-world counterparts. Figure 1c shows the interface with markers positioned where the user believes the real-world blocks are. For target poses, the user places an additional marker and then links two blocks together to signify "this block goes here." If the robot executes poorly or otherwise causes an error, the human can intervene and update the positions of the block markers. With the block poses and target poses, the robot has enough information to perform the task autonomously.

[1]While vision algorithms are sophisticated enough to solve the perception problem for this task, we are purposely excluding a full solution in order to study how our framework performs under imperfect robot perception. Because some people may not have the resources necessary to implement a full vision solution in some cases, we feel that this example of imperfect perception is still informative.

4. CONCLUSION AND FUTURE WORK

We have presented a human-robot system design process that considers communication between humans and robots to compensate for perceptual inadequacies in the robot system. Robots act autonomously within their capabilities, and humans communicate perceptual cues to help the robot progress beyond perceptual roadblocks. We have demonstrated an application of this design process to a manipulation task. Preliminary, subjective evaluations suggest that the perception by proxy user interface can require less time and effort than more traditional HRI approaches.

5. REFERENCES

[1] C.C. Kemp, A. Edsinger, and E. Torres-Jara. Challenges for robot manipulation in human environments [grand challenges of robotics]. *IEEE Robotics & Automation Magazine*, 14(1):20–29, 2007.

[2] C. Liu, W.T. Freeman, et al. Human-assisted motion annotation. In *IEEE CVPR*, pages 1–8, 2008.

[3] C. W. Nielsen. *Using Augmented Virtuality to Improve Human-Robot Interactions*. PhD thesis, Brigham Young University, 2006.

[4] M. Quigley et al. High-accuracy 3d sensing for mobile manipulation: Improving object detection and door opening. In *IEEE ICRA*, pages 2816–2822, 2009.

[5] T. B. Sheridan. *Telerobotics, automation, and human supervisory control*. MIT Press, Cambridge, MA, USA, 1992.