A Panorama Interface for Telepresence Robots

Daniel A. Lazewatsky[*]
dlaz@cse.wustl.edu

William D. Smart[*†]
wds@cse.wustl.edu

[*]Washington University
One Brookings Drive
St. Louis, MO 63130
United States

[†]Willow Garage
68 Willow Road
Menlo Park, CA 94025
United States

ABSTRACT

Telepresence robots are becoming increasingly popular and are increasingly ready to enter use in the real world as stand-ins for remote humans. It is useful, but currently uncommon, to provide the human operator with an approximation of peripheral vision and the ability to saccade around the scene. We have developed an interface which provides peripheral vision to a remote operator by using a motorized pan-tilt camera to create a panorama, and enables the operator to move the camera's gaze within that panorama.

Categories and Subject Descriptors: H.5.2 [Information Interfaces and Presentation]: User Interfaces–Human-Robot Interaction; Graphical User Interface

General Terms: Design; Human Factors

1 Introduction

When interacting with the world, humans have a large amount of sensory information. However, when this interaction is mediated by a robot, that information is limited to what can be gathered by the robot's sensors and sent over the network. More sensors and better sensors would clearly improve the information the operator has about the world, but this solution is likely to incur computational and bandwidth overhead. The interface presented here works to increase the amount of information available to the operator without any additional sensors, while minimizing any extra resource usage.

The interactions possible on a telepresence robot can be broken into three categories: (1) face-to-face interaction with a small number of people while stationary; (2) moving about the environment, controlled by the operator; and (3) interacting with a large number of people while stationary. Here, we focus on the last of these interactions, a common situation for telepresence robots, covering situations such as attending meetings, and presentations, where there is a presenter at a screen or chalkboard, and an audience.

These situations share some properties with situations in which commercial teleconferencing systems are used, but are distinct for several reasons. By virtue of using a mobile robot, there is a one-to-one relationship between remote users and cameras. Due to the mobility not present in teleconference systems, the robot must go through a discovery phase in which it acquires the scene, and must rely on minimal infrastructure such as a wireless network.

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

This interface is appropriate for platforms such as the Willow Garage Texai [8], Anybots QB [1] and other similar telepresence systems. To our knowledge, none of these systems have attempted to provide peripheral vision, either through wide angle cameras or multi-camera setups, or through panoramas, as proposed here.

There are similar solutions in commercial teleconference products, such as the Microsoft Roundtable [5, 2] (later Polycom CX5000 [6]). These provide a stationary, but extreme wide-angle view of the remote location, and automatically saccade to the current speaker. There are also several other similar Pan-Tilt-Zoom (PTZ) based systems [7].

2 Hardware

The target hardware for is the WU Telepresence robot [3] which, like many other telepresence robots, consists of a mobile base, and an approximately human-sized stalk with a screen and motorized pan-tilt camera mounted at the top. Specifically, a Logitech Orbit AF camera is used, which provides 189° pan range and 102° tilt range [4].

A panoramic view can be provided by wide angle cameras, but a narrower-angle camera with motorized pan-tilt was chosen for several reasons. Consumer-grade pan-tilt cameras can be obtained for approximately $100USD, while consumer-grade wide-angle cameras cost, at best, several times more. From an HRI perspective, the motorized pan-tilt conveys valuable information to those interacting with the robot, indicating where the human operator is looking.

3 Interface Design

The interface consists of three sections, described in detail below and shown in Figure 1.

The top section of the screen is devoted to displaying the panorama image. Initially this section is blank, but after scanning the scene, a black and white panorama appears, showing a 250° horizontal and 150° vertical panorama of the robot's surroundings. A live color camera image is overlaid on the panorama at the correct position for the camera's current orientation.

The camera's orientation can be changed by clicking anywhere in the panorama, which specifies the desired position of the center of the camera image. The camera can be zoomed in two ways: either by using the scroll wheel while the cursor is anywhere over the panorama, or double clicking on a target location to simultaneously change the camera position and increase the zoom.

The bottom section of the screen displays the live, unscaled feed from the camera. When the cursor is over the live image, the scroll wheel will zoom the camera, as above.

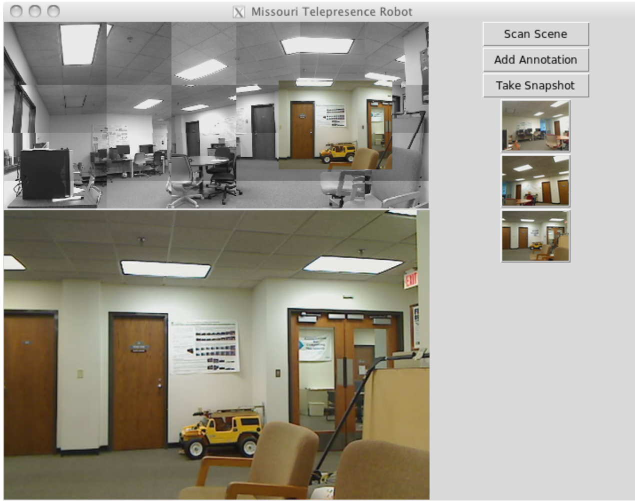

Figure 1: The panorama interface after the discovery phase. Three annotations are shown on the right side of the window.

Additionally, users can control the camera by dragging the live image, specifying the view, allowing the software to decide how to move the camera appropriately.

Several buttons are provided for controlling the interface.

Scan Scene: The *Scan Scene* button directs the camera to scan its full range of motion, taking stills at intervals which are stitched together to create the panorama. This button can be used to initialize the panorama at startup, or reinitialize at any time if the scene has changed significantly, or the robot has moved.

Add Annotation: In the settings that this interface is designed for, it is likely that the scene will stay mostly static; people will be seated in chairs around a table, there will be a fixed whiteboard and projection screen, etc. Many of these fixed locations are likely to be of interest to the human operator, who may wish to quickly switch between them in order to easily follow the flow of conversation. This provides a mechanism for saving a location in the scene and quickly returning to it. Pressing *Add Annotation* adds a new button under the existing control buttons which contains a thumbnail of the current camera image at the current position and zoom level. Pressing this new button causes the camera to rotate to the position it was in when the annotation was added and restore the zoom level.

Take Snapshot: Some situations necessitate taking notes, and in these situations, simply taking a picture of the information source could be more efficient. The *Take Snapshot* saves a full-resolution image of the current camera view to the user's hard drive for off-line viewing.

4 Discussion

This interface provides users of telepresence systems with a visual context of their current environment. Similar interfaces are commercially available as part of larger teleconferencing systems, but these products are expensive, require specialized hardware, and lack features critical for a mobile telepresence robot acting as an avatar for a single operator.

Systems such as Microsoft Roundtable, while providing much of the visual context, lack physical cues which humans use when interacting with each other. Fixed cameras provide no way to estimate the robot's gaze direction, a cue which is vital in human-human interaction.

Because of the need to move the camera to create a panorama during the discovery phase, this interface is not useful in situations when the robot base undergoes translation, but rather is confined to static settings. Many telepresence robots have additional wide-angle cameras mounted lower on the robot for use when driving, providing the necessary peripheral vision using a camera-view better suited for driving.

5 Conclusions & Future Work

A study will be performed to quantify the impact of providing context via peripheral vision for meeting- type situations with a telepresence robot, as well as particular features such as saving annotations to ease saccading to locations of interest. We will also investigate the impact of physically conveying gaze direction. Additionally, we have begun work on characterizing and compensating for the error present in movement of inexpensive pan-tilt cameras, in order to better track the camera pose through time.

Adding a physical cue of *face* in addition to *look at*, as performed by the pan-tilt camera accomplished by adding an action which will cause the robot base to rotate to face the desired point in the scene. Because the movement is pure rotation, the existing panorama can be slid proportional to the angle, maintaining the previous field of view and existing annotations.

6 Thanks

This work was supported in part by NSF award OCI-0753340, and in part by a research gift from Willow Garage, Inc.

7 References

[1] Anybots Inc. Anybots - your personal avatar. `http://anybots.com/#qbLaunch`, August 2010.

[2] R. Cutler, Y. Rui, A. Gupta, J. J. Cadiz, I. Tashev, L. W. He, A. Colburn, Z. Zhang, Z. Liu, and S. Silverberg. Distributed meetings: a meeting capture and broadcasting system. In *Proceedings of the tenth ACM international conference on Multimedia*, MULTIMEDIA '02, pages 503–512, New York, NY, USA, 2002. ACM.

[3] D. A. Lazewatsky and W. D. Smart. An inexpensive robot platform for teleoperation and experimentation. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA 2011)*, 2011.

[4] Logitech. Logitech Orbit AF. `http://www.logitech.com/en-us/webcam-communications/webcams/devices/3480`, 2010.

[5] Microsoft. Microsoft RoundTable. `http://www.microsoft.com/uc/products/roundtable.mspx`, 2010.

[6] Polycom, Inc. Polycom CX5000. `http://www.polycom.com/CX5000`, 2010.

[7] Polycom, Inc. Polycom HDX6000. `http://www.polycom.com/products/telepresence_video/telepresence_solutions/room_telepresence/hdx6000.html`, 2010.

[8] Willow Garage Inc. Texai remote presence system. `http://www.willowgarage.com/pages/texai/overview`, December 2010.