CONSTRAINED IMAGE UNDERSTANDING FOR AN INTERNET ROBOT SUPPORTING TELEPRESENCE

Alexander Ferworn *Wing-Hong Shiu* *Kostas Plataniotis*

The Network-Centric Applied Research Team

School of Comp. Sci. Ryerson Polytechnic University
Comp. & Info. Sci. The University of Guelph
Elec. & Comp. Eng. The University of Toronto

ABSTRACT

The Network-Centric Applied Research Team (N-CART) introduces a distributed model for constrained image recognition for a web-enabled video robot supporting aspects of telepresence.

The recognition engine resides on a network node apart from the actual robot streaming the video. We have found that by employing such a model, the cost and complexity of the robot can be reduced while allowing the recognition services to be reused by other devices.

At the end of the paper, we discuss how we are using the notion of distributed services to deploy video preprocessing for limited capability wireless devices.

1. INTRODUCTION

Broadband is becoming a reality. This fact alone has changed the nature of the Internet. It was a media where virtual collaboration could take place only through the sharing of text messages and other low-bandwidth means. In its turn, the demand for speed has enabled the WWW to become a place challenging Telephone, cable and other providers to keep up with a wide variety of high-bandwidth applications that were inconceivable less than a decade ago--particularly those employing video.

An area of particular interest has been the introduction of Low-cost, limited computing network devices often referred to as Internet Appliances (IAs) [1]. We at N-CART have been experimenting with extending the notion of IAs from mere information tools to the realm of physical interaction. Building on the work of the MAX project [2] we have introduced a form of image understanding to support teleoperation and telepresence over the Internet for a mobile IA.

MAX employs an on-board video camera that streams video JPEGs. Through the use of constrained high-quality images extracted from the video stream, and a remote image recognition server, we are able to identify a common visual artifact, in limited orientations, with reasonable delay, that can provide support for the autonomous performance of relatively complex interaction tasks. In addition such an off-board server can help to support limited resolution IAs as we illustrate in the final section.

2. THE IDENTIFICATION TASK

We have concentrated our efforts on the identification of handicapped access buttons. These buttons are widely deployed and used by wheelchair-restricted individuals to automatically open doors.

Figure 1 Various Handicapped Access Buttons

They have become fixtures in most public and private common areas and passages. Such buttons are the obvious means for providing access to a mobile, teleoperated device such as a robot seeking to traverse the interior spaces of buildings. Unfortunately, finding, maneuvering and finally pressing such a button is a rather difficult task for a device that may not have sufficient onboard resources to deal with the problem. It would be helpful if such a mundane task could be accomplished automatically.

Using the MAX architecture, employing our constrained image understanding technique and a distributed model of control, we describe a solution to this problem that may provide an enhanced telepresence experiences for the users of such devices.

3. THE MAX SERVICES INFRASTRUCTURE

We introduced the MAX server architecture in [3] a modified version of this architecture is shown in figure 2. MAX has been modified to provide an Application Programming Interface (API) to the video and command services employed to stream video and deliver movement commands to the robot.

The Web interface to MAX remains two simple Java Applets [4], one supporting movement commands and the other providing the video stream from the mobile rob[illegible] Through the API however, it has now becom[illegible] provide additional services to the MAX [illegible] illustrated below.

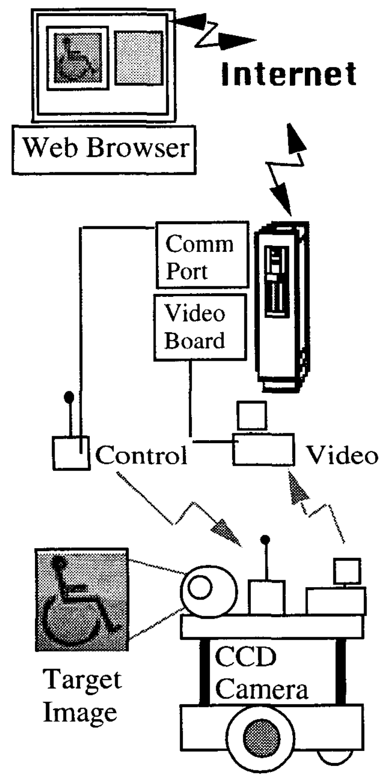

Figure 2 The MAX Services Infrastructure

4. CONSTRAINED RECOGNITION SERVICES

We are in the process of adding a public interface to the MAX services infrastructure. We are currently employing an image minimum error thresholding technique proposed by Kittler and Illingworth [5] and a modified Sobel edge detection technique described in [6] to process raw video images extracted from MAX's video stream to spot door buttons.

We are relying on an implied user model to create our constraint.

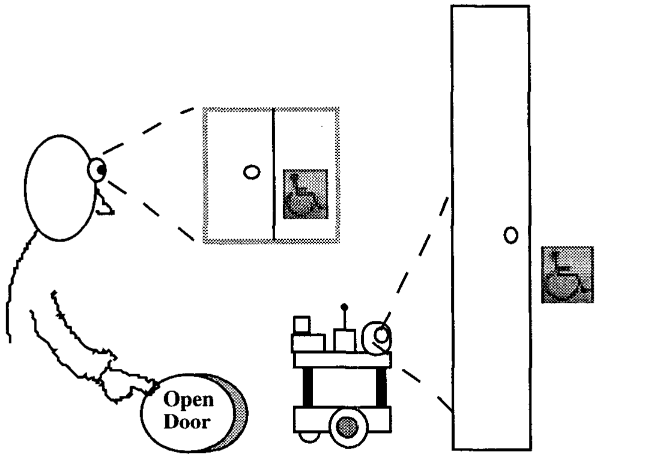

Figure 3 Implied User Model

We have made an assumption that the MAX teleoperator will send a command to MAX to open a door (by pressing a screen button) only when there is a door present. With this assumption, we have a assumed the heuristic that there will also be an access button in the image.

Through numerous test shots taken around access paths from the approximate height of a device similar to MAX (approximately 0.5 M high) we have grown confident that this is in fact the case.

We also rely on the distinct shape of the whee[illegible] and figure within the button to derive heigh[illegible] telemetry information from the image once the[illegible] been isolated.

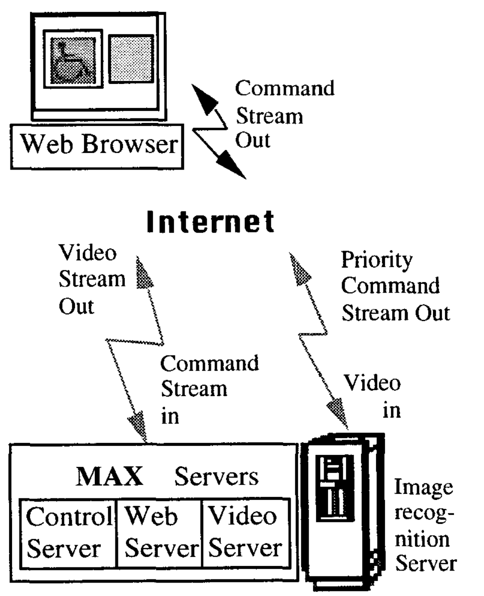

Figure 4 Modified MAX Services Infrastructure

Once the user requests a door be opened, the MAX controller locks the control server queue--effectively locking out the current user from issuing additional commands. The recognition server than performs the necessary recognition task and either issues raw commands to the control server or returns a "no button found" message to the user interface. Once the recognition task is accomplished control is returned to the user. This interaction is illustrated in the figure below.

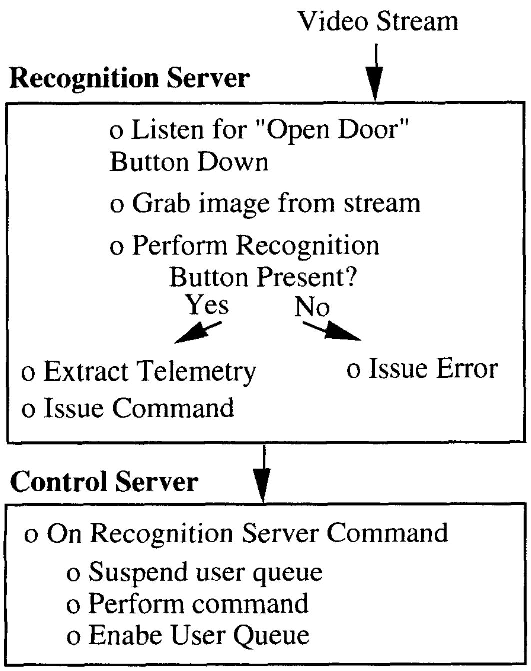

Figure 5 Recognition-Control Algorithm

The commands issued by the recognition server are intended to bring the robot into optimal striking distance of the button. At the moment the button is pressed by collision with it.

The robotic component engaged in a recognition task can be seen in the diagram below.

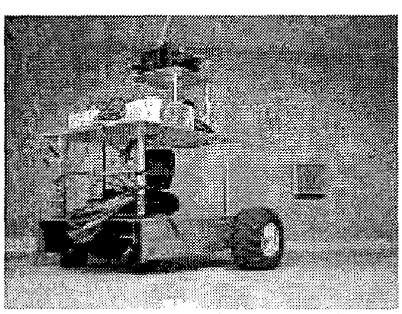
Figure 6 MAX Engaged in a Recognition Task

As a caveat, this is not a foolproof technique nor are there any miracles. Non-existent buttons cannot be found, illumination levels can cause image "ghosts", and various physical constraints could prevent a button from being accurately identified. However, given the nature of the task, a failure to identify an access button results merely in an error condition--presumably causing the user to move the robot and try again.

5. PRELIMINARY RESULTS

The figure below illustrates the process used to perform the recognition task. The top image is that of a poorly illuminated handicapped access button on an interior brick wall. This image has been manually extracted from an image from the video stream for illustration purposes.

The second image is the thresholded button starting the elimination of unwanted artifacts within the image and enhancing the contrasting graphic on the button.

The third image is created using the Sobel edge detection technique. The graphic has been isolated sufficiently to make a decision concerning button presence within the image.

Knowing the height of the camera and employing other image processing techniques, telemetry information can be extracted from the image and appropriate commands issued.

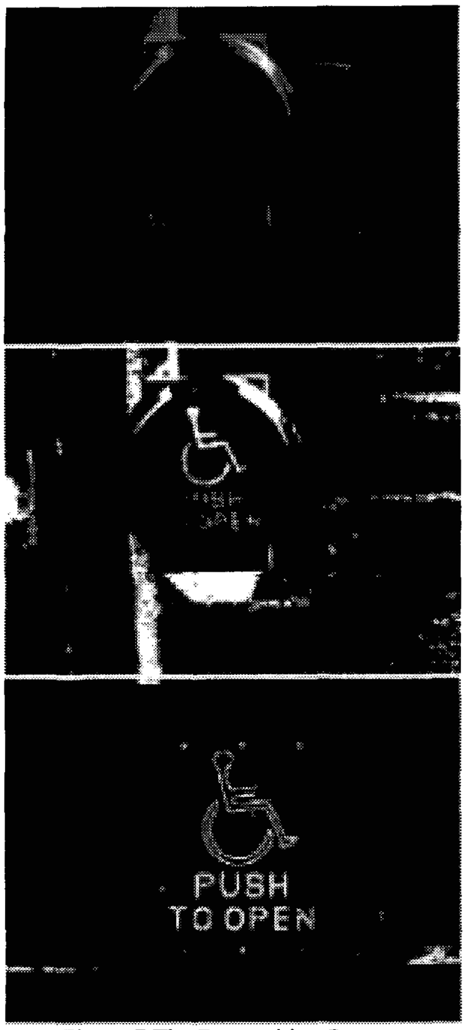

Figure 7 The Recognition Process

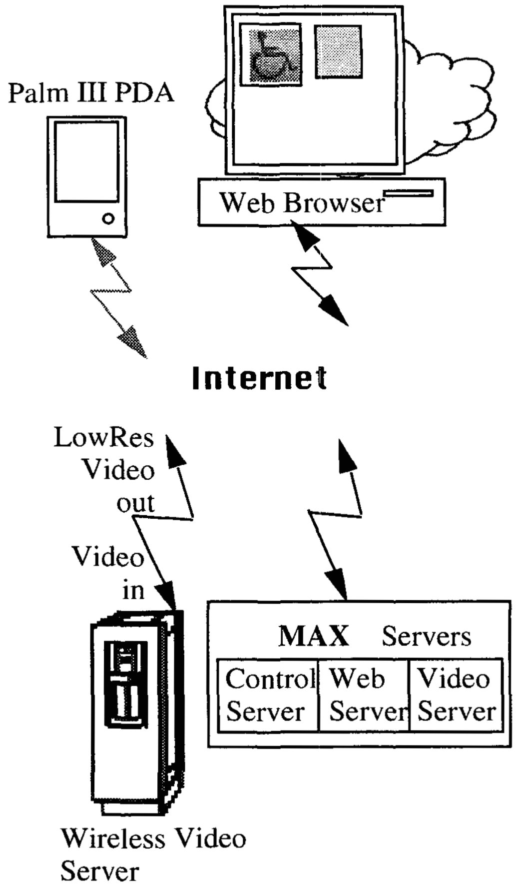

Figure 8 Wireless Video Services

6. LIMITED BANDWIDTH VIDEO SERVICES

Having enjoyed some success in the image recognition task we are using the same system architecture techniques to support far less capable IA devices.

We have been experimenting with Palm Computing Palm III Personal Digital Assistants (PDA) [7] employing a wireless modems to deliver an extremely low bandwidth video stream to the device.

We are currently developing a wireless interface to MAX employing a prototype video server generating low-resolution video from the current MAX video stream. Preliminary results indicated that we can support up to five frames a second employing existing wireless technology.

The figure below is taken from the low-resolution video stream and serves to illustrate the facility of the technique. From the image one can clearly see a handicapped access button in the upper left of the emulator screen.

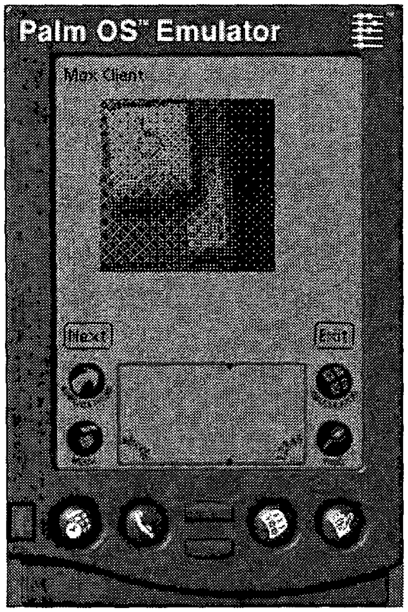

Figure 9 MAX Video on a PDA

Eventually we hope to provide full control capabilities to a PDA for MAX.

11. REFERENCES

[1] Skrzypczak, Milosz, "Consumer Market Convergence" ,The Yankee Group Report Vol. 17, No. 14, September 2000.

[2] A. Ferworn, R. Roque, and I. Vecchia, "MAX: Wireless Teleoperation via the World Wide Web", Proc. Of the 1999 IEEE Canadian Conference on Electrical and Computer Engineering, Edmonton, Alberta, Canada, May 9-12 1999

[3] A. Ferworn, K. Plataniotis, "Effective Teleoperation Over the World Wide Web", Proc. of the IASTED International Conference for Robotics and Applications (RA'99), Santa Barbara, USA, October 28-30, 1999

[4] Control MAX at max.scs.ryerson.ca.

[5] J. Kittler and J. Illingworth, Minimum error thresholding, Pattern Recognition, 19, pp 41-47, 1986.

[6] R. Gonzalez and R. Woods Digital Image Processing, Addison Wesley, 1992.

[7] See www.palm.com