CONSTRAINED IMAGE UNDERSTANDING USING LOSSY COMPRESSED IMAGES

ALEXANDER FERWORN *WING-HONG SHIU* *WEI LU*

The Network-Centric Applied Research Team

School of Comp. Sci.
Ryerson Polytechnic University
350 Victoria St, Toronto, ON,
Canada, M5B2K3

Comp. & Info. Sci.
The University of Guelph
350 Victoria St, Toronto, ON,
Canada, M5B2K3

School of Comp. Sci.
Ryerson Polytechnic University
350 Victoria St, Toronto, ON,
Canada, M5B2K3

ABSTRACT

The Network-Centric Applied Research Team (N-CART) introduces a distributed model for constrained image recognition using lossy compressed images.

The recognition engine resides on a network node apart from where the images are taken. We have found that by employing such a model, the cost and complexity of the system can be reduced while also allowing the flexibility of the recognition services to be used by different devices.

KEY WORDS:

Video Understanding, Mobile Robots, Teleoperation

1. INTRODUCTION

With the cost of broadband Internet connection becoming more and more accessible, a new crop of devices are starting to emerge on the market making use of it. These devices, often referred to as Internet Appliances (IA) [1], are low cost network devices that do simple tasks like providing information or offering email services. We at N-CART, are experimenting with different ways of extending the tasks that can be accomplished by these devices from merely providing information to being able to remotely perform physical tasks. By extending the MAX project [2], we have introduced a form of constrained image understanding using lossy compressed binary images as a reference to our regular image understanding techniques.

MAX is a tele-operated device, which employs an on-board video camera that streams video JPEGs. Through the use of constrained high-quality images extracted from the video stream and with the use of a remote image recognition server, we are able to identify common visual artifacts in limited orientations and with reasonable delay. This provides support for the autonomous performance of relatively complex interactive tasks. We introduce the notion of lossy compressed binary images for image understanding to the existing image understanding system as an intermediate step for quick reference skipping the telemetry extraction step.

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. By identifying and accessing these buttons, MAX would have a way of gaining physical access to areas, which were inaccessible previously.

Figure 1. Various Handicapped Access Buttons

Since these access buttons are common fixtures in most public and private common areas, they make perfect candidates for the target of identification. 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.

By using the existing MAX architecture and adding the constrained image understanding technique with the compressed binary image understanding technique, we feel this is one way of easing the difficulty of using such devices and enhancing the use of such a tele-presence device over the internet.

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 robot. Through the API however, it has now become possible to provide additional services to the MAX infrastructure as illustrated below.

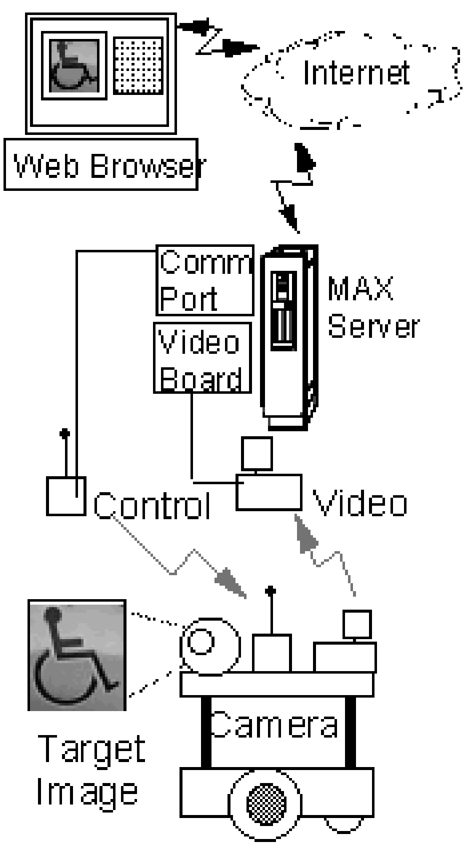

Figure 2. The MAX Services Infrastructure

4. LIMITED BANDWIDTH VIDEO SERVICES

We also have been experimenting with extending MAX control and video services to wireless client devices. We have been using a Palm III Personal Digital Assistants (PDA) [5] employing a wireless modem to deliver an extremely low bandwidth video stream to the device and an even lower bandwidth command stream.

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 per second employing a nominal 14.4 kilobits per second data stream.

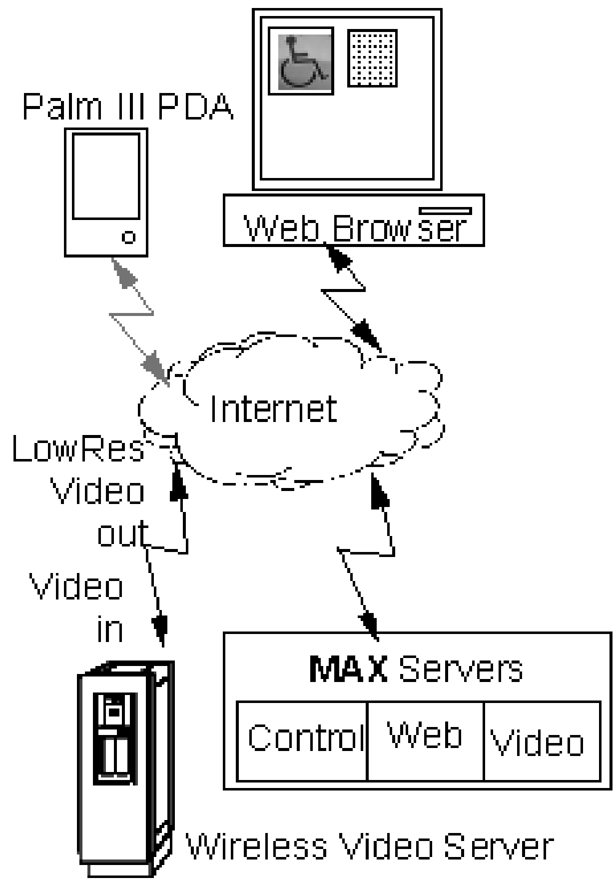

Figure 3. Wireless Video Services

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

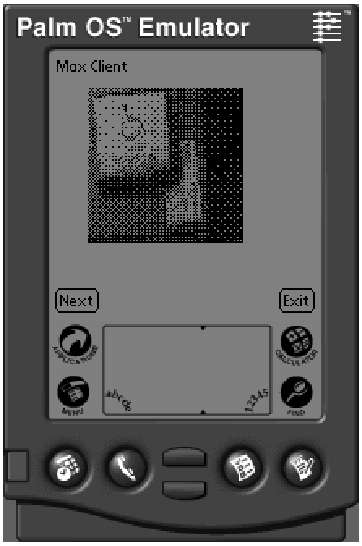

Figure 4. MAX Video on a PDA

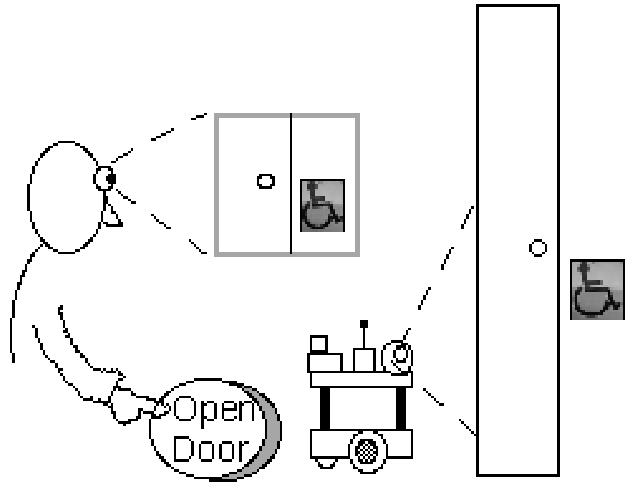

Figure 5. Implied User Model

5. CONSTRAINED RECOGNITION SERVICES

The original constrained image understanding process [6] involves the use of a minimum error thresholding technique proposed by Kittler and Illingworth [7] and a modified Sobel edge detection technique as described in [8] to process raw video images extracted from MAX's video stream to spot the door buttons. The original implied user model assumption to create our constraint still remains. The assumptions were for MAX's tele-operator to send a command to MAX to open a door, by pressing a button on the command screen if there is a door present. We also assume that there will be a handicapped door access button present in the image as well. We also relied on the button being placed in accordance with Ontario building codes which states that "a handicapped button must be placed at a height that is accessible by a handicapped person at a height of approximately 0.5 M high" [9]. Finally, we assume the distinct shape of the wheel chair figure to be present in all of the buttons at a derived height and angle. Telemetry will be derived from the image once the button has been isolated.

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. In the revised scheme of the recognition service, a compressed binary image would be obtained from the PDA image server of MAX to see whether MAX is currently directly in front of the handicapped button. If the button is present and is recognized by the recognition task in the server, a command is sent back to MAX to initiate a process to press the button. If a button is not detected the process defaults back to the original recognition process to once again analyze the image and see if that would improve the likelihood of detecting the handicapped button if present.

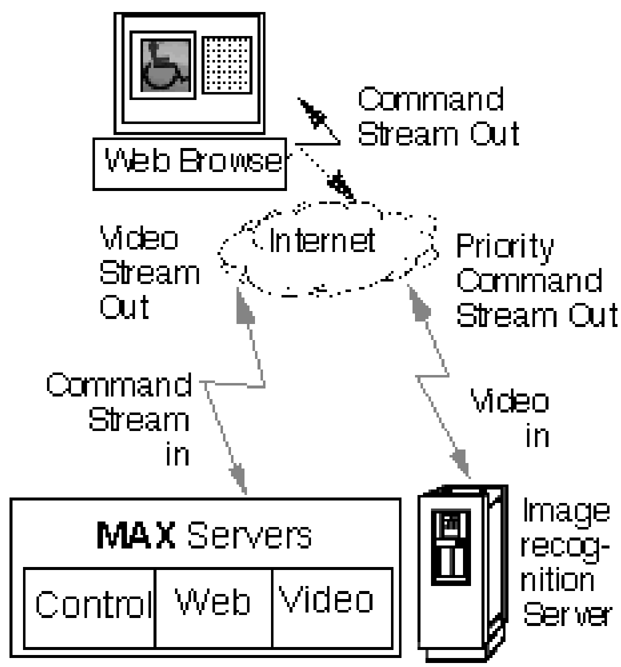

Figure 6. Modified MAX Services Infrastructure

The reason for the addition of this new process being introduced is to provide a faster, more abstracted way of detecting the button on subsequent tries by the user. When the user first initializes the open detection process and the system does not detect the button, we assume the user would try again after the control of MAX is returned to them. On subsequent tries, it is assumed that the user would have moved MAX closer to the button. Using a compressed binary image from the PDA server, there would be less data to transmit to the recognition server and therefore less data for the recognition task to process. The reduced amount of data from the image and the elimination of calculating the telemetry of the button means that this is a faster processing means for MAX to assess whether the button is present or not before processing it in more in detail.

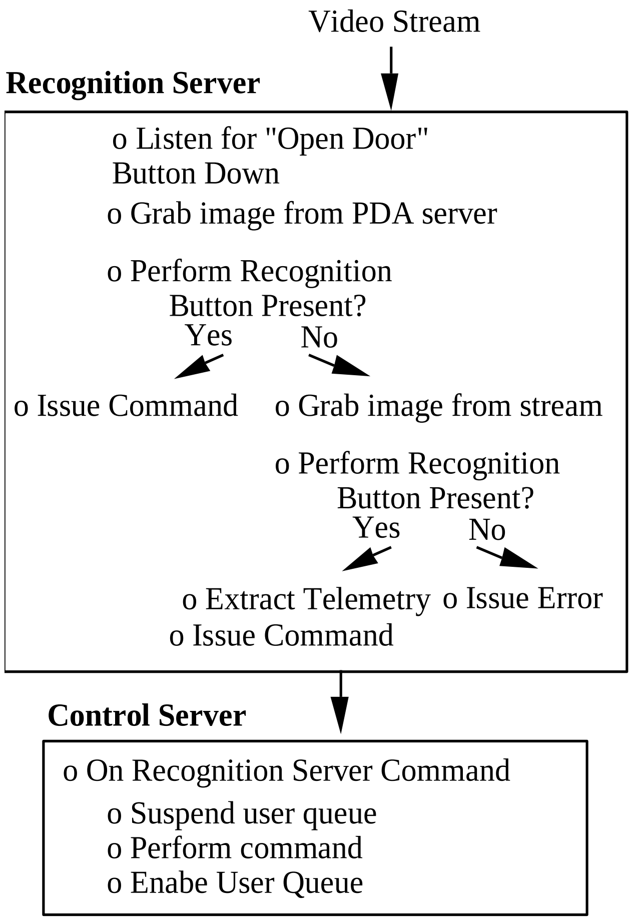

Figure 7. Recognition Control Algorithm

As stated earlier, this is not a foolproof way of detecting the button so there are scenarios where MAX could be located directly in front of the button and still not be able to detect it. This could be because the illumination in the area is poor or the image that was transmitted is corrupted by signal noise in the area. Since a failure in detecting the button would simply return the control of MAX back to the user to try again, the system will fail safely.

Figure 8. MAX Engaged in a Recognition Task

6. Preliminary Results

We are in the early stages of our experimentation. The main problem that we have encountered thus far is the quality of the image acquired from MAX's camera dramatically impacts the compressed binary image that is presented by the PDA server. The top image shows the original image that is captured from the video stream of MAX, as would be seen by the user.

The second image is of the compressed binary image that is sent by the PDA server. As can be seen, the handicapped symbol is barely discernible after the image is compressed. This is most likely due to the fact that the camera that was originally producing the image from MAX gave such a blurry image that most of the details in the image were washed out after the compression.

In the near future, the camera on MAX will be upgraded to a Panasonic GP-CX161-53P CCD camera. We have done some preliminary testing with the new camera and the results look promising. The new camera is able to automatically adjust for lighting conditions better than the old camera--providing a much crisper and easily processed image as can be seen in Fig. 9 (old camera) and Fig. 10 (new camera).

Figure 9. The Recognition Process

Figure 10. Image from MAX using Panasonic GP- CX161-53P camera

7. CONCLUSION

As described, this process of rapid assessment of the viewed object using lossy compressed image can be very useful in helping verify the object before telemetry of the object is calculated. But one of the limitations of this process is that it requires the captured image to be of relatively high quality and also in fair lighting. With the use of higher quality camera and better lighting incorporated into the Max project in the future, we hope we can solve these short comes and make the system successful for this application.

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 Tele-operation 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] See www.palm.com

[6] A. Ferworn, Wing-Hong Shiu and Kostas Plataniotis , *"Constrained Image Understanding for an Internet Robot Supporting Tele-presence"*, Proc. Of the 2001 IEEE Canadian Conference on Electrical and Computer Engineering, Toronto, Ontario, Canada, May 13-16 2001

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

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

[9] Province of Ontario building code, 1999 -