2005 IEEE/RSJ International Conference on Intelligent Robots and Systems

Pano-Presence for Teleoperation

Mark Fiala
Computational Video Group
National Research Council
Ottawa, Ontario, K1A 0R6, Canada
mark.fiala@nrc-cnrc.gc.ca

***Abstract*— Telepresence and *teleoperation* are immersive viewing and control by a user from a remote location. Usual implementations use a standard narrow field of view (FOV) camera and a communications link, and a head mounted display (HMD).**

Teleoperating a robotic vehicle or surveying a scene with such system and a computer monitor is difficult for human operators due to the narrow FOV of standard cameras, the unintuitive interface for directing the camera, and the loss of directional sense. For this reason these systems often use a head mounted display (HMD) instead of a monitor, however this introduces the *HMD pose latency problem* of latency and slow update due to the mechanical motion of the camera and the communications link. Even with good equipment, the experience is disorienting and slow. This paper proposes a *pano-presense* architecture for telepresence for applications such as teleoperation based on panoramic cameras, a communications link, and an HMD. The panoramic camera, capable of capturing light from all azimuth directions, provides a panorama which is be transported over the communications link to a *panorama frame buffer* for viewing in the HMD screen(s). The panorama viewing rate is decoupled from the communications latency so the user can look around freely without experiencing *HMD pose latency problem*, the delay in the HMD's image alignment with the head position. A panorama frame format of an *image cube* is chosen since it can be viewed at full frame rate with the acceleration in consumer graphics cards. Two prototype systems, one telepresence and one teleoperation, using this architecture are described.

***Index Terms*— Teleoperation, Telepresence, Panoramic, Omnidirectional, HMD.**

I. Introduction

Telepresence is the paradigm of a human user at one location receiving stimuli with the assistance of technology as if they were at another. Video and audio is captured at some location and transported in electronic form to a remote location where the user can look around and hear, useful in situations where it is inconvenient or dangerous to be in person such as a security guard surveying a large area, children supervision, military situations, hazardous materials handling, space robotics, underground mining, etc. Telepresence, more than just a closed circuit TV link, attempts to fool the person's senses so they can feel they are present and look around in a natural way. This is achieved today with *head mounted displays* (HMD's) and *cave* environments. The user can look around and see views consistent with that viewing direction at the other location.

Teleoperation is where a user can interact with the remote environment by controlling motors and actuators, together with telepresence is a popular concept in science fiction but has not made it yet to common use due to human factors issues related to the video latency and the HMD's field of view. The *HMD pose latency problem*, when the HMD view image is not from the same pose as the head position, is disorienting and degrades the immersive experience.

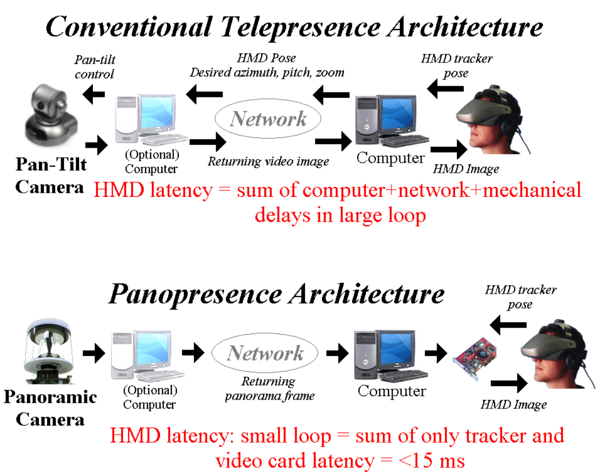

Fig. 1. Using the *panorama frame buffer* paradigm to remove communication latency in the HMD/virtual view loop. With conventional telepresence (top), the HMD view angle is communicated through the network, the image is captured and sent back producing an large delay. With the *HMD view generator* generating a view rapidly from the locally stored *panorama frame buffer*, the latency is removed (bottom).

The author proposes that the largest problem with practical telepresence today is the latency and bandwidth of the video imagery, and that HMD quality (mostly field of view) and price will improve if systems using them are shown to be feasible. The latency issue is partly due to the mechanical response of the narrow FOV camera mount, but mostly due to the communications medium. When a person wearing an HMD looks in a direction, they expect to instantly see the view from a camera with the same pose as the HMD. Today's communication technology, despite its great achievements, does not provide sufficient capabilities to provide a low latency full video stream over equipment available in homes and offices (Section I-A), especially when the teleoperated vehicle or robot is untethered and a wireless link is necessary.

Our solution involves replacing the narrow FOV camera with a panoramic camera, and transmitting *panorama frames* instead of standard image frames. Using a panoramic camera removes the mechanical latency issue and the *HMD view generator* + *panorama frame buffer* architecture eliminates the HMD pose latency problem (Fig. 1).

This *pano-presence* system framework can provide usable telepresence with inexpensive COTS (common off the shelf) equipment and standard computers. Two working systems are demonstrated; a telepresence system with a wired camera, and a teleoperated robotic vehicle.

A. Bandwidth Issues

Raw uncompressed video requires about 80 Mbps (bps=bits/second), MPEG compressed video can achieve 1.5 Mbps for 320x240 imagery (MPEG I) or 4.5 Mbps for 640x480 (MPEG II) but is not always of high quality when compressed in realtime. Wireless digital links are often less than 100 Kbps or less, the public cellular telephone network provide 4.8 kps (analog) or 19.2 kps (CDPD) to modems [12], [11]. COTS video RF links can provide full rate analog video usually over short distances, however only for conventional NTSC cameras, and a pan-tilt platform is necessary providing mechanical latency. Even over wired networks, most network connections provide 10-100 Mbits/second, however with computers and network infrastructure in between, users usually get connections in the order of 150 kbps. One may think that these are minor obstacles to overcome if one has a budget, such as the military, space exploration, or large industry. However high reliability and secure data links end up having similar low data rates, the use of spread spectrum, encryption and other technologies are necessary to ensure communication when there are many users or an enemy attempting to jam an RF signal. Especially for military users such bandwidth limiting measures are necessary, imagine the "inconvenience" soldiers would face if their armed teleoperated vehicle became controlled by the enemy. Even neglecting the issues of a rapid moving camera pan-tilt-roll platform, video transmission using the paradigm of a standard narrow FOV camera and and HMD is not feasible for today.

Even with necessary bandwidth, a communications network will invariably introduce delays. Mania [9], researching virtual reality with NASA reports users notice delays between HMD and view pose as lows as 15 milliseconds. With this as a minimum latency, even the software and operating system within a PC can fail to provide fast enough HMD view generation.

B. Panoramic Image Capture and Viewing

This work is partly motivated by the *Navire* project [1] which seeks to develop virtual navigation of real environments captured with a panoramic camera.

Warping a section of a panoramic image to provide a virtual perspective view is an alternative to a mechanical pan-tilt camera with a standard narrow FOV (less than 50°). Recent work has seen the use of non-perspective image projections for use in capturing images with a wider field of view [7], [2]. A panoramic image captures part of the total light impinging on a 3D point, more than seen with a standard camera. Practical constraints restrict the capture of a full unbroken spherical view, but sensors exist to capture all azimuth angles with an elevation range sufficient for many telepresence and teleoperation applications.

Omni-directional sensors can be built using a mixture of mirrors and lenses in the optical path, so called *catadioptric* cameras that can simultaneously capture light from a wider field of view than a conventional *dioptric* camera consisting of just a lens (or lenses) and a flat image plane. A panoramic camera, one that captures light from 360° along one axis, can be built using the combination of a standard dioptric camera and a rounded mirror. Basu [4], [3], and others [8], [10] have demonstrated such systems.

As shown in Fig. 2, a vertically posed dioptric camera focused on a radially symmetrical mirror can capture light in all azimuth directions, with a viewing range above and below the horizontal horizon plane. Mobile robotics and teleoperated vehicles is one field that can benefit from such imagery presenting a continuous, simultaneous view from all azimuth directions [5], [6].

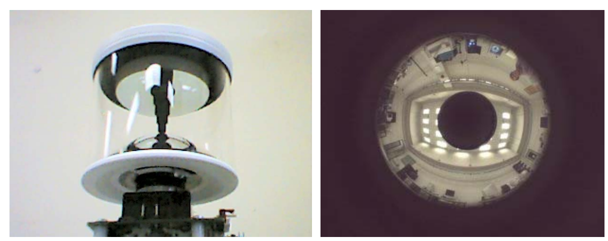

Fig. 2. *A Panoramic Imaging system using Catadioptric optics and a sample panoramic image. The catadioptric camera is created by mounting a* NetVision Assembly B *mirror/lens unit onto a Vitana 1280x1024 IEEE-1394 digital video camera.*

A panoramic image captures part of the total light impinging on a 3D point, a virtual narrow FOV perspective view matching the viewpoint of an HMD screen can be generated from this view (Fig. 3).

II. Proposed Telepresence Paradigm

This paper proposes a simple concept based on a *panorama frame buffer* paradigm where the HMD view update is decoupled from the communications medium by separating the transmission from the display as shown in Fig 1. The view for the HMD is produced by the *HMD view generator*, pixels are warped from a *panorama frame* without waiting for the network. The communications system updates a complete panorama frame at whatever speed the bandwidth enables, and when a whole new frame has arrived the panorama frame buffer that the HMD view generator renders from is switched. This solves one problem, that of the communications latency.

A standard computer (example: PC, laptop) will introduce a latency also and so the design of the HMD view generator is important. Graphics acceleration hardware is present in all modern computers and so this rendering can be done

Fig. 3. *Immersive viewing with an HMD (top left) of a section of a panorama generated from a panoramic camera image (top right). The user sees a perspective projection model image (bottom).*

at the full video speed of the HMD, bringing the latency down to 16ms or less. For this reason the format for the panorama frame is the *image cube* described in Section III. The panorama is represented by a six-sided cube where the sides are interpreted as textures by the rendering engine.

The image cube sides are compressed with standard methods (JPEG, motion-JPEG, MPEG, etc) to reduce the data size for each panorama. Either the full panorama can be transmitted, or a partial panorama containing the directions the user is expected to look is sent. Based on the current HMD view direction and angular velocity, the sub-panorama that the user could possibly see before the next panorama frame is updated can be estimated to further save on transmitted data.

III. Cube format

A panoramic image collects all or part of the light incident on a point in space. People typically think of such a data set as a spherical image, however this does not lend itself to efficient storage and handling. If a six-sided cube format is used instead virtual perspective images can be more readily handled. The cost of an increased storage space (nearly doubled, $\frac{6}{\pi} = 1.9$) over a spherical representation is offset by the benefits of fast rendering with standard graphics hardware and ease of compression and decompression.

The view seen in the HMD screen is a perspective view that can see up to three cube sides at once, the view is rendered with simple texture mapping. Since the cube side images were created from a reprojection of a captured panoramic image or synthetic image, the user is unaware of the joints between cube sides. Fig. 4 shows an HMD view with and without an overlaid grid to visualize the cube sides.

IV. Prototype System

Both a telepresence and teleoperation system was created to test the usability of the proposed panopresence paradigm. The *HMD view generator* implemented as a Windows program creates a view seen on an I-Glasses VGA HMD with a 23° field of view. It connected via TCP-IP to one of two other computers; one connected to a stationary camera on a tripod, and the other to a remote controlled robot platform with a panoramic camera.

The HMD has a resolution of 640x480 pixels and a maximum frame rate of 60 fps. An Intertrax II orientation sensor ([1] provides the HMD pose.

The stationary camera is a stationary Pixelink PL-A360 IEEE-1394 [2] fitted with a Remote Reality NetVision Assembly B panoramic lens/mirror assembly [3] (Fig. 2). It captures a color image of 1280x1024 pixels of which an annular region of 800 pixels diameter contains the panoramic image. Colour cube images of width 500 pixels on a side were sent over a LAN, the cube images were sent uncompressed at 2 Hz, however the HMD screen is updated at 20 fps. Sample images from this are shown in Figs. 3,4

The teleoperation prototype is a panoramic camera mounted on a mobile platform (Figs. 6,7). The platform's motor controls and the panoramic video are wireless allowing it to roam untethered. An NTSC camera was used instead of a higher resolution IEEE-1394 camera such that an inexpensive RF video transmitter could be used [4]. A greyscale NTSC camera was mounted to a Remote Reality NetVision Assembly A panoramic lens/mirror assembly, providing a similar field of view as the stationary camera; an elevation range of 0° to 54° up from the horizontal. Cube images of width 400 pixels are sent across the network at 4 Hz and observed with the HMD from the remote *HMD view generator* software which updates the HMD at 20 fps.

A. Quality of Immersive Experience

Despite the low resolution and poor image quality of the panorama, when users look around they feel comfortable and do not consider the low resolution to overly degrade the experience. When the HMD pose latency problem is addressed, the feeling of realism is quite convincing.

The HMD view generator was configured for the same FOV as the HMD as it was expected this would be necessary to achieve immersive realism. However, showing a wider FOV provides more information and did not seem to affect the realism. Due to the small (23° horizontal) FOV, there was a feeling of "tunnel vision". The user was able to switch

[1] http://www.intersense.com/
[2] http://www.pixelink.com/
[3] http://www.remotereality.com/
[4] www.x10.com

between the correct (23°) and wider (46°) FOV, and all nine users surveyed preferred the wider field of view.

It seems that solving the HMD pose latency problem is more important to achieving realism than having a rendered field of view that correctly matches the HMD, an unexpected result.

The prototype introduced is the first step, but still provide good results. All system components are currently being improved. The current prototype HMD view generator software is being improved to take full advantage of the hardware acceleration of the graphic card and the next version should run at the full frame rate of the HMD. The system will be tried with a ProView(tm) XL50 HMD [5] which has a wider FOV, 40° compared to (23°) for the I-Glasses, and a higher image resolution of 1024x768 pixels. Most importantly to the user experience, we are integrating a new panoramic camera, the Point Grey Ladybug (Section IV-B).

B. Panoramic Camera Hardware and Panorama Quality

Three different omnidirectional cameras were tried in our system. They are all *catadioptric* sensors (contain both lenses and mirrors in the optical path) created by replacing the lens in a digital video camera with a commercial mirror/lens. Two of these were used in the telepresence and teleoperation test of Section IV. The components and their useful image parameters are shown in Figure 8. The best image of the three systems was created using a Pixelink digital video colour camera (http://www.pixelink.com/)) fitted with a Remote Reality NetVision Assembly B panoramic lens/mirror assembly (http://www.remotereality.com/) (Figure 1.). It captures a color image of 1280x1024 pixels of which an annular region of 800 pixels diameter contains the panoramic image. The unused space is due to this model of lens/mirror assembly being designed for both $\frac{1}{2}$ and $\frac{1}{3}$ inch CCD's.

A system of similar pixel resolution was created using a colour 1024x768 pixel dragonfly IEEE 1394 camera, however the image noise was higher giving a poorer subjective perception. Finally an NTSC camera was used which gives the least high quality image, but allows for use with an RF video link with the teleoperated robot.

With the Pixelink and Dragonfly IEEE cameras, providing a useful annular image of diameter 800 pixels, the perspective warp has an equivalent pixel density to a 320x240 image. With the narrow field of view (23° horizontal FOV) of our I-glasses HMD, the low resolution of the panorama is evident. It is the opinion of this author that catadioptric cameras are not available, at the time of this writing, which can provide a level of quality that a consumer system would demand. We are currently integrating a "Ladybug" multi-CCD IEEE 1394 camera from Point Grey Research which provides a nearly seamless hemispherical view with 6 separate cameras enclosed in a small package [6].

[5]http://www.rockwellcollins.com/keo/proviewxl3550.htm
[6]http://www.ptgrey.com/

V. Conclusions

A *Pano-presence* paradigm for *telepresence* and *teleoperation* systems was introduced where *panorama frames* are transmitted instead of standard image frames.

The *HMD pose latency problem*, when the HMD view image is not from the same pose as the head position, is disorienting and degrades the immersive experience. Due to communication delays in practical systems, the system must generate new views corresponding to the HMD pose without waiting for the view to come over the network if the HMD pose latency problem is to be solved. Removing the pose latency appears to be crucial to creating practical telepresence and teleoperation systems that people will tolerate. Subjective experiments showed people did not mind an incorrectly matched field of view, perhaps a consistent HMD/virtual-camera pose with low latency is the most important to achieve immersive realism.

The goal of the pano-presence system framework to provide usable telepresence with inexpensive COTS (common off the shelf) equipment and standard computers was achieved and demonstrated with a total equipment budget of under $4000 (not including computers). Two working system configurations were created; a telepresence system with a wired camera, and a teleoperated robotic vehicle.

References

[1] http://www.site.uottawa.ca/research/viva/projects/ibr.

[2] S. Baker and S. Nayar. A theory of catadioptric image formation. In *IEEE ICCC Conference*, pages 392–197, 1998.

[3] A. Basu and J. Baldwin. *A Real-Time Panoramic Stereo Imaging System and its Applications*. Springer, 2001.

[4] A. Basu and D. Southwell. Omni-directional sensors for pipe inspection. In *IEEE SMC Conference*, pages 3107–3112, Vancouver, Canada, October 1995.

[5] M. Fiala. Linear markers for robot navigation with panoramic vision. In *Proc. of CRV'04 (Canadian Confernence on Computer and Robot Vision*, pages 145–154, May 2004.

[6] M. Fiala and A. Basu. Robot navigation using panoramic landmark tracking. In *Proc. of Vision Interface*, pages 117–124, May 2002.

[7] C. Geyer and K. Daniilidis. Catadioptric projective geometry. In *International Journal of Computer Vision*, volume 43, pages 223–243, 2001.

[8] M. Yamamoto H. Ishiguro and S. Tsuji. Omni-directional stereo for making global map. In *3rd Intl. Conf. Computer Vision*, pages 540–547, 1990.

[9] K. Mania. Perceptual sensitivity to head tracking latency in virtual environments with varying degrees of scene complexity. In *Proceedings of the 1st Symposium on Applied perception in graphics and visualization*, pages 39–47, Los Angeles, USA, 2004.

[10] S. Nayar and T. Boult. Omnidirectional vision systems: Pi report. In *Proceedings of the 1997 DARPA Image Understanding Workshop*, May 1997.

[11] J. Reed. *Software Radio, A Modern Approach to Radio Engineering*. Prentice-Hall, Inc, Upper Saddle River, NJ, 2002.

[12] A. S. Tanenbaum. *Computer networks*. Prentice-Hall, Inc, Upper Saddle River, NJ, 1988.

Fig. 4. Graphics card is actually rendering a 6-sided cube, however the user does not notice (top). Overlaid grid lines show the cube sides (bottom).

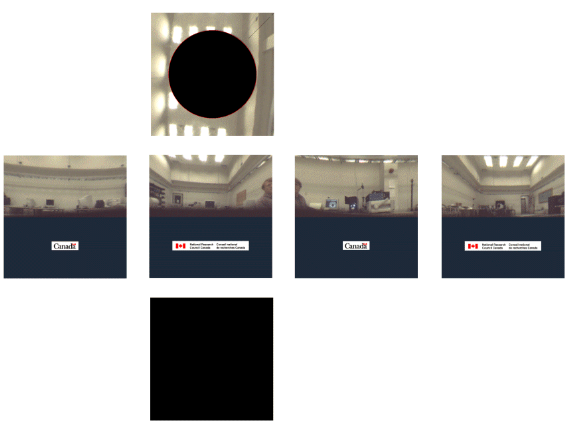

Fig. 5. Cube image *format for panorama storage and handling. Panorama is stored as a standardized 6-sided cube. Top image shows visualization (4 sides only), bottom image shows 6 sides.*

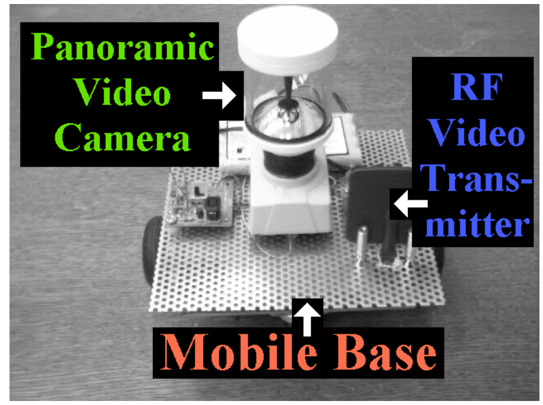

Fig. 6. Teleoperated robot with panoramic camera. Greyscale panoramic image is transmitted back to a host PC with an RF video link. Base motors are controlled by wireless from the host PC's serial port.

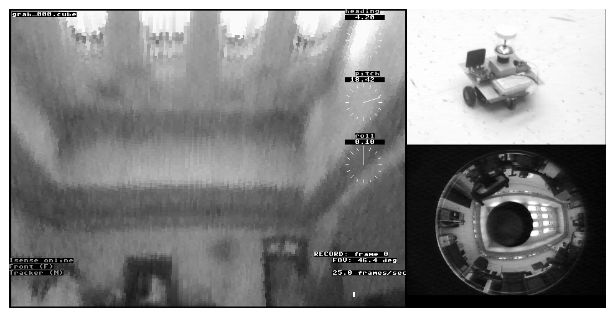

Fig. 7. Teleoperated robot with panoramic camera. Greyscale panoramic image is transmitted back to a host PC with an RF video link. Base motors are controlled wirelessly from the host PC's serial port.

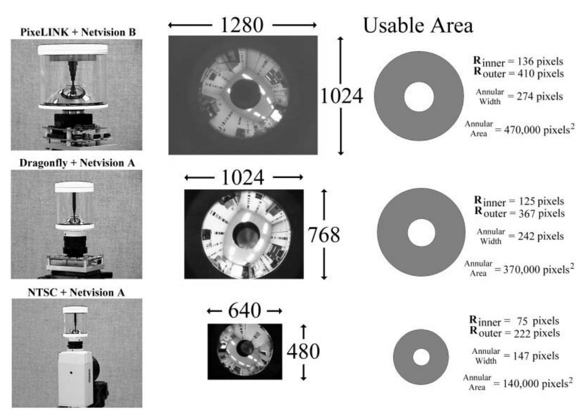

Fig. 8. Resolution with three panoramic camera configurations.