HEAD-COUPLED REMOTE STEREOSCOPIC CAMERA SYSTEM FOR TELEPRESENCE APPLICATIONS

M.T. Bolas
Fake Space Labs
Palo Alto, California 94306

S.S. Fisher
Aerospace Human Factors Research Division
NASA Ames Research Center, Moffett Field, California 94035

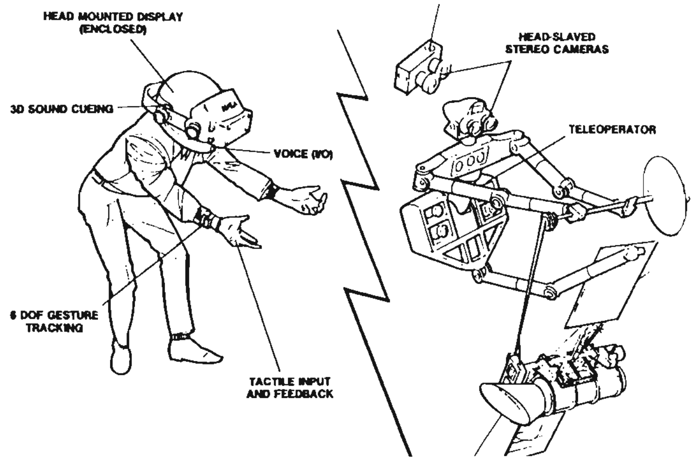
ABSTRACT

The Virtual Environment Workstation Project (VIEW) at NASA's Ames Research Center has developed a remotely controlled stereoscopic camera system that can be used for telepresence research and as a tool to develop and evaluate configurations for head-coupled visual systems associated with space station telerobots and remote manipulation robotic arms. The prototype camera system consists of two lightweight CCD video cameras mounted on a computer controlled platform that provides real-time pan, tilt, and roll control of the camera system in coordination with head position transmitted from the user.

This paper provides an overall system description focused on the design and implementation of the camera and platform hardware configuration and the development of control software. Results of preliminary performance evaluations are reported with emphasis on engineering and mechanical design issues and discussion of related psychophysiological effects and objectives.

1. INTRODUCTION

Although the long term goal of telerobotic systems is virtually autonomous operation, it is generally recognized that human operators will continue to monitor and supervise the systems under normal operations and to intervene under degraded modes. Control of these autonomous and semi-autonomous telerobotic devices and vehicles will require an interface configuration that allows variable modes of operator interaction ranging from high-level, supervisory control of multiple independent systems to highly interactive, kinesthetic coupling between operator and remote system. For remote operations that cannot be performed autonomously, the interface will need capability to quickly switch to interactive control. In this telepresence mode, the operator will require sufficient quantity and quality of sensory feedback to approximate actual presence at the remote task site.



Telepresence with the Virtual Environment Workstation
FIGURE 1

Head-Coupled Remote Stereoscopic Camera System Shown with CCSV
FIGURE 2

Conceptual versions of telepresence have been described by science fiction writers for many decades. Arthur Clarke has described 'personalized television safaris', in which the operator could virtually explore remote environments without danger or discomfort. Robert Heinlein's "waldoes" were similar, but were able to exaggerate certain sensory and dexterous capabilities so that the operator could, for example, control a huge robot. Since 1950, technology has gradually been developing to make telepresence a reality. One of the first attempts at developing a telepresence visual system was done by the Philco Corporation in 1958. With this system an operator could see an image from a remote camera on a CRT mounted on his head in front of his eyes and could control the camera's viewpoint by moving his head [Comeau, 1961]. Further work in anthropomorphic remote viewing systems was done by Argonne National Laboratory in 1965 for use in the nuclear industry [Goertz,1965]. Telepresence research has continued at other laboratories such as NASA Ames in California, the Naval Ocean Systems Center in Hawaii [Pepper, 1984], and MITI's Tele-existence Project in Japan [Tachi, 1985]. Here the driving application is the need to develop improved systems for humans to operate safely and effectively in hazardous environments such as undersea or outerspace.

In the Aerospace Human Factors Research Division of NASA's Ames Research Center, an interactive Virtual Interface Environment Workstation (VIEW) has been developed as a new kind of media-based display and control environment. This system provides a virtual auditory and stereoscopic image surround that is responsive to inputs from the operator's position, voice, and gestures. As a low-cost, multipurpose simulation device, this variable interface configuration allows an operator to virtually explore a 360-degree synthesized or remotely sensed environment and viscerally interact with its components [Fisher, 1986,1988].

A key component of this research has been the development of a remotely controlled stereoscopic camera system that can be used for telepresence research and as a tool to develop and evaluate configurations for head-coupled visual systems associated with space station telerobots and remote manipulation robotic arms. The prototype camera system consists of two compact, lightweight video cameras mounted on a computer controlled platform that provides real-time pan, tilt and roll control in coordination with head position transmitted from the user. This system will also be used for experimental derivation and validation of design specifications for other head-coupled imaging systems [Figures 1,2].

2. SYSTEM DESCRIPTION

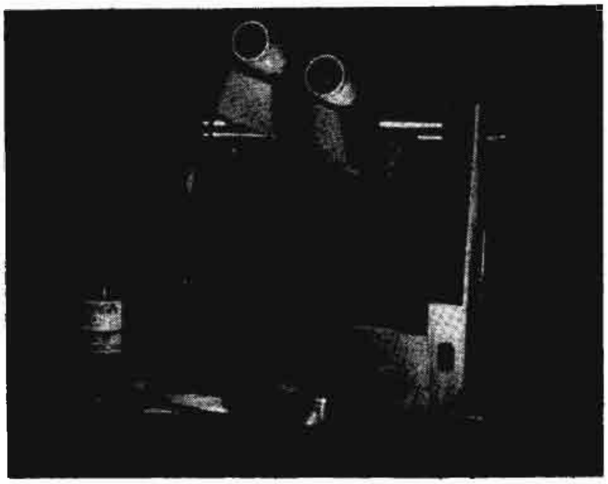
2.1 Cameras and Optics

Over the past three years, a survey of video cameras has been conducted to select an appropriate sensor for the prototype remote camera system. A basic requirement was that the system use CCD chip technology and would output NTSC standard video. A primary consideration was to obtain a minimum of image 'smear' during rapid motion of the sensor. Cameras were also compared on resolution, size, weight, cost, power requirements, sync requirements, and sensitivity. The current system uses two low-cost Sony model XC-38 CCD B/W video cameras. Resolution of this unit is 280(h)x350(v) TV lines from a 384x491 pixel, interline transfer CCD sensor. Weight of each unit is 155 grams including c-mount adaptor, IR cut filter and tripod mount but not including the lens. Each unit draws 2.9W of power and can accept external synch input to ensure proper stereoscopic image presentation. Illumination sensitivity is from about 3 to 400 lux. Use of IR cut filters further reduces minimal image smear.

A specific objective in the development of this camera system has been to match media and computational technology as closely as possible to the perceptual and cognitive capabilities of the human operator in order to achieve a state of telepresence at a remote site. One of the factors that directly influences the achievement of this objective is the ability to create 'orthostereoscopic' displays: the presentation of stereoscopic images that are geometrically correct representations of the depth relationships and perspective of a rendered or captured scene [Kurtz, 1937]. For example, in an orthostereoscopic scene, an object viewed through the remote camera system would subtend the same angle of visual field of view as the object it represents in the real world.

In order to present a geometrically correct stereo image, a unique set of camera and viewing lenses is used to provide unity magnification of the imaged scene. Developed by Eric Howlett, the viewer optics provide a 120 degree field of view to each eye with a 90 degree binocular field overlap when used with a suitable image source [Zientara, 1984]. The objective is to fill the user's field of view as much as possible and to provide an orthostereoscopic image. These lenses are used in both the VIEW lab's helmet-mounted display and a counterbalanced CRT-based stereoscopic viewer (CCSV). The optics introduce two types of distortion: chromatic aberrations along edges of high contrast, and a pincushion radial distortion. To compensate for these distortions, a set of custom lenses were made for the CCD cameras that introduce equal and opposite distortions. The result is an undistorted, wide-angle, life-size image.

The two camera units are attached to a simple metal plate with optical parallel optical. Interocular separation can be adjusted from about 1.5 inches to about 8 inches but typically is set at about 2.5 inches to match average human interocular separation. Set screw adjustments are used to fine tune vertical image alignment.



Head-Coupled Remote Stereoscopic Camera System
FIGURE 3

2.2 Mechanical Hardware

An objective of this telepresence camera system is to couple the remote sensor package in close coordination with the user's head motion. The requirement here is for the displayed scene to also impart correct motion parallax and motion perspective cues in relation to the virtual surroundings as the viewer changes position. This requires a mechanical package that can respond in real-time to orientation and velocity information transmitted from the remote user and that can approximate typical head trajectories. This three degree of freedom mechanical structure is also designed to avoid several problems typically encountered in robotic platforms that incorporate multiple degrees of freedom. Limited travel, cable tangling, pan-tilt-roll order independence, and inertia concerns are all considered as critical design issues [Figure 3].

The camera platform can rotate beyond 360 degrees in all three axes without physical limits imposed by its structure. By using a differential gear cluster, the tilt and roll axes share a symmetric drive arrangement. With this configuration, the roll motor can be placed on the pan axis, shifting mass to the pan motor and away from the tilt motor. In effect, this arrangement achieves a third degree of freedom without decreasing the payload capacity of the system or increasing the cable routing complexity. Accordingly, limits due to cable interference were eliminated by designing cable-ways which run through the center of each axis of rotation. This effectively allows the structure to rotate around the cables. The resulting kinematics of this platform design are independent of pan-tilt-roll ordering. Furthermore, they roughly approximate the translations which occur during a head rotation. This is outlined by Figures 10 and 11 and discussed in section 5.

The camera platform is driven by three DC permanent magnet brush motors. While stepper, brushless, AC, and pancake motors were considered as drives, standard brush style motors were chosen. Although each type of motor offers an array of advantages and disadvantages, the DC motors provide the fine level of motion control required and are readily available, reasonably priced, lightweight, and small. They also offer the ability to be controlled by a number of standard electronic motion control cards.

Because these motors are designed to run at high rotational velocities, they require the use of a speed reduction stage. While a geared reducer could be used, timing belts coupled with different sized pulleys offer a cost effective and acoustically quiet speed reduction solution. Belts also enable the motors to be located off of the axis of rotation, an important practical consideration when planning cable routing and mechanical layout. This layout flexibility allows the payload size of the platform to be changed simply by shifting 2 plates and fitting different sized belts. By choosing fiber reinforced belts with a curvilinear tooth form, typical design concerns of backlash, slip, and stretch were reduced to insignificant amounts.

2.3 Electronic Hardware

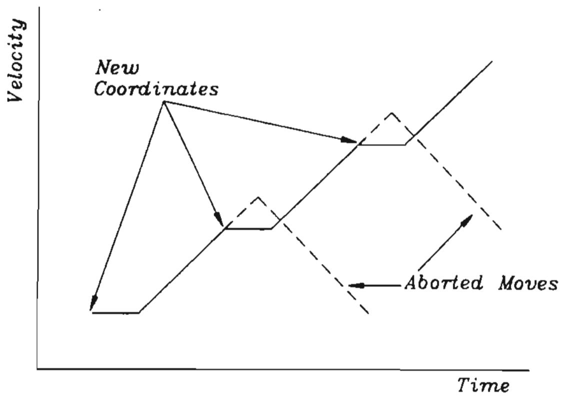
The motors that drive the camera platform are controlled by a set of commercially available motion control cards. These cards are used as standard proportional-integral-differential motion controllers and are synchronized in order to allow for coordinated motion between all three axes. Optical shaft encoders with 1000 line resolution provide positional feedback to the control cards. Although no tachometer is used in the system, the control cards incorporate a 'pseudo' tachometer by measuring the time between counts. This greatly increases the 'stiffness' of the system at low speeds. Switching motor drive amplifiers were used because they offered a small and inexpensive drive solution. However, inductive filters needed to be placed on their output, and all cables were carefully shielded in order to reduce video signal noise.

Although the motion control cards are operated currently as simple point to point controllers, they have the capability for cubic spline path matching. This feature will be used for future experimentation with systems constrained by limited head motion data bandwidth. In the current configuration, the cards are programmed to accept orientation coordinates from the host computer as ASCII strings sent over a serial data line. The cards receive the new coordinates and, as a background process, calculate the motion control parameters required to perform coordinated moves to the new locations.

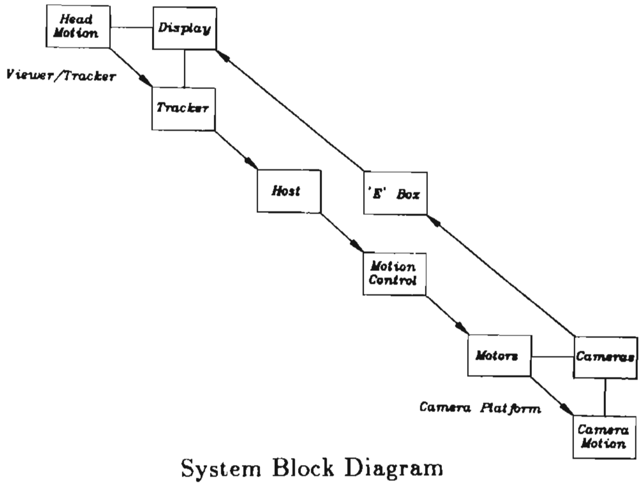
This background processing requires 12 milliseconds during which the platform motion that is currently in progress continues with its current velocity, and is modified to reflect the new coordinates when all calculations are complete. As long as new coordinates are received before the current move nears its target, motion continues smoothly from point to point without any need for handshaking with the host. This proves to be valuable when interfacing the camera platform to systems in which it is difficult to read single bits from I/O ports, as found, for example, in common UNIX-based mini-computers. In addition to the processing delay mentioned above, a delay of one sample period is necessarily introduced due to the 'on the fly' control scheme used here. In other words, the motion control cards must wait to see where they are going, before they can go there [Figure 4].

The camera platform system uses the RS232 serial data standard as the primary means of communication between its components. A major reason for choosing this standard is a longer term requirement for long distance data communications that can be met more easily with serial rather than parallel communication lines. Use of RS232 serial also allows the system to act as a flexible test bed for different display, tracker, and host combinations. In the past year, four different configurations have already been evaluated.

A disadvantage of serial data, however, is system performance degradation due to the time delay it introduces. For example, it takes 22 milliseconds for a block of coordinates to be transferred between the host computer and the motion control cards along a 19200 Baud serial line. The sum total of this and other lag producing elements quickly accumulates to significant amounts.



Command Timing Between Host and Motion Controllers
FIGURE 4



System Block Diagram
FIGURE 5

3. HEAD-COUPLED DISPLAY CONFIGURATIONS AND INTERACTION

3.1 Viewer Configurations

The camera platform is presently configured to be operated by two different interactive display systems under development in the VIEW Lab [Figure 5]. One system couples the platform to a LCD-based, stereoscopic, head-mounted viewer in which head orientation is sensed by an electromagnetic tracking device mounted on the headset [Fisher,1986,1988]. This information is processed and relayed to the camera platform by an HP9000/Model 835 host computer. When the user puts on the display headset, the NTSC video signals from the remote cameras are visible through the wide angle, stereoscopic viewer optics. As the user changes head orientation, the camera moves in correspondence at the remote location [Figure 6].

A second display system consists of a counterbalanced CRT-based stereoscopic viewer (CCSV) that uses transducers at joints of its support linkages to calculate position and orientation of the viewer package. This information is processed and relayed to the camera platform by an IBM AT host computer [McDowall,1990]. In this configuration, the user views the remote scene through a higher resolution, stereoscopic CRT display. In contrast to the head-mounted viewer, viewpoint control is achieved here by grasping handles attached to the suspended viewer package to direct pan, tilt, and roll of the remote camera in real-time. This interaction is similar to using a pair of binoculars that provide a movable, wide-angle window into the remotely viewed space [Figure 7].

3.2 Camera Platform Control Software

When specifying orientation in the form of Euler angles (pan, tilt, and roll) a problematic singularity occurs when the head is oriented to look straight up or down. At these orientations, the roll axis is tilted (because of rotation along the tilt axis) to lie directly in line with the pan axis. In this particular configuration there are an infinite number of ways to specify the orientation of the user's head in terms of Euler angles.

To avoid this issue when using the VIEW lab's head-mounted viewer to interact with computer-generated virtual environments, the electromagnetic tracking system is usually configured to report orientation in the form of a quaternion. Quaternions offer a method to uniquely describe any orientation, and to move in a direct fashion between any two orientations, without having to face these Euler angle problems [Shoemake, 1985]. Using quaternions as a rotational coordinate system has been very successful in the VIEW lab and is incorporated directly into the rotation matrices used in the graphics algorithms of the VIEW lab system software.

However, in order to control the remote camera platform, these quaternions must be mapped into coordinates which match those of the gimbal platform mechanism. Specifically, this is done by first converting the quaternion to pan, tilt, and roll angles, and then reducing the tilt and roll components into angles appropriate for the differential mechanism on the camera gimbal platform. Because the camera gimbal platform is inherently a pan, tilt, and roll device, the software must limit its motion to avoid looking straight up or down for the reasons mentioned above. This is to say that, although a mapping can be made from quaternions to Euler angles, the use of quaternions can not change the kinematic limitations of the camera gimbal mechanism.

An additional problem occurs when using the HP9000 host computer and the electromagnetic tracking system to record the user's head orientation: system performance is I/O bound. Although the HP9000 has ample processor bandwidth to handle the quaternion computations and data formatting requirements of the camera platform, it suffers from the difficulties inherent in performing real-time character I/O on large UNIX-based systems. This presently limits the system frame rate to about 15Hz. This is expected to improve with the installation of special purpose HP hardware to accelerate I/O activity.

In contrast, when interfacing to the CCSV display system, physically correspondent joints on its linkage structure can be directly mapped to the camera platform's gimbal style coordinates. This eases the burden on the host computer in two ways. First, the singularity mentioned above is avoided because the CCSV structure provides a non-ambiguous set of coordinates - the physical joints cannot be two places at once. Second, because the CCSV reports pan, tilt, and roll angles directly, the host computer needs to perform only a few linear transformations before passing the data to the motion control cards.

To handle these relatively minor computational tasks, an IBM AT is a sufficient host computer. In contrast to the I/O limited performance of the HP9000 system mentioned earlier, the CCSV/IBM AT system is computationally bound to a frame rate of about 44Hz. This higher frame rate noticeably increases the responsiveness of the system.

4. SYSTEM PERFORMANCE ISSUES

4.1 System Following Error

An important measure of system performance is the total system following error. This measure represents the difference between orientation of the user's head and orientation of the camera platform while both are in motion.

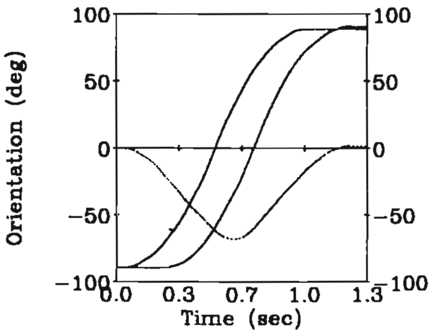
To quantify following error for the camera platform system, a motorized test platform is used that can rotate a viewer/tracker through a prescribed rotation while the camera gimbal system attempts to match this rotation. The motion of the motorized test platform, the data recorded by the head mounted viewer's tracker, and the motion of the camera gimbal are recorded during the move.

By comparing the final location of the motorized test platform with tracker data recorded during the move (not shown on these plots) it is possible to determine that the static overshoot seen on Figures 8(a) and 9(a) is due to tracker orientation errors. For the electromagnetic tracker, the error was about 1 degree in 180. For the CCSV orientation transducers, the error was 8 degrees in 180. The magnetic system error is well within the manufacturer's specifications, and the CCSV error has been corrected by adjusting a gain constant in its software.

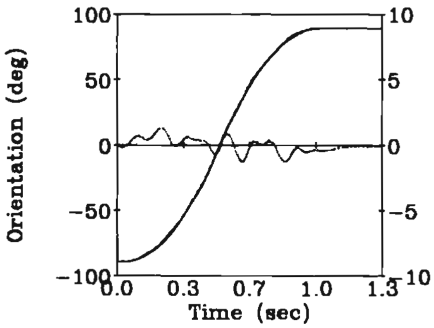
4.2 System Lag

After accounting for this static offset, a new plot was made in which the camera platform motion was shifted back in time until the error signal reached a minimum. In this way, the nature and amount of lag can be investigated. Figures 8(b) and 9(b) illustrate this time-shifted plot and the resulting new error signal. This error is expected due to the finite frame rate and the control system performance. Consequently, it appears that the apparent lag is not velocity or time dependent, but instead is a result of system transport delay. For the head-mounted display combined with the HP9000 host, this delay is 260 milliseconds. For the CCSV display system combined with an IBM AT host, the delay is 140 milliseconds. For comparison purposes, recent evaluation of an F-16C flight simulator indicated an average total system transport delay of 135 milliseconds [Horowitz,1987].

Head-Coupled Remote Stereoscopic Camera System
Shown with Head-Mounted Viewer
FIGURE 6

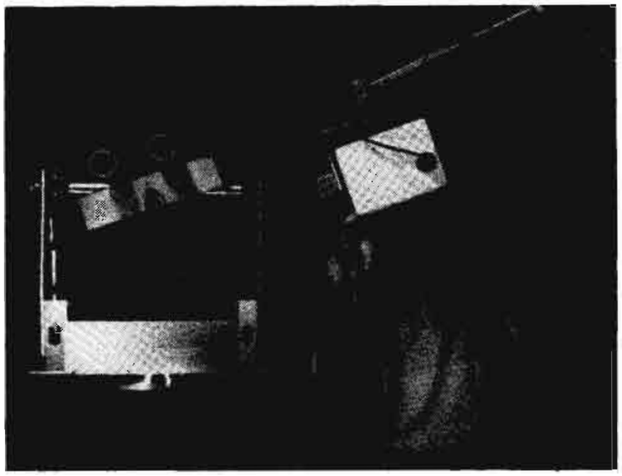


(a) Raw Data

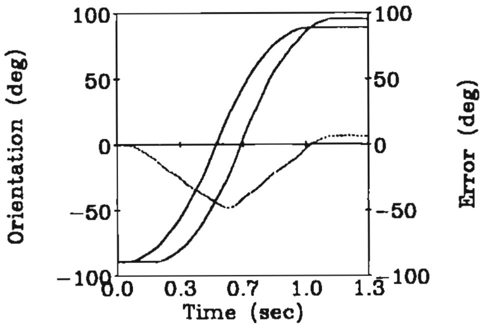


(b) Gain Corrected and Time Shifted Data

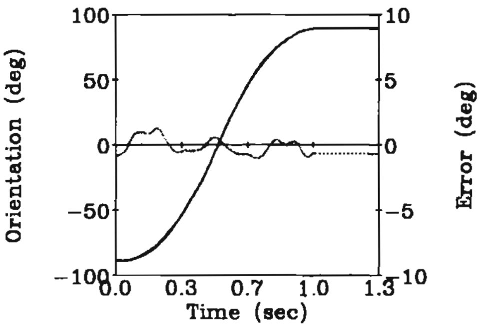
Camera Platform Response - Pan Axis
HP9000/Head-Mounted Display System
FIGURE 8



Head-Coupled Remote Stereoscopic Camera System
Shown with CCSV
FIGURE 7



(a) Raw Data



(b) Gain Corrected and Time Shifted Data

Camera Platform Response - Pan Axis
CCSV/IBM AT System
FIGURE 9

Understanding and reducing this delay is of critical importance for improving the remote camera platform system performance, and appears to be an extremely important factor in meeting the perceptual requirements for achieving a sense of telepresence with the camera platform. Recently, additional research has been initiated in the VIEW lab to identify and evaluate these and other causes of lag not mentioned in this paper [Bryson, 1990].

4.3 System Frame Rate

An important distinction should be made regarding frame rate and lag. Frame rate refers to the number of coordinates sent to the motion control cards per second, whereas lag refers to the time it takes for the camera platform to respond and match head motion. Frame rate is linked to lag in that the frame rate determines the time between successive moves as described previously.

More directly linked to the frame rate is the smoothness of the platform movement. Frame rates below 15 Hz will cause jerky platform motion and unacceptable following error. If the head accelerates at 1000 degrees per second, and the time between samples is 660 milliseconds (15Hz), then an error of 1.8 degrees can be generated due to the granularity of the data.

5. MATCHING HUMAN PERFORMANCE

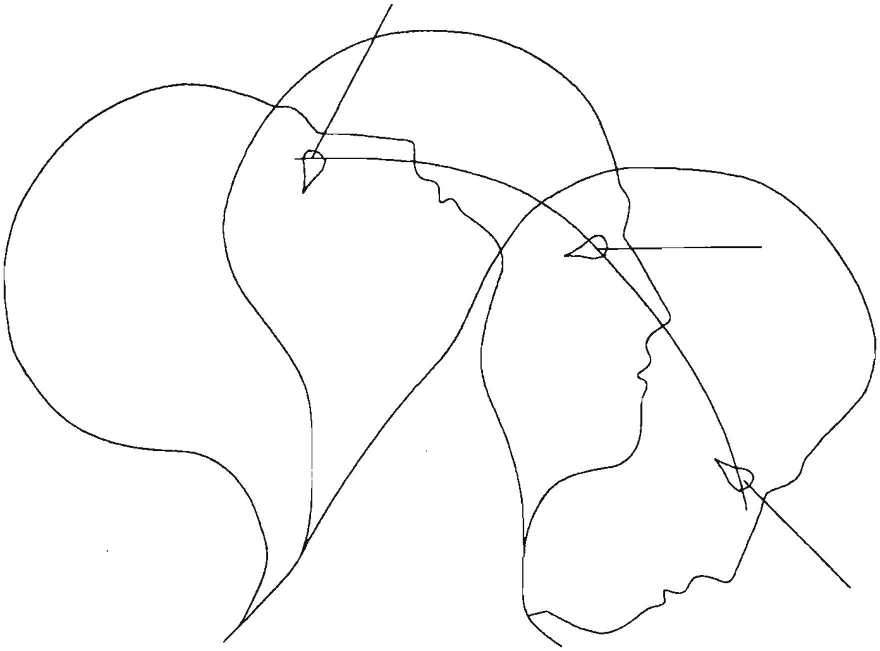
5.1 Roll Axis Requirement

The original design specification for the remote camera platform required only pan and tilt motions. But before making a final design commitment, a simulated camera platform was developed in the Virtual Environment Workstation system. In the simulation, the user visually explores a life-size, computer-generated version of an equipment-filled laboratory. Roll motion is subtracted from the user's head motions while they look around. Without roll, when a user tilts his head sideways, the entire horizon line rolls with him. This effect, coupled with any lag in the system, is quite uncomfortable to view. Qualitatively, users report a harder time locating objects off to their sides and experience an increased amount of queasiness. It appears that when trying to locate objects 'over the shoulder', the subtraction of head roll incorrectly offsets the objects in space from a user's point of view. As a result of this informal testing, roll capability was added to the prototype remote camera platform requirements. In the current viewer/platform configurations, simple software changes can be made to continue more formal roll requirement evaluation.

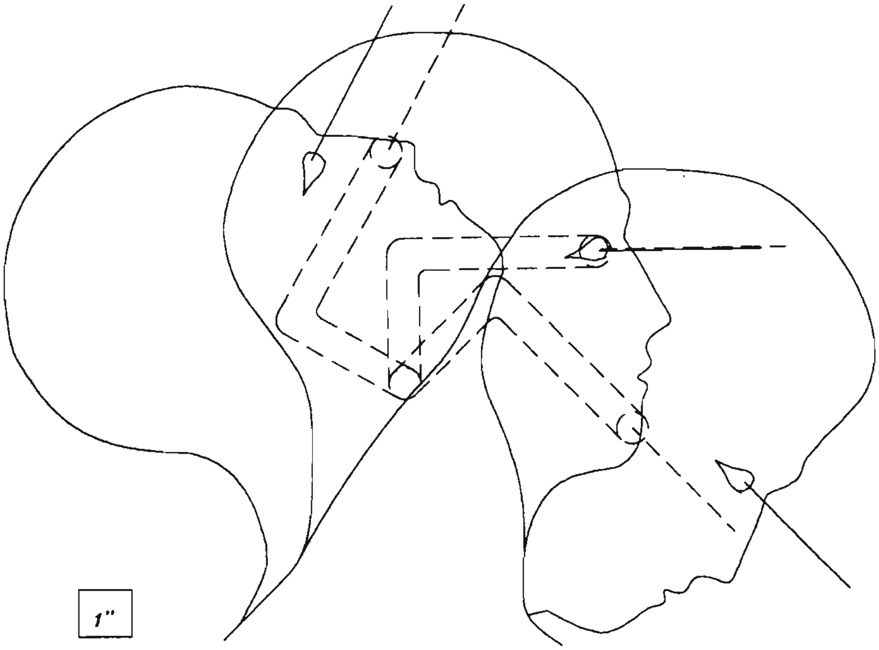
5.2 Matching Typical Head Motions

Even with the addition of roll, however, the camera platform can only approximate human neck rotation and translation. With only three degrees of freedom in the current unit, a small amount of translation is possible by offsetting the camera pair from the pan and roll axis of rotation. Figures 10 and 11 show how this provides a rough approximation of translation corresponding to neck motion. The platform is designed so that the degree of offset can be easily adjusted. This feature will be used to further study the most appropriate location of the camera's sensing element and optical entrance pupil to better match the offset of the eyes above and in front of the pan and tilt axes of the neck.

With regard to matching speed of head motion, current research literature reports a wide range of typical velocities for various activities. In a study done on visual search activity by automobile drivers, maximum velocities for head motion were found to be about 450 deg/sec [Robinson, 1972]. In another study, maximum velocities up to 565 deg/sec were found with accelerations up to 5100 deg/sec [Zangmeister, 1981]. Typical head motions reported in industrial time and motion literature average about 200 deg/sec [Quick, 1962]. Again for comparison purposes, performance specifications for a recently implemented head-slaved projection system for simulation applications were set at 300 deg/sec with acceleration rates of 5,000 deg/sec/sec [Allen,1987].



Head Tilt Motion
FIGURE 10



Head and Canera Platform Tilt Motion Comparison
FIGURE 11

For the rotation data shown in Figures 8 and 9, the platform system was driven to match a turn of 180 degrees in 1 second while following a constant acceleration profile of 1000 deg/sec/sec. The camera platform system can attain rotational velocities in excess of 720 degrees per second with the same acceleration profile and acceptable following errors. In practice, however, it appears that the physical form and mass of the viewing/tracking devices limits users to speeds less than the system's maximum.

5.3 Matching Other Visual System Requirements

Conveying a compelling sense of presence in a remote environment will require many additional display capabilities beyond those provided in this prototype system. Here, the initial effort has been focused on presentation of a geometrically correct stereoscopic image that is spatially and temporally correspondent to typical head orientation dynamics. Additional developments are needed in both the quantity and quality of imagery displayed.

Further improvements toward this objective will require sensors and displays with much greater resolution and luminance to approach real world viewing conditions. This system should have a greater than million pixel color display and sensor resolution that strategically covers a 180-degree field of view. It should also have a capability to present and vary depth of field information to match changes in accommodation and convergence.

6. RESEARCH DIRECTIONS

As planned, the current implementation of the remote camera system will be used to conduct preliminary experimentation on basic requirements for head-coupled remote vision systems. Initial investigations will further examine speed requirements for the camera servo system and will evaluate the effective contribution of the roll axis capability. Other studies will attempt to define interocular and convergence requirements for the stereo camera configuration.

Near term technology development of this remote camera system will include efforts to reduce the size of the package by using miniature color CCD cameras attached to a scaled down replica of the current camera platform. This will also include a motorized camera mount to allow remote control of convergence and interocular separation of the cameras. Currently, the remote camera platform can rotate through 360 degrees in all axes. To expand this capability, the entire unit will be attached to a translation platform to approximate head motion coupled with full body motion. The system will also be remotely located from the user over a long distance video/data link.

7. ACKNOWLEDGMENTS

The authors would like to acknowledge Ian McDowall and Christian Wiedmann for their efforts in the integration and implementation of this system. Also, thanks to the other members of the VIEW lab staff for their support and facilitation of this project, and to Eric Howlett for providing a key element towards the realization of virtual presence.

8. REFERENCES

Bryson,S., S.S.Fisher, "Defining, Modeling, and Measuring system Lag in Virtual Environments," Stereoscopic Displays and Applications, John Merritt and Scott S. Fisher, Editors,,Proc. SPIE 1256, (1990).

Allen,D., B.Tsou, G.Gieske, J.Bien, M.Shipley, J.L.Walker, " System Performance of a Servo-Optical Projections System (SOPS)," Proc. IMAGE Conference IV, (1987).

Comeau,C., J.Bryan, "Headsight Television System Provides Remote Surveillance," Electronics, pp.86-90: Nov.10, (1961).

Fisher,S.S., M.McGreevy, J.Humphries, W.Robinett, 'Virtual Environment Display System', ACM 1986 Workshop on 3D Interactive Graphics, Chapel Hill, North Carolina. October 23-24, (1986).

Fisher,S.S, E.M.Wenzel, C.Coler, M.McGreevy 'Virtual Interface Environment Workstations',Proceedings of the Human Factors Society 32nd Annual Meeting (October 24-28, 1988, Anaheim, California).

Goertz,R., C.Potts, D.Mingesz, J.Lindberg," An Experimental Head-Controlled TV System to Provide Viewing for a Manipulator Operator," Proceedings of the 13th Conference on Remote Systems Technology, pp. 57-60, (1965).

Horowitz, S.J. "Measurements and Effects of Transport Delays in a State-of-the-Art F-16C Flight Simulator," Air Force Human Resources Lab, Brooks AFB. TLSP:Final Report, #AFHRL-TP-87-11, (1987).

Kurtz, H.F., "Orthostereoscopy," Journal of the Optical Society of America, 27(10), 323-339, (1937).

Pepper,R.L. and J.D.Hightower, "Research Issues in Teleoperator Systems," Proceedings of the Human Factors Society 28th Annual Meeting (October 22-26, 1984, San Antonio, Texas).

Quick,J.H, J.H.Duncan, J.A.Malcolm Jr. Work-Factor Time Standards, McGraw-Hill, New York, (1962).

Robinson, G.H. "Visual Search by Automobile Drivers," Human Factors, 14(4), R315-323, (1972).

Shoemake, Ken, (1985) 'Animating Rotation with Quaternion Curves.' Computer Graphics,Proceedings of ACM SIGGRAPH '85, vol.19, no.3: (1985).

Tachi, S., H. Arai, "Study on Tele-Existence (II) - Three-Dimensional Color Display with Sensation of Presence," Proceedings of the International Conference on Advanced Robotics, Tokyo, Japan, pp. 345-353, (1985).

Zangmeister, W.H., A.Jones, and L.Stark, " Dynamics of Head Movement Trajectories," Exp. Neurol. 36, 76-91, (1981).

Zientara, M.," New View on the World - Pioneer Designs 3D Camera Lenses with Micro," InfoWorld, May 1984.