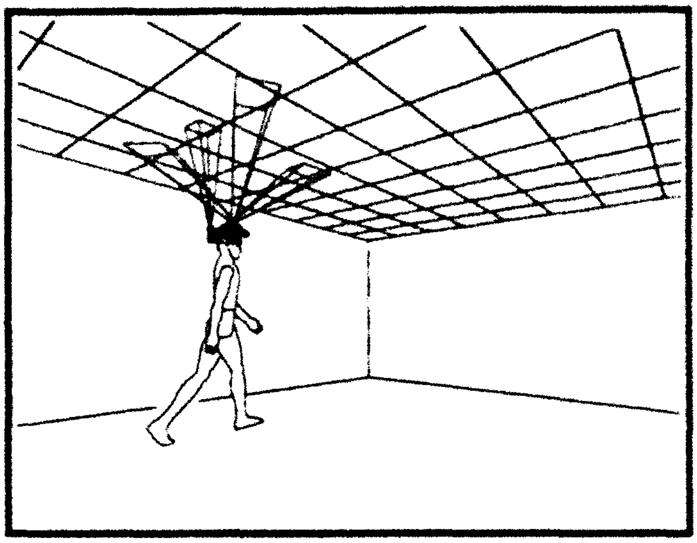
Advanced Technology for Portable Personal Visualization

Report of Research Progress
July 1992 – January 1993

This research is supported in part by
DARPA ISTO Contract No. DAEA 18-90-C-0044

Department of Computer Science
University of North Carolina at Chapel Hill
Chapel Hill, NC 27599-3175
Telephones: 919-962-1931 (Brooks), 919-962-1911 (Fuchs)
FAX: 919-962-1799
E-mail: brooks@cs.unc.edu, fuchs@cs.unc.edu



Principal Investigators: Frederick P. Brooks, Jr., and Henry Fuchs

Faculty: Gary Bishop, Research Associate Professor; Vern Chi, Director, Microelectronic Systems Lab;
Steve Pizer, Kenan Professor; William V. Wright, Research Professor

Staff: Brad Bennett (Systems Programmer), Steve Brumback (Electrical Engineer), Stefan Gottschalk (Programmer),
David Harrison (Hardware Specialist, Video), John Hughes (Hardware Specialist, Force Feedback ARM),
Linda Houseman (Administration), Kurtis Keller (Research Engineer), Jannick Rolland (Optical Engineer),
Kathy Tesh (Secretary), Fay Ward (Secretary)

Research Assistants: Ron Azuma* (Optical Tracker), Mike Bajura** (Ultrasound), Andrew Bell** (Walkthrough),
Devesh Bhatnagar* (Optical Tracker), Andy Brandt** (Ultrasound), David Chen** (Ultrasound),
Jim Chung (HMD—Radiation Treatment Planning), Drew Davidson (HMD—Sound, 3D Modeler),
Erik Erikson (HMD—Foot Tracking, Virtual Tool Kit), Rich Holloway (See-through HMD, error registration),
Terry Hopkins** (HMD—Distortion Correction), Mark Mine* (HMD—Low Latency Rendering),
Ryutarou Ohbuchi* (Ultrasound), Russell Taylor*** (Nanomanipulator), Chris Tector (Ultrasound),
Hans Weber (Architectural Walkthrough)

* Supported by NSF Cooperative Agreement no. ASC-8920219
** Supported by NSF Grant no. MIP-9000894 / DARPA order no. 7510.
*** Supported by NIH grant no. 2-P41_RR02170-09

Advanced Technology for Portable Personal Visualization

1.0 Summary and Goals

Despite the recent avalanche of publicity on virtual reality, the state of the art in head-mounted displays (HMD) is still at the "almost-works" stage. Although extravagant claims are made almost daily on television and in the popular press, there is very little that can be usefully accomplished today with virtual reality technologies.

We at UNC have been working since the early 1970s on aspects of these technologies and have been advancing the state of the art by a "driving problem" approach; we let the needs of selected applications stimulate the direction of the technological developments and then test new results by observing their impact on solving the original application. We are working on three application areas: molecular modeling, 3D medical imaging, and modeling of architectural interiors.

We have developed complete systems that include the head-mounted display device, the display generation hardware, and a head and hand tracker. We have access to a force-feedback ARM that is owned by the GRIP project. We purchase components when available and build components when there was a clear advantage to doing so:

a) We built, under separate major funding by DARPA and NSF, the display-generation hardware (our most recent machine is Pixel-Planes 5, and a new machine, PixelFlow, is now in the design stage).
b) We built an optoelectronic helmet-tracking system (the ceiling tracker) that can determine position and orientation over a large area (currently under a 10- x 12-foot ceiling). To our knowledge, this is the first demonstrated large-area helmet tracker.
c) We purchase commercially-available head-mounted display devices but continue to build head-mounted display devices with see-through capability.

Construction has begun on a larger version of our ceiling tracker. Covering four times the area of the original, this will offer truly wide-range tracking for architectural walkthrough applications and see-through HMD-based applications, such as ultrasound examination. By developing an algorithm to calibrate beacon positions, we greatly reduced the cost of the new ceiling. We plan to have the new system operational in March 1993.

1.1 Goals of the Head-Mounted Display Project

- Demonstrate the usefulness of the head-mounted display (HMD) in real applications.
- Build a system in which virtual objects appear to remain fixed as the user's head moves.
- Improve the hardware subsystems which currently limit performance of HMDs (trackers, HMD optics and displays, real-time graphics generation engines).
- Design and implement software base to support HMD functions, including model-building tools.

- Integrate visual, auditory and haptic (force feedback) displays into a working system.
- Build new input devices and investigate methods of manual control suitable to an immersive synthetic environment.
- Build and investigate the uses and limitations of see-through HMDs (optical and merged-video designs).

1.2 Goals of the See-through Head-Mounted Display Subproject

- Derive and test a theoretical model for the registration error in a see-through head-mounted display (STHMD).
- Build a low-latency system so that virtual objects appear to be stationary when superimposed on the real environment.
- Experiment with methods of registering real and virtual objects.
- Derive a method to calibrate a see-through head-mounted display accurately.
- Assess perception of depth and sizes using a bench prototype STHMD.
- Build and integrate a 60-degree field-of-view STHMD system.
- Design, build, and integrate a video STHMD using two miniature video cameras.

1.3 Goals of the Tracker Subproject

- Develop wide-range trackers for head-mounted displays. Improve our existing tracking system, which can track an HMD inside a small room-sized environment. For the long term, explore technologies that can track multiple users and have unlimited range without requiring modification of the environment.
- Work to improve the other aspects of HMD tracking: latency, speed, accuracy, and resolution.

1.4 Goals of the Walkthrough Subproject

- Develop tools and techniques for building synthetic environment models.
- Explore virtual environments through architectural simulation. Walking through a building is a natural and intuitive process in the real world; the goal of the Walkthrough project is to use this intuitiveness to study virtual-world interaction, especially wayfinding in virtual spaces.
- Serve as a driving problem for HMD, tracker, and graphics architecture research. Walkthrough is a real-world application that pushes the limits of HMD, tracker, and graphics architecture technologies.

1.5 Goals of the Virtual-Environment Ultrasound Scanning Subproject

- Build and operate a 3D ultrasound system which acquires and displays 3D volume data in real time. This design requires significant advances in both 3D data acquisition and 3D volume data display. The former topic is being investigated by a research group at Duke University and is not included in our project. The latter topic is the focus of Prof. Henry Fuchs' ultrasound research at UNC-Chapel Hill. Furthermore, accurate tracking of both the physician's head and hand are required to register the computer reconstructed ultrasound volume properly with the real world.

1.6 Goals of the Nanomanipulator Subproject

- Build a system that a surface scientist can use to perform real-time modification experiments on surfaces under study with a scanning tunneling (STM).
- Gain an understanding of the basic processes for surface modification using an STM by means of bias voltage pulses.
- Use the system to create quantum interference devices, such as quantum transistors.

2.0 Summary of Major Accomplishments

2.1 Head-Mounted Display System

- During this reporting period, we finished the proof of concept of pre-distortion of images to compensate for optical distortion. We also achieved limited success with pre-distortion of only vertex locations of polygons.
- The first stages of a Just-In-Time Pixels system, which incorporates both knowledge about the time of display of a pixel and predictive head-tracking techniques in the computation of the displayed pixel values, has been implemented. The intensity value displayed at each pixel is based upon the estimated viewpoint of the user at the time of its display.
- We generated an animated sequence simulating the difference between images based on Just-In-Time Pixels and conventionally rendered images. We also verified that the distortion due to conventional rendering techniques (and the corresponding lack of distortion of Just-In-Time images) is perceptible in objects moving relative to the user's viewpoint.
- A beam targeting tool to be used for evaluation of the effect of head tracking on targeting radiation treatment beams in cancer therapy has been developed and is being tested in user studies.
- The team developed a serial communications library so that all serial trackers will be able to run over the telephone lines from any of the host machines, providing much greater flexibility in terms of HMD station configurations. This is distinct from the serial device library which operates over the network. The advantage to this approach is that once the phone-line connection is established, the line is dedicated and, therefore, impervious to network loads. This enables us to handle switching of video sources in software from any workstation.
- A 3D color selection survey including slicing through RGB color cubes and HSV color cones was created. It could not be determined whether implemented 3D methods are superior to the 2D methods tried (e.g., HSV sliders) for virtual-worlds interaction.
- Foot tracking became a new area of experimentation. The user's feet are tracked relative to the body in order to derive absolute position (given a known start point). A preliminary paper describing the system and some of its inherent problems was written.

- The sound library for playing sounds in the virtual world was rewritten to increase its responsiveness and to enable the playing of binaural azimuth-localizable sounds.

2.2 See-through Head-Mounted Display

2.2.1 30-Degree Field-of-View System Using Off-the-Shelf Optics

During the last reporting period we built and integrated a first prototype of the 30-degree field-of-view optical and mechanical head-mounted display system. During this reporting period we have made the following accomplishments in research using this see-through head-mounted display:

- We completed a first set of psychophysical studies using the optical bench prototype see-through HMD. Those included size and depth perception in virtual environments, as well as the effect of mismatch between interpupillary distance (IPDs) on the user of the HMD and on the optics and software setups. Part of this research was funded under an NIH grant.

- A first draft of a paper describing the calibration of the system and the results obtained has also been completed.

- Dr. Jannick Rolland presented the results of the experiments in an invited talk at the Optical Society of America by in September 1992 (see Section 5.3).

- We created a simple demonstration application to illustrate the lag problems inherent in see-through head-mounted display (STHMD) systems. This STHMD has been used as a testbed for low-latency rendering software and for low-latency tracker testing; work is continuing in this area.

- The Pixel-Planes 5 graphics software was modified to handle the inverted screen configuration (which complicated the lighting calculations), and an "x-ray vision" visualization program was demonstrated for attendees of the Visualization in Biomedical Computing conference held at UNC. The demonstration features a computed tomography (CT) dataset of a human jaw superimposed onto a styrofoam head as a simulation of a surgical planning tool. The user can make tissue-depth measurements with a virtual ruler.

- The "x-ray vision" demonstration was augmented with virtual menus (from the VTK library described in the previous report) to allow the user to perform various operations, the most important of which is the automatic registration of the virtual and real objects. The user picks "landmarks" (as used by surgeons in current practice) on both datasets, and the system uses a principle components method to align them. Currently, only the centroids are aligned, but work on implementing the rotation portion of the transformation should be completed shortly. The system was also augmented with sound to give the user feedback on menu selections, registration operations, and errors.

2.2.2 60-Degree Field-of-View System with Custom Optics

- A design review of the optics of the 60-degree field-of-view (FOV) optics was conducted at UNC-CH in the presence of a representative of Tektronix, the company that is providing the two miniature color displays. The design review showed that the color plate used in conjunction with the CRTs to provide color needed to be moved close to the CRT; this was due mainly to optical reasons resulting from the folding geometry chosen for the system. The color plate had been moved after the folding prism previously to facilitate the mechanical assembly of the system.

- The reoptimization of the 60-degree FOV optical system was completed in October 1992.

- We also completed the design for the helmet to mount the complicated optics. The helmet has adjustments for the interpupillary distance (IPD) and balanced adjustments for the height of the optics. The design utilizes composites to reduce the mass of the helmet. A unique mount was also designed; This mount creates an oil-tight seal between the lenses and the CRT, and it can also rotate.
- Following the first pass at the mechanical assembly for the 60-degree FOV system, we discovered that pressures exerted by the CRT from thermal expansion could result in a loss in optical image quality. We solved this problem by deciding to couple the CRTs to the optics via a thin film of oil of 0.4mm. Adding this extra thickness of liquid without reoptimizing the system would put the final image out of focus, but we his was compensated for this by having the color plate made slightly thinner.
- We have received a reply from Tektronix that the plate could be made 0.01" thinner, but have not yet received any specific details about which layer of the plate would be made thinner. This information is necessary to proceed with the fine tuning of the optics.

2.2.3 Video See-through Cameras with Custom Optics

- We have completed both the optical and mechanical design for mounting the two miniature video cameras to be used with the Flight Helmet from Virtual Research. The lenses of the video cameras have been designed to compensate for the optical distortion of the LEEP optics used in the Flight Helmet, and the FOVs of the two cameras were designed to match the Flight Helmet FOVs. The required optics designs have been sent to the selected manufacturer.
- Lens mounts, which include the adapters for placing the lenses off center to the cameras, have been designed for the miniature lenses .
- A detailed report of the optics design has been written by graduate student Emily Edwards and Dr. Jannick Rolland.

2.3 Tracking

Background:

We first demonstrated at the ACM SIGGRAPH '91 conference (28 July–2 August 1991) our custom-built optical tracking system, which features a scalable work area that currently measures 10' x 12'. The sensors consist of four head-mounted imaging devices that view infrared light-emitting diodes (LEDs) mounted in standard size (2' x 2') suspended ceiling panels. Photogrammetric techniques allow the head's location to be expressed as a function of the known LED positions and their projected images on the sensors. Discontinuities that occurred when changing working sets of LEDs were reduced by carefully managing all error sources, including LED placement tolerances, and by adopting an overdetermined mathematical model for the computation of head position: space-resection by collinearity. A novel aspect of this system is that the range is not limited to the current configuration. By adding more panels to the ceiling grid, we can scale the system to any desired room size, and we have just begun building a larger version of this system that is four times larger than the original. To our knowledge, this is the first demonstrated scalable tracking system for HMDs. [Ward92]

After its introduction in the summer of 1991, we installed this optical tracking system in our graphics laboratory, where it is now a platform for further tracker research and for running HMD applications. It supports the Walkthrough project, which retired their treadmill input device in favor of our optical tracker. Efforts are underway to integrate see-through HMD capability with this tracker to help the Ultrasound group.

Achievements during this reporting period:

- Construction of the new ceiling has begun. This is a much larger version of the existing ceiling tracker. Spanning about 16' x 30', it covers four times as much area as the existing system. Circuit boards were designed for the panels and supporting electronics. We are now testing prototypes of these boards. On 25 January 1993, construction crews began tearing down office walls to build the new Tracker/Ultrasound laboratory where the new ceiling tracker will reside.

- We refined our calibration algorithm for measuring the locations of the ceiling tracker's LEDs. To demonstrate it, we physically moved some panels in the existing ceiling and then measured the new LED locations. This work was done in conjunction with John F. Hughes of Brown University, and we wrote and submitted a paper detailing this method (see Section 5.2). The ability to measure LED locations in the ceiling means we will no longer have to carefully align the LED positions to a known grid. This greatly reduces the expense needed to build the larger ceiling.

- We implemented a simple prediction method on the ceiling HMD. Accurate prediction requires accurate knowledge of time in the system. To obtain this, we bypass UNIX entirely and directly inject the ceiling tracker data into our graphics engine, Pixel-Planes 5. Clocks on the tracker computers and on Pixel-Planes 5 are synchronized to within a millisecond and are resynchronized every two to three minutes to compensate for drift. More advanced prediction techniques will be explored in the next reporting period.

- Gyros and accelerometers have been acquired for use in aiding predictive tracking and eventually in experimenting with inertial tracking. We tested them individually with a PC-based A/D card, built mechanical mounts for the sensors, and are in the process of integrating them with the ceiling HMD.

- A video camera is mounted above the right eye of the ceiling HMD to provide a primitive video see-through capability in one eye. This will eventually support evaluation of prediction methods and ultrasound applications.

- We built a mount that holds an ultrasound transducer and three optical sensors. With this, the ceiling can track the position and orientation of the ultrasonic sensor, which will permit the Ultrasound group to collect data without worrying about magnetic distortions to which Polhemus and Ascension trackers are vulnerable.

- We have reduced the jitter that occurred when one of the camera views moved off the ceiling. This was accomplished by modifying the code to assign weights to each LED in the ceiling, with progressively smaller weights being assigned to the LEDs closer to the edge of the ceiling. Use of these weights gradually reduces the importance of a view as it moves closer to an edge, thus reducing the jitter caused by its sudden disappearance when it falls off the edge.

- We performed some tests to estimate the magnitude of the distortion in our Polhemus and Ascension trackers. Further end-to-end timing studies were also performed on these and other trackers. The results are being recorded in a technical report.

- We investigated sources of noise in an attempt to get cleaner and more accurate signals from our optical sensors. The investigation did not identify any large error sources.

- A safety mechanism was developed and installed in the standard software to prevent us from firing the LEDs more often than their specified limits permit. We have noticed that about four

LEDs in the existing ceiling appear to be dead, so this new mechanism should protect us from burning out more LEDs.

- Preliminary studies have been done on the feasibility of making a wireless version of the ceiling HMD so that the user will have no cable to drag.
- We submitted a short article describing the ceiling tracker and the requirements demanded of trackers in see-through systems to *Communications of the ACM*, for a special issue on augmented reality systems.

2.4 Interactive Building Walkthrough

- We have reached the stage where interactive radiosity is functional and can be integrated easily into existing applications to provide highly realistic lighting simulations in near-real time.
- We have moved automatic model mesh-generation in the Virtus WalkThrough-to-Pixel-Planes 5 modeling pipeline onto the graphics processors, allowing it to work more tightly with interactive radiosity.
- Significant process has been made in building a highly detailed architectural model to serve as a testbed for user studies on the effectiveness of different types of detail in architectural environments and on wayfinding methods.
- We have continued running user studies with Walkthrough under the ceiling tracker to study the effectiveness and shortcomings of the existing system.

2.5 Virtual-Environment Ultrasound Scanning System

- We implemented a system to enhance visual depth cues by displaying a virtual "pit" within the patient. The walls of the pit are opaque, which allows the physician to judge better where the ultrasound volume lies within the patient.
- The implementation of a software system to reconstruct a series of 2D ultrasound images and their locations and orientations into a 3D regular volume has been completed. This is done off-line.
- The Vlib virtual world management capabilities are now integrated into VVEVOL to allow volume rendering in a virtual environment.
- We integrated ultrasound image acquisition, resampling, and volume rendering into a virtual environment to create one large interactive ultrasound virtual-reality application.
- A system for calibrating both the ultrasound transducer and the see-through video for use under the optical ceiling tracker was developed. The calibration procedure is independent of the type of tracking or position/orientation of the transducer, allowing the system to function under changing conditions. Further work is necessary, but the system shows positive preliminary results.
- Dr. Vern Katz of the UNC Hospitals' Department of Obstetrics and Gynecology used our volume-rendering ultrasound system to scan a fetus inside a pregnant woman.
- We designed and simulated in software a real-time video frame grabber for the Pixel-Planes 5 high-speed ring. The design for the board has gone to the Microelectronics Center of North Carolina for fabrication.

- General Electric Ultrasound Systems has made a generous loan to us of a new ultrasound scanner. This scanner has increased the resolution of our ultrasound images. Furthermore, it has many advanced features lacking on our previous machine.

2.6 The Nanomanipulator: An Atomic-scale Teleoperator

- We have interfaced the system with an HP8131A pulse generator to provide bias pulses under user control. We have used pulses to place mounds of gold onto a surface at locations specified by the user in real time, using our manipulator ARM.
- We have characterized the elements of the feedback loop using a network analyzer. With this information, we designed a new feedback control circuit to optimize system performance.

3.0 Expected Milestones during the Next 12 Months

3.1 Head-Mounted Display System

- Pre-distort images more accurately by dicing large polygons and moving polygon vertices.
- Continue development of the Just-In-Time Pixels renderer. Primary goals will be the incorporation and fine tuning of predictive tracking techniques and the inclusion of the ability to handle more complex scene geometries.
- Complete a study of the effect of head-mounted display on targeting radiation therapy beams. Practicing radiation oncologists, dosimetrists, and radiation physicists are currently being recruited for participation in a user study. This study will make subjective and objective comparison between treatment beam configurations designed with a head-tracked navigation mode and those designed without head tracking.
- Write and publish a paper describing the essential coordinate systems and transformations for a head-mounted display system.

3.2 See-through Head-Mounted Display

3.2.1 30-Degree Field-of-View System Using Off-the-Shelf Optics

- Write the final draft of the paper describing the calibration of the system and psychophysical data obtained.
- Write a research proposal to obtain further funding for this research.
- Procure and calibrate CCD cameras to measure registration error for the 30-degree FOV system.
- Build measurement setup for analyzing registration error for a simple system.
- Derive a model for registration error as a function of error sources (distortion, tracker error, alignment error, etc.).
- Measure registration error for subsets of the STHMD system of ever-increasing size, and compare them with error predicted by the model.
- Complete the calibration procedure for the see-through HMD system.

3.2.2 60-Degree Field-of-View System with Custom Optics

- Complete the fine tuning of the optics design once exact numbers are provided by Tektronix.
- Complete the mechanical design of the system.
- Have the optics and the mechanical parts for the assembly manufactured.
- Assemble the first prototype of the system.

3.2.3 Video See-through Cameras with Custom Optics

- Investigate alternatives for mounting the cameras on the helmet.
- Assemble the cameras after receiving the lenses and the mechanical parts.
- Evaluate the system.

3.3 Tracking

- Finish construction of the new ceiling and get it running. This includes hardware setup and software modifications. Run our calibration method to measure the LED locations in the new ceiling. Successful measurement of LED locations will control discontinuities in the computed position and orientations reported by our tracker. Discover how much this increased range of motion helps in certain applications.
- Mount the gyros and accelerometers on the ceiling HMD, and use them to support advanced prediction methods. Develop, demonstrate, and evaluate prediction methods with both objective measures and user studies.
- Investigate better math techniques for computing the position and orientation of the head with our equipment. Our existing method, space-resection by collinearity, is an iterative method that does not always converge to the correct solution in difficult conditions. Other methods should be explored. One particular alternative flashes only one LED per update, using an incremental Kalman filter to estimate the new position and orientation. It has the potential to improve significantly the accuracy, update rate, and lag of this tracker.
- In order to support exploration of alternate methods, revise the ceiling tracker system software. This will add the ability to record user motion with accurate timestamps and low-level information, such as individual sensor photocoordinates. The programs which control the hardware already offer timestamp support that is accurate to about 30 microseconds. By recording data runs, we can apply alternative mathematical methods to the data to compute position and orientations off-line.
- Integrate the see-through capability for at least one eye on the ceiling HMD in a more rigorous manner than the current implementation. Build a rigid mount to hold the camera, and work on actually achieving reasonable registration of real and virtual objects. See-through capability makes latency much more visible, aiding evaluation of delay compensation techniques.
- In the long run, explore technologies for unlimited range tracking in unstructured environments. An area that needs exploration is *relative-mode* tracking, which measures only the relative differences in position and orientation as the user moves and integrates these differences over time to recover the head's location. Inertial systems are an example of such a tracker. The main

problem with relative-mode trackers is drift, since repeated integration accumulates error over time. A hybrid inertial-optical tracker could prove to be an effective system. Occasional fixes from the optical tracker will prevent drift from accumulating endlessly, and the inertial units will allow much greater freedom of head orientation and immunity to environmental distortions. Such a system could dramatically reduce the number of LEDs required or, for an equivalent cost, greatly expand the range of the tracking system.

3.4 Interactive Building Walkthrough

- Integrate more advanced versions of the Virtus WalkThrough model-building into the system, with support for textures, model partitioning, more complex radiosity emitters, and the replacement of model parts with objects from our model libraries.
- Refine the current interactive radiosity system to increase the accuracy of the lighting model, to lower the lag time required to get an initial solution to the lighting model, and to smooth the transitions as more accurate solutions are obtained.
- Incorporate adaptive and dynamic meshing into the automatic model mesh-generation to provide mesh refinement and to allow the user to change the geometry of the scene.
- Implement an automatic model partitioning scheme by using hierarchy and containment information exported from Virtus Walkthrough.
- Conduct additional user studies.

3.5 Virtual-Environment Ultrasound Scanning System

- Build and integrate a real-time video frame grabber to allow ultrasound image acquisition at a rate of 30 Hz.
- Improve calibration techniques for merging computer-generated and real-world images in a virtual environment. The calibration procedure previously discussed is being designed to solve both video and transducer calibration problems.
- Track the physician's head and hand under the optical ceiling tracker, as opposed to using a magnetic tracker like a Polhemus or Ascension tracker.
- Improve the quality of the image rendering.
- Conduct an experiment in which all resampling and rendering is done off-line to allow the maximum image resolution. This will also drastically cut the perceived lag in the system.
- Conduct experimental trials of the improved system with an Ob-Gyn physician, an ultrasound technician, and several pregnant subjects.

3.6 The Nanomanipulator: An Atomic-scale Teleoperator

- Modify the STM so that the wave form of the tunneling current resulting from voltage pulses can be observed and recorded.
- Work with our collaborator in this research, Professor Stanley Williams of the UCLA Department of Chemistry, to conduct other experiments possible with these new facilities.

4.0 Discussion of Research

4.1 Head-Mounted Display System

Hardware

The current HMD system is used by many projects and subprojects in a variety of configurations. The following is an overview of the current system:

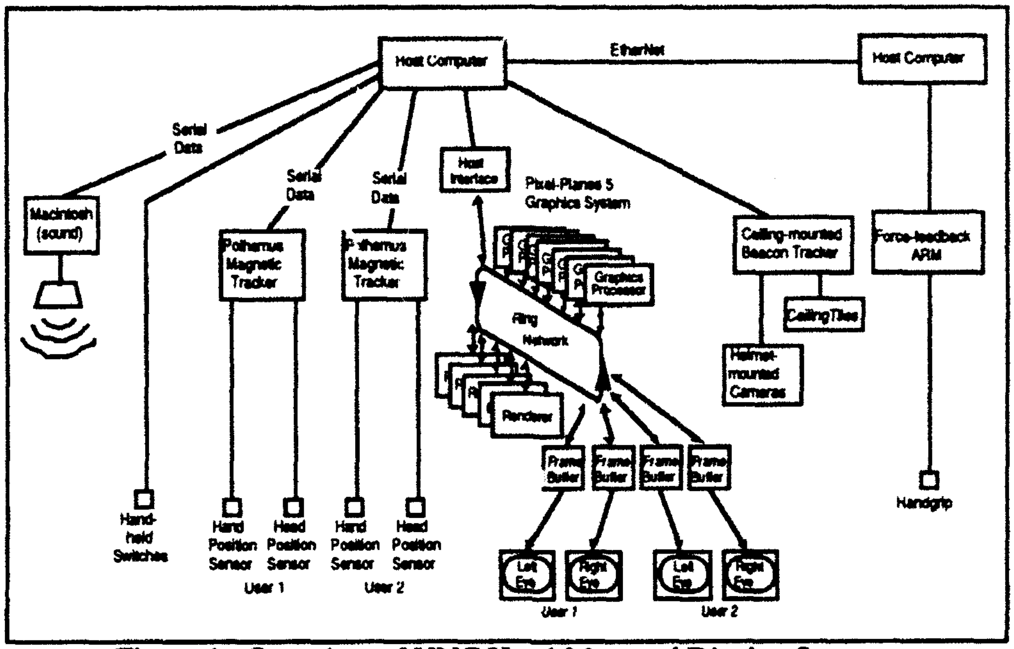


Figure 1. Overview of UNC Head-Mounted Display System.

A description of each component follows:

- Host computer: A Sun 4, which runs the host side of the application and interfaces to the graphics engine.
- Graphics Engine: Pixel-Planes 5, a message-passing multicomputer capable of drawing more than two million Phong-shaded, Z-buffered triangles per second. Multiple applications can run on the system at the same time due to its ability to be split into several subsystems.
- Magnetic trackers: We use Fastrak, 3SPACE, and Ascension magnetic trackers.
- Ceiling tracker: Described in Section 4.3.
- Force-feedback ARM: Simulates forces in the virtual world. It has six degrees of freedom (3 forces and 3 torques) and uses computer-controlled servo motors.
- Manual input devices: The newest input device has a natural handgrip shape with five buttons for user input. We also have hollowed-out billiard balls with 3SPACE or Fastrak sensors mounted inside, and switches on the outside for signaling user actions.

- Sound: Controlled by Macintosh; plays sounds under application control either in earphones attached to HMD, or from separate speakers. The sound Macintosh can be controlled by any machine on our network transparently using sdilib.

In general, an application can run with almost any subset of the above equipment.

Software

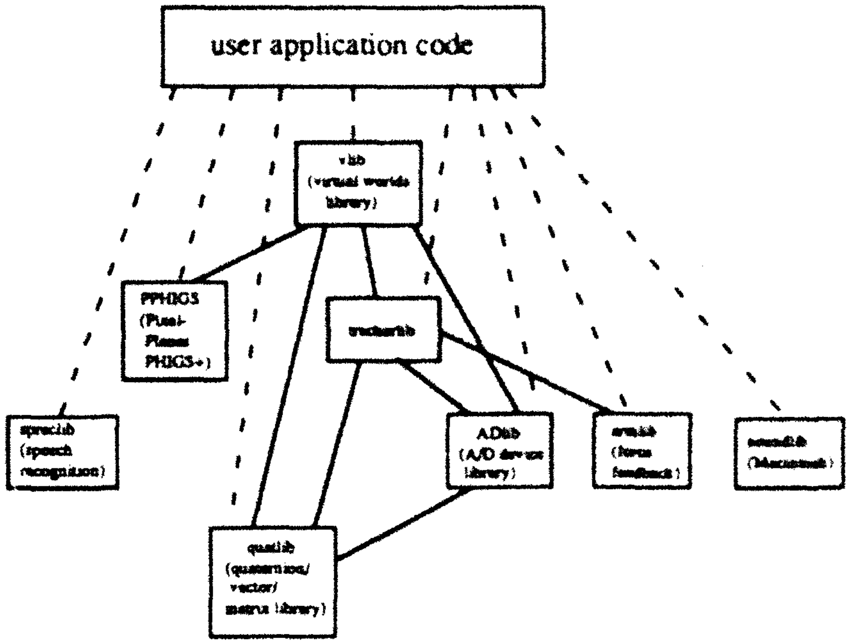


Figure 2. Overview of UNC Virtual-Worlds Software.
(Dashed lines indicate application calls; solid lines indicate inter-library calls.)

The virtual-worlds base software is fairly complete but will continue to expand to be more flexible and powerful and to accommodate new hardware. Most of the libraries support multiple technologies transparently, allowing application code to switch easily between different pieces of hardware by changing only an environment variable. While efforts at improving these libraries will continue, we will work to add new capabilities and develop new applications. The software goals are listed in with the other twelve-month goals for the project in Section 3.1.

4.2 See-through Head-Mounted Display

The search for the perfect display device for HMDs is ongoing. While most commercial research is aimed at large screen TVs, projection systems, HDTV, and large-screen flat panels, only a very small portion of current research deals with the development of small, high-resolution monitors. We continue the evaluation of different technologies and will incorporate those which can extend the usefulness and availability of HMDs.

Research, both in hardware and human factors, will be extended to different types of HMDs. These include the standard, total-immersion variety and see-through (video and optical) HMDs.

Reducing the weight of HMDs is also a large area of concentration of research effort. Due to the highly constrained 60-degree FOV, see-through HMD specification, lighter weight in this design has been sacrificed to some extent in order to achieve a higher image quality and an ideal geometry. Indeed, the geometry is thought to be as important to the comfort of the HMD user as the weight itself. The first prototype will be built using conventional optics since the technology of making lenses from glass or plastic is well understood. An exciting and promising way to reduce the system's weight and its complexity, however, is to explore the usefulness of combining classical optics with binary optics. This seems the logical step to take in the design of the off-axis 60-degree FOV system. Other general weight-saving tricks, such as composite headgear, plastic lens mounts, and wireless video transmission, are under investigation.

The design of a video see-through system using miniature cameras with the Flight Helmet from Virtual Research was completed. The main characteristics of this system are: the cameras' FOVs match the FOVs of the Flight Helmet, the chosen amount of distortion of the cameras' lenses can compensate exactly for the strong distortion of the LEEP optics used in the Flight Helmet, and the offsets of the CCD detectors are used in the cameras to match the centers of distortion of the cameras with those of the LEEP optics. For example, the LCDs within the Flight Helmet are offset with respect to the LEEP optics by 6.5 mm to accommodate for mechanical assembly of the Flight Helmet. We are currently working on how to mount the cameras on the Flight Helmet to minimize any magnification of the real world with respect to the virtual world. To achieve a magnification of 1, the cameras would have to be mounted as if their viewpoint were the eyes of the user's eyepoints. This is not feasible, as explained in the paper written by Edwards and Rolland, but the solution adopted may yet be adequate. One of our goals is to test the system for the adequacy of our chosen solution. If the test results are negative, the next step would be to redesign a complete helmet and video camera assembly.

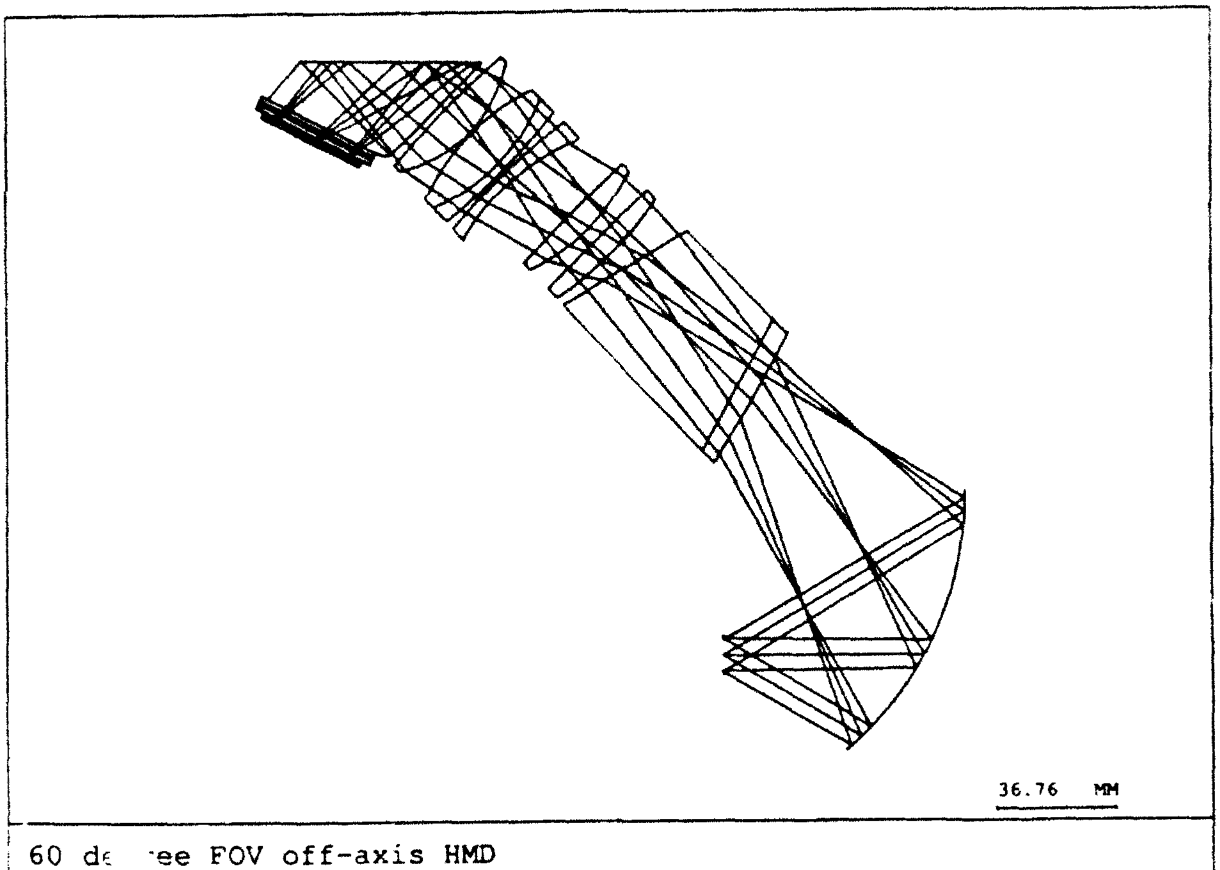


Figure 3. Optical Layout of the Folded See-through 60-Degree Field-of-View HMD.

The goal of the error-model research is to determine the registration error between a real object and a virtual object that are supposed to be in the same location. This error is a function of viewer position and orientation, as well as the head's angular and linear velocities. The theoretical error model will be derived first, and then measurements on a real system will be made to check the model's ability to predict the registration error.

Since the registration is a time-varying visual phenomenon, small cameras positioned where the user's eyes normally would be will be used to record the discrepancy between the real points and their virtual counterparts. We will first build a simple system, derive its theoretical error, and then measure its real error. The system will then be augmented to make it more like a complete STHMD system, and the process will be repeated. In the end, we should have a complete model that predicts the behavior of our existing system as a function of time, position, orientation, and velocity. This model will also include a sensitivity analysis, which will show how an error in any stage of the system will be propagated through the system to affect the final registration error. We will focus our efforts on attacking the most important errors first.

4.3 Tracking

It is well known that trackers for HMDs require high resolution, high update rate, and low latency Range, however, has not received similar emphasis. Many virtual worlds, such as architectural walkthroughs, would benefit from more freedom of movement than is provided by existing trackers. Wide-range trackers would allow larger areas to be explored naturally, on foot, reducing the need to resort to techniques such as "flying" or walking on treadmills. Also, such techniques of extending range work only with closed-view HMDs that completely obscure reality. With see-through HMDs, which we are attempting to build for medical applications, the user's visual connection with reality is intact, and hybrid applications in which physical objects and computer-generated images coexist are possible. In this situation, flying through the model is meaningless. The model is registered to the physical world and one's relationship to both must change simultaneously.

Our optically-based ceiling tracker gives us room-sized tracking capability today and is a vehicle for tracker and HMD application research. The advent of a much larger ceiling will give us truly long-range capability which we hope will improve exploration of architectural walkthroughs and provide a framework for see-through HMD applications. In the long term, however, we would like unlimited range tracking in unmodified environments. No technology today can provide this, but a hybrid inertial-optical system might prove to be an effective solution in the next few years.

Larger Ceiling Tracker

One of the novel aspects of the ceiling tracker is that it can be scaled to any desired room size. The current 10' x 12' system is about the size of a small office or a small dormitory room. While that is large enough to provide some freedom of movement, it really is not very big. For example, some users find it frustrating to explore a virtual kitchen that is larger in area than the tracking system can cover. Thus, we are in the process of building a new ceiling that covers approximately 16' x 30', or about four times the area of the existing system.

The existing ceiling was built to exacting tolerances to maximize the chances of success of the original system. If the locations of the beacons in the ceiling are misplaced by several millimeters, noticeable errors can be observed in the computed position and orientations. Each panel was stretched flat, suspended from an external superstructure, and carefully positioned with a laser level. The panels themselves are based on a sparsely populated circuit board that measures almost

2' x 2'. These measures keep the LEDs to their specified locations within a tolerance of 1-2 mm, but at significant cost.

Since the new ceiling has four times as many panels as the original, several measures were taken to reduce the cost. First, two new circuit boards were designed to replace the original 2' x 2' board. One board, of which there are four per panel, is a long strip to hold the LEDs, and the second holds the controlling circuitry for the panel. Since the area of the two boards is much smaller than that of the original circuit board, they are considerably cheaper. Second, instead of trying to fix the LEDs to a predetermined grid, we place them in a ceiling and then measure where they are using calibration techniques (see the next section). This means that the tolerances on panel construction can be much looser than they were in the original panels. Third, we are dropping the new panels into standard 2' x 2' ceiling grids. This is made possible by a much lighter panel design, adherence to UL and fire codes, and our ability to measure the locations of the LEDs. This means that we can dispense with an external ceiling superstructure that holds the panels in precise locations. To a casual observer, the new ceiling tracker should look indistinguishable from any other ceiling in our building, thus reducing the intrusiveness of the tracker.

Calibration of LED Locations

In collaboration with Professor John F. Hughes of Brown University and Professor Al Barr of California Institute of Technology, we have developed a procedure for measuring the locations of the LEDs in the ceiling. The success of the new ceiling, currently under construction, relies on this calibration. These locations must be known to within a millimeter or two for acceptable tracking performance, in terms of smooth and accurate tracking.

The calibration algorithm begins with a rough estimate of the beacon locations and a large collection of raw HMD sensor data. A relaxation technique is used to adjust the LED position estimates until they are consistent with the raw sensor data. The initial LED position estimates are expected to be accurate to within 1 inch of true value. The raw HMD sensor data are collected while the HMD is being moved about beneath the ceiling. No additional hardware is required, and only practicable precautions need to be taken while collecting the data.

The basic algorithm works as follows: After collecting the raw data (about 45 minutes are required to collect HMD sensor data from 25,000 positions), we use a method called space-resection by collinearity to compute where each of the observations was made. The LED positions are inaccurately known, so the computed observation positions will be inaccurate. The collinearity algorithm, however, has the property of averaging out some of the error. Now consider the observation positions to be fixed, and allow the LED positions to vary. We "fit" each of the LED locations to the fixed observation positions. This too has the property of averaging out random error. This concludes a single iteration of the algorithm. These two steps are repeated until adjustments of both LED positions and observation positions are minimal. This typically takes about 20 iterations, which require 2 hours of computing time on an HP model 700 for 25,000 observations on a ceiling with 960 LEDs. The algorithm is linear in both observation count and LED count.

In the course of developing our algorithm, we discovered two significant sources of error which must be controlled. First, the algorithm is less effective when using raw HMD sensor data obtained while the headmount was in rapid motion. Ideally, the sensor data should be taken when the headmount is stationary, but this would make data collection too lengthy a process: 25,000 observations is a typical observation count needed, and holding the headmount still in 25,000 different locations is not feasible. Therefore, the data is acquired "on the fly," that is, while the HMD is being move around beneath the ceiling. Experience shows that moving the headmount slowly produces better data than moving it fast.

The better solution is to account for movement in the mathematical model of the headmount motion. The current model assumes that the HMD does not move significantly in the 20-millisecond window during which data are taken for each observation. Experience shows, however, that the movement is significant. More sophisticated models would help to reduce this *dynamic error*.

The second source of error is related to the spatial distribution of the observations. Each LED must be seen in many separate observations and from diverse directions. If it is poorly sampled, then the error averaging property of the LEDs' fit to the observations is lost, and the LED will not be properly calibrated. In response to this problem we use goal-directed data collection. We developed an X-Windows application which displays a top view of the ceiling and highlights the locations of poorly sampled LEDs, as well as the LEDs currently visible to the head-mounted cameras. This runs while the operator is collecting data, allowing him/her to move the HMD to those undersampled areas of the ceiling and, thereby, obtaining a better distribution of observations.

As a test of the algorithm's robustness, we tilted three ceiling panels in the existing system by inserting 2 1/4-inch standoffs under one edge. A photograph of the ceiling with these tilted panels is shown in Figure 4. Recall that the panels are 2 feet on a side. The calibration procedure was then conducted to determine the new locations of these LEDs. The initial estimated LED positions were those of the undisturbed ceiling adjusted by up to 1.7 inches in a random direction. It is believed that this is a more strenuous test of the calibration algorithm than the new, larger ceiling will be. The result of the calibration is shown in a computer-generated model of the LED locations, Figure 5. A true measure of the algorithm's accuracy is difficult to obtain, but the algorithm does produce a measure of confidence. In the calibration of the three tilted panels, described above, it is believed that the LEDs positions are known to within about 1 millimeter.

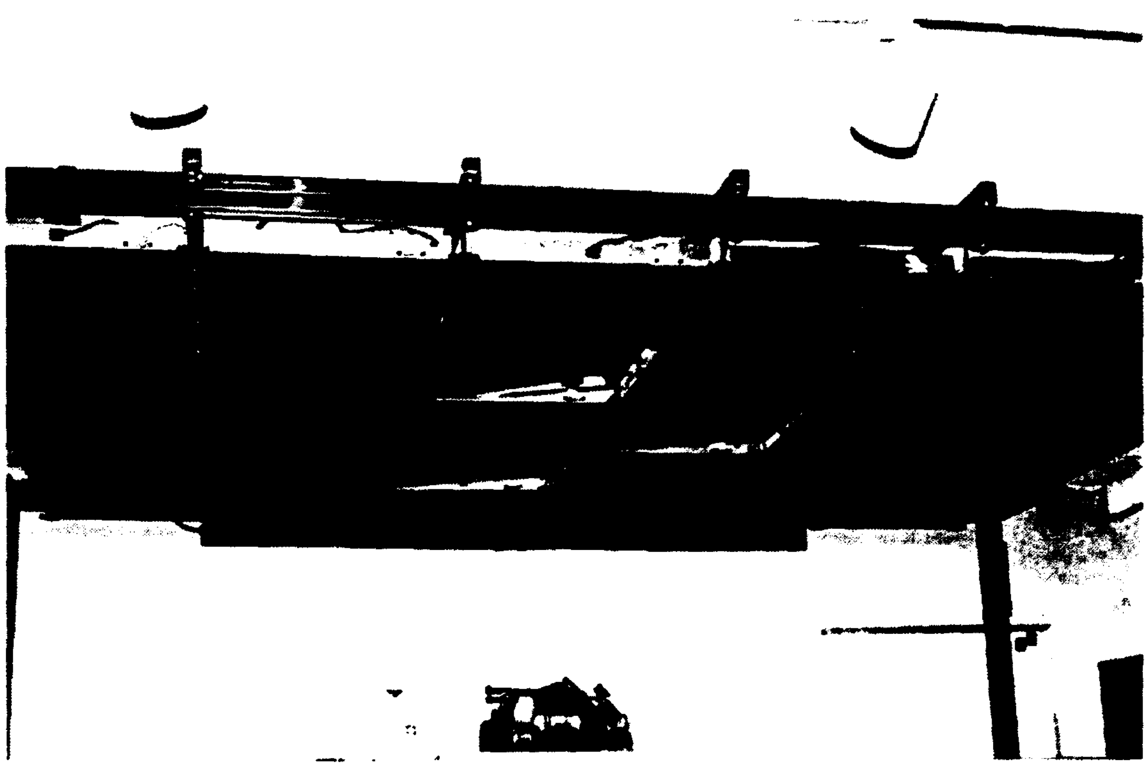


Figure 4. The existing optical ceiling with three panels deliberately tilted to test our calibration algorithm.

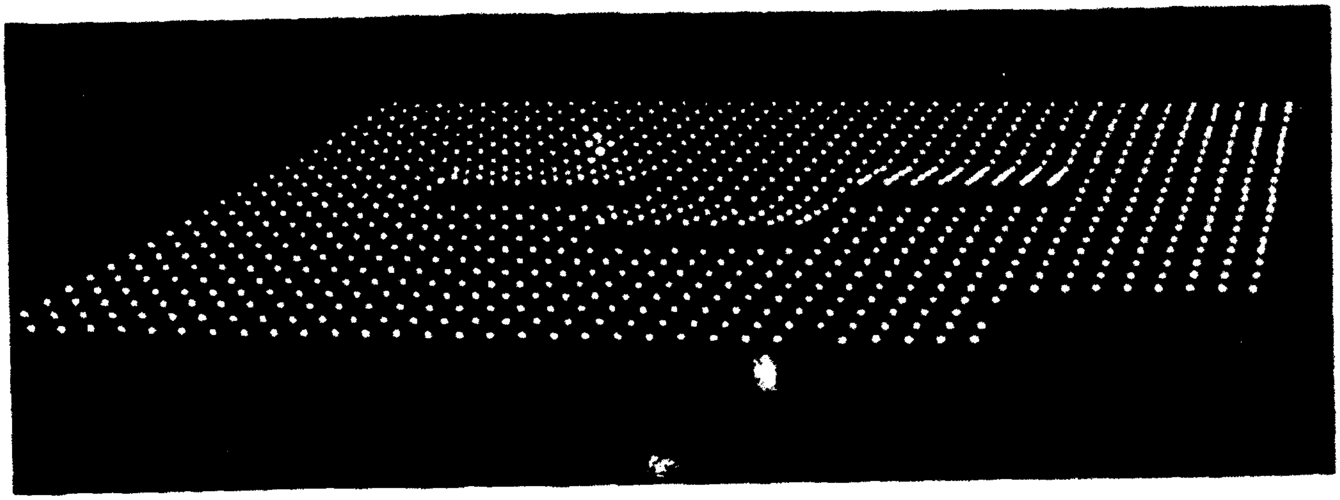


Figure 5. Result of calibration: a picture of the measured 3D positions of all the LEDs in the ceiling. The three tilted panels are clearly evident.

Alternate Mathematical Methods

Space-resection by collinearity [Azuma91] has served us well, but it is computationally-intensive and does not always converge to the correct solution. Two professors in the department, Dinesh Manocha and Jack Goldfeather, are interested in working with us to find direct, noniterative solutions that outperform the current method. Such methods may be inspired by relative motion techniques used in the computer vision field.

In order to improve the accuracy of our tracking, we intend to explore methods that take head motion into account. One alternative is an extension to the present method. Right now, space-resection by collinearity computes a zero-order approximation to the HMD's motion—it finds the position and orientation most consistent with the sensor data. The extension would provide a first-order approximation to the HMD motion—finding the position and orientation *and their rates of change*, which is most consistent to the sensor data.

Another alternative makes use of our ability to get accurate timestamps of when each LED is sampled by applying a Kalman filter to the sequence of sensor data. This method associates with each moment in time a state vector which includes position, orientation, and their first and second derivatives. Since Kalman filters allow scalar measurement updates, it is possible to update the state after sampling a *single* LED, rather than the dozen or so that we currently use before each update. Because sampling an LED takes about one millisecond, this method could potentially reach update rates of up to 1000 Hz. Latency might also be significantly reduced. The main drawback is that this technique cannot double-check its computed answers, so it will have to occasionally run space-resection by collinearity or some other method to insure that it does not drift off course. These alternative methods should also improve the accuracy of ceiling LED calibration.

Predictive Tracking with Inertial Sensors

Latency in virtual environment systems is a major problem. The effects are far more severe in see-through systems than in closed-view systems. In see-through, we try to register virtual objects with the real world. If the virtual objects move with some delay but the real world responds with no delay, however, this registration is destroyed. Even tiny amounts of latency can cause large registration errors. At a rapid head rotation rate of 300 degrees per second, keeping angular registration errors below 0.5 degrees requires a system latency of less than 2 milliseconds!

While system latency can be reduced, it is not currently practical to reduce latency below the threshold of detectability for environments of any significant complexity. Just updating a display at 60 Hz adds 16.67 ms to the system latency. The end-to-end delay from the time our ceiling tracker begins a measurement to the time when the last pixel has been drawn on both images in the ceiling HMD is about 120 ms for a simple environment.

Predictive tracking methods offer the possibility of compensating for much of this delay. Accurate prediction requires a good *understanding* of time in the system. You can't succeed if you don't know how *far* in the future to predict. In an asynchronous system such as ours, this prediction distance varies from iteration to iteration. We have synchronized the clocks in the ceiling tracker with its graphics engine, Pixel-Planes 5, completely bypassing UNIX and all of its non-real-time baggage. Synchronization is accurate to under one millisecond. Communication is done via shared memory. Because the clocks run at slightly different speeds on the two systems, we found that we must resynchronize every two or three minutes.

Accurate prediction requires good sensors. Besides the tracker itself, we will include velocity and acceleration information from inertial sensors. We have acquired rate gyros and linear accelerometers, plugged them into a 16-bit A/D card in an IBM PC, and verified that they work. Mounting rigs were designed and built to attach them to the ceiling HMD or any other HMD. We are currently building the supporting electronics to interface them with the ceiling tracker's A/D equipment.

At the moment, we have a simple prediction method based on a Kalman filter that runs on the ceiling. With the see-through camera activated in one eye, we noticed that this prediction did seem to reduce latency effects but generated other artifacts (such as overshoot). In the next several months, we intend to explore more advanced prediction methods that use inertial sensor information. Also, we must look for ways to evaluate and compare the effectiveness of prediction methods, both in objective terms, such as how much it reduces the average misregistration, and in psychological terms. The latter will require user studies.

We are also interested in using inertial sensors to improve the tracking performance of the ceiling. A hybrid optical-inertial tracker, as discussed earlier, offers many potential advantages. While our gyros are probably good enough to support this, our accelerometers and supporting electronics fall far below the required accuracy specifications. The existing A/D converters, which are accurate to less than 12 bits, will not be sufficient. If using inertial sensors to support delay compensation is successful, then we hope to make a concerted effort in building or acquiring a more accurate inertial tracker to explore the potential of a hybrid system.

Evaluation of Tracker Delays

Measurements have been made of the end-to-end delays in head-mounted display systems currently in use at UNC-Chapel Hill. System components that were tested include the various tracking devices currently in use: the Polhemus 3 Space, the Polhemus FasTrak, the Ascension Bird, the Ascension Flock of Birds, and the UNC Optical Ceiling. These were run on a Sun 4 host computer with our Pixel-Planes 5 graphics engine and some typical display devices (e.g., the

LCD/CRT monitors in use in head-mounted display systems). Significant findings include the reduced delays in certain commercial tracking systems (under 10 ms) and the unanticipated impact of the slow speed of LCD devices (adding as much as a full field time to the system end-to-end delays). We are in the process of writing these results into a technical report.

Tracker Distortion Testing

We attempted to characterize and compare the distortion of commercial magnetic tracking systems due to ambient magnetic interference. Tests consisted of mounting a magnetic sensor on top of a movable tripod and moving it around in the tracker workspace. The reported position values were recorded to a file which was then used to generate the position of spheres in a virtual environment. This in effect yielded a "smoke trail" of the reported tracker positions. The smoke trails of the various tracking systems were then compared visually in a virtual environment. Results showed that the Ascension Technology trackers are substantially less susceptible to distortion of the magnetic field due to the presence of conductors in the workspace than the Polhemus trackers are.

Noise Sources

We collected and analyzed the raw data obtained from the photodiode sensors to identify any noise sources that may be affecting our readings. The data was also analyzed to study the effects of the active gain control done by our software to vary the current used to light each LED and the time-variation of the intensity of the light emitted by the LEDs over the period of each sampling. The results of these analyses showed that our data acquisition is relatively noise free. We did remove some noise from our signals by replacing the power supply for the A/D circuitry.

Self-Tracker

Self-Tracker research has been hindered by the departure from our department of the graduate student who was working on the project. We are seeking a replacement but have been unsuccessful to date. Dr. Gary Bishop, who originated Self-Tracker research with his 1984 thesis [Bishop84], hopes to guide a new student in the continuation of this research.

4.4 Interactive Building Walkthrough

The UNC Interactive Building Walkthrough subproject aims at the development of systems to help architects and their clients explore a building design prior to its construction and, therefore, correct problems on the computer instead of in concrete [Brooks86]. In previous versions of the system, the scenes were shown with radiosity illumination, lights that could be switched on or off with only 100 msec delays, textured surfaces, and near real-time display from changing viewpoints. Update rates realized by the Pixel-Planes 5 graphics scene generator are 30–40 updates/second (15–20 stereo images/second) on scenes consisting of thousands of polygons, many of which are textured. The only downside to the system was that there was a large up-front cost to build and hand tune models and to calculate the radiosity illumination for the scene—the latter cost measured in hours to days. If the user wanted to change the geometry of the model, then the turn-around time was usually measured in days or weeks.

Recently we have moved toward creating a more automatic and interactive system in which changes in the model geometry result in delays of seconds to minutes rather than hours to days, and the radiosity illumination solution is computed in near-real time while the user is viewing the model. The end result is a much more practical and flexible design environment.

Our system uses a head-mounted display and the UNC ceiling tracker to allow the viewer literally to walk freely about a 10' x 12' area of a model. An orange "ribbon" demarcates the boundaries of the tracking volume.

We have long held that virtual-worlds projects need to move forward simultaneously in four dimensions: faster, prettier, handier, and realer. In the past year, we have made progress in the following areas:

• ***Faster:***
We have worked on speeding up the batch preparation processes rather than the run-time calculations. Although initial modeling will still take place off-line, changes to the model, mesh generation, and radiosity are all done interactively. In the next six months we are hoping to extract partitioning information automatically from the new version of Virtus WalkThrough.

• ***Prettier:***
We have begun working on incorporating texture into interactive radiosity.

• ***Handier:***
The big change in this system has been the method of moving the ceiling's virtual space through larger models. The viewer translates the 10' x 12' area in the direction of gaze with the press of a button. Our research suggests that only translating this area, rather than allowing translations and rotations, is less disorienting to the viewer.

We have moved toward a more automatic system of modeling and rendering which can be inserted without significant effort into existing applications and can be used to view existing models not specifically designed for the Walkthrough system.

We have significantly reduced the design iteration cycle time by migrating onto the graphics processors procedures that were previously hand tuned and treated as batch processes.

• ***Realer:***
Real-time mesh generation and interactive radiosity have been incorporated into the system and, as a result, models are beginning to be dynamic entities. The time scale for the modeling pipeline has dropped from days or weeks to seconds or minutes. Changes in the model are no longer limited to corrections of modeling errors; changes in design are encouraged rather than prohibited.

As a product of our collaboration with the Virtus Corporation, we are fast approaching a system which allows for interactive design, with a user under the UNC ceiling tracker making design decisions and a modeler using Virtus WalkThrough to carry them out. Soon the user will be able to explore a model, manipulate objects, suggest modifications, and have those modifications implemented in the model around him/her.

4.5 Virtual-Environment Ultrasound Scanning System

Our current system works in the following way: A video see-through HMD system is used to display ultrasound data in its real-world context with the patient on the examining table. A TV camera is mounted on a conventional HMD to provide video see-through capability. Synthetic images are generated which correspond to the position and orientation of the TV camera as tracked by the HMD system. The synthetic video images and the live (TV camera) images are then combined in a single image. This creates a virtual environment in which synthetic geometry appears fixed in the user's "real world" environment as that user, wearing the HMD, moves around.

The virtual-environment ultrasound scanning system works by gathering ultrasound data and transforming it to match the viewing position of the TV camera mounted on the video see-through HMD. When viewed through the HMD, the transformed ultrasound data appears fixed in its world, its 3D location within a subject. Two-dimensional ultrasound data is input into our system with a frame grabber and a tracking system. Currently we use Polhemus, Ascension and UNC Optical trackers. The frame grabber transfers live 2D ultrasound video images to our image-generation system, Pixel-Planes 5. The system resamples 2D images into a 3D volume that is rendered by Pixel-Planes 5 and then displayed along with live video in the HMD.

The ideal ultrasound system would include interactive display of volume data in a virtual environment and interactive acquisition of 3D ultrasound volumes. We are working along several paths to reach this goal. Current limitations include poor HMD resolution, poor tracking technology, slow rendering speeds, and the lack of a 3D ultrasound machine. UNC is actively pursuing improved HMD and tracking technologies. The ultrasound group is working on automatic calibration of the real world and ultrasound data with the virtual environment. We are also working on improving rendering speed, image quality, and the generation of 3D volumes from a series of tracked ultrasound images.

Ultrasound was chosen as the acquisition medium because it provides real-time data acquisition under the physician's control. Ultrasound scanners are compact and readily available. Furthermore, ultrasound technology is not invasive and is much safer than other imaging technologies such as X-ray. Our collaborators at Duke University and other researchers are working on ultrasound scanners that acquire volume data in real time.

Volume rendering was deemed necessary because the data from ultrasound is inherently noisy. When the 3D ultrasound machine being developed at Duke University becomes available, 3D data can be input directly into our system and the resampling of 2D data will no longer be required. To visualize 3D structures within the body and to allow the user to view these structures from arbitrary directions, we apply volume rendering within the head mounted display. The HMD gives the user the freedom to move around the patient, and the volume data shows the entire body structures and not just cross-sections of them.

4.6 The Nanomanipulator

We are collaborating with Prof. Stan Williams of the UCLA Department of Chemistry to build a nanomanipulator. The nanomanipulator consists of Williams' ultra-low-drift scanning tunneling microscope (STM) driven in real time by our manipulator arm, while the viewer watches real-time reconstruction of the surface, generated on Pixel-Planes 5 and viewed either on a large screen in high resolution or on the head-mounted display. We started this activity in November 1991, got the STM here in Chapel Hill in January 1992, and now have viewing, feeling, and arm-driven STM placement working. The system is also capable of making controlled changes to a surface in real time through the use of bias pulses.

We will attempt pick-up and release of small amounts of matter by short pulses of voltage about five times the normal imaging voltage. Our hope is to etch quantum circuits using this feature. Already we have been successful in placing material on the surface with this technique. Williams has been working hard on this and believes he can accomplish it.

This work is making use of head-mounted display technology developed under this DARPA contract. Our facilities for haptic display with the force-feedback manipulator were developed with support of NIH. An NSF grant under the Small Grants for Exploratory Research (SGER) program has paid for moving the STM here from UCLA and for the salary of the one graduate student currently working on this project.

References:

[Azuma91] Azuma, Ronald and Mark Ward. Space-resection by collinearity: mathematics behind the optical ceiling head-tracker. University of North Carolina at Chapel Hill Department of Computer Science technical report TR 91-048, November 1991.

[Bishop84] Bishop, Gary and Henry Fuchs. "The Self-Tracker: A Smart Optical Sensor on Silicon," *Proceedings of the 1984 MITT Conference on Advanced Research in VLSI* (Dedham, MA: Artech House, January 1984), pp. 65-73.

[Brooks86] Brooks, Jr., F. P. "Walkthrough - A dynamic Graphics System for Simulating Virtual Buildings," *1986 Workshop on Interactive 3D Graphics*, S. Pizer and F. Crow, eds., University of North Carolina at Chapel Hill, October 1986, pp. 9-22.

[Ward92] Ward, Mark, Ronald Azuma, Robert Bennett, Stefan Gottschalk, Henry Fuchs. "A Demonstrated Optical Tracker With Scalable Work Area for Head-Mounted Display Systems." *Proceedings of 1992 Symposium on Interactive 3D Graphics--A special issue of Computer Graphics* (Cambridge, MA, 29 March - 1 April 1992), 43-52.

5.0 Dissemination of Research

5.1 Publications

Bajura, Mike, Henry Fuchs, Ryutarou Ohbuchi. "Merging Virtual Reality with the Real World: Seeing Ultrasound Imagery within the Patient." *Proceedings of SIGGRAPH '92*, (published as *Computer Graphics* 26, no. 2, July 1992, 203–210).

Barrett, H.H., J. Yao, and J.P. Rolland. "Applications of the Hotelling Observer in Medical Imaging." Abstract published in *Proceedings of the Optical Society of America General Meeting*, New Mexico, September 1992.

Fuchs, Henry, Gary Bishop, et al. "Research Directions in Virtual Environments: Report of an Invitational Workshop on the Future of Virtual Environments." *Computer Graphics* 26, no. 3 (August 1992), 153–177. (This workshop, funded by an NSF grant, was held at UNC-Chapel Hill 23–24 March 1992.)

Robinett, Warren, Russell Taylor, Vernon Chi, William V. Wright, Frederick P. Brooks, Jr., R. Stanley Williams, Eric Snyder. "The Nanomanipulator: An Atomic-Scale Teleoperator." Printed in ACM SIGGRAPH'92 Course Notes for the course "Implementation of Immersive Virtual Environments."

Rolland, J.P., and C. Burbeck, "Depth and Size Perception in Virtual Environments." Abstract published in *Proceedings of the Optical Society of America General Meeting*, New Mexico, September 1992.

5.2 Submitted for Publication

Azuma, Ronald. "Head Tracking for Alternate Realities." Submitted for publication in a special issue of *Communications of the ACM* on Embodied Reality and Ubiquitous Computing.

Gottschalk, Stefan, and John F. Hughes (Brown University). "Autocalibration for Virtual Environments Tracking Hardware." Submitted to the SIGGRAPH'93 conference.

Mine, Mark, and Gary Bishop. "Just-in-Time Pixels." Submitted to the SIGGRAPH'93 conference.

Yoo, Terry S., and T. Marc Olano. "Instant Hole: Window onto Reality." Submitted for publication in a special issue of *Communications of the ACM* on Embodied Reality and Ubiquitous Computing.

Taylor, Russell M., Warren Robinett, Vernon L. Chi, Frederick P. Brooks, Jr., William V. Wright, R. Stanley Williams, Erik J. Snyder. "The Nanomanipulator: A Virtual-Reality Interface for a Scanning Tunneling Microscope." Submitted to the SIGGRAPH '93 conference.

5.3 Presentations

By: Mike Bajura
Topic: Merging Virtual Reality with the Real World: Seeing Ultrasound Imagery within the Patient
Event: SIGGRAPH'92 Conference
Place: Chicago, IL
Date: 30 July 1992

By: Gary Bishop
Topic: Head-mounted Display Research at UNC-CH
Event: IBM Training Day for Employees Working in Network Systems
Place: IBM, Research Triangle Park, NC
Date: 3 December 1992

By: Fred Brooks
Topic: Does Virtual Reality Have Any Real Virtue?
Event: Panel at 1992 Visualization in Biomedical Computing conference
Place: Chapel Hill, NC
Date: 14 October 1992

By: Fred Brooks
Topic: Virtual Worlds Research at UNC-Chapel Hill
Event: Invited talk, Workshop on Virtual Reality/Virtual Environments in Army Training
Place: Durham, NC
Date: 28 October 1992

By: Fred Brooks
Topic: Virtual Worlds Research at UNC-Chapel Hill
Event: Invited talk, Lectures in Computer Graphics, Geometry, and Visualization series
Place: Princeton University, Princeton, NJ
Date: 11 November 1992

By: Jeff Butterworth, Stefan Gottschalk, and Russell Taylor
Topic: Head-mounted Display Research at UNC-CH
Event: Lecture Series on Virtual Reality at University of North Carolina at Greensboro
Place: Greensboro, NC
Date: September 1992

By: Henry Fuchs
Topic: Virtual Reality: The Lunatic Fringe or the Ultimate Display
Event: Invited talk, Lectures in Computer Graphics, Geometry, and Visualization series
Place: Princeton, NJ
Date: 23 November 1992

By: Jannick Rolland
Topic: Applications of the Hotelling Observer in Medical Imaging
Event: Invited talk, Optical Society of America general meeting
Place: New Mexico
Date: September 1992

By: Jannick Rolland
Topic: Depth and Size Perception in Virtual Environments
Event: Invited speaker, Optical Society of America general meeting
Place: New Mexico
Date: September 1992

By: Jannick Rolland
Topic: Depth and Size Perception in Virtual Environments
Event: Visit to Philips Medical
Place: Best, The Netherlands
Date: September 1992

5.4 A Partial List of the Visitors Who Observed Graphics Demonstrations, June 1992–January 1993

- Delegation of 13 Hungarian participants in the National Academy of Sciences–Hungarian Academy of Sciences workshop on "Technological Innovation and Growth in Knowledge-Based Economies"
- Kurt Akeley, Silicon Graphics
- 75–90 participants in the "Virtual Reality and Synthetic Environments in Training" conference, sponsored by the Army Research Office in Research Triangle Park, NC
- Professor Dr. sc. Rainer Ortleb, Federal Minister of Education and Science and Deputy Chairman of the Liberal Democratic Party, Germany
- Hans R. Friedrich, Chief of Staff, Ministry of Education and Science's Department of Higher Education and Science, Germany
- ~200 attendees of the 1992 "Visualization in Biomedical Computing" conference that was hosted by our department in October 1992
- Col. John Mentz, Office of the U.S. Secretary of Defense
- David Meisel and Henry Sowizral, Boeing
- Håkon Lie, Norwegian Telecom Research
- 20 members of NC State University's student chapter of ACM
- 9 juniors and seniors in a multimedia class, LeJeune High School, Fayetteville, NC
- Ase Svensson, Klas Odelid, and Sverker Almquist, Lund University, Sweden
- Dr. Brad Walters, UNC Neurosurgery
- Mike McGrath, National Science Foundation

- ~15 8th–12th graders from the Duke University Talent Identification Program's "Visual Math" weekend session
- Donald Boulton, UNC Vice-Chancellor and Dean, Division of Student Affairs, and 17 of his department heads
- Mark Mabry, Lockheed Sanders
- 12 from National Institute of Environment Health Sciences, Research Triangle Park, NC
- Nicolas Ayache, INRIA, France
- 4 from Motorola
- 5 from Virtus Corp., Cary, NC
- Group of 12 computer science majors from Univ. of South Carolina at Spartanburg
- Barbara Herrnstein Smith, Director of the Duke University Center for Interdisciplinary Study in Science and Cultural Theory
- Alan Cox, Honeywell
- Chris Daft, General Electric
- Group of 5 faculty and students from East Tennessee State University Department of Computer Science, Johnson City, TN
- 15 from Glaxo Pharmaceuticals
- 5 from Information Spectrum, Inc. (designers of flight simulations for the U.S. Navy)
- Pete Saraceni, FAA
- Prof. Peter Foldiak, Psychology Department, Oxford University, UK
- David Johnson, NASA Langley, Hampton, VA
- David Stein, Director of NC School of Math and Science, Durham, NC
- 7 from the Florida Community Colleges Academic Technology Committee
- Richard Bland and Cathy Czeto, Lighthouse Low Vision Products
- Mike Levine, Henry Green and J. Pekar, Ocutech (optical company designing products for persons with low vision), Chapel Hill, NC
- Jim Drew, Discovery Place (science museum), Charlotte, NC
- Harry Yae and Mike Booth, University of Iowa, Dept. of Mechanical Engineering
- 13 undergraduates from the Duke Engineering Research Center's Summer Institute
- 16 gifted and talented high school students participating in the Math and Science Ventures Program sponsored by UNC-Charlotte

5.5 Media Coverage of UNC-CH's Research

There is continuing public interest in our department's research accomplishments, particularly virtual worlds research. During this reporting period, information on our research has appeared in a wide variety of media: television, video shown at conferences and exhibitions, newspapers, and magazines.

Television/Video:

The UNC News Service highlighted the department's research in virtual reality as one of only three areas of research included in a new video shown during half-time at televised UNC athletic events.

Video footage of UNC's virtual worlds research was shown at a 10-day exhibition called "Virtual Vertigo" in La Coruña, Spain, and was subsequently shown as part of a documentary on this exhibition for a Spanish television news program.

Video footage of our research was shown at the CNN World Economic Development Congress that was held in Washington, DC, 17–20 September 1992.

WAGA-TV, the CBS affiliate in Atlanta, Georgia, shot footage in the graphics lab and interviewed Gary Bishop for a three-part story on virtual reality that was broadcast on their 6:00 news program 23–25 November 1992.

Newspaper/Magazine Articles:

"Virtual Reality: It's All in the Mind," *Atlanta Constitution*, 29 September 1992.
"Virtual Reality: Exploring the Future," *Furniture Today: The Weekly Business Newspaper of the Furniture Industry*, 5 October 1992. Focuses on architectural walkthrough.
"Virtual Reality: How a Computer-Generated World Could Change the Real World," *Business Week*, 5 October 1992. Includes a quote from Henry Fuchs on the future of virtual reality.
"See-Through View: Virtual Reality May Guide Physicians' Hands," *Scientific American*, September 1992. About the use of head-mounted display in ultrasound exam.

6.0 Appendixes

Appendix A

Fuchs, Henry, Gary Bishop, et al. "Research Directions in Virtual Environments: Report of an Invitational Workshop on the Future of Virtual Environments." *Computer Graphics* 26, no. 3 (August 1992), 153–177.

Appendix B

Bajura, Mike, Henry Fuchs, Ryutarou Ohbuchi. "Merging Virtual Reality with the Real World: Seeing Ultrasound Imagery within the Patient." *Proceedings of SIGGRAPH '92*, (published as *Computer Graphics* 26, no. 2, July 1992, 203–210).

Research Directions in Virtual Environments

An Invitational Workshop on the Future of Virtual Environments

Research Sponsored by the National Science Foundation

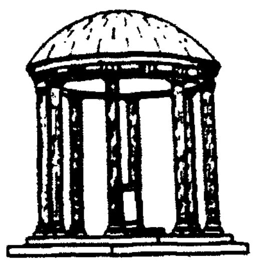
TR92-027

March, 1992

Henry Fuchs, Gary Bishop

Principal Investigators

William Bricken, Frederick P. Brooks, Jr.,
Marcus Brown, Chris Burbeck,
Nat Durlach, Steve Ellis,
Mark Green, James Lackner,
Michael McNeill, Michael Moshell,
Randy Pausch, Warren Robinett,
Mandayam Srinivasan, Ivan Sutherland,
Dick Urban, Elizabeth Wenzel



The University of North Carolina at Chapel Hill
Department of Computer Science
CB#3175, Sitterson Hall
Chapel Hill, NC 27599-3175

Published in ACM SIGGRAPH - Computer Graphics, August 1992, Vol. 26, No. 3, pp. 153-177. The workshop was supported by grant IRI-9213777 from the National Science Foundation.

Research Directions in Virtual Environments

Report of a workshop sponsored by the Interactive Systems Program, National Science Foundation

March 23-24, 1992

University of North Carolina at Chapel Hill

Gary Bishop, UNC-Chapel Hill (co-chair)
William Bricken, U. of Washington, Seattle
Frederick Brooks, Jr., UNC-Chapel Hill
Marcus Brown, U. of Alabama, Tuscaloosa
Chris Burbeck, UNC-Chapel Hill
Nat Durlach, M. I. T.
Steve Ellis, NASA-Ames Research Center
Henry Fuchs, UNC-Chapel Hill (co-chair)
Mark Green, U. of Alberta, Canada
James Lackner, Brandeis University
Michael McNeill, NCSA
Michael Moshell, U. of Central Florida
Randy Pausch, U. of Virginia, Charlottesville
Warren Robinett, UNC-Chapel Hill
Mandayam Srinivasan, M. I. T.
Ivan Sutherland, Sun Microsystems
Dick Urban, DARPA
Elizabeth Wenzel, NASA-Ames Research Center

TABLE OF CONTENTS

Executive Summary

At the request of NSF's Interactive Systems Program, a two-day invitational workshop was held March 23-24, 1992 at UNC Chapel Hill to identify and recommend future research directions in the area of "virtual environments" (VE) *. Workshop participants included some 18 experts (plus 4 NSF officials) from universities, industry, and other leading technical organizations. The two-day schedule alternated between sessions of the entire group and sessions in the following specialty areas, around which the recommendations came to be organized: 1) Perception, 2) Human-Machine Software Interface, 3) Software, 4) Hardware, and 5) Applications. Also, two participants developed a taxonomy of VE applications that is included as an appendix to the report.

Recommendations Summary:

Perception:

Vision

1. Collaborative science-technology development programs should be established at several sites around the country to encourage closer collaboration between developers and scientists.
2. Theoretical research should focus on development of metrics of performance and task demands in VE.
3. Paradigmatic applications and theoretical questions that illustrate the science-technology synergy need identification.

Audition

Spatial Sound

1. Theoretical research should emphasize the role of individual differences in Head-Related Transfer Functions (HRTFs), critical cues for distance and externalization, spectral cues for enhancing elevation and disambiguating the cone-of-confusion, head-motion, and intersensory interaction and adaptation in the accurate perception of virtual acoustic sources. The notion of artificially enhanced localization cues is also a promising area.
2. A fruitful area for joint basic and applied research is the development of perceptually-viable methods of simplifying the synthesis technique to maximize the efficiency of algorithms for complex room modeling.
3. Future effort should still be devoted to developing more realistic models of acoustic environments with implementation on more powerful hardware platforms.

Nonspeech Audio

1. Theoretical research should focus on lower-level sensory and higher-level cognitive determinants of acoustic perceptual organization, with particular emphasis on how acoustic parameters interact to determine the identification, segregation, and localization of multiple, simultaneous sources.
2. Technology development should focus on hardware and software systems specifically aimed at real-time generation and control for acoustic information display.

Haptics

1. Development should be encouraged of a variety of computer-controlled mechanical devices for either basic scientific investigation of the human haptic system or to serve as haptic interfaces for virtual environments and teleoperation.
2. Research programs should be initiated to encourage collaboration among engineers who are capable of building high precision robotic devices and scientists who can conduct biomechanical and perceptual experiments with the devices.
3. Research programs should also be developed to enable collaboration among researchers working on visual, auditory, and haptic interfaces, together with computer specialists who can develop software capable of synchronized handling of all the sensory and motor modalities.

Motion Sickness in Virtual Environments

1. The virtual environment community should be made aware of the sensory-motor adaptation and motion sickness problems to be expected presently because of hardware limitations and in the future as better virtual presence in nauseogenic environments is achieved.
2. Research programs should be initiated to evaluate the incidence and severity of sickness associated with different types of virtual environments, and to assess the kinds of sensory-motor adaptations and aftereffects associated with virtual

---

*By *virtual environments*, we mean real-time interactive graphics with three-dimensional models, when combined with a display technology that gives the user immersion in the model world and direct manipulation. Such research has proceeded under many labels: *virtual reality, synthetic experience*, . etc. We prefer virtual environments for accuracy of description and truth in advertising.

environments.

Evaluation of Virtual Environments
Research should be conducted on the development of psychophysical techniques that measure the level of effort required to achieve a given level of performance, that relate performance on simple tasks with performance in a multi-task situation, and that operate in a systematic and well-defined manner with complex stimulus contexts.

Human-Computer Software Interface:
1. Researchers should focus on the development of new metaphors for VEs and the identification of reusable, application-independent interface components, specifically those which can be encapsulated in software and distributed.
2. NSF should support a software clearinghouse for code sharing, reuse, and software capitalization.
3. We will need to develop metrics to guide the exploration of VE tools, techniques, and metaphors.

Software:
1. The development of new modeling tools for model construction for virtual environments should be supported, especially inside-the-environment modeling tools. These tools need to be developed to the point where their effectiveness can be evaluated..
2. A facility for sharing existing and new models should be established.

Hardware:
Tracking Systems
1. Inertial tracking systems are prime for research activity now because of recent advances in micro-accelerometers and gyros. Inertial adjuncts to other tracking methods for sensing of motion derivatives is also a needed research activity.
2. Research into tracking technologies that allow large working volumes in outside spaces should be encouraged.

Haptic Systems
1. Support basic biomechanical and psycho-physical research on human haptic senses.
2. Support development of interactive force reflecting devices, and devices to distribute forces spatially and temporally within each of the (possibly multiple) contact regions.

Image Generators
1. Research into low latency rendering architectures should be encouraged.
2. Research is needed.into software techniques for motion prediction to overcome inherent system latencies and the errors they produce in registered see-through applications.

Visual Display Devices
NSF should primarily support pilot projects that offer potential for order of magnitude improvement in resolution, brightness and speed. NSF should also investigate display techniques that may offer decreases in latency and to characterize problems with display phenomena such as frame sequential color.

Applications:
1. Applications are needed which provide discriminatory power to evaluate VE technology versus 'through the window' interactive graphics and other similar technologies.
2. Researchers should look toward applications which solve real-world problems. VE must move beyond the stage of an interesting te[illegible]ological toy and begin to solve problems for people where they are.
3. Researchers [illegible] begin work on the probable impact of VE technology on society: Will VEs change the way we work (telecom[illegible] teleconferencing) or our interpersonal interactions? As the technology becomes more readily available, how will society react?
4. Can the use of VEs to communicate between people approach the level of communication we currently experience in person or in a group? What research must be done to move toward that goal? Is it even a desirable goal?

I. Introduction

What is Virtual Environments Research? In 1965, Ivan Sutherland in a paper, "The Ultimate Display", given at the triennial conference of the International Federation of Information Processing Societies, proclaimed a program of research in computer graphics which has challenged and guided the field ever since. One must look at the display screen, he said, as a window through which one beholds a *virtual world*. The challenge to computer graphics is to make the picture in the window look real, sound real, and the objects act real. Indeed, in the ultimate display, one will not look at that world through a window, but will be immersed in it, will change viewpoint by natural motions of head and body, and will interact directly and naturally with the objects in the world, hearing and feeling them, as well as seeing them.

Real-time interactive graphics with three-dimensional models, when combined with a display technology that gives the user immersion in the model world and direct manipulation, we call *virtual environments*. Such research has proceeded under many labels: *virtual reality, synthetic experience*, .etc. We prefer virtual environments for accuracy of description and truth in advertising. Merriam-Webster's New Collegiate Dictionary, Ninth Edition, defines

virtual as "being in effect but not in actual fact", and

environment as "the conditions, circumstances, and influences surrounding and affecting an organism".

Why is VE Research hot now? From 1965 until the mid-1980's, the limited power of computers and of graphical engines meant that Sutherland's vision could only be realized for crude depictions or for painfully slow interactions for many worlds. Many graphics researchers worked on making more faithful visual depictions by solving the problems of perspective, hiding, raster-scanning pictures, shading, or illumination. They got fidelity of motion by animation onto film, computing minutes per frame, giving up interaction. Others worked on real-time motions and interactions in toy worlds of only a few hundred elements.

Advances in technology, in computer and graphics organization, in displays, and in interactive devices now enable us to do in a video frame tasks that used to require batch computing. Digital signal processing algorithms and hardware allow the realistic production of three-dimensional sound cues, and increasingly compact and high performance mechanical sensors and actuators promise realistic simulation of manual interactions with objects. So it is now possible to bring these lines of research together and to approximate Sutherland's vision of interestingly complex worlds with rather good pictures, sounds, and forces, with tantalizingly close to real-time performance.

Though we still have far to go to achieve "The Ultimate Display", we have sufficiently advanced towards the goal that is timely to consider real systems for useful applications:

- What are the characteristics of the applications that will most benefit from such man-machine systems?
- What are the technical barriers that stand in the way of these applications?
- How can these most profitably be addressed? How can NSF (or DARPA) and the VE research community make a coordinated push through these barriers?

II. Overview

In light of the recent surge of interest in Virtual Environments in science, industry, and the media, an invitational workshop was held at the University of North Carolina at Chapel Hill on March 23-24, 1992, at the request of Dr. John Hestenes (Director, Interactive Systems, National Science Foundation). The workshop was chaired by Drs.. Gary Bishop and Henry Fuchs with the purpose of developing recommendations for research directions in this field. Eighteen researchers from the US and Canada spent two days in large and small groups developing a consensus on the recommendations in this report.

The participants divided into five working groups in order to focus on:

1. Perception (chaired by Steve Ellis),
2. Human-Computer Software Interface (chaired by Randy Pausch),
3. Software (chaired by Mark Green),
4. Hardware (chaired by Michael Moshell),and
5. Applications (chaired by Marcus Brown).

Also, two participants, Ivan Sutherland and Warren Robinett, developed a taxonomy of VE applications that is included as an appendix.

The recommendations of each of the groups were reviewed and discussed by all of the participants.

This report summarizes the results of the workshop. These results are organized around the five divisions of the working groups. Each section presents the current status of the sub-area, the perceived needs, and recommendations for future research directions.

III. Perception

Vision

Because of the pervasive, dominant role of vision in human affairs, visual stimuli are without question the most important component in the creation of the computer-based illusion that users are in a virtual environment. There are four aspects of this key role of vision: the characteristics of the visual image, the structure of the visual scene, the visual consequences of manipulative and vehicular interaction with the scene, and the role of visual information for spatial orientation.

Status

Visual image Modern visual psychophysics makes intensive use of computer graphics to synthesize high resolution stimuli for experimental manipulation. Display generation and digital filtering techniques have come to play an essential role in modern laboratories studying human vision. The mathematical and computational techniques used to describe the visual stimuli that are studied have also become the languages in which theories about visual phenomena are phrased (Watson, 1989)

Visual scene Structure in the visual image is automatically identified by biological image processing that segregates foreground from background and spontaneously groups regions together into subparts. Some aspects of this image segregation appear to be the result of parallel processing while other show evidence of sequential processing (Treisman, 1985). Once segregated, the contours and features collected into groups may be interpreted as objects in the space surrounding the observer. The separated patterns of contours and regions may then be interpreted as a surrounding space.

Visual world The spatial interpretation of visual images is highly dependent upon the kinematic characteristics of the image motion, in particular those motions that are consequences of the observer himself (Cutting, 1986). The patterns of image motion that are associated with observers' movements provide much of the necessary information for guidance through a cluttered environment and have provided the basis for development of what J. J. Gibson described as a higher-order psychophysics. In this field, researchers may investigate the natural linkages established between properties of image, or object motion, and complex normal behaviors such a walking or object avoidance.

Just as motion of an observer causes global changes in the pattern of relative motion in the visual image, so to manipulative interaction with visible objects also produces characteristic visible transformations related to the object's position and identity, (e.g. Warren, et al, 1991), which have been extensively studied to provide the bases for psychological and physiological theories of manipulative interaction.

Visual orientation Visual information is not only important for local navigation while traversing an environment but also for global path planning and route selection. These more global tasks have been studied in isolation during scientifically motivated experiments (e.g. in Howard, 1982). But visual orientation is also important for more integrated tasks in which subjects use visual aids such as maps to maintain their internal representation of the surrounding space and assist planning of future activities.

Needs

Visual image Precision visual tasks will require improvements in the image quality of small display systems that provide photopic luminance levels with several arc-minute pixel resolution. Low level visual performance should be assessed with visual parameters likely to be provided by future display systems which may use nonstandard pixel layouts, variable field resolution, and field magnification to optimize allocation of computer graphics processing. Higher resolution inserts in the central visual field may be utilized, but gaze directed control of these fields may not be necessary if they can be made sufficiently large, i.e. to about 30 degrees. Since the presentation of wide fields of view (> 60 degrees monocular), will likely involve some geometric image distortion, studies of the tolerable distortion and characteristics of adaptation will also likely be required for specific tasks. However, because the binocular overlap between the left and right eye images need not be complete, monocular fields exceeding 60! may only rarely be required.

Visual scene Since virtual environment will only be able to present somewhat degraded low level visual cues such as contrast and stereopsis, the capacity for viewers to segregate foreground from background is likely to be less than that with natural images from real environments. Accordingly, visual segregation with degraded image quality and dynamics should be studied and enhancements to overcome difficulties should be developed.

Visual consequences The visual consequences of environmental interactions generally involve intersensory integration and do not quality as strictly visual issues. However there are purely visual consequences of motion in a simulation which are important for perceptual fidelity: a compelling visual simulation will require dynamic as well as kinematic modeling which currently is difficult to carry out at the necessary interactive rates, which ideally should exceed 30 Hz simulation loop frequency. Important work is required on the subjective and objective operator reactions to approximated kinematic and dynamic models of synthetic environments. How far can a simulation deviate from correct dynamical modeling and still appear to be realistic?

Visual orientation Imperfect and slow dynamics of virtual environments can lead to significant difficulties for users to maintain their spatial orientation within a simulated larger environment. Orientation aids to compensate for these difficulties should be developed to allow developers to simulate highly detailed real environments when such detailed simulation is required. These aids amount to enhancements for orienteering within a virtual environment and should assist users in switching between ego- and exocentric frames of reference which will be needed for efficient interpretation and control of objects in the simulated environment.

Recommendations

1. Collaborative science-technology development programs should to be established at several sites around the country to encourage closer collaboration between developers and scientists.

2. Theoretical research should focus on development of metrics of performance and task demands in VE.

3. Paradigmatic applications and theoretical questions that illustrate the science-technology synergy need identification.

Comment: The inherent interdisciplinary nature of VE will benefit from curriculum modifications to improve communication between perceptual scientists and interface designers. Currently, these researchers have significantly different research agenda and goals which can interfere with collaboration. Interface designers are happy with informal, imperfect guidance not the relative truth which the scientists seek.

Audition

Status

Two general areas of acoustic research, spatial sound and the real-time generation of nonspeech audio cues, are critical for virtual environment research and technology development. Speech generation and recognition, also important features of auditory displays, will not be discussed here.

Spatial Sound The simulation of spatial localization cues for interactive, virtual acoustic displays has received the most attention in recent work. Perceptual research suggests that synthesis of purely anechoic signals can result in perceptual errors, in particular, increases in front-back reversals, decreased elevation accuracy, and failures of externalization. These errors tend to be exacerbated when virtual sources are generated from non-personalized Head-Related Transfer Functions, a common circumstance for most virtual displays. In general, the synthesis technique involves the digital generation of stimuli using Head-Related Transfer Functions (HRTFs) measured in the ear canals of individual subjects or artificial heads for a large number of real source (loudspeakers) locations (e.g., Wightman & Kistler, 1989; Wenzel, 1992). Other research suggests that such errors may be mitigated by providing more complex acoustic cues derived from reverberant environments (Begault, 1991). Recently, some progress has been made in interactively synthesizing complex acoustic cues using a real-time implementation of the image model (Foster, et al., 1991)

Nonspeech Audio Following from Gibson's ecological approach to perception, one can conceive of the audible world as a collection of acoustic "objects". In addition to spatial location, various acoustic features such as temporal onsets and offsets, timbre, pitch, intensity, and rhythm, can specify the identities of the objects and convey meaning about discrete events or ongoing actions in the world and their relationships to one another.

One can systematically manipulate these features, effectively creating an auditory symbology which operates on a continuum from "literal" everyday sounds to a completely abstract mapping of statistical data into sound parameters. Principles for design and synthesis can be gleaned from the fields of music (Blattner, Sumikawa, and Greenberg, 1989), psychoacoustics (Patterson, 1982), and higher-level cognitive studies of the acoustical determinants of perceptual organization (Bregman, 1990; Buxton, Gaver, and Bly, 1989). Recently, a few studies have also been concerned with methods for directly characterizing and modeling environmental sounds such as walking sounds (Li, Logan, and Pastore, 1991). Other relevant research includes physically or structurally-based acoustic models of sound source characteristics such as radiation patterns (Morse and Ingard, 1968).

Needs

Spatial Sound It seems clear that simple anechoic simulations of spatial cues will not be sufficient to minimize perceptual errors and maximize perceptual "presence". Dynamic modeling of complex acoustic environments requires enormous computational resources for real-time implementation in a truly interactive (head-tracked) display. Currently it is not practical to render more than the first one or two reflections from a very small number of reflecting surfaces in real-time. However, because of the less stringent requirements of the auditory modality, acoustic digital signal processing is now advanced enough to allow significant strides in our basic understanding of human sound localization. While fully realistic, interactive simulations of a concert hall may not yet be feasible, synthesis techniques are sufficiently developed to allow an unprecedented degree of stimulus control for the purposes of psychophysical studies.

Nonspeech Audio A few cue-generation systems have been specifically integrated for virtual environment applications while some designers are beginning to develop systems intended for data "sonification" However, far more effort should be devoted to the development of sound-generation technology specifically aimed at information display. Perhaps more critical is the need for further research into lower-level sensory and higher-level cognitive determinants of acoustic perceptual organization, since these results will serve to guide technology development. Further, relatively little research has been concerned with how various acoustic parameters interact to determine the identification, segregation, and localization of multiple, simultaneous sources. Understanding of such interaction effects will be critical in any acoustic display developed for both virtual environments and telepresence.

Recommendations

Spatial Sound

1. Theoretical research should emphasize the role of individual differences in HRTFs, critical cues for distance and externalization, spectral cues for enhancing elevation and disambiguating the cone-of-confusion, head-motion, and intersensory interaction and adaptation in the accurate perception of virtual acoustic sources (see Wenzel, 1992). The notion of super-auditory localization, or artificially enhanced localization cues, is also a promising area (Durlach, 1991).

2. A fruitful area for joint basic and applied research is the development of perceptually-viable methods of simplifying the synthesis technique with the goal of maximizing the efficiency of algorithms for complex room modeling (increasing the number and complexity of modeled reflections).

3. In contrast to visual display technology, we are currently much closer to developing truly realistic simulations of auditory environments. Since the research pay-off is likely to be both high and timely, future effort should still be devoted to developing more realistic models of acoustic environments with implementation on more powerful hardware platforms. Some of the issues that need to be addressed are nonuniform radiators, diffuse reflections, scattering reflectors, diffraction and partial obscuration by walls or other objects, spreading loss and high-frequency absorption.

Nonspeech Audio

1. Theoretical research should focus on lower-level sensory and higher-level cognitive determinants of acoustic perceptual organization, with particular emphasis on how acoustic parameters interact to determine the identification, segregation, and localization of multiple, simultaneous sources.

2. Technology development should focus on hardware and software systems specifically aimed at real-time generation and control for acoustic information display, using basic theoretical knowledge as design guidelines.

Haptics

The human haptic system is composed of subsystems that enable tactile and kinesthetic senses as well as motor actions. In contrast to the purely sensory nature of vision and audition, only the haptic system is capable of direct action on real or virtual environments. Being able to touch, feel, and manipulate objects in the environment, in addition to seeing (and/or hearing) them, gives a sense of compelling immersion in the environment that is otherwise not possible. It is quite likely that much greater immersion can be achieved by the synchronous operation of even a simple haptic interface with a visual display, than by large improvements in the fidelity of the visual display alone. Consequently, it is important to develop a wide variety of haptic interfaces to interact with virtual environments. Examples of haptic interfaces that are being used in virtual environment research are joysticks and hand/arm exoskeletons. In general, they measure and display users' body part positions as well as the forces on them. The biomechanical, sensorimotor, and cognitive abilities of the human haptic system determine the design specifications for the hardware and software of haptic interfaces.

Status

Compared to vision and audition, our understanding of haptics is very limited. The recent trend of conducting multidisciplinary haptic studies involving biomechanical, psychophysical and neurophysiological experiments together with computational models has contributed to rapid progress. Due to the availability of powerful computers and high precision mechanical sensors and actuators, it is now possible to exert control over experimental variables as never before.

Biomechanics of Contact: In any task involving physical contact with an object, be it for exploration or manipulation, the mechanics of the contact interface plays a fundamental role. Both in the tactile sensory information flow from the object, and in imposing desired motor action on the object, the contact interface is strongly influenced by the surface and volumetric physical properties of the skin and subcutaneous tissues. Although some data on the in vivo biomechanical properties and several simple computational models of, say, the primate fingerpad are available, they are inadequate at present.

Sensing and Control of Interface Variables: The term interface variables is meant to include the kinematic variables (i.e., the relative positions, orientations, and motions) of various body parts, together with the associated normal and shear forces arising from contact with objects. The kinematic information is conveyed by our kinesthetic sense, whereas the contact forces are sensed by both tactile and kinesthetic systems. In the kinesthetic space, psychophysical phenomena such as anisotropies in the perception of distance and orientation, apparent curvature of straight lines, non-Euclidean distance measures between two points etc., have been reported. In tasks involving physical contact, such as manipulation of objects, it is known that tactile information is crucial in controlling grasp forces. These control actions can range from a fast spinal reflex to a relatively slow, deliberate action. However, even simple questions concerning our abilities, (such as what is our resolution in the sensing and control of interface variables), or the mechanisms, (such as how we perceive joint angles or contact forces), do not yet have unequivocal answers.

Perception of Contact Conditions and Object Properties: The contact conditions perceived through the tactual sense can be broadly classified as static (with or without skin stretch due to shear forces), slipping and vibratory. The object properties that are inferred include both geometric (such as shape), and material properties (such as compliance). The perception of both contact conditions and object properties is based on intensive, temporal, spatial or spatio-temporal stimulus variations, and the associated neural codes. Recent psychophysical and neurophysiological investigations have provided some answers to questions concerning perception and neural coding of roughness, raised features on rigid objects, slip, microtexture, shape, compliance, etc. However, the important connection between the loads imposed on the skin surface within the regions of contact with objects and the corresponding perception has only begun to be addressed.

Needs

In order to enunciate the specifications for the design of haptic interfaces, performance of the human haptic system should be characterized. This includes the determination of (a) the biomechanical properties of skin and subcutaneous soft tissues that govern the mechanics of contact with the interface; (b) the abilities of the human haptic system and the strategies used by human subjects in performing haptic tasks; and (c) evaluation of the effectiveness of haptic interfaces. A major barrier to progress from the perspectives of biomechanics, neuroscience and psychophysics has been the lack of robotic stimulators capable of delivering stimuli under sufficiently precise motion and force control.

Biomechanical Investigations: The tight mechanical coupling between the human skin and haptic interfaces strongly influences the effectiveness of the interface. Therefore, the specifications for the design of sensors and actuators in the interface, as well as the control algorithms that drive the interface, require the determination of surface and bulk properties of, say, the fingerpad. The measurement of force distributions within the contact regions with real objects is needed to determine how a display should be driven to simulate such contacts in virtual environments. In addition, computational models of the mechanical behavior of soft tissues will aid in simulating the dynamics of task performance for testing control algorithms, as well as in determining the required task-specific force distributions for the displays. This requires measurement of the in vivo skin and subcutaneous soft tissue response to time-varying normal and tangential loads.

Psychophysical Investigations: Determination of the human abilities (in terms of resolution, bandwidth, etc.) in sensing and control of net contact force vectors as well as joint angles or end point positions will set the design specifications for haptic interface devices. The rather large body of data available on tactile sensing of vibratory stimuli and the data on spatial localization and resolution, together with additional psychophysical experiments on the perception of contact conditions and object properties will influence directly the design of tactile displays. Theoretical developments concerning the task-specific flow of sensory information and control of motor action are needed to generate testable hypotheses on our haptic interactions with both real and virtual environments.

Haptic Interface Evaluation: Although the haptic interfaces available at present are quite limited in their capabilities, they need to be evaluated from the perspective of human perception. For example, a force-reflecting joystick attached to the floor can be called a grounded display, whereas a force-reflecting exo-skeletal device attached to the user's forearm would be an ungrounded display. (It would, in fact, be grounded at the forearm). The grounding choice affects whether or not the user experiences throughout his/her entire body the stresses induced by contact with a virtual object. The consequences of using an ungrounded display to simulate contact forces which really stem from grounded sources are not known and warrant investigation. Furthermore, the fidelity with which the tactual images have to be displayed and the motor actions have to be sensed by the interface depends on the task, stimulation of other sensory modalities, and interaction between the modalities. Experimenting with the available haptic interfaces, in conjunction with visual and auditory interfaces, helps to identify the necessary design improvements.

Recommendations

1. Development of a variety of computer-controlled mechanical devices for either basic scientific investigation of the human haptic system or to serve as

haptic interfaces for virtual environments and teleoperation should be encouraged. The biomechanical and psychophysical research work detailed in the *Needs* section above should be supported.

2. Research programs should be initiated to encourage collaboration among engineers who are capable of building high precision robotic devices and scientists who can conduct biomechanical and perceptual experiments with the devices.

3. Research programs should also be developed to enable collaboration among researchers working on visual, auditory, and haptic interfaces, together with computer specialists who can develop software capable of synchronized handling of all the sensory and motor modalities.

Motion Sickness in Virtual Environments

Status

Human movement and locomotion involve a dynamic sensory-motor adaptation to the 1G background force of Earth. Disruptions of this relationship, which depends on correlations among patterns of sensory feedback from vision, touch, somatosensation, proprioception, and the semicircular canals and otolith organs of the labyrinth with motor information about ongoing movements, leads to a variety of perceptual and motor errors and often motion sickness as well. Errors and sickness persist until adaptation to the new situation is achieved. Different types of rearrangements differ in terms of how difficult it is to adapt to them and for some situations, adaptation cannot be achieved. There also are great individual differences in terms of susceptibility to motion sickness, ability to adapt, rate of adaptation and retention of adaptation.

Rearrangements can be of many types, including for example, delays in feedback loops as encountered in flight simulators or teleoperator systems, various types of visual or auditory position displacements, alterations of the effective inertial mass of the head or body such as are brought about by wearing headgear or a space suit, changes in the motor commands necessary to achieve movement such as in rotating environments or altered gravitoinertial force fields. Artificial transport of the body in vehicles - cars, trains, boats, aircraft, spaceships - leads to complex alterations in sensory-motor control which typically elicit performance errors and motion sickness until adaptation is achieved. Interestingly, sickness also develops during experienced motion of the body when the patterns of feedback present during voluntary movements are appropriate for the actual rather than experienced motion. This highlights the fact that no general theory exists which predicts which situations will be disruptive, nauseogenic, or difficult to adapt to, or which individuals will be most prone to these problems.

It can be expected, however, that sickness will be a problem both in the early development stages of virtual environments and once they have been perfected. In fact, reports of motion sickness are becoming commonplace. It can be fully expected that as sickness brought about by low resolution displays, visual lags, and improper force feedback disappears because of technical improvements, sickness will be more frequent because of the improved ability to create virtual environments involving "vehicle" motions which are naturally nauseogenic.

Needs

Motion sickness is going to be a highly significant problem in virtual environments. Moreover, it must be recognized that motion sickness is a complex syndrome. Nausea and vomiting represent only one part of motion sickness and often not the most crucial aspect of sickness. Other symptoms include headache, dizziness, eye strain, lethargy, and fatigue. Often a distinction is made between "gut" and "head" symptoms.

Under laboratory conditions, motion sickness is relatively easy to recognize. Experimental studies of motion sickness typically involve highly provocative test situations which rarely last more than an hour or two. Thus, sickness is expected and the personnel carrying out the experiments are highly skilled in recognizing signs and symptoms of acute motion sickness. Often the subjects are trained in identifying the subjective concomitants of sickness as well. Motion sickness is more difficult to recognize under operational conditions. Such conditions tend to be less provocative and to bring on initially more head symptoms than gut ones. A sailor may not realize that his experience of drowsiness is an early sign of developing motion sickness. Highly motivated individuals may be able to "tough out" head symptoms in order to complete a work schedule, but ultimately the generation of work schedules will be constrained.

An attempt should be made to familiarize investigators in the virtual environment area with the primary characteristics of motion sickness. In addition, it would be useful for records to be kept concerning the incidence and characteristics of sickness encountered with different test platforms. Other characteristics of the users, such as motion sickness histories, should also be gathered. In all likelihood, research programs involving the experimental evaluation of sickness with different types of dynamic virtual environments will be necessary as well as programs to study ways of enhancing the rate of acquiring adaptation and of enhancing retention of adaptation. Relevant here, too, will be the assessment of the consequences of adapting to virtual environments with regard to sensory-motor performance on return to the normal environment. We anticipate that functionally significant disruptions of performance will be associated with adaptation to dynamic virtual environments. Their severity, as well as ways of eliminating them, need to be addressed.

Recommendations

1. The virtual environment community should be made aware of the sensory-motor adaptation and motion sickness problems to be expected presently because of hardware limitations and in the future as better virtual presence in nauseogenic environments is achieved.

2. Research programs should be initiated to evaluate the incidence and severity of sickness associated with different types of virtual environments.

3. Research programs should be developed to assess the kinds of sensory-motor adaptations and aftereffects associated with virtual environments.

Virtual Environments for Perception Research

Status

Scientists using computer-generated stimuli and computer-based signal processing have been using the elements of virtual environment technology for years as part of the normal course of research. Computer processing, especially graphics, is regularly used to synthesize sensory images and present experimental stimuli. Though the most obvious presentations have involved visual stimuli, auditory, haptic, and vestibular stimuli are also commonly presented. Typically, these experimental applications are not focused on synthesis of multisensory environments. They are, however, necessarily carefully calibrated and often automated for the conduct of an experiment. Digital signal processing is also extensively used to filter and analyze performance data (See Elkind, Card, and Hochberg, 1989).

Virtual environment (VE) technology is inherently an interdisciplinary field that will be likely to integrate the previous scientific work in perception and manual control (Ellis, 1991; Durlach et al, 1992). There is consequently a great potential benefit from collaboration between perceptual/motor researchers and interface developers. VE can produce physically unrealizable stimulus environments that are uniquely useful for testing theories of perception and manual control, but good/standard calibration techniques are needed to avoid unwitting investigation of display artifacts.

Needs

The natural course of this research will require the development of appropriate evaluation metrics and calibration procedures to be suggested for use by VE system developers. Though anecdotes from successful users indicate significant benefits from the technology, existing VE systems need to be generally more analytically evaluated for specific utility.

VE development forces evaluation of assumptions/better design of tests and can force psycnological and physiological theories to be more precisely stated and integrated and can enable the asking of totally new experimental questions.

Evaluation of Virtual Environments

Status and Needs

VE systems can be meaningfully evaluated only by determining their effects on the human user. No other measure matters. Rigorous, reliable and interpretable measures of these effects require careful and exact control of the stimulus and use of well-designed psychophysical procedures. Although there is a strong well-established tradition of sensory research that conforms to these requirements, evaluation of VE systems presents demands that cannot all be met by existing techniques. New procedures must be developed.

Many of the well-developed experimental techniques focus on the detection of small differences in the stimulus, e.g. in contrast detection, spatial discriminations, motion detection and discrimination and depth discriminations. Among the psychophysical methods currently in use, those that use high contrast stimuli are most likely to be generally useful. Motion detection and velocity discrimination, for example, may prove to be useful indicators of the perceptual stability of the system. Similarly, size discrimination may prove to be a useful indicator of the quality of the depth percept.

Also of interest is the veridicality of the simulated percept. Subjective judgments of absolute size or depth may be useful, but rating judgments are highly context-dependent. They may miss substantial differences between real and virtual environments because each is judged only relative to itself. See-through head-mounted displays that permit combination of virtual and real environments offer a means of determining the veridicality of the simulation by direct comparison to the real environment.

However useful these techniques, none addresses directly the question of how demanding the system is to use. Psychophysical experiments are marked by a strong restriction of task and stimuli. Virtual environments, on the other hand, typically present complex stimuli to observers who make complex judgments. Breaking these complexities down into simple components that are amenable to traditional psychophysical evaluation cannot answer many of the most important questions about the value of the virtual environment because performance depends on how information is acquired in the face of the complexity.

On the other hand, use of complex stimuli can make interpretation of the data difficult. The observer may use stimulus cues other than the one being manipulated intentionally. This problem can be avoided by manipulating only a single localized feature and letting the rest of the stimulus serve as context. In that case, a large number of different contexts or environments must be used to prevent the observer from learning to ignore the context. Use of multiple varied contexts inevitably adds noise to the measured response, requiring more trials, which requires more contexts, etc. Soon the number of images to be stored becomes large enough to be problematical. Restricting the contexts to a class that is appropriate to a single application reduces the variability at the cost of generality. Techniques need to be developed for systematically generating and describing contexts to be used in well controlled studies. A generally accessible library of contexts might add order to this pursuit.

At least as important as stimulus complexity is the problem of task complexity. The accuracy with which an observer can perform a single task tells us little about the resources required to support that performance. For virtual environments to be useful, they cannot place unreasonable demands on the user. If it is difficult or taxing for the observer to collect each desired piece of information, the system may not enhance performance of complex tasks. The information presented in the VE must not only be available, it must be easily available.

Most psychophysical methods do not deal with this aspect of the problem. Some that do exploit the temporal dimension. Reaction times may be useful indicators of the difficulty of a task. This measure has the notable advantage that the task being studied need not be limited to detecting barely discriminable stimulus differences. The major disadvantage to this technique is the need to handle speed-accuracy tradeoffs. Another approach is to limit the exposure time of the stimulus. This avoids the problem of speed-accuracy tradeoffs, and can result in a more systematic representation of the speed of processing. The major weakness of the approach is that for all but contrast detection tasks, a masking stimulus must be presented at termination of the stimulus to stop further processing. Choice of an effective mask is often difficult and because one cannot know whether a given mask is completely effective for a given stimulus, comparison of effects across stimuli is never wholly convincing. Improvements in these techniques and/or creation of new techniques that make use of the temporal dimension to assess task difficulty would be helpful.

More generally, we need to develop ways to assess behavior when the observer is performing more complex tasks, e.g., when he is performing several simple tasks simultaneously or in rapid alternation. Some techniques that have been used in other areas of experimental psychology, e.g., in memory research, may be usefully adopted here.

Recommendation

Research should be conducted on the development of psychophysical techniques that measure the level of effort required to achieve a given level of performance, that relate performance on simple tasks with performance in a multi-task situation, and that operate in a systematic and well-defined manner with complex stimulus contexts.

IV. Human-Computer Software Interface

Status

The interaction techniques for virtual environments have yet to be extensively explored. There are several design principles which are likely candidates for structuring VE interface metaphors, such as:

- natural behavior as an interaction principle
- rapid prototyping for design space exploration
- knowledge-based agents for interactive guidance
- respect for the physiological context of the physical body in VE design
- supernormal capabilities as a metaphor for sensory remapping

These principles are suggestive, but need to be evaluated in task specific application domains.

Current difficulties in the design and construction of virtual worlds include:

- ineffective software, particularly for programming dynamics and interactivity;
- complex hardware configurations and especially hardware latency;
- world modeling and maintenance (version control on worlds is complex);
- no theory of world construction;
- physical literalism (assuming the virtual world is like the physical);
- naive models of interaction in a virtual environment; and
- failure of intuition (physical solutions are not virtual solutions).

Current research challenges include:

- developing software operating systems which facilitate effective, inclusive, real-time interaction, multiple participant, multiple sensory, high bandwidth, low latency VEs;
- determining which aspects of the 2D WIMP and desktop metaphors are generalizable to higher dimensional virtual worlds;

- identifying the core of generic virtual world design tools and application independent interaction techniques;
- integration of multidisciplinary teams to address human, software, and hardware interactions, including design and analysis of experiments; and
- bringing multidisciplinary knowledge to the construction of VEs, including work from database, data fusion, networking, human factors, computer supported collaborative work, and artificial intelligence communities.

Needs

The most pressing needs are:

- software for design of and interaction with virtual worlds that is modular, flexible, and abstract, particularly interpreted languages for rapid prototyping;
- software operating systems and infrastructure to support world design, construction, and interaction, particularly software which reduces latency;
- metaphors which guide the exploration and prototyping of particular tools and techniques for use in VEs;
- measurement techniques and theories for identifying differential effects of world designs on the sense of presence;
- measurement techniques for identifying resource expenditure, cognitive load, transfer of learning, adaptation effects, and other performance parameters of different configuration of VEs; and
- task-specific evaluation of software tools.

Secondary, less general, needs include the development of:

- navigation techniques for high dimensional data and displays;
- location and movement techniques and software tools;
- manipulation techniques and software tools;
- event history, filtering, and recording tools;
- behavioral and dynamic models for determining the dispositions of objects;
- specification languages for world dynamics;
- editing tools for objects and for environmental spaces, including models of inter-object communication, process management, and composition rules;
- a design theory of sensory presentation modes (which sensory suites are best for conveying which tasks?);
- languages and grammars for describing world interactions;
- the virtual body, mapping tools for connecting sensors on the physical body to an accurate physiological model in software, and in turn, to the virtual object being used as a body in the VE; and
- tools that can be used for interaction and construction both inside the VE and outside on a monitor viewing the VE.

Recommendations

Since VE design and interaction is in its infancy, these recommendations are focused on generic rather than specific goals

1. Isolating and evaluating application-independent interaction techniques and metaphors. Researchers should focus on the development of new metaphors for VEs and the identification of reusable, application-independent interface components, specifically those which can be encapsulated in software and distributed. One specific area with high potential is the use of voice input as a parallel input modality.

   While some of this evaluation will be extensive research centered on the human's capabilities, some of it will be rapid, less formal evaluation to help interface designers choose between conflicting alternatives. In one sense, these different objectives underscore the differences between basic science and engineering. We explicitly suggest that NSF recognize the contributions made by evaluations made at various levels of certainty. Fred Brooks refers to these as *certainty shells*, including findings, observations, and rules-of-thumb. This research needs to integrate the diverse skills and styles of multidisciplinary teams.

2. Software capitalization — NSF should support a software clearinghouse for code sharing, reuse, and software capitalization. The cost of having each VE laboratory develop its own infrastructure is prohibitive to the effective conduct of research. We encourage support of world building and maintenance tools, to enable version control, composition of components developed by different design groups (tool portability), ease of customization and configurability, and expressability.

3. Measurement techniques to determine the quality of VEs We will need to develop metrics to guide the exploration of VE tools, techniques, and metaphors. The quality of a VE is likely to be related to specific tasks, physiological comfort, cognitive and perceptual load, resource expenditure, intuitiveness of tools and interactions, ease of reconfiguration, transfer and adaptation effects, and even individual differences and preferences. In this regard, our final, concrete recommendation is that NSF help forge inter-disciplinary collaborations, especially between computer scientists and psychologists.

V. Software

Status

The cost effective development of virtual environments will require a new generation of software tools. These tools must support the special hardware devices that are used in virtual environments and the design of the environments themselves.

A number of software tools have been developed to support the production of virtual environments. These tools can be divided into two broad groups based on whether they are commercial products or the result of a research project.

Most of the commercial products support a particular hardware configuration that is sold by the software vendor. They provide a basic level of software support for the vendor's hardware and allow for the development of simple virtual environments. Most of these tools provide basic support for the hardware devices and limited support for the development of the virtual environments themselves.

The research software tools tend to be more general, since there is no mandate to support a particular hardware platform. These tools support a range of hardware configurations, which facilitates the sharing of research results and the resulting program code. Again, most of these tools only address the device support issues, but in a hardware independent manner. Some tools are beginning to be developed for supporting the production of virtual environments, in addition to supporting the basic hardware. One of the major advantages of this group of software tools is that they allow new researchers to quickly enter the field in a cost effective manner, since they are freely distributed and will likely support whatever hardware configuration the researcher has available.

Needs

Modeling. Modeling is the key software issue at the present time. The modeling problem is not unique to virtual environments, but it is crucial to the success of many VE applications. In many VE applications, there is a tight connection between the user and the underlying model. The user must be able to move through the model, interact with the objects in the model, and change the model interactively. The development of good modeling tools will facilitate the development of new applications and the development of new techniques.

In this report, modeling refers to the data structures that are used to record the geometrical information for the environment. This information includes the shape of the objects in the environment, their moving parts and physical properties, and the behaviors that they can perform (how they interact with other objects in the environment and with the user). This information is not only used to produce the visual presentation of the environment, but it is also used in sound production and tactile feedback.

In a VE application there may be multiple users for the same application. The users could have different needs or be performing different tasks, which would result in a different view of the model for each user. In addition, each user may have personal preferences that dictate a different view of the model. The modeling software most be able to support the different views of the users.

There are also applications where multiple users may be in the same environment at the same time. This implies that the model must be shared among the users. Each user must have a consistent view of the application (which may depend upon how the model is viewed). There must be methods for keeping these models up-to-date and at the same time allow the users to modify individual copies of the model.

The model will be used to drive several different media such as graphics and sound. The model must contain the information required by all of these media, plus the information required to synchronize their presentation. This is not as simple as events in two media occurring at the same time. For example, in the case of lightning, the visual appearance of the lightning must occur before its sound is produced. This problem also occurs in multi-media systems.

Objects in a VE may correspond to objects that occur in the real world. When this happens, we expect these objects to behave in the same way as real world objects behave. In other words, they must follow the traditional laws of physics. When an object is dropped, it should fall until it reaches the ground. It shouldn't be possible for objects to pass through each other. In order to get these effects we need to be able to model at least part of physics. This includes the familiar laws of motion and the ability to detect collisions between objects. These techniques must perform in real-time, so they can be incorporated into the feedback provided by the environment.

Some of the objects in the VE must be able to respond to the user's actions and the actions of other objects. There should be ways of specifying this behavior and assigning these behaviors to the objects in the model. Some of the objects in the environment may also have autonomous behaviors, such as walking. This will involve interactions between the objects and simulated time within the environment. There is some overlap between the techniques required here and the techniques that are being developed in computer animation and simulation.

One of the issues in modeling for VE is whether the modeling should be done inside or outside of the environment. There are clear advantages to both approaches. Modeling in the environment gives the designer a good view of what the user will see when he or she is in the environment. Modeling outside the environment allows the designer to use multiple views of the environment and draw upon the modeling techniques that have been developed over the past few decades.

Current inside-the-environment modeling systems are not as good as the outside-the-environment ones. There are two main reasons for this. First, the I/O devices used in VE are not as well developed as the devices used in traditional interactive computer graphics. For example, head-mounted displays have a much lower resolution than CRT monitors, and most 3D tracking technologies have higher noise levels than 2D devices. Second, the design techniques used in the outside-the-environment modelers have been developed over the past few decades, while there has been very little experience with inside-the-environment modelers. There should be an effort to develop design techniques for inside-the-environment modelers.

One of the main problems with large models is navigating in them. In 2D applications, overview maps of the model can be presented in separate screen windows, but similar techniques haven't been developed for 3D environments. It has been suggested that the environment can be scaled in order to give an overview of it. This will only work if detail can be culled from the model when it is scaled. In order to do this effectively, the model must be constructed in such a way that it is easy to identify the details and the important components of the model. The model should provide assistance to the navigation process.

Modeling is currently a time-consuming process. We need better tools to support this task, particularly modeling tools for VE applications. Reasonable modeling tools have been developed for other domains, most notably computer-aided design and product design. The objects in both of these domains are static, and thus these tools don't address the complete range of VE modeling requirements. An effort should be made to develop better modeling tools that address the problems of VE modeling. The development of these tools will assist in other domains that make heavy use of modeling.

Since the production of good models currently requires a considerable amount of effort, there should be some mechanism for sharing the existing models. There are two aspects to this sharing. First is the determination of a standard format for encoding and transmitting the models. There are numerous formats that have been developed for the interchange of graphical information. One of these formats may be suitable for this purpose, or a new format may need to be developed. The format that is used should be able to encode all the modeling information, including hierarchical structure, physical properties of the objects, sound and behavior. Second, there should be one or more sites that store, maintain and distribute these models. These sites would be responsible for collecting the existing models and new models as they are developed, cataloguing the set of models and distributing them to researchers on request. All of this should be performed over the existing computer networks to guarantee the widest distribution of the models.

There is a need for certain real-time functionality in VE systems. In particular, there is a need for low latency in the processing of requests, the ability to assign priorities to processes, and the management of time. Real-time operating systems exist that meet these requirements, but researchers haven't been willing to use them, since they would need to give up the facilities that are provided by existing main-line operating systems, such as UNIX. As a result, we would like to see these real-time facilities added to UNIX.

Our recommendation is not to directly fund this work under the VE initiative but to encourage other groups to fund this effort under larger projects that are also be interested in this technology. In particular, the high performance computing initiative should be encouraged to fund this work. Other areas that would benefit from this work include robotics and real-time simulation.

Recommendations

1. The development of new modeling tools that meet the requirements of model construction for virtual environments should be supported.

2. A facility for sharing existing and new models should be established. This will involve both the development of standards for model interchange, and the establishment of one or more sites where the models will be maintained.

3. Support the development of inside-the-environment modeling tools. These tools need to be developed to the point where their effectiveness can be evaluated.

VI. Hardware

Hardware for virtual environments is an area of rapid progress, for several reasons. Much of the technology has other applications outside of VE, in areas such as robotics, entertainment and instrumentation. Areas such as high definition display devices are being driven by the general advancement of semiconductor design and fabrication.

Nevertheless, a substantial number of "show stoppers" for VE can be found in hardware. For instance, without some breakthroughs in the bio-mechanical engineering of haptic (force and touch feedback) devices, many of the envisioned applications of VE will remain science fiction.

We will explore the following hardware categories: tracking systems, haptic systems, image generators, visual display devices, audio systems and speech input systems.

Tracking Systems:
This section concerns techniques for determining the location and orientation of the user's head, hands and ultimately of any real-world object of interest to the VE user.

Status.

Five categories of devices exist:

1. Magnetic systems, typified by the Polhemus and Ascension trackers, use magnetic fields emitted by small antennas and detected by one or more receiving antennas. They are unreliable in the presence of metal objects. Today's systems have a range of less than 1m from the emitter; multiple-emitter solutions are possible.

   Latencies (time from initiation of a motion until the new data arrives at the host computer via a parallel link) in current systems are on the order of 50 msec. Rumored improvements may bring this down to 5 msec within a year. Current systems can achieve resolutions of approximately 1mm and 0.03 degrees, and accuracies of approximately 3mm and 0.1 degrees.A wireless version of a magnetic tracker would be technically difficult.

2. Acoustic systems, typified by the Logitech Mouse, use ultrasonic pulses. Range and latency are approximately the same as today's magnetic devices, without susceptibility to magnetic interference. They are susceptible to line-of-sight occlusion, e. g. by the arm of the user. Wireless versions are feasible. Ultimate latency is limited by the speed of sound (about 0.3 m/msec). Accuracy is limited by variations of the speed of sound with changes in air density.

3. Inertial systems have formerly been bulky, expensive, and had too much drift. Advances in micromachines show promise of producing small, inexpensive accelerometers and rate gyros that have sufficient sensitivity to allow tracking via dead reckoning for short periods. Wireless operation would be feasible. These devices are not yet commercially available.

4. Among mechanical systems, the BOOM viewer from Fake Space Labs uses a rigid framework both to support the viewing device and measure position and orientation, with a reported accuracy of 4mm and 0.1 degree resolution at each joint. Because, unlike the magnetic and acoustical devices, no averaging is required, the latency is very small - under 1 msec. Low-cost, non-load-bearing mechanical trackers are also available.

5. Optical systems such as the Optotrack 3000 and the Selspot II determine position of targets via triangulation techniques from cameras at known locations. They determine orientation by observing multiple targets. Without large separations between targets, fine measurement of orientation is very difficult. These systems provide only a small working volume (1 meter cube) and are susceptible to line of sight occlusion.

   Optical systems such as the experimental tracker at UNC mount the cameras on the head and observe targets at fixed locations on the ceiling. This system determines both position and orientation directly and may be extended to cover arbitrary areas but must maintain line of sight and isn't directly suitable for tracking the hands.

Needs.

To support most goals currently under consideration, a position sensor would need to have these characteristics:
- wireless operation
- ability to track multiple users in same space without interference
- range of up to 10m x 10m x 5m with reference to a base unit, perhaps the size of a briefcase;
- no wide-area antenna, ceiling or sensor field required
- latency under 5 msec
- resolution of 1 mm and 0.01 degree
- accuracy of 1 mm and .01 degree in registered see-through applications, 1 cm and 0.1 degree in non-registered applications.
- sampling rate >= 60 Hz
- direct sensing of state AND derivatives

The private sector is making progress towards better tracking systems for VE though the technology hasn't changed significantly for more than 10 years.

Recommendations.

1. Inertial tracking systems are prime for research activity now because of recent advances in micro-accelerometers and gyros.

2. Inertial adjuncts to other tracking methods for sensing of motion derivatives is also a needed research activity.

3. Research into tracking technologies that allow large working volumes in outside spaces should be encouraged.

Haptic Systems:

This section concerns devices which provide the illusion of manual exploration or manipulation of objects in virtual worlds. These devices employ human tactile, kinesthetic and motor systems for interaction with the virtual environment. They perform some or all of two basic functions:

1. measurement of the positions and forces (and time derivatives) of the user's hand and/or other body parts to be used as control inputs to VE, and

2. display of forces and positions and/or their spatial and temporal distributions to the user.

Status. Four areas of work are of interest:

1. Hand position/joint angle measurement: Instrumented wands, joysticks, gloves and exoskeletons that measure hand position/joint angles are available in the market. The major problems are the intrusion the user feels while wearing, say, an exoskeleton, and the ever-present need for improvements in resolutions and ranges.

2. Application of forces and torques: Exoskeletons attempt to both measure hand/arm joint angles and to load these joints with feedback torques around the joint by applying forces at the contact regions. With inputs from research into the teleoperation of robots, some progress has been made on the exoskeletal problem, but the systems are heavy, expensive, clumsy and unreliable. The computational problem of compensating for the mass of the control arm is substantial.

   Simulator motion bases represent the most mature component of this sub-area. The art of providing sustained forces to the subject's whole body through a pilot's seat is well understood. The principal limitations are that the average acceleration provided by a terrestrial motion base must always be at least 1g, and that sustained accelerations greater than 1g are achievable only in centrifuges, which induce other severe problems such as coriolis effects.

3. Tactile Displays. These are two dimensional fields of force applied to the skin, to simulate touching. Both normal and shear forces are necessary to simulate general sensations of touch. The simulation of different coefficients of friction (for various materials and skin conditions: wet, dry, oily) is problematic. A few standalone demonstrations of tactile displays have been built.

4. Other stimulus distributions. These include thermal, several kinds of pain, and possible direct electrical stimuli applied as two dimensional fields to the skin. Pain is the report of a physiological limit's being reached or exceeded. "Deliberate synaesthesia" would use a small vibrating stimulator or small shock (rather than pain) to report that a "virtual limb" or teleoperation actuator device has reached its limit.

Needs.

Mechanical stability is important. Force feedback systems need to have vibration rigorously controlled, probably to less than a few microns amplitude, to prevent false cues. Forces on the order of 10 Newtons are needed, with 10 bit resolution. With respect to force distributions, a spatial density exceeding 1 mm/taxel and a temporal resolution approaching 1 kHz are needed.

Recommendations.

1. Support basic biomechanical and psycho-physical research on human haptic senses.

2. Support development of interactive force reflecting devices, and devices to distribute forces spatially and temporally within each of the (possibly multiple) contact regions.

Image Generators

Status.

The computer graphics industry is rapidly improving its polygon, lighting models and texture rendering capability, and the cost of visual systems continues to drop at a rate >50%/yr for constant performance.. The principal deficit for VE applications concerns latency - the delay between a change to the visual database or viewing parameters and a change in the display.

The latency of fast workstations such as Silicon Graphics products is one frame interval (33ms for 30Hz update rate) to compute the image in the offscreen buffer, plus one refresh interval (14ms at 72Hz refresh). The drive in the private sector towards higher through-put will not automatically solve the latency problem, as pipelined architectures will probably continue to be used. The VE community needs lower latency even at the expense of polygon through-put.

Needs.

Most graphics users do not need low latency (at the levels required by VE), and so most graphics architectures are deeply pipelined. Research in computer architecture is needed to explore ways to reduce latency.

There is no good data on the threshold for perception of latency; research is needed. Simulator sickness studies on the effect of latency have generally explored the domain between 100 and 300 msec, because of limitations in the imaging systems in use.

One way to determine this threshold would be with the use of a panoramic buffer. A pre-stored scene would be "panned" in response to head tracking information. The inherent latency of an NTSC display device is around 16 ms; in order to beat that rate, "racing the beam" would be necessary (i. e. the generation of scan lines immediately before their rendering, based on most current geometry).

The lowest latency VE systems likely to be built in the near future will still have delays from user motion to visual output that cause significant errors in registered see-through systems and perceptual/sickness effects in all VE systems.

Recommendations.

1. Research into low latency rendering architectures should be encouraged.
2. Research into software techniques for motion prediction to overcome inherent system latencies and the errors they produce in registered see-through applications is needed.

Visual Display Devices

Status.

Commercially available LCD displays are presently in the range of 200 x 300 pixels (consisting of a color triad). DARPA is funding development of a true 640 x 512 electroluminescent, and also a color LCD system consisting of color quads (1280 x 1024 in monochrome). The displays are to be ready within 2 years.

DARPA is also funding the development of a reflective system using very small deformable mirrors; around 10**6 moving mirrors. The device will have high speed (10 microsec) and power capabilities because all energy is reflected, not absorbed. Can be used with color by triplex or field seq color.

Optics for head mounted displays are available, but are heavy and expensive. Because the images in VE are totally synthetic, it is possible to perform some transformations (such as correction of some kinds of chromatic aberration) in software which formerly required optical methods.

Needs.

1024 x 1024, color, 60 Hz or faster, in 1" square packaging; lightweight, large visual field optics.

Recommendation

1. NSF should primarily support pilot projects that offer potential for order of magnitude improvement in resolution, brightness and speed.

2. NSF should also investigate display techniques that may offer decreases in latency and to characterize problems with display phenomena such as frame sequential color.

Audio Systems:

Status

Spatial Sound Briefly, the approach is to synthesize externalized, three-dimensional sound cues over headphones using very high-speed digital signal processing devices (see Wenzel, 1992). In general, the synthesis technique involves the digital generation of stimuli using Head-Related Transfer Functions (HRTFs) measured in the ear canals of individual subjects or artificial heads for a large number of real source (loudspeakers) locations (e.g., Wightman & Kistler, 1989). In most current systems (e.g., the Convolvotron), from one to four moving or static sources can be simulated (with varying degrees of fidelity) in a head-stable, anechoic environment by convolution of arbitrary signals with HRTF-based filters chosen according to the output of a head-tracking device. Motion trajectories and static locations at greater resolutions than the empirical data are generally simulated by linear interpolation between the measured impulse responses. Also, in some systems, a simple distance cue can be provided via real-time scaling of amplitude.

Nonspeech Audio The ideal synthesis device would be able to flexibly generate the entire continuum of nonspeech sounds described earlier as well as be able to continuously modulate various acoustic parameters associated with these sounds in real-time. Current devices available for generating nonspeech sounds are based almost exclusively on MIDI (Musical Instrument Digital Interface) technology and tend to fall into two general categories; "samplers", which digitally store sounds for later real-time playback, and "synthesizers", which rely on analog or digital sound generation techniques originally developed for imitating musical instruments. With samplers, many different sounds can be reproduced (nearly) exactly, but substantial effort and storage media are required for accurately pre-recording sounds and there is usually limited real-time control of acoustic parameters. Synthesizers are more flexible in the type of real-time control available but less general in terms of the variety of sound qualities that can be generated. A potential disadvantage of both is that they are not specifically designed for the generation and control of sounds for information display and tend to require that the user have specialized knowledge of musical/production techniques.

Needs

Any generally useful audio hardware system must sample its inputs with at least 16 bits of resolution and 40 to 50 khz sampling rate. To perform the spatial synthesis functions described in the Status section, a computational rate of about 300 MIPs is required. Additional computational power will probably be needed to implement more complex environmental models as well as the real-time generation and control of nonspeech audio cues.

Recommendations:

This seems to be an area where the technology has arrived at a price/performance level which can support rapid progress. Research should now be funded primarily to improve models of perception and techniques for generating acoustical models of the environment and of environmental sounds (acoustical "objects"), rather than for extensive hardware development. However, while not the primary emphasis, some new hardware development will probably be required to accommodate more elaborate and efficient models of complex environments as they develop.

Speech Input:

Status

Inexpensive devices are now available which, with reasonable reliability, can recognize individual words from the speaker who trained the system. Speaker-independent methods are achieving some acceptance.

Needs:

Many selection tasks would be better mediated by voice than by pointing, given the unsatisfactory state of visual pointing in 3D. As the technology of speech recognition improves, textual input in some situations could move from keyboard to voice, which will impact VE-based database activities and other information accessing tasks.

Recommendations:

Support would be appropriate for projects which explore the integration of voice input into the VE user interface. Hardware development is probably not needed.

VII. Applications

Status

For VE systems and technologies with respect to particular applications, there are two basic questions:

1. Can the given application be accomplished with VE technology? Many applications are currently beyond the state of the art.
2. If the application can be accomplished, will it prove superior to other technologies for accomplishing the same task?

While the promise of VE technology has been widely acknowledged, there are very few production-quality applications that are used regularly to solve real-world problems. Many desired applications have requirements which are currently beyond state of the art. For those that are feasible, it remains to be shown that VE provides a superior solution.

Compared to computer models displayed via conventional screen-bound interactive ("through-the-window") graphics, VE offers

- total immersion in the computer model,
- kinesthetic, intuitive viewpoint change, and
- direct manipulation in 3-D.

Thus, one would expect to be able to show comparative superiority in simple applications that exploit these advantages.

These comparative advantages promise an order of magnitude better illusion for the user, thus greater involvement of the user with the task. This promise alone has been the driving force for development of VE technology in the past. With the exception of vehicle simulators and entertainment, most current applications of VE have been developed for driving and testing the technology rather than to accomplish ends related to the applications.

The following application areas stand to gain significantly from developments in VE technology:

1. Data or Model Visualization:

   Scientific visualization techniques are widely accepted as a means for extracting understanding from large data spaces. When VE technologies are coupled with existing visualization systems, the user will experience a fundamental paradigm shift in interacting with the data or model. Stereoscopic imaging combined with intuitive control over point of view frees the user to concentrate on the data rather than the computer interface. The incorporation of other senses into the "display" potentially offers a mechanism for correlation of data features through non-visual means.

   Augmentation is an extension of data visualization: The image presented to the user is a combination of a computer-generated image and the view of the real world around the user.

2. Designing and Planning:

   A primary characteristic of design activities is the iterative *analyze-refine* cycle which takes place as the model evolves. Any method which aids the designer during this process will improve the entire activity. Because a strength of VE is the capability for direct manipulation of objects within the virtual space, design activities should benefit greatly from this technology. Existing applications for architectural design and surgical planning attest to the benefits of VE technology.

3. Education and Training:

   Computer models of real-world systems can be presented through a virtual environment. Given that the appropriate level of fidelity can be provided, the user can interact with the virtual system as though it were the real thing. In this context, the educational implications of this technology are immediately obvious. In addition to simply portraying the real world, the rules governing a virtual world can be altered to allow the user to experience and learn from alternate "realities". In this way, the effect and importance of various physical parameters can be studied, and the educational focus can be on cognitive skills rather than the specifics of a particular environment.

4. Teleoperations:

   VE can be applied to situations where an environment is too hazardous for a human, but a tele-sensor/operator may enter. As an example, the exploration of the sunken Titanic was accomplished primarily through teleoperations. Although the interface was not called 'VE,' this is an obvious extension. (Although it is an interesting problem; research into the robotic teleoperators is beyond the scope of VE research. VE research should focus on the interface between the human and the computer model, not on the model-world interface.)

5. Psychological Test Beds:

   VEs can produce physically unrealizable stimulus environments that are uniquely useful for testing theories of human perception and manual control. This includes research into visual, auditory, and haptic sensory stimuli, and its effects, both short-term and long-term, on the user of a VE.

6. Entertainment and Artistic applications:

   Given the amount of public and media interest in VE technology, there is an economic potential for entertainment which bills itself as "virtual reality". There are existing applications of this type currently installed in major theme-parks. One can only expect to see this market expand in the future as public exposure grows.

7. Artistic Applications

   Artists are often the first individuals to explore new media and technologies. The basic activity of an artist is creating space — whether physical, mental, visual, or emotional. As VE technology becomes more widely accessible, it is reasonable to expect that artists will begin to explore the unique possibilities for expression that virtual environments offer.

8. Communication and Collaboration

   Virtual environments coupled with high-speed networks and distributed computation provide a common neutral space where communication, negotiation, and collaboration can take place. The ability for two or more people in separate locations to meet inside a virtual space and work toward a common goal has the potential to revolutionize the way collaborative work is done.

Needs

To date, most of the major deficiencies of VE have been demonstrated by experimentation with various applications. These deficiencies are not specific to the applications but to VE technology. They include:

1. Hardware, software, and interface issues.
   These have been discussed in detail in an earlier portion of the report.

2. Model engineering.
   Many researchers have found that building a reasonably detailed model for a non-trivial object (i.e., a house) takes an inordinate amount of effort (perhaps a man-year). The need for improvement is obvious.

3. Psychological measurement techniques and metrics.
   Measuring the possible superiority of VE technology often involves measuring human cognitive performance. This type of measurement is extremely difficult.

4. Simulator sickness.

   Experience with space flight suggests that the more realistic VE becomes and the longer people spend immersed in a VE, the more people in VE's may experience various physical reactions, commonly known as simulator sickness. This phenomenon is also discussed in an earlier portion of the report.

Recommendations

1. Researchers in VE should look toward applications which promote measurement. NSF should encourage applications which provide discriminatory power to evaluate VE technology versus 'through the window' interactive graphics and other similar technologies.

2. Widespread development of VE applications are dependent on the availability of a stable and robust software infrastructure. NSF should address the issue of making such a software environment available as new developers enter the field. The alternative is to require each new site to waste valuable time "reinventing the wheel".

2. NSF funding should be directed toward applications which solve real-world problems. VE applications must move beyond simple demonstrations of technology and begin to solve problems for people where they are.

3. If VE applications are to be taken seriously, the causes and effects of sickness from virtual environments must be well understood. To this end, research into "simulator sickness" should be supported. VE also seems to be an excellent platform for related and unrelated psychological/physiological measurements.

4. Interpersonal Interaction as facilitated by VE. Can use of VEs to communicate between people approach the level of communication we currently experience in person, or in a group? NSF funding should support research toward resolving these questions.

5. Researchers should begin work on the probable impact of VE technology on society: Will VEs change the way we work (telecommuting/teleconferencing)? Will they modify our interpersonal interactions? As the technology becomes more readily available, how will society react?

References

Begault, D. R. (1991) Perceptual effects of synthetic reverberation on 3-D audio systems. 91st Convention of the Audio Engineering Society, New York, Preprint 3212 (W-6).

Blattner, M. M., Sumikawa, D. A., & Greenberg, R. M. (1989). Earcons and Icons: Their Structure and Common Design Principles. Human-Computer Interaction, 4, 11-44.

Bregman, A. S. (1990). Auditory Scene Analysis. Cambridge, MA: MIT Press.

Buxton, W., Gaver, W., & Bly, S. (1989). The use of non-speech audio at the interface. (Tutorial No. 10). Presented at CHI'89, ACM Conference on Human Factors in Computing Systems, New York: ACM Press.

Durlach, N. I. (1991) Auditory localization in teleoperator and virtual environment systems: Ideas, issues, and problems. *Perception, 20*, 543-554.

Durlach, N.I., Pew, R.W., Aviles, W.A., DiZio, P.A., and Zeltzer, D.L. (Eds.) (1992). Virtual Environment Technology for Training (VETT). The Virtual Environment and Teleoperator Research Consortium (VETREC). Bolt, Beranek, and Newman Report 7661.

Cutting, James E. (1986) Perception with an eye for motion. Cambridge, MA: MIT Press.

Elkind, Jerome I., Card, Stuart K., Hochberg, Julian, and Huey, Beverly M. (1989) Human performance models for computer-aided engineering, Washington D.C. National Academy Press.

Ellis, Stephen R. (1991) The nature and origin of virtual environments: a bibliographical essay. *Computer Systems in Engineering*, 2, 4, 321-47.

Ellis, Stephen R., Kaiser, Mary K., and Grunwald, Arthur J.(eds) (1991). Pictorial communication in virtual and real environments, *Taylor and Francis, London,and Bristol, NJ.*

Foster, S.H., Wenzel, E.M., and Taylor, R.M. (1991) Real- time synthesis of complex acoustic environments . *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, NY.

Howard, Ian (1982) Human visual orientation, Wiley, New York.

Lackner, J.R. Some aspects of sensory-motor control and adaptation in man. In: R.D. Walk and H.L. Pick, Jr. (Eds.). Intersensory Perception and Sensory Integration, New York, Plenum, 143-173, 1981.

Lackner, J.R. Human sensory-motor adaptation to the terrestrial force environment. In: D. Ingle, M. Jeannerod and D. Lee (Eds.). Brain Mechanisms and Spatial Vision., Nijhoff, Amsterdam, 175-210, 1985.*

Li, X., Logan, R. J., and Pastore, R. E. (1991) Perception of acoustic source characteristics. Journal of the Acoustical Society of America, 90, 3036-3049.

Morse, P. M. and Ingard, K. U. (1968) Theoretical Acoustics. New York: McGraw-Hill.

Patterson, R. R. (1982). Guidelines for Auditory Warning Systems on Civil Aircraft. (Paper No. 82017), London: Civil Aviation Authority.

Treisman, A. (1985) Preattentive processing in vision. Computer Vision, Graphics, and Image Processing, 31 156-177.

Warren William, H., Jr., Mestre, Daniel R., Blackwell, Arshavir W., and Morris, Michael W. (1991) Perception of circular heading from optical flow. Journal of Experimental Psychology, 17, 28-43.

Wenzel, E. M. (1992) Localization in virtual acoustic displays. Presence: Teleoperators and Virtual Environments, 1, 80-107.

Wightman, F. L. & Kistler, D. J. (1989) Headphone simulation of fre~-field listening I: stimulus synthesis. Journal of the Acoustical Society of America, 85, 858-867.

Appendix

Taxonomies for Virtual Environments

On Taxonomies In General Taxonomy can help thinking about a group of ideas. A taxonomy characterizes ideas along several more or less independent axes so as to form a matrix. The taxonomy has two important parts, the axes and the entries in the matrix.

The items listed on each axis are related by whatever characteristic is represented by that axis. If the variable chosen is continuous, the axis will represent a spectrum of values which may be more or less well defined. If the variable chosen has only a few fixed values, then those values divide the space of all ideas considered into categories. For example, a taxonomy of people might include eye color as a variable, characterizing eye color into a small number of common values, e.g. blue, brown, hazel, etc. The several axes of the taxonomy divide the intellectual space of interest into a number of boxes, each of which can be examined in turn.

The boxes in a taxonomy are used to enumerate ideas that have common characteristics in the axes chosen. One can list in each such box a set of ideas that are related to each other by their position on the axis. If all of the axes are continuous, the taxonomy describes a space of possibilities. In such a diagram, each related collection of ideas occupies a region of the overall space instead of a discrete box.

The number of boxes in a taxonomy is, of course, its volume, i.e. the product of the number of values permitted to each variable. The most useful taxonomies generally have fewer than 100 boxes. With more boxes it is difficult to think about each one separately. Some useful taxonomies have only a few boxes; as few as four are sometimes useful. The desire to limit the number of boxes limits the number of values permitted to each variable. Of course if there are more variables, each must be permitted fewer values. A useful discrete variable usually provides 2-5 values; a useful continuous variable generally approximates continuity with small, medium, and large.

There are two parts of building a taxonomy that make building it interesting. First there is the selection of axes and values along them. In forming the axes, one is forced to think of orthogonal dimensions. Which variables are related and which are independent. It often happens that as the taxonomy takes shape one finds that the variables chosen are not really independent. For example, in forming the taxonomy presented here we initially chose a separate axis to distinguish "sensed real world" versus "simulated world." As the taxonomy formed we considered that maybe "simulated" could be included as a single value along the "time" axis. We considered making this change because we found inadequate distinction between boxes differing in this dimension.

The second interesting part of building a taxonomy is filling in the boxes. Here the really interesting results come out by examining combinations of variables that may hitherto not have been considered together. For example, our taxonomy that plots different sense modalities against different physical phenomena leads one to ask what smells look like, how pictur~s sound, and other such ideas that might not otherwise occur to unaided thinking. Making a taxonomy is the only organized way I know to go about creative thought.

Our Taxonomies for Virtual Environments We actually made three taxonomies, which divide up the space of all possible virtual environments in different ways.

<u>Data Flow Between the World and Human.</u> The first taxonomy comes about by observing that a mechanical or electronic device can be interposed between a human being and the real world, as shown in Figure 1 below.

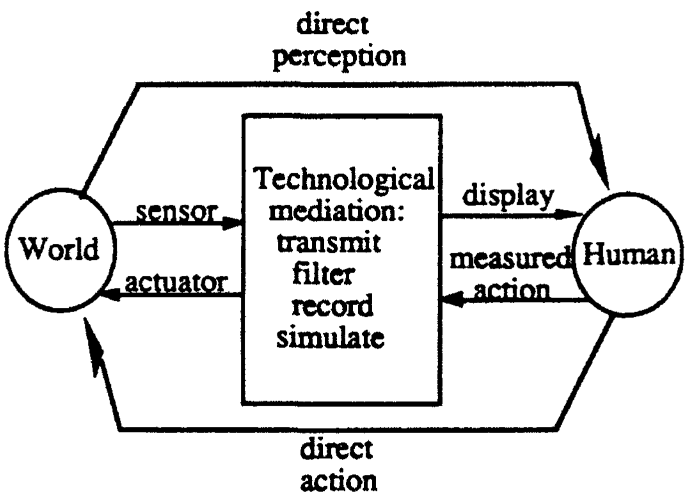


Figure 1. Technologically-mediated experience.

Inputs and outputs from the real world and the human are connected to this black box. This diagram covers all types of technologically-mediated experience, in which some sort of apparatus mediates human perception of and action within the world. The most important subcategories of technologically-mediated experience are recorded, transmitted, and simulated experience. Figure 2 shows that these types of mediated experience are distinguished by different patterns of data flow. The data flow for a robot is also shown in Figure 2, with possible supervision by a human operator. We include within the black box in each of the diagrams the *model* which defines the virtual world perceived by the human.

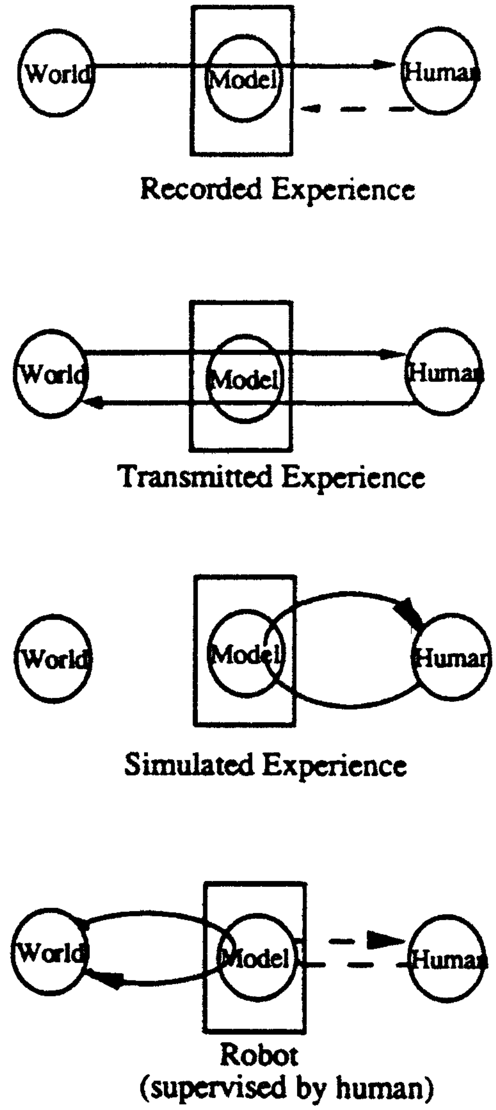


Figure 2. Data flow for types of mediated experience.

Two or more people may communicate or interact within a virtual environment, which leads to the data flow diagrams in Figure 3. The first two diagrams emphasize two different things, communication versus interaction with the virtual world. However all of the data paths exist to do both simultaneously. The last diagram of Figure 3 shows that two people may interact collaboratively with the real world through a shared representation of the real world.

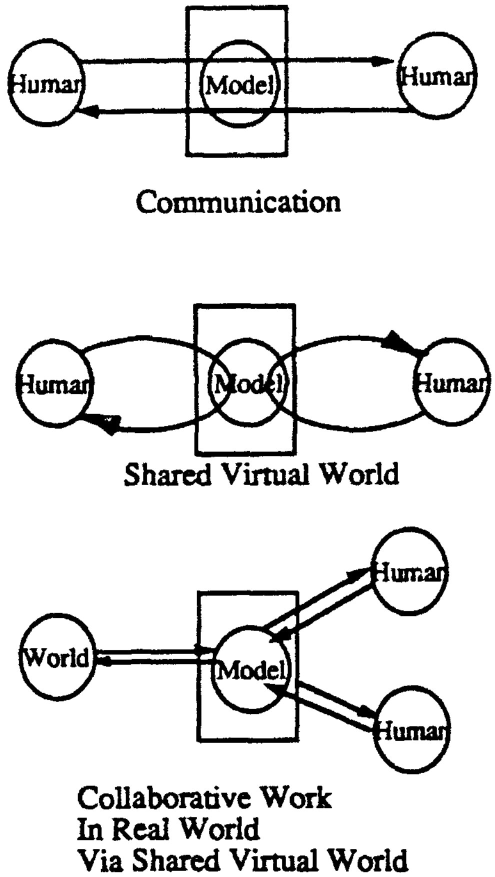


Figure 3. Two people in shared virtual world.

Time and Space Our second taxonomy chooses as axes three variables: time, space, and simulated vs real. In this taxonomy, the dimensions of time and space compare the time and space coordinates of the real world as captured by a sensor versus the actual physical time-space location of the human immersed in a virtual world. For example, in teleoperation, the space coordinates of the virtual world perceived by the human are different from the human's space coordinates.

We have characterized both time and space as "aligned", "displaced", and "scaled". By aligned we mean in real time or in spatial registry. By delayed we mean passed on through time by some kind of recording or displaced through space as in one of the many "tele" kinds of objects. By scaled we mean speeded up, slowed down, enlarged, or minified. Notice that if scaled, aligned does not apply unless two scale factors happen to match.

We have debated how to treat the "simulated versus real" dimension. One of us claims that "simulated" is really another value on the time dimension because in a simulation "real time" is a matter of choice. The other claims that simulations can be intended to model physical systems and therefore that "real time" does have a meaning within a simulation. Entering this debate is instructive in terms of generating ideas. For this report, we have left "simulated versus real" as a separate dimension of the taxonomy.

A issue that comes up in trying to classify simulated virtual worlds is that while some simulations are intended to model the real world, and thus can be classified as real-time or spatially registered, other simulations are completely fanciful and don't match up with any time or place in the real world. Thus "fantasy world simulation" seems to be a separate category of its own in our taxonomy.

We have drawn the taxonomy hierarchically in two stages. The top level is shown in Figure 4, and the two boxes "recording or transmission of real world" and "simulation of real world" are further broken down in Figure 5 along the dimensions of relative time and space. The entries in the boxes of Figure 5 are examples, not category definitions.

	Sense Real World	Simulate
Real World	recording or transmission of real world	simulation of real world
Fabricated World	?	simulation of fantasy world

Figure 4. Top level taxonomy of virtual environments.

Sensed Real World

	Transmit (real-time)	Record (1-to-1 time)	Record (scaled time)
Aligned (spatially)	night vision goggles	ghost (on-site replay of past events)	ghost speeded up
Displaced (spatially)	teleoperation	movie	slow motion instant replay
Scaled (spatially)	micro-teleoperation	micro-movie	micro-movie slowed sown

Simulation of Real Worlds

	Record (1-to-1 time)	Record (scaled time)
real-time in-place sim	historical sim (reenactment on site)	historical sim speeded up
predictive sim in teleoperation	flight sim	sim of trip to Mars speeded up
real-time weather prediction	molecular sim	molecular sim slowed down

Figure 5. Taxonomy of time, space, and simulated vs. real.

Sensors versus Human Senses Our third taxonomy looks at extending human perception within and across sense modalities. Here one axis is just the sensory modalities. The other axis is various phenomena that could be sensed. This matrix is shown in the figure below. Neither axis has a complete list of senses nor phenomena.

	See	Hear	Feel	Vestibular	Smell
Visible Light	spectacles; night vision goggles		Opticon image-to-tactile transducer		
Sound	sonogram	hearing aid			
Force	visualization of strain in special plastic		teleoperation		
Inertial Changes	seismograph	rattle			
Chemical Compostion	gas chromatograph				smell amplifier
Ultrasound	medical ultrasound image	sonar			
Radio Waves	radio telescope	radio			
Infrared	night vision goggles				
X-rays	fluoroscope				
Magnetism	compass				
Radiation		Geiger counter			

Along the main diagonal of this matrix are augmentations and prostheses for the ordinary "built-in" human senses. For example, spectacles appear in the visible light row under vision, and hearing aids are listed at the intersection of sound and hearing. The diagonal entry for smell prompts the question "Could a device for amplifying smell be created?"

The off-diagonal boxes would contain examples of sensory substitution. For example, what do smells look like? What do sounds look like? How do sounds smell? It is the ability of this taxonomy to generate fresh ideas that is interesting.

The lower part of the matrix lists phenomena for which no corresponding human sensory system exists, such as ultrasound. For these imperceptible phenomena, mapping sensors which detect them to human senses expands human awareness by creating "synthetic senses."

Isolated versus Merged Another distinction not captured in any of the earlier taxonomies in this report is that a virtual world may either be portrayed to the human with the real world blacked out, or the virtual world may be superimposed onto the human's direct perception of the real world. This may be done either optically with half-silvered mirrors, or computationally by capturing a representation of the real world with cameras and then mixing the captured model of the real world with another model representing the virtual world.

It is most useful to merge the real and virtual worlds when the virtual world is spatially registered with the real world. However, even when the virtual world models a distant location, seeing through to the surrounding real world may be useful simply to let the human user avoid running into walls and tripping over obstacles.

Acknowledgements

The workshop from which this report proceeded was funded by National Science Foundation Contract Number IRI-9213777 (Interactive Systems Program, Dr. John Hestenes, Director). We gratefully acknowledge the assistance of Ronald Azuma, James Chung, Stefan Gottschalk, Robert Katz, Mark Mine, Jannick Rolland, Shannon Stephenson, Kathy Tesh, Fay Ward, and William V. Wright, as staff assistants.

Merging Virtual Objects with the Real World:
Seeing Ultrasound Imagery within the Patient

Michael Bajura, Henry Fuchs, and Ryutarou Ohbuchi

Department of Computer Science
University of North Carolina
Chapel Hill, NC 27599-3175

Abstract

We describe initial results which show "live" ultrasound echography data visualized within a pregnant human subject. The visualization is achieved by using a small video camera mounted in front of a conventional head-mounted display worn by an observer. The camera's video images are composited with computer-generated ones that contain one or more 2D ultrasound images properly transformed to the observer's current viewing position. As the observer walks around the subject, the ultrasound images appear stationary in 3-space within the subject. This kind of enhancement of the observer's vision may have many other applications, e.g., image guided surgical procedures and on location 3D interactive architecture preview

CR Categories: I.3.7 [Three-Dimensional Graphics and Realism] Virtual Reality, I.3.1 [Hardware architecture]: Three-dimensional displays, I.3.6 [Methodology and Techniques]: Interaction techniques, J.3 [Life and Medical Sciences]: Medical information systems.

Additional Keywords and Phrases: Virtual reality, see-through head-mounted display, ultrasound echography, 3D medical imaging

1. Introduction

We have been working toward an 'ultimate' 3D ultrasound system which acquires and displays 3D volume data in real time. Real-time display can be crucial for applications such as cardiac diagnosis which need to detect certain kinetic features. Our 'ultimate' system design requires advances in both 3D volume data *acquisition* and 3D volume data *display*. Our collaborators, Dr. Olaf von Ramm's group at Duke University, are working toward real-time 3D volume data acquisition [Smith 1991; von Ramm 1991]. At UNC-Chapel Hill, we have been conducting research on real-time 3D volume data visualization.

Our research efforts at UNC have been focused in three areas: 1) algorithms for acquiring and rendering real-time ultrasound data, 2) creating a working virtual environment which acquires and displays 3D ultrasound data in real time, and 3) recovering structural information for volume rendering specifically from ultrasound data, which has unique image processing requirements. This third area is presented in [Lin 1991] and is not covered here.

Section 2 of this paper reviews previous work in 3D ultrasound and Section 3 discusses our research on processing, rendering, and displaying echographic data without a head-mounted display. Since the only real-time volume data scanners available today are 2D ultrasound scanners, we try to approximate our 'ultimate' system by incrementally visualizing a 3D volume dataset reconstructed from a never-ending sequence of 2D data slices [Ohbuchi 1990; 1991]. This is difficult because the volume consisting of multiple 2D slices needs to be visualized incrementally as the 2D slices are acquired. This incremental method has been successfully used in off line experiments with a 3-degree-of-freedom (DOF) mechanical arm tracker and is extendible to 6 degrees of freedom, e.g., a 3D translation and a 3D rotation, at greater computational cost.

Sections 4 and 5 present our research on video see-through head-mounted display (HMD) techniques involving the merging of computer generated images with real-world images. Our video see-through HMD system displays ultrasound echography image data in the context of real (3D) objects. This is part of our continuing see-through HMD research, which includes both optical see-through HMD and video see-through HMD. Even though we concentrate here on medical ultrasound imaging, applications of this display technology are not limited to it (see Section 6.2).

2. Previous Research in 3D Ultrasound

The advantages of ultrasound echography are that it is relatively safe compared with other imaging modalities and that images are generated in real time [Wells 1977]. This makes it the preferred imaging technique for fetal examination, cardiac study, and guided surgical procedures such as fine-needle aspiration biopsy of breast tumors [Fornage 1990]. Ultrasound echography offers the best real-time performance in 3D data acquisition, although slower imaging modalities such as MRI are improving.

The drawbacks of ultrasound imaging include a low signal to noise ratio and poor spatial resolution. Ultrasound images exhibit "speckle" which appears as grainy areas in images. Speckle arises from coherent sound interference effects from tissue substructure. Information such as blood flow can be derived from speckle but in general speckle is hard to utilize [Thijssen 1990]. Other problems with ultrasound imaging include attenuation that increases with frequency, phase aberration due to tissue inhomogeneity, and reflection and refraction artifacts [Harris 1990].

2.1 3D Ultrasound Image Acquisition

Just as ultrasound echography has evolved from 1D data acquisition to 2D data acquisition, work is in progress to advance to 3D data acquisition. Dr. Olaf von Ramm's group at Duke University is developing a 3D scanner which will acquire 3D data in real time [Shattuck 1984; Smith 1991; von Ramm 1991]. The 3D scanner uses a 2D phased array transducer to sweep out an imaging volume. A parallel processing technique called *Explososcan* is used on return echoes to boost the data acquisition rate.

Since such a real-time 3D medical ultrasound scanning system is not yet available, prior studies on 3D ultrasound imaging known to the authors have tried to reconstruct 3D data from imaging primitives of a lesser dimension (usually 2D images). To reconstruct a 3D image from images of a lesser dimension, the location and orientation of the imaging primitives must be known. Coordinate values are explicitly tracked either acoustically [Brinkley 1978; King 1990; Moritz 1983], mechanically [Geiser 1982a; Geiser 1982b; Hottier 1989; McCann 1988; Ohbuchi 1990; Raichelen 1986; Stickels 1984], or optically [Mills 1990]. In other systems, a human or a machine makes scans at predetermined locations and/or orientations [Collet Billon 1990; Ghosh 1982; Itoh 1979; Lalouche 1989; Matsumoto 1981; Nakamura 1984; Tomographic Technologies 1991].

A particularly interesting system under development at Philips Paris Research Laboratory is one of the closest yet to a real-time 3D ultrasound scanner [Collet Billon 1990]. It is a follow on to earlier work which featured a manually guided scanner with mechanical tracking [Hottier 1990]. This near real-time 3D scanner is a mechanical sector scanner, in which a conventional 2D sector scanhead with an annular array transducer is rotated by a stepper motor to get a third scanning dimension. In a period of 3 to 5 seconds, 50 to 100 slices of 2D sector scan images are acquired. Currently the annular array transducer in this system provides better spatial resolution, but less temporal resolution, than the real-time 3D phased array system by von Ramm et al., mentioned above. A commercial product, the *Echo-CT* system by Tomographic Technologies, GMBH, uses the linear translation of a transducer inside a tube inserted into the esophagus to acquire parallel slices of the heart. Image acquisition is gated by respiration and an EKG to reduce registration problems [Tomographic Technologies 1991].

2.2 3D Ultrasound Image Display

One should note that 3D image data can be presented not only in visual form, but also as a set of calculated values, e.g., a ventricular volume. The visual form can be classified further by the rendering primitives used, which can be either geometric (e.g., polygons) or image-based (e.g., voxels). Many early studies focused on non-invasively estimating of the volume of the heart chamber [Brinkley 1978; Ghosh 1982; Raichelen 1986; Stickels 1984]. Typically, 2D echography (2DE) images were stored on video tape and manually processed off-line. Since visual presentation was of secondary interest, wire frames or a stack of contours were often used to render

An interesting extension to 2D display is a system that tracks the location and orientation of 2D image slices with 6 DOF [King 1990]. On each 2D displayed image, the system overlays lines indicating the intersection of the current image with other 2D images already acquired. The authors claim that these lines help the viewer understand the relationship of the 2D image slices in 3D space. Other studies reconstructed 3D grey level images preserving grey scale, which can be crucial to tissue characterization [Collet Billon 1990; Hottier 1989; Lalouche 1989; McCann 1988; Nakamura 1984; Pini 1990; Tomographic Technologies 1991]. [Lalouche 1989] is a mammogram study using a special 2DE scanner that can acquire and store 45 consecutive parallel slices at 1 mm intervals. A volume is reconstructed by cubic-spline interpolation and then volume rendered. [McCann 1988] performed gated acquisition of a heart's image over a cardiac cycle by storing 2DE images on video tape and then reconstructing and volume rendering them. 'Repetitive low-pass filtering' was used during reconstruction to fill the spaces between radial slices, which suppressed aliasing artifacts. [Tomographic Technologies 1991] provides flexible re-slicing by up to 6 planes as well other imaging modes. [Collet Billon 1990] uses two visualization techniques: re-slicing by an arbitrary plane and volume rendering. The former allows faster but only 2D viewing on a current workstation. The latter allows 3D viewing but often involves cumbersome manual segmentation. The reconstruction algorithm uses straightforward low pass filtering.

3. Incremental Volume Visualization

We have been experimenting with volume rendering as one alternative for visualizing dynamic ultrasound volume data. Standard volume rendering techniques which rely heavily on preprocessing do not apply well to dynamic data which must be visualized in real time [Levoy 1988; Sabella 1988; Upson 1988]. We review here an incremental, interactive 3D ultrasound visualization technique which visualizes a 3D volume as it is incrementally updated by a sequence of registered 2D ultrasound images [Ohbuchi 1990, 1991].

Our target function is sampled at irregular points and may change over time. Instead of directly visualizing samples from this target, we reconstruct a regular 3D volume from this time series of spatially irregular sample points. This places a limit on storage and computation requirements which would grow without bound if we retained all the past sample points. The reconstructed volume is then rendered with an incremental volume-rendering technique.

The reconstruction is a 4D convolution process. A 3D Gaussian kernel is used for spatial reconstruction followed by a temporal reconstruction based on simple auto regressive moving average (ARMA) filtering [Haddad 1991]. Time stamps are assigned to each 3D voxel, which are updated during reconstruction. The time stamp difference between a reconstructed voxel and an incoming sample is used to compute coefficients for the ARMA filter. The 3D Gaussian filter is loosely matched to the point spread function of the ultrasound transducer and is a good choice because it minimizes the product of spatial bandwidth and spatial frequency bandwidth [Hildreth 1983; Leipnik 1960].

An image-order, ray-casting algorithm based on [Levoy 1988] renders the final images incrementally. Rendering is incremental and fast only if the viewpoint is fixed and if the updated volume is relatively small. Shading and ray sampling are done only for voxels proximate to incoming data. The ray samples are stored in a 3D array in screen space called a "ray cache" for later use. The ray cache is hierarchical so that a small partial update of the ray cache can be composited quickly (Ohbuchi, 1991). The hierarchical ray cache also allows fast rendering of polygons properly composited with volume data which can enhance the volume visualization [Levoy 1990, Miyazawa 1991]. This incremental volume rendering algorithm is not restricted to ultrasound and is applicable to other problems which update volume data incrementally, e.g. interactive volume modeling by sculpting [Galyean 1991].

To test this visualization technique, we acquired a series of 2D images with a manually guided conventional 2DE scanhead attached to a mechanical tracking arm with 3 DOF (two translations and one rotation). As we scanned various targets in a water tank, their images and their corresponding geometry were stored off-line. We then ran the incremental volume visualization algorithm on a DECstation 5000 with 256 MB of memory using this data. With a reconstruction buffer size of 150 × 150 × 300 and an image size of 256 × 256, it took 15–20 seconds to reconstruct and render a typical image after insertion of a 2D data slice. This time varied with reconstruction, shading, and viewing parameters.

Figure 1 shows 2 out of 90 2D images of a plastic toy doll phantom which is visualized in Figure 2. The 2D images were produced by an ATL Mark-4 Scanner with a 3.5 MHz linear scanhead. The 2D images overlap but are roughly parallel at approximately 2 mm intervals.

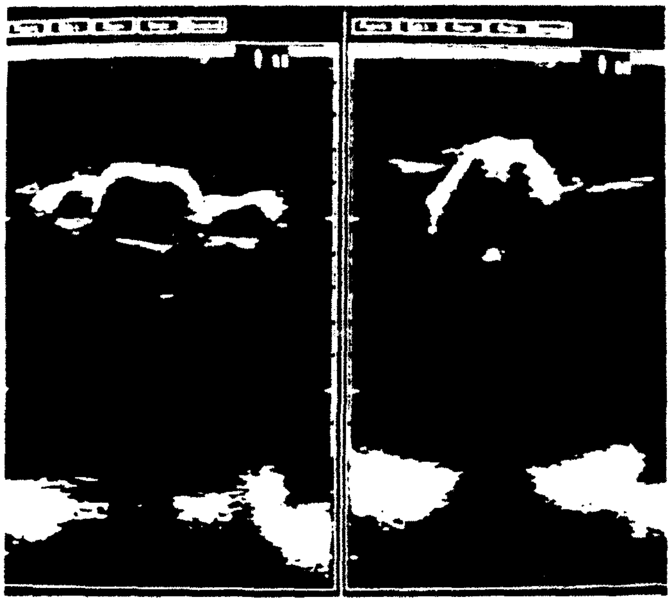


Figure 1. Two of 90 2D ultrasound echography images of a plastic toy doll phantom which was scanned in a water tank. The scans shown are at the torso (left) and at the head (right). The clouds at the bottom of the scans are artifacts due to reflections from the bottom of the water tank.

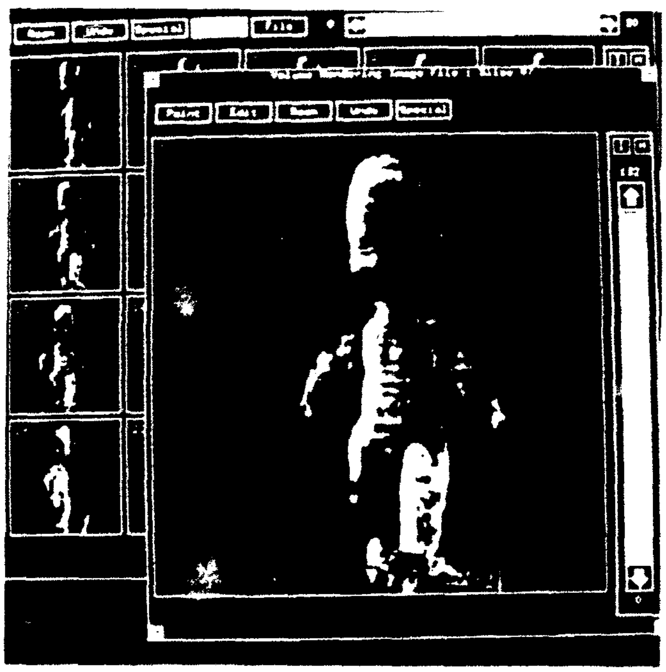


Figure 2. Reconstructed and rendered image of the toy doll phantom using incremental volume visualization.

4. Virtual Environment Ultrasound Imaging

Various medical ultrasound imaging applications require a registration of ultrasound images with anatomical references, e.g., in performing a fine needle aspiration biopsy of a suspected breast tumor [Fornage 1990]. A virtual environment which displays images acquired by ultrasound equipment in place within a patient's anatomy could facilitate such an application. We have developed an experimental system that displays multiple 2D medical ultrasound images overlaid on real-world images. In January 1992, after months of development with test objects in water tanks, we performed our first experiment with a human subject.

Our virtual environment ultrasound imaging system works as follows (note that this is a different system than our older one described in the previous section): as each echography image is acquired by an ultrasound scanner, its position and orientation in 3D world space are tracked with 6 degrees of freedom (DOF). Simultaneously the position and orientation of a HMD are also tracked with 6 DOF. Using this geometry, an image-generation system generates 3D renderings of the 2D ultrasound images. These images are video mixed with real-world images from a miniature TV camera mounted on the HMD. The resulting composite image shows the 2D ultrasound data registered in its true 3D location.

Figure 3 is a block diagram of our system's hardware. There are three major components: 1) an image-acquisition and tracking system, which consists of an ultrasound scanner and a Polhemus tracking system, 2) an image-generation system, which is our Pixel-Planes 5 graphics multicomputer, and 3) a HMD which includes a portable TV camera, a video mixer, and a VPL EyePhone. Each component is described in more detail in Sections 4.1–4.3.

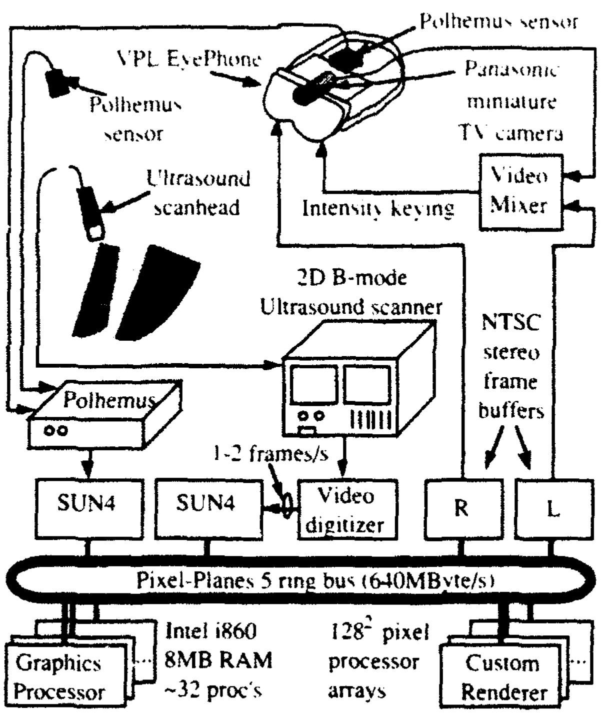


Figure 3. Hardware block diagram for the virtual environment ultrasound system.

4.1 Image Acquisition and Tracking

Two dimensional ultrasound images are generated by an IREX System III echography scanner with a 16 mm aperture 2.5 MHz phased array transducer. These images are digitized by a SUN 4 with a Matrox MVP/S real-time video digitizer and transferred to our Pixel-Planes 5 graphics multicomputer [Fuchs 1989]. The SUN 4 operates as a 2DE image server for requests from the Pixel-Planes 5 system. Images are distributed among the Graphics Processors (GPs) on a round-robin scan-line by scan-line basis. Due to the bandwidth limitations of the SUN 4 VME bus, transfer of the $512 \times 480 \times 8$ bits/pixel images is limited to 2 Hz.

A Polhemus system with one source and two receivers is used for tracking [Polhemus 1980]. One receiver tracks the HMD. The other tracks the ultrasound transducer. The Polhemus system is mounted in non ferrous materials away from magnetic interference sources such as the ultrasound transducer, HMD, and other lab equipment. A calibration procedure is used to relate both the ultrasound transducer to its Polhemus receiver and the HMD TV camera to its Polhemus receiver mounted on the HMD. This calibration procedure is described in Section 4.4.

4.2 Image Generation

Images are generated by the Pixel-Planes 5 system based on geometry information from the tracking system. Pixel-Planes 5 runs a custom PHIGS implementation which incorporates a facility to update display structures asynchronously from the display process. This separates the interactive virtual environment update rate from the 2D ultrasound image data acquisition rate. Images in the virtual environment are registered to the real world within the update-rate limit of the tracking and display system and not within the acquisition-rate limit of the image-acquisition system.

Pixels from the 2D ultrasound images are rendered as small unshaded sphere primitives in the virtual environment. The 2D ultrasound images appear as space-filling slices registered in their correct 3D position. The ultrasound images are distributed among the GPs where they are clipped to remove unnecessary margins and transformed into sphere primitives, which are then sent to the Renderer boards for direct rasterization. Pixel-Planes 5 renders spheres very rapidly, even faster than it renders triangles, over 2 million per second [Fuchs 1985; 1989]. Final images are assembled in double buffered NTSC frame buffers for display on the HMD. To reduce the number of sphere primitives displayed, the ultrasound images are filtered and subsampled at every 4th pixel. Due to the low resolution of the HMD and inherent bandwidth limitation of the ultrasound scanner, this subsampling does not result in a substantial loss of image quality. An option to threshold lower intensity pixels in 2D ultrasound images prior to 3D rendering can suppress lower intensity pixels from being displayed.

4.3 Video See-Through HMD

A video see-through HMD system combines real-world images captured by head-mounted TV cameras with synthetic images generated to correspond with the real-world images. The important issues are tracking the real-world cameras accurately and generating the correct synthetic images to model the views of the cameras. Correct stereo modeling adds concerns about matching a pair of cameras to each other as well as tracking and modeling them. [Robinett 1991] discusses stereo HMD in detail and includes an analysis of the VPL EyePhone.

A Panasonic GP-KS102 camera provides monocular see-through capability for the left eye in our current system. Images from this camera are mixed with synthetic images from the Pixel-Planes 5 system using the luminance (brightness) keying feature on a Grass Valley Group Model 100 video mixer. With luminance keying, the pixels in the output image are selected from either the real-world image or the synthetic image, depending on the luminance of pixels in the synthetic image. The combined image for the left eye and a synthetic image only for the right eye are displayed on a VPL EyePhone.

4.4 Calibration

Two transformations, a "transducer transformation" and a "camera transformation," are needed to calibrate our test system. The transducer transformation relates the position and orientation of the Polhemus tracker attached to the ultrasound transducer to the position and scale of 2D ultrasound image pixels in 3D space. The camera transformation relates the position and orientation of the head-mounted Polhemus tracker to the HMD TV camera position, orientation, and field of view.

Both transformations are calculated by first locating a calibration jig in both the lab (real) and tracker (virtual) 3D coordinate systems. This is accomplished by performing rigid body rotations with the transducer tracker about axes which are to be fixed in both the real and virtual coordinate systems. Two samples from the tracker, each consisting of both a position and an orientation, are sufficient to fix each calibration axis. The transducer transformation is computed by taking an ultrasound image of a target of known geometry placed at a known position on the calibration jig. By finding the pixel coordinates of point targets in the ultrasound image, the world coordinates of pixels in the ultrasound image can be found. From this relationship and the location of the Polhemus tracker attached to the ultrasound transducer at the time the target was imaged, the transducer transformation is derived. Similarly, the camera transformation is found by placing the HMD TV camera at known positions and orientations relative to the calibration jig. The field of view of the TV camera is known from camera specifications. Manual adjustments are used to improve the camera transformation.

5. Experimental Results

In January 1992 we conducted an experiment with a live human subject using the method described above. We scanned the abdomen of a volunteer who was 38 weeks pregnant. An ultrasound technician from the Department of Obstetrics & Gynecology of the UNC Hospitals performed the ultrasound scanning.

Figure 4 is a scene from the experiment. A person looks on with modified VPL EyePhone with the miniature video camera mounted on top and in front. Figure 5 shows the left eye view from the HMD, a composition of synthetic and real images. Figure 6 is another view from the left eye of the HMD wearer which shows several 2D ultrasound images in place within the subject's abdomen.

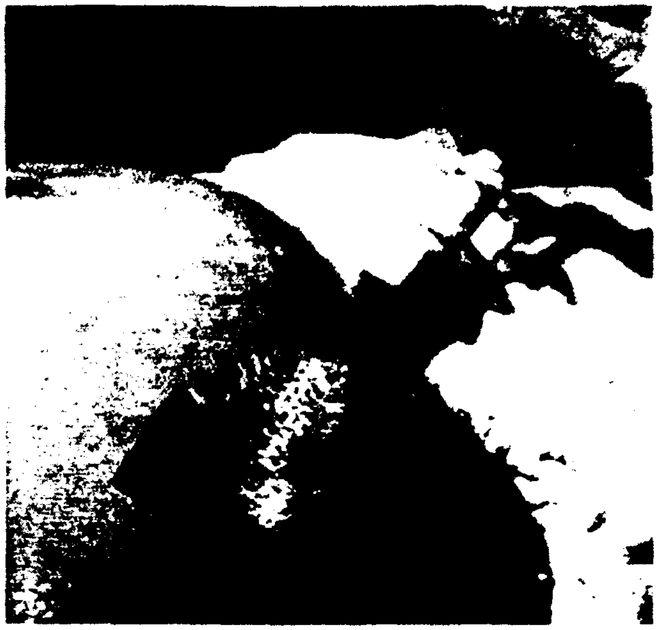


Figure 5. A video image presented to the left eye of the HMD showing a view of the subject's abdomen with a 2D ultrasound image superimposed and registered. Note the ultrasound transducer registered with the image acquired by it. The 2D image is from the antero-inferior view.

Figure 4. An ultrasound technician scans a subject while another person looks on with the video see-through head-mounted display (HMD). Note the miniature video camera attached to the front of the VPL EyePhone HMD.

Figure 6. Another video image presented to the HMD showing several 2D image slices in 3D space within the patient's abdomen. The image slices are from the anterior view.

6. Conclusions and Future Directions

The results presented so far are the initial steps in the first application of what we hope will be a flourishing area of computer graphics and visualization.

6.1 Remaining Technical Problems

1) Conflicting visual cues: Our experiment (Figures 5 and 6) showed that simply overlaying synthetic images on real ones is not sufficient. To the user, the ultrasound images did not appear to be *inside* the subject, so much as pasted on *top* of her. To overcome this problem, we now provide additional es to the user by making a virtual hole in the subject (Figure by digitizing points on the abdomenal surface and constructing a shaded polygonal pit. The pit provides occlusion cues by obscuring the abdomenal surface along the inside walls of the pit. Shading the pit provides an additional cue. Unfortunately, this does not completely solve the problem: the pit hides *everything* in the real image that is in the same location (in 2D) as the pit, including real objects that are closer in 3D than the pit. (Note in Figure 7, the edge of the transducer is hidden behind the pit representation even though it should appear in front of it.)

To solve this problem, the systems needs to know depth information for both the real and synthetic objects visible from the HMD user's viewpoint. This would make it possible to present correct occlusion cues by combining the live and synthetic images with a Z-buffer like algorithm. An ideal implementation of this would require real-time range finding from the viewpoint of the HMD user - a significant technical challenge. Graphics architectures that provide real-time depth-based image composition are already under development [Molnar 1992].

Another remaining problem is the visualization of internal 3D structure in data captured by the ultrasound scanner. Neither our incremental volume rendering algorithm (Section 3) nor multiple explicit image slices in 3-space (Figure 6) solve this problem well. A combination of multiple visualization methods will probably be necessary in the future. We suspect that this problem is difficult because the human visual system is not accustomed to seeing structure within opaque objects, and so our development cannot be guided by the "gold standard" of reality that has been used so effectively in guiding other 3D rendering investigations.

Figure 7. An image showing a synthetic hole rendered around ultrasound images in an attempt to avoid conflicting visual cues. Note the depth cues provided by occlusion of the image slices by the pit walls and shading of the pit. Also note the incorrect obscuration of the ultrasound transducer by the pit wall. (RT3200 Advantage II ultrasound scanner courtesy of General Electric Medical Systems.)

2) System lag: Lag in image generation and tracking is noticeable in all head-mounted displays; but it is dramatically accentuated with see-through HMD. The "live video" of the observer's surroundings moves appropriately during any head movement but the synthetic image overlay lags behind. This is currently one of our system's major problems which prevents it from giving the user a convincing experience of seeing synthetic objects or images hanging in 3-space. A possible solution may be to delay the live video images so that their delay matches that of the synthetic images. This will align the real and synthetic images, but won't eliminate the lag itself. We are also considering predictive tracking as a way to reduce the effect of the lag [Liang 1991]. Developers of some multi-million dollar flight simulators have studied predictive trackingfor many years, but unfortunately for us, they have not, to our knowledge, published details of their methods and their methods' effectiveness. For the immediate future, we are planning to move to our locally-developed "ceiling tracker" [Ward 1992] and use predictive tracking.

3) Tracking system range and stability: Even though we are using the most popular and probably most effective commercially available tracking system from Polhemus, we are constantly plagued by limitations in tracking volume and tracking stability [Liang 1991]. The observer often steps inadvertently out of tracker range, and even while keeping very still the observer must cope with objects in the synthetic image "swimming" in place. We are eagerly awaiting the next generation of tracking systems from Polhemus and other manufacturers that are said to overcomemost of these problems. Even more capable tracking systems will be needed in order to satisfy the many applications in which the observer must move about in the real world instead of a laboratory, operating room or other controlled environment. Many schemes have been casually proposed over the years, but we know of no device that has been built and demonstrated. Even the room-size tracker we built and demonstrated for a week at SIGGRAPH'91 still needs special ceiling panels with infrared LEDs [Ward 1992].

4) Head-mounted display system resolution: For many of the applications envisioned, the image quality of current head-mounted video displays is totally inadequate. In a see-through application, a user is even more sensitive to the limitations of his head-mounted display than in a conventional non-see-through application because he is painfully aware of the visual details he's missing.

5) More powerful display engines: Even with all the above problems solved, the synthetic images we would like to see, for example, real-time volume visualization of real-time volume data, would still take too long to be created. Much more powerful image generation systems are needed if we are to be able to visualize usefully detailed 3D imagery.

6.2 Other Applications

1) Vision in surgery: In neurosurgery, ultrasound is already used to image nearby arteries that should be avoided by an impending surgical incision.

2) Burning buildings: With close-range, millimeter wavelength radar, rescuers may be able to "see through" the smoke in the interior of burning buildings.

3) Building geometry: Geometry or other structural data could be added to a "live" scene. In the above "burning building" scenario, parts of a building plan could be superimposed onto the visual scene, such as the location of stairways, hallways, or the best exits out of the building.

4) Service information: Information could be displayed to a service technician working on complicated machinery such as a jet engine. Even simpler head-mounted displays, ones without head tracking, already provide information to users on site and avoid using a large cumbersome video screens. Adding head tracking would allow 3D superimposition to show, for instance, the location of special parts within an engine, or the easiest path for removal or insertion of a subassembly.

5) Architecture on site: Portable systems could allow builders and architects to preview buildings on site before construction or visualize additions to existing architecture.

With the work presented here and the identification of problems and possibilities for further research, we hope to encourage applications not only of "virtual environments" (imaginary worlds), but also applications that involve an "enhancement of vision" in our real world.

Acknowledgments

We would like to thank the following people: David Chen and Andrew Brandt for experimental assistance; General Electric Medical Systems (and especially R. Scott Ray) for the loan of an ultrasound scanner; Stefan Gottschalk for much assistance with video acquisition, editing, and printing; Professor Olaf von Ramm (Duke University) for donation of the IREX ultrasound scanner; ultrasound technician George Blanchard, RDMS, for scanning the subject; David Harrison and John Hughes for video and laboratory setup; Andrei State for experimental assistance; John Thomas for fabrication of a custom camera mount; Terry Yoo for video tape editing; Vern Katz, MD, for assistance with multiple ultrasound machines and scanning experiments; Nancy Chescheir, MD, for loan of an ultrasound machine and arrangements with the ultrasound technician; Warren Newton, MD, and Melanie Mintzer, MD, for finding our subject; Warren Robinett and Rich Holloway, for consultation with HMD optics and software; Professor Stephen Pizer and Charlie Kurak for consultation on the difficulty of enhancing ultrasound images; David Adam (Duke University) for instruction in the use of the IREX scanner; and our subject and her husband for their time and patience.

This research is partially supported by DARPA ISTO contract DAEA 18-90-C-0044, NSF ERC grant CDR-86-22201, DARPA ISTO contract 7510, NSF grant MIP-9000894, NSF cooperative agreement ASC-8920219, and NIH MIP grant PO 1 CA 47982, and by Digital Equipment Corporation.

References

[Brinkley 1978] Brinkley, J. F., Moritz, W.E., and Baker, D.W. "Ultrasonic Three-Dimensional Imaging and Volume From a Series of Arbitrary Sector Scans." *Ultrasound in Med. & Biol.*, **4**, pp317-327.

[Collet Billon 1990] Collet Billon, A., *Philips Paris Research Lab.* Personal Communication.

[Fornage 1990] Fornage, B. D., Sneige, N., Faroux, M.J., and Andry, E. "Sonographic appearance and ultrasound guided fine-needle aspiration biopsy of brest carcinomas smaller than 1 cm³." *Journal of Ultrasound in Medicine*, **9**, pp559-568.

[Fuchs 1985] Fuchs, H., GoldFeather, J., Hultiquist, J.P., Spach, S., Austin, J., Brooks, Jr., F.P., Eyles, J., and Poulton, J. "Fast Spheres, Textures, Transparencies, and Image Enhancements in Pixel Planes." *Computer Graphics (Proceedings of SIGGRAPH'85)*, **19**(3), pp111-120.

[Fuchs 1989] Fuchs, H., Poulton, J., Eyles, J., Greer, T., Goldfeather, J., Ellsworth, D., Molnar, S., and Israel, L. "Pixel Planes 5: A Heterogeneous Multiprocessor Graphics System Using Processor-Enhanced Memories." *Computer Graphics (Proceedings of SIGGRAPH'89)*, **23**(3), pp79-88.

[Galyean 1991] Galyean, T. A., and Hughes, J.F. "Sculpting: An Interactive Volumetric Modeling Technique." *Computer Graphics (Proceedings of SIGGRAPH'89)*, **25**(4), pp267-274.

[Geiser 1982a] Geiser, E. A., Ariet, M., Conetta, D.A., Lupkiewicz, S.M., Christie, L.G., and Conti, C.R. "Dynamic three-dimensional echocardiographic reconstruction of the intact human left ventricle: Technique and initial observations in patients." *American Heart Journal*, **103**(6), pp1056-1065.

[Geiser 1982b] Geiser, E. A., Christie, L.G., Conetta, D.A., Conti, C.R., and Gossman, G.S. "Mechanical Arm for Spatial Registration of Two-Dimensional Echographic Sections." *Cathet. Cariovasc. Diagn.*, **8**, pp89-101.

[Ghosh 1982] Ghosh, A., Nanda, C.N., and Maurer, G. "Three-Dimensional Reconstruction of Echo-Cardiographics Images Using The Rotation Method." *Ultrasound in Med. & Biol.*, **8**(6), pp655-661.

[Haddad 1991] Haddad, R. A., and Parsons, T.W. *Digital Signal Processing, Theory, Applications, and Hardware*. New York, Computer Science Press.

[Harris 1990] Harris, R. A., Follett, D.H., Halliwell, M, and Wells, P.N.T. "Ultimate limits in ultrasonic imaging resolution." *Ultrasound in Medicine and Biology*, **17**(6), pp547-558.

[Hildreth 1983] Hildreth, E. C. "The Detection of Intensity Changes by Computer and Biological Vision Systems." *Computer Vision, Graphics, and Image Processing*, **22**, pp1-27.

[Hottier 1989] Hottier, F., *Philips Paris Research Lab.* Personal Communication.

[Hottier 1990] Hottier, F., Collet Billon, A. *3D Echography: Status and Perspective*. 3D Imaging in Medicine. Springer-Verlag, pp21-41.

[Itoh 1979] Itoh, M., and Yokoi, H. "A computer-aided three-dimensional display system for ultrasonic diagnosis of a breast tumor." *Ultrasonics*, , pp261-268.

[King 1990] King, D. L., King Jr., D.L., and Shao, M.Y. "Three-Dimensional Spatial Registration and Interactive Display of

Position and Orientation of Real-Time Ultrasound Images." *Journal of Ultrasound Med*, **9**, pp525-532.

[Lalouche 1989] Lalouche, R. C., Bickmore, D., Tessler, F., Mankovich, H.K., and Kangaraloo, H. "Three-dimensional reconstruction of ultrasound images." *SPIE'89, Medical Imaging*, pp59-66.

[Leipnik 1960] Leipnik, R. "The extended entropy uncertainty principle." *Info. Control*, **3**, pp18-25.

[Levoy 1988] Levoy, M. "Display of Surface from Volume Data." *IEEE CG&A*, **8**(5), pp29-37.

[Levoy 1990] Levoy, M. "A Hybrid Ray Tracer for Rendering Polygon and Volume Data." *IEEE CG&A*, **10**(2), pp33-40.

[Liang 1991] Liang, J., Shaw, C., and Green, M. "On Temporal-Spatial Realism in the Virtual Reality Environment." *User Interface Software and Technology, 1991*, Hilton Head, SC., U.S.A., pp19-25.

[Lin 1991] Lin, W., Pizer, S.M., and Johnson, V.E. "Surface Estimation in Ultrasound Images." *Information Processing in Medical Imaging 1991*, Wye, U.K., Springer-Verlag, Heidelberg, pp285-299.

[Matsumoto 1981] Matsumoto, M., Inoue, M., Tamura, S., Tanaka, K., and Abe, H. "Three-Dimensional Echocardiography for Spatial Visualization and Volume Calculation of Cardiac Structures." *J. Clin. Ultrasound*, **9**, pp157-165.

[McCann 1988] McCann, H. A., Sharp, J.S., Kinter, T.M., McEwan, C.N., Barillot, C., and Greenleaf, J.F. "Multidimensional Ultrasonic Imaging for Cardiology." *Proc.IEEE*, **76**(9), pp1063-1073.

[Mills 1990] Mills, P. H., and Fuchs, H. "3D Ultrasound Display Using Optical Tracking." *First Conference on Visualization for Biomedical Computing*, Atlanta, GA, IEEE, pp490-497.

[Miyazawa 1991] Miyazawa, T. "A high-speed integrated rendering for interpreting multiple variable 3D data." *SPIE*, **1459**(5).

[Molnar 1992] Molnar, S., Eyles, J., and Poulton, J. "PixelFlow: High-Speed Rendering Using Image Composition." *Computer Graphics (Proceedings of SIGGRAPH'92)*, ((In this issue)).

[Moritz 1983] Moritz, W. E., Pearlman, A.S., McCabe, D.H., Medema, D.K., Ainsworth, M.E., and Boles, M.S. "An Ultrasonic Techinique for Imaging the Ventricle in Three Dimensions and Calculating Its Volume." *IEEE Trans. Biom. Eng.*, **BME-30**(8), pp482-492.

[Nakamura 1984] Nakamura, S. "Three-Dimensional Digital Display of Ultrasonograms." *IEEE CG&A*, **4**(5), pp36-45.

[Ohbuchi 1990] Ohbuchi, R., and Fuchs, H. "Incremental 3D Ultrasound Imaging from a 2D Scanner." *First Conference on Visualization in Biomedical Computing*, Atlanta, GA, IEEE, pp360-367.

[Ohbuchi 1991] Ohbuchi, R., and Fuchs, H. "Incremental Volume Rendering Algorithm for Interactive 3D Ultrasound Imaging." *Information Processing in Medical Imaging 1991 (Lecture Notes in Computer Science, Springer-Verlag)*, Wye, UK, Springer-Verlag, pp486-500.

[Pini 1990] Pini, R., Monnini, E., Masotti, L., Novins, K. L., Greenberg, D. P., Greppi, B., Cerofolini, M., and Devereux, R. B. "Echocardiographic Three-Dimensional Visualization of the Heart." *3D Imaging in Medicine*, Travemünde, Germany, **F 60**, Springer-Verlag, pp263-274.

[Polhemus 1980] Polhemus. *3Space Isotrak User's Manual.*

[Raichelen 1986] Raichelen, J. S., Trivedi, S.S., Herman, G.T., Sutton, M.G., and Reichek, N. "Dynamic Three Dimensional Reconstruction of the Left Ventricle From Two-Dimensional Echocardiograms." *Journal. Amer. Coll. of Cardiology*, **8**(2), pp364-370.

[Robinett 1991] Robinett, W., and Rolland, J.P. "A Computational Model for the Stereoscopic Optics of a Head-Mounted Display." *Presence*, **1**(1), pp45-62.

[Sabella 1988] Sabella, P. "A Rendering Algorithm for Visualizing 3D Scalar Fields." *Computer Graphics (Proceedings of SIGGRAPH'88)*, **22**(4), pp51-58.

[Shattuck 1984] Shattuck, D. P., Weishenker, M.D., Smith, S.W., and von Ramm, O.T. "Explososcan: A Parallel Processing Technique for High Speed Ultrasound Imaging with Linear Phased Arrays." *JASA*, **75**(4), pp1273-1282.

[Smith 1991] Smith, S. W., Pavy, Jr., S.G., and von Ramm, O.T. "High-Speed Ultrasound Volumetric Imaging System - Part I: Transducer Design and Beam Steering." *IEEE Transaction on Ultrasonics, Ferroelectrics, and Frequency Control*, **38**(2), pp100-108.

[Stickels 1984] Stickels, K. R., and Wann, L.S. "An Analysis of Three-Dimensional Reconstructive Echocardiography." *Ultrasound in Med. & Biol.*, **10**(5), pp575-580.

[Thijssen 1990] Thijssen, J. M., and Oosterveld, B.J. "Texture in Tissue Echograms, Speckle or Information ?" *Journal of Ultrasound in Medicine*, **9**, pp215-229.

[Tomographic Technologies 1991] Tomographic Technologies, G. *Echo-CT.*

[Upson 1988] Upson, C., and Keeler, M. "VBUFFER: Visible Volume Rendering." *ACM Computer Graphics (Proceedings of SIGGRAPH'88)*, **22**(4), pp59-64.

[von Ramm 1991] von Ramm, O. T., Smith, S.W., and Pavy, Jr., H.G. "High-Speed Ultrasound Volumetric Imaging System - Part II: Parallel Processing and Image Display." *IEEE Transaction on Ultrasonics, Ferroelectrics, and Frequency Control*, **38**(2), pp109-115.

[Ward 1992] Ward, M., Azuma, R., Bennett, R., Gottschalk, S., and Fuchs, H. "A Demonstrated Optical Tracker with Scalable Work Area for Head-Mounted Display Systems." *1992 Symposium on Interactive 3D Graphics*, Cambridge, MA., ACM, pp43-52.

[Wells 1977] Wells, P. N. T. *Biomedical ultrasonics*. London, Academic Press.