INTRODUCTION

It is 2010 and you have a very important meeting with your business associates in Chennai. However you have visitors from Japan coming for a mega business deal the same day. Is there any technology by which you can deal with both of them? The answer is yes, and the name of that technology is Tele-Immersion. Tele-Immersion is a technology by which you'll interact instantly with your friend on the other side of the globe through a simulated holographic environment. This technology, which will come along with Internet2, will change the way we work, study and get medical help. It will change the way we live.

Immersion in general means a psycho-physical condition which gives the user the illusion that he acts in an environment which is different from the real one. Thus, from the technical point of view, an immersive system has to meet two requirements. Firstly, the system must be able to isolate the human perception as much as possible (or necessary) from influences of the real word. Secondly, the system has to provide man-machine interfaces which stimulate the human perception in a directive manner such that the desired illusion of a non-existing environment is achieved.

Tele-Immersion (TI) is defined as the integration of audio and video conferencing, via image-based modeling, with collaborative virtual reality (CVR) in the context of data-mining and significant computation. The 3D-effect behind the Tele-Immersion makes it feel like the real thing. The ultimate goal of TI is not merely to reproduce a real face-to-face meeting in every detail, but to provide the "next generation" interface for collaborators, world-wide, to work together in a virtual environment that is seamlessly enhanced by computation and large databases. When participants are tele-immersed, they are able to see and interact with each other and objects in a shared virtual environment.

Tele-Immersion will enable users in different locations to collaborate in a shared, simulated environment as if they were in the same physical room. It's the ultimate synthesis of networking and media technologies to enhance collaborative environments.

To make these new distributed immersive environments a reality, Advanced Network and Services has acted as a catalyst bringing together recognized experts in virtual reality and networking, let by VR pioneer Jaron Lanier, to identify the issues and develop plans to build a national tele-immersive research infrastructure. With goals of accelerating the development of better tools for research and educational collaboration, this plan will be woven into the fabric of many of the Next Generation Internet applications

THE HISTORY

It was in the 1965 that Ivan Sutherland proposed the concept of the ‘Ultimate Display’. It described a graphics display that would allow the user to experience a completely computer-rendered environment. The term Tele-Immersion was first used in October 1996 as the title of a workshop organized by EVL and sponsored by Advanced Network & Services Inc., to bring together researchers in distributed computing, collaboration, VR and networking. At this workshop, specific attention was paid to the future needs of applications in the sciences, engineering and education. In 1998, Abilene, a backbone research project was launched and now serves as the base for Internet2 research. Tele-immersion is the application that will drive forward the research of Internet2.

There are several groups working together on National Tele-Immersion Initiative (NTII) to make this wonderful technology available to the common man.

FIRST FEEL OF TELE-IMMERSION

A Swift investigation revealed that three researchers, led by UNC computer scientists Henry Fuchs and Greg Welch, in May 2000 opened a pair of portals connecting Chapel Hill with Philadelphia and New York. Through these portals, they could peer into the offices of colleagues hundreds of miles away, in life-sized three dimensions and real time. It was as if they had teleported distant chunks of space into their laboratory. The experiment was the first demonstration of tele-immersion, which could radically change the way we communicate over long distances. Tele-immersion will allow people in different parts of the world to submerge themselves in one-another’s presence and feel as if they are sharing the same physical space. It’s the real-world answer to the StarTrek Holodeck, the projection chamber on the Starship Enterprise where crewmembers interact with projected images as if they were real.May’s experiment was the culmination of three years’ work by the National Tele-Immersion Initiative (NTII), a project led by virtual pioneer Jaron Lanier. The test linked three of the members of the group: UNC Chapel Hill, the University of Pennsylvania in Philadelphia, non-profit organization called the Advanced Network and Services in Armonk in New York, were Lanier is Chief Scientist.

At Chapel Hill there were two large screens, hung at right angles above desk, plus projection cameras and head tracking gear. The screens were flat and solid, but once the demo was up and running they looked more like windows. Through the left-hand screen, Welch could see colleagues in Philadelphia as if they were sitting across the desk from him. The right-hand screen did the same for Armonk. When Welch changed point of view, the images shifted in a natural way. If he leaned in, images got larger, if he leaned out, they got smaller. He could even turn his neck to look around the people. To make it work, both target sites were kitted out with arrows of digital cameras to capture images and laser rangefinders to gather positional information. Computers then converted the images into 3D geometrical information and transmitted it to Chapel Hill via Internet2. There, computers reconstructed the images and projectors beamed then into screens.

The images were split and polarized to create a slightly different image to each eye, much like an old-fashioned 3D movie. Welch wore glasses differently oriented polarizing lenses so his left eyes the other, which his brain combined to produce 3D images.

A head-mounted tracker followed Welch's movements and changed the images on the screens accordingly. Like the first transcontinental phone call, the quality was scratchy, also jerky, updating around three times a second rather than 10, the minimum speed needed to capture the full image of facial expressions. It only worked one way: the people in Armonk and Philadelphia couldn't see Chapel Hill.

All this may sound like conventional videoconferencing. But Tele-immersion is much, much more. Where videoconferencing delivers flat images to a screen, Tele-immersion recreates an entire remote environment. Tele-immersion differs significantly from conventional video teleconferencing in that the user's view of the remote environment changes dynamically as he moves his head.

SCIENCE OF TELE-IMMERSION

Tele-immersion essentially takes comprehensive, real-time measures of a person and his surroundings and conveys that information directly to the senses of a person far away. Tele-Immersion (National Tele-immersion Initiative - NTII) will enable users at geographically distributed sites to collaborate in real time in a shared, simulated environment as if they were in the same physical room. It is the ultimate synthesis of networking and media technologies to enhance collaborative environments.

In a tele-immersive environment computers recognize the presence and movements of individuals and objects, track those individuals and images, and then permit them to be projected in realistic, multiple, geographically distributed immersive environments on stereo-immersive surfaces. This requires sampling and resynthesis of the physical environment as well as the users' faces and bodies, which is a new challenge that will move the range of emerging technologies, such as scene depth extraction and warp rendering, to the next level. Tele-immersive environments will therefore facilitate not only interaction between users themselves but also between users and computer-generated models. This will require expanding the boundaries of computer vision, tracking, display, and rendering technologies. As a result, all of this will enable users to achieve a compelling experience and it will lay the groundwork for a higher degree of their inclusion into the entire system.

This new paradigm for human-computer interaction falls into the category of the most advanced network applications and, as such, it is the ultimate technical challenge for Internet2. Further, a secondary objective is to accelerate the development of better tools for research and educational collaboration and promote advances in virtual environment research, which means that this plan will be woven into the fabric of many of the Next Generation Internet applications.

Tele-Immersion is the merging of audio and video conferencing with collaborative virtual reality, data-mining and significant computation systems. As Tele-Immersion often involves distantly located participants all interacting with each other in real-time, it relies heavily on the quality of the underlying networks used to distribute information between participants.

The inevitable issue in any tele-collaborative VR system is a form in which humans and their acting will be represented in the system. In general, there are two ways of representing humans in such systems: using avatars (3D computer graphics models as approximations of humans' embodiments) or using outputs of different 3D acquisition systems (real time 3D "scans" of humans). In order to provide real-time simulation of humans, typical VR system using avatars requires remarkable amount of computational power for the physically based modeling and real-time simulation of the range of human movements and gestures, human skin, hair, textures, clothes and any other detail that might be necessary in particular context. (It would be preferable to have full simulation of humans but some applications might not impose such high requirements. These algorithms are extremely hard and they never cover the entire universe of possible complex gestures and movements so typical for humans.) This approach requires building elaborate 3D models of humans in advance, devising a set of non-trivial algorithms that would perform required simulations, tracking human activities and, at the end, simulating those and presenting them in a visual form to the users, all requested to work in real time. Real-time physically based simulations are very difficult tasks because of their huge computational demands. Because of the limitations present in computer graphics technologies the quality of simulated worlds is still far from real, which is particularly noticeable in simulations of humans. It is extremely hard to represent fine and small details of human appearance and communication clues necessary to achieve a high sense of presence, such as subtle facial gestures, as well as the representations of skin texture and elaborate wrinkles .As a direct consequence people can still easily identify synthetic models and behavior. In a collaborative environment our trust and engagement in communication will depend on whether we believe something represents a real person or not. This, in turn, may have considerable effect on our task performance. In addition to all of this, the process of producing good and inevitably complex 3D model of the real world takes a lot of time and skill, something that is not available everywhere and not with the promptness that is usually needed.

Opposite, 3D acquisition systems invests all computational power not into understanding how the real world works and simulating it, but rather scanning and obtaining 3D visual representation of the same physical world, and reconstructing the closest 3D replica of that world as is. The same system and algorithm is applied on any physical object, and therefore the entire system has more generic approach. Here, again, all stages of the algorithm have to be performed in real time in order to extract 3D geometry from the objects that are moving and changing dynamically over the time. Therefore, 3D models of the world in such systems are acquired on the fly eliminating the need to develop any 3D model a particular application will need in advance. These algorithms, as much as algorithms for simulation of humans, are non-trivial. The real-time requirement always poses additional burden to any system and this makes the task even more difficult. The acquisition systems usually produce certain amount of so-called “3D noise”, is calculated 3D artifacts that occur when the algorithm does not make the correct calculation of the depth information. There is also a possibility of missing parts of 3D information due to the inherent problems that acquisition systems have to deal with (in case of computer vision system, for example, those would be specular features and flat surfaces without texture that might be present on the objects in the real world).

Tele-immersion has an environment called TIDE. TIDE stands for Tele-Immersive Data exploration Environment. The goal of TIDE is to employ Tele-Immersion techniques to create a persistent environment in which collaborators around the world can engage in long-term exploration and analysis of massive scientific data sets. When participants are tele-immersed, they are able to see and interact with other and object in a shared virtual environment. There presence will be depicted by life-like representations of themselves (avatars) that are generated by real-time, image capture and modeling techniques. The environment will persist even when all participants have left it. The environment may autonomously control supercomputing computations, query databases and gather the results for visualization when the participants return. Participants may even leave messages for their colleagues who can replay them as a full audio, video and gestural stream.

All users are separated by hundreds of miles but appear collocated able to see each other as either a video image or as a simplified virtual representation (commonly known as an avatar). Each avatar has arms and hands so that they may convey natural gesture such as pointing areas of interest in the visualization. Digital audio is streamed between the sites to allow them to speak to each other.

TIDE will engage users in CAVEs. ImmersaDesks and desktop workstations around the world connected by the Science and Technology Transit Access Point (STARTAP) –a system of high speed national and international networks. TIDE has three main parts:

- TELE-IMMERSION SERVER (TIS)
- TELE-IMMERSION CLIENT (TIC)
- REMOTE DATA AND COMPUTATIONAL SERVICES

TELE-IMMERSION SERVER

The Tele-Immersion Server's primary responsibility is to create a persistent entry point for the TICs. That is, when a client is connected to the TIS, a user can work synchronously with other users. The environment will persist even when all participants have left it. The server also maintains the consistent state that is shared across all participating TICs. Finally the TIS store the data subsets that are extracted from the external data sources. The data subsets may consists of raw and derived data sets, three dimensional models or images.

TELE-IMMERSION CLIENT

The Tele-Immersion Client (TIC) consists of the VR display device (either CAVE, ImmersaDesk, etc) and the software tools necessary to allow human-in-the loop computational steering, retrieval, visualization, and annotation of data. The TIC also provides the basic capabilities for streaming audio and video, and for rendering avatars to allow participants to communicate effectively to one another while they are immersed in the environment. These capabilities come as part of EVL's Tele-Immersion software framework called CAVERN soft.

REMOTE DATA AND COMPUTATION SERVICES

Remote Data and Computation Services refer to external databases and/or simulations/compute-intensive tasks running on supercomputers or compute clusters that may call upon to participate in a TIDE work session.

The databases may house raw data, or data generated as a result of computations. In most cases the data sets contain too many dimensions and are much too large to visualize entirely. However data mining may be employed to clean the data, to detect specific features in the data, or to extract trends from the data. In some cases as the data mining process may generate models of data, the models can be used to make predictions on missing data points. Furthermore the models can be used to determine which attributes in a multidimensional data-set are the most significant. This is particularly valuable for visualization because the ability to fill the missing data points means a more accurate estimate of the missing data can be made than by simple graphical interpolation. In addition by being able to isolate the most significant attributes, a viewer can prioritize the attributes that they assign to visual features (such as hue, intensity, shape etc) in the visualization. For example Nakayama and Silverman have shown that stereoscopic depth is the most powerful, pre-attentively detected visual feature as compared to other features such as intensity and hue (the features most commonly used in scientific visualizations). This is particularly interesting finding for VR because the medium in which VR resides is inherently stereoscopic. In TIDE the approach taken is to employ data mining algorithms where appropriate as a mean to partition space non-isotropically; to exclude attributes with low significance; to "smart" average attribute values to "summarize" a number of attributes into a single attribute (as a means to reduce dimensionality); and to decimate the data based on the limits of the VR visualization system.

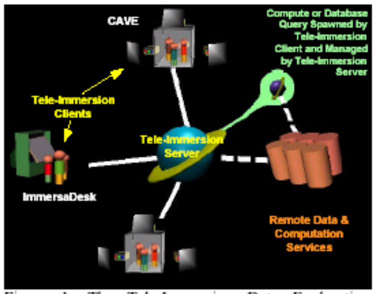


Figure 1. The Tele-Immersion Data Exploration Environment

Initially many of these processes will be controlled on desktop interfaces of PSEs and the resulting decimated data is distributed amongst the collaborators via the Tele-Immersion server. However overtime we will gradually allow an interesting number of these functions to be controlled directly from within the Tele-Immersion environment using three-dimensional interfaces.

To meet the requirements of immersion, it is absolutely necessary to use a large display that covers almost the whole viewing angle of the visual system. In addition, the large display has to be integrated into the usual workspace of an office or a meeting room (see Fig. 2). Thus, the most practicable solution is a desktop-like arrangement with large flat screens like plasma displays with a diagonal of 50 inch and more.Starting from such a desktop-like system and taking into account results from intensive human factors research , further requirements on the presentation of the scene can be formulated as follows :

- conferees are seamlessly integrated in the scene and displayed with at least head,shoulders, torso and arms in natural life-size
- all visual parameters of the scene and the different sources have to be harmonised
- the perspective of the scene is permanently adapted to the current viewpoint of the conferee in front of the display (head motion parallax; look-behind effect)
- eye-contact between two partners talking to each other has to be provided
- gaze from one conferee to another has to be reproduced in a sufficient manner such that everybody can recognise who is looking at whom (e.g.: who is searching for eyecontact)
- voice of a conferee must come from the same direction where he is positioned on the screen

Fig. 2: The Vision of *Immersive Tele-Conference*

Fig. 3: Example for a scene presentation in a immersive conference system

SHARED TABLE ENVIRONMENT

An very attractive way to meet the above requirements is to follow the principle of a shared table environment. It is based on the idea to position the participants consistently in a virtual environment around a shared table. To explain this principle in more detail, Fig. 4 and Fig. 5 show the transmitting and, respectively, the receiving part of a immersive conference terminal.

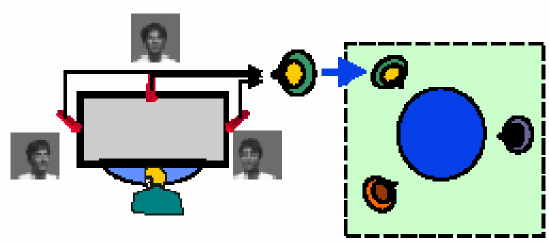


Fig. 4: Signal processing of the transmitting part at an immersive conference terminal

At the transmitting side the conferee in front of the display is captured by multiple cameras and a 3D image of the conferee is derived from this multiview set-up. The 3D images of all participating conferees are then placed virtually around a shared table. Ideally, this is done in a isotropic manner in order to obtain social symmetry. Hence,in the case of three-party conference shown in Fig. 6 the participants form a equilateral triangle. In the case of four parties it would be a square, an equilateral pentagon for a five-party system, and so on.

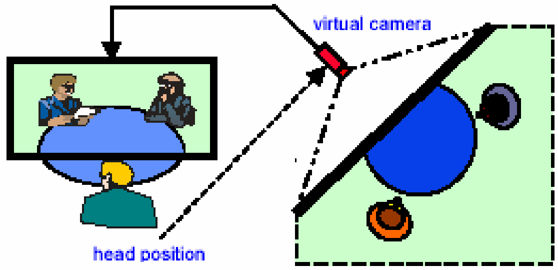


Fig. 5: Signal processing of the receiving part at an immersive conference terminal

At the receiving end this entirely composed 3D scene is rendered onto the 2D display of the terminal by using a virtual camera. The position of the virtual camera coincides with the current position of the conferee's head. For this purpose the head position is permanently registered by a head tracker and the virtual camera is moved with the head. Thus, supposing that the geometrical parameters of the multi-view capture device, the virtual scene and the virtual camera are well fitted to each other, it is ensured that all conferees see the scene under the right perspective view, even while changing their own viewing position. As the consequence, they can also change the view knowingly in order to watch the scene from another perspective, to look behind objects or to look at a previously occluded object. Moreover, all deviations of the conferees' position from a default position, are picked up by the multi-view capture devices. Thus, again supposing well fitted geometrical relations, the 3D image will be moved equivalently in the virtual world and, as a consequence, the other conferees can follow the resulting perspective changes at their displays These circumstances ensure a natural reproduction of eye-contacts and body language in the case of direct face-to-face communication between two partners as well as a natural perspective of this bilateral communication from the position of the third conferee.Last, but not least - the isotropic scene composition and the resulting symmetry enable that the displays can be also placed symmetrically between the partners (i.e. at the middle of the direct viewing axis). Thus, the display works similar to a mirror. Hence, all portrayals appear well balanced in natural life-size at the displays and a psychologically dominance of particular participants is avoided.

From the image processing point of view, the main difficulty of the shared table is to obtain an 3D mage of the conferees for placing it into the virtual environment. It is well known that the 3D shape of an object can be reconstructed if multiple camera views are available and the respective cameras are calibrated . Often, the depth structure is then represented by 3D wire-frames. However, to achieve a natural impression, a wire-frame technique requires a lot of triangles and vertices. Such an detailed wire-frame can only be obtained by complicated image analysis algorithms which suffer from a high computational load. Moreover, real-time rendering of detailed wire-frames is only possible with high-power graphic stations. This all leads to a system complexity which is not desired for the application under study.

A much more attractive approach is a novel view synthesis on the basis of implicit intermediate viewpoint interpolation. Here, the 3D object shape is not reconstructed explicitly, but virtual views are directly calculated from the real camera images by exploiting disparity correspondences. In this context, a very efficient method is the so-called incomplete 3D representation of video objects (IC3D) . In this case, a common texture surface as the one shown in the left image from Fig. 7 is extracted from the available camera views - e.g. the two views from

Fig. 6 -, and the depth information is coded in an associated disparity map as depicted on the right side in Fig. 7. This representation can be encoded like an arbitrarily shaped MPEG-4 video object, where the disparity map is transmitted as an assigned grey scale alpha plane [6]. For synthesis purposes the decoded disparities are scaled according to the user's 3D viewpoint in the virtual scene, and a disparity-controlled projection is carried out. Basically, the original left and right camera views, and also any views from positions on the axis between the two cameras, can be reconstructed. Fig. 8 shows some examples for this synthesis process. Note that the 3D perspective of the person changes with the movement of the virtual camera. One benefit of this technique is its low complexity and high stability compared to algorithms using complete 3D wire-frames. In particular, the rendering of the viewpoint-adapted video object is quite simple and requires a very low and constant CPU time. Due to these properties, it becomes realistic to implement the 3D representation of natural object (e.g. conferees in an immersive tele-conference) as well as virtual view synthesis in real-time.

Fig. 6: Left and right camera view of stereo test sequence CLAUDE

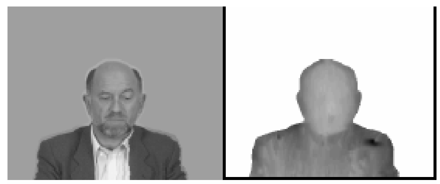


Fig. 7: Texture and disparity maps extracted from stereo test sequence CLAUDE (see **Fig. 6**)

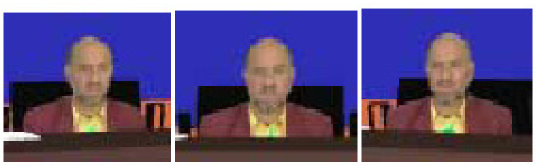


Fig. 8: View-adaptive synthesis in virtual 3D scene (based on representation data in **Fig. 7**)

The conventional IC3D technique has originally been developed for parallel camera set-ups. In this simplified case the vertical component of the disparity vectors is always zero and only horizontal displacements have to be processed. That is the reason why the IC3D approach works with one and not with two disparity maps (see Fig. 6). On the one hand, this limitation to one disparity map is essential as long as MPEG-4 is used for coding because the current MPEG-4 profiles do not support the transmission of more than one disparity map. On the other hand, the restriction to parallel camera set-ups is no longer possible in immersive tele-conferencing scenarios. An immersing system requires large displays and short viewing distances. Therefore the distance between the cameras becomes quite large and the cameras have to be mounted in a strongly convergent set-up in order to capture the same location in front of the display (see Fig. 4). Nevertheless, the techniques of IC3D can be extended to this generalised situation, although disparity correspondences are 2-dimensional in strongly convergent camera set-ups. To explain this IC3D extension in detail, Fig. 9 shows images of a convergent stereo pair. The mounting of the cameras is similar to the one sketched in Fig. 4. The left image refers to a camera at the top of the display, whereas the other has been captured from its left border.

It is well known from epipolar geometry that the disparity correspondences follow so-called epipolar lines which can be derived from the fixed geometry of the convergent camera set-up (see black lines in Fig.9). Due to this epipolar constraint, 2- dimensional disparity correspondences can always be projected onto 1-dimensional displacements along epipolar lines. In addition, it is possible to warp the images in such a way that the epipolar lines become horizontal. Basically, this means that the cameras of the convergent set-up are virtually rotated until they would form a parallel set-up. The effect of this so-called rectification process is shown in Fig. 10.

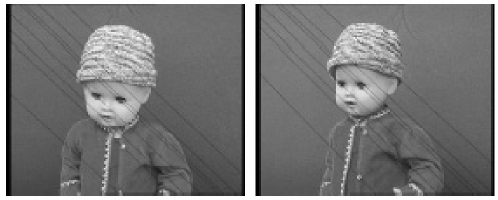


Fig. 9: Two images of a strongly convergent stereo rig referring to a real conference set-up

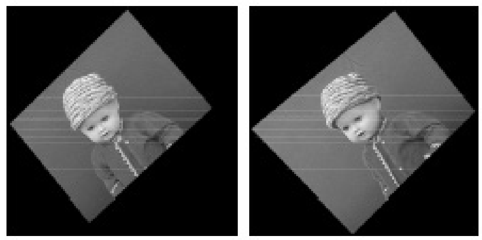


Fig. 10: Rectification of images from **Fig. 9**

Another important extension of IC3D is the usage of tri-linear warping for the novel view synthesis – a quite new interpolation technique which is based on the theory of the trifocal tensor . In contrast to conventional IC3D, this extension allows a synthesis of arbitrary views outside of the baseline between the two cameras. Fig. 11 shows two examples obtained by trilinear warping. Note that the virtual camera has rendered perspectives which are quite different from the original views in Fig. 9.

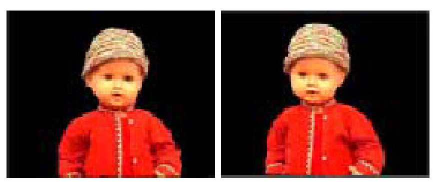


Fig.11: Two virtual views outside the base-line

The above considerations have shown that that 3D representation in immersive tele-conference can efficiently be achieved by using extended IC3D techniques. The results on the synthesis quality are promising and it can be assumed that the whole processing chain including segmentation (background separation), rectification, disparity estimation (along horizontal scan lines), MPEG-4 based IC3D en- and decoding and adaptive view synthesis based on trilinear warping can be implemented in real-time.

VIRTUAL REALITY (VR) DISPLAY DEVICES

As stated earlier VR display devices are mainly four in number. They are listed below

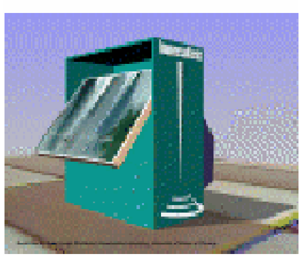
- CAVE
- Immersa Desk
- Desktop Workstation
- Infinity Wall

CAVE

The CAVE virtual reality system is a 10 foot-cubed room that is projected with stereoscopic images creating the illusion that objects appear to co-exist with the user in the room. The CAVE™ is a multi-person, room-sized, high-resolution, 3D video and audio environment. Graphics are projected in stereo onto three walls and the floor. And viewed with stereo glasses. As a viewer wearing a location sensor moves within its display boundaries, the correct perspective and stereo projections of the environment be constantly updated, so the image moves with and surrounds the viewer to achieve immersion.

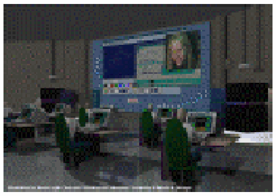
IMMERSA DESK

THE Immersa Desk™ is a drafting-table format version of the CAVE. When folded up, it fits through a standard institutional door, and deploys into a 6'x 8' footprint. It requires a graphics engine of the SGI Onyx or Octane class, one projector, and no architectural modifications to the workspace. The Immersa Desk is software compatible with the CAVE library.



INFINITY WALL

The infinity wall is derivative of the Power Wall, a research effort of Paul Woodward at the University of Minnesota. The Power Wall achieves very high display resolution through parallelism, building up a single image from an array of display panels projected from the rear into a single screen. High-speed playback of previously rendered images is possible by attaching extremely fast disk subsystems, accessed in parallel, to an Onyx. The infinity wall is a simple PowerWall that has tracking and stereo; it is CAVE library compatible.



DESKTOP WORKSTATION

The desktop workstation displays a data-flow model that can be used to construct the visualization that is shared between all the three display devices. The participants in the VR displays can use three-dimensional tools to directly manipulate the visualization.

For example in the CAVE a user is changing the isosurface value in the data-set. These changes are automatically propagated to all other visualization displays. In the meantime the Immersa Desk user, noticing an anomaly in the data-set, inserts an annotation in the data-set as s reminder to return to more closely examine the region. Closer examination of the regions achieved by instructing a remote rendering server consisting of multiple giga-bytes of RAM and tera bytes of disk space, to render the images in full detail as a stereoscopic animation sequence. These animations will take some time to generate and so the users continue to examine other spaces of the data set. Eventually the rendering is complete and the remote server streams the animation to each of the visualization clients for viewing.

CORRESPONDENCES/DEPENDENCIES: TELE-IMMERSIVE DEVICE DESIGN CONCEPTS

A. Motivation for Desktop/Office-Sized VR Display Devices

The VR community needs to conduct research in third-generation VR devices to construct software-compatible,variable-resolution and desktop/office-sized prototypes and products, which evolve over time as technology improves and needs become more demanding; the hardware now needs to be made smaller, higher resolution,and more adaptable to the human and his/her workspace. The recommendation is to procure, evaluate, and integrate a variety of emerging display devices, such as large color plasma displays, LCD projectors, LED panels,Digital Light Valves (DLVs), Grating Light Valves (GLVs), and Digital Micro Mirror Displays (DMDs).To construct the tele-immersive office workspace, one would want affordable wall-sized high-resolution border-less displays with low lag and undiminished image intensity when viewed at an angle. Given that such a display does not exist today, we must rather learn from assembling new VR systems from available components. We must push screen technology development by creating pressure from the computational science and engineering communities with compelling applications projects.
These are the devices which addresses different major issues in the tele-immersion/VR human computer interface:

- ImmersaDesk3 Plasma Panel Desktop VR
- Personal Augmented Reality Immersive System (PARIS)
- Personal Penta Panel (P3)
- Totally Active Workspace (TAWS)
- CyberCeiling
- CAVEscope

B. New Immersive Display Technologies

In the context of building new VR devices, the community needs to investigate the viability, flexibility of operation and breadth of application of the following new display technologies as compared to current 3-tube projector systems:

Liquid Crystal Display (LCD) projectors and panels. They are achieving better resolution now (1280x1024), but have too high lag to be used for stereo unless two projectors are used with shutters.

Digital Micro-mirror Displays (DMDs). These are good resolution (1280x1024), and theoretically fast enough for stereo, but the supplied firmware does not support stereo.

Plasma panel displays. These are low-medium resolution (800x480) but probably fast enough to do stereo with the proper driver electronics. These displays have electronics mounted around their edges that make border-less multi-screen configurations a challenge.

Light Emitting Diode (LED) displays. These are low resolution right now (e.g., 208x272 and 320x192) but bright and border-less, in principle.

Digital Light Valve (DLV) displays. These new desktop projection displays have latency problems for stereo use; they can switch fast enough but do not go to black in the required time. A 2Kx2K resolution version has been built.

Grating Light Valve (GLV) displays. Recently demonstrated in prototype form, this laser-driven microelectromechanical display is capable of HDTV resolution at 96Hz, very promising for VR. Switching speeds are extremely low, allowing a linear array of deflectable ribbon picture elements to scan out an image

C.ImmersaDesk3

C.1.Plasma Panel Desktop Device

1998, The ImmersaDesk3, Electronic Visualization Laboratory, University of Illinois at Chicago

ImmersaDesks and Responsive Workbenches are large because the available 3-tube projection technology has a limit to how small the screen can get approximately 6’ diagonal). Rear projection distances are significant, even when folded with mirrors, and the projector itself is quite large and heavy. Both of these devices are sized for a laboratory, and are too large for a typical faculty office or cubicle. We built a prototype device, called the ImmersaDesk3 to test the plasma panel technology currently available at 640x480 resolution for US$10,000. The ImmersaDesk3 is configured so a user can position the screen at any angle from horizontal to vertical, forward or back, on the desk. The angle can be measured automatically so that the correct perspective view of the computer-generated images for the tracked user is presented. Cameras can be added to this configuration to make image/gesture recognition, tether-less tracking and tele-immersion experiments possible. Given its configuration flexibility, the ImmersaDesk3 is also amenable to the integration of haptic (tactile input/output) devices. We built our system around the Fujitsu PDS4201U-H Plasmavision display panel. The Plasmavision has an active display area of 36x20 inches (in a 16:9 aspect ratio); the entire panel is 41x25x6 inches and weighs 80 pounds. We mounted the Plasmavision on a modified office desk.To accommodate different applications and for greater flexibility, we wanted to be able to position the screen vertically (perpendicular to the desktop), horizontally (flat on the desktop), or at an angle in between. The panel is too heavy for users to shift easily, so we mounted it on hydraulic supports with a hand crank to adjust the angle.

C.2. Problems Encountered with Plasmavision Plasma Panel Displays

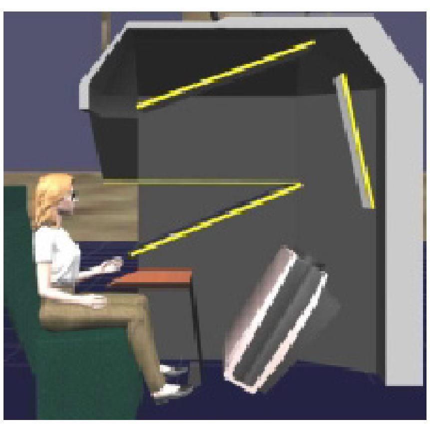
The Plasmavision outputs a 30Hz, interlaced, NTSC resolution image. When this is used for a stereoscopic display,each eye is only seeing a 30Hz signal, and the flicker is very noticeable; prolonged exposure can give many users headaches. Using the NTSC field-interleaved format for stereo yields only 640x240 pixel resolution for each eye’s image. We also found that the red and green phosphors do not decay quickly enough. When we look at a stereo test pattern, which displays separate red, green, and blue color bars for each eye, only the blue bar is sufficiently extinguished; the red and green bars are still visible to the wrong eye. In an informal test of 16 users at SIGGRAPH '98, we noted that 25% of them could fuse the full-color images, while 50% could only fuse the images when the red and green channels were disabled, so that the images were just shades of blue; 25% of the users could not fuse the images at all.

The Plasmavision is electromagnetically noisy, so much so that it interferes with the accuracy of magnetic tracking systems.

Despite these issues, the Plasmavision had advantages. Application developers were very pleased with the size of the display as well as its brightness and color quality, compared to a video projector. Its size gives a much larger angle of view than a conventional monitor, yet it fits well on a desktop, something a 42" monitor or projection system cannot do.Current plasma panel technology has severe limitations as a stereo display device for projection-based VR systems.The inability to easily sync the plasma panel to shutter glasses and the red/green phosphor decay problem preclude clear stereo. The low resolution and 30Hz frame rate also prevent current panels from being serious contenders in this field. Although flat panels can significantly save space, larger display systems would need larger panels or tiled panels. Current plasma panels have borders that prevent seamless tiling.Nevertheless, the concept of a wide-field-of-view, desktop VR system and space-saving flat panel technology for CAVEs and other large displays is appealing. We look forward to improvements in flat panel technology as it evolves.

D. Personal Augmented Reality Immersive System (PARIS)

Twenty years ago, Ken Knowlton created a see-through display for Bell Labs using a half-silvered mirror mounted at an angle in front of a telephone operator. The monitor driving the display was positioned above the desk facing down so that its image of a virtual keyboard could be superimposed on an operator's hands working under the mirror. The keycaps on the operator's physical keyboard could be dynamically relabeled to match the task of completing a call as it progressed. Devices that align computer imagery with the user's viewable environment, like Knowlton's, are examples of augmented reality, or see-through VR. More recently, researchers at the National University of Singapore's Institute of Systems Science built a stereo device of similar plan using a Silicon Graphics' monitor, a well-executed configuration for working with small parts in high-resolution VR . Neither of these systems provides tracking, but rather assumes the user to be in a fixed and seated position.



Artist renderings of PARIS system: Cut-away side view of PARIS system (left) and over-the-shoulder view (right).

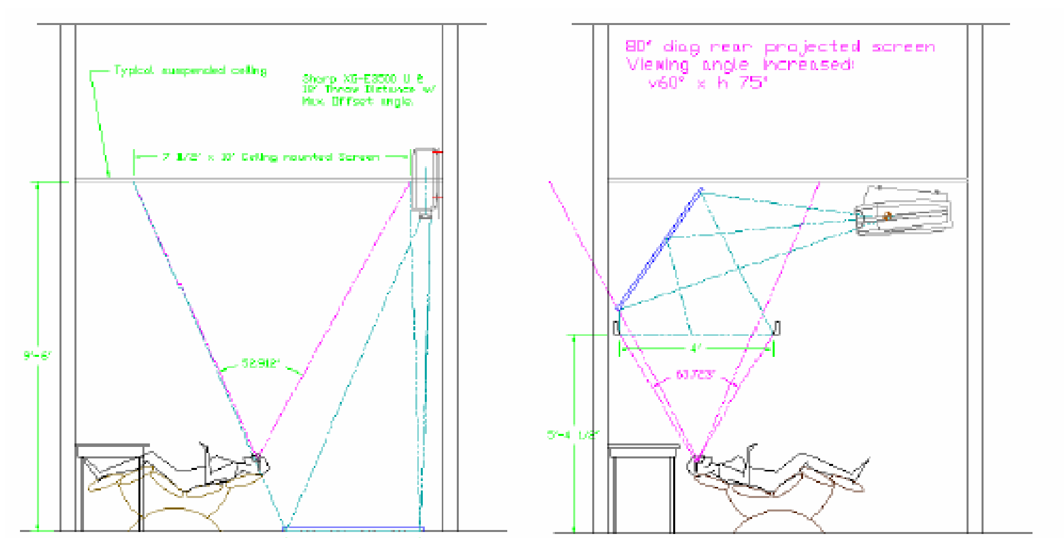
A keyboard is integrated, and that suitably positioned cameras can capture facial expressions and head position. Gesture recognition can come from tracking, as well as processed camera input. Audio support can be used for voice recognition and generation as well as for recording and tele-immersion sessions.

Two 1280x1024 LCD projectors with electronic shutters compatible with active glasses to achieve stereo separation are used.

We can also use PARIS to prototype passive (polarized) stereo since we can polarize the two projector outputs,allowing very inexpensive and lightweight glasses to be incorporated, an important feature for use in museums and schools. If plasma or LED panel displays ever have excellent brightness, stereo speeds, and high-resolution, these would be preferable devices to adapt.

E. CyberCeilings, designed for the Last Unused Projection Surface

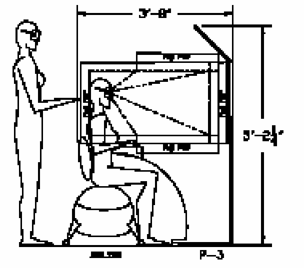
.



Schematic of CyberCeiling

In trying to fit large screens and VR into offices, use of overhead space or the ceiling is conceivable, and has distinct advantages in hospital patient settings, assuming the room can be made dark enough. The drawings below indicate some options for ceiling-mounted front projection with a mirror on the floor, and a smaller, rear projection overhead display. Different lensing can alter the projection distances in the former example. The chair shown is a commercially available executive motorized recliner, but could be replaced by a bed in a hospital setting. This configuration has the benefit that the user may not need to be tracked since body position is fixed and head rotation is not accounted for in projected VR environments

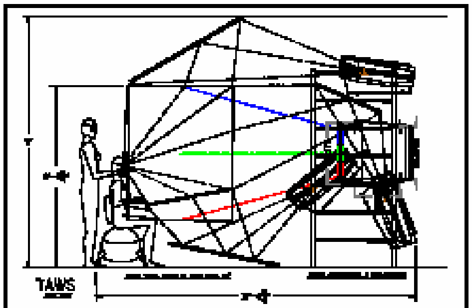
F. Personal Penta Panel (P3) or Dilbert's Dream



P3. Schematic by Greg Dawe, EVL, 1998.

Dilbert's Dream is conceived as an office cubicle whose walls and desk are made from border-less stereo-capable high resolution panels, not,unfortunately, obtainable in the current millenium.Alternatively, we are proposing a "desktop" cubicle. The Personal Penta Panel (P3) is a box made out of 42" diagonal plasma panels. The user places his/her tracked head and hands into the box of screens and is presented with a surround (non-stereo) view. Each panel would have a frame around it, creating seams between screens that would be difficult to eliminate. There are, however, optical methods to relay an image a few inches forward, which could be used to (mostly) eliminate the effects of the frames. Such a device would be useful for all but very close viewing, even in non-stereo, as we wait for the needed technological improvements in panels.

G. Totally Active Work Space (TAWS)



TAWS. Schematic by Greg Dawe, EVL, 1998.

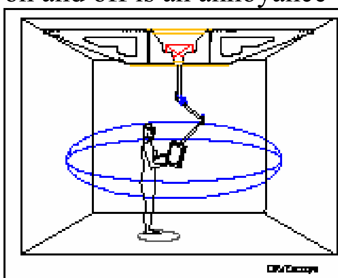
•

Pending the availability of suitable plasma, LCD, or LED panels, we have built screens into a rear-projection desktop structure to simulate the Totally Active Work Space (TAWS)—the ultimate Dilbert's Dream or cubisphere. TAWS is large enough for two colleagues to share the workspace when need be. EVL has been modifying its LCD shutter glasses to run at 160Hz, so that four lenses (in two sets of glasses) can operate almost flicker-free at 40Hz each. This capability, which we call duo-view, allows two tracked users of the same display to see the image in correct perspective and size, essential for sharing a workspace. Research into screen materials is needed because the de-polarization that comes from looking at screens at very oblique angles creates ghosting that is more an issue with duoview than normal stereo.

H. CAVEscope: Simulating Variable Resolution Displays

All projection-based VR devices trade off wide angle of view for resolution.Human vision is acute only for a very narrow angle, the ~five degrees of vision falling on the fovea. It would be desirable, therefore, to have adaptive resolution displays that, given eye tracking, could match human visual acuity in the area of the screen in this five degree angle of view. In stereo, graphics engines currently achieve a resolution of 1280x1024 spread across 5 to 10 feet, a rather less-than-crisp display. Software techniques can be used to render more detail in the area of interest,but resolution itself cannot improve. The projectors now available are not built to handle the dynamic horizontal scanning fluctuations needed for variable resolution display, and neither are the display engines. CAVEscope,however, provides a way to simulate variable resolution in a projection VR setting. We suggest providing a high resolution (e.g., 1024x768 or 1280x1024) LCD display that one can move into the area of detailed interest. Such a display would be like a portal into a higher-resolution space. It would be suspended in the projection-based VR space by a counterweighted mechanism, much like an X-ray machine in a dentist's office. One would navigate in the VR space as normal, with low-resolution surround vision,but pull the CAVEscope into place when high resolution examination is desired. The CAVEscope would be tracked so that it would present the proper perspective projection. Touch screen technology could also be available for user input. A miniature television camera mounted on the CAVEscope could enable tele-conferencing. Users can see and talk to each other using CAVEscopes, or position their devices for coverage relevant to the task at hand. CAVEscope combines the intuitive navigational capabilities of projection-based VR with the detailed view of the LCD portal, all under user control.

CAVEscope should also be usable in an office setting with front projection VR on the office walls and desktop, such as has been proposed by the Advanced Network and Services-sponsored National Tele-Immersion Initiative.Since the wall projections are used mainly for navigation and context, not for detail work, the quality of the projected images could be less than optimal, as long as the CAVEscope image is suitably bright and sharp.Since LCD panel technology does not permit Crystal Eyes-type stereo (due to high lag) at this point, we will need to work with a mono image, pending the availability of a compatible stereo-capable panel in the future. Taking the stereo glasses on and off is an annoyance TAWS



CAVEscope. Schematic by Greg Dawe, EVL, 1998.

Tracked hand-held panels have been suggested as portals into virtual and augmented reality spaces for some time,although, on videotape, the concept is simulated with chroma keying. Discovering where to look in virtual space is a large part of the problem with narrow-angle-of-view devices like panels held at arms length, VR binoculars, or even head-mounted displays. CAVEscope affords the user both the navigational and wide field of view of projectionbased VR with a real-time high-resolution inspection capability. Since CAVEscope has its own rendering engine, the software can be tuned to provide much more detailed rendering in the designated area of interest, which could even be behind or above the user where there are no projected screens!In addition, the user can easily enough freeze the motion and build up the display or re-render it with

ray tracing, a type of successive refinement not normally usable in VR. We believe these notions will provide enhanced performance in accuracy, resolving power, flexibility of operation, user friendliness and navigation for scientists and
engineers using projection-based VR for discovery and observation.

APPLICATIONS OF TELE-IMMERSION

The list of applications of Tele-Immersion is very large. Some of them are:

- Interacting with friends miles away in a simulated holographic environment.
- Tele-Immersion can be of immerse use in medical industry.
- Tele-Immersion also finds its applications in the field of education.

A large set of applications can be managed, depending on the position of infinity wall:

1. Full-scaled model preview in industrial application (aircrafts, cars) when all three modules are lined up (180^0).
2. Flight simulation scenarios visualization when side-modules are in 135^0 layout;
3. Immersive simulations when side-modules are in the “room setup” (90^0).

USES IN EDUCATION

It can be used to bring together students at remote sites in a single environment. With Tele-Immersion students can access data from remote locations. Internet2 will provide accesses to digital libraries and virtual labs.

Exchange of culture is possible without travel.

MEDICAL APPLICATIONS

- 3D surgical learning for virtual operations is possible using this technology.
- In future real surgery can be carried out on real patients.
- It could be life saving if the patient is in need of specific care.
- It gives surgeons the ability to superimpose anatomic images right on their patients while they are being operated on. Surgeons get a chance to learn complex situations before they actually treat their patients.

Figure 2: Local student surgeon in near-field observes and interacts with annotated surgical procedure described by the remote teaching surgeon on the right. (Sketch by Andrei State, UNC-Chapel Hill.)

CHALLENGES OF TELE-IMMERSION

Tele-Immersion has emerged as a high-end driver for the quality of service (QoS), bandwidth, and reservation efforts envisioned by the NGI and Internet2 leadership. From a networking perspective, Tele-Immersion is a very challenging technology for several reasons:

- The networks must be in place and tuned to support high-bandwidth applications
- Low latency, needed for 2-way collaboration, is hard to specify and guarantee given current middleware.
- The speed of light in fiber itself is a limiting factor over transcontinental and transoceanic distances.
- Multicast, unicast, reliable and unreliable data transmissions (called “flows”) need to be provided for and managed by the networks and the operating systems of supercomputer-class workstations.
- Real-time considerations for video and audio reconstruction (“streaming”) are critical to achieving the feel of telepresence, whether synchronous or recorded and played back.
- The computers, too, are bandwidth limited with regard to handling very large data for collaboration.
- Simulation and data mining are open-ended in computational and bandwidth needs-there will never be quite enough computing and bits/second to fully analyze, and simulate reality for scientific purposes.

In Layman's language the realization of Tele-Immersion is impossible today due to:

1. The non-availability of high-speed networks.
2. The non-availability of supercomputers
3. Large network bandwidth requirements.

SOLUTION

The first two basic problems can be overcome when Internet-2 will come into picture later and third problem can be overcome by the fast development of image compression techniques.

ABOUT INTERNET-2

- Internet–2 is not a separate physical network and will not replace the current Internet. It is not for profit consortium consisting of 200 US universities, Industries and is directly under the control of US govt.
- Internet-2 is for developing and deploying advanced network applications and technology, accelerating the creation of tomorrow's Internet.
- Internet-2 enables completely new applications such as digital libraries, virtual laboratories, distance-independent learning and Tele-Immersion.
- A key goal of this effort is to accelerate the diffusion of advanced Internet technology, in particular into the commercial sector.

FUTURE DEVELOPMENTS

The Tele-Immersion system of 2010 would ideally:

- Support one or more flat panels/projections with -ultra color resolution (say 5000x5000).
- Be stereo capable without special glasses.
- Have several built-in micro-cameras and microphones.
- Have tether-less, low-latency, high-accuracy tracking.
- Network to teraflop computing via multi-gigabit optical switches with low latency.
- Have exquisite directional sound capability.
- Be available in a range of compatible hardware and software configurations.
- Have gaze-directed or gesture-directed variable resolution and quality of rendering.
- Incorporate AI-based predictive models to compensate for latency and anticipate user transitions.
- Use a range of sophisticated hap tic devices to couple to human movement and touch.
- Accommodate disabled and fatigued users in the spirit of the Every Citizen Interference to the NTII .

CONCLUSION

Tele-Immersion is a technology that is certainly going to bring a new revolution in the world and let us all hope that this technology reaches the world in its full flow as quickly as possible.

The tele-immersive portal could be characterized as a telephone of the future, a kind of interactive user interface that could be employed in a number of applications that use high bandwidth networks as a platform. Typical applications that will benefit greatly include organizing tele-meetings, a variety of medical applications (preoperative planning, tele-assisted surgery, advanced surgical training, tele-diagnostics),
tele-collaborative design, computer supported training and education, true 3D interactive video and most certainly entertainment.

REFERENCES

Herman Towles, "National Tele-Immersion Initiative: Sharing Presence with Distant Colleagues ", presentation given at Internet2 Spring Meeting held in Washington DC, March 7-9, 2001
Amela Sadagic, "National Tele-immersion Initiative: Towards Compelling Tele-Immersive Collaborative Environments", presentation given at the conference "Medicine Meets Virtual Reality 2001", Newport Beach, CA, Jan 2001

Jaron Lanier, "The National Teleimmersion Initiative", presentation given for US Army, The Pentagon
Amela Sadagic: "National Tele-immersion Initiative: Visions and Challenges", presentation for invited talk at the Second Workshop on VR in Sweden: Future Directions in Virtual Environments - FDIVE II, 11-12 November 1999

Joint presentation by Brown University and NPS: "SOFT - Software Framework for Teleimmersion" and "SOFT Networking and Scalable Virtual Environments"

Ben Teitelbaum, "Internet2 QBone: Building a Testbed for IP Differentiated Services"

Websites

www.electronicsforu.com
www.ieee.org
www.NTII.com
www.advancedorg.tele-immersion.com

www.newscientist.com

APPENDIX

Tele-Immersive Data Flow Types							
Type	Latency	Bandwidth	Reliable	Multicast	Security	Streaming	DynQoS
Control	<30 ms	64 Kb/s	Yes	No	High	No	Low
Text	<100 ms	64 Kb/s	Yes	No	Medium	No	Low
Audio	<30 ms	Nx128 kb/s	No	Yes	Medium	Yes	Medium
Video	<100 ms	Nx5 Kb/s	No	Yes	Low	Yes	Medium
Tracking	<10 ms	Nx128 Kb/s	No	Yes	Low	Yes	Medium
Database	<100 ms	>1 GB/s	Yes	Maybe	Medium	No	High
Simulation	<30 ms	>1GB/s	Mixed	Maybe	Medium	Maybe	High
Haptic	<10 ms	>1 Mb/s	Mixed	Maybe	Low	Maybe	High
Rendering	<30 ms	>1GB/s	No	Maybe	Low	Maybe	Medium

The columns represent flow-type attributes:

- **cy** is the sum of all delays in the system, from the speed light in fiber, to operating system overhead, to tracker settling time and screen refresh.
- **Bandwidth** is the bits/second the system can transmit
- **Reliable** flows are verified and transmitted if bad.
- **Multicast flows** go to more than one site at once.

- **Security** involves encryption overhead that may or may not be warranted or legal.
- **Streaming** data is a constant flow of information over time, as with video, audio and tracking.
- **Dynamic QoS** can provide ways to service bursty high-bandwidth needs on request.

The rows indicate the data flow types:

- **Control Information** consists of data that is used to manage the tele-immersion session, to authenticate the users or processors, to launch processors, to control the display or tracking systems, and to communicate out of band between the world servers and the VR systems.
- **Text** provides simple communication capability within collaborative sessions for simple note taking and passing. Text can also command UNIX processes driving the environments.
- **Audio** gives ambient auditory cues, allows voice communications among users, and is used to issue commands via voice recognition and speech synthesis. A typical application may use multiple audio streams.
- **Video** can allow teleconferencing or remote monitoring displayed within the virtual world. Synthetic 2D animated bitmaps in video format have application as well.
- **Tracking** is achieved with location and orientation sensors, and captures the position and orientation of the user. Typically this data is streamed to the computer responsible for computing the perspective of the scene. Tele-immersion requires tracking data to br shared among sites. Most VR systems had only head and track; future systems will have many more sensors to track more complex posture and body motions.
- **Database** is the heart of tele-immersion application world. The database consists of the graphical models of virtual scenes, objects, and data, and since the database is used to provide the models that are rendered, it must be maintained in a coherent state across multiple sites. Database must be as simple as shared VRML files or as complex as multi-terabyte scientific datasets, VR extensions of video serving, or even Virtual Director recorded sessions. (Virtual Director is a joint EVL/NCSA development project [29].)
- **Simulation** provides the basis for dynamics behaviors, like responding to the users' actions. Small-scale simulations often run on the computer also generating the VR experience, but frequently the simulation will need a dedicated supercomputer [28]. User input is captured and transmitted to the simulation via the network and the simulation will generate an update, which is then propagated to each user site for local.

Table 2: **Access patterns for tele-immersive data**

Writers-readers	Description	Distribution	Example
1-1	Point-to-point connection, with data streamed for a single writer to a single reader.	Socket, stream	HPCresource streams simulation result to a VE server, which transforms them to a polygonrepresentation for use by end-user display machines.
1-m	Single writer to multiple readers.	Multicast, broadcast.	Audio stream from a single person in a VE. Any number of participants may choose to listen to the stream, to hear what the person is saying.
m-n (m<n)	Privileged writers to multiple readers. A small group of writers cooperate to maintain a data structure, which is read by many.	DSM, multicast, RPC/RMI	Two workgroups of users cooperate with a VE. Each group has its own local server. The two servers run-time services used to sync clocks on local machines.
m-n (m=n)	Peer-to-peer sharing of data.	DSM, mobile-object, RPC/RMI	Complex CAD model being manipulated by multiple users within VE.Individual components of the model may be changed by any user. Multiple users may change separate components at the same time. Model is stored within a distributed shared memory, which all users can access.
m-1	Multiple writers accessing a common data structure.	RPC/RMI	A simulation running on a remote HPC resource. Any participant may direct elements of the simulation.