°

Tele-immersion: Tomorrow's Teleconferencing

°

Real-time transmission of real-life environments may be Internet 2's most visionary application

by Steve Ditlea

When Internet 2-level network performance becomes common place at some point in the future, which applications will make the best use of this giant leap in bandwidth? Virtual laboratories, digital libraries, and distance-independent learning are among some of the advanced applications currently being explored. Jaron Lanier, who helped lead in the development of virtual reality during the 1980s, is now guiding an attempt to validate the Net of tomorrow with a nascent technology known as tele-immersion: long-distance transmission of life-size, three-dimensional synthesized scenes, accurately sampled and rendered in real time using advanced computer graphics and vision techniques. Such replication of visual content in large volumes of everyday reality should lead to more naturalistic teleconferencing work environments (and less business travel), greater fidelity in relaying news and entertainment events (high-def will seem positively low-res), and even Star Trek Holodeck-like telepresence in remote locales (beam me up, Jaron).

After three years of effort, Lanier and his colleagues are showing their first proof-of-concept demo, a three-way virtual meeting that makes today's videoconferencing look like the 8mm movies of yesteryear. At the University of North Carolina (UNC) at Chapel Hill, invited visitors can witness on two walls the life-size, real-time images of researchers seated at their desks at the University of Pennsylvania in Philadelphia, and at the Armonk, New York, laboratory of project sponsor Advanced Network & Services, where Lanier serves as the initiative's chief scientist. In what is dubbed the "tele-cubicle," a UNC participant wearing polarizing glasses and a silvery head-tracking device can move around and see a computer-generated 3D stereoscopic image of the other two teleconferencers, whereby the visual content of a block of space surrounding each participant's upper body and some adjoining workspace is essentially reproduced with computer graphics. This results in a more fully dimensional and compressible depiction of such real-world environments than is possible with existing video technology. Though the demo is far from perfect, with two-way transmissions instead of three-way, and jitters visible in the displays, it nonetheless marks the accomplishment of what Lanier characterizes as "the ultimate convergence of the real world and computer graphics."


A University of North Carolina researcher meets virtually with colleagues from Pennsylvania and New York-life-size and in 3D. Tele-immersion's realtime rendering is accomplished through advanced computer vision, graphics, and network techniques.

The National Tele-immersion Initiative (NTII), as this ongoing project is known, was first proposed by Allan H. Weis, founder of Advanced Network & Services, one of the builders of the Internet's original back bone. With the proceeds from the company's sale to America Online, Advanced Network & Services became a research institution, funding work on leading-edge uses of computer network technology. Lanier was then hired and given a staff based in Armonk, along with a budget to provide grants to university re searchers. Principal investigators include Henry Fuchs of UNC, Andries van Dam of Brown University, and Ruzena Bajcsy at the University of Pennsylvania. Researchers at the Naval Postgraduate School, Carnegie Mellon University, Columbia University, the University of Illinois at Chicago, and the University of Southern California were also involved with the project.

Overcoming Technical Issues

"There are quite a few difficulties involved," says Lanier. "The first is, how do you sense a remote place in real time fast enough and with the kind of quality so you can re-render it and make it look good? There you have a mixture of vision problems, graphics problems, and networking problems all in one bundle. Then beyond that, how do you create a physical viewing configuration that supports the illusion of reality?"

Starting with the scene acquisition of a complete three-dimensional representation independent of any one perspective, Lanier and his team opted for vision techniques using a "sea" of multiple video cameras. For the best trade-off of quality versus performance, they employed overlapping trios of video cameras-with more redundancy of scene information allowing fuller coverage of visual surfaces than with pairs. In the advanced teleconferencing application, seven video cameras are arranged in a 120-degree arc in front of each subject, with the cameras sampled in overlapping triads-the optimumarray, given network constraints. Lanier's goal is for higher resolution images with a 60-camera array that can be used in medical tele-mentoring.
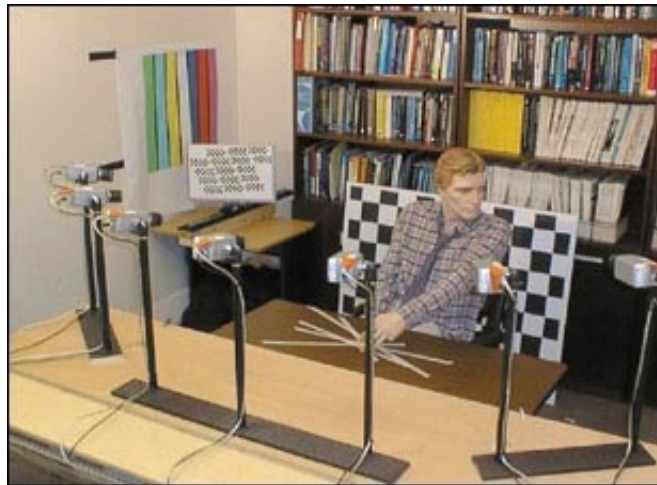
Using an array of seven video cameras, researchers captured visual and dimensional "test" content of a scene containing a human dummy and a checkerboard calibration pattern.

Committed to off-the-shelf parts so other users can deploy tele-immersion, the NTII team chose research-grade Sony digital video cameras with fast IEEE 1394 connectors, also marketed by Apple Computer under the FireWire name. "We want to get to the point where standard configurations are available for tele-immersion," Lanier explains. "One of our next tasks is to write a cookbook, documenting what we're doing for others to follow and build on."

Other Obstacles

Among the graphics problems to be overcome is that of surface ambiguities that the human brain resolves effortlessly but computers have difficulty parsing. Under normal room illumination, such as that from overhead fluorescent bulbs or desktop incandescent lamps, a bare wall displays no surface textures, confounding pattern-recognition software.

In an attempt to accurately register featureless walls, and shiny objects for that matter, the team is exploring a technique developed at UNC called "imperceptible structured light." With this technology, along with a room's existing lighting, a scene is lit by what appear to be additional nor mal spotlights, but embedded with in this illumination are structure-monochromatic geometric pat terns. The patterns are dithered to be imperceptible to people, but a synchronized video camera can pick up these visual calibration patterns so that ambiguities in the shape, color, and reflectivity of objects (such as screens, blank walls, doorways, and even a person's forehead) are eliminated.

The stated goal of real-time scene acquisition and transmission accentuates graphics and networking challenges. For example, sampling rates vary with the complexity and movement in a scene as pattern-recognition algorithms model visual content captured by the video camera arrays; a chosen algorithm's accuracy must be weighed against any lag time it may introduce. To cut down on computational lag, some optimization is performed, such as segmenting a scene so more resources can be devoted to accurately capturing human facial features. Such feature recognition is the underlying technology for Eyematic, a Los Angeles-based start-up for which Lanier is also chief scientist.
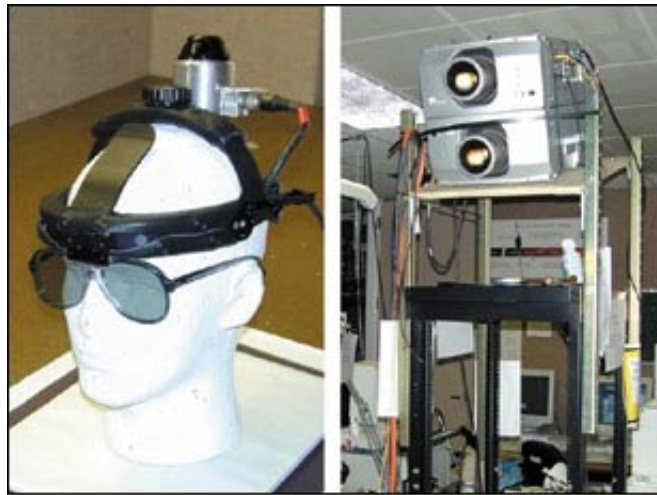
The tele-immersion display employs a headtracker and polarizing glasses to create the illusion of depth from the projected re-rendered left- and right-eye views.

In keeping with the strategy of using commodity resources, the computers processing the acquired visual information for the project are also off-the-shelf. The systems are "from the usual suspects-Dell and IBM," says Lanier. "We fill up racks with them and have them crunch on the problem. A good rule of thumb is one fast processor cluster per camera." Some of the most time-consuming work on this project has been to speed up processor input and output: "It can be incredibly agonizing figuring out why this particular driver is slower than we think it should be or why memory is not freeing up as quickly as it should."

Other considerations for tele-immersion's vision processing include networks' particular limitations such as bandwidth, latency, and protocols. To be determined in different situations is what type of information is best to send over the network. This could be raw feeds from the individual digital cameras compressed for faster transmission or combined signals from a cluster of cameras that are processed for baseline calculation of a scene's parameters and sent out as three-dimensional data. The decision will depend on configuration variables, including the number of cameras. "We want to give people a formula for how to configure a system ideally for a given network," Lanier explains. "That's part of our cookbook project as well."

Once the visual information is received at a remote location, it is re-rendered as computer-generated people and sets by a computer specialized for this task. In the case of the current demo, an SGI Onyx 2 Reality Monster system (employed for high-end virtual reality simulations) was used, though Lanier expects the rendering task soon will be done on less-expensive systems, too. "We've tried some Pentium solutions with gaming cards, so cheap rendering is not that far off," he adds. At this year's Siggraph conference, he was impressed by a cluster of customized Sony Play Stations with extra memory that could handle this task. Still, some problems remain in finding the best techniques for depicting different kinds of objects; for instance, hair renders better as a point cloud, whereas skin renders better as polygons with textures.

Tele-Reality

According to Lanier, the sensation and usefulness of tele-immersion are quite different from videoconferencing. "When you render people properly, they feel real. It's a 'computery' version of them-they're sort of glassy, not quite as filled in as in reality-but your sense of their presence, your ability to make eye contact, your ability to convey your mood and respond to theirs is quite solid because they're life-size, three-dimensional stereoscopic graphics, not small, flat video images."

To enhance the illusion of reality, stereoscopy is essential for depth perception. The display currently used to achieve this in tele-immersion consists of a pair of front projectors for each transmitted scene, with polarized-lens glasses worn by participants to separate reconstituted right- and left-eye views. Lanier is aware that this is hardly the most naturalistic technology for users, as it alters the typical appearance of many participants. "We talked about filtering out the glasses in processing but we never pursued it," he recalls, "because the ultimate solution will be autostereoscopic screens that don't require glasses." One such system, prototyped at NYU Media Research Lab by Ken Perlin, will be incorporated into NTII's next-generation demo.

Accurately establishing a teleconference participant's point of view for proper re-rendering of remote scenes now requires another awkward apparatus: a UNC-developed headtracker based on 3rdTech's HiBall tracker, which sits on

the user's head like a silver salt shaker. It establishes the viewer's physical orientation by sensing position relative to infrared light-emitting diodes embedded in the ceiling. Lanier expects the cost reductions implicit in Moore's Law to apply to video chips as well, ultimately resulting in position tracking by visual sensors imprinted like wallpaper in a room, thus eliminating cumbersome headgear.

In the latest tele-immersion demonstration, researchers manipulate a virtual object, paving the way for real-time rendering of CAD/CAM prototyping and other collaborative work applications.

Currently being implemented in the demo is Lanier's vision of shared workspaces in tele-immersion. In its first configuration, teleconferencers occupy separate display areas. In the next iteration, local participants and remote scenes will over lap for naturalistic communication of formulas on the same virtual whiteboard or manipulation of virtual objects over the network. On an archaeological dig, for example, on-site participants could hand off a virtual representation of a find to a remote expert for closer examination. Also being discussed are virtual objects with special properties-like Columbia researcher Steven Feiner's contribution, the "privacy lamp," a virtual illumination source that could delineate an area where the content would be obscured from one or more participants without the proper levels of clearance.

Though all the systems being developed by NTII are meant to be scalable, for the moment the amounts of visual space that are transmitted in the demo-the volumes surrounding three seated participants-are close to bandwidth limitations, even with Internet 2. This is due to the bottleneck known as the last-mile problem: Connections on university campuses and in research institutions other than supercomputer centers (as well as businesses and homes) are still wired for considerably less bandwidth than the full potential of Internet 2 and next-generation Internet initiatives.

Further limiting the real-time capabilities of tele-immersion are the physics of light itself. Photons passing through fiber optics travel far more slowly than the universal speed of light, causing perceptible transmission delays over long distances. For a wide variety of applications, a lag of 30 to 50 milliseconds is what is acceptable to viewers. In some cases, critical elements like head and hand movements can be tweaked with predictive algorithms that in effect accelerate scene capture beyond the moment in time, inferring the next position of a hand gesture or head movement, but these can only do so much. Perhaps by harnessing super computers, any lag could be eliminated. For now, the maximum distance for tele-immersion to be effective is roughly the width of the United States.

Lanier sees the tele-immersion techniques he's demonstrating as the building blocks of the office of tomorrow, where several users from across the country will be able to collaborate as if they're all in the same room. Scaling up, transmissions could incorporate larger scenes, like news conferences, ballet performances, or sports events. With mobile rather than stationary camera arrays, viewers could establish telepresence in remote or hazardous situations. Far from just a validating application for the next-generation Internet, Jaron Lanier expects tele-immersion to fundamentally change how we view real and virtual worlds.

Steve Ditlea, a New York-based technology journalist, has been covering virtual reality since 1988. He can be reached at ditlea@rcn.com.

Computer Graphics World January, 2001