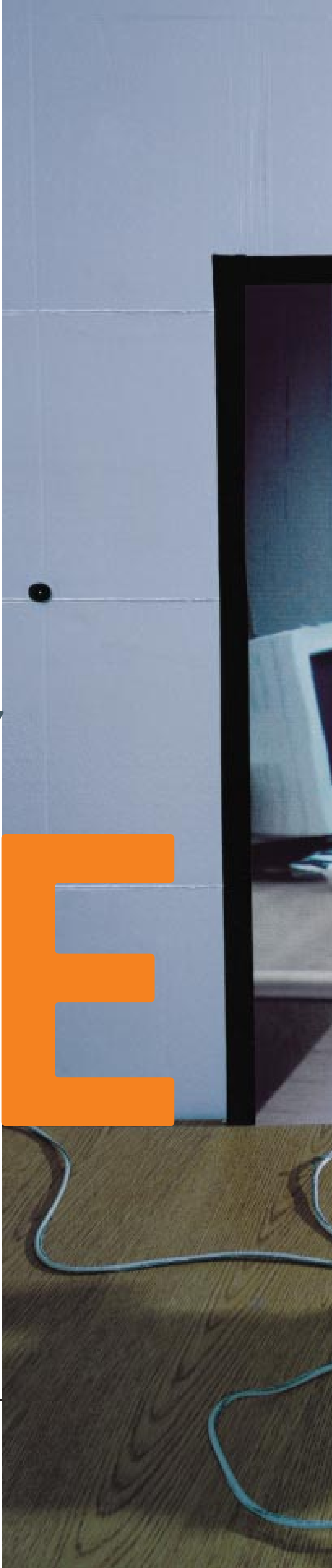
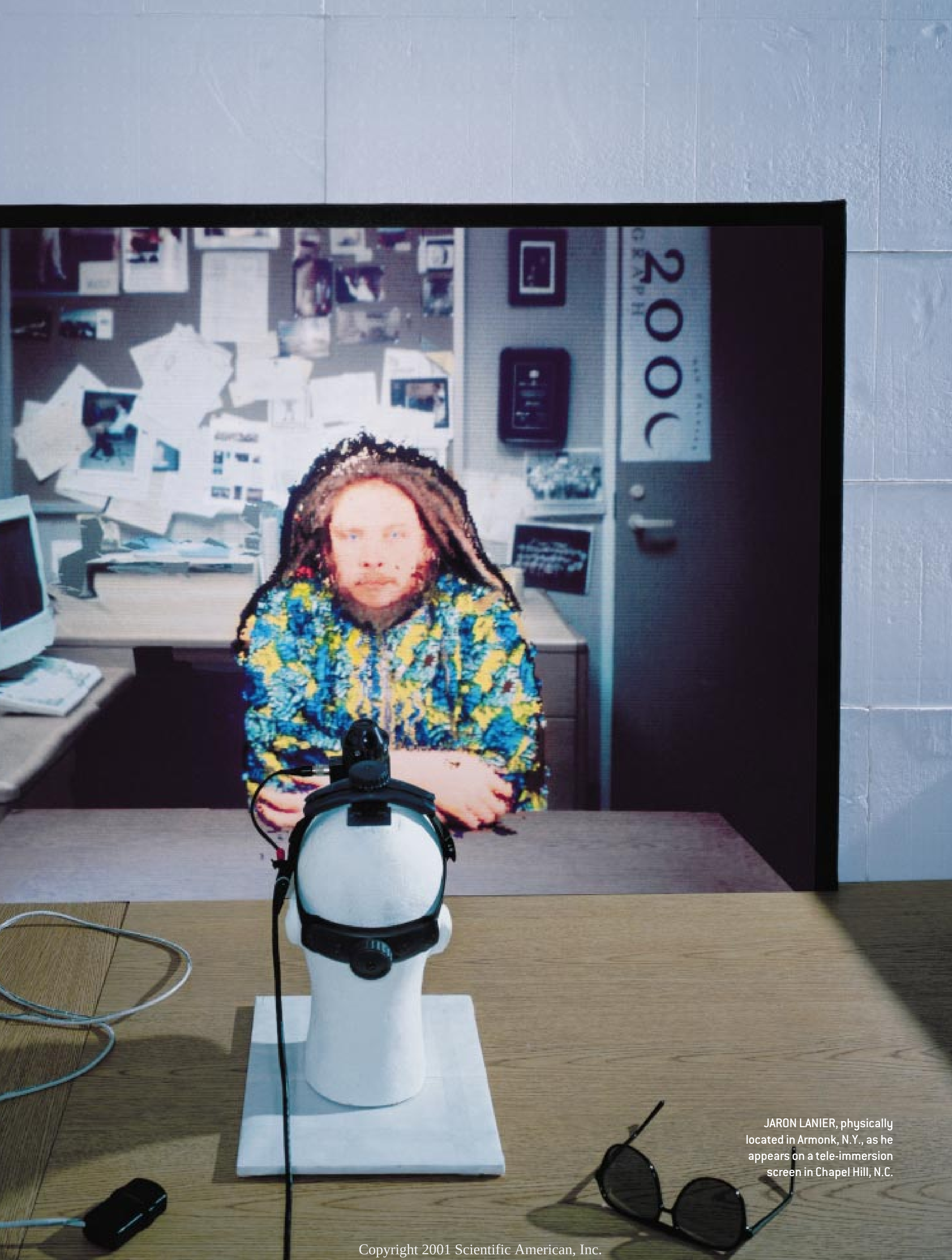
VIRTUALLY THERE

Three-dimensional tele-immersion may eventually bring the world to your desk

BY JARON LANIER

PHOTOGRAPH BY DAN WINTERS





JARON LANIER, physically located in Armonk, N.Y., as he appears on a tele-immersion screen in Chapel Hill, N.C.

Like many researchers, I am a frequent but reluctant user of videoconferencing. **Human interaction** *has both verbal and nonverbal elements, and videoconferencing seems precisely configured to confound the nonverbal ones. It is impossible to make eye contact* properly, for instance, in today's videoconferencing systems, because the camera and the display screen cannot be in the same spot. This usually leads to a deadened and formal affect in interactions, eye contact being a nearly ubiquitous subconscious method of affirming trust. Furthermore, participants aren't able to establish a sense of position relative to one another and therefore have no clear way to direct attention, approval or disapproval.

Tele-immersion, a new medium for human interaction enabled by digital technologies, approximates the illusion that a user is in the same physical space as other people, even though the other participants might in fact be hundreds or thousands of miles away. It combines the display and interaction techniques of virtual reality with new vision technologies that transcend the traditional limitations of a camera. Rather than merely observing people and their immediate environment from one vantage point, tele-immersion stations convey them as "moving sculptures," without favoring a single point of view. The result is that all the participants, however distant, can share and explore a life-size space.

Beyond improving on videoconferencing, tele-immersion was conceived as an ideal application for driving network-engineering research, specifically for Internet2, the primary research consortium for advanced network studies in the U.S. If a computer network can support tele-immersion, it can probably support any other application. This is because tele-immersion demands as little delay as possible from flows of information (and as little inconsistency in delay), in addition to the more common demands for very large and reliable flows.

Virtual Reality and Networks

BECAUSE TELE-IMMERSION sits at the crossroads of research in virtual reality and networking, as well as computer vision and user-interface research, a little background in these various fields of research is in order.

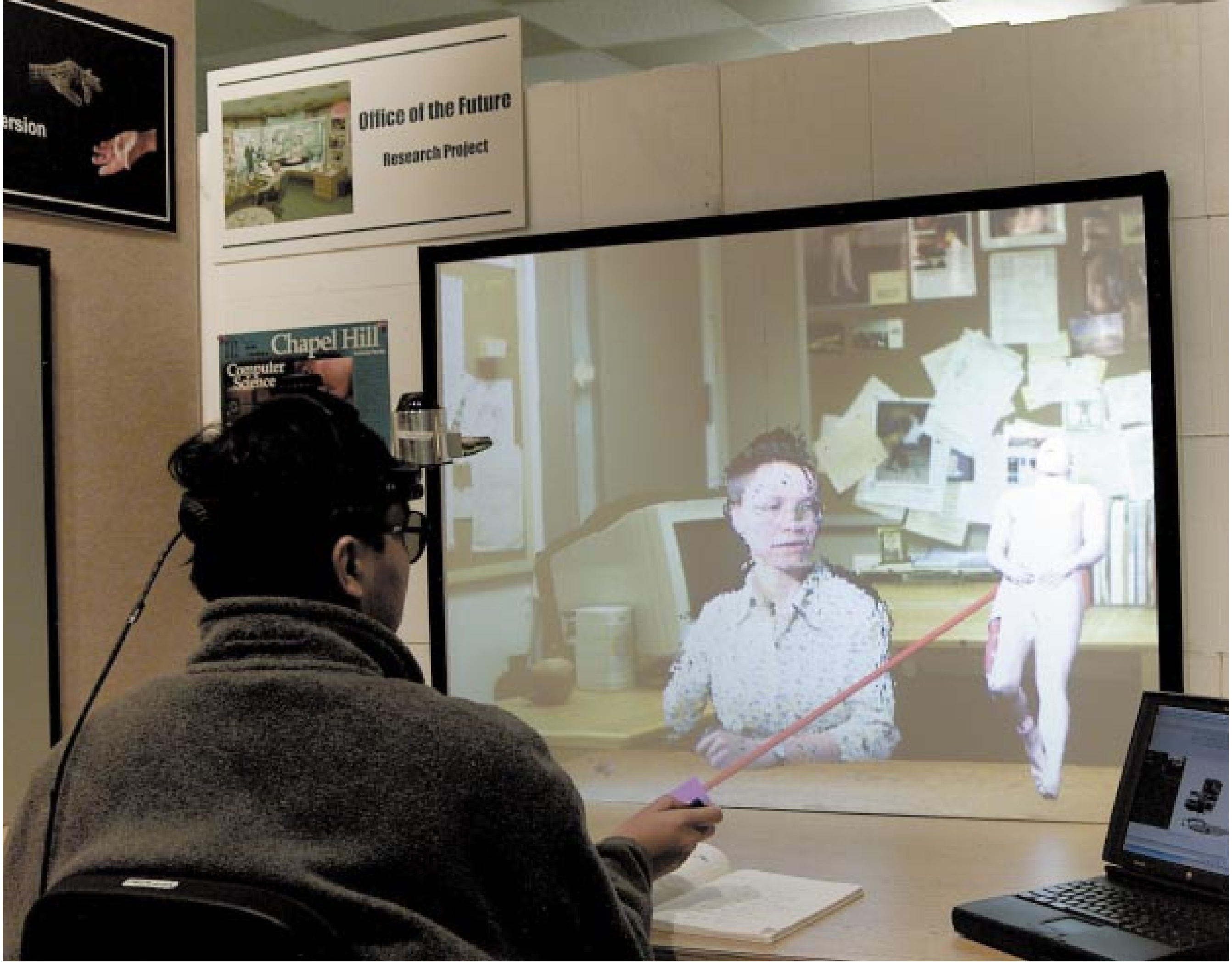
In 1965 Ivan Sutherland, who is widely regarded as the father of computer graphics, proposed what he called the "Ultimate Display." This display would allow the user to experience an entirely computer-rendered space as if it were real. Sutherland termed such a space a "Virtual World," invoking a term from the philosophy of aesthetics, particularly the writings of Suzanne K. Langer. In 1968 Sutherland realized a virtual world for the first time by means of a device called a head-mounted display. This was a helmet with a pair of display screens positioned in front of the eyes to give the wearer a sense of immersion in a stereoscopic, three-dimensional space. When the user moved his or her head, a computer would quickly recompute the images in front of each eye to maintain the illusion that the computer-rendered world remained stationary as the user explored it.

In the course of the 1980s I unintentionally ended up at the helm of the first company to sell general-purpose tools for making and experiencing virtual worlds—in large part because of this magazine. *Scientific American* devoted its September 1984 issue to emerging digital technologies and chose to use one of my visual-programming experiments as an illustration for the cover.

At one point I received a somewhat panicked phone call from an editor who noticed that there was no affiliation listed for me. I explained that at the time I had no affiliation and neither did the work being described. "Sir," he informed me, "at *Scientific American* we have a strict rule that states that an affiliation must be indicated after a contributor's name." I blurted out "VPL Research" (for Visual Programming Language, or

Overview / *Tele-immersion*

- This new telecommunications medium, which combines aspects of virtual reality with videoconferencing, aims to allow people separated by great distances to interact naturally, as though they were in the same room.
- Tele-immersion is being developed as a prototype application for the new Internet2 research consortium. It involves monumental improvements in a host of computing and communications technologies, developments that could eventually lead to a variety of spin-off inventions.
- The author suggests that within 10 years, tele-immersion could substitute for many types of business travel.



TELE-COLLABORATORS hundreds of miles apart consider a computer-generated medical model, which both of them can manipulate as though it were a real object. The headpiece helps the computers locate the position and orientation of the user's head; such positioning is essential for presenting the right view of a scene. In the future, the headpiece should be unnecessary.

PHOTOGRAPH COURTESY OF THE UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL

Virtual Programming Language), and thus was born VPL. After the issue's publication, investors came calling, and a company came to exist in reality. In the mid-1980s VPL began selling virtual-world tools and was well known for its introduction of glove devices, which were featured on another *Scientific American* cover, in October 1987.

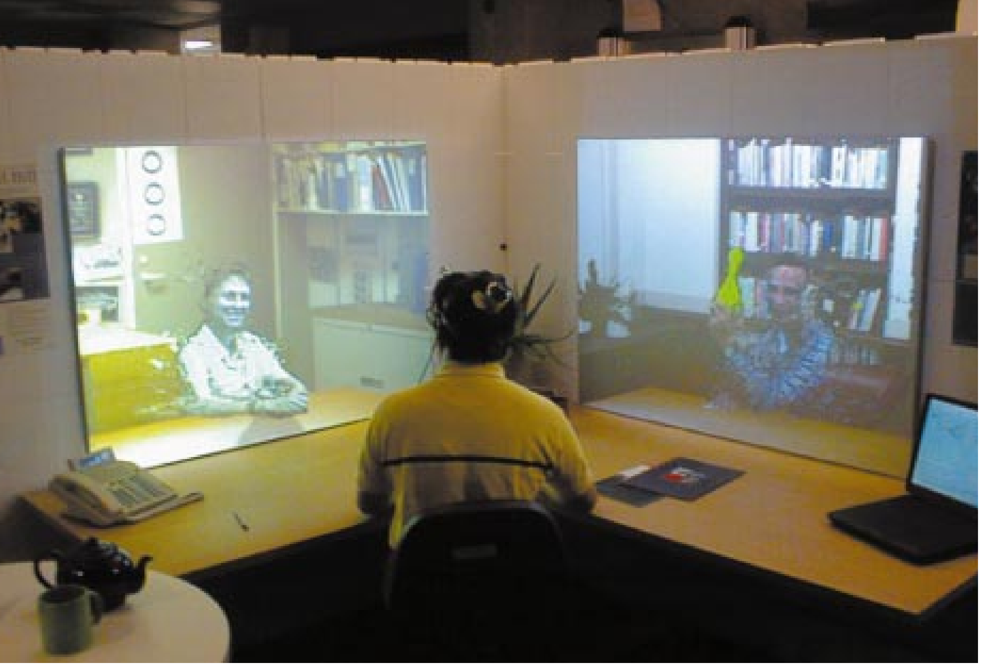
VPL performed the first experiments in what I decided to call "virtual reality" in the mid- to late 1980s. Virtual reality combines the idea of virtual worlds with networking, placing multiple participants in a virtual space using head-mounted displays. In 1989 VPL introduced a product called RB2, for "Reality Built for Two," that allowed two participants to share a virtual world. One intriguing implication of virtual reality is that participants must be able to see representations of one another, often known as avatars. Although the computer power of the day limited our early avatars to extremely simple, cartoonish computer graphics that only roughly approximated the faces of users, they nonetheless transmitted the motions of their hosts faithfully and thereby conveyed a sense of presence, emotion and locus of interest.

At first our virtual worlds were shared across only short physical distances, but we also performed some experiments with long-distance applications. We were able to set up virtual-reality sessions with participants in Japan and California and in Germany and California. These demonstrations did not strain the network, because only the participants' motions needed to be sent, not the entire surface of each person, as is the case with tele-immersion.

Computer-networking research started in the same era as research into virtual worlds. The original network, the Arpanet, was conceived in the late 1960s. Other networks were inspired by it, and in the 1980s all of them merged into the Internet. As the Internet grew, various "backbones" were built. A backbone is a network within a network that lets information travel over exceptionally powerful, widely shared connections to go long distances more quickly. Some notable backbones designed to support research were the NSFnet in the late 1980s and the vBNS in the mid-1990s. Each of these played a part in inspiring new applications for the Internet, such as the

THE AUTHOR

JARON LANIER is a computer scientist often described as "the father of virtual reality." In addition to that field, his primary areas of study have been visual programming, simulation, and high-performance networking applications. He is chief scientist of Advanced Network and Services, a nonprofit concern in Armonk, N.Y., that funds and houses the engineering office of Internet2. Music is another of Lanier's great interests: he writes for orchestra and other ensembles and plays an extensive, exotic assortment of musical instruments—most notably, wind and string instruments of Asia. He is also well known as an essayist on public affairs.



THREE USERS in different cities can share a virtual space thanks to this telecubicle.

PHOTOGRAPH COURTESY OF THE UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL

World Wide Web. Another backbone-research project, called Abilene, began in 1998, and it was to serve a university consortium called Internet2.

Abilene now reaches more than 170 American research universities. If the only goal of Internet2 were to offer a high level of bandwidth (that is, a large number of bits per second), then the mere existence of Abilene and related resources would be sufficient. But Internet2 research targeted additional goals, among them the development of new protocols for handling applications that demand very high bandwidth and very low, controlled latencies (delays imposed by processing signals en route).

Internet2 had a peculiar problem: no existing applications required the anticipated level of performance. Computer science has traditionally been driven by an educated guess that there will always be good uses for faster and more capacious digital tools, even if we don't always know in advance what those uses will be. In the case of advanced networking research, however, this faith wasn't enough. The new ideas would have to be tested on something.

Allan H. Weis, who had played a central role in building the NSFnet, was in charge of a nonprofit research organization called Advanced Network and Services, which housed and administered the engineering office for Internet2. He used the term "tele-immersion" to conjure an ideal "driver" application and asked me to take the assignment as lead scientist for a National Tele-Immersion Initiative to create it. I was delighted, as this was the logical extension of my previous work in shared virtual worlds.

Although many components, such as the display system, awaited invention or refinement before we could enjoy a working tele-immersion system, the biggest challenge was creating an appropriate way of visually sensing people and places. It might not be immediately apparent why this problem is different from videoconferencing.

Seen through polarizing glasses, two walls of the cubicle dissolved into windows, revealing offices with people who ***WERE LOOKING BACK AT ME.***

Beyond the Camera as We Know It

THE KEY IS THAT in tele-immersion, each participant must have a personal viewpoint of remote scenes—in fact, two of them, because each eye must see from its own perspective to preserve a sense of depth. Furthermore, participants should be free to move about, so each person's perspective will be in constant motion.

Tele-immersion demands that each scene be sensed in a manner that is not biased toward any particular viewpoint (a camera, in contrast, is locked into portraying a scene from its own position). Each place, and the people and things in it, has to be sensed from all directions at once and conveyed as if it were an animated three-dimensional sculpture. Each remote site receives information describing the whole moving sculpture and renders viewpoints as needed locally. The scanning process has to be accomplished fast enough to take place in real time—at most within a small fraction of a second. The sculpture representing a person can then be updated quickly enough to achieve the illusion of continuous motion. This illusion starts to appear at about 12.5 frames per second (fps) but becomes robust at about 25 fps and better still at faster rates.

Measuring the moving three-dimensional contours of the inhabitants of a room and its other contents can be accomplished in a variety of ways. As early as 1993, Henry Fuchs of the University of North Carolina at Chapel Hill had proposed one method, known as the "sea of cameras" approach, in which the viewpoints of many cameras are compared. In typical scenes in a human environment, there will tend to be visual features, such as a fold in a sweater, that are visible to more than one camera. By comparing the angle at which these features are seen by different cameras, algorithms can piece together a three-dimensional model of the scene.

This technique had been explored in non-real-time configurations, notably in Takeo Kanade's work, which later culminated in the "Virtualized Reality" demonstration at Carnegie Mellon University, reported in 1995. That setup consisted of 51 inward-looking cameras mounted on a geodesic dome. Because it was not a real-time device, it could not be used for tele-immersion. Instead videotape recorders captured events in the dome for later processing.

Ruzena Bajcsy, head of the GRASP (General Robotics, Automation, Sensing and Perception) Laboratory at the University of Pennsylvania, was intrigued by the idea of real-time seas of cameras. Starting in 1994, she worked with colleagues at Chapel Hill and Carnegie Mellon on small-scale "puddles" of two or three cameras to gather real-world data for virtual-reality applications.

Bajcsy and her colleague Kostas Daniilidis took on the assignment of creating the first real-time sea of cameras—one that was, moreover, scalable and modular so that it could be adapted to a variety of rooms and uses. They worked closely with the Chapel Hill team, which was responsible for taking the "animated sculpture" data and using computer graphics techniques to turn it into a realistic scene for each user.

But a sea of cameras in itself isn't a complete solution. Suppose a sea of cameras is looking at a clean white wall. Because there are no surface features, the cameras have no information with which to build a sculptural model. A person can look at a white wall without being confused. Humans don't worry that a wall might actually be a passage to an infinitely deep white chasm, because we don't rely on geometric cues alone—we also have a model of a room in our minds that can rein in errant mental interpretations. Unfortunately, to today's digital cameras, a person's forehead or T-shirt can present the same challenge as a white wall, and today's software isn't smart enough to undo the confusion that results.

Researchers at Chapel Hill came up with a novel method that has shown promise for overcoming this obstacle, called "imperceptible structured light," or ISL. Conventional lightbulbs flicker 50 or 60 times a second, fast enough for the flickering to be generally invisible to the human eye. Similarly, ISL appears to the human eye as a continuous source of white light, like an ordinary lightbulb, but in fact it is filled with quickly changing patterns visible only to specialized, carefully synchronized cameras. These patterns fill in voids such as white walls with imposed features that allow a sea of cameras to complete the measurements.

The Eureka Moment

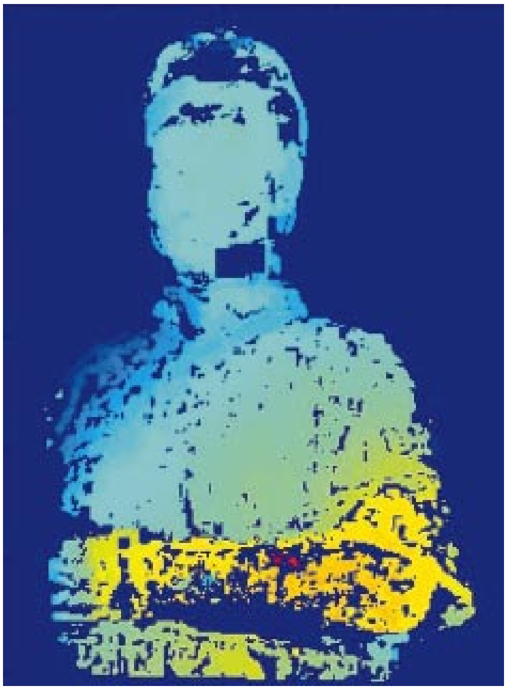
WE WERE ABLE TO demonstrate tele-immersion for the first time on May 9, 2000, virtually bringing together three locations. About a dozen dignitaries were physically at the telecubicle in Chapel Hill. There we and they took turns sitting down in the simulated office of tomorrow. As fascinating as the three years of research leading up to this demonstration had been for me, the delight of experiencing tele-immersion was unanticipated and incomparable. Seen through a pair of polarizing glasses, two walls of the cubicle dissolved into windows, revealing other offices with other people who were looking back at me. (The glasses helped to direct a slightly different view of the scenes to each eye, creating the stereo vision effect.) Through one wall I greeted Amela Sadagic, a researcher at my lab in Armonk, N.Y. Through the other wall was Jane Mulligan, a postdoctoral fellow at the University of Pennsylvania.

Unlike the cartoonish virtual worlds I had worked with for many years, the remote people and places I was seeing were clearly derived from reality. They were not perfect by any means. There was "noise" in the system that looked something like confetti being dropped in the other people's cubicles. The frame rate was low (2 to 3 fps), there was as much as one second of delay, and only one side of the conversation had access to a tele-immersive display. Nevertheless, here was a virtual world that was not a simplistic artistic representation of the real world but rather an authentic measurement-based rendition of it.

In a later demo (in October 2000) most of the confetti was gone and the overall quality and speed of the system had increased, but the most important improvement came from researchers at Brown University led by Andries van Dam. They arrived in a tele-immersive session bearing virtual objects not derived from the physical scene. I sat across the table from Robert C. Zeleznik of Brown, who was physically at my lab in Armonk. He presented a simulated miniature office interior (about two feet wide) resting on the desk between us, and we used simulated laser pointers and other devices to modify walls and furniture in it collaboratively while we talked. This was a remarkable blending of the experience of using simulations associated with virtual reality and simply being with another person.

When Can I Use It?

BEYOND THE SCENE-CAPTURE system, the principal components of a tele-immersion setup are the computers, the network services, and the display and interaction devices. Each of these components has been advanced in the cause of tele-immersion and must advance further. Tele-immersion is a voracious consumer of computer resources. We've chosen to work with "commodity" computer components (those that are also used in common home and office products) wherever

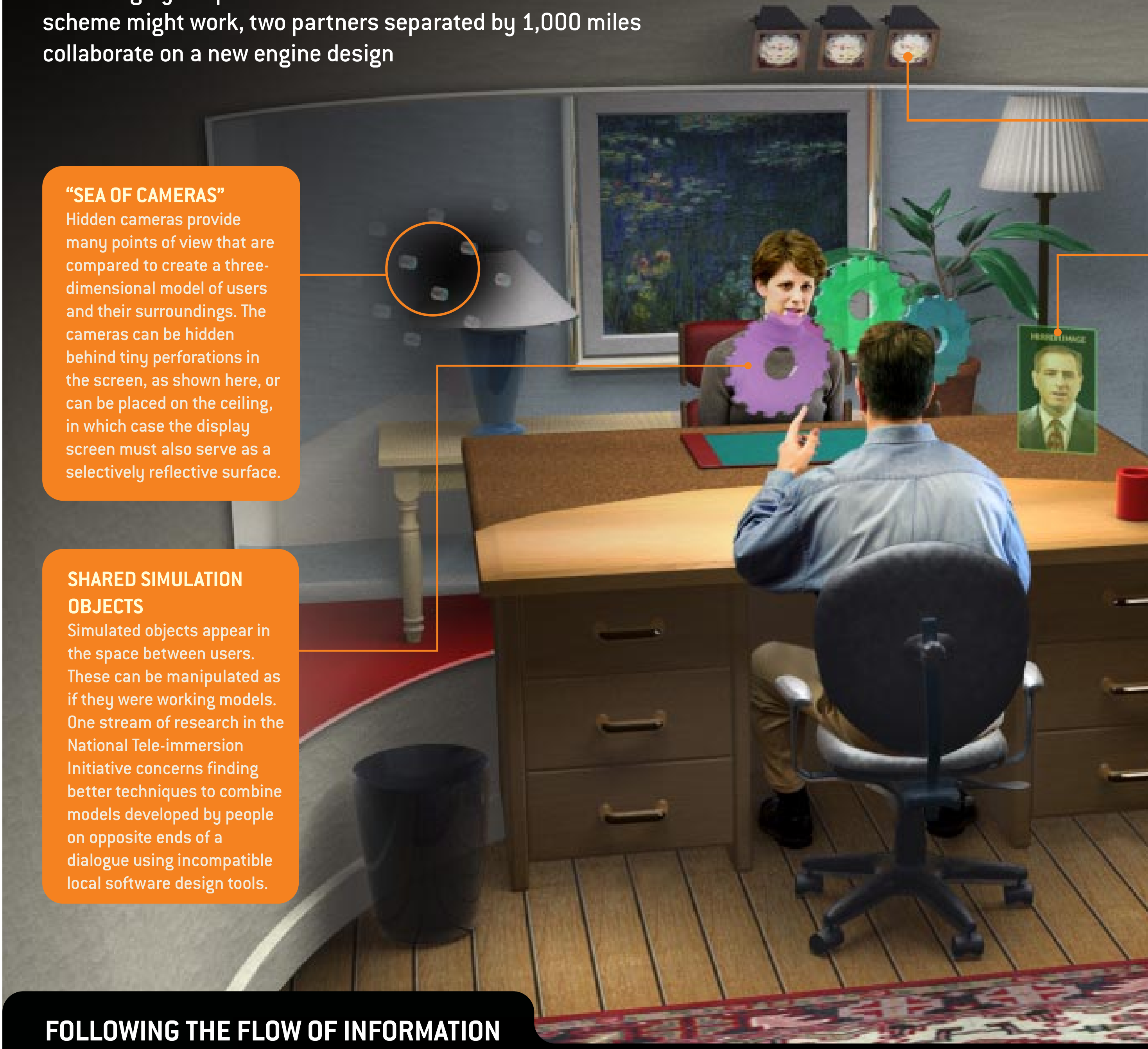


COMPARISON OF TWO VIEWS of a person taken by the tele-immersion cameras yields this image. The colors represent the first rough calculation of the depth of the person's features.

IMAGE COURTESY OF THE UNIVERSITY OF PENNSYLVANIA

HOW TELE-IMMERSION WORKS

In this highly simplified scheme for how a future tele-immersion scheme might work, two partners separated by 1,000 miles collaborate on a new engine design



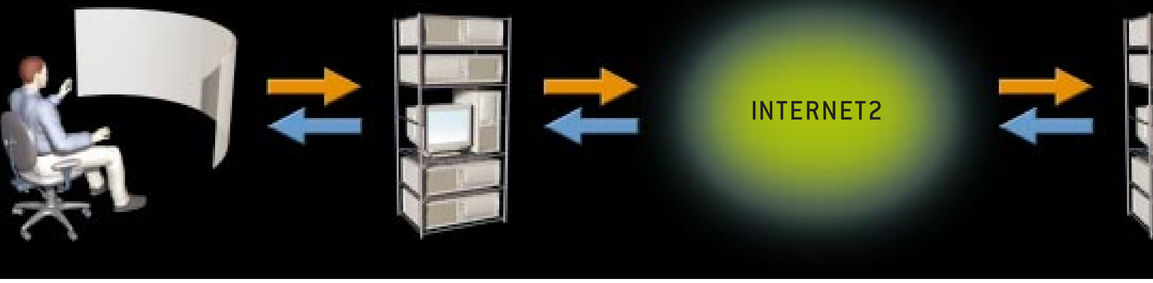
“SEA OF CAMERAS”
Hidden cameras provide many points of view that are compared to create a three-dimensional model of users and their surroundings. The cameras can be hidden behind tiny perforations in the screen, as shown here, or can be placed on the ceiling, in which case the display screen must also serve as a selectively reflective surface.

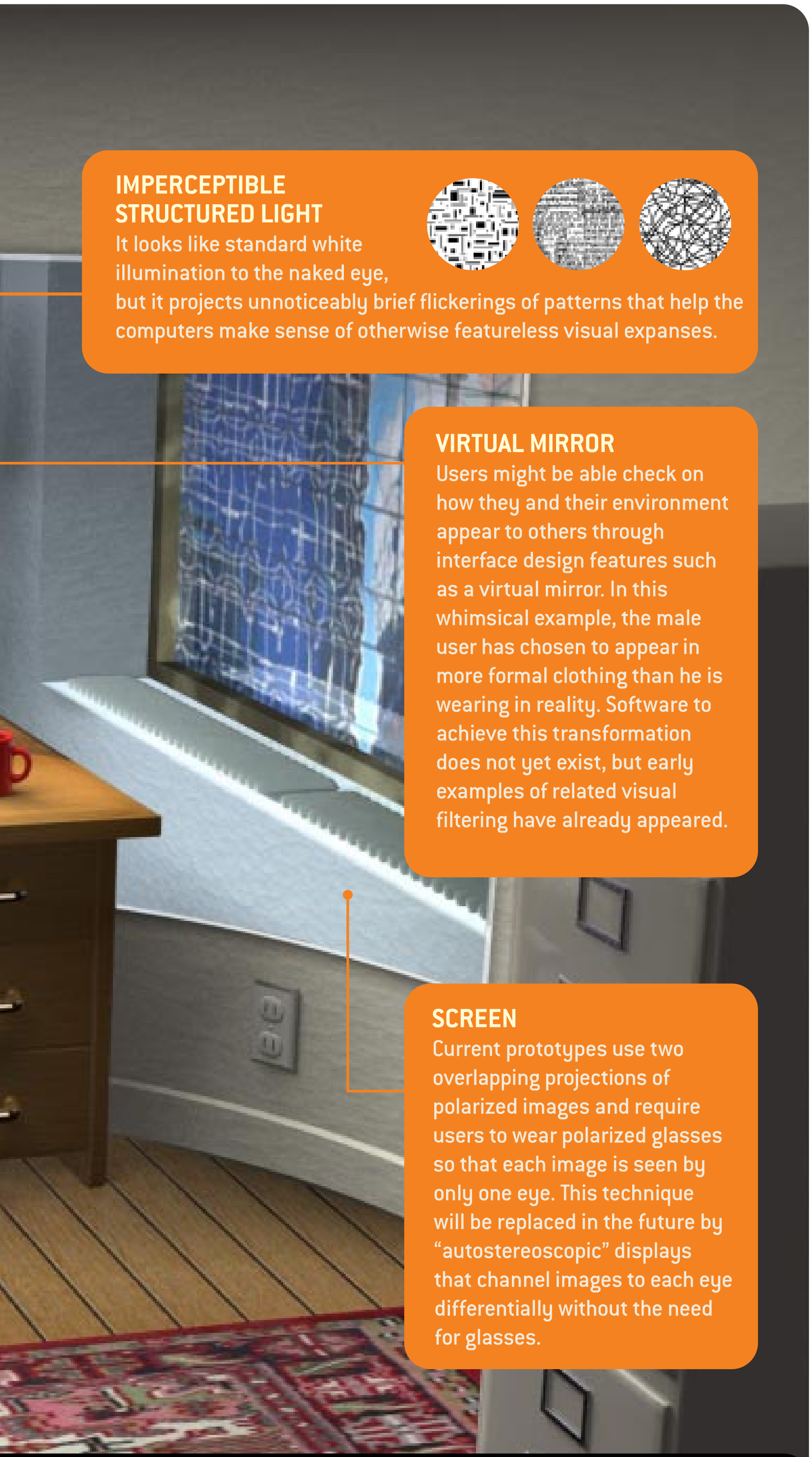
SHARED SIMULATION OBJECTS
Simulated objects appear in the space between users. These can be manipulated as if they were working models. One stream of research in the National Tele-immersion Initiative concerns finding better techniques to combine models developed by people on opposite ends of a dialogue using incompatible local software design tools.

FOLLOWING THE FLOW OF INFORMATION

Tele-immersion depends on intense data processing at each end of a connection, mediated by a high-performance network.

FROM THE SENDER . . .
Parallel processors accept visual input from the cameras and reinterpret the scene as a three-dimensional computer model.





IMPERCEPTIBLE STRUCTURED LIGHT

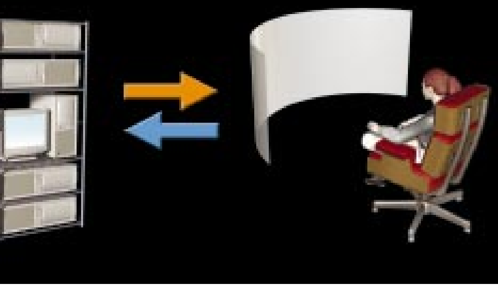
It looks like standard white illumination to the naked eye, but it projects unnoticeably brief flickerings of patterns that help the computers make sense of otherwise featureless visual expanses.

VIRTUAL MIRROR

Users might be able check on how they and their environment appear to others through interface design features such as a virtual mirror. In this whimsical example, the male user has chosen to appear in more formal clothing than he is wearing in reality. Software to achieve this transformation does not yet exist, but early examples of related visual filtering have already appeared.

SCREEN

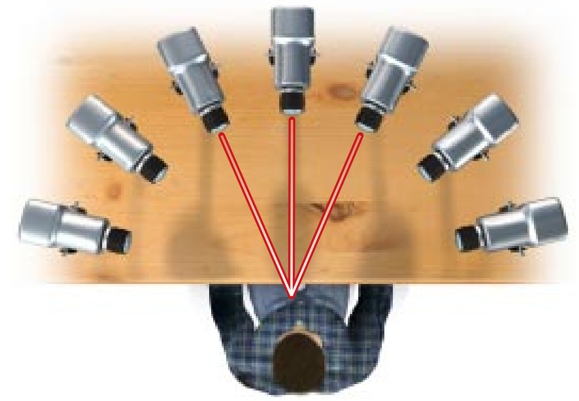
Current prototypes use two overlapping projections of polarized images and require users to wear polarized glasses so that each image is seen by only one eye. This technique will be replaced in the future by "autostereoscopic" displays that channel images to each eye differentially without the need for glasses.



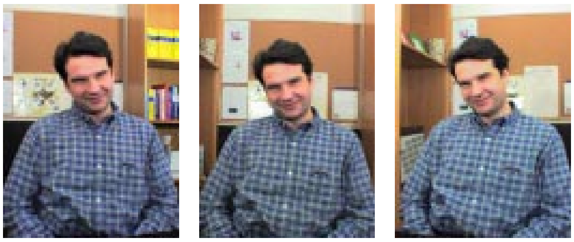
. . . TO THE RECEIVER

Specific renderings of remote people and places are synthesized from the model as it is received to match the points of view of each eye of a user. The whole process repeats many times a second to keep up with the user's head motion.

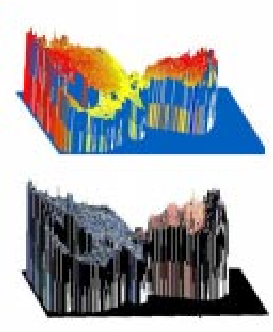
GENERATING THE 3-D IMAGE



1 An array of cameras views people and their surroundings from different angles. Each camera generates an image from its point of view many times in a second.



2 Each set of the images taken at a given instant is sorted into subsets of overlapping trios of images.



3 From each trio of images, a "disparity map" is calculated, reflecting the degree of variation among the images at all points in the visual field. The disparities are then analyzed to yield depths that would account for the differences between what each camera sees. These depth values are combined into a "bas relief" depth map of the scene.

4 All the depth maps are combined into a single viewpoint-independent sculptural model of the scene at a given moment. The process of combining the depth maps provides opportunities for removing spurious points and noise.

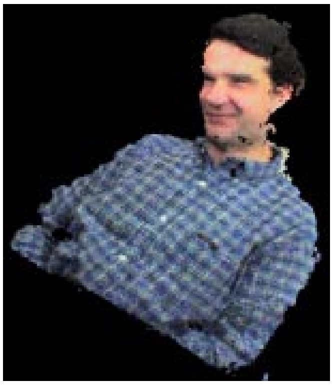
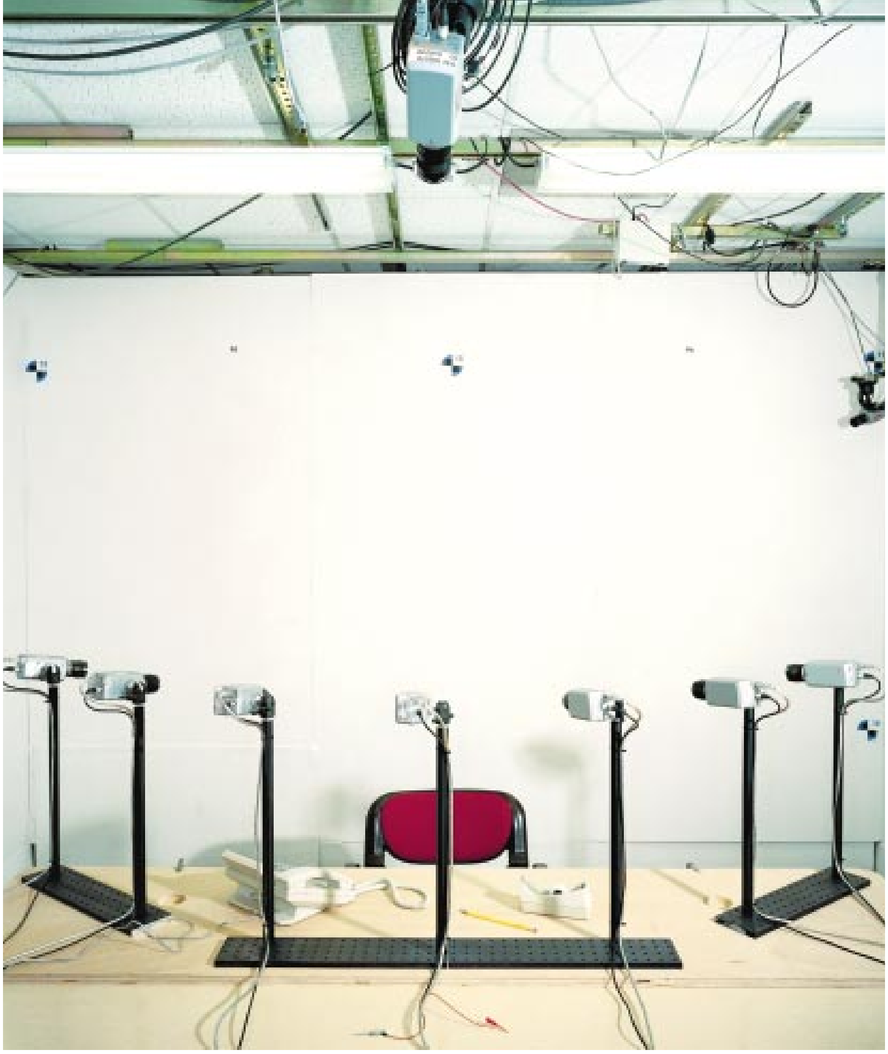


ILLUSTRATION BY JOE ZEFF



SEVEN CAMERAS scrutinize the user in the tele-immersion setup in Chapel Hill.

possible to hasten the day when tele-immersion will be reproducible outside the lab. Literally dozens of such processors are currently needed at each site to keep up with the demands of tele-immersion. These accumulate either as personal computers in plastic cases lined up on shelves or as circuit boards in refrigerator-size racks. I sometimes joke about the number of “refrigerators” required to achieve a given level of quality in tele-immersion.

Most of the processors are assigned to scene acquisition. A sea of cameras consists of overlapping trios of cameras. At the moment we typically use an array of seven cameras for one person seated at a desk, which in practice act as five trios. Roughly speaking, a cluster of eight two-gigahertz Pentium processors with shared memory should be able to process a trio within a sea of cameras in approximately real time. Such processor clusters should be available later this year. Although we expect computer prices to continue to fall as they have for the past few decades, it will still be a bit of a wait before tele-immersion becomes inexpensive enough for widespread use. The cost of an eight-processor cluster is anticipated to be in the $30,000 to $50,000 range at introduction, and a number of those would be required for each site (one for each trio of cameras)—and this does not even account for the processing needed for other tasks. We don’t yet know how many cameras will be required for a given use of tele-immersion, but currently a good guess is that seven is the minimum adequate for casual conversation, whereas 60 cameras might be needed for the most demanding applications, such as long-distance surgical demonstration, consultation and training.

Our computational needs go beyond processing the image streams from the sea of cameras. Still more processors are required to resynthesize and render the scene from shifting perspectives as a participant’s head moves during a session. Initially we used a large custom graphics computer, but more recently we have been able instead to draft commodity processors with low-cost graphics cards, using one processor per eye. Additional processors are required for other tasks, such as combining the results from each of the camera trios, running the imperceptible structured light, measuring the head motion of the user, maintaining the user interface, and running virtual-object simulations.

Furthermore, because minimizing apparent latency is at the heart of tele-immersion engineering, significant processing resources will eventually need to be applied to predictive algorithms. Information traveling through an optical fiber reaches a destination at about two thirds the speed of light in free space because it is traveling through the fiber medium instead of a vacuum and because it does not travel a straight path but rather bounces around in the fiber channel. It therefore takes anywhere from 25 to 50 milliseconds for fiber-bound bits of information to cross the continental U.S., without any allowances for other inescapable delays, such as the activities of various network signal routers.

By cruel coincidence, some critical aspects of a virtual world’s responsiveness should not be subject to more than 30 to 50 milliseconds of delay. Longer delays result in user fatigue and disorientation, a degradation of the illusion and, in the worst case, nausea. Even if we had infinitely fast computers at each end, we’d still need to use prediction to compensate for lag when conducting conversations

MORE TO EXPLORE

National Tele-immersion Initiative Web site: **www.advanced.org/teleimmersion.html**

Tele-immersion at Brown University: **www.cs.brown.edu/~lsh/telei.html**

Tele-immersion at the University of North Carolina at Chapel Hill: **www.cs.unc.edu/Research/stc/teleimmersion/**

Tele-immersion at the University of Pennsylvania: **www.cis.upenn.edu/~sequence/teleim1.html**

Tele-immersion site at Internet2: **www.internet2.edu/html/tele-immersion.html**

Information about an autostereoscopic display: **www.mrl.nyu.edu/projects/autostereo**

PHOTOGRAPH BY DAN WINTERS

across the country. This is one reason the current set of test sites are all located on the East Coast.

One promising avenue of exploration in the next few years will be routing tele-immersion processing through remote supercomputer centers in real time to gain access to superior computing power. In this case, a supercomputer will have to be fast enough to compensate for the extra delay caused by the travel time to and from its location.

Bandwidth is a crucial concern. Our demand for bandwidth varies with the scene and application; a more complex scene requires more bandwidth. We can assume that much of the scene, particularly the background walls and such, is unchanging and does not need to be resent with each frame. Conveying a single person at a desk, without the surrounding room, at a slow frame rate of about two frames per second has proved to require around 20 megabits per second but with up to 80-megabit-per-second peaks. With time, however, that number will fall as better compression techniques become established. Each site must receive the streams from all the others, so in a three-way conversation the bandwidth requirement must be multiplied accordingly. The "last mile" of network connection that runs into computer science departments currently tends to be an OC3 line, which can carry 155 megabits per second—just about right for sustaining a three-way conversation at a slow frame rate. But an OC3 line is approximately 100 times more capacious than what is usually considered a broadband connection now, and it is correspondingly more expensive.

I am hopeful that in the coming years we will see a version of tele-immersion that does not require users to wear special glasses or any other devices. Ken Perlin of New York University has developed a prototype of an autostereoscopic display that might make this possible.

Roughly speaking, tele-immersion is about 100 times too expensive to compete with other communications technologies right now and needs more polishing besides. My best guess is that it will be good enough and cheap enough for limited introduction in approximately five years and for widespread use in around 10 years.

Prospects

WHEN TELE-IMMERSION becomes commonplace, it will probably enable a wide variety of important applications. Teams of engineers might collaborate at great distances on computerized designs for new machines that can be tinkered with as though they were real models on a shared workbench. Archaeologists from around the world might experience being present during a crucial dig. Rarefied experts in building inspection or engine repair might be able to visit locations without losing time to air travel.

In fact, tele-immersion might come to be seen as real competition for air travel—unlike videoconferencing. Although few would claim that tele-immersion will be absolutely as good as "being there" in the near term, it might be good enough for business meetings, professional consultations, training sessions, trade show exhibits and the like. Business travel might be replaced to a significant degree by tele-immersion in 10 years. This is not only because tele-immersion will become better and cheaper but because air travel will face limits to growth because of safety, land use and environmental concerns.

Tele-immersion might have surprising effects on human relationships and roles. For instance, those who worry about how artists, musicians and authors will make a living as copyrights become harder and harder to enforce (as a result of widespread file copying on the Internet) have often suggested that paid personal appearances are a solution, because personal interaction has more value in the moment than could be reproduced afterward from a file or recording. Tele-immersion could make aesthetic interactions practical and cheap enough to provide a different basis for commerce in the arts. It is worth remembering that before the 20th century, all the arts were interactive. Musicians interacted directly with audience members, as did actors on a stage and poets in a garden. Tele-immersive forms of all these arts that emphasize immediacy, intimacy and personal responsiveness might appear in answer to the crisis in copyright enforcement.

Undoubtedly tele-immersion will pose new challenges as well. Some early users have expressed a concern that tele-immersion exposes too much, that telephones and videoconferencing tools make it easier for participants to control their exposure—to put the phone down or move offscreen. I am hopeful that with experience we will discover both user-interface designs (such as the virtual mirror depicted in the illustration on pages 72 and 73) and conventions of behavior that address such potential problems.

I am often asked if it is frightening to work on new technologies that are likely to have a profound impact on society without being able to know what that impact will be. My answer is that because tele-immersion is fundamentally a tool to help people connect better, the question is really about how optimistic one should be about human nature. I believe that communications technologies increase the opportunities for empathy and thus for moral behavior. Consequently, I am optimistic that whatever role tele-immersion ultimately takes on, it will mostly be for the good. SA

Tele-immersion *Team Members*

- **UNIVERSITY OF NORTH CAROLINA, CHAPEL HILL:** *Henry Fuchs, Herman Towles, Greg Welch, Wei-Chao Chen, Ruigang Yang, Sang-Uok Kum, Andrew Nashel, Srihari Sukumaran*
  www.cs.unc.edu/Research/stc/teleimmersion/
- **UNIVERSITY OF PENNSYLVANIA**
  Ruzena Bajcsy, Kostas Daniilidis, Jane Mulligan, Ibrahim Volkan Isler
  www.cis.upenn.edu/~sequence/teleim2.html
- **BROWN UNIVERSITY**
  Andries van Dam, Loring Holden, Robert C. Zeleznik
  www.cs.brown.edu/~lsh/telei.html
- **ADVANCED NETWORKS AND SERVICES**
  Jaron Lanier, Amela Sadagic
  www.advanced.org/teleimmersion.html