



# Getting the Picture

**Joseph A. Paradiso**  
*MIT Media Lab*

Processing power for computer graphics has increased enormously in recent years and will continue to do so as we ride the exponentials derived from Moore's Law. The accelerated transfer of workstation capabilities to plug-in cards and software applications running on standard PCs move detailed rendering outside the domain of high-end machines. Virtual worlds are now literate in the basic properties of Newtonian physics, and animated graphical entities are likewise slowly becoming sentient as researchers build them into intelligent agents endowed with layered, autonomous behaviors. Most other areas of computer graphics (such as modeling software) have also improved.

## Distributed multimodal interfaces

Common input devices and interfaces, however, have stayed relatively static. Most graphics designers and users still use variants of the same mouse, keyboard, and stylus that have been with us since the exodus of the key-punch. These narrow channels of expression limit the growing possibilities. Although we now have much more vivid virtual environments (VEs), most of us still interact with and build them at a single point of contact in a passive, 2D plane. Higher end applications have exploited more expensive and sophisticated interface hardware, such as instrumented gloves and magnetic tracking systems used for motion capture. Although these devices are becoming liberated from hardwired tethers to a base station, users remain "straightjacketed" as they wire their multiple pickups to a belt pack.

In the near term, as manufacturers build increasing processing power, graphics capability, and network bandwidth into conventional PCs, we'll need low-cost, noninvasive, multimodal interface technologies to use many of the revolutionary new applications. (For example, we can't expect users to spend much time in their networked distributed VEs when they have to drive their avatars with a keyboard and mouse. Likewise, it's impractical to expect users to wear anything like today's magnetic trackers when they want to shrug at an intelligent agent.) This need will become even more acute

as tomorrow's computers break out of their current desktop form and reach into the environment. In the upcoming world of smart rooms, intelligent objects, ubiquitous displays, and wearable computing, many channels of a user's actions (visual, auditory, tactile, proximity, physiological, and so on) will be monitored by various embedded sensors. Data from these different systems will converge across local networks and fuse to produce a dynamic, multimodal "input device."

## Exploring interface technologies

The five articles in this special issue explore several interface technologies that spotlight a few keystones along these paths. Some target the graphics designer, while others serve users interacting with responsive environments.

One well-established technology in the graphics community is laser scanning for capturing the texture and geometry of 3D surfaces. Several methods exist to shortcut the expense and complication traditionally associated with these systems, opening up other applications. As an example, we've developed a low-cost phase-measuring laser rangefinder here at the MIT Media Lab<sup>1</sup> that we use for real-time tracking of bare hands in a plane above a projected video wall. This device allows direct, light-insensitive interaction with large-screen graphics.

Two articles in this issue take very different tacks on laser scanners for object capture. The first, by Petrov et al., gives a succinct portrait of state-of-the-art triangulation scanners. They describe the principles behind their Galatea scanner—exceptional for its low cost, high speed, and photorealistic response. In contrast, Borghese et al. take the minimalist approach with their Autoscan system. Here the authors discard the entire scanning mechanism in favor of a simple hand-held laser pointer with which the user "paints" the object of interest while a stereo pair of video cameras observe. A commercial real-time image-processing board finds the laser spot in both images and produces the 3D coordinates. Although much slower than an automated scanner, the Autoscan's expense is potentially minimal

and the scanning details are entirely in the hands of the user, so to speak.

Although the field of machine vision has had many false starts in the past, it continues to make steady progress as more computation power is devoted to image analysis. Since the sensor hardware consists only of a simple video camera—already becoming a stock peripheral in PCs and laptops—the added expense is minimal. Because the hardware is unobtrusive (although potentially invasive from a privacy viewpoint)<sup>2</sup> and so much of human communication is gestural in nature, it's a natural interface for smart rooms and responsive environments. Freeman et al. provide a good picture of the promise and current capability in this area. They describe real-time gesture recognition systems they built using machine vision, including descriptions of interactive applications and special “retinal” hardware they developed to offload front-end processing overhead.

Another potential denizen in the evolving world of smart environments is electric field sensing. In its guise as “capacitive sensing,” it has been with us for nearly a century, manifest in well-known devices such as elevator “touch” buttons, proximity detectors for factory automation, and the Theremin.<sup>3</sup> Electric field sensing has mainly been limited in computer interfaces to applications that require tactile contact—for example, planar touch-pads that employ a dense matrix of sensing electrodes<sup>4</sup> and subcutaneous fingerprint capture systems.<sup>5</sup>

The article by Smith et al. describes work we've done here at the MIT Media Lab on applications of noncontact electric field sensing for interfaces in computer graphics applications. Since the hardware expense is minimal and the low-frequency sensing field is not line-of-sight, these sensors can easily be embedded in a variety of “smart objects,” enabling them to perceive user activity in their vicinity. This article outlines the theory and hardware behind the different modes of electric field sensing. It then gives several application examples, including 3D hand sensors and track pads, proximity-sensing computer monitors, and gesture-sensitive walls for interactive projected graphics.

As the requisite hardware improves in fidelity and affordability, computers increasingly push back at us with tactile and haptic feedback. Applications are ripe in the medical field, for instance, where surgeons can hone their techniques and strategies by practicing a particular operation on simulated cadavers or remotely participate in surgeries via telemedicine. Low-end products, such as inexpensive force-feedback joysticks and haptic “mice” through which the user can feel GUI objects have recently appeared,<sup>6</sup> hinting at a chance for haptics to break into the mass market. Massie's article rounds out this issue by describing haptics applications for computer graphics modelers, which let them “feel” their products taking inspiration from the way a sculptor works with a block of clay.

As we head into the next century, the virtual information world will certainly become more important, requiring us to bring it closer to our sphere of interaction and perception. This will engender a shift to entirely new computer interface paradigms dominated by

input devices very different from today's status quo. The articles in this special issue point toward several promising directions, but we have considerable work to do as ever-increasing computational power demands more input bandwidth and enables a fuller and more cogent digital response. You may have heard the phrase “the future's so bright, we'll have to wear shades,” but now we'll have to build them with embedded displays and eye trackers (and still be unobtrusive)! ■

## References

1. J. Strickon, and J. Paradiso, “Tracking Hands Above Large Interactive Surfaces with a Low-Cost Scanning Laser Rangefinder,” to appear in *Human Factors in Computing Systems (CHI'98): Extended Abstracts*, ACM Press, New York, April 1998.
2. S. Mann, “Humanistic Intelligence,” *FleshFactor: Informationsmacchine mensch* (Ars Electronica Festival 97 catalog), Springer, Vienna/New York, 1997, pp. 217-231.
3. J. Paradiso, “Electronic Music Interfaces: New Ways to Play,” *IEEE Spectrum*, Vol. 34, No. 12, Dec., 1997, pp. 18-30.
4. L. Baxter, *Capacitive Sensors: Design and Applications*, IEEE Press, Piscataway, N.J., 1997.
5. D. Setlak et al., “A Direct Fingerprint Reader,” <http://www.semi.harris.com/fngrloc/flocpapr.htm>.
6. L.B. Rosenberg, “Adding a Realistic Sense of Feel to the Computer Experience,” technical document, Immersion Corp., San Jose, Calif., Oct. 1997, <http://www.force-feedback.com/feelit/white-paper.html>.



**Joseph A. Paradiso** is a research scientist at the MIT Media Laboratory, where he develops new sensors for responsive environments. He is currently the Technology Director for the Things That Think program. He received a BS in electrical engineering and physics from Tufts University in 1977 and a PhD in physics at the Massachusetts Institute of Technology in 1981 as a C.T. Compton Fellow. He has held positions at the Draper Laboratory in Cambridge, Massachusetts and the Swiss Federal Institute of Technology (ETH) in Zurich designing high-energy physics detectors, spacecraft control algorithms, and sensor systems. He has also designed several music synthesizers and electronic musical interfaces. He is a member of the IEEE, American Institute of Aeronautics and Astronautics, American Physical Society, and Sigma Xi.

Contact Paradiso at the Physics and Media Group, MIT Media Lab E15-410, 20 Ames St., Cambridge, MA 02139, e-mail [joep@media.mit.edu](mailto:joep@media.mit.edu), <http://physics.www.media.mit.edu/~joep/>.