New Sensor and Music Systems for Large Interactive Surfaces

Joseph A. Paradiso, Kai-Yuh Hsiao, Joshua Strickon, Peter Rice

Responsive Environments Group
MIT Media Laboratory
20 Ames St. E15-351
Cambridge, MA 02139 USA
+1 617 253 8988
joep@media.mit.edu

ABSTRACT

We have developed several systems for tracking hand and body motion in front of large interactive media surfaces, such as rear-projection screens or video walls. In particular, this paper concentrates on a small, inexpensive, phase-measuring scanning laser rangefinder designed to address this niche. This device, mounted below a corner of a screen, can track the position of multiple bare hands in a cleanly-defined plane just above the projection surface to a range of 4 meters and update all measurements at 30 Hz. Unlike most volumetric vision approaches, such as those commonly used in interactive dance projects, this solution is unaffected by background light and optimized for tracking hands close to the screen, yielding prompt, unambiguous results without the need for image processing. We have used this system in several public installations that blend control of music and graphics while exploring the intersection of close-up personal displays with large-scale, "extroverted" interaction. We also introduce complementary sensor techniques that can detect the user activity away from the screen and tactile interactions at the screen surface. We outline three multimedia installations that have used the laser tracker and suggest future implementations.

1) Introduction and Background

Large video surfaces are becoming commonplace in public environments. The development of new display technologies (e.g., plasma screens, video projectors, multi-monitor video walls, LED matrices, electronic inks) will continue to advance [1], and enable large-screen and wall-sized monitors to further infiltrate our daily lives. Although today's smaller cousin, the video kiosk, is fully interactive, often via a touchscreen interface, large display walls are usually passive output devices with no intrinsic input capability aside from an external keyboard, mouse, or remote control. Various researchers have used computer vision with very large projection walls for experiments in augmented reality [2] and interactive dance [3], but most of this work concentrates on tracking activity in a extended room, rather than monitoring precise gesture at the surface of the screen itself. Close-up interaction at large display walls, however, is a very interesting hybrid between the private dialog that one has with a video kiosk and a large public spectacle, where one's interaction at the screen can be openly viewed by anybody in the vicinity. Many possibilities arise at this juncture, including new families of musical controllers. With moderately precise tracking of hand positions across the screen, additional measurements that indicate the hand's dynamic shape, lower precision response to the position and motion of the hand and body away from the screen, tactile, dynamic response when touching or tapping the screen, and interactive projected graphics to give direct visual feedback, such systems provide fertile ground for multimodal musical expression.

Previous work by the authors and collaborators on the Brain Opera project [4] has explored the use of touchscreens and larger projection displays as musical controllers. Standard touchscreens [5], while convenient, are limited in size and their technologies don't economically scale to large-area displays. The Brain Opera's large-screen interfaces (the "Gesture Walls" [4]) used electric field sensing [6], transmitting from the performer's body to receivers at the display corners. While this captured generic gesture, hand tracking with any precision was only possible when observing strict postural constraints, as the capacitive receivers detected the entire body mass, and not just the hands.

There have been several, mainly vision-based approaches to tracking hand and body movement against interactive walls over the past decades, and most have been tried with interactive music of one sort or another. These have included front-looking systems (such as blue-screen chromakey [7] as used by weather forecasters on television and the Mandala System [8] used for immersive 2D environments) and rear-looking systems that blast IR light through a translucent screen to image the users in front [9]. While effective in some applications, these techniques have limited tolerance to clutter and background light, plus require clear lines of sight. Other strategies look along the edges of the screen, using multiple video cameras [10], which require potentially significant processing to disambiguate image features and similarly constrained lighting.

A related application area is that of smart wallboards, which precisely capture drawings made with coded pens. Again, there are several techniques, exploiting active sonar transponders [11], passive magnetic LC tags [12], and triangulation between a displaced pair of scanning IR lasers [13]. All of these, however, are designed to work with special visually or electronically labeled handheld targets. They do not work with bare hands, as desired for large public installations.

Several optically-based electronic music controllers have been developed to sense hand motion in a plane. Most of these work when back-illumination is shadowed or occluded by the hand. Some classics in this family are CMU's VideoHarp [14] and the LaserHarp [15]; the latter made famous by Jean-Michelle Jarre and Bernard Szajner. While these devices have great aesthetic appeal in performance, they have varying restrictions stemming from background light tolerance, scalability, and, in the case of Jarre's LaserHarp, the need to wear asbestos gloves.

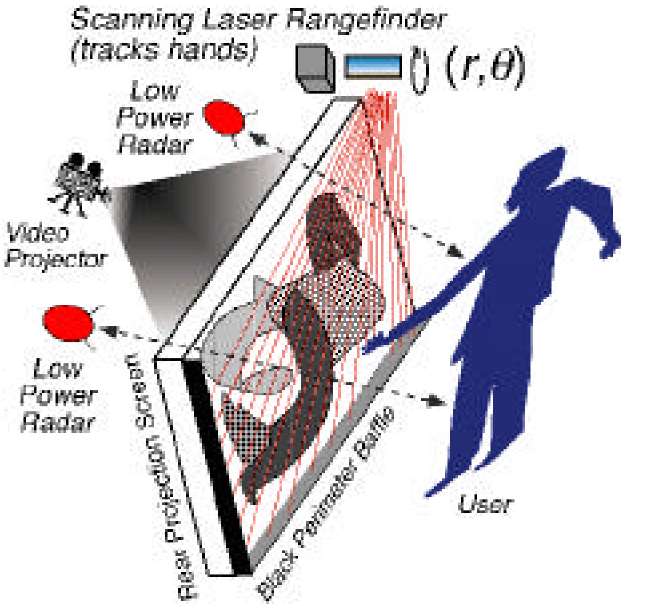

Figure 1: Concept behind the LaserWall System

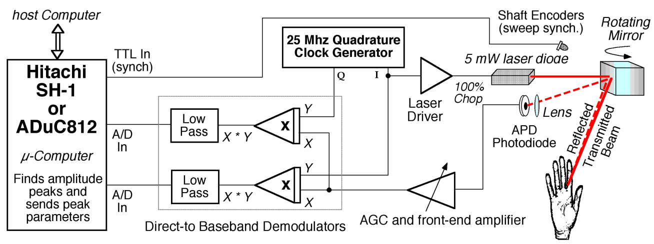

Figure 2: Block diagram of low-cost CW-phase rangefinder

2) The LaserWall

Our solution to this niche, introduced in [16] and detailed in [17], is shown in Fig. 1. We place a compact, scanning laser rangefinder in the corner of a projection screen. The rangefinder scans a plane with a simple rotating mirror, detecting hands by their reflection against a matte-black baffle. As this plane is very cleanly defined, it can be placed as close to the projection screen as desired. The positions of the hands are determined in polar coordinates; θ by the angle of the mirror at the point of reflection and r by measuring range along the beam. By strobing the laser at 25 MHz, we can synchronously detect the hand's reflections, causing our detector to only see the laser illumination and removing sensitivity to background light. By quadrature-demodulating the reflected signal by the laser's strobing clock, we measure the shift in phase between the emitted and reflected light, thus the time it takes the light to reach and return from the hand, hence the hand's range (unambiguously out to 6 meters before the 25 MHz modulation wavelength phase-wraps). Such CW phase-measuring rangefinders have constant accuracy over range (as opposed to triangulation devices, which are asymptotic and less compact), and are easier to implement for fast, accurate measurement than pulsed time-of-flight devices. It should be noted that most optical distance-measuring interfaces designed into today's electronic music controllers (e.g., the various IR Theremins) don't measure true range, but infer it from the magnitude of reflected intensity, which is highly influenced by variations in complexion and aspect angle. As our device is a true rangefinder, it exhibits no such effects. The low-power radars in Fig. 1 detect position and/or motion of people in front of the screen, and are described later.

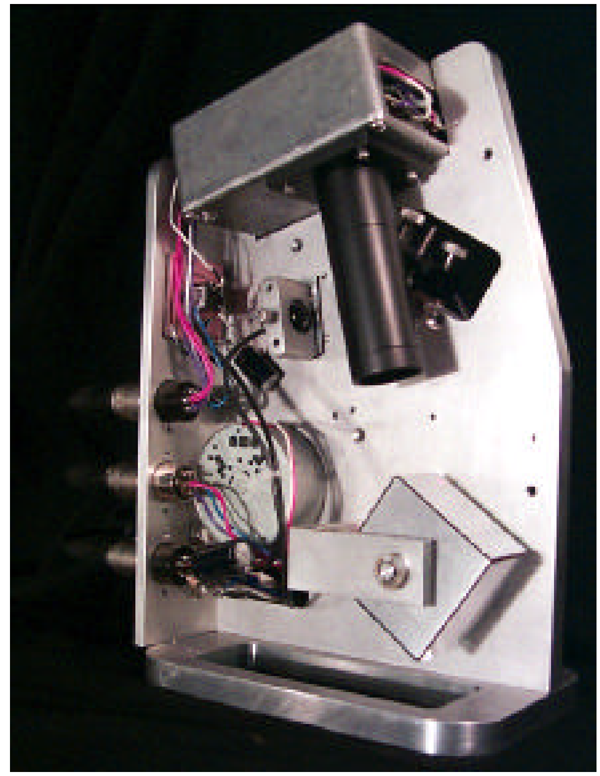

Figure 3: The operational scan head

Although scanning laser rangefinders are commercially available, with applications in areas like survey and robotics [18], they are quite expensive. We accordingly opted to design and build our own low-cost devices, optimized for tracking hand position above large display surfaces. Our design, diagrammed in Fig. 2 and shown in Fig. 3, is very simple. By minimizing the required electronics, using inexpensive and widely-available components, scavenging surplus assemblies from cheap products (e.g., laser pointers and CD players), and fabricating the mirrors ourselves, we have kept the basic price down; each device costs under US$500 in parts, strongly dominated by the avalanche photodiode and its power supply.

We convert directly to baseband with a pair of fast multipliers, and digitize the quadrature pair with a modest microcomputer, which finds the reflection peaks from the hands and fingers (Fig. 4) by subtracting residual baseline reflection from the black baffle (which is responsible for the slow curve in the "Q" trace of Fig. 4) and applying a simple threshold. Our current device scans its full 90° span at up to 30 Hz and transfers the parameters of all peaks to a host PC after each scan is completed. The PC runs the peak amplitudes through an arctangent function and a third-order polynomial calibration to produce Cartesian screen coordinates. No image processing of any sort is necessary. The width and net amplitude of each hand's reflection are also measured, allowing for additional control through hand rotation and manipulation.

Although the rangefinder itself is able to sense further, we've adjusted the demodulation filter and microcomputer sampling rates to enable the detection of hand-sized objects up to 4 meters away with a uniform 1.4 cm resolution when scanning at 30 Hz. Much of the error comes from wobble in our homemade mirror; a 4-point FIR filter removes this and gives very precise tracking. Fig. 5 shows an example of the actual coordinates produced by this system when freehand writing in air. As we use a 5-mw red laser diode, the system is eyesafe at our scan rates and traces a dim red line on the hand for visual feedback.

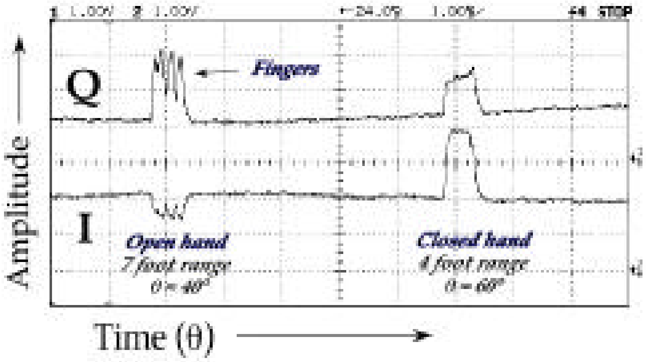

Figure 4: Quadrature response of rangefinder to pair of hands

3) Interactive Media Installations

We have used this system for several public interactive music installations, as seen in Fig. 6. The leftmost example [19] was our first and simplest environment. The graphics routine plotted a red and green rotating square at the location of each detected hand, and changed the background objects as the hands were moved about. Drum loops would start when at least one hand was detected and change with their position. A bass-lead arpeggio would begin when more than one hand was introduced (with tonal range dependent on their mutual distance). Although quite simple, it was a very popular installation at several 1997 Media Lab events.

The middle photo, showing a port of the mouse-driven *Stretchable Music* program [20] written by one of the authors (Rice) to a 4' x 8' projection screen, was considerably more complex. Here, users could grab and stretch objects that sequentially appeared on the screen, each of which exhibited different musical behavior in accordance with the amount of stretch. If more than one hand was detected, the additional hands would draw musical "sparklies", i.e., twinkling points centered on the hand positions that made soft, ethereal sounds. This installation was run for a week at SIGGRAPH 98, where it was likewise very popular with conference visitors. The large-screen interaction engaged the entire body, making it much more compelling than the original GUI version.

Our most recent application (Fig. 6 right) is a LaserWall version of the *VisualWho* graphical database exploration program developed by Donath [21], shown as it was being demonstrated at *Opera Totale 4* in Meistre, Italy during November of 1999 (and to be exhibited at SIGGRAPH 2000 in New Orleans). Here, users could grab graphical anchors coupled into a complicated database, and move them around on the screen, seeing representations of related information attract and drift to the anchor being moved, depending upon their degree of relationship. An interactive music system was also used, with different themes and effects tied to the different anchors and the manner in which they were manipulated. As this complex music system was designed expressly for the LaserWall environment, we provide some detail below.

The music consists of eleven relatively independent layers that work together musically. One of them is a default layer that plays whenever a hand is sensed, and the other ten correspond to anchors used by the *VisualWho* system, and play from the time the anchor is "dropped" onto the active screen region until the anchor is moved back into the inactive "storage" sidebar. Most of the voices in the music system have several behaviors in common. Their volumes vary with the speed of the hand across the screen, regardless of whether the corresponding anchor is being dragged or not. Their volumes are also adjusted by the amount of kinetic energy imparted to the graphical objects. When no hand is present, all voices are generally quieter, and play slowly in the background. Inserting a hand plays all the layers that are present faster and louder and moving the hand quickly around increases the volumes. Removing the hand causes all musical lines to stop and play a chord instead, and then they resume their slow default behavior soon afterwards. Most of the voices will modify their timbres when the corresponding anchor is grabbed and dragged around. Only the voice belonging to the anchor being dragged will have a modified timbre; the others will increase in volume. The number of anchors on the active screen will change the harmonic content between the five harmonies written into the system. The layers themselves range from high-pitched "twinkles", to melodic lines, to background chords and drums.

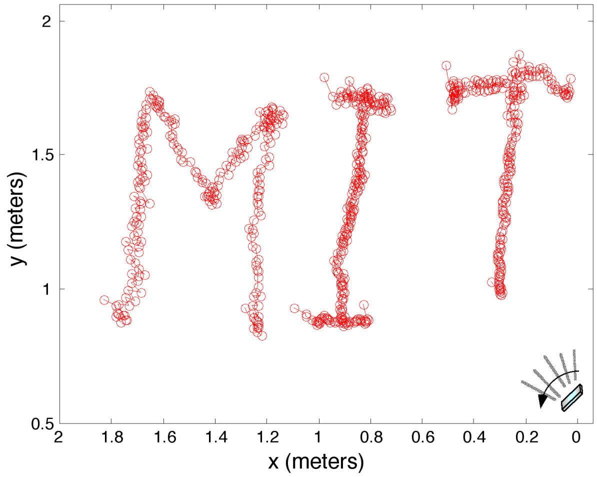

Figure 5: Rangefinder tracking for freehand writing in air

4) Complementary Systems

Although the LaserWall is an engaging interface on its own, its utility can be enhanced when combined with other simple sensor systems that we are developing to detect different kinds of activity in the vicinity. Fig. 1 indicates a pair of "low power radars" behind the screen. These are very simple microwave motion detectors [4] or micropower ranging radars [22], placed out of the light cone of the projector that are able to sense directly through the screen and infer the proximity and degree of bulk motion of the users as they interact. The media content can be then given a response to free gesture away from the screen, in addition to precisely tracking hands when they cross the laser's scan plane.

We have also developed a very simple system [16] to track the position of taps on a large continuous sheet of material such as glass, wood, or plastic. By placing PVDF piezoelectric pickups near the corners of the sheet, and measuring the differential time of the first wavefront arrivals, we obtain an estimate with modest (σ = 3 cm across a 1 m x 1 m sheet of glass) resolution. As the signature of the wavefronts can vary with the type of impact, the structural-acoustic characteristics of the material, and the positions of the pickups, we use an embedded microcomputer to process the pickup signals and infer the timing of the leading edge. In commercial plate glass, for example, we detect slow (450 m/s) flexural waves that emanate from the impact and compensate the timings for the effects of dispersion. We are currently exploring musical content mappings for this interface (e.g., a zoned percussion instrument) and applications in concert with the radar and rangefinder devices, where the user will be sensed away from the screen, just atop the screen, and in contact with the screen.

a) Rotating Cubes '97 b) LaserStretch '98 c) LaserWho '99

Figure 6: Public multimedia installations with the LaserWall

5) Conclusions

Although, at 30 Hz updates and $\sigma = 1.4$ cm resolution, our rangefinder system is currently a bit slow and coarse for a fluent musical interface, the ability to smoothly track hands across large interactive graphical areas opens new possibilities for compelling musical interaction, as the installations with which we've used this device have indicated. New families of interaction are enabled that lay midway between close-up, personal user interfaces and public performance.

A single rangefinder located in one corner of the screen will be able to track multiple hands, which will appear as additional peaks in the scan lines of Fig. 4. If the hands occlude one another (e.g., line up with the rangefinder), however, the system won't be able to separate them. By using contextual information and past data, however, this problem may be minimized (for example, the graphics can respond in a way that discourages overlapping hands). We are currently exploring several solutions to this situation.

We are developing other systems that complement the laser tracker by sensing user activity away from and in contact with the screen. These degrees of freedom promise to provide an immersive musical interaction that works at several levels.

6) Acknowledgements

We thank our research colleagues in the Responsive Environments Group for their assistance in this project, and acknowledge the support of the Things That Think Consortium and other sponsors of the MIT Media Lab.

7) References

1. Werner, K. 1997. "The flowering of flat displays." *IEEE Spectrum*, vol.34, no.5, May 1997, pp.40-9.
2. Maes, P. *et al*. 1997. "The ALIVE System: Full-body interaction with animated autonomous agents," *Multimedia Systems*, vol.5, no.2, pp.105-112.
3. Paradiso, J., Sparacino, F. 1997. "Optical tracking for music and dance performance", in *Optical 3-D Measurement Techniques IV*, A. Gruen, H. Kahmen Eds. pp. 11-18. Heidelberg Germany: Herbert Wichmann Verlag.
4. Paradiso, J. 1999. "The Brain Opera Technology: New instruments and gestural sensors for musical interaction and performance," *Journal of New Music Research*, Vol. 28, No. 2, pp. 130-149.
5. Quinnell R.A. 1995. "Touchscreen technology improves and extends its options." *EDN*, vol.40, no.23, pp.52-64.
6. Paradiso, J., Gershenfeld, N. "Musical applications of electric field sensing." *Computer Music Journal*, 21(3), 1997, pp. 69-89.
7. Wojdala A. 1998. "Challenges of virtual set technology." *IEEE Multimedia*, vol.5, no.1, pp. 50-7.
8. Stanfel, J. 1993. "Mandala: Virtual Cities." *SIGGRAPH 93 Visual Proceedings*, p. 208.
9. Rekimoto, J., Matsushita, N. 1997. "Perceptual surfaces: Towards a human and object sensitive interactive display." In *Workshop on Perceptual User Interfaces* (PUI-97), Banff, October 1997, pp. 30-32.
10. Sharma, R., *et al*. 1996. "Speech/gesture interface to a visual computing environment for molecular biologists," Proceedings *of the 13th IEEE International Conference on Pattern Recognition*, vol. 3, pp.964-8.
11. Yates, C. 1999. "mimio: A whiteboard without the board." *PC Week*, vol. 16, no. 26, p. 50.
12. Murakami, A., *et al*. 1995. "Position Detector." US Patent No. 5,466,896.
13. Eglowstein, H. 1994. "Almost as good as being there." *Byte Magazine*, April 1994, pp. 173-175.
14. Rubine D, McAvinney P. 1990. "Programmable finger-tracking instrument controllers." *Computer Music Journal*, vol.14, no.1, pp. 26-41.
15. See: http://home.swipnet.se/~w-69229/laserharp.htm
16. Strickon, J., Paradiso, J. 1998. "Tracking hands above large interactive surfaces with a low-cost scanning laser rangefinder." *Proc. of CHI '98*: *Summary*, pp. 231-232.
17. Paradiso, J., *et al*. 2000. "Sensor systems for interactive surfaces." To appear in the *IBM Systems Journal*.
18. Everett, H.R. 1995. *Sensors for mobile robots:theory and application*. Wellesley, MA: A.K. Peters, Ltd.
19. Strickon, J.A., 1999. "Design and HCI applications of a low-cost scanning laser rangefinder." MS Thesis, MIT Department of Electrical Engineering and Media Lab.
20. Rice, P.W. 1998. "Stretchable music: A graphically rich, interactive composition system." MS thesis, MIT Media Laboratory.
21. Donath, J. 1995. "Visual Who: animating the affinities and activities of an electronic community." In *Proc. of the 3'rd ACM Int. Conference on Multimedia*, pp. 99-107.
22. Azevedo S, McEwan TE. 1997. "Micropower impulse radar." *IEEE Potentials*, vol.16, no.2, pp.15-20.