Tracking Hands Above Large Interactive Surfaces with a Low-Cost Scanning Laser Rangefinder

Joshua Strickon
MIT Media Laboratory
20 Ames St. E15-495
Cambridge, MA 02139 USA
+1 617 253 0804
strickon@media.mit.edu

Joseph Paradiso
MIT Media Laboratory
20 Ames St. E15-410
Cambridge, MA 02139 USA
+1 617 253 8988
joep@media.mit.edu

ABSTRACT
We have developed an inexpensive scanning laser rangefinder to measure the real-time position of bare hands in a 2-D plane up to distances of several meters. We have used this device to build a precise, multipoint "touch-screen" interface for large video projection systems. In this paper, we describe the concepts and hardware, plus outline an application for an interactive multimedia environment.

Keywords
Laser rangefinder, hand tracker, touchscreen, music interface

INTRODUCTION
Many different technologies have been applied to track the position of objects atop interactive surfaces. The largest commercial market for these items has been in "smart wallboards", that digitally record handwriting. While many of these systems require contact or pressure to be applied against the sensitive surface (working like a large touchscreen), others detect the position of objects just above the wallboard. Most of these employ optical sensing, which enables simple, passive reflecting targets to be easily detected in a sensitive plane defined by a scanning collimated light source, such as a diode laser. A well-known example of this is the "SoftBoard" by Microfield Graphics [1], where a pair of scanning lasers emanate from the two top corners of the board, identifying and tracking coded targets on pens and other objects approaching the whiteboard and intersecting the scanned plane. These sensors are unable to range, thus planar position is determined by triangulating the two angular measurements. To avoid ambiguities in this triangulation, these systems generally allow only one object to be tracked at a time.

Essentially all of the noncontact whiteboard systems require tagged targets of some kind. They are not designed to detect and track generic objects, such as bare hands, as required in our applications for large interactive projection displays. Our earlier work [2] toward this aim resulted in a "Gesture Wall", where we used Electric Field Sensing to track hand and body motion in front of a projection screen that displayed interactive graphics. Here, the user's body was driven by a weak electric field coupled in through the floor; a set of receive electrodes placed about the display perimeter provided signals that corresponded to body distance. Although this system responded well enough to body dynamics and location for its original application, the limited disambiguation from four receive electrodes was unable to result in a repeatable hand tracker without excessively constraining the body's posture and placement.

Other groups have implemented hand trackers using video cameras and computer vision techniques. Some [3] employ IR light sources and cameras behind a translucent rear-projected screen to see hands near the front, while others [4] use multiple cameras to observe a 2D gesture space. Like all vision systems, the performance of these systems can suffer from background light (including light from the display itself in the latter case), image clutter, limited speed of response, and the need for multi-camera correspondence.

THE LASER RANGEFINDER
Our solution to this problem, depicted in Fig. 1, is to use a scanning laser rangefinder at one corner of the display to determine the polar (r,ϕ) coordinates of hands in a plane above the projection surface. Laser rangefinders are commercially-available devices. Triangulation rangefinders can provide very high depth resolution over limited dynamic range, thus are often used in 3D object scanning. Continuous-wave (CW) phase-measuring systems, often used in robot collision avoidance and survey equipment, are well suited to this application, as they can give good resolution (mm-level) across many meters of dynamic range. Such devices are prohibitively expensive, however;

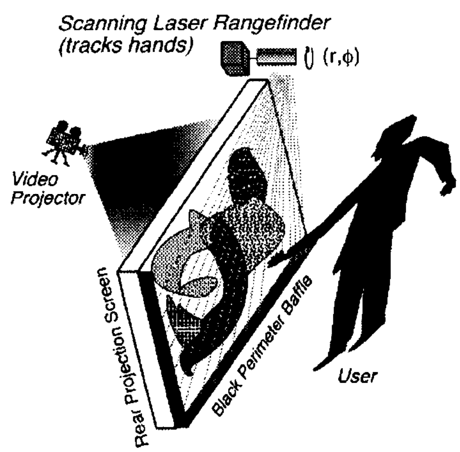

Figure 1: Laser Wall Setup

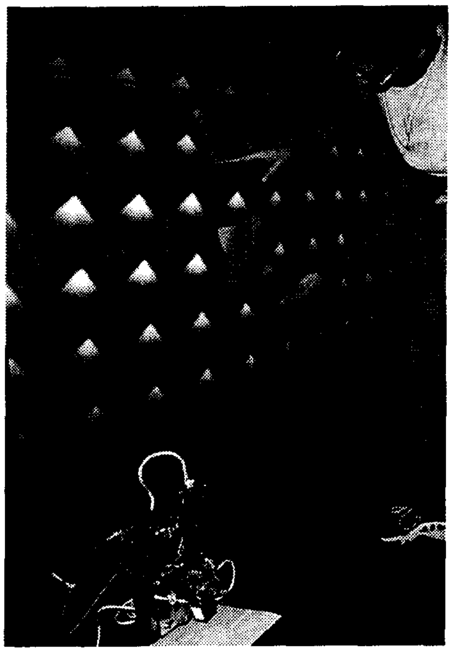

Figure 2: Prototype Rangefinder Tracking 2 Hands

often costing of order US $10,000. As our requirements are comparatively modest (≤1 cm resolution, 4 meters of range, 90° scan coverage, 25 Hz updates), we have opted to design and build our own less costly rangefinder.

Our prototype is visible at lower-left in Fig. 2. A 5 mW red laser diode is driven at 25 MHz; laser light reflecting off a hand is focused onto an avalanche photodiode (APD), which produces a signal that is compensated for differences in skin complexion by a low-noise AGC amplifier. The phase-shift in the detected signal, proportional to the time it takes the laser light to reach the hand and reflect back to the APD camera, is proportional to the hand's range. This is measured in a simple direct-to-baseband synchronous quadrature demodulator, which also rejects essentially all background light; only the laser's reflection is detected. Since they are directed colinearly, the 4-sided polygonal mirror scans both the laser beam and camera's field-of-view at up to 30 Hz. At the beginning of each scan, a free-running counter in an embedded PIC16C73 microcomputer is reset, thereby providing a precise angle measurement. The PIC also continually digitizes the quadrature pair of received signals into 8-bits during each scan. Reflecting objects, such as hands, produce very clear amplitude peaks in the sampled signals (the projection screen is enclosed in a matte-black perimeter baffle to enhance this contrast). The PIC detects these peaks using a simple threshold, and at the conclusion of the scan, reports the angular centroid and width along with the integrated quadrature pair for each peak to a host computer over an RS-232 link.

As the electronics are fairly minimal, this system is quite inexpensive; the parts cost is well under US $500., and this is dominated by the APD, which, with our modest 4 meter range requirement, may be able to be replaced by a much less costly PIN photodiode. Our prototype meets the application requirements sketched above, and as the embedded PIC is able to do the bulk of the needed signal processing (extracting peaks corresponding to hands and calculating their basic parameters), there is little additional processor overhead required to produce (r,ϕ) hand coordinates. Because of the synchronous demodulation, this device is unaffected by background light, as with the video systems described earlier. Only a single scanner is required to provide both r and ϕ, resulting in a simpler system with less ambiguities than either the multiple-camera or SoftBoard solutions. At 100% modulation and 25 Hz scan rate, the current system is eye-safe. Added protection can be introduced, i.e., by shutting down the laser if a large object is detected at too close a range or running at lower power.

Although the rangefinder system is able to track multiple hands, it is subject to occlusion, where one hand "shadows" another. For many applications, this may be addressed in software (i.e., by introducing a tracking filter [4] or "inertia" to the trackpoints), or rigorously by adding a second scanner (since each directly produces both [r,ϕ], there is no correspondence problem; all points are unambiguous).

INTERACTIVE MULTIMEDIA APPLICATION

We designed a simple interactive multimedia environment in order to test this system, connecting our rangefinder to a PC running music and graphics software written in C++, and setting it up to scan a large projected surface, as in Fig. 1. The graphics routine plotted a red and green rotating square at the location of each detected hand, and changed the background 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). Figure 2 shows this environment in use with two hands detected; it has been a very popular installation at several recent Media Lab events.

SUMMARY

We have demonstrated a new, robust interface system for tracking 2D hand positions that is very well suited for interactive projection environments.

ACKNOWLEDGMENTS

We thank Ed Boyden, Matt Reynolds, Neil Gershenfeld and our other Physics and Media colleagues. We acknowledge the support of the Things That Think Consortium and our other sponsors at the MIT Media Laboratory.

REFERENCES

1. Eglowstein, H., Almost as Good as Being There. Byte Magazine, April 1994, pp. 173-175.
2. Paradiso, J. Electronic Music Interfaces. IEEE Spectrum 34(12), December 1997, pp. 18-30.
3. Rekimoto, J. and Matsushita, N. Perceptual Surfaces: Towards a Human and Object Sensitive Interactive Display. In *Workshop on Perceptual User Interfaces* (PUI-97), Banff, October 1997, pp. 30 - 32.
4. Rehg, J. and Kanade, T. Digiteyes: Vision-Based Hand Tracking for Human-Computer Interaction. *Proc. of the Workshop on Motion of Non-Rigid and Articulated Objects*, Austin, Texas, November 1994. IEEE Computer Society Press, pages 16-22.