

# Gestural Interaction and Haptics for a more Robust Reality Based Interface

*John Delaney*

## Abstract

As computers become smaller, more powerful, and more common, they are shedding their keyboards for a bevy of more natural gestural based interfaces. Multi-touch surfaces, screens which are capable of detecting multiple points of user input, are becoming popular tools for researching gestural interfaces. The Media Arts and Technology (MAT) 200C class at University of California Santa Barbara (UCSB) has recently implemented our own multi-touch interface table. In this paper I will discuss the configuration of the table as well two areas of research that would augment the table's current functionality: haptic feedback and single camera tracking.

## 1 Background

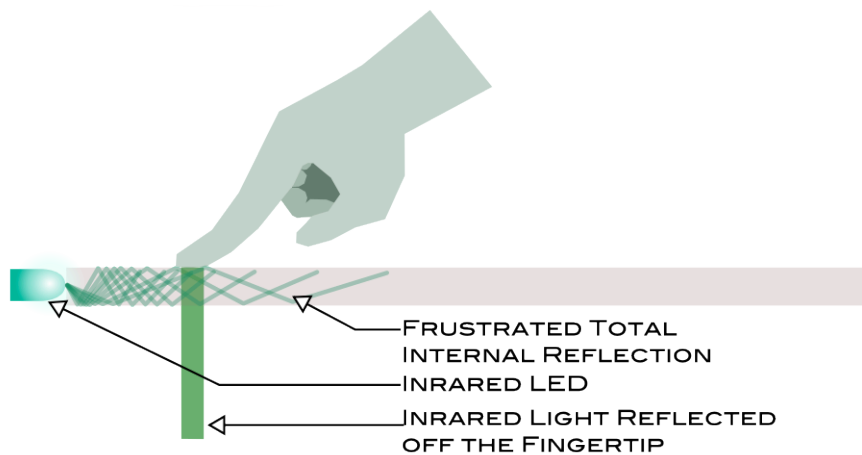


Fig. 1: *Diagram depicting a simplified representation of the physics governing the Frustrated Total Internal Reflection effect*

In the physical world, an interface occurs at the edge where two materials meet, such as between two fluids, air and water. In computer science an interface refers to the interaction between two components such as the graphics card

and the main processor, or the user and the keyboard. The modern computer interface was born out of a technological environment that perceives the natural and built environments as black and white. The modern notion of an interface assumes that the user is interacting with something that isn't interacted rather than interacted with. Computer interfaces have evolved in the 60 years of computing. Once upon a time, scientists in labs had to physically punch programs on a card to interface with computers. Now a majority of the world's population has a computer in their pockets, but since the first Macintosh desktop computer, the desktop interface paradigm has remained unaltered as the defacto interface for most computer devices. With the increasing prevalence of ubiquitous sensor technologies such as accelerometers and digital cameras and processors capable of handling the incoming data, computers are finally able to move into a new era of Reality Based Computing, in which a user can interact with the computer as they would react to their natural environment.

Recently the concept of a Reality Based Interface (RBI) has become popular among researchers [7]. This proposed interface paradigm seeks to move beyond the traditional WIMP (window, icon, menu, pointing device) desktop model and embrace concepts grounded in psychology and humanity. The RBI takes into account four themes: Naive Physics, Body Awareness and Skills, Environment Awareness and Skills, and Social Awareness and Skills. A common problem with traditional WIMP (window, icon, menu, pointing device) based multimedia programs is functionality hidden behind menus and preferences. By mapping gestures or fiducial objects to a programmatic event or functions with intuitive corresponding gestures, the learning curve associated with new application interfaces might be reduced [7]. The previous single camera tracking system is one example of a gesture capturing system. Each gestural input system has its own level of expressive completeness and structural generality and adding layers of gestural input to a surface computing system would pave the way for research in physically based multi-user interfaces that can be interfaced naturally, requiring more from the user physically, and expanding the current computer environment beyond the mouse and keyboard and into the physical and social space around it. For our project in MAT 200c we set out to build a multi-touch table as a research platform with which to research Reality Based Interfaces.

To build our multitouch table we followed the instructions published by Maximum PC [10]. The multi-touch table proposed by this article functions by using computer vision to capture the movements of the users fingertips imaged on a transparent piece of plexiglass. This presents a challenge because the projector must project onto an opaque surface, blocking the visual cues necessary to calculate the user's movements. This problem was solved by the New York University's Media Research Lab [12] by using frustrated total internal reflection (FTIR) of infrared light. Total internal reflection occurs when light passing through a prism strikes the boundary at greater than the critical angle. Light is then trapped inside the prism, continually bouncing off its walls until it reaches a non critical exit point. This phenomenon allows for technology such as fiber-optics which function by transmitting pulses of light [12]. Frustrated

Total internal reflection occurs when a new material comes in contact with the prism at an internal reflection point creating an evanescent wave allowing light to pass through the prism instead of continuing to reflect internally. In the case of the touch table this light then reflects off of our fingers downwards towards an infrared camera. To achieve this end our table utilizes a Unibrain Fire-i firewire web camera with a special lens which lets in infrared light and a linear array of over 100 infrared LEDs directed perpendicularly into the side of a polished a 24x36 piece of plexiglass as depicted in figure 3. For imaging purposes we have a Hewlett Packard projector with a Wide Throw Lens, which projects onto a piece of silicon coated vellum that sits on top of the plexiglass. This system is robust and suitable for research in collaborative interaction and gesture based application interaction.

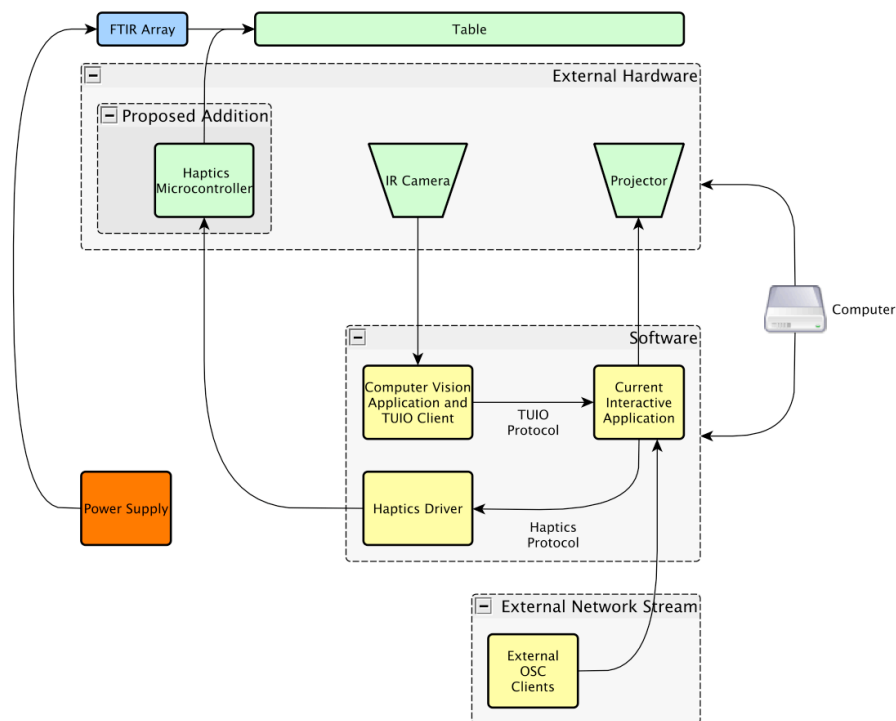


Fig. 2: Graph of the touch table interface in its proposed future configuration. All three dimensional gesture recognition would be handled by the the computer vision client.

To implement object tracking we are using the Community Core Vision (CCV) finger tracking system:

Community Core Vision, CCV for short (aka tbeta), is a open source/cross-platform solution for computer vision and machine sens-

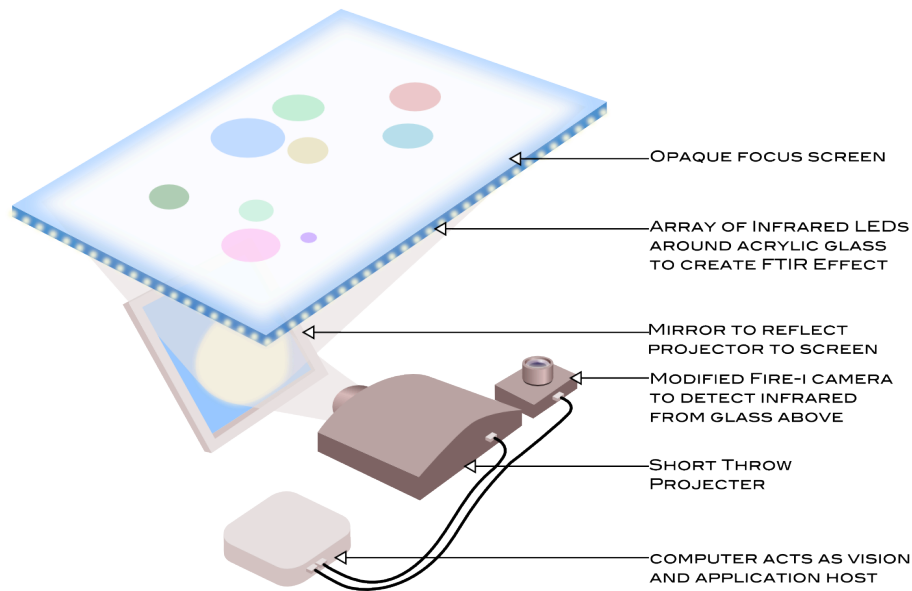


Fig. 3: Schematic diagram of MAT touch table interface

ing. It takes an video input stream and outputs tracking data (e.g. coordinates and blob size) and events (e.g. finger down, moved and released) that are used in building multi-touch applications [2].

The Community Core Vision performs tracks finger movement using a technique called blob detection which is able to track movement of localized luminescent regions, corresponding to the users fingertips. It has a set of controls to improve surface tracking quality such as thresholding, high pass filter, background subtraction, and projection calibration. CCV communicates with other applications via Open Sound Control using the TUIO Protocol. TUIO is a standardized open protocol and API for tangible multi-touch interfaces [3]. Because CCV utilizes Open Sound Control, it is not limited in what applications it can send messages to as long as they can receive open sound control. This frees the application developer from having to master a complicated API so they can focus on analyzing tracked movement and interpreting gestures.

## 2 Single Camera Tracking: Pose Estimation and Gesture Recognition

Gesture refers expressive actions made by the hands, face, and other parts of the body. Gesture recognition is the process by which a receiver interprets these actions [13]. Developing methods of recognizing gesture are paramount to proper implementation of a Reality Based Interface addressing a computers need to respond to a users environmental awareness, bodily awareness and naive

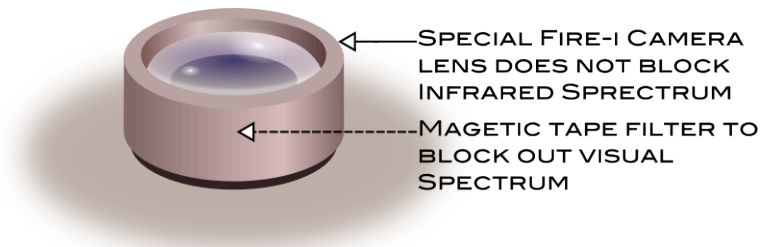


Fig. 4: *Detail of the Fire-i camera lense*

sense of physics. Many computational techniques exist to capture and interpret natural gestures. Here I present two methods by which the MAT table's gestural interface palette can be extended: pose estimation based on tracking an infrared fiducial and pose recognition of natural hand shapes. Because MAT's table uses an infrared camera for multi-touch processing it is naturally a very flexible and reconfigurable interface for research in gestural interaction. With software modifications, it is possible reconfigure the table to interact with active infrared markers. By reconfiguring current research methods to use actively lit infrared markers it would be possible to easily impliment these techniques on MAT's table interface.

Because much of the research occurring at MAT deals with spatial human interaction computing, systems allowing multiple users to interact naturally in a 3D spatial environment are desired. Santos et al. have proposed a system for tracking of a fiducial object in three dimensions with a single camera. In their system, they have combined a unique low cost square fiducial marker with a robust single camera pose estimation system to capture the orientation and distance of an object from the tracking camera. A square fiducial is configured with six reflective markers, four for the corners, one to register the top, and an arbitrary marker is used to determine uniqueness. Markers captured by the camera are spatially hashed and tracked each frame. Their Ptrack pose estimation system then scans the tracked points, first searching for the top orientation row, then the four corners. Once a potential fiducial has been found and its a projection fitted, the scene then checks the location of the marker's unique identifier with the registered fiducials, thus creating a system capable of tracking multiple fiducial objects.

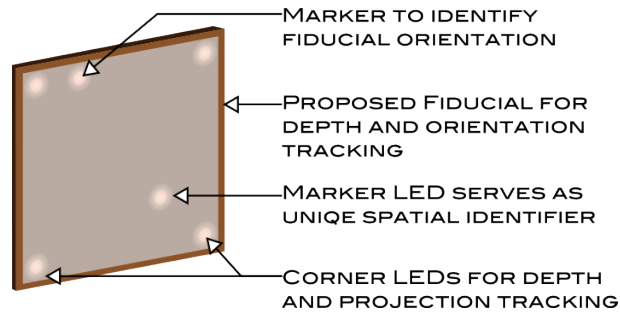


Fig. 5: Proposed fiducial for recognition of three dimensional orientation via pose estimation. The four corner LEDs are used for perspective calculations, the top middle LED is used to calculate orientation and the central LED is a unique ID

There is no reason why a single camera infrared tracking system could not act as a generic platform for gestural motion capture capable of distinguishing between many different forms of gestural communication. Hand-shape is also very important in human gestural interaction. Athitsos and Potamias [6] propose a system of comparing hand poses to a large database of synthetic poses derived from a 3D model. This model allows the storage of many different pose configurations allowing for computer interactions akin to human sign language. Oka and Sato [5] use a model of prediction and correction coupled with a Hidden Markov Model to accurately track multiple fingers. Their research also utilizes a table interface, however, with the camera placed overhead rather than underneath. They have also proposed a system of tracking gestures and distinguishing between symbolic gesture and application manipulation.

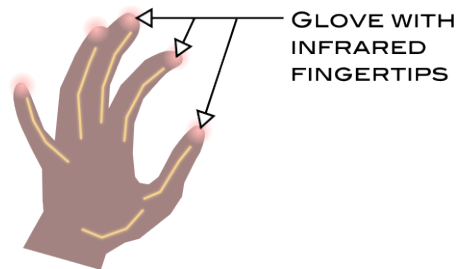


Fig. 6: Proposed glove system for gestural tracking

Marker based single camera tracker systems are attractive because reflective fiducial markers can be replaced with infrared LEDs allowing the fiducials to be seen by the camera through the screen. Additionally for marker-less gesture tracking methods, a glove with embedded IR LEDs could be worn (rendering them no longer markerless). Use of Infrared LEDs would allow fiducials and gestures to be tracked in the same environment as the current FTIR surface

tracking. This method is dependent however on selection of an opaque focus screen capable of passing enough infrared light from the fiducial above while maintaining a good projection of the application environment. Even if the depth and orientation information resolution is limited, the gains in gestural expressiveness are would be worth it. More research needs to be done to incorporate the marker-less gestural recognition systems with the FTIR based tracker system.

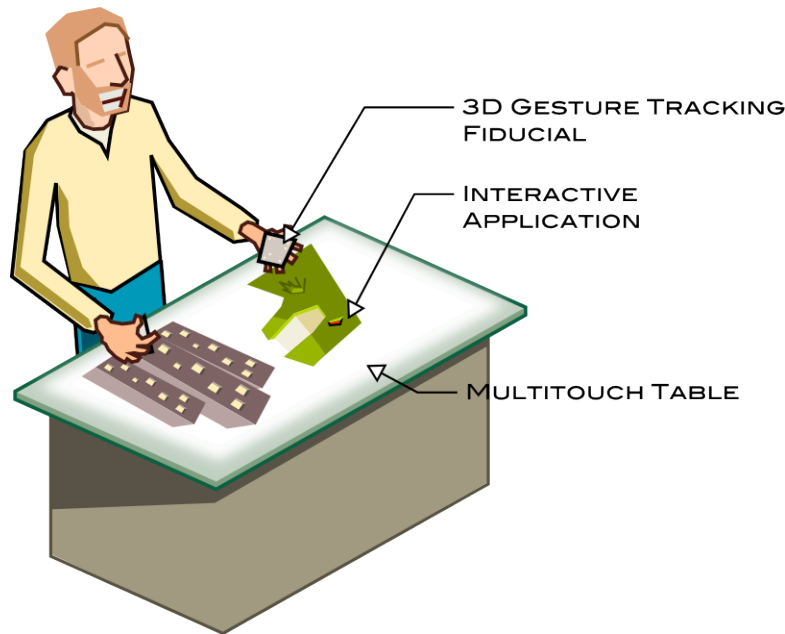


Fig. 7: *The proposed system would layer multiple gesture, FTIR and tracker based interfaces to create a robust gestural interface, potentially removing the menu based layout of traditional multimedia tools.*

### 3 Haptics

The term haptic refers to the sense of touch [8]. Most touch screens for computers lack a tactile quality, which ignores one of the four themes of the RBI concept: Bodily Awareness. Though the immediacy of a touch table appeals to the users sense of (Naive) physics, there is a palpable disconnect between what happens on the screen and what the user sees as a result of their action. I propose bridge the physical gap by installing a simple haptic feedback system into our touch table. The system would comprise of a micro-controller, solenoids, linear motors and a simple OSC protocol for communicating with the micro-controller drivers.

In the past several years a lot of attention has been paid to the use of

haptics in Virtual Reality. Most of this research, however, has occurred in areas such as tele-surgery and flight simulation. Research on haptic feedback in an educational environment has yielded positive results [9]. In a study conducted by Jone's et al [9] three treatment groups (visual only, haptic only, and visual and haptic) were told to explore a series of objects under an atomic force microscope. The results of this study indicate that the groups that received haptic feedback performed better at identifying the object, which is evidence that perhaps haptic feedback might perform a role in the development of spatial understanding [9]. Most haptic research is concentrated in areas of professional application and training. The most successful commercial haptic applications seems to be simple systems such as the vibrator in cell phones or rumble packs in video game controllers, both of which have a very low expressive resolution. There appears to worthwhile research in the area of haptic feedback in multimedia interactivity as it pertains to Reality Based Interfaces.

A haptic system doesn't have to be complicated to provide a greater level of expressive feedback than currently exists. By specifying a protocol in which linear motors, vibrators, and solenoids installed in the underside of a multi-touch surface can be manipulated, application developers can then assign physical responses to programmatic changes or different type of user interactions or gestures. Types of mappings include assigning rate of vibration with finger speed to simulate simple texture or a solenoid thump whenever a the user passes over a an object on the screen. A haptic controller would be very simple to develop and could potentially greatly enrich the end user experience.

## 4 Conclusion

In this paper I have described MAT's multitouch table and outlined two methods in which it can be augmented to greater address the three of the four requirements of a Reality Based Interface: sense of naive physics, environmental awareness, and bodily awareness. A multi-touch surface interface is a relatively simple project but provides a powerful asset for multimedia application development and gesture research. Moving beyond the traditional WIMP interface means searching for modes of interaction that engage a user's sense of reality. Gestural computer interface is an important area of research for both commercial and industrial applications. By developing a system with several layers of gestural interface, researchers can search for intuitive analogues for different gestural modes of input and weight their effectiveness in accomplishing various user interaction tasks with the ultimate goal of developing a robust system that relies not on one technology, but has mappings for many different modes of gestural interaction. Additional research needs to be spent on how haptics enrich the computer interaction experience. A simple system of solenoids and linear motors would allow touch table developers to program a level of physicality into their multi-touch applications allowing the computer to respond to the users gestures as well as provide physical gestures back.

## References

1. Blob Detection. 5 June 2010, 23:48 UTC. In Wikipedia: The Free Encyclopedia. Wikimedia Foundation Inc. Encyclopedia on-line. Available from [http://en.wikipedia.org/wiki/Blob\\_detection](http://en.wikipedia.org/wiki/Blob_detection). Internet. Retrieved 1 June 2010.
2. Community Core Vision Wiki. [http://wiki.nuigroup.com/Main\\_Page](http://wiki.nuigroup.com/Main_Page) Accessed 29 August 2010
3. TUIO Protocol Specification. <http://www.tuio.org/> Accessed 29 August 2010
4. Santos, P., Stork, A., Buaes, A., and Jorge, J. 2006. Innovative geometric pose reconstruction for marker-based single camera tracking. In Proceedings of the 2006 ACM international Conference on Virtual Reality Continuum and Its Applications (Hong Kong, China). VRCIA '06. ACM, New York, NY, 237-244.
5. Kenji Oka, Yoichi Sato, Hideki Koike, "Real-Time Fingertip Tracking and Gesture Recognition," IEEE Computer Graphics and Applications, vol. 22, no. 6, pp. 64-71, Nov./Dec. 2002
6. Potamias, M. and Athitsos, V. 2008. Nearest neighbor search methods for handshape recognition. In Proceedings of the 1st international Conference on Pervasive Technologies Related To Assistive Environments (Athens, Greece, July 16 - 18, 2008). F. Makedon, L. Baillie, G. Pantziou, and I. Maglogiannis, Eds. PETRA '08, vol. 282. ACM, New York, NY, 1-8.
7. Robert Jacob, Audrey Girouard, Leanne Hirshfield, Michael Horn, Orit Shaer, Erin Solovey, Jamie Zigelbaum (2008). "Reality-Based Interaction: A Framework for Post-WIMP Interfaces". CHI '08: Proceedings of the Twenty-Sixth Annual SIGCHI Conference on Human Factors in Computing Systems. Florence, Italy: ACM. pp. 201-210.
8. iPhone Haptics Google Code Page. University of Glasgow. <http://code.google.com/p/iphone-haptics/> Accessed 29 August 2010
9. M.G. Jones, A. Bokinsky, T. Tretter, A. Negishi, A Comparison of Learning with Haptic and Visual Modalities Haptics-e Vol. 3, No. 6, May 3, 2005
10. Inteface. 7 June 2010, 23:48 UTC. In Wikipedia: The Free Encyclopedia. Wikimedia Foundation Inc. Encyclopedia on-line. Available from <http://en.wikipedia.org/wiki/Interface>. Internet. Retrieved 7 June 2010.
11. Build Your Own Multi-touch Surface Computer, April 7th, 2009, Maximum PC © 2010 Future US, Inc. Internet. retrieved 1 June 2010

- 
12. Total Internal Reflection. 10 May 2010, 23:48 UTC. In Wikipedia: The Free Encyclopedia. Wikimedia Foundation Inc. Encyclopedia on-line. Available from [http://en.wikipedia.org/wiki/Total\\_Internal\\_Reflection](http://en.wikipedia.org/wiki/Total_Internal_Reflection) Internet. Retrieved 3 June 2010.
  13. Sushmita Mitra and Tinku Acharya Gesture recognition: A survey (2007) IEEE TRANSACTIONS ON SYSTEMS, MAN AND CYBERNETICS - PART C vol. 37, No. 3, 2007, 311-324