BiDi Screen: A Thin, Depth-Sensing LCD for 3D Interaction using Light Fields

Matthew Hirsch¹  Douglas Lanman²  Henry Holtzman¹  Ramesh Raskar¹
¹MIT Media Lab  ²Brown University

Student Research Competition Grand Finals

Figure 1: 3D interaction with thin displays. We modify an LCD to allow co-located image capture and display. (Left) Mixed on-screen 2D multi-touch and off-screen 3D interactions. Virtual models are manipulated by the user’s hand movement. Touching a model brings it forward from the menu, or puts it away. Once selected, free-space gestures control model rotation and scale. (Middle) Multi-view imagery recorded in real-time using a mask displayed by the LCD. (Right, Top) Image refocused at the depth of the hand on the right; the other hand, which is closer to the screen, is defocused. (Right, Bottom) Real-time depth map, with near and far objects shaded green and blue, respectively.

Keywords: LCD, 3D interaction, light field, 3D reconstruction, depth from focus, image-based relighting, lensless imaging

Problem and Motivation

We transform an LCD into a display that supports both 2D multi-touch and unencumbered 3D gestures. Our BiDirectional (BiDi) screen, capable of both image capture and display, is inspired by emerging LCDs that use embedded optical sensors to detect multiple points of contact. Our key contribution is to exploit the spatial light modulation capability of LCDs to allow lensless imaging without interfering with display functionality. We switch between a display mode showing traditional graphics and a capture mode in which the backlight is disabled and the LCD displays a pinhole array or an equivalent tiled-broadband code. A large-format image sensor is placed slightly behind the liquid crystal layer. Together, the image sensor and LCD form a mask-based light field camera, capturing an array of images equivalent to that produced by a camera array spanning the display surface. The recovered multi-view orthographic imagery is used to passively estimate the depth of scene points. Two motivating applications are described: a hybrid touch plus gesture interaction and a light-gun mode for interacting with external light-emitting widgets. We show a working prototype that simulates the image sensor with a camera and diffuser, allowing interaction up to 50 cm in front of a modified 20.1 inch LCD.

One of the key benefits of our LCD-based design is that it transforms a liquid crystal spatial light modulator to allow both image capture and display in a thin package. Unlike many existing mask-based imaging devices, our system is capable of dynamically updating the mask. Promising directions of future work
include reconfiguring the mask based on the properties of the scene (e.g., locally optimizing the spatial vs. angular resolution trade-off). As higher-resolution sensors and LCD screens become available, our design should scale to provide photographic-quality images, enabling videoconferencing, gaze tracking, and foreground/background matting applications. Higher frame rates should allow flicker-free viewing and more accurate tracking. In order to achieve higher-resolution imagery for these applications, recent advances in light field superresolution could be applied to our orthographic multi-view imagery.

The use of dynamic lensless imaging systems in the consumer market is another potential direction of future work. The BiDi screen offers the unique ability to sense depth in a thin portable package, making it an attractive fit for interaction with mobile devices. The touch plus hover gesture interaction mode of the BiDi screen offers the ability to interact with small-screen devices without obscuring large portions of the display.

Another promising direction is to apply the BiDi screen to high-resolution photography of still objects by using translated pinhole arrays; however, such dynamic masks would be difficult to extend to moving scenes. The ability to track multiple points in free-space could allow identification and response to multiple users, although higher-resolution imagery would be required than currently produced by the prototype. Finally, the display could be used in a feedback loop with the capture mode to directly illuminate gesturing body parts or enhance the appearance of nearby objects [Cossairt:2008], as currently achieved by SecondLight [Izadi:2008].

Light-sensing displays are emerging as research prototypes and are poised to enter the market. As this transition occurs we hope to inspire the inclusion of some BiDi screen features in these devices. Many of the early prototypes discussed in the related work section enabled either only multi-touch or pure relighting applications. We believe our contribution of a potentially-thin device for multi-touch and 3D interaction is unique. For such interactions, it is not enough to have an embedded array of omnidirectional sensors; instead, by including an array of low-resolution cameras (e.g., through multi-view orthographic imagery in our design), the increased angular resolution directly facilitates unencumbered 3D interaction with thin displays.

Background

Design Goals

It is increasingly common for devices that have the ability to display images to also be able to capture them. In creating the BiDi screen we have four basic design goals:

- Capture 3D to enable depth- and lighting-aware interaction.
- Prevent image capture from interfering with image display.
- Support walk-up interaction (i.e., no implements or markers).
- Achieve these goals with a portable, thin form factor device.

Comparison of Design Alternatives
After considering related work and possible image capture options, we believe that the BiDi screen is uniquely positioned to satisfy our design goals. In this section we compare our approach to others.

**Capacitive, Resistive, or Acoustic Modalities:** A core design decision was to use optical sensing rather than capacitive, resistive, or acoustic modalities. While such technologies are effective for multi-touch, they cannot capture 3D gestures. Some capacitive solutions detect approaching fingers or hands, but cannot accurately determine their distance. Nor do these technologies support lighting-aware interaction. Optical sensing can be achieved in various ways. In many prior works, cameras image the space in front of the display. The result is typically a specially-crafted environment, similar to g-speak, where multiple cameras track special gloves with high contrast markers; or, the display housing is enlarged to accommodate the cameras, as with Microsoft's Surface.

**Cameras Behind, To the Side, or In Front of the Display:** Another issue is the trade-off between placing a small number of cameras at various points around the display. A camera behind the display interferes with backlighting, casting shadows and causing variations in the display brightness. Han's FTIR sensor, SecondLight, and DepthTouch all avoid this problem by using rear projection onto a diffuser, at the cost of increased display thickness. If the camera is located in front of the display or to the side, then it risks being occluded by users. Cameras placed in the bezel, looking sideways across the display, increase the display thickness and suffer from user self-occlusion. Furthermore, any design incorporating a small number of cameras cannot capture the incident light field, prohibiting certain relighting applications and requiring computationally-intensive multi-view stereo depth estimation, rather than relatively simple depth from focus analysis.

**Photodetector Arrays:** In contrast, our approach uses an array of photodetectors located behind the LCD (Figure 2). This configuration will not obscure the backlight and any attenuation will be evenly-distributed. Being behind the display, it does not suffer from user self-occlusion. The detector layer can be extremely thin and optically transparent (using thin film manufacturing), supporting our goal of portability. These are all design attributes we share with multi-touch displays being contemplated by Sharp and Planar. However, we emphasize that our display additionally requires a small gap between the spatial light modulating and light detecting planes. This critical gap allows measuring the angle of incident light, as well as its intensity, and thereby the capture of 3D data.

**Camera Arrays:** Briefly, we note that a dense camera array placed behind an LCD is equivalent to our approach. However, such tiled cameras must be synchronized and assembled, increasing the engineering complexity compared to the bare sensor in a BiDi screen. In addition, the sensors and lenses required by each camera introduce backlight non-uniformity. Designs incorporating dense camera arrays must confront similar challenges as the BiDi screen, including light absorption (by various LCD layers) and image flicker (due to switching between display and capture frames).

**Related Work**

**Multi-touch and 3D Interaction**

Sharp and Planar have demonstrated LCDs with integrated optical sensors co-located at each pixel for inexpensive multi-touch interaction. The Frustrated Total Internal Reflection (FTIR) multi-touch wall [Han:2005], TouchLight [Wilson:2004], Microsoft Surface, Oblong Industries g-speak, Visual Touchpad

Recent systems image directly through a display surface. [Izadi:2008] introduce SecondLight as a rear-projection display with an electronically-switchable diffuser. In their design, off-screen gestures are imaged by one or more cameras when the diffuser is in the clear state. While supporting high-resolution image capture, SecondLight significantly increases the thickness of the display, placing several projectors and cameras far behind the diffuser. Similarly, DepthTouch[Benko:2009] places a depth-sensing camera behind a rear-projection screen. While producing inferior image quality, the BiDi screen has several unique benefits and limitations with respect to such direct-imaging designs. Foremost, with a suitable large-format sensor, the proposed design might eliminate the added thickness in current projection-vision systems, at the cost of decreased image quality.

Sending Depth

A wide variety of passive and active techniques are available to estimate scene depth in real-time. Our prototype records an incident light field using attenuating patterns equivalent to a pinhole array. A key benefit is that the image is formed without refractive optics. Similar lensless systems with coded apertures are used in astronomical and medical imaging to capture X-rays and gamma rays. [Zomet and Nayar:2006] describe a system composed of a bare sensor and several attenuating layers, including a single LCD. [Liang:2008] uses temporally-multiplexed attenuation patterns, also displayed with an LCD, to capture light fields. [Zhang:2005] recover a light field by translating a bare sensor. [Levin:2007] and [Farid:1997] use coded apertures to estimate intensity and depth from defocused images. [Vaish:2006] discuss related methods for depth estimation from light fields. In a closely-related work, [Lanman:2008] demonstrate a large-format lensless light field camera using a family of attenuation patterns, including pinhole arrays, conceptually similar to the heterodyne camera of [Veeraraghavan:2007]. We use the tiled-broadband codes from those works to reduce the exposure time in our system. Unlike these systems, our design exploits a mask implemented with a modified LCD panel. In addition, we use reflected light with uncontrolled illumination.

Lighting-Sensitive Displays

Lighting-sensitive displays have emerged in the market in recent years; most portable electronics, including laptops and mobile phones, use ambient light sensors to adjust the brightness of the display depending on the lighting environment. [Nayar:2004] proposes creating lighting-sensitive displays by placing optical sensors within the display bezel and altering the rendered imagery to accurately reflect ambient lighting conditions. [Cossairt:2008] implement a light field transfer system, capable of co-located capture and display, to facilitate real-time relighting of synthetic and real-world scenes. [Fuchs:2008] achieve a passive lighting-sensitive display capable of relighting pre-rendered scenes printed on static masks. Unlike their design, our system works with directional light sources located in front of the display surface and can support relighting of dynamic computer-generated scenes.

Contributions
We demonstrate that a BiDi screen can recognize on-screen as well as off-screen gestures. We also demonstrate its ability to detect light-emitting widgets, showing novel interactions between displayed images and external lighting.

**Future Designs**

The emphasis of this paper is on demonstrating novel techniques for optical sensing enabled when an LCD and diffuse light-sensing grid are placed proximate to each other. As such devices are currently being developed for commercial deployment, one goal is to influence the design of these displays by exploring design choices and illustrating additional benefits and applications that can be derived.

**Thin, Depth-Sensing LCDs**

Earlier light-sensing displays focused on achieving touch interfaces. Our design advances the field by supporting both on-screen 2D multi-touch and off-screen, unencumbered 3D gestures. Our key contribution is that the LCD is put to double duty; it alternates between its traditional role in forming the displayed image and a new role in acting as an optical mask. We show that achieving depth- and lighting-aware interactions requires a small displacement between the sensing plane and the display plane. Furthermore, we maximize the display and capture frame rates using optimally light-efficient mask patterns.

**Lensless Light Field Capture**

We describe a thin, lensless light field camera composed of an optical sensor array and a spatial light modulator. We evaluate the performance of pinhole arrays and tiled-broadband masks for light field capture from primarily reflective, rather than transmissive, scenes. We describe key design issues, including: mask selection, spatio-angular resolution trade-offs, and the critical importance of angle-limiting materials.

**Unencumbered 3D Interaction**

We show novel interaction scenarios using a BiDi screen to recognize on- and off-screen gestures. We also demonstrate detection of light-emitting widgets, showing novel interactions between displayed images and external lighting.

**Dynamic**

Because the mask, whether composed of pinholes or a tiled-broadband code, is formed on an LCD, we can dynamically vary the size and density of such periodic patterns.

**Benefits and Limitations**

The BiDi screen has several benefits over related techniques for imaging the space in front of a display. Chief among them is the ability to capture multiple orthographic images, with a potentially thin device, without blocking the backlight or portions of the display. Besides enabling lighting direction and depth measurements, these multi-view images support the creation of a true mirror, where the subject gazes into her own eyes, or a videoconferencing application in which the participants have direct eye contact [Rosenthal 1947]. At present, however, the limited resolution of the prototype does not produce imagery competitive with consumer webcams.

The BiDi screen requires separating the light-modulating and light-sensing layers, complicating the display design. In our prototype an additional 2.5 cm was added to the display thickness to allow the placement of the diffuser. In the future a large-format sensor could be accommodated within this distance, however the current prototype uses a pair of cameras placed about 0.5 m behind the diffuser - significantly increasing the device dimensions. Also, as the LCD is switched between display and capture modes, the proposed design will reduce the native frame rate. Image flicker will result unless the display frame rate remains above the flicker fusion threshold [Izadi 2008]. Lastly, the BiDi screen requires external illumination, either from the room or a light-emitting widget, in order for its capture mode to function. Such external illumination
results the displayed image contrast. This effect may be mitigated by applying an anti-reflective coating to the surface of the screen.

Results

Designing a BiDi Screen

Figure 2 Image capture and display can be achieved by rearranging the optical components within an LCD. A liquid crystal spatial light modulator is used to display a mask (either a pinhole array or equivalent tiled-broadband code). (Left) The modulated light is captured on a sensor array for decoding. (Right) With no large-area sensor available, a camera images a diffuser to simulate the sensor array. In both cases, LEDs restore the backlight function.

Our BiDi screen is formed by repurposing typical LCD components such that image capture is achieved without hindering display functionality. We begin by excluding certain non-essential layers, including the CCFL/light guide/reflector components, the various brightness enhancing films, and the final diffuser between the LCD and the user.

In a manner similar to [Lanman 2008], we then create a large-aperture, multi-view image capture device by using the spatial light modulator to display a pinhole array or tiled-broadband mask. Our key insight is that, for simultaneous image capture and display using an LCD, the remaining backlight diffuser must be moved.
away from the liquid crystal. In doing so, a coded image equivalent to an array of pinhole images is formed on the diffuser, which can be photographed by one or more cameras placed behind the diffuser. The backlight display functionality is restored by including an additional array of LEDs behind the diffuser.

**Interaction Modes**

**Figure 4** Additional interaction modes. (Left) A virtual world navigated by tracking a user’s hand. Moving the hand left, right, up, and down changes the avatar’s heading. Reaching towards or away from the screen moves. The layers of the prototype are shown in the circled inset, including from left to right: the decorative cover, LCD (in a wooden frame), and diffuser. (Middle) A relighting application controlled with a real flashlight. The flashlight is tracked using the captured light field. A similar virtual light is created, as if the real flashlight was shining into the virtual world. (Right) A pair of cameras and multiple LEDs placed 1 m behind the diffuser.

**Multi-Touch and 3D Interaction:** The BiDi screen supports on-screen multi-touch and off-screen gestures by providing a real-time depth map, allowing 3D tracking of objects in front of the display. As shown in Figure 1, a **model viewer** application is controlled using the 3D position of a user’s hand. Several models are presented along the top of the screen. When the user touches a model it is brought to the center of the display. Once selected, the model is manipulated with touch-free “hover” gestures. The model can be rotated along two axes by moving the hand left to right and up and down. Scaling is controlled by the distance between the hand and the screen. Touching the model again puts it away. As shown in Figure 4, a **world navigator** application controls an avatar in a virtual environment. Moving the hand left and right turns, whereas moving the hand up and down changes gaze. Reaching towards or away from the screen affects movement. As shown in the supplementary material, more than one hand can be tracked, allowing multi-handed gestures as well.
Lighting-Sensitive Interaction: Another interaction mode involves altering the light striking the screen. A model lighting application allows interactive relighting of virtual scenes (Figure 4). In this interaction mode, the user translates a flashlight in front of the display. For a narrow beam, a single pinhole (or MURA tile) is illuminated. Below this region, a subset of light sensors is activated. The position of the pinhole, in combination with the position of the illuminated sensors, determines the direction along which light entered the screen. A similar light source is then created to illuminate the simulated scene - as if the viewer was shining light directly into the virtual world.

Project Status

The BiDi Screen was first presented at SIGGRAPH 2009 as a short talk and poster. Though our technical paper was rejected from SIGGRAPH 2009, it was ultimately accepted to SIGGRAPH Asia 2009 after further results were demonstrated.

We were thrilled to achieve second place in the Student Research Competition for SIGGRAPH 2009 based on our poster presentation. The positive feedback we received from the SRC award inspired us to push forward with a second, more portable, prototype. This prototype was submitted and accepted to SIGGRAPH Asia 2009 as an Emerging Technology demonstration. Through the additional talk and exposure provided by the SRC award at SIGGRAPH 2009, we were able to identify the importance of the gestural interface aspect of the project, and subsequently focused primarily on this aspect during the ETech demo.

Media Coverage: Our Emerging Technology demo received significant media attention in December 2009, which we feel is a good demonstration of the potential impact of our work on the public. A summary of the media coverage is shown here:

- MIT Press
- Wired
- New Scientist
- EE Times
- Mumbai Mirror
- Engadget
- Hackaday
- Gizmodo
- Hardocp
- ZDNet
- Venture Beat
- Information Week
- Slashdot

Inspiring Future Research Questions

The BiDi screen project suggests many diverse research directions. Advances in coupling optical sensor hardware with displays will be necessary to commercialize the technology. The search for efficient methods of producing large area sensors will lead to novel new hardware structures with applications beyond interactive displays. At the other end of the spectrum, the availability of a portable low-cost gestural interface will give user experience researchers the opportunity to craft new and better interfaces for all manner of devices. The marriage of image capture and display will provide researchers with the tools to build the next generation of communication and entertainment devices.