Projector-Based Location Discovery and Tracking

Johnny Chung Lee
May 2008
CMU-HCII-08-102

Human Computer Interaction Institute
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

Thesis Committee:
Scott E. Hudson (Chair), Carnegie Mellon University
Jodi Forlizzi, Carnegie Mellon University
Luis Von Ahn, Carnegie Mellon University
Paul Dietz, Mitsubishi Electric Research Labs

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy

This research was supported in part by the National Science Foundation under grants CCR-03244770, IIS-0121560, IIS-0205219, IIS-0325351, IIS-0329090, IIS-9800597, IIS-9980013, Mitsubishi Electric Research Labs, Intel Research Council and General Motors.

Keywords: Image projection, projector calibration, augmented reality, location tracking, motion capture, high-speed projection, infrared projection, multi-projector applications, projector alignment, incremental tracking.

Abstract

Today, the primary use of projection technology is for creating large flat displays that provide a shared viewing experience for presentations or entertainment applications. While research projects have explored the powerful ability for projected light to create illusions that can reshape our perception and our interaction with surfaces in the environment, very few of these systems have had success in terms of commercial and consumer adoption. Part of this limited adoption can be attributed to the lack of practicality in the cost-of-operation due to the complexity of installation and reliability of execution. Often these systems require expert knowledge to perform system setup and calibration between the projected image and the physical surfaces to make these illusions effective. In this thesis, I present a technique for inherently adding object location discovery and tracking capabilities to commercial projectors. This is accomplished by introducing light sensors into the projection area and then spatially encoding the image area using a series of structured light patterns. This delivers a unique pattern of light to every pixel in the projector's screen space directly encoding the location data using the projector itself.

By unifying the image projection and location tracking technologies, many of the difficult calibration and alignment issues related to interactive projection and projected spatial augmented reality applications can be eliminated simplifying their implementation and execution. Furthermore, by creating a hybrid visible light and infrared light projector, a single calibration-free device can perform invisible location tracking of input devices while simultaneously presenting visible application content. I present a detailed description of the projector-based location discovery and tracking technique, a description of three prototype

implementations, and a demonstration of the effectiveness of this simplification by re-implementing, and in some cases improving upon, several location-sensitive projector applications that have been previously executed using external calibration and tracking technologies.

Acknowledgements

The primary influences in this work most certainly are my thesis advisor, Scott Hudson, and research mentor, Paul Dietz.

Scott provided the research guidance and outlook that has taken me from a naïve new graduate student to the researcher I have become today. When I began the doctoral program at Carnegie Mellon University, I had little knowledge and appreciation of what research should be and a lack of confidence that I could make a meaningful contribution to the field. Scott has been integral in helping me gain perspective, to select problems of substance, and taught me the approach to develop solutions that are practical. These two skills have become essential in defining my path of work as a researcher.

Paul provided me with an exposure to a sense of technical creativity and personal enjoyment in invention that I have not found in any other individual. As mentor, he has provided me with a perspective on life and the impact that work, research, and creativity can have on the world outside the typical academic boundaries. For this, I will be eternally grateful

In addition to these two individuals, I have had the pleasure to work with and be in the company of brilliant individuals that have inspired me on both professional and personal levels. This includes my co-authors Ramesh Raskar, Desney Tan, Dan Maynes-Aminzade, Jay Summit, Chris Kyriakakis, and Darren Leigh. My classmates, colleagues, and friends James Fogarty, Daniel Avrahami, Darren Gergle, Luis Von Ahn, and Laura Dabbish have been an incomparable set of individuals to have begun my research career with. I felt sad to see them graduate and move on, but I now look forward to joining them "on the other side".

My brother, David Lee, was the one who first got me interested in Human-Computer Interaction when he was taking classes at University of Virginia with

Randy Pausch. Once I entered college, my first exposure to research was with Denny Proffitt who gave a young undergrad an unusual opportunity to participate in a large research project with Carnegie Mellon. It was through this work, that I was able to develop a research relationship with Randy Pausch and Jodi Forlizzi. Without these individuals, I most certainly would not be where I am today.

Thank you.

Table of Contents

Index of Figures

1: Introduction

The earliest known drawing capturing the idea of projecting a drawn image onto a wall was created by a man named Johannes de Fontana in 1420. The sketch was of a monk holding a lantern behind a translucent drawing of the devil. The light from the lantern passing through the translucent window would have created a rough replication of the drawing on the wall. Though crude, this drawing became the basis and inspiration for countless inventors over the following centuries involved with development of optics, light sources, and image rendering technologies that have shaped the modern video projector. Today, projectors have become part of our daily lives in classrooms, business presentations, movie theaters, and consumer televisions. Projectors have become a staple display technology in the contemporary world of visual media and computing.

Today, the most common use of projection technology is to create physically large displays on flat surfaces providing a shared viewing experience that can accommodate a large number of simultaneous observers. This usage makes it ideal for giving presentations or entertaining an audience. However, using projectors to create flat passive displays similar to other display technologies such as liquid-crystal displays (LCD), and plasma displays is perhaps the simplest use of projection technology. A unique property of projection is that the desired image is not visible by looking at the device itself but rather by looking at the light that is reflected and diffused off of a surface of our choice. Unlike LCD or plasma technologies whose display size and shape are rigidly determined by the manufacturer, projectors are capable of rendering images on a wide range of surfaces as selected by the end user and place very few constraints in terms of display surface location, material, shape, or size.

However, with this versatility comes the responsibility of positioning and orienting the projector in a manner that is appropriate for the application. Modern projectors have little or no knowledge of their physical relationship to surfaces in the environment. While certain assumptions have been made in projector designs to make them slightly easier to use in the most common applications, projectors are still quite naïve and the vast majority of the responsibility for proper set up falls upon the shoulders of the end user. Even once this process is complete, the resulting image is passive. There is no way to interact directly with the projected image. Input must be accomplished using some other device such as a mouse, keyboard, or game controller which provide relative input and control data. When using touch sensitive surfaces, the projection system must be given some knowledge about the location and orientation of the projected image relative to the sensing surface. Furthermore, researchers have developed a host of visionary applications exploring how projected light can be used to create illusions that reshape and alter our perception of surfaces in the environment that reach well beyond simply creating a large display. However, like touch input, these applications require knowledge of surface locations relative to the projected image. Often, this information is manually entered into the system by a knowledgeable user.

Research Summary and Components

This dissertation presents a technique that endows projectors with the inherent ability to discover the location of photosensitive objects within the image area. By doing this, we can greatly simplify the implementation and execution of these location-sensitive projector applications. This is accomplished using a technique developed in partnership with my colleagues at Mitsubishi Electric Research Labs (MERL). There are four major components of this research work presented in this dissertation:

- First, a novel method of performing projector-based location discovery using embedded light sensors is developed.
- Second, a prototype exploring techniques for increasing the speed and reducing the perceptibility of patterns is developed.
- Third, a prototype for a hybrid visible and infrared light projector capable of simultaneously providing visible application content and invisible location tracking is developed.
- Fourth, a series of conceptual applications demonstrating the possibilities provided by a high-speed projector are presented.

Before describing the details of this work, I would first like to present to the reader examples of related work in projector applications from the human-computer interaction and computer graphics communities. This description of visionary applications defines the landscape in which this work applies and will hopefully illustrate the value in simplifying the implementation and accessibility of such projects. The relevance of these applications to my specific contributions will be highlighted throughout the paper as the details of this work are described.

Location-Sensitive Projector Applications

As mentioned previously, one basic application requiring knowledge about the location of the projected image is for direct, or *in situ,* interaction with projected content. Products from Mimio (mimio.com), SMART Technologies (smarthtech.com), and Polyvision (polyvision.com) are examples of after-market devices that can be added to a projection system to track the location of an input stylus and map the input to pixel locations in a calibrated projected image. These products require a tracking technology to discover and update the location of either the user's finger or an instrumented stylus. While the actual tracking technology varies, they all require a manual calibration process where users are asked to tap a series of marks to find the correspondence between the tracking data and pixels in the projected image. Devices like the Diamond Touch table from Mitsubishi Electric Research Labs [Dietz, 2001], Smart Skin from Sony CSL [Rekimoto 2002], the Frustrated Total Internal Reflection Multi-Touch system from New York University [Han, 2005], the TouchLight [Wilson, 2004] system, PlayAnywhere [Wilson, 2005], and the Planar Manipulator Display [Rosenfeld 2004] use a variety of techniques to track un-instrumented human hands supporting multiple simultaneous touches, multiple simultaneous users, and area touching creating extremely rich touch input surfaces. However, once again, these systems require a manual calibration using a similar set of controlled sample touches to align the tracking data with the image. Additionally, the physical relationship between the projector and the interactive surface must remain rigid for the illusion to be compelling. This level of sensitivity toward alignment and calibration often results in significant increases in system cost and operation complexity. The technique presented in this dissertation offers a simplification that can dramatically reduce this cost and overhead.

Because projection technology does not share many of the physical constraints on display size, shape, or material expressed by other display technologies such as cathode ray tubes, liquid crystal displays, or plasma screens, researchers have found it to be an attractive option for exploring applications that reach well beyond 2-dimensional touch input. Tangible Bits [Ishii, 1997] and Augmented Surfaces [Rekimoto, 1999] are examples of digital workbench applications that create extremely rich table-top interaction experiences by using instrumented, location-tracked objects, placed in the image area to interact with projected content. The appearance of these objects can still be augmented with projected content even though they have been placed on top of the primary display surface. The Illuminated Clay project [Piper, 2002] takes this concept even further by allowing users to sculpt and shape the display surface with their hands. This is accomplished by using deformable materials such as clay or sand. In this system, a high-speed 3-dimensional laser scanner acquires the updated geometry and the projected overlay responds accordingly allowing the users to see contour changes or run simulations based on the shape of the surface. However, these systems also relied on either manual physical alignment of the projector to the tracking system or used a structured collection of sample points, similar to touch calibration, to perform software-based alignment. The technique presented in this dissertation offers that ability to perform automatic calibration and alignment.

By leveraging the ability to easily create physically large displays, researchers have used projectors to change the appearance of entire walls in a room transforming the environment to better suit the needs of a task as described by [Raskar, 1998]. The VideoWindow system [Bellcore, 1989] is a wall-sized display system that is designed to give the appearance that a room extends into another space creating the illusion of a shared-room telecollaboration experience. Teleport system [Gibbs, 1998] is a similar technology but provides motion parallax cues to improve the illusion. The Cave Automated Virtual Environment (CAVE) [Cruz-Neira, 1993] is an extension of this concept to include every wall as well as the floor and ceiling to provide a fully immersive experience of being in a virtual environment. The Focus+Context display [Baudisch, 2001] takes a different approach and combines a physically large projected image with a strategically placed high-resolution display creating a variable resolution surface. This provides support for a large immersive experience while also supporting a high level of detail in a small work area. However, for these illusions to work, both of these systems require precise alignment of the projected images and knowledge about the location of the viewer's point of view. Again, these have relied on accurate manual physical positioning of the projector and expertise in the system software to ensure all the components are aligned to create a compelling effect.

Unlike other display technologies, projection does not impose physical borders or boundaries around the visible image. This has made it an attractive candidate for creating tiled and overlapped displays. [Li, 2000] describes a system that utilizes large arrays of projectors (8-24) in a tiled configuration to create a single high-resolution display wall. Accomplishing this requires extremely precise alignment among the projected images such that, when blended together at the edges, they merge into a single high-quality seamless display. Often, building high-resolution displays from many lower resolution commercially available projectors is a far more economical solution than creating a single high-resolution custom projector, particularly if the surfaces are irregular or curved [van Baar, 2003; Raskar, 2003]. Some multi-projector applications do not stitch the images to create a larger display, but instead overlap the images entirely to create layered content. Examples of such systems include polarized stereoscopic projected displays, increasing the brightness of the image, light field display, and dynamic shadow elimination [Sukthankar, 2001]. Dynamic shadow elimination, sometimes referred to as virtual rear-projection [Summet, 2003], simulates the shadow-less appearance of a rear-projected display by using two or more front-projected images and computer vision techniques to identify the location of a user's shadow. The system then fills in the shadows created by one projector with the light from another projector at an un-occluded vantage point. The 3D TV project [Matusik, 2004] uses a large array of projectors aligned to the same area on a lenticular screen to create a large-scale multi-user auto-stereoscopic display by approximating the light field passing through that surface. The MultiView system [Nguyen, 2007] uses overlapped projection and retro-reflective film to provide a perspective correct view for each person in a video conferencing system preserving accurate eye contact and gaze which was shown to be important for establishing trust in social collaboration tasks. Similar to the previous projector applications, aligning multi-projector displays is often an exclusively manual and physical process. One previous approach to this problem in this domain has been to use computer vision feedback mechanisms capable of achieving high-quality results without human involvement [Wallace, 2005; Raskar 1999]. However, these algorithms tend to be highly-specialized and as a result are typically application and system specific.

Another property that is unique to projection technology is that the optical path of the projected light can be folded and re-directed through the use of mirrors. The Everywhere Displays project [Pinhanez, 2001] uses a computer controlled pan-tilt mirror placed in front of the projection lens allowing the image to be placed on many surfaces throughout the environment rather than simply being confined to a static area. Using knowledge about the geometry of surfaces in the room relative to the projector, these steerable projection systems can transform registered surfaces into displays providing location-sensitive information. However, registering surfaces requires system expertise to manually enter room geometry.

The Shader Lamps work [Raskar, 2001] takes a step away from the domain of flat displays and explores how spatially modulated light can be used to alter the appearance of complex 3-dimensional surfaces. By using the projector as a sophisticated illumination tool combined with detailed knowledge of the location, geometry, and reflective properties of the object, it is possible to change the apparent surface colors and material properties. This technique can be combined with a six degree of freedom tracker to create the illusion of painting on handheld objects using only projected light [Badyopadhyay, 2001]. Though the object geometry must be manually entered or acquired using a scanning device, it does not need to be updated so long as the object is not deformed. However, the registration of the projector pose must be carefully re-entered when either the projector or the object is moved. Similarly, this was also done manually using a projected cursor to enter calibration points and required detailed system expertise.

These products and research projects demonstrate the how projection technology can be used to create rich interactive experiences and illusions that reach far beyond simply projecting a physically large display. This is the power gained from endowing the projection system with knowledge about its orientation and the image location relative to the display surface or the location of objects placed within the projection area. To achieve these effects, all of these systems require a calibration process where the correspondence between the features of interest in the physical world and the projector's screen space is entered. Without an accurate registration, these illusions will not be compelling or effective.

In nearly all of systems described above, this correspondence was accomplished through a manual calibration process typically in the form of asking the user or system developer to register the location of projected markers relative to some other tracking technology. In many cases, this process can be extremely lengthy, tedious, error prone, and requires expert knowledge of the system. Additionally, re-calibration may be necessary if even very small changes occur to the projection setup. As a result, the practicality of these projector applications is severely undermined causing adoption to remain low despite their visionary interaction techniques and obvious utility.

In this dissertation, I present a technique of using the projector itself to perform location discovery and tracking of photosensitive objects without the need for an external tracking technology. By unifying the image projection and location tracking, we are able to eliminate the need to obtain the correspondence between the tracking data and the projection image. This significantly simplifies the implementation and execution of many of the commercial products and research projects described above. In the following *Background and Approach* section, I will introduce the foundation concepts that this technique is built upon and discuss its relationship to other current approaches to the same problem. I will discuss the first prototype implementation that uses an unmodified consumer projector to perform low-speed location discovery and demonstrate the projector applications for which it is useful. Then, I will present two additional implementations that I developed that represent different approaches toward achieving rapid location discovery sufficient for interactive tracking and the corresponding applications that they simplify. I also briefly discuss a system developed by my collaborators which optimizes the same concept to achieve very high-speed location tracking. However, this ability comes at expense of image projection transforming it into a purely motion tracking system which places it slightly outside the goals of this work. Lastly, I present a series of concept applications that would be enabled by a projector with integrated interactive tracking.

2: Background and Approach

A modern multimedia projector can be described as an electrically addressable spatial light modulator whose working volume is defined by the frustum of light emanating from the projection lens. The modulation of the light within this volume occurs in a plane that is perpendicular to the optical axis of the projector. This plane is divided into a grid of discrete regions, called pixels, which the projector is able to independently vary in intensity and color. If a flat diffuse surface is intersected with this volume parallel to the grid plane, we get an image. However, the modulation of pixels does not necessarily create an image that is coherent to a human observer. In the early 1980's researchers in the range finding community began exploring how projecting highly structured patterns onto non-planar surfaces combined with a camera can be used to quickly acquire scene geometry. These structured light patterns, typically regular grid or stripe patterns, uniquely encode position information for each pixel in the projection image. Decoding these patterns as seen by the camera provides the correspondence between the projector pixels and camera pixels. Combined with knowledge about the camera pose relative to the projector, this pixel correspondence map becomes an index of ray intersection pairs, and the geometry of the surface can then be calculated using triangulation [Posdamer, 1982; Depiero, 1996]. By synchronizing high-speed projection with a high frame rate camera, researchers have demonstrated the potential to create a real-time 3D scanning system [Raskar, 1998].

While the ability to perform 3-dimensional scene capture is not essential to most of the projector applications described in the previous section, the ability to find the correspondence between the projector pixels and objects in the physical world to an external location discovery/tracking technology is essential. As a result, we are able to use the same structured light patterns to encode each pixel to quickly identify the location of objects in terms of the projector's screen space. The approach presented in this dissertation does so without the use of a camera. This is accomplished by embedding individual light sensors at the locations of interest in the target surface. Each light sensor then detects the pattern of light it receives from the projector yielding the pixel location of that sensor, which is then reported back to the computer for use in an application. By embedding the photo-sensing device directly into the projection surface or target object, the location data needed for the applications described previously can now come from the projector itself rather than external location discovery/tracking system. As a result, it eliminates the need to discover the correspondence between the location data and the projection image since the projector-based location data inherently matches the pixel coordinates in the projected image. Projector-based location discovery and tracking simplifies the implementation of these systems by removing the need for an external tracking system and simplifies usage by removing the need for correspondence discovery.

Automatic Projector Calibration

Previous approaches to simplifying the construction and implementation of complex projector applications have predominantly relied on computer vision techniques. While there are a few instances of other approaches toward automatic projector calibration that do not use a camera, they are relatively limited in their capabilities.

For example a few commercial projectors include sensors to either detect the direction of acceleration due to gravity or detect the length of the support legs [Wood 2005]. These provide the projector with a measurement of pitch which corresponds to a certain degree of vertical keystone distortion in the projected image given a number of assumptions about the orientation of surfaces in the room. These sensors provide an automatic-keystone correction mechanism that is likely to produce an undistorted image in a typical business presentation environment. However if these assumptions are violated, then the automatic-keystone correction will be incorrect.

In more sophisticated location-sensitive projector applications, the vast majority of work on automatic calibration has focused on computer vision techniques. Since projected imagery is a visible medium, using a camera to provide feedback about the location of the image is a natural choice. In some systems such as [Rekimoto, 1999; Wilson, 2005], computer vision is already in use as the object tracking and recognition technology. Thus, using a vision-based automatic calibration technique would be an efficient use of existing resources. Other projects, such as [Wallace, 2005; Raskar, 1999], have elected to add a camera system solely for the purposes of calibration and projector alignment. This has been shown to be reasonably successful in certain applications, especially in multi-projector stitching applications where sub-pixel alignment accuracy can be achieved by carefully analyzing the aliasing effects of straight lines on a known surface or by utilizing a pan-tilt-zoom camera system [Bimber 2005]. Cameras also have certain geometric similarities to projection which make it an attractive pairing of technology. In theory, camera-based calibration also does not require that the display surface or objects to be augmented, potentially easing implementation. However, in practice, specialized tags [Rekimoto, 1999; Wilson, 2005] or light-emitting dots [Yotsukura, 2002] are necessary to reliably locate and track objects using vision techniques unless the scene is very simplistic [Raskar, 2001]. A number of systems which use computer vision for tracking require specialized near infrared (IR) illumination/reflectivity [Han, 2005; Wilson, 2005] to achieve good performance and IR filters to eliminate interference from visible projected content. To use the camera system for automatic calibration, these IR filters must be physically removed.

These adaptations highlight one of the largest weaknesses in computer vision-based calibration – feature recognition and background separation. While computer vision algorithms have advanced significantly over the past 40 years, the ability to segment and properly identify un-augmented objects still remains an open problem. Similarly, uncontrolled lighting conditions and surfaces of varying or unknown reflectance can be difficult for computer vision systems. This is one reason controlled IR illumination and uniform reflectance surfaces are frequently employed to achieve robust and accurate vision tracking. Even when conditions are ideal, the recognition algorithm must still be tweaked and adjusted for image distortion and object geometry [Wilson, 2005], which may vary greatly between individual system implementations or even among different objects within a single system. By using an embedded light sensor approach, points in a projection surface which may not be distinguishable by a camera can be tracked without issue. As will be described in more detail later, sensors can be embedded slightly beneath the top layer of a uniform surface to discover the locations of visually featureless points. Embedded light sensing uses only local illumination at each sensor and each sensor can adapt to its own lighting conditions independently.

Additionally, it would be possible to employ light sensors that use modulated light communication which eases the issue of signal segmentation even further.

Another disadvantage of camera-based calibration is the added computational and bandwidth costs. The data rate from a camera is equal to *number of pixels* × *bits per pixel* × *frames per second* and the data must typically be analyzed at the pixel level to perform accurate tracking and recognition. For example, a 640x480 resolution 8-bit camera capturing 30 frames per second generates over 70Mbits/sec of data and still must be processed by a high-speed PC to perform vision processing. That same data rate could support tracking over 120,000 light sensors simultaneously in the same 640x480 area at 30Hz. The data would be processed locally at each sensor using a low-cost microcontroller in a manner somewhat similar to distributed computation. For example, the camera-based automatic calibration system described in [Wallance, 2005] is able to color calibrate and align an impressive 24 projector system to create a single seamless display. However, the process requires nearly 10 minutes and utilizes a pan-tilt-zoom camera system that can focus on various parts of the screen. In contrast, a grid of color sensitive light sensors could achieve similar results in seconds.

Camera-based tracking also has difficulty in scalability with respect to the number of objects that can be simultaneously tracked. Tracking large numbers of objects becomes difficult as object density approaches the limits of the camera resolution and ambiguity of object identity also increases. Projector-based tracking supports an unlimited number of light sensors, whose location and identity are discovered in constant time. The time necessary to broadcast location data from the projector is a function only of the resolution of the projector and is independent of the number of sensors being tracked. If the projection area is very large, more than one light sensor may reside within a single pixel without issue. Additionally, identity is inherent to the sensor that collected the data. As described earlier, the bandwidth requirements for transmitting sensor location back to the host computer are minimal. However, if the bandwidth is unavailable, it can be compressed or processed locally at the sensor location.

While computer vision approaches to automatic projector calibration and alignment have certain advantages, it typically comes at the cost of increased complexity and algorithmic challenges. By placing light sensors in the projection area, many of the difficulties related to computer vision are avoided entirely. Additionally, the correspondence between tracking data and the projected image is inherent, further simplifying system implementations. Further discussion on performance scalability differences between camera-based tracking and projector-based tracking can be found in *Comparison to Computer Vision Approaches*.

Structured Light

The structured light patterns used in this work are a time-multiplexed binary Gray-coded stripe pattern introduced to the range finding community in 1984 [Inokuchi, 1984]. However, the history of Gray code sequences reaches several centuries back. The naming credit belongs to Frank Gray, a researcher at Bell Labs, who was granted a patent in 1953 [Gray, 1953] for the application of the sequences in communications. However, these codes were mentioned as early as 1550 by a mathematician named Cardan as the solution to a puzzle called the *Chinese Ring Puzzle*. The puzzle itself dates back to the 2nd century AD [Gardner, 1986]. The Gray-code patterns are a variation of the basic binary patterns used by [Posdamer, 1982], which is a sequence of black and white patterns that progressively divide the projection area along one axis into smaller and smaller regions using binary division. To resolve both x and y coordinates, this sequence is run twice, once horizontally and once vertically. The Gray-coded variation improves upon this by ensuring that the stripe boundaries never occur in the same location ensuring that the Hamming distance between two adjacent regions is only one providing signal stability in spatial encoding applications. This prevents the catastrophic decoding error that might occur in typical binary division if a photosensor were to straddle a division boundary of a high order bit. The end result of this property is that Gray-coded patterns limit the error from boundary events to +/- 1 pixel. It also increases robustness against noise and defocusing of the image, an important aspect when projecting onto surfaces that are oblique to the focus plane. These binary patterns also have an $O(\log_2(n))$ relationship between the necessary number of patterns to uniquely encode each pixel and the number of pixels, n. Specifically, due to their axis aligned nature, the number of patterns necessary is $\lceil \log_2(width) \rceil + \lceil \log_2(height) \rceil$. Every pixel in an XGA resolution projector (1024x768) can be uniquely identified with only 20 binary patterns. Further, this approach scales nicely for future technologies with vastly higher resolutions. For example, only 60 binary images would be necessary to resolve the entire continental United States to millimeter accuracy. To resolve each pixel in a 32x32 pixel area requires 10 patterns. The Gray-coded binary pattern sequence can be seen in Figure 1.

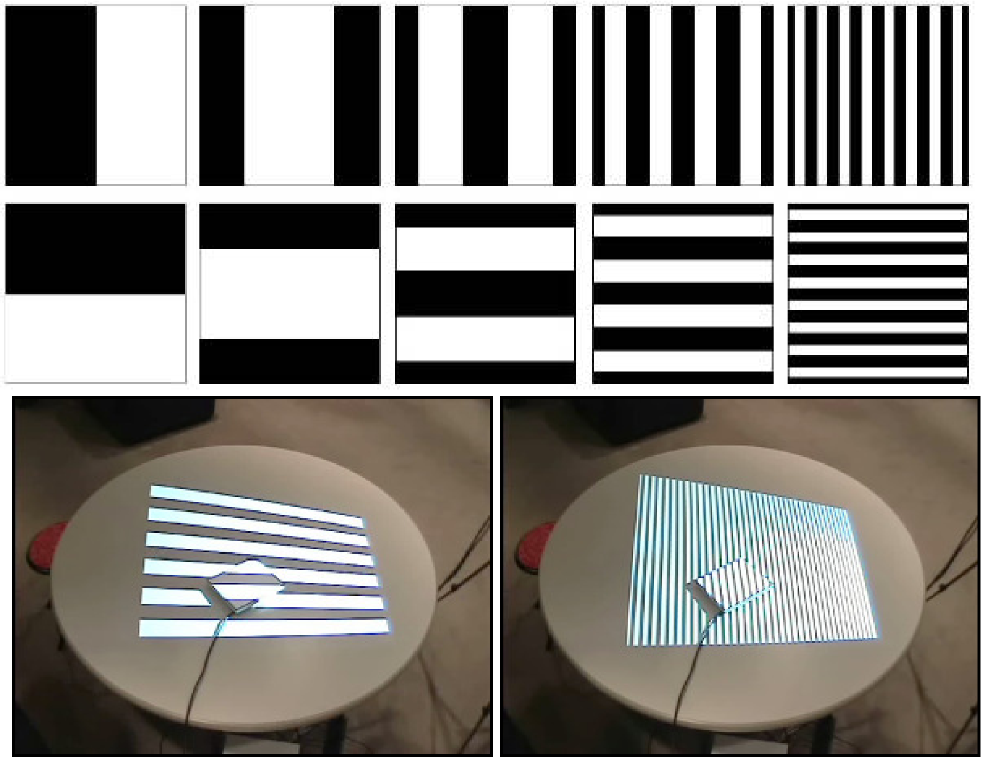


Figure 1. This set of horizontal and vertical Gray-coded binary patterns (top) are sequentially projected (bottom) to discover the x and y pixel coordinates of each sensor.

While a number of other structured light patterns have been created [Salvi, 2004], these variations in coding strategies have largely been to optimize performance in the range finding domain when utilizing a camera. Thus, these variations employ spatial neighborhood and color sensitivity not typically available in point sample photo-sensors or to minimize artifacts resulting from surface folding. As a result, many of these patterns are not applicable to this approach. Of the applicable alternatives reviewed, there was no pattern variation that provided an obvious improvement over a binary Gray-coded pattern when using a discrete pixel projector. Even *n*-ary Gray-codes which use grey levels or colors to shorten the sequence length by increasing the number of bits per image do so at the cost of decreasing the ease of correctly segmenting the coded patterns.

Binary patterns provide the highest level of robustness against signal interference and support a wider variety of transmission strategies, which will be discussed later.

In [Sugimoto 2005], researchers explored projecting structured light patterns to transmit arbitrary data to photosensitive objects within the projection area. This context of use was to provide remote control commands to toy vehicles, but the concept could be expanded to allow visible light communication to a variety of objects within the projection area. This is somewhat similar to a visible light version of [Nii 2005]. However, arbitrary data transmission has difficulties in scaling up in region density due to data corruption resulting from sensors landing on discrete region boundaries.

3: Projector-Based Location Discovery

In the first prototype implementation of projector-based location discovery [Lee, 2004], we explored applications that can be implemented using low-speed or one-time location discovery using an unmodified commercial projector. The refresh rate of most commercial projectors is 60Hz, or 60 distinct images per second. When using binary patterns, this translates to 60 bits of location data per second per pixel. While projector resolutions vary, a common image resolution is 1024 pixels wide by 768 pixels tall sometimes referred to as XGA (eXtended Graphics Array) resolution. Since the number patterns necessary to uniquely encode each pixel using binary patterns is $\lceil \log_2(width) \rceil + \lceil \log_2(height) \rceil$, the number of patterns required is 20, 10 horizontal patterns and 10 vertical patterns, resulting in a minimum location discovery time of approximately 333ms, or 1/3rd of a second, with a 60Hz projector. In practice, we prefix the location encoding sequence with an all white and an all black pattern to provide an asynchronous start bit and allow each sensor to acquire an appropriate threshold level for decoding the following

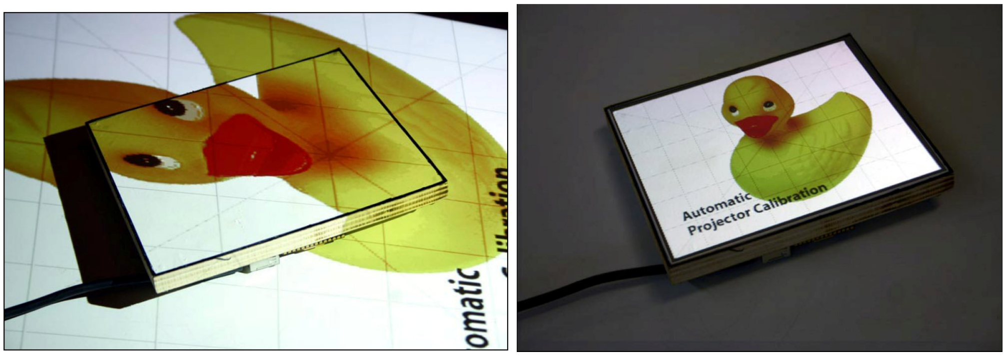


Figure 2. Screen calibration application for project-based location discovery

location bits. We also use a following stop bit allowing dynamic bit timing rather than fixating on 60Hz. When lighting situations are difficult, gathering more samples per bit can also help filter out noise.

Perhaps the simplest application of locating points in the projection area is to locate the four corners of a rectangular target screen. Then the image can be projected to precisely match the boundaries of the target surface, shown in Figure 2. This is accomplished by pre-warping the image electronically such that when projected, the result is a seemingly undistorted image matching the physical boundaries of the display. This warping transform is called a homography and is computed using the four coordinate pairs from the four embedded light sensors. The homography matrix is bound to the orientation of the sensors. Even if the optical path is folded using a series of planar mirrors the matrix will automatically rotate and flip the image as needed to maintain its orientation relative to the display surface, Figure 3. The resulting effect is the illusion of a fully functional display which has been simulated using projected light on a passive surface. The surface can be made of very light-weight material such as wood, medium density fiber (MDF), foam core, or even stretched canvas. These light-weight simulated displays can be used in applications where it would be physically or economically prohibitive to use real displays. Similarly, a single projector can be used to

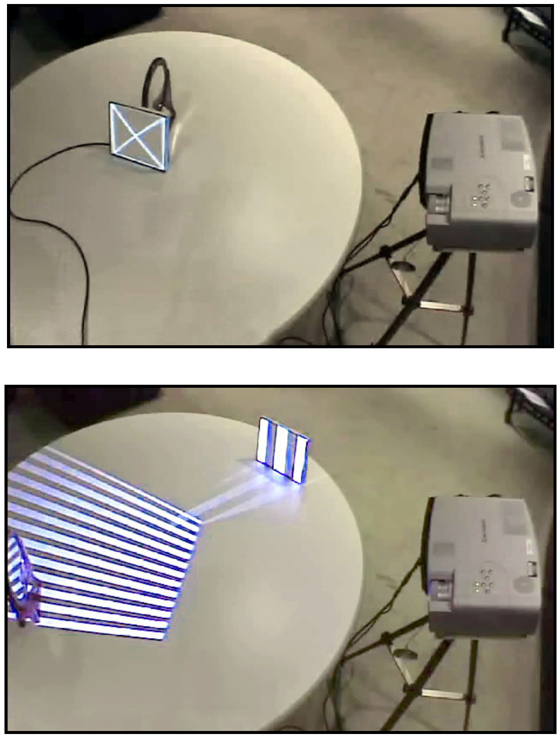


Figure 3. Location discovery of sensor continues to work even under very shallow projection angles (top) and when the optical path is folded using a mirror (bottom).

simulate multiple small screens simultaneously, thereby creating several "picture-frame"-like displays that can be haphazardly scattered on a bookshelf, mantle, or desk. The basic geometric relationship of surfaces in the environment can also be captured by temporarily placing sensor frames on desks, floors, and walls for multi-surface projection applications such as the Everywhere Displays project [Pinhanez, 2001].

The instrumentation of the target surface can be seen in Figure 4. We used optical fibers to channel the light energy from each corner to a sensor board, which then relays the data to a host PC. The optical fiber is made of toy grade plastic and costs just a few cents per meter. Because each fiber is only 1mm in

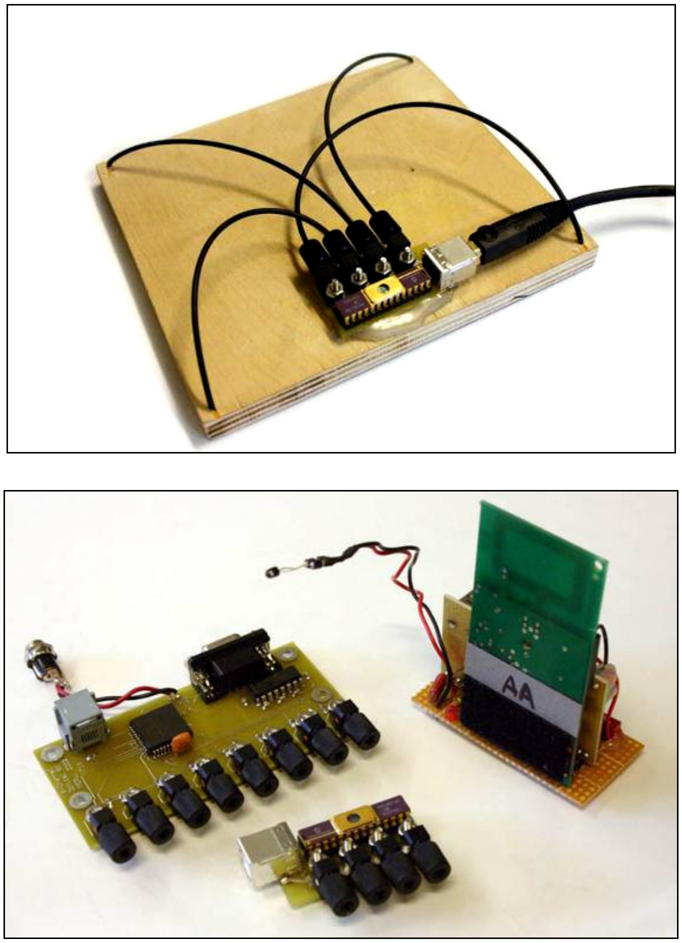


Figure 4. Top: A rear view of the instrumented target surface showing the optical fibers and sensor package. Bottom: an 8-channel and 4-channel wired sensor packages and a single wireless RF sensor tag.

diameter, they minimize the physical presence of the sensor at the projection surface. Additionally, fibers allow the use of a small centralized electronics package placed in a convenient location regardless of screen size or geometry simplifying surface instrumentation. The fibers are installed at each corner such that the tip lies just beneath the front white surface of the screen. This hides any visual evidence of the fiber, seen in Figure 2, and also provides a light diffuser that helps bounce the light into the fiber even at very shallow projection angles. This prototype was able to provide successful location discoveries even when the projection angle was less than 2 degrees, Figure 3 (left). This loosens the constraint on projector pose, and provides the ability to create shallow front-projected displays using commodity commercial projectors without the need for specialized optics.

Since this prototype had a wired USB connection to the PC, both power and bandwidth were available to stream sensor data over the connection. An early prototype of a wireless tag as well as an 8-sensor and 4-sensor wired package can be seen in Figure 4. The number of light sensors varies depending on the desired application. The bandwidth and power requirements of the sensors are easily supported using contemporary wireless technologies.

Applications of Projector Based-Location Discovery

Examples of applications which use four sensors to define a quadrilateral include display simulation as described above, overlaying and aligning multiple projectors onto the same screen for stereoscopic projection, boosting image brightness, dynamic shadow elimination [Sukthankar, 2001], and creating a 3D TV [Matusik, 2004]. By performing location discovery sequentially for each projector, we can co-align as many projectors as the application demands quickly and accurately as shown in Figure 5. The sensors do not necessarily have to be used to define the boundaries of the image but simply define points contained within the overlapping image area to find the correspondence between each projector. When used in this manner, more than four sensors can be used to acquire a best-fit solution resulting in sub-pixel alignment accuracy. Similarly, this technique can be applied to multi-projector stitching applications. By positioning sensors in a regular grid, projector stitching becomes a simple extension of single-display warping. In Figure 5, two projectors are stitched together using a surface

Figure 5. Multi-projector applications: image layering (top) and image stitching (bottom)

containing six sensors (one in each corner and a shared pair at the midpoints of the top and bottom edges). Each non-overlapping set of projectors is patterned sequentially, warped, and then blended. This stitching technique easily scales to larger numbers of projectors and can also be applied to non-planar surfaces such as a planetarium assuming the curvature of the screen is known in advance.

Touch calibration can also be done with just four sensors since the same homography used for single display warping can be used to map 2-dimensional tracking data to the projected image. However, often having more than four

Figure 6. A commercial calibration-free interactive whiteboard system.

sensors is useful for both sub-pixel calibration and robustness against placement of the projection area. In these applications, the sensor location is typically used solely for touch calibration and the projected image remains unwarped. For automatic touch calibration to be successful there must be a known physical relationship between the sensors and the touch technology. Touch surfaces such as the Diamond-Touch table and electronic white board systems from SmartBoard and Polyvision, where the tracking technology is integrated into the display surface, can provide a known physical relationship to the light sensors. However, attached devices such as the Mimio tracking system would not benefit greatly from embedded light sensing without permanent installation of the locator, or using specialized calibration tags. To prototype touch calibration, we installed 8 fiber sensors into a Diamond-Touch table – 4 fibers in each corner of the touch sensitive area and 4 fibers in the interior of the sensing area defining a rectangle inset by 10 cm on each side. This allowed the Diamond-Touch table to either perform image warping to match the projected image directly to the touch sensing location data or calibrate the touch location data to an un-warped projected image

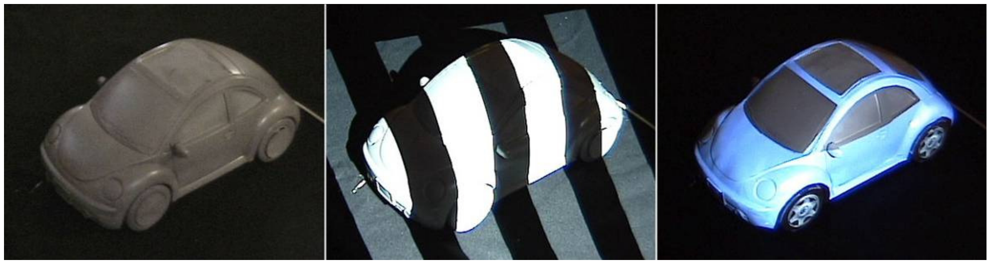


Figure 7. Auto-calibrating Shader Lamps - appearance augmentation using projected light

that may be entirely contained within the touch sensing area. A major commercial interactive whiteboard manufacturer has licensed this technique and released a new line of self-calibrating electronic whiteboard systems shown in Figure 6. This manufacturer elected to embed 16 or more fibers throughout their whiteboard to ensure a high probability that any usefully large projected image would encompass at least 4 points. The locations of the fibers relative to the touch sensing surface is discovered at the factory and then stored in the whiteboard system memory. This allows an end user to begin interacting with the image from a newly positioned projector immediately after a brief automatic touch-calibration period.

Shader Lamps [Raskar, 2001] is a method for using projected light to dynamically decorate physical objects. Surface textures, material properties, illusion of movement, and different lighting conditions can all be simulated with projected light. However, this illusion requires extremely accurate registration of the projected image onto the physical surfaces to be effective. This was previously achieved through a tedious manual process lasting 15-20 minutes and must be entirely redone if either the model or projector is moved or adjusted even slightly. We reconstructed a demo from the Shader Lamps work using projector-based location discovery and embedded light sensors, shown in Figure 7. The gray model car is given a new paint job, a sunroof, and hubcaps. Eight optical fibers are embedded at key registration points around the model, connected to an

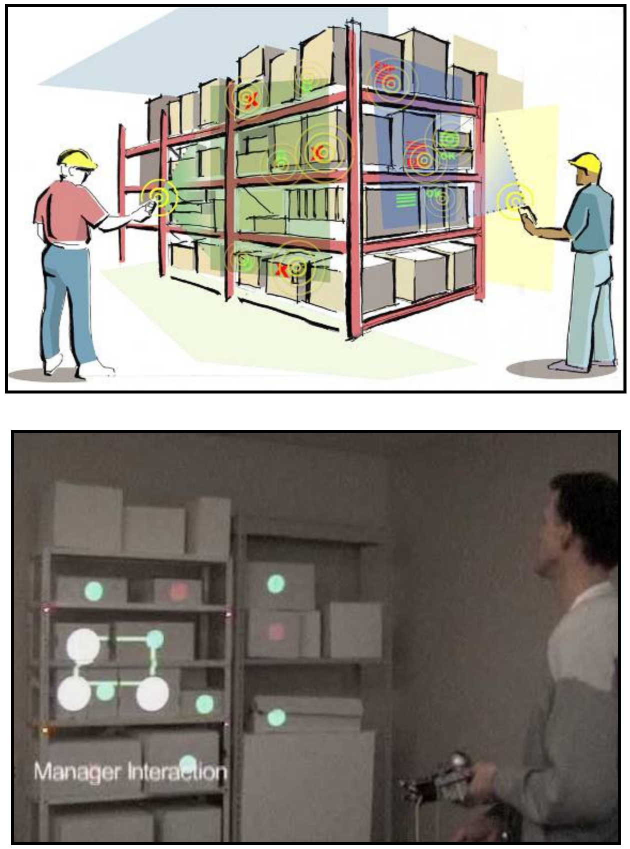


Figure 8. Using photosensitive RFID tags and a handheld projector to point, select, and digitally annotate physical objects that have been visually augmented.

8-sensor board shown in Figure 4. These registration points are used to discover the projector pose with respect to the physical model, given knowledge of the model geometry and sensor locations.

In the RFIG (radio frequency identification and geometry) tag project [Raskar, 2004], this technology was used in the development of a hand-held projection system that used structured light projection to discover the location of wireless photosensitive RFID tags, Figure 8. This system combined a small portable projector, an RFID reader, and a camera for tracking objects and surfaces at interactive rates allowing the projector to be used as a real-world pointing and annotation device. The photosensitive tags scattered in the environment can be scanned to discover both their identity and location. The handheld projector can then interact with these tags and manipulate the data contained within them in a spatially aware manner. One possible usage scenario of this system is in warehouse inventory control environments, where a worker could use the handheld projector to point at, highlight, select, and annotate RFIG tags attached to boxes on a shelf. The annotations and neighborhood location data are stored locally on the tags so that they can be recalled later by another worker using a similar handheld projection system. The second worker would scan the RFIG tags for their data and locations and then highlighted annotations would be projected directly onto each tagged object indicating actions to be taken by the worker or indicate if the boxes had been moved. While the locations of tags were discovered using projector-based location discovery, interactive movements and pointer tracking were accomplished using a camera and visually distinct grounding markers. The 60Hz refresh rate of a standard commercial projector was not sufficient to perform full screen tag tracking at interactive rates. However, as alternative projection applications become increasingly common, the demand for novel projector designs will increase. For this project, I developed a micro-projector prototype, shown in Figure 9, to demonstrate the feasibility of small handheld projectors. This fully functional projector has a total volume of approximately 1 cubic inch and a 640x480 resolution with a 60Hz refresh rate. Commercial prototypes of micro-projectors have just begun to emerge as of the writing of this dissertation such as Symbol's Laser Projection Display (LPD) [Whittenberg, 2007], Microvision's PicoP projector [Microvision, 2008], and 3M Micro-projector [Hunter, 2008]. Other projector prototypes I have built, described later, optimize for other factors such as non-visible light and higher frame-rates allowing projector-based photo sensor tracking interactive rates in a visually acceptable manner.

Limitations and Discussion

It is important to note that when the warping is done electronically, the projected image must be resampled at a lower resolution to achieve proper physical alignment. As the magnitude of warping increases, the resulting image quality decreases. This degradation primarily impacts the readability of small text and the appearance of fine lines, though larger fonts and images maintain a reasonable appearance. Image filtering does improve appearance of down-sampled video, but we are ultimately subject to the physical limitations of the projector and the inherent resolution loss due to down sampling. It is possible to perform the warping process optically rather than digitally, thereby avoiding resampling of the image. However, such units require six-degree of freedom manipulation of the optical elements which causes them to be prohibitively expensive and impractical for widespread use. Pan-tilt-zoom capabilities could be used to locate target displays and zoom in to preserve pixel density, but these devices can also be expensive and complex.

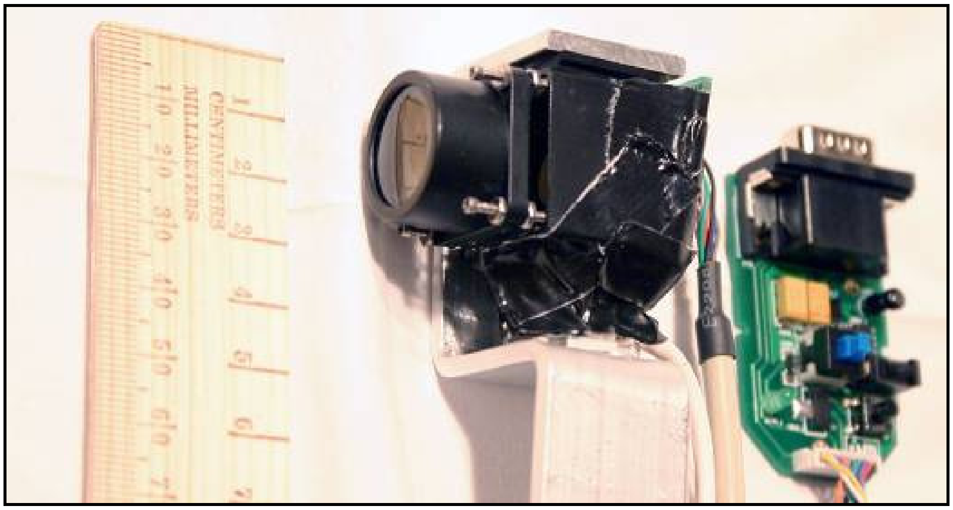


Figure 9. Miniature handheld projector prototype.

While the Gray-code pattern sequence used in this prototype yielded robust performance, there are a few modifications that could be made to further improve the quality of the demodulation of the location data. This prototype uses a simple threshold to determine the value of each bit from each pattern. However, a more robust approach would be to project each pattern followed by its inverse and then determine the value of each bit based on the difference in the intensity of light between each pattern pair. The disadvantage is that this doubles the patterning time. If this time cost is unacceptable, a small improvement can be gained by selecting a better threshold level. While an acceptable threshold was obtained by averaging the sensor output from an all white pattern and an all black pattern, this midpoint makes the assumption that the light sensor has a linear output response. To help control for non-linearity in the sensor, a pixel-size checkerboard pattern and its inverse provides samples near the midpoint grey-level response of the sensor. Averaging these two values, would provide a slightly better threshold level for demodulating the location patterns.

In this first prototype, I focused on core implementation issues and explored applications that can be simplified using this technique with an unmodified commercial projector. The issues included pattern sequence protocol, location demodulation, sensor packaging, fiber optic installation, homography calculation, and image warping. The applications described thus far have involved target surfaces whose physical relationship to the projector remains relatively static. A one-time location discovery is needed for the initial setup to create the illusion and re-discovery is necessary only if the display surface or projector is moved. One cause for the static nature of the discussed applications is the low-frame rate of commercial video projectors prohibiting support for interactive tracking of photosensitive objects. Additionally, unmodified projectors can only render patterns using visually distinct color values which are visible to human observers. These patterns tend to be distracting and produce visual strain on a human observer when presented at 60Hz. In the second implementation, I develop a prototype that addresses both of these issues.

4: Projector-Based Tracking

Inspired by my previous work, Summet began rough exploration of tracking a single sensor in [Summet, 2005]. Summet experimented with low-resolution geometric patterns in an effort to detect the direction of sensor movement. The pattern was composed of a hexagonal ring divided into smaller regions, each with a unique light encoding, and was placed around the discovered location of the sensor. This provided a method of detecting small movements by decoding which area the sensor had entered and then re-centering the pattern over the predicted sensor location. Due to its geometric nature, the pattern was relatively large, required a complex encoding scheme due to number of neighboring regions, was not robust to inter-region sensor placement, was only able to coarsely resolve the detected offset, and resulted in extremely frenetic visual activity that distracted from other projected content.

In the second implementation, I expanded upon the previous prototype and adopted a similar incremental tracking approach. However, I use different tracking patterns to address the issue of pattern visibility, encoding simplicity and robustness, higher tracking rates, and higher positional resolution. To explain how this is accomplished, I must first briefly describe how consumer grade projectors work.

Anatomy of a Projector

With only a few exceptions, most modern video projectors have three major components: a bright light source, a device to modulate the light to create an image, and optics to scale the resulting image onto a display surface. The current dominant light modulation technologies used in front projection systems are liquid-crystal display (LCD), liquid crystal on silicon (LCOS), scanned beam laser displays, and Digital Light Projection (DLP) technology from Texas Instruments. There are many properties of DLP technology that make it attractive for projector-based location discovery and tracking and thus we have focused most of our attention on this technology. However, the general concept of projector-based location discovery and tracking will continue to apply regardless of future display technology. In some cases, a simple timing technique could be used to discover the location of light sensors. However, this would require frame synchronization with the projector.

Digital Light Processing refers to consumer video projection devices that use a Digital Micro-mirror Device (DMD) for light modulation. A DMD is a very high-density array of computer controllable microscopic mirrors that can be directed either to reflect light away from, or toward, the projection optics creating black and white pixels respectively. Each mirror corresponds to a single pixel in the projected image. To create grey pixels, each mirror rapidly moves back and forth faster than 50,000 times per second using a pulse-width modulation (PWM) style encoding. The actual performance of the DMD mirror remains confidential but is speculated to be significantly faster. The human visual perception system then interprets these high-frequency flashes of varying duty cycles as varying levels of gray. To create color, a rotating color wheel is placed in front of the light source to rapidly cycle between red, green, and blue light. The DMD processes each separate color channel of the source image sequentially. Typically, the color wheel spins at either two or three times the base refresh rate of the video signal of 60Hz. The human vision system then integrates the images together to create the appearance of a single color image [Yoder, 1997]. While some recent implementations of DLP technology have greater sophistication, this still remains the basic approach for single DMD chip projectors. There are two key features of DLP technology that make it attractive for location discovery and tracking approach presented in this thesis: the very high potential frame rates for binary images and the ability to modulate invisible near infrared (IR) light. Since DMDs utilize physical mirrors, DMDs are able to modulate frequencies of light that fall outside the visible spectrum unlike liquid-crystal technologies. My second prototype utilizes the pulse-width modulation encoding of light intensity from a modified commercial projector to reduce the perceptibility of the tracking patterns. Exploring the high-speed and infrared capabilities of DMDs using a custom DMD-based projector was done in the third implementation and is described later in this dissertation.

Low-Perceptability Tracking Patterns

The high-contrast location discovery patterns used in the previous prototype created a brief but rather caustic visual experience. While short infrequent bursts is less of an issue for some applications, a constant stream of flashing patterns to perform location tracking presents a significant usability issue. The pure white and black patterns used in [Lee, 2005; Summet, 2005] delivered a data stream to each light sensor in a manner analogous to an amplitude modulation (AM) transmission. The amplitude modulation corresponds to the difference in white, 1, and black, 0, intensities and the effective carrier frequency is the cumulative frequencies found in the bandwidth of light emitted by the projector. With such a wideband carrier in the visible light spectrum and a modulation rate of 60Hz, these patterns are manifested as high-contrast black-and-white stripes which are easily visible by a human observer. One solution to this problem is to remove the low-frequency components, either temporally or spatially. To do this we can use a higher frequency modulated (FM) data transmission technique rather than a slow AM transmission. In other words, rather than use the presence or absence of light, we can use rapid light flashes at different frequencies to encode each bit resulting in a less caustic visual experience. To accomplish this, I take advantage of the pulse-width modulation color rendering used in commercial DLP projectors to achieve a FM transmission alternating between carrier frequencies of 180Hz and 360Hz. The result is a tracking pattern that appears to be solid gray to a human observer but in actuality contains rapidly changing location data detectable only by a light sensor. The data modulation rate still remains 60Hz, but our human vision system is not able to detect a difference between the two carrier frequencies thus making the modulation imperceptible.

To achieve the FM transmission described above, I removed the color wheel from an InFocus X1 DLP projector, which contains an 800x600 pixel (SVGA) resolution DMD. This creates a gray-scale only projector, and flattens the original color space into a single luminosity dimension. If we select two colors that have similar DMD duty cycles, they will appear to be identical shades of grey to a human observer but may be rendered using very different signal patterns which are detectable by a light sensor. Specifically, a pure bright red color and a medium gray color when rendered by the modified projector have an identical grey appearance to the human eye, but are manifested as a 180Hz signal and a 360Hz signal respectively, Figure 10.

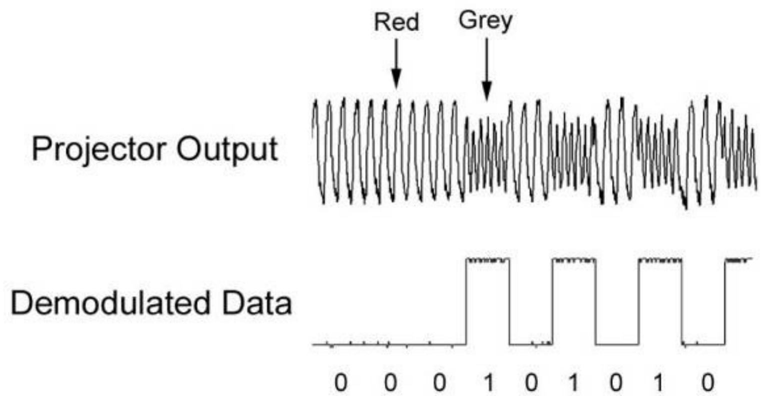


Figure 10. Frequency modulated pattern transmission using two colors

By using these two colors, I can hide the tracking patterns in what appear to be solid gray squares. In this implementation, the gray regions retain a very slight perceptible flicker. This is an artifact introduced by the projector's internal color processing system managing the transition between the two colors resulting in a minor deviation from the carrier frequencies between frames. As a result, the transitions appear momentarily brighter or darker than either base color. However, the flicker is very subtle and is not likely to be a noticeable visual distraction when performing a task.

Achieving Interactive Tracking Using 60Hz Refresh Rate

As described earlier, the number of Gray-coded binary patterns necessary to resolve the location of a light sensor to a single pixel in a projection area is bound by $\log_2$(*number of pixels*). Thus, an SVGA (800x600) projector requires 20 images yielding a maximum update rate of 3Hz using 60Hz modulation. We can improve upon this update rate by using an initial full-screen location discovery step followed by localized tracking using smaller patterns. Once we discover the absolute position of each sensor, we can project smaller tracking patterns over their locations to obtain incremental offsets. Smaller patterns require fewer divisions to resolve down to a single pixel. Therefore, we can acquire incremental offsets much faster than absolute positions. Additionally, small, localized tracking patterns liberate the rest of the projection area for application content.

In this implementation, I use square axis-aligned tracking patterns centered over each sensor that subdivides the contained space horizontally five times and vertically five times using Gray-coded FM binary patterns. This creates a 32x32 unit grid centered over the previous sampled location of the sensor. Once the offset is found, the tracking pattern is then re-centered over the updated location. The number of subdivisions for the localized tracking patterns was chosen primarily for its even division into 60Hz yielding an *x-y* coordinate pair update rate of 6Hz. Finer or coarser tracking patterns could be selected for speed and accuracy depending on the needs of the target application. However, there is a limitation on the minimum number of divisions a particular implementation can support due to system latency. Since incremental tracking uses the most recent location offset to reposition the tracking pattern, the location data must be decoded and propagated through the entire feedback loop before the next set of patterns can begin projecting. In this implementation, the average loop time was approximately 60ms which corresponds to 3-4 frames at 60 frames per second. Since I only use 10 frames per tracking update, a latency of 4 frames is a substantial increase to the overall sensing time. A large portion of this latency was caused by task scheduling within the operating system of the host PC and is not inherent to the tracking technique. This latency comes for the graphics pipeline that renders the patterns and the communication pipeline that return the data from the sensors to the software application.

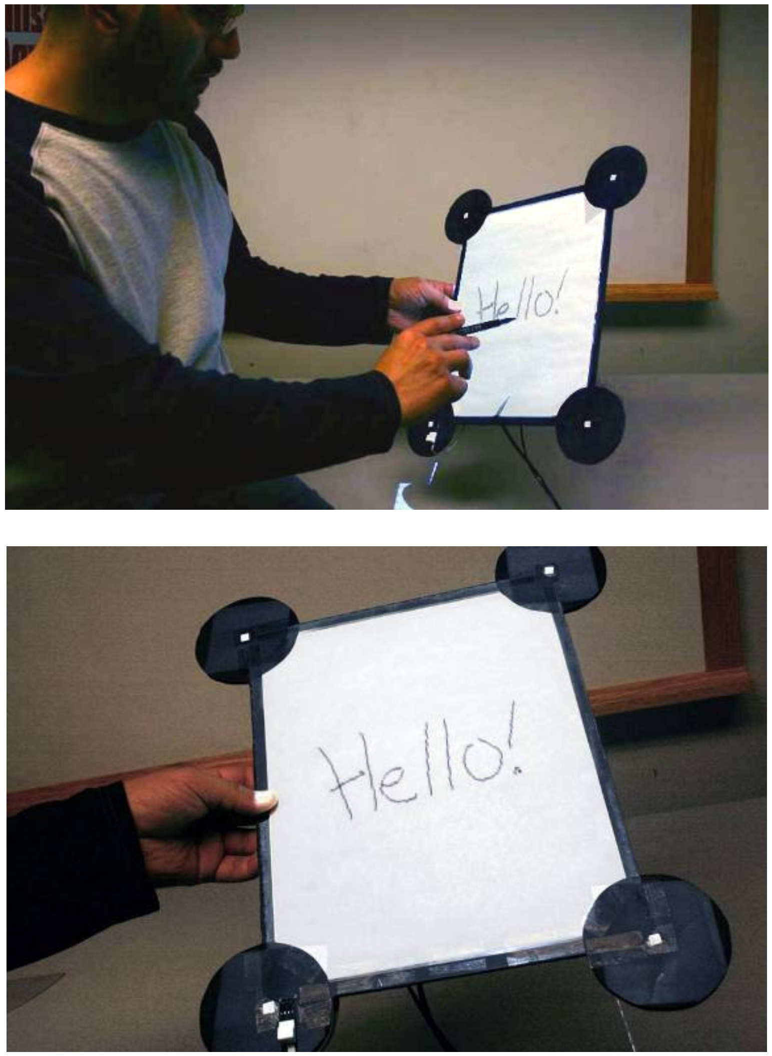


Figure 11. A handheld surface containing optical sensors and a touch sensitive surface is tracked using project-based tracking. The location data is then used to project content that matches the surface movements simulating an ultra-lightweight tablet display at low-cost.

To prevent this latency from severely impacting the tracking rate, I take advantage of the Gray-coded patterns ability resolve the *x* and *y* offsets independently. This axis independence allows me to use an interleaved tracking technique. This effectively pipelines the tracking operations allowing me to transmit the tracking patterns for one axis while waiting for the result from the other axis to propagate into the tracking pattern software. Since the feedback latency is less than 4 frames and the patterning time for a single axis is 5 frames, I can retain 100% utilization of the projector's tracking capability. The end result is a tracking update rate of 12Hz alternating between each axis. It is important to note that though I was able to find a reasonable solution using grouped Gray-coded patterns, feedback latency places a substantial constraint on the usage of alternative patterns that may utilize recent sensor data to improve tracking performance. Tracking algorithms that require instantaneous or near instantaneous feedback from sensors are not likely to be executable in practice.

Localized Pattern Size and Shape

The size and shape of the localized tracking patterns play a critical role in determining the range of movements supported by this tracking technique. If the sensors move outside of the tracking pattern boundaries within the sampling period, the sensor will become lost requiring a full-screen sensor re-discovery process. This requires a momentary interruption (0.367secs in the implementation) of an application's projected content and thus should be avoided. The size, shape, and sample rate of the localized patterns determine the maximum sensor velocity the system can continuously track without error.

I have described the tracking patterns thus far as resolving to an offset within a 32x32 unit grid using five horizontal patterns and five vertical patterns. In the simplest implementation, this grid might be mapped to a 32x32 pixel area in the projected image. This may provide an acceptable coverage of movements for applications that primarily focus on tracking objects in the image plane or tracking single sensors. However, if the distance between the sensors and the projector is allowed to change substantially, a fixed pixel dimension of the patterns will result in a wide variation in the maximum supported tracking velocity in terms of meters per second. This can be problematic and confusing to the user, for example, when moving surfaces that are meant to be hand-held such as a simulated tablet, shown in Figure 11.

For these applications, I use a fixed physical size for the tracking patterns to maintain a consistent maximum tracking velocity regardless of distance from the projector. This is accomplished by using the known geometry of the display surface and the currently observed locations of the corners. Using fixed physical dimensions also maintains the relative size of the tracking patterns with respect to the physical display as well as the projected content. Additionally, it produces a variable pixel accuracy behavior based on distance. As the display moves farther from the projector, the tracking patterns will shrink in pixel space resolving down to a single pixel. As the display moves closer to the projector, the pixel density increases making pixel-perfect alignment less important and the accuracies of the tracking patterns reduce accordingly.

The shape of the tracking patterns I use in this implementation are simple squares aligned to the image plane of the projector. I use this shape because of the axis-aligned nature of the Gray-code patterns. Elongated shapes could be used to permit a higher range of movement in one particular direction for applications such as a projected slider widget. Similarly, a variety of pattern geometries could be used to track specialized sensors that have restricted or expected ranges of movement for application specific tasks or interaction techniques. However for general purpose tracking in two-dimensions, a shape with a greater degree of radial symmetry, allowing a similar freedom of movement in any direction, is more appropriate.

Motion Modeling

It is possible to soften the maximum supported tracking velocity constraint by modeling the motion of the sensors to predict likely future locations. Since physical motions exhibit a high degree of temporal continuity, recent motion history can be used to generate a strong prediction of likely positions in the near future. The model I use consists of a moving average of recent velocity, acceleration, and jerk (derivative of acceleration). Combining these values and the most recent sampled position, we can calculate a probable path for the sensor and then center the tracking pattern accordingly. Fortunately, the predicted locations do not need to be exact since the tracking patterns search over an area giving the system a relatively large acceptable margin of error. By using a motion model, we can adjust the locations of the tracking patterns to dramatically increase the range of movements the system can successfully track. The motion constraint is then moved to the third derivative of position, jerk. The model can be made to include further derivatives or otherwise be made more complex. However, in our exploration this simple model provided a good balance between the coverage of the motions used in these test applications and tracking errors due to mis-prediction. Mis-predictions are an inherit risk of any predictive model, since no model can accurately account for all the complexities of the physical world or the intentions of the user. Motion models can be selected and tweaked to adjust the balance between freedom of movement and tracking failures. The appropriate balance will be application and implementation specific.

Tracking Loss Strategies

Tracking loss can occur for several reasons including exceeding the supported motion constraints, model mis-predictions, and corrupt or unavailable tracking data. In some cases, circumstances may allow the system to reacquire the sensor from a momentary tracking loss through chance. However, if a sensor is identified as being conclusively lost, a fallback strategy is necessary to re-discover the sensor locations. This may be triggered manually through user input, or by a pre-defined timeout for lack of sensor data, or possibly signaled by a sequence of erratic improbable offsets (sometimes a symptom of interference). There are several options that can be employed for recovering lost sensors, each having their own advantages and disadvantages with no clear choice as to which is the best overall behavior for all applications. In this section, I describe recovery strategies when tracking only a single sensor. If multiple sensors with a known geometric relationship are tracked simultaneously, this information can be used to make informed predictions and will be discussed later in *Occlusion Detection and Behavior*.

The simplest option is to perform a full screen discovery process to search the entire projection area for lost sensors. The downside is that the entire projection area becomes gray, interrupting any projected application content. However, the upper bound on the recovery time can be as short as 1/3rd of a second assuming the sensors remain in the projection area. If the conditions of use result in relatively infrequent sensor loss, this may be a reasonable strategy and is the one I use in the current implementation.

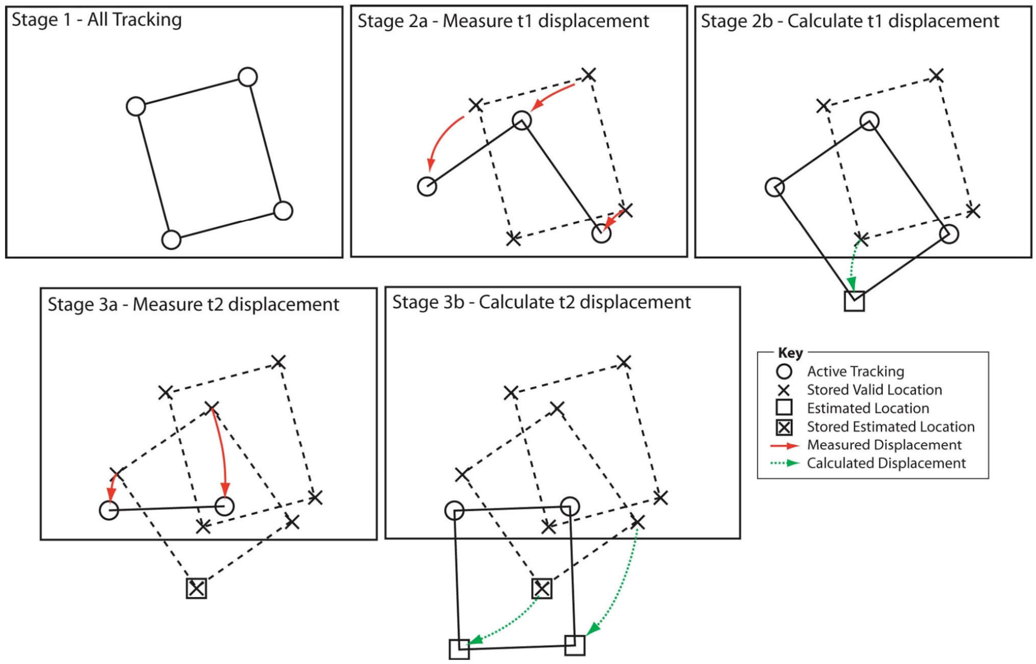


Figure 12. Illustration of the sensor location estimation process when dealing with 0, 1, and 2 unavailable sensors.

Another approach described in [Summet 2005] is to grow the tracking patterns around the last known valid location until it contains the sensor again shrinking back to normal size after the correct location has been discovered. This has the benefit of searching only a small region of the projection area yielding a potential recovery time shorter than 1/3rd of a second as well as causing a minimal amount of obstruction to any projected content. However, the upper bound on the recovery time is determined by the growth function and may result in an average performance substantially longer than the time needed to perform a full-screen discovery. Additionally, the expansion and contraction increases the visual saliency of the tracking patterns, which may potentially be more distracting and detrimental than a momentary gray screen. Alternatively, historical or statistical approaches can be employed to determine probable locations of a lost sensor. However, these techniques also suffer from high upper bounds on recovery time and increased visual saliency caused by frenetic pattern movement. Preferable behavior will likely depend on the application, usage environment, and the specifics of the implementation.

Occlusion Detection and Behavior

In addition to reducing the perceptibility of the tracking patterns, FM based transmission also improves our ability to detect sensor occlusion over our previous AM based transmission. In an AM transmission, it is often impossible to distinguish the difference between signal loss and a long sequence of ‘0’ bits. When using FM, the lack of either carrier signal signifies that the connection has been lost. Additionally, the FM technique uses very narrow band carrier frequencies when compared to the white and black image AM transmissions used in our prior work. This makes it easier to filter out interference and reject corrupted bits. These properties allow us to detect occlusions and other signal errors on a per-bit basis providing highly robust behavior. When using projector based tracking for interactive surfaces, sensor occlusions may occur frequently. Per-bit detection of signal loss allows an occlusion to occur at any point in the tracking period without resulting in a tracking failure due to corrupted data. Though reasonably robust detection of signal loss can be accomplished with AM transmission using trailing check bits [Summet 2005], this additional data reduces the overall update rate and does not guarantee detection.

To properly demodulate an FM transmission typically requires either analog filtering electronics or sufficient computing power to perform real-time signal processing. However, these substantially increase the cost and complexity of the sensor design. In this implementation, I use a simple software demodulation scheme that tracks signal amplitude and edge counts. Though a crude approximation of proper FM demodulation, it can be run on a low-cost microcontroller with minimal external components and has worked effectively in our explorations. A transmission error is defined as a sudden drop in signal amplitude, insufficient signal amplitude, or invalid edge count. These errors are able to flag signal loss due to occlusions or leaving the projection area and some limited forms of signal interference. The carrier frequencies of 180Hz and 360Hz generate 6 and 12 edges respectively every frame period. Valid edge counts (using a +/- 1 margin) are converted into 0's and 1's while invalid edge counts are flagged as errors. These error flags are transmitted back to the host computer with the decoded bit string.

Once we are able to reliably identify these transmission errors, we must decide what policy to use in the behavior of the tracking patterns when the sensor location is unavailable. One policy is to simply discard the data and reuse the last known valid position of the sensor. The resulting effect is that the tracking pattern does not move if an occlusion occurs. When tracking individual sensors, this may be the most appropriate policy. In our exploration, we informally observed that many occlusions occur when the user is attempting to interact with other objects rather than moving the sensor itself, such as pointing at the projected content, drawing on the touch sensitive surface, or just walking in front of the projector. Thus, the likelihood that a sensor remains stationary during an occlusion is reasonably high. If we are tracking multiple sensors simultaneously in a known geometric configuration, such as the simulated tablet application shown in Figure 11, we can use the displacement of the available sensors to generate an estimated location of any occluded or off-screen sensors. With respect to the execution of this estimation technique, there is no functional difference between sensor occlusion and a sensor moving out of the projection area. Thus, for the purposes of explanation, I will describe this process in the context of a tablet exiting the projection area as illustrated by Figure 12.

In Stage 1, all sensors are visible by the projector and no estimations are necessary. If one sensor moves outside of the projection area, Stage 2a, we store a snapshot of the last valid locations for all four sensors and then measure the displacement of the three tracked sensors. These six offset values can be used to calculate the top six values in a 3x3 affine transformation matrix, t1. The estimated location is then the last valid location of the lost sensor multiplied by t1, Stage 2b. This affine transform encapsulates translation, rotation, scale, and skew providing a very strong estimate of the lost sensor's location. Though tracking may be impossible if the estimated location is outside of the projection area, this estimated point can still be used to preserve the geometry of the projected content. When a second sensor is lost, Stage 3a, another snapshot is taken of all four sensor locations (tracked or estimated) at the time of disappearance. Then the displacement of the remaining two sensors from their respective snapshot locations is used to generate another transform t2. However this transform is significantly simpler than t1 encapsulating only two dimensions of translation, one degree of rotation, and one degree of scale. As expected, the strength of the estimation becomes progressively weaker as we have fewer sensors to compute the transformation. If an additional sensor is lost and we are left with a single actively tracked sensor, we are limited to only updating the translation of the geometry to motion-match the remaining corner. The screen must be brought back into the projection area at a similar orientation if the tracking patterns are to re-acquire the lost sensors. In our exploration, we found this occlusion behavior to be effective at estimating sensor locations under typical usage. However, performing complex movements when tracking data is scarce will cause the estimations to be incorrect. If this occurs, a full-screen discovery or another fallback strategy described in *Tracking Loss Strategies* must be performed.

A closer look at Stage 3b shows that final estimated location of the first lost sensor is actually the result of two transformations, t1 and t2, from the last known valid location. This is significant because the estimation error of each transform is compounded defining a relationship between likelihood of estimation error and the order in which a sensor was lost. Additionally, you can see in Stage 3b that we specifically transform a stored snapshot of the estimated location rather than calculate the final estimated value dynamically using t1 and t2. The reason for doing this is because we are not guaranteed to have LIFO ordering of sensor loss and re-acquisition. Otherwise, we could simply implement a matrix stack for each lost sensor and push and pop matrices as needed. However, when LIFO ordering is not maintained, matrices may have to be deleted or modified in the middle of the stack. Using location snapshots simplifies the implementation and accounting tasks required to support non-LIFO ordering of sensor loss and reacquisition.

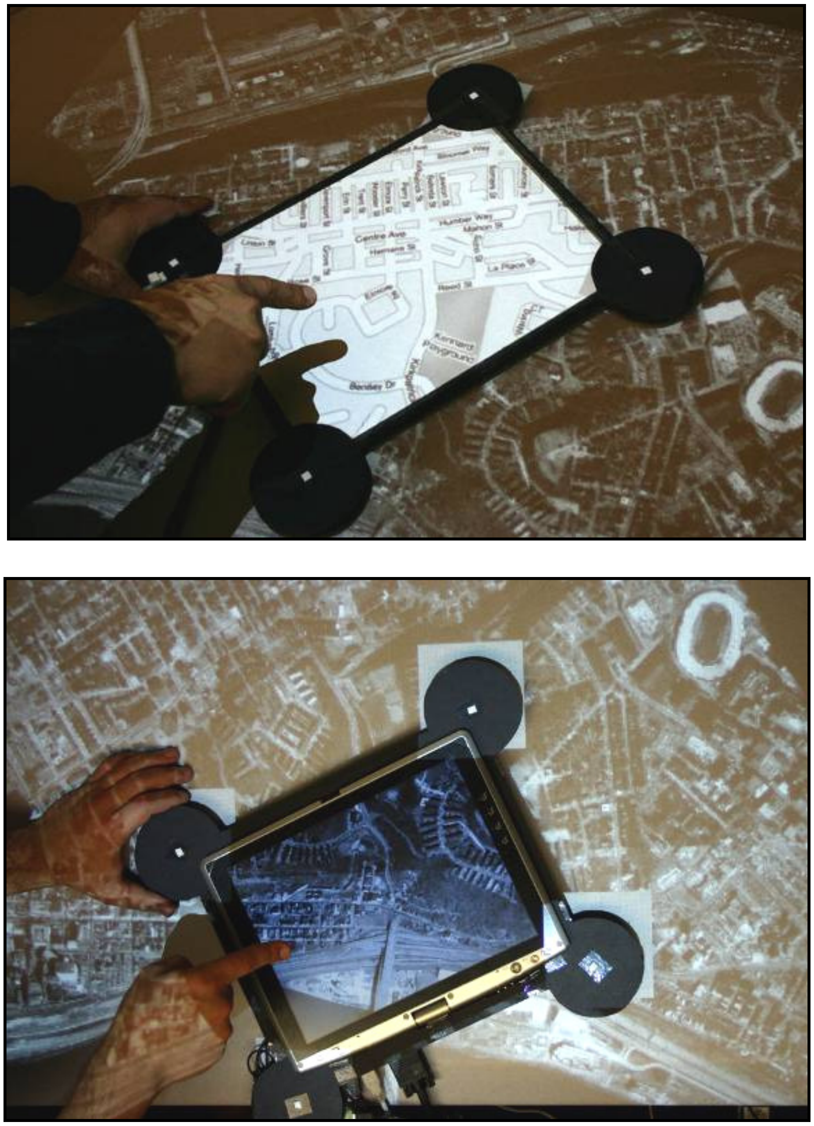


Figure 13. Magic Lenses using only a projector (top) and with a high-resolution display (bottom)

Applications of Projector-Based Location Tracking

In addition to supporting the set of applications enabled by low-speed projector based location discovery, interactive motion tracking allows the simulation of fully functional and interactive tablet-like hand-held displays. As before, we project onto a rectangular surface containing optical sensors in each corner, shown in Figure 11. A homography is computed from the sensor locations to pre-warp projected content to fit the physical boundaries of the tablet. I also added a touch-sensitive surface allowing the user to interact directly with the projected content. A user can use their finger or a stylus to create free-hand drawings, take notes, or interact with a graphical user interface such as a webpage just they would with a tablet PC. This effectively allows the creation of fully functional tablet-like surfaces that are very inexpensive and weigh only slightly more than a typical clipboard. If high speed projection is available, a light sensitive pen could also be tracked using the projector eliminating the need for the added touch-sensitive surface. By displacing the display technology, it is possible to reduce costs by using a few ceiling mounted projectors to simulate hand-held displays in a private work environment or public space such as a museum where tablets may be given to visitors. If the surfaces are damaged, lost, or stolen, they can be easily replaced at minor expense. An environment such as a medical office might use a very large number of these surfaces to physically manage information similar to clip boards or file folders with the benefits of computerized tablet displays, but without the additional cost or weight. Though the performance of this prototype is far from being able to render modern tablet PCs obsolete, improved engineering could reduce this performance gap making this a viable and practical alternative in some application scenarios.

Magic Lenses [Bier, 1993; Ishii, 1997; Ulmer, 1997] are an elegant technique for allowing users to easily explore two-dimensional data sets containing multiple layers. For example, geographical information system (GIS) data contains aerial photographs, street data, and topography information and a hand-held surface could be used to physically and dynamically explore different layers of the data. They can be also be used to create transparent/translucent tools or visual filter lenses. We can use the projection area outside the boundary of the moveable surface to display one view of the map data while the inner area provides a window into an alternative view, Figure 13. To explore a different region of the map, the user can simply move the surface over the new area of interest. Alternatively, we can easily substitute the passive white projection surface with a high-resolution LCD display creating a moveable version of the Focus plus Context display [Baudisch, 2001]. We use four optical sensors to discover and track the corner locations of the LCD and modify the displayed content accordingly. In addition to allowing the user to choose an alternative view of the data as described before, the high-resolution display also provides a much greater level of detail than the projected image. In this implementation, I used an SVGA InFocus X1 projector and a tracked Toshiba Portege M200 tablet PC, which provided a 10:1 ratio in pixel density.

Though these applications thus far have been described using a single display surface, a projector can easily simulate more than one moveable display simultaneously, Figure 14. Each surface is tracked independently and the content for each display is warped appropriately. Additionally, because both displays are tracked using the same projector, the geometric relationship between the displays is also readily available. This information can be used to adapt the content of the two displays for display interactions such as self-orienting display spanning or intelligent transferring of objects between screens [Hinckley, 2003; Hinckley, 2004; Rekimoto, 1999; Streitz, 1999].

This technique can also be used to track individual sensors, Figure 15. These sensors are packaged in black foam-board magnetic “pucks” that can be manipulated in a physical manner on a whiteboard or digital workbench. These input points can then be used to define the ends of a multi-handed physical input tool such as a map measuring tool, similar to [Ishii, 1997; Ulmer, 1997], or to physically manipulate control elements in a planning task or simulation (e.g. a particle flow system), similar to [Ben-Joseph, 2000].

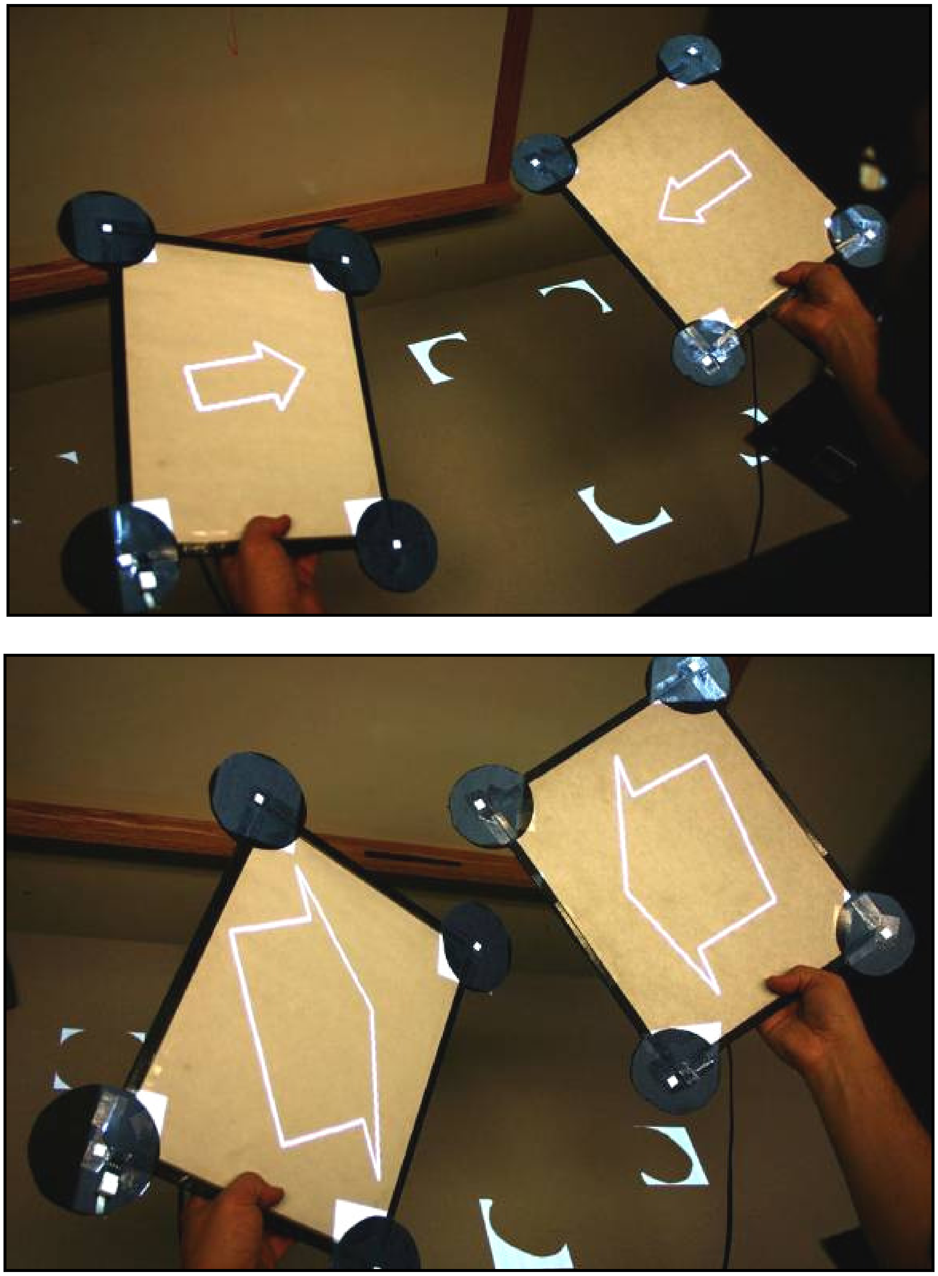


Figure 14. Location sensitive multi-display interaction. The projected arrows reflect the direction and distance to the other surface.

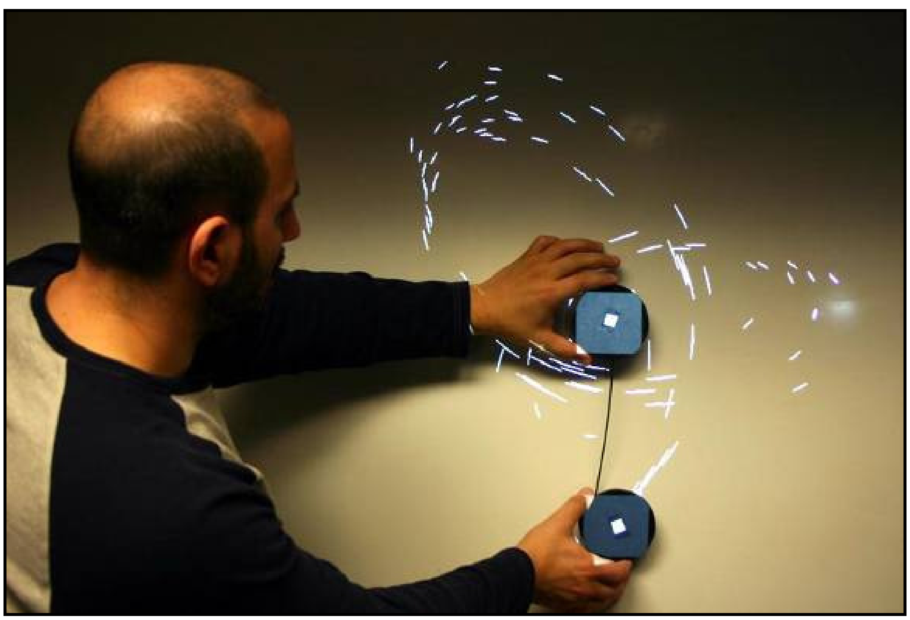


Figure 15. Physical input “pucks” for use in a mapping tool or interacting with a simulation.

5: Hybrid Infrared-Visible Light Projection

The previous tracking implementation used an off-the-shelf projector which encoded pixel locations using visible light patterns. This resulted in tracking patterns that could be seen by human observers and also consumed a portion of the projection area reducing the number of pixels available for application content. While the previous work had success in reducing the perceptability of the tracking patterns using high-frequency visible light patterns, the long term goal was to create a projector capable of projecting both visible images for application content and invisible infrared images for location discovery and tracking. This would allow the location tracking to occur without the user's awareness and would not interfere with application content. In this section, I describe a proof-of-concept implementation of such a device.

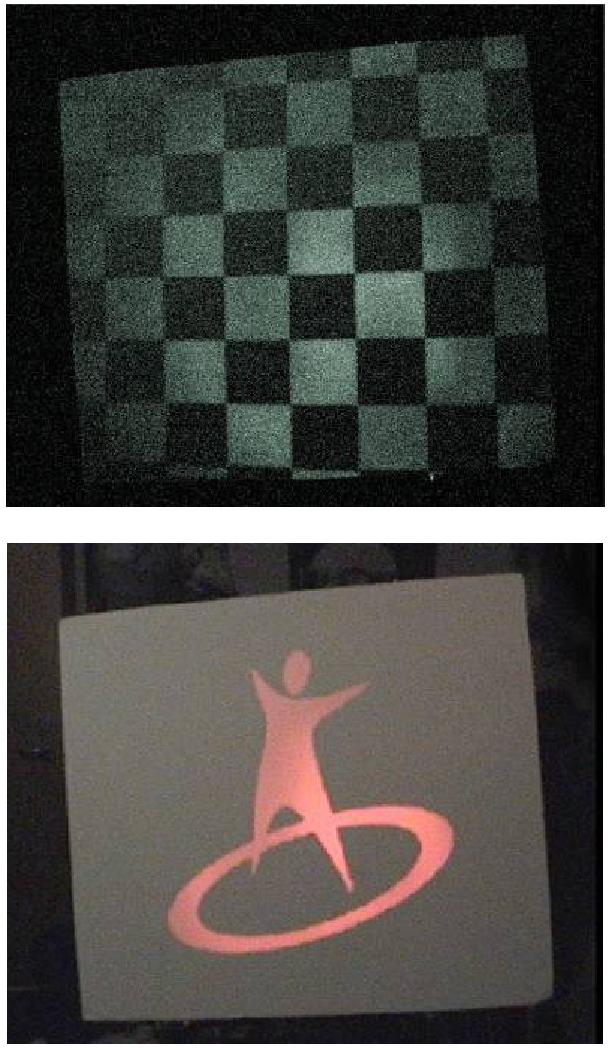


Figure 16. Two views of the hybrid projector output: a test pattern seen in infrared (top) and a visible light image (bottom)

Creating the Hybrid Infrared and Visible Light Projector

As mentioned earlier in the *Anatomy of a Projection* section, one of the primary components of a typical modern projection system is a bright light source. At the time of the writing of this proposal, the vast majority of commercial projection systems use Xenon or Metal-Halide gas bulbs which are high-output, wide spectrum, visible light sources. However recently, due to reasons of cost, size, power efficiency, and life span manufacturers have begun to explore high-output light emitting diode (LED) arrays as an alternative light source. While still relatively new to the market, this approach allows lower cost, more compact, cooler running, lower maintenance, and simpler projector designs. LED illumination has been most successful thus far in rear-projection television systems, but as the technology improves it is a likely replacement for the costly fragile gas bulbs currently used in front projection devices. LEDs also have the advantage that they can be manufactured to emit red, green, and blue light as well as non-visible infrared (IR). Color images can be created by electronically cycling each group of LEDs on and off rapidly in synchrony with the DMD rather than use a mechanical spinning color wheel. Similar to cycling between LED colors, we can use an LED light source to project both visible and non-visible IR images using a single projector as shown in Figure 16.

The light source is composed of 24 high-output visible light red LEDs and 24 high-output near infrared LEDs shown in Figure 17. Because the goal was to only create a proof-of-concept device, we did not target color support in this implementation. However, a commercial manufacturer could add a fourth IR color group to the RGB color arrays used in their existing design. A microcontroller is used to rapidly switch each group of LEDs on and off. A culminating lens is placed directly in front of the LED array to focus the light onto the DMD. Despite the semi-banded layout of the LED array, this optical configuration yielded reasonably even illumination with only moderate vignetting.

To spatially modulate our light source, I used a DMD with a 1024 by 768 array of mirrors. This is part of the DMD Discovery 1100 Kit from Tyrex Services that allows binary images to be sent from a PC via a USB 2.0 connection to the mirror array. Due to the limitations of the development kit, I could only

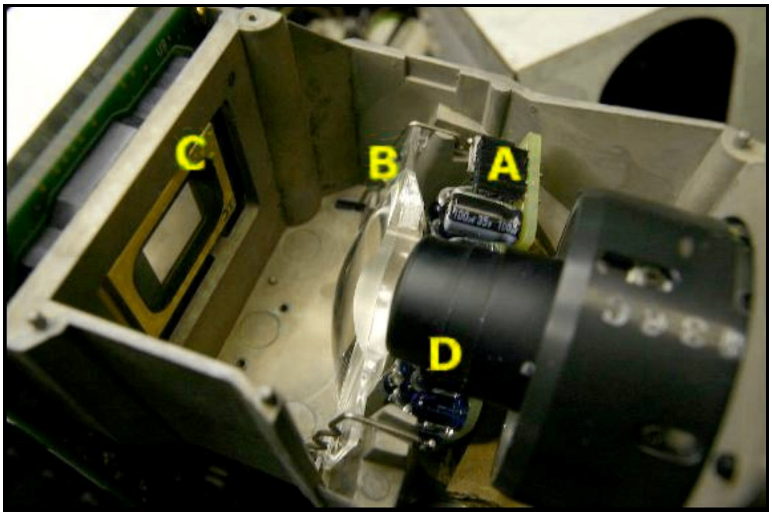


Figure 17. Top: Inside the projector A) LED light source B) culminating lens C) DMD device and D) projection lens. Bottom: light source of 24 red (clear) and 24 infrared (dark) LEDs

send 180 binary images per second. While this is far below the capabilities of the DMD itself, it allows us to explore the principle of the approach.

The projection lens and component housing used in this prototype were taken from an InFocus X1 DLP projector. This simplified the implementation as it allows us to reuse the mountings and lens system from a commercial projector providing the necessary physical relationship between each component to ensure proper optical alignment. A view of the components inside our prototype is shown in Figure 17.

Invisible Location Discovery

Once we have a functioning projector prototype capable of emitting both visible and infrared images, we can use a series of Gray-coded binary patterns to discover the locations of sensors without the user's awareness.

The light sensors we use are Vishay 56KHz IR receivers. These are low-cost receivers frequently used in remote controls. One benefit of using a modulated IR light is that it reduces interference from ambient IR sources and increases the effective range.

Due to the nature of wireless communication, the receivers have a built-in automatic gain control (AGC) which governs how much the incoming signal should be amplified before it is interpreted as digital information. This important feature allows the receiver to continue working properly in the presence of ambient noise and varying signal strength. However, the AGC can accidentally interpret long uninterrupted transmissions of the target signal as background noise resulting in de-amplification of the data stream until the signal is lost. To mitigate this behavior, we modulate the 56 KHz carrier wave during the tracking period with an alternating data pattern of "01010101…" at 2 KHz. This prevents the ACG from accommodating and ensures our IR signal will be detected by the receiver. To spatially modulate the amount of IR light each pixel location receives, we use our DMD. The projector can operate in an open-loop mode broadcasting location data without the need for feedback from the sensors. It is worth noting that the DMD is not a perfect modulator. A small amount of IR light still escapes even when the mirrors are set to reflect light away from the lens. This is caused by back-scattered light within the projector housing and other limitations of the DMD development kit. We observed that the ACG within the IR receivers would periodically detect this signal leak causing the sensors to misinterpret the location data resulting in tracking instability. It would be possible to create IR receivers with a software controllable gain to eliminate the artifacts resulting from erratic AGC behavior.

On the sensor side, we use a PIC microcontroller to look for the presence of the 2 KHz data signal to determine location. Using a series of 20 gray coded binary images, we can resolve the location of the IR receiver to the nearest pixel in a 1024x768 area. The DMD kit we are using is capable of rendering 180 binary images per second allowing up to 6 location samples per second. Our actual performance is slightly less due to synchronization overhead. As mentioned before, a production DMD unit with dedicated high-speed memory buffers is capable of rendering more than 50K binary images per second which could yield over 2500 location updates per second. In practice, manufactures would want to use the majority of the DMD duty cycle to create visible light images rather than perform location tracking. However, it would be possible to achieve 60Hz tracking using less than 2.5% of the DMD duty cycle. Location discovery could be performed in less than 400 microseconds between each visible frame providing seamless real-time input interaction with the projected content with negligible impact on the visual quality of the image.

The core concept of an infrared projector is not novel. Dynamic Infrared Scene Projectors (DIRSP) have existed for a few years, developed primarily for military thermal imaging applications. Because of the reflective nature of DMD chips, they can be used to modulate very long wave infrared light creating artificial thermal scenes [Lane, 1998]. In the non-military domain, the Smart Light System [Nii, 2004] was a very-low resolution IR projector prototype which uses an LED array to directly generate pixels in the projected image. While optically simple and offers the ability to project low-resolution images at very high-speed, it does not easily scale to arrays containing millions of pixels. The

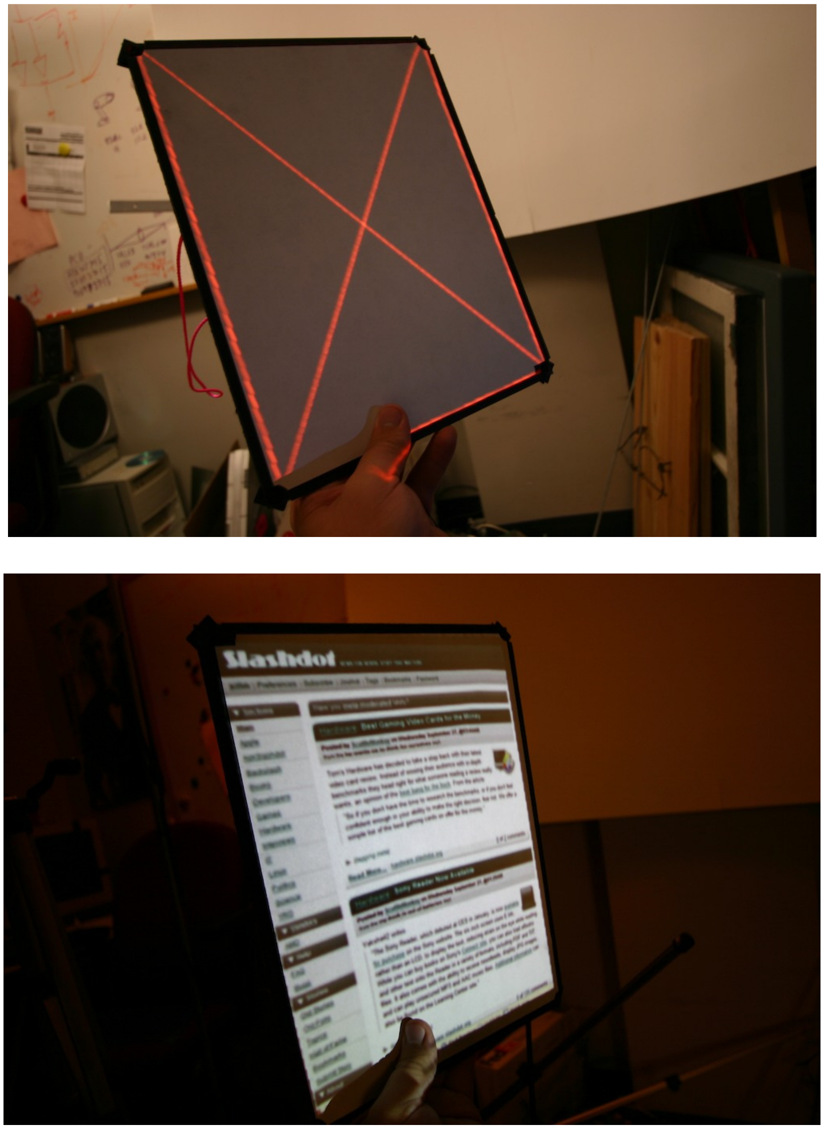


Figure 18. Tracking the location of a hand-held surface containing four IR receivers (top) and then projecting visible application content onto that surface to simulate an active display(bottom).

prototype I developed is the first to attempt at high-resolution hybrid infrared and visible light projection.

A hybrid infrared and visible light projector greatly simplifies sensor tracking over the previous prototype by eliminating the issues related to incremental tracking such as pattern size, speed, recovery from tracking loss, and interleaved updates. Since location discovery is occurring over the entire projection area in the non-visible spectrum, there is no interference with visible application content and no instability in tracking due to a failed incremental update. Similarly, pattern projection is once again an open-loop process - location data can be broadcasted without requiring feedback from sensor locations.

This prototype device successfully demonstrates that a single projector can be used to discover the locations of sensors placed in the projection area using non-visible infrared light as well as project visible application content. By unifying the location tracking and projection technology into a single device we can greatly simplify the implementation and execution of many interactive projected applications. By performing the location discovery process using non-visible light, we can track objects without the user's knowledge, preserve 100% of the projection area for application content, and search the entire projection area for sensors eliminating the issues related to incremental tracking discussed earlier. Since this prototype is limited in frame rate, for the purposes of demonstration, we simulate the output of a commercially manufactured system by coupling it with another projector to assist in displaying visible application content. This coupling can be done with either a half-silvered mirror to align the two projection frustums or using existing software techniques.

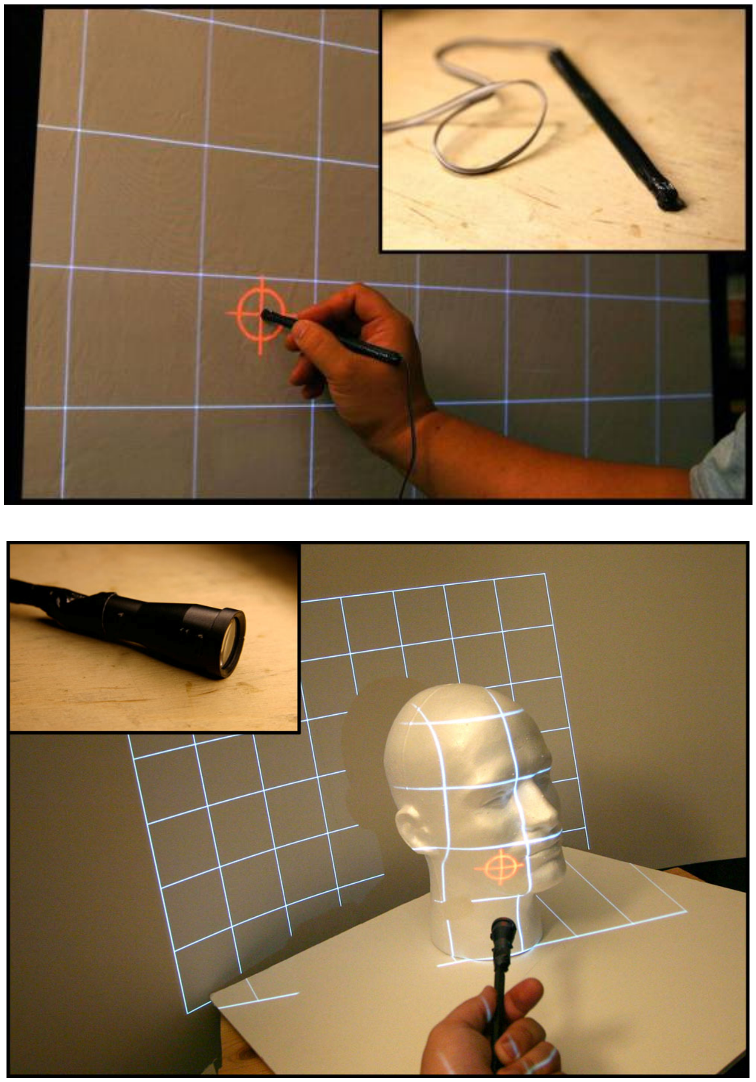


Figure 19. Top: A stylus with a light sensing tip (insert) user to interact with a rear-projected display. Bottom: a stylus with a focusing lens mounted on the tip (insert) for distant pointing on non-planar and discontinuous surfaces.

Applications of Hybrid Projection

The simulated display shown in Figure 18 is accomplished by simultaneously tracking four sensors, one placed in each corner, and then warping application content to fit the defined quadrangle. This allow us to simulate an active display on a light-weight, low-cost surface. By adding a touch sensitive film to the hand-held surface, we can simulate tablet-pc like interaction [Lee 2005].

Since the entire projection area is available for application content, the user now has an unobstructed view of the projected image allowing precise interactions using a stylus input device with a light sensor on the tip. These light pens can be used on both front and rear-projection surfaces without the need for calibration. Additionally, an unlimited number of pens can be used simultaneously without any impact on the ambiguity of stylus identity or tracking time. By placing a focusing lens on the tip of the stylus, the stylus becomes a short distance pointer on front projected displays. Both of these prototypes are shown in Figure 19. The geometry of the display surface does not need to be known and pointer tracking continues to work even if the surface is non-planar and discontinuous. This is difficult or impossible to accomplish using alternative tracking technologies. By inherently coupling the image data with location data, we can discover the pixel location of sensors despite significant distortions and modulations to the optical path.

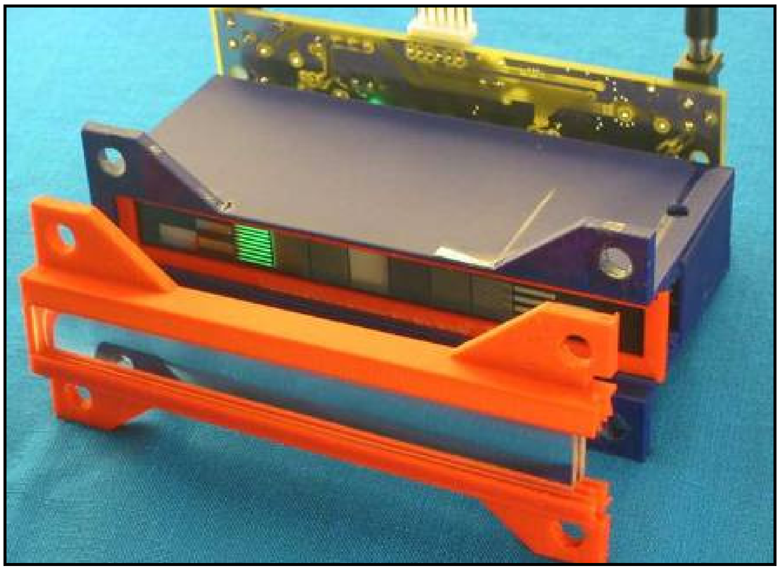
Long distance pointing technologies such as the Nintendo Wii controller utilize external IR LED emitters and an integrated blob-tracking IR camera (manufactured by PixArt Imaging) for tracking. These emitters must be placed in proximity to the display and is not sensitive to display size resulting in a relative pointing system. An IR capable projector can place multiple IR dots directly within the projected image without obscuring application content creating an absolute pointing system as well as support many spatially distributed IR dots or complex patterns such as 2D barcodes or AR tags [Kato 1999] allowing 3D recovery of the camera position and automatic screen identification in a multi-screen environment.

By embedding sensors into small objects, we can track interactive physical widgets on a table top surface similar to [Rekimoto 1999, Ullmer 1997]. Multiple sensors in a single physical widget can be used to detect rotational orientation and also perform Shader Lamp techniques [Raskar 2001]. Further reaching applications include location dependent data delivery [Nii 2005, Sugimomto 2005] and real-time range finding [Cotting 2004]. By projecting the patterns in IR light combined with an IR camera it is possible to capture depth data of a user's face or body in real-time without the user's awareness.

As mentioned before, a production DMD unit would be capable of inserting the tracking patters between each visible frame providing seamless interactive input in exchange for minor impact on image brightness. The design modification necessary to support this rich set of interactive capabilities on all LED-based, DLP projection systems currently coming to market would be very small. Even existing DLP projectors would be capable of presenting these patterns in visible light at sufficient speed to be imperceptible. Projecting each pattern followed by its inverse would eliminate non-uniformity in light distribution providing a uniform appearance. Thus, many of these applications could be supported with a slight firmware change.

6: High-Speed Motion Tracking

My co-authors Raskar, et. al. at MERL, explored a variation of this work which sacrifices visible image projection for the sake of achieving very high-speed location discovery. The approach is quite simple: use a collection of infrared LED illuminated slide projectors where each projector is dedicated to a single Gray-code pattern shown in Figure 20. By using 8-10 miniature projectors in a co-linear stacked configuration combined with high-speed LED illumination, they were able achieve very rapid spatial encoding along a single axis. This approach demonstrated the potential of achieving tracking rates as high as 20KHz using high-bandwidth IrDA (Infrared Data Association) receivers. This implementation also demonstrated some of the advantages provided by using modulated light for transmit location data. As a result, this tracking system was very robust against a wide variety of indoor, outdoor, and dynamically changing illumination. By exploiting the epipolar geometry between emitters, multiple projectors of known displacement can obtain 2-dimensional and 3-dimensional location data. This simply involves triangulating the sensor position using two or more projection units. While the approach shows promise as a potential alternative to current motion capture technologies, the loss of visible light projection transforms it into a pure location tracking technology. As a result of being an external tracking technology, it does not provide any significant advantages to the location-sensitive projector applications described earlier. However, it does have a number of advantages over camera-based motion capture systems which is discussed in *Comparison to Computer Vision Approaches.* This projection-based location tracking technology does overcome some of the frame rate limitations we encountered in our prototypes. Thus, for the purposes of exploring supported interaction techniques, we did explore combining these space labeling projectors with a visible light projector to mimic the effective performance of a commercially manufactured tracking projector. Unfortunately, the stability and resolution of the prototype available to us was not sufficient to create spatially augmented reality applications. This instability came primarily from improper optical alignment of the lenses in the prototype, the low-contrast ratio provided by the film slides, and the erratic automatic gain control behavior of the IR receivers. All of these contributors could be addressed with further engineering effort. However, doing so was outside the scope of this dissertation work. Thus, an alternative tracking technology was used to prototype high-speed concept applications.



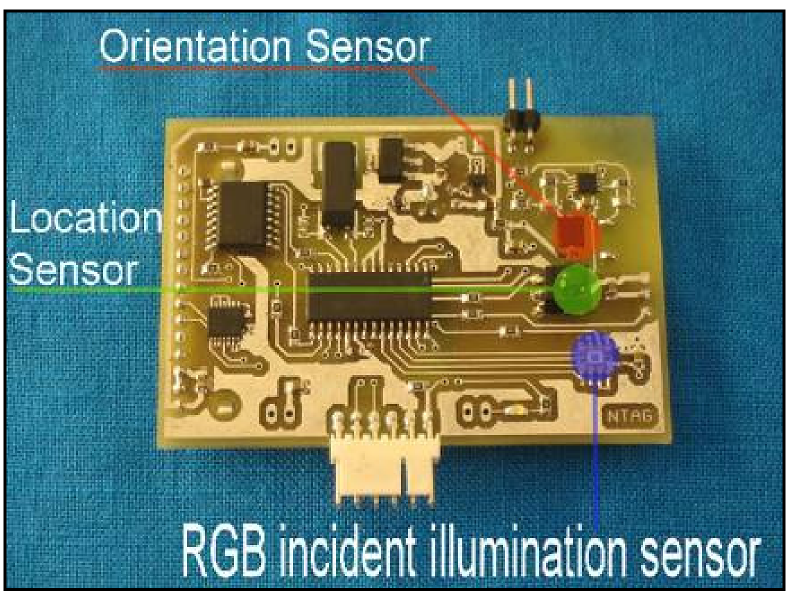


Figure 20. Top: co-linear stacked space-labeling projector Bottom: compound receiver tag.

7: Foldable Interactive Surfaces

Since we are projecting on passive surfaces, the materials do not need to be either rigid and or rectilinear. They can be flexible and bendable. Many of the displays we see in hand-held devices today are small LCD displays of fixed shape and size. In this respect, they are insensitive to the user desires and the needs of an application. Ideally, we would like displays that we can dynamically reshape or resize to suit our desired usage, similar to the way we might read a newspaper, or simply so that we are able to fit a large display into our pocket. In this section, I explore this concept of interactive foldable displays and create a number of working prototypes.

Emerging technologies such as electronic paper and organic light emitting diode (OLED) displays are expected to provide some degree of flexibility. However, current prototypes remain quite rigid and are typically rectilinear. This prevents them from becoming truly foldable in the sense that we think of paper as

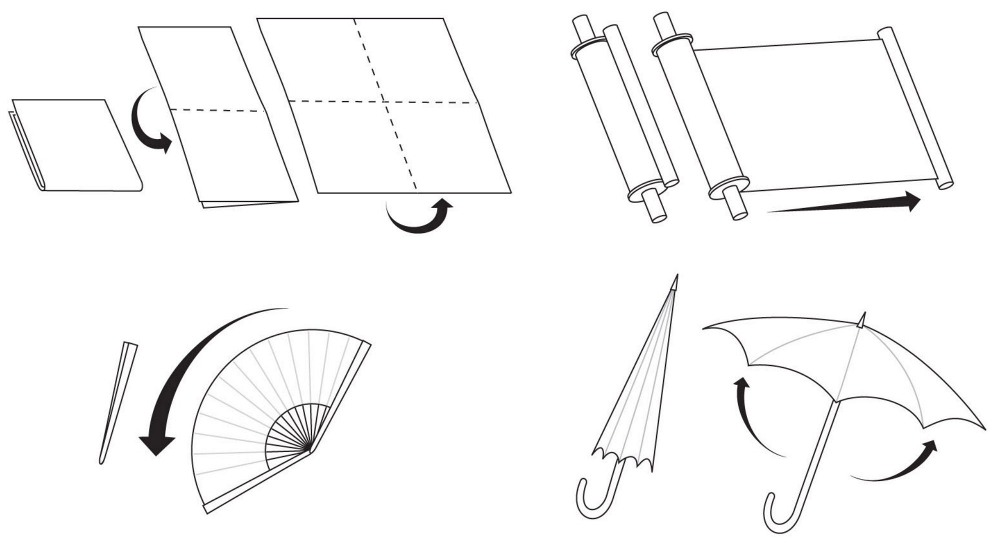


Figure 21. Foldable display shapes (left to right): newspaper, scroll, fan, and umbrella.

being foldable. Additionally, performing input on such flexible displays is an entirely separate technological hurdle. Projected spatial augmented reality explores the use of passive display surfaces whose appearance is augmented with image projection. This allows us to combine the flexibility and minimal weight of plain paper or fabric with the dynamic content capabilities of a computer display creating a coherent and fully functional user experience. This is somewhat similar to the approach used by PaperWindows, which explored interaction techniques with sheets of paper as if they were digital displays [Holman 2005]. PaperWindows focused mostly on flat paper interaction techniques in a tabletop environment and relied on a high-cost Vicon motion tracking system for location discovery.

Tracking

Due to the limited performance provided by the DMD discovery kit, this exploration utilized a camera based approach to simulate the behavior that would be achievable using high-speed projector-based tracking. In fact, using a camera-based approach introduced algorithmic complexities and geometric limitations that would not have been present using projector-based location discovery. The camera used in this exploration is an integrated motion tracking camera manufactured by PixArt technologies. This camera is most widely accessible in the Nintendo Wii remote. Camera is capable of tracking up to four points simultaneously with a spatial resolution of 1024 by 768 at 100Hz. While it is upsampling from a lower resolution sensor, the exact specifications remain confidential. However, when using bright light sources the upsampling is quite good. Four points are sufficient to match content onto planar surfaces or reconstruct the orientation of non-planar surfaces of known geometry [Horaud 1989]. Since IR blob tracking is done automatically in hardware on the remote, this is a very low-cost, easy to implement solution that provides high-resolution, low-latency tracking. However, camera based tracking has its limitations in terms of the number of distinct points that can be reliably tracked simultaneously, the inability to provide point identity, and requires manual calibration with the projected image for alignment. This can increase the complexity of supporting more complex foldable geometries. In contrast, projector-based tracking does not have such limitations. A large number of points can be tracked with unambiguous identification and without the need for calibration.

The LEDs in the display surface run for several hours using a small rechargeable battery pack. Since infrared LEDs emit non-visible light, the LEDs appear as small black dots 5mm in diameter. The LEDs can also be placed beneath a translucent surface to hide their visual presence entirely.

Foldable Shapes

In this section, we will present four foldable display designs. This is, of course, not an exhaustive list. However, we believe they present a number of expansion and collapsing behaviors likely to be used in a typical foldable display. For each description, please refer to Figure 21 for an illustration and Figure 22 for images of the working prototypes.

Newspaper

One of the most common formats in which we interact with large sheets of printed material is a typical newspaper. Sometimes referred to as the broadsheet format, these large sheets of paper are folded in half vertically and then again horizontally allowing a variable visual area ratio of 4 to 1. Additional folds can be added to further increase the magnitude of variability in visual area. In our prototype, we use two folds to support viewing half a page up to two full pages side-by-side. The user can gracefully increase or decrease the viewing area simply by unfolding or folding the display. The presence or absence of tracking points cues the computer to which faces are currently visible. The image is warped for each face of the surface such that when projected, the content appears undistorted even if the surface is not held perfectly flat.

Scroll

While less common today, possibly due to the mechanical support that must accompany each document, large printed material was once transported and viewed in the format of scrolls that could be unrolled. This allowed individuals to not only customize the amount of visible area, but also the location of that area within a long document – hence the concept of “scrolling” a window in a typical GUI environment. By creating a digital display scroll we can change the size and aspect ratio of the viewable area quickly and easily exposing more of the application content. This design can also be collapsed into a relatively small form factor for storage.

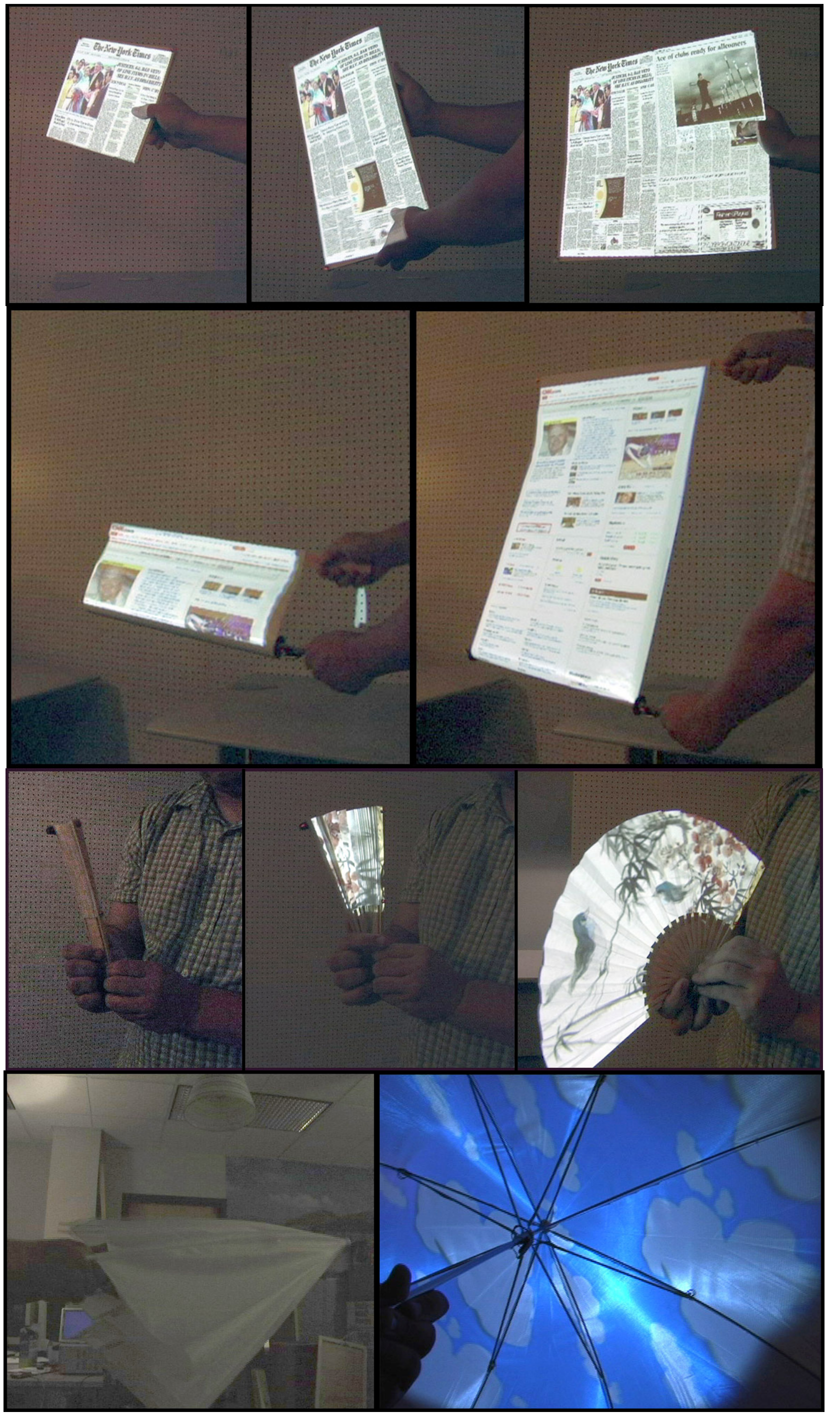


Figure 22. Foldable display prototypes at various stages of expansion (top to bottom): newspaper, scroll, fan, and umbrella

Fan

Folding fans are perhaps one of the best examples of a device that must be very large in surface area to be effective, usable by one hand so that the other hand is free to perform a task, and collapsible for easy storage in a pocket. Coincidentally, these properties are also desirable in a mobile display technology. As a result, this design may be one of the most practical for foldable displays in a mobile scenario. In this prototype the ratio of display area from a fully expanded to fully collapsed configuration is approximately 20 to 1 ranging from 100 square inches to a small strip. The elongated strip can be used to display status messages, rolling text, or progress bars similar to a portable music player. Some folding fans designs allow full 360 degree expansion creating a circular display area. The fan format can either be used in full or partial expansion to vary the amount of screen area desired. While we used a pleated folding fan for this prototype, folding fans can also be composed of parallel slats resulting in a nearly planar display surface minimizing distortion.

Umbrella

Another common example of expanding and collapsing a large surface is a parasol or umbrella. These surfaces can frequently be operated by one hand using spring loaded designs and can produce a very large surface area very quickly. Depending on the culture of origin and intended purpose, umbrella and parasol designs vary from parabolic bell shapes, to conical, to nearly planar. Distortions due to non-planar surfaces can be compensated for if the geometry is known before hand. An umbrella design may perhaps not be the most ideally suited shape for interactivity due to the central perpendicular column of the handle. However, the surface area change ratio is very dramatic making it potentially attractive for certain applications and lends itself to rotational input. The handle also provides an optically adequate location for a projection and tracking device for true mobility [Hashimoto 2006].

Orientation Sensitivity

Since we have tracking data of points on the surface for the purposes of projection, we can also use that orientation information to trigger different behaviors in the display surface. The following behaviors are illustrated in Figure 23, and images of the working prototypes are shown in Figure 24.

While we can display content on one side of the foldable surface, we can also detect when the display has been flipped based on the visibility of LEDs and motion modeling. This allows us to create double-sided display surfaces by projecting different content on each side. Flipping the surface in different directions can trigger different behaviors. For example, flipping left and right may switch between documents. Continuously flipping in one direction may step forward or backward through a sequence pages while flipping up and down might take the user to the table of contents and the index respectively [Chen 2007]. It is also possible to react to more subtle tilting movement of the surface altering the view depending on the angle at which is it held. We refer to this as a simulated lenticular display. In a multi-user tabletop scenario, this tilting behavior may correspond to the implicit privacy use of the document. For example, a display placed flat on the table can be considered *public* since it is visible to all users while picking up the display and tilting it such that only you are able to see it can trigger a *private* display state. Similarly, tilting the display away such that everyone except for you can see the content can trigger an *excluded* state possibly used for presentations or game scenarios when other players know something about you that you do not.

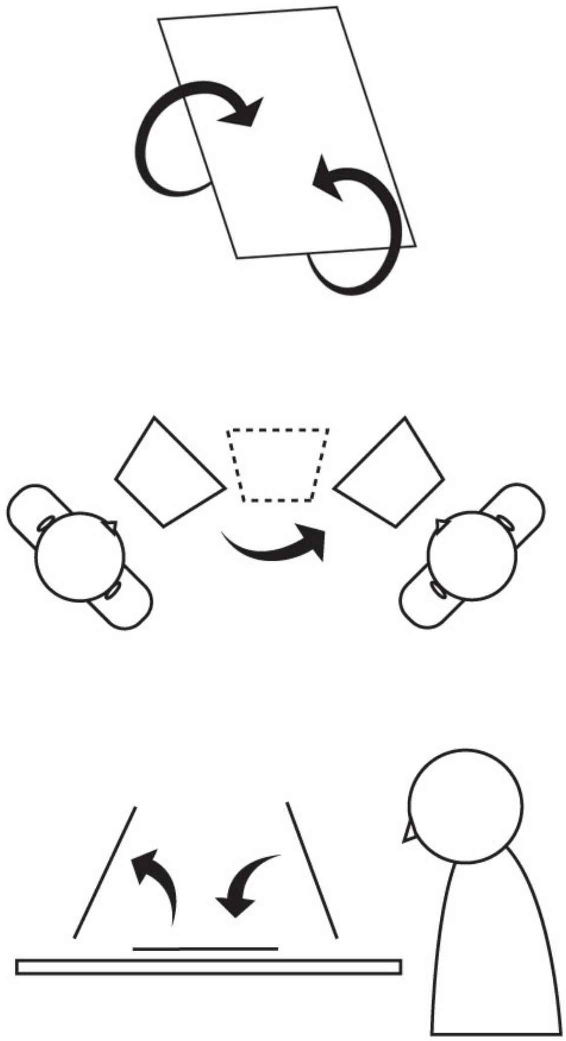


Figure 23. Orientation sensitivity behaviors (left to right): double-sided display surfaces can react differently depending on the direction they are flipped, simulated lenticular can change the document view depending on the angle of viewing in a hand-held display, or a tabletop scenario where tilt angle may correspond to different privacy states: *private, public, excluded.*

In addition to reacting to vertical tilting, we can also respond to horizontal rotation allowing the system to be aware if a display is being used by only one person, is being shared, or has been passed to another user. This may be useful if

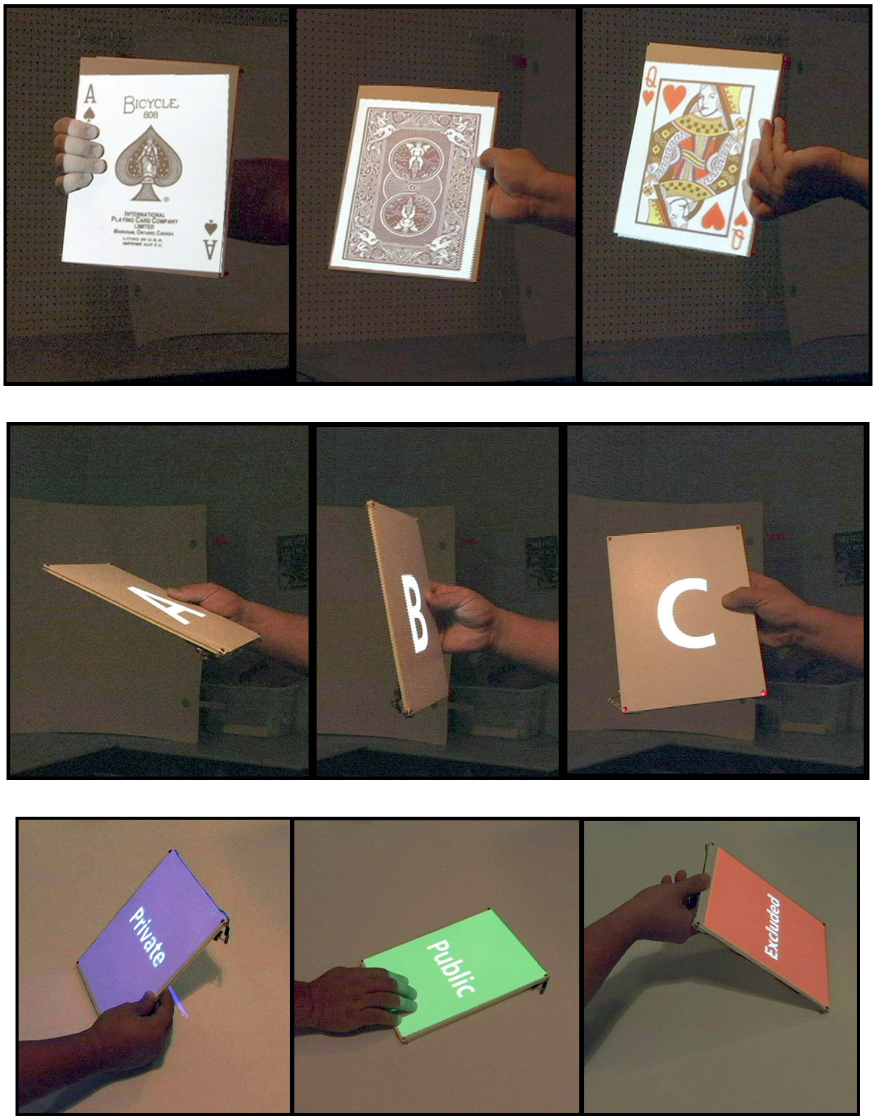


Figure 24. Orientation sensitivity prototypes: (top) double-sided flip direction, (middle) vertical and horizontal simulated lenticular in a hand-held surface, and (bottom) a tabletop scenario where tilt correlates to different privacy states.

partners are working on a shared task but have differing specialties or interests warranting different views of the material. The shared state would contain a summary or transitional view helpful in supporting communication between partners.

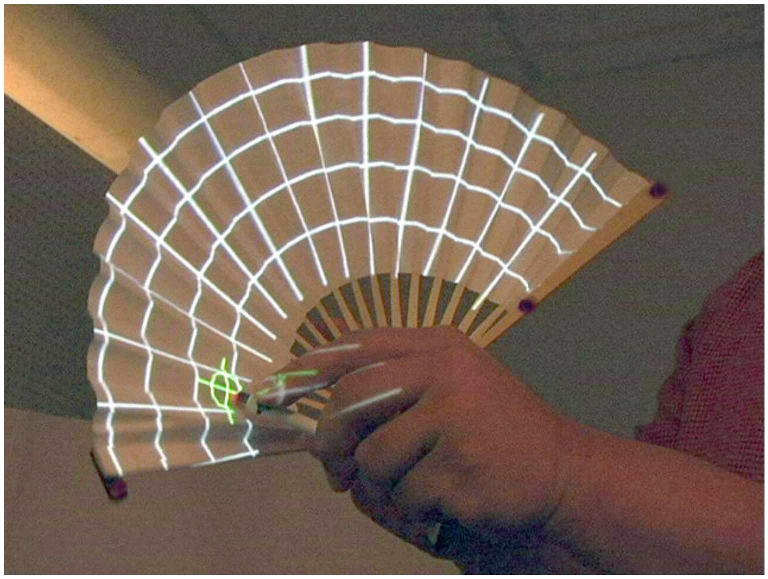


Figure 25. Interactive foldable display

While it would be possible to have the system react to an additional degree of rotational freedom as well as three degrees of translational movement, having such complex display behavior would be highly application dependent. For example, six degree-of-freedom display tracking would be appropriate for creating a view portal into a virtual 3D environment but unnecessary for many 2D GUIs.

Interactivity

By tracking additional dots over those embedded into the display surface, we can track a stylus for input shown in Figure 25. We add a button to the stylus to activate the LED providing a passive method of detecting clicking and dragging. The additional point only appears during mouse down events. This technique provides an easy way to obtain interactivity on all of the surfaces described including their double-sided variants. While it is possible to support multiple cursors, using a camera based approach may be difficult due to segmentation reasons. However, projector-based tracking would allow many cursors to be used simultaneously without ambiguity.

10: Comparison to Computer Vision Approaches

As described in the *Background and Approach* section, exploration in the area of location-sensitive projection has generally relied on manual alignment in static applications or an external tracking technology for interactive applications. Previous approaches toward automating this registration with the surface or calibration with the tracking system have predominately focused on computer vision techniques. Since projected displays are a visible medium and share many geometric similarities to a camera, computer vision is an attractive approach to the problem. In fact, a camera-projector pair provides a number of very sophisticated capabilities such as detailed range finding, radiometric compensation or enhancement, and passive input sensing. While particularly well suited for camera-projector pairs, these applications lie outside of the scope of this discussion. This discussion applies only to the uses of computer vision techniques to identify and track the locations of instrumented objects. While this may seem to severely limit the relevance of this discussion at first, very few robust vision-

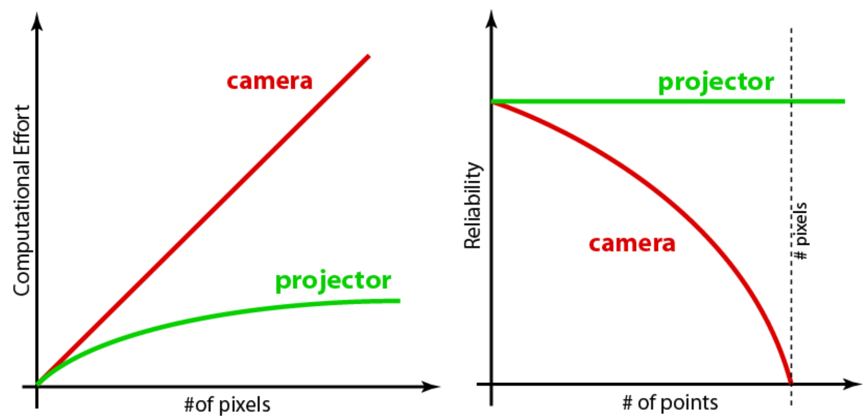


Figure 26. Scalability comparisons between camera-based and projector-based tracking for computational effort (left) and multi-point tracking (right).

based tracking systems available today fall outside of this scope. Computer vision provides the theoretical potential to identify and track un-instrumented objects passively. However, in practice, reliable tracking is only possible with objects that are highly regular in features or shape, tagged with high-contrast markers, or respond to specialized illumination. General object recognition and tracking remains an open research problem in computer vision.

One of the primary differences between camera and projector-based tracking is that camera-based tracking is able to utilize passive markers and features. Having unique colors, shapes, or visible markings are sufficient to identify and locate objects. No active tags or markers on the objects are necessary where as projector-based tracking requires active sensors and feedback communication with the projection system. This is admittedly a significant disadvantage of projector-based tracking. However, there only a few domains where completely passive computer vision systems have been widely successful in using passive markers such as barcode scanning, chroma keying, face tracking, and optical character recognition. Even in these domains, the recognition algorithms must be heavily crafted toward each application and a large number of constraints are placed on the image capturing process. For example, a face tracking system will not readily work for tracking cars on a highway. Performance is highly dependent on environmental conditions. Insufficient background separation, uncontrolled illumination, orientation, surface reflectivity and partial visibility all can severely impact recognition performance. Reliable recognition and accurate location tracking of passive features is only possible in heavily controlled circumstances. Passive markers also suffer from two other limitations: 1) they must be large enough to be seen by the camera which increases the minimum physical size of objects and decreases the maximum number of objects that can be simultaneously tracked and 2) markers are typically also visible to human observers which can interfere with the application.

To work around these performance issues and size limitations, active tags using light emitting diodes (LED) are often employed to provide a distinctly colored, bright, point light source to the camera. This significantly eases the segmentation and location tracking of small markers. When very high performance is needed from a camera-based motion tracking systems, this is the approach typically employed. This is also the most comparable to projector-based tracking both in terms of instrumentation and tracking performance. Optically, projector tracking is the exact reverse. Instead of several active markers emitting light toward a central receiver, we have several active markers receiving light from a central emitter.

In scenarios where instrumentation of the surface or object is acceptable, projector-based tracking provides a number of advantages over camera-based approaches with respect to resolution, speed, point count, and point identity. Since computer vision algorithms often have to look at every pixel in the image to determine if it belongs to an object of interest, the computational effort typically scales linearly with camera resolution. This corresponds to a linear increase in tracking latency. Projector-based tracking uses Gray-coded binary patterns to encode each pixel in the image and the number of patterns necessary to do so has a logarithmic relationship to screen resolution, Figure 26. Thus, projector-based tracking scales much more easily to very high resolutions with far less impact on computational effort and thus overall speed performance. For example, only 60 patterns would be necessary to resolve the entire continental United States to millimeter accuracy. The limiting factor in speed is determined by how quickly the projection display technology can present the binary patterns. DMD technology has already demonstrated the ability to present well over 50,000 binary patterns per second. Thus, extremely high-speed high-resolution motion tracking is achievable using contemporary display technology.

Since each sensor is responsible for decoding its own location, the already minimal computational effort is distributed at each sensor and unambiguous sensor identity is available with the returned data. This allows the projector to simply broadcast the location data in an open-loop manner independent of sensor count. In contrast, computer vision approaches suffer from point ambiguity particularly as large numbers of points enter the scene. While some identity can be transmitted using blinking tags, the speed at which this data can be transmitted is limited by the frame rate of the camera and does not address the ambiguity issue when large numbers of points are visible. The ability of computer vision techniques to resolve each point reliably is significantly reduced as the number of simultaneous points approaches the number of pixels in the camera image, see Figure 26. Thus, projector-based tracking will provide much more reliable performance in applications requiring unambiguous identity and/or large numbers of points.

Even when comparing against active markers for camera-based tracking, projector-based tracking is still more robust against background complexity, surface reflectivity, and irregular illumination. Interference from light sources which mimic the light from active markers may be common and confuse camera systems. However, interfering light sources which mimic the Gray-code patterns of the projector are far less likely, especially in the presence of predetermined start and stop patterns. Such sources would likely generate erroneous data incongruous with the data reported from other sensors allowing an easy method of detecting invalid data. Additionally, projector-based tracking lends itself to the use of modulated light transmission, such as remote control infrared communication, much more than current camera-based technologies. This further increases the robustness to light inference. If a future camera design incorporates integrated high-speed demodulation of light sources, this could reduce this performance gap dramatically increasing the performance attainable from a camera-based tracking system. But, such a design is currently only theoretical.

Furthermore, if the application combines the location data with projected imagery, using the projector to provide both location discovery and application content is an inherent simplification that eliminates the need for an external tracking technology and any related calibration process. While not solely related to performance scalability, it is a noteworthy advantage of this approach. A tabulated summary of the feature comparisons presented in this section are presented in Figure 27.

It is worth mentioning that a hybrid approach could be created by placing a camera adjacent to the projector and retro-reflective markers on the target surface. The same Gray-code patterns could be projected allowing the camera to see flashes from each reflective marker representing their location. This would ease the segmentation and resolution limitations of computer vision as well as minimize surface instrumentation. However, retro-reflective makers would have a visible instantiation, would be subject a minimum size to be effective, and would not provide marker identity.

	Camera Tracking	**Projector Tracking**
Resolution Scalability	Linear: O(n)	Logarithmic: O(log(n))
Sensor Count Scalability	Resolution limited	Physically/spatially limited
Data Bandwidth	pixels * depth * frame rate * # cameras	log2(pixels) * tracking rate * # sensors
Marker Complexity	Passive or active emitter	Active sensor with transmitter
Marker Size	Resolution limited	Component size limited
Marker Identity	Unavailable or must be transmitted within limits of camera frame rate	Inherent to each sensor
Background Complexity	Must be easily separable from target	Insensitive
Motion Complexity	Algorithmic tuning for rotation and distance variation	Insensitive
Surface Complexity	Limited with respect to the visible features and camera resolution	Broader support for complex and shallow projection angles
Surface Material	Limited to prevent interference during vision recognition	Insensitive
Dynamic & Shadowed Illumination	Very sensitive – variations in illumination create interference with marker recognition	Fairly insensitive – local illumination sensing and projected modulation reduce likelihood of interference
Modulated Light	Not available with current technology	Currently available at low-cost
Application Content	Requires external display device and calibration	Can be provided by the projector and is calibration free

Figure 27. Tabulated comparison of features between Camera tracking and Projector tracking

11: Error Modeling of Moving Sensors

Previous work exploring the use of Gray-coded binary patterns for spatial encoding have made the assumption that the sampling point remains stationary during the presentation of the entire sequence. This assumption is valid for some applications using low-speed, location discovery described earlier in this dissertation. However, it is not valid for the interactive applications involving moving surfaces or moving styli. Thus, analysis of the error introduced in encoding behavior of a non-stationary sampling position would be useful for understanding the limitations of tracking performance.

Since attempting to physically move the sensor at different rates and empirically measuring the tracking performance becomes impractical at higher speeds, I can use a mathematical simulation of sampling data that would be obtained from moving a sensor along a known path through a presentation of Gray-coded binary patterns. This simulation allows us to vary velocity, origin, path geometry, and bit depth providing a structured manner to analyze tracking performance. A simulation was created in Matlab of a sensor moving through a sequential presentation of Gray-coded binary patterns, starting with the highest order bit.

In the first simulation, I analyze linear sensor movement. The velocity of the sensor was gradually increased from 0% to 100% of the pattern width per digitization period. For example, a velocity of 0% corresponds to a stationary sensor and a velocity of 100% indicates that the sensor moved across the entire projection area within one Gray-code patterning time, or one digitization period. The origin of movement was randomized to obtain more realistic performance measures but was bounded to ensure the sensor would not exit the digitization region during the digitization period. Offset error was defined as the decoded simulated sensor data minus the midpoint of position of the motion (ideal) divided by the total pattern width. The choice of the ideal position is somewhat subjective, being the midpoint, starting point, or end point of motion. But in this case, this did not affect the results significantly. The patterning depth was 32-bits. The offset errors for the simulation can be seen in Figure 28. As expected, the size of the error increases proportionately with velocity. The normalized view, which divides the error by the size of the sensor movement, shows a fairly consistent offset error of approximately 40%. The skewed distribution of points in the normalized plot is an artifact of the normalization process from the original data samples. An interesting property of the error is that the variance of the offset is quite small. It is consistent with the direction motion. Thus for a given velocity, the expected error falls into a relatively small window that could be modeled and then compensated in software, further increasing accuracy. The striation in the data points occurs as the velocity of the sensor begins to shear across a higher order Gray-code pattern causing an additional bit error.

In the second simulation, I analyze sinusoidal sensor movement which may be slightly more representative of real world motions in interactive applications. The phase and frequency were varied to cover a wide range possible motion. Again, the offset error is defined as the distance between the decoded location and the midpoint of the motion (t=0.5) divided by the pattern width. The patterning depth was 32-bits. The amplitude of the movement is one pattern width. A sensor velocity of 1 represents when the frequency of movement reaches 1 cycle per digitization period. The offset errors are shown in Figure 29. As expected from the previous simulation, the error increases proportionately with sensor velocity. As before, the mean motion error is approximately half of the movement in a period. Though not reflected in this plot, at lower frequencies the data exhibits the same low-variance offset as was shown in the linear movement simulation. However, once the sensor movement period exceeds the digitization period, it moves back and forth too quickly for the Gray-code patterns to be meaningful and the decoded location becomes random.

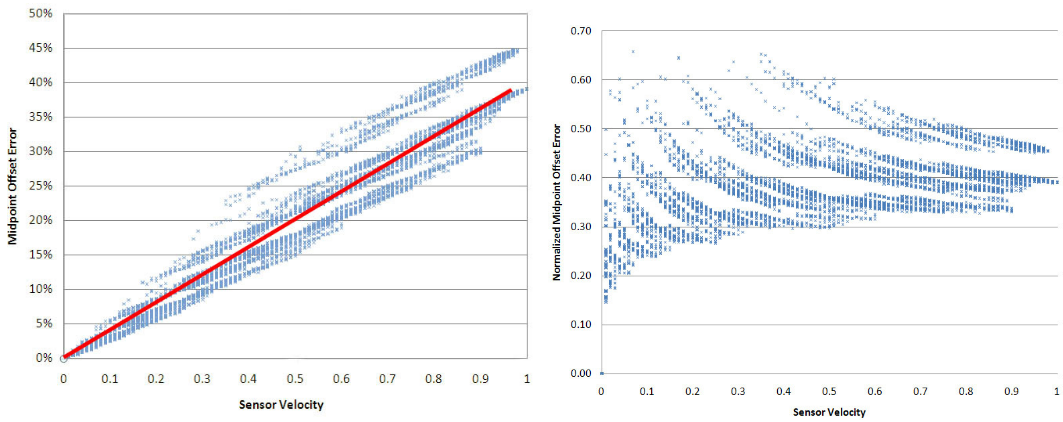


Figure 28. Offset encoding error of a sensor moving linearly in a Gray-coded projection area relative to the total pattern width (left) and normalized to the size of the movement (right).

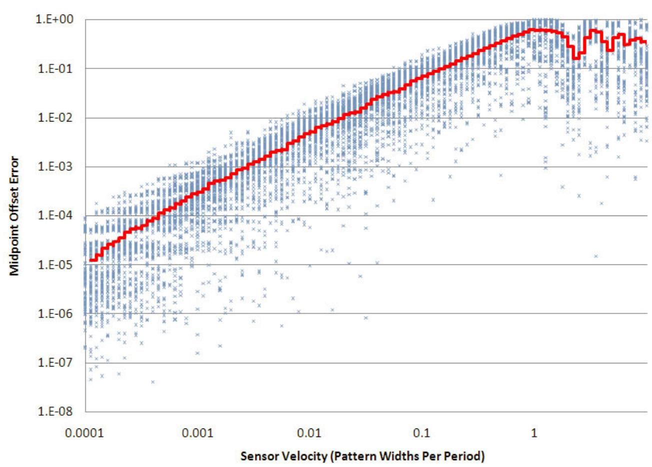


Figure 29. Offset encoding error of a sensor moving sinusoidally in a Gray-coded projection area.

12: Summary and Conclusion

Research work within the human-computer interaction and computer graphics domains have explored the powerful capability projection technology has to transform the functionality of surfaces throughout our environment. These applications include interactive white boards, computational worktables, large scale immersive environments, spatially augmented reality, variable resolution displays, and hand-held projection. However, executing these concepts remains complex and costly largely due to the precise calibration, alignment, and tracking issues involved in generating a compelling interactive experience. As a result, these ideas have had difficulty in achieving widespread adoption outside of the labs that originally created them.

While previous attempts to solve this problem using computer vision techniques have had some success, the projector-based location discovery and tracking system presented in this dissertation offers an alternative solution that provides a number of significant advantages. This approach discovers the location of photo sensors placed in the projection area using a series of structured light patterns that uniquely encode each pixel in a projection area. The result is a robust, scalable, low-cost, high-speed method for location discovery and positional input using a projection system.

This work covered three implementations that explored design considerations of both the sensors and projector, in particular, methods of increasing speed and reducing the visibility of tracking patterns. The first implementation explored slow-speed applications of projector-based location discovery using an unmodified projector. These applications include screen-calibration, multi-projector stitching or alignment, automatic touch calibration, and the Shader Lamps methods. The second implementation explored techniques for increasing tracking speed through the use of incremental tracking patterns and a reduction in pattern perceptibility through frequency modulated transmission of Gray-code patterns from a modified projector. This work also included algorithmic considerations of tracking pattern behavior and surface modeling to accommodate sensor motion and occlusion. The applications enabled by this prototype include tablet-PC simulation, physical magic lenses, moveable focus plus context displays, and tangible input devices. The third implementation was a proof-of-concept hybrid visible and infrared light projector able to provide both full screen location discovery patterns and full screen application content. This allowed a seamless experience for interactive applications such as multi-stylus input, distant pointing on non-planar discontinuous surfaces, and multi-display identification. The last component of this dissertation explored concept applications that would be enable by high-speed projection such as interactive foldable displays, multi-sided displays, and simulated lenticular lenses. Throughout this work, each step of added capability was demonstrated within relevant application concepts. In total, over fifteen classes of projector applications are either simplified or improved upon through the use of this technique. In cases where performance was limited due to access to development resources, expected performance was simulated using an external tracking device.

When using a projection technology to track the location of light sensors and display visual application content on interactive surfaces, occlusion becomes a key issue. Like any projection system, hands and bodies positioned between the projector and the surface will result in shadows. In these areas, not only will the application content be lost but also the tracking information. Thus for an interactive system where hands and body are likely to be in close proximity to the surface, it is beneficial to consider projector placements that will minimize occlusion. Example solutions for vertical surfaces would include shallow-front projection, which uses either specialized lenses or curved mirrors to reduce image throw distance reducing the amount of shadow cast by a proximate object. Alternatively, multiple projectors at different angles, similar to [Sukthankar, 2001], reduces the impact of occlusions through redundancy. If a rear projected configuration is possible, this removes the user from the optical path avoiding the occlusion problem entirely providing an uninterrupted interaction experience. In horizontal table-top configurations, severe occlusions tend to be less common due to the natural orientations of the hand and body. A stylus would be typically held at an angle such that application content is still visible at the pen tip. This would also mean that tracking data is also available at the tip. However, if higher quality experiences are required, similar strategies of shallow, multiple, or rear projection configurations could be employed.

Since this work first began in 2004, this technique for projector-based location discovery and tracking has been licensed for use within two commercial products with nearly a dozen other inquiries from companies exploring applications of projection technology in advertizing, immersive environments, stereoscopic movies, theatrical lighting production, interactive art, and televisions. While there are several near term applications of this technique relevant today, many of the more exciting interactive projection applications are still emerging. Over the next 5-10 years, the need to combine location tracking with image projection will become increasingly essential to computing experiences. Within the research community, over 20 publications related to motion tracking, interactive paper, immersive displays, ubiquitous computing, and augmented reality have referenced this work or its derivatives. A larger number of individuals are able to explore novel concepts in this domain furthering the state of research using the capabilities provided by the technique presented in this thesis.

Integrated tracking of light sensors in projection systems enables a large number of interactive display applications and reduces the cost of executing existing research concepts. As projection technology continues to evolve in terms of brightness, size, resolution, and affordability, the ability to augment the appearance of objects and surfaces in our environment will become increasingly ubiquitous. This may come in the form of micro projection in our mobile devices, or as computing spaces where image projection is provided as a service, or perhaps in some other form. In these scenarios, it would then be possible to simply carry within our pockets a surface for reflecting an image rather than the display technology itself. This material could be light-weight, flexible, collapsible, or even disposable radically changing the manner in which we live with displays. Rather than carry the bulky rigid rectilinear surfaces we tolerate today, it would be possible to summon interactive displays of any shape or size on arbitrary surfaces in our environment at our convenience to suit the needs of the application. These surfaces may vary from a small handheld digital map unfolded from a pocket, to large conference table augmented workbench supporting multiple people simultaneously, a community billboard-like status display in a busy workplace, or a creating a temporary digital note pad on the bus stop bench. While this does not quite capture the mid-air holographic displays portrayed by science fiction films, it would be capable of providing a very similar experience.

Future Work

The interactive prototypes presented in this dissertation were limited in terms of tracking speed performance relative to potential performance offered by today's commercial Digital Micro-mirror Device projectors. Working with Texas Instruments to develop a full speed, high performance prototype or integrating this technique in the upcoming generations of DLP projectors is definitely of significant interest. Some progress has been made with technical visits and conversations with the engineering staff, but this is ultimately a business decision. Such inherent support for motion tracking with projection system would greatly benefit the progress of research in this domain and the development of applications.

Once the hardware tracking capabilities are readily available, there becomes a need for software toolkits that can take advantage of the location data. Commercial hardware platforms such as the Microsoft Surface or the Perceptive Pixel display provide software APIs that take the first steps toward enabling rich geometric manipulation of application widgets and interaction objects. However, these toolkits are generally limited to two-dimensional manipulations of data. When projecting onto moving surfaces in 3 dimensions, several additional degrees of freedom are added to the movement and distortion of the displayed content and the captured input on the surfaces. Similarly, the number of control points may be quite large. For our own development purposes, we have made small toolkit written in Microsoft C# and DirectX for rendering content onto moveable projected displays tracked using four points. However, this toolkit is quite limited allowing only simple 2D and 3D geometry rendering. There is room for significant improvement and further development. Such a toolkit would be a necessary component in creating a generally usable platform for spatially augmented reality displays.

Conclusion

By unifying the image projection and location tracking technologies, many of the difficult calibration and alignment issues related to interactive projection and projected spatial augmented reality applications can be eliminated simplifying their implementation and execution. Using either high-speed projection or hybrid visible and infrared light projection, a single calibration-free device can perform invisible location tracking of photosensitive input devices while simultaneously presenting visible application content. In this dissertation, I have presented a detailed description of the projector-based location discovery and tracking technique, a description of three prototype implementations, and a demonstration the effectiveness of this simplification by re-implementing, and in some cases improving upon, several location-sensitive projector applications.

Bibliography

Badyopadhyay, D., Raskar, R., Fuchs, H., “Dynamic Shader Lamps: Painting on Real Objects.” IEEE ACM Int. Symposium on Augmented Reality (ISAR), New York, 2001.

Bellcore. The VideoWindow Teleconferencing Service Model. Special Report SR-ARH-001424. Morristown, NJ: Bell Communications Research, 1989.

Baar, J van.; Willwacher, T.; Rao, S.; Raskar, R., "Seamless Multi-Projector Display on Curved Screens", Eurographics Workshop on Virtual Environments (EGVE), ISBN: 3-905673-00-2, pp. 281-286, May 2003

Baudisch, P., Good, N., and Stewart, P. “Focus Plus Context Screens: Combining Display Technology with Visualization Techniques.” In proceedings of ACM UIST ’01, Orlando, FL, November 2001.

Ben-Joseph, E., Ishii, H., Underkoffler, J., Piper, B., and Yeung, L. “Urban Simulation and the Luminous Planning Table: Bridging the Gap between the Digital and the Tangible.” In Journal of Planning in Education and Research, Volume 21, pp.195-202. 2000.

Bier, E., et al., "Toolglass and Magic Lenses: The Seethrough Interface." In proceedings of ACM SIGGRAPH 1993, Anaheim, CA., Aug. 1993,

Bimber, O., Raskar, R. "Spatial Augmented Reality: Merging Real and Virtual Worlds" Publisher: A K Peters, Ltd. (July 31, 2005)

Chen, N., et al. “Enhancing Document Navigation Tasks With a Dual Display Electronic Reader”, Proceedings of ACM Symposium on User Interface Software and Technology, October 2007.

Cotting, D.; Naef, M.; Gross, M.; Fuchs, H., "Embedding Imperceptible Patterns into Projected Images for Simultaneous Acquisition and Display." Third IEEE and ACM International Symposium on Mixed and Augmented Reality, 02-05 Nov. 2004 Page(s):100 109.

Cruz-Neira, Carolina, Daniel J. Sandin, and Thomas A. DeFanti. “Surround-Screen Projection-Based Virtual Reality: The Design and Implementation of the CAVE,” Computer Graphics, SIGGRAPH 1993.

DePiero, F.W., and Trivedi, M.M., “3-D Computer Vision using Structured Light: Design, Calibration, and Implementation Issues,” Advances in Computers(43), 1996, Academic Press, pp.243-278

Dietz, P., Leigh, D., “DiamondTouch: a multi-user touch technology”, Proceedings of the 14th annual ACM symposium on User interface software and technology, November 2001, Orlando, Florida

Gardner, M. "The binary Gray code", in Knotted donuts and other mathematical entertainments, F. H. Freeman and Co., NY, 1986

Gibbs, Simon, Constantin Arapis and Christian J. Breiteneder. “TELEPORT-Towards Immersive Copresence,” ACM Multimedia Systems Journal, 1998.

Gray, F., “Pulse Code Communications” United States Patent 2632058, issued 1953.

Han, J., “Low-cost multi-touch sensing through frustrated total internal reflection”, Proceedings of the 18th annual ACM symposium on User interface software and technology, October 2005, Orlando, Florida

Hashimoto, S., Matsumoto, T., "Pileus: The Interenet Umbrella". http://www.pileus.net 2006

Hinckley, K., “Bumping Objects Together as a Semantically Rich Way of Forming Connections between Ubiquitous Devices.” UbiComp 2003

Hinckley, K., Ramos, G., Guimbretiere, F., Baudisch, P., Smith, M., "Stitching: Pen Gestures that Span Multiple Displays." In Proceedings of ACM Advanced Visual Interfaces. 2004

Holman, D., Vertegall, R., Altosaar, M., Troje, N., Johns, D., "Paper windows: interaction techniques for digital paper" Proceedings of the SIGCHI conference on the human Factors in computing systems, 2005.

Horaud, R. Conio, B. Leboulleux, O. Lacolle, B.," "An analytic solution for the perspective 4-point problem" Proceedings of Computer Vision and Pattern Recognition, June 1989.

Hunter, T. N., Hagmeier, K., “3M Revolutionizes Mobile Displays: Miniature Projection Technology Available Now” 2008, http://www.3m.com/mpro/techrelease.pdf

Inokuchi, S., Sato, K., Matsuda, F., “Range imaging system for 3-D object recognition” Proceedings of the International Conference on Pattern Recognition, 1984, pp. 806–808.

Ishii, H., and B. Ullmer. "Tangible bits: towards seamless interfaces between people, bits and atoms." In proceedings of ACM CHI 1997.

Kato, H., Billinghurst, M. (1999) Marker Tracking and HMD Calibration for a video-based Augmented Reality Conferencing System. In Proceedings of the 2nd International Workshop on Augmented Reality (IWAR 99). October, San Francisco, USA.

Lane, R.G. Heath, J. "Innovations in infrared scene simulator design." AUTOTESTCON '98. IEEE Systems Readiness Technology Conference., 1998 IEEE

Lee, J., Dietz, P., Aminzade, D., and Hudson, S. “Automatic Projector Calibration using Embedded Light Sensors”, In Proceedings of ACM UIST ’04. October 2004.

Lee, J., Hudson, S., Summet, J., and Dietz, P. "Moveable Interactive Projected Displays Using Projector Based Tracking", Proceedings of the ACM Symposium on User Interface Software and Technology, October 2005.

Li, K. et al. “Building and Using a Scalable Display Wall System”. IEEE Computer Graphics and Applications, vol 20(4), pp 671-680, 2000.

Lionel M Ni, Yunhao Liu, Yiu Cho Lau, and Abhishek Patil, "LANDMARC: Indoor Location Sensing Using Active RFID", ACM Wireless Networks Volume 10, Issue 6, November 2004,

Matusik, W.; Pfister, H., "3D TV: A Scalable System for Real-Time Acquistion, Transmission and Autostereoscopic Display of Dynamic Scenes", ACM Transactions on Graphics (TOG) SIGGRAPH, ISSN: 0730-0301, Vol. 23, Issue 3, pp. 814-824, August 2004.

Microvision, Pico Projector, 2008 http://www.microvision.com/pico_projector_displays/index.html

Nguyen, D., and Canny, J. MultiView: Improving Trust in Group Video Conferencing through Spatial Faithfulness. Proceedings of the 2007 Conference on Human Factors in Computing Systems (CHI’07). San Jose, CA

Nii, H., Sugimoto, M., And Inami, M. "Smart light ultra high speed projector for spatial multiplexing optical transmission." Procams, 2005.

Pinhanez, C. “The Everywhere Displays Projector: A Device to Create Ubiquitous Graphical Interfaces.” In proceedings of Ubiquitous Computing 2001.

Piper, B., Ratti, C., and Ishii, H., “Illuminating clayA 3-D Tangible Interface for Landscape Analysis.” In proceedings of ACM CHI 2002.

Posdamer, J. L., Altschuler, M. D. ,”Surface measurement by space-encoded projected beam systems”, Computer Graphics and Image Processing 18 (1) (1982) 1–17.

Raskar, R., Fuchs, H. Welch, G., Lake, A., Cutts, M., “3D Talking Heads : Image Based Modeling at Interactive rate using Structured Light Projection,”University of North Carolina at Chapel Hill, Dept of Computer Science, Technical Report TR98-017, 1998

Raskar, R., Welch, G., Cutts, M., Lake, A., Stesin, L., and Fuchs, H. “The office of the future: A unified approach to image-based modeling and spatially immersive displays.” In proceedings of ACM SIGGRAPH '98. ACM Press, 1998.

Raskar, R., Brown, M., Yang, R., Chen, W., Welch, G., Towles, H., Seales, B., Fuchs, H. “Multi-Projector Displays using Camera-Based Registration.” IEEE Visualization 99, Oct 1999, San Francisco, CA

Raskar, R., Welch, G., and Low, K.L. “Shader Lamps: Animating real objects with image-based illumination.” In proceedings of Eurographics Workshop on Rendering, 2001.

Raskar, R.; Beardsley, P.A., "A Self-Correcting Projector", IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), ISSN: 1063-6919, Vol. 2, pp. 504-508, December 2001

Raskar, R.; van Baar, J.; Beardsley, P.A.; Willwacher, T.; Rao, S.; Forlines, C., "iLamps: Geometrically Aware and Self-Configuring Projectors", ACM Transactions on Graphics (TOG), ISSN: 0730-0301, Vol. 22, Issue 3, pp. 809-818, July 2003

Raskar, R.; Beardsley, P.; van Baar, J.; Wang, Y.; Dietz, P.H.; Lee, J.; Leigh, D.; Willwacher, T., “RFIG Lamps: Interacting with a Self-Describing World via Photosensing Wireless Tags and Projectors.” In proceedings of ACM SIGGRAPH. August 2004.

Rekimoto, J., Saitoh, M., "Augmented Surfaces: A Spatially Continuous Work Space for Hybrid Computing Environments." In proceedings of ACM SIGCHI 1999, pp.378-385.

Rekimoto, J., "SmartSkin: An Infrastructure for Freehand Manipulation on Interactive Surfaces" In Proceedings of ACM SIGCHI 2002

Rosenfeld, D., "Planar manipulator display" Computer Graphics and Applications, IEEE Volume 24, Issue 1

Salvi, J., Pages, J., Batlle. J., "Pattern Codification Strategies in Structured Light Systems." Pattern Recognition 37, April 2004

Streitz, N.A., Geibler, J., Holmer, T., Konomi, S., Müller-Tomfelde, C., Reischl, W., Rexroth, P., P. Seitz, R.,Steinmetz, "i-LAND: An interactive Landscape for Creativity and Innovation." In proceedings of ACM CHI'99.

Sugimoto, M., Kojima, M., Nakagawa, A., Kagotani, G., Nii, H., and Inami, M. "Augmented coliseum: An augmented game environment with small vehicles." SIGGRAPH2005 Emerging Technologies

Sukthankar, R., Cham, T., and Sukthankar, G. “Dynamic Shadow Elimination for Multi-Projector Displays.” Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2001.

Summet, J. et al. “Increasing the Usability of Virtual Rear Projection”. IEEE International Workshop on Projector-Camera Systems (PROCAMS) held in conjunction with the International Conference on Computer Vision (ICCV 2003), Nice, France, October 2003.

Summet, J., Sukthankar, R., “Tracking Locations of Moving Hand-held Displays Using Projected Light.” In proceedings of Pervasive 2005. Munich, Germany. pp 37-46

Ullmer, B., Ishii, H., “The metaDESK: models and prototypes for tangible user interfaces.” Proceedings of ACM UIST ’97. October 1997.

Wallace, G. et al. “Tools and Applications for Large-Scale Display Walls.” IEEE Computer Graphics and Applications, July 2005.

Ward, A., Jones, A., Hopper. A., "A New Location Technique for the Active Office." IEEE Personal Communications, Vol. 4, No. 5, October 1997.

Wilson, A. TouchLight: An Imaging Touch Screen and Display for Gesture-Based Interaction, International Conference on Multimodal Interfaces, 2004.

Wilson, A. PlayAnywhere: A Compact Tabletop Computer Vision System, Symposium on User Interface Software and Technology (UIST), 2005

Wittenberg, C. "Ultra-miniature projector: A high resolution, battery powered laser display", 2007. http://www.sid.org/chapters/ne/sid_ne_jan2007_wittenberg.pdf

Wood, J., Taber, R., Ramachandran, G., “Automatic keystone correction system and method” United States Patent 6877863. Issued April 2005.

Yoder, L. “The Digital Display Technology of the Future.” Texas Instrument white paper library, (www.dlp.com). Presented at Infocomm, 1997.

Yotsukura,T. et al.,"Hyper Mask - Talking Head Projected onto Real Object". The Visual Computer, vol. 18(2):111-120, 2002.