Projector-Based Location Discovery and Tracking

Johnny Chung Lee
Carnegie Mellon University
5000 Forbes Ave, Pittsburgh, PA 15213
johnny@cs.cmu.edu

ABSTRACT

While a number of projects within the computer graphics, computer vision, and human-computer interaction community have explored the powerful ability for projected light to create illusions that can reshape our perception and interaction with surfaces in the environment, very few of these systems have had success in terms of commercial and consumer adoption. Often these systems require expert knowledge to perform system setup and calibration between the projected images and the physical surfaces to make these illusions effective. In this thesis work, 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. I present three evolutionary prototype systems of increasing capability and demonstrate the applications enabled and simplified during each stage of development. By unifying the image projection and location tracking technology into a single device I can greatly simplify the implementation of previous systems and enable interaction capabilities which are difficult or impossible using alternative approaches.

Author Keywords

Infrared Projection, Projector-based Tracking, Augmented Reality, Simulated Displays, Physical Interaction

ACM Classification Keywords

H5.2 [Information interfaces and presentation]: User Interfaces. H5.1 [Multimedia Information Systems]: Augmented Reality.

INTRODUCTION

While the most common uses of projectors today are limited to simple tasks such as business presentations and entertainment, there exists a large body of research work in the field of human-computer interaction, computer vision, and computer graphics that have explored how projectors can provide sophisticated capabilities such as 3D scanning [Dep96], office augmentation [Ras98], display simulation [Lee05], contextually located displays [Pin01], appearance augmentation [Ras01], and augmented worktables [Rek99,Ulm97,Wil05]. While far from a complete list, these projects demonstrate the significant value projectors provide as a data acquisition device or interactive tool rather than simply providing a large passive display. However to obtain these sophisticated capabilities, these systems have typically relied on coupling the projector with external tracking technologies for obtaining user input or discovering the locations of objects. As a result, this adds significant costs both in terms of equipment and developer expertise necessary to integrate and calibrate multiple discrete technologies into a single coherent user experience. This cost and complexity has limited the adoption of such systems outside the research lab that initially created them.

In this work, I present a technique for using the projector itself to obtain user input and object tracking in a calibration-free manner eliminating much of this cost and complexity. Additionally, by unifying the image projection and location tracking technology into a single device I can also enable interaction capabilities which are difficult or impossible using previous techniques. To describe this work, I will detail the applications and implementations of three prototype systems that I created - each of which builds upon the capabilities of the previous prototype. The first prototype is a low-speed location discovery system using a standard commercial projector. The second prototype uses a modified projector to increase the speed of location discovery sufficient to support motion tracking and reduces the perceptibility of the projected patterns. The third implementation is a hybrid projector capable of emitting both visible images as well as invisible infrared patterns at high speed for the simultaneous presentation of application content and motion tracking of objects and user input.

APPROACH

To transform a projector into a location discovery device, we must take advantage of the projector's capability to deliver a time-varying sequence of light energy to different locations in space. By selecting a set of structured light patterns, we can uniquely encode each pixel in the projector's screen space. Light sensors placed in the projection area can recover their locations to the nearest

UIST'07, October 7–10, 2007, Newport, Rhode Island, USA.
ACM 978-1-59593-679-2/07/0010.

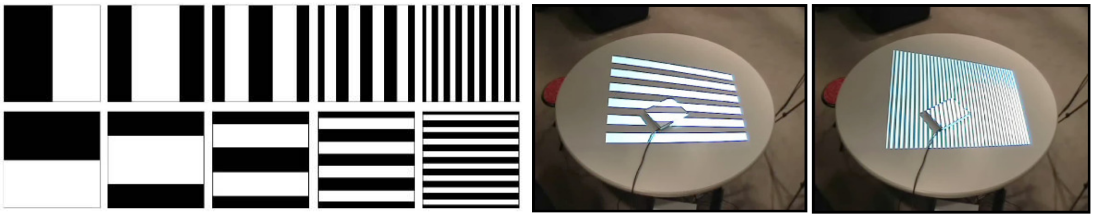

Figure 1. This set of horizontal and vertical Gray-coded binary patterns (*left*) are sequentially projected to discover the *x* and *y* pixel coordinates of sensors embedded into a rectangular surface (r*ight*).

pixel in a fast and robust manner. This robustness comes from the spatial encoding behavior of Gray-coded binary patterns which have been frequently used in the range finding community. These patterns also have a $\log_2(n)$ relationship between the number of patterns needed to resolve a given number of pixels n [Ino84]. The location data decoded by each sensor is inherently bound to the pixel coordinates that correspond to the sensor location. Sensors can be placed in pointer/controller form factors to obtain calibration-free input from the user or embedded into surfaces to shape and define the behavior of projected imagery.

Previous approaches to this problem have overwhelmingly focused on computer vision techniques to identify and locate important locations in the projection area. However, generalized object recognition, segmentation, and tracking remain an open and difficult problem in computer vision. A given vision algorithm has limitations on scene, target, motion, and illumination complexity. By using a projector-based approach to location discovery, many of these issues and limitations are avoided by simply relying on varying light energy emitted by the projector. This reduces system complexity, significantly reduces computational requirements, increases robustness, and eliminates the need for camera-projector calibration which typically requires user involvement. Additionally, the inherent binding of the location data with pixels in the visible light images provides capabilities unattainable using other approaches.

LOCATION DISCOVERY

The first implementation of this technique projected these Gray-code binary patterns as black and white visible light images using an unmodified commercial projector [Lee04]. While individual sensors can be used to find the *x*-*y* coordinates of objects placed in the projection area, using multiple sensors in a known geometric relationship yields more sophisticated applications. Four points can be used to automatically perform touch calibration in a touch-table or electronic white board system or can be used to define the boundaries of a simulated display, Figure 2. Light sensors can be embedded beneath the top layer of display surfaces to eliminate any visible evidence of their existence. This can also improve performance at shallow angles by providing a light diffuser in front of the sensor. This prototype was capable of performing accurate location discovery even when the projection angle was less than 2 degrees. Additionally, the orientation of the image is bound to the orientation of the surface which results in inherent and automatic reversal and adaption of projected images when the optical path is manipulated using mirrors or lenses. These capabilities are impossible using a camera-based approach.

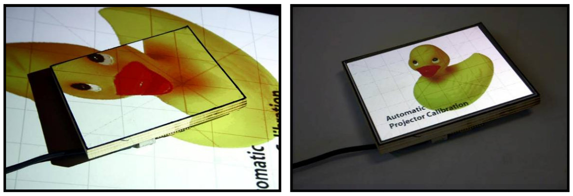

Figure 2. Fitting a target surface for display simulation

Figure 3. Multi-projector stitching (*top)* and appearance augmentation of 3D objects (*bottom)*.

Multiple projectors can be stitched together using an array of sensors placed in a grid layout. Each projector is calibrated individually and the overlapped area is blended yielding a single seamless display, Figure 3. In contrast to the computationally intensive and time consuming computer vision techniques developed for this problem, high quality alignment can be obtained within about one second. Multiple projectors can also be quickly and easily aligned to the same rectangular area providing layered imagery to create view angle dependent displays [Ngu07] or shadow reduction/elimination [Suk01].

Embedded sensors can also be used to register the orientation of arbitrary three-dimensional objects. Using the known surface geometry, the appearance of the object can be altered or customized, Figure 3. Recognition and orientation detection of arbitrary 3D objects, such as this, is still an open problem in computer vision. Similarly, temporarily placed sensor frames can be used to profile multi-planar environments, such as an office, used by steerable projection systems that can place static and dynamic displays on any surface in the environment [Pin01].

A 1024x768 resolution projector requires 20 images to uniquely identify every pixel in the projector's screen space. Since most commercial projectors have a 60Hz refresh rate, this translates to a maximum location discovery rate of 3Hz assuming 100% utilization of the projection channel for location discovery leaving no time for application content. As a result, this prototype lends itself better toward applications that only have momentary object or projector movement. Additionally, the black and white patterns used for location encoding create a caustic visual experience for human observers. My second implementation addresses these issues of location discovery speed and pattern perceptibility.

INTERACTIVE MOTION TRACKING

The second implementation of this technique uses a slightly modified Digital Light Processing (DLP) projector to render the binary Gray-coded binary patterns in a low-perceptibility manner. This prototype also moves toward an incremental tracking approach which supports higher speed location discovery and simultaneous application content projection. These capabilities provide real-time interaction with an application using moveable input sensors.

To reduce pattern perceptibility, we used a frequency shift keyed (FSK) or frequency modulated (FM) approach to encoding the binary images. Rather than using regions of black-and-white, we use regions of two different flashing frequencies whose difference is detectable by a light sensor but appear identical to a human observer. This is accomplished by removing the color wheel from a DLP projector which flattens the color space. Since DLP projectors render different colors using high-speed modulation of a digital micro-mirror display (DMD), each color has a unique pattern of high-frequency flashes resulting from mirror actuation. Removing the color wheel produces a monochrome projector, but which is able to render a single shade of grey using different flash patterns. To a sensor, these different frequencies are clearly visible and the binary data can be demodulated. But to a human observer, these tracking patterns appear to be static grey squares.

Incremental tracking involves first discovering the sensor locations using full-screen patterns and then projecting very small tracking patterns over the area around the last known locations. Incremental offsets are obtained and the location of the pattern is updated. A smaller area means fewer patterns, which in turn means faster updates. The result is an interleaved 12Hz update rate. For additional details regarding performance enhancements and tracking loss strategies associated with incremental tracking, please refer to [Lee05]. Smaller patterns also mean that the remaining pixels can be used for application content. This creates a fully usable interactive system where both input and output are provided by a single projector.

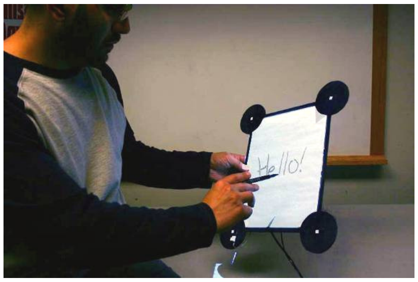

Figure 4. Hand-held tablet-PC simulation

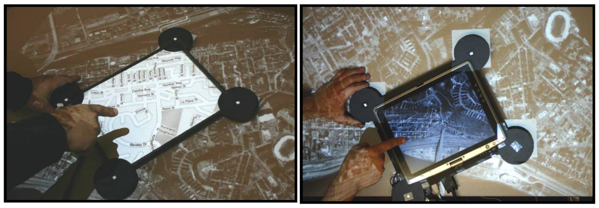

Figure 5. Magic lenses (left) and moveable Focus+Context (right)

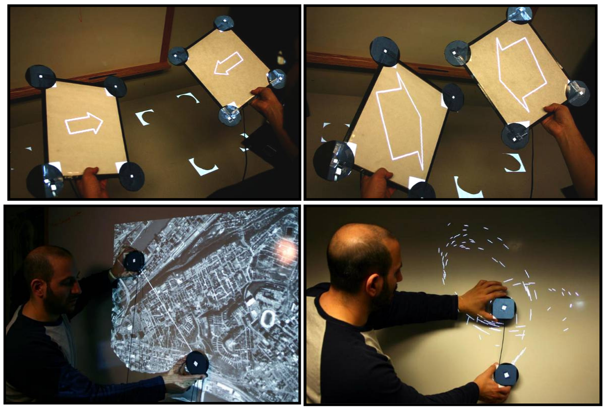

Figure 6. Location sensitive displays (*top*) and bi-manual physical input tools (*bottom*)

By tracking the location of four sensors, we can simulate the experience of a lightweight hand-held active display by projecting perspective corrected content onto the surface. Using a resistive film for stylus input, this provides a full tablet-PC like experience using a surface that weighs less than a legal pad, Figure 4. By projecting both inside and outside of the sensor frame, we can define a physical magic lens style window providing an alternative view of application data or overlay tools [Bie93]. Similarly, we can replace the passive surface with a high-resolution display creating a variable resolution, moveable Focus+Context system [Bau01], Figure 5. Multiple surfaces can be tracked and projected onto simultaneously using a single projector. Each surface is aware of its relative distance and orientation yielding a platform for location sensitive displays, Figure 6. The incremental cost of each surface is only $10. A large number of simulated displays can be used in environments where it would not be safe, economical, or physically possible for traditional display technologies. Single sensors provide a free mechanism for tracking physical bi-manual input controls that can be used in a tabletop or electronic whiteboard scenario.

However, incremental tacking consumes pixels that cannot be used for applications and also adds instability in the form of tracking loss when sensors move too fast. These issues are addressed in my third implementation.

HYBRID INFRARED AND VISIBLE PROJECTION

To eliminate the visibility of the Gray-coded binary patterns entirely, we can use invisible infrared (IR) light and low-cost IR sensors used in remote controls. However, no existing commercial projector is currently capable of emitting modulated IR light necessary for reliable data transmission. Additionally, a dedicated IR projector would provide location data, but an external device would be needed to provide visible application content. This is would negate many of the significant benefits gained from unifying the location discovery and visible projection technology into a single device. The approach we developed solves both of these issues and can be easily integrated into the upcoming generation of DLP projectors. Our solution utilizes the benefits of light-emitting diode (LED) illumination.

In addition to red, green, and blue visible light, LEDs are capable of emitting IR light that can be modulated at high-frequencies for the purposes of data transmission. Commercial manufactures have already developed three color LED DLP projector prototypes. A four color LED projector would require only a minor design change in light source design yet provide significant added value to future DLP projectors. In addition providing and improving upon the numerous capabilities described already, a hybrid infrared-visible light projector provides inherent support for input "light pens" containing an IR sensor in the tip. Stylus tracking is calibration-free and supports large numbers of simultaneous users without any increase in location or identity ambiguity. Similarly, pointer tracking continues to work even when projected onto non-planar discontinuous surfaces of unknown geometry, Figure 7. Both of these are difficult or impossible using alternative tracking technologies.

CONCLUSION

By unifying the image projection and location tracking technologies into a single device, many of the difficult calibration and alignment issues related to interactive projector applications can be eliminated. This approach removes the need for an external tracking system reducing system cost and complexity. Since, the location data is bound to projected pixels, a number of interaction capabilities are inherently provided which are difficult or impossible using previous techniques.

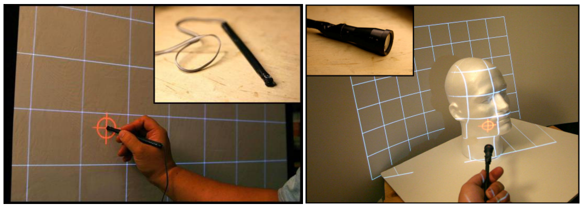

Figure 7. Inherent calibration-free multi-user stylus on a rear-projected display (*left*) and on non-planar discontinuous surfaces of unknown geometry (*right*).

REFERENCES

1. 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.
2. Bier, E., et al., "Toolglass and Magic Lenses: The Seethrough Interface." In proceedings of ACM SIGGRAPH 1993, Anaheim, CA., Aug. 1993,
3. 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
4. Lee, J., Dietz, P., Aminzade, D., and Hudson, S. "Automatic Projector Calibration using Embedded Light Sensors", In Proceedings of ACM UIST '04. October 2004.
5. 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. pp. 63-72
6. 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.
7. 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
8. Pinhanez, C. "The Everywhere Displays Projector: A Device to Create Ubiquitous Graphical Interfaces." In proceedings of Ubiquitous Computing 2001.
9. 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.
10. 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.
11. Rekimoto, J., Saitoh, M., "Augmented Surfaces: A Spatially Continuous Work Space for Hybrid Computing Environments." CHI'99, pp.378-385.
12. 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.
13. Ullmer, B., Ishii, H., "The metaDESK: models and prototypes for tangilble user interfaces." Proceedings of ACM UIST '97. October 1997.
14. Wilson, A. PlayAnywhere: A Compact Tabletop Computer Vision System, Symposium on User Interface Software and Technology (UIST), 2005.