MITSUBISHI ELECTRIC RESEARCH LABORATORIES
http://www.merl.com

Pokey: Interaction Through Covert Structured Light

Paul Beardsley, Yui Ivanov, Biliana Kaneva, Shoji Tanaka, Christopher Wren

TR2008-087 January 2008

Abstract

In this paper we describe a method to support interaction with a cellphone based projector-camera system. We describe a novel approach that uses a technique known in Computer Vision as structured light. It is based on projecting a pattern of light with known geometric properties onto a scene while imaging it with a camera. The distortions of the known pattern in the resulting image are due to the scene geometry which can be readily estimated. The main contribution of this paper is that the structure is created as consequence of the way raster-scan, laser-based micro-projectors operate, and is in fact invisible to the user. The structure of the projected light is sensed through careful synchronization within the camera-projector system and is imperceptible to the user. In this paper we describe the technique, and test it with a cell-phone based application that exploits this method while providing a natural interactive environment with no additional special equipment. The system enables manual interaction with a projected application using only the rasterizing projector and camera that will be part of next generation cell phones.

Tabletop 2008

This work may not be copied or reproduced in whole or in part for any commercial purpose. Permission to copy in whole or in part without payment of fee is granted for nonprofit educational and research purposes provided that all such whole or partial copies include the following: a notice that such copying is by permission of Mitsubishi Electric Research Laboratories, Inc.; an acknowledgment of the authors and individual contributions to the work; and all applicable portions of the copyright notice. Copying, reproduction, or republishing for any other purpose shall require a license with payment of fee to Mitsubishi Electric Research Laboratories, Inc. All rights reserved.

Copyright © Mitsubishi Electric Research Laboratories, Inc., 2008
201 Broadway, Cambridge, Massachusetts 02139

Pokey: Interaction through Covert Structured Light

Paul Beardsley[*] *Yuri Ivanov, Biliana Kaneva*[†] *Shoji Tanaka*[‡] *Christopher R. Wren*
Mitsubishi Electric Research Laboratories
201 Broadway, Cambridge, MA 02139

Abstract

In this paper we describe a method to support interaction with a cellphone based projector-camera system. We describe a novel approach that uses a technique known in Computer Vision as structured light. *It is based on projecting a pattern of light with known geometric properties onto a scene while imaging it with a camera. The distortions of the known pattern in the resulting image are due to the scene geometry which can be readily estimated. The main contribution of this paper is that the structure is created as consequence of the way raster-scan, laser-based micro-projectors operate, and is in fact invisible to the user. The structure of the projected light is sensed through careful synchronization within the camera-projector system and is imperceptible to the user. In this paper we describe the technique, and test it with a cell-phone based application that exploits this method while providing a natural interactive environment with no additional special equipment. The system enables manual interaction with a projected application using only the rasterizing projector and camera that will be part of next generation cell phones.*

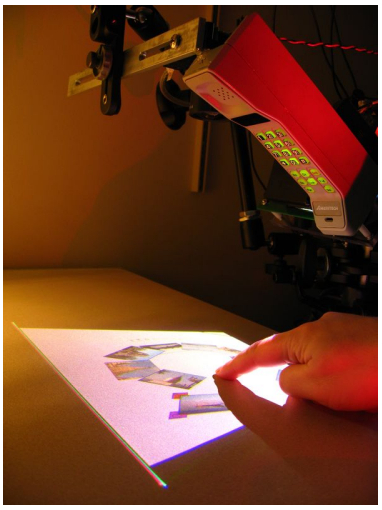


Figure 1. A prototype cellphone-camera-projector device displaying a touch interface.

1. Introduction

The growing popularity and portability of cell phones makes them an ideal, omnipresent computational platform and information appliance. Development of User Interfaces (UI) to be deployed on cell phones is complicated by two competing requirements - quality of the display, versus keeping its size small enough to fit in a pocket. One solution to these seemingly incompatible requirements involves utilizing a projector to move the screen off the device to enable high fidelity, possibly multi-user interactions, e.g. see [1]. Cell phone and display manufacturers are currently developing projection devices to fit inside the cellphone body as well, [2]. This interest signifies a new trend and new opportunity for UI designers and technologists, making the coexistence of projectors and cameras in cellphones easier.

One problem that projector-equipped phones inevitably will have to address is that of a variable focus. A solution to this problem is readily available in the form of laser-based projection systems, eg. [3]. The images projected with this technology are in sharp focus everywhere, regardless of the surface geometry. This feature makes the laser-based devices very attractive for the cellphone use, as they remove the restriction on the projection surface being flat and at a fixed angle to the optical axis.

We thus anticipate that cell phones will soon be equipped with both a camera and a laser-based projector. We present a way to link these two systems together to create a robust and compact interaction modality. The goal is to create an interactive projection display that is tolerant of poor light, dirty and uneven surfaces, and projector placement. Furthermore, all the necessary hardware must be contained within the phone. Since the projector is intended to show application imagery, we also wish to avoid creating any ar-

[1] now at the Disney Laboratory in Zurich
[2] now at the Massachusetts Institute of Technology
[3] at Mitsubishi Electric ITC in Japan

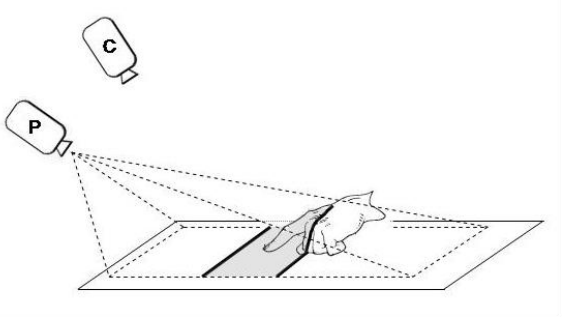


Figure 2. Structured light in a projector-camera system distored by the presence of a hand.

tifacts that are visible to the user. By carefully synchronizing the camera to the laser projector we can create a sensing modality for touch-surface interaction that satisfies these goals.

2. Sensing Theory

Recovery of 3D scene information from a single camera has a long history (see [4] for references). This approach to recovery of 3D geometry assumes that the surface under consideration has some known reflectance properties, such as Lambertian, Hapke, etc. and does not typically include formalisms describing the properties of scratched-up beer-stained table top of a local pub, which is our target environment.

Due to its robustness to reflectance and geometry of imaged surfaces, the method of structured light has received significant attention in the past 20 years. For example, [6] describes a method for rapid active ranging using color-coded structured light. [5] describes a multi-stripe inspection application. More recently, the advances in computing power of desktop computers made 3D scanning applications an everyday reality with the idea of a structured lighting taking a prominent role. The main idea of the approach is to project a pattern of light with known geometric properties onto a scene in question. Then the analysis of distortions induced on the pattern as seen from a camera reveals the scene geometry, e.g [8].

2.1. Structured Light Approach

Our work addresses the problem of generating and displaying a structured light pattern onto a scene in a way that is not obtrusive to the user. The goal of the method is to extract 3D information about the surface during the interaction such that fast update is possible for near-real time tracking of user hands and other objects.

An example of a camera-projector setup that is capable of inferring a 3D geometry of the scene is shown in figure 2. The figure shows a projector (P), projecting a horizontal stripe onto the surface. The edges of the stripe should be seen by the camera (C) as straight lines if projected onto a planar surface. If an edge falls onto an elevated object (hand, as shown in the picture), the distortion in the appearance of the edge of the stripe is related to the amount of elevation at every point.

2.2. Aliasing and Invisible Shadows

While the structured light approach is widely accepted as one of the simplest ways to recover 3D scene information with a monocular imager, it defies the purpose of the camera-projector setup in a cell phone: the user wants to see the interface, not bands of light. In our work we exploit the aliasing effect that occurs when both the display and the imaging device use frame rasterization. Since the frame is painted by the display one line at a time, opening and closing the camera shutter in a rapid sequence during the acquisition of a single frame will result in some lines of the projected image being "blanked out". This creates artificial horizontal stripes when the projected image is captured by such camera. This has no effect on the user's experience as it does not affect the projected image. The camera, however, gains full use of this artificial striping to analyze the straight edge distortions and thus recover the 3D geometry of the scene.

It should be noted that this effect is frequently seen when a video camera is pointed at a TV screen. Generally regarded as a nuisance, in this case it serves a useful purpose of creating a covert structured light pattern. It is the principle on which we base the work presented in this paper.

The aliasing effect can be seen in the figure 3. The figure shows what a scene containing a planar background and a user's hand looks like to the user (a), and the camera (b). Since the aliasing effect will always generate a straight line, any distortions in the line seen by the camera will be due to the 3D geometry, effectively playing the role of the structured light device.

Our solution is completely implemented in the low-level hardware through synchronization. This is different from the method proposed by Raskar *et al.*[7] which requires a higher frequency projector and tight cooperation between the camera and the software generating the projected image.

3. Hardware Implementation

We built an prototype system to serve as a proof of concept. The system is depicted in Figure 4. It consists of a video camera, a miniature RGB laser projector and a

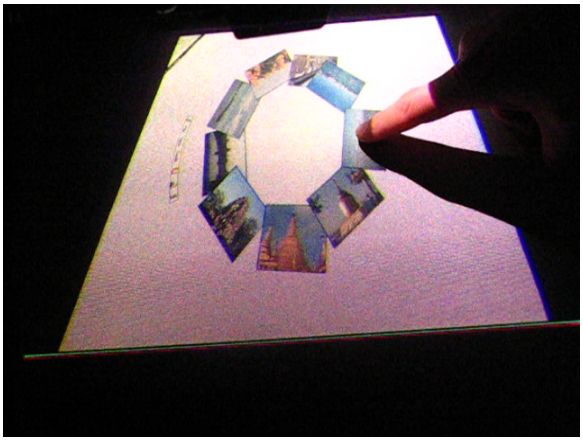
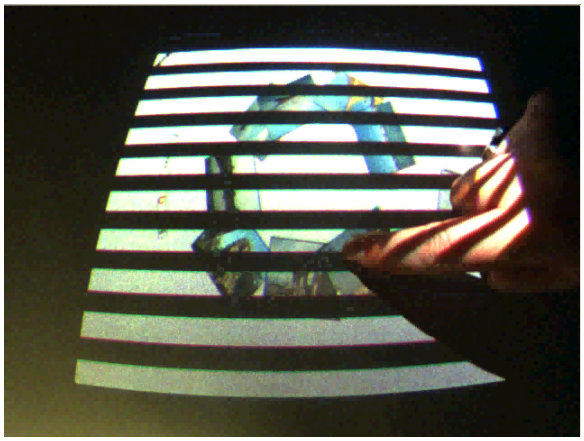
a)  b) 

Figure 3. Synchronized projector and camera to generate the stripes of the structured light. The figure shows the user's hand as seen by a human(a) and a camera(b).

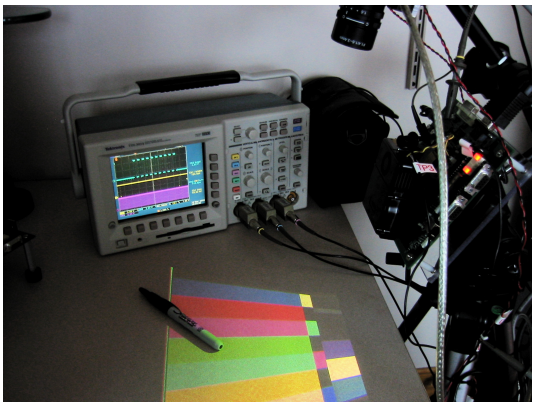


Figure 4. Hardware Protoype with Color Projector.

general-purpose micro-controller board used for accurate synchronization.

The laser-based RGB projector was provided to us by Symbol Technologies and is capable of projecting full-color 640x480 size images at 60 fps. Commercially manufactured projectors will target this refresh rate since it is the minimum required to insure that the display will be flicker-free. Techniques that require faster projectors are not practical since they imply significant additional engineering that serves no purpose except to enable interaction. We've found that manufacturers are not willing to incur significant additional cost to implement interaction.

We use the Firefly 2 video camera from PointGrey Research and run it in a shutter-triggered mode (Mode 5). This mode allows us to control the electronic shutter, and to supply a frame integration count. The latter specifies to the camera the number of times the shutter opens and closes before the charge is read off the imaging device.

The PIC-based synchronization board accepts frame- and a line-sync signals from the projector and outputs a shutter control signal to drive the camera shutter in precise synchrony with the projector. CMOS imagers provide this level of control as a consequence of their implementation, even though not all camera modules expose this functionality to the application developer. The Firefly 2 is an expensive camera only because it is sold to a specialty market in small quantity.

4. Software Implementation

The multi-band surface scanning method described in the previous section does not affect the user's viewing experience. However, as sensed by the camera, it is effectively the same as dividing the original image into a set of horizontal bands, setting alternate bands to black, projecting, and detecting the projected image at the camera. At each time step we reverse the phase of the bands, first the odd bands are dark, then the even bands, and so on. Subtracting these two phases allows us to ignore any environmental lighting.

An automatic calibration at startup (taking a few seconds) is used to detect the boundary of the projected image and the placement of the horizontal bands, as seen on the camera image plane. There are two types of information that can be reliably extracted from the camera images, relating to the presence of a 3D object above the projection surface, and to cast-shadows on the surface.

Detecting a light area on the camera image plane within an expected dark band is evidence of a 3D object above the projection surface - the object is intercepting projected light

and reflecting it towards the camera. It is also possible for a 3D object to be present without obtaining this signal if there happens to be a dark band projecting onto it. But for our typical usage conditions, this cue is quite sufficient to detect a hand with pointing finger.

In a similar way, detecting a dark area within an expected light band is evidence of a shadow being cast. However there is a complication in this simple model. There may be projected image texture within the light band that is itself black. We avoid this by enforcing design rules on the projected image such that the brightness is above a threshold. It is possible that providing the vision system access to the content of the projected image could help loosen this restriction, but we have not yet explored this option.

Thus every pixel in the camera image can be labeled as (A) evidence of a 3D object above the projection surface, (B) evidence of a cast-shadow on the surface, (C) the remainder. We compute the moments of the pixels in class A to determine their center-of-gravity and the major axis of the distribution. A hand with pointed finger generates a major axis that roughly aligns with the finger. We then examine the two intersections of the major axis with the boundary of the camera image plane, to determine on which side of the center-of-gravity the hand is entering the image plane. Finally we compute the extremal pixel in class A on the opposite side from hand-entry—this is our estimated position F of the finger tip. The same major axis and direction is used to find the extremal point of the pixels in class B—this is our estimated position S of the cast-shadow of the fingertip. This step of the analysis is similar to the method used in PlayAnywhere[9]. The difference lies in how the shadows are generated.

We now have sufficient information to determine if the fingertip is on or off the surface. If $distance(F, S)$ is less than a threshold then the fingertip is on the surface. This corresponds to the observation that the physical fingertip moves into coincidence with the shadow fingertip as the finger approaches the surface. The prior calibration gives the position of the fingertip in the coordinate frame of the original image, and we have all the information necessary for touch surface interaction.

5. Discussion and Applications

We tested the proposed technique with in application that enables the user to manipulate digital images stored on the "phone" (Figure 1) and projected onto the surface of the desk (Figure 3). Tracking a finger tip in the presence of dynamic, projected light is a challenging problem. Indeed, with traditional tracking methods this task would be nearly impossible due to the fact that the dynamic projection would invalidate scene models and projected light falling on the finger itself would reduce the visual distinctiveness of the finger relative to the scene. As can be seen in the supplemental video, it is possible with our technique to select, drag, and release the photos, moving them around the scene. These actions change the projected scene during manipulation. Since the photos are dragged by the fingertip, the greatest changes are near the fingertip, where the most sensitive part of the analysis is attempting to determine the height of the fingertip above the table to distinguish touching from hovering. The method is also demonstrably robust to wide changes in illumination because the method operates on the difference between even and odd projected bands.

6. Conclusion

In this paper we present a method for touch surface interface technology the uses raster-scanning, laser-based projectors. These projectors are ideal for mobile touch surfaces because they can be very compact, have no focus depth problems, and they impart a strong temporal structure to the projected light. This structure can be revealed through simple, cheap, low-level hardware synchronization. Once revealed, this structure makes the vision task significantly easier by leveraging the extensive literature on structured light techniques.

7. Acknowledgments

The authors would like to express our gratitude to Dr. Chia Chen, Dr. Clifton Forlines, and Dr. Daniel Wigdor for sharing thier surface interaction expertise. We would also like to thank Dr. William Yerazunis for making sure we didn't burn the place to the ground.

References

[1] `http://www.1800mobiles.com/su2.html`.
[2] `http://www.i4u.com/article10387.html`.
[3] `http://www.physorg.com/news99759115.html`.
[4] B. Horn. *Robot Vision*. McGraw-Hill, 1986.
[5] J. A. Jalkio, R. Kim, and S. Case. Three dimensional inspection using multistripe structured light. *Optical Engineering*, pages 966–974, 1985.
[6] K.L.Boyer and A. Kak. Color-encoded structured light for rapid active ranging. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 9(1):14–28, 1987.
[7] R. Raskar, G. Welch, M. Cutts, A. Lake, L. Stesin, and H. Fuchs. The office of the future : A unified approach to image-based modeling and spatially immersive displays. In *SIGGRAPH*, pages 179–188. ACM, 1998.
[8] R. Valkenburg and A. McIvor. Accurate 3d measurement using a structured light system. *Image and Vision Computing*, 1998.
[9] A. Wilson. Playanywhere: A compact tabletop computer vision system. In *Symposium on User Interface Software and Technology (UIST)*, 2005.