LampTop: Touch Detection for a Projector-Camera System Based on Shape Classification

Doney Alex
Dept. of Elec. Engg.
Indian Institute of Science
Bangalore, 560012, India
doneyalex@gmail.com

Kalpathi Ramakrishnan
Dept. of Elec. Engg.
Indian Institute of Science
Bangalore, 560012, India
krr@ee.iisc.ernet.in

Pushkar Gorur
Dept. of Elec. Comm. Engg.
Indian Institute of Science
Bangalore, 560012, India
pushkar.gorur@gmail.com

Bharadwaj Amrutur
Dept. of Elec. Comm. Engg.
Indian Institute of Science
Bangalore, 560012, India
amrutur@ece.iisc.ernet.in

ITS '13, Oct 06-09 2013, St Andrews, United Kingdom.
ACM 978-1-4503-2271-3/13/10.
http://dx.doi.org/10.1145/2512349.2514921

Abstract
The LampTop enables an effective low cost touch interface utilizing only a single camera and a pico projector. It embeds a small shape in the image generated by the user application (e.g. a touch screen menu with icons) and detects touch by measuring the geometrical distortion in the camera captured image. Fourier shape descriptors are extracted from the camera-captured image to obtain an estimate of the shape distortion. The touch event is detected using a Support Vector Machine. Quantitative results show that the proposed method can effectively detect touch.

Author Keywords
Projector - Camera Systems; Augmented workspaces; Touch detection; Fourier shape descriptor.

ACM Classification Keywords
H.5.2 [Information Interfaces and Presentation (e.g., HCI)]: User Interfaces

Introduction
A subjective study of projector-camera interfaces for office spaces [5] showed that users have a strong preference to interact using pointing gestures and direct finger touch involving minimal body movement. In this paper we propose one such system: LampTop which is essentially a

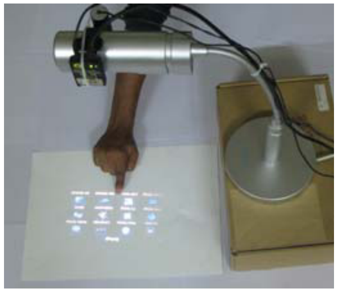

Figure 1: Setup of the lamptop

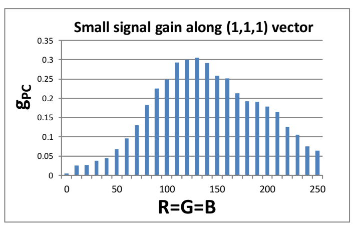

Figure 2: Small signal gain plotted along the (1,1,1) axis in the RGB color space

table lamp which doubles up as an effective touch interface. LampTop consists of a table lamp augmented with a pico projector and a single low cost camera as shown in Fig. 1. The system allows the user to interact with a computer/laptop by detecting finger position and touch events over the active projection surface.

A large number of HCI systems for table tops have been proposed over the last decade. TinkerSheets [8] uses physical tokens to indicate menu selection. Bonfire [2] uses accelerometers in the laptop to determine tapping and flicking gestures on the surface. PlayAnywhere [6] uses shadow shape analysis to determine a touch event. However, we find that in a number of situations, the shadows are not visible due to the camera and light source positions. Takeoka et al. [4] use Multilayered infrared (IR) laser planes which are synchronized with shutter signals from a high-speed camera. Recent researches have adopted Stereo cameras and Kinect depth sensors to detect touch [7].

The work closest to the proposed approach is the system introduced by Jingwen et al. in [1]. In this method, a structured code pattern is embedded into the entire projected image. The binary code recovered from the image over the finger tip is used to detect touch. However, Jingwen et al. mention that the structured code pattern causes flickering when the embedded intensity was increased. The method used an expensive CCD camera with a very low distortion lens. We found that the poor light performance of low cost CMOS sensor based web cameras (relative to the superior performance of expensive CCD cameras) necessitates a large embedded intensity which would cause significant image degradation. In contrast, LampTop embeds a small shape only over the tip of the finger and analyzes the camera-captured image to detect touch. Since LampTop requires only a single camera, the overall cost of the system is lower compared to kinect/stereo camera based systems. The touch detection method developed for LampTop will also be particularly useful when extended to small form-factor smart phones which cannot accommodate depth sensors.

Touch detection

(I) Shape embedding and retrieval
Skin color segmentation (using a linear classifier in the HSV color space) is performed on the camera-captured image to obtain the segmented hand region of the user. Template matching is used to determine the position of the finger tip in the camera plane. The mapping of coordinates in the camera imaging plane to the projector plane is a homography. The homography is estimated initially during system startup by projecting a checker board pattern. The projective transformation parameters obtained are used to compute the position of the finger tip in the projector plane.

The R,G & B values of pixels in a circular region over the finger tip are incremented in the projected image (i.e. image generated by the user application) by an offset δ_+. A second successive image is projected in which the pixel values (in the same circular region) are decremented by an offset δ_-. The difference of the corresponding camera-captured images is converted to Gray scale and thresholded to recover the embedded shape (shown in Fig. 3 e,f). R, G & B values of pixels directly over the finger are set to 0 and 255 (in the two projected images). Robust segmentation of the other pixels (in the circular region) requires the small signal gain g_{PC} (i.e. the ratio of change in camera-captured pixel value and the change in projector pixel value) to be large. The gain g_{PC} determined along the $(1,1,1)$ vector in the RGB color

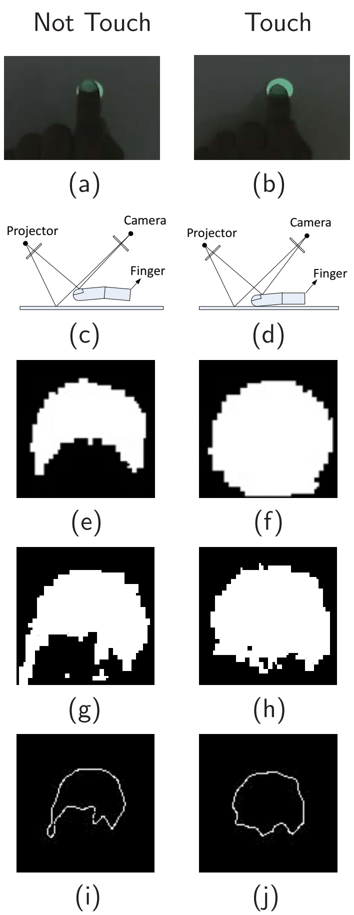

Figure 3: Figure shows (a), (b) Camera captured images (c), (d) Geometry (e), (f) Typical segmentation samples (g), (h) Noisy segmentation samples (i), (j) Shape reconstructed with 30 Fourier coefficients

space is plotted in Fig. 2. We observe that g_{PC} is low when R, G and B values lie close to the boundaries (i.e. 0 and 255) due to the non-linearity of the projector camera transfer function. Hence, if the pixel value in the application image is in the low gain regions, we increase the offsets by a value δ_{inc} in the projected image to ensure that $(\delta_+ + \delta_- + 2\delta_{inc}) * g_{PC}$ is greater than the minimum signal d_{min} required for robust segmentation. The low gain regions are determined initially during system startup (d_{min} was set to 5 based on experimentation).

(II) Touch Detection based on shape classification
The epipolar geometry of the camera and the projector system is shown in Fig. 3 c,d. When the finger is on the surface, the projected circle is recovered completely with very little distortion as shown in Fig. 3 f. However, when the finger is above the surface, the angle subtended by the cone of light rays reduces. Hence, the circle is distorted as can be seen in the segmented image Fig. 3 e. Due to noise in the segmentation process (Fig. 3 g,h shows two noisy segmented images), simple shape coverage thresholding methods cannot be used to classify the segmented image. Hence, we propose to detect the distortion in the shape using Fourier descriptors to determine whether the finger has touched the surface.

Fourier descriptors have been shown to describe shape compactly and accurately [3]. The general form of the shape is represented by the low frequency Fourier coefficients and information of the finer details are contained in the coefficients with high index. In this application, discarding higher frequencies removes the finer noise in the image boundaries. However, the lower frequency coefficients are required to represent the overall shape of the image. We choose 30 low frequency indices as the feature for shape classification. Fig. 3 i,j shows that the noisy segmentation at the boundaries is removed but the overall shape is retained. We note that the magnitude of the high frequency Fourier indices increase when the finger is above the surface (due to the edges introduced in the shape). Hence, reducing the high frequency coefficients in the undistorted shape i.e. in the original projected image will increase the discriminability of the classifier. With this in view, we choose the circle as the shape to embed in the projected image. We train a Support Vector Machine (SVM) using a Radial Basis Function (RBF) kernel with 1000 samples of 'Touch' and 'Not touch' images. The trained model is used to classify the segmented images during testing.

Experimental evaluation and results

The lamptop was placed at a height of 50cm from the surface creating a 26x18cm active touch area. The algorithms were implemented in C++ using the OpenCV library. The program was executed in single threaded mode on an Intel Core i3 processor with 8GB RAM running at 3.1GHz. Three subjects were invited to participate in the experiment and were briefed about the system. Number of trials (touching the surface) per person was 500. Fig. 4 shows the True Positive Rate (TPR) plotted against the offset δ (set equal to δ_+ & δ_-). We can observe that the accuracy is greater than 95% when δ is equal to 25. Reducing the value of δ results in poor shape reconstruction due to errors in segmentation, hence reducing the TPR. In [1], lower values of δ were found to be sufficient. However, we note that the proposed system uses CMOS web cameras unlike the high quality CCD camera and lens modules that are used in [1].

The number of false alarms (across multiple trials) is determined for different distances of the finger from the

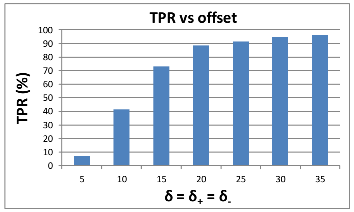

Figure 4: Average TPR (%) vs δ (Across 3 different users)

Height (inches)	False alarm rate
0.5	24.6%
1	0.6%
1.5	0%

Table 1: Number of False alarms with varying height of finger above surface

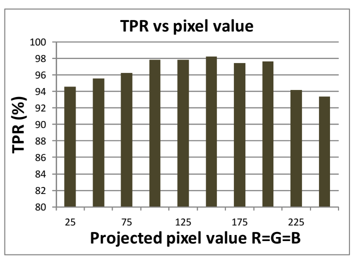

Figure 5: TPR (%) vs Projected pixel value

touch surface. Table 1 shows that the proposed method produces no false alarms when the finger was placed at a distance of 1.5 inches from the surface. A False Alarm Rate (FAR) of 0.6% was observed when the finger was placed at a height of 1 inch from the surface. This can be attributed to the reduced distortion between the projected and camera captured shapes. The FAR can be reduced be increasing the distance between the projector and camera. However, we observed that users typically do not hover the finger very close to the surface. Hence, we do not notice any false alarms when users are allowed to interact using the proposed system.

To quantitatively measure the accuracy of the proposed encoding method on different projected images (provided by the application), we have measured the TPR for different values of the projected pixels. Fig. 5 shows that the True Positive Rate does not significantly reduce when the projected pixel values are varied. The total computation time required for a single application image was measured to be 12.4ms. Hence, it can be easily ported to a low cost embedded platform.

Conclusion

We have presented LampTop, a computer vision based table lamp augmented with a pico projector to display interactive content on the table. A single web camera is used to determine touch. A shape is embedded in the projected image over the tip of the finger. The touch detection algorithm accurately classifies the recovered shape using compact Fourier descriptors. Computation time was measured to be 12.4ms/frame when executed in single-threaded mode. Results show that the average TPR of the proposed approach is 95.4%. Unlike in [1], we only embed the shape over the finger tip. Hence, the distortion in the projected image is lower. The current system does not detect 'drag' (touch-detection while moving). In the future, we plan to improve LampTop to support this.

Acknowledgment

We gratefully acknowledge the funding support from the Robert Bosch Center for Cyber Physical Systems at the Indian Institute of Science, Bangalore.

References

[1] Dai, J., and Chung, R. Making any planar surface into a touch-sensitive display by a mere projector and camera. In *Proc. CVPRW* (2012), 35–42.

[2] Kane, S. K., Avrahami, D., Wobbrock, J. O., Harrison, B., Rea, A. D., Philipose, M., and LaMarca, A. Bonfire: a nomadic system for hybrid laptop-tabletop interaction. In *Proc. UIST '09*, ACM (2009), 129–138.

[3] Poppe, R., and Poel, M. Comparison of silhouette shape descriptors for example-based human pose recovery. In *Proc. FGR* (2006), 541–546.

[4] Takeoka, Y., Miyaki, T., and Rekimoto, J. Z-touch: an infrastructure for 3d gesture interaction in the proximity of tabletop surfaces. In *Proc. ITS '10*, ACM (2010), 91–94.

[5] Voida, S., Podlaseck, M., Kjeldsen, R., and Pinhanez, C. A study on the manipulation of 2d objects in a projector/camera-based augmented reality environment. In *Proc. CHI '05*, ACM (2005), 611–620.

[6] Wilson, A. D. Playanywhere: a compact interactive tabletop projection-vision system. In *Proc. UIST '05*, ACM (2005), 83–92.

[7] Wilson, A. D. Using a depth camera as a touch sensor. In *Proc. ITS '10*, ACM (2010), 69–72.

[8] Zufferey, G., Jermann, P., Lucchi, A., and Dillenbourg, P. Tinkersheets: using paper forms to control and visualize tangible simulations. In *Proc. TEI '09*, ACM (2009), 377–384.