BrightShadow: Shadow Sensing with Synchronous Illuminations for Robust Gesture Recognition

Jun Rekimoto
Interfaculty in Information Studies,
The University of Tokyo
7-3-1 Hongo, Bunkyo-ku,
Tokyo 113-0033 Japan

Interaction Laboratory, Sony
Computer Science Laboratories, Inc.
3-14-13 Higashigotanda,
Shinagawa-ku, Tokyo 141-0022
Japan.
rekimoto@acm.org

Copyright is held by the author/owner(s).
CHI 2008, April 5-10, 2008, Florence, Italy.
ACM 978-1-60558-012-8/08/04.

Abstract
We introduce a new sensor architecture for robust gesture recognition that uses a combination of a high-speed camera and synchronous LED illumination. This sensor looks at shadows cast by a user's hand for recognizing position. The position of the hand can be robustly recognized by independently tracking multiple shadows and by using multiple light sources with time-synchronous modulation with the camera. We also developed a multi-finger tracking system that uses similar modulated illumination from multiple light positions. We expect that these sensing configurations can be naturally integrated into our daily environments as LED lighting becomes more commonplace.

Keywords
Sensing, Interaction, Shadow, Synchronous Illumination, Gesture Recognition

ACM Classification Keywords
H5. Information interfaces and presentation (e.g., HCI): H5.2 User Interfaces – Input devices and strategies (e.g., mouse, touchscreen)

Introduction

Recognizing hand gestures without wearing any cumbersome devices is the foundation for many ubiquitous computing and smart environment applications, such as remote-controllers for digital appliances. Many camera-based gesture recognition systems have been proposed, but most of them are affected by ambient light changes, noisy background images, or even a user's body image itself. A system that is robust enough to be used in daily life is still difficult to achieve.

This paper describes a new sensor configuration for camera-based gesture recognition called "BrightShadow," which uses synchronously modulated light sources with a high-speed camera.

BrightShadow sensor configuration

The BrightShadow sensor configuration has three features:

- *"Time-synchronous illumination"* to make image recognition significantly easier. Time-synchronous means the light sources' "on" and "off" are triggered by each camera frame. The system knows which frame is illuminated and which frame is not. Then it becomes very simple to extract the effect of the light-source by simply subtracting an image frame with light "on" from the next image frame with light "off". Because the other ambient light does not change during these two consecutive image captures, its effect can be cancelled.

- Instead of directly recognizing hand images, the camera looks at "shadows" cast from a user's hand. Shadows have higher contrast than hand images. Thus, they can be more easily and more robustly tracked. If multiple shadows are separately recognized, the finger positions in the 3D space can also be calculated.

- *"Multiple light sources"* instead of multiple cameras can be used for simpler recognition. Because each light source can have a distinctive illumination sequence, illumination effects of multiple light sources can be distinguished.

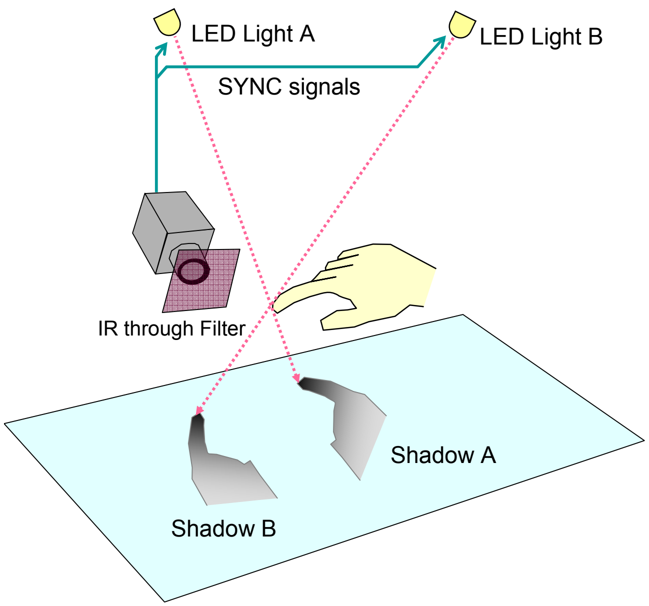

figure 1. The BrightShadow sensor configuration: Multiple LED light sources are synchronized with a high-speed camera's frame rate to illuminate a hand alternatively. The camera can independently recognize shadows to track the hand position in 3D space.

Configuration 1: Sensing Multiple Shadows for 3D Hand Tracking

Figure 1 shows an example BrightShadow sensor configuration. Two LED light sources are used to illuminate the hand. These LEDs are synchronized with the video camera; three combinations of LED on/off states (all off, LED A on and LED B off, and LED A off and LED B on) are repeated synchronously with camera frames (Figure 2).

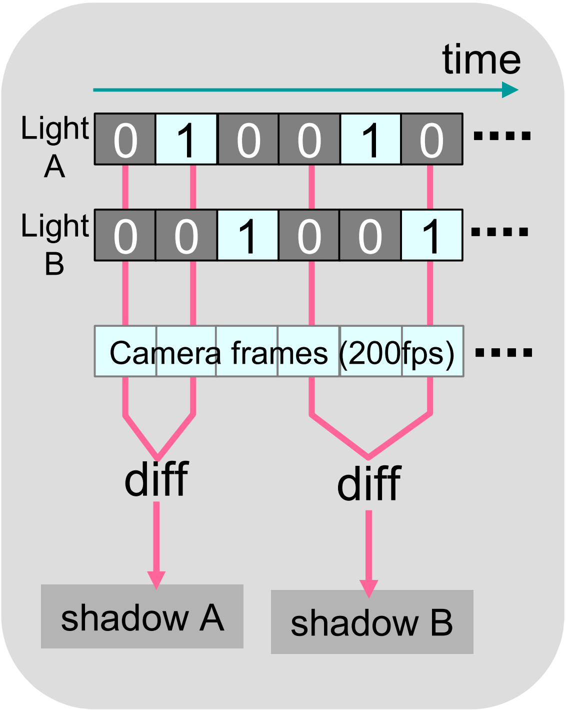

figure 2. The Light on-off sequences for independently identifying shadows from two light sources.

Then, using simple camera image subtractions enables separately extracting shadows cast by two light sources. If the positions of both Light A and Light B are known, a single camera can be used to estimate the 3D position of the user's fingertips. The system also recognizes simple finger gestures, such as pointing with an index finger and pinching with an index finger and a thumb.

Figure 3 shows actual images taken by the system. Two shadows are clearly separated.

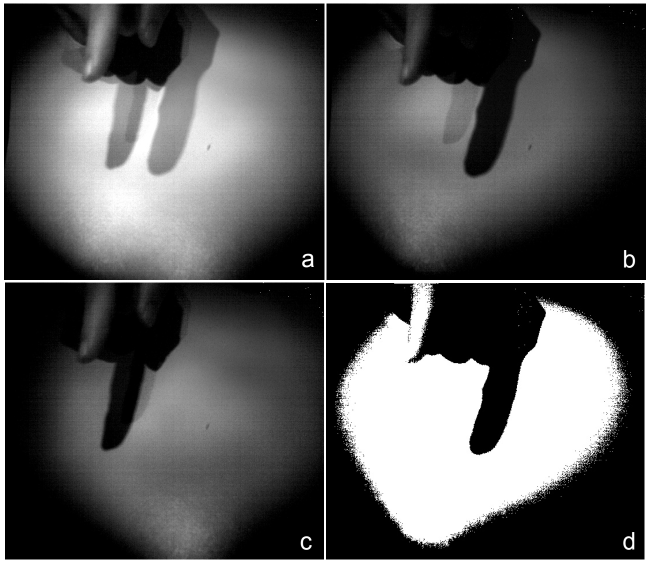

figure 3. Captured images: (a: normal image without light synchronization. b,c: results of frame subtraction for detecting a single shadow, and d: the result of brightness binarization for fingertip extraction.)

With a high-speed camera that takes 200 frames per second, the system performs this 3-frame processing 66 times per second. This performance rate is sufficient interactive systems.

Configuration 2: Light-wall for robust multi-finger tracking

A variation of the BrightShadow sensing is our second sensor configuration, which is called the "Light-Wall". In the Light-Wall system, the similar synchronous LED light sources are placed on both sides of the person's hand (Figure 4).

The sensing principle is similar to the previous one, although reflections are used instead of shadows in this case. When a user sticks his/her finger(s) toward the camera, the LED lights on both sides alternatively illuminate their fingertips. Because the both sides of a fingertip are illuminated by different light sources, detection of single and multiple fingertips is possible if the system looks for two nearby connected regions from two images (with different light sources).

Figure 5 shows examples of captured images. The background image and light can also be easily canceled (note the very shiny background in an unilluminated image in Figure 5(A)). Only fingertips become bright in the camera's view. Applying connected-region (blob) extraction then enables the system to track multiple finger points, even though the background has other moving objects such as the person's body or ambient light such as sunlight.

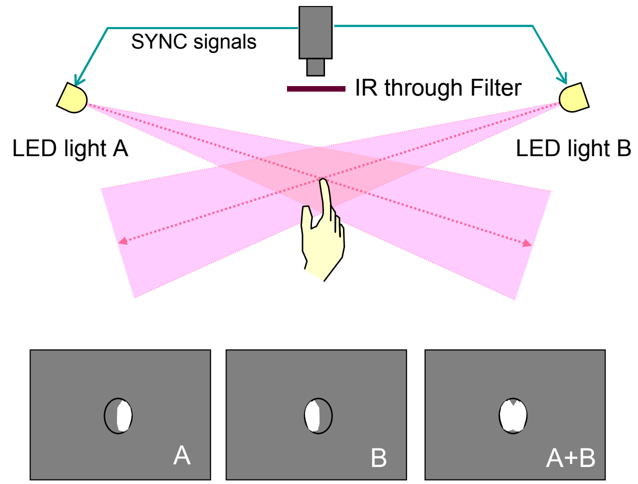

figure 4. "Light-Wall" sensor configuration for fingertip tracking.

figure 5. Images taken by the "Light-Wall" configuration. (A: original image with an extremely shiny background, B,C: results of frame subtraction, D: after binarization)

We chose a vertically placed array of LEDs as light sources. These lights create a “wall of light” in front of the person, and the wall recognized the location whenever he or she pokes his/her fingers into the wall.

This sensor configuration can be used to support single and multiple finger-points interactions. Unlike other "multi-touch" systems (such as [2,5,6]), a user does not have to approach and touch the display. He or she can control it from a distance, like a remote controller can.

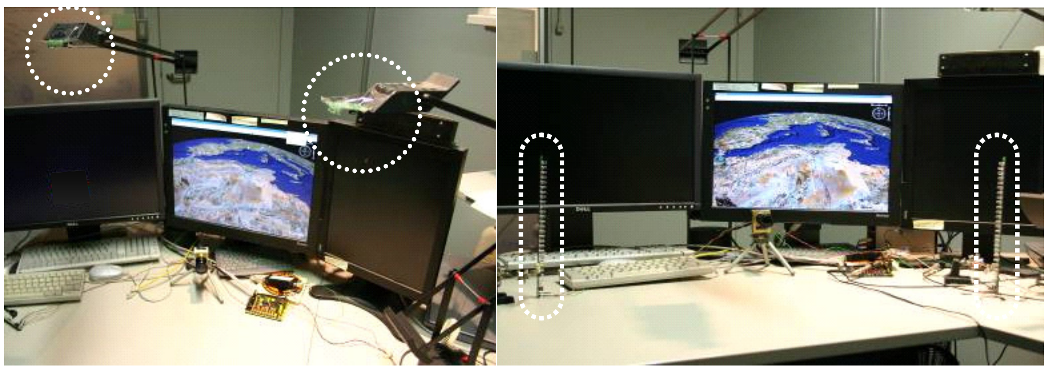

figure 6. Light Source configurations of current prototypes

Combining with Environmental Illumination

Although our initial prototype uses explicit infrared light sources (as shown in Figure 6), visible lights can also be installed in a room for normal illumination. As white LEDs become brighter and more commonplace, our environmental lighting will be soon be replaced by such LEDs. Then, each LED should shine with a different illumination pattern, synchronized with the camera frame rate for recognition. The BrightShadow sensor architecture should enable the system to estimate someone’s pose in such an environment, by separately recognizing shadows cast from multiple LED light sources installed in an environment as normal illumination (Figure 7).

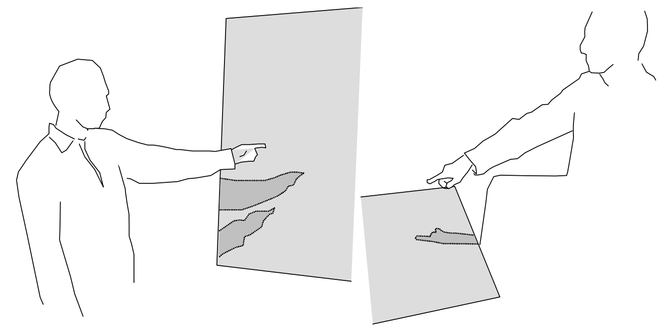

figure 7. Possibility of sensing natural shadows in everyday environments

Related Work

The use of time-synchronized illumination is also effective for distinguishing shadows used by our system for sensing from shadows created by other light sources. This is the major advantage over previous systems that require controlled (dark) environments for silhouette recognition [4].

HoloWall [5] is our previous interactive surface system that uses a combination of infrared lights and a camera. The light sources are placed behind the rear-projection screen for better recognition robustness. Although a semi transparency of a rear-projection sheet separates the light reflection of fingertips from the background image, the system is still affected by the ambient light change. Our proposed sensing architecture greatly reduces this problem by introducing synchronous light modulation.

Casssinelli et al.‘s system uses a high-speed camera for finger tracking by illuminating a fingertip with an active laser [1]. While it uses motor-controlled lasers, BrightShadow just requires multiple fixed lights.

BrightShadow’s synchronous illumination architecture should also greatly improve the robustness of other optical-based multi-touch systems, such as Frustrated Total Internal Reflection [2], LCD displays with photo sensors [3,7], and the Microsoft Surface tabletop system [6].

Conclusion and Ongoing Work

A new sensor configuration for robust gesture recognition was introduced. The proposed BrightShadow system can track finger positions in 3D space by using multiple time-synchronous light sources with a high-speed camera and by focusing on multiple-shadows cast from a user’s hand.

We are currently working on prototype systems based on this sensing principle. One is an augmented table system with visible illumination mounted on the ceiling as sources of light. This table system recognizes multiple hands gestures by focusing on the table surface. We are planning to investigate the possibility of using people’s body shadows cast on the wall to recognize their standing point near the wall. This could be used during activities such as a discussion in front of the whiteboard. Finally, we are also interested in integrating this idea with an LCD display and a photo-diode array [7].

References.

[1] Cassinelli, A., Perrin, S., and Ishikawa, M., *Smart laser-scanner for 3D human-machine interface*, CHI’05 extended abstracts, pp.1138-1139, 2005.

[2] Hann, J., *Low-cost Multi-touch Sensing through Frustrated Total Internal Reflection*, ACM UIST'05, pp.115-118, 2005.

[3] Hodges, S., Izadi, S., Bulter, A., Rrustemi, A., Buxton, B., *ThinSight: versatile multi-touch sensing for thin form-factor displays*, ACM UIST'07, pp.259-268, 2007.

[4] Krueger, M.W., *Artificial Reality II*, Addison-Wesley, 1990.

[5] Matsushita, M., Rekimoto, J., *HoloWall: Designing a Finger, Hand, Body, and Object Sensitive Wall*, Proceedings of UIST'97, 1997.

[6] Microsoft Surface, http://www.surface.com

[7] Toshiba Matsushita, *LCD with Finger Shadow Sensing*, http://www3.toshiba.co.jp/tm_dsp/press/2005/05-09-29.htm