Mid-Air Interactive Display Using Modulated Display Light

Zoran Zivkovic and Hendriek Groot Hulze
Trident Microsystems (Europe), The Netherlands

***Abstract*--A low cost solution is presented for transforming a regular display into an interactive display. The display light is modulated using frequencies invisible to human eye. A camera captures images and the known light modulation is used to isolate the display generated light. The resulting demodulated images do not depend on the ambient light conditions. Furthermore, the objects close to the display are clearly visible and easy to detect/track. Touch-screen like mid-air interaction with the display can be realized by robustly detecting human hands when they are close to the display.**

I. INTRODUCTION

Touch-free technologies for user-display interaction are becoming increasingly interesting for the consumers. Instead of swiping your fingers over the display, you can simply swipe the air, leaving your display grease and germ free.

Low cost touch-free interaction can be realized by adding just a cheap web camera. The camera images are analyzed and the user hands/fingers detected [1]. However, robust and reliable user/object detection using a camera is difficult. The major reason is that complex algorithms are required to detect the objects of interest in highly variable environments such as a typical living room. Furthermore, the highly variable light conditions that can be expected in the viewing environment make the detection even more difficult.

A number of systems were presented that control the light conditions by using an invisible infra-red (IR) light source and an IR camera [2]. In the recently presented system [3][4] the IR light source and camera are placed behind the transparent OLED display using a special lens system The strength of the light source was such that only the objects close to the display are illuminated and in this way easy to detect and track.

This paper describes how similar "invisible" controlled light conditions can be achieved by temporal modulation of the light coming from the display itself. As result, a mid-air interactive system robust to ambient light conditions can be realized using a standard display and a webcam. The main blocks of the system are illustrated in Figure 1 and explained in more detail next.

II. LIGHT MODULATION/DEMODULATION

The display light can be modulated/demodulated in various ways. The frequency of the modulation should be high, i.e. higher than 60Hz, such that it is invisible to the human eye. For different types of display technologies there are different ways to introduce the high frequent light modulation. Light pulses will be used here. Let $I_{on}(t)$ denote the amount of light observed during the light pulse at time *t* by the camera at a certain pixel corresponding to an object in the scene. A part

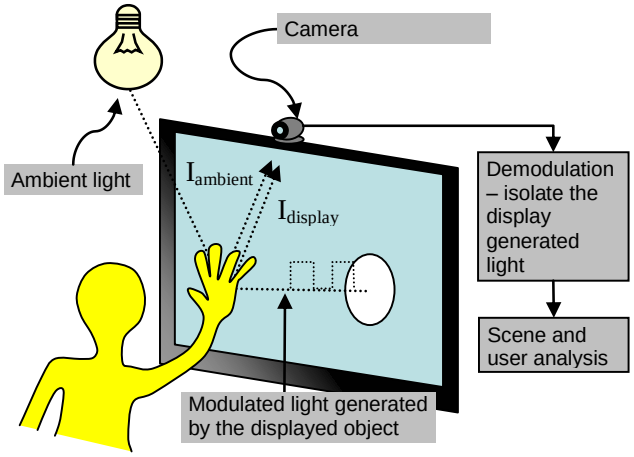


Fig. 1. Interactive display system using modulated display light.

of the observed reflected light will be from the ambient light sources $I_{ambient}(t)$ and another part will be the reflected display light $I_{display}(t)$, see Figure 1:

$$I_{on}(t) = I_{ambient}(t) + I_{display}(t). \quad (1)$$

After the light pulse at time *t+dt* the ambient part is observed:

$$I_{off}(t+dt) = I_{ambient}(t+dt). \quad (2)$$

If the time period *dt* is very short and the objects in the scene are moving slowly, the influence of the ambient light can be removed by subtraction, which is a simple demodulation:

$$I_{display}(t) \approx I_{on}(t) - I_{off}(t+dt). \quad (3)$$

III. A PRACTICAL SYSTEM REALIZATION

Low cost cameras operate in rolling shutter mode. The image rows are captured sequentially. Figure 2 presents four images captured by a low cost camera. Once the last row of one image is captured, the camera starts capturing the first row of the next image. The images are stacked on top of each other to illustrate this "rolling" nature of the capture process. The images are captured at 120 frames per second. The person was standing in front of an LCD display. The display backlight was controlled to emit light in 90Hz square. The square pulses are depicted in Figure 2 aligned to the time where image lines were captured. Each image row was captured by integrating light during the exposure time of the camera. To visualize the effects of the pulses and the exposure time on the observed

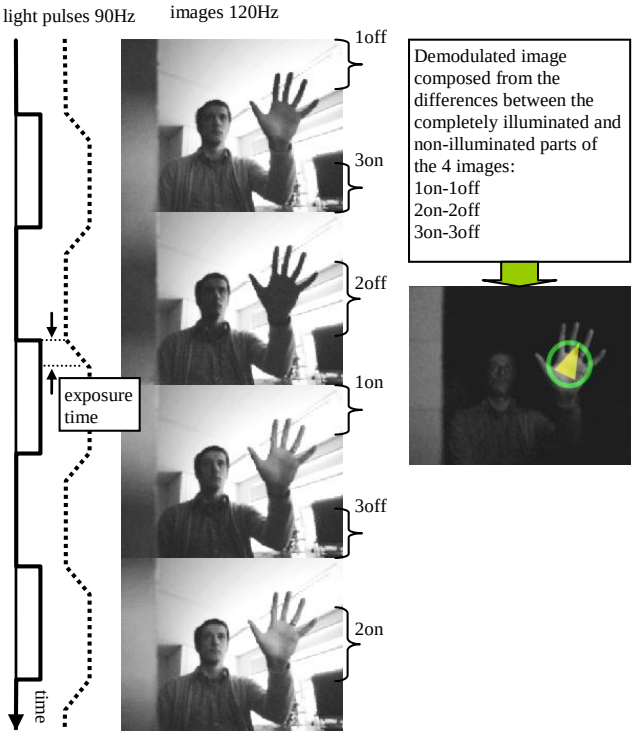


Fig. 2. Illustration of modulation/demodulation using 90Hz square pulses and a 120Hz rolling shutter camera. The square pulses in time are presented on the left synchronized with the time where image lines were captured. The dotted line illustrates the influence of the camera exposure time.

images, a single color object is placed in front of the camera, visible on the right side of the captured images in Figure 2. Because of the "rolling" exposure time, the sharp edges of the light pulses are observed as linear segments, illustrated by the dotted line in Figure 2. The captured images have some lines measuring the maximum of the display emitted light by integrating during the "on" period of the light pulses. Some image lines will measure only the ambient light, indicated as "off" period lines in Figure 2. Subtracting the "off" image line values from the "on" image lines, see equation (3), an image is obtained where the influence of the ambient light is removed, see Figure 2. It can be shown that for 90Hz pulses and 120Hz image capture a complete "demodulated" image can be reconstructed from 4 original captured images.

IV. HAND DETECTION

The demodulated images are passed through an image segmentation algorithm to detect blobs that could be the user hands close to the display. Shape of each detected blob is described by some shape descriptors. In our experiments we used the Fourier descriptors [6]. The shape descriptors are used to filter only the hand shaped blobs. A Support Vector Classifier trained on the example hand shape data is used for the filtering. Detected hand and its estimated orientation are illustrated in Figure 2 by the overlaid circle and the triangle.

Three blob detection algorithms are evaluated where the same classifier is used to filter out only the hand shaped ones. Simple threshold and connected region detection is tested first. This already gives reasonable recognition results, see Figure 3.

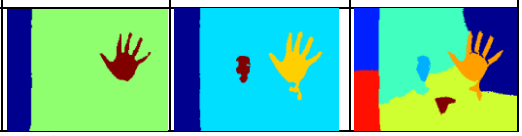
Segmentation algorithm	Threshold	Mean shift [7]	Graph cut [8]
Processing time	1ms	5ms	5000ms
Recognition accuracy	96.1%	99.4%	99.3%
Example image segmentation			

Fig. 3. Evaluation of image segmentation algorithms for the hand detection. The set of videos used for the evaluation contained 5000 images of 5 different people in different light conditions and 5000 random images.

However, this method is sensitive to the proper choice of the threshold. Using mean-shift segmentation [7] increases the computation costs but the recognition is better. Complex graph cut segmentation [8] does not improve the results.

V. CONCLUSIONS

If the display light is modulated, the known modulation can be used to reconstruct the scene as if it was illuminated only by the display. The strength of the light from the display is such that only the objects close to the display are illuminated and in therefore easy to detect and track.

A practical realization is described here. The modulation is introduced by controlling the backlight of a standard LCD display to emit light in square pulses. A low cost camera was used to capture the images. Demodulation scheme based on 4 images is described. The hands close to the display are detected in the demodulated images. Various blob detection algorithms were tested and evaluated, showing that reliable hand detection can be achieved.

The presented system can be used for mid-air interaction with the display similar to much more expensive systems based on IR light [2][3][4]. Recently, the low cost IR light depth cameras were introduced [5] offering rich touch-free interaction possibilities. The system presented here will detect only objects close to the display and offer only interaction similar to a touch screen. However, the costs of the system presented here are potentially much lower since only a standard camera is used.

REFERENCES

[1] W.T. Freeman, C.D. Weissman, "Television Control by Hand Gestures", IEEE Intl. W. on Automatic Face and Gesture Recognition, June, 1995.
[2] A. D. Wilson, "Touch Light: An imaging touch screen and display for gesture-based interaction", Proceedings of the 6th ACM International Conference on Multimodal interfaces, pp. 69 – 76, 2004.
[3] M.J. Large, T. Large, A.R.L Travis, "Parallel Optics in Waveguide Displays: A Flat Panel Autostereoscopic Display," Journal of Display Technology, vol.6, no.10, pp.431-437, Oct. 2010.
[4] Microsoft Applied Sciences, "The Wedge: Seeing Smart Display". 2010. http://www.microsoft.com/appliedsciences/content/projects/
[5] Microsoft Xbox Kinect, www.xbox.com/kinect
[6] M.F. Wu, H.T. Sheu, "Contour-based correspondence using Fourier descriptors", Vision, Image and Signal Processing, IEE Proc., 1997.
[7] D. Comanicu, P. Meer: "Mean shift: A robust approach toward feature space analysis". IEEE Trans. Pattern Anal. Machine Intell., May 2002.
[8] J. Shi, J. Malik, "Normalized Cuts and Image Segmentation", IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(8), 888-905, 2000.