Journal of Information Display
Vol. 13, No. 2, June 2012, 67–71

Taylor & Francis
Taylor & Francis Group

Aspects of a head-mounted eye-tracker based on a bidirectional OLED microdisplay

Judith Baumgarten[a]*, Tobias Schuchert[b], Sascha Voth[b], Philipp Wartenberg[a], Bernd Richter[a] and Uwe Vogel[a]

[a]*Center for Organic Materials and Electronic Devices Dresden (COMEDD), Fraunhofer IPMS, Maria-Reiche-Straße 2, 01109 Dresden, Germany;* [b]*Autonomous Systems and Machine Vision, Fraunhofer IOSB, Fraunhoferstraße 1, 76131 Karlsruhe, Germany*

(*Received 23 December 2011; Revised 13 February 2012; Accepted for publication 4 March 2012*)

In today's mobile world, small and lightweight information systems are becoming increasingly important. Microdisplays are the base for several near-to-eye display devices. The addition of an integrated image sensor significantly boosts the range of applications. This paper describes the base-building block for these systems: the bidirectional organic light-emitting diode microdisplay. A small and lightweight optic design, an eye-tracking algorithm, and interaction concepts are also presented.

Keywords: microdisplay; eye-tracking; head-mounted display (HMD); OLED

1. Introduction

Currently, microdisplays are the main parts of electronic viewfinders or head-mounted video glasses. The advantages of microdisplays based on organic light-emitting diodes (OLEDs) are high brightness and contrast ratios without the use of backlight. The OLED-on-CMOS (complementary metal oxide semiconductor) technology allows the integration of the display and image sensors on one CMOS chip and hence extends their application range. As an example, a system that allows head-mounted eye-tracking in combination with augmented reality is presented in this paper. The following sections deal with the system's basic hardware and software components.

2. Microdisplay

A bidirectional microdisplay combines camera and display functionalities into one application-specific integrated circuit (ASIC) [1]. It is an active-matrix OLED display that can also capture images. As such, the camera and display matrices intersect with each other and are thus both located within the active area of the microdisplay. Figure 1(a) shows a realized arrangement of the display pixels and photodiodes (PDs) in a full-color display. Each pixel consists of a red (R), a green (G), and a blue (B) subpixel and additionally a white (W) subpixel or PD. Four pixels are grouped into 2×2 clusters where the subpixels are permuted in such a way that the four white subpixels or PDs form the center of the cluster. Each fifth cluster in the horizontal and vertical directions contains PDs. As such, this arrangement applied at a 640×480 RGB display resolution ends up in a monochrome 128×96 camera resolution. Due to the current limitations in OLED processing, the colors are achieved by color filters above a white OLED.

A critical point in combining the display and camera in one ASIC is the crosstalk between both parts of the functionality. On the one hand, there is the optical crosstalk between the emitting and capturing pixel cells and, on the other hand, there is the electrical crosstalk in the subsidiary CMOS circuitry. If not covered correctly, the quality of the captured images will significantly decrease. One possible way of avoiding crosstalk is by applying time-sequential operation. Figure 1(b) shows the timing diagram for the camera and display parts in the parallel (V1) and sequential (V2) operation modes. In the parallel operation mode, the exposure phase as well as the emitting and programming of the display take place at the same time. This results in high crosstalk from the display to the captured image. In the sequential mode, the exposure starts when the display is completely programmed and the emitting phase is over. After the pre-defined emitting time, the OLED is switched off so that optical crosstalk could not occur. To avoid electrical crosstalk, the complete exposure time has to be between the last valid pixel data and the vertical synchronization signal so that there would be no electrical events in the display part of the CMOS circuitry. The drawback of the sequential approach is the need for a higher pixel clock for managing the same frame rate. The required speed-up directly depends on the ratio of the camera and display resolutions as well as on the maximum exposure time.

*Corresponding author. Email: judith.baumgarten@ipms.fraunhofer.de

ISSN 1598-0316 print/ISSN 2158-1606 online

http://dx.doi.org/10.1080/15980316.2012.674981
http://www.tandfonline.com

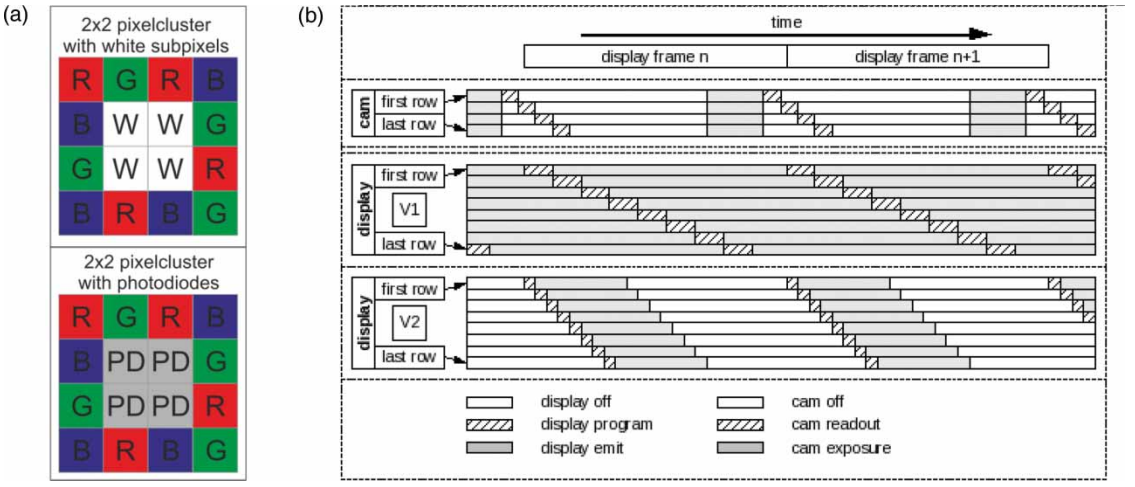

Figure 1. (a) Pixel layout. (b) Timing schemes for the parallel (V1) and sequential modes (V2).

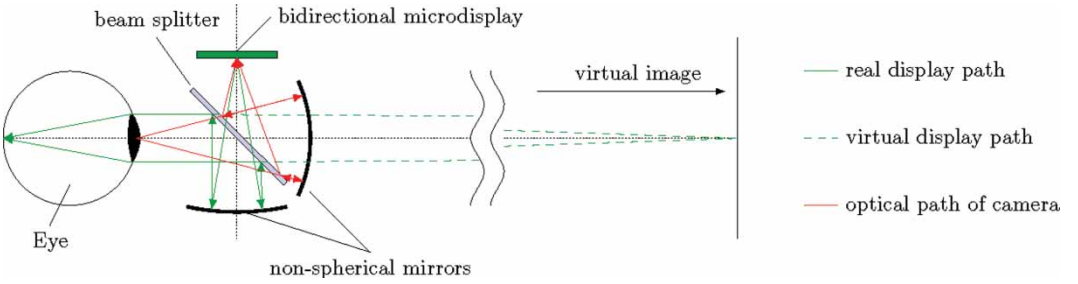

Figure 2. Bidirectional optic for HMD application.

3. Optical system

A bidirectional microdisplay as described above needs an optical system that handles both optical paths: that of the display and that of the camera. One difficulty therein lies in the necessity of two focal planes. The display path has to be focused on the retina, whereas the camera path has to be focused on the eye surface to provide optimal images for the eye-tracking algorithm. To avoid noise, it is further recommended that separate wavelengths be used for both paths. As the display path, of course, needs to be transmissive for visible light, this area of the spectrum is not feasible for the camera path. Due to the noxious effects of ultraviolet light [2], near-infrared (NIR) illumination of the eye scene is commonly used in eye-tracking systems [3]. Another advantage of NIR illumination with respect to eye-tracking is the improved contrast between the important characteristics of the eye [4]. Morimoto *et al*. [5] showed that the use of different arrangements of NIR diodes results in different appearances of the pupil within the captured image.

With respect to the facts discussed previously, the optical system shown in Figure 2 was designed for HMD application. The system consists of two non-spherical mirrors, a beam splitter, and the microdisplay. The non-spherical mirrors have a dichroic coating reflecting either visible light (380–780 nm) in the display path or NIR light (780–1100 nm) in the camera path. The visible light of the OLED passes the beam splitter and is reflected by the down-side non-spherical mirror and the bottom side of the beam splitter to the eye. The user perceives the arising virtual image straight ahead, at a distance affected by the curvature of the non-spherical mirror. Additionally, the visible light of the environment passes the non-spherical mirror at the front as well as the beam splitter, so that the system allows the projection of a virtual image within the natural vision of the user. The NIR diodes for illuminating the eye scene are placed outside the optical system and thus also outside the optical axes. Therefore, the pupil does not reflect much of the light and thus appears darker than the rest of the captured image [5]. As the diodes emit light at an 850-nm wavelength, the contrast between the pupil and the iris is improved [4]. After the reflection on the eye surface, the NIR waves pass the beam splitter and are reflected by the non-spherical mirror at the front. The top side of the beam splitter finally redirects the waves to the microdisplay.

A see-through optic based on the design described above was produced, wherein the virtual image is projected at a distance of 1 m and takes a 32 × 24°-visual-angle field of view. The overall system, including the optic, microdisplay, and head band, weighs only about 295 g.

4. Eye-tracking algorithm

The drawback of the lightweight and compact design of the previously described bidirectional microdisplay is the rather low resolution of the recorded eye scene compared with the standard eye trackers. Therefore, the most significant tasks for eye-tracking and gaze analysis are to improve the image quality and to make the subpixel detection of the pupil and the corneal reflections more robust over time.

The eye-tracker has two functions, as shown in Figure 3(a): the pre-processing of the input data and the subsequent eye-tracking. The first step of the pre-processing is gray-value correction. This step was added to be able to cope with the side effects of the bidirectional microdisplay's pixel layout (cmp. Section 2). Figure 3(b) shows that the gray values of the PDs depend on their position in the clusters. Therefore, a high-pass filter was introduced, which copes with the mean gray-value shifts according to the position in the cluster. That is, the mean gray value for each of the four positions in the clusters is calculated and is afterwards subtracted from the appropriate gray values. Figure 3(b) shows the images after gray-value correction. A gray-value offset was added to make the changes more visible. This offset led to an image that seemed to have a lower contrast compared with the input image. This, however, was only an illusion and did not affect the further computation.

To reduce the noise and to enhance the edges, the corrected image was smoothed in the next pre-processing step. The experiments showed that a 3×3 median filter yields good results. The contrast of the images was still rather low, however, making edge detection, which is essential for the pupil detection algorithm, difficult. Therefore, histogram equalization was applied to adapt the contrast and to improve pupil detection. Image regions containing artifacts (e.g. light reflections and borders) were excluded from the equalization process because they heavily influence the equalization results. The result of the histogram equalization is shown in Figure 4(a).

In the next step, the eye-tracking algorithm detected the pupil and corneal reflections to allow robust gaze tracking. First, the corneal reflections were detected by a row-based thresholding algorithm. One characteristic feature of the NIR-LED arrangement is that the two reflections are almost on the same row in the eye image. Therefore, the algorithm looks for the rows with the brightest gray values in the image, under two constraints: there must be a minimum row pixel distance and a maximum column pixel distance between the two detected reflections. The identification of the correct center of the reflections is also difficult due to the blurring coming from the curvature of the eye and the tear fluid. Figure 3(b) shows a typical input image without

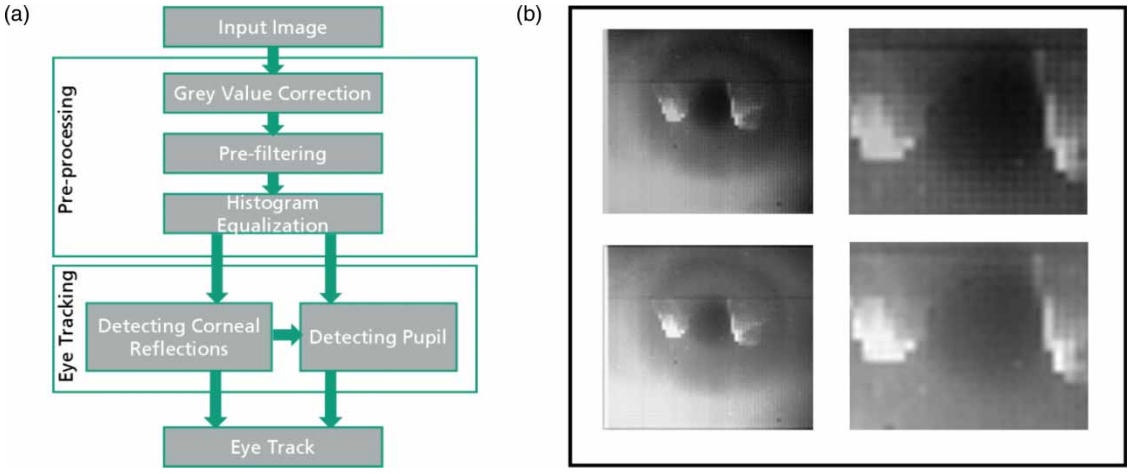

Figure 3. (a) Flow diagram for eye-tracking. (b) Effect of gray-value correction: upper row – input image and close-up view and bottom row – gray-value-corrected image and close-up view.

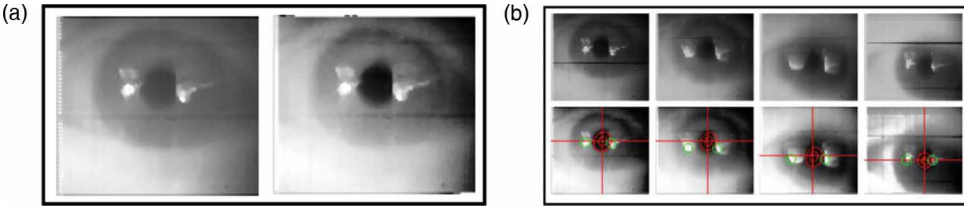

Figure 4. (a) Effect of histogram equalization. (b) Input images of the four scenarios and output images with pupil (red) and corneal-reflection (green) detection.

pre-processing. The blurred reflections are clearly visible. The algorithm estimates the centers by first calculating the smallest distance between the two reflections and then calculating the center of mass of the brightest pixels in a pre-defined neighborhood.

The detected corneal reflections define the window for pupil detection. That is, the center of the pupil is assumed to be between the two corneal reflections. This assumption is feasible due to the arrangement of the NIR-LEDs. An adaptive threshold yields the most likely pupil contours. Starting from these contours, a variant of the Starburst algorithm [6] was applied. The Starburst algorithm was adapted for low-resolution images, and the following changes were made: (1) the detected contours from the previous thresholding were first used; (2) the gradients were computed based on the optimized filters; and (3) only the most likely edge candidates were used for pupil estimation. These steps improve pupil detection and allow more robust subpixel accuracy pupil center estimation. The algorithm was tested in four scenarios: in blue and brown eyes and with the eyes focused and not focused on the eye-tracker.

Figure 4(b) shows the input images of the four scenarios and the corresponding results of the eye-tracking algorithm. The pupil (red circle) and the two corneal reflections coming from the NIR-LEDs (green circles) were detected in all the scenarios. The eye-tracking algorithm showed good accuracy and allowed the use of the gaze direction depending only on the relative positions of the pupil and the corneal reflections for the interaction tasks, without calibration.

5. Eye-tracking-based interaction concepts

A wide variety of eye-tracking applications exist in areas such as psychology, marketing, and scientific and medical research. These eye-tracking applications can be classified into two types: active and passive. Active eye-tracking allows the user to interact with a device through eye movements or blinks. The typical, well-known active applications are eye typing or gaze writing, where the user can formulate words and sentences only with the use of his or her gaze direction combined with fixations. Passive eye-tracking applications, on the other hand, analyze the unconscious movements of the user's eyes, based on which it is possible to gain information about the user's actual condition, such as his or her fitness or mood.

Compared with other head-mounted units, the bidirectional device offers the possibility of interacting with the device itself and its projected augmented reality, based on eye-tracking.

In the following, three interaction concepts are presented, which can be used to design active eye-tracking applications.

5.1. *Split-/full-screen control*

Figure 5(a) shows the four video streams via a so-called *split screen*. Here, the screen is divided into four equal parts, where every part shows one stream. The fact that the user focuses either on the display screen or the reality is important for this interaction concept. By focusing on the reality through the display optics, his or her pupil stays as far as possible on the middle of the screen. In this case, it is assumed that he or she is not interested in one of the split-screen video streams. If the focus is shifted to one of the small image streams, the eyes move to one of the corners. If the user's focus stays there for a pre-defined time, interest in the corresponding video stream is assumed, and the display switches to a full-screen display of this video stream. To get back to the split screen, the user has to focus on one of the outer display borders.

5.2. *Text scrolling*

To be able to read an extensive multiline text on the display, it is necessary to split the text into several parts, which are shown on the display. The number of visible lines is limited by the available display resolution and information density. It is a fact that as humans naturally read a text, their gaze moves along the line that is being read. In some situations, the user's gaze jumps from pre-scanning the scene to looking for a new target. In the case of text reading, this behavior is limited. With this fact in mind, it is possible to use this eye behavior as a passive interaction input. The display shows a scrollable text that moves vertically when the user is looking at the upper and lower borders of the screen. The speed of scrolling is determined by the distance between the gaze and the display center. That is, the normal reading behavior is scrolling the text slowly in the reading direction. If the user wants to skip several text paragraphs, his or her gaze jumps to the end or beginning of the text, and the text movement speed increases. Furthermore, the text movement speed and minimal-distance adaption rates, which start the text scrolling, have to be appropriate so that the natural reading behavior and experience

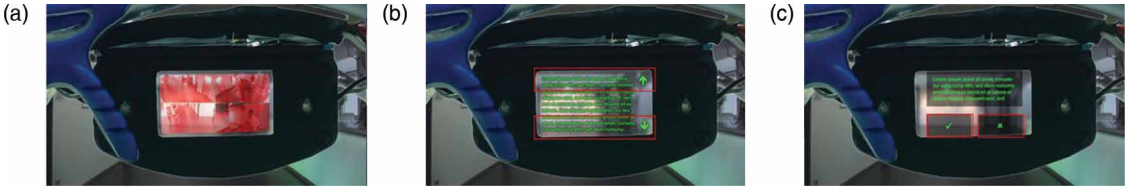

Figure 5. Interaction examples: (a) split-/full-screen control; (b) text scrolling; and (c) exception handling.

will not be disrupted. Figure 5(b) shows this interaction concept.

5.3. Exception handling

To prompt the user to make a decision in the case of an exception, a combination of text display and active eye activation can be used. The upper part of the display shows the information as text, and the lower part visualizes two virtual buttons. After reading the upper exception text, the user can focus on one of the lower virtual buttons to signal his or her choice to the system. Figure 5(c) shows an example of this dialog type.

6. Conclusion

In this paper, the main parts of a mobile eye-tracking system were described. The problems arising from the usage of a bidirectional microdisplay were discussed, and solutions to such problems were presented. Furthermore, an optic concept for head-mounted displays was explained with respect to the base aspects of the described eye-tracking algorithm. Finally, some interaction concepts were presented.

The future work will include the improvement of eye-tracking on low-resolution near-to-eye data (e.g. by modeling the structure of the corneal reflections) and analysis of the eye-tracking and gaze-tracking accuracy. Furthermore, the resolution of the display and camera of the bidirectional microdisplay will be increased and more compact optic concepts will be evaluated.

References

[1] B. Richter, U. Vogel, R. Herold, K. Fehse, S. Brenner, L. Kroker, and J. Baumgarten, Bidirectional OLED Microdisplay: Combining Display and Image Sensor Functionality into a Monolithic CMOS Chip, ISSCC DIG TECH PAP I, 314 (2011).

[2] F.J.G.M. van Kuijk, Environ. Health Persp. **96**, 177 (1991).

[3] D.W. Hansen and Q. Ji, IEEE Trans. Pattern Anal. **32**, 478 (2010).

[4] E.C. Lee and K.R. Park, Mach Vision Appl. **20**, 319 (2008).

[5] C.H. Morimoto, D. Koons, A. Amir and M. Flickner, Image Vision Comput. **18**, 331 (2000).

[6] D. Li, D. Winfield, and D.J. Parkhurst, Starburst: A hybrid algorithm for video-based eye-tracking combining feature-based and model-based approaches, Workshop CVPR IEEE 3, 79 (2005).