CMOS image sensor for high speed and low latency eye tracking

Junichi Akita^{a)}

Kanazawa University,

Kakuma, Kanazawa, Ishikawa 920–1192, Japan

a) akita@is.t.kanazawa-u.ac.jp

Abstract: Eye tracking, or detecting where the user is looking at, is expected as a new type of user interfaces, with including the phenomenon of rapid eye movement, so called saccade. However, real-time tracking of saccade is difficult with the conventional image processing systems for their long processing time and latency against the speed of saccade. In this paper, we describe the design of the CMOS image sensor for eye tracking with high speed and low latency capability, as well as its evaluation results.

Keywords: CMOS image sensor, eye tracking, saccade, column parallel architecture

Classification: Integrated circuits

References

- [1] E. A. Hoffman and J. V. Haxby: “Distinct representations of eye gaze and identity in the distributed human neural system for face perception,” *Nat. Neurosci.* **3** (2000) 80 (DOI: [10.1038/71152](https://doi.org/10.1038/71152)).
- [2] J. Triesch, *et al.*: “What you see is what you need,” *J. Vis.* **3** (2003) 86 (DOI: [10.1167/3.1.9](https://doi.org/10.1167/3.1.9)).
- [3] J. Watanabe, *et al.*: “Study of remote saccade detection technique based on retroreflective features of the retina,” *J. Virtual Reality Soc. Japan* **9** (2004) 105 (DOI: [10.18974/tvrsj.9.1_105](https://doi.org/10.18974/tvrsj.9.1_105)).
- [4] Tobii AB: Tobii Pro X2-60, <https://www.tobii.com/en/product-listing/tobii-pro-x2-60/>.
- [5] nac Image Technology, Inc.: EMR-9, <http://www.eyemark.jp/product/emr-9/>.
- [6] Tobii AB: Tobii Pro Spectrum, <https://www.tobii.com/en/product-listing/tobii-pro-spectrum/>.
- [7] T. Takegami, *et al.*: “An algorithm for model-based stable pupil detection for eye tracking system,” *Syst. Comp. Jpn.* **35** (2004) 21 (DOI: [10.1002/scj.10544](https://doi.org/10.1002/scj.10544)).
- [8] J. Akita, *et al.*: “Column-parallel vision chip architecture for high-resolution line-of-sight detection including saccade,” *IEICE Trans. Electron.* **E90-C** (2007) 1869 (DOI: [10.1093/ietele/e90-c.10.1869](https://doi.org/10.1093/ietele/e90-c.10.1869)).
- [9] S. Yamamura and J. Akita: “Design and preliminary evaluation of high-speed line-of-sight calculation CMOS image sensor,” *Electron. Commun. Jpn.* **101** (2018) 47 (DOI: [10.1002/ecj.12062](https://doi.org/10.1002/ecj.12062)).

- [10] H. Kawakami, *et al.*: “Column-parallel architecture for line-of-sight detection image sensor based on centroid calculation,” *ITE Trans. Media Technol. Appl.* 2 (2014) 161 (DOI: [10.3169/mta.2.161](https://doi.org/10.3169/mta.2.161)).

1 Introduction

Eye tracking, or detecting where the user is looking at, is expected as a new type of user interfaces [1]. The eyeball often moves very rapidly, which is called as saccade, and it is also expected to be applied for another type of user-interfaces [2]. However, real-time tracking of saccade is difficult with the conventional image processing systems for their long processing time and latency against the speed of saccade, whose speed reaches up to 700 [deg/sec], which required frame rate of over 200 [fps] to capture [3]. Most of the commercial eye tracking system capture the eye image at the video rate of 60 [fps] [4], that is too slow to track the saccade. There are several eye tracking systems at a high frame rate to capture the saccade, however, their output latency, or the processing time to obtain the result, is several frames [5, 6], and they cannot track the saccade in real time.

In this paper, we describe the design of the CMOS image sensor for eye tracking that realize the high speed and low latency capability for the real time saccade tracking. It employs the column-parallel processing architecture to achieve the high speed processing, the low latency, and the high resolution. We also describe the evaluation results of eye tracking using the designed CMOS image sensor.

2 Design of CMOS image sensor for eye tracking

2.1 LoS calculation algorithm

Fig. 1. Example of an infrared eye image

Eye tracking is performed by calculating the line of sight (LoS), or the eye direction. The LoS can be calculated from the position of the pupil in the infrared image of the user's eye, which is observed as a black area [7, 8], as shown in Fig. 1. The position of the pupil, whose shape is a circle, can be defined as a centroid of the pupil (digitized black) area at the sub-pixel accuracy. The centroid of the area, $(\bar{x}, \bar{y})$, with the binary flag of p_{xy} for the pixel composing the pupil area, can be calculated as follows.

$$(\bar{x}, \bar{y}) = \left(\frac{\sum_x \sum_y x p_{xy}}{\sum_x \sum_y p_{xy}}, \frac{\sum_x \sum_y y p_{xy}}{\sum_x \sum_y p_{xy}} \right) \quad (1)$$

As shown in eq. (1), the position of the pupil can be calculated from the values of S , SX and SY as follows.

$$S = \sum_x \sum_y p_{xy}, \quad SX = \sum_x \sum_y x p_{xy}, \quad SY = \sum_x \sum_y y p_{xy} \quad (2)$$

Thus, we can obtain the LoS by using the values of S , SX , and SY . The dividing calculation for SX and SY by S requires low computing power since it is performed just once in each frame. Therefore, we focus on the high-speed calculations of S , SX , and SY , which require the high computing power to realize high speed, low latency eye tracking.

Note that the actual LoS, where the user is looking at, is calculated from the center of the pupil with the calibration parameters measured in advance to the operation.

2.2 Image sensor architecture for LoS calculation

Fig. 2. Column-parallel image sensor architecture for eye tracking

Fig. 2 shows the CMOS image sensor architecture for high speed, low latency, and high resolution LoS calculation [9]. The architecture is composed of image sensor (pixel array) part and the column-parallel LoS calculation circuit, PE (processing element). The pixel array is the standard pixel array similar to the conventional CMOS image sensor whose size is 6.0 [μm], which is suitable for realizing practical high resolution (VGA (640 $\times$ 480 [pixels]) or above). The column-parallel LoS calculation circuit, PE, is placed at each column. One PE processes the signals from all the pixel in one column in order.

Fig. 3 shows the PE architecture of one column. PE is composed of S , SX , and SY , that perform to the y -directional accumulation, $\sum_y$, of S , SX , and SY , respectively, in eq. (2). PE performs the calculation of $\sum_y$ by using the flag, p_{xy} generated by the comparator, x coordinate value of the column, and y coordinate value of the pixel accessed. The PE's component of S , SX , and SY are composed of single-bit accumulator with read-out buffer as shown in Fig. 4, to form the ripple-carry accumulator [9]. Although the processing time for one column is proportional

Fig. 3. PE architecture of one column

Fig. 4. Circuit of single-bit PE of (a) S, and (b) SX and SY.

to the size of row, each PE processes each column's signal in parallel at high speed with using the conventional digital circuit implementation, such as 10 [MHz]. This realizes the total high processing frame rate, such as 500 [fps] or above so that the saccade can be tracked.

After all the pixels are accessed, the x -directional accumulation, $\sum_x$ of S , SX , and SY , respectively, in eq. (2) of PE in each column, by the accumulator connected to PEs in Fig. 2. Here, we obtain the calculation results of S , SX , and SY in eq. (2) in the accumulator.

Fig. 5. Operation timing

Fig. 5 shows the operation timing of the designed image sensor. The access to each pixel is composed of the integration operation, the generating electrons from the received photons, followed by the read out and the digitize operations by the comparator. The y -directional operation finishes after the pixel read out operations, followed by the x -directional accumulation operations. These operations finishes in single frame. We obtain the calculation result at the end of each frame, thus, the possible lowest output latency of single frame is achieved, which is smaller than the conventional image processing-based eye tracking systems.

2.3 Design of CMOS image sensor for eye tracking

The author has already tried to design the CMOS image sensor for eye tracking based with the proposed column-parallel architecture. However, the fabricated CMOS image sensor has some problems in operation [9]. Based on this first version, we designed the second version with improvements to solve the problems in operation.

The most important point in the design revision is to apply the clock tree for PEs. In the evaluation of the first version, the PEs output the random values, and we estimated its reason as the meta-stable operation of flip flops for lack of the proper clock buffering. We inserted the clock tree as shown in Fig. 6 to obtain the proper clock signals in each flip flop. The number of clock buffers and the drive capability of each clock buffer are determined based on the load capacitance estimated from the signal wire length and the number of driven flip flops¹.

Fig. 6. Inserted clock tree

Fig. 7. Fabricated CMOS Image sensor for eye tracking

Fig. 7 shows the fabricated CMOS image sensor for eye tracking using CMOS 0.18 μm CIS process. The number of pixels is 640×480 , and the chip size is $4 \times 5 \text{ mm}$.

3 Experiment on eye tracking

3.1 Experimental setup

We have carried out the eye tracking experiment using the fabricated CMOS image sensor. Fig. 8 shows the designed evaluation board for the test on the fabricated CMOS image sensor. The control signals are generated by the FPGA, and the calculated results of S , SX , and SY are transferred to PC. The raw image output is also captured by A/D converter for debugging.

¹This work is supported by VLSI Design and Education Center (VDEC), the University of Tokyo in collaboration with Cadence Design Systems, Inc. and Mentor Graphics, Inc.

Fig. 8. Evaluation board

Fig. 9. Experimental setup

Fig. 9 shows the experimental setup for eye tracking. The subject is asked to put his chin to the chin-rest. The camera is set to capture his eye area, and three IR-LEDs are used for illumination. Fig. 10 shows the captured raw images, where the center of the pupil is indicated by the cross point of red lines. The size of the pupil is also indicated as the box of the green lines, as the bounding box whose size is calculated from the pupil area, S . As shown in Fig. 10, the position and the size of the pupil are obtained by using the fabricated CMOS image sensor. Here, we manually adjusted the threshold voltage of the comparator for digitizing the pupil to obtain the adequate digitized image for the pupil.

Fig. 10. Captured eye image and the calculated pupil's position and size

Note that there are some black regions other than the pupil, that result in the errors in the calculation results of the pupil's center. The region of interest (ROI) should be adequately configured so as to access to the eye region. In this experiment, the ROI is manually configured. It is also notable that we can calibrate combining the pupil's center (including errors) and the actual LoS without configuring ROI to obtain the correct LoS. [10].

3.2 Experimental results on eye tracking

Fig. 11. Recorded eye motions. The x -coordinate of the pupil (a) and the angular velocity of the eye (b).

We carried out the eye tracking experiment at the frame rate of 315 [fps], with exposure time of 980 [μ s]. The subject is asked to read the sentences for three lines, and the position of the pupil is recorded during this task. Fig. 11(a) shows the obtained x -coordinate of the pupil, calculated as SX/S . Note that the smaller x -coordinate value corresponds to the right direction of the eye. We can observe three pairs of the left-to-right eye motion, that correspond to reading three lines at 2000~4000 [ms], 6000~8000 [ms], and 10000~12000 [ms], where some saccade motions are included. We also observed two sudden eye motions from left-to-right followed by right-to-left at 8500 [ms] and 13000 [ms], which is expected to be corresponding to the blinks.

We also calculated the angular velocity of eye motion from the recorded eye motion. The radius of our eye ball is almost constant, approximately 12 [mm] for adults, regardless of the age and the race. The relation of the pixel size and the

Fig. 12. Model of eye motion angle calculation

actual size around eye can be calculated from the number of the pixels in the captured image and the measured physical size of the eye area. In the experimental setup in Fig. 9, one pixel size is corresponding to 0.1 [mm] around the eye area. We can calculate the eye motion angle from the motion of the the pupil center as shown in Fig. 12 as follows.

$$\Delta\theta [\text{rad}] = \tan^{-1} \frac{\Delta x}{R_e} \sim \frac{\Delta x}{R_e} \quad (3)$$

Here, $\Delta\theta$, Δx , and R_e are the eye motion angle, the difference of the x -coordinate of the pupil center, and the size of eye ball, respectively. Note that Δx and R_e are represented in the unit of the pixel, and R_e equals to 120 [pix] for 12 [mm]. Fig. 11(b) shows the calculated angular velocity of eye motion with the average of five successive samples.

Fig. 13. Magnified eye motions in Fig. 11. (a)(b) The x -coordinate of the pupil and (c)(d) the angular velocity of the eye.

Fig. 13(a) and (b) show the magnified samples of the pupil positions in Fig. 11, and Fig. 13(c) and (d) show their corresponding angular velocities of eye motion in Fig. 13(a) and (b), respectively. We can observe the angular velocity of eye motion which larger than 100 [deg/s], up to 250 [deg/s], which we can judge them as saccade motions.

4 Conclusion

In this paper, we described the design of the CMOS image sensor for eye tracking that realize the high speed and low latency capability for saccade tracking. We also described the evaluation results of eye tracking using the fabricated CMOS image sensor, and shown the saccade tracking result at single frame latency.