SICE Annual Conference 2007
Sept. 17-20, 2007, Kagawa University, Japan

Development of Head-Mounted Eye Tracking System achieving Environmental Recognition Ability

Tetsuro Sugita[1]*, Satoshi Suzuki[2], Julian Kolodko[3] and Hiroshi Igarashi[4]

[1]Department of Electronic Engineering, Tokyo Denki University, Tokyo, Japan
(Tel : +81-3-5280-3281; E-mail: sugita@ab.ham.coe.dendai.ac.jp)
[2]Department of Robots and Mechatromics, Tokyo Denki University, Tokyo, Japan
(Tel : +81-3-5280-3916; E-mail: ssuzuki@fr.dendai.ac.jp)
[3]Company Name Undisclosed(E-mail:juliankbox-contact@yahoo.com.au)
[4]Department of Electrical Electronic Engineering, Tokyo Denki University, Tokyo, Japan
(Tel : +81-3-5280-3281; E-mail: igarashi@isl.d.dendai.ac.jp)

Abstract: This paper presents development of a new head-mounted eye tracking system for Human Adaptive Mechatronics(HAM) system. Estimation of human skill level is important to assist human suitability. Eye movement indicates human insight such as changing attention and intention, hence this paper focuses on design of a headgear to reduce operator's blind side and to extend scene camera image which superimposes point of gaze. By using a robust pupil locator based entirely on integer calculations for easy real-time implementation in FPGA hardware, an algorithm estimating pupil's center against ill-conditional images contaminated by eyelid and eyelash is proposed.

Keywords: Eye tracking device, FPGA

1. INTRODUCTION

Most existing human-machine systems didn't consider human skill level. Because a human has to learn to operate a machine and to acquire skills by themselves, it was difficult to obtain the best performance. To improve performance in the human-machine systems, new intelligent Mechatronics called HAM (Human Adaptive Mechatronics)[1][2] which adapts to human skill level was proposed. HAM system needs estimating human skill level so that the machine supports the operation and adapts human skill level. For this aim, human state, especially attention and intention, should be estimated.

It is known that an eye movement directly indicates human attention and intention. Hence, observation of eye movement is effective for HAM. In machine operation, the human obtains surrounding environment information by using peripheral field. In vehicle operation, which is one of typical machine operation, there are differences on eye movement between beginner and expert. When the human pays one's attention to some object in peripheral field, human changes one's gaze to watch the objects. Therefore recording peripheral field and measurement of line-of-sight (LOS) in wide angle is needed to investigate human skill level. But most exiting eye-tracking-systems focus on improvement measurement accuracy of estimating pupil center, and recording peripheral field and measurement of LOS in wide angle was not considered sufficiently.

When eye moves in a wide range, the pupil is hidden partly by the eyelash and the eyelid. To make eye tracking system be useful interface, an algorithm estimating the pupil center wants to be easy-to-use and robust. Our final goal is to build a very compact and light weight system that can easily be moved from place to place. The design will allow the user maximum freedom of mobility, we hope that it minimizes fatigue and it will have a minimal impact on the user's filed of view. We assume that processing of the visual data is done in real-time with the majority of processing occurring in a compact FPGA based processing system. We show that pupil tracking can be performed entirely using simple, integer based arithmetic which leads to a simpler FPGA based implementation.

2. EYE TRACKING DEVICE

Figure.1 shows designed for this head gear, and Fig.2 shows the developed eye tracking device. The system consists of eye tracking system and an environment recording system. To reduce user's burden, both of systems use CCD cameras because of its light weight and small size. Eye tracking system obtains infra-red image reflected by a hot mirror. Using the hot mirror reduces obstruction interrupting user's eye sight. The head gear is designed to measure wide area equivalent to one's eyesight, by using two environment-cameras. In short, we designed the headgear by paying attentions about the following points.

- lightweight
- robust measurement to varying environment light
- less obstacle of one's eyesight

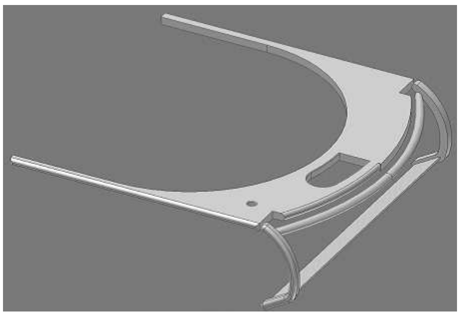

Fig. 1 3D-CAD model of a headgear

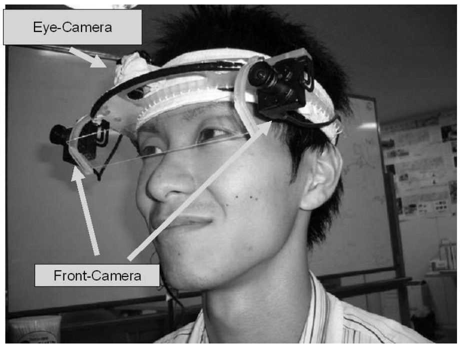

Fig. 2 Headgear mounted on cameras and hot-mirror

3. IMAGE PROCESSING OF EYE TRACKING CAMERA IMAGE

Most eye tracking systems developed as practical application use a technique called Purkinje image tracking. When a point source of light is shone onto the eye, the reflected image is called the Purkinje image. The relationship between the pupil center and the center of the Purkinje image can be used to determine eye-gaze angle. Purkinje images, however, fail if the reflection does not fall on the spherical part of the cornea. And the images do not appear when the eye moves widely from the center location. Hence, in our work, we have chosen to use a 'dark pupil method'. This method utilizes the property such that the pupil is a dark area in the eye tracking camera image. We expect that the eye image is processed by the dark pupil method to find the pupil center in wide area and with high speed. Further we consider to directly map the pupil center location to the environment-camera image. The reason are (a) we are not interested in gaze angel (as given by the Purkinje based methods) but rather we want to know where is the user looking in the environment image, and (b) the dark pupil method allows us to skirt issues of reflection and lens distortion - these factors can be built directly into the model mapping the pupil center location to the environment-camera coordinate system.

Using an FPGA as our processing platform also places certain practical constraints on the choice of algorithm. Implementing floating point arithmetic is significantly more difficult than integer or fixed point arithmetic so algorithms that minimize the use of floating point arithmetic is preferred. For similar reasons, in addition, subtraction and multiplication operations are preferred over trigonometric and other complex mathematical functions such as exponentiation. Finally, implementing processing functions in FPGA gives us the ability to exploit parallelism in ways that is not possible with ordinary DSPs thus parallel algorithmic structures are preferred where available. These points are taken into consideration on the developmental the later algorithms.

3.1 Algorithm to detect pupil's center

This section describes image processing to estimate a position of the pupil center, the algorithm is shown in Fig.3. First stage is to quickly find the pupil area to decide

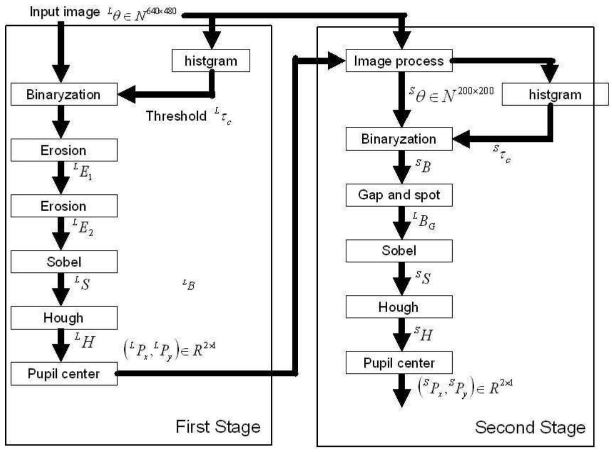

Fig. 3 Algorithm flow chart for pupil estimation

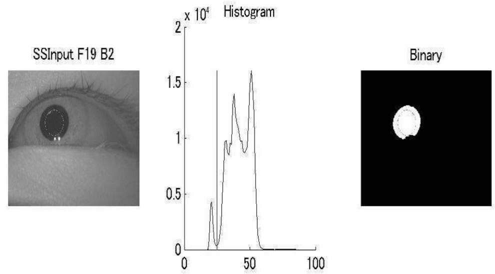

Fig. 4 Automatic tune to obtain binary image.From left to right, Input image(left), histogram (64 bins) & threshold (vertical line)(center) and resulting binary image(right).

small tracking window covering pupil area. By using the dark-pupil technique, the pupil is the darkest region in the image. Thus we can roughly locate the pupil simply by appropriate threshold. This binary image is described as ^LB in Fig.3. Then, binary image ^LB is applied to erosion twice to reduce noise. This image is described as $^LE_2$ in Fig.3. Horizontal and vertical Sobel filters are applied to the image$^LE_2$, in order to obtain edge image ^LS. Next, hough transform is applied to ^LS in order to obtain ^LH, then we acquire pupil center ($^LP_x, ^LP_y$) by using pupil center step. Above mentioned image processing flow is named the first stage.

3.2 Thresholding

Our approach to find a threshold begins by making an assumption regarding the minimum and maximum area for the pupil in the image. We then construct an image brightness histogram and search (starting from the bin containing the darkest pixels) for the first bin where (a) the cumulative sum is greater than the minimum area and (b) the cumulative sum is greater than the maximum area. The threshold τ_c is chosen to lie between these two bins as shown in Fig.4.

3.3 Refined binary image

3.3.1 Erosion

Nevertheless acquiring appropriate threshold, some noises such as eyelid and eyelash are included in the bi-

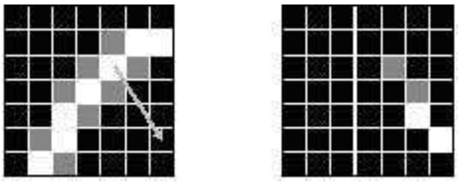

Fig. 5 Ellipse's contour on the binarized edge image ${}^{S}H$(left).The normal line drawn in a center accumulator array C(right).

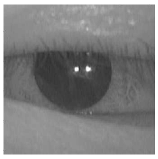

Fig. 6 eye image

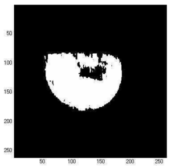

Fig. 7 binary image with gap and spot

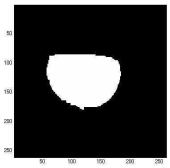

Fig. 8 refind image

nary image. To reduce these noises, when there are five or more black pixels in the analyzing area that is the 3 3 area surrounding each pixel, the corresponding pixel is set to black, otherwise it remains white. This processing is only done in the first stage.

3.3.2 Spot removing and gap filling

Even though the image is processed by erosion, spots and gaps cracks deteriorate estimation accuracy of pupil center by ill-effected the eyelid and the eyelash(Fig.6). Therefore, spots are removed by analyzing the 3 3 area surrounding each pixel. If there are eight or more white pixel in the analyzing area, the corresponding pixel is set to white, otherwise it remains black. Figure 7 is the example of spot removed.

Gap filling is performed using a simple algorithm based on scanline. The binary image is first scanned horizontally one line at a time. Any run of less then several black pixels that are bounded by white pixels is replaced by white pixel. The same process is then repeated in the vertical direction. Figure 8 is the refined image that is applied the Spot removing and gap filling to Fig.6 .

3.4 Pupil Center Estimation

Now, we make the assumption that pupil will appear as an ellipse in the image and that the normal lines to contours-curves of the ellipse converge to a center of the ellipse. To utilize this property, we construct a center accumulator array $C \in R^{200\times200}$ having one element for each pixel in the original image. An estimation process of the pupil center is mentioned below. First, the array C is initialized to zero. By using the binarized edge image ${}^{S}H$ that is already compensated by the spot removing and the gap filling, we draw a line in the corresponding direction of each white pixel of the ellipse's contour (see left of Fig.5). This line is normal line to the ellipse's contour and is determined by the Bresenham line drawing algorithm. The normal line is drawn in C, not on the original image ${}^{S}H$, and it is drawn by incrementing the elements of C that fall on the line(see right of Fig.5). We limit the length of the line by using upper (R_{max}) and lower (R_{min}) bounds for the pupil major axis, calculated using the limits on the pupils area defined earlier. As we move along the line in C, we do not begin incrementing the accumulator elements until we have stepped over R_{min} elements and we stop incrementing elements when we have stepped over a total of R_{max} elements.

4. EXPERIMENT

4.1 Experiment of Pupil Center Estimation

In the experiment, a subject instructed to watch 6×5 markers orderly. Image is captured by 33 frames per second. The image had been obtained for one second after getting the subject's cue, and the pupil's center is calculated by the image processing for one second. Means of calculated position are used to detect the position of pupil for the evaluation of proposal method. i is defined as raw number, and j is defined to tell the each marker as column number.

The proposal algorithm is able to estimate pupil's center when pupil's shape is ellipse (as shown in Fig.9) and the lack of image caused by the eyelid (as shown in Fig.10). However, when the subjects gaze at lower area, the pupil is not always estimated correctly. Figure 11 and Fig. 12 is pupil image when subject gazes at 6th raw-4th column. Figure 11 shows the good result because appropriate threshold could be obtained. Binary image in Fig. 12 is failed case, because appropriate threshold is not obtained to separate eyelash area from pupil area.

4.2 Mappimg from coordinate system of the eye tracking camera image to coordinate system of environment-camera image

We will develop system that recognizes the area where a subject gazes at. Therefore, we need to know a the relation between the coordinate system of the eye tracking cameras image and the other coordinate system of the environment-camera image. The lens distortion of a camera image is generally expressed as the following quadratic surface.

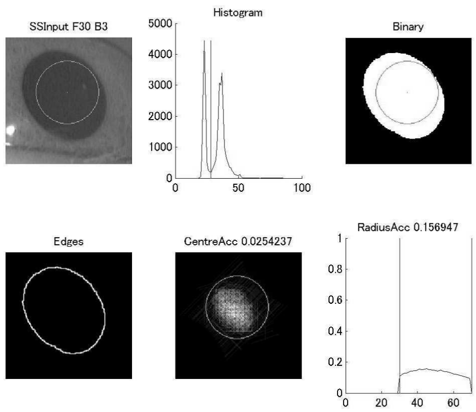

Fig. 9 Estimation of pupil center when pupil's shape is ellipse(good case)

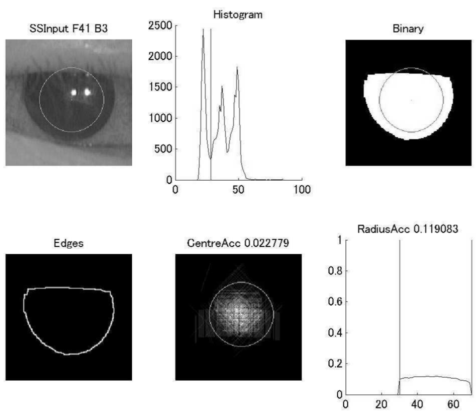

Fig. 10 Estimation of pupil center when pupil is lack of a part(good case)

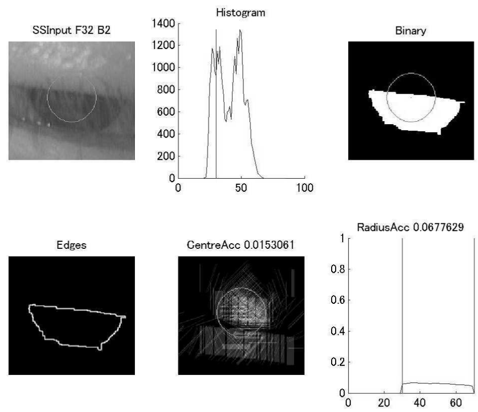

Fig. 11 Estimation of pupil center with appropriate threshold(good case)

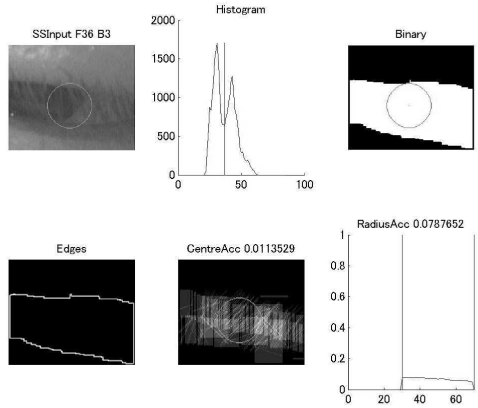

Fig. 12 Estimation of pupil center with inappropriate threshold(bad case)

$$\begin{bmatrix} u_c \\ v_c \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & a_3 & a_4 & a_5 & a_6 \\ a_7 & a_8 & a_9 & a_{10} & a_{11} & a_{12} \end{bmatrix} \begin{bmatrix} u_e^2 & u_e^2 \\ u_e & u_e \\ u_e v_e & u_e v_e \\ v_e^2 & v_e^2 \\ v_e & v_e \\ 1 & 1 \end{bmatrix}, \quad (1)$$

where $[u_c, v_c]^T \in N^{2\times 1}$ is a center position of marker on coordinate system of the environment-camera image, $[u_e, v_e]^T \in N^{2\times 1}$ is a center position of a pupil on the coordinate system of eye tracking camera image, and $a_i(i = 1, \cdots, 12)$ is mapping parameters that should be estimated by extracting from real image data. These parameters are identified by using data of which position is known correct $apriori$. Now, defining F as

$$F := \begin{bmatrix} u_{c11} & u_{c12} & \cdots & u_{cij} \\ v_{c11} & v_{c12} & \cdots & v_{cij} \end{bmatrix} \quad (2)$$

and E as

$$E = \begin{bmatrix} u_{e11}^2 & u_{e11}^2 & \cdots & u_{eij}^2 & u_{eij}^2 \\ u_{e11} & u_{e11} & \cdots & u_{eij} & u_{eij} \\ u_{e11}v_{e11} & u_{e11}v_{e11} & \cdots & u_{eij}v_{eij} & u_{eij}v_{eij} \\ v_{e11}^2 & v_{e11}^2 & \cdots & v_{eij}^2 & v_{eij}^2 \\ v_{e11} & v_{e11} & \cdots & v_{eij} & v_{eij} \\ 1 & 1 & \cdots & 1 & 1 \end{bmatrix}, \quad (3)$$

the mapping parameters shown in Eq.(1) are calculated by the least-square method as follows.

$$A = EF^T(FF^T)^{-1}, \quad (4)$$

where

$$A = \begin{bmatrix} a_1 & a_2 & a_3 & a_4 & a_5 & a_6 \\ a_7 & a_8 & a_9 & a_{10} & a_{11} & a_{12} \end{bmatrix}. \quad (5)$$

Then, point of gaze is estimated using calculated mapping parameters A and pupil center position. A circle which is estimated point of gaze superimposed on environment-camera image. Figure 13 shows the final result. Pupil center estimation is good condition from the 4th-raw to the 5th-raw. But, pupil center can not be estimated when subject gazes at upper area. In this case, the 6th-raw markers couldn't be estimated because appropriate threshold was not obtained, because about 70%

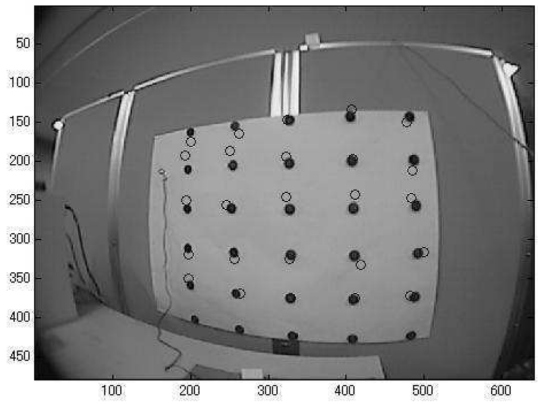

Fig. 13 experimental result

of edge of ellipse was lost by effect of the narrow eye. We have to improve this issue in future work.

5. CONCLUSION

In this paper, we proposed the algorithm of detection of pupil center and declared availability of proposal algorithm. This algorithm can estimate pupil center, although pupil's shape is almost hemicycle. This algorithm is suitable to develop of a new head mounted eye tracking system for HAM system.

REFERENCES

[1] K.Furuta, "Control of Pendulum -From Super Mechano-System to Human Adaptive Mechatronics ". IEEE Conf.Decision and Control,pp1498-1507,2003

[2] Harashima F., "Human Adaptive Mechatronics" Plenary Talk,Int Conf.Control,Automation & Systems,Bangkok,26 Aug 2004

[3] Julian Kolodko,Satoshi Suzuki,Fumio Harashima, "Eye-Gaze Tracking:An approach to PUpil Tracking targeted to FPGAs" IEEE/RSJ Conf.Intelligent Robots and Systems Trans.pp.2942-2947,2005.

[4] Suzuki S.,Tomomatsu N.,Harashima F.,Furuta., "Skikk Evalution based on State-Trasition Model for Human Adaptive Mechatronics (HAM)" 30th Annual Conf. of IEEE Industrial Electronics Society (IECON'04), Nov2-6, Busan, Korea, 2004

[5] Nicklin D., "FPGAs-Enabling DSP in Real-Time Video Processing", http://www.xlinx.com/publications/product/sp2e /xcell_pdf/xc_dspvid43.pdf,July 2002.

[6] Glenstrup A.J., Engell-Nielsen T., "Eye Controlled Media: Present and Future State", Ph.D. Thesis, University of Copenhagen, DIKU (Institute of Computer Science), Denmark, 1st June 1995.

[7] Ohno T., Mukawa N., Yoshikawa A., "FreeGaze: A Gaze Tracking System for Everyday Gaze Interaction", Proceedings of the symposium on ETRA 2002: eye tracking research &applications symposium, pp125-132, 2002.

[8] Shih, S-W., Liu J., "A Novel Approach to 3D Gaze Tracking using Stereo Cameras" Technical Report TR-M3LAB-2002-001,Multimedia Man-Machine Interface Laboratory,Department of Computer Science and Infomation Engineering, National Chi Nan Universiy, http://www.csie.ncnu.edu.tw/ stone/, 2002

[9] Morimoto C. H., Koons D., Amir A., Flickner M., "Pupil Detection and Tracking Using Multiple Light Sources", Image and Vision Computing , vol 18, pp331-335, 2000.

[10] Amir A et al, "An embedded system for an eye-detection sensor" Computer Vision and Image Understanding, vol 98, pp104-123, 2005.

[11] Atherton T.J., Kerbyson D.J., "Size Invariant Circle Detection", Image and vision computing, vol 17, pp795-803, 2005.

[12] Colin Flanagan, "The Bresenham Line-Drawing Algorithm" http://wwww.cs.helsinki.fi/group/goa/mallinnus /lines/bresenh.html,Cited Dec 2004.

[13] Young I.T., Gerbrands J.J., van Vliet L.j., "Image Processing Fundamentals: Morphology-based Operations" http://www.ph.tn.tudelft.nl/Courses/FIP/noframes/fip-Morpholo.html, Cited Dec 2004.

[14] Robyn Owens, "Binary Images" http://homepages.inf.ed.ac.uk/rbf/CVonline/ LOCAL_COPIES/OWENS/LECT2/node3.html,1997. Cited Dec 2004.

[15] Fitzgibbon A.W.,Pilu M.,Fisher R.B., "Direct Least Squares Fitting of Ellipses", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol 21(5),pp476-480,May 1999.