

DUAL CAMERAS ACQUISITION FOR THREE-DIMENSIONAL EYE-MOTION TRACKING

Sunu Wibirama¹, Supan Tungjitkusolmun², Chuchart Pintavirooj³, Kazuhiko Hamamoto⁴

^{1,2,3}Department of Electronics, Faculty of Engineering,
King Mongkut's Institute of Technology Ladkrabang, Thailand

⁴Department of Information Media Technology,
School of Information and Telecommunication Engineering,
Tokai University, Japan
Email: kpchucha@kmitl.ac.th

ABSTRACT

Three-dimensional eye-motion tracking using dual cameras is presented in this paper. The system consists of two arbitrary orientation USB cameras mounted on tripod. The two-dimensional positions of pupil center were obtained using template matching technique. The three-dimensional positions of pupil center were then computed using Direct Linear Transformation technique. The algorithm was validated by comparing an eye model movement in world Euclidean coordinate system and DLT system. The system yielded error 0.2072, 0.1675, and 0.3797 pixel for X,Y, and Z axes respectively with range of error from 0 to 1 pixel.

1. INTRODUCTION

Eye-motion tracking is an extensive research area in biomedical engineering and computer vision. Recent advances in understanding of the motor control of eye movements and new findings in the anatomy of extraocular muscles have greatly increased the interest in a complete three-dimensional (3D) description of eye position [1-5].

Scleral Search Coil method has been used successfully to measure horizontal, vertical and torsional components of the eye [1,2]. However, this technique is expensive and invasive, limiting the measurement duration to about 30 mins because of discomfort during recording process [3]. Parker *et al.* [4] proposed a method to measure 3D components based on digital signal processing. The measurement was done by analyzing marker on several printed eye images. Although the result was promising, this method did not support real time processing. Shih *et al.* [5] proposed a satisfactory 3D eye tracking system using multiple infrared reflections and stereo CCD cameras. The 3D gaze was measured by comparing infrared mark and pupil center. Patients were required to fix their heads to get accurate tracking result from table-mounted eye-motion tracking system. The drawbacks of such system are that it is complicated to use and associated high costs.

In this research, we introduce a real time 3D eye-motion tracking system without infrared reflections. We attempted to reduce the cost by designing an eye-motion tracking system using two images acquired from two arbitrarily oriented USB cameras. The 3D coordinates of pupil center were then estimated using DLT technique proposed in [6].

This work is organized as follows: Section II explains the proposed system and method. Experimental results are provided in Section III. Discussions and conclusions are given in Section IV.

2. PROPOSED SYSTEM AND METHOD

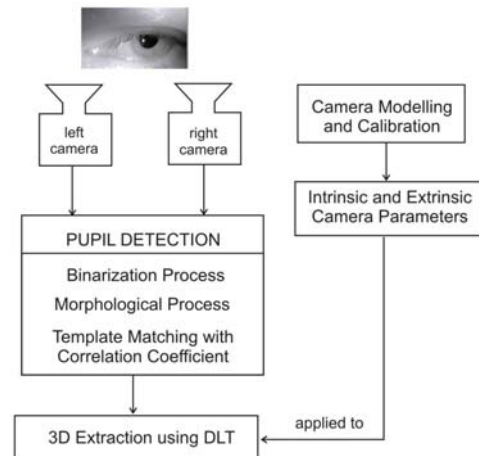


Figure 1. 3D eye coordinates acquiring method using two arbitrarily oriented cameras.

The eye-motion tracking system consists of 3 parts: (i) two arbitrarily oriented USB cameras, (ii) personal computer (PC) and (iii) the 3D eye tracking software created using Visual C++ and OpenCV library [7]. We used Oker 177 web camera (Shenzhen Golden Tiger & Dragon Technology Develop Co., Ltd, Guangdong, China) with sampling rate of 60 Hz, focus range of 30 mm to infinitive, and resolution of 2 megapixels. The two cameras were installed on small tripod in order to provide stable and comfortable environment during the experiment process. The 2.2 GHz Intel Pentium PC with Windows XP operating system was used in the system.

The 3D eye coordinates acquiring method is shown in Fig. 1. First, we applied camera modeling and calibration to obtain intrinsic and extrinsic camera parameters which would be used in DLT algorithm. The pupil detection process using template matching was then conducted to obtain 2D positions of pupil center. The 3D coordinates of pupil center were then extracted using DLT technique.

2.1. Camera Modeling and Calibration

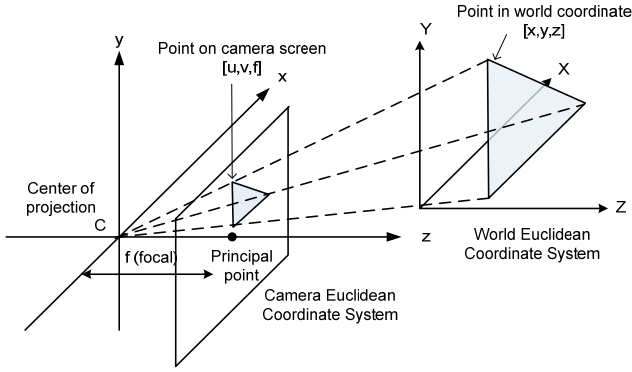


Figure 2. Perspective transformation from world Euclidean coordinate system to camera Euclidean coordinate system.

A point located in real 3D space (world Euclidean coordinate system) can be expressed as point $[x \ y \ z \ 1]^T$. A point $[x \ y \ z \ 1]^T$ taken by a camera undergoes a linear transformation from 3D projective space to 2D projective space as shown in Fig.2. Point $[u \ v \ w]^T$ in camera coordinate system is obtained by examining the following homogeneous coordinate formula:

$$\begin{bmatrix} u \\ v \\ w \end{bmatrix} = K [I_3 \ | \ 0_3] \begin{bmatrix} R & -T \\ 0_3^T & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix},$$

$$\text{or } \begin{bmatrix} u \\ v \\ w \end{bmatrix} = M \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}, \text{ where } M = [KR / -KRT]. \quad (1)$$

Matrices K and $[I_3 \ | \ 0_3]$ in (1) are intrinsic parameters which define physical parameters of the camera. Matrix K is camera calibration matrix defined as

$$K = \begin{bmatrix} fa & 0 & u_0 \\ 0 & fb & v_0 \\ 0 & 0 & 1 \end{bmatrix}, \quad (2)$$

where f is focal length in camera projection; a and b are conversion factor from physical unit to pixel unit in the x and y directions respectively; and (u_0, v_0) is a principal point in camera projection.

$$\text{Matrix } [I_3 \ | \ 0_3] \text{ is defined as } \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

Matrix $\begin{bmatrix} R & -T \\ 0_3^T & 1 \end{bmatrix}$ consists of extrinsic parameter

which defines camera location and orientation in world Euclidean coordinate system. Rotation matrix R expresses three elementary rotations of camera Euclidean coordinate axes which respect to the world Euclidean coordinate system. Rotation along x , y , and z are termed as pan, tilt, and roll respectively. Translation vector T gives three elements of the translation of the origin of the world Euclidean coordinate system with respect to camera Euclidean coordinate system. Matrix 0_3^T is vector $[0 \ 0 \ 0]$. To obtain matrix M , observing each known point $X=[x \ y \ z \ 1]^T$ and its corresponding 2D image point $[u \ v \ w]^T$ will yield an equation:

$$\begin{bmatrix} x & y & z & 1 & 0 & 0 & 0 & 0 & -ux & -uy & -uz & 1 \\ 0 & 0 & 0 & 0 & x & y & z & 1 & -vx & -vy & -vz & 1 \\ & & & & & & & & M & & & \end{bmatrix} \begin{bmatrix} m_{11} \\ m_{12} \\ M \\ m_{34} \end{bmatrix} = 0$$

$$\text{or } AM = 0. \quad (3)$$

If n such points in world euclidean coordinate system are available, A will be of size $2n \times 12$. M can be solved by performing Singular Value Decomposition (SVD) of A to derive $A = UDV^T$. U is left singular vector, D is diagonal matrix consists of singular values of A , and V is right singular vector of A . The last column of V is the solution for M . To separate extrinsic parameter, observe that M in (1) can be written as $M=[KR/-KRT]=[A/b]$. Then we can obtain translation matrix T by solving equation $T = -A^{-1}b$. To determine R , we decompose A into a product of two matrices K and R using QR decomposition.

2.2. Pupil Detection

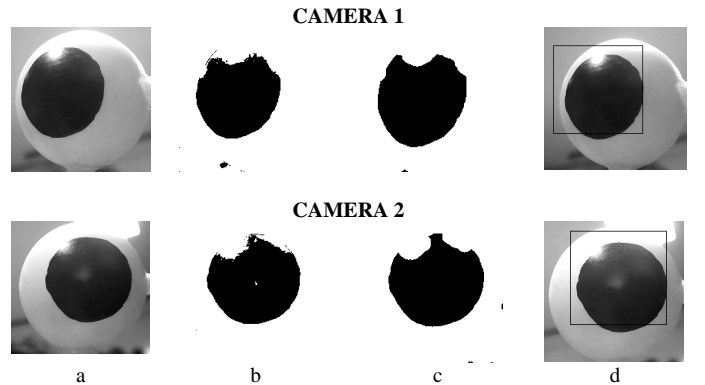


Figure 3. Pupil detection using plastic eye model: (a) original images, (b) segmentation using thresholding, (c) removing small artifact at pupil's boundary using morphology, (d) result of template matching with correlation coefficient.

The pupil detection process is shown in Fig. 3. The image frames were captured from two different videos of two cameras with arbitrary orientation. The captured frames shown in Fig. 3.a were first segmented to binary images using thresholding. The thresholding results were shown in Fig. 3.b. The remaining artifact at the pupil's border was removed using basic opening-closing morphological process. The morphological results were shown in Fig. 3.c.

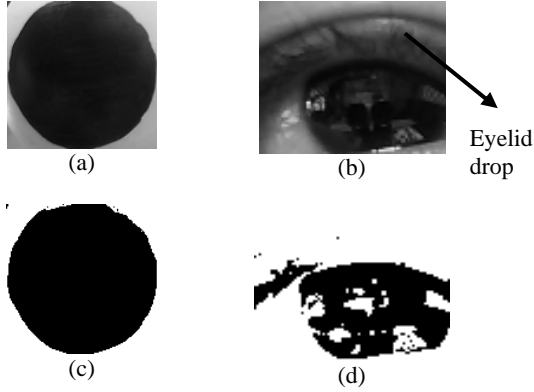


Figure 4. Template image used in experiment: (a) template of plastic eye model; (b) template of real human eye with eyelid interference; (c) template of plastic eye model after thresholding; (d) template of real human eye after thresholding.

To locate the pupil, template matching technique was then applied. The pupil template for tracking plastic eye model was circular disk as shown in Fig.4.a. The pupil template for tracking real human eye was a half-circular disk to include the case of matching in the presence of eyelid drop where the pupil appeared as half-circular shape as shown in Fig. 4.b. The templates were first converted to binary images (Fig.4.c and Fig.4.d) before used in matching process. The template image was then slid through the binary image of captured frames during the matching process. The similarity measurement used in the matching procedure was correlation coefficient $\hat{\rho}_s(X, Y)$ defined as

$$\hat{\rho}_s(X, Y) = \frac{\hat{C}_s(X, Y)}{\sqrt{\hat{\sigma}_x^2 \hat{\sigma}_y^2}}, \quad -1 \leq \hat{\rho}_s(X, Y) \leq 1, \quad (4)$$

where X is the eye image and Y is the template, $\hat{C}_s(X, Y)$ is covariance, $\hat{\sigma}_x^2$ and $\hat{\sigma}_y^2$ are variances. In this experiment, the best match was found when the template matching result achieved the maximum value of $\hat{\rho}_s(X, Y)$. The template matching result was a region inside the black rectangle as shown in Fig.3.d.

2.3. 3D Coordinates Extraction

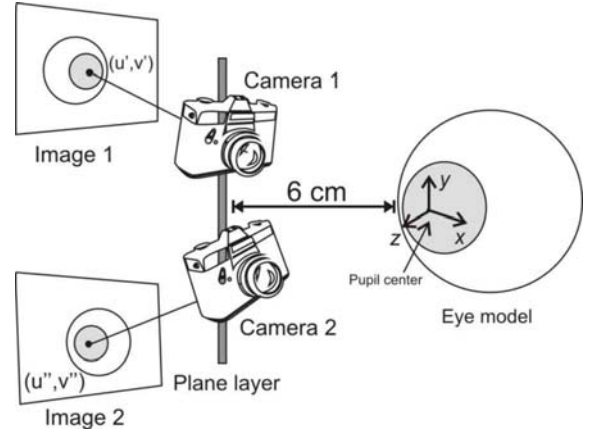


Figure 5. 3D coordinates extraction

Direct Linear Transformation is a method that extracts 3D coordinates of the object from 2D coordinates in multiple photographs taken at arbitrary poses around the object. In this research, we used two cameras which were located approximately 6 cm in front of plastic eye model as shown in Fig.5. Camera modeling and calibration process yielded intrinsic and extrinsic parameters. Parameters for the first and second cameras were inserted into matrix M' and M'' respectively. Note that (u', v') and (u'', v'') are 2D coordinates resulted from image 1 and image 2 respectively. Thus, to estimate 3D coordinates (x, y, z) of pupil center, the following homogenous equation can be solved:

$$\begin{bmatrix} u' m_3'^T - w' m_1'^T \\ v' m_3'^T - w' m_2'^T \\ u'' m_3''^T - w'' m_1''^T \\ v'' m_3''^T - w'' m_2''^T \end{bmatrix} X = AX = 0 \quad (5)$$

where m_i^T are the row i^{th} of the projection matrix M , while X is matrix $[x \ y \ z \ 1]^T$.

3. EXPERIMENTAL RESULT

To validate the 3D positions extracted from each video frame, we compared real and computed movements of plastic eye model in X, Y , and Z axes. Real movement is an eye model movement in world Euclidean coordinate system. We measured manually the X, Y , and Z movement using a ruler. Computed movements were derived from DLT technique. The experiments were repeated 5 times for each axis. From experimental results, we found that every 1 mm movement in world Euclidean coordinate system yielded ± 1 pixel movement in DLT system.

TABLE I. COMPARISON RESULT OF X-AXIS MOVEMENT

Exp.	Real X-Axis Movement (mm)	Computed X-Axis Movement (pixel)	Error (pixel)
1 st	5	5.1453	0.1453
2 nd	5	5.2989	0.2989
3 th	5	5.2972	0.2972
4 th	5	5.0181	0.0181
5 th	5	5.1246	0.1246
Root Mean Square of Error			0.2072

TABLE II. COMPARISON RESULT OF Y-AXIS MOVEMENT

Exp.	Real Y-Axis Movement(mm)	Computed Y-Axis Movement (pixel)	Error (pixel)
1 st	5	4.7414	0.2586
2 nd	5	5.2460	0.2460
3 th	5	5.1098	0.1098
4 th	5	4.9731	0.0269
5 th	5	4.9882	0.0118
Root Mean Square of Error			0.1675

TABLE III. COMPARISON RESULT OF Z-AXIS MOVEMENT

Exp.	Real Z-Axis Movement(mm)	Computed Z-Axis Movement (pixel)	Error (pixel)
1 st	5	5.6912	0.6912
2 nd	5	5.3576	0.3576
3 th	5	5.0199	0.0199
4 th	5	5.2827	0.2827
5 th	5	5.1872	0.1872
Root Mean Square of Error			0.3797

The experimental results show that the system yielded error approximately 0.2072, 0.1675, and 0.3797 pixel for X, Y, and Z axes respectively with range of error 0 to 1 pixel. The error are considered within acceptable tolerance. Most error were caused by inaccurate pupil detection which supplies wrong 2D coordinate in DLT system.

The second experiment was system implementation using real human eye as shown in Fig.6. We were interested in testing the performance of pupil detection algorithm when the subject fully opened his eye (Fig. 6.a.1 and Fig. 6.a.2), closed his eye (Fig. 6.e.1 and Fig. 6.e.2), and opened only a half of his eye (Fig. 6.b.1, Fig. 6.b.2, Fig. 6.c.1, and Fig.6.c.2). We also asked the subject to wear eyeglasses to check whether the algorithm worked in the presence of interference caused by eyeglasses and some reflections of unwanted light sources (Fig. 6.f.1, Fig.6.f.2, Fig.6.g.1, Fig.6.g.2, Fig.6.h.1, and Fig.6.h.2). The subject was also asked to blink his eye and to lean his head towards left and right direction. The fail cases were mainly caused by large ambiguity of black area resulted by tresholding segmentation process The algorithm detected the eyebrow or the other areas instead of the pupil.

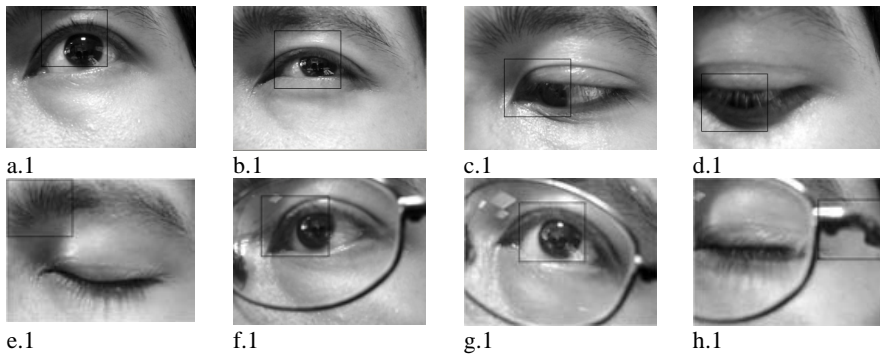
4. CONCLUSION AND DISCUSSION

Three-dimensional eye-motion tracking system has been presented in this paper. The DLT techniques were adopted to derive dynamic 3D information from arbitrary orientation cameras. The 3D eye tracking system was then validated in X, Y, and Z directions by comparing real and computed movements. The system was also tested using real human eye. The proposed system successfully detected 3D coordinates within acceptable error tolerance. Future work will be done to improve accuracy of pupil detection process.

REFERENCES

- [1] H.Collewijn, J.van der Steen, L.Ferman, & T.Jansen, "Human ocular counterroll: Assesment of static and dynamic properties from electromagnetic scleral coil recordings," in *Exp. Brain Res.*, vol. 59, no.1, pp. 185-96, 1985.
- [2] D.A. Robinson, "A Method of measuring eye movement using a scleral search coil in a magnetic field," in *IEEE Trans. Biomed. Electron*, vol. 10, pp. 137-45,1963.
- [3] J.P. Ivins, J. Porrill, & J.P. Frisby, "Deformable model of the human iris for measuring ocular torsion from video images," in *IEEE Proc.Vision, Image, Signal Processing*, pp. 213-20, 1998.
- [4] J.A. Parker, R.V. Kenyon, & L.R. Young, "Measurement of Torsion from Multitemporal Images of the Eye Using Digital Signal Processing Techniques" in *IEEE Transactions on Biomedical Engineering*, vol. BME-32, no.1, pp 28-36, 1985.
- [5] S. Shih & Jin Liu, "A Novel Approach to 3-D Gaze Tracking Using Stereo Cameras", in *IEEE Transactions on System, Man, and Cybernetics*, vol. 34, no.1, pp. 234-245, 2004.
- [6] Y.I. Abdel-Aziz and H.M. Karara, "Direct linear transformation into object space coordinates in close-range photogrametry," in *Proc. Symp.Close-Range Photogrametry*, pp. 1-18, 1971.
- [7] <http://opencv.willowgarage.com>

CAMERA 1



CAMERA 2

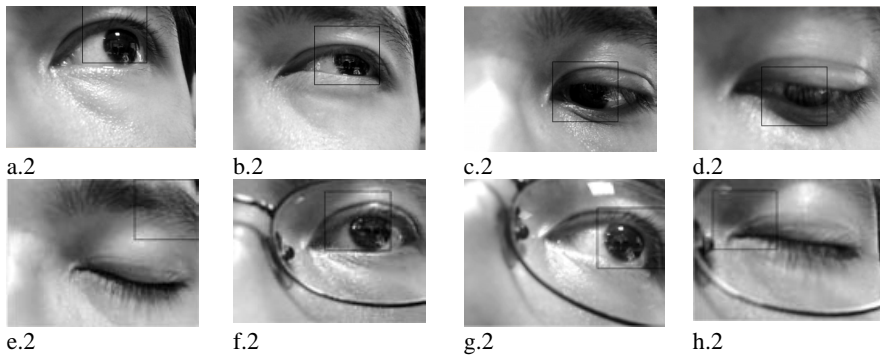


Figure 6. 2D tracking using arbitrary orientation cameras: success cases on tracking full eye (a.1 and a.2), half-closed eye (b.1 ,b.2,c.1, and c.2), almost closed eye (d.1 and d.2). Fail cases on tracking full closed eye (e.1 and e.2). Success cases on tracking full eye interfered by eyeglasses (f.1,f.2,g.1, and g.2) and fail cases on tracking full closed eye interfered by eyeglasses (h.1 and h.2).