Webcam-Based 2D Eye Gaze Estimation System By Means of Binary Deformable Eyeball Templates

Jinwoo Kim, *Member, KIMICS*

Abstract— **Eye gaze as a form of input was primarily developed for users who are unable to use usual interaction devices such as keyboard and the mouse; however, with the increasing accuracy in eye gaze detection with decreasing cost of development, it tends to be a practical interaction method for able-bodied users in soon future as well. This paper explores a low-cost, robust, rotation and illumination independent eye gaze system for gaze enhanced user interfaces. We introduce two brand-new algorithms for fast and sub-pixel precise pupil center detection and 2D Eye Gaze estimation by means of deformable template matching methodology. In this paper, we propose a new algorithm based on the deformable angular integral search algorithm based on minimum intensity value to localize eyeball (iris outer boundary) in gray scale eye region images. Basically, it finds the center of the pupil in order to use it in our second proposed algorithm which is about 2D eye gaze tracking. First, we detect the eye regions by means of Intel OpenCV AdaBoost Haar cascade classifiers and assign the approximate size of eyeball depending on the eye region size. Secondly, using DAISMI (Deformable Angular Integral Search by Minimum Intensity) algorithm, pupil center is detected. Then, by using the percentage of black pixels over eyeball circle area, we convert the image into binary (Black and white color) for being used in the next part: DTBGE (Deformable Template based 2D Gaze Estimation) algorithm. Finally, using DTBGE algorithm, initial pupil center coordinates are assigned and DTBGE creates new pupil center coordinates and estimates the final gaze directions and eyeball size. We have performed extensive experiments and achieved very encouraging results. Finally, we discuss the effectiveness of the proposed method through several experimental results.**

Index Terms— **Eye Tracking, Pupil Center Detection, 2D Gaze Estimation, Deformable Template Matching.**

I. INTRODUCTION

In our everyday lives, people perceive the outer reality by means of their eyes. The eyes are the windows of the soul opening to the interaction with others. In this regard, the eyes are actually an interaction bridge between human and the outer reality. Therefore, the eyes have been always considered as a mysterious organ which can carry various kinds of information to be used in daily life. One important aspect of the eyes is also PoR (Point of Regard) which can be used a proxy for handling the users' attention or intention as an interaction input between human and the computer (HCI) where gaze information can be used as a form of input instead of keyboard and the mouse. Eye gaze as a form of input was primarily developed for users who are unable to use usual interaction devices such as keyboard and the mouse; however, with the increasing accuracy in eye gaze detection with decreasing cost of development, it tends to be a practical interaction method for able-bodied users in soon future as well. This paper explores a low-cost, robust, rotation and illumination independent eye gaze system for gaze enhanced user interfaces.

Eyes are extremely rapid. For this reason, eye movements are faster than other current input media. However, the signals are noisy. In order to develop a gaze enhanced media with gaze inputs, first of all, signals have to be normalized and they have to be continuous, not broken. However, people are not accustomed to operating devices by simply moving their eyes. Actually, this fact brings the Midas Touch Problem within itself. In addition, selection of target is difficult by means of current techniques such as: dwell-time clicks, eye blinks, mouse clicks or keyboard inputs. Therefore, every single eye gaze tracking system has a different visual angle accuracy which is important for the size of selection targets such as: buttons, icons, pictures, texts etc. The systems with lower visual angle accuracy are more precise and robust in terms of selection of targets and this is the most important factor to classify the eye gaze systems. This paper proposes several brand-new algorithms in terms of eye tracking, pupil detection, eye movements detection and the gaze tracking with satisfactory visual angle accuracy by using low-cost web camera without Infra-Red filter with low-resolution images. Current algorithms for robust eye gaze detection require calibration and training process which is not applicable in everyday life. And most of them use head-mounted cameras with Infra-Red technology which are expensive and not ergonomic for everyday use. Besides this, current algorithms are all pupil center based algorithms which is not robust in various lightening conditions because of corneal reflection or glint on the

---

Manuscript received August 3, 2010; revised September 6, 2010; accepted September 10, 2010..

Jinwoo Kim is with the Department of Information & Communication, Kyungsung University, Busan, 607-736, Korea (Email: jinwoo@ks.ac.kr)

eyeball. Therefore, the ideal system must be cheap, ergonomic and fully-automated in terms of usability and workload.

A great deal of computer vision research is dedicated to the implementation of systems designed to detect user movements and facial gestures [1]-[7]. In most of the cases, such systems are created with the specific goal of providing a way for people with disabilities or limited motor skills to be able to use computer systems, albeit in much simpler applications [1], [6], and [7]. The motivation for the system proposed here is to provide an inexpensive, low-cost and unobtrusive means for disabled people to interact with simple computer applications in a meaningful way that requires minimal effort. In fact, gaze tracking has been used in studies dating back to 1935 [8] and 1967 [9]. During these early days of eye tracking systems were cumbersome, invasive and not very accurate. However, with recent advances in eye tracking technology, we have a system that is remote, not burdened, non-invasive and accurate to within 0, 5 - 1, 0 degree of accuracy [10].

A variety of eye trackers based on image processing have been described in the literature. Deng et al. [11] presented a region-based deformable template method for locating the eye and extracting eye features. A system based on a dual state model for tracking eye features is proposed in [12]. Both of these approaches require manual initialization of eye location. Witzner and Pece [13] also model the iris as an ellipse, but the ellipse is locally fitted to the image through an EM and RANSAC optimization scheme. They propose a likelihood model that incorporates neighboring information into the contour likelihood model and furthermore also avoids explicit feature detection. This method allows for multiple hypotheses tracking using a particle filter. The aim is to use the method in cases where thresholds are difficult to set robustly.

A blink detection algorithm for human computer interaction has been proposed by Morris [14], in which the initialization step requires motion detection to locate the eyelids. Any other significant movement on the face, such as that induced by speaking, may cause the system to fail in detecting eyelids. De la Torre et al. [15] describes a similar system for driver warning based on principal component analysis (PCA) and active appearance models (AAM), which requires a training period for each user. Although this method is robust to translation and rotation, it may fail if the face view is not fully frontal. Recently, a system based on feature detection using one camera has been suggested by Smith [16]. Noureddin et al. [17] suggest a two camera solution where a fixed wide angle camera uses a rotating mirror to direct the orientation of the narrow angled camera. They show that the rotating mirror speeds up acquisition in comparison to a pan-tilt setup. Artificial Neural Network (ANN) also has been applied into eye gaze estimation in the literature. This approach employs cropped images of eyes to train regression functions, as seen in multi layer network [18].

This paper claims that gaze information and the point where users look at the screen can be captured by means of deformable template matching based eye tracking system even the eye movements are noisy and pupil center is not known exactly. Besides this, eye pupil center detection is more robust than current systems by means of darkest pixel searching in an eye socket by modeling an arbitrary gray-scale eyeball with center of mass and deformable template windows. For robust pupil center detection, there will be no necessity for infra-red head-mounted camera. Proposed algorithm can determine where the user's interest is focused automatically. Finally, a sample dwell-time based virtual keyboard will be used in order to measure the accuracy of the proposed system and a benchmarking of several main-stream algorithms with our algorithm will be performed at the end.

First, we detect the eye regions by means of Intel OpenCV AdaBoost Haar-cascade classifiers and assign the approximate size of eyeball depending on the eye region size. Secondly, using DAISMI (Deformable Angular Integral Search by Minimum Intensity) algorithm, pupil center is detected. Then, by using the percentage of black pixels over eyeball circle area, we convert the image into binary (Black and white color) for being used in the next part: DTBGE (Deformable Template based 2D Gaze Estimation) algorithm. Finally, using DTBGE algorithm, initial pupil center coordinates are assigned and DTBGE creates new pupil center coordinates and estimates the final gaze directions and eyeball size. We have performed extensive experiments and achieved very encouraging results. Finally, we discuss the effectiveness of the proposed method through several experimental results.

II. DEFORMABLE ANGULAR INTEGRAL SEARCH ALGORITHM BY MINIMUM INTENSITY(DAISMI)

Start

-Get the image from camera.

-Convert the image into gray scale.

-Because the pupil is shiny due to environmental lightening effects, in order to get rid of it, pass the image through mean filter: 3X3 convolution matrix: { {1/9f, 1/9f, 1/9f}, {1/9f, 1/9f, 1/9f}, {1/9f, 1/9f, 1/9f} }.

-Detect the face by OpenCV AdaBoost Haar Cascade Classifier and get the facial features: Up-Left corner coordinates, face width and face height.

-If there is at least one face detected, then detect the eye sockets by OpenCV Adaboost Haar Cascade Classifier and get the eye region features: Up-Left corner coordinates, eye socket width and eye socket height.

-If there is at least one eye socket detected, then determine which eye socket is for left eye and which is for right eye by establishing a simple algorithm such that if eye socket resides in the left part of the middle of the face, then it is left; otherwise right eye.

-For each pixel in the eye socket, find the intensity difference (***d***) between the each pixel's intensity and black color. Because the image is grayscale, the minimum intensity difference will yield the pupil center in theory.

-The closest difference (***closestDiff***) can be maximum 255 when the pixel's intensity value is 255. Because black color's intensity is 0, difference is 255-0=255.

```
Start Loop (Iterate through each pixel in the eye socket)
    If d < closestDiff Then
        -Assign the x and y coordinates of the iterated pixel in the loop to the eyeball as its center coordinates.
        -Assign the eye socket's height divided by 3 to the eyeball as an approximate diameter. (Experiments confirmed this approximation by trial and error methodology.)
        -Draw 8 deformable template windows around the eyeball. Window size is the radius of the eyeball.
        Start Loop (Iterate through the each pixel in 8 deformable template windows. The number of the pixels is 8 multiplied by window size that is 8 X eyeball radius.)
            -Create and initiliaze the variable Counter=0 at the begining of each loop.
            -Scan all the pixels in eye ball circle by angular integral search algorithm.
            -Two nested loops in which iterators are i and j. (0<=i<=8) and (0<=j<=radius).
            -Center of the eye ball is defined with center_x and center_y.

            point_x=abs (int (center_x + j*cos (PI*i/4)));
            point_y=abs (int (center_y + j*sin (PI*i/4)));

            Intensity of the pixels in windows is:
            c_win=intensity (point_x, point_y).
            Intensity of the pixel in the eyeball center is:
            c_center=intensity (center_x, center_y).

            If c_win > c_center Then
                Counter=Counter+1;
            End If
        End Loop

        If Counter >= (window_size X 7) Then (In theory, it must be window_size X 8 but we used some tolerance for better results.)
            ClosestDiff = d;
        End If
    End If
End Loop
```

-Do the same angular integral search by minimum intensity algorithm for both two eye regions.

-Convert the image into binary (black and white) by using auto-thresholding algorithm in which percentage of black pixels over the eye ball determines the optimum thresholding by means of p-tile thesholding algorithm.

Stop

Fig. 1. Eye ball detection in various positions by DAISMI.

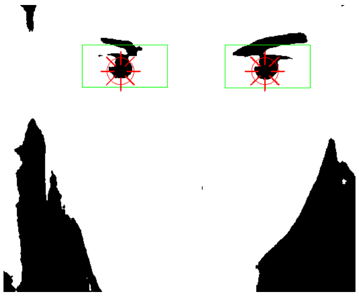


Fig. 2. Binary image conversion for deformable template matching based 2D Gaze Estimation Algorithm.

III. DEFORMABLE TEMPLATE BASED 2D GAZE ESTIMATION ALGORITHM (DTBGE)

After getting the binary eyeball images, we can step into the second part of the algorithm. At this part, a binary deformable eyeball template will be modeled and 2D Gaze Estimation will be performed with respect to the distance in eye movements. At the beginning of the algorithm, the position of the cursor is assumed as the center of the screen. At any horizontal and vertical eye movement, the distance vector will be estimated as a means of frame by frame pixel-wise distance among eye movements.

This part of the algorithm consists of 4 steps as follows:

A. The basic configuration

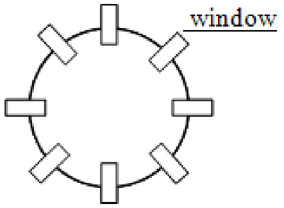


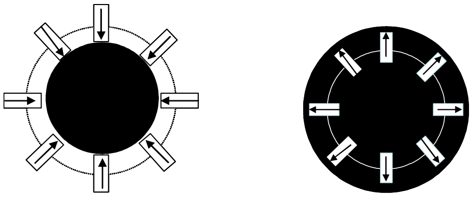
Fig. 3. Deformable template windows.

The deal is to find the location of the basic deformable model of the eyeball shown in Figure 3. Assumption is the fact that any kind of deformable eyeball has a virtual radius in terms of the size of the circle. The arbitrary deformable eyeball circle has 8 windows. In theory, the best precision can be taken if the number of windows is in the number of circumference of the circle. However, the deal is the speed problem. We have tested the algorithm with many window numbers; best results have been taken with 8 windows. Therefore; the number of the windows in the deformable eyeball template is assumed as 8 in this algorithm. Because the image is binary, the black part of the image is ('0 ') and white part is ('255'). Using earned value through the window object orientation and size of the pan is adjusted. First of all, its energy for the window (expression) can be defined as in Eq.(1).

$$E_w = \frac{\sum(\text{ intensity value of each pixel within the window})}{\text{number of pixels in the window}} - \frac{255}{2} \quad (1)$$

B. Adjustment of the size of object

As defined above by the energy values of the window, the size of the object can be defined by means of angular energy values. Finally, $E_W(\theta)$ passes through the center of the object and the resulting value is the energy of the window with angle θ.



(a) Shrinking eyeball (b) Expanding eyeball

Fig. 4. Resizement of the deformable object.

Fig. 4 (a) as an object, the size of the eye (appearing as black part) is greater than the size of the sum of the amount of energy. In contrast, Fig. 4 (b) the size of the window is less than the sum of the energy. In other words,

$$E_{resize} = \int_0^{2\pi} E_w(\theta)d\theta \quad (2)$$

Eq. (2) implies the fact that the size of the object will increase if the sum of angular energy values of windows less than ‘0’, conversely; it will decrease if the sum of angular energy values of windows greater than ‘0’.

C. Horizontal Movement of the Object

In order to find the exact location of the eyes, deformable object must move from side to side. Cosine function between 90 ° ~ 270 ° has a negative value. By using the Eq. (3), the direction of the horizontal movement is determined.

$$E_{move(x)} = \int_0^{2\pi} E_w(\theta)\cos\theta d\theta \quad (3)$$

D. Vertical Movement of the Object

In order to find the exact location of the eyes, deformable object should move up and down as well. Sine function between 0 ° ~ 180 ° has a positive value. By using the Eq. (4), the direction of the vertical movement is determined.

$$E_{move(y)} = \int_0^{2\pi} E_w(\theta)\sin\theta d\theta \quad (4)$$

In the system, the calibration procedure is very simple. All the user needs to do is to look at the center of the monitor for just one time. This is a promise between user and the system. When the user moves his/her eyes, the amount of pixel-wise pupil displacement estimated by means of DTBGE will be oriented to the screen relative to the screen size assuming the simple matter of the fact that the initial position of the gaze is the center of the screen.

IV. EXPERIMENTAL RESULTS

The system was primarily developed and tested on a Windows XP PC with an Intel Pentium Dual CPU (T2390) with 1.86 GHz processor and 1.75 GB RAM. The monitor size was 12.1" (307 mm) with 9.72" (247 mm) width and 7.28" (185 mm) height with 4:3 aspect ratios. Video was captured with a Logitech Quickcam Pro 4000 webcam at 30 frames per second. All video was processed as binary images of 160 X 120 pixels using various utilities from the Intel OpenCV library[19].

A. Eye Tracking Experiment

The experiment was conducted to test the eye tracking system whether it can successfully locate the pupils of the users under different conditions. We collected a data set consisting of 300 images taken under varying lighting conditions, head positions and with complex backgrounds. The experimental result shows that the performance rate can reach 100% at FPS (Frame per Second) =15 while the camera input image size was 160 x 120.

TABLE I
COMPARISON OF DTBGE WITH SEVERAL MAIN-STREAM 2D GAZE ESTIMATION ALGORITHMS

Infra Red	Training and Calibration	Pupil Center Based	Camera and Video Size	Accuracy (deg)	Algorithms
Yes	No-No	Yes	Head-Mounted (320 × 240)	2 – 4	Active Contours [13]
Yes	No-Yes	Yes	Head-Mounted (320 × 240)	2.9	Image Processing [17]
No	Yes-No	Yes	Head-Mounted (320 × 240)	1.5	ANN [18]
No	No-No	No	Webcam (160 × 120)	2	DTBGE

B. 2D Gaze Estimation and Mouse Control Experiment

The second experiment was to ask six users to move the cursor to the positions of virtual keyboard buttons which were generated relative to the screen size. Each user was given a brief introduction to how the system works and then directly conducted the experiment without any practice. Each user conducted the experiment five times.

During the experiment, users could hit the expected button in different durations. Even they cannot succeed in the true hit; their resulting gaze was so close to the expected gaze in terms of unit error distance. Therefore, unit error distance values cannot be used in terms of gaze accuracy evaluation. That's why most of the gaze trackers are evaluated with respect to their visual angle accuracy. For our system, we estimated the visual angle accuracy as around 2°.

V. CONCLUSIONS

In this paper, a shape and intensity based deformable eye pupil center detection algorithm and a deformable template matching based eye movements and pupil size detection algorithm has been presented in terms of developing a low-cost 2D Gaze Estimation and Cursor Control system regardless of the robust position of pupil center. In many other algorithms which are based on PoR (Point of Regard) information, robust pupil center position is required in order to get rid of noisy eye movements. However, unlikely; our algorithm DTBGE uses a new deformable template matching based eye movement detection algorithm in binary source images regardless of knowing the pupil center. The approximate position of pupil is enough for getting robust eye movement detection and corresponding pupil center displacement. This algorithm is very useful for those who want to build 2D Gaze Estimation and Mouse Control systems with low resolution source images such as simple web-cams without Infra-Red filter. Even the pupil location data is noisy, it is not important for robust pupil displacement by means of DTBGE. Experimental results showed very good results such that the visual angle accuracy was around 2° while the speed was 15 fps.

The comparison of our algorithm with other algorithms in the same section is as follows:

In Table I, DTBGE has no calibration, no training process, no infra-red filtering with head-mounted camera and it is the only algorithm which doesn't need pupil center for robust 2D Gaze Estimation even it uses low-resolution image (160 × 120) and the success rates are so close to each other. For this reason, DTBGE is an ideal option for those who cannot afford expensive trackers, does not like calibration and wants to get high accuracy even the system is with a simple web-cam. Finally, DTBGE can be used with more performance if the auto-thresholding can be provided more robustly. The only deal in the algorithm is the ratio between white and black pixels in template windows and this is all about thresholding. Therefore, better thresholding algorithms can yield better results.

ACKNOWLEDGMENT

This research was supported by Kyungsung University Research Grants in 2010.

REFERENCES

[1] M. Betke, J. Gips, and P. Fleming, "The camera Mouse: Visual tracking of body features to provide computer Access for people with severe disabilities," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 10, no. 1, pp. 1-10, Mar. 2002.

[2] M. Betke, W. Mullally, and J. Magee, "Active detection of eye scleras in real time," *Proceedings of the IEEE CVPR Workshop on Human Modeling, Analysis and Synthesis (HMAS 2000)*, Jun. 2000.

[3] R. L. Cloud, M. Betke, and J. Gips, "Experiments with a camera-based human-computer interface system," *Proceedings of the 7th*

ERCIM Workshop, User Interfaces For All (UI4ALL 2002), pp. 103-110, Oct. 2002.

[4] S. Crampton, and M. Betke, "Counting fingers in real time: A webcam based human-computer interface with game applications," *Proceedings of the Conference on Universal Access in Human-Computer Interaction (affiliated with HCI International 2003)*, pp. 1357-1361, Jun. 2003.

[5] C. Fagiani, M. Betke, and J. Gips, "Evaluation of tracking methods for human-computer interaction," *Proceedings of the IEEE Workshop on Applications in Computer Vision (WACV 2002)*, pp. 121-126, Dec. 2002.

[6] J. Lombardi, and M. Betke, "A camera based eyebrow tracker for hands-free computer control via a binary switch," *Proceedings of the 7th ERCIM Workshop, User Interfaces For All (UI4ALL 2002)*, pp. 199-200, Oct. 2002.

[7] J. J. Magee, M. R. Scott, B. N. Waber, and M. Betke, "Eyekeys: A real-time vision interface based on gaze detection from a low-grade video camera," *Proceedings of the IEEE Workshop on Real-Time Vision for Human-Computer Interaction (RTV4HCI)*, Jul. 2004.

[8] G. T. Buswell, "How People Look at Pictures: A Study of The Psychology of Perception in Art," *The University of Chicago Press*, 1935.

[9] A. L. Yarbus, "Eye Movements and Vision," *Plenum Press*, 1967.

[10] Tobii Eye-tracker, Tobii Technology Japan, Ltd 3-4-13, Takanawa, Minato-ku, Tokyo, 108-0074, Japan. Home Page [Online]. Available: http://www.tobii.com.

[11] J. Deng, and F. Lai, "Region based template deformation and masking for eye-feature extraction and description," *Pattern Recognition*, vol. 30, no. 3, pp. 403-419, 1997.

[12] Y. Tian, T. Kanade, and J. Cohn, "Dual-state parametric eye tracking," *Proc. Of Conf. On Automatic Face and Gesture Recognition*, pp. 110-115, 2000.

[13] D. Witzner Hansen and A. E.C. Pece, "Eye tracking in the wild," *Computer Vision and Image Understanding*, vol. 98, no. 1, pp. 182–210, Apr. 2005.

[14] T. Morris, F. Zaidi, and P. Blenkhorn, "Blink detection for real-time eye tracking," *J.Networking and Computer Applications*, vol. 25, no. 2, pp. 129-143, 2002.

[15] F. De la Torre, C. Garcia Rubio, and E. Martinez, "Subspace eyetracking for driving warning," *Proc. Of ICIP*, vol. 3, pp. 329-332, 2003.

[16] P. Smith, M. Shah, and N. Da Vitoria Lobo, "Determining driver visual attention with one camera," *Trans. on Intelligent Transportation Systems*, vol. 4, no. 4, pp. 205-218, 2003.

[17] B. Noureddin, P.D. Lawrence, and C.F. Man, "A non-contact device for tracking gaze in a human computer interface," *Computer Vision and Image Understanding*, vol. 98, no. 1, pp. 52–82, 2005.

[18] Shumeet Baluja and Dean Pomerleau, "Non-intrusive gaze tracking using artificial neural networks," *Advances in Neural Information Processing Systems*, vol. 6, pp. 753–760, 1994.

[19] OpenCV Library Home Page [Online]. Available: http://sourceforge.net/projects/opencvlibrary/

Jinwoo Kim received the B.S degree in Electrical Engineering from Myongji University in 1992 and the M.S and Ph.D. degrees in Electronic Engineering and System design Engineering from Fukui University, Fukui, Japan, in 1996 and 1999, respectively. From 2000 to 2003, he was a contract Professor in the Department of Information Communication and Computer Engineering at Hanbat National University, Daejeon, Korea. Since 2003 he has been with the Department of Information and Communication Engineering at Kyungsung University, Busan, Korea, where he is currently an associate professor. From Dec., 2007 to Mar., 2009, he was a visiting researcher in the Department of Bioengineering at Tokyo University, Japan. His current research interests include image processing, pattern recognition, and medical imaging technology.