A Method for Localizing the Eye Pupil for Point of Gaze Estimation

By Narayan Panigrahi, Kiran Lavu, Sai Krishna Gorijala, Peter Corcoran, and Saraju P. Mohanty

Estimating gaze point through localization of the pupil in the image of the eye through optical eye-tracking during opto-ocular motion of the eye possess many challenges. This article discusses a method to estimate point of gaze on a screen using non-intrusive eye tracking. A modified web camera is used to obtain centre of pupil and is processed further to obtain the point of gaze.

1. INTRODUCTION

Estimation of point of gaze in a scene presented in a digital screen has many applications such as fatigue detection and attention tracking. Some popular applications of eye tracking through gaze estimation is depicted in Figure 1. For estimation of point of gaze it is required to identify the visual focus of a person within a scene. This is known as eye fix or point of fixation. Finding the point of gaze involves tracking of different features of human eyes. Various methods are available for eye tracking. Some of them use special contact lens [1] while some others use electrical potential measurement [2]. Optical tracking of eyes is a non-intrusive tracking technique which uses a sequence of image frames of eyes recorded using video capturing devices. This technique is popularly known as video oculography (VOG). Different techniques used in VOG for the purpose of eye tracking are pupil-corneal reflection vectors or Purkinje images [3], pupil-eye corner vectors [4].

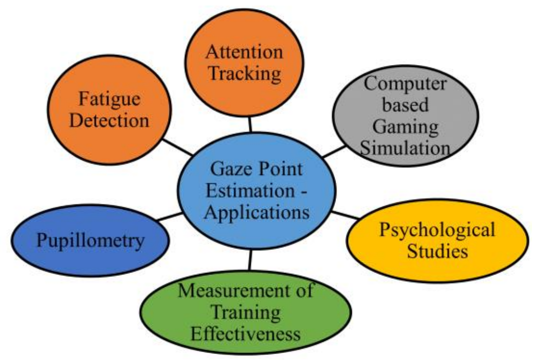


FIGURE 1. Applications of gaze point estimation.

All of these techniques successfully work only if two or more features of eyes are detected accurately. This increases the computational complexity thereby limiting the speeds of eye tracking systems. Some commercial eye tracking equipments available in the market are Eyelink-1000, SMI, Tobii glasses. The cost of these systems is quite high and therefore can be used only for research purposes. In this article a low cost system is proposed that uses a modified web camera to obtain fixation points of subjects within a scene using centre of pupil. The system is capable of generating 25-30 frames per second which is sufficient to compute the position and size of pupil in the pixels. Eye tracking and gaze position estimation has many applications.

2. PREVIOUS RESEARCH ON GAZE ESTIMATION

The first application of eye-tracking for an eye-controlled computer was reported in [5]. The eye-gaze as an input for human-computer interaction is shown in [6]. Different eye-tracking algorithms, applications and methods were reported in [7, 8, 9,10]. Many interesting applications of real-time eye gaze tracking for gaming design and Consumer Electronics Systems are available in [11]. Eye tracking can be done using methods such as 2D regression and 3D modelling of eye. The 3D modelling methods implement a geometric model of eye to estimate the point of intersection of visual axis and the scene. These systems typically use single camera [12]. It is observed that if the number of cameras and light sources can be increased to reduce the number of parameters to be determined during calibration [13, 14]. Models utilize two or more cameras have been explored in [15, 16, 17, 18]. However, these models might require multiple cameras or complex mathematical models for estimation of centre of cornea and estimate visual axis. The other popular method is 2D regression technique. In this method, the pupil – glint vector is mapped to the gaze coordinates by using a mapping function. This mapping function can be resolved using polynomial regression or artificial neural networks. methods used for calibration and deducing mapping functions using regression technique are presented in [19, 20]. A system that implement artificial neural networks for determining the mapping function used for translating pupil – glint vector to gaze points is available in [21]. All of these regression techniques depend on accurate detection of two or more features of the eye for accurate tracking. In [11], a system that detects human face in an image followed by detection of eyes and estimate centre of pupil from the image has been proposed. Due to this, the image of eye should be processed for multiple times to detect all the required features. The accuracy of the system depends on accurate identification of these features. This uses a considerable amount of computational effort leading to slow performance of the system. Through the current paper, we propose a system that uses only a single feature, the centre of pupil, to estimate point of gaze.

3. A SPECIFIC EYE TRACKING METHOD

The first step of obtaining point of gaze is to obtain coordinates of the centre of pupil from snapshots of eye of a subject. There are two popular ways of performing that in the snapshots: (a) Passive Imaging (b) Active Imaging [22-23].

Passive imaging uses visible light available in the surroundings to capture images of eyes as depicted in Figure 2. These images are processed to detect various features of eyes such as eye corners and pupil.

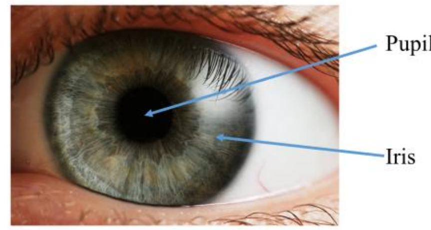


FIGURE 2. Image of eye captured using passive imaging.

Active Imaging uses an illumination source along with the imaging device. If the illumination source emits light in visible spectrum, it causes discomfort to the user and limits duration for which the experiment can be conducted. If the source is coaxial to the optical axis of eye, IR light gets reflected from retina and illuminates pupil leading to bright pupil effect as depicted in Figure 3(a). If the illumination source is offset from optical path, the retro reflection from retina is directed away from the camera causing dark pupil effect as depicted in Figure 3(b). The bright pupil effect is difficult to obtain due to the requirement that the illumination source is to be coaxial to optical path. This requires specialized IR illumination source. On the other hand, dark pupil effect can be produced easily and hence it is widely used technique for eye tracking.

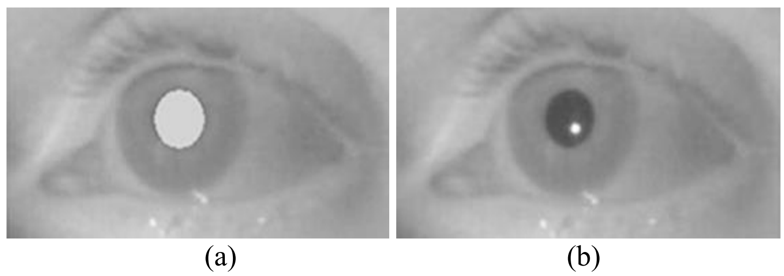


FIGURE 3. (a) Bright Pupil effect. (b) Dark Pupil Effect.

3.1. DETECTION OF PUPIL

The main feature required for any eye tracker is to detect the (x, y) coordinates of centre of pupil. However, to identify the relative motion of pupil, at least one other feature that stays fixed during eye movements is required. This feature can be an eye corner or a glint formed by the IR illumination source. The vector of pupil-glint or pupil-eye corner can be used to obtain point of regard on a screen allowing slight freedom in movement of user head. However, if the head pose of the viewer is stabilized, (x, y) coordinates of centre of pupil alone is sufficient to obtain the eye fix. This reduces the computational complexity required to identify other features of eye such as glints and eye corners.

Detection of location of pupil is a key feature of any eye tracking mechanism. The accuracy of the system largely depends on accurate detection of centre of pupil for all positions of eyes. Detection of pupil is easier when using bright or dark pupil effect (refer Figure 3(b)). For simplicity, the discussion is limited to dark pupil effect. The first step is to acquire an image of eye using IR camera. The dark regions in this frame consist of eye lashes, a part of eye brow and the pupil. Of all these dark regions, pupil has a well-defined shape and is the largest connected component in the entire image. The second step is to obtain a binary image which consists of dark regions only. This can be achieved by applying an intensity threshold transformation to the image. A threshold value is chosen which represents the value of intensity of colour in each pixel which is to be considered as dark. If all the three values of any pixel i.e., R, G and B are less than that threshold value, then that pixel is considered as dark. Thus all the pixels with RGB values above the

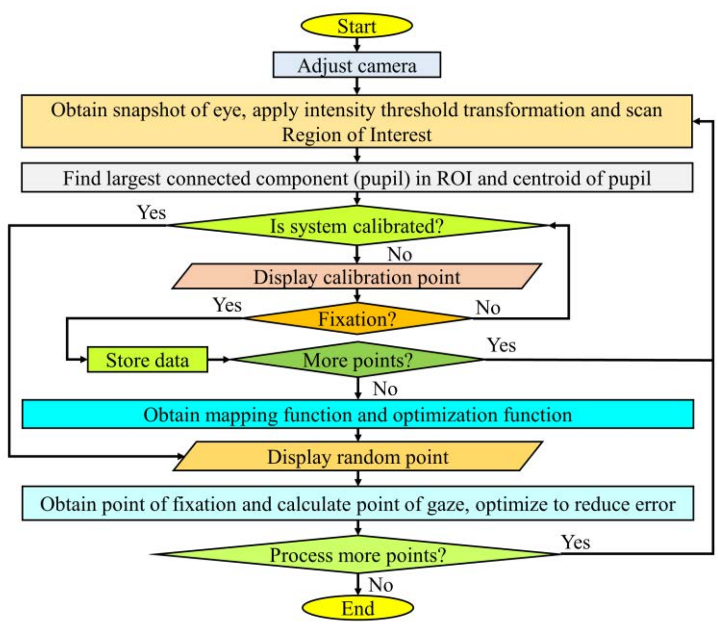


FIGURE 4. The proposed method for pupil detection.

threshold are painted in white and all the pixels with RGB values less than or equal to threshold are painted in black. Now, the binary image contains pupil and a few other dark spots from lashes and eye brows.

The next step is to find the largest connected component i.e., the pupil in this image. One way is to process entire image to detect a largest connected component. This method is suitable when head pose is not fixed. However, if head pose is stable, a small region can be chosen within which the pupil can be located and process that region only in every frame. This reduces a significant amount of computational effort and time required to detect the pupil. To find the largest connected component Scan plus Array based Union Find (SAUF) algorithm [23] is used. Once the pupil is detected, the centroid of pupil will give the (x, y) coordinate of centre of pupil. The bright pupil effect follows same principle except for the fact that pupil appears bright and hence the threshold effect is to be applied in such a way those only bright regions remain. The rest of the procedure is same as that of dark pupil effect. The proposed method is depicted in the flow-chart in Figure 4.

3.2 OBTAINING POINT OF GAZE FROM PUPIL COORDINATE

The pupil coordinate will provide the information on location of pupil in any captured frame. However, the goal of detecting pupil coordinate is to calculate a point of gaze from it. For this purpose, a personal calibration routine is run for every session. A calibration routine typically displays a few points or dots in a particular order or in random manner. The users are instructed to fixate on those points and while doing so, the corresponding pupil coordinate for each point is recorded. This data is used to obtain a modelling function which gives the relation between pupil coordinates and actual screen coordinates. If (S_x, S_y) are the screen coordinates and (e_x, e_y) are the pupil center coordinates, then two separate functions can be used to relate screen coordinates to pupil coordinates:

$$S_x = f_x(e_x, e_y) \quad \text{---} \quad \text{(i)}$$

$$S_y = f_y(e_x, e_y) \quad \text{---} \quad \text{(ii)}$$

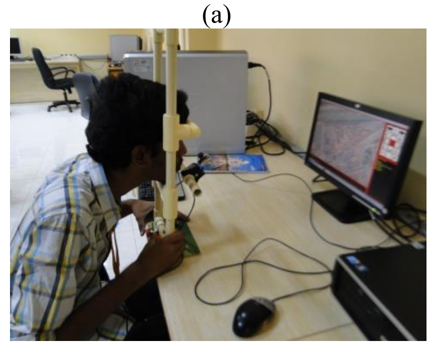
The functions f_x and f_y are modelled as polynomials in e_x, e_y and are represented in equations (iii) and (iv) [24]:

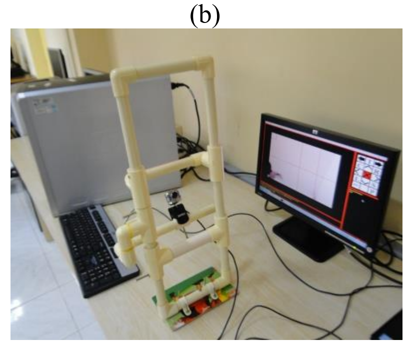
$$S_x = a_0 + a_1 e_x + a_2 e_y + a_3 e_x e_y \quad \text{---} \quad \text{(iii)}$$

$$S_y = b_0 + b_1 e_x + b_2 e_y + b_3 e_x e_y \quad \text{---} \quad \text{(iv)}$$

To compute the actual coordinates (S_x, S_y) for corresponding pupil coordinates (e_x, e_y), we have computed the value of the polynomial coefficients (a_0,..a_3) and (b_0,.. b_3) using regression techniques. Two most popular regression techniques are Polynomial Affine Transformation [24] using matrix inversion and Least squares technique. In polynomial affine transformation, four equations are obtained by substituting the recorded data by displaying four different points on the screen and the corresponding pupil coordinates. These equations are solved using matrix inversion method to obtain unknowns i.e. (a_0, ..a_3) and (b_0,..b_3). The other method is to display any number of screen coordinates and pupil coordinates and obtain the values of coefficients (a_0, a_3) and (b_0, b_3) using least squares method. The value of computed coefficients are used, to estimate a new points of gaze on screen (S_x, S_y) corresponding to any pupil coordinate (e_x, e_y). A detailed implementation and mathematical rigor of PAT (Polynomial Affine Transformation) for registration of satellite images with map by removing geometric errors is discussed in [24]. The actual experimental setup is shown in Figure-5(b), 5(c) and 5(d).

(a)


(b)


(c)

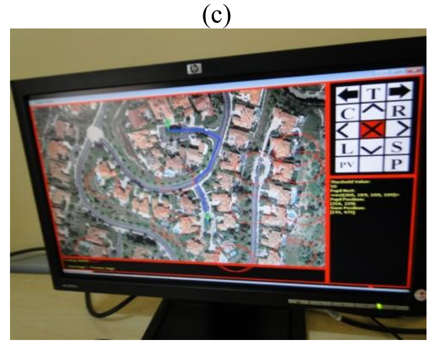
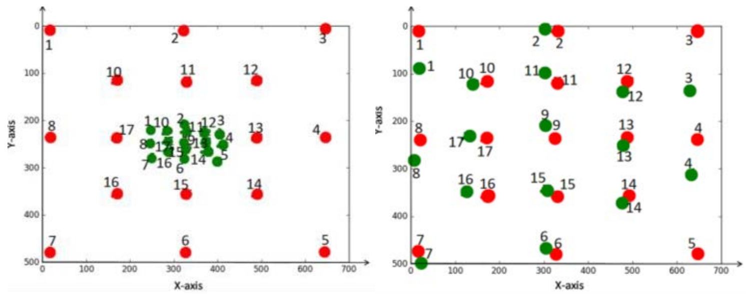

(d)

FIGURE 5: Experimental setup for gaze estimation. (a) Illustration, (b) The Actual Experimental Setup (c) The Head Fix with Camera (d) The GUI with eye-fixations graph.

4. A CASE STUDY

The experimental setup consists of a screen displaying the scene in which fixation graph is to be identified. A customized web camera was mounted on a chin rest to obtain snapshots of eyes. The camera was modified to detect infrared light by removing the IR blocking filter from its lens assembly. The chin rest acts as a support to stabilize head movement of the subject. The screen is placed at a distance of 50 cm from subject's eye as in Figure 5. The overall display resolution for the scene is 800x600 pixels. The scene is 22 cm in height and 35 cm wide. This provides a visual cone with 25^{O} vertical and 39^{O} horizontal field of view and hence the resolution of visual angle is 26 pixels per degree vertically and 21 pixels per degree horizontally. IR LEDs were used to illuminate eyes of the subject.

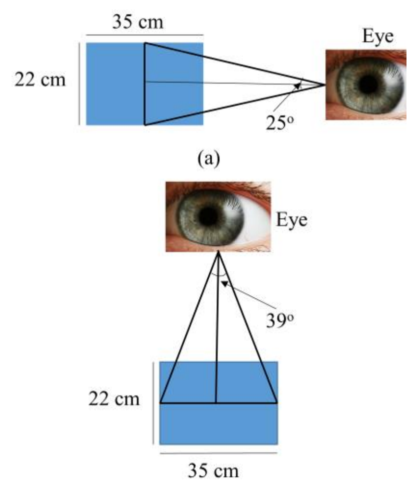


(a) (b)

FIGURE 6. (a) Phase 1 in calibration. (b) Phase 2 in calibration.

First step in calibration was to match the coordinate system of eye with coordinate system of screen. In order to achieve this, a crosshair is displayed on screen. The subject is instructed to fixate on centre of crosshair and the camera is adjusted such that centre of pupil is exactly at centre of crosshair displayed on the screen. Then, calibration is done by repeating the fixate process using a set of 17 random points in the screen in two phases. During first phase, these points are displayed on the screen as red coloured dots. The subject is instructed to fixate on these points and whenever a fixation was detected, the corresponding pupil position is recorded. The red coloured dots in the Figure 6 represent actual screen coordinates and green dots represent their corresponding pupil coordinates.

These coordinate pairs as recorded in the first phase are used to establish the PAT by computing the coefficients of the polynomial. These polynomial coefficients are used to estimate a point of gaze on the screen using equations (iii) and (iv). The second phase is added to improve accuracy of this estimation. During this phase, the estimated points of gaze and the actual screen coordinates computed using the PAT are used in polynomial regression to obtain a function to optimize estimated point of gaze. The red coloured dots in Figure 6 represent actual screen coordinates and green coloured dots represent estimated points of gaze.



(b)

FIGURE 7. (a) Vertical visual cone. (b) Horizontal visual cone.

In this specific case study, the screen was placed 50 cm away from the subject's eye. The overall display resolution for displaying the scene was 800x600 pixels. The scene was 22 cm high and 35 cm wide. The angles subtended by the eye vertically and horizontally with the scene were measured (Figure 7). The vertical angle was 25° and the horizontal angle was 39°. This provides a visual cone with resolution of 24 pixels per degree vertically and 21 pixels per degree horizontally.

To test the system, five random points were displayed and their corresponding points of gaze are calculated. The average error for each session was calculated by:

$$E = \sum \sqrt{(xs - xc)^2 + (ys - yc)^2} \text{---} \quad \text{(v)}$$

where xs, ys are actual screen coordinates displayed on screen and xc, yc are calculated points of gaze by the software. The experiment was performed using matrix inversion technique and least squares method to deduce the mapping function [24]. The average error in both the cases is presented in Table 1. The average error was found to be 48 pixels. This corresponds to 2.105° of visual angle. A sample of experimental data can be seen in Figure 8. The performance of the proposed method is compared with state of the art methods as presented in Table 2.

Table 1. Error Analysis of the Pupil Localization.

	Matrix Inversion	Least Squares
Horizontal Error	2.8° (59 Pixels)	1.81° (38 Pixels)
Vertical Error	3.2^0 (77 Pixels)	2.4° (58 Pixels)
Mean Error	3.0° (68 Pixels)	2.10° (48 Pixels)

5. CONCLUSIONS

The presented system uses only one feature of eye i.e., centre of pupil of eye of the subject to obtain point of gaze on a screen. This reduces computational effort required for detecting other features such as eye corners, corneal reflections and glints, enabling for high speed gaze tracking. Also, noise in detection of these additional features is eliminated thereby increasing the confidence and reliability of the data acquired from our system. As the proposed system uses a simple web camera, the cost of the system is lower than many other eye trackers while achieving accuracies similar to that of expensive methods. The future directions of this research can be multi front [25]. Specifically, the future research includes integration of the proposed system in a smart healthcare framework [26]. The point of gaze on a screen can be fused with time interval when the beta wave of the EEG is high of the subject to ascertain the point of fix on the screen.

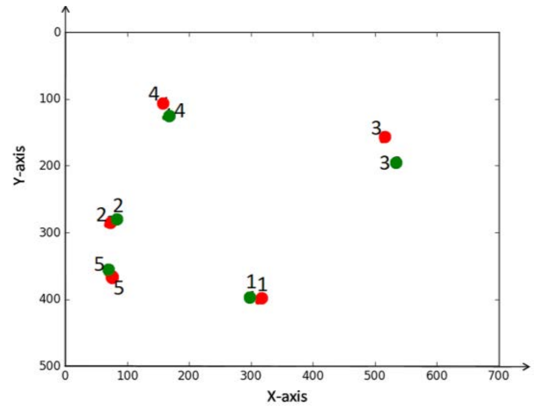


FIGURE 8. Test case for gaze estimation.

ABOUT THE AUTHORS

Narayan Panigrahi (pani@cair.drdo.in) is a scientist in the Centre for Artificial Intelligence and Robotics (CAIR), Bengaluru. He is an inventor of four Indian patents and has authored four books and fifty research articles in peer reviewed conferences and journals.

Kiran Lavu (kiranlavu9@gmail.com) graduated Electronics and Communication Engineering (ECE) from Vignan's University, Vadlamudi, India in the year 2017.

Sai Krishna G. (saikrishnagorijala@gmail.com) graduated Electronics and Communication Engineering (ECE) from Vignan's University, Vadlamudi, India in the year 2017.

Peter Corcoran (dr.peter.corcoran@ieee.org) is a Fellow of IEEE, the Founding Editor of IEEE Consumer Electronics Magazine and holds a Personal Chair in Electronic Engineering at the College of Engineering & Informatics at NUI Galway. He is co-author on 300+ publications and co-inventor on 300+ granted US patents.

Saraju P. Mohanty (saraju.mohanty@unt.edu) is a Professor at the University of North Texas. His research is in Smart Electronic Systems. He is an author of 250 research articles, and 3 books. He is the Editor-in-Chief of the IEEE Consumer Electronics Magazine. He serves as the Chair of Technical Committee on VLSI, IEEE Computer Society.

REFERENCES

[1] D. A. Robinson, "A method of measuring eye movement using a scleral search coil in a magnetic field," *IEEE Transactions on Bio-Medical Electronics*, Vol. 10, No.4, October 1963, pp. 137–145.

[2] A. Bulling, "Eye Movement Analysis for Activity Recognition Using Electrooculography," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Volume 33, Issue 4, April 2011, pp. 741-753.

[3] H. D. Crane and C. M. Steele, "Generation-V dual-Purkinje-image eyetracker", *Applied Optics*, Vol. 24, No. 4, 1985, pp. 527–537.

[4] J. Zhu and J. Yang, "Evaluation of pupil centre-eye corner vector for gaze estimation using a web cam", in *Proceedings of the Symposium on Eye Tracking Research and Applications*, 2012, pp. 217-220.

[5] J. L. Levine, "An eye-controlled computer", IBM Thomas J, Watson Research Center, Research Report RC-8857, Yorktown Heights, N. Y., 1981.

[6] A. T. Duchowski , "A breath-first survey of eye-tracking applications." Behavior Research Methods, Instruments, & Computers, vol. 34, no. 4, 2002, pp. 455-470.

[7] J. G. Wang, and E. Sung, "Study on eye gaze estimation," *IEEE Trans. Syst., Man, and Cybernetics - Part B*, vol. 32, no. 3, Jun. 2002, pp. 332-350.

[8] D. H. Yoo, and M. J. Chung, "A novel non-intrusive eye gaze estimation using cross-ratio under large head motion," *Computer Vision and Image Understanding*, vol. 98, no. 1, Apr. 2005, pp. 25-51.

[9] A. Sugioka, Y. Ebisawa, and M. Ohtani, "Noncontant video-based eye-gaze detection method allowing large head displacements," in *Proceedings of the International Conference on Medicine and Biology Society*,1996 , pp. 526-528.

[10] H. C. Lee, D. T. Luong, C. W. Cho, E. C. Lee, and K. R. Park, "Gaze Tracking System at a Distance for Controlling IPTV" *IEEE Transactions on Consumer Electronics*, vol.56, no.4, November 2010, pp.2577-2583.

[11] P. M. Corcoran et al.: Real-Time Eye Gaze Tracking for Gaming Design and Consumer Electronics Systems, Vol. 58, No. 2, May 2012, pp. 347-355.

[12] A Meyer, M. Böhme, T. Martinetz, and E. Barth, "A single-camera remote eye tracker", Percept. Interact. Technol., vol. 4021, pp. 208– 211, 2006.

[13] E. D. Guestrin and M. Eizenman, "General theory of remote gaze estimation using the pupil centre and corneal reflections," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 6, pp. 1124- 1133, June 2006.

[14] C. Hennessey, B. Noureddin, and P. Lawrence, "A single camera eye-gaze tracking system with free head motion," Measurement, vol. 1, no. March, pp. 27–29, 2006.

[15] C. C. Lai, S. W. Shih, and Y. P. Hung, “Hybrid method for 3-D gaze tracking using glint and contour features,” IEEE Trans. Circuits Syst. Video Technol., vol. 25, no. 1, pp. 24–37, 2015.

[16] T. Ohno and N. Mukawa, “A free-head, simple calibration, gaze tracking system that enables gaze-based interaction,” Proc. Eye Track. Res. Appl. Symp. Eye Track. Res. Appl. - ETRA’2004, pp. 115–122, 2004.

[17] D. Beymer and M. Flickner, “Eye gaze tracking using an active stereo head,” 2003 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 2003 Proc., vol. 2, no. 2, p. II-451-8, 2003.

[18] Zhiwei Zhu and Qiang Ji, “Novel Eye Gaze Tracking Techniques Under Natural Head Movement,” IEEE Trans. Biomed. Eng., vol. 54, no. 12, pp. 2246–2260, 2007.

[19] P. Blignaut, “Mapping the Pupil-Glint Vector to Gaze Coordinates in a Simple Video-Based Eye Tracker,” J. Eye Mov. Res., vol. 7, no. 1, pp. 1–11, 2013.

[20] Z. R. Cherif, a Nait-Ali, J. F. Motsch, and M. O. Krebs, “An adaptive calibration of an infrared light device used for gaze tracking,” Imtc 2002 Proc. 19th Ieee Instrum. Meas. Technol. Conf. Vols 1 2, no. May, pp. 1029–1033, 2002.

[21] C. Jian-nan, Z. Chuang, Y. Yan-tao, L. Yang, and Z. Han, “Eye Gaze Calculation Based on Nonlinear Polynomial and Generalized Regression Neural Network,” in Proceedings of the 5th International Conferenc on National Computing, pp. 617–623, 2009.

[22] W. Hansen and Q. J. Dan, "In the Eye of the Beholder: A Survey of Models for Eyes and Gaze", *IEEE Trans. Pattern Anal. Mach. Intell*, Vol. 32, No. 3, March 2010, pp. 478–500.

[23] N. Panigrahi N., B K. Mohan., G. Athithan, “Pre-Processing algorithm for Rectification of Geometric Distortions in Satellite Image*”*, Defence Science Journal, Vol. 61, No. 2, pp. 174-179, 2011.

[24] N. Panigrahi and S. Tripathy, “Image Registration using Polynomial Affine Transformation,” *Defence Science Journal*, Vol. 52, No.3, July 2002, pp. 5253-259.

[25] A. Kar and P. Corcoran, "A Review and Analysis of Eye-Gaze Estimation Systems, Algorithms and Performance Evaluation Methods in Consumer Platforms”, *IEEE Access*, Volume 5, 2017, pp. 16495-16519.

[26] P. Sundaravadivel, E. Kougianos, S. P. Mohanty, and M. Ganapathiraju, “Everything You Wanted to Know about Smart Healthcare”, *IEEE Consumer Electronics Magazine (CEM)*, Volume 8, Issue 1, January 2018, pp. 18-28.

Table 2. Comparison of various eye tracking algorithms

Method used and Reference paper	Hardware setup	Features required	Accuracy
3D Geometrical model [12]	2 IR LEDs and one camera	Glints and pupil centre	About 1^o (on simulated data)
3D Geometrical model [14]	Multiple IR illumination sources, one camera	Pupil contour using bright pupil effect and glints using dark pupil effect	0.90^o
3D Geometrical model [15]	2 cameras and 4 IR LEDs	Pupil contour and glints	1.18^o to 1.43^o
3D Geometrical model [16]	2 cameras for face detection and 1 camera on pan and tilt mechanism	3D facial orientation, glints and pupil contour	About 1^o
3D Geometrical model [17]	1 camera for face detection and 2 cameras with mirrors on pan and tilt mechanism	3D facial orientation, glints and pupil contour	0.6^o
3D Geometrical model [18]	2 cameras with multiple IR illumination sources	Pupil and glints	1.6^o
2D Regression [19]	1 camera, 1 IR LED	Pupil and glint	1.17^o
2D Regression [20]	Photo transistors, Pulsed IR illumination	No image is used	2.5^o
2D Regression [21]	2 concentric rings of IR LEDs, 1 camera	Pupil and glints	5^o – 8^o
2D Regression [22]	2 loops of IR LEDs and 1 camera	Pupil and glint	20 pixels
2D Regression, The proposed Method	Modified web camera with built in IR LEDs and IR blocking filter removed	Pupil Centre only	2.10^o