Motion Estimation of Handheld Optical Coherence Tomography System using Real-Time Eye Tracking System

Abira Bright B., Lakshmi Parvathi M., and Vani Damodaran*

Department of Biomedical Engineering, SRM Institute of Science and Technology, Kattankulathur 603203, Tamil Nadu, India

***e-mail:** vanid@srmist.edu.in

Abstract. Optical coherence tomography (OCT) is the clinical golden standard for cross-sectional imaging of the eye. The majority of clinical ophthalmic OCT systems are table-top devices that need the patient to align with the chinrest in order to capture a motion-free image. Portable OCT devices are used to perform retinal imaging on infants or patients who are confined to beds. Eye movements and relative motion between the patient and the imaging probe make interpretation and registration challenging and become a barrier to high-resolution ocular imaging. Thus, an OCT scanner with an automated real-time eye tracking system and a movement mapping for correction mechanism is required to overcome such motions. The aim of this work is to develop an algorithm to track pupil motion and allow motion-corrected imaging of the retina without the requirement of chinrest, fixation of the target, or seating chair and to minimize the requirement of skillset to operate and to correct motion artifacts. Two algorithms based on landmark and threshold were developed, capable of identifying and monitoring eye movements. The acquired output value of both algorithms was compared with the manually calculated actual center value of the pupil. The average deviation from the actual location was found to be 0.2~0.6 for the landmark and 0.4~0.9 for the threshold-based algorithm. In this study, it is observed that iris localization and gaze direction estimation is more accurate in the landmark-based system compared to the threshold-based eye-tracking system.

Keywords: optical coherence tomography (OCT); depth Camera; eye tracking; gaze tracking.

Paper #8306 received 28 Feb 2023; revised manuscript received 11 Aug 2023; accepted for publication 16 Aug 2023; published online 15 Sep 2023. doi: 10.18287/JBPE23.09.030314.

1 Introduction

Optical Coherence Tomography (OCT) is a non-invasive, high-resolution imaging system [1] that provides structural data on biological tissues [2]. It is the gold standard for imaging retinal [3]. Both the anterior and posterior portions of the eye are imaged using it [4]. However, major OCT systems are tabletop devices that are only available in imaging rooms in eye care facilities. Additionally, to operate and align these systems, qualified ophthalmic personnel are required. Only patients who can sit straight, use a chinrest arrangement, and fixate their attention on a target can use these devices. As a result, OCT imaging cannot be performed outside the clinic and for elderly, infant, and bedridden patients. Handheld OCT probes [5] and image registration methods [6] can be used to overcome these barriers of table top systems.

Handheld OCT systems suffer from two issues, i.e. operator stability and patient movement. Operator stability refers to the stability of the probe holding, and patient movement, refers to eye and head movement. Eye movements can be classified into Saccadic movement and Smooth pursuit movement [7]. Humans perform multiple saccades to align their visual axis with the subject of interest, and rapid eye movement help, to

This paper was presented at the International Conference on Nanoscience and Photonics for Medical Applications – ICNPMA, Manipal, India, December 28–30, 2022.

correct the position error between the eye's visual axis and the subject. Smooth pursuit eye movements help stabilize the projection of the moving subject onto the fovea and correct any velocity error between the eye and the subject [8]. These movements can cause distortion while acquiring retinal images which directly affects the image quality [9, 10]. Image registration, a sort of digital motion correction, helps to eliminate motion artifact's caused by operator tremors and/or patient movements. These image registration techniques, however, are unable to repair large-scale motion distortions that result in data loss owing to misalignment, necessitating the use of steady-handed, highly qualified operators [11]. The most effective way to overcome the issues is to design. An eye-tracking system to track eye movements and generate input to the OCT scanner for motion compensation. This feature would make the imaging system more robust and user-friendly.

Numerous real-time eye-tracking systems are available, typically electrooculography, contact lens techniques, limbus tracker, dual Purkinje, and video-based pupil tracking. However, video-based tracking is more advantageous over other methods due to its high resolution, flexibility, precision, and accuracy tracking techniques [12]. As cost-effective alternatives, these systems are employed with linguistic tools like OpenCV and Visual C [13]. The quality of the camera and the speed of the programme rely on software performance time [10, 11].

In this work, the development of a pupil-tracking system using a depth camera is described. The pupil tracking was performed using two algorithms. The outcomes are assessed and compared to identify the better performing algorithm.

2 Methodology

Eye movements are unavoidable during OCT imaging, especially in live subjects. Eye tracking allows real-time monitoring of these movements and helps correct motion artifacts, ensuring high-quality images even in the presence of patient movements. With eye tracking, OCT systems can work more efficiently. The system can automatically follow the eye's movements, reducing the need for manual adjustments and potentially speeding up the imaging process. In order to track the eye movement, it is important to use high resolution camera to detect very small changes. An eye tracker continuously records eye movements, which will be given as the input to the scanning system of the OCT system in order to align with the eye precisely. The tracking system involves image acquisition using a depth camera and the eye is isolated from the detected face. The acquired eye images are pre-processed to remove morphological noises caused by light reflection in the iris. The iris is then detected using threshold and landmark based algorithm. The detected iris region is then tracked to estimate gaze as shown in Fig. 1.

2.1 Image Acquisition

The eye is imaged using the Intel Real Sense Depth Camera D435i. This camera consists of a cutting-edge stereo-edge system (stereo depth baseline 1080p RGB), equipped with an inertial measurement unit (IMU) for detecting rotations and motions. The specification of the camera is presented in Table 1.

Table 1 Specification of Real Sense Camera D435i.

Pixel	1280 × 800
Resolution	720 p
Frame Rate	30 fps
(L × B × H)	36.5 × 19.4 × 10.5 mm
Focal Length	1.93 mm
Interface	USB

The required software packages and wrappers, namely the Intel RealSense SDK (Software Development Kit) from Github and Python wrappers were downloaded from the Python Platform. Real-time data acquisition and image processing were performed in Python.

The pupil is the dark circular aperture in the center of the eye that controls the amount of light entering the eye. By analyzing the position and movement of the pupils, eye-tracking systems can infer valuable information about a person's gaze direction. In this work, two eye-tracking methods are adopted, one is threshold based and the other is landmark-based. The subject is instructed to sit at approximately 7 cm (the minimum distance for the depth camera being utilized) from the camera, and both algorithms are used to assess eye movement in various directions.

In order to verify both the algorithm's performance, the pupil movement in each frame is measured. The number of pixel needs to be mapped with distance in mm. In order to estimate what a pixel represents in mm, a graph sheet is used. The graph sheet is placed at 7 cm from the camera and imaged. The number of pixels occupying one column (1 cm width) is estimated using Euclidean distance and is found to be 70 pixels as shown in Fig. 2 and each pixel represents 0.148 mm.

All the images were assessed manually to locate the center and movement of the eye in a different direction, this was used as ground truth to analysis the accuracy of the algorithm.

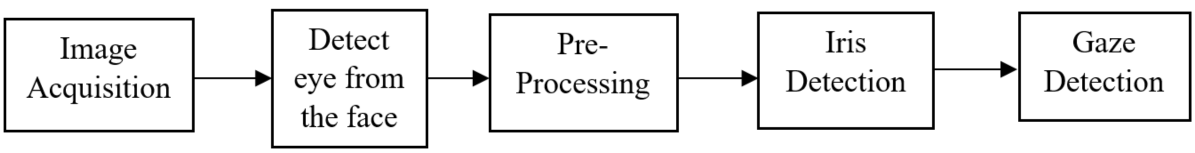


Fig. 1 Overview of gaze tracking algorithm.

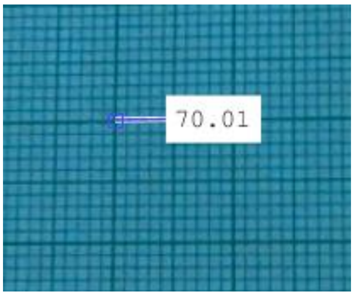


Fig. 2 Pixel to distance calculation using graph sheet.

2.2 Threshold-Based Eye Tracking

In threshold based eye tracking method, real-time RGB images are acquired using RealSense Depth D435i Camera. The face is detected using Haar features [9]. Haar features are a key component in the Viola-Jones face detection algorithm, which is a widely used and efficient method for detecting faces in images. The Viola-Jones algorithm is based on the principle of cascaded classifiers and uses Haar-like features to represent regions of an image and distinguish between facial and non-facial regions [14]. The threshold-based eye tracking algorithm flowchart is shown in Fig. 3.

A rectangular bounding box is marked surrounding the face region and eye region as shown in Fig. 4(a). Bounding boxes are used for region localization and isolation of the entire the region of interest (ROI) from an image as shown in Fig. 4(b). Once the bounding box is placed around the eye region, the pupil is tracked. The RGB is converted to a grayscale image as shown in Fig. 4(c). The grayscale is then converted to a binary image by thresholding (a threshold value of 70 is used) as shown in Fig. 4(d). All the pixel values greater or less than the threshold value are assigned to the maximum or minimum value respectively. The thresholding step clearly separates the ROI, i.e differentiating the sclera from the iris.

Once the pupils are detected, their positions are localized within the image. This allows the system to determine the precise position of the pupil. To track the pupils over time, the system establishes associations between the detected pupils in consecutive frames. After the pupils are continuously tracked, the eye-tracking system analyses the movement patterns and infers the person's gaze direction.

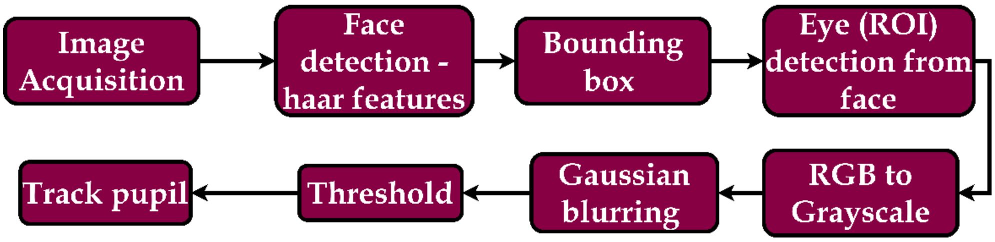


Fig. 3 Threshold based eye tracking algorithm flowchart.

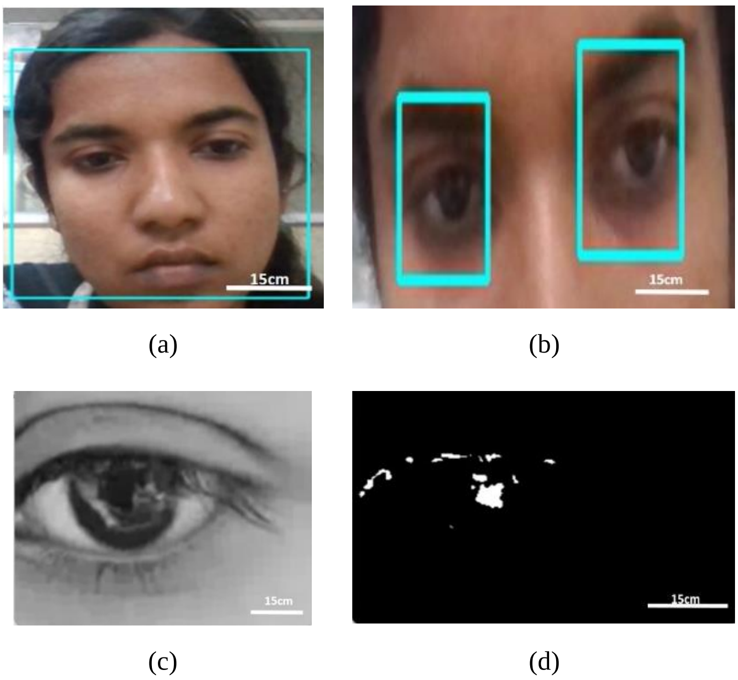


Fig. 4 Threshold based eye tracking (a) face detection using Haar features, (b) bounding box for eye detection, (c) RGB to gray scale image, and (d) thresholding for iris detection.

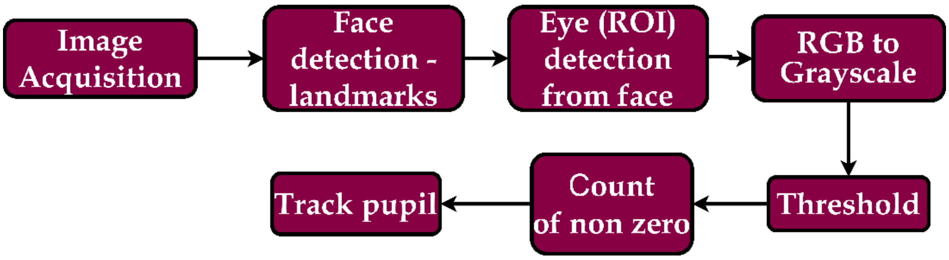


Fig. 5 Landmark based eye tracking algorithm flowchart.

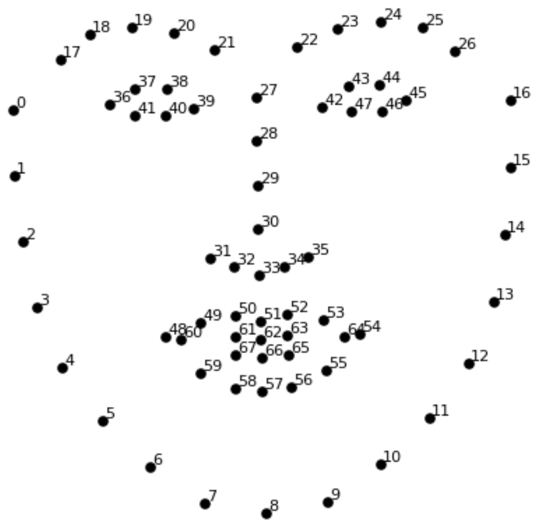


Fig. 6 Facial landmarks with 68 points from the machine learning library Dlib.

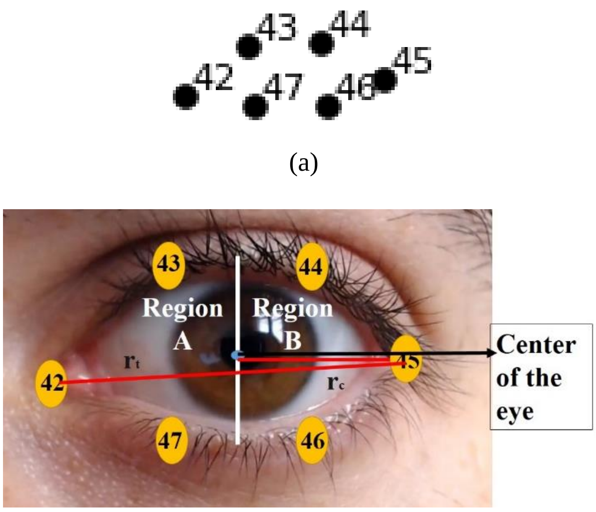
2.3 Landmark-Based Eye Tracking

In Landmark based method, a mesh model is used to identify different features in the face. The data is acquired from the Real Sense Depth D435i Camera. The video data is converted to frames, the RGB frames are converted to Grayscale. The overall algorithm flowchart is shown in Fig. 5.

The machine learning library Dlib [15] based 68 landmark mesh model is used in this work to identify eye as shown in Fig. 6. This model processes the grayscale frame and returns 68 landmarks for each detected face. These landmarks are used to ensure accuracy irrespective of the age of the patient and the size of the face. The ROI (region of interest) is extracted by selecting the landmarks which denote the eye region ((right eye (36–41), left eye (42–47))).

This method measures angular eye position relative to the face movement using facial landmarks and measures eye position relative to these landmarks. It creates a mesh of the approximate eye region and separates the eye region from the original frames by using media Dlib. In order to identify the iris, the ROI region is thresholded. A threshold of 70 was chosen after multiple trials. The grayscale images are converted into binary images based on the chosen threshold.

For processing only one eye is taken (left eye) as shown in Fig. 7(a). Eye landmarks 43 and 44 which are in the upper region, and 46 and 47 from the bottom region are averaged and a vertical line is drawn to form the reference center point. The line divides the eye region as left and right region as shown in Fig. 7(b). To localize the iris position, the non-zero pixels are counted in each region. When the eye looks to the right the non-zero value is higher on the left side and vice versa. This help in detecting the direction in which the eye has moved.



(a)

(b)

Fig. 7 (a) Right eye landmarks, (b) eye movement detection by dividing eye into regions and calculating non-zero value in both regions.

An iris position ratio is computed in this study to precisely locate the iris. As a starting point, the eye's right landmark is used. The distance of the center of the eye from this reference point is measured and denoted as r_c as shown in Fig. 7(b). The r_t denotes the measurement of the distance between the left and right landmarks of the eye. In order to determine whether the iris is close to or far from the right landmark of the eye, i.e. to estimate if the iris is left, center, or right, the ratio of the two measured distances is used. The iris position ratio k value is represented in Eq. (1). Three categories comprise the calculated ratio value: non-zero values in $A < B$ correspond to the left, $A = B$ to the center, and $A > B$ to the right. These values are derived based on measurements of various eye images.

$$k = \frac{r_c}{r_t}. \tag{1}$$

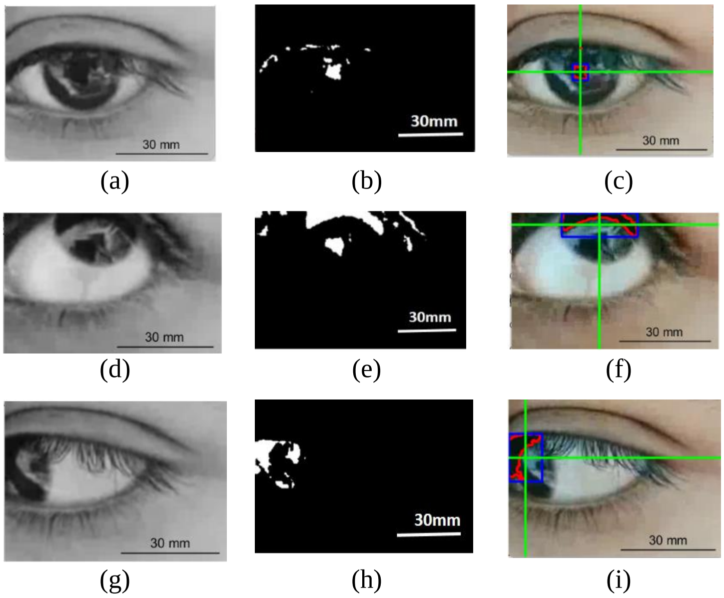


Fig. 8 Threshold-based eye tracking (a, d, g) extracted eye regions showing eyes in different positions; (b, e, h) extracted eye image is thresholded to identify the direction of gaze from (a), (d), and (g), respectively; (c, f, i) pupil is detected and gaze direction is determined to see if the eye is the center, up and right direction, respectively.

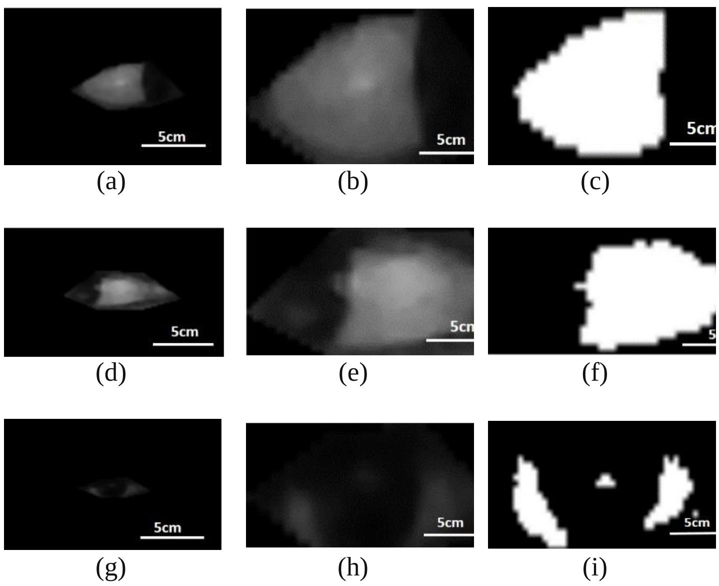


Fig. 9 Landmark-based real-time eye tracking system (a, d, g) the eye region is extracted from the face; (b, e, h) eye region landmarks are used to exclude the eyelashes and extract only the eye region from the frame (a), (d), and (g) respectively; (c, f, i) the extracted eye region is contoured to identify the direction of the eye movement and the gaze direction is identified as right, left and center respectively.

3 Result and Discussions

Two eye-tracking systems were developed and the performance of both algorithms on gaze tracking were assessed. Both the eye tracking algorithm were tested on 10 frames. The images include different positions of the iris and the precise location and tracking of these movements using threshold based algorithm and landmark based algorithm are as shown in Fig. 8 and Fig. 9, respectively.

In the threshold-based method, once the eye region is separated from the face, thresholding is used to isolate the iris alone. The center of the iris is then estimated and shown as a marker in the original image for live tracking of the eye.

The tracking output of the landmark-based algorithm was tested in different eye directions and under different light conditions as shown in Fig. 9. The eye is first extracted from the face image as shown in Fig. 9(a). Using the landmarks of the eye, the lashes are eliminated and only the eye region is extracted as shown in Fig. 9(b). Based on thresholding, the iris and the surrounding sclera are identified and the total number of non-zero pixels to the left and right of the center of the iris is calculated in order to identify position of the eye A < B eyelooking left, A = B eye is at the center, A > B eye looking right) as shown in Fig. 9(c).

Table 2 Comparison of average deviation observed for threshold, landmark, and manual method in the estimation of center of iris.

Frames	Threshold	Landmark	Actual center
1	1.6	1.2	1.0
2	0.9	0.6	0.4
3	1.2	1.0	0.9
4	0.85	0.6	0.5
5	1.0	0.5	0.4
6	0.7	0.4	0.4
7	0.65	0.5	0.5
8	0.9	0.5	0.4
9	1.1	1.0	0.95
10	0.85	0.7	0.7

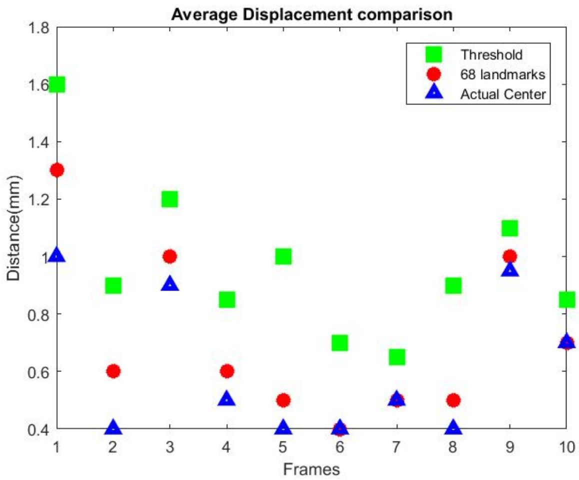


Fig. 10 Average deviation observed for threshold and landmark-based methods.

The individual frames are also assessed manually to identify the actual position of the iris by locating the center of the iris. The average deviation of these measured values is used as ground truth to assess the accuracy of the threshold and landmark based algorithms as shown in Table 2.

For landmark-based displacement, the average deviation was found to be in the range of approximately 0.2~0.6 whereas for the threshold-based system, the deviation was approximately 0.4~0.9. Hence, the iris localization and gaze direction estimation are more accurate in landmark-based system than threshold-based eye tracking systems. A graphical representation of the average deviation is shown in Fig. 10. It can be observed that the landmark-based system has a higher correlation with the actual center compared to threshold-based algorithm.

4 Conclusion

In order to automatically align the handheld imaging system, tracking of the eye is important. Two algorithms were developed and tested for iris detection and gaze tracking. The threshold-based algorithm uses a bounding box technique and the landmark-based algorithm uses 68 landmarks to precisely locate the iris. The movement of the iris is tracked to estimate the position of the eye. A total of 10 frames were used to analyze the efficiency of the algorithms. The average deviation from the manually calculated iris position is estimated for both algorithms and found to be 0.2~0.6 and 0.4~0.9 for threshold and landmark respectively. It is clearly observed that the landmark-based algorithm tracks the iris better. Future work on improving the facial landmark system and implementation of the same on a handheld OCT system is underway.

Acknowledgement

This work was supported by the SERB-funded Startup-grant project No: SRG/2020/002076.

Disclosures

The authors declare no conflict of interest.

References

1. V. Mazlin, P. Xiao, K. Irsch, J. Scholler, K. Groux, K. Grieve, M. Fink, and A. C. Boccara, "Optical phase modulation by natural eye movements: application to time-domain FF-OCT image retrieval," Biomedical Optics Express 13(2), 902–920 (2022).
2. J. A. Izatt, M. R. Hee, E. A. Swanson, C. P. Lin, D. Huang, J. S. Schuman, C. A. Puliafito, and J. G. Fujimoto, "Micrometer-Scale Resolution Imaging of the Anterior Eye In Vivo With Optical Coherence Tomography," Archives of Ophthalmology 112(12), 1584–1589 (1994).
3. E. A. Swanson, J. A. Izatt, M. R. Hee, D. Huang, C. P. Lin, J. S. Schuman, C. A. Puliafito, and J. G. Fujimoto, "In vivo retinal imaging by optical coherence tomography," Optics Letters 18(21), 1864–1866 (1993).

4. U. Schmidt-Erfurth, S. Klimscha, S. M. Waldstein, and H. Bogunović, "A view of the current and future role of optical coherence tomography in the management of age-related macular degeneration," Eye 31(1), 26–44 (2017).
5. W. Jung, J. Kim, M. Jeon, E. J. Chaney, C. N. Stewart, and S. A. Boppart, "Handheld Optical Coherence Tomography Scanner for Primary Care Diagnostics," IEEE Transactions on Biomedical Engineering 58(3), 741–744 (2011).
6. N. D. Shemonski, F. A. South, Y.-Z. Liu, S. G. Adie, P. S. Carney, and S. A. Boppart, "Computational high-resolution optical imaging of the living human retina," Nature Photonics 9(7), 440–443 (2015).
7. P. J. Rosenfeld, M. K. Durbin, L. Roisman, F. Zheng, A. Miller, G. Robbins, K. B. Schaal, and G. Gregori, "ZEISS AngioplexTM Spectral Domain Optical Coherence Tomography Angiography: Technical Aspects," in OCT Angiography in Retinal and Macular Diseases, F. Bandello, E. H. Souied, and G. Querques (Eds.), S. Karger AG, 56, 18–29 (2016).
8. H. Singh, J. Singh, "Human eye tracking and related issues: A review," International Journal of Scientific and Research Publications 2(9), 1–9 (2012).
9. 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," in IMTC/2002. Proceedings of the 19th IEEE Instrumentation and Measurement Technology Conference (IEEE Cat. No.00CH37276), 2, Anchorage, AK, USA, 1029–1033 (2002).
10. W. Wang, Y. Huang, and R. Zhang, "Driver gaze tracker using deformable template matching," In Proceedings of 2011 IEEE International Conference on Vehicular Electronics and Safety, Beijing, China, 244–247 (2011).
11. N. H. Cuong, H. T. Hoang, "Eye-gaze detection with a single WebCAM based on geometry features extraction," In 2010 11th International Conference on Control Automation Robotics & Vision, Singapore, 2507–2512 (2010).
12. S. Das, S. K. Swar, S. Laha, S. Mahindar, S. Halder, H. Koushik, and S. Deb, "Design approach of Eye Tracking and Mind Operated Motorized System," International Journal of Innovative Research in Science, Engineering and Technology (An ISO 3297: 2007 Certified Organization) 5(8), 14349–14357 (2016).
13. I. García, S. Bronte, L. M. Bergasa, N. Hernández, B. Delgado, and M. Sevillano, "Vision-based drowsiness detector for a realistic driving simulator," In 13th International IEEE Conference on Intelligent Transportation Systems, Funchal, Portugal, 887–894 (2010).
14. B. S. Kim, H. Lee, and W. Y. Kim, "Rapid eye detection method for non-glasses type 3D display on portable devices," IEEE Transactions on Consumer Electronics 56(4), 2498–2505 (2010).
15. D. E. King, "Dlib-ml: A machine learning toolkit," The Journal of Machine Learning Research 10, 1755–1758 (2009).