Gaze Estimation by 2D Geometrical Analysis of Low Resolution Images

SANAULLAH MEHRAN UJJAN*, DILEEPKUMAR SOOTHER*, NIZAM-UL-HAQ*, AND TARIQUE RAFIQUE MEMON*

* Department of Electronic Engineering, Mehran University of Engineering & Technology, Jamshoro, Pakistan

Authors E-Mail: (dileepkalani1994@gmail.com, mehran.ujjan@gmail.com,nizam.prince3@gmail.com, memon.tariq@gmail.com)

ABSTRACT

In this research, an eye gaze detection method is developed to realize a HCI (Human Computer Interface) through determining the position of the Iris of eye as an input. This investigation is conducted for utilizing the ordinarily available components such as a low resolution webcam which is built-in any electronic gadgets such as laptop, smart phone, and tablet avoiding the need to buy extra hardware to interact with the machine. A complete gaze detection mechanism needs three basic steps to be taken, detection of eye socket, location of iris along with its center point estimation, and a reference setting mechanism through which point of gaze could be identified. Multiple approaches are investigated to do three mentioned steps and develop a less computationally expensive and effective gaze detection mechanism. For computational simplicity, extracted eye image is considered a 2D (Two-Dimensional) plane and different ways of setting reference points for gaze point identification are explored. Solving issues along the way, research is concluded by developing a remote gaze detection method based on the 'Chehra' which is an open source facial landmark localization technique. The results depicted 'Chehra' based method as better than the other investigated methods. The approach yielded results with 83% accuracy.

Key Words: Human Computer Interface, Eye Gaze Detection, Gaze Estimation, Facial Landmark Localization.

1. INTRODUCTION

Gaze detection is the estimation of the human eye position or movement through predicting the different gaze actions of the eyeball. Previously, various researches have been conducted on the gaze detection which involves measuring the motion of an eye by putting electrode around the eyelids that measures the potential difference with the movement of the eyeball and tracking the eye motion using image processing techniques [1-6]. One of the most popular methods among all is gaze detection by image processing techniques using multiple ranges of the cameras for the detection of the gaze movement and interpolating them for the estimation of the gaze movement. The most recent research works use tracking of the eyeball using IR (Infra-Red) camera. As the IR camera works by transmitting IR light to the cornea of the eye by its light refection the location of the gaze is estimated [7-8].

Various sophisticated devices are available in the market such as Tobi Eye X and Gaze Pointer, etc. which are enabling humans to interact with the electronic gadgets such as LED TV, smart phone, and laptop which use the movement of the eyeball for the control of various electronic devices [6,9-11]. Eye tracking devices mostly employ bright pupil dark pupil techniques and active IR illuminations [12-13] in them to address darker inside of the eye socket, eye lid problem and get more precise estimation. Contrary to these approaches, in this research use of IR cameras is avoided owing to their high cost. Also IR cameras are not readily available with machines so user is forced to add some extra hardware with its machine. A typical webcam usually attached to the machine has very low resolution so out method should be able to work low resolution images.

Ince and Kim [14] have developed two separate algorithms, first one does pupil center localization based on DAISMI, where as a deformable template based 2D gaze estimation technique is established to obtain proper eye movement detection. However, cursor control experiment conduct over a virtual keyboard shows that developed mechanism is highly unreliable for resolution sensitive tasks. Comparatively, in this investigation we employ a computationally simple approach even though it has less degrees of freedom and present a simple HCI mechanism which could be used to have a joystick type control with a machine (up, down, right, left).

This research proposes a simple less computationally expensive geometrical computation based algorithm for detecting the eye movements by configuring an open source algorithm that is "Chehra" of the MATLAB software. For this, the points for matching the co-ordinates of the iris are fixed for determining the direction of the gaze. This algorithm keeps tracking the location of the iris for implementing this algorithm in real time; the computation of it is kept as simple as possible. The remainder of this divided as follows: Section 2 discusses the tools and techniques used for the eye gaze detection system; Section 3 discusses the methods used in this research. Section 4 reports and discusses the results, and Section 5 concludes the paper.

2. TOOLS AND TECHNIQUES

The system is developed using a webcam and conducting a geometrical analysis of the low resolution images captured. A simple webcam of 640*480 pixels resolution is employed in this research specification are given in Table 1. The software used for this investigation is MATLAB and its different toolboxes such as image acquisition toolbox, image processing tool boxes, and computer vision toolbox [15]. Multiple approaches for extracting the eye image, locating the iris, and gaze estimation are investigated.

TABLE 1. SPECIFICATIONS OF THE WEBCAM

Specifications	Details
Type	PK-753F
Connectivity	USB-2.0
Sensor	1/6-CMOS of 640*480 pixels
FPS	30
Focal lengths of lens	min=2.4, max=3.5
Angle of observation	60 degrees

3. METHODOLOGY

In this research, three different gaze detection techniques are employed and compared in order to select best possible approach. These techniques are employed to perform 2D image analysis. Fig. 1 depicts the operation of the developed setup. Task of gaze detection is possible by going through these steps:

(i) Detection of Eye socket in a low resolution image
(ii) Localization of Iris
(iii) Gaze detection mechanism

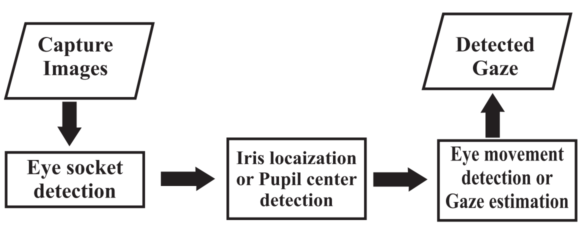

FIG. 1. FLOW DIAGRAM OF THE DEVELOPED SYTEM

Various algorithms to perform these steps are investigated. Each technique is discussed as follows.

3.1 Hough Transform and Viola Jones Based Approach

For the detection of Eye socket first Viola Jones algorithm is used. It is an established and one of the most used techniques for face feature detections. And it is done by going through four important steps, Firstly Haar features are selected and based on that an integral image is created, and finally through Ad-boost training facial features are determined and classified. Haar features works on the basic assumption that typically a face is constituted of similar features in which their orientation with respect to each other remains same for e.g. Eye, Nose, Lips and Hair. Eyes area is a bit darker, nose looks like a bridge and Cheeks are a bit brighter comparatively. These attributes help distinguish different parts of the face.

For iris localization Hough transform is employed; considering it as a circle to find center co-ordinates of it, and radius [16]. Hough transform helps detect varying shapes in images for e.g. ellipses, circles and lines of varying slope and curve.

Hough transform utilizes different ways to detect varying shapes in an image, for circle detection it carries out different steps. An accumulator space is first created, and then every pixel in the image is considered a center point of circle of a specified radius. Then all points that satisfy the Equation (1) are detected.

$$(x-f)^2 - (y-g)^2 = r \qquad (1)$$

Every combination of x and y that satisfy the equation of circle of center points (f,g) and radius r is detected as a circle. This method has the ability to detect a circular shape whose circumference or the set of (x,y) may lie outside of the accumulator space. Hence Eye lid problem does not necessarily effects the detection of Iris by this method. Fig. 2 shows eye socket image detected with Viola Jones algorithm with extracted eye socket in top left corner, whereas Fig. 3 shows iris localization by using Hough transform while looking at up, down, right and left positions.

FIG. 2. DETECTED EYE SOCKET VIA VIOLA JONES

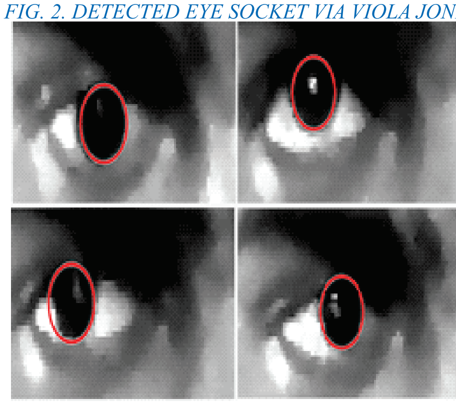

FIG. 3. IRIS LOCALIZATION USING HOUGH TRANSFORM

As center points are detected through Hough transform which is used for real time localization of Iris. Hard coding and setting references points for gaze detection mechanism would be unreliable. For choosing reliable reference points we introduced a training phase, through which user was asked to look at white box appearing at different locations on laptop screen for 3 seconds. Fig. 4 shows the different positions at which white box appeared and user was asked to look at.

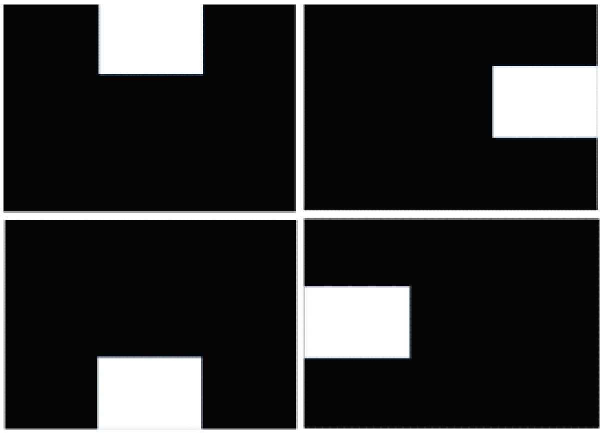

FIG. 4. WHITE BOX APPEARING AT DIFFRENT LOCATIONS

All the center values of iris estimated through Hough transform were captured and stored during 3 seconds. Then reference points are generated by taking a mean of these values. The flow diagram of the testing phase. Fig. 5 shows the set reference points. Equation (2) shows how reference points are calculated. Where, m represents x,y co-ordinates of iris center, and subscript n represents number of instances of iris center captured during the span of 3 seconds for each position.

$$M_1 = \left(\frac{m_1 + m_2 + \ldots\ldots + m_n}{n} \right) \quad (2)$$

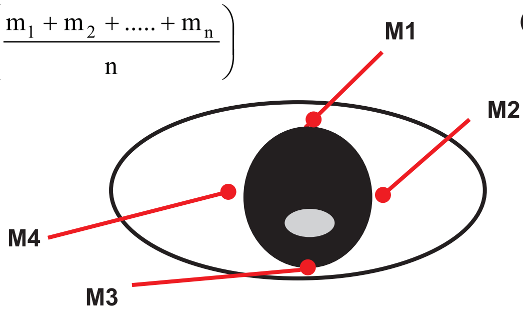

FIG. 5. REFERENCE POINTS ON EYE IMAGE

Having set reference points, point of gaze was identified using mean distance formula which is given in Equation (3).Where MnX and MnY represent x,y values of reference points while measured x and measured y are iris center estimation during real time.

$$d = \sqrt{\left(\left(MnX - \text{measured } x\right)^2\right) + \left[\left(\left(MnY - \text{measured } y\right)\right)\right]^2} \quad (3)$$

The condition of gaze identification as depicted in Fig. 6(a) can be stated such as any Mn for which d is minimum in the direction of gaze". Whereas, Fig. 6(b) depicts the flow chart of training process.

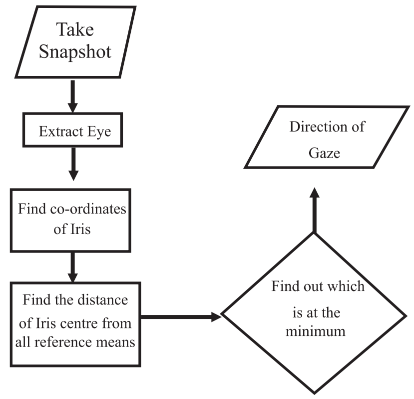

FIG. 6(a).TESTING PROCESS

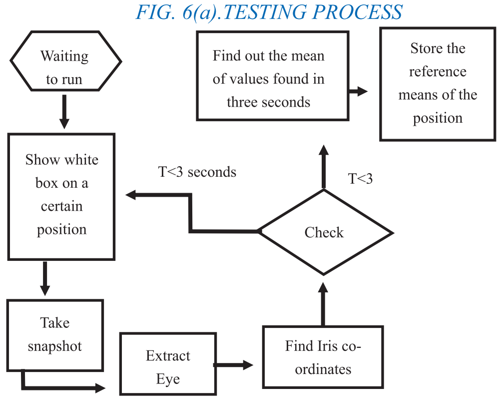

FIG. 6(b). TRAINING PROCESS

In this approach, eye image was extracted from full image. Then to locate the center of iris. Hough transform was used. For the whole gaze detection mechanism a training phase and mean distance formula are used to set the reference points and gaze detection respectively. The method is unable to produce acceptable results due to scaling issues of captured eye image and susceptibility to the noise of Haar like features.

3.3 Head Mounted Setup Using Hough Transform

This approach is employed to counter the scaling issues offered by Viola Jones algorithm (Haar like features). Most of the times captured eye socket varied in its size therefore hindering to locate proper reference points. Hence to reduce the reliability on Viola Jones a head mounted setup was developed. Camera was mounted on a frame of goggles to capture direct image of eye. A train and test phase were also used to counter the noise and increase the reliability. The setup is shown in Fig. 7.

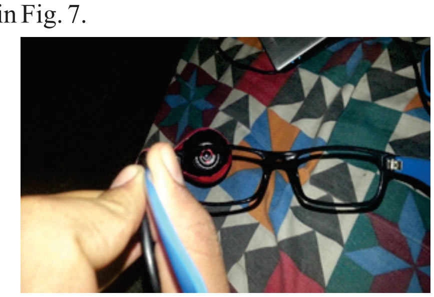

FIG. 7. WEBCAM MOUNTED ON THE FRAME OF GLASS

In this approach, process of extracting the eye image was bypassed by using a head mounted setup. Hough transform is used to track the iris. The training phase and the flow chart of this method used would be similar to the first approach. It has only difference of the snapshots of eye which are directly taken, now it does not require extracting eye from the image. This approach performs better in terms of accuracy but due to the need of wearing a hardware setup on eye this was deemed as an intrusive approach.

3.4 'Chehra' Based Approach

The 'Chehra' is a regression based model to localize facial features in an image, and is made open access to everyone for further experimentation. This algorithm was developed by Asthana et. al. [17] at the Imperial College London. Through incrementally updated the generic model Chehra seems to provide a good concentration on accuracy for real time localization of facial features. It has outperformed various facial feature detection methods.

The 'Chehra' algorithm is more accurate than Viola Jones in localizing the facial landmarks hence increasing the accuracy of captured eye image. Method of gaze computation is also changed. This time eye image is divided into 6 segments. This method does not require testing and training process and it is also non-intrusive because it is able to remotely detect gaze. The operation of the "Chehra" for eye gaze detection is depicted in Fig. 8.

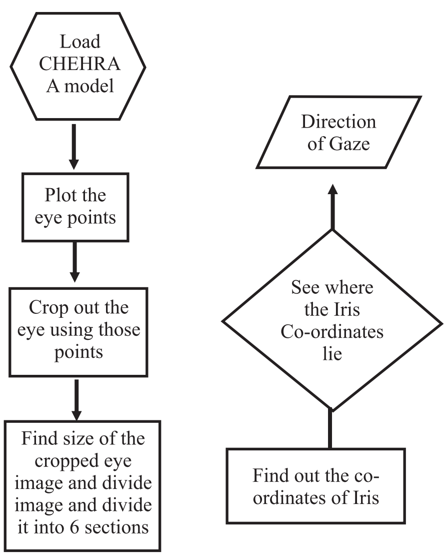

FIG. 8. FLOW DIAGRAM OF THE "CHEHRA" ALGORITHM FOR EYE GAZE DETECTION

In this method, camera is connected to the laptop so that processing can be done using "Chehra" algorithm in MATLAB. This countered the scaling and rotational problem and this solved the problem of extracting the eye from the video recorded by the camera through Viola Jones algorithm.

Some modifications are made in the open source code and the system is capable of detecting eye points out of it. As the point of interest of this research is the effective localization of eye boundaries, that is why "Chehra" algorithm is employed to detect eye movements through image. In this research, a distant method of detection is made possible, besides some changes are also made in the computational method.

Eye socket captured through 'Chehra' is less prone to scaling and rotational issues. Through the algorithm, the cropped eye image was divided into six sections and the center of the eye was sorted out to be the point at which the person is gazing on as shown as Fig. 9. After cropping out the eye, size of the image is found. Consider, if x-axis is the horizontal dimension of eye image and y-axis is the vertical dimension. The x-axis is divided into three parts, in which center section is comparatively shorter than the adjacent sections, and y-axis is divided into equal half sections.

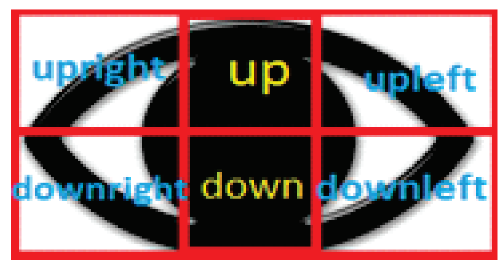

FIG. 9. SIX REGIONS OF EYE

After cropping out the eye, x and y components of the image were divided into three sections, respectively.

4. RESULTS AND DISCUSSION

The Viola-Jones based approach did not yielded desirable results owing to scaling issues of captured eye image and susceptibility to the noise of Haar like features. That's why other methods are investigated to get desired results. Average testing accuracy of this approach was 50%. All the approaches were tested on three different people. Accuracy was calculated as in Equation (4).

$$\text{Accuarcy} = \frac{\text{Number of Correctly Identified Gaze Position}}{\text{Total Number of Trails}} \times 100 \quad (4)$$

The head mounted camera approach was used to overcome the limitations of first approach. However, it was intrusive. Average Testing accuracy of this model was 70% that is greater than first approach. To overcome limitations of the aforementioned two approaches, 'Chehra' was employed. It did not need any extensive training and testing process. Moreover, it is a non-intrusive technique. The result of single eye socket detection is shown in Fig. 10(a-b).This result was achieved by localizing the eye boundaries. The results that are generated are scaled and rotational invariant. In the computational method of the algorithm, the cropped eye image was divided into six sections and the center of the eye was sorted out to be the point at which the person was gazing on as shown in Fig. 10(a-b).

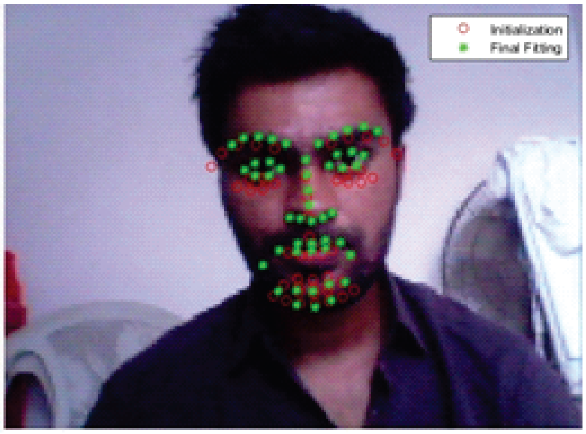

FIG. 10(a). FACIAL FEATURES DETECTED BY "CHEHRA

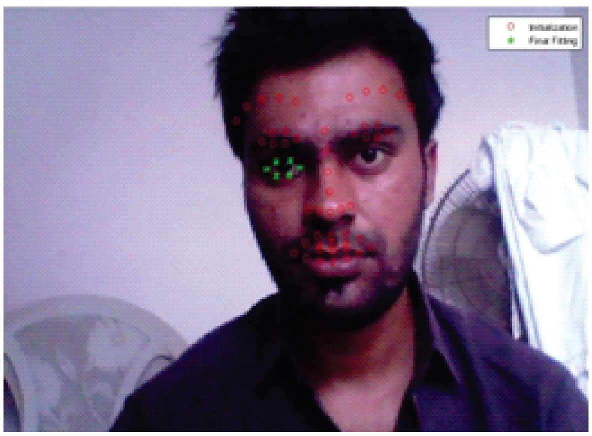

FIG. 10(b). EXTRACTED EYE POINTS FROM "CHEHRA"

In order to validate this algorithm for the Eye socket detection, few configuration steps were taken to get effective results. This algorithm was altered for operating in a single iris. Third approach can detect different eye movements can provide a foundation for HCI for various commercial devices available. It can function in almost all consumer computers irrespective of operating system (Windows/Linux/Android) and which possess a built-in camera. The developed system will open the doors to such technologies which do not need sophisticated and overpriced devices to interact with the machine.

The 'Chehra' based approach efficiently detected the six movements of eye with the accuracy of 83% without any training process and head mounted camera. However, some hindrances were faced while detecting the movements so therefore a condition was set to examine the direction of the gaze three times based on the reference values as shown in the above table. What makes this proposed algorithm different from the previous "Chehra" algorithm is its precision. Previous algorithm had lesser accuracy rate, while this one has 83%.

5. CONCLUSION

In recent decades, various researches have been conducted related to image processing. In this paper, an eye gaze detection system was presented with minimal and affordable resources such as webcam and "Chehra" algorithm. The "Chehra" algorithm was employed to analyze the images and to detect the particular 6 movements of eye. The system was able to remotely detect the eye movements with greater accuracy than the other investigated methods. The results revealed increase in the accuracy that is 83%.In addition, this method is non-intrusive as it did not require any extensive training setup. The method results have confirmed the effectiveness of the method detecting the eye gaze.

6. FUTURE WORK

This paper suggest for future work, to develop a standalone system could be developed which assist the paralyzed patients. In future, the degrees of freedom can be increased which can serve as even better means of HCI.

ACKNOWLEDGEMENT

Authors of this paper would like to acknowledge Engr. Khuhed Memon, Department of Electronic Engineering, Mehran University of Engineering & Technology, Jamshoro, Pakistan, for supporting their work and guiding the group members during development of this project.

REFERENCES

[1] Kar, A., and Corcoran, P.J., "A Review and Analysis of Eye-Gaze Estimation Systems, Algorithms and Performance Evaluation Methods in Consumer Platforms", IEEE Access, [DOI: 10.1109/ACCESS.2017.2735633] Volume 5, pp. 16495-16519, 2017.

[2] Tonsen, M., Steil, J., Sugano, Y., " Invisible Eye: Mobile Eye Tracking Using Multiple Low-Resolution Cameras and Learning-Based Gaze Estimation", Proceedings of ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, Volume 1, pp. 1-21, 2017.

[3] Lu, F., Sugano, Y., Okabe, T., and Sato, Y., "Adaptive Linear Regression for Appearance-Based Gaze Estimation",IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 36, No. 10,October, 2014.

[4] Zhu, D., Mendis, B.S.U., Gedeon, T., Asthana, A., and Goecke, R., "A Hybrid Fuzzy Approach for Human Eye Gaze Pattern Recognition", International Conference on Neural Information Processing, pp. 655-662, Springer, 2008.

[5] Zhang, X., Sugano, Y., Fritz, M., and Bulling, A., "It's Written All Over Your Face: Full-Face Appearance-Based Gaze Estimation", Proceedings of IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 51-60, 2017.

[6] Hansen, D.W., and Ji, Q., "In the Eye of the Beholder: A Survey of Models for Eyes and Gaze",IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 32, pp. 478-500, 2010.

[7] Zhu, Z., Ji, Q., Fujimura, F., and Lee, K., "Combining Kalman Filtering and Mean Shift for Real Time Eye Tracking Under Active IR Illumination", Proceedings of IEEE Conference on

Object Recognition Supported by User Interaction for Service Robots, Volume 4, pp. 318-321, 2002.

[8] Timm, F., and Barth, E., "Accurate Eye Centre Localisation by Means of Gradients", Proceedings of 6th International Conference on Computer Vision Theory and Applications, Vilamoura, Algarve, Portugal, 5-7 March, 2011.

[9] Jyoti, S., and Dhall, A., "Automatic Eye Gaze Estimation Using Geometric & Texture-Based Networks", IEEE 24th International Conference on Pattern Recognition, pp. 2474-2479, 2018.

[10] Xia, Y., Lou, J., Dong, J., Qi, L., Li, G., and Yu, H., "Hybrid Regression and Isophote Curvature for Accurateeye Center Localization", Multimedia Tools and Applications, Volume 79, pp.805–824, 2020.

[11] Kim, J., Stengel, M., Majercik, A., DeMello, S., Dunn, D., Laine, S., McGuire, M., and Luebke, D.,"An Anatomically-Informed Dataset, for Low-Latency, Near-Eye Gaze Estimation", Proceedings of Conference on Human Factors in Computing Systems Proceedings, pp.1-12, 2019.

[12] Youmaran, R., and Adler, A., "Using Infrared Illumination to Improve Eye andFace Tracking in Low Quality Video Images", IEEE 24th Biennial Symposium on Communications, pp. 391-394, 2008.

[13] Zhu, Z., and Ji, Q.,"Robust Real-Time Eye Detection and Tracking Under Variable Lighting Conditions and Various Face Orientations", Computer Vision and Image Understanding, Volume 98, pp. 124-154, 2005.

[14] Ince, I.F., and Kim, J.W., "A 2D Eye Gaze Estimation System with Low-Resolution Webcam Images", Journal on Advances in Signal Processing, Volume 1, August, 2011.

[15] www.mathworks.com (Last Accessed: 16th May, 2020).

[16] Lupu, O.G., and Ungureanu, F., "A Survey of Eye Tracking Methods and Applications", E-Health and Bioengineering Conference, Volume 3, pp. 72-86, 2013.

[17] Asthana, A., Zafeiriou, S., Cheng, S., and Pantic, M., "Incremental Face Alignment in the Wild", IEEE Proceedings of Conference on Computer Vision and Pattern Recognition, pp. 1859-1866, 2014