Proc. 1st International Conference on
Machine Learning and Data Engineering (iCMLDE2017)
20-22 Nov 2017, Sydney, Australia
ISBN: 978-0-6480147-3-7

Visual Attention Detection: What is the target person looking at?

Partha Pratim Debnath[1], Md. Golam Rashed[2] and Dipankar Das[3]

[1]Lecturer, Bangladesh Army University of Engineering and Technology, Natore, Bangladesh
[2]Assistant Professor, University of Rajshahi, Rajshahi, Bangladesh
[3]Professor, University of Rajshahi, Rajshahi, Bangladesh
parthapratim.ice10@gmail.com, {golamrashed, dipankar} @ ru.ac.bd

Abstract
In this paper, we have described an approach to detect the visual focus of attention (VFOA) of a person from his/her gaze direction with varying lighting conditions and varying distance from the camera and evaluate their performance. The continuous video of the target person is captured and fed into an expert system for further processing. From frame by frame analysis, the head and eyeball of the target person is detected using vector field of image gradient. If the person changes his/her gaze, the corresponding coordinate of pupil also changes. The frontal view of the person is divided into three regions corresponding to three target objects and the gaze direction are detected based on in which region the coordinate of pupil is located in the eye area. Our technique also combines the both Rapid Eye Movement (REM) and head rotation detection, which provides an efficient tool for VFOA tracking. With a low cost camera, it is very hard to work at night and cloudy days because of lack of proper lighting. So considering these limitations, we have developed a system that can successfully track visual focus of attention of the target person by means of sustained and transient attention or distraction and finally control the attention by deploying an external signaling or alarming system.

Keywords: *Visual focus of attention, eye center localization, gaze detection, sustained attention.*

1. INTRODUCTION

Detection and controlling of human's visual focus of attention (VFOA) is a part of active research for many years [1]. However, the researchers of computer vision and artificial intelligence seek for an accurate method for fixation tracking based on object [2] detection by which human attention level may be realized and diversified attention may be controlled. Basically the term "Fixation" refers to a process of maintaining the visual gaze at a single location. Visual fixation is never perfectly steady, fixational eye movement occurs involuntarily, especially for the case of short-term stimuli/attention. There are many real life situations those need the tracking of VFOA of the target person to avoid the critical circumstances. For example, if we can detect the attention/ distraction of the driver in the vehicle [3], we can develop an automated alarming system to alert the driver and bypass fatal accidents. The results of VFOA detection are also an important tool to generate attention patterns of the particular driver and compare it with the others to select the best ones. To accomplish it, most of the vision researchers used wearable sensor based strategies [4]. However, most of the wearable sensors based strategies have a number of limitations for application in detecting and controlling human's VFOA in real world environments. For example, in the context of public social spaces, people may fix their VFOA at particular location based on their own interests and may not be interested in actively engaging with the technology [5]. Thus, in our proposed technique, we are interested in making use of human's VFOA tracking for a wearable-free solution where human do not need to attach markers to themselves or carry special devices so that we may observe them in an unrestricted manner. Here, in this work, we have used low cost USB camera in front of the target person that captures continuous video frames. Vision researchers formerly used solely head tracking [6] or eyeball tracking [7] [8] techniques independently to meet our objectives.

However, in our approach, we have used head tracking and eyeball-tracking techniques in combination for proper detection of visual focus comprising of transients depends on the tracking of the frequency and time of the head rotation as well as rapid eye movement (REM) realization. The accuracy of our technique is highly sensitive to the choice of proper illumination and right distance from camera-especially for the case of low cost camera deployment. In our present study, we conducted experiments in controlled environment and results show that the proposed approach can yield good performance of detecting visual attention of a target human in general situations that offers transient/sustained attention/ destruction and offers optimum cost efficiency.

2. METHODOLOGY

To detect the visual focus of attention of the target person, we need a camera installed at a suitable position so that image can be captured very easily. These continuous images are feed into our proposed system for frame by frame further processing. The overall process is described as follows-

2.1 Head Pose Detection

The main goal of head pose detection is to track the head from the continuous image whether or not the head is in movement. In our work, we have used the Seeing Machine's faceAPI to detect and track the head pose, h_p of the target person. To detect the head, we have used haar cascade classifier.

2.2 Face Points Extraction by Active Shape Model

Our modeling method works by examining the statistics of the coordinates of the labeled points in the head rectangle. In order to be able to compare equivalent points from different shapes, they must be aligned with respect to a set of axes. We achieve the required alignment by scaling, rotating and translating the shape so that they correspond as closely as possible. In this technique, we aim to minimize a weighted sum of squares of distances between equivalent points on different shapes.

2.3 Iris Center Detection

A multistage approach [9] [10]] has been used to detect the iris. A 3-D head tracker [11] detects the head position and the rectangular area in the image. Then the facial feature points are extracted from the active shape model [12]. These points are used to roughly detect the eye regions from the face. The vector field of image gradient (VFIG) is used to detect the iris center.

The VFIG iris center detection technique is described as follows-

Let I_c be the possible iris center and I_{gi} be the gradient vector in position I_{xi}. If I_{di} is the normalized displacement vector, then it should have some absolute orientation as the gradient I_{gi}. We can determine the optical center I_c* of the iris (darkest position of the eye) by computing the dot products of I_{di} and I_{gi} and finding the global maximum of the dot product over the eye image:

$$I_c^* = \text{argmax}\, I_c \{\frac{1}{N}\sum_{i=1}^{N}(P\ )\}$$

Where, $P = (I_{di}{}^T I_{gi})^2$ and $I_{di} = (I_{xi} - I_c) / (\| I_{xi} - I_c \|_2)$

i = 1, 2,..., N and the displacement vector I_{di} is scaled to unit length in order to obtain an equal weight for all pixel position in the image.

2.4 VFOA Detection

After the successful detection of the head pose and pupil of the eye, the attention/ distraction of the target person is tracked with the help of head and eye rectangle. A subdivision method proposed in the paper [13], is used here. To achieve the optimum efficiency, we have used the "Unequal Partitioning of eye rectangle leaving the uncovered regions" method to track the movement of the eyeball within the eye area. "Partitioning of the head rectangle" technique is also used to detect the head movement.

Now the speed of the eyeball movement and head movement provides us a clue to detect the short or long term attention/ distraction of the target person. We classify the inattention of the target person in two major criteria-

a) Transient

These short term stimuli occur in the subconscious mind. The target person is not aware of it. It happens unwillingly to provide some rest for the brain.

b) Sustained

This occupies a long time [14]. In our experiments we have set it around 8 seconds. The target person offers this type of inattention willingly.

To detect the transient and sustained attention of the target person efficiently in different environments, two different parameters- right distance from the camera and proper lightening needs to be ensured. It is very important to measure the right distance from the camera. The participant should sit within this range of distances so that the images can be efficiently captured by the camera. Moreover, any deviation from this range of distances can be used as a clue to detect the proper inattention (such as drowsiness). For example, video analysis shows that prior to any road accidents, the driver tends to bow down his head towards the steering of the car. Apparently, the driver comes out of the threshold value of the right distance. To track the threshold value of right distance, we have used the trial and error strategy. The right distance is selected based on the highest efficiency of head and eye center detection. Any deviation from the threshold efficiency provides a clue to inattention.

Proper illumination of eye tracking is also a very important factor of gaze detection. The reason behind this fact is that, to trace the eye pupil, our system tries to find out the darkest point within the face rectangle using the vector field of image gradient. In that case, for the different illumination, the efficiency of head and eye center detection varies. Generally, the efficiency increases with the increase of illumination. But we know that the radius of eyeball decreases with the increase of light intensity after a certain value so that minimum amount of light can enter through the eyeball. So, at the higher illumination level the efficiency of eye center tracking also decreases. The threshold value of proper illumination is selected based on the highest efficiency of head and eye center detection

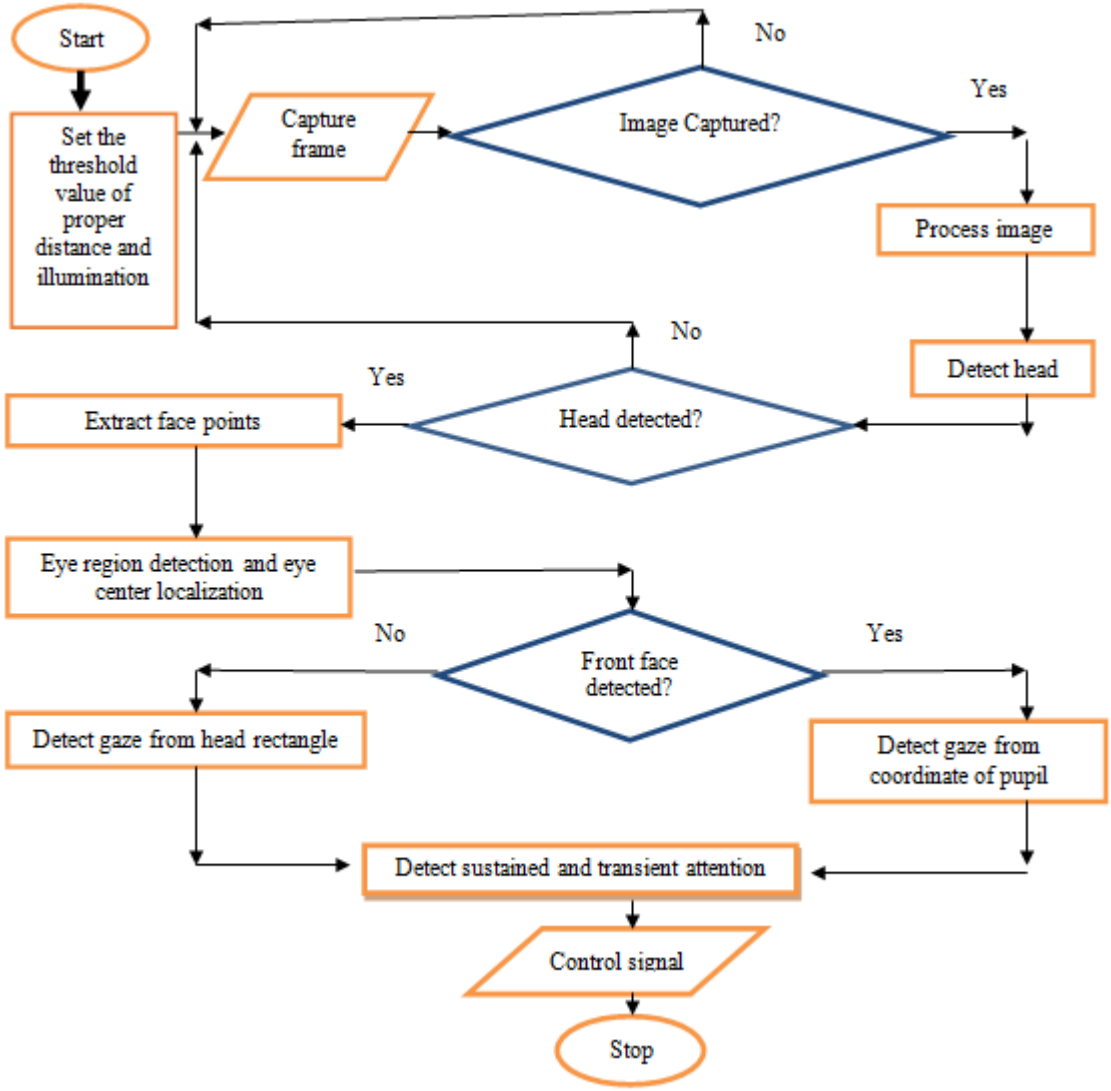

Figure1. Flow chart of the proposed system

3. EXPERIMENTS

3.1 Experimental Setup

We conducted separate experiments applying different techniques under different lighting conditions and distances from camera under a controlled environment. Data were taken using different participants. There were total 3 male nonpaying participants with the age 25, 24 and 22 years respectively. For real time frame analysis, we used Microsoft Visual Studio 2010 and OpenCV 2.4.9. The camera specification is given in the following Table 1.

Table 1: Camera specification

Model	Logitech c170
Photo quality	5 MP
Video quality	HD 720p
Focus type	40 cm and beyond
Auto light correction	Premium

3.2 Data Collection

Data were collection in successive steps which are discussed below-

a) Head and Eye Center Detection

In head and eye pupil tracking experiments, the participants were asked to seat at different distances from the camera under different lighting conditions to set the threshold value. They sat at 30cm, 40cm, 50cm, 60cm, 70cm, 80cm, 90cm and 100cm away from the camera. Different lighting conditions were provided by deploying different number of (1, 2, 3, and 4) 32 Watt energy bulbs. The provided illumination was 50, 100, 150, and 200 Lux respectively (measured by light sensor). The area of our room was $4m^2$. The average video length was 2 minutes.

b) Visual Focus Detection

For gaze detection through applying different techniques, these participants were asked to seat 70 cm away from the camera and to look at different target objects. At first they just moved their eyeball to look at these objects without moving their head. And secondly, they rotated their head to look. These objects are situated 0.5m apart from each other and 1.5m from the target person. One of the three objects is located in the central field of view (CFV) that means just in front of the target person and the other two objects in the near peripheral field of view (NPFV) that means to the left and right side of the target person. The average video recording time was 2 minutes. The provided illumination was 200 Lux.

c) Sustained and Transient Focus of Attention Detection

In sustained and transient attention detection experiments, the participants were asked to look (just move their eyeball) at the left object and the right object for a very short duration so that they may be considered as transients. The duration of transients was also varied (1, 2, 3 seconds) during different experiments. Transients were also detected with the varying head rotation time (1, 2, 3 seconds). The average video recording time was 2 minutes. The provided illumination was 200 Lux. The threshold value for the right distance from the camera was set 70 cm. Figure 2 illustrates some sample frames that were captured during the data collection procedure.

Figure 2. VFOA detection of the target person

3.3 Performance Evaluation Matrices

The efficiency of VFOA tracking under different lightening conditions and different distances from the camera is evaluated based on the following formulae-

Accuracy =

$$\frac{\text{Total number of trials in which head and eye center is detectd}}{\text{Total number of trials at a particular distance from camera}} \times 100$$

%.....(1)

Under the threshold value of right distance from the camera and proper lighting, the efficiency of the transient detection is based on the following formulae-

Transient Detection Accuracy =

$$\frac{\text{Total number of transient attention detected}}{\text{Total number of transient attention occured}} \times 100\ \%.....(2)$$

And similarly the efficiency of head rotation detection is expressed as-

Head Rotation Detection Accuracy =

$$\frac{\text{Total number of head movement detected at a particular direction}}{\text{Total number of head movement occured at that particular direction}} \times 100\%.....(3)$$

4. EXPERIMENTAL RESULTS

Based on Equation 1, we derived the efficiency for head and eye center tracking which are plotted in Figure 3 and 4 respectively. The transient attention detection accuracy is derived based on Equation 2 and shown in Figure 5. Finally, we get the Figure 6 that illustrates the head rotation detection accuracy based on Equation 3. Figure 7 illustrates the transient attention pattern of the target person for typical situations on the time span of about 1 minute.

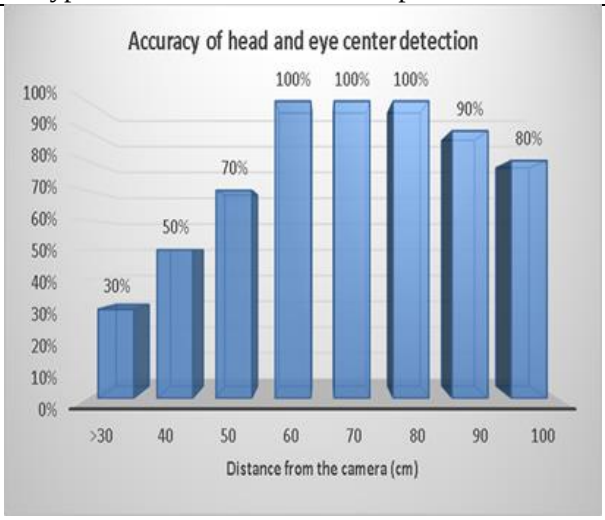

Figure 3. Accuracy of the head and eye center detection with varying distances from the camera

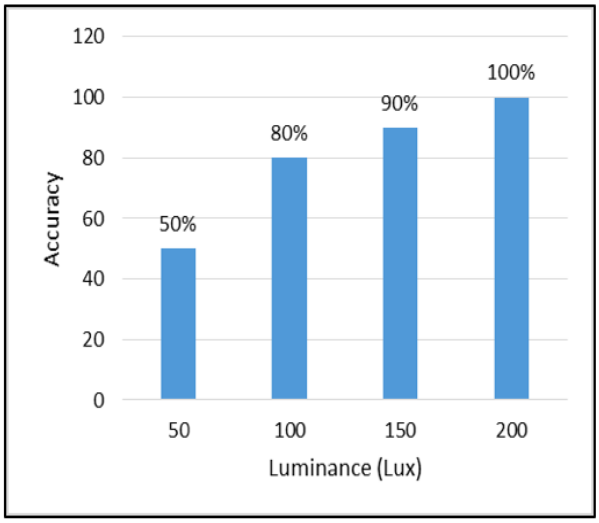

Figure 4. Accuracy of the head and eye center detection with varying lighting conditions

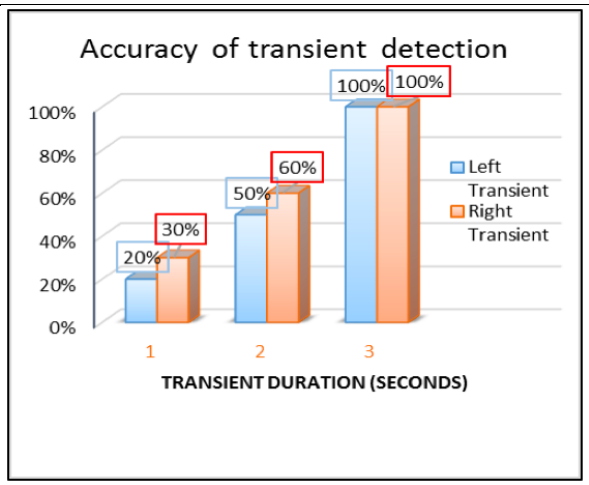

Figure 5. Accuracy of transient detection (eyeball rotation only)

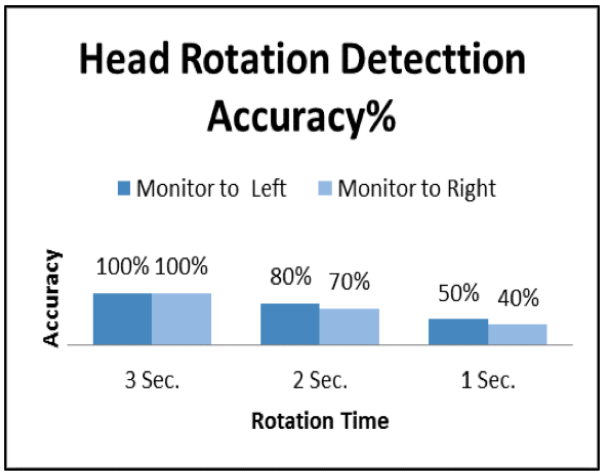

Figure 6. Accuracy of transient detection (head rotation only)

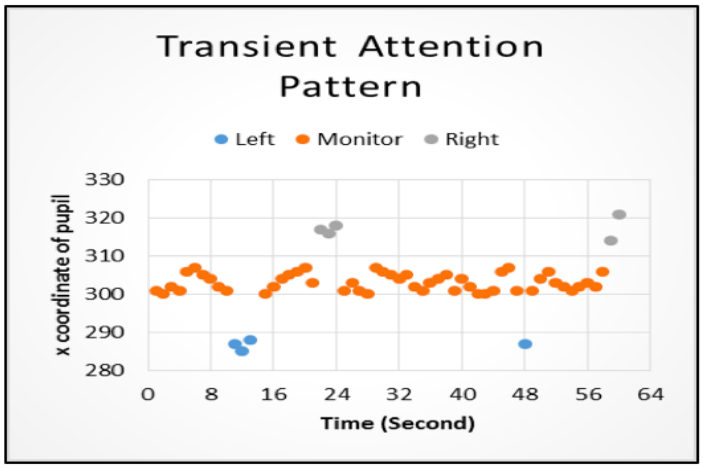

Figure 7. Transient attention pattern of the target person

5. CONCLUSION

In the work, we tried to track the head and the eye center from the captured continuous image with variable distances from the camera and in different lighting conditions. We see that if the distance from the camera is between 60cm to 80cm, we get the best tracking performance. Moreover, for a proper illumination, the number of 32W energy bulb is 4 for a $4m^2$ room that provided illumination of 200 Lux. To optimize the efficiency, this information provides us a clue to set the proper threshold value of illumination and right distance from the camera. In gaze detection combining with head pose we conducted different experiments with variable head rotation time. It is seen that, if the head rotation time is 3 seconds then the gaze tracking performance is the best.

In sustained and transient focus of attention detection experiments, our system can track the sustained attention with 100% accuracy. However, the transient tracking performance depends on the transient duration. We see that if the transient time is 3 seconds or more, the tracking performance is the best.

REFERENCES

[1] Judd T., Ehinger K., Durand F., and Torralba A., "Learning to Predict Where Humans Look," in IEEE 12th international conference on Computer Vision, pp. 2106-2113, IEEE, September 2009.
[2] Alexe B., Deselaers T., and Ferrari, "What is an object," V. 2010. CVPR, pp. 73-80.
[3] Cal H., Lin Y., and Mourant R., "Evaluation of Drivers' Visual Behavior and Road Signs in Virtual Environment," in proceeding of HFES 51st annual meeting, Baltmore, vol. 5, USA, pp. 1645–1649, 2007.
[3] L. Itti and C. Koch, "Computational modelling of visual attention,"Nature Reviews Neuroscience, vol. 2, no. 3, pp. 194–203, Mar 2001.
[4] Zhang H., Smith, M. R., and Witt, G. J, "Identification of Real-Time Diagnostic Measures of Visual Distraction with An Automatic Eye-Tracking System," Human Factors: The Journal of the Human Factors and Ergonomics Society, 48(4), pp. 805-821, 2006.
[5] Md. Golam Rashed, Royta Suzuki, Takua Yenezawa, Antony Lam, Yoshinori Kobayashi, and Yoshinori Kuno," Robustly Tracking People with LIDARs in a Crowded Museum for Behavioral Analysis" Institute of Electronics, Information and Communication Engineers (IEICE) Transactions on Fundamentals of Electronics, Communications and Computer Sciences, Vol E100, No. 11, PP 2458-2469, Nov 2017
[6] Murphy-Chutorian, E., & Trivedi, M. M. (2009). Head Pose Estimation in Computer Vision: A Survey, IEEE Transactions on Pattern Analysis and Machine Intelligence, 31(4), pp. 607- 626.
[7] Kim S., Chung S. T., Jung S., Kim, J., and Cho S. "Multi-Scale Gabor Feature Based Eye Localization," World Academy of Science, Engineering and Technology 21, pp. 483-487, 2007.
[8] Kroon B., Hanjalic A., and Maas S. M., "Eye Localization for Face Matching: Is It Always Useful and Under What Conditions?" in Proceedings of the 2008 international conference on Content-based image and video retrieval, pp. 379-388, ACM, July 2008.
[9] Asteriadi, S, Nikolaidis N, HajduA & Pitas I. (2006, March). An Eye Detection Algorithm Using Pixel to Edge Information. In Int. Symp. on Control, Commun. and Sign. Proc.
[10] B. S and O. J. M, "From camera head pose to 3d global room head poseusing multiple camera views," inIn Proc. Int'l. Workshop Classificationof Events Activities and Relationship, 2007.
[11] An K. H. & Chung M. J. (2008, September). 3d Head Tracking and Pose-Robust 2d Texture Map-Based Face Recognition Using A Simple Ellipsoid Model. In 2008. IROS 2008. IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 307-312. IEEE.
[12] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shapemodels—their training and application,"Comput. Vis. Image Underst., vol. 61, no. 1, pp. 38–59, Jan. 1995. [Online]. Available:http://dx.doi.org/10.1006/cviu.1995.1004
[13] P. P. Debnath, A. F. M. R. Hasan, and D. Das, "Detection and controllingof drivers' visual focus of attention," in2017 International Conferenceon Electrical, Computer and Communication Engineering (ECCE), Feb2017, pp. 301–307.
[14] David Cornish M. and Dukette D, "The essential 20: twenty components of an excellent Health Care Team," Dorrance Publishing