Towards Webcam-based Face Direction Tracking To Detect Learners' Attention Within Asynchronous E-Learning Environment

Nehad T.A. Ramaha[1], Ismail R. Karas[2], Elanur GÜL[3], Maşidenur Rumeysa Bozkurt[3], Rabia Yayvan[3]

1 Dept. of Computer Engineering, Karabuk University, Demir Celik Campus, 78050 Karabuk/Turkey - nehadramaha@karabuk.edu.tr

2 Dept. of Computer Engineering, Karabuk University, Demir Celik Campus, 78050 Karabuk/Turkey - ismail.karas@karabuk.edu.tr

3 Dept. of Computer Engineering, Karabuk University, Demir Celik Campus, 78050 Karabuk/Turkey - elanurgul29@gmail.com, bzkrtrmys@hotmail.com, rabia.yayvan@hotmail.com

KEY WORDS: COVID-19, E-Learning, Computer Vision, Face Direction, Asynchronous e-learning, Attention Detection.

ABSTRACT:

Recently, as a consequence of COVID-19 pandemic, the delivery of education at most of the educational institutions depended mainly on e-learning. So, the researchers give more attention for both synchronous and asynchronous e-learning. Although from an economical perspective, asynchronous e-learning seems to be the best e-learning option for institutions, still one of the biggest challenges is how to keep learners motivated for the entire learning process. One of important motivational factors that drives the success of the learning process is the learner attention. Therefore, to retain the learners' attention during the asynchronous e-learning process, we need first to detect their loss of attention. Accordingly, more studies started to focus on detecting learners' attention. However, those studies can't be widely used for attention detection within asynchronous e-learning environments, as the used approaches tend to be inaccurate, and complex for the design and maintain. In contrast, in this study, we explore the possibility to find a simple way that can be widely used to detect learners' attention within the asynchronous e-learning environments. Therefore, we used webcams which are available in almost every laptop, and computer vision tools to detect learners' attention by tracking their faces. Thereafter, we evaluated the accuracy of our suggested method, the result of this evaluation showed that our method is efficient.

1. INTRODUCTION

E-learning has gained a lot of popularity over the past decades and are now being offered to millions of learners around the world. Actually, there are many applications, platforms, and websites that could be used to deliver the classes online using one or both of the communication methods; asynchronous and synchronous communication. Synchronous e-learning method involves direct real time communication between the teacher and the learners. While, asynchronous e-learning method allows the learners to get the knowledge on their pace; at any time and in any place. Examples of synchronous e-learning are video conferences and real-time web chats. And examples of asynchronous e-learning are instructional websites, email, and forum. (Hrastinski, 2008)

Since the start of COVID-19 pandemic in 2020 most educational institutions over the world transformed the classes from traditional to online education. Although, online education considers the safest alternative for the traditional education, teachers and learners at different institutions show a negative feeling about distance education (Bruscato and Baptista, 2021a, Ramaha, 2021). According to Ebner and his colleagues (2020) before COVID-19, online learning have often been seen as a supporting method for the face to face classes, and rarely found pure e-learning for more than 30 years. As mentioned by Bruscato and Baptista (2021b) both synchronous and asynchronous online learning seem to be equally effective the educational process. However, asynchronous online learning seems economically the best option for institutions.

Learners' motivation playing an important role in the learning process, since many studies proved that motivation is one of the major elements of the educational process that drives the learners' performance (Weiner, 1985; Ames, 1992; Anderman and Maehr, 1994; Lin, Wu, and Wang, 2010; Ramaha, Basha, Ismail, and Umar, 2015). According to Keller's (1987) ARCS model of motivation, there are four motivational factors: Attention, Relevance, Confidence, and Satisfaction. Within the synchronous e-learning environment teacher interact directly with his/her learners, this allows the teacher to detect learners' attention problems and communicate with them to avoid those problems. In contrast, within the asynchronous e-learning environment the interaction between teacher and learners is indirect, this make it difficult for the teacher to detect learners' attention problems (Keller and Suzuki, 2004; Ramaha and Karas, 2021). Therefore, detecting learners' loss of attention during an asynchronous e-learning process is particularly important, as this can help applying some motivational tactics to return the learners' attention to the lesson.

Recently, more research studies have started to focus on detecting learners' attention. These researches fall into three main categories; first category focused on (i) utilizing physical sensors such as Body Pressure Measurement System (BPMS) (e.g. D'Mello and Graesser, 2010), Heart Rate (HR), Skin Conductance (SC), brain waves (EEG) (e.g. Derbali, Chalfoun, and Frasson, 2011a, 2011b; Abujayyab et al., 2017), and eyetracker (e.g. Veliyath et al., 2019), these types of researches can't be widely used for attention detection within asynchronous e-learning environments as it need special equipments. The second category focused on (ii) analyzing the data generated from the interaction between the learner and the learning system (e.g., de Vicente, 2003; Qu. and Johnson, 2005; Ramaha, 2012; Ramaha, 2017; Abujayyab and Karaş, 2019), these data are usually saved by the system in a log file, these types of researches applicable for attention detection in e-learning systems. However, those research approaches still

inaccurate, and are complex to design and maintain. The third category focused on (iii) detect learners' attention utilizing webcams that available currently in almost all laptops (Bixler and D'Mello, 2016; Laddi and Prakash, 2019). Most of this research category tried to detect learners' inattention or loss of focus based on eye tracking and eye gaze data. However, those research approaches also still inaccurate, and are complex to design and maintain. Therefore, in this study, we explore the possibility to find a simple approach that can be widely applied to detect learners' attention within asynchronous e-learning environments. Thus, this study using a webcam-based face direction tracking to detect the learners' attention.

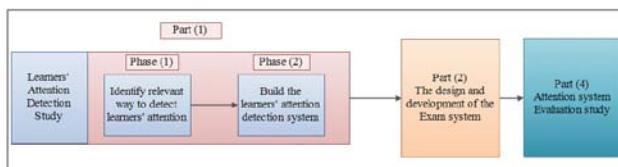## 2. MATERIAL AND METHODS

2.1 Research Framework

Figure 1. Research Framework.

Figure 1 shows the research framework of this study, which consists of three parts:

1. Learners' attention detection: This part of the study includes two phases:
- Phase 1: During this phase and after literature review, we suggested to use webcam-based face direction tracking to detect the learners' attention.
- Phase 2: Through this phase we used computer vision (CV) tools and techniques to track the learners' face directions using the webcam.
2. The development of exam system: This part of the study includes a development of a simple exam system to be used with the learners' attention detection system. The aim of this system to test and evaluate the attention detection and its accuracy.
3. Evaluation study: During this part we compare the learners' attention detection values from the system with the detection values come from the teachers for the same faces.

2.2 Face Direction Tracking

Face direction tracking is a challenging problem in computer vision because of the various steps required to solve it. Figure 2 shows the steps of the face direction tracking process. First step is to detect the face in the video frame. Thereafter, in the second step we identify the different facial landmarks such as the mouth, eyes, and nose. After that, the points of each landmark will be converted to 3D coordinates. Then, in the next step we resolve the Perspective-n-Point (PnP) and extract the Euler Angles to find the inclination of the face. The last step is to find the face direction based on the Euler angles.

2.2.1 Pre-Processing: In computer vision, Face direction estimation specifically implies the relative orientation of the face with respect to the camera. This estimation usually referred to as a Perspective-n-Point problem (PnP). Firstly, we need to detect the face from the camera frame and then predict the face landmarks. For these two steps we used frontal face detector and the shape predictor functions from dlib library.

2.2.2 Tracking process: After detecting the face and predicting its landmarks, now we can start the face direction tracking process. The tracking process starts with solving the PnP to find the inclination of the face. To accomplish this step we used solvePnP function from OpenCV library. This function helps to find face landmarks pose from 2D point correspondences, and then it will return the rotation and the translation vectors that transform a 3D point expressed in the face coordinate frame (the real world) to the camera coordinate frame. Next equation defines the PnP problem solution:

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & \gamma & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \quad (1)$$

The left side of equation 1 $s[u\ v\ t]^t$ denotes the 2D image taken by the camera. The right side of the equation has 3 parts. The first part of the right side is the camera matrix and it's looks like an upper triangular matrix, in this matrix f(x, y) is the focal lengths γ is the skew parameter which we will leave as 1 in our code, while (u_0, v_0) are the center of the image. The middle part, which is the matrix with r and t are the desired 3D rotation and 3D translation of the camera. The last part denotes the 3D model of the face. Figure 3 shows the representation this equation.
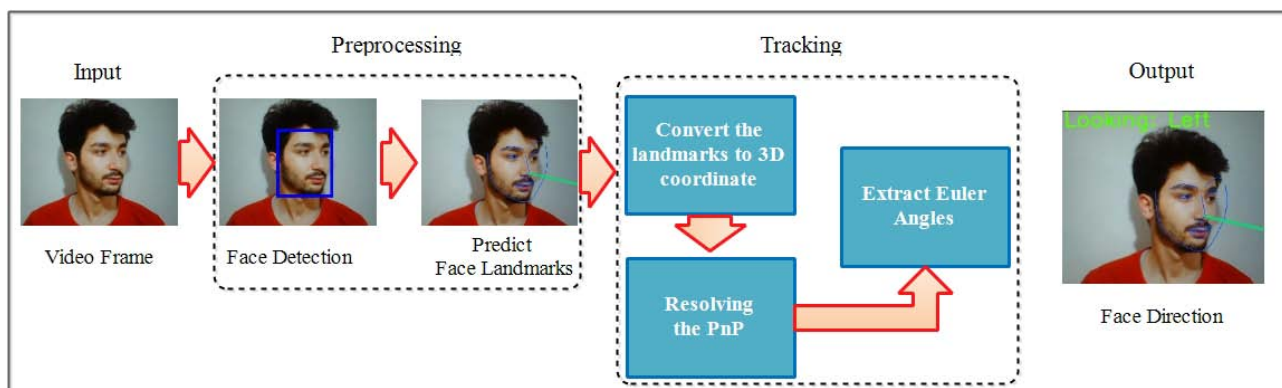

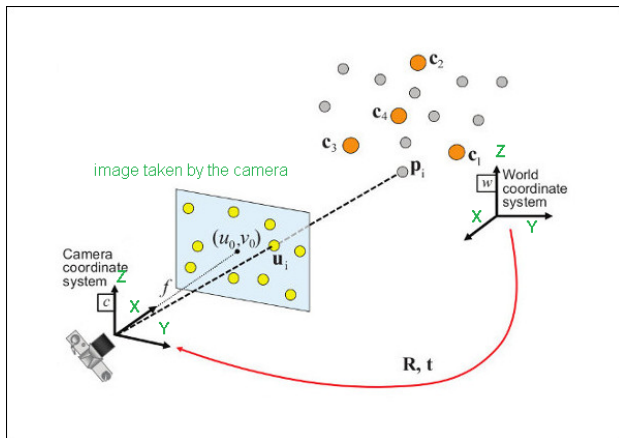**Figure 2**. Face Direction Tracking Process.

Figure 3. Representation of PnP Equation (OpenCV, 2021).

Finally, by resolving PnP we get rotation vector , and translation vector matrix. Thereafter, we can extract Euler Angles using RQDecomp3x3 function from OpenCV library. However, RQDecomp3x3 function takes a rotation matrix rather than a rotation vector as a parameter. Therefore, we used Rodrigues to convert a rotation matrix to a rotation vector. After getting the Euler Angles we can check if the face looking to the left or the right. However, to check if the face looking up, down, or forward we draw a line from the nose (e.g Figure 4) and calculate the angle of the line's incline.
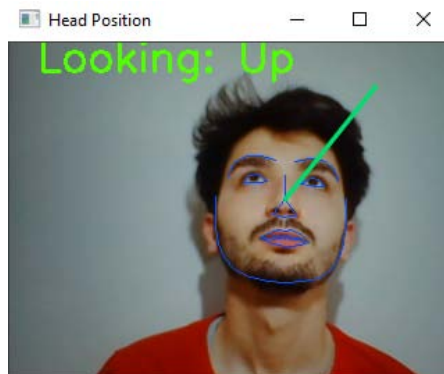

Figure 4. Face Direction.

2.3 Data Collection

In order to evaluate the suggested method to detect learners' attention, we developed a simple exam system with face tracking ability, Figure 5 shows the developed system. Then we asked 20 learners to use the system and answer the quiz. The learners used their own devices with different types of web cams. The system collects a random data in its database about the tracking of the learners' face as seen in Figure 6. Moreover, the system takes the learners' face snapshot that related to the saved tracks data for evaluation purpose.
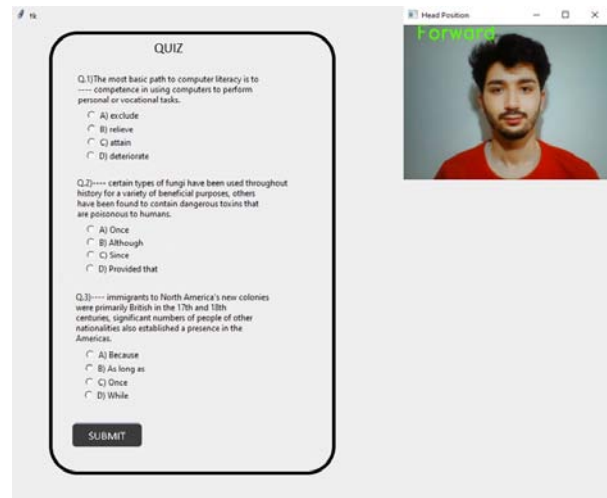

Figure 5. The Exam System.

Figure 6. Face Direction Tracking Database.

2.4 Evaluation study

The evaluation study procedures that were followed during this part of the research can be summarised as follows:

1.   Obtain the learners' for the study.
2.   Meeting with the learners' to arrange the study schedule and introduce the system.
3.   The learners' used the system according to the schedule.
4.   The system collected a random data in its database about the tracking of the learners' face direction and took a snapshot for their faces.
5.   Obtain the teachers who will detect the learners' face direction from the saved faces' snapshots.
6.   During the detection process the teachers were asked to label the face direction of each photo.
7.   The evaluation process has been done by comparing the label from the system with the labels from the teachers.

3.   RSULTS

We have 10 teachers participated in the evaluation study. Each participant was asked to label each of 15 faces' photo by the face's direction; there are 5 choices for the labels right, left, up, down, or forward. Table 1 presents the obtained replies of each photo. The columns represent: the name of the face's photo, the direction label of each of the photos from the system output (this column hid for the participants), the 5 choices of the labels given to participants, and the number of times each of the labels was selected.

Table 1. Face Direction Tracking Evaluation Study Results.

System output		Participants' choices				
Face	Label	Right	Left	Up	Down	Forward
F1	Up	0	2	8	0	0
F2	Left	0	10	0	0	0
F3	Left	0	9	1	0	0
F4	Right	9	0	0	1	0
F5	Up	1	0	9	0	0
F6	Left	0	8	1	0	1
F7	Up	1	0	9	0	0
F8	Down	0	0	0	10	0
F9	Right	8	0	1	0	1
F10	Right	9	0	0	0	1
F11	Up	0	2	8	0	0
F12	Forward	0	0	0	0	10
F13	Left	0	9	1	0	0
F14	Left	0	8	2	0	0
F15	Down	1	0	0	9	0

4. DATA ANALYSIS

One out of the 5 choices for labelling the faces' direction that given to the participants is the same as the output of the system, and was considered as the "Agree" label. The other four options were considered "Disagree" labels. For each of the faces, a chi-square goodness of fit test was performed. The column p in Table 2 represents the probability of the distribution obtained in the evaluation study. Those faces in which the participants showed a statistically significant preference ($p < 0.05$) for the "Agree" label were considered as correct label from the system.

Table 2 shows the results of applying the chi-square goodness of fit test to the system output; the faces' direction labels. The columns represent: the name of the face, number of participants, number of agree replies, number of disagree replies, and the probability of replies assuming the null hypothesis.

Table 2. Chi-square Results For System Output.

System output		The Result			
Face	Label	n	Agree	Disagree	p
F1	Up	10	8	2	0.0017
F2	Left	10	10	0	0.0000
F3	Left	10	9	1	0.0001
F4	Right	10	9	1	0.0001
F5	Up	10	9	1	0.0001
F6	Left	10	8	2	0.0017
F7	Up	10	9	1	0.0001
F8	Down	10	10	0	0.0000
F9	Right	10	8	2	0.0017
F10	Right	10	9	1	0.0001
F11	Up	10	8	2	0.0017
F12	Forward	10	10	0	0.0000
F13	Left	10	9	1	0.0001
F14	Left	10	8	2	0.0017
F15	Down	10	9	1	0.0001

5. SUMMARY AND DISCUSSION

In this study, we have examined the challenge of tracking the learners' face direction to detect their attention while using an asynchronous e-learning environment using a regular webcam. This can help the asynchronous e-learning systems developers to enhance their systems with the ability of detecting learners' loss of attention during the learning process, as this can help returning the learners' attention to the lesson. Before deciding our way to detect learners' attention we did a comprehensive literature review. As a result of that review, we found that the main drawback of most of those previous studies is that can't be widely used. Therefore, during this study we explored the possibility to find an appropriate way that can be widely applied to detect learners' attention during the use of asynchronous e-learning environments. Thus, we decided to use the webcam-based face direction tracking to detect the learners' attention. In this context, we used two open-source libraries with Python, dlib and OpenCV. For this purpose, we developed an exam system that has the ability to track learners' face direction during the use. Thereafter, the system was used by 20 learners and interaction data have been collected. After that, the collected data have been evaluated with the participation of 10 teachers.

The results from the evaluation study showed that the suggested way to track the learners' face direction is efficient. As the analysis of the results showed that in all the cases of detecting the learners' face direction by the system, there was a significant agree from the participants with the system outputs. However, there was some limitation in our study, such as the number of participants that was low because of COVID-19 pandemic restrictions. The time also was another restriction, as we need more time to make a more comprehensive comparison between different methods of face direction tracking.

We have already started planning the future steps to enhance our suggested method for detecting learners' attention and to make a more comprehensive evaluation of our method. However, beyond the pure technical issues, we still have many additional interesting privacy and pedagogical research questions such as: Would our suggested method be accepted by the learners? Why yes, and why not? What is the impact of our method on the learners' attention, engagement, and performance?

REFERENCES

Abujayyab, S. K., Ahamad, M. S. S., Yahya, A. S., Ahmad, S. Z., & Aziz, H. A. (2017, October). Automating an integrated spatial data-mining model for landfill site selection. *In AIP Conference Proceedings* (Vol. 1892, No. 1, p. 130001). AIP Publishing LLC.

Abujayyab, S. K., & Karaş, İ. R. (2019). Automated prediction system for vegetation cover based on modis-ndvi satellite data and neural networks. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 42, 9-15.

Ames, C., (1992). Classrooms: Goals, structures, and student motivation. *Journal of educational psychology*, 84(3), 261.

Anderman, E.M. and Maehr, M.L., (1994). Motivation and schooling in the middle grades. *Review of educational Research*, 64(2), 287-309.

Bixler, R. and D'Mello, S., (2016). Automatic gaze-based user-independent detection of mind wandering during computerized reading. User Modeling and User-Adapted Interaction, 26(1), 33-68.

Bruscato, A.M. and Baptista, J., (2021a). Teaching Modalities in Brazilian and Portuguese Universities: A Case Study on the Perception of Students and Professors in Times of Covid-19. *Revista Brasileira de Educação*, 26.

Bruscato, A.M. and Baptista, J., (2021b). Synchronous and asynchronous distance learning of anaphora in foreign languages: an experimental study. *Texto Livre: Linguagem e Tecnologia*, 1-18.

D'Mello, S., and Graesser, A. (2010). Mining bodily patterns of affective experience during learning. *Paper presented at the Educational Data Mining* 2010.

de Vicente, A., (2003). Towards tutoring systems that detect students' motivation: an investigation (Doctoral dissertation, University of Edinburgh).

Derbali, L., Chalfoun, P., and Frasson, C., (2011a). Assessment of learners' attention while overcoming errors and obstacles: An empirical study. *Paper presented at the Artificial Intelligence in Education*.

Derbali, L., Chalfoun, P., and Frasson, C., (2011b). A theoretical and empirical approach in assessing: From serious games to an ITS. *Paper presented at the FLAIRS*.

Ebner, M., Schön, S., Braun, C., Ebner, M., Grigoriadis, Y., Haas, M., Leitner, P. and Taraghi, B., (2020). COVID-19 epidemic as E-learning boost? Chronological development and effects at an Austrian university against the background of the concept of "E-Learning Readiness". *Future Internet*, 12(6), 94.

Hrastinski, S., (2008). Asynchronous and synchronous e-learning. *Educause quarterly*, 31(4), 51-55.

Keller, J. (1987). Development and use of the ARCS model of instructional design. *Journal of instructional development*, 10(3), 2-10.

Keller, J. and Suzuki, K., (2004). Learner Motivation And E-Learning Design: A Multinationally Validated Process. *Journal of educational Media*, 29(3), 229-239.

Laddi, A. and Prakash, N.R., (2019). Eye Gaze Tracking Based Directional Control Interface For Interactive Applications. *Multimedia Tools and Applications*, 78(22), 31215-31230.

Lin, K.C., Wu, T.K. and Wang, T.B., (2010), August. Using competitive digital game-based learning to improve learning motivation. *In IET International Conference on Frontier Computing. Theory, Technologies and Applications* (128-132). IET.

OpenCV: *Finds an object pose from 3D-2D point correspondences.* url: https://docs.opencv.org/4.1.0/d9/d0c/group__calib3d.html (visited on 10/02/2021).

Qu, L. and Johnson, W.L., (2005). Detecting the learner's motivational states in an interactive learning environment. *In Proceedings of the 2005 conference on artificial intelligence in education: Supporting learning through intelligent and socially informed technology,* 547-554.

Ramaha, N.T., (2017). *Extraction, Validation, and Evaluation of Motivational Tactics Rules in a Web-Based Intelligent Tutoring System (WITS)* (Doctoral dissertation, Universiti Sains Malaysia).

Ramaha N.T., (2021). Towards an Interactive Web-Based Tutoring System(IWTS) To Retain The Students' Motivation. *International Journal of Engineering and Information Systems,* 5(9), 15-25.

Ramaha, N.T., Basha, A.D., Ismail, W.M.F.W., and Umar, I.N., (2015). Students' Motivation Is "Realistic Challenge": Online Courses. *International journal of scientific & engineering research*, 6(6), 1383- 1386.

Ramaha, N.T. and Ismail, W.M.F.W., (2012). Assessment of learner's motivation in web based e-learning. *International journal of scientific & engineering research*, 3(8), 1-5.

Ramaha, N. T., and Karas, I. R. (2021). Maintain Learners' Motivation within Asynchronous E-Learning Environments: How Can Interactive Avatars Help. *European Journal of Advances in Engineering and Technology*, 8(9), 9-14.

Veliyath, N., De, P., Allen, A. A., Hodges, C. B., and Mitra, A. (2019). Modeling Students' Attention In The Classroom Using Eyetrackers. *In Proceedings of the 2019 ACM Southeast Conference* (pp. 2-9).

Weiner, B., (1985). An Attributional Theory Of Achievement Motivation And Emotion. *Psychological review*, 92(4), 548.