Towards a new method of estimating the student attention based on the eye gaze

Tarik Hachad[1], Abdelalim Sadiq[1], Fadoua Ghanimi[2], Lamiae Hachad[3], Ahmed Laguidi[4,5]
[1]Department of Computer Science, Faculty of Sciences, Ibn Toufail University (UIT), Kenitra, Morocco
[2]Laboratory of Electronic Systems, Information Processing, Mechanics and Energy, National School of Chemistry,
Ibn Toufail University (UIT), Kenitra, Morocco
[3]Laboratory of Signals Systems and Components, Faculty of Sciences and Technologies,
Sidi Mohamed Ben Abdellah University (USMBA), Fez, Morocco
[4]Laboratory of Networks, Computer Science, Telecommunication, Multimedia (RITM), High School Technology,
Hassan II University of Casablanca, Casablanca, Morocco
[5]Laboratory of Modelling Applied to Ecnomics and Management (MAEGE), High School Technology,
Faculty of law, Economic and Social Sciences Ain Sbaâ, Hassan II University of Casablanca, Casablanca, Morocco

Article Info

Article history:

Received Jan 6, 2022
Revised Jul 5, 2022
Accepted Jul 26, 2022

Keywords:

Attention detection
Computer vision
Convolutional neural network
Eye gaze estimation
Student attention

ABSTRACT

This paper presents a new system for automating the monitoring and estimation of student attention during the course session. The followed approach is based on the analysis of the student's gaze to predict his state of attention. A simple hardware device consisting of a camera and a pc was used in this study. Existing machine learning algorithms were used for the student gaze estimation. The principles of homography were used to ensure the transformation from an image coordinates system to a real-world coordinates system. 5 students took part in this experiment and whose gaze was detected and analyzed during 10 minutes of the class session in order to analyze their states of attention and inattention.

Corresponding Author:

Tarik Hachad
Department of Computer Science, Faculty of Sciences, Ibn Toufail University (UIT)
Kenitra, Morocco
Email: Tarik.hachad@uit.ac.ma

1. INTRODUCTION

The human gaze is a natural cue that provides rich information on the attention of individuals in social interactions. Human beings receive and communicate various information through their eyes. Indeed, an eye points to the object to be analyzed during a recognition or learning operation, pointing to an interlocutor expresses an interest in the discussion. Based on this reality, and with the objective of improving the teaching and learning experience, we have chosen to analyze an aspect of the human eye which is the direction of gaze to extract information on the attention of students in the class during the course [1]. The operation must be carried out under normal course conditions, so the used devices must to be not distractive or invasive for the students [2]. To achieve this, we have chosen to use a single camera as a source of information and to deploy a method for 3D gaze estimation based only on the images of the faces provided by the camera.

The present paper proposes a system for monitoring and measuring the student's attention during a normal class session, based on his point of gaze. This system uses an inexpensive camera and a computer to gauge the student's attention throughout the course. The field of gaze analyzed is the plane which contains the display board, so a student is said to be attentive if his gaze points to an area in this plane. Analysis of the information collected will determine the extent to which the teaching materials and the teaching style are attracting the attention of the student. It will also make it possible to identify distraction objects and the moments of loss of attention of the students.

Our paper is organized as shown in: In next section 2, a review of existing gaze estimation approaches and data sets are presented. In section 3, we explain the proposed approaches with details related to our system. Next, we discuss the results obtained by our system compared with those of similar systems. Finally, we end with some conclusions and perspectives.

2. LITERATURE REVIEW

In recent years, a wide variety of remote gaze tracking algorithms have been reported in the literature. For our use case, the general knowledge base can be divided into three categories: appearance-based, model-based and cross-ratio based.In this section, we present the three categories of gaze estimation and the datasets that were created for this purpose.

2.1. Gaze estimation: appearance-based methods

Choi *et al*. [3] used convolutional neural networks (CNNs) to perform head pose estimation with categorization of driver gaze areas (central rear-view mirror, left and right part of the windscreen, left window). They have built their own data set of men and women drivers, including situations of wearing glasses, their system achieves an accuracy of 95%.

Konrad *et al*. [4], gaze tracking was carried out in a very constrained environment, the camera was placed at a distance of 51 cm from the individual’s face. To train their CNN neural networks, they built a data set implemented in their particular setup and composed of images of 5 subjects. The results are promising, however, the CNN network needs a lot of data to be properly trained.

George and Routray [5], the proposed algorithm consists in detecting the faces present in the image using a modified version of the Viola-Jones algorithm, the rough eye region is obtained using geometric relations and facial landmarks. Then, a convolutional neural network is used for gaze direction classification. This algorithm was tested on the Eye Chimera data set and gave good results in terms of computational complexity which makes it a good choice for smart devices.

However, Vora *et al*. [6] use two distinct CNN architectures (AlexNet and VGG16) to classify the driver’s gaze into seven zones. This two CNNs have been fine tuned on the dataset created for this purpose and which contains 47,515 images naturalistic driving tests of 11 drives, driven by 10 subjects in two different cars and labeled with 6 gaze zones. This research study submit a comparison of the performance of the two architectures and the authors found that VGG16 outperformed AlexNet due to the small size of the core (3×3) in the convolution layer. Also, they proved that using the upper part of the face as input works better than the whole face.Their system achieved an accuracy of 93.36%.

A new low-cost gaze-based text entry method has been proposed in Zhang *et al*. [7]. This approach aims to help people with disabilities communicate by text using eye movement. Indeed, the authors classified the gaze in 9 directions responding to the input method on a T9 keyboard. The confirmation of the letter entered is done by blinking of the eyes. To achieve this goal, they built a convolutional neural network (CNN) to estimate gaze in 9 directions. The CNN was trained on a large-scale data set they created with images of the eyes of 25 people. According to the results, this model can estimate the gaze of different people in various lighting conditions, with an accuracy of 95.01%.

2.2. Gaze estimation: model-based methods

Model-based techniques perform gaze estimation by combining the geometric model of the eye with eye features, such as cornea reflection and pupil center [8], [9]. These methods attempt to estimate, using a geometric model of the eye, the center of the cornea, the optical and visual axes of the eye. The direction of gaze is then determined by the visual axis, which goes through the center of the cornea and the fovea. Unlike the old methods which used infrared illuminations and high resolution cameras to extract features (eyeball), recent methods, rely on machine learning approaches which allow them to high features extraction accuracy from a simple webcam images , and under varying lighting conditions [10].

2.3. Cross-ratio based gaze estimation

The cross-ratio (CR) based methods are invariant to head pose changes. They perform gaze estimation by projecting a known rectangular pattern of near-infrared (NIR) lights on the eye of the user and using invariant property of projective geometry. Yoo and Chung [11] have achieved an interesting work on the experimental verification of cross-ratio based methods.

The various methods mentioned above have advantages and disadvantages. The model-based methods allow the gaze estimation under the constraint of the user's head movement. But, they require the use of specific materials (multiple cameras and several light sources).

Appearance-based methods learn to map directly the eye appearance to human gaze. This has become possible by the advancements of deep learning techniques allowing the extraction of the shape and texture properties of the eyes, and by the creation of several eye gaze estimation data sets. These methods have low hardware requirements which make them suitable for implementation on platforms without a high-resolution camera or additional light sources [12]. Despite, they are not able to model effectively under conditions of varying head positions and illumination changes. This is because the appearance of the eyes may look similar under different head poses and gazing directions. Changes in illumination (under the same pose) can change the appearance of the eye and affect the gaze estimation accuracy [12], [13].

Cross ratio based methods don't require a hardware calibration mapping the position of the camera to the monitor and allow free head motion. However, the distance from the user greatly affects their performance and the projection of infrared light for a long time can tire the user [14], [15].

2.4. Data sets for gaze estimation

Several large-scale gaze estimation datasets have been created in recent years, among which a good portion are publicly available. Some of these datasets have been constructed using images captured in labs under particular setups, while others were acquired in outdoor environments. Table 1 summarizes a comparative study of the common datasets used in gaze estimation work.

Table 1. Summary of some common gaze estimation datasets

Datasets	Year	Total	Subjects	Purpose	Configuration
CAVE-DB [13]	2013	5880	56	Gaze estimation	Collected in laboratory conditions; 5 differents head poses horizontally; 21 gaze directions for each subject and head pose.
Eyediap [16]	2014	94 videos	16	Gaze estimation	Collected in laboratory conditions; free head pose.
UT Multiview [17]	2014	64000	50	Gaze estimation	Collected in laboratory; 8 head poses and 20 gaze directions per head pose.
MPIIGaze [18]	2015	213659	15	Evaluating gaze tracking methods	Images are captured by laptops cameras in daily life; Free head pose and variable illumination conditions.
GazeCapture [19]	2016	2445504	1474	Gaze estimation	Images are captured by a mobile phone camera under different variation in head pose and illumination
MPIIFaceGaze [20]	2017	45k	15	Appearance-based gaze estimation	Collected using laptop camera, free head pose.
RT-Gene [21]	2018	123k	15	Appearance-based gaze estimation	Collected in laboratory; free head pose; annotated with mobile eye tracker; use GAN to remove the eye-tracker in face images [22].
Nvgaze [23]	2019	4.5M	30	Near-eye gaze estimation	Collected under laboratory conditions, using infrared illumination.
ETH-XGaze [24]	2020	1.1M	110	Gaze estimation	Collected in the laboratory, high resolution images, different poses of the head and different gazes.
EVE [25]	2020	4,2k videos	54	Gaze estimation	Collected in laboratory; different gazes, different head poses with annotation.

3. OUR APPROACH

The system we present in this article is essentially based on a low-cost camera and a software that implements our algorithms and methods for monitoring the student gaze point. The proposed system consists of a camera placed above a display board which serves as a means of illustration or as a projection surface (see Figure 1). The students sit at a distance of 150 cm in front of the blackboard and follows the teacher's explanations. The goal is to follow each student's gaze and detect cases where he looks away from the projection or illustrations displayed on the board.

3.1. Distance from the camera estimation

The camera captures images of the student at a rate of 30 frames per second with a resolution of 1280/720 px. The system first detects the student's face and calculates the distance between the student and the camera which is supposed to be the center of the world coordinate system used to precisely define the student's gaze point. Then, it tries to extract the 2D image coordinates of the region of interest features such as the iris, the internal and external corner of the eye to predict the center of the eyeball and calculate the gaze vector. Finally, a transformation of the gaze vector coordinates from the image coordinates to the world coordinates to find the point of intersection of the gaze vector with the plane which contains the camera and the display board. This process is illustrated in the Figure 2.

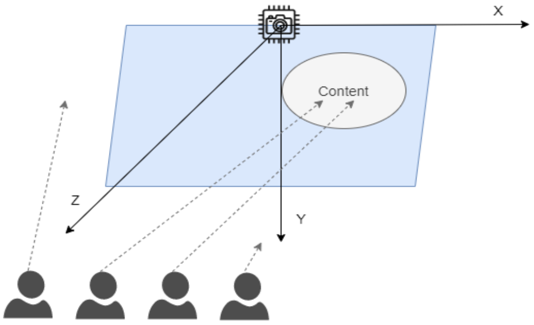

Figure 1. General overview of the concept of monitoring the students gaze in the classroom using a simple camera

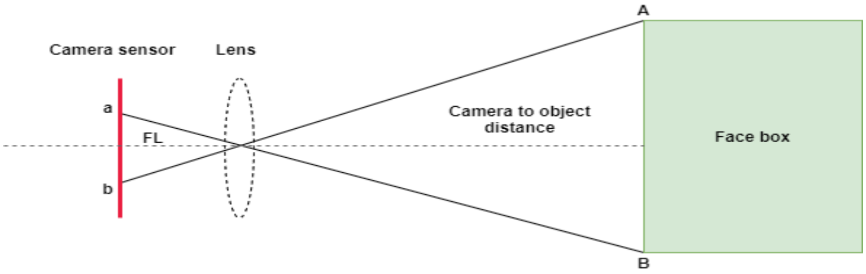

Figure 2. Principle of pinhole camera

In this article, we use a method based on a single camera to calculate the distance between the student and the stationary camera. Figure 2 shows the experimental setup and the operating principle of this method based on triangle similarity. The segment AB represents the width in centimeters of the object placed in front of the camera, ab is the width in pixels of the reflection of the object on the complementary metal–oxide–semiconductor (CMOS) of the camera and Alfa is the angle between the optical axis and the A end of the object. Camera calibration is required, it should be done using images of an object captured by the camera at different angles in a 3-dimensional plane (x, y, z). In our scenario, it's the face of the student which will be detected using an appropriate detection algorithm. The values of AB, ab, and Distance are measured, it remains to deduce the value of the focal length (FL) using:

$$tan\alpha = \frac{AB}{2\ distance}\ and\ tan\alpha = \frac{ab}{2\ FL}$$

$$\frac{AB}{2\ distance} = \frac{ab}{2\ FL}$$

$$FL = \frac{ab\ x\ distance}{AB}$$

Note that the distance and measurement AB are in centimeters and ab is in pixels. Once the camera is calibrated and the FL is calculated, we can calculate the distance from the pupil to the camera using the triangle similarity, so:

$$Distance = \frac{FL\ x\ AB}{ab} \qquad (1)$$

3.2. Face detection and eye region localization

As in most computer vision systems that deal with issues related to facial emotion, head pose, or gaze estimation, the first stage of our system is face detection. There are a multitude of techniques performing face detection [26]. However, the best performing approaches are those based on CNN's because they have shown very good results in extracting features from convolution layers [27]. CNN made it possible to jointly perform face detection and facial features extraction to reduce computation time and improve extraction accuracy. In Figure 3, we present the steps followed by our proposed system. We have used Dlib [28] for the detection and extraction of facial features. This library is open and has shown remarkable performance in detecting faces and extracting face landmarks.

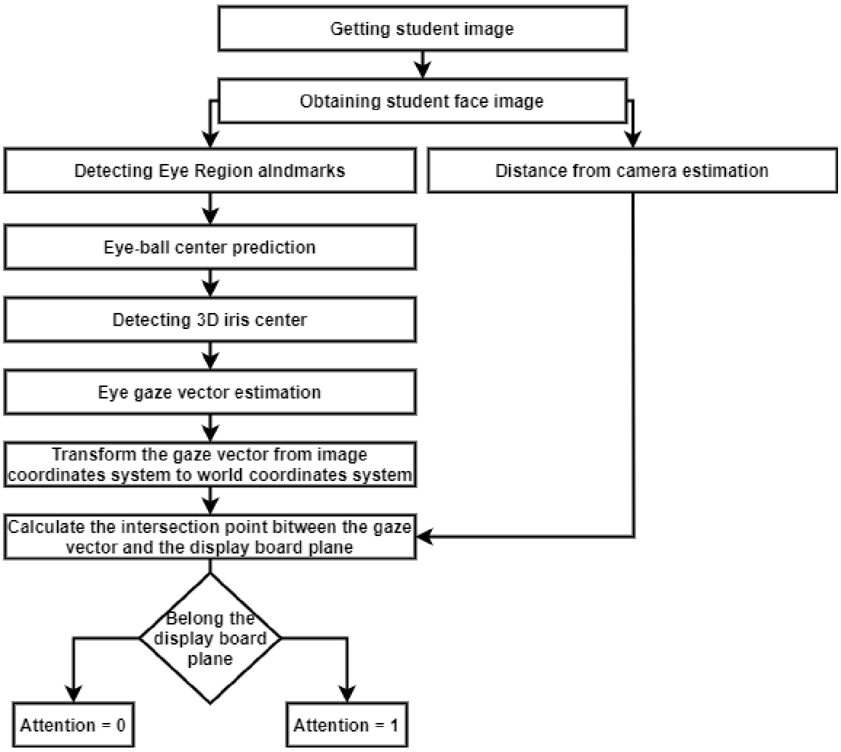
Getting student image
Obtaining student face image
Detecting Eye Region alndmarks
Distance from camera estimation
Eye-ball center prediction
Detecting 3D iris center
Eye gaze vector estimation
Transform the gaze vector from image coordinates system to world coordinates system
Calculate the intersection point bitween the gaze vector and the display board plane
Belong the display board plane
Attention = 0
Attention = 1

Figure 3. Flow diagram of the proposed system

3.3. Gaze estimation

Since gaze estimation is a crucial step in our system, we have reviewed a number of existing methods that best match our use case and material constraints. We have chosen to use the method of Park *et al.* [29] which outperforms the state of the art results on real-world eye images. The main idea behind this method is to set up an accurate eye landmarks detector, which will eventually allow the estimation of the gaze. The data used to generate test, and evaluate the eye landmarks detection model come from 4 datasets EYEDIAP, MPIIGaze +, UT Multiview, and Columbia. The detection of facial and skeletal joint landmarks is a well-researched subject. Indeed, several studies have proposed architectures of deep convolutional neural networks to solve the problem of facial landmarks and skeletal joints detection [30], [31]. The authors have adapted the hourglass architecture [32] for the facial landmark detection task, this architecture was originally applied to the human pose estimation, with the aim to solve the recurrent issue of occlusion of a part of the body by a hand or an object.

The region of interest is the eye, which contains fewer overall structuring elements than in pose estimation. This peculiarity allowed the detection with reasonable precision of the center of the eyeball and the iris edge landmarks in cases of occlusion. Using the eye landmarks that the system provided, the authors propose two scenarios for gaze estimation:

3.3.1. Feature-based gaze estimation

The features vector is made up of 17 coordinates: 8 of the limbus, 8 of the edge of the iris, 1 of the center of the iris and a 2D gaze direction. These landmark coordinates are normalized by the width of the eye c2-c1 which represents respectively the outer and inner corner of the eye in a coordinate system centered at c1. The 2D gaze direction is obtained by subtracting the center of the eyeball from the center of the iris. A support vector regression (SVR) is then trained using the 36 landmark-based features to produce a model that estimates a 3D gaze direction representing the pitch and yaw of the eyeball.

3.3.2. Model-based gaze estimation

Two intersected spheres were used to model the human eyeball, the first is large and the second is small to represent the corneal bulge. Two intersected spheres were used to model the human eyeball. The first one is large and the second is small to represent the corneal bulge. The data available to the system are the 8 landmarks of the edge of the iris predicted from the eye image, the eyeball center landmark and the iris center landmark. The radius of the eyeball is estimated in pixels from the coordinates of the eyeball center and those of the iris. Thus, the coordinates of the iris points as shown in:

$$uij \;=\; xij \;=\; xc \;-\; rxy\, cos\theta j'\, sin\phi j'$$

$$vij \;=\; yij \;=\; yc \;+\; rxy\, sin\theta j'$$

For model-based gaze estimation, θ, ϕ, δ, γ which are the gaze direction (θ, ϕ), δ angular iris radius and the angular offset γ equivalent to eye roll are unknown. The already detected landmarks are used to solve this problem. Thus, authors proposed the use of an iterative optimization method such as the conjugate gradient.

The gaze vector in the image coordinate is then deduced, its starting point is the center of the iris u (x, y) in pixel and its ending point is v (x, y) where:

$$xv = xu + c * cos\theta sin\varphi$$

$$yv = yu + c * sin\theta$$

The coordinates obtained are in pixels in an image system coordinate. To get the point of gaze, we have first to transform the coordinates of the gaze vector from the image coordinates system (xp, yp) to the real-world coordinates system (X, Y, Z), whose origin O is the camera and the display board is a plane in (X, Y). The Z axis constitutes the depth between the plane (board, camera) and the student. The gaze point is then obtained by calculating the intersection between the gaze vector and the plane (display board).

3.3.3. Camera calibration

The cameras are based on a model-based imaging system called pinhole or perspective projection. Indeed, this allows the projection of the PW points of a 3D scene onto a 2D plane made up of pixels, which is the image. This can be expressed by:

$$\mathrm{M} \;=\; \mathrm{K} * \;[\mathrm{R}|\mathrm{t}]$$

K is the matrix that contains the intrinsic parameters and the extrinsic matrix [R|t] which are respectively the rotation matrix and the translation vector that define the coordinates changes from the real-world coordinate system to that of the camera.

$$K = \begin{bmatrix} fx & \gamma = 0 & cx \\ 0 & fy & cy \\ 0 & 0 & 1 \end{bmatrix}$$

The different components of the intrinsic parameter matrix are as follows fx and fy represent the focal length on the x and y axes, expressed in pixels, γ is the skew between the axes, in general it is equal to 0, cx and cy are the coordinates in pixels of the intersection of the optical axis and the image plane. So, the relation between a point Pw in the world coordinate system (X, Y, Z) and its image projection Pc (xc, yc) in the image coordinate system is given by:

$$Pc \begin{bmatrix} xc \\ yc \\ 1 \end{bmatrix} = MPw \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$$

Determining the set of camera parameters that describe the mapping between the 3D reference coordinates and those of the 2D image is the primary goal of an operation called camera calibration [33]. In fact, it involves the image analysis of the projection of a series of characteristic points, which are characteristics inherent in an object whose three-dimensional coordinates are known with great precision [34]. The literature offers several calibration methods such as, calibration pattern, geometric clues and deep learning based, in this study we chose to use a calibration method belonging to the calibration pattern family, because of its high precision and given that it has been implemented by several development environments like Python in its OpenCv library.

The camera calibration algorithm is as:

a. a chessboard of (10.7) squares is used as a calibration pattern, a series of 40 captures of the chessboard were taken under different angles see Figure 4.
b. Detect the chessboard and locate the pixel coordinates of the black boxes' corners in the images.
c. Resolve the intrinsic and external parameters of the camera.

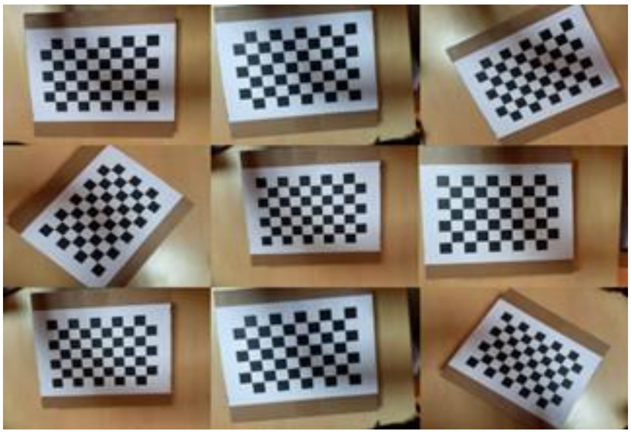

Figure 4. Sample of the chessboard images used in camera calibration

3.4. Estimating gaze point coordinate in the world coordinate system

After the camera calibration step, which allowed us to get the camera Intrinsic and extrinsic parameters, we can now, calculate the coordinates in the real-world coordinates system of the vision vector. Thus, we can determinate the point of intersection between the gaze vector and the plane that contains the display board.

As shown in Figure 1, in our system we assume that the camera is the center O of the real-world coordinate system, the display board is contained in the XOY plane, the Z axis and the optical axis of the camera are the same and that the positive direction is towards the front of the camera. Therefore, the 2D coordinates of the gaze point on the XOY plane can be obtained by switching from the image coordinates system to the real-world coordinates system by checking that Z= 0.

$$Ux = -(xic - cx) * \frac{D}{fx}$$
$$Uy = (yic - cy) * \frac{D}{fy}$$
$$Vx = -s * \sin\theta + ux$$
$$Vy = s * \sin\varphi + vy$$

Where,

$$s = D/\cos\theta \qquad (2)$$

3.5. Student attention quantification

A student during the class session can only be in one of two states: attention and inattention. Thus, the student's attention can be modeled by the following formula: $A(t) \in \{0,1\}$. We consider a student in a state of attention {1,0} according to his gaze point, if his gaze point is in the plane P whose center is the camera, the X and Y axes are respectively the length and width of the display board. In this case, the level of attention is 1, the opposite case 0, where the gaze is completely outside the plane P.

$$\forall Gt(x,y) \in P \; so \; A(t) = 1 \; else \; 0$$

4. EXPERIMENTAL RESULTS

In order to corroborate our point of gaze estimation method, we have developed a test set up. A standalone application that implements the various functionalities of the system has been developed using the Python language and the OpenCV library. It processes images (10 frames/second with a resolution of 1280/720 px) and returns results in real-time without delay on a workstation equipped with an Intel (R) Core (TM) i7-8665U CPU and memory 16 GB RAM.

Three students participated in this experiment. Measurements of the width of the face of each student were made for calibration purposes. The students are seated at a distance of 1.5 m in front of the display board Figure 1. They were asked to point out randomly 9 locations marked on the board with blue dots during 10 seconds. The coordinates of the 9 points on the (X, Y) coordinates system were measured to allow comparison with the estimated gaze points.

The Figure 5 shows the difference between the estimated positions and the actual positions for the 9 points. The gaze point was considered to be the mean of the left and right eye gaze's points coordinates (xi, yi). To simplify the readability of the results, 1/2 the length of the display board has been added to xi in order to obtain positive data. The estimation error was calculated for each axis (X, Y), Euclidean distance was used to calculate the distance between the actual values of the points displayed on the display board and their estimates.

Analysis of the experimental data shows an average error of 4.6 cm and 4.3 cm on the two X, Y axes which largely meets our expectations. Our system is able to find the 2D coordinates of the student's point of gaze on the slideshow plane at a distance of 1.5 m with an acceptable precision. We reiterate the objective of this system is to be able to identify, during the course lecture, the loss of attention of the student, which can be expressed by moving the student's gaze away from the display board or by a stagnant gaze that indicates a thinking state or a having the head elsewhere state.

The accuracy of our method can be comparable to that of Wan *et al*. [35] in terms of distance error on the X and Y axes, and the distance which separates the subject from the projection plane. However, the hardware device they used is more sophisticated than ours. They used a stereo camera and an infrared light source to calculate the cornea center. The use of infrared light can be impractical in outdoor environments or unhealthy under certain conditions [36]. The method of Gutiérrez *et al*. [37] employs an intrusive device, since it is necessary to set up a configuration that blocks the head movement and the subject must wear a pair of glasses on which the camera is placed. This method achieves a good accuracy, but on a very small distance that does not exceed 40 cm.

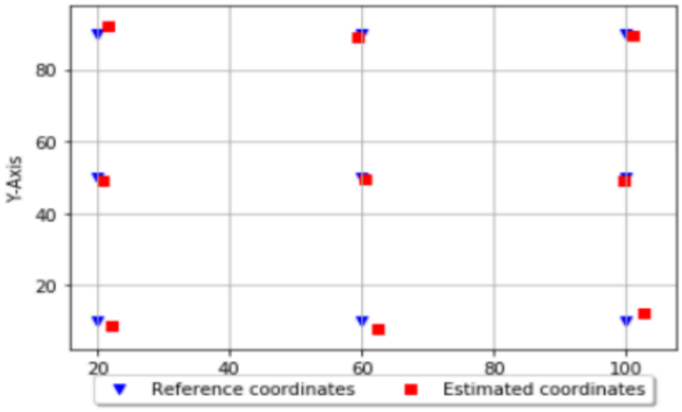

Figure 5. Student's gaze point estimation accuracy

We present, in Table 2, a comparison of these three previously cited methods and ours. As you can see our method achieves considerable accuracy, noting that this precision is achieved without the use of sophisticated equipment or infrared lighting and the experimentation has been realized under real conditions. It should also be noted that, the good precision depends on the camera calibration and of the stage of calculating the distance between the camera and the student.

Table 2. Average error in cm of our method with that of the comparison methods

Method	Materiel	Distance From the camera in cm	Gaze Error X	Gaze Error Y
Mlot *et al*. [37]	A camera and a motorized linear rail system to adjust the distance from the camera.	Variable from 20 to 40	0.29	0.38
Wan *et al*, [35]	Sterio camera and near infrared light	80-400 (150)	0.6	0.6
Mlot *et al*. [37] and Wang *et al*. [38]	camera	86	0.25	0.2
Our proposition	Single camera	150	4.6	4.3

For the rest of the experiment, the participants were invited to follow a 10-minute class session without instructions. A presentation was projected in front of the students at the location defined as plane P. The images of the students were captured with a frequency of 10 frames/second to lighten the calculation. The direction of gaze is estimated for each student and the state of attention is detected. Figure 6 shows 3 minutes of the recorded signal of one participant's detected attention and distraction states.

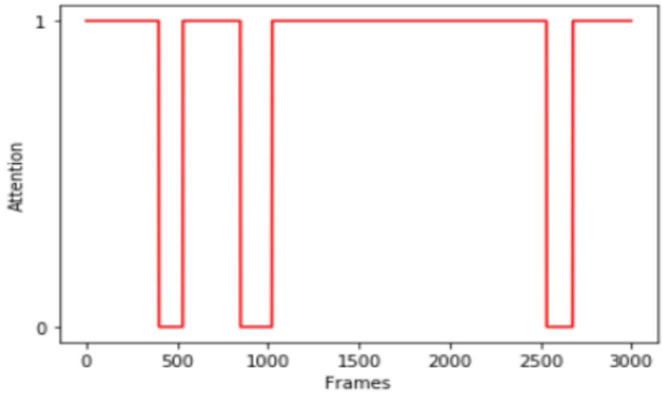

Figure 6. Predicted student's signal attention based on his gaze point

5. CONCLUSION

In this article, we have presented a successful method of tracking student attention during the classroom session by relying on the information the gaze can disclose. Our approach is based on the estimation of the student gaze point on the display board or slideshow to determine in a precise way if the student follows the explanations given by his teacher or is distracted by other elements present in the classroom. In future work, we will try to combine gaze point tracking with the head pose estimation in order to be able to analyze more attention situations such as the case when the student reading his notes or writing them, turn towards a comrade or even look at the ceiling.

REFERENCES

[1] T. Hachad, A. Sadiq, and F. Ghanimi, "A new big data architecture for real-time student attention detection and analysis," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 8, pp. 241–247, 2020, doi: 10.14569/IJACSA.2020.0110831.

[2] T. Hachad, A. Sadiq, F. Ghanimi, and L. Hachad, "A Novel architecture for student's attention detection in classroom based on facial and body expressions," *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 9, no. 5, pp. 7357–7366, Oct. 2020, doi: 10.30534/ijatcse/2020/68952020.

[3] I.-H. Choi, Y. G. Kim, and T. B. H. Tran, "Real-time categorization of driver's gaze zone and head pose -using the convolutional neural network," in *HCI Korea 2016*, Jan. 2016, pp. 417–422, doi: 10.17210/hcik.2016.01.417.

[4] R. Konrad, S. Shrestha, and P. Varma, "Near-eye display gaze tracking via convolutional neural networks," Thesis, Standford Univ., Standford, CA, USA, Tech. Rep, 2016.

[5] A. George and A. Routray, "Real-time eye gaze direction classification using convolutional neural network," in *2016 International Conference on Signal Processing and Communications (SPCOM)*, Jun. 2016, pp. 1–5, doi: 10.1109/SPCOM.2016.7746701.

[6] S. Vora, A. Rangesh, and M. M. Trivedi, "On generalizing driver gaze zone estimation using convolutional neural networks," in *2017 IEEE Intelligent Vehicles Symposium (IV)*, Jun. 2017, pp. 849–854, doi: 10.1109/IVS.2017.7995822.

[7] C. Zhang, R. Yao, and J. Cai, "Efficient eye typing with 9-direction gaze estimation," *Multimedia Tools and Applications*, vol. 77, no. 15, pp. 19679–19696, Aug. 2018, doi: 10.1007/s11042-017-5426-y.

[8] A. Kar and P. Corcoran, "A review and analysis of eye-gaze estimation systems, algorithms and performance evaluation methods in consumer platforms," *IEEE Access*, vol. 5, pp. 16495–16519, 2017, doi: 10.1109/ACCESS.2017.2735633.

[9] D. W. Hansen and Qiang Ji, “In the eye of the beholder: a survey of models for eyes and gaze,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 3, pp. 478–500, Mar. 2010, doi: 10.1109/TPAMI.2009.30.

[10] M. L R D and P. Biswas, “Appearance-based gaze estimation using attention and difference mechanism,” in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Jun. 2021, pp. 3137–3146, doi: 10.1109/CVPRW53098.2021.00351.

[11] 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, pp. 25–51, Apr. 2005, doi: 10.1016/j.cviu.2004.07.011.

[12] A. A. Akinyelu and P. Blignaut, “Convolutional Neural network-based methods for eye gaze estimation: a survey,” *IEEE Access*, vol. 8, pp. 142581–142605, 2020, doi: 10.1109/ACCESS.2020.3013540.

[13] B. A. Smith, Q. Yin, S. K. Feiner, and S. K. Nayar, “Gaze locking,” in *Proceedings of the 26th annual ACM symposium on User interface software and technology*, Oct. 2013, pp. 271–280, doi: 10.1145/2501988.2501994.

[14] J.-B. Huang, Q. Cai, Z. Liu, N. Ahuja, and Z. Zhang, “Towards accurate and robust cross-ratio based gaze trackers through learning from simulation,” in *Proceedings of the Symposium on Eye Tracking Research and Applications*, Mar. 2014, pp. 75–82, doi: 10.1145/2578153.2578162.

[15] F. L. Coutinho and C. H. Morimoto, “Improving head movement tolerance of cross-ratio based eye trackers,” *International Journal of Computer Vision*, vol. 101, no. 3, pp. 459–481, Feb. 2013, doi: 10.1007/s11263-012-0541-8.

[16] K. A. F. Mora, F. Monay, and J.-M. Odobez, “EYEDIAP,” in *Proceedings of the Symposium on Eye Tracking Research and Applications*, Mar. 2014, pp. 255–258, doi: 10.1145/2578153.2578190.

[17] Y. Sugano, Y. Matsushita, and Y. Sato, “Learning-by-Synthesis for appearance-based 3d gaze estimation,” in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2014, pp. 1821–1828, doi: 10.1109/CVPR.2014.235.

[18] X. Zhang, Y. Sugano, M. Fritz, and A. Bulling, “Appearance-based gaze estimation in the wild,” in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2015, vol. 07-12-June, pp. 4511–4520, doi: 10.1109/CVPR.2015.7299081.

[19] K. Krafka *et al.*, “Eye Tracking for Everyone,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2016, vol. 2016-Decem, pp. 2176–2184, doi: 10.1109/CVPR.2016.239.

[20] X. Zhang, Y. Sugano, M. Fritz, and A. Bulling, “It’s written all over your face: full-face appearance-based gaze estimation,” *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, vol. 2017-July, pp. 2299–2308, 2017, doi: 10.1109/CVPRW.2017.284.

[21] T. Fischer, H. J. Chang, and Y. Demiris, “RT-GENE: real-time eye gaze estimation in natural environments,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11214 LNCS, 2018, pp. 339–357.

[22] Y. Cheng, H. Wang, Y. Bao, and F. Lu, “Appearance-based Gaze Estimation with deep learning: a review and benchmark,” *arXiv preprint arXiv:2104.12668*, 2021, [Online]. Available: http://arxiv.org/abs/2104.12668.

[23] J. Kim *et al.*, “NVGaze: An anatomically-informed dataset for low-latency, near-eye gaze estimation,” in *Conference on Human Factors in Computing Systems - Proceedings*, May 2019, pp. 1–12, doi: 10.1145/3290605.3300780.

[24] X. Zhang, S. Park, T. Beeler, D. Bradley, S. Tang, and O. Hilliges, “ETH-XGaze: A large scale dataset for gaze estimation under extreme head pose and gaze variation,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 12350 LNCS, 2020, pp. 365–381.

[25] S. Park, E. Aksan, X. Zhang, and O. Hilliges, “Towards end-to-end video-based eye-tracking,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 12357 LNCS, 2020, pp. 747–763.

[26] A. Kumar, A. Kaur, and M. Kumar, “Face detection techniques: a review,” *Artificial Intelligence Review*, vol. 52, no. 2, pp. 927–948, Aug. 2019, doi: 10.1007/s10462-018-9650-2.

[27] R. Liu *et al.*, “An intriguing failing of convolutional neural networks and the CoordConv solution,” *Advances in Neural Information Processing Systems*, vol. 2018-December, pp. 9605–9616, 2018.

[28] D. E. King, “Dlib-ml: A machine learning toolkit,” *Journal of Machine Learning Research*, vol. 10, pp. 1755–1758, 2009.

[29] S. Park, X. Zhang, A. Bulling, and O. Hilliges, “Learning to find eye region landmarks for remote gaze estimation in unconstrained settings,” in *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications*, Jun. 2018, pp. 1–10, doi: 10.1145/3204493.3204545.

[30] Y. Sun, X. Wang, and X. Tang, “Deep Convolutional network cascade for facial point detection,” in *2013 IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2013, pp. 3476–3483, doi: 10.1109/CVPR.2013.446.

[31] A. Toshev and C. Szegedy, “DeepPose: Human pose estimation via deep neural networks,” *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 1653–1660, 2014, doi: 10.1109/CVPR.2014.214.

[32] A. Newell, K. Yang, and J. Deng, “Stacked hourglass networks for human pose estimation,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9912 LNCS, 2016, pp. 483–499.

[33] G. D’Emilia and D. Di Gasbarro, “Review of techniques for 2D camera calibration suitable for industrial vision systems,” *Journal of Physics: Conference Series*, vol. 841, no. 1, p. 012030, May 2017, doi: 10.1088/1742-6596/841/1/012030.

[34] A. De la Escalera and J. M. Armingol, “Automatic chessboard detection for intrinsic and extrinsic camera parameter calibration,” *Sensors*, vol. 10, no. 3, pp. 2027–2044, Mar. 2010, doi: 10.3390/s100302027.

[35] Z. Wan, X. Wang, L. Yin, and K. Zhou, “A method of free-space point-of-regard estimation based on 3D eye model and stereo vision,” *Applied Sciences*, vol. 8, no. 10, p. 1769, Sep. 2018, doi: 10.3390/app8101769.

[36] F. Mulvey, A. Villanueva, D. Sliney, R. Lange, S. Cotmore, and M. Donegan, *Exploration of safety issues in Eyetracking*. COGAIN EU Network of Excellence, 2008.

[37] E. G. Mlot, H. Bahmani, S. Wahl, and E. Kasneci, “3D Gaze estimation using eye vergence,” in *Proceedings of the 9th International Joint Conference on Biomedical Engineering Systems and Technologies*, 2016, pp. 125–131, doi: 10.5220/0005821201250131.

[38] R. I. Wang, B. Pelfrey, A. T. Duchowski, and D. H. House, “Online 3D gaze localization on stereoscopic displays,” *ACM Transactions on Applied Perception*, vol. 11, no. 1, pp. 1–21, Apr. 2014, doi: 10.1145/2593689.

BIOGRAPHIES OF AUTHORS

Tarik Hachad received received the Bachelor's degree in Computer Science in 2008 and the Master's degree in Big Data Engineering in 2018 at the Faculty of Sciences of Rabat, Mohammed 5 University of Rabat. Currently, he is a Phd candidate in the Computer Science Research laboratory of the Faculty of Sciences, Ibn Tofail University, Kenitra. His research interests include computer vision, deep learning, sentiment analysis, data mining, big data and internet of things. He can be contacted at email: tarik.hachad@uit.ac.ma.

Prof. Dr. Abdelalim Sadiq received a B.S in software engineering from Sciences and technologies Faculty Moulay Ismail University, Errachidia in 1999, DESA degree in computer network and telecommunication from National School of Computer Science and Systems Analysis (ENSIAS) Mohammed V University, Rabat, Morocco in 2002 and Ph.D. degree in Computer Science from ENSIAS, Mohammed V University, Rabat, Morocco, in 2007. He is currently a full professor in computer science Department of Sciences Faculty, Ibn Tofail University, Kenitra, Morocco. His research interests include multimedia information retrieval and processing, sentiments analysis, IoT and data science. He has served as a reviewer for several international conferences and journals. He can be contacted at email: a.sadiq@uit.ac.ma.

Prof. Dr. Fadoua Ghanimi is a Professor of Computer Science at Ibn Tofail University, where he has been since 2000. From 2013 to 2020, she worked at Hassan II University, eventually as a Professor of Computer Science. Her research interests span both Artificial Intelligence and applications. Much of her work has been on improving the understanding, design, and performance of artificial intelligence-based systems, mainly through the application of machine learning, deep learning to NLP and IoT. Professor GHANIMI is co-author and is the author of many papers. He has given numerous invited talks and tutorials. She can be contacted at email: ghanimi_fadoua@yahoo.fr.

Dr. Lamiae Hachad was born in Errachidia, Morocco. After a degree in computer science at FSTE (Faculty of Science and Technologies, Errachidia), and a Master in Intelligent Systems and Networks at FSTF (Faculty of Science and Technology, Fez), she joined the Signals Systems and Components laboratory (FSTF) in December 2012. She defended her doctoral thesis in June 2018. Professor of Higher Education at the EMSI (Moroccan School of Engineering Sciences). Her research interests include: antenna processing, MIMO systems, cognitive radio, next generation networks (4G, 5G, 6G, IoT), and artificial intelligence. She can be contacted at email: lamiae.hachad@usmba.ac.ma.

Ahmed Laguidi received the B.Sc. degree in software engineering from the Faculty of Sciences and Technologies (FSTE), Moulay Ismail University, Errachidia, in 2005, the M.Sc. degree in computer science and networks from the Faculty of Sciences and Technologies (FSTG), Cadi Ayyad University, Marrakech, in 2008 and the Ph.D. degrees in telecommunications and computer networks from the National School of Electricity and Mechanics (ENSEM), Hassan II University, Casablanca, Morocco, in 2019. Currently, he is a Professor at the Department of mathematics and computer science, Hassan II University. His research interests include 5G network, D2D communications, cloud computing, cloud RAN, and the Internet of things. He can be contacted at email: laguidi.ahmed@mail.com.