Journal of Physics: Conference Series

PAPER • **OPEN ACCESS**

The application of image processing for monitoring student's attention level during online class

To cite this article: J W Ong *et al* 2021 *J. Phys.: Conf. Ser.* **2120** 012028

The application of image processing for monitoring student's attention level during online class

J W Ong, W J Chew*, and S K Phang

School of Computer Science and Engineering, Taylor's University, No. 1 Jalan Taylor's, 47500 Subang Jaya, Selangor, Malaysia.

*E-mail: WeiJen.Chew@taylors.edu.my

Abstract. With the COVID-19 pandemic still causing the world to be quarantined in their house to prevent the spread of the virus, this means online classes are still the main method of conducting classes. This project aims to help lecturers monitor the students during class as they are having problems checking whether the students are paying attention or not. This project uses the student's facial features to determine their attention level using two different coding algorithm Viola-Jones and Sobel edge. These two algorithms help to determine what kind of facial expression that the students are making. The Viola-Jones algorithm detects and captures the student's facial features such as eyes and mouth while the Sobel edge algorithm detects the edges of the facial features to determine whether the eyes and mouth are open or closed. With the data collected it will run through the database to determine the student's attention level and inform the lecturer.

1. Introduction

The year 2020 has seen the world fighting against the COVID-19 pandemic. One of the main methods to fight the COVID-19 is to social distance to prevent the spread of the virus. This means people are not allowed to gather in large crowds which results in many activities being stopped at the moment. One such activity is conducting face to face classes in a university. With everyone being quarantined in their homes, this mean that classes will have to be conducted online. These types of classes have been previously explored and the current situation accelerated the need for platforms to help distribute the contents online [1,2]. However, with the changes of physical class to online class, the lecturers are not able to determine whether the students are paying attention to the classes or not.

Therefore, lecturers have to come up with a few methods in order to determine the student's attention level during the online classes. The quick solution is to prompt students during class to check whether they are paying attention or not. The lecturer may call out a student to ask them questions regarding the classes by giving the student a small exercise or quiz to check whether the students are paying attention in class. However, this method is not suitable when used in a large class as it is impossible to question every student and it may take a long time for the students to answer, which will slow down the pace of the class. Another method that the lecturers can use to check a student's attention is to ask the students to turn on their video camera to enable the lecturer to monitor the students during classes. However, even with the camera turned on, it is still difficult to determine whether the students are paying attention or not. This is because a student is able to minimize the online class program and do something else with the computer while the class is still ongoing.

Therefore, multiple studies that uses camera-based tools have been conducted to help determine student's attention level during online classes. One of the method uses an electroencephalogram (EEG) meter to determine a student's attention level [3]. With the EEG meter equipped on the user, the camera will take a picture of the user and let the picture be examined by the algorithm in the system to determine the attention level. After the system have determine the attention level, the system will report the attention level and the cycle continues. Another method suggested using electrocardiography (ECG) and skin conductance response (SCR) to determine human emotions [4]. However, both methods are not suitable to be used in an online classroom setting since each student has to acquire an EEG or ECG meter in order for the data to be captured and examined by the algorithm.

The goal of this project is to determine how facial expression and eye movement can be used to determine the attention level of the students. Questions like which expression indicates that the student is having low attention level, which position of the eye determines the attention level or how the lighting in the student's room will affect the accuracy of the data collection will be explored. The main objective is to figure out how to determine the student's attention level by reading the facial expression and position of the eyes. Most research papers are either using only facial expression to determine the student's attention level or by only using the eye movement of the students to determine their attention levels. Therefore, the objective of this project is to create an algorithm to determine the attention levels of the students using both facial expression and eye movement. The algorithm will be able to be implemented in other third-party software to enable the algorithm to be more accessible for other users.

This project is intended to be used during an online class session where each student's camera will be turned on. Therefore, the proposed algorithm will monitor the student through their webcam image, which consists of a front facing head. It is assumed that each input video will only have one person in it with a relatively bland background. The facial expression and eye movement of that one person will be monitored throughout the class session.

2. Literature Review

In the year 2020, facial detection systems are one of the most widely used system in modern technology to improve the lifestyle of most people. The facial detection system is commonly used as a security measure, with example such as the facial lock that is used in smartphones. Therefore, currently there are some studies that focuses on facial detection to determine a person's attention level during a meeting as the current COVID-19 pandemic has forced all meetings and classes to be conducted online. One of the system that is being used to determine a person's attention level is the VADS system. The VADS is a system that collects the 3D geometric model of a person using the smart-phone camera.

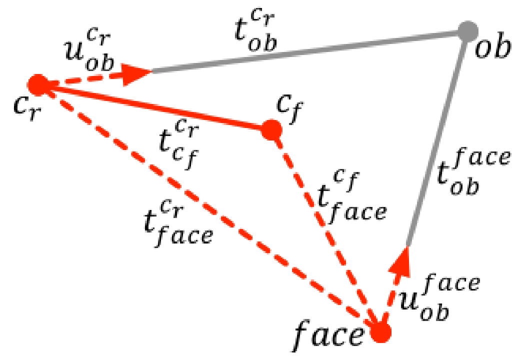

Figure 1. The angle calculation using VADS and smart-phone camera [5]

Table 1. Definition of variables in VADS calculation [5]

Symbol	Description
$navi$	The local-navigational frames
c_r, c_f	The rear and front camera-centered coordinate system (frame)
$face$	The face coordinate system (frame)
ob	The intended object
t_p^{fx}	The position (vector) to p in fx-frame
u_p^{fx}	The unitary directional vector towards p in the fx-frame
R_{fo}^{fn}	The rotation matrix which transform the coordinate system from the fo-frame to fn-frame

The smartphone have been implemented with an advance facial detection algorithm to calculate the distance and angle using the smart-phone front camera while the VADS will determine the direction of the users gazes. Once all the parameters are collected, the VADS will form a virtual binocular vision to locate the direction of the eyes and determine the user's attention level on that specific object, as shown in Figure 1 and Table 1. The main issue in this study is obtaining an accurate 3D model for the facial features as the smartphone will have difficulties in obtaining high quality images. Also, in order for the VADS to create the 3D model, the user's face must be directly facing the camera [5].

Another study determines the attention level by reading the facial and body features by using a kinect one sensor. In this experiment, the student's behavior will be recorded during the lecture. The recording will take place in a physical classroom in order for the kinect one sensor to collect the body movement of the students. The data is collected using MATLAB scripts provided from the Kin2 Toolbox. From the data, five parameters will be monitored in order to determine the attention level of the student. These parameters are Writing Wu(t), Yawning Yu(t), Supporting head Su(t), Leaning back Lbu(t) and Person's gaze Gu(t) .

$$A_w(l) = \frac{1}{N_u * N_s} \sum_{u=1}^{N_u} \sum_{t=1}^{N_s} W_u(t) | A_u^r(t) = l \quad (1)$$

Once all the five parameters are collected, the script will apply it in the equation shown in Equation 1 where high attention level is associated with writing and body leaning forward, medium attention level is associated with body leaning forward, head supported by hand and observing the slides, low level attention is associated with leaning backwards, rubbing the neck, scratching their head, yawning and looking away. The flaw for this study is that all the movement will not be able to be implemented in the proposed project as the classroom will be online. It will be difficult to record the body of the students as every student will need to own a kinect one senor for the data collection [6]

Another method to determine a person's attention level is by the facial expression that the users are expressing. In this experiment, the camera will first determine the user's gender and collect the user in a neutral expression. Once the gender is determined, the system is fed with a set of data in order to determine which expression the user is showing, as shown in Figure 2.

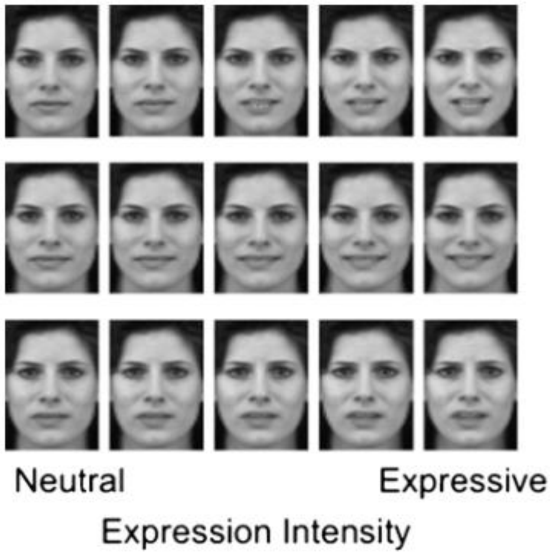

Figure 2. Different expressions of participant [7]

With the data collected, it can conclude that when the users express an angry expression, this shows that the attention level is swaying slightly while a user expressing happiness has the least attention being swayed. However, the data also shows that users that have a fearful expression has the highest attention being swayed. This proves that the expression can be used to determine the attention level of the users. The issue with this method in an online class setting is that most students are just paying attention to the classes, therefore the students generally would not display any emotion during the class [7].

Another study that uses facial expression to determine how much attention were lost is using Ebbinghaus as a measuring tool to understand how the facial expression will work alone with the size perception to determine the attention level of the users. In this experiment, all participants are female undergraduates and undergo the SADS (Social Avoidance and Distress Scale) to measure the participant's anxiety level during a social interaction. This is to determine the mental state of the participants to maintain consistency of the results. Using the Ebbinghaus tool to capture the faces of the participants in a quiet laboratory, the results show that in a short period condition, the participants that perceive a large angry face have a higher rate of attention being swayed while the one that perceive a large happy face will have a lower rate of attention being swayed. However, they were unable to collect any solid data that determine whether the large neutral face have any huge effect on the participants attention level [8].

With all the drawbacks presented in the studies above, this projects aim is to create an improved algorithm using both facial expression and eye movement to determine a student's attention level.

3. Research Methodology

With the information gathered from multiple studies, the first step will be to capture the faces of the students during the online classes. To solve the issue of knowing a student's attention level, this project will be using MATLAB software to create an algorithm with the function of facial recognition and attention level detection. Figure 3 shows the flowchart of the proposed system.

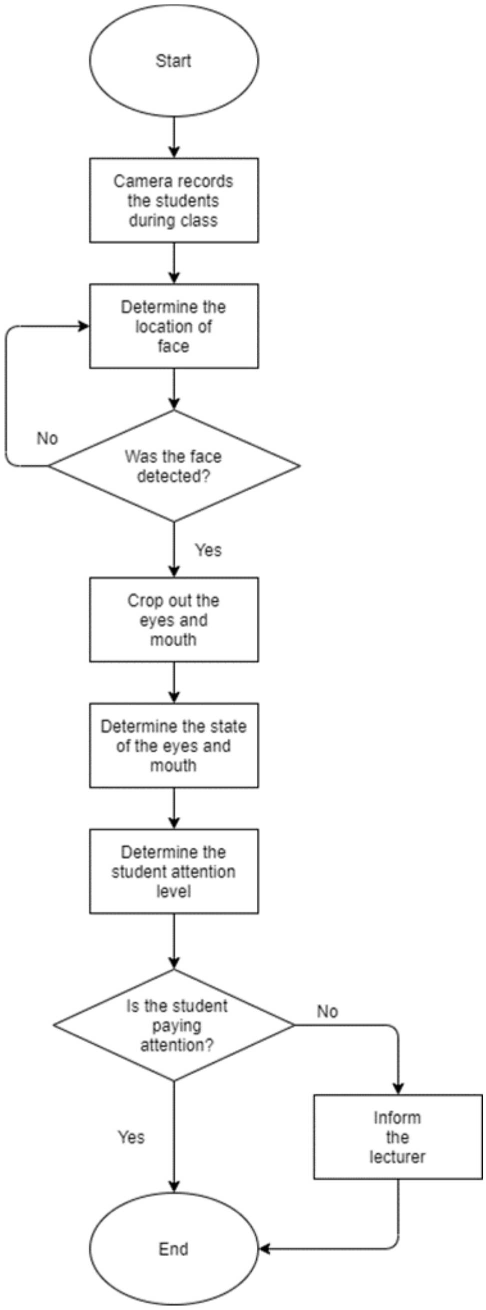

Figure 3. Flow chart of the complete system

3.1 Viola-Jones algorithm.
The Viola-Jones face detection algorithm is the base algorithm for this project.

$$\sum_{1\le i\le N}\sum_{1\le j\le N} I(i,j)1_{P(i,j)} - \sum_{1\le i\le N}\sum_{1\le j\le N} I(i,j)1_{P(i,j)} \quad (2)$$

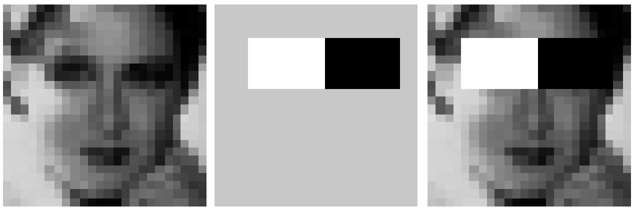

Figure 4. Examples of using Haar-like features [9]

Viola-Jones algorithm uses a technique called the Haar-like features to determine the position of each facial features. Equation 2 shows the formula of how the Haar-like features determine the position of the face by letting *I* and *P* be the variable of the image and pattern while the letting *N* be the size of the image. The shape of the image must be a square to implement the Haar-like features. Figure 4 shows the photos to be in black and white. This is to compensate for the lighting effect on the photo allowing the algorithm to determine the face. The black and white box are placed to help segregate the different parts of the face and determine the position of the facial parts such as eyes, nose, mouth and cheeks.

Figure 5. Different sizes and shapes in the Viola-Jones algorithm [9]

As shown in Figure 5, there are multiple size and shapes for the Viola-Jones algorithm to detect the facial features. It detects the shade values and the total amount of pixels in the image to create the correct sized boxes to use in order to capture the facial features.

3.2 Hough transform algorithm.

The second algorithm, Hough transform algorithm, is used to determine the eye movement. The main function of the Hough transform is using Canny detector to convert the edges of an image to lines in order to detect the shapes of the content in the image. The algorithm works by first building a parameter space for the line slope of M and intercept C and create an array to accumulate the M and C values. Next, the Canny detector will find and extract the edges of the image to determine the length of the slope. Once the length of the slope is determined, the data gathered will be applied in the Hough transform algorithm to determine the shape of the image. After extracting the edges of the image using Canny detector, a 3D accumulator array for the x, y and radius values is created [10].

3.3 Sobel Edge algorithm

Sobel edge algorithm is a detection algorithm that takes the input picture and converts it into multiple edges in order to determine whether the eyes and mouth are open or closed.

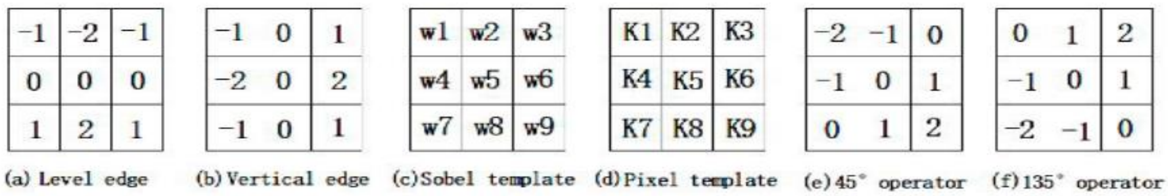

Figure 6. Sobel edge 3*3 detection template [11]

$$G_x = f_x(x,y) = f(x-1,y+1) + 2f(x,y+1) = f(x+1,y+1) - f(x-1,y-1) - 2f(x,y-1) - f(x+1,y-1) \quad (3)$$

$$G_y = f_y(x,y) = f(x+1,y-1) + 2f(x+1,y) + f(x+1,y+1) - f(x-1,y-1) - 2f(x-1,y) - f(x-1,y+1) \quad (4)$$

$$\mathrm{G} = \sqrt{\mathrm{G}_x^2 + \mathrm{G}_y^2} \quad (5)$$

Sobel edge detects the 3x3 pixel area and compares the preset threshold values to determine the gradient values shown in Figure 6 to detect the vertical, horizontal and angle of the picture. By comparing the values with the preset threshold in the algorithm, the gradient values can be calculated. With the values determined as segregated, the algorithm will take the value and calculate the gradient of the pixels shown in Equation 3 to 5 [11].

4 Results and Discussion

The proposed method was tested by using a webcam to capture a student during an online class and have the algorithm extract out the frames from the webcam for testing. Each frame was input into the algorithms to determine the student's attention level.

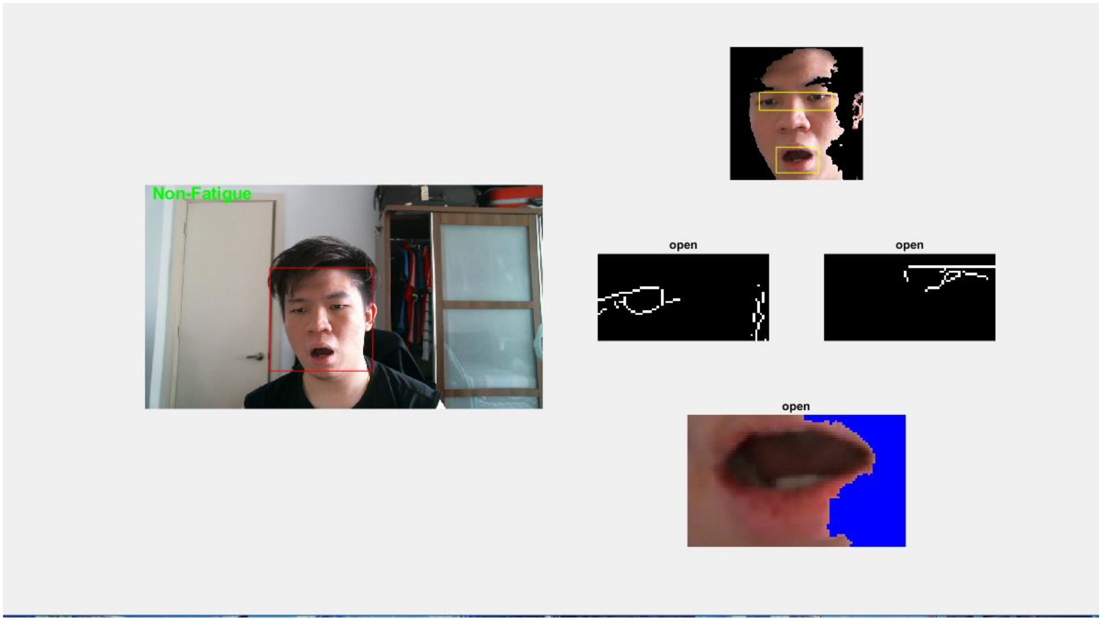

Figure 7. Results using webcam with no error

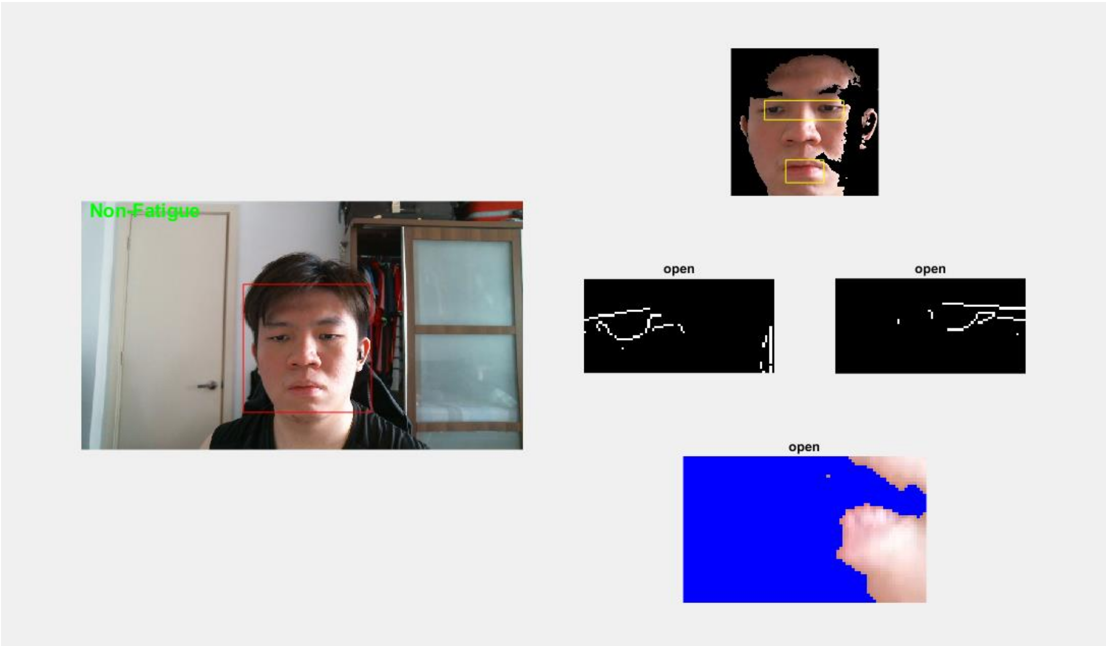

Figure 8. Results using webcam with errors

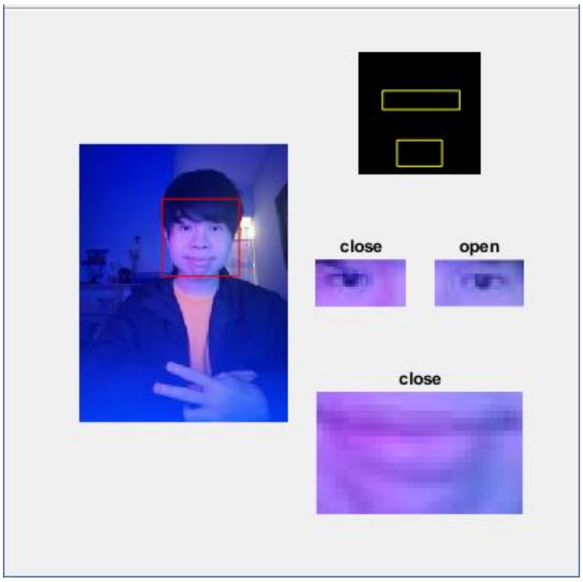

Figure 9. Results in a dark lighting condition

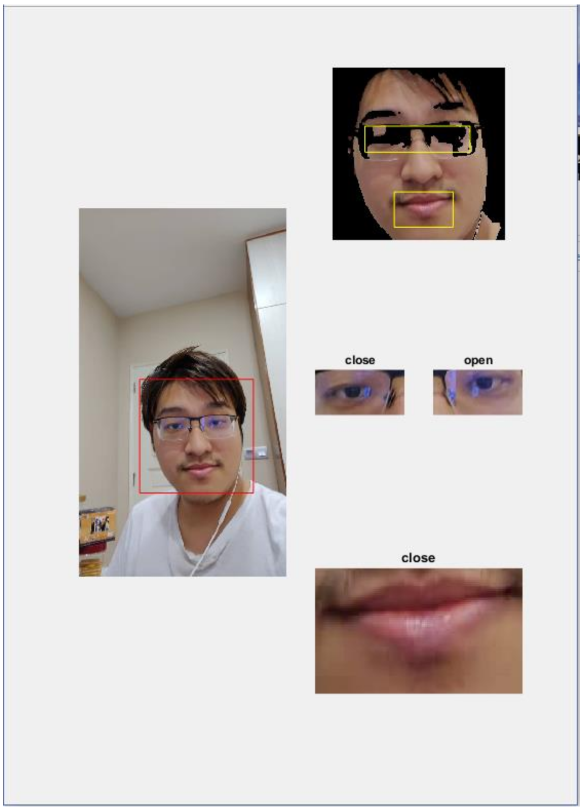

Figure 10. Results with the student wearing glasses

The Viola-Jones algorithm was used to detect the face of the student. The code uses a build in function called vision. Cascade Object Detector to detect the face using the webcam. In the code, this function was used three times. First, it was used to crop out the face from the full image and the other two was to detects the eyes and mouth. With all the facial features detected and categorized, next, the Sobel edge algorithm was used to determine whether the eyes and mouth were open or closed. After converting the different facial features to Sobel edge forms, the code compares it with a pre-build database that contains multiple samples of the open/closed eyes and mouth to determine whether the eyes and mouth are closed or open. Once the results were determined, each part of the results were split into three sections. The results will be used to create vector values of the eyes and mouth to conclude whether the student is paying attention or not.

Figure 7 and Figure 8 shows the results of the code in different condition to check the accuracy of the facial detection algorithm using images taken from the webcam while Figure 9 and Figure 10 shows the results using pre-taken photos of students. Figure 7 shows the code was able to detect the facial features and was able to determine the state of the eyes and mouth with no errors. However, sometimes some errors do occur. For example, Figure 8 indicates the mouth is open even though in the image the mouth is closed. Another error is shown in Figure 9 where it was tested in a dark lighting environment. The error is the incorrect statement of the eyes since the student's right eye is open but it is indicated as closed. Results from Figure 10 is tested with a student wearing glasses and have similar error with Figure 9. When the student's right eye is open, the code still indicates that the eye is closed. This error could be caused by the sample size being too small to train the Sobel edge detection or the Sobel edge algorithm is reading the eyes incorrectly as the eyes does not have the same shapes.

5 Conclusion

The results obtained shows that the Viola-Jones algorithm was working as intended and was able to extract the face, eyes and mouth. However, there were some inconsistencies using the Sobel edge algorithm to determine whether the eyes are close or open.

With the current technologies available, this system can be further improved in future studies. This product can be used for healthcare purposes such as online therapy sessions or online consultation sessions. Companies can also use this product in their online training sessions since the COVID-19 pandemic is still ongoing. This can be used to keep track of the employees so that they are paying attention during the training.

References

[1] C. M. K. Chew, F. L. Ooi, M. Al-Atabi, S. Namasivayam 2013 Innovative teaching and learning tools for foundation in engineering education. *Journal of Engineering Science and Technology* vol 8, pp. 93-104.
[2] L. P. Hasugian, D. Rusmana 2021 Development of UX Orbit design online courses platform in the field of user interface (UI) and user experience (UX). *Journal of Engineering Science and Technology* vol. 16, no. 3, pp. 2454-2463.
[3] C. H. Tseng and Y. H. Chen 2018 A camera-based attention level assessment tool designed for classroom usage. *J. Supercomput.* vol. 74, no. 11, pp. 5889–5902.
[4] K. N. Minhad, S. Ali, M. I. Reaz 2017 A design framework for human emotion recognition using electrocardiogram and skin conductance response signals. *Journal of Engineering Science and Technology* vol. 12, no. 11, pp. 3102-3119.
[5] Z. Jiang et al. 2016 VADS: Visual attention detection with a smartphone. *IEEE INFOCOM 2016 - The 35th Annual IEEE International Conference on Computer Communications* pp. 1-9.
[6] J. Zaletelj and A. Košir 2017 Predicting students' attention in the classroom from Kinect facial and body features. *Eurasip J. Image Video Process. 2017* vol. 80.
[7] J. Gómez-Cuerva and J. E. Raymond 2011 Perception of facial expression depends on prior attention. *Psychon. Bull. Rev.* vol. 18, no. 6, pp. 1057–1063.
[8] J. W. Choi, K. Kim, and J. H. Lee 2016 The influence of attention toward facial expressions on size perception. *Psychol. Res.* vol. 80, no. 1, pp. 69–75.
[9] Y. Q. Wang 2014 An Analysis of the Viola-Jones Face Detection Algorithm. *Image Process. Line* vol. 4, pp. 128–148.
[10] A. S. Hassanein, S. Mohammad, M. Sameer, and M. E. Ragab 2015 A Survey on Hough Transform, Theory, Techniques and Applications. *CoRR abs/1502.02160*
[11] K. Zhang, Y. Zhang, P. Wang, Y. Tian, and J. Yang 2018 An improved Sobel edge algorithm and FPGA implementation. *Procedia Comput. Sci.* vol. 131, pp. 243–248.