This is the Pre-Published Version.

The following publication Y. Liu, L. F. Cheung, W. L. Lam and H. C. B. Chan, "Detection of Online Student Behavior Using Emotion and Eye/Head Movement," 2022 IEEE International Conference on Teaching, Assessment and Learning for Engineering (TALE), Hung Hom, Hong Kong, 2022, pp. 264-269 is available at https://doi.org/10.1109/TALE54877.2022.00051.

Detection of Online Student Behavior Using Emotion and Eye/Head Movement

Yunfei Liu, Long Fai Cheung, Wa Lap Lam and Henry C. B. Chan
Department of Computing
The Hong Kong Polytechnic University
Hong Kong
yunfei.liu@connect.polyu.hk

***Abstract*—During the COVID-19 pandemic of the past few years, online/hybrid teaching has been used around the world, posing challenges for teachers and students alike. One challenge is related to monitoring online student behavior. Facial recognition technologies offer a promising solution, providing useful references for teachers. In this paper, we present our initial work on using emotion, and eye and head movement to detect online student behavior. In particular, we study how these methods can be used to detect five common classroom behaviors: reading slides, writing notes, thinking, checking phones, and engaging in classroom activities, through test cases with the aim of identifying key characteristics. By using the aforementioned methods collectively, more accurate detection results can be achieved. The findings (e.g., key characteristics) should provide valuable insights into understanding online student behavior, and future machine learning work in particular.**

***Index Terms*—online learning, hybrid teaching, facial recognition, classroom behavior**

I. Introduction

During the COVID-19 pandemic, global universities have been using online/hybrid teaching instead of traditional classroom teaching. Teachers use video conference software to conduct online lectures. This new teaching/learning environment brings challenges for both teachers and students. According to Jeffery and Bauer [1], concentration is a key issue in online lectures, affecting student learning. For example, attending an online class may be accompanied by many distractions. Based on a study by Bunce et al. [2], it was found that student concentration time should typically be kept within 9.5-12 minutes, and the attention span for online classes may be further reduced. According to Bolkan and Goodboy [3], there is a strong correlation between adequate teaching and high levels of student concentration. However, it is challenging to evaluate student understanding and attentiveness in online classes due to various factors (e.g., delay and remoteness). In an online environment, teachers can only see students through a web camera, making it more difficult to observe the overall class reactions and performance, as in a traditional class. In an online lecture, when student concentration levels decrease, learning effectiveness and efficiency will be influenced. Thus, teachers should check/monitor the overall attentiveness during online lectures. Furthermore, concentration levels may also be linked to lecture content. If many students lose concentration, the teaching materials presented at that time may need to be enhanced. Student concentration level reports help teachers identify which part of their lecture caused the problem. As a result, student learning effectiveness can also be improved.

With the advent of Artificial Intelligence (AI), facial recognition can be used to predict human behavior and emotions through facial expressions and movements. In the context of this paper, AI can be used to better understand student behavior. Here are some examples of related works. Emotional analysis can be used to check student concentration levels. Based on a study by Kinnunen [4], negative emotions will reduce self-efficacy and learning motivation. Therefore, it is desirable to monitor students' positive emotions. Panahi and Duraisamy [5] studied six types of emotions: anger, fear, happiness, neutral, sadness and surprise, to analyze student engagement. Similarly, another study by Meriem et al. [6] computed a concentration index based on facial emotions. The computation is performed according to the weighted average of the probabilities of different emotions. Sharma et al. [7] performed an emotion assessment to evaluate concentration levels using Microsoft Azure – Emotion API. In many studies, it was found that student emotions in class remain neutral most of the time. Hence, more effective emotion detection methods based on the minority emotion states are required. Zakka and Vadapalli [8] proposed considering emotional states with the top two highest probabilities for better analysis. Furthermore, the low-value emotional states can be measured with two new coefficients, called arousal coefficient and valence coefficient, to evaluate student concentration levels, regardless of the neutral emotion [9]. Eye and head features are other crucial indicators that can analyze student concentration levels in real-time. Alrawahneh and Safei [10] used eye tracking and head pose detection to evaluate concentration levels by using the OpenCV library. Cha and Kim [11] performed concentration analysis based on face shape (i.e., considering facial length, centre of the face, and vertical width of the open eyes). Krithika and Lakshmi Priya [12] used head rotation and eye movement to categorize concentration levels. Horvat and Jagušt [13] evaluated student attentiveness in real time through emotional intensity and eye gaze estimation. Combining eye, head, and mouth characteristics can lead to more accurate results in general (e.g., detecting whether a student is sleepy). Roy et al. [14] found that the eye aspect ratio (EAR) and mouth aspect ratio (MAR) can categorize whether a student is yawning or sleeping. In a related paper, Maheswari et al. [15] used the combination of face aspect ratio (FAR), EAR, MAR, and certain head and hand gestures to achieve comprehensive detection. Although the scenario was not education-related, the methodology can be applied in a classroom environment. In addition to facial features, behavioral data can also be used to evaluate class attentiveness. Xie and Cao [16] studied three categories of in-class behaviors, including organized actions in the classroom, teacher/student speech interaction, and interactive class activities. Based on the study by Yang et al. [17], looking at mobile phones, drinking, studying, writing and hand movements are common in-class behaviors. Li et al. [18] broadened the study by considering sitting postures and head movements to include more in-class behaviors. Apart from using APIs, machine learning algorithms can be used to develop facial recognition models. For instance, Linear Regression, Naïve Bayes, Support Vector Machine, K-means and K-Nearest Neighbors with Ensemble Classifiers are effective and accurate algorithms for image classification purposes [19].

In general, existing works typically focus on feature extraction to analyze student concentration levels using facial and behavioral data. Adopting an integrated or hybrid approach can enhance performance. Relatively little work has been done in this area. Inspired by the above related works, this short paper aims to present our initial work on detecting students' facial expressions and behavior through emotions and movements in eye and head. Compared to the previous works, our focus is to study three approaches together (i.e., based on emotions, and eye and head movement), identify key characteristics, and evaluate their effectiveness in monitoring several common student activities. The current study demonstrates the feasibility of the aforementioned approaches and provides the basis for our future study. In our future work, machine learning will be incorporated to analyze data collected from a hybrid learning environment.

The remaining sections of the paper are organized as follows. Section II presents the methodology and implementation. Section III discusses the test cases and results. Section IV presents the conclusion.

II. Methodology and Implementation

Fig. 1 shows the methodology. Video frames of online students are taken (e.g., from recorded videos or streaming videos). Based on Microsoft Azure Face Application Programming Interface (API), emotional state, eye and head movements can be determined. The API provides the emotion index in a JSON file as well as in various facial landmark coordinates. Based on the landmark coordinates, eye movements can be estimated. Furthermore, the API provides angle-based data for estimating head movement. Details are discussed in the following subsections.

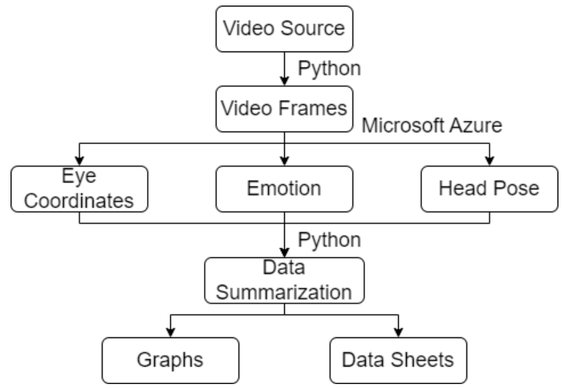

Fig. 1. Methodology

A. Emotional Analysis

Emotional analysis can be used to evaluate student involvement or response in a class. For example, if a teacher finds students in general are feeling negative emotions, the teaching may need to be adjusted. The API detects and analyzes eight types of emotions: anger, contempt, disgust, fear, happiness, neutral, sadness, and surprise. After the analysis, it returns the analysis result through a JSON file. The JSON file contains the probabilities or confidence scores of the eight types of emotions. If an emotion has a high confidence score, it means that the recognized face is more likely to be associated with that emotion.

B. Eye Movement Tracking

Eye movement can be used to check student concentration and other behaviors. For example, if a student is concentrating, he/she is more likely to focus on the screen. However, for a student doing something else such as checking his/her mobile phone, the eyes will have more irregular movements. To facilitate eye movement tracking, we use the Microsoft Azure API which allows real-time facial recognition for pictures. The API provides 10 facial landmark points that are relevant to the eyes, including pupil, inner and outer point, top and bottom point, which are shown as (x_1, y_1), (x_4, y_4), (x_2, y_2), (x_3, y_3) and (x_5, y_5) in Fig. 2, respectively for the left eye. Similar points are also defined for the right eye.

To facilitate the above task, we introduce two indexes: horizontal index and vertical index. These two indexes can show the relative movement of the pupil of an eye.

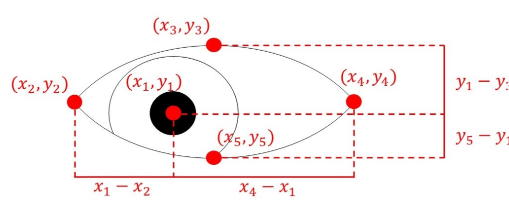

Fig. 2. Coordinates for the Left Eye

The horizontal index of the left eye is defined as

$$H.I. = \frac{(x_1 - x_2) - (x_4 - x_1)}{(x_1 - x_2) + (x_4 - x_1)} \quad (1)$$

And the vertical index is defined as

$$V.I. = \frac{(y_5 - y_1) - (y_1 - y_3)}{(y_1 - y_3) + (y_5 - y_1)} \tag{2}$$

For the right eye, similar indexes can be defined. The overall horizontal/vertical index can then be determined by computing the average of the horizontal/vertical indexes of the left eye and right eye. According to the definition of horizontal/vertical indexes, when the pupil moves to the left, the horizontal index will be close to -1. When the pupil moves to the right, the horizontal index will be close to 1. The vertical index is similar to the horizontal index. When the pupil moves up and down, the vertical index will be close to 1 and -1, respectively. When the pupil is precisely in the middle of the eye, both the horizontal and vertical indexes will be 0.

C. Head Pose Detection

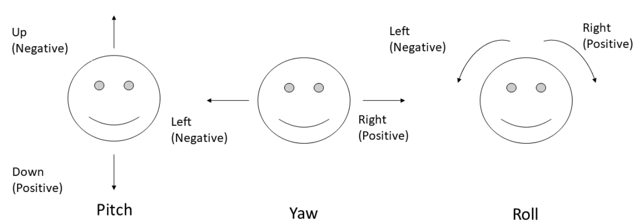

Fig. 3. Head Pose Estimation – Yaw and Pitch Angles

The head pose is a three-dimensional attribute provided by the API, which covers head movement based on three angles: yaw angle, pitch angle and roll angle, corresponding to moving left/right, moving bottom/above and tilting the head to the left/right, respectively. The yaw and roll angles range between –180 and 180 degrees while the pitch angle ranges between –90 and 90 degrees. Fig. 3 shows the yaw, pitch and roll angles. Therefore, the head pose attribute is useful for detecting whether a student is looking at the center of the screen. When a student is looking at the center (e.g., the screen of an online class), the angle should be close to zero. Hence, how long the angle remains at zero can be an indication of a student's concentration level.

Fig. 4 shows the system implementation. There are six key modules. First, each video frame from a recorded or streaming video is captured by the video interface module. Working in conjunction with the API interface module, the video processing module performs basic facial recognition tasks. The API interface module communicates with the API to support the tasks. The outputs are further processed by the emotion processing, eye movement and head movement modules. For example, for eye movement, the horizontal/vertical indexes are computed based on the facial landmark coordinates. Finally, the results (i.e., emotions, eye and head movement) can be displayed in graphs and stored in data files for further analysis.

III. Test Cases and Discussion

To conduct an initial evaluation of the above methods, we have evaluated five test cases using simulated videos (15

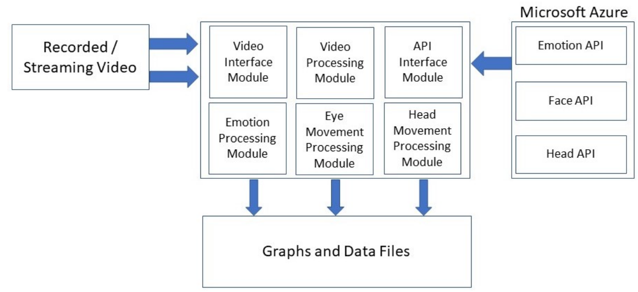

Fig. 4. System Implementation

seconds each) of a person representing different behaviors, including reading slides, taking notes, thinking, checking phones and engaging in classroom activities. Here, the cases/results are presented. The aim is to investigate the effectiveness of the aforementioned methods and discover general behaviors (i.e., study the relevant characteristics and how to differentiate activities based on emotions, eye and head movement). It provides the foundation for our future work. Our long-term goal is to use them in conjunction with machine learning. The test cases/results are discussed in the following subsections.

A. Test Case 1: Reading Slides

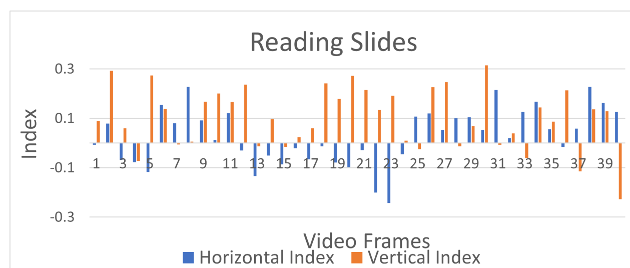

Fig. 5. Horizontal/Vertical Indexes – Reading Slides

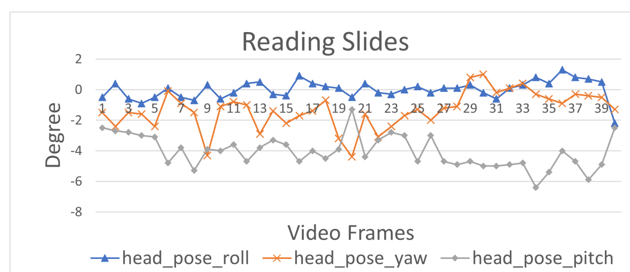

Fig. 6. Head Pose – Reading Slides

Fig. 5 shows the eye movement when a student is reading slides. It can be seen that the vertical index is mostly positive. Also, it is found that the pupils tend to stay in the upper part of the eyes. This is related to the video camera position. For the horizontal index, there are more variations with both positive and negative values, indicating that the eyes move left and right. Fig. 6 shows the head movement. It can be seen that the yaw and pitch angles of the head pose have small changes

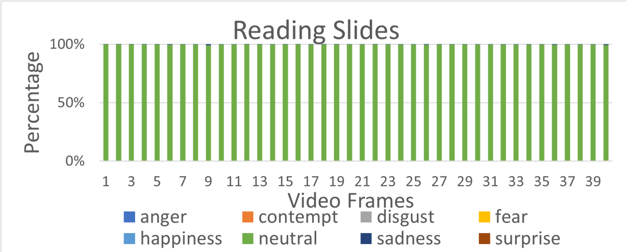

Fig. 7. Emotions - Reading Slides

(i.e., a few degrees). Also, both angles show a similar trend pattern. This indicates that in this case, there are slight changes in head movement as expected. Fig. 7 shows that the emotion is almost 100 percent neutral.

B. Test Case 2: Writing Notes

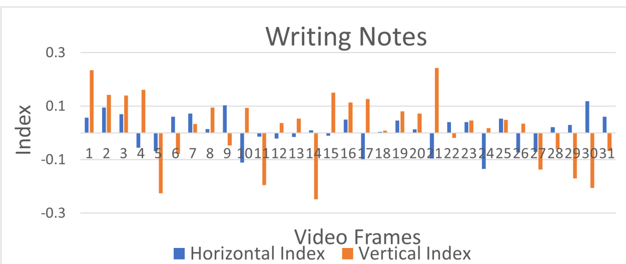

Fig. 8. Horizontal/Vertical Indexes – Writing Notes

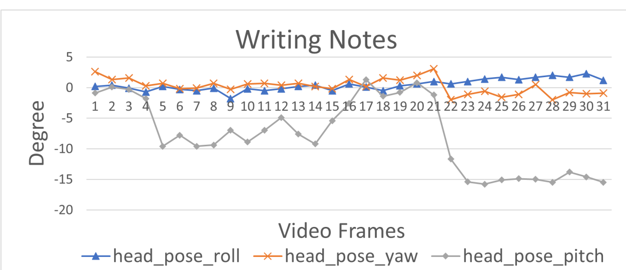

Fig. 9. Head Pose – Writing Notes

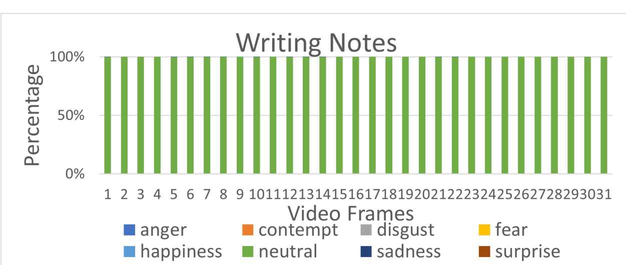

Fig. 10. Emotions - Writing Notes

Fig. 8 shows the eye movement when a student is taking notes. Compared to reading slides, there are more balanced positive and negative values for the vertical index. This is because of significant head movement as shown in Fig. 9 (i.e., the head moves up/down, with more downward movement as expected). This also causes variation in the vertical index. For the horizontal index, there is less fluctuation. Fig. 10 shows the emotion is almost 100 percent neutral.

C. Test Case 3: Thinking

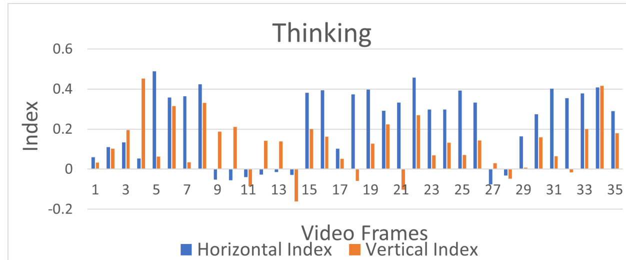

Fig. 11. Horizontal/Vertical Indexes – Thinking

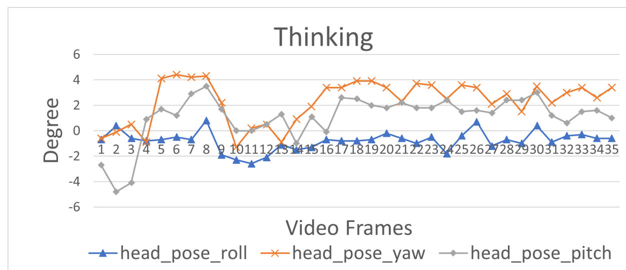

Fig. 12. Head Pose – Thinking

Fig. 11 shows the eye movement when a student is thinking. In this case, the pupils become stationary (i.e., almost no movement), resulting in steady horizontal/vertical indexes (e.g., nearly all positive values in the example). Fig. 12 shows the head movement for thinking. It can be seen that the yaw and pitch angles of the head pose show mostly positive values with an irregular trend. This indicates that in this case, the head is moving slightly. Fig. 13 shows the emotion is almost 100 percent neutral.

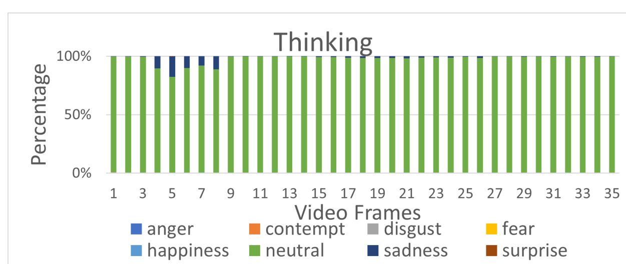

Fig. 13. Emotions - Thinking

D. Test Case 4: Checking Phones

Fig. 14 shows the eye movement when a student is continuously checking his/her phone, with a mostly positive vertical index. Note that although the eyes look down, the head also

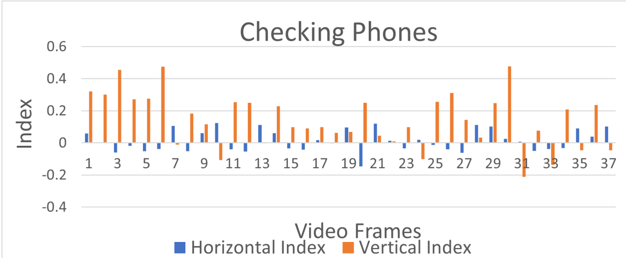

Fig. 14. Horizontal/Vertical Indexes – Checking Phones

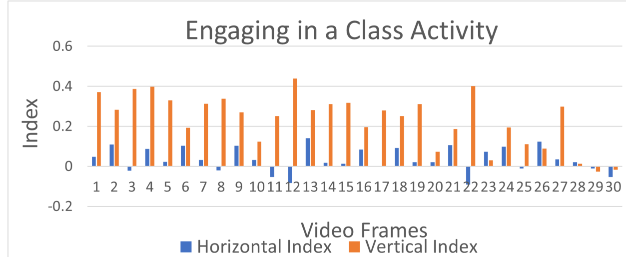

Fig. 17. Horizontal/Vertical Indexes – Engaging in a Class Activity

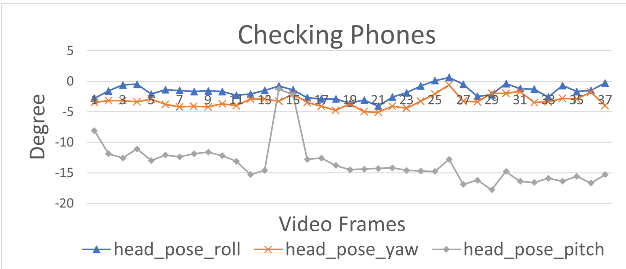

Fig. 15. Head Pose – Checking Phones

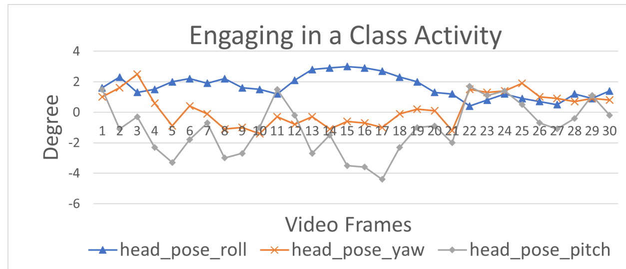

Fig. 18. Head Pose – Engaging in a Class Activity

moves down significantly, resulting in a positive horizontal index (i.e., relatively, the pupils stay within the upper part of the eyes). Fig. 15 shows the head movement. It can be seen that the pitch angle of the head pose indicates a large negative value most of the time. This means that the head mostly stays downward as expected. While it is similar to taking notes, the time periods are different (i.e., checking phones tend to have a longer downward period). Fig. 16 shows that the emotion is almost 100 percent neutral. However, this may depend on the phone content or messages.

E. Test Case 5: Engaging in a Class Activity

Fig. 17 and Fig. 18 show the eye and head movement, respectively when a student is engaging in an in-class activity (e.g., answering an online question). The graphs indicate that the pupils are stationary (i.e., steady horizontal/vertical index) and there is little head movement (i.e., small degree). Fig. 19 shows the emotions. Unlike the previous test cases, there are more mixed emotions (i.e., instead of mostly neutral).

F. Test Case Summary

Table I summarizes the key characteristics of the five cases. The table indicates it is more difficult to detect activities based on one single method (i.e., based on emotion, eye movement or head movement). As some activities produce similar results, teachers could consider using all methods, to distinguish them. In other words, a hybrid detection method should perform better. For example, case 2 (writing notes, positive engagement) and case 4 (checking phones, negative engagement) have similar emotions and head poses. By considering eye movement as well, they can be distinguished more clearly.

IV. Conclusion

In conclusion, motivated by online/hybrid teaching amid COVID-19, we have presented initial work on detecting online student behavior based on emotion, eye and head movement. Microsoft Azure API is used to provide emotion analysis and facial landmark data for further analysis. Table I summarizes the test results or key characteristics for five test cases (i.e., covering common class activities). By considering the methods collectively, more accurate detection can be achieved. The

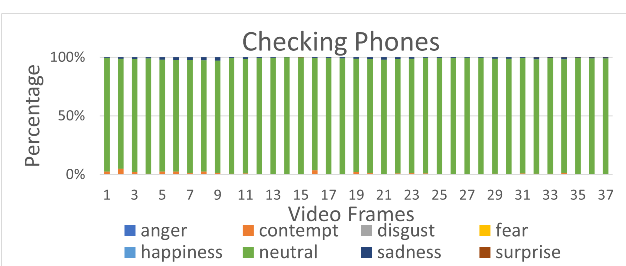

Fig. 16. Emotions - Checking Phones

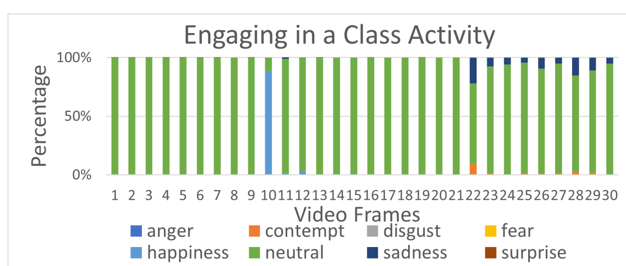

Fig. 19. Emotions - Engaging in a Class Activity

TABLE I
TEST CASES COMPARISON: (1) READING SLIDES (2) WRITING NOTES (3) THINKING (4) CHECKING PHONES (5) ENGAGING IN A CLASS ACTIVITY (VI: VERTICAL INDEX; HI: HORIZONTAL INDEX)

Case	Emotion	Eye Movement	Head Pose
(1)	Mostly neutral	Mostly +ve VI with obvious variation of HI	Little movement
(2)	Mostly neutral	More variation of VI than HI	Obvious downward movement
(3)	Occasional change	Steady VI/HI in one direction	Little movement
(4)	Mostly neutral	Mostly +ve VI with little variation of HI	Obvious downward movement
(5)	Mixed emotions	Mostly +ve VI with little variation of HI	Little movement

proposed methods can be used to enhance teaching/learning in various ways. For example, they can provide a better understanding of student behaviors in a class. Furthermore, by studying student behaviors, teaching methods can be adjusted and learning materials can be enhanced. However, the proposed solution has some limitations as accuracy may be affected by different factors such as webcam quality. Also, students may hide their actual activities by pretending to carry out certain behaviors. As future work, machine learning will be incorporated into the proposed method to provide more accurate detection and discover more insights to enhance future teaching/learning experiences.

ACKNOWLEDGMENT

This work was supported by a Special Grant for Strategic Development of Virtual Teaching and Learning of The Hong Kong Polytechnic University under account number 89Q7.

REFERENCES

[1] K. A. Jeffery and C. F. Bauer, "Students' responses to emergency remote online teaching reveal critical factors for all teaching," *Journal of Chemical Education*, vol. 97, no. 9, pp. 2472–2485, 2020.

[2] D. M. Bunce, E. A. Flens, and K. Y. Neiles, "How long can students pay attention in class? a study of student attention decline using clickers," *Journal of Chemical Education*, vol. 87, no. 12, pp. 1438–1443, 2010.

[3] S. Bolkan and A. K. Goodboy, "Transformational leadership in the classroom: Fostering student learning, student participation, and teacher credibility.," *Journal of Instructional Psychology*, vol. 36, no. 4, 2009.

[4] P. Kinnunen, "Dealing with emotions – engineering teachers' observations of students' emotional reactions to receiving feedback on their work," in *2020 IEEE Frontiers in Education Conference (FIE)*, pp. 1–5, 10 2020.

[5] A. Panahi and P. Duraisamy, "Automated classroom engagement evaluation using machine learning for 180 degree camera environment," in *2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, pp. 01–05, IEEE, 2021.

[6] B. Meriem, H. Benlahmar, M. A. Naji, E. Sanaa, and K. Wijdane, "Determine the level of concentration of students in real time from their facial expressions," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 1, 2022.

[7] P. Sharma, M. Esengönül, S. R. Khanal, T. T. Khanal, V. Filipe, and M. J. Reis, "Student concentration evaluation index in an e-learning context using facial emotion analysis," in *International Conference on Technology and Innovation in Learning, Teaching and Education*, pp. 529–538, Springer, 2018.

[8] B. E. Zakka and H. Vadapalli, "Estimating student learning affect using facial emotions," in *2020 2nd International Multidisciplinary Information Technology and Engineering Conference (IMITEC)*, pp. 1–6, 2020.

[9] P. Takáč, P. Sinčák, and M. Mach, "Lecture improvement using students emotion assessment provided as sas for teachers," in *2016 International Conference on Emerging eLearning Technologies and Applications (ICETA)*, pp. 347–351, 2016.

[10] A. Alrawahneh and S. Safei, "A model of video watching concentration level measurement among students using head pose and eye tracking detection," *Journal of Theoretical and Applied Information Technology*, pp. 4305–4315, 2021.

[11] S. Cha and W. Kim, "Concentration analysis by detecting face features of learners," in *2015 IEEE Pacific Rim Conference on Communications, Computers and Signal Processing (PACRIM)*, pp. 46–51, IEEE, 2015.

[12] L. Krithika and G. Lakshmi Priya, "Student emotion recognition system (sers) for e-learning improvement based on learner concentration metric," *Procedia Computer Science*, vol. 85, pp. 767–776, 2016.

[13] M. Horvat and T. Jagušt, "Emerging opportunities for education in the time of covid-19-adaptive e-learning intelligent agent based on assessment of emotion and attention," in *Central European Conference on Information and Intelligent Systems*, pp. 203–210, Faculty of Organization and Informatics Varazdin, 2020.

[14] M. L. Roy, D. Malathi, and J. D. Jayaseeli, "Students attention monitoring and alert system for online classes using face landmarks," in *2021 IEEE 4th International Conference on Computing, Power and Communication Technologies (GUCON)*, pp. 1–6, IEEE, 2021.

[15] V. U. Maheswari, R. Aluvalu, M. P. Kantipudi, K. K. Chennam, K. Kotecha, and J. R. Saini, "Driver drowsiness prediction based on multiple aspects using image processing techniques," *IEEE Access*, vol. 10, pp. 54980–54990, 2022.

[16] Z. Xie and X. Cao, "Construction three-dimensional observation system of teacher-student interaction behavior in classroom and practice," in *2010 2nd International Conference on Education Technology and Computer*, vol. 4, pp. V4–271–V4–274, 2010.

[17] S. Q. Yang, Y. H. Chen, Z. Y. Zhang, and J. H. Chen, "Student in-class behaviors detection and analysis system based on cbam-yolov5," in *2022 7th International Conference on Intelligent Computing and Signal Processing (ICSP)*, pp. 440–443, 2022.

[18] L. Li, M. Liu, L. Sun, Y. Li, and N. Li, "Et-yolov5s: Toward deep identification of students' in-class behaviors," *IEEE Access*, vol. 10, pp. 44200–44211, 2022.

[19] U. V. Maheswari, R. Aluvalu, and K. K. Chennam, "Chapter 5 application of machine learning algorithms for facial expression analysis," in *Machine Learning for Sustainable Development*, pp. 77–96, De Gruyter, 2021.