COMP-8790 Applied Artificial Intelligence

AI-Powered Classroom Monitoring

Prithvika Babu (110127176)

Nour ElKott (104982352)

Dhwani Patel (110028937)

December 10, 2023

Contents

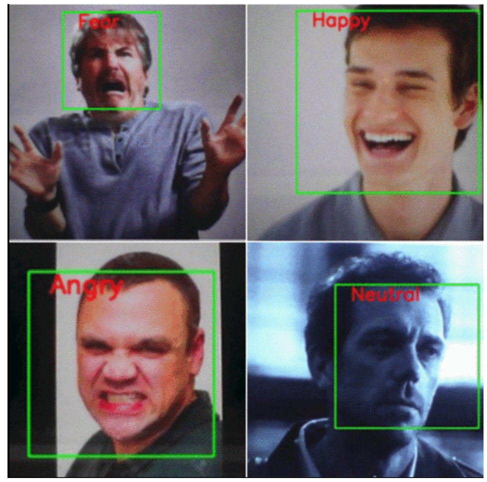
1 Abstract

This project builds upon the model developed by Vishnumolakala et al. in their paper titled ”In-Class Student Emotion and Engagement Detection System (iSEEDS)”. Their model employs emotion detection and eye tracking to assist teachers in monitoring students’ emotional states and engagement levels, in order to facilitate responsive teaching. To enhance their model, we incorporated emotion detection, eye tracking, and head pose estimation. This extension achieves the goal of tracking students’ concentration levels during a lesson, recognizing the significance of head movements in ensuring focused attention. Once our program ends, a detailed report is generated, summarising the time spent by the student on each emotion, eye gaze, and head position. By integrating artificial intelligence into the learning environment, our project provides educators with insightful reports, offering a deeper understanding of students’ engagement. This paper outlines our project’s objectives, methodology, experimentation, results, and analysis, to provide proof of the potential of AI-powered solutions for responsive teaching.

2 Introduction

A 2023 survey by Microsoft revealed that 60% of teachers believed teaching methods should adapt to leverage modern digital tools, and 80% expressed a need for additional tools to manage their workload [5]. Given that teachers should focus on creating and teaching lessons, it is burdensome for them to also ensure every student remains focused, which may potentially disrupt the learning experience for others. Hence, offering teachers the flexibility to concentrate solely on teaching would be beneficial to both the teachers and their students. For a teacher to identify and address boredom in the classroom is crucial for enhancing students’ academic progress. Boredom negatively impacts students’ success and signifies a need for increased engagement. Our survey, conducted at the beginning of the semester, showed that 80% of students attribute their boredom to unenthusiastic teaching methods or lack of professor engagement. Traditional observation methods are impractical, with only 27% of students recognizing their boredom and 47% being unsure. This aligns with our project’s objective — a program designed to monitor students’ emotional and concentration levels, generating detailed reports for each student. These reports serve a dual purpose: individual reports enable teachers to discuss concentration levels with each student and their parents, while collective reports will help teachers track overall emotional responses. For instance, if a majority of students display confusion, teachers can adapt their teaching methods to enhance the overall learning experience. Our project implements emotion detection, eye gaze tracking, and head pose estimation. By implementing more appropriate teaching methods, the students will not only be more interested and enthusiastic in classes but will also stop correlating a particular class with a negative emotion, which will be useful in self-development and reflect positively on their academics.

(a) Output of iSEEDS eye tracking model



(b) Output of iSEEDS emotion detection model

Figure 1: The figures above are the generated iSEEDS eye tracking and emotion detection algorithms, performed separately.

3 Literature Review

ISeeds [8] focused on extracting the concentration levels of students using eye tracking (Figure 1a) and emotion detection (Figure 1b), and making use of the data generated. As shown in Figure 1, both algorithms are run separately, a limitation we aimed to extend in our project. The libraries implemented in iSeeds for the algorithms include OpenCV and the models used involve Convolutional Neural Networks (CNN). By determining the eye tracking movements and the emotion displayed most throughout the duration of the recording, a conclusion is drawn on how concentrated the student is. However, this paper mentions head-pose detection as a limitation since it wasn't implemented, which is one of the most important features to include, and we have extended this project to include.

4 Methodology

The project utilises a multifaceted approach incorporating emotion detection, eye-tracking, and head pose estimation. Each technique contributes to the total understanding of students' emotional and concentration levels. The project employs deep learning frameworks, Dlib and MediaPipe for facial detections and PyTorch for model development, computer vision libraries like OpenCV for image analysis, and real-time data software including Google Colab for simulations. The methodologies used for three models are described below:

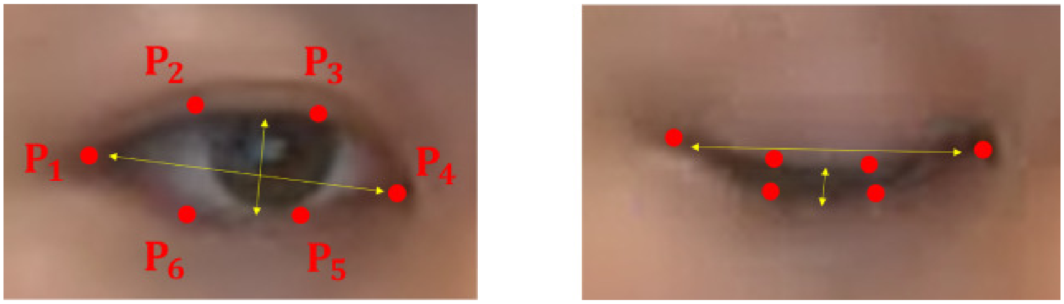


Figure 2: Landmarks found on points of each eye, detected through dlib, are used to detect if eyes are opened or closed.

4.1 Emotion Detection

Face Detection: This model makes use of a face cascade classifier from OpenCV (cv2.CascadeClassifier) to detect the faces in each frame, and draw a 'rectangular box' on the face.

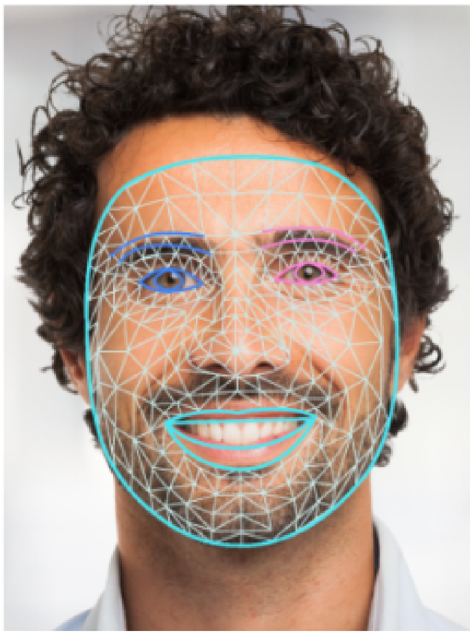
Emotion Detection: To determine the facial landmarks, Dlib is used, and these landmarks are then used to extract a region of interest (ROI). The ROI is preprocessed before being fed into an open-source pretrained emotion detection model. This emotion detection model uses ratios for each of the facial landmark coordinates to 'predict' the emotion faced e.g. lower coordinates on beginning of eyebrow, and upper coordinate on ends of eyebrows could mean the person is angry. Video Input Processing: Keras library is used for video processing, while the other functions act on the preprocessed input video/webcam, to divide them into frames and act on each frame accordingly.

4.2 Eye Gaze Tracking

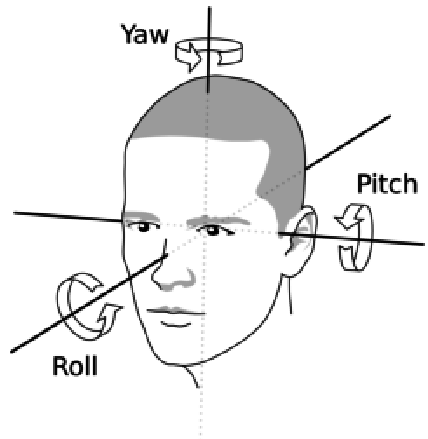
Facial Landmarks Detection: The model utilizes Dlib's pre-trained face detector to locate faces in the video frames. Once a face is detected, Dlib's facial landmarks predictor is employed to identify specific points on the face, including the corners of the eyes.

Eye Aspect Ratio (EAR) Calculation: EAR is a crucial metric for discerning whether the eyes are open or closed. Figure 2 illustrates the landmarks employed to determine the state of the detected eyes. It is calculated independently for the left and right eyes based on the distances between facial landmarks.

Gaze Direction Detection: The horizontal position of the eyes is used to estimate the direction of gaze. Looking left, right, or straight is determined based on the average horizontal position If the average horizontal position is less than 0.4, it is considered as looking left. If greater than 0.6, it's considered as looking right. Otherwise, it's categorized as looking straight.



(a) Face mesh created using facial landmarks identified by MediaPipe.



(b) Diagram of roll, pitch, and yaw on a head.

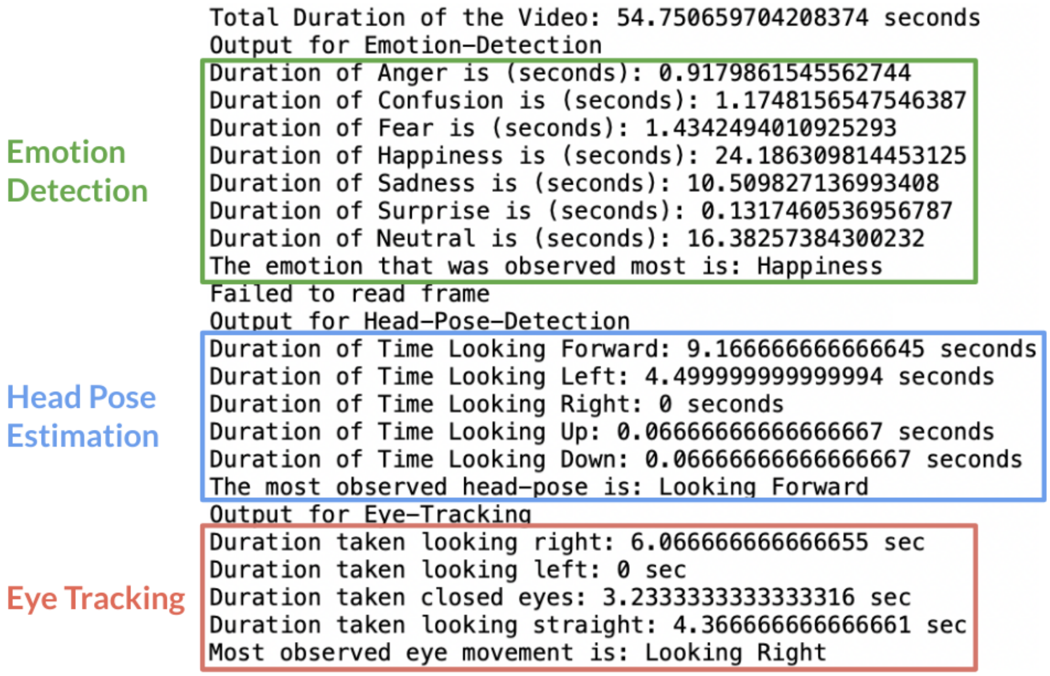
Figure 3: The figures above show the important concepts needed in head-pose estimation.

4.3 Head Pose Estimation

MediaPipe: The MediaPipe library is used to detect the facial landmarks in the inputted video/ camera feed. It is programmed using machine learning (ML) models to output 3D facial landmarks and a face mesh, as shown in Figure 3a. These landmarks are used to track head movements.

Pose-n-Perspective: Using OpenCV's *solvePnP* function, we project the facial landmarks identified by Mediapipe from a 2D plane onto a 3D plane. This process is essential for translating these landmarks into real-world representations of head movements, given that we live in a 3D environment.

Head Pose Classification: In the context of a head's orientation, roll, pitch, and yaw describe rotational movements around different axes (Figure 3b). Roll involves tilting the head from side to side, pitch involves nodding the head up and down, and yaw refers to turning the head from left to right. Together, these terms capture the full range of rotational motions. However, for the application of this project, we omitted roll since it would not impact a student's concentration.




```
Total Duration of the Video: 54.750659704208374 seconds
Output for Emotion-Detection
Duration of Anger is (seconds): 0.9179861545562744
Duration of Confusion is (seconds): 1.1748156547546387
Duration of Fear is (seconds): 1.4342494010925293
Duration of Happiness is (seconds): 24.186309814453125
Duration of Sadness is (seconds): 10.509827136993408
Duration of Surprise is (seconds): 0.1317460536956787
Duration of Neutral is (seconds): 16.38257384300232
The emotion that was observed most is: Happiness
Failed to read frame
Output for Head-Pose-Detection
Duration of Time Looking Forward: 9.166666666666645 seconds
Duration of Time Looking Left: 4.499999999999994 seconds
Duration of Time Looking Right: 0 seconds
Duration of Time Looking Up: 0.06666666666666667 seconds
Duration of Time Looking Down: 0.06666666666666667 seconds
The most observed head-pose is: Looking Forward
Output for Eye-Tracking
Duration taken looking right: 6.066666666666655 sec
Duration taken looking left: 0 sec
Duration taken closed eyes: 3.2333333333333316 sec
Duration taken looking straight: 4.366666666666661 sec
Most observed eye movement is: Looking Right
```

Figure 4: A sample report generated by the program.

4.4 Final Report Generation

At the end of the program, a final report is generated. The report consists of the time spent on each:

- Emotion (ie. happiness, saddness, surprise, confusion, etc.),
- Eye gaze direction, and
- Head pose.

5 Discussion

5.1 Results and Analysis

Our extension and implementation of the In-class Student Emotion and Engagement Detection System (iSEEDS) demonstrated successful outcomes across its integrated modules. Emotion detection, utilizing OpenCV and a pre-trained dlib 5-point landmarking model, accurately classified a range of emotions based on facial expressions. The Eye-tracking module, employing Dlib for face detection and Eye Aspect Ratio (EAR) calculation, effectively tracked eye movements, providing insights into durations of open and closed eye states as well as gaze directions. Head Pose Estimation, using cv2.solvePnP and the Face Mesh Model, successfully classified head poses, contributing to a comprehensive understanding of student engagement. The system's real-time feedback and multi-modal approach offer valuable insights for instructors to adapt teaching methods and enhance classroom interactions. Despite notable successes, considerations include the speed of the emotion detection module and the system's current design for single-person detection.

5.2 Implications and Significance

The implications of this system are significant in the context of responsive teaching. By understanding students' emotional states and concentration levels, instructors can tailor their teaching methods to enhance engagement and address potential challenges. The real-time feedback generated by iSEEDS allows for immediate interventions, fostering a more interactive and adaptive learning environment.

5.3 Limitations

While the system demonstrated effectiveness, certain limitations should be considered. The speed of the emotion detection module and the sequential execution of algorithms may impact the real-time processing capabilities. Additionally, the model is currently designed for single-person detection, and factors such as facial interference and dynamic teacher movement may influence the accuracy of head pose estimation.

6 Conclusion

Our model would be used to enhance students' concentration levels through the integration of Emotion Detection, Eye-tracking, and Head Pose Estimation. While our model has limitations, it successfully captures facial expressions, eye movements, and head poses. Moreover, The model also generates a text file with the output on the duration of each parameter, and the most observed parameter in this frame. This data can be used to analyse how concentrated a student is. Hence, the project concludes with an emphasis on how it effectively tackled the challenges of monitoring student concentration levels and provided instructors with valuable insights.

7 Future Work

Our model currently works on one person at a time, so future works could include being able to detect and function on multiple faces, since classrooms would generally invest on one camera per class. In addition, considering the limitation of some people having natural facial features which may be misinterpreted by the model, we can add in more data to the model for preprocessing, and the data would contain the pictures of some students as well as people of different ethnicities. Another inclusion would be to analyse the accuracy of the current model using some machine learning algorithms in order to determine the locations for improvement. The current model generates a text file containing the data as an output, but one thing to be considered for the future is to generate graphical representation of the processed data, so it is more comprehensible and user-friendly.

8 References

1. Akihiro Kuwahara, Kazu Nishikawa, Rin Hirakawa, Hideaki Kawano, Yoshihisa Nakatoh, Eye fatigue estimation using blink detection based on Eye Aspect Ratio Mapping (EARM), Cognitive Robotics, Volume 2, 2022, Pages 50-59, ISSN 2667-2413, https://doi.org/10.1016/j.cogr.2022.01.003.

2. Jantunen, T., Mesch, J., Puupponen, A., & Laaksonen, J. (2016). On the rhythm of head movements in Finnish and Swedish Sign Language sentences. Speech Prosody, 850-853. DOI: 10.21437/SpeechProsody.2016-174.

3. José, I. (2018, June 28). Facial mapping (landmarks) with Dlib + Python. Towards Data Science. Medium. https://towardsdatascience.com/facial-mapping-landmarks-with-dlib-python-160abcf7d672.

4. Krishna, N. (2022, January 27). Camera Calibration with Example in Python. Medium. https://towardsdatascience.com/camera-calibration-with-example-in-python-5147e945cdeb.

5. Microsoft Canada Inc. (2023, September 13). K-12 Teachers Say Classroom Models Need to Evolve to Prepare Canadian Students for the Future. Newswire. https://www.newswire.ca/news-releases/k-12-teachers-say-classroom-models-need-to-evolve-to-prepare-canadian-students-for-the-future-850316188.

6. Rosebrock, A. (2017, April 3). Facial landmarks with dlib, OpenCV, and Python. PyImageSearch. https://pyimagesearch.com/2017/04/03/facial-landmarks-dlib-opencv-python/.

7. Saheb, T. (2023). Ethically contentious aspects of artificial intelligence surveillance: a social science perspective. AI Ethics, 3, 369–379. https://doi.org/10.1007/s43681-022-00196-y.

8. S. K. Vishnumolakala, V. S. Vallamkonda, S. C. C, N. P. Subheesh, and J. Ali, "In-class Student Emotion and Engagement Detection System (iSEEDS): An AI-based Approach for Responsive Teaching," 2023 IEEE Global Engineering Education Conference (EDUCON), Kuwait, Kuwait, 2023, pp. 1-5, doi: 10.1109/EDUCON54358.2023.10125254.