International Multidisciplinary Research Journal
ISSN: 2424-7073
Available Online at https://imrjournal.info/
Volume: III; December 2024; Page No.1-8
DOI: https://doi.org/10.47722/imrj.2001.31

Research Article

ENHANCING CLASSROOM ENGAGEMENT THROUGH AI-POWERED EMOTIONAL, HEAD POSE, AND GAZE TRACKING: A NOVEL APPROACH TO RESPONSIVE TEACHING

Kalyani Selvarajah[1], Nour ElKott[2], Prithvika Babu[3], Dhwani Patel[4]
School of Computer Science, University of Windsor, Canada

ARTICLE INFO

Article History:

Received 15.08.2024
Accepted 15.10.2024
Published 15.12.2024

Keywords:

AI in Education,
Emotion Detection,
Eye Tracking,
Head Pose Estimation

ABSTRACT

Active participation of students in classroom is crucial for enhancing the learning process. Their emotional state significantly influences not only the content they grasp but also their level of engagement during lessons and their overall academic performance. Emotions impact how motivated students are to study, concentrate, and manage their learning. Monitoring students' emotions in the classroom and handling them properly are important for a better learning experience. However, it can be an added challenge for teachers who also need to focus on creating and teaching high-quality lessons. To support responsive teaching, we have developed an AI powered classroom monitoring tool that detects emotions and head-pose, and tracks students' eye movement so that the teachers can monitor students' emotional states and engagement levels. In this research, we address one of the limitations in the existing work, head-pose estimation to improve the model accuracy. This model includes the following steps: (1) analyzes students' emotional states — such as confusion, happiness, and more — during the lesson, (2) tracks their gaze direction to determine if their focus is on the instructor, to their sides, or if their eyes are shut completely, and (3) monitors head orientation to identify where students spend most of their time looking. After completing the analysis over a specified span of time, the AI powered tool generates a detailed report on student focus and emotional status to present educators with statistics that can be used to tailor their teaching strategies whether it's online or in a classroom setting. As a result, teacher can improve the teaching materials for better content delivery and support adaptive teaching methods.

I. Introduction

In today's rapidly evolving educational landscape, teachers face increasing challenges to keep students engaged amidst distractions and diverse learning needs. A majority of teachers recognize the need for modern tools to enhance both teaching methods and classroom management. According to the survey in 2023 by Microsoft, 60% of teachers believed that

teaching methods should adapt to use modern digital tools, and 80% expressed a need for additional tools to manage their workload (Microsoft Canada Inc., 2023). The primary duty of teachers is to conduct lessons, and the added responsibility of consistently maintaining student focus can be challenging, which may lead to disruptions in the overall learning experience for the class. Hence, offering teachers the flexibility to concentrate solely on teaching would be beneficial to both the teachers and their students. To enhance students' academic progress, it is crucial to identify and address boredom, confusion, and other negative emotions, as well as monitor student engagement in the classroom.

Emotions are complex internal states that include feelings, cognitive, physical reactions, expressions, and motivations that happen when an individual achieve or are blocked from achieving a goal (D'Errico F 2016). Recognizing students' emotions is essential, as these emotions can significantly influence their motivation and ability to study effectively. While the Microsoft survey shows the growing reliance on digital tools in education, recent research by Vishnumolakala et al. 2023 focuses on examining the concentration levels of students through emotion detection and corresponding eye tracking analysis which serve as reliable indicators of stress, fatigue, or lack of understanding. They discuss that head-pose detection as a limitation in their work to have a precise concentration metric. The ***head-pose detection*** is the process of analyzing the orientation and position of a student's head to infer their level of attention and engagement in the classroom. Therefore, this research addresses the head-pose detection as it is one of the crucial features to include in the context of attention monitoring. Additionally, ***eye gaze tracking*** which involves monitoring the direction and focus of a student's eyes to determine students' attention level and engagement with classroom material, is considered. This aligns with this research's objective — an AI tool designed to monitor students' emotional and concentration levels, then generate a detailed report 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 help teachers track overall emotional and concentration trends. For instance, if the majority of students display negative emotions such as confusion, teachers can modify their teaching methods to improve the overall learning experience. This project incorporates three key components: emotion detection, eye gaze tracking, and head pose estimation. Each of these elements contributes to understanding students' emotional and concentration levels. By addressing these limitations through the integration of advanced AI technologies, this research aims to significantly enhance responsive teaching methods to ultimately create a more dynamic and personalized learning environment.

II. Materials and Methods

This research incorporates emotion detection and corresponding eye tracking analysis and head pose estimation, with each technique contributing to a comprehensive understanding of students' emotional and concentration levels. Figure 1 displays three sample frames from the live feed input; the simultaneous operation of emotion detection, eye gaze tracking, and head pose

estimation processes shows how these components interact during data collection. The approach employs the following tools:

- **Dlib**
  - used for facial landmark detection and emotion analysis.
- **MediaPipe**
  - used for real-time face and pose tracking.
- **PyTorch**
  - used for model development and for training deep learning models.
- **OpenCV**
  - handles image analysis and processing tasks.
- **Google Colab**
  - the chosen platform for simulating and testing the models.

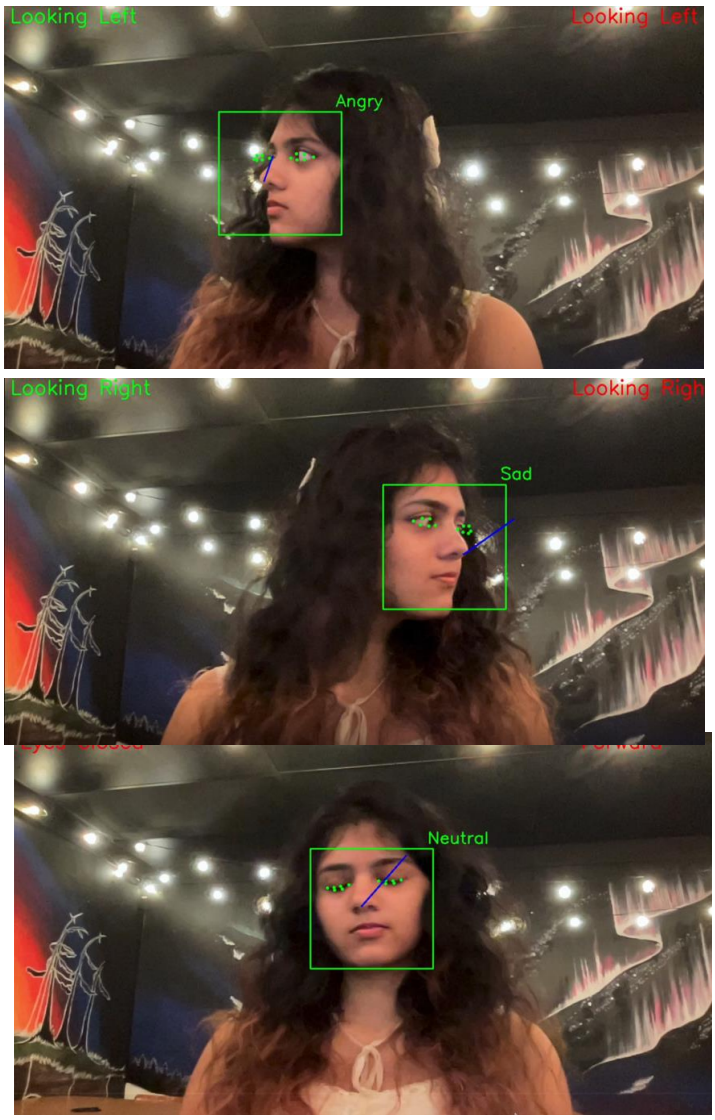

Figure 1: All three processes: (a) emotion detection (green box, corresponding green emotion), (b) eye gaze tracker (top left text), and (c) head pose estimator (top right text)—run simultaneously.

A. Emotion Detection

In this research, the emotion detection model begins with face detection using a face cascade classifier from OpenCV, is a machine learning method, where a cascade function is trained from a lot of positive and negative images so that it can find objects in different pictures. It identifies faces in each frame and marks them with a rectangular box. Next, Dlib is employed to determine facial landmarks, extracting a region of interest (ROI) from the detected face (José, 2016). This ROI undergoes preprocessing before being input into a pretrained emotion detection model, which uses the ratios derived from the coordinates of facial landmarks to predict the person's emotional state as shown in Figure 2. For example, lower coordinates at the beginning of the eyebrows and upper coordinates at the ends of the eyebrows can indicate anger. Video input processing is run using the Keras library, which segments the preprocessed input video or webcam feed into frames, allowing each frame to be processed accordingly.

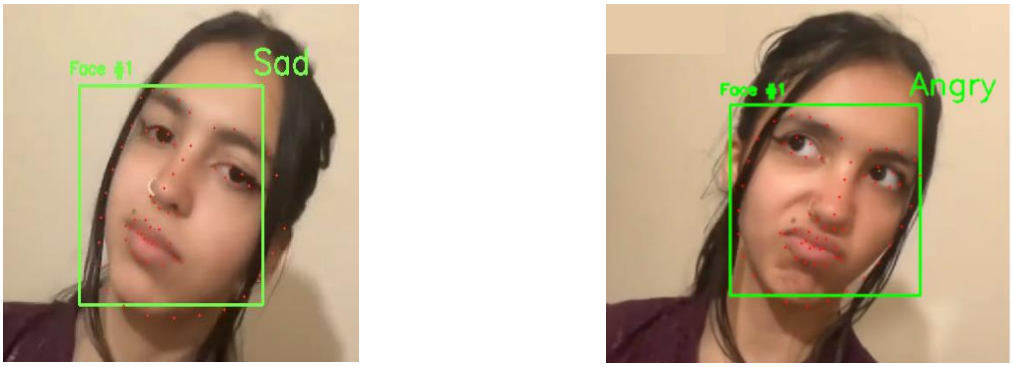

Figure 2: Sample frames from the emotion detection process. Frames display the red facial landmarks used to determine the emotion; the green box and green text identify the face and emotion displayed.

B. Eye Gaze Tracking

Eye gaze tracking, as discussed in Vishnumolakala et al. 2023, involves monitoring the direction and focus of a student's eyes to determine their attention level and engagement with classroom material. By analyzing where a student is looking, the system can infer whether they are following along with the lesson or becoming distracted. Just as with the emotion detection, the eye gaze tracking model starts with facial landmarks detection using Dlib's pre-trained face detector. Once a face is detected, Dlib's facial landmarks predictor identifies specific points on the face, including the corners of the eyes. The Eye Aspect Ratio (EAR) is then calculated independently for the left and right eyes based on the distances between facial landmarks; this determines whether the eyes are open or closed. Figure 3 illustrates the landmarks used for this calculation. To estimate gaze direction, the model assesses the horizontal position of the eyes. If the average horizontal position is less than 0.4, the gaze direction is set to left; if greater than 0.6, it is considered looking right; otherwise, it is categorized as looking straight.

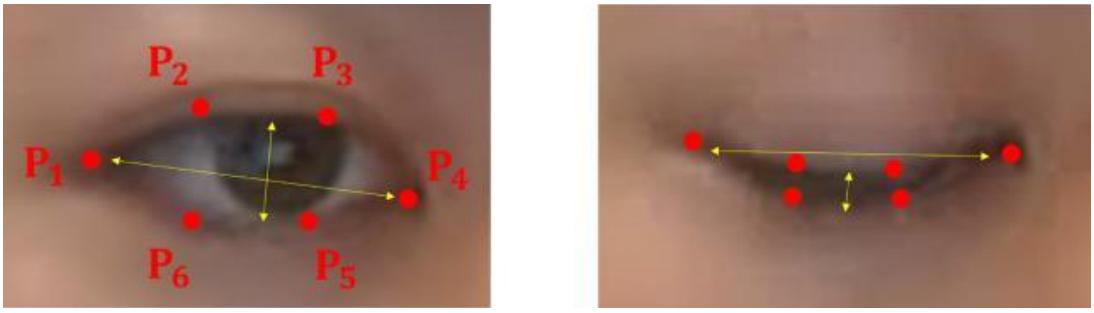

Figure 3: Landmarks found on points of each eye, detected through dlib, are used to detect if eyes are opened or closed.

C. Head Pose Estimation

The direction of someone's head tells a lot about what they're thinking or feeling. For instance, when people talk, they often move their heads to signal when it's their turn to speak or who they're talking to. Nodding shows agreement, while other gestures can show disagreement, confusion, or agreement (Asperti et al. 2023). In computer vision, head pose estimation means figuring out which way a person's head is facing compared to the camera's view.

In this research, the head pose estimation process begins with using the MediaPipe library to detect facial landmarks within the input video or camera feed (Rosebrock, 2021). MediaPipe outputs 3D facial landmarks and a face mesh, as shown in Figure 4 (a), crucial for tracking head movements. To translate the 2D facial landmarks identified by MediaPipe into real-world representations of head movements in a 3D environment, we utilize OpenCV's solvePnP function. This step accurately projects the landmarks onto a 3D plane (Krishna, 2022). Head pose classification involves determining the head's orientation based on roll, pitch, and yaw — rotational movements around different axes depicted in Figure 4 (b) (Jantunen, 2016). Roll pertains to the tilting of the head from side to side, pitch involves the nodding of the head up and down, and yaw refers to the turning of the head from left to right. For the purposes of this research, roll has been excluded as it does not necessarily represent a student's concentration.

(a) (b)

Figure 4: For the head-pose estimation: (a) Face mesh created using facial landmarks

identified by MediaPipe. (b) Diagram of roll, pitch, and yaw on a Head (Jantunen, 2016).

D. Final Report Generation

After the video or live camera feed stops, a summary of all the analysis is produced. This summary includes the time spent on each of the following:

1. Emotion (ie. happiness, sadness, surprise, confusion, etc.),
2. Eye gaze direction, and
3. Head pose.

A sample of the type of report generated is shown in Figure 5.

III. Results and Discussion

In this research, we have addressed the limitation of iSEEDS (Vishnumolakala et al. 2023). Our model has yielded successful outcomes across all the integrated modules. In the Emotion Detection module, OpenCV and a pre-trained Dlib 5-point landmarking model were able to accurately classify a range of emotions based on facial expressions. The Eye-Tracking module, which implement Dlib for face detection and Eye Aspect Ratio (EAR) calculation, effectively tracked eye movements, tracking the durations of open and closed eye states as well as gaze directions.

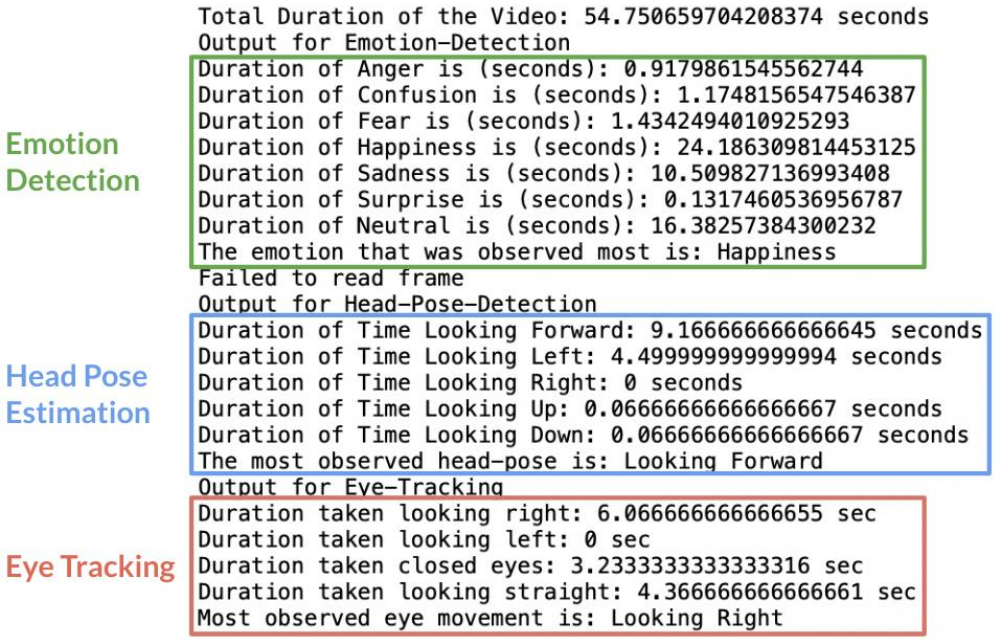

Figure 5: A sample report generated by the program.

Finally, the Head Pose Estimation module uses OpenCV's solvePnP and the Face Mesh Model to successfully classify head poses, which tracks student engagement.

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. This personalized approach has the potential to improve educational outcomes and a more dynamic learning environment.

While the system is effective, certain limitations should be considered. The speed of the emotion detection module and the sequential execution of algorithms may impact 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.

IV. Conclusion

This research successfully integrates emotion detection, eye gaze tracking, and head pose estimation to provide a comprehensive analysis of students' emotional and concentration levels in the classroom. By leveraging advanced AI tools such as Dlib, MediaPipe, and OpenCV, the system can accurately detect and classify emotions, monitor eye movements, and estimate head poses. These capabilities enable a deeper understanding of student engagement, allowing teachers to tailor their instructional strategies to better meet individual and collective needs.

The results demonstrate the effectiveness of this AI-driven approach in enhancing responsive teaching methods. The generated reports offer valuable insights into each student's emotional and attentional states, facilitating meaningful discussions with students and their parents and enabling teachers to adjust their methods based on real-time data. This system holds the potential to create a more dynamic, personalized learning environment that can address the diverse needs of students, ultimately contributing to improved educational outcomes.

However, the study also highlights certain limitations, such as the impact of processing speed on real-time application and challenges related to single-person detection and accuracy in the presence of facial interference or teacher movement. These factors suggest areas for future improvement, including optimizing the system for multi-person detection and enhancing its robustness in dynamic classroom settings. To address the limitation of natural facial features potentially being misinterpreted by the model, more diverse data will be included for preprocessing, incorporating images of students from different ethnic backgrounds (Saheb, 2023). To avoid issues of facial obstructions, images of these obstructions can be added to the training datasets to improve the model performance. Another area for future improvement is to analyze the accuracy of the current model using various machine learning algorithms to identify areas for enhancement. Finally, while the current model generates textual data output, future versions should aim to provide graphical representations of the processed data, making it more comprehensible and user-friendly. Despite these challenges, this research represents a significant step toward the development of intelligent educational tools that support both teachers and students in achieving better academic results.

V. References

Kuwahara, A., Nishikawa, K., Hirakawa, R., Kawano, H., & Nakatoh, Y. (2022). Eye fatigue estimation using blink detection based on Eye Aspect Ratio Mapping (EARM). Cognitive Robotics, 2, 50-59. ISSN 2667-2413. https://doi.org/10.1016/j.cogr.2022.01.0033

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.

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

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

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.

Rosebrock, A. (2021, April 3). Facial landmarks with dlib, OpenCV, and Python. PyImageSearch. https://pyimagesearch.com/2017/04/03/facial-landmarks-dlib-opencv-python/.

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.

Vishnumolakala, S. K., Vallamkonda, V. S., C, S. C., Subheesh, N. P., & Ali, J. (2023). 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, 1-5. doi: 10.1109/EDUCON54358.2023.10125254.

D'Errico F, Paciello M, Cerniglia L.
When emotions enhance students' engagement in e-learning processes. Journal of e-Learning and Knowledge Society. 2016 Sep 27;12(4).

Asperti, Andrea, and Daniele Filippini. "Deep learning for head pose estimation: A survey." *SN Computer Science* 4.4 (2023): 349.

How to cite this article:
Selvarajah K. "et al." (2024) 'Enhancing Classroom Engagement Through AI-Powered Emotional, Head Pose, and Gaze Tracking: A Novel Approach to Responsive, *International Multidisciplinary Research Journal*, Volume: III; December 2024; Page 1-8.
DOI: https://doi.org/10.47722/imrj.2001.31