Real Time Detection and Analysis of Facial Features to Measure Student Engagement with Learning Objects

Conor Cohen Farrell
National University of Ireland Maynooth

Charles Markham
National University of Ireland Maynooth

Catherine Deegan
Technological University Dublin

Real Time Detection and Analysis of Facial Features to Measure Student Engagement with Learning Objects

Conor Cohen Farrell, Charles Markham, Catherine Deegan
Maynooth University, TU Dublin Blanchardstown

Abstract

This paper describes a software application that records student engagement in an on-screen task. The application records in real time the on-screen activity and simultaneously estimates the emotional state and head pose of the learner. The head pose is used to detect when the screen is being viewed and the emotional state provides feedback on the form of engagement. The application works without recording images of the learner. On completing the task, the percentage of time spent viewing the screen and statistics on emotional state (neutral, happy, sad) are produced. A graph depicting the learner's engagement and emotional state synchronised with the screen captured video is also produced. It is envisaged that the tool will find application in learning activity and learning object design.

Keywords: Student Engagement, Keras, Dlib, Face Tracking.

1 Introduction

Student engagement measures the time and effort that students devote to study and other educationally purposeful activities. Student engagement also involves the design of resources for learning that promote learning itself. This is a large area of research and is regularly carried out on a National level [HEA, 2018]. Measuring engagement in education is a broad, multi-level topic that includes accessing overall engagement within the University environment, down to engagement in individual learning objects. The aim of this work was to develop a means of measuring student engagement in an on-screen task without the need for advanced hardware. Many approaches to this problem make use of specialist equipment such as eye trackers and RGB-D cameras [Bidwell et al., 2011][Zaletelj et al., 2017]. Other approaches record lower fidelity data making combined use of questionnaires, mouse and key loggers [Brown et al., 2014]. Recent advances in machine vision have provided software tools suitable for measuring engagement using a simple sensor such as the camera integrated into most laptops. There has been previous work done in this area. The OpenFace library has been used to record head pose and Facial Action Units relating to emotion of a class group watching an educational video [Chinchu et al. 2017]. Similar technologies have been used to measure classroom engagement [Etherington, 2019]. Rather than develop a pervasive technology, the aim of this work was to develop a tool to record a single student's response to an individual learning activity. Figure 1 shows themes considered by the HEA Survey of Student Engagement in Ireland [HEA, 2018]. Although this report considers engagement at a higher level, the effective teaching, learning strategies, and supportive environment themes do inform the design of the tool.

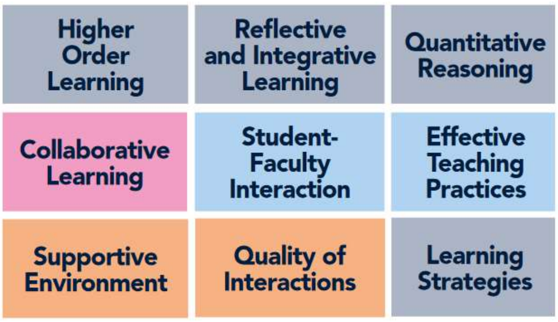


Figure 1: Indicators relating to engagement [HEA, 2018]

2 State of the Art

There are many options open to researchers wishing to implement face detection and expression using machine vision. Paul Ekman, perceived as the father of facial expression analysis, published the seminal paper on the subject describing human facial expression in terms of the activation of action units (groups of muscles) in the face [Ekman et al., 1976][Ekman et al., 1996]. This technique is known as Facial Action Coding System (FACS). This can be done manually by viewing the subject. More recently, computational methods allow FACS coding of expression [Hamm, 2011]. This approach provides detailed information about subtle changes in facial expression.

The Haar Cascade Classifier proposed by Paul Viola and Michael Jones has been used for a long while for detecting faces and has been integrated in OpenCV for several iterations [Viola & Jones, 2001]. This approach returns the location of a face in a scene independent of scale. While effective and computationally efficient to find faces in a scene, it works best with faces directed towards the camera. It does not provide a simple means of measuring head pose or facial expression.

A more recent approach to face detection is to use a Deep Neural Network (DNN) e.g. TensorFlow and Caffe. These are accessible through API frame works such as PyTorch and Keras.

3 Application development

The application was built around the OpenCV library for Python; an overview of the system developed is shown in figure 2.

The Linear Support Vector Machine (SVM) provided by Dlib was used to identify landmark points on the 2D image of the face. To determine head pose a rotation vector and translation vector was obtained by passing the camera matrix, model points (3D landmark points on an idealised head), and image points (2D landmark points) into the OpenCV 'solvePnP' method. The resulting rotation vector was converted to a rotation matrix (using Rodrigues method). The resulting translation and rotation matrices were combined using OpenCV method 'hcomcat'. The matrix formed was then decomposed to provide Yaw, Pitch and Roll values using 'decomposeProjectionMatrix'.

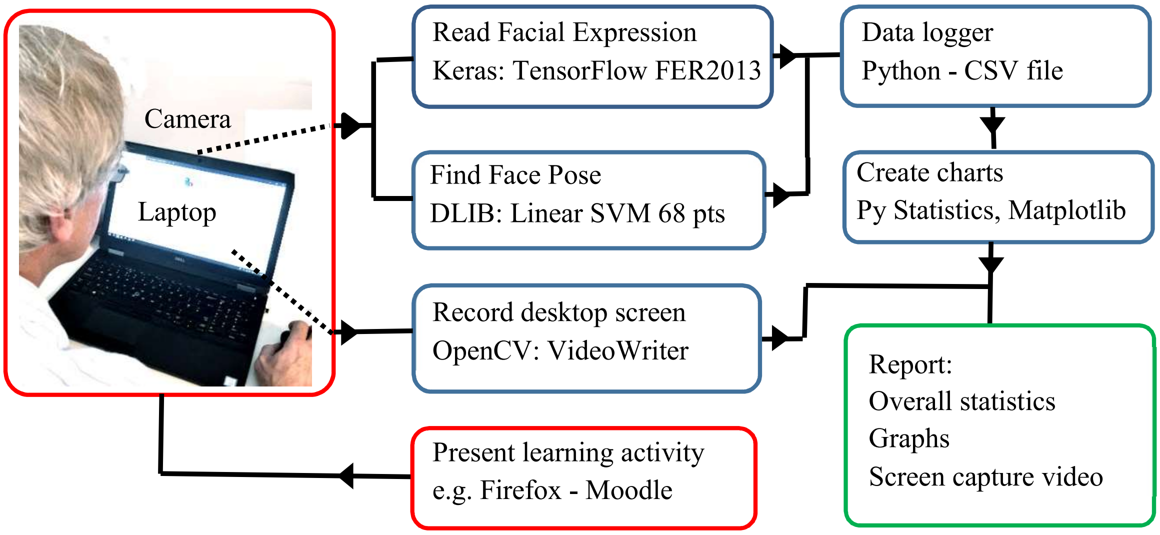


Figure 2: System overview

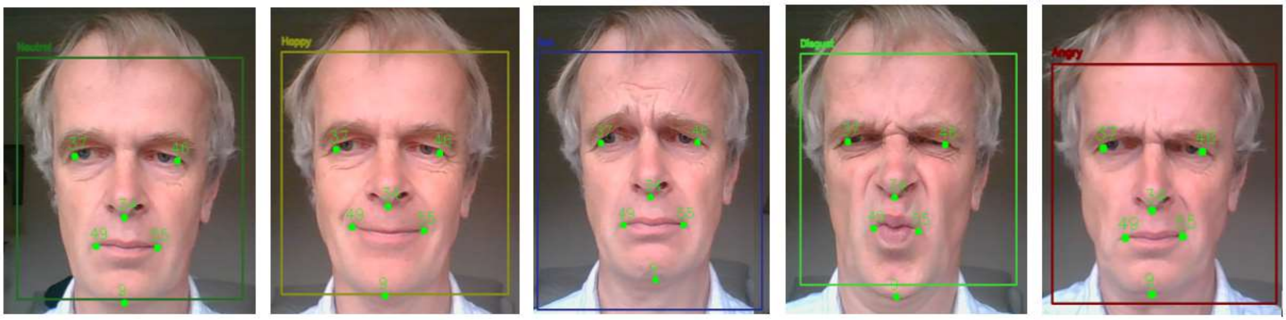
Facial expressions were detected using Keras (Python API for TensorFlow), TensorFlow (Machine Learning framework) and FER2013 (training data for facial expression). The FER2013 model is loaded. The face is located in the scene, extracted as a single image, converted to grey-scale and resized to 64x64 pixels. The 'emotion_classifier.predict' method was used to identify the relative level of each of the emotional states (happy, sad, angry, disgust, fear, surprise, and neutral). The highest scoring emotional state was returned as the emotion identified.

The screen recording was achieved using the OpenCV methods built around 'VideoWriter'. This approach could achieve a frame rate of about 4fps on the Laptop and 10fps on a higher performance desktop computer. The user interface was developed using the PyQt API.

The application generated graphs of system response as a function of time using the Matplotlib Python 2D plotting library. The graphs produced were save to a local folder as image files (png), along with CSV data and a video of the screen.

4 Results

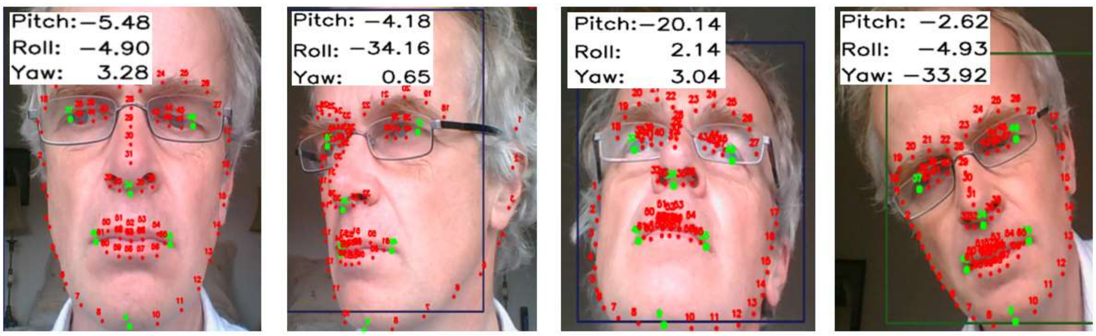
To commission the application a diagnostic feature was added to the program to mirror the analysis to the user in real time. The application did not record face data; the images shown are cropped screen shots of the application running. Figure 3 shows the response of the emotion detector to different acted facial expressions. This commissioning procedure may be an important feature to add in future work as lighting affects the performance of the emotion classifier.



Neutral Happy Sad Disgust Angry

Figure 3: Diagnostic display showing 5pt facial landmarks attached to face and emotion estimate based on five acted facial expressions.

Figure 4 shows the response of the head pose estimator in response to changes in head position. This component of the system worked well with and without glasses. Yaw, pitch or roll values exceeding a pre-set threshold were used to detect when user engagement with the laptop screen was lost.



Looking Forward Looking to right Looking up Rotating head

Figure 4: Head pose measurements yaw, pitch and roll (red 68pt mask and green 5pt landmarks).

To validate the systems response to a learning object a user was asked to view a PowerPoint presentation and progress through it (pressing a key to proceed to next page) as the system recorded their response, see Figure 5.

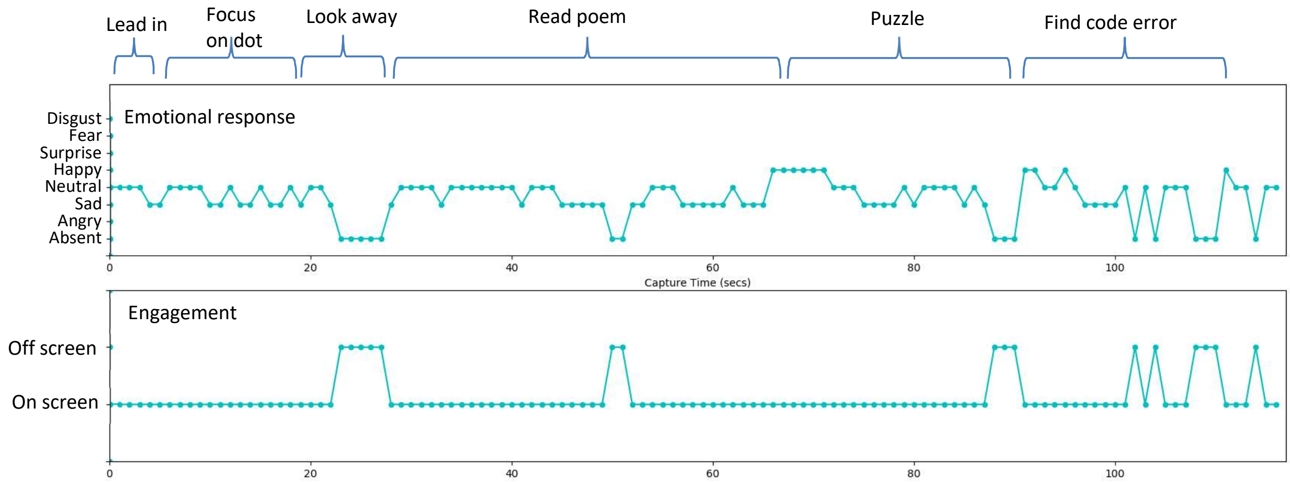


Figure 5: Engagement and emotional response recoded over time produced while interacting with PowerPoint presentation, overall engagement (in terms of time) was 85%, the emotional responses were - neutral 45%, absent 13%, happy 9% and sad 33%.

Before the presentation begins the application needs to be started and then minimised (this is labelled as lead in time above). The user is then asked to focus on a dot in the centre of screen, the emotion switches between neutral and sad in this period and there is full engagement. The next slide asked them to look away from the screen until they hear a beep. During this time the emotion is absent and no engagement is detected. The user then reads the poem ("The Road Not Taken" by Robert Frost) and in this case shows happiness at the end. The puzzle asks them to count the number of triangles in a simple drawing and finally they are asked to identify the error with the java program (missing semicolons).

5 Conclusions

This paper has demonstrated the feasibility of recording head pose and human emotion at the same time as producing a screen recording of student interaction with a learning object on a standard laptop. The head pose (engagement) was found to be reliable functioning in variable lighting and when glasses were worn. The emotion sensor needed much more care to get functioning as it required ideal lighting conditions (avoiding shadows). The tool developed does not record or retain images of the users face. The application will require ethical approval for use in a wider study of student engagement with different forms of learning object. The effectiveness of the tool for collecting this information will be the focus of future work.

References

[Bidwell, Jon & Fuchs, Henry, 2011], *Classroom Analytics: Measuring Student Engagement with Automated Gaze Tracking.* 10.13140/RG.2.1.4865.6242.

[Brown, L & Howard, A.M.. 2014], *A real- Time model to assess student engagement during interaction with intelligent educational agents*. ASEE Annual Conference and Exposition, Conference Proceedings.

[Chinchu Thomas, Dinesh Babu Jayagopi, 2017], *Predicting Student Engagement in Classrooms using Facial, Behavioral Cues*, MIE'17, November 13, 2017, Glasgow, UK

[Ekman, P., & Friesen, W. V., 1976], *Measuring Facial Movement.* Environmental Psychology and Nonverbal Behavior.

[Ekman, P., & Rosenberg, E., 1997], *What the face reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS). What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS) (2nd ed.).* https://doi.org/10.1109/76.49983

[Etherington, C, 2019] *Computer Vision for Classroom "Engagement Detection": Is it Ethical?,* January 19, 2019. https://news.elearninginside.com/computer-vision-for-classroom-engagement-detection-is-it-ethical/

[Hamm, J., Kohler, C. G., Gur, R. C., & Verma, R., 2011], *Automated Facial Action Coding System for dynamic analysis of facial expressions in neuropsychiatric disorders.* Journal of Neuroscience Methods, 200(2), 237–256. https://doi.org/10.1016/j.jneumeth.2011.06.023

[HEA] The (2018) Irish Survey of Student Engagement. https://hea.ie/assets/uploads/2018/11/ISSE-Report-2018-TEXT-Tag-A.pdf

[Pei Li, Adam Jourdan, 2017] *Sleepy pupils in the picture at high-tech Chinese school*, Reuters, Technology News, May 2017.

[Viola, P, Jones, M, 2001], *Rapid Object Detection Using a Boosted Cascade of Simple Features*, CVPR (1) 1 (511-518), 3

[Zaletelj, Janez, 2017], *Estimation of students' attention in the classroom from kinect features.* 220-224. 10.1109/ISPA.2017.8073599.