An Affect Computing based Attention Estimation

Amey Desai[1], Samarth Jain[2], Lohit Marodi[3], Sreejith V[4]
Department of Electrical and Electronics Engineering[1]
Department of Computer Science and Information Systems[2,3.4]
BITS-Pilani K. K. Birla, Goa campus, Goa, India
Email IDs: {f20140758[1], f20160018[2], f20150073[3],}@goa.bits-pilani.ac.in, srevin@gmail.com[4]

***Abstract*—Advancement in ubiquitous system has the potential to change the way we interact with the environment around us. In current smart systems, body language and visual cues form an important part of input to analyse and interpret the human perception. Understanding the emotional state of a human will help in generating a personalized response accordingly. Even-though there are multitude ways of capturing the affect of a person, common tools are found in computer vision techniques. In this paper, we present an attention measurement technique in a classroom setting using computer vision. Understanding the cognitive state of a student attending the class will help in improving the quality of education. Results indicate that the proposed system can be used to detect affective state automatically.**

***Index Terms*—Human attention, Smart systems, Affective computing**

I. Introduction

With the advancement in technology and more ubiquitous devices in place, users expect computers to take care of their needs rather than to take decisions themselves. Recent advancement of affective computing with smart system give the computer more intelligence, a human-like capabilities to observe, analyse and interpret the affect features. Affective computing builds an affect model of the subject based on different emotional aspects sensed via different sensors. This includes sensing the emotion in the speech, facial expression, movement (or body gestures) etc.

Application of affective computing is mainly focusing on determining a user emotional state and hence building an affect model. The application includes understanding the user attention state, gaming, e-learning, healthcare etc. Two research focuses [1] of affective computing include (i) acquisition and processing of human centered signals for recognition and integration (ii) combining multiple signals (multimodal) to determine the emotional affect.

Automated learning analysis and hence identifying the quality of learning by measuring the cognitive state of student is becoming an active topic of research in educational community. This is particularly needed for a large classroom where the students have difficulty in viewing and listening to the lecture. It is also challenging for a teacher to monitor individual student to find out their understanding about the lecture. A smart way of calculating the student engagement is to detect the affective state via facial expressions, body posture, face gazes etc [1]. Ubiquitous nature also enables such technology to take into account contextual information. Since, most human interaction involves a subtle understanding of context, these improvements make interactions with technology more human-like or natural. As an example, gaze is an important but subtle indicator of attention within a conversation. The beginning of individual conversations often requires that both individuals make eye contact. The engagement of an audience is often judged by their collective gaze.

Drawing on these principles, we have designed a system to measure student attention in classrooms using computer vision. In this paper, we use gaze as the primary indicator of attention and gauge student engagement based on gaze fixture. A system such as this is inexpensive and scalable and can provide real-time feedback to the instructor, which is crucial in such an environment and can greatly improve classroom dynamics. This helps the teachers to improve the classroom engagement by adjusting their teaching plans.

II. Background

Automated way of detecting students attention will help the teacher to have a better understanding of the class engagement. Traditionally, such information is collected manually. Three traditional ways of capturing attention include *filling questionnaire*, *conducting class test* and *direct observation*. These methods are time consuming and resource constrained and is difficult to conduct in a large classroom. To address these issues, a smart attention capturing system was proposed in different literature.

Depending on the type of sensors used, an attention capturing system can be classified into three categories. *i) Contact-based ii)vContact-less iii) hybrid(a combination of contact-based and contact-less* In *contact based techniques*, a wearable sensor or a sensor attached to the individual is used the attention. In contact-less techniques, a sensor which is present in the environment such as an RGB camera will capture the and hence do the detection. A *hybrid approach* uses a combination of *contact based* and *contact less* approaches.

Most of the computer vision based affect detection system uses a camera and face recognition models to capture the attention level [2]. Eye tracking is used in some literature to detect attention. Electroencephalography (EEG) based devices are used to capture to capture the attention of the user by directly recording the activity of the human brain [1]. Zhang et al., propose to use wearable devices with sensors such an accelerometer and gyroscope to recognize the students' behavior [3].

Vision based attention detection systems were designed to detect the human movements such as head rotation, eye gaze direction and understanding the facial features to infer human attention.

III. Proposed Approach

The goal of this paper primary focus on developing low complex, robust and low cost system. We propose the use computer vision and off-the-shelf hardware to implement the system. There are three main stages in estimating a subject's attention from the camera feed. The first stage employs *Facial Recognition* to tag individuals in the camera feed. Each subject detected in the feed is then further processed to estimate their line-of-sight. A subject is assumed to be paying attention if their gaze is on the point of interest. The point of interest include the possible location that the subject should be looking. This includes the blackboard, projector or the location of the teacher. The second stage involves *extraction of a set of facial landmarks*. We choose a set of six landmark points on the face. These features are used to calculate the *rotational matrix* of the face in stage 3, which helps estimate the deviation of line-of-sight from the camera.

A. Facial Recognition and landmark extraction

Several pre-trained facial recognition models exist [4], and we choose one such model for our implementation.The face recognition helps to identify several facial landmark.

We use the following set of points to calculate line-of-sight:

1) Nose tip
2) Left Eye left corner
3) Right Eye right corner
4) Mouth right corner
5) Mouth left corner
6) Chin tip

B. Line of sight estimation

The extracted landmarks are mapped to a set of points in the camera's frame of reference. Given the focal length of the camera, we can now estimate the transformation matrix using a linear matrix transform and the resulting rotational parameters give the deviation of gaze from the camera.

Let $\begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$ be the relative coordinates of the facial points and $\begin{bmatrix} U \\ V \\ W \\ 1 \end{bmatrix}$ be the real-world coordinates.

The relation between the two sets of coordinates is as follows:

$$s \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} r_{00} & r_{01} & r_{02} & \dots & t_x \\ r_{10} & r_{11} & r_{12} & \dots & t_y \\ r_{20} & r_{21} & r_{22} & \dots & t_z \end{bmatrix} \begin{bmatrix} U \\ V \\ W \\ 1 \end{bmatrix} \quad (1)$$

where s is the scale factor.

C. Attention Estimation

*1) **Top View**:* The scenario of students sitting in the class with the teacher is given in Figure 1. α defines the viewing angle. α depends only on the top view coordinates x and y. *CaseI*: Student looks at teacher

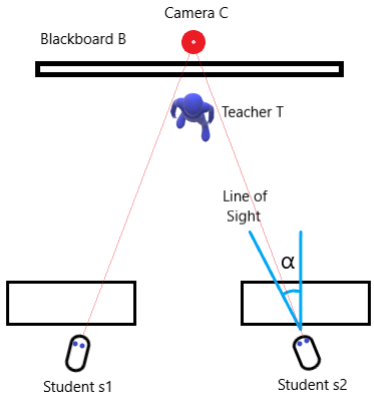

Fig. 1. Top View

$$\alpha_0 = \tan^{-1} \frac{X_0 - X_i}{Y_0 - Y_i} \quad (2)$$

Assumption: α can have a margin of 5.
Therefore, for students to look at a teacher

$$\tan^{-1} \frac{X_0 - X_i}{Y_0 - Y_i} - 5^\circ < \alpha < \tan^{-1} \frac{X_0 - X_i}{Y_0 - Y_i} + 5^\circ \quad (3)$$

Case II: Student is taking notes so he/she looks towards the blackboard

$$\alpha_1 < \alpha < \alpha_2$$

$$\tan^{-1} \frac{X_A - X_i}{Y_A - Y_i} < \alpha < \tan^{-1} \frac{X_B - X_i}{Y_B - Y_i} \quad (4)$$

*2) **Side View**:* The scenario of students sitting in the class with the teacher is given in Figure 2.

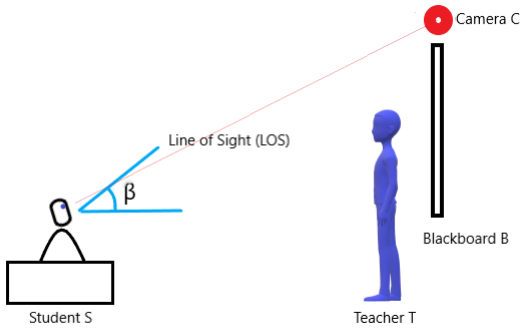

Fig. 2. Side View

Depends on y and z axis

Case 1: Student looks at teacher's face

$$\beta_0 = \tan^{-1}\frac{Z_0 - Z_i}{Y_0 - Y_i} \tag{5}$$

A buer of 5 can be given if the student is seeing hand gestures, etc. Therefore for a student to look at the professor

$$\tan^{-1}\frac{Z_0 - Z_i}{Y_0 - Y_i} - 5^\circ < \beta < \tan^{-1}\frac{Z_0 - Z_i}{Y_0 - Y_i} + 5^\circ \tag{6}$$

Case 2: Student looks at blackboard

$$\beta_2 < \beta < \beta_1$$

$$\tan^{-1}\frac{Z_A - Z_i}{Y_A - Y_i} < \alpha < \tan^{-1}\frac{Z_B - X_i}{Z_B - Y_i} \tag{7}$$

Net analysis is a mixture of both $\alpha and \beta$

Case I: Student looks at professor

$$\tan^{-1}\frac{X_0 - X_i}{Y_0 - Y_i} - 5^\circ < \alpha < \tan^{-1}\frac{X_0 - X_i}{Y_0 - Y_i} + 5^\circ \tag{8}$$

and

$$\tan^{-1}\frac{Z_0 - Z_i}{Y_0 - Y_i} - 5^\circ < \beta < \tan^{-1}\frac{Z_0 - Z_i}{Y_0 - Y_i} + 5^\circ \tag{9}$$

Case 2: Student looks at blackboard

$$\tan^{-1}\frac{X_A - X_i}{Y_A - Y_i} < \alpha < \tan^{-1}\frac{X_B - X_i}{Y_B - Y_i} \tag{10}$$

and

$$\tan^{-1}\frac{Z_A - Z_i}{Y_A - Y_i} < \beta < \tan^{-1}\frac{Z_B - X_i}{Z_B - Y_i} \tag{11}$$

It is thus possible to localize each student, the instructor and the blackboard. The α and β for each student will be given by the camera. If the α and β obtained for a student lies in either of above two cases, it means he/she is paying attention in the class.

D. Other Application

E. Learning Environment

This system can be used in typical classroom environments to provide feedback to instructors. It can also potentially be used in other learning environments such as laboratories as a safety tool, alerting instructors to potentially hazardous situations.

1) Advertising: A system such as this can help ascertain attention catching features in advertisements. Visual and other features of interest can then be used to create more effective content.

2) Live Performances: Live performances such as in theatre can use this tool to gauge audience engagement as well as the effectiveness of various artistic modalities in capturing attention. This could even contribute to audience directed plays, where the performance could take one of several possible routes depending on the audience's preference.

3) Driver Safety: As part of a larger safety system, this tool could help detect inattention or drowsiness on the road and prevent accidents. We envisage a system where apart from gaze, other data such as heart-rate and breathing rate are used to measure the driver's alertness and automatically respond in case of danger.

IV. Implementation

A. Tools

Software

- **Python** is an interpreted high-level programming language for general-purpose programming.
- **OpenCV** is a library of programming functions mainly aimed at real-time computer vision.
- **Dlib** is a general purpose cross-platform software library written in the programming language C++.

Hardware

- HD Webcam Logitech C270
  The webcam has a field of view of 60 degrees and focal length of 4 millimeter and was used to obtain image feed.

Dlib model is used for face recognition. This model suits our objective since it has built-in support for feature extraction. The Dlib model identifies several facial landmarks. Experiments were conducted in a closed class room environment with nine subjects to detect the attention they pay in the class. The location of teacher and blackboard with the location of the students are used to identify the relative gaze where the students are paying attention. We also used a BCI headset, Neurosky module to further understand the attention level of the student. Such hardware, can help in improving the quality of attention capturing along with the computer vision methods.

V. Conclusion

We present a proof-of-concept of an attention measuring system in this paper. The proposed system can enhance classroom learning experience by providing feedback to instructors. The proposed system can also be used in several other areas such as advertising, performing arts and automobile safety. The proposed system can be enhanced with additional hardware such as BCI module to increase the accuracy of attention detection.

References

[1] Gogia, Y., Singh, E., Mohatta, S., & Sreejith, V. (2016, November). Multi-modal affect detection for learning applications. In 2016 IEEE Region 10 Conference (TENCON) (pp. 3743-3747). IEEE.

[2] Asteriadis, S., Karpouzis, K., & Kollias, S. (2011, May). The importance of eye gaze and head pose to estimating levels of attention. In 2011 Third International Conference on Games and Virtual Worlds for Serious Applications (pp. 186-191). IEEE.

[3] Zhang, X., Wu, C. W., Fournier-Viger, P., Van, L. D., & Tseng, Y. C. (2017, June). Analyzing students' attention in class using wearable devices. In 2017 IEEE 18th International Symposium on A World of Wireless, Mobile and Multimedia Networks (WoWMoM) (pp. 1-9). IEEE.

[4] M. H. Yang, D. J. Kriegman, N. Ahuja, Detecting faces in images: a survey, IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(1):34-58,2002