An Intelligent System Framework for Measuring Attention Levels of Students in Online Course Environments

Omer Useche and Eman El-Sheikh
Department of Computer Science, University of West Florida, Pensacola, Florida, USA

Abstract – *With the growing popularity of online education, new challenges arise. One of those challenges, which results from the loss of face-to-face interaction, is the inability to track students' attentional response to instructional content and resources included in online courses. This paper describes an intelligent system that attempts to address this challenge, by monitoring student attention while participating in a Massive Open Online Course (MOOC). Educators and instructional designers can use the feedback gathered from the system to evaluate both the individual attentional needs of students and the effectiveness of certain instructional content and resources used in an online course, and adapt the course accordingly.*

Keywords: intelligent system, student modeling, attention measurement, computer vision, online education, MOOC.

1 Introduction

Significant recent advances in educational technology often come along with statements along the lines of "education is going to drastically change." These statements were often inspired from the advent of some new technology such as radio, film, television, personal computers and more recently, the Internet. Nevertheless, while these technologies were successful in changing other aspects of society, education remained largely unchanged. However, a growing number of online educational content and e-learning based courses have begun to popularize online instruction. This is evident in a survey conducted in 2013, which showed that at least 32% of students were taking at least one online course at the time [1]. Not only are universities offering fully online courses, many are also publishing Massive Open Online Courses (MOOCs), e-learning courses that anyone with a computer and access to the Internet can enroll in. These courses often make use of a varied number of instructional resources such as content modules, quizzes, exercises, videos, and online forums where instructors and students can collaborate and ask questions to provide a complete learning package. Moreover, many MOOCs provide relatively easy to use platforms for anyone to create their own online courses. But even with these advances, however, concerns regarding the effectiveness of online courses are still being raised.

Education has for years been interpreted as an activity that takes place in a classroom, with rows of students sitting on uncomfortable desks while directing their attention to an instructor standing in front of a whiteboard. However, while traditional methods and styles of education may seem outdated to many, there is at least one element that MOOCs have still not been able to address: the feedback that instructors get from their students based on everyday, face-to-face interaction. In fact, many of those who still oppose the use of MOOCs as an effective method of instruction stress "the lack of face-to-face tutoring as one of the main weaknesses of online courses" [2].

It could be easily argued that the job of an instructor is to facilitate the learning process. This process is driven by a complex system made up of a series of components and subsystems that determine what is to be learned and how effectively we learn. Information that must be learned by a student is interpreted as relevant or irrelevant by an executive subsystem called the decider, which makes decisions as to what is worth directing attention to based on received information, and on what needs to be encoded in the brain as knowledge [3]. Attention is an important part of this process so it follows that facilitating the learning process would require capturing students' attention.

Frequent face-to-face interactions in classroom settings inform educators of the educational needs and learning preferences of their students, and more importantly, the degree to which their instructional methods are effective based on the perceived levels of attention and motivation of their students. This type of feedback is important, as attention is necessary for acquiring new information [4]. Attention, in particular, is largely considered by cognitive psychologists, and more recently by a branch of neuroscience concerned with understanding the learning process in the brain, as a key to learning. Attention can be understood as a "mechanism that can flexibly control the flow of information from the environment to the organism and through the organism's various stages of neural processing" [5]. As such, its importance to education cannot be undermined. Indeed, the educational psychologist Robert Gagne, one of the most cited authors in instructional design and education literature, considered attention to be the first step of a series of nine instructional events that are essential in the leaning process

[6]. Sylwester and Choo advise that "teachers should adapt their instruction to the built-in biases and limitations of their students' stable attentional mechanisms" and "use imaginative teaching and management strategies to enhance the development of their students' adaptable attention processes" [7]. Yet, attention is often subjectively measured by observing students' behavior in face-to-face settings, and recommendations and strategies for gaining students' attention are often devised with the classroom in mind.

In the context of online educational content, how can instructors and instructional designers assure that they are grabbing students' attention in order to assure effective learning? Moreover, how can attention in an online learning environment be measured when, more often than not, attention seems to be a subjective interpretation of a student's behavior in class?

In an effort to provide a potential solution to this problem, we propose an intelligent system that uses computer vision technology to track and measure students' attention levels while engaging in instructional content through a MOOC. The information gathered by the system can then be used as an indication of student engagement and as feedback that can be used to improve online educational content. This system can be used by educators and instructional designers who wish to explore solutions that could bridge the gaps that exist between online and face-to-face education and enhance online learning.

2 Related work

The study of attention can be traced back to the beginnings of experimental psychology in the middle of the 19th century, when psychologists and scholars attempted to gain an understanding of attention through means of observation and cognitive analysis in order to further understand human behavior. At the beginning of the 20th century, Geisler reviewed a series of methods that were considered for measuring attention through: 1) changes of *peripheral vision*, which becomes more narrowed when a subject concentrates on a particular image, 2) changes in *muscular strength*, by correlating muscular tension with attention, 3) *liminal and differential sensitivity*, by asking subjects to rate how much they were able to notice different types of stimuli, 4) *reaction time*, which inversely correlates attention to retardation of attention, 5) *accuracy of work*, which correlates quality or quantity of work to the degree of attention that is directed to a particular task, and 6) a semantic attempt at measuring attention by using different *graded distractors* or varied degrees of stimuli to manipulate the perceived levels of attention on a subject [8].

As technology has continued to be refined, so have the methods used for tracking, identifying and/or measuring attention in different settings. Of particular interest for the development of our proposed system are a number of methods that measure attention based on the position of a student's head while he or she is performing a particular task, using technology to capture head pose lean as an indication of eye gaze. For instance, Ba and Ordobez attempted to recognize the Visual Focus of Attention (VFOA) in the context of a meeting by tracking head pose as an indication of visual focus through the use of a geometric model that allows their system to correlate head pose to visual gaze [9]. By doing this they could determine what participants were directing their attention to during meetings. Similarly, Ishii et al. studied the relation between eye gaze and attentional focus during conversations using eye gaze duration, eye movement and pupil size as effective variables to track for this purpose [10].

Even more relevant to our methods for measuring attention, Doshi and Trivedi found head pose to be a clear indication of directed attention by asking participants to describe where their attention was going to be directed at a particular moment, and by stimulating unconscious attention [11]. In this process, the researchers found that head movements often preceded eye gaze when shifting attention between different objects and stimuli. Moreover, Asteriadis, Karpouzis and Kollias conducted a study that highlights the importance of taking into consideration both head pose and eye gaze when tracking attention by exploring the ability of intelligent systems to replicate a human's perception of attention [12]. In their paper they studied a series of annotations from the University of Boston dataset, which was gathered from a series of participants on perceived levels of attention from a number of pictures of subjects who were engaged in a particular task. Based on their exploration of this subject, they concluded that both eye gaze and head pose play an important role in determining attention. For instance, large head pose and small eye gaze were associated with low levels of attention, as opposed to large head pose and large eye gaze, a combination that was associated by participants as an indication of high levels of attention.

While these studies were not conducted in the context of an online course taken individually by a student, they provide relevant information that can be used for the design of our system. For instance, Stanly conducted a study in which he attempted to predict user attention by using the Microsoft Kinect to capture a wide array of variables, including body lean, head pose, position of hands and audio. For the head pose variable, he used yaw, pitch and roll angles to demine eye gaze and what the users were looking at while performing a test [13]. These variables were found to be effective in determining whether users were attentive or inattentive after comparing test results to perceived focus of attention.

3 Research Tools and Methods

The proposed system will be an extension of Open edX, the open source platform behind edX. EdX is a non-profit online site created by Harvard and MIT that offers online courses and MOOCs from some of the world's most renowned universities and institutions [14]. Because Open edX is open sourced, developers can extend the source code and add new features to any course for use by instructors and educators who use edX's free platform for building their own interactive online courses.

For tracking head pose and estimating eye gaze, we will use the Intel RealSense camera and developer kit. The Intel RealSense camera is able to capture x, y and z coordinates, and the development kit provides a set of libraries that can be used to track a number of inputs, which includes head pose, as illustrated in Figure 1. Another important reason for using the Intel RealSense camera is the fact that Intel plans to incorporate the same technology in laptops and mobile devices in the future, which would allow systems such as the one proposed in this paper to be commonly used to enhance the online learning experience in MOOCs in the future.

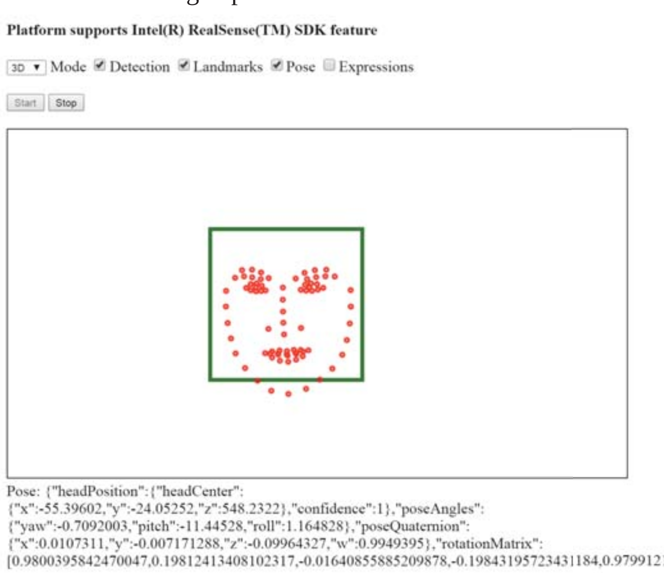

Figure 1. Intel RealSense Image Processing Example

Gathering of head pose data will be accomplished using JavaScript and JQuery frameworks provided by Intel for image processing. The data gathered from the camera will be sent to a virtual machine server that contains all back end code used by Open edX, which uses Python and Django to process the data in order to generate reports on attention levels for each student.

It is important to note that, while Open edX provides instructors and course creators with a wide array of tools that can be used to deliver instruction in several formats, for this system, we will focus on gathering attention data only when students watch instructional videos that are part of an Open edX course developed for this study. This will help instructors judge the quality and value of the instructional videos they decide to create and/or use for their own courses.

Each report generated by the attention tracking system will contain the following information:

1. URL and title of the instructional video that is played while data on attention levels is gathered.
2. Perceived levels of attention, based on perceived eye gaze using head pose data gathered from the code running in the browser. Data visualization techniques will be used in order to provide a historic, graphical representation of attention for individual students enrolled in a particular class. This information will inform instructors of individual attention levels for each student, which could help identify students who may have difficulties understanding certain content.
3. Group attention data, also represented using data visualizations tools and techniques. This will allow instructors to compare the effectiveness of different instructional content based on the data gathered from all students enrolled in a course.

Attention will be calculated based on whether it is perceived

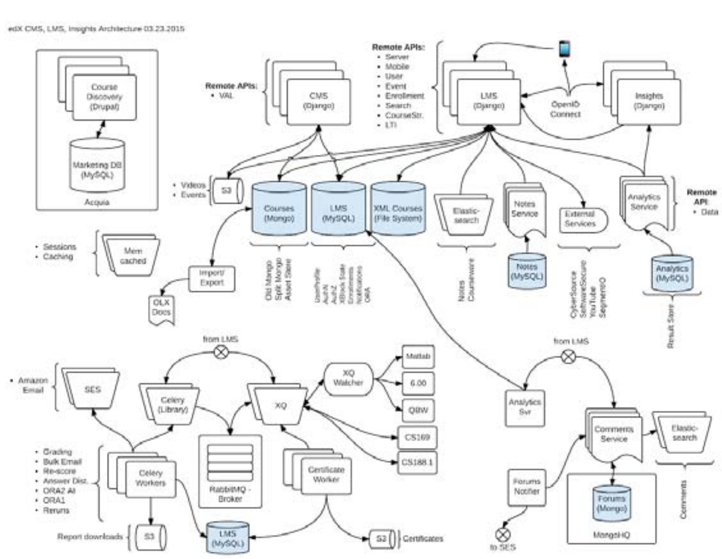

Figure 2. edX CMS, LMS, Insights Architecture [15]

that a student is directing her/his eye gaze towards an instructional video. Because the values for yaw, pitch and roll detected by the Intel RealSense camera all equal 0.0 when users are staring directly at the camera, attention will be calculated based on how these values deviate from 0.0 during the duration of instructional videos. Additional calculations may be required during development.

Lastly, a custom Open edX course is currently in development, which will be designed to work with and used to evaluate our system. The course will contain a series of instructional videos, each followed by a short test for evaluating the content taught in each of those videos.

4 System Architecture

Open edX is composed of multiple components, each serving a different purpose for the creation, storage and delivery of online courses as shown in Figure 2. Our goal is to build our system as an extension to these components in the Learning Management System (LMS) as depicted in Figure 3. The system will be modeled to fit the existing Open edX architecture, which uses a server-client architecture to deliver course content.

The architecture will rely on the following:

1) Browser modules that make use of Intel's RealSense JavaScript libraries for capturing head pose data throughout the duration of instructional videos. The data captured will then be sent in JavaScript Object Notation (JSON) format to the Django server for processing. Other modules in the browser will be responsible for visually processing attention data from the server (as charts) in an instructor module containing attention reports.
2) An xBlock module that processes the head pose data sent from the browser. The data is processed in the Django LMS server and stored as part of a student's performance history in a MySQL database.
3) Head pose data and attention levels are associated with videos watched while data was gathered in a MySQL database. The server extracts attention and head pose data from this database every time a client requests an attention report.
4) All course data is extracted from a Mongo database.

There will be two client modules responsible for preparing head pose data for processing. One module will primarily make use of the RealSense camera to capture raw head pose data. This module will then pass data to a filter module that will determine whether head pose values should be recorded or not based on variance values that will be established in the early stages of development. A calibration video, which will ask students to go through a series of steps (such as directing their gaze in different directions) will also be developed and used to establish what changes in head pose values should be considered relevant and how often head pose data should be saved for processing at the server level. This is done to limit the amount of data that will need to be sent to the server. We are exploring options for measuring attention data at the client (browser) or server level based on criteria including impact on performance and what happens when a user changes the page in the middle of a video. A possible machine learning module could also adjust head pose values as needed based on test scores and history data. Once data is passed to the server, the system will estimate attention levels and record results in a database and return an adjusted standard deviation score. A separate data visualization module then displays attention reports to instructors.

Because the LMS is built with Django, our system will rely on the Model View Controller (MVC) pattern to extend the functionality of edX as a unit called xBlock, a term used by edX to describe Python classes that are used to build small edX web applications.

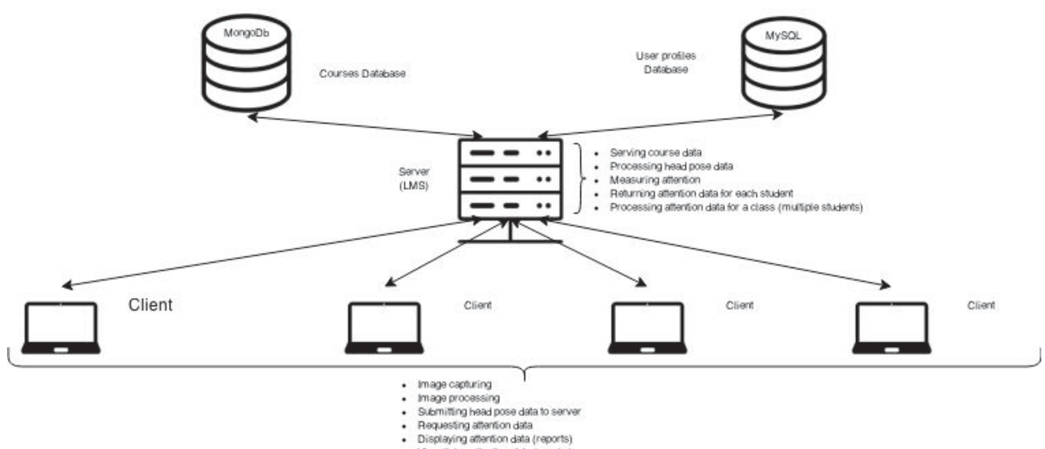

Figure 3. High-level System Architecture

5 Additional Features

The following features will be added to the system as time permits:

- Attention could be tracked while students use other course content or resources, such as when taking quizzes or going though practice exercises for subjects like math and computer science.
- Add a learning agent that further individualizes results by enhancing the accuracy of perceived attention levels based on historical data for each student and comparing perceived levels of attention with grades for individual modules.

6 Discussion

Capturing attention data with respect to specific online educational content allows instructors to further understand the needs of their online students. This could in turn improve the online education experience for teachers and students, in particular for students whose needs prevent them from attending a physical classroom. For instance, a report regarding attention levels gathered for an entire class may show that a majority of students seem to be more than 75% attentive for a particular video. In this case, it is to be expected that the overall scores for the evaluation associated with that video will be high. However, if only a small number of students show less than 50% of attention and low scores for the video in question, the instructor may determine that the content in that video may not be useful for those particular students. In this situation the instructor may then recommend that those students review additional supporting materials or use different instructional content that could potentially engage students more effectively.

Furthermore, we expect that the information that our system provides will be used as an indication of the quality and appropriateness of certain educational content. For instance, if an instructor sees that the overall attention levels for a particular video are below 50%, and the scores for the associated evaluation are low as well, the instructor may deem the video ineffective or not appropriate for the content that is being evaluated. In this case, the instructor may decide to create a new video that will be able to capture students' attention more effectively.

The above scenarios will of course require the judgment of the instructor when determining whether the instructional content is a cause for low attention levels, if some students require additional help to understand and follow that content, or if the attention data is not reflective of the quality of instruction at all. Critical judgment will require consideration of attention data and evaluation scores for individual students and for a class a whole. However, it would be helpful to add a machine learning component that makes this type of determination for the instructor. The component could advise the instructor on whether a video or instructional resource is ineffective or if one or more students need further help to understand specific content. Such a machine learning component is a potential direction for future work.

7 References

[1] Allen, I. E., & Seaman, J. (2013). *Changing Course: Ten Years of Tracking Online Education in the United States*. Sloan Consortium. PO Box 1238, Newburyport, MA 01950.

[2] Muzafarova, T., & Kaya, E. (2015). Survey of Awareness of Massive Open Online Courses (MOOC)–a Case of International Black Sea University Students, Georgia. *Journal of Education*, *3*(2), 15-19.

[3] Hays, R. T. (2006). *The science of learning: A systems theory approach.* Boca Raton, FL: BrownWalker Press.

[4] Halpern, D. F., & Riggio, H. R. (2003). *Thought and knowledge: An introduction to critical thinking*. Mahwah, NJ: Lawrence Erlbaum Associates.

[5] Geisler, W. S., & Cormack, L. K. (2011). Models of overt attention. *The Oxford Handbook of Eye Movements*. Oxford, NY: Oxford University Press.

[6] Tuckman, B., & Monetti, D. (2012). *Educational Psychology with Virtual Psychology Labs.* Belmont, CA: Cengage Learning.

[7] Sylwester, R., & Cho, J. Y. (1993). What Brain Research Says about Paying Attention. *Educational Leadership*, *50*(4), 71-75.

[8] Geissler, L. R. (1909). The measurement of attention. *The American Journal of Psychology*, 473-529.

[9] Ba, S. O., & Odobez, J. M. (2009). Recognizing visual focus of attention from head pose in natural meetings. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, *39*(1), 16-33.

[10] Ishii, R., Shinohara, Y., Nakano, T., & Nishida, T. (2011). Combining multiple types of eye-gaze information to predict user's conversational engagement. In*2nd workshop on eye gaze on intelligent human machine interaction*.

[11] Doshi, A., & Trivedi, M. M. (2009, June). Head and gaze dynamics in visual attention and context learning. In *Computer Vision and Pattern Recognition Workshops, 2009. CVPR Workshops 2009. IEEE Computer Society Conference on* (pp. 77-84). IEEE.

[12] Asteriadis, S., Karpouzis, K., & Kollias, S. (2011, May). The importance of Eye Gaze and Head Pose to estimating levels of Attention. In *Games and Virtual Worlds for Serious*

Applications (VS-GAMES), 2011 Third International Conference on (pp. 186-191). IEEE.

[13] Stanley, D. (2013). Measuring attention using Microsoft Kinect. *Rochester Institute of Technology.* Retrieved from http://scholarworks.rit.edu/cgi/viewcontent.cgi?article=5772&context=theses.

[14] About Open edX. Retrieved April 14, 2015, from https://open.edx.org/about-open-edx.

[15] edX CMS, LMS, insights architecture [Online image]. (2015). Retrieved April 12, 2015, from https://open.edX.org/contributing-to-edX/architecture.