

OING: OpenING the third eye of the lecturer

Jungkook Park, Youngjae Chang, Taegyeong Lee, Hyunwoo Kim

1. Introduction

Delivering a lecture is easy: any kind of talk in front of audience with some information could be called as a lecture. However, delivering a good one is totally different story. To convey knowledge well, lecturers should consider a number of factors, including the attention level of the students. The problem is that, the lecturers are usually overwhelmed by the lecturing activity itself, and so they do not have enough time and effort to care about the class participants. This leads to a tragedy where everyone except for the lecturer is sleeping throughout the lecture.

Our goal is to provide an adversary view of the audience thus lessening the burden of the lecturer. For example, a simple visualization of the students' attention could help lecturer to keep in mind the timely change of the audience engagement. Glancing at the visualization would require less effort and attention of the lecturer than observing the whole class to find out the attentive level.

To achieve this goal, we should be able to capture the attention signals which is not a trivial job. A student can take various actions along the lecture: looking at the lecturer, reading the textbook, or taking some notes. On the other hand, student can be in a daze, chat with others, or surf the web with his/her personal device. We should also carefully formulate this noisy information on attention level to an useful infographic the lecturer can easily digest. The lecturer should be able to perform an appropriate action based on the output, for example making a joke or move to the low-attentive region.

Here, we propose Oing, a system to measure the attention of the class in real time and provide a visualization of the attention as a heatmap to help lecturer make appropriate action to handle low-attentive region of the classroom.

2. Related Works

In this section, we describe related work in two distinct research areas: *vision-based attention analysis* and *lecture analysis*, and how our research contributes to these areas.

2.1. Vision-based attention analysis

Gaze detection have been used as a primary approach on measuring classroom attention. Bidwell (2011) measured student engagement using gaze tracking. In his work, a third grade classroom was recorded using five cameras and four Kinects. Gaze targets of each student were extracted, and expert observed each student every 20 seconds to determine her engagement. Based on the data, student engagement classifier was trained using Hidden Markov Model. However, it ended up with unsatisfactory result requiring more observation data and a revised set of states for Hidden Markov Model.

Recent technological development suggests the possibility of a gaze tracking of multiple users using only a single camera. Sugano (2016) presented AggreGaze which collectively estimates audience attention on public displays. AggreGaze used face angles, locations of audience and pupil positions to predict the region gaze are settling on the public display. The gaze prediction of each viewer are later collectively analyzed to reduce the deviation of erroneous prediction. The heatmap generated from such collection shows high accuracy comparable to other intrusive approaches.

Raca have conducted several research on students attention analysis using the vision system. His focus was on measuring the quality of the lecture; Raca (2014) have suggested the students' timing of body motion as a measure of attention. He have proposed the idea of sleeper's lag, where the motion of the students is lagged compared to the others, the attention of according student is low.

2.2. Lecture analysis

Delivering a good lecture was always an important topic for the lecturers. Reid (1948) pointed out some considerable factors to improve lecture: lecturer's personality, examples, delivery of the lecture, presentation forms and student comments. From those points, he emphasized the importance of drawing students' attention and appealing to students. To achieve it, he suggested using good introduction to the lecture, using specific examples, making the lecture more audible, preparing lecture outlines, and taking students' comment.

Bain (1993) suggested similar points. He mentioned that the some of the main advantages of lecture are using various channels, including verbal communication, facial expressions, and gestures, and opportunity for students to ask questions.

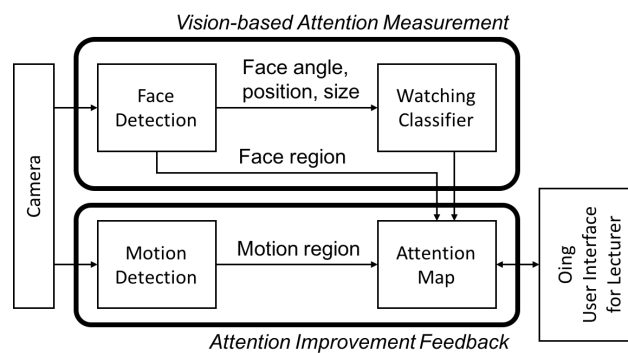
For better delivery of the lecture, the lecturer should exploit these points by increasing interaction with students. He also suggested some possible approaches, for example, providing outline notes, preparing in-class pair activities, and raising questions for the main point of the lecture.

3. Preliminary Study

There were several protocols to observe and analyze the lecture quality. For example, Sawada et al. suggested Reformed Teaching Observation Protocol (2002) and Hora et al. suggested Teaching Dimensions Observation Protocol (2013). Though most of them were focused to the lecturer's skillset, they suggested some common sign of attentive students: taking notes and looking at the lecture material or the lecturer. So, we tried to use those behavior as a sign of attention.

Also, to get some meaningful insight about the lecture on the lecturer's side, an interview with researcher at Center for Excellence in Learning and Teaching was done. Her main focus was also on the lecturer, not students. Still, she suggested that lecturer using space well, for instance, speaking in front of the students with low attention, could improve the attention of the students significantly.

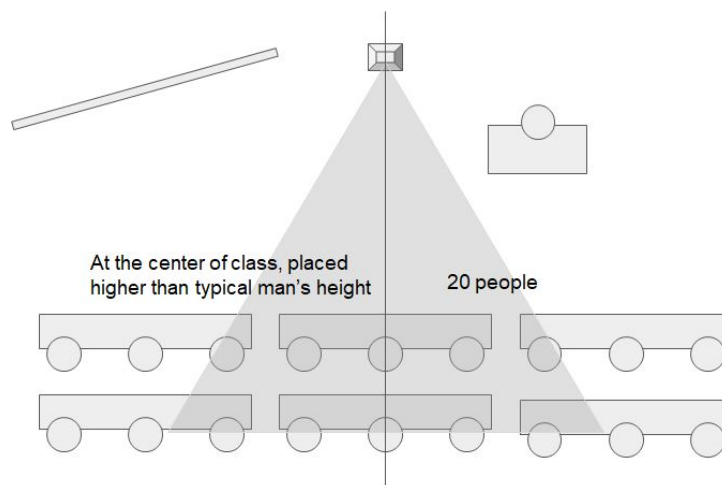
4. Oing: Vision-based Attention Tracker



[Figure 1] System Architecture

Now we present Oing, a real-time vision-based attention tracking system for the lecturer. As shown in Figure 1, Oing consists of two part, (1) vision-based attention measurement and (2) attention improvement.. Oing measures students' attention by using a single camera in front of the classroom and visualize classroom dynamics and attention distribution according to classroom space. The following parts explain about each part of the Oing system.

4.1 Installation environment



[Figure 2] The camera will be placed at the top center of the classroom, giving a sight of at least 20 students.

We have intentionally limited the target environment of the system to typical lecture halls which holds up to hundreds of students. The lecture hall setting provides us two benefits. Firstly, the fixed configuration of the classroom limits the actions that students can perform during the lecture. This makes it easier to capture the attention signals of the students. Secondly, the pitched floor provides a better view of the audience, thus we can capture more students with a single camera.

4.2 Vision-based Attention Measurement

As we mentioned in Section 3, the student's behavior of looking at the lecture material or the lecturer is one of the common signs of their attention. Therefore, we adopted students' faces in the classroom as an attention signal. As shown in Figure 1, Oing utilizes a camera located at the top-front of the classroom and captures a picture of the students once for every 1 second during the lecture. The captured image is fed into the face detection algorithm of Google Cloud Vision API to extract the features of face angle, face size, and face location in the image.

Based on the results from the face detection model, we found that there are many non-attentive students whose faces are detected. Therefore, we designed a "watching" classifier to classify whether the detected faces are actually looking at the lecture materials or the lecturer or not, which is the attention signal we target. The watching classifier utilizes features extracted from face detection and trained with 30,000 of manually labeled face data. As a result, our random-forest based watching classifier successfully classified watching faces with 0.89 accuracy.

4.3 Attention Improvement

Based on the attention signals detected through vision processing, we then designed the interface to promote the attention level of the students. We first enumerated commonly used techniques teachers use during the lecture. Most of the techniques focused on the improvement of contents, hence we decided to give feedback on the behavioral activity of the lecturer.

We decided to build an interface that encourages the lecturer to move around the lecture hall during the lecture. As we mentioned in the preliminary study (Section 2), this movement is especially effective on large classes. Changes in lecturing position help re-focusing students' attention, yet this movement should be purposeful not to distract them.

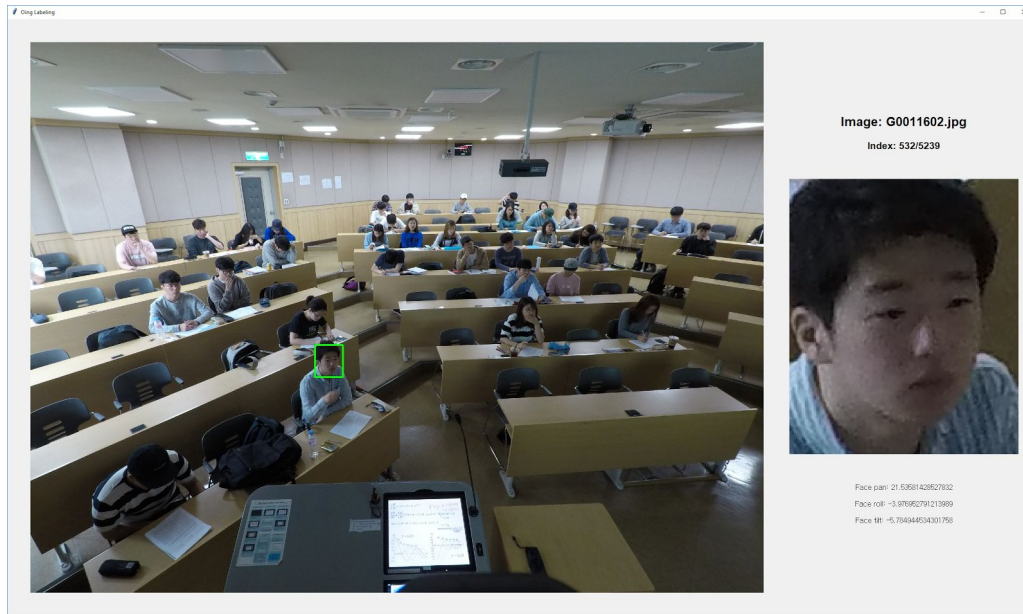
Aided by the position of watching faces, we generated the attention heat map upon the live video of the classroom. The lecturer will be directed to positions where students are showing low attention for certain amounts of time. We fine-tuned the system to highlight the region where attention is being dropped the most. The detailed implementation of the system will be discussed in Section 6.

5. Data Collection

To make the system able to collect attention signals from students in the classroom, we first collected training data, which consists of several pictures of the classroom captured at the front of the room at every second, from a real lecture environment. We installed one GoPro Hero 5 camera above the lecturer's table in the room in KAIST E11 building. After all data collection was done, we applied the face detection algorithm to all collected pictures.

Data was collected from three classes. The first one was a major required course at 9:00, with 68 students enrolled. The whole class was recorded, producing 4,257 pictures with 35,830 faces. The second one at 10:30 was a basic required course with 86 enrolled students. The lecture was recorded about 45 minutes, producing 2,687 pictures with 37,150 faces. The third one at 14:30 was also a basic required course, with 92 students. The lecture was recorded for about 26 minutes, giving 1,582 pictures, with 12,604 faces.

To generate ground truth data for students' attention, the detected faces of the students were analyzed manually. (Figure 3) If the student was seeing the lecturer or the lecture slide, she was classified as an attentive student. In other cases, she was classified as not attentive. About 30,000 faces were classified manually.



[Figure 3] Data Labeling Interface

6. Implementation

OING consists of two parts - measuring part and improving part. In the first measuring part, we design a system to measure attention signals from captured from a single off-the-shelf camera in real time. In this part, we use vision API provided by Google Cloud to detect faces from the pictures, and transform detected faces to attention signal by applying “watching classifier” trained by manually labeled real-classroom dataset. In the second part, we design two different systems to give feedback to the lecture by utilizing attention signals measured from the first measuring part.

6. 1. Measuring part

The main objective of measure part is to capture attention signals from pictures gathered from a single off-the-shelf camera. From the preliminary study, we found students’ face as the main source of attention signal, and therefore we developed the system that extract students’ face in the classroom using face detection algorithm.

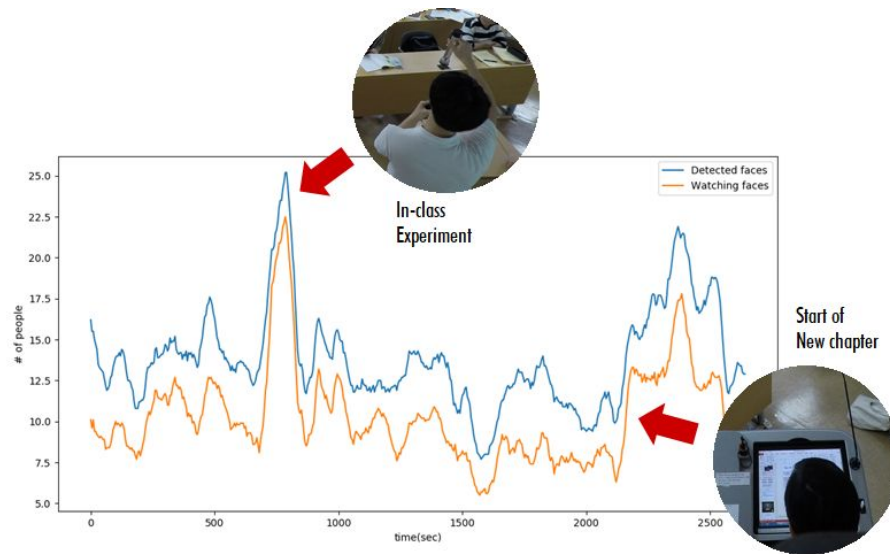
We first tried to implement face detection algorithm using the Haar Feature-based Cascade Classifier provided by OpenCV projects, however, it was not successful due to the physical characteristic of classroom - the classroom is horizontally wide and thus only side of face is captured - that haar feature-based algorithms are not good at. We also tried various types of algorithm and tweaks, and finally decided not to implement face detection algorithm by ourselves, rather, make use of Google Cloud Vision API which provides deep neural network based face detection algorithm.

After detecting the faces from the pictures, we manually labeled each of detected faces whether they are watching lecture materials or not. These labels are used to train a classifier to determine whom “watching” the lecturer. We tried SVM classifier and Random Forest, and concluded on the random forest for since the model fits better to our case. We then perform a Grid Search to find the best model parameters. The chosen random forest classifier consists of 33 trees with the depth of at most 9.

feature	x	y	box_x	box_y	yaw	roll	pitch
importance	0.116	0.352	0.124	0.118	0.083	0.103	0.105

[Table 1] Feature Importance of the Random Forest Classifier

Table 1 shows the feature importance of each feature from the random forest classifier. It suggests that the y-coordinate of the face is important. Low y position of the face means that the student is sitting at the front. Also, students at front are more likely to pay attention to the lecture. So, high importance of y-coordinate would be justifiable. All other factors has similar importance. Those factors are somewhat related to the position of the student in the same row, the similar importance could be interpreted.



[Figure 4] Validity of Attention signals

We qualitatively validated our data with our intuition to check the validity of the gaze as an attention signal. Figure 4 shows the number of faces detected from Google Cloud Vision API (blue line) and the number of faces classified as “watching” (orange line). It was observed that the two numbers have similar tendency. Also, there were two peaks in the graph, meaning large number of people seeing front. The first one corresponded to an in-class experiment, and the second one corresponded to the start of a new chapter. Both events would cause high attention of students. This could be an evidence to the correlation between the attention level and the number of faces watching front.

6. 2. Improving part

After successfully measuring the attention signals, we built two prototype systems to give a feedback to lecturers. The objective of feedback is mainly on giving spatial-temporal information where the lecturer should pay more attention to regain overall classroom attention. We adopted two different feedback approaches on each system, the first approach is to give information about high-attentive students, and the other approach is to give low-attentive students.



[Figure 5] Prototype interface visualizing who are paying attention for more than 2 minutes.

The first approach gives an overlay view of the classroom to the lecture by highlighting on high-attentive students, while hiding low-attentive or irrelevant regions. This gives a quick understanding on the temporal attention level of overall classroom, so that lecture can catch the insight about when and where he/she should refresh the classroom to regain the attention. To enable this approach, we implemented heatmap-based masking algorithm on attention signal data, and also motion-detection based making algorithm to hide irrelevant regions (such as desks, walls, etc.).

However, to improve the attention of the class, the lecturer should try to regain the attention of students losing their attention. So, we changed our design approach to help lecturer focus more on those audience.



[Figure 6] Prototype interface visualizing a low-attentive region.

In the second approach, students losing attention are visible. In detail, if a student showed attention signal at least once during last 10 minutes, and they did not show attention signal during last 3 minutes, their face is visible on the screen. Students not paying attention at all, or students actively listening the lecture are not shown. We thought the lecturer did not need to do anything, and could not do anything for those students.

7. Conclusion

In conclusion, we developed a simple way to estimate class' attention from image obtained from a single camera. First, the class attention was estimated as the number of people watching front. Its plot versus time showed that our proposed attention signal -- faces watching the front -- could be used as an estimator for class attention. The number of faces watching front was calculated using random forest classifier trained by 30,000 faces.

Also, we suggested a simple visualization to induce lecturer's focus to students losing their attention based on the attention measurement. Using the visualization, we expect lecturers to care for their students drifting away from the lectures more easily.

Our system, however, have some limitations. To measure the attention in real time, the trained classifier should be integrated with the real time images from the camera. Also, the classifier should be trained with more data to make it less sensitive to the camera environment, including the position and angle of the camera installed. Integrating multiple cameras would be required to detect all students in the classroom. Then, students out of the camera angle and students whose faces were not detected due to camera angle could also be detected, leading to higher confidence. Better visualization of the students' attention, reflecting the user's opinion, would be an essential part for fully functional Oing system.

We believe Oing can be a good initial step towards the interface for offline groups. Offline gatherings often happens in our everyday lives, but we have no tools to help we succeed on such gatherings. A lecture or a seminar is a good example of such event; huge audience gather on small physical location, but the instructor is the only one to lead them all during the lecture hours. Giving them a support with computing technology will improve the whole experience.

8. References

- [1] Bidwell, Jonathan, and Henry Fuchs. "Classroom analytics: measuring student engagement with automated gaze tracking." *Behav Res Methods* 49 (2011): 113.
- [2] Sugano, Yusuke, Xucong Zhang, and Andreas Bulling. "AggreGaze: Collective Estimation of Audience Attention on Public Displays." *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*. ACM, 2016.
- [3] Raca, Mirko, Roland Tormey, and Pierre Dillenbourg. "Sleepers' lag-study on motion and attention." *Proceedings of the Fourth International Conference on Learning Analytics And Knowledge*. ACM, 2014.
- [4] Reid, L. R. How to Improve Classroom Lectures. *Bulletin of the American Association of University Professors*, 34(3), pp. 576-584. http://www.jstor.org/stable/40220313?origin=JSTOR-pdf&seq=1#page_scan_tab_contents
- [5] Bain, K. How Do I Improve My Lectures. Retrieved from <https://www.montclair.edu/academy/resources/teaching-resources/how-do-i-improve-lectures/>
- [6] Sawada, Piburn, Judson, Turley, Falconer, Benford & Bloom. Measuring Reform Practices in Science and Mathematics Classrooms: The Reformed Teaching Observation Protocol. *School Science and Mathematics*, 102(6), pp. 245-253. DOI: 10.1111/j.1949-8594.2002.tb17883.x
- [7] Hora MT, Oleson A, Ferrare JJ. Teaching Dimensions Observation Protocol (TDOP) User's Manual. Madison: Wisconsin Center for Education Research, University of Wisconsin–Madison; 2013. <http://tdop.wceruw.org/Document/TDOP-Users-Guide.pdf>

