Heat Metaphor for Attention Estimation for Educational VR

David Broussard   Christoph W. Borst

CACS VR Lab — University of Louisiana at Lafayette

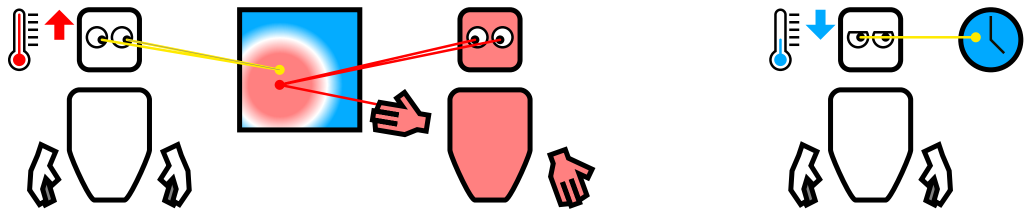

Figure 1: Heat metaphor: The teacher transfers heat to a looked-at object and students heat or cool based on where they are looking. Attentive students' heat can also transfer to objects (not illustrated). Heat is not directly visualized to users.

Abstract

We prototype a technique, for educational VR applications, to estimate each student's level of attention in real time. Our system attaches scores to both students and objects, which change in response to eye-tracked gaze intersections. Compared to a simple angle-based approach, our system provides a dynamic and granular representation of object importance and frees the lesson designer from having to fully define objects of interest and timings. Our system takes into account simultaneous behaviors of multiple students and filters out brief behavioral deviations of attentive students. The results may help a teacher or a virtual agent better guide students.

Index Terms: Human-centered computing—Virtual reality; Applied computing—Education

1 Introduction

In educational VR, it can be challenging for a teacher to remain aware of the condition of students [8]. In person, teachers observe cues such as eye gaze, body movement, and facial expressions to gauge audience awareness and understanding. However, most VR environments provide coarse information. Increasingly, VR devices include eye tracking and other sensors, supporting increased cues. Automated and simplified ways of observing these inputs may improve teacher guidance or automated aids for students.

Our approach manages a score or "temperature" for each object and user to help estimate if a student is distracted. The student's own score is dynamically adjusted based on the target of their gaze. This estimated attention score is combined with information about actions the student is performing, and the aggregated information could be used to control the display of visual indicators for a teacher [2], the presentation of attention-restoring cues to students [9], or an autonomous educational agent [6].

2 Related Work

There is substantial research using eye-tracking to detect student disengagement, distraction, and inattention. Much of it incorporates desktop eye tracking into non-immersive tutoring and distance learning software. D'Mello et al. designed an intelligent tutoring system that attempts to detect boredom and disengagement using student gaze tracking [3]. The authors note the impossibility of defining a set of acceptable gaze behaviors for every possible state, so the system separates the screen into several zones (one for the tutor, one for an image, etc.) and triggers a gaze-reactive intervention when the student looks away from the tutor or image for more than 5 seconds while the tutor is speaking. Wang et al. describe an "eye-aware" educational agent using eye movements and pupil dilation for broader inferences about the state of a student [7], which follows from research suggesting that pupil dilation can occur in response to increased attentional effort [4].

There are fewer prior approaches in the context of educational VR. Gaze-reactive hotspots were generalized by Khokhar et al. to allow a virtual agent to detect and respond to inattention based on gaze angle [6]. Such hotspots must be set up and positioned at objects of interest by a lesson designer in advance, although it is noted briefly that they could be partially automated based on teacher actions like pointing. We build on this observation by automatically deriving gaze targets and heat from the behavior of the teacher (and in some cases, sufficiently attentive students).

Deep learning approaches to detect distraction from eye tracking in VR have suggested an accuracy of about 90% [1] [5]. Although this produced promising results offline with individual student sessions, it was not validated in real time for classes with multiple students. Our system, in contrast, uses an intuitive rule-based algorithm that is both computationally cheap and easily extended to real-time data from many students. Our system takes into account the current behavior of other attentive peers when scoring students.

3 Scoring

The key component of our system is a dynamic attention score, estimating student attention level. The underlying technique is based on what a student is looking at. For instance, looking at the teacher or a whiteboard likely indicates attentiveness, while staring at walls or outside a window could indicate distraction.

The system assigns to each object a dynamic relevance score, which can be thought of like a temperature (Figure 1.). The teacher, for instance, is always relevant, and thus an active source of "heat". The teacher can transfer relevance to an object by looking at or pointing to it, causing it to quickly "heat up". Otherwise, objects will naturally "cool off" with a gradual exponential decay. Objects, in turn, transfer this "heat" to students that are looking at them, thus increasing their attention scores. Students, too, "cool off" when looking at irrelevant objects. Students with a high attention "temperature" are considered attentive, and below a certain temperature

This is an author-formatted version. Original publication: D. Broussard and C. W. Borst, "Heat Metaphor for Attention Estimation for Educational VR," 2023 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), 2023, pp. 683-684, doi: 10.1109/VRW58643.2023.00184.

threshold, students could be considered potentially distracted.

Note that the relevance information maintained by each object is not intended to be directly visualized: It is maintained internally to support the final per-student estimates, which may be visualized.

This transfer of relevance is a two-way process, but the influence of a single student is generally not enough to significantly affect an object's relevance. An exception is that when multiple mostly-attentive students look at the same object, the object's relevance can increase substantially. This behavior is intended to prevent a teacher from having to stop to explicitly look at or point at objects to mark them as relevant as long as most of the class is following along to the teacher's guidance. It is unlikely that a large number of students will suddenly simultaneously look at an object that the teacher does not intend as relevant. But, if multiple students become distracted simultaneously by a single source, it becomes marked as relevant. Thus, in general, the relevance and attention scoring system deals with attentional anomalies (deviation from the norm), even if the behavior of the class is not necessarily desired by the teacher.

Per-object relevance score alone results in some drawbacks; very large objects must have a single relevance score corresponding to the entirety of the object, even if only a part of it is relevant. For instance, a whiteboard might potentially take up a large amount of space, but only has a single currently important region edited or referenced by a teacher. Older regions may no longer be relevant, but under this system, students looking at these regions would still be looking at an object which is, in total, "relevant".

A more robust solution is to store heatmaps across all object surfaces to track how relevance accumulates over individual regions of each object's surface. When relevance is transferred to an object, instead of the entire object "heating up", only a small region centered on the gaze intersection position increases in relevance, and the relevance of the rest of the object continues to decrease over time. Using a heatmap-based system of object relevance generalizes fine-grained relevance scoring to all sizes of objects, ensuring very large objects never become fully relevant inadvertently without requiring the environment designer to consider lesson specifics ahead of time.

As a simple approximation, our prototype does not maintain a full heatmap for each surface; rather, each object is given a single dynamic "hotspot" and a small radius of relevance around it. Only when a student's gaze vector is within a certain angle of the vector to the hotspot are they considered to be looking at the object, and the hotspot moves in response to sources of relevance.

4 Extensions to Scoring

In addition to solely using each student's estimated gaze target for the purposes of attention scoring, the speed of saccades is detected and used as an input to the attention scoring algorithm. Related literature suggests that slower saccades may indicate distraction or mind-wandering behavior, and thus faster saccades may indicate possible higher attentiveness. Thus, saccade speed is used to modulate the attention decay time of the student. Additionally, pupil dilation is reported by the eye tracker, and dilation is known to sometimes occur in response to attentional effort.

Attention scoring is further augmented with the addition of a per-object relevance history. Each object keeps track of a list of the most recent student gaze intersections, specifically recording the length of each student's continuous gaze and the attention score of each student. With relevance history, objects start out losing relevance quickly, but as more attentive students look at the object, the object retains its relevance score for longer.

Further improvements can be made by leveraging an additional per-student attention history, which records the gaze target and gaze length for each recent gaze intersection. This could be used, for instance, to decrease the influence that an object's relevance has on a single student's attention score if the student has been staring at it for too long. A student with a history of paying attention could even be given more of an influence over object relevance.

5 Conclusions and Future Work

Our gaze-based student scoring system uses a real-time, dynamic model of object relevance to maintain an estimated attention score for students, without the need of a teacher or lesson director to plan in advance which objects are important at which times. The response of our system to gaze deviations has only been informally validated during development, using mostly single or simulated students. Thus, there is still a need to evaluate the approaches with real class-like deployments.

The use of tracking-based attention estimation raises ethical concerns about data privacy and potential misuse. The former can be mitigated by allowing students to control what types of sensor information (if any) they share, and by ensuring they understand how it can be used. The latter might be mitigated by careful design of any monitoring or visual indicator system that recieves or presents student scores, to encourage teachers to only use filtered or aggregated information as an instructional aid rather than rating or punishing "misbehaving" students. For autonomous educational agents, the potential for such misuse can be reduced.

We are conducting research to quantify temperamental information using additional sensing such as the physiological sensors and EEG. We anticipate this information could be integrated into our broader educational VR framework and scoring system, to enable a more complete consideration of student conditions.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. 1815976.

References

[1] S. M. Asish, E. Hossain, A. K. Kulshreshth, and C. W. Borst. Deep learning on eye gaze data to classify student distraction level in an educational vr environment. In *ICAT-EGVE 2021 - International Conference on Artificial Reality and Telexistence and Eurographics Symposium on Virtual Environments*, 2021. doi: 10.2312/egve.20211326

[2] D. M. Broussard, Y. Rahman, A. K. Kulshreshth, and C. W. Borst. An interface for enhanced teacher awareness of student actions and attention in a vr classroom. In *2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 284–290. IEEE, 2021.

[3] S. D'Mello, A. Olney, C. Williams, and P. Hays. Gaze tutor: A gaze-reactive intelligent tutoring system. *International Journal of human-computer studies*, 70(5):377–398, 2012.

[4] B. Hoeks and W. J. Levelt. Pupillary dilation as a measure of attention: A quantitative system analysis. *Behavior Research methods, instruments, & computers*, 25(1):16–26, 1993.

[5] A. Khokhar and C. W. Borst. Towards improving educational virtual reality by classifying distraction using deep learning. In *ICAT-EGVE 2022 - International Conference on Artificial Reality and Telexistence and Eurographics Symposium on Virtual Environments*, pp. 85–90, 2022. doi: 10.2312/egve.20221279

[6] A. Khokhar, A. Yoshimura, and C. W. Borst. Pedagogical agent responsive to eye tracking in educational vr. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 1018–1019. IEEE, 2019.

[7] H. Wang, M. Chignell, and M. Ishizuka. Empathic tutoring software agents using real-time eye tracking. In *Proceedings of the 2006 symposium on Eye tracking research & applications*, pp. 73–78, 2006.

[8] A. Yoshimura. Remote instruction in virtual reality: A study of students attending class remotely from home with vr headsets. *Mensch und Computer 2020-Workshopband*, 2020.

[9] A. Yoshimura, A. Khokhar, and C. W. Borst. Eye-gaze-triggered visual cues to restore attention in educational vr. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 1255–1256. IEEE, 2019.