Using Eye Gaze to Differentiate Internal Feelings of Familiarity in Virtual Reality Environments: Challenges and Opportunities

Trevor Chartier, Iliana Castillon, Videep Venkatesha,
Anne Cleary and Nathaniel Blanchard

October 18, 2024

Using Eye Gaze to Differentiate Internal Feelings of Familiarity in Virtual Reality Environments: Challenges and Opportunities

Trevor CHARTIER, [a,1], Iliana CASTILLON [a]
Videep VENKATESHA [a]Anne M. CLEARY [a,2]
Nathaniel T. BLANCHARD [a]
[a] *Colorado State University, Fort Collins CO, 80523, USA*
ORCiD ID: Trevor Chartier https://orcid.org/0009-0003-6751-1905, Iliana Castillon https://orcid.org/0009-0000-0001-9762, Videep Venkatesha https://orcid.org/0009-0000-4635-3010, Anne M. Cleary https://orcid.org/0000-0003-1856-9070, Nathaniel T. Blanchard https://orcid.org/0000-0002-2653-0873

Abstract. Our group previously reported a feasible approach to detect the internal state of familiarity with eye-gaze features [1]. Utilizing an existing paradigm [2], we examined participants' feelings of familiarity during immersion within virtual reality (VR) scenes, some of which had had their spatial layout familiarized through prior presentation of a different scene with the same configuration. While immersed in a test scene, participants indicated the onset of familiarity via a button press on a handheld controller, then verbally indicated whether they could state the source of the familiarity or not. A potential issue is that machine learning models may have detected eye-gaze features reflecting the act of pressing the button rather than features associated with the internal state of familiarity. Although in [1] we addressed this challenge by including a buffer period between the button press and the window of data used for model training, it remains uncertain within what time frame features associated with the button press may persist. Here, we introduce an approach for potentially overcoming the confounding effects of the button-press by holding it constant. We examine machine learning models' ability to detect whether a scene's layout had been experimentally familiarized among only instances where subjective familiarity was reported. We then repeat this method for instances where no familiarity was reported. Finally, we examine experimentally familiarized scenes where familiarity was reported to detect recall-success vs. recall-failure for the familiarity's source.

Keywords. Machine Learning, Internal State Detection, Familiarity, Virtual Reality, Intelligent Virtual Tutoring Systems

[1]Corresponding Author: Trevor Chartier, trevor.chartier@colostate.edu
[2]Corresponding Author: Anne M. Cleary, anne.cleary@colostate.edu

1. Introduction & Related Works

Research on internal state detection centers on determining the reliability of machine learning models at identifying whether an individual is in a particular cognitive state based on the external physiological qualities they exhibit. Much of this research examines the automatic detection of mind wandering, an internal state often characterized by task-unrelated-thought. These methods have been mostly successful using data from various physiological features, including eye-gaze, skin conductance, and heart rate, with eye-gaze features being the most effective [3,4].

Some research has examined the detection of another internal state–the subjective sense of familiarity–from eye-gaze features [1]. Familiarity is a common subjective experience that is tied closely to recognition. It is thought to trigger a recall search in memory where individuals will attempt to identify the source of familiarity [5]. Our group previously employed popular machine learning algorithms to detect when participants were consciously experiencing familiarity within virtual reality (VR) scenes and found above-chance accuracy results [1]. The current work aims to expand upon our prior findings, including addressing some methodological challenges we previously encountered.

Establishing a ground-truth for familiarity experiences is a non-trivial task that requires some form of subjective report from participants. Often used in mind wandering studies [6], probe-based reports prompt participants throughout the experiment to periodically answer whether they are currently experiencing the given cognitive state. However, this method fails to capture the precise moment of onset for a cognitive state. We aimed to capture the moment of onset of familiarity by having participants indicate, via a button-press on a VR hand-controller, the moment something feels familiar in a VR scene. This study design posed unique challenges. Namely, eye-gaze features may correlate with the action of pressing the hand-controller button, thereby introducing a confounding variable to extracting the eye-gaze data that reflect the sensation of familiarity. Training models on this data set initially produced accuracy levels too high to fit in with prior research regarding cognitive state detection, suggesting that the models may have been picking up on differentiable eye-gaze patterns of intent to press the button in the moments preceding a button press. To negate this effect in our previous study, we introduced a 500 millisecond (ms) buffer period between the button press and the two second window of eye-gaze data collected. While this is a viable approach, there is currently little research to inform the correct buffer size to use in VR settings for removing eye-gaze features related to the button-press. As such, it is possible that residual gaze features correlated with the button press persisted beyond the 500ms buffer period and into the window used for model training; it is also possible that valuable eye-gaze data reflective of the state of familiarity were removed from model training.

In the current work we present another approach to dealing with this issue. We hold the VR hand-controller button-press constant while training the model on experimentally familiarized vs. unfamiliarized VR scenes. Based on prior research [2,7,8], a VR scene was considered experimentally familiarized if an otherwise novel scene containing its same spatial layout had been presented earlier in the experiment. First, we conducted a model search using eye-gaze features coming from the moments preceding a familiarity report in a scene to predict the experimentally-familiarized status of the given scene. Next we conduct a model search using eye-gaze features coming from randomly selected moments from within scenes that no familiarity was reported to detect, again, the experimentally-familiarized status of the given scene. Finally, we trained the models on experimentally familiarized test scenes where familiarity was reported via the button-press to compare instances where recall of the corresponding configurally similar study scene succeeded vs failed.

Beyond addressing the button-press challenges inherent in using a button-press to detect internal states via eye-gaze patterns in VR, we aim to show that distinct patterns in eye-gaze features emerge from different *levels* of familiarity when the subjective state is reported. Familiarity strengths are known to vary, where "a high degree of overlap between the features of the current situation and the features of previous experiences in memory produces a relatively strong familiarity signal" and"a low degree of overlap produces a relatively weak familiarity signal" [9,10]. As such, we might expect higher levels of familiarity with experimentally familiarized scenes than those that have not been experimentally familiarized, even among VR scenes for which the button was pressed to indicate a sense of familiarity.

Machine learning models have been shown to be able to identify previously seen images from unseen images using eye-gaze features, even in cases where participants failed to consciously recognize the previously studied images [11]. However, unlike previous research that examined when a participant viewed an image identical to one previously seen, we aim to detect when participants are immersed in novel scenes containing the same spatial layout as a previously viewed scene. Additionally, the current approach allows us to assess whether machine learning models are able to make this separation not only when participants consciously felt that the scene was familiar, but also when participants felt that it was unfamiliar, allowing for a direct comparison of model performance in each case.

2. Methods

We used the same dataset as [1], in whose study 26 undergraduate students from Colorado State University participated for course credit. As described in [1], the procedure from which the data were obtained closely resembled that of [2].

2.1. Data Collection

Eye-gaze data was measured at 120hz by the HTC Vive Pro Eye headset and collected using the SRanipal software development kit (SDK) version 1.3.6.8. Data was stored in comma-separated-values (CSV) files along with participants' recorded verbal responses and annotations, which included columns for familiarization/study status (whether a configurally identical scene was studied) and recall status (whether the studied scene was recalled or not following a report of familiarity).

2.2. Data Preprocessing and Model Training

We extracted eye-gaze data from 1, 2, and 3 second windows before each familiarity indication. Four buffer periods, ranging from 0 ms to 1000 ms, were included between the familiarity indication and the window of data extracted, consistent with [1]. Trials in which a familiarity indication occurred too soon into scene's onset to extract the given buffer and window size were excluded. For each positive report of familiarity, a negative instance of the same window size is extracted, from the same participant, from a scene in which no familiarity was reported. No buffer period is included for the negative instances, as no button press occurs during these. Eye-gaze features of interest were extracted using PyTrack. The result was 24 distinct datasets of eye-gaze features, with 12 datasets for positive instances and 12 for negative instances, for model training and comparison.

Across all detection tasks, model search was conducted with Hyperopt for hyper parameter optimization using random search over 300 training evaluations. Model algorithms used included AdaBoost, Naive Bayes, Logistic Regression, Support Vector Classifier, Random Forest, and K-Nearest Neighbors. Training and evaluation was done using Leave-One-Participant-Out-Cross-Validation. Cohen's Kappa was used to evaluate the model accuracy; specifically, we report the average across the held participants.

3. Results

During instances of positive familiarity reports, the Ada Boost model algorithm preformed best at differentiating between scenes that were experimentally familiarized and scenes that were not experimentally familiarized. This model resulted in a Cohen's Kappa score of 0.16 from training on 3 seconds of eye-gaze data, temporally separated from the button press by a 500ms buffer. Model results for all buffer and window sizes tested can be found in Table 1.

The same detection task, preformed on eye-gaze data extracted from scenes where no familiarity was reported, also found the best performance from the Ada Boost algorithm combined with a 3 second window of eye-gaze data. However, the model achieved a lower Cohen's Kappa score of 0.09. Model results for all window sizes tested can be found at the bottom of Table 1.

During experimentally familiarized test scenes where familiarity was reported, the Naive Bayes model was able to detect a participant's recall status with a Cohen's Kappa score of 0.25. This best performing model resulted from training on 3 seconds of eye-gaze data, temporally separated from the button press by a 250ms buffer. Model results for all buffer and window sizes tested can be found in Table 2.

4. Discussion

Among instances where participants reported consciously feeling a sense of familiarity with a VR scene (thus the button-press was held constant as it was

Table 1. Detecting the experimentally familiarized status of scenes using eye-gaze features — a comparison of model performance with various buffer and window sizes (Standard Deviation in parenthesis).

Buffer	Window	Model	Cohen's Kappa	F1 Score	Familiarity Reported
0 ms	1 sec	Naive Bayes	0.11 (0.17)	0.49 (0.13)	✓
	2 sec	Ada Boost	0.08 (0.16)	0.61 (0.11)	✓
	3 sec	Ada Boost	0.06 (0.17)	0.60 (0.10)	✓
250 ms	1 sec	Random Forest	0.08 (0.18)	0.62 (0.13)	✓
	2 sec	Ada Boost	0.13 (0.18)	0.63 (0.15)	✓
	3 sec	Ada Boost	0.13 (0.21)	0.60 (0.13)	✓
500 ms	1 sec	Ada Boost	0.08 (0.25)	0.62 (0.15)	✓
	2 sec	SVC	0.13 (0.16)	0.63 (0.12)	✓
	3 sec	**Ada Boost**	**0.16 (0.22)**	**0.64 (0.12)**	✓
1000 ms	1 sec	Ada Boost	0.08 (0.17)	0.62 (0.14)	✓
	2 sec	Random Forest	0.10 (0.18)	0.62 (0.15)	✓
	3 sec	Ada Boost	0.09 (0.18)	0.60 (0.11)	✓
N/A	1 sec	SVC	0.05 (0.14)	0.56 (0.17)	✗
	2 sec	Ada Boost	0.01 (0.14)	0.56 (0.14)	✗
	3 sec	**Ada Boost**	**0.09 (0.21)**	**0.63 (0.15)**	✗

Table 2. Detecting participant's recall status preceding positive reports among experimentally familiarized scenes — a comparison of model performance with various buffer and window sizes (Standard Deviation in parenthesis).

Buffer	Window	Model	Cohen's Kappa	F1 Score
0 ms	1 sec	Ada Boost	0.16 (0.27)	0.64 (0.17)
	2 sec	Ada Boost	0.14 (0.26)	0.53 (0.21)
	3 sec	Naive Bayes	0.23 (0.18)	0.64 (0.12)
250 ms	1 sec	Ada Boost	0.14 (0.24)	0.64 (0.13)
	2 sec	Naive Bayes	0.10 (0.24)	0.54 (0.19)
	3 sec	**Naive Bayes**	**0.25 (0.21)**	**0.66 (0.12)**
500 ms	1 sec	Naive Bayes	0.06 (0.18)	0.45 (0.16)
	2 sec	Ada Boost	0.19 (0.18)	0.66 (0.12)
	3 sec	Naive Bayes	0.23 (0.28)	0.65 (0.14)
1000 ms	1 sec	Ada Boost	0.16 (0.23)	0.65 (0.15)
	2 sec	Naive Bayes	0.18 (0.25)	0.60 (0.15)
	3 sec	Random Forest	0.17 (0.23)	0.66 (0.15)

always pressed in these cases), detectable differences in eye-gaze patterns occurred between experimentally familiarized and unfamiliarized scenes. In contrast, when participants did not consciously experience familiarity with a scene (thus the button-press was held constant in being unpressed), our models were unable to detect any significant differences in the eye-gaze patterns among experimentally familiarized vs. unfamiliarized scenes. These patterns may suggest that eye-gaze patterns reflective of experimental familiarization of scene layout is only detectable when participants are experiencing a sense of familiarity for the scene. If so, an implication might be that a subjectively detectable sense of familiarity must be present in the experiencer in order for variations in familiarity intensity to be externally detectable. However, the lower model performance among negative instances might reflect not having a precise moment (the button-press) to reference. Unlike in positive instances where data corresponds to the time point of the button-press, for negative instances, we sampled a random window of data from within the scene. Future research should continue to try to tackle this challenge. One method might be to yoke each positive instance with a random negative instance from the same participant, using the same time point in the negative instance as when the button was pressed in the positive instance.

Finally, among experimentally familiarized scenes for which the button was pressed, we also observed differences in eye-gaze patterns between instances of familiarity accompanied by recall success vs. failure. This suggests that when a VR scene feels familiar, eye-gaze patterns differ in some fundamental way when the source of the familiarity is identified compared to when it is not, or, alternatively, that there is a detectable gaze pattern difference between recognizing a scene based on recall vs. based on familiarity alone.

Methodologically, this work presents a potential means of addressing issues surrounding measuring eye-gaze in VR when there is a hand-controller button-press to indicate a subjective mental state. An alternative method for research could be designing the experiment so that a button-press is always required (both for positive and negative instances).

References

[1] Castillon I, Chartier T, Venkatesha V, Okada NS, Davis A, Cleary AM, et al. Automatically Identifying the Human Sense of Familiarity Using Eye Gaze Features. In: International Conference on Human-Computer Interaction. Springer; 2024. p. 291-310.

[2] Okada NS, McNeely-White KL, Cleary AM, Carlaw BN, Drane DL, Parsons TD, et al. A virtual reality paradigm with dynamic scene stimuli for use in memory research. Behavior Research Methods. 2023:1-24.

[3] Bixler R, Blanchard N, Garrison L, D'Mello S. Automatic detection of mind wandering during reading using gaze and physiology. In: Proceedings of the 2015 ACM on international Conference on Multimodal Interaction; 2015. p. 299-306.

[4] Blanchard N, Bixler R, Joyce T, D'Mello S. Automated physiological-based detection of mind wandering during learning. In: 12th International Conference on Intelligent Tutoring Systems (ITS). Springer; 2014. p. 55-60.

[5] Cleary AM, Irving ZC, Mills C. What Flips Attention? Cognitive Science. 2023;47(4):e13274.

[6] Bixler R, D'Mello S. Automatic gaze-based user-independent detection of mind wandering during computerized reading. UMUAI. 2016;26:33-68.

[7] Cleary AM, Brown AS, Sawyer BD, Nomi JS, Ajoku AC, Ryals AJ. Familiarity from the configuration of objects in 3-dimensional space and its relation to déjà vu: A virtual reality investigation. Consciousness and cognition. 2012;21(2):969-75.

[8] Cleary AM, Claxton AB. Déjà vu: An illusion of prediction. Psychological Science. 2018;29(4):635-44.

[9] Cleary AM. Recognition memory, familiarity, and déjà vu experiences. Current Directions in Psychological Science. 2008;17(5):353-7.

[10] Cleary AM, McNeely-White KL, Huebert AM, Claxton AB. Déjà vu and the feeling of prediction: An association with familiarity strength. Memory. 2021;29(7):904-20.

[11] Nishimura G, Faisal A. Déjà vu: Classification of memory using eye movements. 2015.