Eye Movement Biometrics Using a New Dataset Collected in Virtual Reality

Dillon Lohr
Texas State University
San Marcos, Texas, USA
djl70@txstate.edu

Samantha Aziz
Texas State University
San Marcos, Texas, USA
sda69@txstate.edu

Oleg Komogortsev
Texas State University
San Marcos, Texas, USA
ok11@txstate.edu

ABSTRACT

This paper introduces a novel eye movement dataset collected in virtual reality (VR) that contains both 2D and 3D eye movement data from over 400 subjects. We establish that this dataset is suitable for biometric studies by evaluating it with both statistical and machine learning–based approaches. For comparison, we also include results from an existing, similarly constructed dataset.

CCS CONCEPTS

• **Security and privacy → Biometrics**.

KEYWORDS

eye tracking, biometrics, authentication, virtual reality

ACM Reference Format:
Dillon Lohr, Samantha Aziz, and Oleg Komogortsev. 2020. Eye Movement Biometrics Using a New Dataset Collected in Virtual Reality. In *Symposium on Eye Tracking Research and Applications (ETRA '20 Adjunct), June 2–5, 2020, Stuttgart, Germany*. ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3379157.3391420

1 INTRODUCTION

The study of eye movement–driven authentication has largely focused on authenticating users based on 2D eye movement features, and promising results have been achieved in the past [Friedman et al. 2017; George and Routray 2016].

We believe that eye tracking will become ubiquitous in virtual reality (VR) devices with the inclusion of foveated rendering, a technique which can significantly reduce computational requirements by utilizing gaze information. One benefit of VR is that it makes it easy to elicit vergence responses by changing the depth of a stimulus. Since convergence and divergence responses have been shown to vary between people [Tyler et al. 2012], the addition of vergence features may improve authentication rates.

In the present study, we improve upon the work of Lohr et al. [2018] by using a significantly larger dataset collected at a higher sampling frequency, employing more recent biometric approaches, and comparing our results against another dataset. Although we do not explore the use of vergence features in the present study, our dataset will enable future research in that area.

ETRA '20 Adjunct, June 2–5, 2020, Stuttgart, Germany

ACM ISBN 978-1-4503-7135-3/20/06.
https://doi.org/10.1145/3379157.3391420

2 DATASETS

For our analysis, we used two different datasets. The first dataset, which we call round 1 of the virtual reality eye movement database (VREM-R1), is part of a new, long-term data collection procedure that we started several months ago. The second dataset is the SBA-ST dataset used by Friedman et al. [2017], which we included for comparisons against VREM-R1.

VREM-R1 consists of eye movement data from 458 participants (206 male, 248 female, 4 other) ages 18-58 (median: 20), each recorded twice with a 5-minute break between recording sessions. Each recording session lasted around 30 minutes, for a total experiment duration of around 1 hour. Participants were recorded with SMI's tethered system based on the HTC Vive (ET-HMD). The embedded eye-tracking device by SMI tracks both eyes simultaneously with a sampling rate of 250 Hz. A work-in-progress report by Lohr et al. [2019] found that the ET-HMD has a binocular spatial accuracy around 0.67° (based on an average across 12 subjects). Five tasks—three guided and two unguided—were presented during each recording session. The guided tasks were each designed to elicit a specific type of eye movement (namely saccades, smooth pursuit, and vergence), while the unguided tasks captured a variety of eye movements as participants watched a short video clip or read an excerpt of text from *National Geographic*.

SBA-ST consists of eye movement data from 335 participants (178 male, 157 female) ages 18-46 (median: 21). Participants were recorded monocularly at 1000 Hz with SR Research's EyeLink 1000. More information about the SBA-ST dataset is given by Friedman et al. [2017]. We modeled the data collection procedure used in VREM-R1 after the SBA-ST procedure so that direct comparisons could be made for some tasks.

This study focused only on the reading data, because eye movements during reading have been used to achieve some of the best biometric performance [Friedman et al. 2017]. There were 422 participants from VREM-R1 and 322 participants from SBA-ST that had reading data for both sessions; all other subjects were excluded from this study. We assessed the biometric performance achieved with VREM-R1 relative to SBA-ST using one statistical and one machine learning–based approach. These approaches were chosen because they both yielded high biometric performance on data collected with the EyeLink 1000.

3 EYE MOVEMENT CLASSIFICATION

We classified each eye movement signal using the MNH algorithm [Friedman et al. 2018]. Since the MNH (with default parameters) expects 1000 Hz signals with relatively low noise (high spatial precision), we interpolated the data in VREM-R1 to 1000 Hz using MATLAB's `pchip` function. The ET-HMD produces much noisier

monocular signals than the EyeLink 1000, so low-pass filtering would be necessary to smooth the signals enough that the MNH would not over-classify noise. Instead of exploring different filtering options, we simply used the heavily-filtered binocular signal from the ET-HMD. There are downsides to using the binocular signal instead of a low-pass filtered version of a monocular signal, such as catch-up saccades during smooth pursuit being virtually non-existent in the binocular signal, but such problems are beyond the scope of this study.

We removed subjects with either 5% or more missing samples or 18% or more samples classified as noise by the MNH. These thresholds were chosen empirically. From VREM-R1, 66 total subjects were removed in this way, leaving 356 subjects for our analysis. From SBA-ST, 114 total subjects were removed in this way, leaving 208 subjects for our analysis.

4 STATISTICAL APPROACH

Our statistical approach, which we refer to as STAT, closely followed the data analysis procedure used by Friedman et al. [2017]. A brief summary follows. Over 1000 features from fixations, saccades, and post-saccadic oscillations were extracted. A subset of these features was selected based on normality, redundancy, and intra-class correlation. Principal component analysis (PCA) was used for dimensionality reduction. Cosine distance was used for computing similarity measures.

When checking normality, we used the Shapiro-Wilk test instead of the Pearson Chi-Square test. We determined the best subset of features, their normality transformations, and the optimal number of principal components by minimizing equal error rate (EER) on a training set of subjects, then we employed them when evaluating on a disjoint set of subjects.

For both datasets, we randomly selected 75 subjects for training and used another randomly-selected 75 subjects for evaluation. We repeated training and evaluation 100 times for each dataset with different samples of subjects to get a distribution of EER. We chose to use 75 subjects each time based on the methodology of George and Routray [2016].

5 MACHINE LEARNING APPROACH

Our machine learning approach, which we'll refer to as RBFN, closely followed the procedure used by George and Routray [2016]. A brief summary follows. We extracted 12 features from each fixation and 46 features from each saccade. Two radial basis function (RBF) networks were created, one for fixations and one for saccades. Each network contained 32 neurons per subject and had one output node per subject. Each neuron used a Gaussian activation function

$$\phi_{\mu,\sigma}(x) = e^{-\beta\|x-\mu\|^2}, \text{ where } \beta = \frac{1}{2\sigma^2} \tag{1}$$

with parameters μ and σ determined using k-means clustering. For a given subject, each event was fed through the network and the outputs were averaged across all events. The fixation and saccade outputs were averaged together to produce the final match scores.

Occasionally, a neuron's activation had a non-finite β, and when that happened we simply set $\beta = 1$ to allow that neuron to still activate for nearby samples. Also, when optimizing the weights of our models, we used the Moore-Penrose pseudo-inverse instead of gradient descent. We did not remove any features from our models.

For both datasets, we randomly selected 75 subjects, trained on data from session 1, and evaluated on data from session 2. We repeated training and evaluation 100 times for each dataset with different samples of subjects to get a distribution of EER.

6 RESULTS

Table 1 reports the mean and standard deviation of our achieved EER with each dataset and with each biometric approach.

Table 1: EER achieved with each dataset and approach.

Dataset	Approach	EER (%)	
		Mean	SD
VREM-R1	STAT	9.98	2.39
	RBFN	14.37	1.67
SBA-ST	STAT	2.04	1.32
	RBFN	5.12	0.74

7 DISCUSSION

Compared to SBA-ST, the EER with VREM-R1 was nearly 5-times worse using the STAT approach and about 3-times worse using the RBFN approach. There are two main reasons for why the VREM-R1 performed worse. First, and perhaps most importantly, the MNH was specifically designed for data from the EyeLink 1000, so there were several problems with its classification of signals from the ET-HMD. Second, the binocular signals that we used were heavily filtered, so features like peak saccade velocity would have been heavily affected. Fine-tuning the MNH parameters, using some low-pass filtered version of a monocular signal instead of the binocular signal, and removing data after the end of reading would all improve biometric performance when using the VREM-R1 dataset.

In the future, we are interested in exploring the inclusion of vergence-related features, because these should be uncorrelated with the existing features set and should further help distinguish between individuals. Additionally, as we expand the VREM-R1 dataset with additional rounds of recordings, we plan to assess the temporal persistence of features over longer periods of time.

ACKNOWLEDGMENTS

We thank Dr. Lee Friedman for his statistical guidance. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-1144466. The study was also funded by 3 grants to Dr. Komogortsev: (1) National Science Foundation, CNS-1250718 and CNS-1714623, www.NSF.gov; (2) National Institute of Standards and Technology, 60NANB15D325, www.NIST.gov; (3) National Institute of Standards and Technology, 60NANB16D293. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation or the National Institute of Standards and Technology.

REFERENCES

Lee Friedman, Mark S. Nixon, and Oleg V. Komogortsev. 2017. Method to assess the temporal persistence of potential biometric features: Application to oculomotor, gait, face and brain structure databases. *PLoS ONE* 12, 6 (jun 2017), e0178501. https://doi.org/10.1371/journal.pone.0178501

Lee Friedman, Ioannis Rigas, Evgeny Abdulin, and Oleg V. Komogortsev. 2018. A novel evaluation of two related and two independent algorithms for eye movement classification during reading. *Behavior Research Methods* 50, 4 (aug 2018), 1374–1397. https://doi.org/10.3758/s13428-018-1050-7

Anjith George and Aurobinda Routray. 2016. A score level fusion method for eye movement biometrics. *Pattern Recognition Letters* 82 (oct 2016), 207–215. https://doi.org/10.1016/j.patrec.2015.11.020 arXiv:1601.03333

Dillon Lohr, Samuel-Hunter Berndt, and Oleg Komogortsev. 2018. An implementation of eye movement-driven biometrics in virtual reality. In *Eye Tracking Research and Applications Symposium (ETRA)*. https://doi.org/10.1145/3204493.3208333

Dillon J. Lohr, Lee Friedman, and Oleg V. Komogortsev. 2019. Evaluating the data quality of eye tracking signals from a virtual reality system: Case study using SMI's eye-tracking HTC Vive. arXiv:1912.02083

Christopher W. Tyler, Anas M. Elsaid, Lora T. Likova, Navdeep Gill, and Spero C. Nicholas. 2012. Analysis of human vergence dynamics. *Journal of Vision* 12, 11 (oct 2012), 21–21. https://doi.org/10.1167/12.11.21