JDS

Estimating Fun Based on Eye Behavior Collected from HMD

MAYU AKATA, Univercity of Tsukuba, Japan
YOSHIKI NISHIKAWA, Univercity of Tsukuba, Japan
TOSHIYA ISOMOTO, LY Corporation, Japan
BUNTAROU SHIZUKI, Univercity of Tsukuba, Japan

Review Article

Code and data links:
https://github.com/borisveytsman/acmart
https://zenodo.org/link

Keywords: gaze-based interaction, user intent, hands-free, target selection, eye-tracker

BT and GKMT designed the study; LT, VB, and AP conducted the experiments, BR, HC, CP and JS analyzed the results, JPK developed analytical predictions, all authors participated in writing the manuscript.

Authors' addresses: Mayu Akata, akata@iplab.cs.tsukuba.ac.jp, Univercity of Tsukuba, Tsukuba, , Japan, 305-0006; Yoshiki Nishikawa, nishikawa@iplab.cs.tsukuba.ac.jp, Univercity of Tsukuba, Tokyo, , Japan, 305-0006; Toshiya Isomoto, r.t.isomoto@gmail.com, LY Corporation, Tokyo, , Japan; Buntarou Shizuki, shizuki@cs.tsukuba.ac.jp, Univercity of Tsukuba, Tsukuba, , Japan, 305-0006.

ABSTRACT

This study shows a method for estimating users' emotions in Virtual Reality (VR) spaces through the collection of eye behaviors. In our method, we use eye-related information available from the Head Mounted Display (HMD), including the direction vector of the gaze, coordination of the pupil, pupil diameter, and the eyelid opening width, to estimate whether the user is having fun or feeling others emotions. Using the LightGBM algorithm, the estimation accuracy resulted in an AUC of 0.84 and an accuracy of 0.78.

INTRODUCTION

Incorporating user emotions into interfaces can provide users with richer interaction, such as automated video scoring (as shown in Fig. 1) and a human communication support system (e.g., [1]). Some of the most established emotion estimation methods are camera based, such as those that read facial expressions (e.g., [8, 11]). However, facial recognition can be challenging for emotion estimation in situations where the camera cannot capture the entire face. For instance, in Virtual Reality (VR) context, which requires users to wear a Head Mounted Display (HMD), parts of the face (e.g., forehead, eyes, and nose) are obscured, making it difficult to use camera-based methods to estimate expressions.

Research has been conducted on estimating users' emotions using eye behavior (e.g., [2, 4, 7, 9, 10, 12]). Past research has often used mouth information, such as smiles, to detect feelings of having fun (e.g., [3]). On the other hand, in this study, we use the eye behaviors sampled through the eye-tracker built into an HMD, which makes it possible to use the eye behaviors that can not be captured through an external camera for emotion estimation. Our research, while still a work in progress, is groundbreaking, as it is the first to attempt to estimate a user's emotion by only using eye behaviors sampled through an eye-tracker. Through an experiment, we collected users' eye behaviors and ground-truth emotions during a movie-watching task. We then developed a machine-learning (ML) model to estimate whether users feel like they are having fun, or experiencing other emotions, by using eye behaviors; the estimation results in AUC of 0.84. Furthermore, we show an automated video scoring system to provide an example of an application.

EMOTION ESTIMATION METHOD

Our emotion estimation method is based on eye behaviors and an ML model. As the features of the ML model, we calculate statistical data from four types of eye behaviors sampled through an eye-tracker built into an HMD: the direction vector of the gaze, coordination of pupil, pupil diameter, and eyelid opening width.

To develop the ML model, we conducted a pilot study to collect the eye behaviors and ground-truth emotions in a movie watching task involving ten of our laboratory students. We used HTC VIVE PRO EYE as the HMD to sample eye behaviors at 90 Hz. During the movie watching task, we asked the participants (one female, mean age = 22.3) to indicate when they felt like they were having fun while watching a 30 min Japanese comedy displayed in an HMD. To label the emotion, participants pushed a key on the keyboard positioned in their hands. In total, we collected 235 labels indicating fun.

Before calculating the features, we first excluded the outliers (0.1% of the whole data) that were caused by the eye-tracking system errors. By using the eye behaviors and the labels, we calculated the features for the ML model; We referred to features used in previous research [5], as it consisted of only eye behaviors for ML-based interaction. We used 2,000 ms of eye behaviors as a positive label; we used the eye behaviors of -500 ms–1500 ms based when the participants pushed the key (i.e., indicating fun). As the negative label, we used randomly sampled eye behaviors, except for the 2,000 ms that were used as the positive label. Using the 2,000 ms of eye behaviors (i.e., 180 samples = 90 Hz x 2 sec), we first downsampled them into 20 samples by taking the average every 100 ms (i.e., 9 samples). We then calculated relative values based on the first sample for the 20 samples. These processes were conducted to minimize the eye-tracking noise and the dependency on the surroundings and individual differences. Lastly, we calculated 84 features for the relative values, as shown in Table 1. That is, we used 84 features and one label as one dataset.

Using the datasets, we trained and tested an ML model. To make the ratio between positive and negative labels 50:50, we randomly picked 235 datasets from the negative labels and eye behaviors. Of these 235 datasets, we first split them into train and test datasets of 188 and 47, at a ratio of 4:1, and adopted a 5-fold cross-validation. For the ML algorithm, we compared Random Forest, Support Vector Machine, Perceptron, and LightGBM([6]); which had the highest accuracy, was adopted. We conducted no hyperparameter tuning and used the default settings of scikit-learn. Using the trained model and test datasets, we then calculated the performance of our ML model, which were AUC of 0.84, an accuracy of 0.78, and an F1 score of 0.77.

As an example of an application that could use our method, we show an automated video scoring system, as shown in Fig. 1. Currently, common metrics for automated video scoring video content often rely on view counts, number of likes, and comment counts. In contrast, our method can expand the automated scoring to consider user emotions, especially reflecting users' implicit emotions, for each time sample.

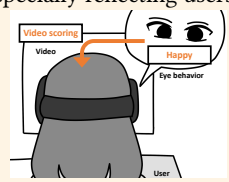

Fig. 1. Automated video scoring system.

Table 1. The 84 features used in our system.

Data	Feature	Total
Gaze direction vector Pupil position Pupil diameter	Mean, Last Value, Amplitude, SD, Min, Kurtosis, Max, Skewness	24 (3 × 8) 40 (5 × 8) 8 (1 × 8)
Eyelid opening width	Mean, SD, Amplitude, Max, Min, Last Value	12 (2 × 6)

CONCLUSION

In this paper, we described a method for estimating two emotions (fun and others) using ML and eye behavior in an HMD interaction. We plan to enhance this work-in-progress model by increasing the number of estimated emotions and improving estimation performance by using variable experimental conditions and by exploring suitable features and ML algorithms for emotion estimation.

REFERENCES

[1] Min Chen, Ping Zhou, and Giancarlo Fortino. 2017. Emotion Communication System. *IEEE Access* 5 (2017), 326–337. https://doi.org/10.1109/ACCESS.2016.2641480

[2] Hong Feng and Xunbing Shen. 2022. A Random Forest Algorithm-Based Emotion Recognition Model for Eye Features. In *Proceedings of the 3rd International Symposium on Artificial Intelligence for Medicine Sciences* (Amsterdam, Netherlands) *(ISAIMS '22)*. Association for Computing Machinery, New York, NY, USA, 148–152. https://doi.org/10.1145/3570773.3570851

[3] Anurag Goswami, Ganjigunta Ramakrishna, and Rajni Sethi. 2021. Review on Smile Detection. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology* (2021), 577–583.

[4] Zhuoluo Huang and Yafang Li. 2023. Machine Learning-Based Study of Eye Features under the Emotion of Anger. In *Proceedings of the 2022 6th International Conference on Electronic Information Technology and Computer Engineering* (Xiamen, China) *(EITCE '22)*. Association for Computing Machinery, New York, NY, USA, 1631–1635. https://doi.org/10.1145/3573428.3573716

[5] Toshiya Isomoto, Shota Yamanaka, and Buntarou Shizuki. 2022. Dwell Selection with ML-Based Intent Prediction Using Only Gaze Data. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 3, Article 120 (2022), 21 pages. https://doi.org/10.1145/3550301

[6] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. 2017. LightGBM: a highly efficient gradient boosting decision tree. Curran Associates Inc., Red Hook, NY, USA.

[7] Jia Zheng Lim, James Mountstephens, and Jason Teo. 2020. Emotion Recognition Using Eye-Tracking: Taxonomy, Review and Current Challenges. *Sensors* 20, 8 (2020). https://doi.org/10.3390/s20082384

[8] Paweł Tarnowski, Marcin Kołodziej, Andrzej Majkowski, and Remigiusz J. Rak. 2017. Emotion recognition using facial expressions. *Procedia Computer Science* 108 (2017), 1175–1184. https://doi.org/10.1016/j.procs.2017.05.025 International Conference on Computational Science, ICCS 2017, 12-14 June 2017, Zurich, Switzerland.

[9] Hao Wu, Jinghao Feng, Xuejin Tian, Edward Sun, Yunxin Liu, Bo Dong, Fengyuan Xu, and Sheng Zhong. 2020. EMO: Real-Time Emotion Recognition from Single-Eye Images for Resource-Constrained Eyewear Devices. In *Proceedings of the 18th International Conference on Mobile Systems, Applications, and Services* (Toronto, Ontario, Canada) *(MobiSys '20)*. Association for Computing Machinery, New York, NY, USA, 448–461. https://doi.org/10.1145/3386901.3388917

[10] Xu Yan, Li-Ming Zhao, and Bao-Liang Lu. 2021. Simplifying Multimodal Emotion Recognition with Single Eye Movement Modality. In *Proceedings of the 29th ACM International Conference on Multimedia* (Virtual Event, China) *(MM '21)*. Association for Computing Machinery, New York, NY, USA, 1057–1063. https://doi.org/10.1145/3474085.3475701

[11] Ce Zhan, Wanqing Li, Philip Ogunbona, and Farzad Safaei. 2006. Facial Expression Recognition for Multiplayer Online Games. In *Proceedings of the 3rd Australasian Conference on Interactive Entertainment (IE '06)*. Murdoch University, Murdoch, AUS, 52–58.

[12] Lim Jia Zheng, James Mountstephens, and Jason Teo. 2020. Four-class emotion classification in virtual reality using pupillometry. *Journal of Big Data* 7, 43 (2020), 1–9.