

Eye Fixation Versus Pupil Diameter as Eye-Tracking Features for Virtual Reality Emotion Classification

Lim Jia Zheng

Graduate researcher, Evolutionary
Computing Laboratory
Faculty of Computing and Informatics,
Universiti Malaysia Sabah
Kota Kinabalu, Sabah, Malaysia.
limjiazheng94@gmail.com

James Mountstephens

Senior lecturer, Evolutionary
Computing Laboratory
Faculty of Computing and Informatics,
Universiti Malaysia Sabah
Kota Kinabalu, Sabah, Malaysia.
james@ums.edu.my

Jason Teo*

Professor, Evolutionary Computing
Laboratory
Faculty of Computing and Informatics,
Universiti Malaysia Sabah
Kota Kinabalu, Sabah, Malaysia.
jtwteo@ums.edu.my
(*corresponding author)

Abstract— The usage of eye-tracking technology is becoming increasingly popular in machine learning applications, particularly in the area of affective computing and emotion recognition. Typically, emotion recognition studies utilize popular physiological signals such as electroencephalography (EEG), while the research on emotion detection that relies solely on eye-tracking data is limited. In this study, an empirical comparison of the accuracy of eye-tracking-based emotion recognition in a virtual reality (VR) environment using eye fixation versus pupil diameter as the classification feature is performed. We classified emotions into four distinct classes according to Russell's four-quadrant Circumplex Model of Affect. 360° videos are presented as emotional stimuli to participants in a VR environment to evoke the user's emotions. Three separate experiments were conducted using Support Vector Machines (SVMs) as the classification algorithm for the two chosen eye features. The results showed that emotion classification using fixation position obtained an accuracy of 75% while pupil diameter obtained an accuracy of 57%. For four-quadrant emotion recognition, eye fixation as a learning feature produces better classification accuracy compared to pupil diameter. Therefore, this empirical study has shown that eye-tracking-based emotion recognition systems would benefit from using features based on eye fixation data rather than pupil size.

Keywords—emotion classification, eye-tracking, fixation, pupil diameter, virtual reality, support vector machines

I. INTRODUCTION

Emotion classification is the method of determining and identifying the emotional states from a human behavioral response by utilizing statistical methods including machine learning algorithms. Machine learning is an approach that utilizes the data and information or the collected datasets obtained to predict the accuracy of emotion recognition [1]. Emotion detection technique has been used in multiple areas for development such as artificial intelligence, computer vision, and signal processing. Several applications using emotion detection have been developed such as human mental health sensing systems [2] and driver monitoring systems [3].

The analysis of emotional states is useful for both the understanding of human behavior and the integration of human factors into artificial systems. Thus, researchers are putting significant efforts into the study of emotion recognition for human-computer interaction (HCI) to improve the machine understanding of human emotions.

Some advanced technologies have been developed for enhancing the interaction and communication between humans and machines. In early works, most of the emotion recognition experiments were done based on external expressions such as body gestures and human speech. With the rapid development of more advanced technologies, several types of information and data from internal expressions can be utilized to classify human emotional states such as galvanic skin response (GSR) signal, electroencephalography (EEG), electromyography (EMG), and electrocardiography (ECG).

Additionally, the usage of eye-tracking technology has become popular and has recently attracted significant attention from the research community. The study of eye movement has become increasingly popular in the fields of psychology, neuroscience, cognitive science, and advertising. Hence, eye-tracking technology can be similarly utilized and deployed in the emotion recognition system. However, existing studies on emotion recognition using eye-tracking data solely is limited. Several eye signals can be utilized to classify emotions, where pupil size and eye fixations are the most commonly used features from the eye-tracking data. Therefore, the purpose of this paper is to classify four classes of emotions and to conduct a comparison between the use of these two features, which are pupil diameter and eye fixation.

To the best of our best knowledge, there is currently no study that compares the performance of emotion recognition based on the features of pupil size and eye fixations. In this paper, we classified emotions according to Russell's emotion model and made a comparison of the performance from the eye features, which are pupil size and eye fixations using Support Vector Machine (SVM) in a virtual environment. The introduction was presented in the first part of this paper. Section 2 presented the related works and background of emotion, eye-tracking, and VR. The next section is the methodology of this experiment including the experimental setup and procedure. The results and discussion were presented in Section 4 and the last section is the conclusion of this paper.

II. BACKGROUND

A. Emotion

Emotion refers to a state of feeling which is closely linked to the human nervous system that influences their behavior through physical and psychological changes [4].

The definition of emotion currently does not have a consensus since they have many definitions and theories from different researchers. Emotion classification is a distinction between several complex types of emotions and distinguishes one from another. A popular example of basic emotions comes from Ekman's model, one of the most influential works from his findings, which introduced the 6 primary emotions: anger, happiness, sadness, disgust, surprise, and fear [5]. His theory is then supported by Plutchik's model, which developed the "wheel of emotions", introducing 8 primary emotions and adding 2 emotions into Ekman's six basic emotions, which are trust and anticipation [6]. The circumplex model of emotion is one of the most prominent models for emotion classification [7]. It has been widely used by researchers for stimuli testing, including emotional phrases, affective states, and facial expressions. A four-quadrant of emotions results from the combination of the valence and arousal dimensions in this model, and each of the quadrants represents the respective emotions. Different emotions are evoked according to the level of the dimensions, which is the high or low arousal and negative or positive valence.

B. Emotion Classification using Eye-tracking and Related Works

Eye-tracking refers to a method of determining and calculating the eye positions and eye motions of a user. Eye-tracking becomes a hot topic in the field of computer science research such as online reading and image classification. Eye-tracker is a sensor technology for capturing and recording an individual's eye properties then extracting them into a form of data. Several features can be extracted from the eye-tracking data to classify emotions such as pupil position, eye fixation, and pupil size. The activity of the pupil is strongly related to the human nervous system and there is a correlation between pupil size and emotion [8]. Fixation is defined as a motionless gaze and it refers to the focus point in time. The act of fixating refers to the stage between two saccades at which the pupils are relatively stationary and practically all visual feedback occurs.

Eye movement signals can be utilized in the study of emotion classification since the eye data contain emotional-relevant features. However, numerous studies on emotion recognition rely on multiple modalities, whereas the study using eye-tracking data solely is fewer. For the combination modality, the most commonly used is brain signal, which is EEG signals combined with pupillary responses to recognize emotions. Several features can be extracted from the eye-tracking data such as pupil diameter, which is the most commonly used feature as the single modality for emotion classification. A neural network approach for emotion recognition uses gaze position to classify emotions and obtained a promising result [9]. Except for pupil size and pupil position, there are also other eye features that have been utilized for emotion recognition by some investigators such as eye fixation [10] and electrooculography (EOG) signals [11]. There is a study classified 3-class emotions based on saccade duration and pupil diameter [12]. Last but not least, there are emotional studies that use a novel feature from eye-tracking data to detect emotions such as motion speed of the eye [13], and distance between sclera and iris [14]. Currently, the study on emotion recognition based on eye features solely is very

limited and there are no previous studies that used fixation position as the eye feature for emotion classification. Therefore, the objective of this paper is to classify four classes of emotions and compare the accuracy performance based on pupil diameter and fixation position from eye-tracking data using SVM machine learning with parameter tuning in VR.

C. Emotion Classification in Virtual Reality

VR is a synthetic environment that can be equivalent to the real world or a totally different virtual world that can be generated by a computer program. VR can be contributed to many domains such as social science, medicine, entertainment, and engineering. In emotion classification, Immersive Virtual Environments (IVE) can be used to evoke an individual's emotional responses through VR experiences. A high emotion recognition performance is obtained using biometric data with machine learning in an immersive VR experience [15]. Nowadays, many VR headsets are integrated with eye-tracking technology, hence eye-tracking data can be obtained easily. The eye-tracker is attached within the VR headset, the process of data capturing and recording is running simultaneously when the 360° videos are presented. VR creates a simulated environment and the user is immersed in 360° video presentations within the VR headset. Within the VR, distraction and outside environment influences can be minimized. VR provides the senses of vision and auditory, the user is immersed and fully controlled, therefore an authentic emotional response can be obtained from the user.

III. METHODOLOGY

A. Participants

All participants were given an explanation and procedure regarding the experiment. 30 participants (16 males, 14 females) with ages between 21 to 29 years old volunteered in this experiment. All participants were given a brief explanation about the experimental procedure before the experiment started. The participant was also given guidance to wear the VR headset and the sitting position before the experiment began. Figure 1 showed the experiment setup when a participant wears the VR headset during the experiment.



Fig. 1. The VR headset used by the participant during the experiment.

B. Experiment Protocol and Setup

A presentation of 360° videos was set up with VR to evoke the emotions of a user. Eye-tracking data was captured and recorded using the Pupil Labs, an add-on eye-tracker within an HTC Vive VR headset. The contents of the video presentation consist of 4 quadrants of emotions based on the valence arousal model from the Circumplex Model of Affect. There was a total of 4 sessions for the video presentation regarding a specific quadrant of emotions, namely calm, scared, happy, and bored. Figures 2 to 5 showed examples of VR contents according to each video quadrant stimulation. The total duration of the video was about 6 minutes. The protocol of the experiment is shown in Figure 6. A 5-second VR startup was given before the first session stimulation started. Each video stimulation represents one quadrant of emotions and it was about 80 seconds for each session. A 10-second resting period was given before the next session started.



Fig. 2. A happy scene represents quadrant 1 stimulation.



Fig. 3. An scared scene represents quadrant 2 stimulation.



Fig. 4. A bored scene represents quadrant 3 stimulation.



Fig. 5. A calm scene represents quadrant 4 stimulation.

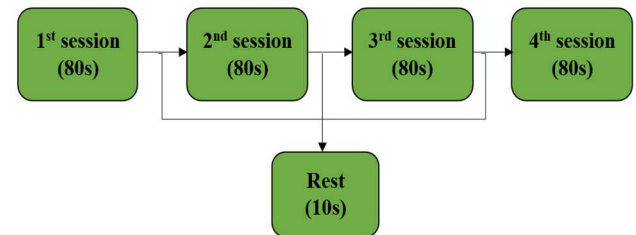


Fig. 6. Flow of video presentation in experiment.

C. Data Collection and Feature Extraction

The capturing and recording of eye data were conducted simultaneously while the 360° videos were presented. The video presentation was done by using Unity with a recording script in the programming language of C#. The eye data were collected and recorded using the Pupil Labs application, called Pupil Capture. The collected data is then exported to the form of data and saved in a CSV file using Pupil Player. Figure 7 shows the interface of detecting and extracting the eye fixation of a user. Pupil diameter and fixation position were chosen as the features for emotion classification. The pupil diameter was extracted by 3D pupil detection and scaled to millimeters (mm) with perspective calibration from 2D extraction in pixels. The data were corrected for perspective. The automatic measurement of pupil diameter is based on the Pupil Datum Format from Pupil Labs software with the 3D detector using Python. The eye fixation is detected and calculated using the implementation of dispersion-based algorithms. Fixation is described as groups of successive points within a specific dispersion. Dispersion is the mechanism by which a wave's phase velocity relies on its frequency. Group velocity dispersion (GVD) is defined as a function of a dispersive medium, most commonly used to assess how the medium affects the length of an optical pulse passing through it. Fixations generally have at least 100 milliseconds (ms) of duration, hence a minimum duration threshold of 100-200ms is often used in dispersion-based identification (I-DT) techniques to help reduce system variability. I-DT algorithm checks the dispersion, D by summing up the differences between the maximum and minimum of x and y values.

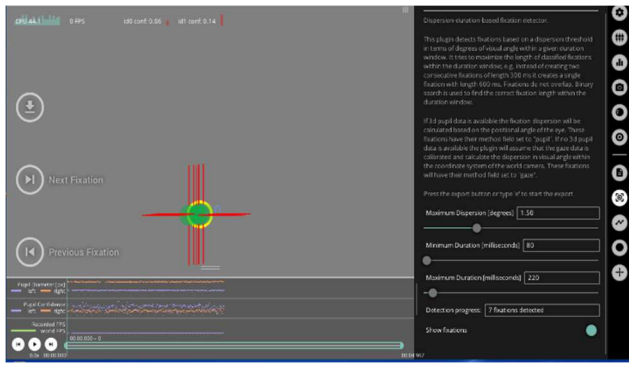


Fig. 7. Interface of detecting and extracting fixation data.

D. Classification

For the classification tasks, the classifier used in this project is SVM with Radial Basis Function (RBF) kernel. Python is the machine learning language to classify emotions. Three separate experiments will be conducted to compare the performance using SVM with parameter tuning (gamma value) and determine the highest accuracy obtained.

IV. RESULTS AND DISCUSSION

The results of emotion recognition performance in accuracy are presented in charts. Figures 7 and 8 presented the relationship between average pupil diameter and the emotional quadrant, as well as the relationship between average fixation duration and the emotional quadrant respectively. The accuracy of emotion classification is compared based on the features of pupil diameter and eye fixation in 3 separate experiments. Figures 9 to 10 displayed the graph of accuracy comparison using pupil diameter and eye fixation for each subject in the three experiments.

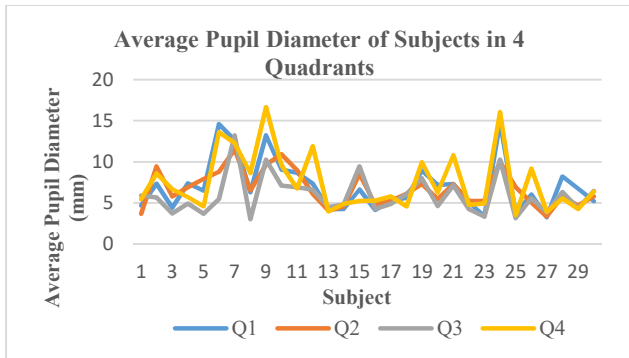


Fig. 8. Average pupil diameter of subjects.

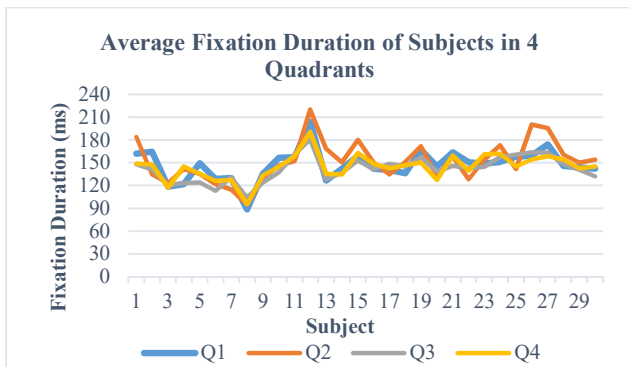


Fig. 9. Average fixation duration of subjects.

Figure 8 illustrates the relationship between average pupil diameter and the emotional quadrants. From the findings, pupil diameter is the largest in quadrant 4, which is the emotion of calm. There are 40% of subjects showed the largest pupil size. While, the results showed that the pupil diameter is the smallest in quadrant 3, which is the emotion of bored. It showed that there is a big variation of pupil size in the low arousal level. Figure 9 illustrates the variation of the average fixation duration depending on the quadrant's emotions. The results showed that it has the longest fixation duration when the subject was stimulated in quadrant 2, which is the scared emotion. While, it has the shortest fixation duration in quadrant 3, the bored emotion. It showed that it has a big variation of fixation duration in the negative valence. The findings showed that there is a correlation between both eye features and the emotional quadrants.

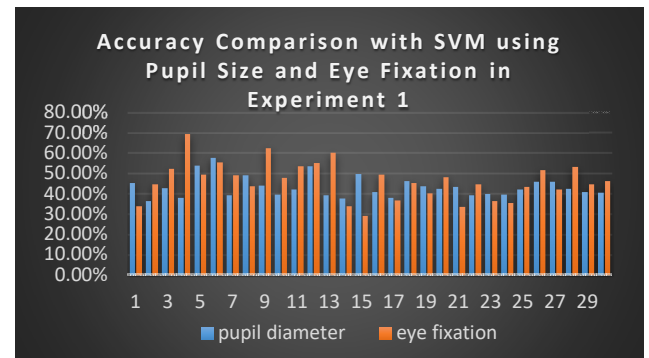


Fig. 10. The accuracy comparison of each subject in experiment 1.

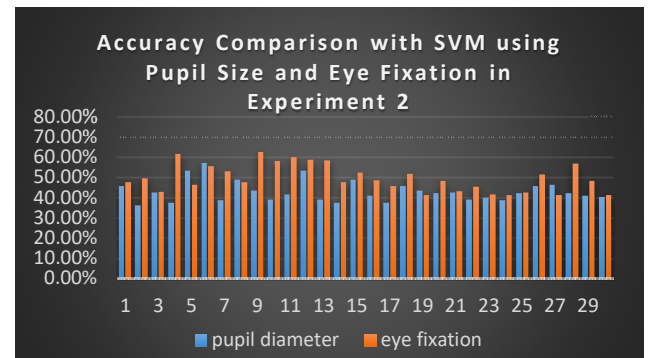


Fig. 11. The accuracy comparison of each subject in experiment 2.

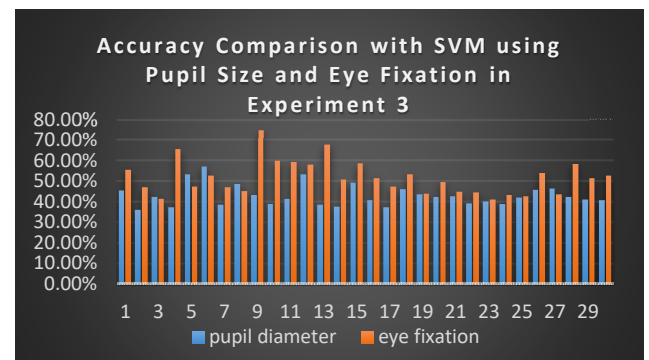


Fig. 12. The accuracy comparison of each subject in experiment 3.

The three classification experiments were conducted using the SVM machine learning algorithm with RBF kernel to classify four distinct classes of emotions by running the Python script. Each of the experiments used different parameters in SVM by setting the range of gamma values to determine the best accuracy obtained. In

experiment 1, the gamma value is set to the range of 0.1 to 100, the gamma value in experiment 2 is ranging from 1 to 2000, while experiment 3 is 10 to 2000. The acquired eye-tracking data is then processed with the machine learning algorithm to obtain the accuracy performance.

From Figure 10, the highest accuracy obtained from pupil diameter was 57.66% while fixation was 69.23% in experiment 1. In experiment 2, pupil diameter achieved the highest accuracy of 57.05%, while fixation achieved the 62.50%. In experiment 3, the accuracy of 56.98% was the highest performance obtained from pupil diameter, and the highest accuracy obtained from fixation data was 75%. From the results, the performance using pupil diameter in experiment 1 showed the highest accuracy, which is 57.66% among the 3 experiments. While the highest performance using eye fixation was 75% in experiment 3. The findings showed that the fixation data has a better performance in emotion classification compared to pupil diameter. From the findings, most of the fixation data showed that it has a higher performance in a higher gamma value. While pupil diameter obtained a better performance when the gamma value is smaller. It showed that it is a promising result for emotion classification using only a single feature from the eye-tracking data solely, which pupil diameter achieved accuracy close to 60% and fixation data has an accuracy close to 80% from a four-class random emotion classification.

The contents of our VR videos are stitched with 16 videos, which contain positive emotions and negative emotions, that are closely related to the four quadrants of emotions. Each session of the video presentation consists of 4 video clips and has a high impact to evoke the intended emotions. For example, a video showing a teaching lesson in a classroom, which corresponds to the boring emotions from quadrant 3. It is tough to determine a specific emotion from a group of complicated emotions. Hence, we classified emotions into four distinct quadrants according to the arousal valence model. The weakness of the experimental setup is the limitation of freedom for the participant when they are watching the video presentation. Participants are not allowed to stand up or move around due to the short-wired VR headset.

V. CONCLUSION AND FUTURE WORK

In this study, the objective is to compare the emotion recognition performance using the features of pupil diameter and fixation position from eye-tracking data in VR. We used the two eye features to classify emotions into four distinct quadrants according to Russell's Circumplex Model of Affect in VR stimuli with 30 subjects. SVM with RBF kernel is used as the classifier in this investigation. Three separate experiments were conducted using parameter tuning by setting the gamma value in order to attempt the attainment of the highest accuracy of emotion classification. From the findings, emotion recognition using fixation data has a better performance, which achieved an accuracy of 75% compared to pupil diameter, which is 57%. Findings showed that it is a promising result since an emotion classification accuracy using eye-tracking data solely could attain up to 75% accuracy from a four-class classification problem. For future work, more eye-tracking features will be attempted to be extracted from the eye-tracking data and used for emotion classification hopefully

to obtain an even better performance than the results obtained here thus far.

ACKNOWLEDGMENT

This work was supported by Ministry of Science, Technology and Innovation (MOSTI) [IF0318M1003], Malaysia [grant number ICF0001-2018].

REFERENCES

- [1] E. Cambria, "Affective Computing and Sentiment Analysis," *IEEE Intelligent Systems*, vol. 31, no. 2, pp. 102–107, 2016, doi: 10.1109/MIS.2016.31.
- [2] R. Guo, S. Li, L. He, W. Gao, H. Qi, and G. Owens, "Pervasive and unobtrusive emotion sensing for human mental health," *Proceedings of the 2013 7th International Conference on Pervasive Computing Technologies for Healthcare and Workshops, PervasiveHealth 2013*, pp. 436–439, 2013, doi: 10.4108/icst.pervasivehealth.2013.252133.
- [3] Y. L. Wu, H. Y. Tsai, Y. C. Huang, and B. H. Chen, "Accurate Emotion Recognition for Driving Risk Prevention in Driver Monitoring System," *2018 IEEE 7th Global Conference on Consumer Electronics, GCCE 2018*, pp. 796–797, 2018, doi: 10.1109/GCCE.2018.8574610.
- [4] M. Cabanac and M. Cabanac, "What is emotion? What is emotion?," vol. 6357, no. December, pp. 69–83, 2016, doi: 10.1016/S0376-6357(02)00078-5.
- [5] P. Ekman, "Basic Emotions," *Encyclopedia of Personality and Individual Differences*, pp. 1–6, 1999, doi: 10.1007/978-3-319-28099-8_495-1.
- [6] R. Plutchik, "The nature of emotions," *Philosophical Studies*, vol. 52, no. 3, pp. 393–409, 2001, doi: 10.1007/BF00354055.
- [7] J. A. Russell, "A circumplex model of affect," *Journal of Personality and Social Psychology*, vol. 39, no. 6, pp. 1161–1178, 1980, doi: 10.1037/h0077714.
- [8] E. H. Hess, "ATTITUDE AND PUPIL SIZE," *Scientific American*, vol. 212, no. 4, pp. 46–55, Nov. 1965.
- [9] C. Aracena, S. Basterrech, V. Snasel, and J. Velasquez, "Neural Networks for Emotion Recognition Based on Eye Tracking Data," *Proceedings - 2015 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2015*, pp. 2632–2637, 2016, doi: 10.1109/SMC.2015.460.
- [10] A. Gomez-Ibanez, E. Urrestarazu, and C. Viteri, "Recognition of facial emotions and identity in patients with mesial temporal lobe and idiopathic generalized epilepsy: An eye-tracking study," *Seizure*, vol. 23, no. 10, pp. 892–898, 2014, doi: https://doi.org/10.1016/j.seizure.2014.08.012.
- [11] S. Paul, A. Banerjee, and D. N. Tibarewala, "Emotional eye movement analysis using electrooculography signal," *International Journal of Biomedical Engineering and Technology*, vol. 23, no. 1, pp. 59–70, 2017, doi: 10.1504/IJBET.2017.082224.
- [12] Y. Wang, Z. Lv, and Y. Zheng, "Automatic emotion perception using eye movement information for E-healthcare systems," *Sensors (Switzerland)*, vol. 18, no. 9, 2018, doi: 10.3390/s18092826.
- [13] V. Raudonis, G. Dervinis, A. Vilkauskas, A. Paulauskaite, and G. Kersulyte, "Evaluation of Human Emotion from Eye Motions," *International Journal of Advanced Computer Science and Applications*, vol. 4, no. 8, pp. 79–84, 2013, doi: 10.14569/ijacsa.2013.040812.
- [14] Rajakumari and S. Selvi, "HCI and Eye Tracking: Emotion Recognition Using Hidden Markov Model," 2015.
- [15] J. Nam, H. Chung, Y. ah Seong, and H. Lee, "A new terrain in HCI: Emotion recognition interface using biometric data for an immersive VR experience," *arXiv*, 2019.