


SHORT REPORT

Open Access

Four-class emotion classification in virtual reality using pupillometry



Lim Jia Zheng, James Mountstephens and Jason Teo* 

*Correspondence:
jtwteo@ums.edu.my
Faculty of Computing
and Informatics, Universiti
Malaysia Sabah, Jalan UMS,
88400 Kota Kinabalu, Sabah,
Malaysia

Abstract

Background: Emotion classification remains a challenging problem in affective computing. The large majority of emotion classification studies rely on electroencephalography (EEG) and/or electrocardiography (ECG) signals and only classifies the emotions into two or three classes. Moreover, the stimuli used in most emotion classification studies utilize either music or visual stimuli that are presented through conventional displays such as computer display screens or television screens. This study reports on a novel approach to recognizing emotions using pupillometry alone in the form of pupil diameter data to classify emotions into four distinct classes according to Russell's Circumplex Model of Emotions, utilizing emotional stimuli that are presented in a virtual reality (VR) environment. The stimuli used in this experiment are 360° videos presented using a VR headset. Using an eye-tracker, pupil diameter is acquired as the sole classification feature. Three classifiers were used for the emotion classification which are Support Vector Machine (SVM), k-Nearest Neighbor (KNN), and Random Forest (RF).

Findings: SVM achieved the best performance for the four-class intra-subject classification task at an average of 57.05% accuracy, which is more than twice the accuracy of a random classifier. Although the accuracy can still be significantly improved, this study reports on the first systematic study on the use of eye-tracking data alone without any other supplementary sensor modalities to perform human emotion classification and demonstrates that even with a single feature of pupil diameter alone, emotions could be classified into four distinct classes to a certain level of accuracy. Moreover, the best performance for recognizing a particular class was 70.83%, which was achieved by the KNN classifier for Quadrant 3 emotions.

Conclusion: This study presents the first systematic investigation on the use of pupillometry as the sole feature to classify emotions into four distinct classes using VR stimuli. The ability to conduct emotion classification using pupil data alone represents a promising new approach to affective computing as new applications could be developed using readily-available webcams on laptops and other mobile devices that are equipped with cameras without the need for specialized and costly equipment such as EEG and/or ECG as the sensor modality.

Keywords: Emotion classification, Eye-tracking, Pupil diameter, Virtual reality, Machine learning

Introduction

Emotion classification is the task of detecting human emotions, mostly from using facial expressions [8], verbal expressions [5], and physiological measurements. Several applications using emotion classification techniques have been developed and applied to real-world solutions such as driver fatigue monitoring [3] and mental health monitoring [12]. However, most studies on emotion classification based on physiological signals are obtained from electroencephalography (EEG) and electrocardiography (ECG) [18, 22]. Subsequently, much less is known regarding the use of eye-tracking as a sensor modality for detecting emotions. Therefore, the aim of this paper is to report on the results of recognizing emotions using eye-tracking data only in the form of pupil diameter without any other additional modality.

Eye-tracking refers to the method of tracking eye movements and identifying where the user is looking at as well as recording other eye-related attributes such as pupil diameter. Eye-tracking can be utilized in many domains such as marketing research, health-care [13], education [7], psychology, as well as video gaming [2]. Eye-tracking technology can be widely deployed in the near future since it only needs a camera to acquire the required data. As such, it requires significantly fewer sensors to implement in the recording device.

Additionally, most emotion classification studies use movies, images or music as their stimulation tool to evoke the user's emotions. Much fewer studies have used attempted to use Virtual Reality (VR) to present emotional stimuli. VR provides a virtual environment that is highly similar to the real world and as such could potentially evoke stronger emotional responses from the user compared to the other stimulation tools. The user can be immersed in a real-world experience by watching the 360° videos using a VR headset. The user will have less distractions and will be able to focus more on the stimuli in virtual environment.

Three machine learning classifiers were used in the experiment, which are Support Vector Machine (SVM), k-nearest neighbor (KNN), and Random Forest (RF). These methods are suitable for classification tasks. SVM algorithm analyzes data for the analysis of regression and classification. It maps data into a high-dimensional function space such that datasets can be classified even if the data cannot be separated linearly. KNN is a machine learning algorithm that uses data and identifies new datasets according to similarities. It works to evaluate the k-nearest neighbors based on the minimal distance from the test samples to the training dataset. Random forest is a method that generates and fuses multiple decision trees randomly into one "forest". It builds a multitude of decision trees and outputs the class which is the classification or regression of the individual trees. Three different models of classifier are used to determine which machine learning algorithm could obtain the best performance in recognizing the emotions in four quadrants.

Background

Emotion classification

Emotion is a brain-related mental state containing three specific components, which are physiological response, subjective experience, and behavioral response [9]. Emotion

reflects the feeling and thoughts of an individual as well as the degree of pleasure/displeasure. Ekman's model proposed six basic emotions from his research works, which are fear, happiness, anger, surprise, disgust, and sadness [10]. These six basic emotions are then extended to eight emotions, anticipation and trust are added to the list of Plutchik's model [21]. Emotion classification refers to the task of recognizing an individual's emotions and classify an emotion from their reactions and responses. Emotion classification is defined as the categorization of emotions and attempts to differentiate one emotion from another. Russell's Circumplex Model of Affects [25], which contains arousal and valence dimensions with four quadrants of emotions as a result of combining these two dimensions, is among the most commonly adopted emotion model by emotion researchers to test users and attempt to classify their emotions according to these four quadrants. Each quadrant represents the respective emotional states according to the combinations of a high/low arousal (HA/LA) together with a positive/negative valence (PV/NV). Quadrant 1 is a combination of HA/PV and represents the emotions of happy, excited, elated, and alert; Quadrant 2 is a combination of HA/NV and represents the emotions of tense, nervous, stressed, and upset; Quadrant 3 is a combination of LA/NV and represents the emotions of sad, depressed, confused, and bored; while Quadrant 4 is a combination of at LA/PV and represents the emotions of contented, serene, relaxed, and calm. Since there are many complex emotions that occur at each of these quadrants, it is very difficult to determine a particular specific emotion based on the user's responses and reactions. Hence, this letter attempts to classify the emotional analysis by distinguishing the emotions based on the respective quadrant information according to Russell's model of emotions.

Eye-tracking

Eye-tracking is an advanced technology that is used to measure the eye movement or the point of view of an individual. Eye-tracking technology has been applied in many fields such as in various medical research, cognitive psychology studies, as well as in Human-Computer Interaction (HCI) research [17]. Eye movement signals provide the vision localization of an individual and thus enables the direct observation and accurate pinpointing of what is attracting their attention. Eye movement signals can be utilized as an indication of the individual's behaviors and some previous works have used eye signals to investigate the attention of users in reading [24].

Eye-tracking in emotion classification

The eye features such as pupil diameter contain some emotional-relevant characteristics, hence the eye-tracking data can be utilized in emotion classification. There are many eye features can be used to conduct emotion classification such as fixation duration of the pupil, motion speed of the pupil, pupil position, and pupil size. There are studies on the analysis of eye movements for human behavior recognition [14, 16]. However, studies that focus specifically on emotion classification using eye-tracking data alone is very limited as most of such studies incorporate other sensor modalities such as EEG and ECG. There is a study that focuses on the emotional eye movement analysis using electrooculography (EOG) signals [20]. There have also been studies that rely on other types of eye-tracking data such as fixation duration and pupil position [1, 23]. Most of the papers that

rely on eye-tracking data solely only classify the arousal dimension or valence dimension separately or basic 3-class emotions such as positive, neutral, and negative [4, 27]. There is a recent report which reviews recent papers on emotion detection using eye-tracking providing a taxonomy as well as current challenges in this field [19]. To the best of our knowledge, there has thus far not been any systematic study conducted on using eye-tracking data exclusively for four-class emotion classification. Therefore, this letter attempts to perform four-class emotion classification according to the four quadrants from Russell's model.

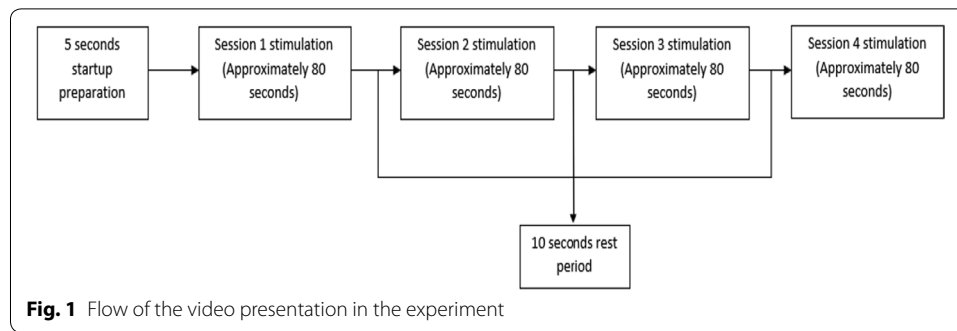
Emotion classification in VR

Through the use of VR technologies, the user is fully immersed in a virtual environment that very closely resembles the real world. As such, this provides a greater sense of reality for the user when they are experiencing visual stimuli through a high level of VR immersion. Moreover, since the user is fully enclosed by their head-mounted display (HMD), this would block out external secondary stimuli which may distract the user from the immediate primary stimuli being experienced by the user. Hence, the user should be more connected to the stimuli being presented in the virtual environment and hence a more direct and real emotional response will be evoked through a VR presentation. A past study has reported that Immersive Virtual Environments (IVE) can indeed be utilized effectively as a presentation tool for emotion inducement [11]. A number of VR HMDs now have the option of integrating eye-tracking devices into their HMDs. As such, eye-tracking data can now be obtained easily with the add-on eye-trackers that are placed into the VR HMD. There is a recent study on emotion classification using a wearable EEG headband and machine learning in a virtual environment by utilizing 360° videos [26]. There are also authors that presented their classification investigations using eye-tracking and VR in facial expressions [6, 15]. Currently, there is no study on emotion classification using eye-tracking data solely in a virtual environment. Therefore, we employ this approach for our work on emotion classification using eye-tracking in VR environments.

Methods

Experiment setup

In this study, VR is used as our emotional presentation stimuli. The HTC Vive VR headset with a pair of earphones was used to stimulate the participant's emotions. The experiments were conducted with the presentation of emotional 360° videos consisting of four distinct emotions according to Russell's model of emotions. A total of ten subjects (9 males and 1 female) participated in our experiment. The age range is 21–28. An explanation was given to all participants before the experiment started. The experiment was conducted with the presentation of a series of 360° videos lasting a total of about 6 min. The eye-tracking data of participants were recorded using an add-on eye-tracker from Pupil Labs for the VR headset. The flow of the video presentation in the experiment is as shown in Fig. 1. There were four separate sessions of video stimulation lasting 80 s each according to each of the quadrants of emotion. A 10-s rest period is given before the next video stimulation session from the following quadrant is commenced.

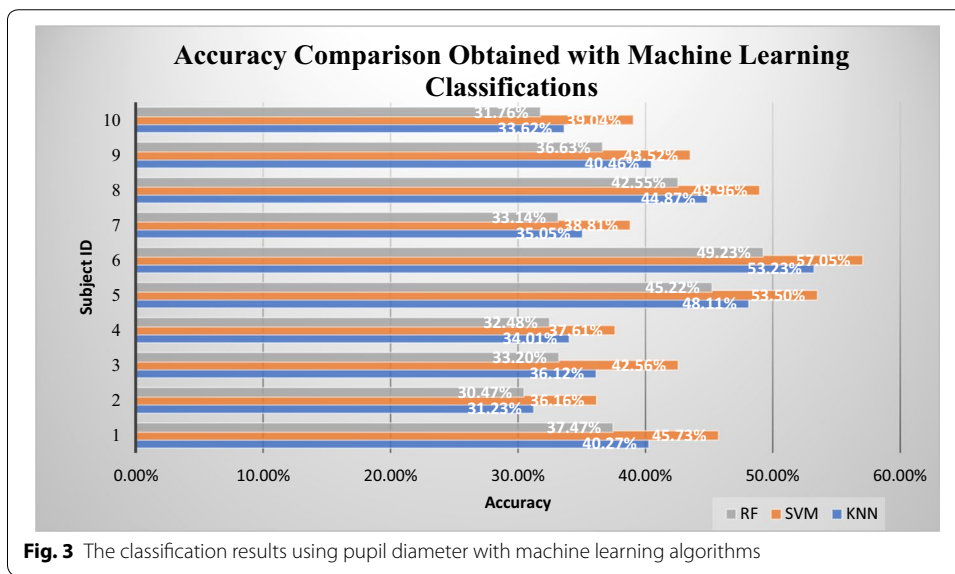
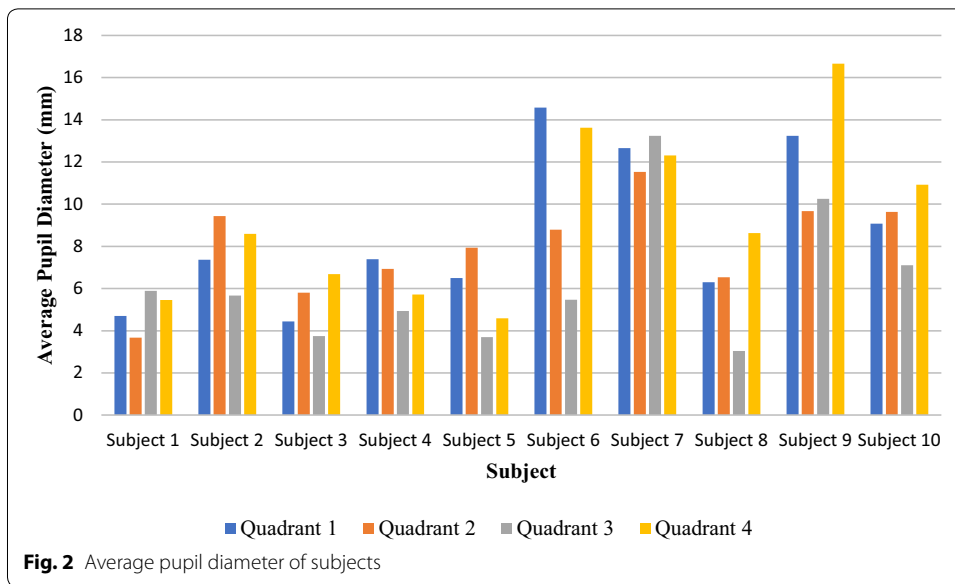


Data collection and classification methods

The eye-tracking data were collected using the Pupil Labs application. In the data collection process, eye calibration is conducted for each of the participants. At first, all participants will wear the VR headset with the add-on eye-tracker. The pupil data was recorded using Pupil Capture. The presenting video and data recording are started simultaneously. The video is presented using Unity, a platform for 360° videos, with the recording script in C# programming language. Data recording will be stopped simultaneously when the video has completed playing. The recording directory is then transferred to the Pupil Player for the data visualization process. These recorded data were the exported using the raw data exporter in Pupil Player and saved to a CSV file format. There are several types of eye-tracking data that can be exported from Pupil Player such as gaze data, fixations, and pupil data. Pupil diameter was chosen as the eye feature in this experiment. The data is capture through pupil detection and the diameter of the pupil is estimated in millimeters (mm) based on the diameter of the anthropomorphic average eyeball. This study utilized stimuli prepared using VR-based content. This dataset was specifically collected using a VR environment from each participant and has approximately 70,000 datapoints with a timestamp included for every second of acquisition. The machine learning tasks were done by using Python. Three types of machine learning classifiers were used in this experiment, namely Support Vector Machine (SVM), K-nearest neighbor (KNN), and Random Forest, to classify the emotions. The SVM classifier was used with the Radial Basis Function (RBF) kernel in this experiment while the range the for k value in the KNN classifier was set to 5.

Results and discussion

From Fig. 2, the results show that most of the pupil diameter is largest in Quadrant 4 while it is the smallest in Quadrant 3. It also showed that the pupil diameter has the biggest changes in low arousal level. The outcomes also showed that the pupil diameter is smaller in Quadrants 2 and 3 which is located at the negative valence. These characteristics show that pupil diameter does indeed exhibit changes with certain emotions, hence we can extract such emotional-relevant features to conduct classification using machine learning algorithms to attempt to distinguish between the different quadrants of emotions.



The classification results using pupil diameter were obtained from three different machine learning algorithms, which were the KNN, SVM, and RF classifiers. From Fig. 3, the SVM classifier showed the best performance for emotion classification when comparing against the KNN and RF classifiers. The highest accuracy achieved was 57.05% by the SVM classifier while the highest accuracy obtained from the KNN and RF classifiers was 53.23% and 49.23% respectively.

Tables 1, 2, 3 shows the confusion matrices of each emotion classification for participant 5 and 6, which were chosen because the classification results from these two particular participants showed the highest average classification accuracy across the four classes. The highest accuracy achieved from when comparing across all the classifier’s confusion matrix was for Quadrant 3 in participant 6, which was 69.86% for

Table 1 Confusion matrices of participants 5 and 6 using SVM

Class	1	2	3	4
1	1051 (51.21%)	292	346	301
2	376	1196 (58.85%)	322	93
3	146	336	1031 (47.66%)	157
4	479	208	464	772 (58.35%)
Class	1	2	3	4
1	1258 (42.28%)	106	209	404
2	704	209 (46.03%)	843	189
3	169	62	3413 (69.86%)	103
4	844	77	420	605 (46.50%)

Table 2 Confusion matrices of participants 5 and 6 using KNN

Class	1	2	3	4
1	1161 (44.63%)	247	292	290
2	485	1081 (56.98%)	292	129
3	309	343	754 (43.25%)	264
4	646	226	405	646 (48.60%)
Class	1	2	3	4
1	1283 (42.17%)	208	185	301
2	693	452 (31.71%)	639	161
3	195	565	2863 (70.83%)	123
4	871	200	355	520 (47.05%)

Table 3 Confusion matrices of participants 5 and 6 using RF

Class	1	2	3	4
1	1003 (43.72%)	262	303	422
2	469	945 (53.42%)	321	252
3	285	331	704 (41.12%)	350
4	537	231	384	771 (42.95%)
Class	1	2	3	4
1	1142 (40.64%)	260	187	388
2	661	433 (28.13%)	595	256
3	227	605	2572 (69.64%)	342
4	780	241	339	586 (37.27%)

SVM, 70.83% for KNN, and 69.64% for RF. The next highest classification rate was observed in participant 5 for Quadrant 2, which was 58.85% for SVM, 56.98% for KNN, and 53.42% for RF. Both these observations suggest that pupil diameter appears to be a promising approach to use when attempting to identify LA/NV and HA/NV emotions, which are located in the negative valence quadrants. This appears to be a novel finding which shows some relationship between pupil diameter and negative valence stimuli as pupil diameter was only previously correlated to arousal levels.

Conclusion

In this letter, we classified emotions according to Russell's four-quadrant Circumplex Model of Emotions using eye-tracking data from the presentation of VR 360° video stimuli to the participants. We collected the eye-tracking data from an eye-tracker that was mounted inside the VR headset and pupil diameter was chosen as the eye feature for emotion classification in this experiment. We used three different machine learning algorithms to conduct classification tasks. The findings showed that Support Vector Machine (SVM) had the best average accuracy of 57.05% across all four quadrants compared to the other two classifiers which are K-nearest neighbor (KNN), and Random Forest (RF). From the analysis of the confusion matrix, it was also observed that the accuracy for correctly predicting emotions resulting from the LA/NV quadrant of emotions was the highest at around 70% for all three classifiers for a particular participant. For future work, this study will attempt to compare the performance of four-class inter-subject emotion classification as well as to investigate the use of deep learning.

Abbreviations

ECG: Electrocardiography; EEG: Electroencephalography; HA: High arousal; HMD: Head-mounted display; KNN: k-nearest neighbor; LA: Low arousal; NV: Negative valence; PV: Positive valence; RBF: Radial basis function; RF: Random forest; SVM: Support vector machine; VR: Virtual reality.

Acknowledgements

This work was supported by the Ministry of Energy, Science, Technology, Environment and Climate Change (MESTECC), Malaysia [Grant Number ICF0001-2018].

Authors' contributions

LJZ—writing—original draft. JM—writing—review & editing, supervision. JT—writing—review & editing, supervision, funding acquisition. All authors read and approved the final manuscript.

Funding

This work was supported by the Ministry of Energy, Science, Technology, Environment and Climate Change (MESTECC), Malaysia [Grant Number ICF0001-2018].

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Received: 30 March 2020 Accepted: 21 June 2020

Published online: 06 July 2020

References

1. Alhargan A, Cooke N, Binjammaz T. Multimodal affect recognition in an interactive gaming environment using eye tracking and speech signals. In: *ICMI 2017—proceedings of the 19th ACM international conference on multimodal interaction*; 2017. p. 479–86. <https://doi.org/10.1145/3136755.3137016>.
2. Almeida S, Mealha Ó, Veloso A. Video game scenery analysis with eye tracking. *Entertain Comput*. 2016;14:1–13. <https://doi.org/10.1016/j.entcom.2015.12.001>.
3. Alsibai MH, Manap SA. A study on driver fatigue notification systems. *ARPN J Eng Appl Sci*. 2016;11(18):10987–92.
4. Aracena C, Basterrech S, Snasel V, Velasquez J. Neural networks for emotion recognition based on eye tracking data. In: *Proceedings—2015 IEEE international conference on systems, man, and cybernetics, SMC 2015*; 2016. p. 2632–7. <https://doi.org/10.1109/SMC.2015.460>.
5. Basu S, Chakraborty J, Aftabuddin M. Emotion recognition from speech using convolutional neural network with recurrent neural network architecture. In: *Proceedings of the 2nd international conference on communication and electronics systems, ICCES 2017, 2018-Jan (Icces)*; 2018. p. 333–336. <https://doi.org/10.1109/CESYS.2017.8321292>.
6. Bekele E, Bian D, Zheng Z, Peterman J, Park S, Sarkar N. Responses during facial emotional expression recognition tasks using virtual reality and static IAPS pictures for adults with schizophrenia. In: *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 8526 LNCS (PART 2); 2014. p. 225–35. https://doi.org/10.1007/978-3-319-07464-1_21.

7. Busjahn T, Begel A, Orlov P, Sharif B, Hansen M, Bednarik R, Shchekotova G. Eye tracking in computing education categories and subject descriptors. In: ACM: proceedings of the tenth annual conference on international computing education research; 2014. p. 3–10. <https://doi.org/10.1145/2632320.2632344>.
8. Chanthaphan N, Uchimura K, Satonaka T, Makioka T. Facial emotion recognition based on facial motion stream generated by kinect. In: Proceedings—11th international conference on signal-image technology and internet-based systems, SITIS 2015; 2016. p. 117–124. <https://doi.org/10.1109/SITIS.2015.31>.
9. Damasio AR. Emotion in the perspective of an integrated nervous system. *Brain Res Rev*. 1998;26(2–3):83–6. [https://doi.org/10.1016/S0165-0173\(97\)00064-7](https://doi.org/10.1016/S0165-0173(97)00064-7).
10. Ekman P. Basic emotions. *Encyclopedia of personality and individual differences*. Cham: Springer; 1999. p. 1–6. https://doi.org/10.1007/978-3-319-28099-8_495-1.
11. Gorini A, Mosso JL, Mosso D, Pineda E, Ruiz NL, Ramírez M, et al. Emotional response to virtual reality exposure across different cultures: the role of the attribution process. *CyberPsychol Behav*. 2009;12(6):699–705. <https://doi.org/10.1089/cpb.2009.0192>.
12. Guo R, Li S, He L, Gao W, Qi H, Owens G. Pervasive and unobtrusive emotion sensing for human mental health. In: Proceedings of the 2013 7th international conference on pervasive computing technologies for healthcare and workshops, PervasiveHealth 2013; 2013. p. 436–9. <https://doi.org/10.4108/icst.pervasivehealth.2013.252133>.
13. Henneman EA, Marquard JL, Fisher DL, Gawlinski A. Eye tracking: a novel approach for evaluating and improving the safety of healthcare processes in the simulated setting. *Simul Healthcare*. 2017;12(1):51–6. <https://doi.org/10.1097/SIH.0000000000000192>.
14. Hess EH. The tell-tale eye: How your eyes reveal hidden thoughts and emotions. In *The tell-tale eye: How your eyes reveal hidden thoughts and emotions*. Oxford: Van Nostrand Reinhold; 1975.
15. Hickson S, Kwatra V, Dufour N, Sud A, Essa I. Eyemotion: classifying facial expressions in VR using eye-tracking cameras. In: Proceedings—2019 IEEE winter conference on applications of computer vision, WACV 2019; 2019. p. 1626–1635. <https://doi.org/10.1109/WACV.2019.00178>.
16. Isaacowitz DM, Wadlinger HA, Goren D, Wilson HR. Selective preference in visual fixation away from negative images in old age? An eye-tracking study. *Psychol Aging*. 2006;21:40–8. <https://doi.org/10.1037/0882-7974.21.1.40>.
17. Jacob RJK, Karn KS. Eye tracking in human-computer interaction and usability research: ready to deliver the promises. *Mind's Eye*. 2003. <https://doi.org/10.1016/B978-044451020-4/50031-1>.
18. Ko KE, Yang HC, Sim KB. Emotion recognition using EEG signals with relative power values and Bayesian network. *Int J Control Autom Syst*. 2009;7(5):865–70. <https://doi.org/10.1007/s12555-009-0521-0>.
19. Lim JZ, Mountstephens J, Teo J. Emotion recognition using eye-tracking: taxonomy, review and current challenges. *Sensors (Switzerland)*. 2020;20(8):1–21. <https://doi.org/10.3390/s20082384>.
20. Paul S, Banerjee A, Tibarewala DN. Emotional eye movement analysis using electrooculography signal. *Int J Biomed Eng Technol*. 2017;23(1):59–70. <https://doi.org/10.1504/IJBET.2017.082224>.
21. Plutchik R. The nature of emotions. *Philos Stud*. 2001;52(3):393–409. <https://doi.org/10.1007/BF00354055>.
22. Rattanyu K, Ohkura M, Mizukawa M. Emotion monitoring from physiological signals for service robots in the living space. In: ICCAS 2010—international conference on control, automation and systems; 2010. p. 580–583. <https://doi.org/10.1109/ICCAS.2010.5669914>.
23. Raudonis V, Dervinis G, Vilkauskas A, Paulauskaite A, Kersulyte G. Evaluation of human emotion from eye motions. *Int J Adv Comput Sci Appl*. 2013;4(8):79–84. <https://doi.org/10.14569/ijacsa.2013.040812>.
24. Rayner K. Eye movements and attention in reading, scene perception, and visual search. *Quart J Exp Psychol*. 2009;62(8):1457–506. <https://doi.org/10.1080/17470210902816461>.
25. Russell JA. A circumplex model of affect. *J Pers Soc Psychol*. 1980;39(6):1161–78. <https://doi.org/10.1037/h0077714>.
26. Teo J, Suhaimi NS, Mountstephens J. Augmenting EEG with inertial sensing for improved 4-class subject-independent emotion classification in virtual reality; 2019. p. 1–8. <https://doi.org/10.4108/eai.18-7-2019.2287946>.
27. Wang Y, Lv Z, Zheng Y. Automatic emotion perception using eye movement information for E-healthcare systems. *Sensors (Switzerland)*. 2018;18(9):2826. <https://doi.org/10.3390/s18092826>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)
