

CorrFeat

Correlation-based feature extraction algorithm using skin conductance and pupil diameter for emotion recognition

Zhang, Tianyi; Ali, Abdallah El; Wang, Chen; Zhu, Xintong; Cesar, Pablo

DOI

[10.1145/3340555.3353716](https://doi.org/10.1145/3340555.3353716)

Publication date

2019

Document Version

Final published version

Published in

ICMI 2019 - Proceedings of the 2019 International Conference on Multimodal Interaction

Citation (APA)

Zhang, T., Ali, A. E., Wang, C., Zhu, X., & Cesar, P. (2019). CorrFeat: Correlation-based feature extraction algorithm using skin conductance and pupil diameter for emotion recognition. In W. Gao, H. M. Ling Meng, M. Turk, S. R. Fussell, B. Schuller, B. Schuller, Y. Song, & K. Yu (Eds.), *ICMI 2019 - Proceedings of the 2019 International Conference on Multimodal Interaction: Proceedings of the 2019 International Conference on Multimodal Interaction* (pp. 404-408). (ICMI 2019 - Proceedings of the 2019 International Conference on Multimodal Interaction). Association for Computing Machinery (ACM).
<https://doi.org/10.1145/3340555.3353716>

Important note

To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

CorrFeat: Correlation-based Feature Extraction Algorithm using Skin Conductance and Pupil Diameter for Emotion Recognition

Tianyi Zhang

Centrum Wiskunde & Informatica
Delft University of Technology
Amsterdam, The Netherlands
tianyi.zhang@cwi.nl

Abdallah El Ali

Centrum Wiskunde & Informatica
Amsterdam, The Netherlands
abdallah.el.ali@cwi.nl

Chen Wang

Future Media Convergence Institute
Xinhuanet
Beijing, China
wangchen@news.cn

Xintong Zhu

Future Media Convergence Institute
Xinhuanet
Beijing, China
zhuxintong@news.cn

Pablo Cesar

Centrum Wiskunde & Informatica
Delft University of Technology
Amsterdam, The Netherlands
p.s.cesar@cwi.nl

ABSTRACT

To recognize emotions using less obtrusive wearable sensors, we present a novel emotion recognition method that uses only pupil diameter (PD) and skin conductance (SC). Psychological studies show that these two signals are related to the attention level of humans exposed to visual stimuli. Based on this, we propose a feature extraction algorithm that extract correlation-based features for participants watching the same video clip. To boost performance given limited data, we implement a learning system without a deep architecture to classify arousal and valence. Our method outperforms not only state-of-art approaches, but also widely-used traditional and deep learning methods.

CCS CONCEPTS

• **Computing methodologies** → **Machine learning**; • **Human-centered computing** → *Human computer interaction (HCI)*.

KEYWORDS

Emotion recognition; Skin conductance response; Pupil diameter; Machine learning; MAHNOB-HCI database

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

ICMI '19, October 14–18, 2019, Suzhou, China

© 2019 Association for Computing Machinery.

ACM ISBN 978-1-4503-6860-5/19/10...\$15.00

<https://doi.org/10.1145/3340555.3353716>

ACM Reference Format:

Tianyi Zhang, Abdallah El Ali, Chen Wang, Xintong Zhu, and Pablo Cesar. 2019. CorrFeat: Correlation-based Feature Extraction Algorithm using Skin Conductance and Pupil Diameter for Emotion Recognition. In *2019 International Conference on Multimodal Interaction (ICMI '19), October 14–18, 2019, Suzhou, China*. ACM, New York, NY, USA, 5 pages. <https://doi.org/10.1145/3340555.3353716>

1 INTRODUCTION

Emotion recognition using physiological signals has been widely applied to many fields such as mental health care [15], engagement evaluation [12] and driver monitoring [34]. Although previous work has obtained high recognition accuracy, most of these methods need to use non-wearable devices such as EEG [27], fMRI [6] or a face camera [20]. With the rapid development of wearable (biosensing) technologies, there are several wearable devices to measure the skin conductance (SC) and pupil diameter (PD). For example, SC can be measured by Empatica E4 wristband without attaching electrodes to the skin, which makes it less intrusive than EEG and ECG. The pupil diameter, which carries abundant information regarding cognitive activities [30], can be measured using wearable eye trackers such as Tobii glasses. Both SC and PD are related to the attention level to visual stimuli [9, 17, 19], which makes them suitable for emotion recognition during video watching. Thus, we focus only on SC and PD to recognize emotions.

Feature extraction from multiple physiological signals has several challenges. One of the major problems is that the lengths of physiological signals from different sessions when subjects watch different videos are always variable. It is also challenging to build a standard machine learning pipeline to automatically extract features from data with varied sizes.

Most previous work [14, 23, 31] solve this problem by manually designing handcrafted features across different sensors and fusing them at decision level [29], which is time-consuming and often leads to low accuracies. Other works [1, 28, 33] segment or pad the signals to let them have fixed lengths and train the data with neural networks. These kinds of methods could lead to over-fitting problems because of mislabeled segmentations and limited amount of data.

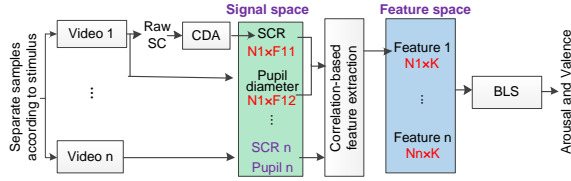


Figure 1: Flow chart of the proposed algorithm

To solve this problem, we present a correlation-based feature extraction algorithm (*CorrFeat*) to automatically extract features from SC and PD. As shown in Figure 1, the samples are first separated into different groups according to the video clips the subjects watched. Then, the skin conductance response (SCR) is extracted from the raw SC signals using Continuous Deconvolution Analysis (CDA) [2]. The joint features from SCR and PD are then extracted by a correlation-based feature extraction algorithm. The extracted features are finally inputted into a Broad learning System (BLS) [7] to classify valence and arousal.

2 ALGORITHM

The extraction of SCR

The signal of SC can be separated into two activities: a slowly varying tonic activity, which is known as skin conductance level (SCL), and a fast varying phasic activity, which is known as skin conductance response (SCR) [4]. SCR is conceived to carry the essential information of sympathetic activity, which can reflect stimulus-specific responses [5]. SCL however, is often related to non-stimulus-specific events which varies from different individuals. To extract the stimulus-specific responses from the SC signal, we implement CDA [2] on the SC signal.

The raw SC signal is first filtered by a Hanning window to reduce high frequency noisy fluctuations of the raw signal [24, 32]. After that, SCR can be extracted from the SC signal using deconvolution:

$$SCR = Driver_{phasic} * \xi = (SC *^{-1} \xi - Driver_{tonic}) * \xi \quad (1)$$

where the $*^{-1}$ stands for the standard deconvolution operation. The $Driver_{tonic}$ and $Driver_{phasic}$ are sudomotor nerve activities caused by non-stimulus-specific and stimulus-specific events respectively. $Driver_{tonic}$ can be estimated using the

inter-impulse data detection [3]. $\xi = e^{-t/\tau_1} - e^{-t/\tau_2}$ ($\tau_1 = 0.75, \tau_2 = 2$) is the trigger function.

Correlation-based feature extraction

The purpose of the correlation-based feature extraction is to extract joint features with the unified data structure, which can be easily input into any machine learning systems without further pre-processing. To manage this, we first separate all samples into different groups according to the stimulus the subjects have watched. The SC and PD for the same visual stimulus should have the same length. Moreover, we hypothesize that the same stimulus will trigger relatively similar valence and arousal among different subjects. Thus, our algorithm is about to learn different transformation matrices on different groups.

Suppose SCR_{mn} and P_{mn} stands for SCR and the inverse of PD of the subject m watch the video n . $n \in [1, N]$, $m \in [1, M]$, where N and M is the number of subjects and video stimulus respectively. $SCR_m \in SCR = [SCR_1, SCR_2, \dots, SCR_M]$ and $P_m \in P = [P_1, P_2, \dots, P_M]$ is the group of signals which different subjects watch the same video. The target of the feature extraction is to find two linear projections ω_{m1}, ω_{m2} which can maximize the correlation between $SCR_m \omega_{m1}$ and $P_m \omega_{m2}$ with the constrain $\omega_{m1}^T S_{11} \omega_{m1} = \omega_{m2}^T S_{22} \omega_{m2} = 1$:

$$(\omega_{m1}, \omega_{m2}) = \arg \max(\omega_{m1}^T S_{12} \omega_{m2}) \quad (2)$$

where S_{11} and S_{22} is the covariance of SCR_m and P_m respectively. S_{12} is the cross-covariance of SCR_m and P_m . To solve equation (2), we follow the Singular Value Decomposition (SVD) method in [21]. The mean values of each column of SCR_m and P_m are first subtracted to remove the subject bias. After that, we calculate the covariance and cross-covariance of the two signals. Then, we implement the SVD on the equation: $[U, D, V] = \text{SVD}(V_s D_s V_s^T \cdot S_{12} \cdot V_p D_p V_p^T)$, where D_s and D_p are diagonal matrices whose diagonal elements are the k biggest non-zero eigenvalues of S_{11} and S_{22} respectively, where $D_s = \text{diag}(\frac{1}{\sqrt{D_{s1}}}, \frac{1}{\sqrt{D_{s2}}}, \dots, \frac{1}{\sqrt{D_{sk}}})$ and D_p has the same format. $V_s = [V_{s1}, V_{s2}, \dots, V_{sk}]$ is composed of the k corresponding eigenvectors of $[D_{s1}, D_{s2}, \dots, D_{sk}]$ respectively. The V_p are calculated using the same method. Now, the two linear projection $(\omega_{m1}, \omega_{m2})$, which map the two signals into feature space can be calculated by:

$$\omega_{m1} = V_s D_s V_s^T \cdot U', \quad \omega_{m2} = V_p D_p V_p^T \cdot V' \quad (3)$$

where U' and V' are consisted of the first K columns of U, V respectively. At last, the joint feature of SCR_m and P_m can be obtained by: $H_m = [SCR_m \omega_{m1}, P_m \omega_{m2}]$. We implement the above procedure on all the M groups. At last, we map all signals into a feature space: $H = [H_1^T, H_2^T, \dots, H_M^T]^T \in R^{MN \times 2K}$, where K is the dimension of the feature space. In our experiment, we choose $K = 25$ because most features whose index greater than 25 are approximate equal to zero.

BLS for classification

After mapping the signals into the feature space, BLS [7] is used to classify the valence and arousal. Compared with deep learning systems, BLS is less time-consuming because it does not have a big number of hyperparameters and complicated structures [8]. BLS maps original training data into two high dimensional nodes (i.e., feature nodes and enhancement nodes). Instead of using back propagation to calculate the weights between the nodes and labels, BLS calculate the weights by pseudoinverse, which make the classification process to be faster and have less chances to be over-fitted.

Suppose $H' \in R^{L \times 2K}$ is the training set selected from the correlation based features $H \in R^{MN \times 2K}$. We first normalize H' to have mean of 0 and standard variance of 1 using z score normalization [18]. Then, the feature node A and the enhancement nodes B can be calculated using random representation and non-linear mapping. The last step of BLS is to calculate the weights between the input nodes and labels by the pseudoinverse of the input nodes:

$$W = \frac{E^T \cdot E + I^{L \times L} \cdot C}{E^T \cdot y} \quad (4)$$

where $E = [A, B]$, $I^{L \times L}$ is the unit matrix with the dimension of $L \times L$. $C = 2^{-30}$ is the regularization parameter for sparse regularization. After that, the network has been established. If a new sample E_t comes, the output y_t it can be obtained by $y_t = E_t \cdot W$.

3 DATASET

To evaluate the performance of the proposed algorithm, we test it on the MAHNOB-HCI database [26]. MAHNOB-HCI database contains self-labeled physiological signals and eye-tracking data of 27 subjects when they were watching 20 video clips. We choose this database because: 1) this database is the only existing public database which contains both eye-tracking data and SC signals with emotion labels, 2) all the signals in this database are strictly synchronized by specialized hardware and 3) there are more subjects participating in the experiment of MAHNOB-HCI database compared with some other emotion database [13, 16, 22].

4 RESULTS AND ANALYSIS

Leave-one-subject-out cross-validation

To test the performance of the proposed algorithm, we run the leave-one-subject-out cross-validation on MAHNOB-HCI database. The test was ran on a laptop with NVIDIA 1060 GPU and 16 GB RAM. We select the data from one subject as validation set and the rest of them as the training set (no test samples are used for training). We do not implement double cross validation as this can indeed introduce bias.

Then we run the test on the 27 subjects to test the subject-independent performance of the algorithm. For the binary classification (high-low), we use the labels from 1-9 scaling manikin. For the 3-class classification (high-neutral-low), we use the labels from both emotion keywords and 1-9 scaling manikin. Table 1 shows the mapping between emotional keywords, SAM scaling and different classes [26].

Table 1: The mapping of the emotion keywords and three classes of arousal and valence[26]

class	Emotional keywords	SAM ¹	SAM ²
Low ³	sadness, disgust, neutral	1-4	1-3
Neutral ³	joy and happiness, amusement	-	4-6
High ³	surprise, fear, anger, anxiety	5-9	7-9
Low ⁴	fear, anger, disgust, sadness, anxiety	1-4	1-3
Neutral ⁴	surprise, neutral	-	4-6
High ⁴	joy and happiness, amusement	5-9	7-9

1. Binary classes. 2. Three classes. 3. arousal. 4. valence

As shown in Figure 2, we obtain overall accuracy of 73.12% and 89.22% for arousal and valence using the labels from SAM. For the 3-class classification, the accuracy using the emotional keywords (82.90% for arousal and 82.10% for valence) are higher than the accuracy using 1-9 SAM scaling (59.81% for arousal and 68.60% for valence). This result demonstrates that the accuracy of our algorithm does not rely on the selection of training data, which shows a good performance of subject-independence.

Comparison with ML and DL methods

In this section, we compare the performance of our method with widely-used traditional machine learning (ML) and deep learning (DL) methods. For the ML methods, we select the mean, standard variance, average root mean square, maximum amplitude, average amplitude and mean of the absolute values for SCR, SCL, SC, PD, first and second differential of SC and PD according to [25]. We also test these methods through the *CorrFeat* generated from our method. For the deep learning methods, we pad both these signals of zeros [33] to let them have the same size for training.

Table 2 shows the result of the comparison. All validation and training approaches are the same as section 4. Our method outperforms both the traditional ML and DL methods. In addition, *CorrFeat* also performs well on other classifiers, however the recognition accuracy is not as good as the one using BLS. From the comparison with deep learning methods, we find that making the network deeper does not lead to an increase in accuracy. This is due to over-fitting when the network does not have enough data for training.

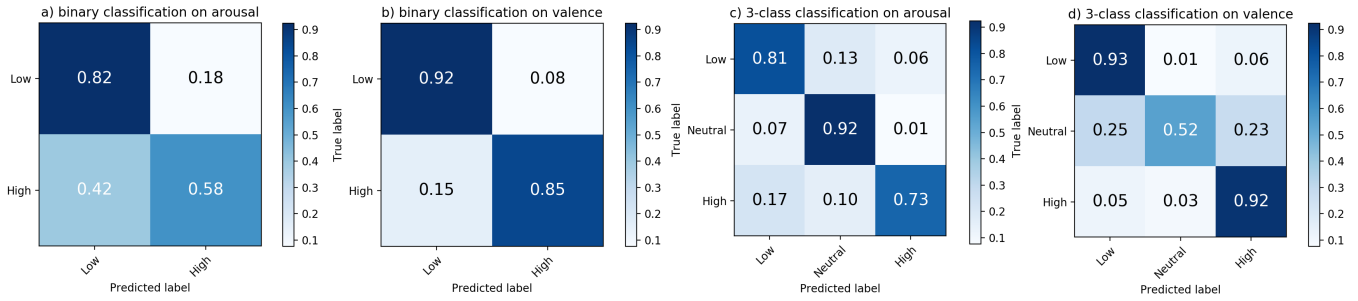


Figure 2: Confusion matrix for a) binary classification for arousal (high-low), b) binary classification for valence (high-low) c) 3-class classification for arousal (high-neutral-low), d) 3-class classification for valence (high-neutral-low)

Table 2: Comparison between traditional machine learning methods on MAHNOB-HCI (accuracy %)

	Correlation-based features					Manually selected features					Deep learning methods			CorrFeat	
	SVM ⁴	KNN ⁵	LDA ⁶	RF ⁷	NB ⁸	SVM	KNN	LDA	RF	NB	BLS	CNN ⁹	CNN ¹⁰		LSTM
Arousal¹	66.7	65.7	67.3	64.8	63.6	57.1	53.9	59.5	56.6	39.3	67.8	61.5	60.8	49.7	73.1
Valence¹	88.9	86.6	88.9	87.3	85.7	63.2	61.1	65.1	64.5	43.5	72.5	66.7	69.2	55.3	89.2
Arousal²	54.9	42.4	51.3	52.6	49.3	36.2	36.5	43.9	41.6	27.0	49.3	42.3	42.9	37.0	59.8
Valence²	42.4	59.5	63.3	60.4	63.6	36.5	41.8	47.7	47.7	35.9	56.5	36.6	32.2	35.6	68.6
Arousal³	78.4	80.4	80.6	78.9	77.3	47.2	39.2	53.3	48.7	37.9	56.5	42.8	43.2	35.2	82.9
Valence³	82.0	77.4	81.8	81.3	78.1	45.0	39.1	51.9	51.0	40.4	57.1	46.4	47.7	37.0	82.1

1.Binary classification using 1-9 scaling values as ground truth. 2. Three-class classification using 1-9 scaling values as ground truth. 3. Three-class classification using 9 emotional keywords as ground truth. 4. Support Vector Machine. 5.K-Nearest Neighbor. 6.Linear Discriminant Analysis. 7.Random Forest. 8.Naive Bayes. 9. 1D-CNN with 2 conv layers. 10. 1D-CNN with 4 conv layers.

Moreover, the padding also limits the performance of these methods by breaking the time-dependent feature of signals.

Table 3: Comparison with the state-of-art on MAHNOB-HCI

	Year	Signals	Accuracy(%) F1-score			
			Aro.	Val.	Aro.	Val.
Soleymani et al [26]	2012	EEG, eye gaze	67.7	76.1	0.62	0.74
Ferdinando et al [11]	2017	ECG	64.1	69.6	-	-
Gui et al [14]	2018	PD	73.0	78.5	0.72	0.77
Ferdinando et al [10]	2018	ECG,SC	81.9	78.7	-	-
CorrFeat (proposed)	2019	SC,PD	82.9	82.1	0.83	0.81

Comparison with the state-of-art methods

Table 3 demonstrates the comparison between the state-of-art methods which use the emotional keywords as labels for 3-class classification on MAHNOB-HCI database. Our method achieves the highest accuracy among other methods with only SC and PD. Gui et al [14] also obtain good result with only PD. However, they only validate their method on 23 subjects among all 27 in MAHNOB-HCI database. All the remaining methods include ECG signals to classify the

arousal and valence. However, the measurement of ECG signal requires attaching electrodes to the skin, which makes it more intrusive than measuring SC. Our method also has the highest F1-score among all 3 methods. It is worth noting that our method does not result in lower F1-score because of the sample imbalance reported by Gui et al [14], which shows *CorrFeat* can better generalize the distribution of data according to their correlations.

5 CONCLUSION AND FUTURE WORK

In this paper, we present a new feature extraction algorithm for emotion recognition using only SC and PD. The validation result on MAHNOB-HCI database shows that it outperforms both the state-of-art methods and other machine learning methods using only two sensors, which provides a new baseline for emotion recognition using wearable sensors. Our future plan is to run an experiment in mobile environments to collect data using wearable eye trackers and SC sensors to find out if our method can still perform well across mobile and ubiquitous environments.

ACKNOWLEDGMENT

This work was supported by the Joint PhD Program between Xinhuanet and Centrum Wiskunde & Informatica.

REFERENCES

- [1] Nese Alyuz, Eda Okur, Ece Oktay, Utku Genc, Sinem Aslan, Sinem Emine Mete, Bert Arnrich, and Asli Arslan Esme. 2016. Semi-supervised model personalization for improved detection of learner's emotional engagement. In *Proceedings of the 18th ACM International Conference on Multimodal Interaction*. ACM, 100–107.
- [2] Mathias Benedek and Christian Kaernbach. 2010. A continuous measure of phasic electrodermal activity. *Journal of neuroscience methods* 190, 1 (2010), 80–91.
- [3] Mathias Benedek and Christian Kaernbach. 2010. Decomposition of skin conductance data by means of nonnegative deconvolution. *Psychophysiology* 47, 4 (2010), 647–658.
- [4] Wolfram Boucsein. 2012. *Electrodermal activity*. Springer Science & Business Media.
- [5] Jason J Braithwaite, Derrick G Watson, Robert Jones, and Mickey Rowe. 2013. A guide for analysing electrodermal activity (EDA) & skin conductance responses (SCRs) for psychological experiments. *Psychophysiology* 49, 1 (2013), 1017–1034.
- [6] Gao-Yi Chao, Chun-Min Chang, Jeng-Lin Li, Ya-Tse Wu, and Chi-Chun Lee. 2018. Generating fMRI-Enriched Acoustic Vectors using a Cross-Modality Adversarial Network for Emotion Recognition. In *Proceedings of the 2018 on International Conference on Multimodal Interaction*. ACM, 55–62.
- [7] CL Philip Chen and Zhulin Liu. 2018. Broad learning system: An effective and efficient incremental learning system without the need for deep architecture. *IEEE transactions on neural networks and learning systems* 29, 1 (2018), 10–24.
- [8] Hongtian Chen and Bin Jiang. 2019. A review of fault detection and diagnosis for the traction system in high-speed trains. *IEEE Transactions on Intelligent Transportation Systems* (2019).
- [9] Frank D Colman and Allan Paivio. 1969. Pupillary response and galvanic skin response during an imagery task. *Psychonomic Science* 16, 6 (1969), 296–297.
- [10] Hany Ferdinando and Esko Alasaarela. 2018. Enhancement of Emotion Recognition using Feature Fusion and the Neighborhood Components Analysis.. In *ICPRAM*. 463–469.
- [11] Hany Ferdinando, Tapio Seppänen, and Esko Alasaarela. 2017. Enhancing Emotion Recognition from ECG Signals using Supervised Dimensionality Reduction.. In *ICPRAM*. 112–118.
- [12] Julien Fleureau, Philippe Guillotel, and Izabela Orlac. 2013. Affective benchmarking of movies based on the physiological responses of a real audience. In *Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference on*. IEEE, 73–78.
- [13] Michael Grimm, Kristian Kroschel, and Shrikanth Narayanan. 2008. The Vera am Mittag German audio-visual emotional speech database. In *2008 IEEE international conference on multimedia and expo*. IEEE, 865–868.
- [14] Dongdong Gui, Sheng-hua Zhong, and Zhong Ming. 2018. Implicit Affective Video Tagging Using Pupillary Response. In *International Conference on Multimedia Modeling*. Springer, 165–176.
- [15] Rui Guo, Shuangjiang Li, Li He, Wei Gao, Hairong Qi, and Gina Owens. 2013. Pervasive and unobtrusive emotion sensing for human mental health. In *Pervasive Computing Technologies for Healthcare (Pervasive-Health), 2013 7th International Conference on*. IEEE, 436–439.
- [16] Jennifer Healey, Rosalind W Picard, et al. 2005. Detecting stress during real-world driving tasks using physiological sensors. *IEEE Transactions on intelligent transportation systems* 6, 2 (2005), 156–166.
- [17] Eckhard H Hess and James M Polt. 1960. Pupil size as related to interest value of visual stimuli. *Science* 132, 3423 (1960), 349–350.
- [18] Anil Jain, Karthik Nandakumar, and Arun Ross. 2005. Score normalization in multimodal biometric systems. *Pattern recognition* 38, 12 (2005), 2270–2285.
- [19] JW Kling and Harold Schlosberg. 1961. The Uniqueness of Patterns of Skin-Conductance. *The American journal of psychology* 74, 1 (1961), 74–79.
- [20] Chuanhe Liu, Tianhao Tang, Kui Lv, and Minghao Wang. 2018. Multi-Feature Based Emotion Recognition for Video Clips. In *Proceedings of the 2018 on International Conference on Multimodal Interaction*. ACM, 630–634.
- [21] Nick Martin and Hermine Maes. 1979. *Multivariate analysis*. Academic press London.
- [22] Gary McKeown, Michel F Valstar, Roderick Cowie, and Maja Pantic. 2010. The SEMAINE corpus of emotionally coloured character interactions. In *2010 IEEE International Conference on Multimedia and Expo*. IEEE, 1079–1084.
- [23] Prateek Panwar and Christopher M Collins. 2018. Detecting negative emotion for mixed initiative visual analytics. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, LBW004.
- [24] Rosalind W. Picard, Elias Vyzas, and Jennifer Healey. 2001. Toward machine emotional intelligence: Analysis of affective physiological state. *IEEE transactions on pattern analysis and machine intelligence* 23, 10 (2001), 1175–1191.
- [25] Lin Shu, Jinyan Xie, Mingyue Yang, Ziyi Li, Zhenqi Li, Dan Liao, Xiangmin Xu, and Xinyi Yang. 2018. A Review of Emotion Recognition Using Physiological Signals. *Sensors* 18, 7 (2018), 2074.
- [26] Mohammad Soleymani, Jeroen Lichtenauer, Thierry Pun, and Maja Pantic. 2012. A multimodal database for affect recognition and implicit tagging. *IEEE Transactions on Affective Computing* 3, 1 (2012), 42–55.
- [27] Tengfei Song, Wenming Zheng, Peng Song, and Zhen Cui. 2018. EEG Emotion Recognition Using Dynamical Graph Convolutional Neural Networks. *IEEE Transactions on Affective Computing* (2018), 1–1. <https://doi.org/10.1109/TAFFC.2018.2817622>
- [28] Goran Udovičić, Jurica Đerek, Mladen Russo, and Marjan Sikora. 2017. Wearable emotion recognition system based on GSR and PPG signals. In *Proceedings of the 2nd International Workshop on Multimedia for Personal Health and Health Care*. ACM, 53–59.
- [29] Gyanendra K Verma and Uma Shanker Tiwary. 2014. Multimodal fusion framework: A multiresolution approach for emotion classification and recognition from physiological signals. *NeuroImage* 102 (2014), 162–172.
- [30] Chin-An Wang and Douglas P Munoz. 2015. A circuit for pupil orienting responses: implications for cognitive modulation of pupil size. *Current opinion in neurobiology* 33 (2015), 134–140.
- [31] Wei-Long Zheng, Bo-Nan Dong, and Bao-Liang Lu. 2014. Multimodal emotion recognition using EEG and eye tracking data. In *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, Chicago, IL, 5040–5043. <https://doi.org/10.1109/EMBC.2014.6944757>
- [32] Wan-Hui Wen, Guang-Yuan Liu, Nan-Pu Cheng, Jie Wei, Peng-Chao Shangguan, and Wen-Jin Huang. 2014. Emotion recognition based on multi-variant correlation of physiological signals. *IEEE Transactions on Affective Computing* 1 (2014), 1–1.
- [33] A Tianyi Zhang and B Olivier Le Meur. 2018. How Old Do You Look? Inferring Your Age From Your Gaze. In *2018 25th IEEE International Conference on Image Processing (ICIP)*. IEEE, 2660–2664.
- [34] Yu-Dong Zhang, Zhang-Jing Yang, Hui-Min Lu, Xing-Xing Zhou, Preetha Phillips, Qing-Ming Liu, and Shui-Hua Wang. 2016. Facial emotion recognition based on biorthogonal wavelet entropy, fuzzy support vector machine, and stratified cross validation. *IEEE Access* 4 (2016), 8375–8385.