Towards Objective Assessment of Movie Trailer Quality Using Human Electroencephalogram and Facial Recognition

Qing Wu*, Wenbing Zhao*, Tessadori Jacopo†
*Department of Electrical Engineering and Computer Science, Cleveland State University, Cleveland, OH 44115, USA
† Department of Advanced Robotics, Istituto Italiano di Tecnologia, GE, 16163, Italy
Email: q.wu55@vikes.csuohio.edu, wenbing@ieee.org, jacopo.tessadori@iit.it

***Abstract*—In this paper, we propose a novel framework to objectively evaluate the quality of movie trailers by fusing two sensing modalities: (1) Human Electroencephalogram (EEG), and (2) computer-vision based facial expression recognition. The EEG sensing data are acquired via a cap instrumented with a set of 4-channel EEG sensors from the OpenBCI Ganglion board. The facial expressions are captured while a user is watching a movie trailer using a regular webcam to help establish the context for EEG analysis. On their own, facial expressions reveal how engaged a user is while watching a movie trailer. Additionally, facial expression data help us identify situations where noises caused by muscle movement in EEG data. Using a shallow neural network, we classify facial expressions into two categories: positive and negative emotions. A quarter-central decision making strategy model is used to analyze EEG signals with a low pass filter activated by time stamp when large human movements are detected. A small human subject test showed that the adaptive analysis method can achieve higher accuracy than that obtained via EEG alone. Besides for movie trailer evaluation, this framework can be utilized in the future towards remote training evaluation, wearable device personalization, and assisting paralyzed people to communicate with others.**

Keywords: Facial Recognition; Affective Computing; Wearable Computing; EEG; Sensor Fusion

I. Introduction

Previously we proposed using EEG to objectively evaluate the quality of movie trailers [1]. We believe this line of research is urgently needed. The US movie box office revenue is over $10 billion per year [2]. However, investing in movie production is highly risky. According to [3], only 36% of the movies produced between 2000 and 2010 in the US had a box office revenue higher than the production cost. Usually, movie producers pay a lot of money to make movie trailers [4] and marketing researchers use them to gauge future box office earnings [5]. Among all factors to evaluate a movie trailer, engagement [1] and emotional changes are essential parameters. Especially, EEG has been extensively used in the past 10 years for measuring engagement of people watching video [6].

In this paper, we extend our previous work by incorporating another sensing modality, i.e., computer-vision-based facial express recognition, into the objective assessment framework for movie trailers. Using a conventional webcam with the OpenCV library, we facilitate human pupil recognition and use a shallow neural network to differentiate positive facial expressions from negative expressions. This sensing modality helps establish the context for EEG sensing. Furthermore, it facilitates a consumer device to adapt its interfaces and features based on the emotional state of the user. Current related research includes audio analysis to detect human emotions [7], eye blink [8] for Brain-Machine Interface (BMI), and facial expressions as a communication channel between human and computer-based systems [9].

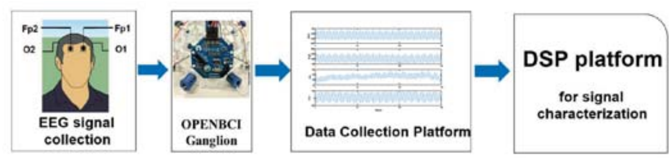

Figure 1. The system layout and the details of the Cleveland State University EEG cap (CEC). CEC is designed to capture four EEG Channels (i.e. FP1, FP2, O1, and O2) by OpenBCI Ganglion hardware.

Facial recognition has been investigated for many years based on image processing and machine learning algorithms [10]–[19]. However, few studies have paid attention to the use of the facial recognition modality to establish context for other types of sensing modalities. Our framework not only offers a metric to evaluate emotional changes of a user, but also differentiates EEG from electromyography (EMG) and other bio-signals. Via a small human subject test, we found that by fusing the two sensing modalities, our framework can achieve higher consistency between the objective assessment result and the self-reported ratings via surveys regarding the quality of movie trailers.

Besides movie trailer evaluation, the proposed framework can be utilized in the future towards distance education evaluation, such as online courses and remote training, wearable device customization, and assisting paralyzed people to communicate with others.

II. Methods

A. Subjects and Test Data Descriptions

This study utilized four EEG channels (i.e. Fp1, Fp2, O1, O2 as shown in Figure 1) and the OpenBCI Ganglion Board. All participants were instructed with a basic outline on how these tests would unfold and on the fact that they could quit

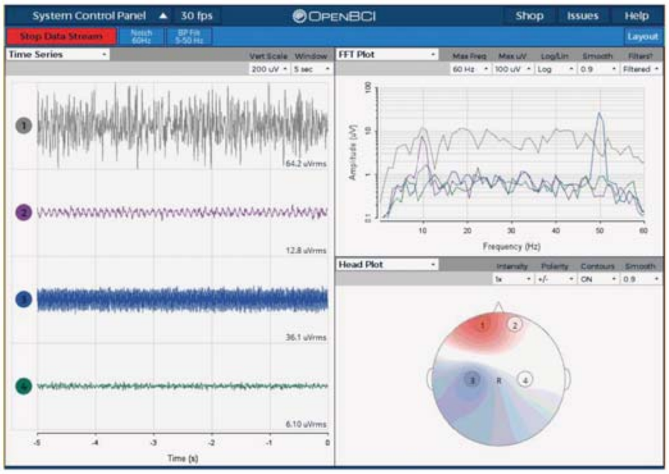

Figure 2. OpenBCI real-time human EEG capture platform in four channels.

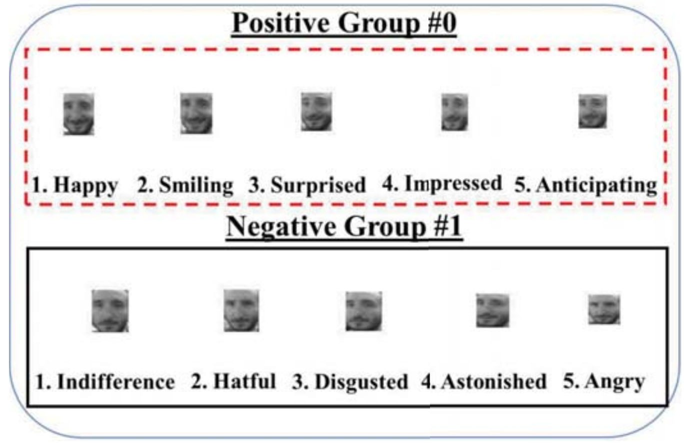

Figure 3. A supervised machine learning entropy degradation method (EDM) is used to classify a participant's facial expression into two categories.

the tests at anytime. Four different movie trailers have been shown to four participants from different countries (three men and one woman). The experiments took place at the IEEE International Conference on Systems, Man, and Cybernetics (SMC) 2017 Brain Hackathon Competition in October.

Two of the selected movie trailers are 60 seconds in length, while the other two movie trailers are around 150 seconds long. These trailers are all publicly available on YouTube, namely "2016 Oscar Nominated Short Films: Animated Trailer HD", "AMAZING MR HUBLOT! Animation test", "Underdogs US Release TRAILER 1(2015)-Bella Thorne, Katie Holmes Animated Movie HD", "Home Official Trailer #1 (2015)-Jennifer Lopez, Rihanna Animated Movie HD", respectively.

The EEG signals of the four participants were recorded in real-time while they were watching different movie trailers. A short survey is included for them to score on these four movie trailers. A set of integer numbers is used by participants to score. Each score is only applied once. In this study, the score range is set from 1 to 4, where 1 presents the movie that the participant would least like to watch in a cinema, while 4 represents the most interesting movie that the participant would pay to watch. These surveys are taken as ground truth for later comparisons.

We used the OpenBCIGUI and OpenBCIHub software to capture the EEG signals of the participants in realtime, as shown in Figure 2. These SDK tools are offered from the github open source platform and the OpenBCI website (http://openbci.com/index.php/downloads). We also implemented EEG signal analysis using MATLAB, such as the figure displayed, filters applied, and normalization. During EEG data recording, any large movements of the participants were logged along with their time stamps.

B. Human Facial Detection

Other than using EEG signals, a webcam can detect the facial expressions of participants and track their eye movements [20]–[22]. Based on the same approach, we can detect whether a participant is paying attention to the movie trailer or not. In the worst case (i.e. if the subject has closed his/her eyes during the whole test), the EEG data recorded would be considered invalid.

We divide this part into two scenarios. One is eye detection, while the other is focus definition. To detect eye close/open, we use OpenCV/JavaCV to control the camera and detect a human face. It can control the webcam for taking a picture and recognize a person's face from a possible confound. The eye detection software can first recognize a human face in a small rectangle area, and it iterates an even smaller rectangle area for human eye segmentation. Test benches are implemented using Java/Eclipse.

C. Machine Learning Based Data Analysis Method

In this paper, we applied the shallow neural network entropy degradation method (EDM) [23] to recognize the facial expressions of the participants. In Figure 3, different pictures were taken to train the library with labels. These pictures were firstly categorized into two parts, namely as Positive Group Number 0 and Negative Group Number 1. In this project, human expressions are tagged as happy, smiling, surprised, impressed, and anticipating for Group 0, and indifference, hateful, disgusted, astonished, and angry for Group 1. Figure 3 indicates how that this algorithm can detect whether an input color image belongs to Group 0 or 1. Each facial expressions changes from Group 1 to Group 0 will be recorded with time stamp.

The adaptive EDM algorithm transforms the vectorized gray level images of each training set into a score. The score is further converted by probabilities through a logistic function, a cost function where the differences between the transformation and labels are calculated and later be fed back by back-propagation stage [24]–[27]. It was designed by a score-probability policy. A color image is fed into a properly trained EDM to be classified. Results of the probabilities calculated for this input indicate which group it belongs to (i.e. more likely to belong to the positive expression group or the negative expression group). Specifically, we define the $ML - QW()$ function, inside which we initialize the points from the outputs of function $ICA - QW()$ [23].

$$a_{i,j} = log(\sum_{j}\sum_{i}^{N}\overrightarrow{point}(p = j) \times \overrightarrow{x^{j}})+ \\ log(\sum \overrightarrow{point}(p = j))/(\sum\sum(\overrightarrow{point})) \tag{1}$$

where j = 0 or 1, (p=j) is an indicator function (i.e. if p and j are the same, the output is 1; otherwise is 0). i ranges from 1 to the size of input x. N is the total size of input x.

In function $ICA-QW()$, we basically apply the Basic Bell-Sejnowski ICA algorithm. We use the estimated maximum entropy signals "$Y = cdf(y)$". The output of the previous step is to calculate its value by function h as below,

$$h = log(|det(W)|) + \frac{1}{N}\sum(log(e + 1 - Y^{2})) \tag{2}$$

where W is initialed as an identity 5x5 matrix, and later be used in the gradient descents as following,

$$W = W + eta \times (g) \tag{3}$$

to an updated W to increase h in the next loop, where eta is a small number defined to control convergence speed. N is the number of input x, and g is the matrix of gradients [23].

In this research, we apply an adaptive quarter-central decision making strategy for EEG signal analysis. The previous algorithm is applied to 4 different EEG-channel signals, Fp1, Fp2, O1 and O2, shown in Figure 1, into dot energies calculated by three high frequency bands, namely α (8-12 Hz), β (12-32 Hz), and γ(32-72 Hz) [1], [28]. The core idea of this algorithm is to roughly estimate how the participants are feeling while watching two different movie trailers. A fuzzy model could be set up by this algorithm [1], since Fp1 and Fp2 represent human decision making center, while O1 and O2 represent visual stimulation which leads to entertainment.

The energy of EEG signals is calculated in the frequency domain using short window FFT in MATLAB R2016b. A normalized energy is summed up by different frequency bands, α, β, and γ respectively. The power spectrum density (PSD) of one channel c with sampling frequency of f_s and frequency band b = [b_1, b_2] is calculated as:

$$P^{(c)}(k) = \sum_{k=0}^{N-1}\left|x_{k}^{(c)}e^{-j2\pi k/f_s}\right|^{2} \tag{4}$$

where x represents the time domain data of channel c with N samples (i.e. the small window length) [1]. P is defined to compute the PSD of channel c signal in as follows [28]:

$$P_{b}^{(c)} = \frac{\sum_{k=b1}^{k=b2}P^{(c)}(k)}{\sum_{k=0}^{k=f_s/2}P^{(c)}(k)} \tag{5}$$

We adapted this algorithm to add a low pass filter to certain frames when large human movements without facial expression changes are detected using time stamps. We also treat facial expression group changes as certain incidents to validate our EEG analysis results.

Table I
Predicted by our algorithm

Movie Number and Person Number	P1	P2	P3	P4
M2 > M1	Y	Y	Y	N
M4 > M3	Y	N	Y	N

Table II
Ground Truth by Census

Movie Number and Person Number	P1	P2	P3	P4
M2 > M1	Y	Y	Y	N
M4 > M3	Y	Y	Y	N

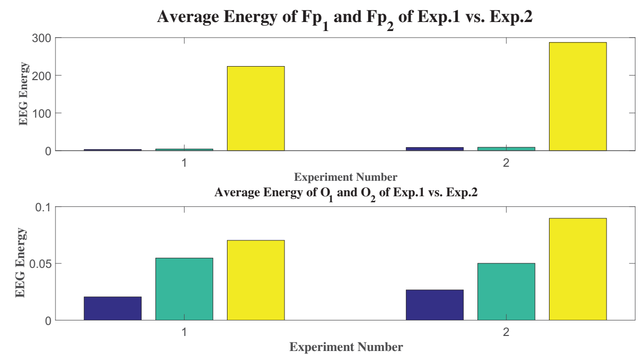

Figure 4. Total energy in high frequency bands of EEG for P3 in movie trailer 1 vs. movie trailer 2. This result indicates the subject is enjoying movie trailer 2 better than movie trailer 1.

III. Results and Discussions

As shown in Figure 2, four-channel EEG signals are displayed by OpenBCI in real time. In the time period, each spike represents a certain incident discovered by the subject. For example, when a fat cat first shows in "Home Official Trailer", three out of four participants had facial expression changes from negative to positive. This change also relates to an EEG spike in the same time stamp, since a spike can be either characterized by amplitude or frequency changes. Usually, higher EEG amplitudes and more dynamics indicate that the subject paid more attention [1].

Figure 4 shows that the total energy in high frequency bands of EEG for P3 in movie trailer 2 are higher than movie trailer 2. It indicates that the participant (P3) is enjoying movie trailer 2 better than movie trailer 1 (similar results shown in Table II). Based on the EEG data of the four participants and by comparing the two sets of similar movie trailers, our classification made 1 mistake (accuracy around 87.5%) as shown in Table II compared to the ground truth table as shown in Table II. As a comparison, using an EEG alone without the facial recognition modality resulted in only 62.5% accuracy.

IV. Conclusions and Future Work

In this paper, we reported a framework that combines two sensing modalities for movie trailer evaluation. The EEG sensing was done with a cap instrumented with 4-channel EEG sensors and the OpenBCI Ganglion signal capture hardware. The computer vision sensing modality is done via a conventional webcam with the OpenCV image processing library. We conducted tests with four participants to validate the feasibility of our approach to fuse the EEG and computer vision modalities for movie trailer evaluation. By differentiating EEG from EMG and other bio-signals, and by recognizing concentration/engagement level for participants with a redundant modality, the proposed framework achieved better accuracy than our previous work. Thus, the new framework can offer better service capabilities for consumer devices, especially wearable devices, to automatically adapt its interfaces and features based on the emotional state and usage patterns of individual users. In the future, besides for movie trailer evaluation, the proposed framework can also be applied to evaluate remote education, and to assist communications with paralyzed people [29]. We plan to add more computational functions into this framework for different applications [30].

Acknowledgment

The authors thank AJ Keller and Conor Russomanno from OpenBCI for their valuable input and support. This study is supported in part by a faculty research development (FRD) award and a FRD-Internet of Things award from Cleveland State University.

References

[1] Q. Wu, W. Zhao, H. Sharma, and T. Qiu, "Movie trailer quality evaluation using real-time human electroencephalogram," in *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2017, pp. 3553–3557.

[2] T. numbers, "Domestic theatrical market summary for 2016," http://www.the-numbers.com/market/2016/summary, 2016.

[3] M. T. Lash and K. Zhao, "Early predictions of movie success: The who, what, and when of profitability," *Journal of Management Information Systems*, vol. 33, no. 3, pp. 874–903, 2016.

[4] K. M. Johnston, E. Vollans, and F. L. Greene, "Watching the trailer: Researching the film trailer audience," *Participations*, vol. 13, no. 2, pp. 56–85, 2016.

[5] Y. Tonga and Y. Songb, "Large data based research on the editing of film trailer," *Sociology*, vol. 5, no. 1, pp. 23–28, 2015.

[6] C. Herba and M. Phillips, "Annotation: Development of facial expression recognition from childhood to adolescence: Behavioural and neurological perspectives," *Journal of Child Psychology and Psychiatry*, vol. 45, no. 7, pp. 1185–1198, 2004.

[7] A. A. Zhilenkov and D. Denk, "Based on mems sensors man-machine interface for mechatronic objects control," in *Young Researchers in Electrical and Electronic Engineering (EIConRus), 2017 IEEE Conference of Russian*. IEEE, 2017, pp. 1100–1103.

[8] J. Ma, Y. Zhang, A. Cichocki, and F. Matsuno, "A novel eog/eeg hybrid human–machine interface adopting eye movements and erps: Application to robot control," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 3, pp. 876–889, 2015.

[9] M. S. Bartlett, G. Littlewort, I. Fasel, and J. R. Movellan, "Real time face detection and facial expression recognition: Development and applications to human computer interaction." in *Computer Vision and Pattern Recognition Workshop, 2003. CVPRW'03. Conference on*, vol. 5. IEEE, 2003, pp. 53–53.

[10] Y. Luo, C.-m. Wu, and Y. Zhang, "Facial expression recognition based on fusion feature of pca and lbp with svm," *Optik-International Journal for Light and Electron Optics*, vol. 124, no. 17, pp. 2767–2770, 2013.

[11] A. L. Benton, "The neuropsychology of facial recognition." *American Psychologist*, vol. 35, no. 2, p. 176, 1980.

[12] A. Yates, "Intersection of longest paths in graph theory and predicting performance in facial recognition," 2017.

[13] A. Benton and M. Van Allen, "Impairment in facial recognition in patients with cerebral disease," *Cortex*, vol. 4, no. 4, pp. 344–IN1, 1968.

[14] D. Tranel and A. R. Damasio, "Knowledge without awareness: an autonomic index of facial recognition by prosopagnosics," *Science*, vol. 228, pp. 1453–1455, 1985.

[15] E. K. Warrington and M. James, "An experimental investigation of facial recognition in patients with unilateral cerebral lesions," *Cortex*, vol. 3, no. 3, pp. 317–326, 1967.

[16] M. A. O. Vasilescu and D. Terzopoulos, "Multilinear image analysis for facial recognition," in *Pattern Recognition, 2002. Proceedings. 16th International Conference on*, vol. 2. IEEE, 2002, pp. 511–514.

[17] G. Sachs, D. Steger-Wuchse, I. Kryspin-Exner, R. C. Gur, and H. Katschnig, "Facial recognition deficits and cognition in schizophrenia," *Schizophrenia research*, vol. 68, no. 1, pp. 27–35, 2004.

[18] S. J. Simonian, D. C. Beidel, S. M. Turner, J. L. Berkes, and J. H. Long, "Recognition of facial affect by children and adolescents diagnosed with social phobia," *Child Psychiatry & Human Development*, vol. 32, no. 2, pp. 137–145, 2001.

[19] F. Philipp-Wiegmann, M. Rösler, P. Retz-Junginger, and W. Retz, "Emotional facial recognition in proactive and reactive violent offenders," *European Archives of Psychiatry and Clinical Neuroscience*, pp. 1–9, 2017.

[20] S. Meimon, J. Jarosz, C. Petit, E. G. Salas, K. Grieve, J.-M. Conan, B. Emica, M. Paques, and K. Irsch, "Pupil motion analysis and tracking in ophthalmic systems equipped with wavefront sensing technology," *Applied Optics*, vol. 56, no. 9, pp. D66–D71, 2017.

[21] R. Andersson, L. Larsson, K. Holmqvist, M. Stridh, and M. Nyström, "One algorithm to rule them all? an evaluation and discussion of ten eye movement event-detection algorithms," *Behavior research methods*, vol. 49, no. 2, pp. 616–637, 2017.

[22] T. Endo, N. Ukita, H. Tanaka, N. Hagita, S. Nakamura, H. Adachi, M. Ikeda, H. Kazui, and T. Kudo, "Initial response time measurement in eye movement for dementia screening test," in *Machine Vision Applications (MVA), 2017 Fifteenth IAPR International Conference on*. IEEE, 2017, pp. 262–265.

[23] Q. Wu and W. Zhao, "Small-cell lung cancer detection using a supervised machine learning algorithm," in *Proceedings of the International Symposium on Computer Science and Intelligent Controls*. IEEE, 2017, pp. 88–91.

[24] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[25] G. Wang, "A perspective on deep imaging," *IEEE Access*, vol. 4, pp. 8914–8924, 2016.

[26] X. Luo, Y. Xu, W. Wang, M. Yuan, X. Ban, Y. Zhu, and W. Zhao, "Towards enhancing stacked extreme learning machine with sparse autoencoder by correntropy," *Journal of the Franklin Institute*, 2017.

[27] A. Vedaldi and K. Lenc, "Matconvnet: Convolutional neural networks for matlab," in *Proceedings of the 23rd ACM international conference on Multimedia*. ACM, 2015, pp. 689–692.

[28] J. Birjandtalab, M. B. Pouyan, and M. Nourani, "Nonlinear dimension reduction for eeg-based epileptic seizure detection," in *Proceedings of the IEEE-EMBS International Conference on Biomedical and Health Informatics*. IEEE, 2016, pp. 595–598.

[29] W. Zhao, X. Luo, and T. Qiu, "Smart healthcare," *Applied Sciences*, vol. 7, no. 11, p. 1176, 2017.

[30] X. Liu, Q. Wu, W. Zhao, and X. Luo, "Technology-facilitated diagnosis and treatment of individuals with autism spectrum disorder: An engineering perspective," *Applied Sciences*, vol. 7, no. 10, p. 1051, 2017.