IOSR Journal of Electronics and Communication Engineering (IOSR-JECE)
e-ISSN: 2278-2834,p- ISSN: 2278-8735.
PP 70-75
www.iosrjournals.org

Human Disposition Detection using EEG signal and Facial Expression: A Survey

Anuja R. Bhagwat[1], A. N. Paithane [2]
[1,](UG student E&TC dept., Rajarshi Shahu College Of Engineering, Pune, India.)
2(Faculty of E&TC dept., Rajarshi Shahu College Of Engineering, Pune, India)

ABSTRACT : *Multimedia content is made to induce emotions and be emotionally expressive, the need and importance of automatic emotion recognition has grown with increasing role of human computer interface applications. Recent advances on affect detection are focused on detecting emotions continuously. In this paper, for the first time, we continuously detect valence from electroencephalogram (EEG) signals and facial expressions in response to human expression. Power spectral features from EEG signals as well as facial fiducial points are used as features to detect valence levels for each frame continuously. We study the correlation between features from EEG and facial expressions with continuous valence. We have also verified our model's performance for the emotional highlight detection using emotion recognition from EEG signals. Finally the results of multimodal fusion between facial expression and EEG signals are presented. Having such models we will be able to detect spontaneous and subtle affective responses over time. So basically we trying to built an automatic emotion monitoring system. Due this we can easily treat physiological patient. This will best contribution for physiological field.*

Keywords *- Emotion Recognition, EEG signal, Facial Expression, Long Short Term Memory Neural Networks.*

I. INTRODUCTION

Emotions are the reactions or perceptions that a person has of specific situation. Affective features of multimedia are therefore an invaluable source of information for multimedia indexing and recommendation [1]. Given the difficulty of collecting emotional self-report to multimedia from users, emotion recognition is an effective way of collecting users' emotional feedback in response to multimedia for the purpose of multimedia indexing [2]. Emotion recognition could be done from the text, speech, facial expression or gesture. In this paper, we concentrate on recognition of "inner" emotions from electroencephalogram (EEG) signals and to maintain the robustness of the system by using Facial expression.

To study human emotional experience and expression in more detail and on a scientific level, and to develop and benchmark methods for automatic recognition, researchers are in need of rich sets of data of repeatable experiments. The richness of the human emotional expressiveness poses both a technological as well as a research challenge [3].

II. Existing System

Researchers have proposed a number of methods and several approaches have been proposed previously as given below.

Yisi Liu, Olga Sourina, and Minh Khoa Nguyen, They proposed paper 'Real-time EEG-based Emotion Recognition and its Applications'. In this paper, they concentrate on recognition of "inner" emotions from electroencephalogram (EEG) signals. They propose real-time fractal dimension based algorithm of quantification of basic emotions using Arousal-Valence emotion model. Two emotion induction experiments with music stimuli and sound stimuli from International Affective Digitized Sounds (IADS) database were proposed and implemented. Finally, the real-time algorithm was proposed, implemented and tested to recognize six emotions such as fear, frustrated, sad, happy, pleasant and satisfied. Real-time applications were proposed and implemented in 3D virtual environments [4]. The user emotions are recognized and visualized in real time on his/her avatar adding one more so-called "emotion dimension" to human computer interfaces. An EEG-enabled music therapy site was proposed and implemented. The music played to the patients helps them deal with problems such as pain and depression. An EEG-based wehb-enable music player which can display the music according to the user's current emotion states was designed and implemented.

Disadvantages of this system are not an integration of their tools in Co-Spaces on the Web targeting entertainment industry.

Mohammad Soleymani, Guillaume Chanel, Joep J. M. Kierkels And Thierry Pun they are proposed 'Affective Characterization Of Movie Scenes Based On Content Analysis And Physiological Changes'. In this paper, an affective characterization method for movie scenes is proposed based on emotions that are felt by spectators. Physiological responses of participants were recorded while watching movie scenes and key features were extracted from these responses. By computing correlations between these key physiological features and the users' self-assessment of arousal and valence, it was identified which physiological features are essential for accurate determination of valence-arousal. Such accurate determinations provide us with a continuous assessment of affect which can serve as a ground truth for affect determination. For example Zygomaticus EMG signals which represent smile and laughter have high correlation with valence. Furthermore, content based multimedia features were extracted from the movies scenes. Their correlations with both physiological features and users' self-assessment of valence-arousal were shown to be significant [5]. A procedure was proposed to actually estimate user's affect in response to movie scenes based on selected multimedia content features. Predicting user's affect opens the door to many novel applications. One is personalized content delivery systems with configurable emotional-based preferences. Users will watch a training set of short movie clips; after configuration, the system will be able to predict the users' response to new movie scenes either from physiological signals or multimedia content. A similar strategy is applicable to neuromarketing where consumers' reactions to marketing stimuli could be predicted.

The movie scenes did not necessarily correspond to very strong emotions; some of them contained just mild and tranquil scenes. These were intentionally selected because the final application was not only to characterize affect, but also to show the ability to estimate different amplitudes of emotions. The final application will have to index all types of different movie scenes from highly intense ones to calm and fairly neutral [6].

Disadvantages of this Felt emotions from the movie scenes where determined without any a priori assumptions on valence-arousal values. It would however be possible to use the genre of movies (e.g., drama, comedy, etc.) or the temporal sequences of the emotional events as prior knowledge for better affect determination [7].

Chung-Yeon Lee and Seongah Chin, they proposed work on 'Facial Expression Mirroring-based Classification of Emotions is using Electroencephalogram Signals'. In this paper, they present a mechanism to do analysis of electroencephalogram signals originated from emotional impulses and to carry out classification of the emotion. Also they have validated the methods to foresee usability for brain-computer interface. The partial derivatives of EEG are taken as features for some training data set gained from facial expression mirroring. Four emotions including neutral, anger, happiness, and surprise have been classified using the support vector machines [15, 16]. The experimental results can be extended in providing innovative potential in the areas including games, virtual reality, agents, and various entertainments.

Disadvantage of this brain computer interface cannot be used in providing intelligent services for users. For instance, if we understand genuine emotions of customers for entertainment content services, more customized taking into account the current emotions intelligence services for the users can be utilized.

We presented a complete framework for continuous prediction of human emotions based on features characterizing head movements, face appearance and voice in a dynamic manner by using log-magnitude Fourier spectra. We introduced a new correlation-based measure for feature selection and evaluated its efficiency and robustness in the case of possibly time-delayed labels. We proposed a fast regression framework based on a supervised clustering followed by a Nadaraya-Watson kernel regression that appears to outperform, for the aimed task, Support Vector Regression. Our fusion method is based on simple local linear regressions and significantly improves our results. Because of the high power of generalization of our method, we directly learned our fusion parameters using our regressors outputs on the training set [9]. In order to improve the fusion for methods that are morre prone to over-fitting, we would have to learn these parameters in cross-validation. Our system has been designed for the Audio/Visual Emotion Challenge (AVEC'12) which uses Pearson's correlation as evaluation measure. Therefore, every step of our method has been built and optimized to maximize this measure. The SEMAINE database on which our system has been learned and tested contains videos of natural interactions but recorded in a very constraint environment. A perspective for adapting these kinds of human emotion prediction systems to real conditions, as for assistance robotics, would be to learn the system on "in the wild" data [15,16,17].

Disadvantage of this system an accurate system for everyday interactions would need to be efficient in terms of correlation but also in terms of Root-Mean-Square Error (RMSE). Some modifications on their system would be needed to increase its performance regarding this measure.

III. Proposed System

We performed statistical analyses to identify the relationship between EEG signals and facial expressions. We were particularly interested to identify how much of the emotion detection using EEG signals can be attributed to the electromyogenic artifacts caused by facial expressions. The contributions presented in this paper are as follows. First, to the best of our knowledge, this is the first attempt in detecting continuous emotions, in both time and dimension, using EEG signals. Second, we detect continuous emotions from facial expressions and provide the multimodal fusion results. Third, we study the correlation between the EEG power spectral features that we used for emotion recognition and continuous valence annotations to look for the possible effect of muscular artifacts. Finally, we apply the models trained with the continuously annotated data on EEG responses that could not be interpreted due to the lack of facial expressions from the users. In this work, the emotional responses visible in the frontal camera capturing facial expressions are annotated continuously in time and valence and arousal dimension by five annotators [11].

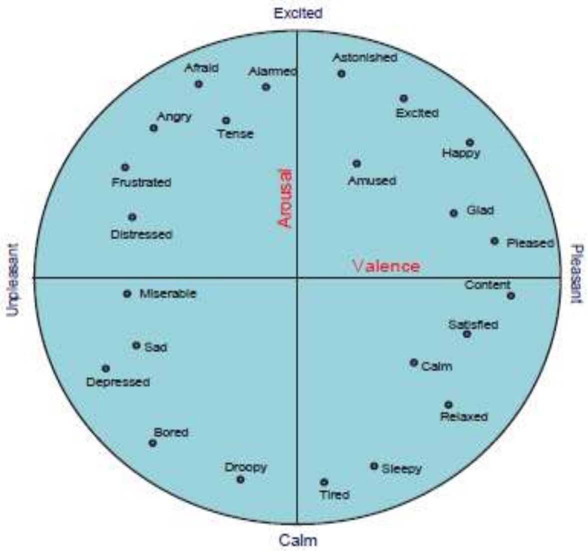

Fig.1. Arousal/Valence space description of emotions

The averaged annotations served as a ground truth to be detected from facial expression analysis and EEG signals. Different regression models, utilized in the similar state of the art studies, were tested, and the performance of continuous emotion detection was evaluated using a 10-folding cross validation strategy.

Methodology Used

A. EEG signals

The EEG is an electrical waveform that is recorded from the brain by using electrodes appropriately placed on the head, then amplifying and displaying the electrical signal using a computer, or other suitable instrument. EEG signals were available at 256Hz sampling rate. EEG signals were re-referenced to the average reference to enhance the signal-to-noise ratio. Average re-referencing is performed when there is no reference electrode by subtracting the average amplitude of EEG signals from all the electrodes from every EEG signal recorded from any electrode. This averaged signal includes noise and artifacts that can be detected on the scalp but are not originated from the brain, e.g., electric potentials from cardiac activities. The power spectral densities of EEG signals in different bands are correlated with emotions. The power spectral densities were extracted from 1 second time windows with 50% overlapping. We used all the 32 electrodes for EEG feature extraction. The logarithms of the PSD from theta (4Hz $< f <$ 8Hz), alpha (8Hz $< f <$ 12Hz), beta (12Hz $< f <$ 30Hz) and gamma (30Hz $< f$) bands were extracted to serve as features. In total number of EEG features of a trial for 32 electrodes and 4 bands is $32 \times 4 = 128$ features.

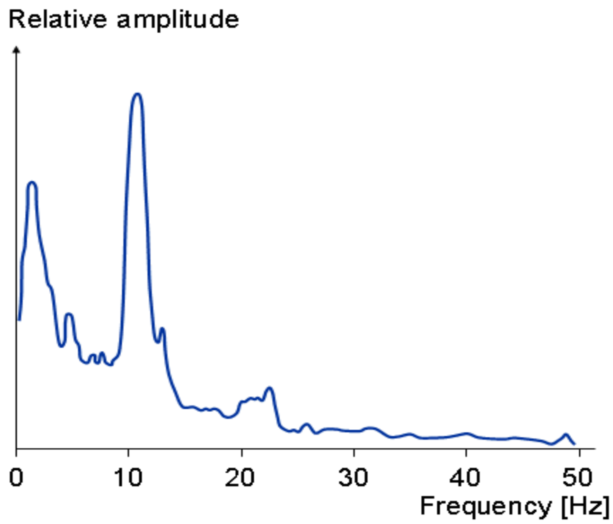

Fig.2. Frequency spectrum of normal EEG.

B. Analysis of facial expressions

Face detection is based on identifying and locating a human face in the image, regardless of position, size, and condition.

Its applications are widely used in artificial intelligence, surveillance video, identity authentication and human machine interaction. An active appearance model face tracker was employed to track 40 points [12] (see Figure 3). The facial points were extracted after registering the face to a normalized face and correcting the head pose. A reference point was generated by averaging the inner corners of eyes and points on the subjects' nose which assumed to be stationary. The distances of 33 point including eyebrows, eyes, lips and iris to the reference point were calculated and averaged to be used as features.

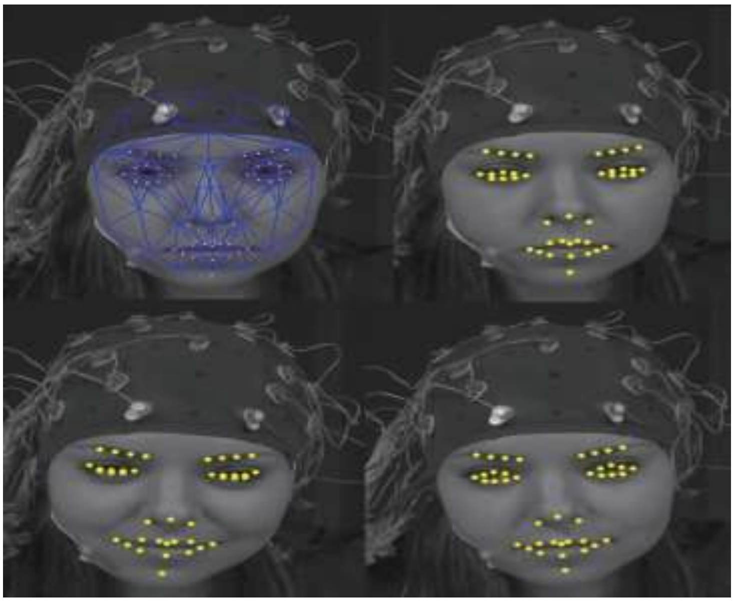

Fig. 3. Examples of the recorded camera view including tracked facial points. The top left image shows the active appearance model that is fit to the face.

C. Dimensional affect detection

Four commonly used regression models for similar studies were utilized for continuous emotion detection, namely, multi-linear regression (MLR), support vector regression (SVR), conditional random fields (CCRF) and long short-term memory recurrent neural networks (LSTM-RNN).

1. Long Short Term Memory Neural Networks
LSTM-RNN has shown to achieve top performances in emotion recognition studies for audio-visual modalities [8, 6]. LSTM-RNN is a network which has one or more hidden layers including LSTM cells. These cells contain a memory block and some multiplicative gates which will determine whether the cell stores, maintains or resets its state. In this way, the network learns when to remember and forget information coming in a sequence over time and therefore it is able to preserve long-range dependencies in sequences. Recurrent Neural Networks are able to remember the short term input events through their feedback connections. LSTM adds the ability to also remember the input events from a longer period using the gated memory cell.

2. Continuous Conditional Random Fields
Conditional random fields (CRF) are frameworks for building probabilistic models to segment and classify sequential data. Unlike hidden Markov models (HMM), they do not assume that the observations are conditionally independent and therefore are good alternatives for cases where there is a strong dependency between observations. Continuous conditional random fields (CCRF) [14] are developed to extend the CRFs for regression.

IV. Artifacts And Their Effect

There is often a strong interference of facial muscular activities and eye movements in the EEG signals. The facial muscular artifacts and eye movements are usually more present in the peripheral electrodes and higher frequencies (beta and gamma bands). We, hence, expected that the contamination from the facial expressions in the EEG signals to contribute to the effectiveness of the EEG signals for valence detection. In this Section, the EEG features were extracted from all the 32 electrodes (128 features from 4 bands). To study this assumption, we used a linear mixed effect model to test the effect of EEG features on estimating valence (annotated from the expressions) given the information from eye gaze and facial expressions. Linear mixed-effect model enables us to model the between participant variations in a random effect term while studying the effect of the independent variables (EEG, face and eye gaze) on the dependent variable, valence. The eye movements were taken from the data recorded by the Tobii eye gaze tracker at 60Hz which we resampled to 4Hz to match the other modalities. The facial point movements were defined by how much they moved from one sample to the next. We calculated this in our feature set.

V. Conclusion

We presented a study of continuous detection of valence and arousal using EEG signals and facial expressions. Promising results are obtained from EEG signals. We expects the results from facial expressions to be superior due to the bias of the ground truth towards the expressions, i.e., the ground truth was generated based on the judgment of the facial expressions. However, the results from LSTM-RNN showed that EEG modality performance is not far inferior to the one of facial expressions. The analyses of the correlation between the EEG signals and the ground truth showed that the higher frequency components of the signals carry more important information regarding the pleasantness of emotion and the informative features from EEG signals are not completely due to the contamination from facial muscular activities.

References

[1] M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic, “A multimodal database for affect recognition and implicit tagging,” IEEE Trans. Affective Computing, vol. 3, pp. 42–55, 2012.
[2] H. Joho, J. Staiano, N. Sebe, and J. Jose, “Looking at the viewer: analysing facial activity to detect personal highlights of multimedia contents,” Multimed. Tools. Appl., vol. 51, no. 2, pp. 505–523, 2010.
[3] F. Silveira, B. Eriksson, A. Sheth, and A. Sheppard, “Predicting audience responses to movie content from electro-dermal activity signals,” in ACM UbiComp’13, 2013, pp. 707–716.
[4] K. R. Scherer, “What are emotions? And how can they be measured?,” Social Science Information, vol. 44, no. 4, pp. 695–729, 2005.
[5] J. A. Russell and A. Mehrabian, “Evidence for a three factor theory of emotions,” J. Research in Personality, vol. 11, no. 3, pp. 273–294, 1977.
[6] H. Gunes and B. Schuller, “Categorical and dimensional affect analysis in continuous input: Current trends and future directions,” Image and Vision Computing, vol. 31, no. 2, pp. 120–136, 2013.

[7] M. W¨ollmer, F. Eyben, S. Reiter, B. Schuller, C. Cox, E. Douglas-Cowie, and R. Cowie, "Abandoning emotion classes-towards continuous emotion recognition with modeling of long-range dependencies.," in INTERSPEECH, 2008, pp. 597–600.

[8] B. Schuller, M. Valster, F. Eyben, R. Cowie, and M. Pantic, "AVEC 2012: the continuous audio/visual emotion challenge," in ACM ICMI, 2012, pp. 449–456.

[9] M. Soleymani, G. Chanel, J. J. M. Kierkels, and T. Pun, "Affective Characterization of Movie Scenes Based on Content Analysis and Physiological Changes," Int'l J. Semantic Computing, vol. 3, no. 2, pp. 235–254, June 2009.

[10] M. Soleymani, S. Koelstra, I. Patras, and T. Pun, "Continuous emotion detection in response to music videos," in IEEE Int' Conf. Automatic Face Gesture Recognition (FG), march 2011, pp. 803–808.

[11] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Y. Patras, "DEAP: A database for emotion analysis using physiological signals," IEEE Trans. Affective Computing, vol. 3, pp. 18–31, 2012.

[12] M. Soleymani, M. Larson, T. Pun, and A. Hanjalic, "Corpus development for affective video indexing," IEEE Trans. Multimedia, 2014, in press.

[13] S. Koelstra and I. Patras, "Fusion of facial expressions and eeg for implicit affective tagging," Image and Vision Computing, vol. 31, no. 2, pp. 167–174, 2013.

[14] J. Nicolle, V. Rapp, K. Bailly, L. Prevost, and M. Chetouani, "Robust continuous prediction of human emotions using multiscale dynamic cues," in ACM ICMI, 2012, pp. 501–508.

[15] Paithane, A. N., and D. S. Bormane. "Analysis of nonlinear and non-stationary signal to extract the features using Hilbert Huang transform." *Computational Intelligence and Computing Research (ICCIC), 2014 IEEE International Conference on*. IEEE, 2014.

[16] Paithane, A. N., and D. S. Bormane. "Electrocardiogram signal analysis using empirical mode decomposition and Hilbert spectrum." *Pervasive Computing (ICPC), 2015 International Conference on*. IEEE, 2015.

[17] Paithane, A. N., D. S. Bormane, and Sneha Dinde. "Human Emotion Recognition using Electrocardiogram Signals." International Journal on Recent and Innovation Trends in Computing and Communication,ISSN: 2321-8169,Volume: 2 Issue: 2,194-197.