Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 8 Issue 10, October-2019

Emotion Recognition based on Multimodel: Physical - Bio Signals and Video Signal

Vo Thi Huong, Nguyen Thi Khanh Hong, Le Huu Duy
The University of Danang - University of Technology and Education,
Vietnam

Abstract- **Today, detecting and classifying emotions has become an important item of research and life. Emotion detection and classification are becoming more detailed and accurate thanks to development of various fields such as electronics, sensors or computer engineering. Emotion recognition methods are studied using different data collection methods and one of the most popular and effective methods is the use of physical – bio sensors. Physical – bio sensors -based approaches can provide accurate, sustainable biological information with external influences and interferences. Especially when compared with other approaches such as image processing, video processing. In this paper, we describe a method of classifying and assessing emotions based on a combination of signals collected from physical – bio sensors, video collection and machine learning methods. Specifically, we will describe the platform of a physical – bio signal collection system, the process of collecting information and the processing information system used to identify how emotional behavior is characterized. We have also shown that a combination of physical – bio sensors acquisition systems, video collection and machine learning methods can provide identification performance with an accuracy of 83.2%.**

Key words- Emotion Recognition; physical – bio sensor; machine learning

I. INTRODUCTION

The problem of classifying, evaluating and recognizing human emotions is currently taken an attention in research. Emotion recognition problem is described as an identification problem in which the input information is physical – bio signals such as heart rate, oxygen concentration, blood pressure, or objective comments collected from images, cameras, or reviews from other factors such as doctors and specialists [1]. The result after processing the collected information is conclusions about emotional state such as happiness, sadness, anger. And the problem of identifying and classifying emotions is a complex problem because of many different affected factors [2]. In addition, the assessment of emotional recognition is mostly based on logical thinking from the input information, but emotions work according to the illogical element [3]. For example, a person who has extremely good emotions, is extremely comfortable, still cannot describe the definition of good emotions consistently with others. So the emotional recognition project is still a challenging problem.

Despite many difficulties and challenges, the problem of classification and emotional identification is focused on research for following reasons:

- By understanding human emotions, current information systems can increase the level of interaction with people, creating entirely new user experiences. For example, the sound system can reduce music, reduce volume during periods of extreme stress or fear, or suggest suitable movies according to the mood of the viewer. Computers and entertainment systems can also identify users' reactions like interest or annoyance when an entertainment content is randomly shown. So entertainment and information systems seem to become more companions rather than tools to serve people every day[4].
- Emotion evaluation and recognition can also be used as a means of physical - biological monitoring to self-assessment and monitoring of human emotional states. The results of physical - biological monitoring have many benefits such as a method to improve communication, or to evaluate emotional behaviors that can have negative effects on people and society. From there can give appropriate responses and methods to limit, or improve it [5].

As mentioned previous part, there are many methods used to collect emotional signal information such as physical – bio signals (heart rate, blood pressure, etc.) or factors such as facial expressions and rhythm words, gestures... or objective reviews from experts and doctors [6]. Each method of collecting emotional signal information has different advantages and disadvantages. In this paper, we will focus on what is current research. It is a method that uses physical – bio sensors such as Electromyography (EMG); Electrocardiogram (ECG) ; Electrodermal activity (EDA) to collect biomedical information. The method of using biological sensors has many outstanding advantages compared to other methods. For example, measuring devices are becoming more and more compact as the built-in sensors are becoming more and more compact. Therefore, the device that collects signals using a physical – bio sensor can be as compact as wearable devices, even jewelry, giving users a sense of privacy and individuality over methods. In other segments, when users are subject to "supervision" by cameras or other audio and video recording devices [7] [8] .

Our goal is to collect the physical – bio signals of a person under different conditions of real life to detect emotions automatically. We propose a method for a multimodal detection of emotions using physical – bio signals. The paper is structured as follows. In section 2, a brief state of the art on the multimodal recognition of emotions and different methods to merge signals are shown. In section 3, all the steps of the proposed methods are explained in details; later on in section 4, a comparison between obtained results of related work and ours proposed modal; finally, conclusion and future work are reported in section 5.

II. RELATED WORKS

In general, an emotion recognition system are based on three fundamental steps:

- **Acquisition of the signals,**
- **Features Extraction**
- **The detection of emotions.**

Some research has focused on the detection of emotions using facial expressions, vocal expressions or physiological signals [9], [10], [11]; however fewer studies are focused on the multimodal recognition [12] of emotions.

In addition, multimodal approach presents not only to enhance the recognition rate but also more strength to the system when unimodalities are acquired in a noisy environment and less accuracy [13].

In theory, there are three methods [14] to merge the signals from various sensors: fusion at the signal level (fusion of physiological signals), feature level fusion (fusion of features) and decision level fusion (fusion of decisions) [15], [16].

III. METHODOLOGY

In this section, we propose a multimodal and automatic method of emotion recognition based on the fusion of the above decisions. Our method is divided into two major phases namely: the Training phase and the Recognition phase

A. *Training phase*

This phase consists of three steps (preprocessing signal, feature extraction and training) in order to provide a training base which will then be used in the Recognition phase for the automatic detection of emotions as Figure 1.

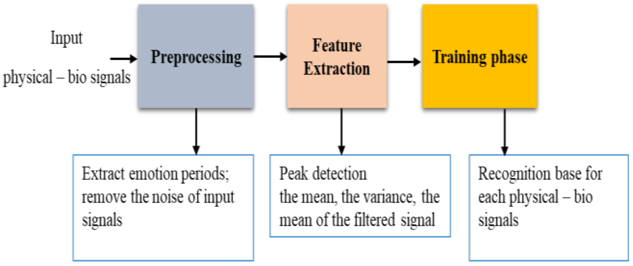
Input
physical – bio signals
Preprocessing
Feature Extraction
Training phase
Extract emotion periods; remove the noise of input signals
Peak detection the mean, the variance, the mean of the filtered signal
Recognition base for each physical – bio signals

Figure 1. Flowchart of the training phase

- *Preprocessing*

In this step, after having acquired the physical - bio signal (here we directly measure these signals and named UTE-EMOTICA database). We isolate the part of the signal corresponding to a given emotion because we have information on the period in which each emotions is expressed. We filter it to remove the noise of the useful signal, which will facilitate the extraction of the features. We have opted for the convolution method for filtering, which consists in convoluting the signal in the spatial domain with different filters (Hanning filter is chosen). This method is less computationally expensive in calculations.

- *Feature Extraction*

We continuously proceed to the detection of peaks, which is done by calculating the gradient of the signal and then finding the sign changes in the gradient, because it is rare to find points in discrete signals where the gradient is zero. A maximum is shown by the passage of a positive gradient to a negative gradient, a minimum by the passage of a negative gradient to a positive gradient. To detect and isolate a peak, our method should detect a minimum followed by a maximum followed by a minimum.

Once a peak is isolated, we calculate a feature vector composed of five features that are: the mean, the variance, the mean of the filtered signal, the variance of the filtered signal, and the amplitude of the peak.

In case of video signal, the procedure of feature extraction for face emotion is consist of 3 steps. Faces are detected from input image at step 1, then transformed and aligned by using Facial Landmark method at step 2 before feeding them into the trained model at final step.

The first two steps are very important for preparing input data to the CNN in final.

At step 1, we used Multi-Task Cascaded Convolutional Neural Network (MTCNN) which is considered as state-of-the-art face detection. This model consists of 3 separate networks: Proposal Network (P-Net), Refine Network (R-Net), and Output Network (O-Net) as depicted in Figure. 2.

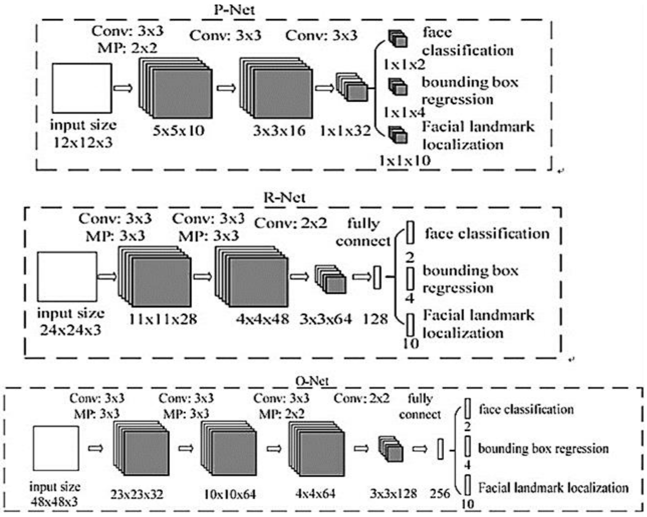

Figure 2. Multi-Task Cascaded Convolutional Neural Network

After detection step, face alignment step is highly recommended to be applied before moving to feature extraction phase. As recognition system depends upon how the face is aligned towards the camera, accuracy rate of face recognition can be in-creased by aligning the face by translation, rotation, and scale.

At final step, 128-D feature vector are extracted for every frame. To extract representation vector for each frame, a 6 layered Convolutional Neural Network (CNN) are used. We used pre-trained CNN model trained from public Kaggle emotion dataset.

- *Training phase*

We have 7 emotions that we are predicting namely (0=Angry, 1=Disgust, 2=Fear, 3=Happy, 4=Sad, 5=Surprise, 6=Neutral), so we have 7 labels.

There are 280 samples of 7 emotions in UTE-EMOTICA database

At each frame, 5 features are extracted from one bio-signal. With 3 types of bio-signal, we have totally 15 feature at each frame. Then we concatenate it with 128-D vector extracted from frame at same time step. At the end, we have 142-D feature vector.

After extraction of the features, linear SVM classifier is used for training recognition model. The Support Vector Machine (SVM) has been widely used in various pattern recognition tasks. It is believed that SVM can achieve near-optimal separation between classes[15]. The embedded data from the Representation step was used as inputs of SVM classifier to train on each identity. In linear SVM classifier, a data point is viewed as an n-dimensional vector, in n-dimensional space and the SVM's goal is separating such points with an (n − 1) dimensional hyperplane, see [16]. An SVM training algorithm builds a model of data points in space so that the data points of the separate classes are divided by a clear gap that is as wide as possible [16]. Given a training set of labeled samples:

$$D = \{(x_i, y_i) | x_i \in R^n, y_i \in \{-1,1\}\}_{i=1}^{p} \qquad (1)$$

A SVM tries to find a hyperplane to distinguish the samples with the smallest errors [15].

$$\mathbf{w} \cdot \mathbf{x} - \mathrm{b} = 0 \qquad (2)$$

B. *Recognition phase*

In recognition process, new input examples are mapped into that same space and predicted to belong to a class based on which side of the gap they fall on [16]. The SVM returns the label with the maximum score, which represents the confidence to the closest match within the trained data.

IV. RESULTS

For these results, we use as database the physical – bio signals made by the UTE-EMOTICA. For this database, we used maximum 3 physical – bio sensors were used: EMG, ECG, EDA and facial video capture. During this collection, 7 emotions were taken into account, which are "Angry, Disgust, Fear, Happy, Sad, Surprise, Neutral" and every emotion was maintained for five minutes per day. The results obtained by our algorithm when the unimodal recognition of emotions approach is used are grouped on the histogram below. This approach allows having a mean recognition rate of 63.52%.

We have 2 modals:

Modal 1: there two physical-bio signals: EMG and ECG and video capture.

Modal 2: there three physical-bio signals: EMG, ECG and EDA and video capture.

As shown in the Figure 3, certain emotions are better detected with certain modalities than others. Indeed the EMG modality allow to better detect the " Angry" and "Disgust" emotions, while the ECG modality better detects the 'Disgust' and 'surprise' emotions. The EDA modality rather allows a better detection of the emotions "Angry" and "Fear". These characteristic of modalities is very crucial because it will allow putting weight on each of the modalities, depending on whether it can better detect an emotion or not for the purpose of a more efficient detection.

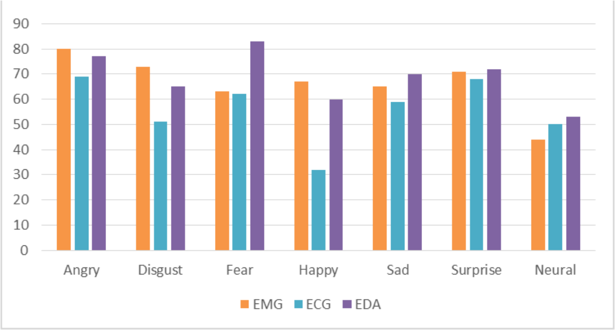

Figure 3. Monomodal recognition rate

Subsequently, we have enhanced our method to the multimodal approach to increase the emotion recognition rate. Indeed, this multimodal approach allowed having a recognition rate of 74.3% with modal 1 and 83.2% with modal 2, which is a considerable improvement of the recognition rate compared to the unimodal approach which presented a recognition rate of 63.52%.%.

Table 1. Accuracy of two modal

Emotion	Accuracy %	Accuracy %
	Modal 1	Modal 2
Angry	86.02	94.11
Disgust	77.63	89.89
Fear	62.4	73.2
Happy	69.6	80.3
Sad	71.67	77.45
Surprise	80.2	87.4
Neural	72.6	79.89

The results grouped in the above table present a average recognition rate. Furthermore, we find out that our method allows to detect each of the seven emotions with a good recognition rate, where the minimum of 62.4% is obtained for the emotion "Fear" and the maximum is obtained for the emotion" Angry" with a good classification rate of 94.11%. The table 2 below allows doing a comparison between our results and the different results of the methods of the state of the art that allow an instantaneous detection of emotions.

Table 2. Comparison with related works on Recognition rate

Methods	Recognition rate (%)
Kim's Project [17]	**61.2**
Fusion based HHT (fusion of 4 physiological signals) [18,19]	**62**
Fusion of 4 physiological signals [18]	**71**
Method of Chaka Koné [20]	**81.69**
Proposed Method (UTE-Emotica)	**83.2**

V. CONCLUSION AND PERSPECTIVES

We have proposed an enhanced method of multimodal recognition of emotions based on the processing of physical – bio signals. These signals of 2 modalities were applied for the recognition of 7 basic emotions (Angry, Disgust, Fear, Happy, Sad, Surprise, Neutral). Our proposed method for multimodal recognition based on the machine learning has been deployed and developed. The different results illustrate a remarkable improvement in the emotion recognition rate.

In our future work, we will create more sample for UTE- EMOTICA database for both physical – bio signals and video acquisition platforms in order to generate our recognition algorithm. Besides, a complete platform with more flexible and convenient for the acquisition of physical – bio signals for emotions detection. Moreover, our proposed system will be improved with more appropriate recognition base for a various people.

ACKNOWLEDGMENTS

1."This research is funded by Funds for Science and Technology Development of University of Technology and Education under project number T2018-06-86".

2. "This research is funded by Funds for Science and Technology Development of University of Danang under project number B2017-ĐN06-03"

REFERENCES

[1] TRangaraj M. Rangayyan; Biomedical Signal Analysis – A Case-Study Approach; IEEE Press 2002

[2] Barreto A., Heimer M., and Garcia M., "Characterization of Photoplehtysmographic Blood Volume Pulse Waveforms for Exercise Evalution," Proceedings 14th Southern Biomedical Engineering Conference, Shreveport, Louisiana, April, 1995, pp. 220-223

[3] Chrisite, "Multivariate Discrimination of Emotion-specific Autonomic Nervous System Activity", Master Thesis in Science of Psychology, Virginia

[4] Picard R.W.; Toward computers that recognize and respond to user emotion; IBM Systems Journal; Vol 39, Nos 3&4, 2000

[5] Healy J. and Picard R., "Digital processing of Affective Signals", ICASSP 98

[6] Cowie R., "Describing the emotional states expressed in speech", ISCA workshop on speech and emotion, Belfast 2000

[7] Juang B.H & Soong F.K., Hands-free Telecommunications; HSC 2001, pp.5-10; Kyoto, Japan.

[8] Pentland A.; Perceptual Intelligence; Communications of the ACM; Volume 43, Number 3 (2000), Pages 35-44.

[9] E. Monte-Moreno, M. Chetouani, M. Faundez-Zanuy et J. SoleCasals Maximum likelihood linear programming data fusion for speaker recognition. Speech Communication, 51(9):820–830, 2009. 68.

[10] A. Mahdhaoui et M. Chetouani : Emotional speech classification based on multi view characterization. In IAPR International Conference on Pattern Recognition, ICPR, 2010. 51

[11] Ammar Mahdhaoui. Analyse de Signaux Sociaux pour la Modelisation de l'interaction face à face.Signal and Image processing. Université Pierre et Marie Curie - Paris VI, 2010. French. <tel-00587051>

[12] Nicu Sebe, Erwin Bakker, Ira Cohen, Theo Gevers et Thomas Huang.Bimodal Emotion Recognition, 2005.

[13] Shan , Shaogang Gong , Peter W. McOwan : Facial expression recognition based on Local Binary Patterns: A comprehensive study Caifeng in Image and Vision Computing 27 (2009) 803–816

[14] E. Monte-Moreno, M. Chetouani, M. Faundez-Zanuy et J. SoleCasals Maximum likelihood linear programming data fusion for speaker recognition. Speech Communication, 51(9):820–830, 2009. 68

[15] Hamza Hamdi. Plate-forme multimodale pour la reconnaissance d'émotions via l'analyse de signaux physiologiques : Application à la simulation d'entretiens d'embauche. Modeling and Simulation. Université d'Angers, 2012. French. <tel-00997249>

[16] R. Sharma, V.I. Pavlovic, and T.S. Huang. Toward multimodal human-computer interface. roceedings of the IEEE, 86(5) :853–869, 1998. 29, 30, 32, 167.

[17] K. H. Kim, S.W. Bang, S.R. Kim Emotion recognition system using short-term monitoring of physiological signals in medical & biological engineering and computing 2004, vol 42, page 419-427, 17 February 2004

[18] Imen Tayari Meftah. Modélisation, détection et annotation des états émotionnels à l'aide d'un espace vectoriel multidimensionnel. Artificial Intelligence. Université Nice Sophia Antipolis,2013. French. <NNT : 2013NICE4017>. <tel-00908233>.

[19] C. Zong, M. Chetouani Hilbert Huang transform based Physiological signals analysis for emotion recognition in signal processing and information technology(ISSPIT),2009 IEEE International Symposium on, page 334-33, 14 Decembre 2009

[20] Koné C., Tayari I.M., Le-Thanh N., Belleudy C. (2015) Multimodal Recognition of Emotions Using Physiological Signals with the Method of Decision-Level Fusion for Healthcare Applications. Inclusive Smart Cities and e-Health. ICOST 2015. Lecture Notes in Computer Science, vol 9102. Springer, Cham

[21] Chen, Jun-Cheng, Vishal M. Patel, and Rama Chellappa. "Unconstrained face verification using deep cnn features." 2016 IEEE winter conference on applications of computer vision (WACV). IEEE, 2016.