COMPUTER METHODS AND PROGRAMS IN BIOMEDICINE 122 (2015) 149–164

journal homepage: www.intl.elsevierhealth.com/journals/cmpb

Reliable emotion recognition system based on dynamic adaptive fusion of forehead biopotentials and physiological signals

Mahdi Khezri [a,*], *Mohammad Firoozabadi* [b], *Ahmad Reza Sharafat* [a]

[a] *Department of Electrical and Computer Engineering, Tarbiat Modares University, Tehran, Iran*
[b] *Department of Medical Sciences, Tarbiat Modares University, Tehran, Iran*

A R T I C L E I N F O

Article history:
Received 6 March 2015
Received in revised form
11 July 2015
Accepted 13 July 2015

Keywords:
Emotion recognition system
Dynamic adaptive fusion of classification units
Forehead bioelectric signals
Physiological signals
Human computer interactions

A B S T R A C T

In this study, we proposed a new adaptive method for fusing multiple emotional modalities to improve the performance of the emotion recognition system. Three-channel forehead biosignals along with peripheral physiological measurements (blood volume pressure, skin conductance, and interbeat intervals) were utilized as emotional modalities. Six basic emotions, i.e., anger, sadness, fear, disgust, happiness, and surprise were elicited by displaying preselected video clips for each of the 25 participants in the experiment; the physiological signals were collected simultaneously. In our multimodal emotion recognition system, recorded signals with the formation of several classification units identified the emotions independently. Then the results were fused using the adaptive weighted linear model to produce the final result. Each classification unit is assigned a weight that is determined dynamically by considering the performance of the units during the testing phase and the training phase results. This dynamic weighting scheme enables the emotion recognition system to adapt itself to each new user. The results showed that the suggested method outperformed conventional fusion of the features and classification units using the majority voting method. In addition, a considerable improvement, compared to the systems that used the static weighting schemes for fusing classification units, was also shown. Using support vector machine (SVM) and k-nearest neighbors (KNN) classifiers, the overall classification accuracies of 84.7% and 80% were obtained in identifying the emotions, respectively. In addition, applying the forehead or physiological signals in the proposed scheme indicates that designing a reliable emotion recognition system is feasible without the need for additional emotional modalities.

1. Introduction

Emotion/affect is an internally mental perception of an object or event that can be associated with an expressive overt behavior. Incorporation of emotional intelligence in computers can improve their interactions with humans. Since human computer interactions (HCIs) have become unavoidable, more accurate HCIs that are comparable with human–human interactions are more desired and beneficial for users. Emotional skills power computers to understand and recognize human emotional states and give an appropriate response [1,2]. For example, a person who is working with a computer in a state of impatience or boredom expects the computer's response to be faster and more accurate [1]. In addition, for a student

* *Corresponding author*. Tel.: +98 21 82884331.
E-mail addresses: mahdi_khezri_ee@yahoo.com (M. Khezri), pourmir@modares.ac.ir (M. Firoozabadi), sharafat@modares.ac.ir (A.R. Sharafat) .

http://dx.doi.org/10.1016/j.cmpb.2015.07.006
0169-2607/

in an online learning environment, learning efficiency would increase if the computer were able to detect his or her emotions and create an appropriate situation according to this state [1]. Various applications for computers and systems identify different emotions, such as human-like communications, robotics, learning environment, entertainment, and games, providing assistance or feedback to a person, human–computer mediated systems, medicine, mental health, etc. [1–3].

Given the importance of emotions in human life, a new research field related to emotional phenomena was introduced by Rosalind Picard: affective computing [1]. The goal of affective computing is to design systems that are emotionally intelligent. To design an emotional system that detects different emotions, methods utilized by humans in their everyday communications such as facial expressions, speech and sounds, gestures, and body movements [4–11] as well as physiological patterns can be used [12–23]. For example, Motta and Picard [9] and D'Mello and Graesser [10] analyzed children's physical movements with the use of a body pressure measuring system (BPMS) to evaluate their interest in a learning environment and working with computers.

It has been shown that specific physiological patterns in each emotion are created. Ekman and Levenson presented the first findings that considerable changes in the autonomic nervous system (ANS) are created in accordance with emotional scenarios [24]. Since physiological signals are the result of ANS activity, they cannot be easily imitated. These signals are the same among people with different languages, cultures, and even gender and age [12,13]. In one of the first studies that used physiological signals, Picard and her colleagues recognized eight emotional states using blood volume pressure (BVP), skin conductance (SC), respiration rate (RR), and facial muscle activity [14]. They used personalized imagery to elicit the desired emotions from an actor and achieved overall recognition accuracy of 81% using hybrid linear discriminant classification. Nasoz et al. applied galvanic skin response (GSR) and heart rate variability (HRV) to recognize six types of emotions induced by selected movie clips [17]. The researchers achieved the best recognition accuracy rate of 83% with the Marquardt backpropagation algorithm. Yuan Lin and his colleagues identified four music induced emotional states (joy, anger, sadness and pleasure) from electroencephalogram (EEG) signals. They attained an average recognition accuracy of 82% using the SVM classifier [25]. In another study, Frantzidis et al. proposed a two-step classification procedure for discriminating emotional states from EEG signals [26]. The first classification level involved arousal discrimination, and then valence discrimination was performed. Using Mahalanobis distance (MD) and SVM classifiers, overall classification rates of 79.5% and 81.3%, respectively, were obtained [26]. AlZoubi et al. designed an emotion recognition system for detecting eight emotional states (e.g., boredom, confusion, curiosity, delight, flow/engagement, surprise, and neutral) during interactions between students and a tutoring system [20]. The researchers used three physiological signals (electrocardiogram (ECG), electromyogram (EMG), and GSR) and combinations. In that study, single-channel and three-channel multimodal models were generally more diagnostic than two-channel models. Recently, Khosrowabadi et al. applied a six-layer biologically inspired feedforward neural network to discriminate human emotions from EEG [27]. In that study, EEG data were collected from participants while they were subjected to audio-visual stimuli. Overall classification accuracy of 70.8% and 71.4% was obtained for arousal and valence discrimination, respectively.

Since the human body system may utilize a combination of emotional methods to represent a specific emotion, multimodal emotional systems have been proposed by researchers investigating affect [11–14,17,20,22,23,28–32]. An important consideration when designing a multimodal system is combining or creating fusion between the signals information. Fusion methods are usually implemented at the level of the extracted features or the results from individual classification units [33–37]. Fusing input signals due to a lack of the same time resolution usually is not considered.

In the design of a multimodal emotion recognition system, as we described some of them, the conventional method of feature fusion as well as simple fusion of classification units has been applied more frequently [11–23,28–32]. For example, Kim et al. applied the simple feature fusion scheme by concatenating the features extracted from the physiological signals to identify four types of emotional states [13]. Koelstra et al. implemented a statistical weighting scheme at the decision-level fusion of the classification units to determine the contribution of each emotional modality. They applied EEG and peripheral physiological signals as emotional measures with multimedia content analysis (MCA) [32]. Their system using fusion of all the modalities recognized arousal, valence, and liking/disliking ratings of music video-induced emotions, with F1-scores of 0.616, 0.647, and 0.618, respectively. In addition, Wagner et al. evaluated different statistical decision-level fusion such as weighted majority voting, weighted average, maximum, minimum and median rules, and more as well as cascading specialist, for designing a multimodal emotion recognition system using facial, gestural, and speech signals [11]. They obtained the best average accuracy of 55% using an emotion-adapted decision-level fusion scheme for identifying four categories of emotions: positive-high, positive-low, negative-high, and negative-low.

Although in feature-level fusion complete information on the signals is available, the system's response rate is decreased due to the high-dimensional features set [11]. In addition, the created feature set may contain redundant information that is not necessary for the system. In classifier or decision-level fusion, the signals create several independent classification units, and then the individual subsystem results are combined in a predetermined manner. In this type of fusion, if some of the classifiers produce incorrect results, other subsystems compensate and ultimately create the desired output. In addition, each unit applies only the features of some of the signals; therefore, the response rate of the system can be higher than the feature-level fusion scheme [11]. Despite all these advantages, in the classifier fusion schemes some of the signals may not be sufficiently informative, and therefore, their corresponding classification units do not act properly. This can undesirably influence the final accuracy of the ensemble system. Therefore, because of the varied performances of the modalities, their acceptability in the multimodal system is not the same [11,32]. The fusion process can be improved by using