Contents lists available at ScienceDirect

Digital Signal Processing

www.elsevier.com/locate/dsp

Emotion recognition based on time–frequency distribution of EEG signals using multivariate synchrosqueezing transform

Ahmet Mert [a,*], Aydin Akan [b]

[a] *Department of Mechatronics Engineering, Bursa Technical University, Yildirim, Bursa, Turkey*
[b] *Department of Biomedical Engineering, Izmir Katip Celebi University, 34320, Cigli, Izmir, Turkey*

ARTICLE INFO

Article history:
Available online xxxx

Keywords:
Emotion recognition
Electroencephalogram
Multivariate synchrosqueezing transform
Wigner–Ville distribution
Independent component analysis

ABSTRACT

This paper investigates the feasibility of using time–frequency (TF) representation of EEG signals for emotional state recognition. A recent and advanced TF analyzing method, multivariate synchrosqueezing transform (MSST) is adopted as a feature extraction method due to multi-channel signal processing and compact component localization capabilities. First, the 32 participants' EEG recordings from DEAP emotional EEG database are analyzed using MSST to reveal oscillations. Second, independent component analysis (ICA), and feature selection are applied to reduce the high dimensional 2D TF distribution without losing distinctive component information in the 2D image. Thus, only one method for feature extraction using MSST is performed to analyze time, and frequency-domain properties of the EEG signals instead of using some signal analyzing combinations (e.g., power spectral density, energy in bands, Hjorth parameters, statistical values, and time differences etc.). As well, the TF-domain reduction performance of ICA is compared to non-negative matrix factorization (NMF) to discuss the accuracy levels of high/low arousal, and high/low valence emotional state recognition. The proposed MSST-ICA feature extraction approach yields up to correct rates of 82.11%, and 82.03% for arousal, and valence state recognition using artificial neural network. The performances of the MSST and ICA are compared with Wigner-Ville distribution (WVD) and NMF to investigate the effects of TF distributions as feature set with reduction techniques on emotion recognition.

1. Introduction

Electroencephalogram (EEG) signal is a non-invasive recording of the electric potentials generated by neuronal activity in the brain. The multi-channel EEG signals are analyzed to extract information about the various status of the brain including diagnosis and brain computer interface (BCI). EEG analyzing techniques are the first preferred method for epilepsy and sleep disorder diagnosis [1,2], while BCI applications using EEG are widely studied to explore hidden information in the brain to control any peripheral [3]. Emotion state recognition using EEG signals is also a recent and popular research to explore how signals changes are affected by emotional state [4,5].

Cognitive emotion in 2D map is an approved approach to identify the states by psychologists [6], and widely used in emotion recognition applications. Arousal and valence scales in 2D cognitive emotional state are ranged from negative to positive to identify the emotion. While arousal scale quantifies activity or inactivity from excited to bored emotion, valence is used to define positive (pleasant) or negative emotion (e.g., from happy to sad). The basic approach to recognize these scores in 2D map is based on analyzing spectral properties of EEG signals. Thus, spectral powers in alpha (8–13 Hz) and beta (14–30 Hz) bands, and their ratios are widely studied by researchers. Alpha band is correlated to inactivity of the brain, while beta band is the indicator of the active brain state. On the other hand, spectral asymmetry between the EEG channels from right to left hemisphere is reported to be associated with the valence score [4,7–9]. It is assumed that inactivation in the left lobe estimated by spectral asymmetry indicates negative emotion. However, complex brain functions make emotion recognition using EEG signal a challenging study, and it attracts many researchers.

The first step of the emotion recognition is to obtain correctly labeled EEG signals stimulated by picture, music or video clips. After a visual stimuli is demonstrated to a participant, multi-channel EEG signals are recorded, and then the signals are labeled depending his/her ratings. One of the comprehensive research and database called "A database for emotion recognition analysis using physiological signals (DEAP)" consists of EEG signals, which are widely used by researchers for emotion recognition in 2D

* Corresponding author.
E-mail address: ahmet.mert@btu.edu.tr (A. Mert).

https://doi.org/10.1016/j.dsp.2018.07.003

scale [4]. 32 channel EEG recordings of 32 participants were acquired for 40 video clips stimuli, so each participant's 40 emotional states were classified using training states belonging to others. This scheme is called single trial classification, and the validation technique is named leave-one-participant-out cross-validation (LOPO-CV). Since, LOPO-CV is a participant or user independent classification task, many researchers implement LOPO-CV for benchmarking. In the original study of DEAP database, theta, alpha, beta, and gamma spectral densities (PSDs) with their asymmetrical PSDs (APSDs) of 32 EEG channels have been analyzed to extract feature vectors. Consequently, high/low arousal and valence classification with Bayes classifier (BC), LOPO-CV, and the proposed PSD based feature vectors have accuracy rates of 62% and 57.6%, respectively. Using the same validation and feature extraction on the DEAP recordings, principal component analysis (PCA) reduction and deep learning machine (DLM) performances are evaluated for benchmarking purpose of the proposed method with rates of 53.42%, and 52.03% for high/low/neutral valance & arousal states classifications [10].

Dual-tree complex wavelet packet transform (DT-CWPT) based feature extraction is also performed on DEAP database. The sub-band distributions of EEG signals are assumed as 2D signal, and then reduced using singular value decomposition (SVD) before support vector machine (SVM) classifier. The results of LOPO-CV are reported as accuracy rates of 66.9%, and 65.3% for high/low arousal, and high/low valence, respectively [11]. Empirical mode decomposition is also recent, advanced and adaptive method to analyze the signal in sub-bands [12]. The decomposed oscillations called intrinsic mode functions (IMFs) are band-limited, and they provide to analyze sub-bands in oscillatory mode [13]. Multivariate extension of the EMD called multivariate empirical mode decomposition (MEMD) [14] is used to obtain IMFs of the multi-channel EEG signals. Several signal analyzing methods are applied to extract feature vector for emotional state recognition [9]. IMFs power, Hjorth's parameters, PSDs and their asymmetrical differences from left to right lobe channels with ICA based feature reduction are the features of the study. k- nearest neighbor (k-NN) and artificial neural network (ANN) with two hidden layers using LOPO-CV yield mean accuracy rates of 51.01%, 67%, 75%, and 72.87% for high/low arousal and valence states recognition, respectively. In addition, multi-channel decomposition capability and high computational complexity were reported. In another emotion recognition based on EMD [15], time series differences (TSD), phase differences (PD) and normalized energy (NE) of the IMFs are analyzed for feature extraction, while EEG channel selection and IMF selection are investigated to reduce computational complexity without losing high accuracy rates up to 71.99% and 69.10%. In [16], EMD and Hilbert transform combination called Hilbert–Huang Transform (HHT) based features are obtained, and then SVM algorithm is trained and tested with selected feature sub-space. It has been reported that 73.61% and 84.83% of correct rates are resulted using different data and cross-validation. On the other hand, there are several studies to focusing on selection of the best feature subset to reduce complexity. In [17], several time and frequency-domain methods (statistical properties, band power, Hjorth parameters, and fractal dimension) are applied to each channel, and then mutual information based minimum-Redundancy-Maximum-Relevance (mRMR) method is used to select best feature set. Thus, 60.72% and 62.39% of accuracy rates are obtained using 35 and 20-dimensional space for high/low arousal and valence classification, respectively. In another comprehensive study on feature extraction, selection, and classification of emotional EEG, more features in time-domain (Statistical features: mean, maximum, minimum, power, standard deviation, 1st difference, normalized 1st difference, standard deviation of 1st difference, 2nd difference, standard deviation of 2nd difference, normalized 2nd difference, quartile 1, median, quartile 3, quartile 4. Hjorth's parameters: activity, mobility and complexity. Fractal dimension. Non-stationary index. Higher order crossing), in frequency-domain (power spectrum, logarithm of power spectrum), and in time–frequency domain (Power of DWT, RMS of DWT, recursive energy efficiency (REE) of DWT, and Abs (log (REE)) from gamma, theta, alpha, beta.) are performed on DEAP dataset to evaluate 5 population-based heuristic search algorithms, namely: Differential evolution (DE), genetic algorithm, ant colony optimization, particle swarm optimization, and simulated annealing. 45 feature extraction methods with selection using DE has the maximum accuracy rate of 67.474% for 4 classes (four quadrants in 2D emotional map) using probabilistic neural network (PNN). There are also several studies with different EEG recordings [18], validation scheme and approaches. Speech analysis [19], chaotic behavior of galvanic skin response (GSR) [20], forehead signal analysis of music induced emotional state [7], and wireless low-cost EEG & ECG signals acquisition [21] can be categorized as different types of emotion recognition. Especially, portable and low-cost feasibility of the emotion recognition is investigated in a recent study [21]. Instead of using expensive medical devices, EEG and ECG wireless signal acquisition is adopted to affective computing in everyday application. After common reference, PSDs of the 14 channels EEG, heart rate variability (HRV) and PQRST features from ECG signal are applied to SVM. Accuracy rates of 62.49%, and 62.32% are also reported for high/low valence, and high/low arousal classification, respectively.

In this study, our purpose is to analyze EEG signals in time–frequency (TF) domain. Unlike the combination of various signal processing and analyzing methods used for feature in previous studies, our proposed method is based on exploring the TF domain of multi-channel EEG signals for emotional state recognition. Thus, class separating properties of the well-known and successful methods (e.g., PSDs, energy, sub-band energy, asymmetrical PSDs, and IMFs properties in EMD domain) will be included into one TF distribution. From this perspective, the traditional Wigner–Ville distribution (WVD), and the recent and advanced method, namely synchrosqueezing transform (SST) [22] are used to extract feature vectors of the proposed emotion recognition. SST is a recent and advanced analyzing method to generate highly localized TF representation of amplitude and frequency modulated signals. It is a post-processing method on continuous wavelet transform (CWT) to overcome the limitations of the WVD, short-time Fourier Transform (STFT), and CWT. Thus, it is capable of concentrating energy coefficients around the instantaneous frequency curves. In addition to these localized TF representation, multivariate extension of the SST method called multivariate synchrosqueezing transform (MSST) is studied to process multi-channel signals, and to obtain compact TF distribution covering the mutual components among the channels [23]. Consequently, our approach to analyze the TF distribution of the EEG signals to cover all energies of sub-bands and components for emotional feature extraction. First, WVD and MSST based TF representations in 2D are extracted, and then ICA is used to reduce dimensionality by revealing linear combinations of the TF distributions. Second, machine learning algorithms are performed on emotional state recognition. The same approaches especially the LOPO-CV procedures of the original DEAP study are applied for benchmarking purpose of the proposed method.

The remainder of the paper is organized as follows: Section 2 provides a short description of the DEAP emotional EEG database, TF-domain analysis, and reduction methods. The proposed TF representation based emotion recognition is presented in Section 3. In Section 4, simulation results of the proposed method are examined, and discussed. Consequently, the conclusions are drawn in Section 5.

Please cite this article in press as: A. Mert, A. Akan, Emotion recognition based on time–frequency distribution of EEG signals using multivariate synchrosqueezing transform, Digit. Signal Process. (2018), https://doi.org/10.1016/j.dsp.2018.07.003