Biomedical Signal Processing and Control 47 (2019) 26–40

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

F-SVD based algorithm for variability and stability measurement of bio-signals, feature extraction and fusion for pattern recognition

Anil Hazarika [a,*], Mausumi Barthakur [b], Lachit Dutta [a], Manabendra Bhuyan [a]

[a] Department of Electronics and Communication Engineering, Tezpur University, Tezpur, Assam, India
[b] Department of Neurophysiology, Guwahati Neurological Research Centre, Guwahati, Assam, India

ARTICLE INFO

Article history:
Received 7 June 2017
Received in revised form 24 May 2018
Accepted 8 August 2018

Keywords:
Singular value decomposition (SVD)
Canonical correlation analysis (CCA)
Motor unit action potential (MUAP)
Collective correlation coefficient matrix (CCCM)
Electromyography (EMG)

ABSTRACT

A support system with efficient learning framework helps eliciting complete knowledge of underlying phenomena of interest. It makes the analysis less-onerous, time-consuming and error-prone and thus promotes large scale applications. Such modeling requires profound understanding of available information and its appropriate utilization. Albeit success of electromyogram (EMG) support systems, challenges still exits specifically in early phase of design mainly due to inherent variations and complex data distribution patterns of signals. In this article, a frame singular value decomposition (F-SVD) based method-generalizing Canonical correlation analysis for automatic classification of EMG signals to diagnose amyotrophic lateral sclerosis (ALS), myopathy and normal subjects, is proposed. At first, signals are decomposed to formulate a set of vectors and performed subspace transformation to demonstrate the variability and stability of signals base on correlations between pairs of vectors. Besides, discrete Wavelet transformation is applied on generated vectors and correlation analysis is performed. Afterwards, taking highly correlated statistical measures a set of compact feature distributions are estimated and fused via two recently proposed parallel and serial feature fusion models. Finally two global descriptors for effective classifications of various EMG patterns are proposed. The efficacy of derived feature space is validated by intuitive, graphical and statistical analysis. The model performances are investigated over two datasets. It achieves accuracy of 98.10% and 97.60% over two and three-class groups of first dataset receptively. Accuracy over second dataset is 100% with a specificity of 100% and sensitivities of 100%. This is first time that F-SVD is employed for automatic classification of EMG. Experiments results on various datasets evince adequacy of our method. Further comparison of performance with state-of-the-art methods depicts that our method comparable or superior in terms of various performance metrics.

1. Introduction

An efficient modeling to derive a viable support system for medical applications relies on profound understanding of available information and its proper utilization [1]. However, wide variability and complex nature of information are inherent to many engineering domains [2]. The feature data may contain many redundant, irrelevant and superfluous contents. Therefore, removal of such contents and estimation of an appropriate features for effective learning are indispensable provisions. In such cases, information fusion, specifically feature fusion provides an effective solution [3].

Objective of feature fusion is to transform the input space to an well-defined coordinate system, from where latent features are derived based on some optimization criteria [4]. Such features have much robustness in real-life applications especially when best feature sets are unknown [5]. Principal components analysis (PCA) well approximates the distribution of original space without considering the class information. Linear discriminant analysis (LDA) [6] incorporates the class structure to enhance the separation margin among multi-group features. Other classification paradigms include uncorrelated LDA [7], non-linear LDA and self-organizing feature map [4]. However, LDA fails to preserve the common relation between two feature vectors. Recently, canonical correlation analysis (CCA) [8,9] has become popular among the research communities. It finds unique feature based on the degree of proximity between two input feature views or vectors extracted from same or different objects. Such features can inherently characterize the object and thus it can better explore the insights and finer details of the problems at hand.

* Corresponding author.
E-mail address: anilhazarika8@gmail.com (A. Hazarika).

https://doi.org/10.1016/j.bspc.2018.08.010
1746-8094/

Over the decade, a number of such applications are reported. Sun et al. [10] adopted a CCA-based feature fusion for handwritten characters and face recognition. Measuring the correlation between two groups of feature vectors, highly correlated features were evaluated and used for learning. Shen et al. [11] addressed an orthogonal multi-CCA to avoid small size problem and to improve the recognition rate. Pong et al. [12] employed a generalized version for face recognition using image information at various level of resolutions to enhance quality of feature space and reported promising results. In [13], generalized projected vectors were used for image recognition. In [14,15], a multiset integrated CCA (MICCA) framework was introduced for multi-set problems which distinctly represents integral correlation among multi-set features. However, it is an iterative approach, which reduces the performance. Most of the aforementioned studies include highly correlated vectors as features, however orthogonal components among variables that may also play significant role for recognition were not considered. In [16], a local discrimination CCA was employed to combine local properties and discrimination between various classes in addition to correlated features among the class variables. Shen et al. [11] introduced a two-view semi-supervised model to include a new sparse based label propagation algorithm to infer label information for unlabeled data. Allan [17] addressed multi-group variables problems, however integral relation among the multi-set variables was not reported. Moreover, various other applications of CCA include multimodal fusion [18], myoelectric interface [19], artifact removal [20], frequency recognition [21] and brain imaging modalities study [22]. Thus, promising results of aforementioned studies promote to implement such strategy for bio-medical signals analysis such as an EMG.

Various EMG support system were reported in literature [23–25]. Some approaches utilized direct signal features and others employed morphological features of motor unit action potentials (MUAPs) [26,27]. In [28], dominated MUAPs features were employed to the learning model. However, for wide variability and mix MUAP patterns in signals, it is difficult to identify the correct MUAP pattern for decision. Recently, wavelet based methods achieved remarkable success [29–36]. For instance, Subasi et al. [37] proposed a synchronized version of discrete WT (DWT) and autoregressive coefficients (i.e., DWT+AR). Fattah et al. [38] cited two direct methods, one utilized the higher values of DWT coefficients while other employed ten maximum DWT coefficients extracted from frames of EMG signal for classification. In [39], the statistical measures of DWT coefficients were utilized to classify various classes of neuromuscular diseases. Use of DWT and its statistical measures greatly reduce the dimensionality issue, nonetheless, it often requires the manual intervention [7]. Use of coefficients from each level may not be feasible for all MUAPs to categorize various classes. Further, MUAP based methods process all available MUAPs equally while ignoring the non-stationary MUAPs that could lead to false findings. In general, existing methods utilize either specific set of features or their statistical measures. Although it simplifies the models, but may fail in learning of wide-scale applications. Intuitively, modeling multiple views or features can bring more explicit information about the disease associated with the subjects to the learning process.

Fig. 1 gives a schematic outline of the proposed method for EMG classification using frame singular value decomposition (F-SVD)-generalizing CCA. At first, EMG signals are decomposed into set of multiple vectors. Pair wise correlations are then computed to measure variability of signal profiles and significant sets of vectors are estimated. DWT is performed on multiple vectors followed by correlation measures to estimated second set of vectors. Afterwards, a set of discriminant are derived and fused through two fusion to derive global descriptors. This approach does not require prior knowledge and also free from assumptions. To the best of author knowledge, classification of EMG in F-SVD domain is yet to be reported in the literature. The performances are evaluated and compared with state-of-the-art-methods. The rest of the article is organized as follows. Section 2 discusses experimental data used in this work. Section 3 gives a brief description of theoretical approach. Section 4 provides feature extraction and selection procedure. Section 5 describe feature fusion and classification strategy and results and discussion is presented in Section 6, comparison analysis in Section 6.3. Finally, we conclude with Section 7.

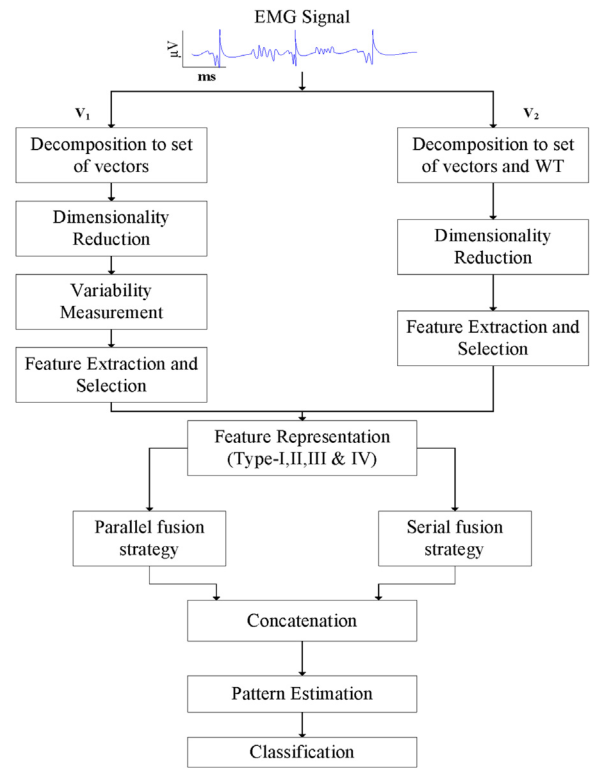

Fig. 1. Flowchart diagram of proposed method.

2. Materials

The experiment is carried out using wide-spread used public database [40] and Guwahati Neurological Research Centre hospital database [41]. Data sets consist of three group of subjects, *Amyotrophic lateral sclerosis (ALS)* and *Myopathy (MYO)* and *Normal (NOR)* groups. First two groups are categorized as neuromuscular disorders (NDs). None in the control group had signs or history of NDs. First data set contain a total of 250 recordings (50 ALS, 50 MYO and 150 NOR) [40] and 60 recordings (20 from each) in the second set [41]. The study includes both male and female subjects. NOR, ALS and MYO include 6 males, 4 females of age 21–37 yrs (mean 27.7 ± 4.5), 4 males, 4 females of age: 35–67 yrs (mean 52.8 ± 11.8) and 5 males, 2 females of age 19–63 (mean 36.3 ± 14.6) respectively. Second data set includes four ALS subjects – 3 females and 1 males of age 38–52 yrs (mean 43.5 ± 7), 4 cases of MYO – 2 males and 2 females of age 42–59 yrs (mean 47.5 ± 7.8) and 4 NOR – 2 females and 2 males of 26–34 yrs (mean 29.3 ± 3.4). Recording were performed under proper supervision of expert neurologist. Duration of each recording is 11.5 s and was sampled at 23,437.5 Hz (f_s) with 16 bit resolution. Concentric needle electrode with a leading-off area of 0.07 mm^2 was used to record the signals by inserting at various muscle sites with a surface ground electrode placed on the limb position. Recording sites are (a) biceps brachii (BB); (b) abductor polliciesbrevis (AP); (c) tibialis anterior (TA); and (d) vas-