2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Combined Dynamic Time Warping and Spatiotemporal Attention for Myoelectric Control

Milad Jabbari[1], Rami N. Khushaba[2] and Kianoush Nazarpour[3]

Abstract— **The success of pattern recognition based upper-limb prostheses control is linked to their ability to extract appropriate features from the electromyogram (EMG) signals. Traditional EMG feature extraction (FE) algorithms fail to extract spatial and inter-temporal information from the raw data, as they consider the EMG channels individually across a set of sliding windows with some degree of overlapping. To tackle these limitations, this paper presents a method that considers the spatial information of multi-channel EMG signals by utilising dynamic time warping (DTW). To satisfy temporal considerations, inspired by Long Short-Term Memory (LSTM) neural networks, our algorithm evolves the DTW feature representation across long and short-term components to capture the temporal dynamics of the EMG signal. As such the contribution of this paper is the development of a recursive spatio-temporal FE method, denoted as Recursive Temporal Warping (RTW). To investigate the performance of the proposed method, an offline EMG pattern recognition study with 53 movement classes performed by 10 subjects wearing 8 to 16 EMG channels was considered with the results compared against several conventional as well as deep learning-based models. We show that the use of the RTW can reduce classification errors significantly, paving the way for future real-time implementation.**

Index Terms— **Deep learning, electromyography, myoelectric control, warping.**

I. INTRODUCTION

Machine learning is a promising approach for mapping the electromyogram (EMG) activity to a set of control commands to control a powered prosthetic [1]–[3]. Research in this direction has shown that a number of factors may impact the accuracy of EMG-based control, these include all of the electrode location shift, varying muscle contraction efforts, muscle fatigue, forearm orientation, contraction intensity and their combinations [4]–[6]. The key factor determining the robustness of this control scheme, denoted as myoelectric control, is the separability of the extracted EMG features. The quality of the features has been shown to have a greater influence on the performance of EMG decoding than the choice of the classifier [4].

Current EMG feature extraction (FE) methods are criticized for their two intrinsic limitations. First, they are cross-sectional; that is, they cannot extract the inter-temporal information and dependencies that may exist between FE windows, even though these windows may be overlapping [7], [8]. Second, they merely concatenate features extracted from individual channels [5], [9]–[11]; not capturing the synergistic and spatial patterns of muscles activity.

To address the first limitation, a number of hand-crafted [5], [12] as well as deep learning (DL) based approaches including long short-term memory (LSTM) neural networks [13]–[15] were proposed. While such methods excel in extracting the temporal information, they merely rely on the concatenation of features extracted from individual channels and hence may overlook the spatial information aspect, that is the relationship between different muscles' activation patterns.

Similarly, recent literature offer methods that extract spatial features of the EMG signals; by capturing the relationship between spatially distributed EMG sensors, via traditional [5], [10], [16] as well as DL models, e.g. convolutional neural networks (CNN) [17], [18]. The idea of extracting features from spatially filtered EMG signals was also applied with various time-domain (TD) features showing significant benefits in terms of the classification accuracy for both offline analysis [3], [10] and real-time implementation [19]. However, muscles activation patterns have temporal components that might be overlooked when using traditional spatial information extraction tools. In addition, conventional structures used within DL models are typically blind to, within-channel, inter-temporal information of the EMG signals. Therefore, there has been a surge of research in combining CNNs and LSTM neural networks to capture spatio-temporal features the EMG signals [20], [21]. Not surprisingly, a limiting factor in the real-time implementation of such approaches for myoelectric control is their prohibitively large number of model parameters [22].

Our hypothesis is that a mixture of bespoke hand-crafted features that extract inter-temporal and spatial components of the EMG signals can compete with or possibly outperform data-driven methods, such as DL. To test this hypothesis, we developed an LSTM-type FE paradigm, named recursive temporal warping (RTW), to combine the long-and short-term spatial characteristics of the EMG signals. For this purpose, dynamic time warping (DTW) was employed to efficiently capture the nonlinear similarity between the EMG signals from multiple spatially distributed sensors. Also, we developed an attention mechanism to limit the length of the short-term memory component; enabling future real-time translation.

This work was supported by the Engineering and Physical Sciences Research Council (EPSRC), U.K., under Grant EP/R004242/1.

[1]Milad Jabbari was with Iran University of Science and Technology, Tehran, Iran. jabbari_milad@yahoo.com

[2]Rami Khushaba is with the Australian Centre for Field Robotics (ACFR), The University of Sydney, Chippendale, NSW 2008, Australia rami.khushaba@sydney.edu.au

[3]Kianoush Nazarpour is with School of Informatics, The University of Edinburgh, Edinburgh EH8 9AB, United Kingdom Kianoush.Nazarpour@ed.ac.uk

II. MATERIALS AND METHODS

A. Dynamic time warping (DTW)

Assume two temporal sequences of length l, $\mathbf{a} = \{a_1, a_2, \cdots, a_l\}$ and $\mathbf{b} = \{b_1, b_2, \cdots, b_l\}$, representing the EMG signals from two channels. We define $\mathbf{D}(\mathbf{a}, \mathbf{b})$ as an $l \times l$ distance matrix between $\mathbf{a}$ and $\mathbf{b}$, with $\mathbf{D}_{ij} = (a_i - b_j)^2$. In the case of DTW, the distance is calculated through a different warping path $\mathcal{P}$ generated by traversing the matrix $\mathbf{D}$ along ordered pairs of positions:

$$\mathcal{P} =< (e_1, f_1), (e_2, f_2), \cdots, (e_l, f_l) > \tag{1}$$

where $e_i \in [1 : l]$ and $f_i \in [1 : l]$ are the positions that make the warping path through the two time series $\mathbf{a}$ and $\mathbf{b}$. A valid warping path must satisfy the conditions $(e_1, f_1) = (1, 1)$, $(e_l, f_l) = (l, l)$, $0 \leq e_{i+1} - e_i \leq 1$ and $0 \leq f_{i+1} - f_i \leq 1$ for all $i < l$.

To find the DTW distance, a constraint on the amount of warping allowed is usually imposed on the distance measurements, that is $|e_i - f_i| \leq w.l$, $\forall (e_i, f_i) \in P^*$. The value of w is the maximum amount the warping path is allowed to deviate from the diagonal. The distance D to any path $\mathcal{P}$ is given with

$$D_{\mathcal{P}}(\mathbf{a}, \mathbf{b}) = \sum_{i=1}^{s} p_i \tag{2}$$

where $p_i = \mathbf{D}_{e_i, f_i}$ is the distance between element at position e_i of $\mathbf{a}$ and at position f_i of $\mathbf{b}$ for the i^{th} pair of points in a proposed warping path P. The total number of elements in the warping path is represented by s. Considering a space of all possible paths as $\mathcal{P}$, then the DTW path P^* is the one that has the minimum distance, i.e.,

$$P^* = \min_{P \in \mathcal{P}} \mathbf{D}_{\mathcal{P}}(\mathbf{a}, \mathbf{b}) \tag{3}$$

which can be found exactly with a dynamic programming formulation.

B. Recursive Temporal Warping (RTW)

Applying DTW across each two EMG channels results in a feature vector that is denoted with DTW. We multiply the feature vectors extracted from the current windows, DTW_t, with that extracted from the previous window DTW_{t-n}. To account for the long-term memory (LTM), we borrow the cell state concept of the LSTM method. This is the information highway that passes through all the cells across all time steps to evolve a recursive representation of the features. At the same time, the contents of each cell are updated with a weighted contribution of the cell state through the parameter β (is selected empirically to achieve the best performance).

$$features_t = DTW_t \odot DTW_{t-1} + \beta \times LTM, \tag{4}$$

where $\odot$ denotes element-wise multiplication (short-term memory (STM)), and $features_t$ represents the extracted features from the current time step (t). An attention mechanism is used twice: 1) within each cell to normalize the current windows features, and 2) before adding it to the aforementioned features to normalize the long-term memory component. The final features representation is given by

$$features_{ti} = \frac{features_{ti}}{\sum_i features_{ti}} + \log\left(1 + \frac{LTM_i}{\sum_i LTM_i}\right) \tag{5}$$

A subscript i was added to clarify the normalization applied across features of the current time step t. The structure of the proposed method is shown in Fig. 1(a).

C. Data Description

We used database-5 (DB5) of Ninapro repository [23] that included EMG data from ten intact-limbed subjects performing 52 movement classes and rest (Totally 53 classes). Each movement was repeated for six times.

D. Feature Sets and classifiers

The following features were extracted from overlapping windows of 150 ms at 50 ms increment:

- HTD: Hudgins' TD feature set [7].
- AR-RMS: The 6th-order auto-regressive (AR) model parameters and the RMS of the EMG signal from each channel;
- LSF9: The lower sampling rate features defined in [9].
- ATD: Combined AR with TD as defined in [24].
- STFS: Spatiotemporal features from [25].
- fTDD: The fusion of TD features from [8].
- TSD: The temporal spatial-descriptors from [5].

For classification, the following conventional classifiers were chosen:

- Linear discriminant analysis (LDA)
- Extreme learning machine (ELM) (one hidden layer and 1250 neurons)
- k-nearest neighbour (KNN) ($k = 5$).
- Support vector machines (SVM).
- Long short term memory (LSTM).
- Convolutional neural network (CNN).
- Combination of CNN and LSTM (CNN+LSTM)

For the deep networks, raw EMG signals were processed using the root-mean-square (RMS) values, generating NC scalar values for per each analysis window, where NC is the number of EMG channels in each data set. These RMS values were turned into pseudo-images by multiplying each generated vector of size ($NC \times 1$) by its transpose ($1 \times NC$), resulting in images of size ($NC \times NC$). These images were then scaled logarithmically and then provided as inputs to the CNN and CNN+LSTM models. For the LSTM model, the raw EMG samples for each 150 ms window of the NC channels were provided as input. The layered structure of the utilized DL methods are shown in Fig. 1(b).

III. RESULTS

To test the performance of the proposed FE algorithm, leave one out cross validation scheme was used and average classification errors were calculated for each fold. The Wilcoxon signed rank test was applied to investigate

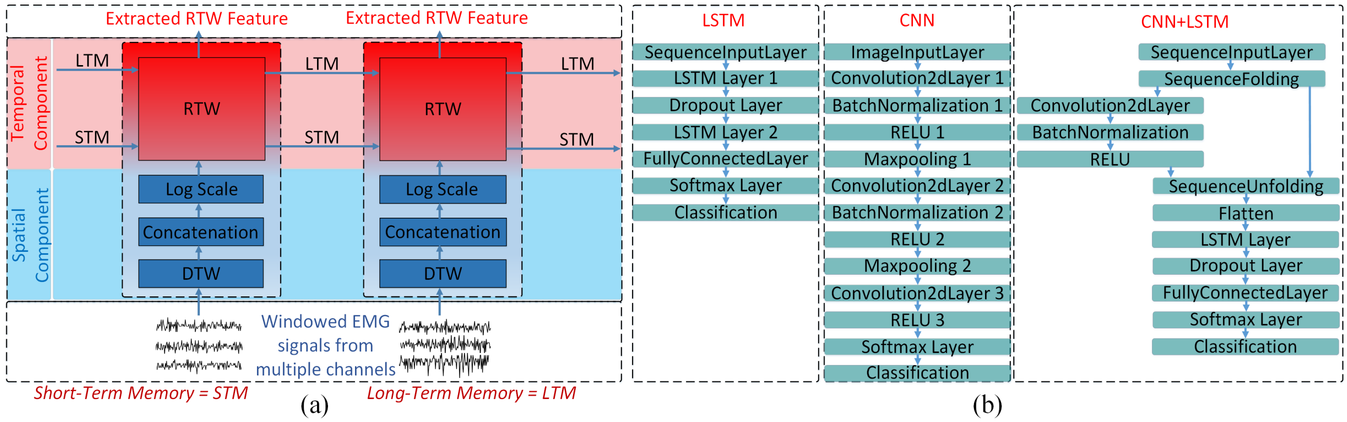

Fig. 1. Block diagram of the proposed RTW FE method (a) and the utilized DL models (b).

the statistical significance. Moreover, the size of differences observed was measured using Cohen's effect size d for paired samples of the achieved results. Fig. 2(top) illustrates the box plots of average classification errors for seven classifiers using raw EMG for DL models and RTW feature for a number of traditional classifiers. In terms of the DL models, the CNN+LSTM setting demonstrated a better performance with average classification error of 32.11±4.70% than the individual models of CNN and LSTM with 35.64±5.02% and 41.30±4.07%, respectively (p <0.001 for both tests, d=3.15 for CNN vs. CNN+LSTM and d=3.55 for LSTM vs. CNN+LSTM). This in turn indicates that combination of CNN+LSTM captures more information than the individual CNN and LSTM, while the spatial information captured by CNN appears more important on these data sets than the temporal information captured by LSTM. Moreover, it is demonstrated that RTW feature set combining with common traditional classifiers has achieved the lowest error. Among them, the KNN and SVM classifiers showed the best performance with average errors of 16.94±2.47% and 16.94±2.44%, respectively (no statistically significant differences were found between KNN and SVM, d=0.0032, and p=0.4325).

We compared the performance of the RTW with that of the recently developed features by the other groups using SVM in all cases. Fig. 2(bottom) shows that the RTW achieves the lowest error (16.94±2.44%). The fTDD is the second best feature with an average error of 30.02±4.15%. Statistical evaluations show there is significant difference between RTW and fTDD (p <0.001, and d=2.9921). Moreover, confusion matrices are illustrated in Fig. 3(top, left), (top, right), (bottom, left), and (bottom, right) for RTW+SVM, AR_RMS+SVM, RTW+KNN, and fTDD+KNN, respectively. The matrices show that the performance of RTW is significantly better than AR_RMS and fTDD in both SVM and KNN classifiers. Additionally, Fig. 4(top) shows the average classification errors with the proposed RTW with respects to varying the window sizes: from 50 ms to 250 ms. The results show the error rates reduce with larger windows, as expected. Finally, the standard deviation (STD) of the

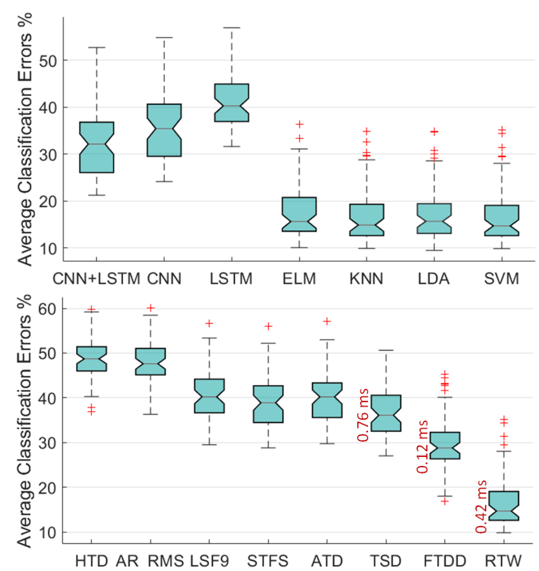

Fig. 2. Box plots of average classification errors achieved by raw EMGs as input for DL models and RTW for traditional classifiers (top), and average errors with seven traditional features and RTW using SVM (bottom). Times required to extract RTW, fTDD, and TSD are illustrated with red values. (The averages are calculated across 6 folds)

classification errors for each feature set are shown in Fig. 4(bottom). As it is visible, RTW has the lower variability among all features.

IV. Concluding Remarks

We presented the RTW method, a new temporal-spatial FE algorithm. It adopts the long and short-term memory concepts to two challenges of conventional FE methods including being cross-sectional and the lack of ability to capture spatial patterns of the EMG signals. The proposed RTW evolves the spatial similarity measures with an attention mechanism considering the cross-attention between different EMG signals. An attention mechanism has lately been used to improve neural machine translation by selectively focusing on parts of the source sentence during translation. The novelty of this paper lies in the development of a temporal-

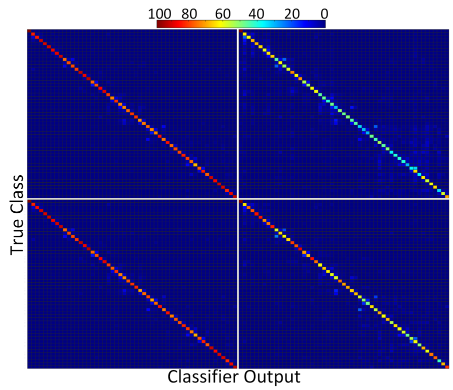

Fig. 3. Confusion matrices of 53 classes for RTW+SVM (top, left), AR_RMS+SVM (top, right), RTW+KNN (bottom, left), and fTDD+KNN (bottom, right). Color map represents classification accuracy scores.

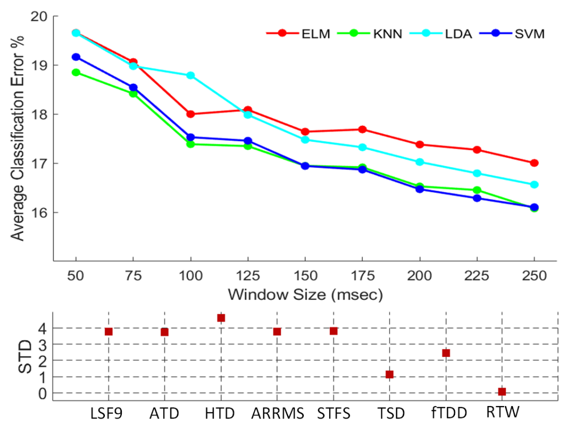

Fig. 4. Average RTW classification errors with different windows sizes (top), and STD of the classification results by the different features across the different classifiers (bottom).

spatial attention mechanism. Our findings illustrate that the RTW algorithm can reach the best performance compared to other traditional feature sets. Results show that time required to extract RTW with 0.42 ms is within the range of acceptable time as it false near the sanest value achieved by other FE algorithms and would be practicable in future real-time implementations. Using RTW as input for traditional classifiers significantly outperformed DL models with raw EMGs as input while showing lower computational expenses. Feeding DL models with the RTW instead of raw EMG signals can be considered to test the possibility of increasing the accuracy in future works.

References

[1] K. Nazarpour, *Control of Prosthetic Hands: Challenges and Emerging Avenues*. London: IET, 2020.

[2] A. Phinyomark and E. Scheme, "EMG pattern recognition in the era of big data and deep learning," *Big Data and Cognitive Computing*, vol. 2, no. 3, p. 21, 2018.

[3] A. Krasoulis and K. Nazarpour, "Myoelectric digit action decoding with multi-output, multi-class classification: an offline analysis," *Sci. Rep.*, vol. 10, p. 16872, 2020.

[4] L. J. Hargrove, K. Englehart, and B. Hudgins, "A comparison of surface and intramuscular myoelectric signal classification," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 5, pp. 847–853, 2007.

[5] R. N. Khushaba, A. H. Al-Timemy, A. Al-Ani, and A. Al-Jumaily, "A framework of temporal-spatial descriptors-based feature extraction for improved myoelectric pattern recognition," *IEEE Trans. Neural Syst. Rehab. Eng.*, vol. 25, no. 10, pp. 1821–1831, 2017.

[6] R. N. Khushaba, A. Al-Timemy, S. Kodagoda, and K. Nazarpour, "Combined influence of forearm orientation and muscular contraction on EMG pattern recognition," *Expert Syst. Appl.*, vol. 61, pp. 154–161, 2016.

[7] B. Hudgins, P. Parker, and R. N. Scott, "A new strategy for multifunction myoelectric control," *IEEE Trans. Biomed. Eng.*, vol. 40, no. 1, pp. 82–94, 1993.

[8] R. N. Khushaba, A. Al-Ani, A. Al-Timemy, and A. Al-Jumaily, "A fusion of time-domain descriptors for improved myoelectric hand control," in *2016 IEEE Symposium Series on Computational Intelligence (SSCI)*, 2016, pp. 1–6.

[9] A. Phinyomark, R. N Khushaba, and E. Scheme, "Feature extraction and selection for myoelectric control based on wearable EMG sensors," *Sensors*, vol. 18, no. 5, p. 1615, 2018.

[10] A. A. Al Taee, R. N. Khushaba, and A. Al-Jumaily, "Spatially filtered low-density EMG and time-domain descriptors improves hand movement recognition," in *Proc. IEEE EMBC*, 2019, pp. 2671–2674.

[11] E. Campbell, A. Phinyomark, and E. Scheme, "Current trends and confounding factors in myoelectric control: Limb position and contraction intensity," *Sensors*, vol. 20, no. 6, p. 1613, 2020.

[12] A. H. Al-Timemy, R. N. Khushaba, G. Bugmann, and J. Escudero, "Improving the performance against force variation of EMG controlled multifunctional upper-limb prostheses for transradial amputees," *IEEE Trans. Neural Syst. Rehab. Eng.*, vol. 24, no. 6, pp. 650–661, 2015.

[13] A. Olsson, N. Malešević, A. Björkman, and C. Antfolk, "Exploiting the intertemporal structure of the upper-limb sEMG: Comparisons between an lstm network and cross-sectional myoelectric pattern recognition methods," in *Proc. IEEE EMBC*, 2019, pp. 6611–6615.

[14] C. Wang, W. Guo, H. Zhang, L. Guo, C. Huang, and C. Lin, "sEMG-based continuous estimation of grasp movements by long-short term memory network," *Biomed. Signal. Proces. Control.*, vol. 59, p. 101774, 2020.

[15] M. Jabbari, R. N. Khushaba, and K. Nazarpour, "EMG-Based hand gesture classification with long short-term memory deep recurrent neural networks," in *Proc. IEEE EMBC*, 2020, pp. 3302–3305.

[16] M. Stachaczyk, S. F. Atashzar, and D. Farina, "Adaptive spatial filtering of high-density EMG for reducing the influence of noise and artefacts in myoelectric control," *IEEE Trans. Neural Syst. Rehab. Eng.*, vol. 28, no. 7, pp. 1511–1517, 2020.

[17] W. Geng, Y. Du, W. Jin, W. Wei, Y. Hu, and J. Li, "Gesture recognition by instantaneous surface EMG images," *Sci. Rep.*, vol. 6, no. 1, pp. 1–8, 2016.

[18] Q. Gao, J. Liu, and Z. Ju, "Hand gesture recognition using multimodal data fusion and multiscale parallel convolutional neural network for human–robot interaction," *Expert Systems.*, p. e12490, 2020.

[19] A. Krasoulis, S. Vijayakumar, and K. Nazarpour, "Multi-grip classification-based prosthesis control with two EMG-IMU sensors," *IEEE Trans. Neural Syst. Rehab. Eng.*, vol. 28, no. 2, pp. 508–518, 2019.

[20] P. Xia, J. Hu, and Y. Peng, "EMG-based estimation of limb movement using deep learning with recurrent convolutional neural networks," *Artif. Organs.*, vol. 42, no. 5, pp. E67–E77, 2018.

[21] Y. Hu, Y. Wong, W. Wei, Y. Du, M. Kankanhalli, and W. Geng, "A novel attention-based hybrid cnn-rnn architecture for sEMG-based gesture recognition," *PloS One*, vol. 13, no. 10, p. e0206049, 2018.

[22] U. Côté-Allard, C. L. Fall, A. Drouin, A. Campeau-Lecours, C. Gosselin, K. Glette, F. Laviolette, and B. Gosselin, "Deep learning for electromyographic hand gesture signal classification using transfer learning," *IEEE Trans. Neural Syst. Rehab. Eng.*, vol. 27, no. 4, pp. 760–771, 2019.

[23] S. Pizzolato, L. Tagliapietra, M. Cognolato, M. Reggiani, H. Müller, and M. Atzori, "Comparison of six electromyography acquisition setups on hand movement classification tasks," *PloS One*, vol. 12, no. 10, p. e0186132, 2017.

[24] A. Krasoulis, I. Kyranou, M. S. Erden, K. Nazarpour, and S. Vijayakumar, "Improved prosthetic hand control with concurrent use of myoelectric and inertial measurements," *J. Neuroeng. Rehabil.*, vol. 14, no. 1, pp. 1–14, 2017.

[25] O. W. Samuel, M. G. Asogbon, Y. Geng, X. Li, S. Pirbhulal, S. Chen, N. Ganesh, P. Feng, and G. Li, "Spatio-temporal based descriptor for limb movement-intent characterization in EMG-pattern recognition system," in *Proc. IEEE EMBC*, 2019, pp. 2637–2640.