Kent Academic Repository

Full text document (pdf)

Citation for published version

Lin, Yuzhou and Ramaswamy, Palaniappan and De Wilde, Philippe and Li, Ling (2020) A normalisation approach improves the performance of inter-subject sEMG-based hand gesture recognition with a ConvNet. In: 2020 42nd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC).

DOI

https://doi.org/10.1109/EMBC44109.2020.9175156

Link to record in KAR

https://kar.kent.ac.uk/83026/

Document Version

Author's Accepted Manuscript

A normalisation approach improves the performance of inter-subject sEMG-based hand gesture recognition with a ConvNet

Yuzhou Lin, Ramaswamy Palaniappan, Philippe De Wilde, Ling Li*

Abstract— **Recently, the subject-specific surface electromyography (sEMG)-based gesture classification with deep learning algorithms has been widely researched. However, it is not practical to obtain the training data by requiring a user to perform hand gestures many times in real life. This problem can be alleviated to a certain extent if sEMG from many other subjects could be used to train the classifier. In this paper, we propose a normalisation approach that allows implementing real-time subject-independent sEMG based hand gesture classification without training the deep learning algorithm subject specifically. We hypothesed that the amplitude ranges of sEMG across channels between forearm muscle contractions for a hand gesture recorded in the same condition do not vary significantly within each individual. Therefore, the min-max normalisation is applied to source domain data but the new maximum and minimum values of each channel used to restrict the amplitude range are calculated from a trial *cycle* of a new user (target domain) and assigned by the class label. A convolutional neural network (ConvNet) trained with the normalised data achieved an average 87.03% accuracy on our *G. dataset* (12 gestures) and 94.53% on *M. dataset* (7 gestures) by using the leave-one-subject-out cross-validation.**

I. Introduction and Related Work

Hand gestures, as a natural form of human expression, play an important role in body language. Simply using fingers and arms allows us to create a natural and intuitive interaction modality. Surface electromyography (sEMG) refers to the collective electrical signals from muscles that is usually collected by non-invasive electrodes [1]. The sEMG-based hand gesture classification has great potential to use in the fields of rehabilitation and assistance, for example, voluntarily controlling a robotic exoskeleton system by a stroke patient or manipulating a hand prosthesis for a hand amputee. Recently, the inter-subject sEMG-based hand gesture recognition has drawn more attention since the non-requirement of training data from the user is more attractive in the real-life applications. However, this is a challenging research problem since sEMG signals are highly subject specific [2]. In other words, the difficulty comes from the subject variability, which could also be referred to as the problem of domain shift, i.e., the distribution of the target domain is different from the source domain.

To address this problem, Kerber et al. [3] employed a normalisation approach (w.r.t the highest measured peak) to the segmented sEMG and computed 10 features from the normalised data to train a SVM. An overall 95% accuracy was reported for classifying 5 hand gestures with a majority vote for an active segment. However, the user-independent classification accuracy drops to about 50% when the number of hand gestures is increased to 12. Khushaba [4] proposed a canonical correlation analysis-based framework to compute a series of style-independent features and an average 82.96% accuracy with majority vote was reported for classifying 12 finger movements (including a resting state) by training data from normal subjects and testing with data from amputees.

Recently, many deep learning approaches, such as the Convolutional Neural Network (ConvNet), have been proposed to improve the robustness of a classifier. Furthermore, the transfer learning (TL) has been widely used to solve the domain shift problem in the field of deep learning. Côté-Allard et al. [5] proposed a ConvNet employing raw sEMG as input and a 97.39% accuracy was obtained to classify 7 hand gestures including neutral by applying transfer learning (TL) with four *cycles* of training data from the target domain where a *cycle* is defined as the recording of all hand gestures required to be classified.

In this paper, we propose a referencing min-max normalisation approach to re-weight the source domain sEMG data to simply reduce the domain shift. Three datasets, where differences due to many factors such as recording positions, acquisition setups, the number of hand gestures and subjects, are used to validate the inter-subject gesture recognition performance of our proposed method with the leave-one-subject-out cross-validation (LOSOXV). Finally, the result is compared to the state-of-the-art transfer learning approach [5].

II. EMG Datasets

A. G. Dataset

This dataset has been introduced in [6], [7]. Here we refer this dataset as *G. Dataset* since both non-invasive electrodes and amplifiers used to for recording are from Guger Technologies, Graz, Austria. The data were recorded by 5 intact right-handed subjects with the 16-channel sEMG electrodes placed over 12 muscles located in the upper arm, forearm and hand (details can be found in [6]). During the acquisition, the subjects were asked to repeat 12 hand activities including extension, flexion, pronation, supination, clench, stretch fingers, rotate hand (clockwise), rotate hand (anticlockwise), grip, thumb up, thumb and index up 4 times in two alternating conditions between extension (*G.0*) and relaxation (*G.1*) with the elbow resting on the arm of a chair. Here, each *cycle* includes 12 sEMG onsets, which were determined visually. At the stage of data pre-processing, all sEMG was first down-sampled by the scale of 4, i.e., 1200

*Corresponding author
Y. Lin, R. Palaniappan, P. De Wilde, L. Li are with School of Computing, University of Kent, Kent CT2 7NF, UK {yl339,R.Palani,P.Dewilde,C.Li}@kent.ac.uk

Hz to reduce the processing time but also keep sufficient information since the usable energy of sEMG is mainly distributed from the 0 to 500 Hz in terms of frequency range [8]. Furthermore, a 250 ms sliding window was applied with an overlap of 225 ms (90% $\times$ 250ms) to split the sEMG onsets, which allows to make a prediction under the real-time usage constraints ($\leq$300 ms [9], [10]).

B. M. Dataset

The *M. Dataset* [5] is a recent public sEMG dataset used for classifying hand gestures. In this project, a sub-dataset called 'Test0', refers to as *M.0* here, is used to evaluate the performance of inter-subject sEMG-based hand gesture recognition with our proposed method. The 8-channel sEMG signals were recorded with one Myo armband from 17 intact subjects doing 7 hand gestures (Neutral, Hand close, Wrist extension, Ulnar deviation, Hand open, Wrist flexion, Radial deviation) for 4 times by standing up and having their forearm parallel to the floor. Here the data were segmented by applying a 260ms sliding window with an overlap of 235ms as suggested in [5].

III. Methods

A. Hypotheses

We hypothesized that:

1) The distribution of signal from a channel
   - varies little when an individual is repeating a hand gesture many times;
   - varies a lot across subjects when they are performing the same hand gesture.
2) Considering a hand gesture, the maximum (max) and minimum (min) sEMG values of a channel signal collected from a user could be used to help to shift and rescale the distribution of the same channel data recorded from other subjects.

To further prove our hypotheses, we employed a Gaussian probability density function (PDF), shown in Eq. 1, to present how dense the probability is close to a discrete sEMG magnitude. Furthermore, we analysed the first channel sEMG signal labeled with Gesture 1 (Extension) from *G.1*. Fig. 1 shows that the distributions of the first channel signal collected by the first subject (Sb0) only vary little among different sessions while significant differences of the distributions of the same channel signal across different subjects could be found, shown by the green lines in Fig. 2. It proves our first hypothesis. The red dotted and dashed plots in Fig. 2 show that the differences of the same channel signal distributions across subjects could be reduced by applying normalisation, which proves our second hypothesis. The details of this normalisation approach is presented in Section III-B.

$$f(x) = \frac{e^{-(x-\mu)^2/(2\sigma^2)}}{\sigma\sqrt{2\pi}} \tag{1}$$

where x is a channel sEMG onset, μ and σ are the mean and standard deviation of x.

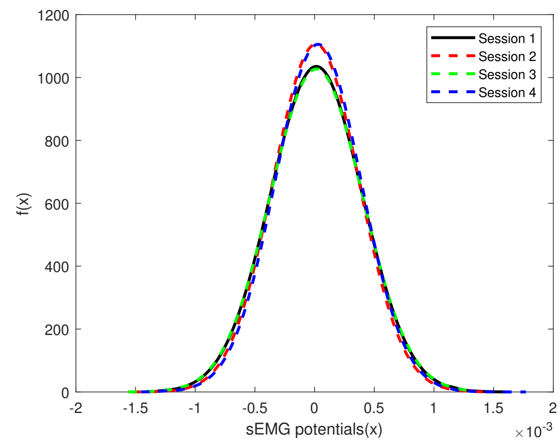

Fig. 1. PDFs of the sEMG from same subject but different sessions

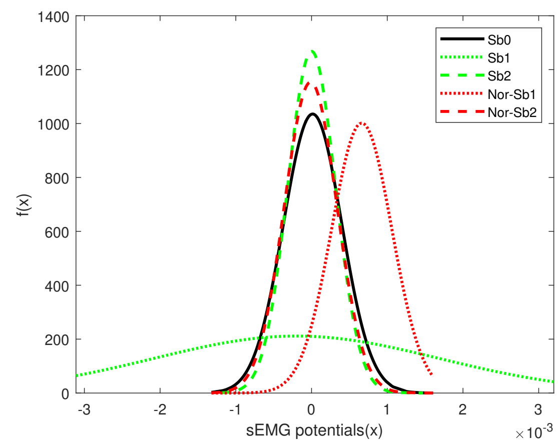

Fig. 2. PDFs of the sEMG from same session but different subjects

B. Referencing normalisation

We propose a referencing normalisation method to reduce the domain shift problem of the inter-subject sEMG-based hand gesture classification. We assume one *cycle* has been collected from a testing subject, which is referred to as the reference data. The max and min sEMG potentials across all channels for each class would then be calculated from this *cycle* shown as $\max(\mathbf{X}^t_{i(k)})$ and $\min(\mathbf{X}^t_{i(k)})$ in Eq. 2. Moreover, all *cycles* from source domain would be normalised according to Eq. 2 to reduce the domain shift.

$$\mathbf{X}^{s'}_{i(k)} = \frac{\mathbf{X}^s_i - \min(\mathbf{X}^s_i)}{\max(\mathbf{X}^s_i) - \min(\mathbf{X}^s_i)} \times (\max(\mathbf{X}^t_{i(k)}) - \min(\mathbf{X}^t_{i(k)})) + \min(\mathbf{X}^t_{i(k)}) \tag{2}$$

where $\mathbf{X}'$ is a normalised sEMG onset; $\mathbf{X}^s$ and $\mathbf{X}^t$ represent a raw sEMG onset in Source and Target domain; i and k stand for the channel number and class label.

C. The ConvNet architecture

The ConvNet architecture implementation (in PyTorch v.1.1.0 and Python 3.7.3) for *M. Dataset* is as described in [5]. It needs to be modified for the *G. Dataset* since the input and output are different. The input data corresponds to 16 $\times$ 300 (i.e. the number of channels $\times$ the number of sample points) and 8 $\times$ 52 while the output refers to the 12 and 7 hand gestures for *G. Dataset* and *M. Dataset* respectively. The details of the modified ConvNet could be found in Table I. It consists of two convolutional layers and one fully connected layer using the recent techniques such as Batch Normalisation (BN) [11], Parametric Rectified Linear Unit (PReLU) activation function [12] and ADAM [13]. The first and second convolutional layer has 32 and 64 filters of size 3 $\times$ 5 respectively. The cross-entropy loss function is used here and the hyperparameters: learning rate, batch size and epoch number are selected as 0.023, 256, 100, respectively.

TABLE I
PROPOSED CONVNET ARCHITECTURE FOR *G. Dataset*

Layer Order - Type	Output Shape	Parameters
L1-Conv2d	[-1, 32, 14, 296]	512
L2-BatchNorm2d	[-1, 32, 14, 296]	64
L3-PReLU	[-1, 32, 14, 296]	32
L4-Dropout2d	[-1, 32, 14, 296]	0
L5-MaxPool2d	[-1, 32, 14, 98]	0
L6-Conv2d	[-1, 64, 12, 94]	30,784
L7-BatchNorm2d	[-1, 64, 12, 94]	128
L8-PReLU	[-1, 64, 12, 94]	64
L9-Dropout2d	[-1, 64, 12, 94]	0
L10-MaxPool2d	[-1, 64, 12, 31]	0
L11-Linear	[-1, 500]	11,904,500
L12-BatchNorm1d	[-1, 500]	1,000
L13-PReLU	[-1, 500]	500
L14-Dropout	[-1, 500]	0
L15-Linear	[-1, 12]	6,012

D. System structure

An overview working scheme for inter-subject sEMG-based hand gesture recognition with a ConvNet using the referencing normalisation approach that we proposed could be seen in Fig. 3. The first step is to calculate the max and min sEMG potentials across all channels for each class from a *cycle* collected from a testing subject. The next step is to normalise each sEMG onset in source domain before applying the sliding window for segmentation. After finishing training with the normalised segmented data, all other *cycles* except the *cycle* 1 in the target domain are simply segmented to test the classification performance with the trained ConvNet.

IV. RESULTS

To show the robustness of the results, we used LOSOXV to evaluate the performance and the average classification accuracy is presented without executing a majority vote. In other words, the recognition accuracy presents how many test segmented windows are classified correctly. Note that, all experiments are reported as an average of 20 runs. We first investigated the performance of inter-subject sEMG-based hand gesture classification with our proposed method according to Fig. 3 on *G.0*, *G.1*, and *M.0*. The results ('RNor') have been compared with using a standard normalisation

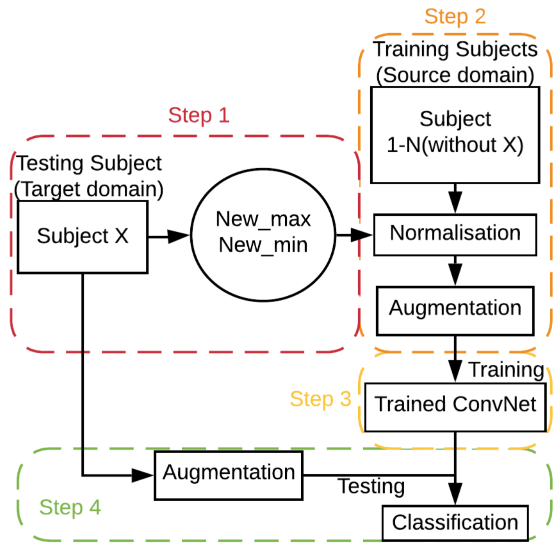

Fig. 3. A high-level diagram illustrating the steps for inter-subject sEMG-based hand gesture classification

('Nor') in Table II. Here a standard normalisation refers to normalising each input data, i.e., each segmented channel sEMG in both source and target domain, to [0,1].

TABLE II
CLASSIFICATION ACCURACY OF THE CONVNETS ON THREE DATASETS WITH RESPECT TO THE NORMALISATION METHODS APPLIED

Dataset		*G.0*	*G.1*	*M.0*
Nor	mean	33.06%	30.56%	73.76%
	std	11.03%	9.61%	9.98%
	p0	4.08e-09	7.11e-09	4.12e-12
RNor	mean	**85.09%**	**88.97%**	**94.53%**
	std	6.96%	7.95%	4.79%
	p0	3.70e-05	1.13e-05	3.80e-16
p1		3.90e-18	3.90e-18	1.78e-57

p0: The p-value for the Shapiro-Wilk test.
p1: The p-value for the non-parametric Wilcoxon signed-rank test.

Since each result deviated from the normal distribution shown using a Shapiro-Wilk test (n = 100, 340 for *G.* and *M. datasets*) where all p-values (p0) were less than 0.05, a non-parametric Wilcoxon signed-rank test was then used to compare the classification accuracies between the normalisation methods. It can be seen that all p-values (p1) are less than 0.05 as in Table II, which implies that the average recognition accuracy using our proposed normalisation was significantly higher compared to standard normalisation across different datasets.

Additionally, we further investigated the performance of gesture classification using the recently proposed transfer learning method [5]. The first three *cycles* from training subjects were used to pre-train the ConvNet and then the last *cycle* was used for the validation. This pre-trained model is called *Source Network*, where the parameters are frozen except for BN parameters to ensure the ability to update domain-related information from the target domain. Furthermore, a *Second Network* was trained by *cycles* from the target subject, and then be merged with the *Source Network* by an element-wise summation. The different number of *cycles* (i.e., *cycle* 1; *cycle* 1, 2; *cycle* 1, 2, 3) from the target subject were used to train the *Second Network* in turn in order to investigate how the classification performance could be affected by the increased training data from the target domain. To test the performance of the trained ConvNet using TL, the *cycle* 4 from the target subject was used.

TABLE III
CLASSIFICATION ACCURACY OF THE CONVNETS ON THREE DATASETS WITH RAW DATA AND TRANSFER LEARNING

No. of Training *Cycles*		1	2	3
G.0	mean	12.33%	63.59%	**73.11%**
	std	6.81%	24.10%	22.37%
G.1	mean	13.7%	44.05%	**92.50%**
	std	8.44%	39.58%	12.50%
M.0	mean	94.61%	97.41%	**98.77%**
	std	5.39%	3.79%	2.36%

It can be noticed from Table III on the better classification accuracy that is achieved with the increasing of the number of training *cycles*. However, with the limited number of training *cycles*, the TL performs poorly on *G. Dataset*. It implies that the TL requires more training *cycles* from the target domain to update domain-related information as the difficulty of hand gesture recognition task increases.

Overall, Table II and Table III show that the best performances achieved on *G.1* are higher than *G.0* regardless of the methods applied. It implies that the muscle variability across subjects could be reduced if all subjects were performing a hand gesture in a stable way, i.e., with their elbow on rather than off the chair arm.

Furthermore, it is difficult to compare our referencing normalisation approach to the state-of-the-art TL directly due the different number of training and testing *cycles* involved even for the same dataset. However, in terms of the inter-subject sEMG-based hand gesture recognition, applying our referencing normalisation approach could achieve similar results when three *cycles* from the target subjects are used in the training process with the state-of-the-art TL.

V. CONCLUSION AND FUTURE WORK

This paper presented a referencing normalisation approach to reduce domain shift when considering the inter-subject sEMG-based hand gesture recognition. The performance of classification using our method was investigated with a ConvNet on three sub datasets by LOSOXV. The 85.09%, 88.97% and 94.53% subject-independent classification accuracy were achieved on *G.0* and *G.1* when predicting 12 hand gestures not including rest and on *M.0* to classify 7 hand gestures including rest. Additionally, we applied a state-of-the-art transfer learning method to investigate its ability to solve the domain shift problem. Our results showed that it required only one training *cycle* on *M.0* but three training *cycles* on *G.1* for TL to update the domain-related information so that achieving better classification accuracy than our proposed normalisation method.

However, inter-subject sEMG-based hand gesture recognition usually does not practically allow the target domain data to be involved in the training process. Compared to the TL, our proposed referencing normalisation method only requires one *cycle* from a user to extract a little prior information but not used for training. This is easy to implement with a low computation cost. Results showed that generally high inter-subject sEMG-based hand gesture recognition accuracy could be obtained in terms of different number of hand gestures, subjects, recording positions and acquisition setups. Nevertheless, our method may not work well if a user is performing a hand gesture in a different way in real world. In the future, we would test the application of our proposed normalisation method for inter-session recognition to investigate the robustness of long-term hand gesture recognition.

REFERENCES

[1] R. Chowdhury, M. Reaz, M. Ali, A. Bakar, K. Chellappan, and T. Chang, "Surface electromyography signal processing and classification techniques," *Sensors*, vol. 13, no. 9, pp. 12431–12466, sep 2013.

[2] M. Halaki and K. Ginn, "Normalization of emg signals: To normalize or not to normalize and what to normalize to?," in *Computational Intelligence in Electromyography Analysis*, Ganesh R. Naik, Ed., chapter 7, p. 175. IntechOpen, 2012.

[3] F. Kerber, M. Puhl, and A. Krüger, "User-independent real-time hand gesture recognition based on surface electromyography," in *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services*, New York, NY, USA, 2017, MobileHCI '17, pp. 36:1–36:7, ACM.

[4] R. N. Khushaba, "Correlation analysis of electromyogram signals for multiuser myoelectric interfaces," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 22, no. 4, pp. 745–755, 2014.

[5] 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 Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, no. 4, pp. 760–771, April 2019.

[6] Y. Lin, P. De Wilde, R. Palaniappan, and L. Li, "Muscle connectivity analysis for hand gesture recognition via semg," in *2018 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, Nov 2018, pp. 848–852.

[7] L. Li, D. Looney, C. Park, N. U. Rehman, and D. P. Mandic, "Power independent emg based gesture recognition for robotics," in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Aug 2011, pp. 793–796.

[8] J. Wang, L. Tang, and J. E. Bronlund, "Article: Surface emg signal amplification and filtering," *International Journal of Computer Applications*, vol. 82, no. 1, pp. 15–22, November 2013, Full text available.

[9] B. Hudgins, P. Parker, and R. N. Scott, "A new strategy for multifunction myoelectric control," *IEEE Transactions on Biomedical Engineering*, vol. 40, no. 1, pp. 82–94, Jan 1993.

[10] K. Englehart and B. Hudgins, "A robust, real-time control scheme for multifunction myoelectric control," *IEEE Transactions on Biomedical Engineering*, vol. 50, no. 7, pp. 848–854, July 2003.

[11] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *CoRR*, vol. abs/1502.03167, 2015.

[12] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," *CoRR*, vol. abs/1502.01852, 2015.

[13] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," *International Conference on Learning Representations*, 12 2014.