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

sEMG-Based Hand Movement Regression by Prediction of Joint Angles With Recurrent Neural Networks

Philipp Koch, Kamran Mohammad-Zadeh, Marco Maass, Mark Dreier,
Ole Thomsen, Tim J. Parbs, Huy Phan, and Alfred Mertins

Abstract— **This work takes a step towards a better biosignal based hand gesture recognition by investigating the strategies for a reliable prediction of hand joint angles. Those strategies are especially important for medical applications in order to achieve e.g. good acceptance of hand prostheses among amputees. A recurrent neural network with a small footprint is deployed to estimate the joint positions from surface electromyography data measured at the forearm. As the predictions are expected to be not smooth, different post processing methods and a regularisation term for the objective function of the network are proposed. The experiments were conducted on publicly available databases. The results reveal that both post processing strategies and regularisation have a positive impact on the results with a maximal relative improvement of $6.13\,\%$. On the one hand post processing strategies introduce an additional delay, consequently, the improvement is analysed in context of the caused delay. On the other hand the regularisation strategy does not cause a delay and can be adjusted easily to cope with different ground truths or compensate for certain problems in the hand tracking.**

I. Introduction

The advantages of a hand gesture recognition system based on biosignals are multifarious. Such a system could easily be used in mobile devices and provides an intuitive human-machine interface that can of course be used in numerous applications [1], [2]. One of the most obvious and probably most important application areas are hand prosthetic and exoskeletons [3], [4]. Furthermore, a hand gesture recognition system can also be used to interact with a virtual reality or to control exoskeletons, and in many other use cases.

In the context of this publication, we focus especially on medical applications, because that is where the requirements and also the benefits for the users are the greatest. In general, the requirements on gesture recognition systems are very similar. To allow for a satisfying user experience three major aspects have to be covered. First, the usage has to be intuitive for the user, meaning the amputees do not have to learn certain muscular contraction patterns to control their prosthesis, but they can rather perform the natural gesture and the system recognises it. Secondly, the response time and the introduced delay should be unnoticeable or at least relatively small. A delayed reaction of a human machine interface leads to low level of acceptance among the users as it reduces the usability significantly. Finally, performed gestures have to be recognised robustly and accurately.

Usually, hand gesture recognition systems solve a classification problem, meaning that only a limited number of gestures can be recognised. In recent years the number of recognisable gestures got as high as 50 to allow for a more intuitive control by allowing the user to choose from a large set of gestures. With the introduction of deep learning methods it became possible to distinguish between this many gestures more and more reliably. In recent years the hand crafted feature extraction as used in [5], [6], [7] were replaced by one trained in data driven fashion. By using small recurrent neural networks (RNNs) [8], [9], [10] or significantly larger convolutional neural networks (CNNs) [11], [12]. With the small RNNs, it became possible to reduce the delay while achieving remarkable classification accuracies even for a large number of different gestures.

Consequently, the next step to improve intuitive control and usability of such a human machine interface – and the aim of this work – is to replace the classical classification by a regression scheme, meaning the system will not distinguish between predetermined gestures but predict the angle of each join. As a result, the user can (hopefully) move the hand freely while the system still predicts its movement correctly.

Unsurprisingly, the tool used for the prediction is also a network. The approaches based on rather small RNNs achieve state-of-the-art performance, are suitable for deployment on embedded systems (in contrast to the CNN based networks with their significantly larger footprints), and also have shown their general suitability for regression tasks even though only voltage signals instead of join angles were predicted [13]. Consequently, the proposed network is inspired by those small RNNs and features just a single RNN cell followed by one dense/fully-connected layer.

The networks predictions are expected to be rather noisy resulting in joint angle sequences that cannot match the smoothness of biological movements. Therefore, different post processing strategies are introduced. One strategy utilises a median filter while the other is based on a total-variation approach.

Experiments were conducted on the publicly available databases one and two of the Ninapro project [5]. Note that

Philipp Koch, Mark Dreier, and Alfred Mertins are with the German Research Center for Artificial Intelligence (DFKI), AI in Biomedical Signal Processing, 23562 Lübeck, Germany {philipp.koch, mark.dreier, alfred.mertins}@dfki.de

Marco Maass, Ole Thomsen, Tim J. Parbs, and Alfred Mertins are with the Institute ,for Signal Processing, University of Lübeck, 23562 Lübeck, Germany {marco.maass, o.thomsen, t.parbs, alfred.mertins}@uni-luebeck.de

Kamran Mohammad-Zadeh is with the University of Lübeck, 23562 Lübeck, Germany kamran.mohammadzadeh@student.uni-luebeck.de

Huy Phan is with the School of Electronic Engineering and Computer Science, Queen Mary University of London, Bethnal Green, London, E1 4NS, United Kingdom h.phan@qmul.ac.uk

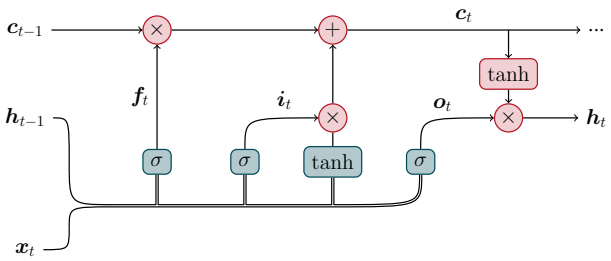

Fig. 1. Schematic illustration of the basic principle of an LSTM cell.

the sampling frequencies of the two databases are different. Database IX contains the corresponding joint angles needed for the regression [14]. The predictions, the effects of the different post processing approaches, as well as potential weaknesses of the ground truth are discussed. The results reveal the effect of post processing and the quality of the predictions.

II. The Regression Problem

In this work the joint angles of a hand should be regressed given sEMG signals. As most of the potential applications require a prediction of every joint angle for every window, a sequence-to-sequence formulation of the problem is chosen. This means, before processing the next window, all joint angles for the current one have to be predicted, and for the prediction, information of the predecessor windows can be used. Let us assume $\boldsymbol{x}_t \in \mathbb{R}^D$ is the D-dimensional representation (possibly the raw signal) of the window corresponding to the current time step t. To leverage the sequential nature of the data the regression problem is stated as follows

$$\mathcal{R} : (\boldsymbol{x}_t \in \mathbb{R}^D, \boldsymbol{x}_{t-\eta}, \boldsymbol{x}_{t-2\eta}, \cdots) \longmapsto \boldsymbol{y}_{t+\tau} \tag{1}$$

where η denotes the step size and τ the prediction offset. In this work η is chosen to be equal to the window length resulting in non-overlapping consecutive windows between which no information is lost. The prediction offset τ is set to half of the window length meaning the predictions correspond to the end of the current window rather than to the center. This formulation allows for the use of RNNs and enables the regressor to try to compensate for positive or negative time delays of the signal compared to the hand movement.

As the human motions are assumed to be smooth, a post processing step is introduced since the predictions are expected to be rather noisy and not smooth. Let $L \in \{2k+1 | k \in \mathbb{N}_0\}$ be the filter length used for the median filtering then the smoothing is given by

$$\mathcal{P} : (\boldsymbol{y}_{t-\frac{L-1}{2}\eta}, \boldsymbol{y}_{t-\frac{L-2}{2}\eta}, \ldots, \boldsymbol{y}_t, \ldots, \boldsymbol{y}_{t+\frac{L-1}{2}\eta}) \longmapsto \boldsymbol{y}_t. \tag{2}$$

Note that such a post processing step introduces an delay of $\frac{L-1}{2}\eta$ samples.

III. Proposed Network

A. Architecture

The network used in this paper features the long short-term memory (LSTM) cell [15]. Fig. 1 illustrates the general principle of this RNN cell. Characteristic for the LSTM cell is its state that allows information to persist over time within the cell. The state is updated in each time step via the forget gate $\boldsymbol{f}$ and input gate $\boldsymbol{i}$. Together with the so-called output gate $\boldsymbol{o}$ the modified state is used to calculate the cells output.

In [13] a simple network has proven its suitability for regression. Hence, in this work a network with just two layers is used. The first layer, a LSTM cell, performs the time variant data processing while the final fully connected layer does the mapping necessary for the actual regression of the different joint angles.

B. Loss Function

As objective function the Huber loss was used. It is a combination of ℓ_1 and ℓ_2 loss:

$$L_\delta(\hat{\boldsymbol{y}}_t^j, \boldsymbol{y}_t^j) = \begin{cases} \frac{1}{2}(\hat{\boldsymbol{y}}_t^j - \boldsymbol{y}_t^j)^2, & |\hat{\boldsymbol{y}}_t^j - \boldsymbol{y}_t^j| \leq \delta \\ \delta|\hat{\boldsymbol{y}}_t^j - \boldsymbol{y}_t^j| - \frac{1}{2}\delta, & \text{otherwise} \end{cases} \tag{3}$$

The overall loss of a training sequence is calculated by the summation of all joints and all time steps. In the following this sum will be referred to as the value of the loss.

C. Optimization

To train the network the well-known Adam optimizer [16] was used. To improve the training procedure mini-batches were used as well as dropouts.

IV. Postprocessing

Since initial experiments revealed that a sequence of predicted joint angles is not smooth even though the RNN performed the predictions in a sequence-to-sequence manner, a post-processing is applied. After studying the provided ground truth two rather simple techniques were chosen: the median filter, and a one-dimensional total variation (TV) denoising approach [17]. Again, for simplicity, each joint is treated independently, though with the same hyperparameters.

Since the median filtering is covered in (2), in the following only the TV denoising step is briefly introduced. For any discretized time signal $s(n)$ of length N, the optimal TV-denoised solution $\hat{s}(n)$ can be found by

$$\underset{\hat{\boldsymbol{s}} \in \mathbb{R}^N}{\text{minimize}} \quad \frac{1}{2}\sum_{n=0}^{N}(s(n) - \hat{s}(n))^2 + \beta \sum_{n=0}^{N-1} |\hat{s}(n+1) - \hat{s}(n)|, \tag{4}$$

where $\beta > 0$ denotes the regularisation parameter. An optimal solver can be found for the problem in (4) by using the algorithm from [17]. The optimization problem can require up to the entire sequence. Thus, the concept of delay of conventional filter cannot easily be transferred to TV denoising. However, as mentioned, the ground truth signals appear to have a small total variation making this TV denoising a near optimal solution for post processing.

In order to find the suitable parameters for the two methods, a grid search is performed. In the case of the median filter, only the filter length is varied, while in the case of TV denoising, the hyperparameter β was varied in the range of $[10^{-2}, 10^3]$.

TABLE I

HYPERPARAMETERS OF THE NETWORK FOUND VIA GRID SEARCH.

δ	λ	Batch size	State size	Learning rate	Dropout rate
1.0	10^{-6}	200	128	0.0008	0.5

V. OBJECTIVE FUNCTION WITH REGULARISATION

As the TV denoising is expected to achieve the best results in post processing, it should be investigated whether a similar constraint can be introduced to the loss function for the network. The advantage of such a regularisation is that unlike the TV denoising and median filtering no additional delay is introduced. Consequently, the Huber loss is extended by a regularisation term

$$L^R_{(\delta,\lambda)}((\hat{\boldsymbol{y}}^j_t, \boldsymbol{y}^j_t) = L_\delta(\hat{\boldsymbol{y}}^j_t, \boldsymbol{y}^j_t) + \lambda||\nabla\hat{\boldsymbol{y}}^j_t||_1 \tag{5}$$

with $\lambda \in \mathbb{R}_+$ being a weighting factor. The regularisation term, called estimated gradient regularisation (EGR), should penalise rapid unnatural motion predictions. In this study the norm of the gradient was approximated by

$$||\nabla\hat{\boldsymbol{y}}^j_t||_1 = ||\hat{\boldsymbol{y}}^j_{t-1} - \hat{\boldsymbol{y}}^j_{t+1}||_1. \tag{6}$$

Note that the approximation of the gradient is different form those in (4). For the regularisation the central difference quotient was used as the gradient should be estimated for a sampled value and not between two sampled values.

VI. EXPERIMENTS

The experiments were conducted using dabases DB I and DB II of the ninapro project [5] whereas the ground truth could be found in DB IX [14]. The raw EMG data were fed into the network. In case of DB I the window length and the hop was set to $100\,\mathrm{ms}$ while for DB II $5\,\mathrm{ms}$ were used. The prediction offset τ was set to half of the window size to avoid delays. The network's parameters, found via a quick grid search, are shown in Table I. In training, sequences with a fixed length of $1\,\mathrm{s}$ were used and the network was optimised for 20 epochs. As an error measure the mean absolute error (MAE) was used. If not stated otherwise the reported results are averaged along all joints and across all subjects of one database.

A. Results

The naive network trained with the Huber loss (3) can predict the joint angles with an average MAE error of $6.93°$ and $5.72°$ for DB I and DB II, respectively.

In Table II the obtained results for the various smoothing techniques for DB I are shown. The network optimised with the EGR achieves a slightly better MAE compared to the default one that does not use any post-processing or regularisation. The median filtering on the other hand leads to an improved MAE at the cost of an additional delay of $700\ \mathrm{ms}$ when aiming for the biggest improvement. Also, on DB II the usage of EGR outperforms the default method without causing additional delay as can be seen in Table III. Here, as well, the median filtering leads to the best result at the cost of an additional delay of $625\ \mathrm{ms}$. As expected the TV denoising leads to the best result for both databases (see Tables II and III). However, as mentioned, the results cannot be compared with those for the other methods as the delay is uncertain, making the TV denoising impractical for real world applications. Note, to the best of our knowledge there are no other publications to compare with. But as the MAE is well below $7°$ which corresponds to a displacement of just a couple of millimeters depending on hand size, the experiments let us conclude that regressing the hand movement is possible even with decent accuracy.

TABLE II

RESULTS OF THE PROPOSED SMOOTHING TECHNIQUES FOR DB1.

	Default	Median filter	TV denoising	EGR
MAE in degrees	6.93°	6.66°	6.56°	6.87°
Rel. MAE decrease	–	4.05 %	5.64 %	0.87 %
Additional delay	0 ms	700 ms	–	0ms

TABLE III

RESULTS OF THE PROPOSED SMOOTHING TECHNIQUES FOR DB2.

	Default	Median filter	TV denoising	EGR
MAE in degrees	5.72°	5.41°	5.39°	5.53°
Rel. MAE decrease	–	5.73 %	6.12 %	3.43 %
Additional delay	0ms	625 ms	–	0 ms

For both databases in Fig. 2 the relation between MAE and delay is illustrated. To put things in perspective the results for the default network, the EGR, and TV denoising are added as a constant line. The general behaviour is similar for both databases. Up to a delay of $700\,\mathrm{ms}$, the more precises regression result can be trade of against an increased delay. Beyond the $700\,\mathrm{ms}$ mark the results decrease again. This makes sense as after nearly three quarters of a second, the hand movement has changed significantly. The EGR provides the better results if a delay lower than about 100 ms and 200 ms is allowed for DB I and DB II, respectively. Consequently, as in real-world applications the acceptable delays are rather small, EGR appears to be the proper choice for such use cases.

To illustrate the effects of the different smoothing methods in Fig. 3 the true joint angles as well as the predicted joint angles using various methods over time are shown, exemplary for one randomly chosen sensor. As expected, the predicted movement of default network is rather noisy. Both the median post processing and the EGR leads to significantly smoother prediction curves that appear to be closer to a natural movement. The TV denoising leads to predictions with characteristics close to those of the ground truth. In a nutshell, a post processing can (obviously) alter the prediction but also a regularisation term added to the objected function can be used to train the network.

B. Ground Truth – Bringing it in perspective

With the results in mind, a brief discussion of the ground truth is necessary. The general shape of the curve is not

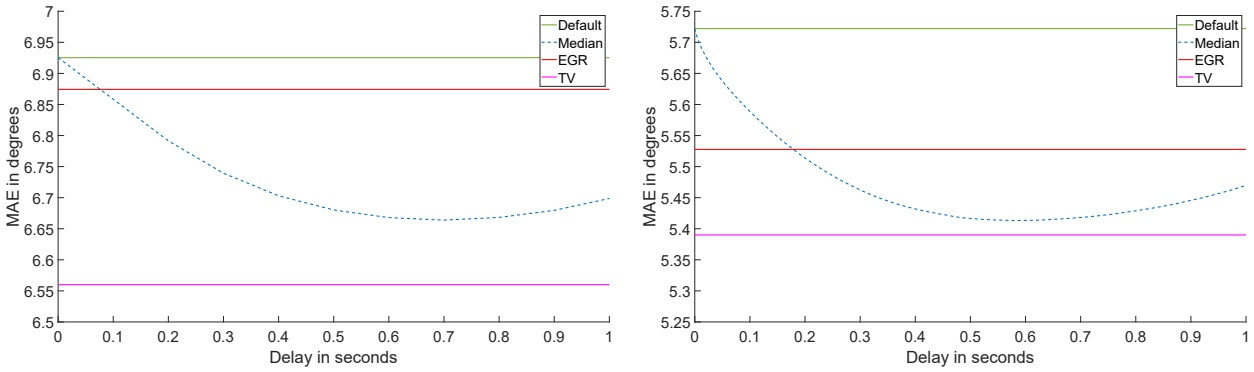

Fig. 2. MAE on test data over the corresponding delay. The MAE curves for DB I are displayed in (a), for DB II in (b). As the default and median have no additional delay and the delay of TV cannot be determined a priori, the results are plotted as horizontal lines.

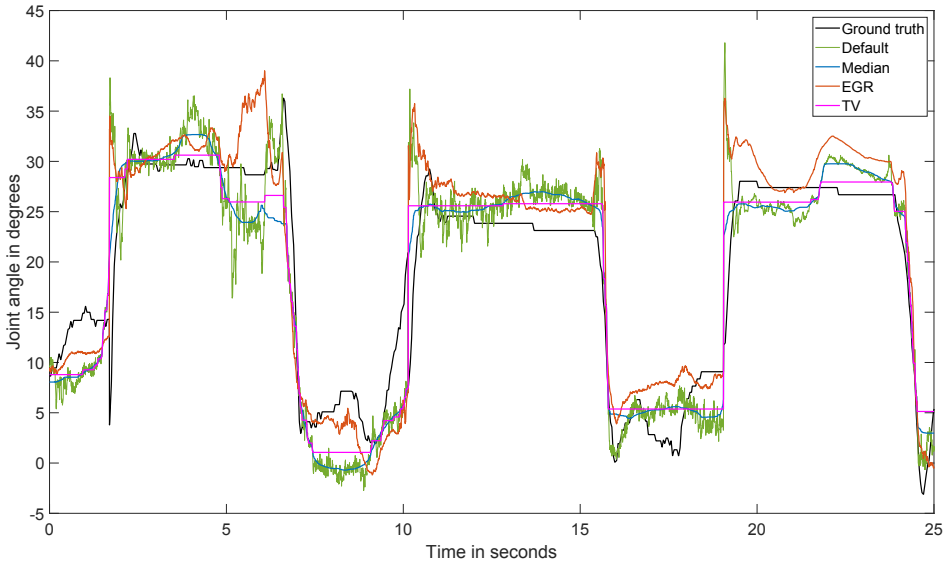

Fig. 3. Exemplary illustration of ground truth and corresponding predictions (using best parameters for each method) from DB II. Note that the results of median are shown without their offsets.

what the authors of this paper would expect. Intuitively, the changes in a natural movement are not sudden so that the curve should be smooth and free of rapid changes in joint angles. It would lead too far to reason about why the ground truth has its characteristics. But nevertheless by adjusting the methods used for regularisation and post processing to the characteristics of alternative hand tracking data should be possible. As well, it can be assumed that the prediction quality would improve as there is potentially a stronger correlation between the EMG signal and the data to be regressed. Also a wisely chosen regularisation or post processing might even be capable of compensating some of the shortcommings of the hand gesture recognition/ground truth.

VII. Conclusion

Generally, it is possible to estimate the joint angles given sEMG data surprisingly well when using simple RNN. With an average error less than 7° the positional error is, depending on the individual hand size, just a few millimeters. With post processing, especially with the TV denoising, even lower MAEs can be achieved at the cost of a significant delay. In contrast the regularisation added to the networks' loss function does not add a delay but still improves the predictions. Like the post processing, the regularisation can be varied in order to achieve a certain behaviour of the output. As this work focused mostly on minimising the error and considered the naturalness of movement only marginal, future research could focus on that in order to improve results and compensates for problems with ground truth and hand tracking.

Acknowledgement

This work has been supported by the KI-LAB Lübeck funded by the Federal Ministry of Education and Research (BMBF) under the Grant No. 01IS19069.

REFERENCES

[1] J. Cheng, X. Chen, Z. Lu, K. Wang, and M. Shen, "Key-press gestures recognition and interaction based on sEMG signals," in *Proc. Int. Conf. Multimodal Interact. and Mach. Learn. Multimodal Interact.*, 2010.

[2] F. Muri, C. Carbajal, A. M. Echenique, H. Fernández, and N. M. López, "Virtual reality upper limb model controlled by EMG signals," *J. Phys. Conf. Ser.*, vol. 477, 2013.

[3] J. Rosen, M. Brand, M. B. Fuchs, and M. Arcan, "A myosignal-based powered exoskeleton system," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 31, no. 3, pp. 210–222, 2001.

[4] K. Kiguchi and Y. Hayashi, "An EMG-based control for an upper-limb power-assist exoskeleton robot," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 42, no. 4, pp. 1064–1071, 2012.

[5] M. Atzori, A. Gijsberts, C. Castellini, B. Caputo, A.-G. Mittaz Hager, S. Elsig, G. Giatsidis, F. Bassetto, and H. Müller, "Electromyography data for non-invasive naturally-controlled robotic hand prostheses," *Sci. Data*, vol. 1, no. 140053, 2014.

[6] X. Zhang, X. Chen, Y. Li, V. Lantz, K. Wang, and J. Yang, "A framework for hand gesture recognition based on accelerometer and EMG sensors," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 41, no. 6, pp. 1064–1076, 2011.

[7] K. Englehart and B. Hudgins, "A robust, real-time control scheme for multifunction myoelectric control," *IEEE Trans. Biomed. Eng.*, vol. 50, no. 7, pp. 848–854, 2003.

[8] P. Koch, H. Phan, M. Maass, F. Katzberg, and A. Mertins, "Recurrent neural network based early prediction of future hand movements," in *Proc. IEEE Eng. Med. Biol. Soc. (EMBC)*, July 2018.

[9] P. Koch, H. Phan, M. Maass, F. Katzberg, R. Mazur, and A. Mertins, "Recurrent neural networks with weighting loss for early prediction of hand movements," in *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, September 2018.

[10] P. Koch, M. Dreier, M. Maass, H. Phan, and A. Mertins, "Rnn with stacked architecture for semg based sequence-to-sequence hand gesture recognition," in *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, January 2021.

[11] M. Atzori, M. Cognolato, and H. Müller, "Deep learning with convolutional neural networks applied to electromyography data: A resource for the classification of movements for prosthetic hands," *Front. Neurorobot.*, vol. 10, no. 9, 2016.

[12] 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. 36571, 2016.

[13] P. Koch, M. Dreier, A. Larsen, T. J. Parbs, M. Maass, H. Phan, and A. Mertins, "Regression of hand movements from semg data with recurrent neural networks data," in *Proc. IEEE Eng. Med. Biol. Soc. (EMBC)*, July 2020.

[14] N. J. Jarque-Bou, M. Atzori, and H. Müller, "A large calibrated database of hand movements and grasps kinematics," *Scientific Data*, 2019 (submitted).

[15] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[16] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. ICLR*, December 2014.

[17] Laurent Condat, "A direct algorithm for 1-D total variation denoising," *IEEE Signal Processing Letters*, vol. 20, no. 11, pp. 1054–1057, 2013.