

# Hand Gesture Recognition Based on Surface Electromyography

Ali-Akbar Samadani<sup>1</sup> and Dana Kulić<sup>1</sup>

**Abstract**—Human hands are the most dexterous of human limbs and hand gestures play an important role in non-verbal communication. Underlying electromyograms associated with hand gestures provide a wealth of information based on which varying hand gestures can be recognized. This paper develops an inter-individual hand gesture recognition model based on Hidden Markov models that receives surface electromyography (sEMG) signals as inputs and predicts a corresponding hand gesture. The developed recognition model is tested with a dataset of 10 various hand gestures performed by 25 subjects in a leave-one-subject-out cross validation and an inter-individual recognition rate of 79% was achieved. The promising recognition rate demonstrates the efficacy of the proposed approach for discriminating between gesture-specific sEMG signals and could inform the design of sEMG-controlled prostheses and assistive devices.

## I. INTRODUCTION

Conventionally, the work on automatic human gesture recognition is based on observable manifestations of the gestures represented in terms of kinematic trajectories, which are captured by video cameras and/or optical motion capture systems. Advancements in physiological sensing technology have enabled access to underlying electrical muscular activities (surface Electromyography (sEMG)) associated with a displayed gesture. Therefore, sEMG provides an alternative source for automatic recognition of bodily movements including hand gestures. The automatic recognition of hand gestures based on sEMG signals is a promising approach for controlling prostheses and in rehabilitation applications (e.g., [1]–[4]).

The majority of the work on sEMG-based gesture recognition to date is done based on user-specific training of the classifier (e.g., [2], [5]), and/or requiring at least one training input from every new user [6], [7]. Inter-individual recognition of hand gestures from sEMG signals is difficult due to the noisy nature of sEMG signals, differences in sEMG sensor placement and contact conditions, and inter-individual differences in performing a gesture. Furthermore, sEMG signals belong to the class of sequential observations and therefore, there are phase and length variations between different sEMG patterns associated with a gesture, which impede the application of feature-based discriminative techniques. To address the first problem (noisiness of sEMG), in this work, the sEMG signals are preprocessed to obtain their smoothed shape (envelopes). To account for the interpersonal differences, a hidden Markov model (HMM) with a mixture of Gaussian outputs is proposed to enable encoding gestures with a multi-modal distribution in sEMG space. In such a

modeling, exemplars from a gesture belonging to different modes (kinematic or kinetic) of that gesture are assigned to distinct mixtures. Another advantage of HMMs is their robustness to phase and/or length differences, eliminating the need for length normalization and landmark alignment for the sEMG signals. The proposed approach constructs gesture-specific HMM models and uses them to perform maximum likelihood recognition for a given gesture. The performance of the proposed approach is evaluated with leave-one-subject-out cross validation using a dataset of various hand gestures performed by 25 different subjects.

## II. METHODOLOGY

The raw sEMG signals are noisy and contaminated by environmental, measurement, and motion artifacts. The sEMG signals also include significant spatial-temporal variability, which impedes feature-based discriminative analysis due to variable-length and spatially-unaligned signals. In addition, there are subject-specific sources of variabilities (e.g., subcutaneous fat, muscle fibre composition, amount of hair [8]) in sEMG signals that need to be accounted for to enable inter-individual recognition.

### A. Preprocessing

Signal envelopes are first generated by: 1) centering raw sEMG signals to remove any DC offset in the signals, 2) rectifying the centered signals (taking the absolute value of the signals), 3) low-pass filtering of the centered and rectified signals [9](see Fig. 1).

Next, the resulting sEMG envelopes are segmented to extract the active region of the signals. In this work, an adaptive segmentation approach is applied, which computes an amplitude threshold for each combination of sEMG channel, gesture, and subject. The amplitude threshold for  $c^{th}$  channel in the  $k^{th}$  gesture performed by  $s^{th}$  subject is computed as the ratio of the maximum of the sEMG signal in the  $c^{th}$  channel to the  $s^{th}$  subject rest gesture median in the  $c^{th}$  channel. Therefore, the segmentation requires a rest gesture for each subject. The resulting thresholds are then used to segment the active regions of the sEMG signals.

### B. HMM-based gesture modeling and recognition

HMM models a sequential observation as a stochastic process whose dynamics are described by a discrete hidden state variable. The hidden state varies between  $N$  hidden state values based on a state transition matrix  $A$  of size  $N \times N$ . The observation variables are described by a vector of size  $C$  ( $C$  represents the number of sEMG channels in our case). The distribution of the observations for each hidden state is modeled as a mixture of  $M$  multivariate Gaussians

<sup>1</sup>A Samadani and D. Kulić are with the Electrical and Computer Engineering Department at the University of Waterloo, Waterloo, ON N2L 3G1 CANADA (e-mail: asamadani@uwaterloo.ca, dkulic@uwaterloo.ca).

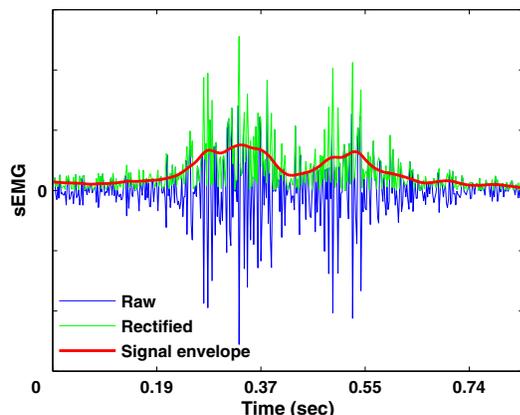


Fig. 1. An example of sEMG envelopes.

and is denoted as  $\mathfrak{B}$ . For a gesture  $\mathbf{O}$  of length  $T$ , the observation probability for the  $i^{\text{th}}$  HMM state is defined as

$$b_i(\mathbf{O}_t) = \sum_{j=1}^M w_{ij} N(\mathbf{O}_t, \mu_{ij}, \Sigma_{ij}) = \sum_{j=1}^M w_{ij} \frac{1}{(2\pi)^{C/2} |\Sigma_{ij}|^{1/2}} e^{-1/2(\mathbf{O}_t - \mu_{ij})^T \Sigma_{ij}^{-1} (\mathbf{O}_t - \mu_{ij})}, \quad (1)$$

where  $\mathbf{O}_t$  is the gesture at time  $t$ , and  $w_{ij}$ ,  $\mu_{ij}$ , and  $\Sigma_{ij}$  are the weight, mean, and covariance of the  $j^{\text{th}}$   $C$ -dimensional Gaussian at state  $i$ . Furthermore, there is an initial state probability  $\pi_i \mathbb{1}_{i=1}^N$  for each hidden state. Therefore, an HMM model  $\lambda$  consists of  $\lambda(A, \mathfrak{B}, \pi)$ . Efficient algorithms exist for estimating the model parameters  $A$ ,  $\mathfrak{B}$ , and  $\pi$  (e.g., the Baum-Welch algorithm, an expectation-maximization algorithm) and evaluating the likelihood that a new observation was generated from the model (e.g., the forward algorithm). A detailed review of HMMs can be found in [10].

In this work, an HMM with a mixture of Gaussian outputs is trained to encode sEMG signals associated with each hand gesture (gesture-specific HMM). Different individuals might perform a gesture differently. These differences could be the result of kinematic differences in performing a gesture or individual differences in muscle recruitment. Furthermore, different individuals might exert a different amount of force in performing a hand gesture (kinetic differences e.g., a weak fist versus a strong fist), which might result in different sEMG patterns associated with a gesture. The mixture of Gaussian outputs used in the gesture-specific HMMs allows for encoding different modes of a gesture in a single HMM; observations belonging to different modes of a gesture populate distinct Gaussians at each hidden state in the associated HMM.

The trained gesture-specific HMMs are then used to perform maximum-likelihood classification (ML) of the gestures. This is done by computing the likelihood that a testing gesture  $\mathbf{O}$  is generated by the trained HMMs ( $\lambda_k, k = 1, \dots, K$ , for  $K$  gestures) and classifying the testing gesture as belonging to the gesture class with the highest likelihood ( $\arg \max_{1 \leq k \leq K} P(\mathbf{O} | \lambda_k)$ ).

### III. EXPERIMENTAL SETUP

The performance of the proposed recognition approach is evaluated with leave-one-subject-out cross-validation (LOSOCV). In each fold of LOSOCV, a subject is left out (testing subject) and the models are trained using the remaining subjects (training subjects) and then, the trained models are tested using the testing subject.

#### A. Dataset

A prototype device from Thalmic Inc.<sup>1</sup> containing 8 sEMG sensors was used to capture the sEMG signals from the surface of the forearm of 25 subjects, while performing a set of 21 hand gestures<sup>2</sup>. The sEMG sensors are arranged side-by-side along an armband that wraps around the subject's forearm, so that there is no specific assignment between a particular sEMG sensor and a muscle or muscle group. The subjects were asked to display each gesture within 2 seconds after an audio cue was played. Five trials of each gesture were collected from each subject.

#### B. HMM Initialization and model selection

To reduce subject-specific variabilities and enhance inter-individual recognition, the signal envelopes for a gesture performed by a subject are normalized by the maximum of all trials of the gesture by that subject. Since all the gestures start from a known pose (resting position), the hidden state sequence for the HMMs always starts at state 1; hence, the hidden state priors:  $\pi_i = 1$  for  $i = 1$ , and  $\pi_i = 0$ , otherwise.

To initialize the training process, the signal envelopes are divided into  $N$  (number of hidden states) equal segments, and  $k$ -means clustering is run within each segment to identify  $M$  clusters ( $M$  mixtures). Then, the means and covariances of the identified clusters within each segment (hidden state) are used as the initial means and covariances of the output mixture of Gaussians associated with that hidden state. Models with full and diagonal covariances at output Gaussians are tested. In each segment, a mixture weight is initialized as the ratio of data points in the cluster to the total number of data points in that segment. We have tested models with left-to-right and full transition matrices (ergodic models) and found that left-to-right models perform better on average. This is due to the progressing nature of the gestures from their start points to end points without cyclic transitions.

The choice of HMM parameters, particularly the number of mixtures, is based on the dataset to be modeled, and therefore must be experimentally selected. The presented recognition approach is therefore suitable for generalization as the number of mixtures can be adapted based on the data. In this work, the best configuration of the HMMs in terms of number of states and mixtures is selected to optimize LOSOCV testing gesture ML recognition.

<sup>1</sup>Thalmic Labs, [www.thalmic.com](http://www.thalmic.com)

<sup>2</sup>Rest, fist, paddle in, paddle out, snap, spider man, gun, thumb and index finger tap, thumb and middle finger tap, thumb and ring finger tap, thumb and pinky finger tap, point with index, point with index and middle, point with index-middle and ring, thumbs up, talking hand, thumb flick, index finger flick, middle finger flick, ring finger flick, pinky finger flick.

#### IV. RESULTS AND DISCUSSION

The best HMM configuration is a left-to-right HMM with seven states and five mixtures with diagonal covariance for Gaussian outputs and results in a LOSOCV recognition rate of  $49\% \pm 6\%$ <sup>3</sup>. The recognition rate is above chance (4.8%) indicating that the proposed approach is capable of distinguishing gestures based on their sEMG. However, we observed that some gestures are better discriminated than others. For instance, there is a large number of confusions between similar gestures (e.g., pointing with index and middle versus pointing with index, middle and ring). In addition to confusion between kinematically similar gestures, there exist confusions between kinetically similar gestures (gestures performed with a similar amount of force exerted). For instance, a strong pointing gesture or a thumb up gesture have a high chance of confusion with fist gestures due to the fist-like motion of fingers not involved in the pointing or thumb up. In addition, subject-specific differences are more apparent in some gestures than others (e.g., snap), which impede the inter-individual recognition of these gestures. In general, we observed that gestures with unconstrained hand fingers (e.g., thumb in a gun gesture) are more prone to misclassification. This is because each subject tends to recruit the unconstrained fingers differently, which results in significantly different sEMG patterns for different subjects.

To further explore the efficacy of sEMG for hand gesture recognition, we constructed three reduced gesture sets consisting of gestures with a minimum of 45% LOSOCV recognition rate in the test on the full gesture-set. Tables I, II, and III show confusion matrices for the recognition of 4 gestures, 6 gestures, and 10 gestures, respectively, using the proposed approach.

TABLE I  
LOSOCV TESTING CONFUSION MATRIX (%) AND INTER-INDIVIDUAL RECOGNITION VARIANCE FOR 4 GESTURES.

	<b>Fist</b>	<b>P-in</b>	<b>P-out</b>	<b>T-R Tap</b>
<b>Fist</b>	96±1	3	0	1
<b>P-in</b>	5	94±3	0	1
<b>P-out</b>	3	0	97±1	0
<b>T-R Tap</b>	14	7	3	76±13

P-in/out: paddle in/out, T-R Tap: Thumb-Ring tap.

TABLE II  
LOSOCV CONFUSION MATRIX (%) AND INTER-INDIVIDUAL RECOGNITION VARIANCE FOR 6 GESTURES.

	<b>Rest</b>	<b>Fist</b>	<b>P-in</b>	<b>P-out</b>	<b>Spider</b>	<b>T-R tap</b>
<b>Rest</b>	99±0	0	0	0	0	1
<b>Fist</b>	0	94±3	3	0	2	1
<b>P in</b>	0	4	90±4	0	6	0
<b>P out</b>	0	2	0	95±2	3	0
<b>Spider</b>	0	3	7	5	67±10	19
<b>T-R Tap</b>	0	13	7	3	15	62±16

P-in/out: paddle in/out, Spider: Spider man, T-R Tap: Thumb-Ring tap.

LOSOCV recognition rates of 79%, 85% and 91% were achieved on the gesture sets with 10, 6, and 4 gestures, respectively. These LOSOCV recognition rates are high and demonstrate the suitability of the proposed approach for

<sup>3</sup>We omit the confusion matrix for the full gesture-set due to limited space.

inter-individual recognition of hand gestures. In general, there are gestures which are better recognized than others, indicating their distinct sEMG patterns that remain consistent between different subjects (e.g., paddle out). However, the results from the reduced-set experiment also show similar recognition deficiencies as those observed in the full-set experiment: 1) confusion between similar gestures (e.g., thumbs up and fist in Table III), and 2) low recognition rates for gestures with unconstrained fingers (e.g., thumb-ring tap in Tables I and II). We have also observed that gestures with lower LOSOCV recognition rates have higher LOSOCV recognition variance, which could indicate a higher degree of inter-individual differences or a higher degree of variability in the gesture.

We hypothesize that there are within-subject variabilities between different trials of a single gesture. To test this hypothesis, an additional 10 trials for all the gestures in the dataset were collected for one subject and the performance of the proposed approach in discriminating between gestures from the subject is tested using stratified 10-fold cross validation (10-FCV: equal proportions of gestures are presented in the training set in each fold). An HMM model with seven states and three mixtures was selected using cross validation to optimize within-subject recognition rates. Table IV shows the within-subject confusion matrix for a set of 5 gestures.

TABLE IV  
10FCV WITHIN-SUBJECT CONFUSION MATRIX (%)

	<b>Fist</b>	<b>P-in</b>	<b>P-out</b>	<b>Spider</b>	<b>Gun</b>
<b>Fist</b>	80	0	0	10	10
<b>P-in</b>	20	70	0	10	0
<b>P-out</b>	0	0	90	0	10
<b>Spider</b>	10	0	0	90	0
<b>Gun</b>	0	0	0	0	100

P-in/out: paddle in/out, Spider: Spider man, T-R Tap: Thumb-Ring tap.

As can be seen from Table IV, while some gestures are easily recognized (e.g., Gun), there are within-subject confusions. These confusions might stem from two sources: 1) within-gesture variabilities, and 2) between-gesture similarities. We hypothesize that a careful selection of the number of mixtures of Gaussians in a gesture-specific HMM can address the first source of variability, as each mixture models a variation of the gesture. To test this hypothesis, we have increased the number of mixtures for fist and paddle-in to 4 and 5, respectively. The results show improvements in recognition of these gestures (fist: 90%, and paddle-in: 80%). However, a further increase in the number of mixtures did not improve the recognition. Therefore, it is important to tune the number of mixtures for each gesture separately for optimized recognition. With regard to the second source of variability, a visual-inspection of the sEMG signals for the misclassified gestures revealed between-gesture similarities (for example, see Fig. 2). The between-gesture similarities impede sEMG-based recognition of the gestures as there will be conflicts between similar gestures.

#### V. CONCLUSIONS AND FUTURE DIRECTIONS

An sEMG-based hand gesture recognition approach is presented that encodes gesture-specific sEMG signals in an

TABLE III  
LOSOCV CONFUSION MATRIX (%) AND INTER-INDIVIDUAL RECOGNITION VARIANCE FOR 10 GESTURES.

	Rest	Fist	P-in	P-out	Snap	T-I tap	T-P tap	Point I M	Thumbs Up	Thumb Flick
Rest	99±0	0	0	0	0	0	0	0	0	1
Fist	0	73±13	0	0	7	0	1	0	19	1
P-In	0	3	88±6	0	3	0	0	1	5	0
P-Out	0	2	0	92±3	1	4	0	0	1	2
Snap	0	1	0	0	86±9	0	3	0	0	10
T-I tap	0	0	0	5	2	75±12	9	3	0	6
T-P tap	0	0	0	2	2	6	66±15	8	13	2
Point I M	0	0	0	1	1	5	8	74±15	5	7
Thumbs Up	0	14	5	0	2	0	4	2	64±14	10
Thumb Flick	0	2	1	1	1	9	2	2	9	73±12

P-in/out: paddle in/out, T-I/P Tap: Thumb-Index/Pinky tap, Point I M: point with index and middle fingers.

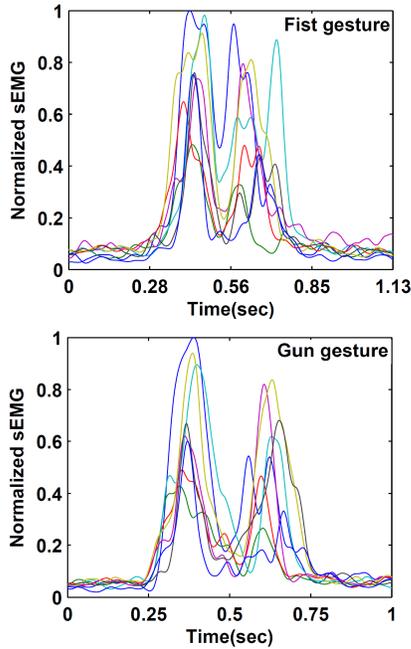


Fig. 2. An example of between-gesture similarities. A fist gesture misclassified as gun (top), a typical gun gesture (bottom). Each color-coded line in these figures corresponds to 1 of 8 sEMG channels.

HMM model with a mixture of Gaussian outputs. The mixture of Gaussian outputs allows for encoding kinematically or kinetically dissimilar sEMGs associated with a gesture in a single HMM. The proposed recognition approach is tested with leave-one-subject-out cross validation and a dataset of various hand gestures performed by 25 subjects. Recognition rates of 79%, 85%, 91% were achieved for gesture sets with 10, 6, and 4 gestures, respectively. The resulting LOSOCV rates are promising and demonstrate the suitability of the proposed approach for inter-individual gesture recognition based on sEMG.

There are gestures that are more difficult to recognize from sEMG signals and are confused with one another due to inter-individual differences (e.g. a gun with thumb up and a gun with thumb down). We hypothesize that gestures involving all the fingers or those in which motions of all the fingers are constrained have a better chance to be recognized. We will further explore this hypothesis in our future experiments. We have also found within-subject differences in performing a gesture. A careful tuning of the number of Gaussian

outputs enables modeling the within-gesture variabilities. In the future, we will investigate a systematic approach for selecting a proper number of Gaussian outputs. There were also confusions between kinematically or kinetically similar gestures. The between-gesture similarities and within-subject variabilities along with inter-individual variabilities make the task of automatic gesture recognition based on sEMG signals a challenging one. Therefore, in developing an sEMG-based recognition model that can be used to control assistive devices and/or prosthetic limbs, care should be taken in choosing the gesture set. In particular, for prostheses application, a set of non-conflicting gestures (dissimilar gestures in sEMG space) should be used. Finally, we hypothesize that a more anatomically-driven sEMG sensor placement could help recognition of the hand gestures, but may impede device usability due to the need for careful positioning.

## REFERENCES

- [1] A. D. C. Chan and K. Englehart, "Continuous myoelectric control for powered prostheses using hidden markov models," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 1, pp. 121–124, 2005.
- [2] M. Yoshikawa, M. Mikawa, and K. Tanaka, "A myoelectric interface for robotic hand control using support vector machine," in *IROS*, 2007, pp. 2723–2728.
- [3] R. Khushaba, A. Al-Ani, and A. Al-Jumaily, "Orthogonal fuzzy neighborhood discriminant analysis for multifunction myoelectric hand control," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 6, pp. 1410–1419, 2010.
- [4] L. Hargrove, E. Scheme, K. Englehart, and B. Hudgins, "Multiple binary classifications via linear discriminant analysis for improved controllability of a powered prosthesis," *IEEE Trans. Neural Syst. Rehab. Eng.*, vol. 18, no. 1, pp. 49–57, 2010.
- [5] P. Shenoy, K. Miller, B. Crawford, and R. P. N. Rao, "Online electromyographic control of a robotic prosthesis," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 3, pp. 1128–1135, 2008.
- [6] T. Matsubara and J. Morimoto, "Bilinear modeling of emg signals to extract user-independent features for multiuser myoelectric interface," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 8, pp. 2205–2213, 2013.
- [7] F. Orabona, C. Castellini, B. Caputo, A. Fiorilla, and G. Sandini, "Model adaptation with least-squares SVM for adaptive hand prosthetics," in *ICRA*, 2009, pp. 2897–2903.
- [8] A. Burden and R. Bartlett, "Normalisation of EMG amplitude: an evaluation and comparison of old and new methods," *Medical Engineering & Physics*, vol. 21, no. 4, pp. 247 – 257, 1999.
- [9] K. Peter, "The ABC of EMG—a practical introduction to kinesiological electromyography," *Noraxon INC. USA*, 2005.
- [10] L. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition," *Proc. IEEE*, vol. 77, no. 2, pp. 257–286, 1989.