

Human Hand Movement Recognition based on HMM with Hyperparameters Optimized by Maximum Mutual Information

Ruoshi Wen*, Qiang Wang*, Xiang Ma* and Zhibin Li†

* Harbin Institute of Technology, Harbin, China

E-mail: ruoshiwen@gmail.com, wangqiang@hit.edu.cn, maxianghit@163.com

† The University of Edinburgh, Edinburgh, U.K.

E-mail: zhibin.li@ed.ac.uk

Abstract—Performing dexterous and versatile movements is essential for multi-finger manipulators for human-robot collaboration, and designing effective control methods for the robotic manipulator is challenging. To recognize human hand movements, we used surface electromyography (sEMG) for sensing myoelectric activity due to its portability and low-cost compared to cameras, and proposed a hidden Markov model (HMM) based method to characterize the transition of action primitives. For building HMMs for hand movements, the hyperparameters, including features, the window length and the number of states, are optimized by the maximum mutual information (MMI) criterion. The optimal features - marginal Discrete Wavelet Transform (mDWT) and mean value - are extracted from multichannel signals acquired from 12 electrodes. Our proposed method is validated by recognizing 40 hand movements from activities of daily living (ADL) in the second NinaPro database. Using MMI as the optimization criterion for hyperparameters, we have improved the average recognition accuracy over 40 subjects in the database from 92.02% to 97.32%.

Index Terms—Hand movement recognition, sEMG, HMM, MMI

I. INTRODUCTION

Demonstration-based robotic control and teleoperation enable robots to learn from humans' dexterous and versatile movements and improve the collaboration capability between both sides [1], [2], [3]. Before learning from demonstration or executing a command, robots should understand the action intention, which makes the recognition problem one of the major research areas in robotic control [4], [5]. Surface electromyography (sEMG) is a non-invasive perception method for the detection, recording and interpretation of electric activities of skeletal muscles [6]. Moreover, it provides planning and execution information for movement decoded from signals generated by the brain, thus being a direct information source for humans' movement intention. sEMG-based human hand movement recognition can be modeled as a time-series classification problem, which can be solved by three approaches: feature-based, distance-based and model-based methods.

Feature-based methods classify the sequence by building a classifier with the extracted feature vector that represents the whole time-series. Features from both the time domain and the frequency domain, such as mean absolute value (MAV), histogram, zero crossings (ZC), and waveform length (WL), have been utilized. By combining with classifiers such as artificial neural networks (ANN) [7], [8], support vector machine (SVM) [9], and linear discriminant analysis (LDA) [10], feature-based

classification has achieved good results [11], [12], [13]. In those studies, the largest number of subjects from whom sEMG was acquired was fifteen [9], and the number of hand movements ranges from three [8] to twelve [14].

Although feature-based methods obtained the accuracy of around 90%, the number of subjects and movements in the aforementioned studies is limited. Moreover, due to the variance of the sEMG quality, the lack of hand gestures and the number of subjects, it is difficult to benchmark the performance of these feature-based classifiers. In 2014, a benchmark database Ninapro (Non-Invasive Adaptive Prosthetics) published the datasets acquired by the state-of-the-art instruments, including multichannel sEMG signals from 67 intact subjects when performing more than 50 hand movements of ADL [15], [16], [17]. In their work, a feature-based method combined with the Kernel Regularized Least Squares (KRLS) algorithm is applied to classify hand movements. It treats the feature vectors extracted from the fixed-length windows segmented from an entire sequence as inputs and the movement class to which that sequence belongs as labels. Although higher accuracy is obtained using longer windows, bigger delays are inevitably introduced to the control system [18]. Models built on the features extracted from parts of the sequence also suffer from the temporal information loss. To understand how two movements differ both spatially and temporally, it is desirable to model the movement recognition problem on the entire sequence. To this end, distance-based and model-based methods are considered.

Distance-based methods measure the information similarity between sequences from different movements. K-nearest neighbours algorithm (KNN) is the simplest method which classifies a sequence by the majority vote of its neighbours [19], and a multi-dimensional dynamic time wrapping (MD-DTW) based method successfully extracted the entire trajectories for hand movement recognition with speed variations [20]. Despite that no training process is required in KNN, the computational complexity to classify a new sequence increases with the size of the training set.

A challenging problem of hand movement recognition is the variations in velocity, shape, and duration even when executing the same movement, where model-based methods play a crucial role. Naive Bayes sequence classifier is the simplest generative model for sequence classification, which has been widely applied in text classification [21] and genomic sequences classification [22]. However, the conditional independence assumption required by Naive Bayes is often violated. The hidden Markov model (HMM) can characterize the statistical properties of spatial-temporal sequences and model the transition between hidden

states. Compared to Naive Bayes, the independence requirement does not have to be strictly met. The successful applications in HMM-based hand movement recognition [23], [24], and sign language recognition [25] validated the feasibility of model-based methods .

One difficulty in training HMMs is the optimization of hyperparameters, especially how to choose the number of hidden states. In speech recognition, those parameters are chosen empirically (five hidden states are usually used to represent a word). However, in HMM-based hand movement recognition, information on how to choose the number of hidden states can be barely found. In this work, we applied the maximum mutual information (MMI) criterion to select the optimal set of models from candidates [26].

For the given Ninapro benchmark, we have shown that model-based methods are superior to feature-based and distance-based methods, and will proceed to solve the challenging aspects. Besides providing an overview of the methodologies, we assessed the performance improvement of the HMM-based classification using the technique of hyperparameters' optimization.

Our contributions can be summarized as follows:

1) We proposed a hand movement recognition framework based on ergodic HMM which can model the myoelectric activity transitions with multichannel signals of sEMG, achieving an excellent performance despite limited training data.

2) We proposed using recognition accuracy as a metric to optimize the length of the sliding window for the segmentation.

3) We proposed the maximum mutual information criterion for hyperparameters' optimization - the number of hidden states and feature selection, which were chosen by experience before.

The remaining of our paper is organized as follows. We first overview the scope of hand movement recognition methods and review the features as well as mathematical preliminaries to train HMMs. We choose the feature combination and the number of hidden states through cross-validation accordingly, based on two metrics - the mutual information and the classification accuracy. Then, we evaluate the performance of HMMs optimized by these two metrics on the test set and analyse the results. Afterwards, we discuss the internal reasons for the success and limitations of the proposed method. In the end, we draw conclusions and future directions in the last section.

II. METHODOLOGY

A. Overview

An overview of our proposed hand movement recognition method is shown in Fig. 1. Multichannel signals of EMG and accelerometer are recorded by electrodes placed on the surface of the muscles which are involved with hand movements. The proposed method trains an HMM for each hand movement with multichannel signals, so the number of HMMs is the same as that of the hand movements. The hand movement is recognized by selecting the model with the highest probability. Each phase in the framework will be detailed in the following subsections.

B. Splitting, Preprocessing, and Windowing

We first split the multichannel signals into a training set and test set and standardized sEMG signals with statistics from the training set. Then, we used a sliding window to segment the standardized sEMG and raw triaxial accelerometer signals. We also conducted experiments to find the optimal sliding window length, where it was set to 200, 400 and 800 *ms* respectively, with the increment of 100 *ms*.

C. Feature Extraction

One of the critical factors that influence the recognition accuracy is the feature selection. The features should not only

be able to extract the most useful information, but computationally efficient. We considered three popular features for sEMG signals: the root mean square (RMS), the histogram, the marginal Discrete Wavelet Transform (mDWT), and one feature for accelerometer signals: the mean value.

1) *Root Mean Square*: RMS has a quasi- or curvilinear-relationship with force exerted by muscles [6], which can be easily implemented in both digital and analogue systems as:

$$RMS = \sqrt{\frac{1}{N} \sum_{n=1}^N x_n^2}, \quad (1)$$

where N denotes the number of samples within each sliding window and x_n denotes the n th signal sample.

2) *Histogram*: The range of data is divided into several disjoint intervals (bins) within which the number of observations is counted. The bin edges used to capture the outliers on both sides were set to $[-\infty, -3, \dots, +3, \infty]$ and the number of bins was set to 20, thus resulting in a histogram feature vector of 240 dimensions. Compared to a smaller number of training samples, the dimensionality of the histogram feature vector is large, suffering from the curse of dimensionality. Observations in the bins located at double tails barely change during the entire movement, which can be removed for dimensionality reduction. Therefore, we only kept the four central bins for the histogram feature. By reducing the feature dimensionality of each channel from 20 to 4, we improved the recognition accuracy from lower than 50% to above 80%.

3) *Marginal Discrete Wavelet Transform*: Discrete Wavelet Transform (DWT) as a time-frequency feature is generally utilized to decompose signals at different resolution levels with wavelet types such as Haar, Daubechies wavelets and so on. In this study, we used the 7th order Daubechies wavelet and the marginal coefficients up to the third level, considering its excellent performance and feature dimensionality reduction. We implemented the marginal DWT by calculating marginal coefficients [27] on the decomposition algorithm proposed by Mallat [28] using the wavelet package in Matlab (*Mathworks*).

4) *Mean Value*: The mean value was extracted from the triaxial accelerometer signals.

The feature candidates, which include individual features aforementioned and their combinations, were extracted within each sliding window and stacked to a feature vector. The vector sequences obtained from all sliding windows will be called observation sequences in the following sections about HMM training. Then the performance of models built with different feature candidates was evaluated. Finally, the optimal feature (combination) was selected to build HMMs for hand movement recognition.

D. Choice of HMM Topology

There are mainly two types of HMM topology as illustrated in Fig. 2: ergodic (or fully connected) and left-right model. In an ergodic HMM, every state could reach every other state in a single step; while in a left-right topology, the state index increases (or stays the same) as time increases. Considering that some action primitives may reappear during certain tasks, the ergodic one shown in Fig. 2(b) is selected in our study.

E. Optimization of Hyperparameters

Before training, hyperparameters consisting of two factors need optimizing. The factor of observation vector representation involves window length and feature selection and the other is the number of hidden states. Because little information on the optimization of these two factors was found in related works, we present and compare two criteria by maximizing the average

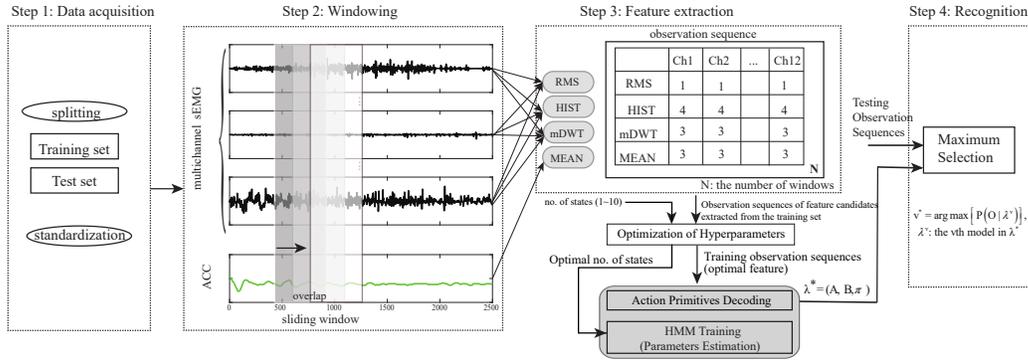


Fig. 1. Overview of the framework of the proposed method for HMM-based hand movement recognition.

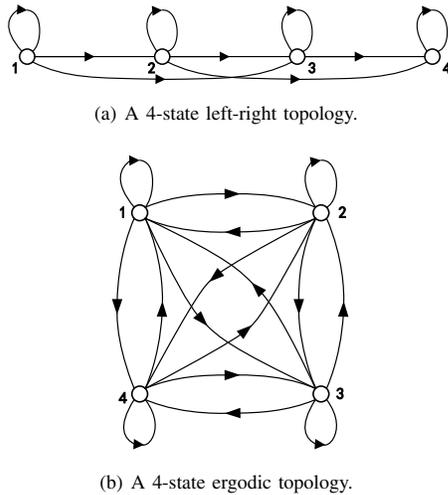


Fig. 2. Illustration of two classic HMM topologies.

classification accuracy and mutual information, while applying k-fold cross-validation with grid search on the training set.

F. HMM Training

An HMM with continuous observations can be characterized by a compact notation:

$$\lambda = (A, B, \pi), \quad (2)$$

where A denotes the state transition probability distribution, B denotes the observation symbol probability distribution, and π denotes the initial state distribution.

1) *Action Primitives Decoding:* Each movement can be decoded into several action primitives. In the training process, the observations were first decoded to uncover the hidden part of the model and Gaussian mixture distribution was applied to build a map from the continuous observation sequence to hidden states, formulated as [29]:

$$b_j(O) = \sum_{m=1}^M c_{j,m} \mathcal{N}[O | \mu_{j,m}, \Sigma_{j,m}], \quad 1 \leq j \leq N \quad (3)$$

where O is the observation vector to be decoded, $c_{j,m}$ is the mixture coefficient for the m th mixture in state j and \mathcal{N} is Gaussian probability density function, with mean vector $\mu_{j,m}$

and covariance matrix $\Sigma_{j,m}$ for the m th mixture component in state j . The mixture coefficients constraint is given by:

$$\sum_{m=1}^M c_{j,m} = 1, \quad 1 \leq j \leq N, \quad (4)$$

$$c_{j,m} \geq 0, \quad 1 \leq j \leq N, \quad 1 \leq m \leq M. \quad (5)$$

2) *Model Parameters Estimation:* Given the observations in the training set, the unknown model parameters are estimated by maximum likelihood approach, which is implemented by Baum-Welch algorithm [30]. This procedure is equivalent to the constrained optimization as:

$$\bar{\lambda}^* = \underset{\bar{\lambda}}{\operatorname{argmax}} Q(\lambda, \bar{\lambda}) \quad (6)$$

where the re-estimated model is defined by $\bar{\lambda} = (\bar{A}, \bar{B}, \bar{\pi})$, and $Q(\lambda, \bar{\lambda}) = \sum_Q P(Q|O, \lambda) \log [P(O, Q|\bar{\lambda})]$ is Baum's auxiliary function. The stochastic constraints of the HMM parameters are given by:

$$\sum_{i=1}^N \bar{\pi}_i = 1, \quad (7)$$

$$\sum_{j=1}^N \bar{a}_{ij} = 1, \quad 1 \leq i \leq N, \quad (8)$$

$$\sum_{k=1}^M \bar{b}_j(k) = 1, \quad 1 \leq j \leq N. \quad (9)$$

G. HMM-based Recognition

After the training process, the optimal parameters for each HMM model were obtained. For hand movement recognition with the trained HMMs, the probability of an observation sequence given the model $P(O|\lambda)$ was calculated by the forward algorithm, defined as:

$$\alpha_t(j) = P(O_1 O_2 \dots O_t, q_t = S_j | \lambda), \quad (10)$$

where the forward variable $\alpha_t(j)$ is solved iteratively as:

Step1: Initialization

$$\alpha_1(i) = \pi_i b_i(O_1), \quad 1 \leq i \leq N. \quad (11)$$

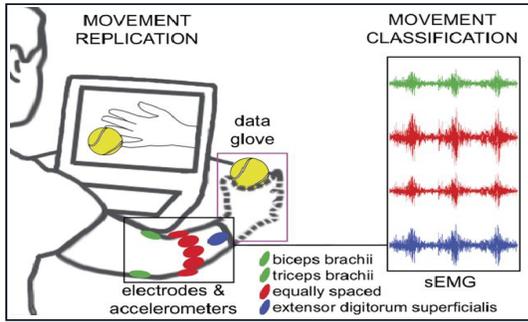


Fig. 3. Placement of EMG electrodes: 8 electrodes were equally spaced around each participant’s forearm at the radio-humeral joint, two at the EDC and FDS muscles, and another two at the biceps and triceps muscles [15]

Step2: Iteration

$$\alpha_{t+1}(j) = \left[\sum_{i=1}^N \alpha_t(i) a_{ij} \right] b_j(O_{t+1}), \quad (12)$$

$$1 \leq t \leq T - 1, 1 \leq j \leq N.$$

Step3: Termination

$$P(O|\lambda) = \sum_{i=1}^N \alpha_T(i). \quad (13)$$

Each observation sequence in the test set can be recognized by the maximum probability selection, given by:

$$v^* = \underset{v}{\operatorname{argmax}} [P(O|\lambda^v)], \quad 1 \leq v \leq V \quad (14)$$

where v^* denotes the index of the recognized movement.

III. EXPERIMENTAL SETUP

A. Data Acquisition

To evaluate the performance of the proposed method, we utilized the second version of the publicly available database from the Non-Invasive Adaptive Prosthetics project [15] - the largest sEMG database to the best of our knowledge. In DB2, 40 intact subjects performed six repetitive tasks of 40 hand movements following the videos displaying their right hands as shown in Fig. 3. The signal acquisition setup is a Delsys™ Trigno Wireless system® consisting a base station and twelve wireless electrodes. The accelerometer signals were acquired by the triaxial accelerometers integrated on the electrodes sampled at a rate of 148 Hz (being upsampled to the sampling rate of sEMG of 2 kHz using interpolation afterwards). The position of the electrodes is shown in Fig. 3.

The sEMG signals were filtered using a Hampel filter [31] from 50 Hz (and its third and fifth harmonics) power-line interference before synchronizing with the accelerometer signals. Following the protocol of splitting data into the training set and test set for classification [17], [15], the second and fifth repetitions of each movement were used for testing, while the remaining four for training. Statistics from the training set of each subject was used to standardize the training set and test set.

B. Hyperparameter Optimization

The training set includes signals acquired from four repetitions of each movement. To optimize hyperparameters (the observation vector representation and the number of hidden states), we applied four-fold cross-validation on the training set with grid search, i.e., three repetitions for training the model and the left one for validation.

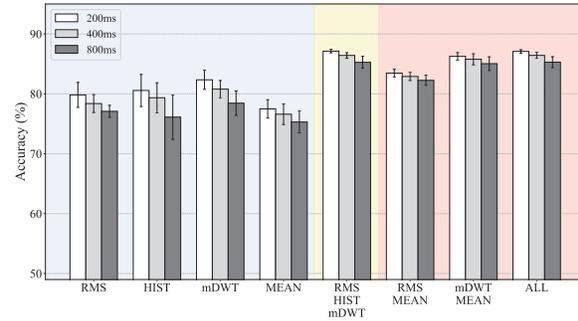


Fig. 4. Average classification accuracy over 40 subjects: 1) three sEMG features and one ACC feature individually, 2) the combination of three sEMG features, and 3) the combinations of sEMG and ACC features, extracted from window lengths of 200, 400, 800 ms. The error bars indicate unit standard deviation.

1) *Observation Vector: Window length and feature selection are two main factors for observation vector representation. Under each of three window lengths (200, 400 and 800 ms), features in terms of three cases - three sEMG features and one ACC feature individually, the combination of sEMG features, the combinations of sEMG and ACC features - were extracted. The performance was measured by the average classification accuracy over 40 subjects, as shown in Fig. 4.*

Regarding the window length, results in Fig. 4 show that a window length of 200 ms outperforms 400 and 800 ms for all feature candidates, hence a sliding window of 200 ms was chosen to segment multichannel signals.

For feature selection, from all individual sEMG features, it can be observed that mDWT achieves better performance than RMS and HIST. Because of the more extracted information, the accuracy achieved by the combination of sEMG features shows a noticeable increase compared to each sEMG feature. It also shows that the inclusion of ACC modality to RMS and mDWT improves the accuracy towards a higher level, which verifies the suggestion in [17] that the ACC and sEMG signals are complementary sources. Three feature candidates (RMS, HIST and mDWT; mDWT and MEAN; RMS, HIST, mDWT and MEAN) achieve the best performance without significant differences ($p = 0.384$, ANOVA). Moreover, these candidates have far less standard deviations than the others, resulting in more reliable performance, thus being selected to represent the observation vector.

2) *Number of hidden states: To optimize the number of hidden states, we conducted experiments and evaluated the performance under each number by two criteria - average recognition accuracy and mutual information over k folds, following Algorithm. 1 and Algorithm. 2.*

Average recognition accuracy of the selected feature candidates under different numbers of hidden states (from 1 to 20) is shown in Fig. 5. The accuracy of the feature candidate mDWT and MEAN decreases from 87.2% to 80.5% as the number of hidden states increases, while the accuracy of the other two fluctuated around 87% as the number of states increases from 1 to 15, and then drops to about 85%. The combinations of sEMG features (RMS, HIST, and mDWT) and all features (RMS, HIST, mDWT, and MEAN) achieve nearly the same good performance over most numbers of states. Following Occam’s Razor principle, the combination of sEMG features (RMS, HIST, and mDWT) was chosen. The number of hidden states was optimized to two.

Besides fitting the model to the training data by the maximum likelihood, we also need to consider that the model should distinguish itself from the others which represent the other

Algorithm 1 Average recognition accuracy computed over k-fold cross-validation.

Require:

Observation sequences in the training set.

Ensure:

Average recognition accuracy.

- 1: Split the observation sequences into the validation set and training set;
- 2: **for** the number of hidden states from 1 to 20 **do**
- 3: **for** each feature combination **do**
- 4: **for** each fold **do**
- 5: Train 40 HMMs using maximum likelihood;
- 6: Compute log probability matrix P (where $P_{i,j}$ represents the log probability of the i th observation being generated by the j th model) using Eq.(11),Eq.(12), Eq.(13);
- 7: Recognize movements in the validation set and compute accuracy under each fold;
- 8: **end for**
- 9: Compute the average recognition accuracy over k folds.
- 10: **end for**
- 11: **end for**

Algorithm 2 Average mutual information computation over k-fold cross-validation.

Require:

Observation sequences in the training set.

Ensure:

Average mutual information.

- 1: Split the observation sequences into the validation set and training set;
- 2: **for** the number of hidden states from 1 to 20 **do**
- 3: **for** each feature combination **do**
- 4: **for** each fold **do**
- 5: Train 40 HMMs using maximum likelihood;
- 6: Compute log probability matrix P ;
- 7: Compute mutual information using Eq.(16) under each fold;
- 8: **end for**
- 9: Compute the average mutual information over k folds.
- 10: **end for**
- 11: **end for**

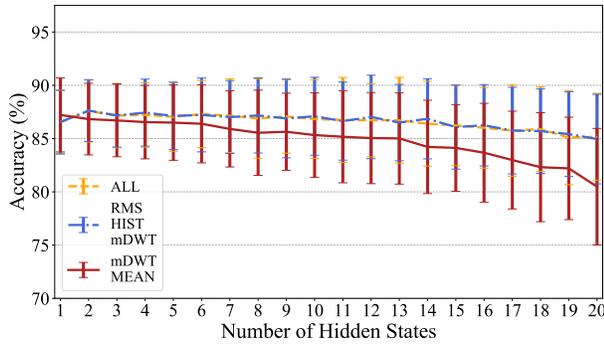


Fig. 5. Average accuracy of recognition achieved by three feature candidates (mDWT and MEAN; RMS, HIST and mDWT; RMS, HIST, mDWT and MEAN) under each number of hidden states.

movements. The criterion for the mutual information I between the observation sequence O^v and the complete set of models $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_V)$ is implemented as:

$$I_v = \log P(O^v|\lambda_v) - \log \sum_{w=1}^V P(O^v|\lambda_w). \quad (15)$$

By maximizing (15), the best model λ_v on the training sequences O^v is selected. By maximizing the sum over all training sequences, hyperparameters are optimized simultaneously by (16):

$$I = \sum_{v=1}^V \left[\log P(O^v|\lambda_v) - \log \sum_{w=1}^V P(O^v|\lambda_w) \right]. \quad (16)$$

Average mutual information under different numbers of hidden states is illustrated in Fig. 6. The mutual information of two feature candidates (RMS, HIST and mDWT; RMS, HIST, mDWT

and MEAN) is nearly the same under all numbers of states and increases gradually by the number of hidden states. However, the feature combination of mDWT and MEAN obtained much larger mutual information than the other two. Following the maximum mutual information criterion, the feature candidate of mDWT and MEAN was chosen as the representation for the observation vector, and the number of hidden states was optimized to six.

IV. RESULTS

For each subject in the database, HMMs were trained on the training set with hyperparameters optimized by two metrics mentioned above. The performance was measured by the average accuracy of the recognized movement in the test set.

A. Evaluation Based on Accuracy of Recognition

Based on the recognition accuracy, we extracted three sEMG features mentioned above from multichannel signals and set the number of hidden states to 2. The extracted observation sequences were used to train ergodic HMMs with two hidden states, which were then used to recognize the hand movements in the test set. Fifty experiments were repeated to validate the accuracy convergence of our proposed method, where the initial HMM parameters were randomized. The average recognition accuracy and standard deviation over repetitive experiments for 40 subjects are shown in Fig. 7. The robustness of our method was validated by the small standard deviation within each subject shown in Fig. 7. However, there is performance variation among subjects, i.e., the recognition accuracy of 30% subjects is higher than 95%, while that of the other nine subjects is below 90%. Results show that there exist individual muscular differences, thus suggesting that the tuning of HMM parameters should be subject-specific.

B. Evaluation based on maximum mutual information

Based on MMI, the feature combination of mDWT and MEAN was extracted, and the number of hidden states was set to 6. The experiments were also repeated for fifty times to test the reliability of results. As illustrated in Fig. 7, there is a substantial increase regarding recognition accuracy compared to the first criterion. The accuracy exceeds 90% for all subjects and is above 95%

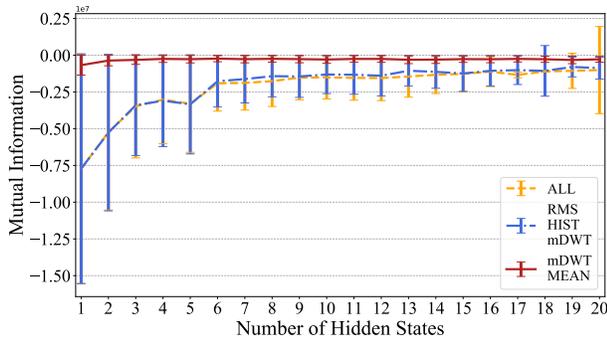


Fig. 6. Average mutual information of three feature candidates (mDWT and MEAN; RMS, HIST and mDWT; RMS, HIST, mDWT and MEAN) under each number of hidden states.

for 90% subjects. Even taken the standard deviation within each subject into consideration, the performance of the latter criterion is much better. It is conclusive that applying the MMI criterion, instead of experience or recognition accuracy via cross-validation, is an advantageous way to optimize hyperparameters.

V. DISCUSSION

A. The Success of HMM-based Hand Movement Recognition

HMM-based recognition method obtained the accuracy over 95% for 90% subjects when classifying 40 hand movements, which cover the majority of ADL. The confusion matrix is shown in Fig. 8, where each row is normalized. The element in the diagonal corresponds to the total number of the correctly recognized sequences from the test set. The proposed method achieved the accuracy of more than 90% for all hand movements, where more than 20% movements were recognized accurately, more than 60% movements obtained the accuracy of above 98%, and more than 80% of the movements had the recognition accuracy over 95%. The success of our proposed method is analysed as follows.

While a person is moving the hand, the involved muscles undergo a transition from rest to activation and deactivation. Here, HMM can be successfully applied to describe the state transition in time-series, which is ideal for modelling the transition of action primitives for different hand movements, which are realized by the coordination of the involved muscles (extension and flexion of wrists, palms and fingers). Different coordination modes, which is characterized by the number, excitation time and amplitude of the involved muscles, can be measured by multichannel sEMG signals.

In terms of feature selection, the inclusion of ACC in sEMG features significantly increases the recognition accuracy. The accelerometer signals contain more accurate positional/spatial information than sEMG. Therefore, while combining sEMG and ACC features, the uncertainty of sEMG due to noise, cross-talk and semi-stochastic property, is reduced to a lower level.

Feature candidates selected by the maximum likelihood procedure achieved no significant performance difference regarding the recognition accuracy. So we present the metric - mutual information - to evaluate the performance of different hyperparameters from a different perspective. The mDWT, which extracts information from both time and frequency domain, has lower dimensionality and computational complexity compared to the feature combination of RMS (time-domain) and HIST (frequency-domain). Under this number of hidden states, the model's ability to distinguish observation sequences from incorrect model candidates is also maximized.

Our proposed HMM-based method also outperforms distance-based methods, where multidimensional dynamic time wrapping (MD-DTW) was used to measure the similarity between multi-channel signals of sEMG [20]. The performance was compared on the same database, and in terms of recognition accuracy, the result of our work shows an advantage (See Fig. 3 in [20], where no movement was predicted with 100% accuracy, only 5% movements' accuracies exceed 98%, less than 10% movements achieved the accuracy of 95%, half movements obtained the accuracy of over 90%, and 4 movements had the accuracy of less than 80%). Moreover, in terms of prediction, our method has an advantage in the computational complexity, i.e., $\mathcal{O}(M)$ where M is the total number of movements, whereas MD-DTW has the computational complexity of $\mathcal{O}(N^2)$ where N is the number of samples.

B. Limitation of Recognition Using sEMG and Accelerometer

As indicated in Fig. 8, six movements got an accuracy slightly below 95% and the worst performance was 90%, locating in the 22nd and 30th. The misclassification happens mostly in three movements highlighted by squares. Mostly the 22nd movement is misclassified to the 18th and 20th, and the 19th movement is also prone to be misclassified to the 18th.

For the group in the red square, four fingers were flexing at the same time, while the thumb moved differently. To recognize this group, the HMMs should be able to classify different thumb movements. However, sEMG signals do not contain the electric activity of muscles that control the thumb movement because physiologically, extensor pollicis longus, abductor pollicis longus and the other two brevis, responsible for thumb's extension and flexion, are deeper muscles which cannot be sensed by surface EMG. Moreover, accelerometers attached on the arm have no measurement of the thumb position at all. We can also observe that the other two misclassifications (29th/30th, and 28th/33rd) are also due to different thumb movements. Therefore, additional sensors should be introduced to distinguish different hand movements caused by the thumb gesture.

VI. CONCLUSION AND FUTURE WORK

In this paper, we proposed an HMM-based hand movement recognition method with multichannel signals of sEMG and accelerometers and evaluated the method on the NinaPro database. Then, we introduced the maximum mutual information (MMI) criterion combined with 4-fold cross-validation to optimize hyperparameters. Although misclassification exists in several groups of movements due to the limitation of sensors, experimental results indicated using HMM with multichannel signals of sEMG and accelerometer is an advantageous and reliable method for hand movement recognition.

In future work, to shorten the process to optimize hyperparameters via cross-validation, non-parametric Bayesian method can be applied. Moreover, it is time-consuming to obtain enough labelled data, especially when conducting experiments on people with amputation for designing prosthetics. Future work should also focus on adapting existing models on new subjects to improve the efficiency of the training process.

ACKNOWLEDGEMENT

This work was supported by the National Natural Science Foundation of China under Grant 61876054.

REFERENCES

- [1] B. Xi, S. Wang, X. Ye, Y. Cai, T. Lu, and R. Wang, "A robotic shared control teleoperation method based on learning from demonstrations," *Int. J. Adv. Robo. Syst.*, vol. 16, no. 4, p. 1729881419857428, Jul. 2019.

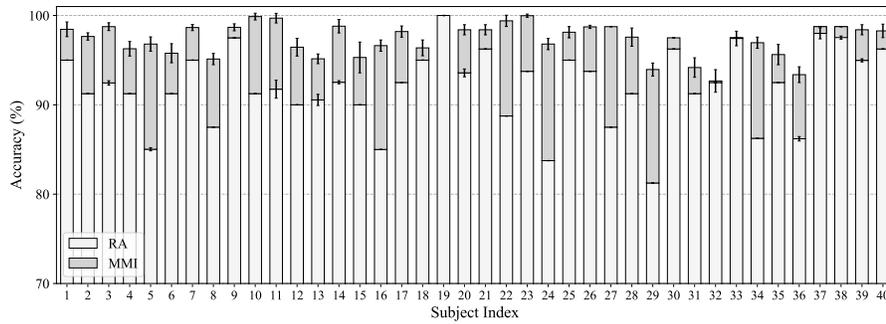


Fig. 7. Average recognition accuracy over 50 repeated experiments of 40 subjects evaluated based on the metrics of accuracy of recognition and mutual information.

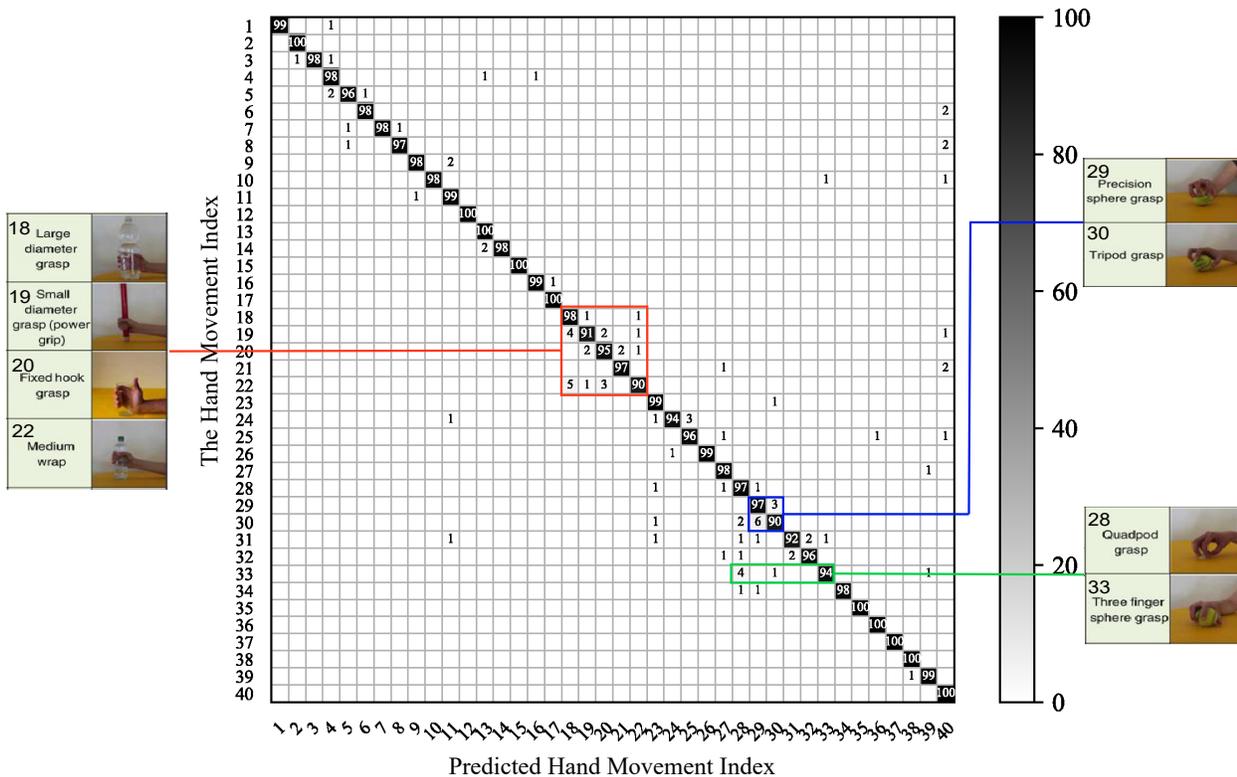


Fig. 8. Confusion matrix for hand movement recognition over 40 subjects.

[2] L. Yang, Z. Li, Q. Lei, J. Xu, Y. Deng, and Y. Zhong, "Teaching a robot to draw: hand gesture demonstration based on human-robot interaction," in *12th Int. Conf. Mach. Vision (ICMV 2019)*, vol. 11433, 2020, p. 114330M.

[3] R. Wen, K. Yuan, Q. Wang, S. Heng, and Z. Li, "Force-guided high-precision grasping control of fragile and deformable objects using semg-based force prediction," *IEEE Robot. Autom. Lett.*, vol. 5, no. 2, pp. 2762–2769, 2020.

[4] X. Feng, Q. Song, Q. Guo, D. Liu, Z. Zhao, and Y. Zhao, "Hand gesture recognition with ensemble time-frequency signatures using enhanced deep convolutional neural network," in *2019 Asia-Pac Sig. Inf. Process. Assoc. Annu. Summit. Conf., APSIPA ASC*. IEEE, Nov. 2019, pp. 1602–1605.

[5] Y. Li, Y. Wang, Y. Jiang, and L. Zhang, "Action recognition using convolutional neural networks with joint supervision," in *2019 Asia-Pac Sig. Inf. Process. Assoc. Annu. Summit. Conf., APSIPA ASC*, Nov. 2019, pp. 2015–2020.

[6] E. Criswell, *Cram's Introduction to Surface Electromyography*. London, U.K.: Jones & Bartlett Learning, 2010.

[7] I. Batzianoulis, S. El-Khoury, E. Pirondini, M. Coscia, S. Micera, and A. Billard, "EMG-based decoding of grasp gestures in reaching-to-grasping motions," *Rob. Auton. Syst.*, vol. 91, pp. 59–70, May 2017.

[8] S. Mane, R. Kambli, F. Kazi, and N. Singh, "Hand motion recognition from single channel surface EMG using wavelet & artificial neural network," *Procedia Comput. Sci.*, vol. 49, pp. 58–65, 2015.

[9] J. Liu, "Adaptive myoelectric pattern recognition toward improved multifunctional prosthesis control," *Med. Eng. & Phys.*, vol. 37, no. 4, pp. 424 – 430, Apr. 2015.

[10] 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. Rehab.*, vol. 14, no. 1, p. 71, Jul. 2017.
- [11] A. Bhattacharya, A. Sarkar, and P. Basak, “Time domain multi-feature extraction and classification of human hand movements using surface EMG,” in *2017 4th Int. Conf. Adv. Comput. Commun. Syst. (ICACCS)*. IEEE, Jan. 2017.
- [12] A. Subasi, L. Alharbi, R. Madani, and S. M. Qaisar, “Surface EMG based classification of basic hand movements using rotation forest,” in *2018 Adv. Sci. Eng. Technol. Int. Conf. (ASET)*. IEEE, Feb. 2018.
- [13] D. R. Freer, J. Liu, and G.-Z. Yang, “Optimization of EMG movement recognition for use in an upper limb wearable robot,” in *2017 IEEE 14th Int. Conf. Wearable Implant. Body Sens. Netw. (BSN)*. IEEE, May 2017.
- [14] N. Malesevic, D. Markovic, G. Kanitz, M. Controzzi, C. Cipriani, and C. Antfolk, “Decoding of individual finger movements from surface EMG signals using vector autoregressive hierarchical hidden Markov models,” in *2017 Int. Conf. Rehab. Robo (ICORR)*. IEEE, Jul. 2017.
- [15] M. Atzori, A. Gijsberts, C. Castellini, B. Caputo, A.-G. Mittaz Hager, S. Elsig, G. Giatsidis, F. Bassetto, and H. Mller, “Electromyography data for non-invasive naturally-controlled robotic hand prostheses,” *Sci. Data*, vol. 1, 12 2014.
- [16] M. Atzori, A. Gijsberts, I. Kuzborskij, S. Elsig, A.-G. M. Hager, O. Deriaz, C. Castellini, H. Muller, and B. Caputo, “Characterization of a benchmark database for myoelectric movement classification,” *IEEE Trans. Neural Syst. Rehab. Eng.*, vol. 23, no. 1, pp. 73–83, Jan. 2015.
- [17] A. Gijsberts, M. Atzori, C. Castellini, H. Mller, and B. Caputo, “Movement error rate for evaluation of machine learning methods for sEMG-based hand movement classification,” *IEEE Trans. Neural Syst. Rehab. Eng.*, vol. 22, no. 4, pp. 735–744, Jul. 2014.
- [18] T. R. Farrell and R. F. Weir, “The optimal controller delay for myoelectric prostheses,” *IEEE Trans. Neural Syst. Rehab. Eng.*, vol. 15, no. 1, pp. 111–118, March 2007.
- [19] K.-Y. Lian, C.-C. Chiu, Y.-J. Hong, and W.-T. Sung, “Wearable armband for real time hand gesture recognition,” in *2017 IEEE Int. Conf. Syst. Man Cyber. (SMC)*. IEEE, Oct. 2017.
- [20] M. AbdelMaseeh, T.-W. Chen, and D. W. Stashuk, “Extraction and classification of multichannel electromyographic activation trajectories for hand movement recognition,” *IEEE Trans. Neural Syst. Rehab. Eng.*, vol. 24, no. 6, pp. 662–673, Jun. 2016.
- [21] G. Feng, J. Guo, B.-Y. Jing, and T. Sun, “Feature subset selection using naive Bayes for text classification,” *Pattern Recognit. Lett.*, vol. 65, pp. 109–115, Nov. 2015.
- [22] S. Douglass, S.-W. Hsu, S. Cokus, R. B. Goldberg, J. J. Harada, and M. Pellegrini, “A naive Bayesian classifier for identifying plant microRNAs,” *Plant J.*, vol. 86, no. 6, pp. 481–492, Jun. 2016.
- [23] N. Malešević, D. Marković, G. Kanitz, M. Controzzi, C. Cipriani, and C. Antfolk, “Vector autoregressive hierarchical hidden Markov models for extracting finger movements using multichannel surface EMG signals,” *Complexity*, vol. 2018, pp. 1–12, Feb. 2018.
- [24] A. Samadani, “EMG channel selection for improved hand gesture classification,” in *2018 40th Annu. Int. Conf. IEEE Eng. Med. Bio. Soc. (EMBC)*. IEEE, Jul. 2018.
- [25] X. Ma, L. Yuan, R. Wen, and Q. Wang, “Sign language recognition based on concept learning,” in *2020 IEEE Int. Instrum. Meas. Technol. Conf. (I2MTC)*, 2020, pp. 1–6.
- [26] L. Bahl, P. Brown, P. de Souza, and R. Mercer, “Maximum mutual information estimation of hidden markov model parameters for speech recognition,” in *IEEE Int. Conf. Acoust. Speech Sig. Process. ICASSP*. Institute of Electrical and Electronics Engineers, 1986.
- [27] M.-F. Lucas, A. Gaufriau, S. Pascual, C. Doncarli, and D. Farina, “Multi-channel surface EMG classification using support vector machines and signal-based wavelet optimization,” *Biomed. Signal Process. Control*, vol. 3, no. 2, pp. 169–174, Apr. 2008.
- [28] S. Mallat, “A theory for multiresolution signal decomposition: the wavelet representation,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 7, pp. 674–693, Jul. 1989.
- [29] L. R. Rabiner, “A tutorial on hidden markov models and selected applications in speech recognition,” *Proc. IEEE*, vol. 77, no. 2, pp. 257–286, 1989.
- [30] L. Baum, “An inequality and associated maximization technique in statistical estimation of probabilistic functions of a Markov process,” *Inequalities*, vol. 3, pp. 1–8, 1972.
- [31] I. Kuzborskij, A. Gijsberts, and B. Caputo, “On the challenge of classifying 52 hand movements from surface electromyography,” in *2012 Annu. Int. Conf. IEEE Eng. Med. Bio. Soc.* IEEE, Aug. 2012, pp. 4931–4937.