Published in final edited form as:

Conf Proc IEEE Eng Med Biol Soc. 2017 July ; 2017: 2968–2971. doi:10.1109/EMBC.2017.8037480.

Decoding Complex Imagery Hand Gestures

Seyed Sadegh Mohseni Salehi², Mohammad Moghadamfalahi, Fernando Quivira, Alexander Piers, Hooman Nezamfar, and Deniz Erdogmus¹

¹Electrical and Computer Engineering Department, Northeastern University, Boston

²ssalehi@ece.neu.edu

Abstract

Brain computer interfaces (BCIs) offer individuals suffering from major disabilities an alternative method to interact with their environment. Sensorimotor rhythm (SMRs) based BCIs can successfully perform control tasks; however, the traditional SMR paradigms intuitively disconnect the control and real task, making them non-ideal for complex control scenarios. In this study we design a new, intuitively connected motor imagery (MI) paradigm using hierarchical common spatial patterns (HCSP) and context information to effectively predict intended hand grasps from electroencephalogram (EEG) data. Experiments with 5 participants yielded an aggregate classification accuracy–intended grasp prediction probability- of 64.5% for 8 different hand gestures, more than 5 times the chance level.

I. INTRODUCTION

Brain Computer Interfaces have shown promises in providing individuals with alternative interaction methods via brain activities. Major electroencephalogram (EEG) based BCI systems can be categorized into three different categories based on the type of brain activity they utilize, event-related potential (ERP) based, steady state visually evoked potential (SSVEP) based, and motor imagery (MI) based.

ERP and SSVEP are more popular in BCI systems where the short response time and the availability of several options are of concern. Controlling a wheelchair [1] and typing with characters [2] are two examples that rely on the above features, respectively. In control and navigation applications SSVEP-based BCI systems are widely used due to their fast response. SSVEP systems are fast, but they require the visual stimulation and vision capabilities [3] not suitable for individuals with severely impaired vision.

MI, the process of imagining a physical movement without execution, can be categorized as imagining moving different limbs (simple MI) and imagining different movements of the same limb (compound MI)[4]. MI tasks desynchronize the stable, resting alpha/mu (7 – 12 Hz) and beta (12 – 30 Hz) rhythms, that can be measured through EEG potentials on the scalp[4], [5]. The spatial distribution of brain activity measured by EEG signals in response to these imaginations depend on the complexity of the MI task; simple MI (imagination of left hand vs. right leg) produces activity in different spatial locations while compound MI (open palm vs. closed fist) produces activities in similar spatial cortical areas [6]. The high spatial correlation of compound MI coupled with the volume conduction effect, the spatial

distortion of EEG signals on the scalp caused by travelling through tissue and fluid, make different MI tasks difficult to distinguish.

One particular application of BCIs is the control of prosthetics or exoskeletons for individuals with motor control loss. From the BCI perspective, a brain controlled hand requires rapidly (real-time) and accurately classified MI that maps directly to the motion of the target limb. Simple MI is not sufficiently intuitive, and compound MIs are difficult to differentiate. Hence classifying EEG potentials suitable to control a robotic limb is a challenge.

Several studies have classified gestures with moderate success, but these gestures are not sufficiently intuitive to be used in a BCI prosthetic. These studies investigated classifying simple MI by simply distinguishing imagination of right and left hand movements with 80% – 93% accuracy[6], [7], [8]. Others have completed binary classifications on different same-limb movements, such as distinguishing between wrist flexion, extension, supination, pronation, and finger movement[6], [9], [10] resulting in accuracies between 58 – 82%. However, the higher accuracies are only achieved when wrist extension has been one of the binary options; without wrist extension classification accuracies tend to run closer to the lower side of the reported range. Binary classification does not provide enough degrees of freedom to control a prosthetic hand, and the imaginary movements used for classification are not necessarily intuitive hand motions. A recent study [11] classified four hand states (wrist flexion, extension, supination, and pronation) using source imaging analysis [12] to improve signals, separability. Even with this combination, the imagined wrist motions do not map directly or intuitively enough to natural hand movements. In this paper, we propose classification of four grasps on each hand totaling to 8 options. The grasps mimic four of the required grasps to complete everyday tasks. The goal is to improve classification accuracy and time to control a prosthetic hand, utilizing the natural choice of grasps.

II. Method

A. Experiment design

Classifier parameters are estimated using data collected during a supervised calibration session. The classifier is meant to discriminate EEG signals corresponding to imagination of four gestures shown in Fig. 1.a. Prior to data collection, participants get familiarized with gestures through executing each gesture two times, for 5 seconds. During the calibration session, participants are presented with the *target* gesture for 2 seconds during which they are asked to focus on that particular gesture. Then, upon the presentation of the word *start* on the display participants imagine the target gesture for 5 seconds (Fig. 1.b). Here, each MI task for a gesture is called a *trial*. Inter trial interval in the calibration task is arbitrary and is decided by the user based on their required time to rest. Each participant performs 20 trials of each gesture, summing to a total of 160 MI trials.

B. EEG feature extraction

Prior to feature extraction, EEG signals are filtered with a 3–30 Hz FIR band-pass filter to retain the alpha/mu and beta frequency bands, as they show the most activity in response to

imagination and execution of hand gestures [7]. Assuming the band-passed EEG signal follows a Gaussian distribution within the time window, and source activity constellations between different class pairs are independent, we can design a spatial filter (spatial transform) that maximizes the signal's variance for one class and minimizes for the other. We used the common spatial pattern (CSP) algorithm [13] to calculate the spatial transforms. CSP algorithm is used to find the transformation matrix between class pairs. CSP finds the best coordinates between binary classes. In the multiclass case, we propose a Hierarchical Common Spatial Pattern (HCSP) to utilize a series of binary transformations found by CSP.

In HCSP, classes with common properties are placed in one category. For example, first level contains two categories, motions on the right hand and motions on the left hand. Similarly, extension or flexion of fingers and abduction or adduction of thumbs are considered different categories in the second and the third level respectively. Fig 2 outlines the HCSP for classification of different hand gestures. The goal of each classifier is to classify the categories not classes. By cascading the category classifiers, all 8 classes can be detected.

The normalized spatial covariance matrix is estimated as

$$C_j = \frac{1}{N_j} \sum_{i=1}^{N_j} \frac{E_i^{(j)} E_i^{(j)'} }{\text{trace}(E_i^{(j)} E_i^{(j)'})} \quad (1)$$

where $E_i^{(j)}$ is an $m \times t$ matrix representing filtered windowed EEG evidence corresponding to the i^{th} trial of the j^{th} class. Additionally, m is the number of channels, t is the number of samples in a time window, $'$ is the transpose operator, $\text{trace}(\cdot)$ is the sum of the diagonal elements of the matrix, and N_j is the number of trials for the j^{th} class.

In order to find the spatial transformation matrix at each level, V_L , the optimization problem in equation 2 should be solved.

$$V_L^* = \underset{V_L}{\operatorname{argmax}} \frac{V_L^T \sigma_{L_l = -1} V_L}{V_L^T (\sigma_{L_l = -1} + \sigma_{L_l = +1}) V_L} \quad (2)$$

where $\sigma_{L_l = -1}$ and $\sigma_{L_l = +1}$ are the average covariance matrices - where ± 1 represent one of the categories - in level L and are calculated by

$$\sigma_{L_l} = \frac{\sum_{j \in \text{category}(L_l)} N_j C_j}{\sum_{j \in \text{category}(L_l)} N_j} \quad (3)$$

$\sum_{j \in \text{category}(L_l)} N_j$ is the number of trials for each category at each level.

Solving the optimization problem represented in equation 2 is identical to solving generalized eigenvalue problem

$$V_L' \sigma_{L_l = -1} V_L = D \wedge V_L' (\sigma_{L_l = -1} + \sigma_{L_l = +1}) V_L = I \quad (4)$$

where D is a diagonal matrix.

According to each spatial filter V_L , computed for level L , the projected data P_{L_l} corresponding to the i^{th} trial of the j^{th} class, $E_i^{(j)}$, can be calculated as

$$P_{L_l} = V_L' E^{(j \in category(L_l))} \quad (5)$$

In equation 4 the k smallest and largest eigenvalues in D correspond to k leftmost/rightmost columns in spatial filter V respectively. These values yield to the smallest variance in category L_{k-1} and simultaneously largest variance in category L_{k+1} and vice versa. To extract the features, the first and the last k rows of P_{L_l} are considered.

Using $P_{L_l}^{(K)}$ ($K = 1 \dots 2k$) which maximizes the difference of variances between two categories, $f_{L_l}^{(K)}$, features for each trial of the corresponding category at each level are calculated as

$$f_{L_l}^{(K)} = \log \left(\frac{\text{var}(P_{L_l}^{(K)})}{\sum_{q=1}^{2k} \text{var}(P_{L_l}^{(q)})} \right) \quad (6)$$

where, i is the trial index in each the category.

C. EEG Likelihood Probability Extraction

For proper estimation of the classifier performance a leave-one-out cross validation approach is used. At each level, the corresponding feature vector, $f_{L_l}^{(K)}$, is calculated for each trial. The classifier is trained in two steps. First, Fisher LDA algorithm is used to extract fisher scores by

$$F_{L_l}^{(K)} = (w_L^{(K)})' f_{L_l}^{(K)} \quad (7)$$

where

$$w_L^{(K)} \propto \left(\Sigma_{L_l = -1}^{(K)} + \Sigma_{L_l = +1}^{(K)} \right)^{-1} \left(\mu_{L_l = +1}^{(K)} - \mu_{L_l = -1}^{(K)} \right) \quad (8)$$

and $\Sigma_{L_l = -1}^{(K)}, \mu_{L_l = -1}^{(K)}$ and $\Sigma_{L_l = +1}^{(K)}, \mu_{L_l = +1}^{(K)}$ are the category's covariance and mean.

Then, the likelihood probability densities are calculated for the fisher scores in each category $P(F_{L_l}^{(K)} | L = l)$. Multivariate Gaussian model is considered at each level for density estimation due to the nature of EEG signals. Classification accuracy is defined as the probability of the correct answer.

Different integer window lengths from 1 to 5 seconds are considered for feature extraction.

D. Graphical model

The graphical model in Fig 3 is proposed to probabilistically merge the category classifiers and employ the inter-level information to calculate the posterior probability of each gesture. This model represents the generative model of the collected EEG data in epoch e and the time required for the model to make a decision (1–5 s). The goal of this graphical model is to estimate the next state s_e with incorporation of inter-level prior information. The inter-level information is the prior probability of a gesture at a level given the state at the higher level. This information is gathered based on the combination of gestures used over time for different tasks. For instance, the probability of the thumb being closed would be increased if closed fingers is detected at the higher level. Specifically in this graphical model: s_e is the decision between all classes at epoch e , L_e^j represents the decision from the level j classifier, and $\epsilon_{e_i}^j$ is the extracted CSP feature from EEG evidence at time i for decision L_e^j .

Using the graphical model above, the posterior probability of each state given the observed random variables $\left(\left\{ \epsilon_{e_i}^1 \right\}_{i=1}^t, \left\{ \epsilon_{e_i}^2 \right\}_{i=1}^t, \left\{ \epsilon_{e_i}^3 \right\}_{i=1}^t \right)$, can be calculated to construct the PMF over all gestures, required to make a decision. The posterior probabilities are calculated as follows:

$$\begin{aligned}
& P\left(s_e \left| \left\{ \epsilon_{e,i}^1 \right\}_{i=1}^t, \left\{ \epsilon_{e,i}^2 \right\}_{i=1}^t, \left\{ \epsilon_{e,i}^3 \right\}_{i=1}^t \right.\right) \quad (9) \\
& = P\left(L_e^1, L_e^2, L_e^3 \left| \left\{ \epsilon_{e,i}^1 \right\}_{i=1}^t, \left\{ \epsilon_{e,i}^2 \right\}_{i=1}^t, \left\{ \epsilon_{e,i}^3 \right\}_{i=1}^t \right.\right) \\
& \propto P\left(\left\{ \epsilon_{e,i}^1 \right\}_{i=1}^t, \left\{ \epsilon_{e,i}^2 \right\}_{i=1}^t, \left\{ \epsilon_{e,i}^3 \right\}_{i=1}^t \left| L_e^1, L_e^2, L_e^3 \right.\right) \\
& \times P\left(L_e^1, L_e^2, L_e^3\right) \\
& \propto P\left(\left\{ \epsilon_{e,i}^1 \right\}_{i=1}^t \left| L_e^1\right.\right) P\left(\left\{ \epsilon_{e,i}^2 \right\}_{i=1}^t \left| L_e^2\right.\right) P\left(\left\{ \epsilon_{e,i}^3 \right\}_{i=1}^t \left| L_e^3\right.\right) \\
& P\left(L_e^3 \left| L_e^2\right.\right) P\left(L_e^2 \left| L_e^1\right.\right) P\left(L_e^1\right)
\end{aligned}$$

In this set of equations: $P\left(\left\{ \epsilon_{e,i}^M \right\}_{i=1}^t \left| L_e^M\right.\right)$, represents EEG observation likelihood for particular gesture at level M , estimated using the calibration data; $P\left(L_e^M \left| L_e^{M-1}\right.\right)$, is the probability of each command at level M given the inter-level information; and $P\left(L_e^1\right)$, represents the prior probability of each hand being target.

E. Decision Criteria

One decision is made in every epoch. Decisions are made using a Maximum-a-Posteriori (MAP) estimate.

$$\hat{s}_e = \arg \max_{s_e} P\left(s_e \left| \left\{ \epsilon_{e,i}^1 \right\}_{i=1}^t, \left\{ \epsilon_{e,i}^2 \right\}_{i=1}^t, \left\{ \epsilon_{e,i}^3 \right\}_{i=1}^t \right.\right) \quad (10)$$

At each time interval t , if the posterior probability of $\hat{s}_e$ is greater than the system *Confidence Threshold*, it will be chosen as the final decision of the epoch. If after N time interval (here $N=5$) the posterior probability is not greater than the system *Confidence Threshold*, the maximum posterior probability of $\hat{s}_e$ across all time intervals will be chosen as the final decision of the epoch. Analysis and results

III. Analysis and Results

Five healthy participants (3 males and 2 females) in the range of 20–30 years old consented and participated in a data collection session following an approved protocol by Northeastern University's IRB office. Participants were not under the influence of any chemicals, such as caffeine.

Before starting the experiment, different gestures were described to the participants; participants executed these gestures to get more familiar with the motions to be imagined during the imagination task. Participants were asked to imagine the entire progress of the

gestures, not only the final position. Between each trial (2 seconds preparation, 5 second gesture imagination), participants rested until they felt comfortable to continue. On average, the experiment consisting of 160 trials, 4 different gestures on each hand took less than 45 minutes.

All possible number of features (K) and window lengths ($t \geq 3s$) combinations were evaluated offline to find the parameter values leading to the highest classification accuracy. Figure 4 shows the total system accuracy as a function of the window length and number of features. As expected, more EEG evidence (larger t) results in higher accuracy. The average maximum accuracy of 64.5% (71% the highest) is achieved using $K = 6$ and the window length of $t = 5$ seconds using uniform inter-level prior probability. The average confusion matrix is shown in figure 5. The confusion matrix shows the system accurately classifies between left and right hand (level 1), hence the two approximately zero blocks on the top-right and bottom-left of the matrix (Green squares). The highest confusion is between classes that differ in thumb position (level 3), as shown by the highlighted red blocks. The confusion matrix represented in figure 5 shows gestures with differences in details are not as separable as the more general levels like left and right hand.

IV. DISCUSSION

With the goal of utilizing intuitive and natural gestures towards building motor imagery based classifiers. The natural gestures are more complex they reflect on the same region and require more sophisticated classifiers as opposed to the tasks performed or imagined on the similar limbs on the different sides. We proposed a hierarchical model combining binary classifiers on different levels taking advantage of the similarities in different gestures. Binary CSP is used to extract features at each level. The results from each classifier is probabilistically merged, and inter-level probabilities are employed to achieve a multiclass gesture classifier. In a study with 5 healthy participants our method resulted in an average accuracy of 64.5% among 8 complex hand gestures, more than 5 times the chance level.

Acknowledgments

Our work is supported by NSF (IIS-1149570, CNS-1544895), NIDLRR (90RE5017-02-01), and NIH (R01DC009834). Relevant code and data will be disseminated via Northeastern University Digital Repository Service, in collection NEU/COE/ECE/CSL (permalink: <http://hdl.handle.net/2047/D20199232>).

References

- [1]. Nezamfar H, Sinyukov D, Orhan U, Erdogmus D, and Padir T, "Brain interface to control a tele-operated robot," in Proceedings of the fifth international brain-computer interface meeting, 2013.
- [2]. Nezamfar H, Salehi SSM, Moghadamfalahi M, and Erdogmus D, "Flashtype: A context-aware c-vep-based bci typing interface using eeg signals," IEEE Journal of Selected Topics in Signal Processing, vol. 10, no. 5, pp. 932–941, 2016.
- [3]. Nezamfar H, Salehi SSM, and Erdogmus D, "Stimuli with opponent colors and higher bit rate enable higher accuracy for c-vep bci," in 2015 IEEE Signal Processing in Medicine and Biology Symposium (SPMB). IEEE, 2015, pp. 1–6.
- [4]. Yi W, Qiu S, Qi H, Zhang L, Wan B, and Ming D, "Eeg feature comparison and classification of simple and compound limb motor imagery," Journal of neuroengineering and rehabilitation, vol. 10, no. 1, p. 1, 2013. [PubMed: 23336711]

- [5]. Pfurtscheller G and Da Silva FL, "Event-related eeg/meg synchronization and desynchronization: basic principles," *Clinical neurophysiology*, vol. 110, no. 11, pp. 1842–1857, 1999. [PubMed: 10576479]
- [6]. Vu kovi A and Sepulveda F, "A two-stage four-class bci based on imaginary movements of the left and the right wrist," *Medical engineering & physics*, vol. 34, no. 7, pp. 964–971, 2012. [PubMed: 22119365]
- [7]. Pfurtscheller G, Neuper C, Flotzinger D, and Pregenzer M, "Eeg-based discrimination between imagination of right and left hand movement," *Electroencephalography and clinical Neurophysiology*, vol. 103, no. 6, pp. 642–651, 1997. [PubMed: 9546492]
- [8]. Hamed M, Salleh S-H, Noor AM, and Mohammad-Rezazadeh I, "Neural network-based three-class motor imagery classification using time-domain features for bci applications," in *Region 10 Symposium*, 2014 IEEE. IEEE, 2014, pp. 204–207.
- [9]. Vuckovic A and Sepulveda F, "Delta band contribution in cue based single trial classification of real and imaginary wrist movements," *Medical & biological engineering & computing*, vol. 46, no. 6, pp. 529–539, 2008. [PubMed: 18418635]
- [10]. Mohamed A, Marwala T, and John L, "Single-trial eeg discrimination between wrist and finger movement imagery and execution in a sensorimotor bci," in *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE. IEEE*, 2011, pp. 6289–6293.
- [11]. Edelman BJ, Baxter B, and He B, "Eeg source imaging enhances the decoding of complex right-hand motor imagery tasks," *Biomedical Engineering, IEEE Transactions on*, vol. 63, no. 1, pp. 4–14, 2016.
- [12]. Edelman B, Baxter B, and He B, "Discriminating hand gesture motor imagery tasks using cortical current density estimation," in *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE. IEEE*, 2014, pp. 1314–1317.
- [13]. Ramoser H, Muller-Gerking J, and Pfurtscheller G, "Optimal spatial filtering of single trial eeg during imagined hand movement," *Rehabilitation Engineering, IEEE Transactions on*, vol. 8, no. 4, pp. 441–446, 2000.

(a)

(b)

Fig. 1.

a) gestures for each hand. b) The experimental paradigm

Fig. 2.
Hierarchical classification model

Fig. 3.

System graphical model in epoch e . The final state s_e is composed of the state in each level L_e^M and thus deterministically related (dashed lines). States are calculated from the observed random variables and extracted EEG features of each level $\epsilon_{e_i}^M$. States and random variables are probabilistically related (solid lines).

Fig. 4. Average system accuracy using different number of features (x-axis) and different time-window lengths of EEG evidence.

Fig. 5.

Average confusion matrix for 5 participants with using $K = 6$ features and the window length of $t = 5$ seconds