Electromyogram (EMG) Feature Reduction Using Mutual Components Analysis for Multifunction Prosthetic Fingers Control

Rami N. Khushaba, *Member, IEEE,* Sarath Kodagoda, *Member, IEEE,* Dikai Liu, *Member, IEEE,* and Gamini Dissanayake, *Member, IEEE*

***Abstract*— Surface Electromyogram (EMG) signals are usually utilized as a control source for multifunction powered prostheses. A challenge that arises with the current demands of such prostheses is the ability to accurately control a large number of individual and combined fingers movements and to do so in a computationally efficient manner. As a response to such a challenge, we present a combined feature selection and projection algorithm, denoted as Mutual Components Analysis (MCA). The proposed MCA algorithm extends the well-known Principal Components Analysis (PCA) by pruning the noisy and redundant features before projecting the data. To implement the feature selection step, the mutual information concept is utilized to implement a new information gain evaluation function. The performance and significance of the proposed MCA is demonstrated on an EMG dataset collected for the purpose of this research from eight subjects with eight electrodes attached on their forearm. Fifteen classes of fingers movements where considered in this paper with MCA achieving >95% accuracy on average across all subjects.**

I. Introduction

Surface Electromyogram (EMG) signals recorded from an amputees residual muscles have been extensively investigated as a source of control for prosthetic devices [1], [2]. In such a scheme of control, denoted as myoelectric control, various feature sets and classification methods were utilized in the literature proving the feasibility of such an approach [3], [4]. Nowadays, prostheses increasingly incorporate humanoid robotic features with many degrees of freedom for which many researchers acknowledge the growing need for controlling these artificial multi-fingered dexterous hands [**?**], [5], [6].

Previous research on prosthetic fingers control proved the feasibility of classifying individual fingers movements using EMG signals. As an example, both Peleg *et al.* [7] and Tsenov *et al.* [8] employed two surface EMG electrodes to identify which individual finger is activated with various features extracted from the EMG signals. Tenore *et al.* [9] further extended the idea of EMG based finger control into movements that consisted of flexion and extension of all the fingers individually and of the middle, ring and little finger as a group achieving $\geq$ 98% accuracy with thirty-two electrodes [9], [10] and with fifteen electrodes [11]. According to Weir *et al.* [12], it can be difficult to obtain more than three or four stable, sufficiently uncorrelated control signals on a residual limb using surface EMG electrodes. Thus, a reduction in the number of electrodes, without compromising the classification accuracy, would significantly simplify the requirements for controlling a state of the art powered prosthetic.

Andrews *et al.* [13] targeted the optimal electrodes locations for EMG based finger control suggesting that a seven channel configuration results were not significantly different from that of three channels except for few subjects achieving <50% accuracy. This in turn suggests that either the extracted features were not informative enough for fingers movement classification across all subjects or the selected three electrodes cannot generalize well on different subjects. Additionally, no systematic approach for channel reduction was suggested as a brute-force-attack approach using all possible combinations was utilized. On the other hand, Cipriani *et al.* [14] performed real-time experiments on both amputees and able-bodied subjects using eight pairs of electrodes to classify seven fingers movements. These included two classes of combined fingers movements with an average accuracy of 79% (on amputees subjects)-to-89% (on able-bodied subjects) using the k-nearest neighbour (kNN) classifier. However, no experiments were conducted to validate the need for the total eight pairs of electrodes upon that of a smaller combination. Thus, the problem of selecting the most relevant channels, and therefore the features extracted from these channels, for fingers movement classification requires more investigation. Additionally, the number of important channels is also to be identified as using a large number of channels will result in a huge set of extracted features which in turn necessitates dimensionality reduction. However, not all of extracted features need to be considered when implementing the dimensionality reduction step as the multichannel approach might result in redundant features that may not correlate well with the class label. Therefore, to boost the performance of the dimensionality reduction methods, it is then necessary to remove the dimensions, which are unreliable for classification [15].

In this paper, a new algorithm for feature selection and projection is developed using a mutual information (MI for short) based evaluation function and the well-known Principal Components Analysis (PCA) and denoted as Mutual Component Analysis (MCA). The MI-based function evaluates and approximates the information content of each subset of features with regard to the output class. The approximated MI-based evaluation is considered as a starting point of a pruning algorithm that extracts a subset of relevant features from an initial set of available features and then further reduces their dimensionality using PCA. An impor-

All of the authors are within the School of Electrical, Mechanical, and Mechatronic Systems, Faculty of Engineering and Information Technology, University of Technology, Sydney (UTS). Email: Rami.Khushaba, Sarath.Kodagoda, Dikai.Liu, Gamini.Dissanayake@uts.edu.au

tant factor that justifies the applicability of PCA to EMG classification problems is caused by the large variance of the EMG signal. Englehart [16] proved that, in situations when the information is liberally dispersed amongst the original feature set, PCA can consolidate this information in a more effective manner when compared to feature selection methods. However, the fact that PCA does not consider the class label in the projection process limits the performance of PCA when compared to other projection methods (that all proved to outperform PCA), including PCA combined with a self-organized feature map (PCA+SOFM) [17] and Uncorrelated Linear Discriminant Analysis (ULDA) [18]. Thus, to provide some sort of class relevance to PCA, the MI evaluation function is proposed to rank the features according to their relevance to the problem and to encourage more independence with respect to the extracted features.

The structure of this paper is as follows: Section II reviews the concept of MI first and then describes the proposed MCA based feature selection and projection algorithm. Section III describes the data collection procedure and the utilized hardware. Section IV presents the experimental results and finally, conclusions are provided in Section V.

II. BACKGROUND

A. *Mutual Information Evaluation Function*

In information theory, the concept of MI is defined as the reduction of uncertainty about a random variable due to the knowledge of another random variable [19], [20]. The MI between two random variables X and Y, denoted as $I(X;Y)$, measures the amount of information in X that can be predicted when Y is known and is given as

$$I(X;Y) = \sum_{y\in Y}\sum_{x\in X} p(x,y)\log\frac{p(x,y)}{p(x)p(y)} \tag{1}$$

where $p(x,y)$ is the joint probability distribution function of X and Y, and $p(x)$ and $p(y)$ are the marginal probability distribution functions of X and Y respectively. Shannon entropy, which is a measure of uncertainty of random variables, is usually used to represent mutual information according to the following formula

$$\begin{aligned} I(X;Y) &= H(X) - H(X|Y) \\ &= H(Y) - H(Y|X) \\ &= H(X) + H(Y) - H(X,Y) \end{aligned} \tag{2}$$

where $H(X)$ and $H(Y)$ are the entropy of X and Y respectively, $H(X,Y)$ is their joint entropy, and $H(X|Y)$ and $H(Y|X)$ are the conditional entropies of X given Y and of Y given X, respectively.

In a learning task, X and Y can be any two features, i.e., f_1 and f_2, and $I(f_1;f_2)$ is used to reflect the amount of information *Redundancy* between the two features. When two features highly depend on each other, the respective class-discriminative power would not change much if one of them was removed. Alternatively, either f_1 or f_2 could be replaced by the class label C and $I(C;f_1)$ or $I(C;f_2)$ is used as a measure of *Relevance*, i.e., how relevant f_1 or f_2 is to the problem at hand that is characterized by the decisions in the class label. In terms of mutual information, given a set $\mathscr{F}$ of n features $\{f_1, f_2, ...f_n\}$, the purpose of feature selection is to find a feature subset $\mathscr{L}$ with p features $\{f_1, f_2, ...f_p\}$, where $p < n$, which jointly have the largest dependency on the target class C. To this end, various MI-based feature selection algorithms were proposed in the literature, these can be divided into three main categories of methods including: 1) methods that rank the features according to their individual relevance to the problem only using $I(C;f_i)$ [21], [22], 2) methods that continuously estimate the MI between each subset of features and the class label, i.e., $I(C;\mathscr{L})$ [23], [24], [25], [26], and 3) methods that consider both of the relevance and redundancy estimates when ranking the feature subsets to implement a "minimal-redundancy-maximal-relevance (MRMR)" criterion [27], [28], [29].

An incremental search procedure is usually utilized to find the near-optimal feature subset as defined by one of the above criteria. Subsets ranked according to the second category are usually more promising than that of other categories, but the computational cost of this category of methods grows very large especially when considering large feature subsets. In practice, the MRMR criterion shows better performance than that of individual feature ranking and less computational cost than that of feature subset ranking. Thus, we propose a new evaluation measure based on the MRMR category.

B. *Mutual Components Analysis (MCA) Algorithm*

Unlike most of the attempts from the literature, which utilize the MI concept in feature selection, we propose here a combination of MI-based feature ranking and PCA feature projection to implement a new feature projection method as shown in Fig.1. The first objective of the proposed MCA method is to rank the features according to their discriminative power to find a sufficient feature subset $\mathscr{L}$, given $\mathscr{L} \subseteq \mathscr{F}$, that can preserve the learning information

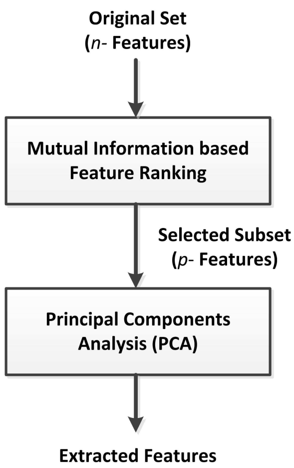


Fig. 1. Block diagram of the proposed MCA algorithm.

observed in the original feature set $\mathscr{F}$ with minimal losses, i.e., $I(C;\mathscr{F}) \approx I(C;\mathscr{L})$. Further, the second objective of MCA is induced by the high variance of the EMG features. Thus, to compact the variance of the ranked features, a PCA-projection step is implemented to map such features into a new domain with lower dimensionality to produce a set of uncorrelated features.

The steps of the MCA algorithm is given bellow. There, $I(C;f_i,f_j)$ is the mutual information between two features and the class and g is the proposed evaluation function for a given subset. All other parameters were defined earlier.

Algorithm-1: Mutual Information-Based Feature Reduction

- **Step-1**: Choose the feature $f_i \in \mathscr{F}$ that produces the maximum value of $I(C;f_i)$; set $\mathscr{L} \leftarrow f_i$; $g(\mathscr{L}) = I(C;f_i)$.
- **Step-2**: For each feature $f_i \in \mathscr{F}$, $f_i \notin \mathscr{L}$:
  - Compute: $m(f_i) = g(\mathscr{L}) + \lambda\ I(C;f_i)$, where λ represents the information gain.
  - Choose the feature f_i that maximizes m.
  - Set $\mathscr{F} \leftarrow \mathscr{F} \setminus \{f_i\}$, $\mathscr{L} \leftarrow \mathscr{L} \cup f_i$; $g(\mathscr{L}) = m(f_i)$.
- **Step-3**: If $|\mathscr{L}| < |\mathscr{F}|$ go to step 2.
- **Step-4**: $g(\mathscr{F}) = g(\mathscr{L})$.

In the first step, $g(.)$ is initialized to the maximum mutual information between a single feature and the class label $I(C;f_i)$. At the same time the subset of chosen features, $\mathscr{L}$ is initialized to $\{f_i\}$. Step 2 defines the intermediate function m of feature f_i , which is the summation of the latest value of $g(.)$ and its mutual information with the class label multiplied by λ as defined below.

$$\lambda = \frac{2}{1+exp(-\alpha D)} - 1, \tag{3}$$

$$D = \min_{f_j \in \mathscr{L}} \left[\frac{H(f_i) - I(f_i;f_j)}{H(f_i)}\right] \times \tag{4}$$
$$\frac{1}{|\mathscr{L}|} \sum_{f_j \in \mathscr{L}} \exp\left(\frac{I(C;f_i) + I(C;f_j) - I(C;\{f_i,f_j\})}{H(C)}\right)$$

where α is a small constant chosen (here, $\alpha = .3$ is empirically chosen) to make $\lambda \approx 0$, and $|\mathscr{L}|$ is the cardinal of $\mathscr{L}$.

This is an approximation of the amount of information added to $g(.)$ when choosing f_i. If f_i is completely redundant with respect to the already selected features then $H(f_i) = I(f_i;f_j)$ and $\lambda = 0$. On the other hand, if f_i is independent from all the features chosen so far then $(H(f_i) - I(f_i;f_j))/H(f_i) = 1$ as $I(f_i;f_j)$=0, and the second term will be equal to $(I(C;f_i) + I(C;f_j))/H(C)$ and λ is close to 1. If f_i is partially dependent upon any feature in $\mathscr{L}$ then λ will vary between 0 and 1.

- **Step-5**: normalize the resultant vector of gains yielded by adding each feature $m(f_i)$ by the corresponding subset length, i.e., $m(f_i) = m(f_i)/|\mathscr{L}|$. This is done to remove the subset size effect from the estimated gain.
- **Step-6**: select the first p features ($p < n$) with their corresponding $m(f_i)$ values being $> \frac{1}{|\mathscr{L}|}\sum_i m(f_i)$.
- **Step-7**: multiply the dataset with the selected subset of p-features by the eigenvectors of their corresponding co-variance matrix, i.e., apply PCA for feature projection.

For the testing phase, only the selected p features will be multiplied by the precomputed eigenvectors of the training data covariance matrix.

III. Data Collection

Eight subjects, six males and two females, aged between 20-35 years were recruited to perform the required fingers movements. The subjects were all normally limbed with no neurological or muscular disorders. All participants provided informed consent prior to participating in the study. Subjects were seated on an armchair, with their arm supported and fixed at one position. The datasets were recorded using eight EMG channels (DE 2.x series EMG sensors) mounted across the circumference of the forearm and processed by the Bagnoli desktop EMG system from Delsys Inc., as shown in Fig.2. A 2-slot adhesive skin interface was applied on each of the sensors to firmly stick them to the skin. A conductive adhesive reference electrode, dermatrode reference electrode, was placed on the wrist of each of the subjects during the experiments. The collected EMG signals were amplified using a Delsys Bagnoli-8 amplifier to a total gain of 1000. A 12-bit analog-to-digital converter (National Instruments, BNC-2090) was used to sample the signal at 4000 Hz; the signal data were then acquired using Delsys EMGWorks Acquisition software. The EMG signals were then bandpass filtered between 20-450 Hz with a notch filter implemented to remove the 50 Hz line interference.

Fifteen classes of movements were collected during this experiment including: the flexion of each of the individual fingers, i.e., Thumb (T), Index (I), Middle (M), Ring (R), Little (L) and the combined Thumb-Index (T-I), Thumb-Middle (T-M), Thumb-Ring (T-R), Thumb-Little (T-L), Index-Middle (I-M), Middle-Ring (M-R), Ring-Little (R-L), Index-Middle-Ring (I-M-R), Middle-Ring-Little (M-R-L), and finally the hand close class (HC) as shown in Fig.3. When collecting data, the subjects were asked to perform each of the aforementioned fifteen movements, and hold that movement for a period of 5 seconds in each trial. Six trials, or repetitions, of each movement were collected. Four trials from each movement data were allocated for training and the remaining 2 trials were allocated for testing.

IV. Experiments and Results

In the proposed EMG-pattern recognition system shown in Fig.4, an analysis window size of 128 msec that was incremented by 64 msec was utilized when extracting the features. Various features in time and frequency domain were extracted from each of the analysis windows to represent the EMG activity. These included all of the followings (reader can refer to [3] for details): number of zero crossings (1

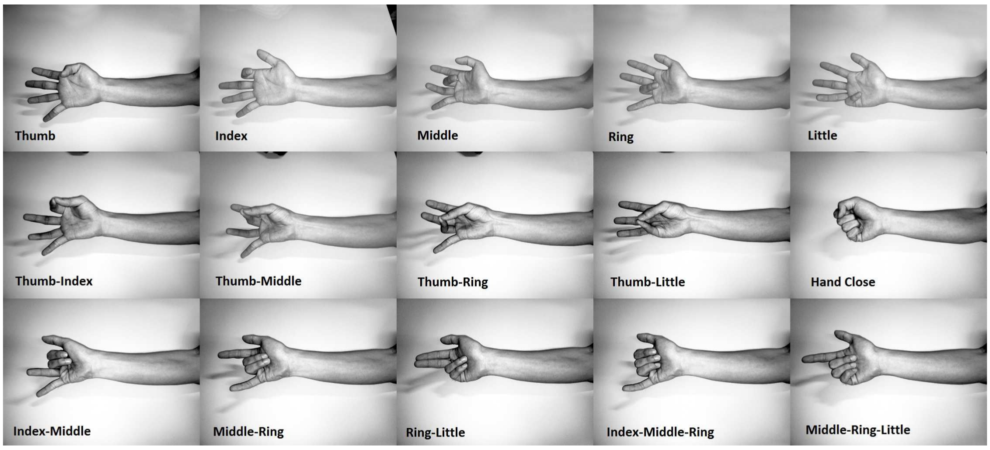
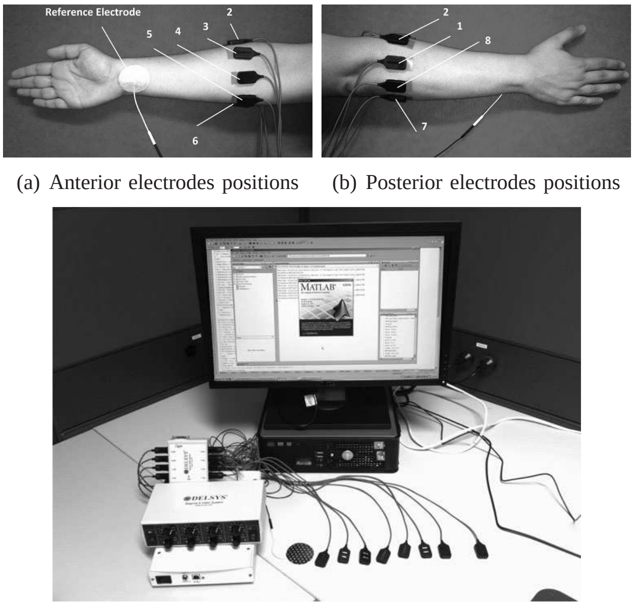


Fig. 3. Different classes of individual and combined fingers movement considered in this paper.



(a) Anterior electrodes positions (b) Posterior electrodes positions

(c) The EMG System hardware utilized in experiments.

Fig. 2. Electrodes placement on the right forearm

feature), waveform length (1 feature), number of slope-sign changes (1 feature), skewness (1 feature), root-mean-square (1 feature), mean absolute value (1 feature), integral absolute value (1 feature), parameters of an autoregressive (AR) model with an order of 11 providing significant enhancements upon smaller model orders (11 features), and the Hjorth time-domain parameters (3 features) totaling 21 features/channel. In a problem of 8 channels, the total number of extracted features is 168 features (21 feature/channel $\times$ 8 channels = 168 features).

In the classification step, different classifiers were utilized in the experiments including: Support Vector machine (SVM) with the LIBSVM implementation [30], k-Nearest neighbor kNN classifier, and the Extreme Learning Machine (ELM) classifier [31]. The LIBSVM classifier parameters were optimized as: (SVM type: C-SVC), cost parameter for C-SVC was $c = 8$ and kernel type was set to a radial basis function with $\gamma = 12/\text{Number of features}$. On the other hand, the number of neighbors in kNN was set to 5 (selected experimentally) and the number of hidden nodes in ELM was set to 50. In the final step, a majority vote postprocessing step was utilized to smooth the output of the classifier and to further enhance the classification accuracy [18], [32]. For a given decision point d_i, the majority vote decision smooths the classifier output by also considering the previous m decisions (m=8 in this experiment). The value of d_{mv} is simply the class label with the greatest number of occurrences in this point window of the decision stream.

In the rest of this section, we investigate the significance of the proposed MCA algorithm when selecting a subset of features from different channels and then projrcting them with PCA. During the experiments, only the first 20 principal components are utilized to minimize the computational cost as using larger number of principal components provided nearly the same results with no statistical significant difference.

The performance of the proposed MCA algorithm is compared against all of the: traditional PCA, discriminant analysis based feature projection methods (including LDA and ULDA), and sparse PCA (denoted as SPCA) [33]. SPCA was included here as the proposed MCA also attempts to function in a similar manner to SPCA. SPCA employs a least angle selection and shrinkage operator (LASSO) or elastic net approach to solve an optimization problem that penalizes the l_1-norm and l_2-norm of the model error while implementing a feature selection step. Thus, SPCA is also considered as a combination of feature selection and projection. In order to demonstrate the performance of the proposed mutual information measure, we employ both PCA and SPCA in the second stage. The combination of the MI-based feature ranking function with PCA will be denoted as MCA1 while that with SPCA will be denoted as MCA2.

The classification error rates achieved by the proposed MCA algorithm in comparison to PCA, SPCA, LDA and ULDA are shown in Fig.5 using three different classifiers.

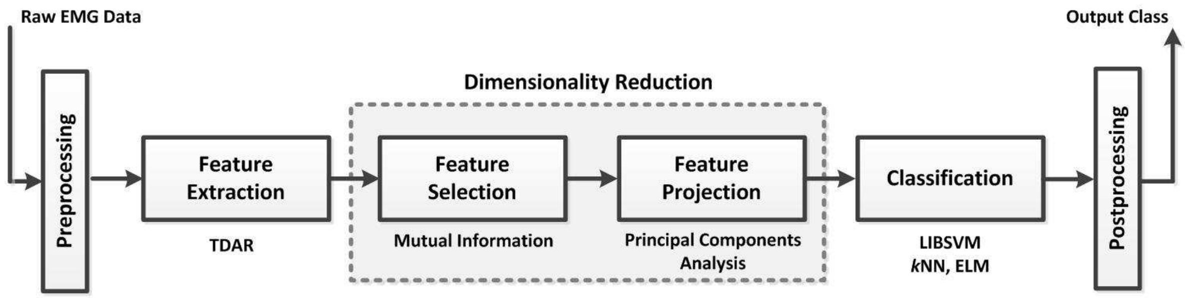


Fig. 4. Block diagram of the proposed EMG classification system.

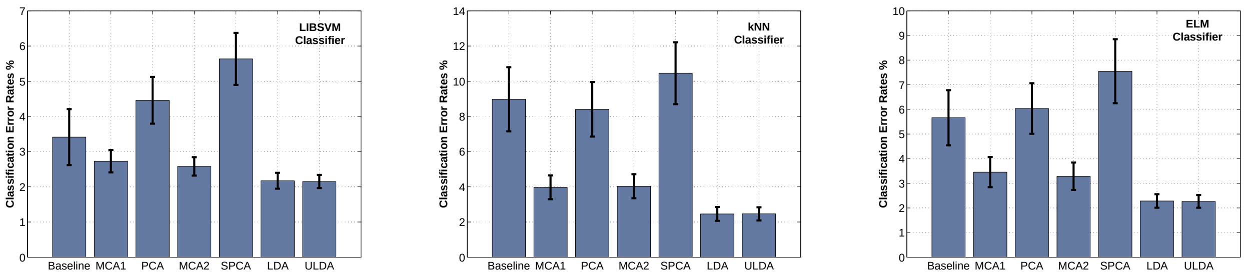
One can clearly notice that in the current EMG classification problem, canonical PCA achieved on average, across different subject and classifiers, better performance than that exhibited by SPCA. Such a performance might be justified by the feature selection approach utilized within SPCA which might not be very powerful when applied on EMG signals due to the high variance nature of EMG. In SPCA, the loadings, i.e., the weights of the projection matrix are found by a regression-type optimization problem that continuously shrinks the coefficients toward zero while attempting to minimize the representation error [33]. In simple words, the sparsity of SPCA is oriented toward selecting a projection matrix with minimal parameters that guarantee accurate representation of the original data matrix. However, SPCA, just like PCA, pays no attention to the distribution of the class information. In such a case, selecting feature subsets that do not convey the same amount of class relevant information as the original feature set will result in large misclassification rates which explains the low performance of both PCA and SPCA in comparison to Baseline. On the other hand, our MI-based feature ranking scheme attempts to select small subsets with features that best interact together to increase the class relevance information while minimizing redundancy. This was proved by applying the MI-based feature ranking with SPCA. In such a case, the feature matrix that SPCA works on represents the most relevant subset of features and SPCA only attempts to minimize the number of coefficients used to represent such a subset of features. Thus, both MCA1 and MCA2 achieved nearly similar performance with MCA2 being slightly better than MCA1 and both outperforming PCA and SPCA and the Baseline (Baseline represent the use of all features in classification without feature projection).

On the other hand, it is clearly obvious that, in terms of classification error rates, LDA and ULDA achieved better performance than all of PCA, SPCA, MCA1, and MCA2. However, one should note two important issues related to LDA and ULDA. The first that LDA fails without a proper regularization step or without preprocessing the features first with PCA to remove the correlation that produces singularity. On the other hand, ULDA employs the singular value decomposition twice which is computational very expensive in comparison to PCA with Eigen-decompositions.

In order to validate the significance of the achieved results with MCA1 and MCA2 upon that of the rest of the methods, we performed a Bonferroni corrected t-test of the hypothesis that the classification performance of MCA1 and MCA2 is not significantly different from that of the rest of the methods. The desired experiment-wise significance level was set to ALPHA = 0.05 with the Bonferroni criteria effectively dividing ALPHA by the number of t-tests being performed. Experimental results with H=0 indicates that the null hypothesis of similar performance by the compared methods cannot be rejected at the 0.05 significance level; while H=1 indicates that the null hypothesis can be rejected at the 0.05 level. The test results are given in Table.I with p representing the p-value. These results indicate that, on average across different classifiers, there was no significant difference between the performance of MCA1 and MCA2 upon that of LDA and ULDA. Thus, using the features selected by mean of MI-maximization (in the first step) with PCA and SPCA (as a second step) achieved a similar performance to that of LDA and ULDA without direct use of the class label in the projection process, with H=0 in the tests of MCA1 and MCA2 against LDA and ULDA. On the other hand, H=1 indicated the significant differences between the performance of MCA1 and MCA2 upon that of PCA and SPCA that combine all of the features in the projection step. This in turn proves the effectiveness of selecting the most relevant subset of features before the projection process.

However, the advantage of MCA1 and MCA2 is not in the classification performance only, but also in the associated computational cost of the MCA1 and MCA2. As best and worst cases, a computational time of 0.449 sec was required for MCA1, including mutual information computation and PCA stage, compared to 6.991 sec for ULDA on a PC with 3 GHz CPU and 2 GB of RAM on Matlab [1]. The MCA1 and MCA2 algorithms were able to compete with LDA and ULDA by using only a portion of the total number of features that LDA and ULDA acted upon in the projection step. In order to observe number of retained features in the MI-based subset selection step we present in Table.II the number of retained features from each subject's dataset out of the original 168 extracted features. It is clear that MCA1 and MCA2 acted only upon a small portion of the features

[1]Mutual information implementation acquired from http://penglab.janelia.org/software/



(a) Using LIBSVM classifier (b) Using kNN classifier (c) Using ELM classifier

Fig. 5. Classification error rates achieved using TDAR feature set with different feature projection methods and classifiers, with bars indicating standard error.

TABLE I
BONFERRONI CORRECTED T-TEST RESULT ON THE CLASSIFICATION PERFORMANCE OF MCA1 AND MCA2 VERSUS THE REST OF THE PROJECTION METHODS

Classifier	Method	**MCA1**	**MCA2**	**PCA**	**SPCA**	**LDA**	**ULDA**
LIBSVM	**MCA1 vs. →**		H= 0 p= 0.8242	H= 1 p= 0.0091	H=1 p=1.6677e-005	H=0 p=0.3984	H=0 p=0.3805
	MCA2 vs. →	H= 0 p= 0.8242		H= 1 p= 0.0048	H=1 p=6.4902e-006	H=0 p=0.5334	H=0 p=0.5123
kNN	**MCA1 vs. →**		H= 0 p= 0.9664	H= 1 p= 0.0023	H=1 p=1.1357e-005	H=0 p=0.2927	H=0 p=0.2938
	MCA2 vs. →	H= 0 p= 0.9664		H= 1 p= 0.0027	H= 1 p= 1.3568e-005	H= 0 p= 0.2738	H= 0 p= 0.2749
ELM	**MCA1 vs. →**		H= 0 p= 0.8666	H= 1 p= 0.0103	H= 1 p= 5.9388e-005	H= 0 p= 0.2406	H= 0 p= 0.2336
	MCA2 vs. →	H= 0 p= 0.8666		H= 1 p= 0.0063	H=1 p=3.0253e-005	H=0 p=0.3143	H=0 p=0.3058

extracted originally from each subject's EMG data while all of the PCA, SPCA, LDA, ULDA, and Baseline acted upon all 168 features. However, not all of the features might be effective for classification. Thus, the selection of the most important subset that maximizes the class relevance while minimizing redundancy is of great importance to the accurate performance of an EMG-driven system.

TABLE II
NUMBER OF RETAINED FEATURES IN THE FEATURE SELECTION STEP

Subjects	S1	S2	S3	S4	S5	S6	S7	S8
No. of features	43	43	44	43	43	42	42	43

Given that small subsets of features proved very significant to the current problem, then the next step in the experiments is to discover the regions from which the MI-based feature selection function ranked the extracted features as being of high importance to the problem. Observing the features ranked by the MI-based function indicated that most of the features ranked very important for the current classification problem were obtained from surface EMG electrodes placed over the extensor digitorium, extensor carpi ulnaris, and fexlor carpi ulnaris muscles. In comparison to the results given by Hargrove *et al.* [34], these muscles, in addition to others, were also shown to be the most promising regions for performing ten classes of forearm movements. However, the proposed single and combined fingers movements classes are more complicated as they involve movements of the same finger individually and in a subset with other fingers, which makes the separation between the classes more complicated than that of gross arm movements. In such a case, the proposed method was capable of achieving $>95.61\%$ accuracy on average across all subjects with ≥ 4 EMG channels proving the effectiveness of the proposed system. Additionally, although the location of the optimal channels was not the same between all subjects, it should be noted that there was some variation in the manner the surface electrode was applied between subjects due to the differences in the physical dimensions of the subjects forearm. However, the aforementioned regions were in general the most informative regions providing accurate discriminant information for the separation of the 15 classes of fingers movements investigated in this paper.

V. CONCLUSION

A new algorithm for EMG feature reduction was proposed in this paper. The proposed MCA algorithm selects the most promising feature subset using a mutual information based evaluation function that attempts to maximize the subset relevance to the problem while minimizing information redundancy. The selected features were then projected using either PCA or SPCA to further reduce the dimensionality of the extracted features. Both of the MCA1 (with PCA in the second step) and MCA2 (with SPCA in the second step) provided significant enhancements upon the original PCA and SPCA with no feature selection. This in turn proves the effectiveness of selecting a subset of the original features before the feature projection step. The justification is that projection based methods combines the effect of all features together; however, certain features might be irrelevant for the problem at hand and their existence might decrease the classification accuracies if considered in a subset with the most promising features. On the other hand, the performance of both MCA1 and MCA2 proved to be of no significant difference upon that of LDA and ULDA with MCA1 and MCA2 using less than half of the feature set utilized by LDA and ULDA. Finally, the channel selection problem using the proposed MI-based channel ranking and selection algorithm proved that only four channels were required to achieve accurate classification results across eight subjects performing 15 classes of individual and combined fingers movements.

REFERENCES

[1] R. Merletti and P. A. Parker, ”Electromyography Physiology, Engineering, and Noninvasive Applications”, IEEE Press Series in Biomedical Engineering, John Wiley and Sons Inc, Hoboken, New Jersey, 2004.

[2] K. Englehart and B. Hudgins, ”A Robust, Real-Time Control Scheme for Multifunction Myoelectric Control”, *IEEE Transactions on Biomedical Engineering*, vol. 50, no. 7, pp. 848-854, 2003.

[3] M. A. Oskoei and H. Hu, ”Myoelectric control systems-A survey”, *Biomedical Signal Processing and Control*, vol. 2, no. 4, pp. 275-294, 2007.

[4] M. A. Oskoei and H. Hu, ”Support Vector Machine-Based Classification Scheme for Myoelectric Control Applied to Upper Limb”, *IEEE Transactions on Biomedical Engineering*, vol. 55, no. 8, pp. 1956-1965, 2008.

[5] S. Adee, ”The revolution will be prosthetized”, *IEEE Spectrum*, pp. 37-40, Jan. 2009.

[6] R. J. Smith, D. Huberdeau, F. Tenore, and N. V. Thakor, ”Real-Time Myoelectric Decoding of Individual Finger Movements For a Virtual Target Task”, Proceedings of the 31st Annual International Conference of the IEEE EMBS, pp. 2376-2379, 2009.

[7] J. J. Baker, E. Scheme, K. Englehart, D. T. Hutchinson, and B. Greger, ”Continuous Detection and Decoding of Dexterous Finger Flexions With Implantable MyoElectric Sensors”, *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 18, no. 4, 2010.

[8] D. Peleg, E. Braiman, E. Yom-Tov, and G. F. Inbar, ”Classification of Finger Activation for Use in a Robotic Prostheses Arm”, *IEEE Transactions on Biomedical Engineering*, vol. 10, no. 4, pp. 290-293, 2002.

[9] G. Tsenov, A. H. Zeghbib, F. Palis, N. Shoylev, V. Mladenov, ”Neural Networks for Online Classification of Hand and Finger Movements Using Surface EMG signals”, Proceedings of the 8th Seminar on Neural Network Applications in Electrical Engineering, pp. 167-171, 2006.

[10] F. Tenore, A. Ramos, A. Fahmy, S. Acharya, R. Etienne-Cummings, and N. V. Thakor, ”Toward the Control of Individual Fingers of a Prosthetic Hand Using Surface EMG Signals”, Proceedings of the 29th Annual International Conference of the IEEE EMBS, pp.6146-6149, 2007.

[11] F. V. G. Tenore, A. Ramos, A. Fahmy, S. Acharya, R. Etienne-Cummings, N. V. Thakor, ”Decoding of Individual Finger Movements Using Surface Electrmyography”, *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 5, pp. 1427-1434, 2009.

[12] R. J. Smith, F. Tenore, D. Huberdeau, R. Etienne-Cummings, and N. V. Thakor, ”Continuous Decoding of Finger Position from Surface EMG Signals for the Control of Powered Prostheses”, Proceedings of the 30th Annual International Conference of the IEEE EMBS, pp. 197-200, 2008.

[13] R. Weir, P. R. Troyk, G. DeMichele, D. A. Kerns, J. F. Schorsch, and H. Maas, ”Implantable myoelectric sensors (IMESs) for intramuscular electromyogram recording”, *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 1, pp. 159-171, 2009.

[14] A. Andrews, E. Morin, and L. Mclean, ”Optimal Electrode Configurations for Finger Movement Classification using EMG”, Proceedings of the 31st Annual International Conference of the IEEE EMBS, pp. 2987-2990, 2009.

[15] C. Cipriani, C. Antfolk, M. Controzzi, G. Lundborg, B. Rosen, M. C. Carrozza, and F. Sebelius, ”Online Myoelectric Control of a Dexterous Hand Prosthesis by Transradial Amputees”, *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. PP, no. 99, DOI:10.1109/TNSRE.2011.2108667, 2011.

[16] X. Jiang, ”Linear Subspace Learning-Based Dimensionality Reduction”, *IEEE Signal Processing Magazine*, vol. 28, no. 2, pp. 16-26, 2011.

[17] K. Englehart, ”Signal representation for classification of the transient myoelectric signal”, PhD Thesis, Department of Electrical and Computer Engineering, University of New Brunswick, 1998.

[18] J. U. Chu, I. Moon, and M. S. Mun, ”A real-time EMG pattern recognition system based on linear-nonlinear feature projection for a multifunction myoelectric hand”, IEEE Transactions on Biomedical Engineering, vol. 53, no. 11, pp. 2232-2239, 2006.

[19] A. D. C. Chan and G. C. Green, ”Myoelectric control development toolbox”, In Proceedings of the 30th Conference of the Canadian Medical and Biological Engineering Society, Toronto, Canada, M0100, 2007.

[20] G. J. Klir, ”Uncertainty and Information: Foundations of Generalized Information Theory”, John Wiley and Sons, Inc, New Jersey, USA, 2006.

[21] T. M. Cover and J. A. Thomas, ”Elements of information theory, second edition”, John Wiley and Sons, Inc., New Jersey, USA, 2006.

[22] H. Wang, D. Bell, and F. Murtagh, ”Feature Subset Selection based on Relevance”, *Vistas in Astronomy*, vol. 41, no. 3, pp. 387-396, 1997.

[23] H. Wang, D. Bell, and F. Murtagh, ”Axiomatic Approach to Feature Subset Selection Based on Relevance”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 3, pp. 271-277, 1999.

[24] N. Kwak and C. H. Choi, ”Input Feature Selection by Mutual Information Based on Parzen Window”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 12, pp. 271-277, 2002.

[25] T. W. S. Chow and D. Huang, ”Estimating Optimal Feature Subsets Using Efficient Estimation of High-Dimensional Mutual Information”, *IEEE Transactions on Neural Networks*, vol. 16, no. 1, pp. 213-224, 2005.

[26] V. G. Verdejo, M. Verleysen, and J. Fleury, ”Information-theoretic feature selection for functional data classification”, *Neurocomputing*, vol. 72, no. 16-18, pp. 3580-3589, 2009.

[27] M. A. Kerroum, A. Hammoucha, and D. Aboutajdine, ”Textural feature selection by joint mutual information based on Gaussian mixture model for multispectral image classification”, *Pattern Recognition Letters*, vol. 31, no. 10, pp. 1168-1174, 2010.

[28] R. Battiti, ”Using Mutual Information for Selecting Features in Supervised neural Net Learning”, *IEEE Transactions on Neural Networks*, vol. 5, no. 4, pp. 537-550, 1994.

[29] A. Al-Ani, M. Deriche and J. Chebil, ”A new mutual information based measure for feature selection”, *Intelligent Data Analysis*, vol. 7, no. 1, pp 43-57, 2003.

[30] P. A. Estevez, M. Tesmer, C. A. Perez, and J. M. Zurada, ”Normalized Mutual Information Feature Selection”, *IEEE Transactions on Neural Networks*, vol. 20, no. 2, pp. 189-201, 2009.

[31] C. C. Chang and C. J. Lin, ”LIBSVM: A library for support vector machines”, Available Online: http://www.csie.ntu.edu.tw/ cjlin/libsvm/, 2001.

[32] G. B. Hunag, Q. Y. Zhu, and C. K. Siew, ”Extreme Learning Machine: Theory and Applications”, Neurocomputing, vol. 70, pp. 489-501, 2006.

[33] A. D. C. Chan and K. B. Englehart, ”Continuous myoelectric control for powered prostheses using hidden Markov models”, *IEEE Transactions on Biomedical Engineering*, vol. 52, no. 1, pp. 121-124, 2005.

[34] H. Zou, T. Hastie and R. Tibshirani, ”Sparse Principal Component Analysis”, *Journal of Computational and Graphical Statistics*, vol. 15, no. 2, pp. 265-286, 2006.

[35] L. Hargrove, K. Englehart, and B. Hudgins, ”A comparison of surface and intramuscular myoelectric signal classification”, IEEE Transactions on Biomedical Engineering, vol. 54, no. 5, pp. 847-853, 2007.