Support Vector Machine-Based Classification Scheme for Myoelectric Control Applied to Upper Limb

Mohammadreza Asghari Oskoei*, *Student Member, IEEE*, and Huosheng Hu, *Senior Member, IEEE*

***Abstract*—This paper proposes and evaluates the application of support vector machine (SVM) to classify upper limb motions using myoelectric signals. It explores the optimum configuration of SVM-based myoelectric control, by suggesting an advantageous data segmentation technique, feature set, model selection approach for SVM, and postprocessing methods. This work presents a method to adjust SVM parameters before classification, and examines overlapped segmentation and majority voting as two techniques to improve controller performance. A SVM, as the core of classification in myoelectric control, is compared with two commonly used classifiers: linear discriminant analysis (LDA) and multilayer perceptron (MLP) neural networks. It demonstrates exceptional accuracy, robust performance, and low computational load. The entropy of the output of the classifier is also examined as an online index to evaluate the correctness of classification; this can be used by online training for long-term myoelectric control operations.**

***Index Terms*—Classification, data segmentation, entropy, feature selection, myoelectric control, support vector machine (SVM).**

I. Introduction

Since most disabled people have problems manipulating current assistive robots and rehabilitation devices that employ traditional user interfaces, such as a joystick and/or keyboard, they need more advanced hands-free human–machine interfaces. Surface myoelectric signals (MESs), which are collected from the skin covering muscles, contain rich information that can be used to recognize neuromuscular activity in a noninvasive manner. A myoelectric control system (MCS), which uses MES as a reference input signal to manipulate assistive robots or rehabilitation devices, has the potential to become a competent alternative to traditional body-powered user interfaces [28]. A pattern-recognition-based MCS, functions by recognizing different pretrained signal patterns, and applies corresponding predefined motion commands to electric motors and/or actuators. It provides hands-free and proximal function control based on neuromuscular activity. The success of myoelectric control depends greatly on classification accuracy. Effective feature extraction and classification methods are crucial to achieving high classification performance in pattern recognition.

Manuscript received September 13, 2007; revised November 16 and December 27, 2007. *Asterisk indicates corresponding author.*

*M. A. Oskoei is with the Department of Computing and Electronic Systems, University of Essex, Colchester CO4 3SQ, U.K.(e-mail: masgha@essex.ac.uk).

H. Hu is with the Department of Computing and Electronic Systems, University of Essex, Colchester CO4 3SQ, U.K. (e-mail: hhu@essex.ac.uk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TBME.2008.919734

The application of pattern recognition to myoelectric control schemes was first introduced in the 1960s–1970s [1], [2]; however, due to limited acquisition instruments and computing capacity at that time, real-time control was not feasible. Hudgins *et al.* [3] were pioneers in developing a real-time pattern-recognition-based MCS. Using time-domain (TD) features and a multilayer perceptron (MLP) neural network, they succeeded in classifying four types of upper limb motion, with an accuracy of approximately 90%. This work was continued over the last 15 years, by employing various classifiers, such as linear discriminant analysis (LDA) [4], [5], MLP/radial basis function (RBF) neural networks [6], time-delayed artificial neural network (ANN) [7], fuzzy [8], [9], NEURO-fuzzy [10], fuzzy ARTMAP networks [11], fuzzy-MINMAX networks [12], Gaussian mixture models (GMMs) [13]–[15], and hidden Markov models (HMMs) [16].

Vuskovic and Du [11] introduced a modified version of a fuzzy ARTMAP network to classify prehensile MESs. Englehart *et al.* [4] showed that LDA, outperforms MLP on time-scale features that are dimensionally reduced by PCA. In addition, significant results were achieved using probabilistic approaches. Chan and Englehart [16] applied an HMM to discriminate six classes of limb movement based on a four-channel MES. It resulted in an average accuracy of 94.63%, which exceeded an MLP-based classifier used in [5] (93.27%). Furthermore, Huang *et al.* [13] and Fukuda *et al.* [14] developed a GMM as a classifier in their MCS; the former showed an accuracy of approximately 97%. Englehart *et al.* [5] introduced a continuous classification scheme that provided more robust results for a shortened segment length of signal, and high-speed controllers. In this paper, we apply a support vector machine (SVM) as a classifier to recognize upper limb motion from surface MES patterns.

The SVM is a kernel-based approach with a strong theoretical background, which has become an increasingly popular tool for machine learning tasks involving classification and regression. It has recently been successfully applied to several applications, ranging from face identification [17] and text categorization [18] to bioinformatics and database mining [19]. An SVM constructs an optimal separating hyperplane in a high-dimension feature space of training data that are mapped using a nonlinear kernel function. Therefore, although it uses a linear learning machine method with respect to the nonlinear kernel function, it is in effect a nonlinear classifier. The use of a nonlinear kernel function greatly increases the power of learning and generalization. However, it can also increase the risk of overfitting, which may lead to bad generalization. Therefore, the flexibility of the kernel-induced feature space is controlled by setting an upper band for generalization risk. Having control on the generalization performance improves the robustness against noisy data such as MES. Training an SVM involves the optimization of a convex cost function; there are relatively few free parameters to adjust, and the architecture does not have to be found via experimentation.

The rest of this paper is organized as follows. In the next section, we introduce nonlinear SVM, and then, multiclass classification. Section III presents methodologies used for data acquisition, experiments, analysis, and evaluation. Section IV provides experimental study and statistical analysis in which the impact of segmentation methods, features, classifiers, and postprocessing are examined with respect to classification performance. In addition, the validity of entropy as an index for classification correctness is investigated. Finally, in Section V, conclusions are presented.

II. Background

A. Binary SVM

In SVM, training is reformulated in such a way as to obtain a quadratic programming (QP) problem. The solution to this QP problem is global and unique. For the empirical data $(x_1, y_1), \ldots, (x_m, y_m) \in R^N \times \{\pm 1\}$ that are mapped by $\phi : R^N \mapsto F$ into a "feature space," the linear hyperplanes that divide them into two labeled classes can be shown as

$$w \times \phi(x) + b = 0 \quad w \in R^N, \quad b \in R. \tag{1}$$

To construct an optimal hyperplane with maximum-margin and bounded error in the training data (soft margin), one solves the following QP problem

$$\begin{gathered} \min_{w,b} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{m} \xi_i \\ y_i(w \times \phi(x_i) + b) \geq 1 - \xi_i, \quad i = 1, \ldots, m. \end{gathered} \tag{2}$$

The first term in cost function (2) makes maximum margin of separation between classes, and the second term provides an upper bound for the error in the training data. Due to the noise, which regularly accompanies the MES, error in the training data (i.e., the samples that are placed on the wrong side of the hyperplane shown by ξ_i) is practically inevitable. Meanwhile, the bounded error in the training data can prevent overfitting. The constant $C \in [0, \infty)$ creates a tradeoff between the number of misclassified samples in the training set and separation of the rest samples with maximum margin. A way to solve (2) is via its Lagrange function. Given kernel $k(x_i, x_j) = \phi(x_i).\phi(x_j)$, the Lagrange function of (2) is simplified to

$$\max_{\alpha} \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{m} \alpha_i \alpha_j y_i y_j k(x_i, x_j) \tag{3}$$

$$w = \sum_{i=1}^{m} y_i \alpha_i \phi(x_i), \sum_{i=1}^{m} \alpha_i y_i = 0, \quad 0 \leq \alpha_i \leq C \quad \forall i. \tag{4}$$

The (4) shows that the optimal hyperplane, in feature space, can be written as the linear combination of training samples with $\alpha_i \neq 0$. These informative samples, known as support vectors, construct the decision function of the classifier based on the kernel function

$$f(x) = \operatorname{sgn}\left(\sum_{i=1}^{m} y_i \alpha_i k(x, x_j) + b \right). \tag{5}$$

There are different kernels, which are often selected based on the data structure and type of the boundaries between classes, the most popular ones are: linear: $k(x_i, x_j) = x_i.x_j$; polynomial: $k(x_i, x_j) = (\gamma x_i.x_j + r)^d$; RBF: $k(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$; and sigmoid: $k(x_i, x_j) = \tanh(\gamma x_i.x_j + r)$. Here, $r, \gamma, d > 0$ are the kernel parameters. Generally, RBF is suggested for use in unknown applications. More details about the SVM are available in [21] and [27].

B. Multiclass SVM

The SVM, inherently, is a binary classifier, while many problems we are interested in solving (such as the classification of limb motions), are multiclass. An SVM performs very well for binary problems; it is desirable to extend its capabilities to multiclass problems. There are two approaches to a multiclass SVM. One involves directly considering all data in a single optimization problem, while the other involves constructing and combining several binary classifiers. The literatures show that the former approach does not offer any additional advantages over the combined binary SVMs [22], [23].

Probably the simplest scheme for the k-class classification problem is to train k independent binary classifiers that are each trained to distinguish training samples for one class with regard to remaining classes. This scheme is referred to as the "one-against-all" or OAA. It is very simple to implement, relatively fast running, obvious, and produces results that are often as/more accurate as/than other methods. Another method is the "one-against-one" or OAO. In this method, $k(k–1)$ binary classifiers are trained to separate a pair of two classes. To classify a new sample, a class that gains most votes of the binary classifiers is chosen as the final output. This method same as OAA has a simple conceptual justification and can be implemented quickly. The advantage of OAO is that it conducts binary classifications on all pairs of classes, and computes the probability for each class [24]. This supports the analytic concept for generalization and certainty. Given that r_{ij} is an estimate for the probability of the output of a pair wise classifier between class i and j (i.e., $r_{ij} \approx P(y = i | y = \{i, j\}, x)$, $r_{ij} + r_{ji} = 1$), and that p_i is the probability of the i$^{\text{th}}$ class, the class probability $p = (p_1, \ldots, p_k)$ can be derived via a QP problem (6)

$$\min_{p} \sum_{i=1}^{k} \sum_{j:j \neq i} (r_{ij} p_j - r_{ji} p_i)^2, \sum_{i=1}^{k} p_i = 1, \quad p_i \geq 0, \quad \forall i. \tag{6}$$

This paper employs library for support vector machines (LIBSVM) [20] as the core of an SVM classifier, and conducts multiclass classifications using C-SVM and the OAO method.

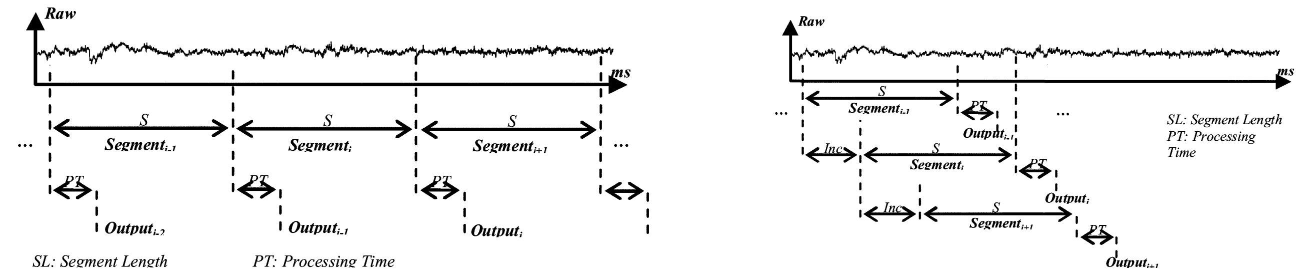

Fig. 1. Disjoint (left) and overlapped (right) segmentation.

III. Methodology

This section introduces methods applied for data collection, experiments, and analysis. It is comprised of different sections for data acquisition, data segmentation, feature selection, classification, postprocessing, and evaluation.

A. Data Acquisition

A four-channel MES was collected from four locations on a forearm (i.e., biarticulate wrist flexor and triarticulate and biarticulate wrist extensor muscles), using bipolar active electrodes (Biometrics, Ltd., SX230). An active electrode has a preamplifier with gain 1000, which can differentiate between a small signal of interest and much larger interference signals that are present on the skin. It also has a very high input impedance to cope with mismatches in skin contact resistance. Signals are passed through a high-pass filter with a cutoff frequency of 20 Hz to remove dc offsets due to membrane potentials, and to minimize interference due to electrode movement. A low-pass filter is used to remove unwanted frequencies above 450 Hz, and a notch filter used to remove unwanted line-frequencies (50/60 Hz). An electrode was also placed on the wrist, providing a common ground reference. Signals were sampled at 1000 Hz using a 12-b A/D converter.

Data were collected from 11 healthy subjects. Each subject performed five limb motions, and rest to provide six distinct states (i.e., classes). The motions were isotonic and comprised of flexion, extension, abduction, adduction, and keeping the hand straight. Two sequences of six motions in which each motion was held fixed for the five seconds are called a block. Four blocks of data were gathered from subjects in each session. Two sessions were conducted for each subject, and in each session, the accuracy of classification was computed using a fourfold cross-validation method.

B. Data Segmentation

A segment is a sequence of data limited in a time slot, which is used to estimate signal features. A short length of segment leads to bias and variance in feature estimation; while a long one imposes high computational load and a likely failure to perform real-time operation. Real-time constraints enforce a delay time of less than 300 ms between the onset of muscle contraction made by a subject, and a corresponding motion in a device [3]. It should be noted that the minimum interval between two distinct contractions is approximately 200 ms [29], [30]. This means that a segment of MES data with a length of 200 ms (or more) contains enough information to estimate a motion state of the hand. A segment length equal to or less than 200 ms leaves enough time (at least 100 ms) for the computation of features, classification, and generation of control commands, plus a device response time to maintain a real-time smooth motion control scheme.

However, a segment larger than 200 ms necessitates overlapped segmentation [1], [2] in order to avoid failure in real-time operation. The application of overlapped segmentation facilitates the employment of large segments (greater than 200 ms) for real-time control; however, computational load is still a matter of fact. Two methods of segmentation, namely, disjoint and overlapped segmentation, are illustrated in Fig. 1. Disjoint segmentation is associated with segment length, while overlapped segmentation is associated with length and increment. The increment is the time interval between two consecutive segments. It should be less than the segment length, and more than the processing time.

The first experiment in which the accuracy of classifications with a segment length of 50, 100, 150, 200, 300, and 500 ms were examined, investigates the influence of the segment length on classification (for different features). For segment lengths of 50, 100, 150, and 200 ms, disjoint segmentation was applied, while for lengths of 300 and 500 ms, overlapped segmentation with an increment of 200 ms was applied. Since an MES has an undetermined state between two levels of contraction, most classification error belongs to the transition period between classes. Hence, to avoid contradictory data during transition between motions, one segment of the transition period was eliminated from the training data. This means that classification relies on a steady state of muscle contraction.

C. Feature Selection

Because of the significance of features during classification, feature selection is an essential stage in myoelectric control design. A significant amount of literatures investigates or compares various features of TD, frequency-domain (FD), and time-scale, for myoelectric control [1], [2], [25]. Features should be capable of presenting the characteristics or properties of a signal for different limb motions. Computational load should also be considered in the real-time applications. In this work, the relative performance of various single features and feature sets (multifeatures) is determined in the context of an SVM-based classifier.

A surface MES is formed via the superimposition of individual action potentials (APs), generated by irregular discharges of active motor units (MUs) in a muscle fiber. Therefore, amplitude and frequency both represent the level of activity of motor units in the fiber, and signal features are mostly chosen from both time and frequency domains. The MES is a nonstationary stochastic process, with approximately zero mean and varying variance. Its amplitude, variance, energy, and frequency vary depending on contraction level. Time-scale features are excluded in this investigation, since MES can be assumed stationary in short-time muscle contractions. In addition, time-scale features, which are powerful in transient states, do not outperform TD features during a steady state. However, they impose a huge computational load [5], [13].

A wide range of features are considered individually and in-group in this paper. The involved features of the TD are mean absolute value (MAV), root mean square (RMS), waveform length (WL), variance (VAR), zero crossings (ZCs), slope sign changes (SSC), William amplitude (WAMP), and two types of modified mean absolute values (MAV1 and MAV2). A mathematical definition of features is presented in Table II. Considered FD features are a power spectrum (PS), autoregressive coefficients of order 2 and 6 (AR2 and AR6), and the mean and median of signal frequencies (FMN, FMD). In addition, four sets of features are examined in this work. The first set, which is called the TD feature set, was used by Hudgins *et al.* [3], and includes MAV, WL, ZC, and SSC. The second feature set, recommended by Huang *et al.* [13], consists of RMS and AR6. The third and fourth sets are derivations of the first and second sets. The third set is formed by MAV and WL, while the fourth set was constructed from RMS and AR2. Hereafter, these sets are named as a "multifeature."

The features of each channel were extracted from segments with various lengths, and then, concatenated together and fed into a classifier. The dimension of a feature space depends on the type and number of features; for example, the single features MAV and AR6, as well as multifeature RMS $+$ AR6, produce 4, 24, and 28-dimensional feature vectors, respectively.

D. Classification

In pattern-recognition-based myoelectric control, MES patterns corresponding to each motion are recognized and classified into distinctive classes. Hence, the classifier is an important module, and, ideally, it should perform correct classification. The following experiment compares SVM with two well-known classifiers, including LDA and MLP neural networks. Moreover, four popular kernels, including RBF, linear, polynomial, and sigmoid were examined in this investigation. A comparison was conducted based on selected features, including a single-feature MAV, multifeatures MAV $+$ WL $+$ ZC $+$ SSC, and RMS $+$ AR6, with a segment length of 200 ms.

SVM and MLP both require adjustment before application. Hence, model selection was applied to adjust parameters of the SVM for each subject/session and feature/segment, individually. Parameter adjustment was applied before offline training process. It uses the data collected for training or probably testing the classifier, and does not need extra data. A grid-search was employed as a method of model selection to adjust SVM parameters [20]. In this method, the performance of an SVM was examined based on a wide range of parameters; before those with the best performance were picked. A fivefold random cross-validation scheme was used to evaluate the parameters. Since performing a complete grid-search is time consuming, it was applied in two stages via coarse grids, and then, fine grids. In the coarse grids, the range of parameter C was $\{1, 10, 20, 50, 80, 100, 200, 600, \text{and } 1000\}$ and the range of r and γ was $\{0.16, 0.5, 2, 5, 8, 10, \text{and } 50\}$. In the fine grids, the parameters were examined in a range of $\pm 5\%$, $\pm 10\%$, $\pm 15\%$, and $\pm 20\%$ of their selected values.

A fivefold random cross-validation scheme yields the mean accuracy of five individual classifications when the whole data (i.e., training and test set) is randomly divided into five subsets of which part is chosen as a test set, while the rest is used as a training set. Cross validation as well prevents overfitting during classification. The layout of the MLP was also adjusted before utilization. Several layouts were tested and the one that yielded the best result was selected. The range of nodes in each hidden layer of MLP was $\{4, 5, 6, 8, 10, 12, 15, \text{and } 20\}$. Since a back propagation algorithm begins randomly initialized, the performance for each session was averaged over four iterations, and then, examined. All experiments in this work are based on using an adjusted classifier for each subject/session and feature/segment, except where it is mentioned specifically.

E. Postprocessing

Englehart *et al.* [5] showed that the application of overlapped segmentation accompanied by majority voting (MV) as a postprocessing mechanism improves accuracy and prevents degradation, by shortening segment size. The following experiment is used to examine this idea when applied to SVM-based classifiers. Overlapped segmentation is a technique that generates a dense stream of output for myoelectric control. The "Increment," as shown in Fig. 1, is a time slot, which separates the start of a new segment from a proceeding segment. Its lower bound is a time portion during which the processor needs to process the last segment, while its upper bound is the segment length. An overlapped segmentation employs an unused time portion of the processing time between two consecutive segments.

The MV is a postprocessing that eliminates transient jumps, and produces smooth output. It counts the estimated classes in the most recent $2\ m + 1$ estimations about a considered estimation (m-estimations before and m-estimations after), and outputs the value that occurs most as a corresponding estimation. Since MV uses the next m-estimation to produce the current output to avoid any failure in real-time control, the total delay in output should be less than 300 ms. Hence, real-time constraints impose

$$m \times \text{Increment} \leq 300\,\text{ms}.$$

The idea of shortening the segment length helps the development of a fast controller that requires a short response time for the inputs. In this experiment, the performance of classification with and without MV was examined. In addition, it compared classification performance over disjoint segments with a length of 50 ms and overlapped segments, with those with a length of 200 ms and an increment of 50 ms. Different single features and multifeatures were considered in comparison. As previously mentioned, considering an "Increment" equal to 50 ms, implies that the processing time is less than 50 ms. To avoid real-time failure, MV with $m = 4$ that imposes a delay of 200 ms was applied.

F. Evaluation

In all experiments mentioned in this section, accuracy is used as the main index to illustrate the performance of classification. Accuracy is defined as the rate of correct classification to all data in a test set. The accuracy of each session is computed using fourfold cross validation. In each fold, a block is chosen as the test set, and the remainder employed as a training set. The accuracy of each session was calculated as the average accuracy of four classifications. In addition, statistical analyses are applied to interpret the experimental results. As previously mentioned, the experiments were founded on data collected over 17 sessions from 11 subjects. Sufficient time (i.e., more than 20 days) was taken between sessions that had common subjects to avoid dependency between sessions. Therefore, each experiment had 17 independent observations, with an identical distribution. The purpose of statistical analysis is to find statistically meaningful differences over observations with certain significance.

Due to a relatively low rate of observations and their unknown distribution, nonparametric approaches were strongly suggested. Wilcoxon rank-sum and Kruskal–Wallis are two nonparametric [26] statistical methods that were adopted in this work. Wilcoxon rank-sum is a two-sided test for two groups of data with independent samples and an identical distribution, and can be used to recognize whether their medians are equal. Kruskal–Wallis is an extension of the Wilcoxon rank-sum for the data with more than two groups. The critical p-value, which determines whether a result is judged "statistically significant," was chosen as 0.05.

The applied SVM outputs the probability of the classes for each classification. It is obvious that the estimated class should have the highest probability in comparison to other classes. Given $p_i(n)$ as the probability of assigning the nth estimation in class i of a k-class classification problem, the entropy of the nth segment of data, is defined as

$$E(n) = -\sum_{i=1}^{k} p_i(n) \log[p_i(n)]. \tag{7}$$

Recently, the entropy has been noted as an internal measure to evaluate the correctness of classification for online training [14]. In MCS, it can be used as an unsupervised evaluation index during manipulation. A final experiment was used to examine the correlation of entropy and accuracy. It investigated the validity of entropy as an internal measure to evaluate the correctness of classification.

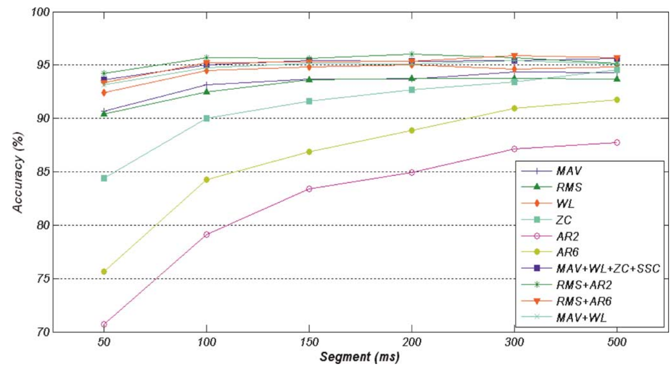

Fig. 2. Classification performance for FD single features decreases by shortening the segment length, while it is almost stable for TD single features and all multifeatures.

IV. Results and Discussion

This section presents the results of the experiments in five parts. In the first part, the effect of segment length is examined. The next part evaluates candidate single and multifeatures. The third part compares a SVM applied using different kernels with other classifiers. MV postprocessing is discussed in the fourth part, while the fifth and final part is used to investigate the validity of entropy in evaluation.

A. Segment Length

The purpose of this experiment was to investigate the dependence of features individually (single-features) or in-group (multifeatures), on segment length in MES classification. As discussed in the previous section, theoretically, a segment length of 200 ms contains enough information to estimate motion states of hand, while maintaining real-time constraints. Fig. 2 illustrates the performance of classification in different segment lengths. Six preselected single features (i.e., MAV, WL, RMS, ZC, AR2, and AR6) and four multifeatures (i.e., MAV + WL + ZC + SSC, RMS + AR2, RMS + AR6, and MAV + WL) were classified using an SVM with the RBF kernel and preadjusted parameters.

As can be seen, for all single features, accuracy decreases by shortening the segment length; this is because a shorter segment yields more bias and variance in feature estimation. TD single features (i.e., MAV, WL, RMS, and ZC) are more stable than FD single features (i.e., AR2 and AR6) to changes in segment length. The drop in accuracy of AR features is about 14% when the segment is reduced from 200 to 50 ms, while it is about 3% for TD single features. This is reasonable since the methods applied to compute FD features (i.e., AR and PSD) are highly sensible to the number of samples used in computation. In addition, ZC is too sensitive to changes in segment length; it is though named as a TD feature, represents changes in the frequency of a signal.

The performance of TD single-features had no considerable improvement with an increase in segment length. This supports the idea that a desired contraction is held for 200 ms, and longer segments merely decrease variance without a change in bias in feature estimation. WL, with almost fixed accuracy in all tested lengths, yields the most stable performance among the single

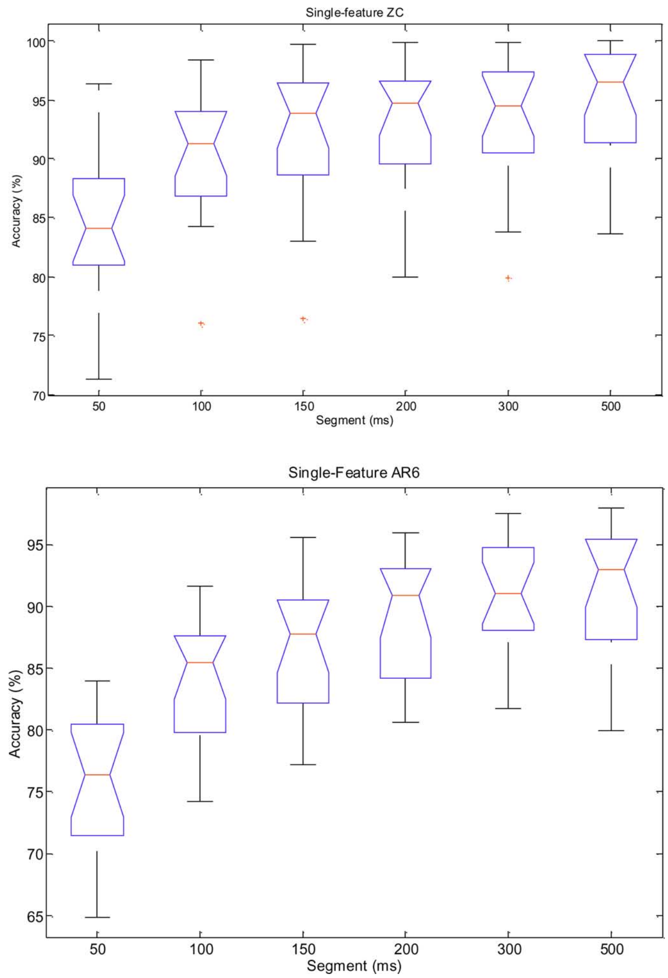

Fig. 3. Classification performance of ZC/AR6 drops significantly by decreasing the segment length from 200 to 50 ms.

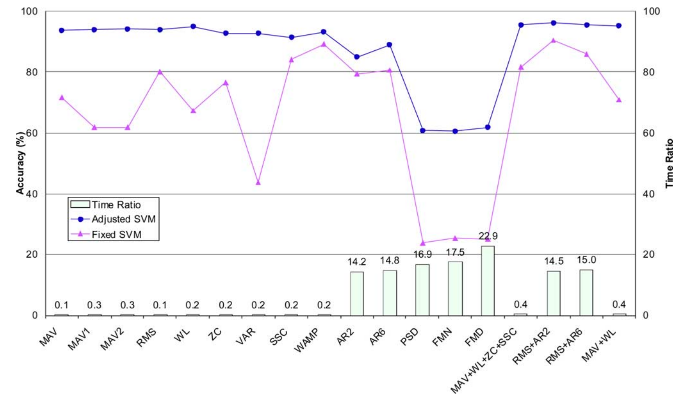

Fig. 4. Classification performance and time ratio of computational load of 14 single features and four multifeatures. Performance is for an SVM before and after parameter adjustment.

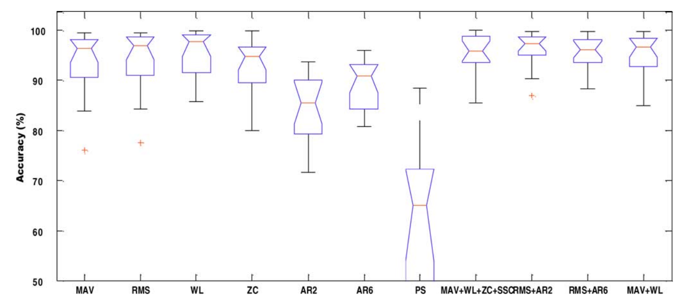

Fig. 5. Multifeatures and TD single features outperform significantly FD single features.

features. For AR features, the performance improves by 3% when segment length is increased from 200 to 500 ms.

Fig. 2 also depicts the performance of four multifeatures in different segment lengths. It can be seen that the multifeatures, which include a combination of different single features are more stable than whole single features to changes in segment length. Average accuracy reduction emanated from a shortening of segment length is about 1.9%, while it is more than 10% for single features. In addition, there is no considerable improvement in accuracy as segment length becomes longer.

Statistical analysis applied to recent results, has shown that the performance of classification of ZC, AR2, and AR6 drops significantly by decreasing the segment length to 50 ms, while for the rest of features, it does not differ significantly by changing the segment length between 50 and 500 ms. Fig. 3 shows the estimated median and confidence interval of accuracy for ZC and AR6 in two box-plots. It shows, for instance, that the confidence interval of the accuracy of AR6 for a segment of 200 ms is 85%–93%, while for a segment of 50 ms, it is 72%–80%.

This experiment suggests the use of TD single-features (MAV, WL, and RMS) or any multifeatures for fast control schemes in which a myoelectric controller should respond over a short time, and consequently, in which data segments would be short. With respect to the computational load required for AR features (more details in next subsection), a TD multifeature set (i.e., MAV + WL + ZC + SSC) is recommended for fast control schemes, without a considerable decline in accuracy. It loses 1.7% and 0.3% accuracy for segments of 50 and 100 ms, respectively. For single features, WL outperforms others, and accuracy drops by approximately 2.5% and 0.5% for 50 and 100 ms, respectively.

B. Feature Selection

Evaluation of MES features has formed in two stages. A preliminary experiment was conducted to preselect effective features from the list that mentioned in the last section, and then, an analytical comparison was applied. Fig. 4 shows the performance of features classified by an SVM before and after adjusting its parameters as well as a time ratio that indicates the comparative load of computation for each feature. It demonstrates the considerable difference between the accuracy of classification before and after parameter adjustment, and indicates its necessity for all the features. It also reveals that FD features (i.e., AR, PS, FMN, and FMD) impose relatively high load of computation compared to TD features (i.e., MAV, WL, . . .).

Having a perfect explanation and statistical comparison, necessitates extenuating the range of considered features via preselection. FMN and FMD features were excluded from further study, because of their weak performance during classification. This meant that the pattern of the mean and median of signal

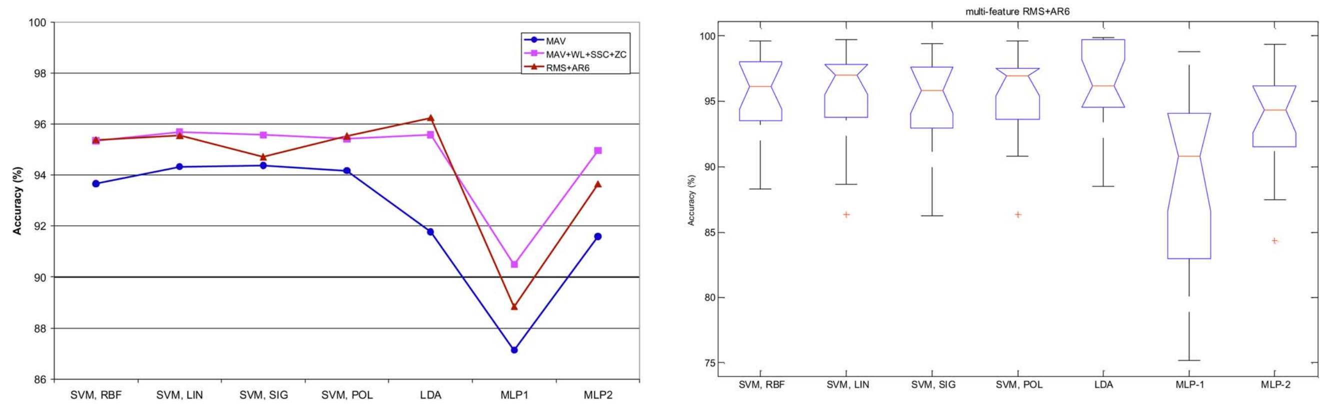

Fig. 6. Comparison of SVM with kernels of RBF, linear, sigmoid, and polynomial, LDA, and MLP with one and two hidden layers (Left) for feature sets: MAV, MAV + WL + ZC + SSC and RMS+AR6 (Right) box-plot for feature set: RMS + AR6.

frequencies were not sufficiently discriminative for the classification of limb motions. The modified MAV features (MAV1 and MAV2) were also excluded since they work the same as MAV. This means that the modifications do not provide any more discriminative information than MAV. VAR is also removed, due to its similarity to RMS. In comparison, VAR results in weaker performance in classification than RMS in all segment lengths. WAMP, SSC, and ZC are close together in the rate of performance, as well as the method of computation. Although WAMP yields better performance, ZC was selected for further comparison because of its popularity in literature.

The result of main experiment, which includes an analytical comparison of seven single features and four multifeatures, is illustrated in Fig. 5. It is in the form of box-plot yielded after statistical analysis. The features were computed from disjoint segments with length 200 ms (200 samples in a segment), and classified using an adjusted SVM with a kernel RBF. It can be seen that the selected TD single features (i.e. WL, RMS, MAV, and ZC) provide more discriminative information than FD single features (i.e. AR6, AR2, and PS). As the best instances of each group, WL and AR6 offer a 95% and 89% rate of accuracy, respectively. Fig. 5 also shows there is no noticeable difference in the performance of four tested multifeatures, and all outperform the single features by accuracy around 95.5%. Furthermore, the multifeatures provide much lower discrepancy in accuracy during several independent attempts across different subjects. For instance, the confidence interval of observations for the single feature WL is about 7.5%, while it is about 3.5% for the multifeature RMS + AR2. A confidence interval means a range that accommodates the rate of accuracy with a probability of more than 95%. This means that the range of classification accuracy for multifeatures, in several independent observations, is notably narrower than the range of classification accuracy for single features. This can be named as the major advantage of using multifeatures instead of single features in myoelectric classification.

In conclusion, the multifeature set RMS + AR2 that present high median accuracy and low discrepancy, can be named as the best feature set for a segment length of 200 ms. When considering the computational load for feature extraction and stability of performance in various segment sizes, the TD multifeature sets, (i.e., MAV + WL + ZC + SSC or MAV + WL) are the best options for MES classification. Among the single features, WL performs the best in accuracy, stability, and computation load.

C. Classifiers

The following experiment compared the accuracy of the SVM with LDA and MLP. Experimental results are shown in Fig. 6 (left). The classifiers were examined one-by-one over a single-feature MAV, multifeature MAV + WL + SSC + ZC, and multifeature RMS + AR6, individually. The first four items of the graph are SVM-based classifiers, with RBF, linear, sigmoid, and polynomial kernels, respectively. As the graph shows, the four applied kernels perform similarly over considered features. This can be interpreted that the boundaries between classes are almost linear. The average accuracy for all kernels is approximately $95.5 \pm 3.8\%$. The LDA is placed after SVM with the average performance of $94.5 \pm 4.9\%$. It performs the same as the SVM, probably because of the existence of linear boundaries between classes. The last two items of the graph (i.e., MLP1 and MLP2) belong to the result of MLPs with one and two hidden layers, respectively. As can be seen, the accuracy of the MLP with one hidden layer drops approximately 6%, while the MLP with two hidden layers performs with similar accuracy to the SVM and LDA.

Statistical analysis depicts that there is no meaningful difference between the performance of the considered classifiers over an MAV and multifeature MAV + WL + ZC + SSC; while a MLP performs significantly weaker with the multifeature RMS + AR6 Fig. 6 (right). The training process for a MLP (i.e., back propagation) takes longer than for an SVM or LDA. Experimental results show the time needed to train a MLP with one hidden layer is about ten times that needed to train an SVM. The training process for the MLP was nonrepeatable since it was initiated from random initial weights, and sought local minimum errors rather than global ones. It was for the SVM repeatable and fast. The training process for the LDA is fast, and, in contrast with SVM and MLP, it does not need parameter adjustment before application. The SVM settles to a global minimum error after training. The SVM is fast, as well, in classification. Experimental results recorded an average

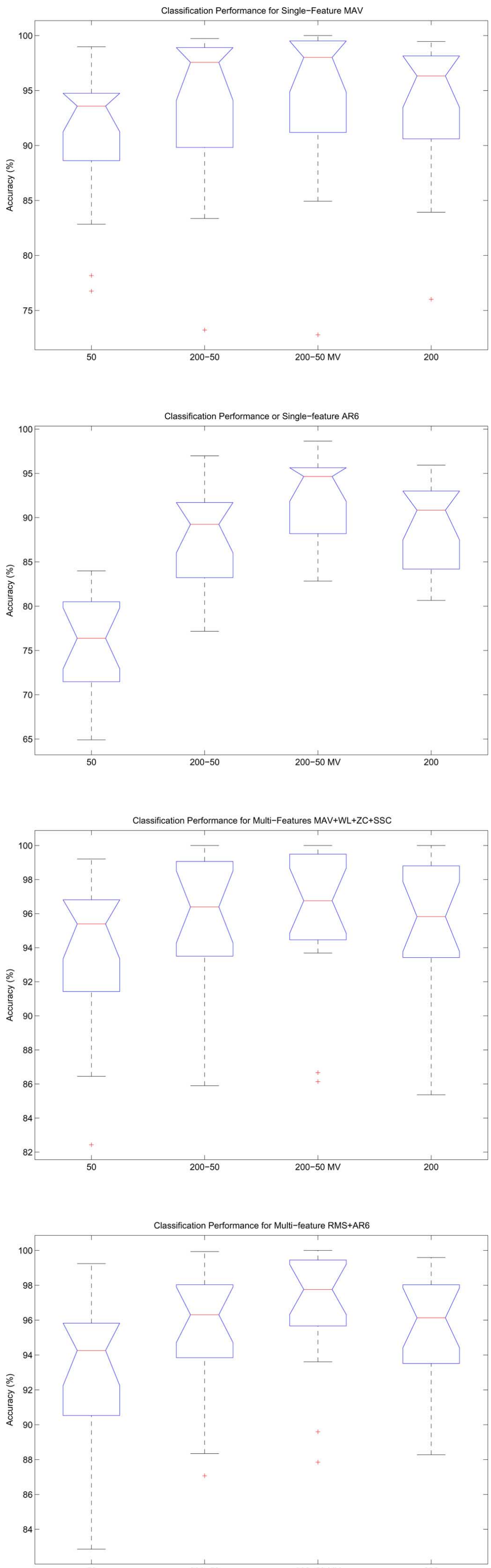

Fig. 7. Performance of classification applied on disjoint segments of 50 and 200 ms, and overlapped segments (length 200–50 ms increment) with/without MV for single features MAV and AR6 and multifeatures MAV + WL + ZC + SSC and RMS + AR6.

TABLE I
AVERAGE ACCURACY OF TWO SELECTED MULTIFEATURES

Feature	Segment	No MV	MV
MAV+WL+ZC+SSC	Disjoint	95 ± 4 %	96 ± 3 %
	Overlapped	96 ± 4 %	96 ± 3 %
RMS+AR6	Disjoint	95 ± 3 %	97 ± 3 %
	Overlapped	96 ± 4 %	97 ± 3 %

process time less than 50 ms including time required to TD multifeatures extraction and SVM classification. This is crucial in overlapped segmentation and postprocessing coming in the next subsection.

D. Postprocessing

As described previously, shortening segment length, which is occasionally offered as a means to decrease response time and make a controller faster, leads to degradation in classification accuracy. The following experiment shows the effect of overlapped segmentation and MV on the classification performance.

Fig. 7 illustrates the results of statistical analysis applied to experimental results. It compares the accuracy of short segments (50 ms), overlapped segments (200–50 ms), postprocessed overlapped segments (200–50 ms MV), and standard segments (200 ms), over features of MAV, AR6, MAV + WL + ZC + SSC, and RMS + AR6. A comparison of the performance for disjoint segments of 200 ms (200), with overlapped segments 200 ms and increment 50 ms (200–50), reveals the fact that the overlapped segmentation makes a controller four times faster, without a noticeable degradation in accuracy. In contrast, the MV that makes the controller slow has no remarkable improvement on the performance of the overlapped segmentation. In addition, overlapped segmentation and MV decrease notably the discrepancy of accuracy over different sessions. The range of accuracy for TD multifeature changes from 86%–99% to 94%–99%, after an application of overlapped segmentation and MV. This occurs because postprocessing makes the performance more robust. The experimental results lead us to use overlapped segmentation to increase the speed of the controller, and to apply MV to produce more stable output during classification.

Table I indicates the average performance for two multifeatures. The achieved result for RMS+AR6 is the same as the result achieved by Huang *et al.* [13] applying GMM on four-channel MES to recognize six limb motions. This suggests that TD multifeature set (MAV + WL + ZC + SSC) works same as RMS + AR6, but with remarkably lower computational load.

E. Entropy

The final experiment examined the validity of entropy as an evaluation index for correctness of classification. This experiment utilized an adjusted SVM with the kernel RBF on single feature MAV with segment length 200 ms. More than 14 000 independent observations (approximately 95% correct and 5% noncorrect classifications) are examined. The results showed a significant difference between correct and noncorrect classifications (as shown in Fig. 8). The entropy of correct classifications

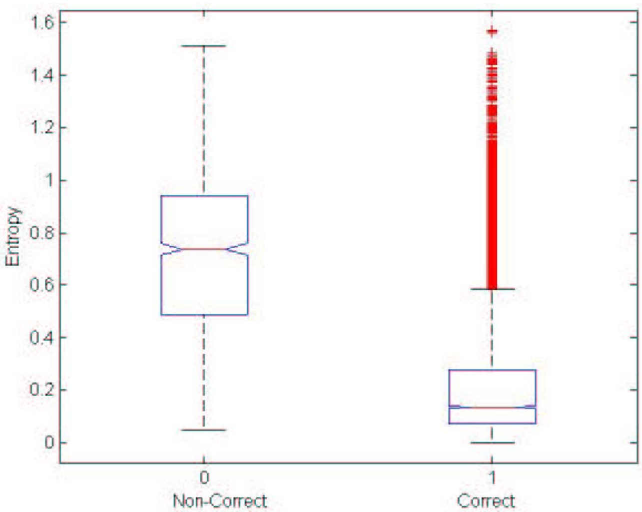

Fig. 8. Entropies of correct and noncorrect classification outputs are significantly distinct.

TABLE II
MATHEMATICAL DEFINITION OF FEATURES, GIVEN x_i AS THE iTH SIGNAL, AND N THE NUMBER OF SAMPLES IN A SEGMENT, AND p_j AS THE LINE (BAND) jTH OF SIGNAL POWER SPECTRUM

$MAV2=\frac{1}{N}\sum_{i=1}^{N}w_i\lvert x_i\rvert \;,\; w_i=\begin{cases}1 & 0.25N\le i\le 0.75N\\ 4i/N & 0.25N\ge i\\ 4(i-N)/N & 0.75N\le i\end{cases}$	$WL=\frac{1}{N}\sum_{i=1}^{N-1}\lvert x_{i+1}-x_i\rvert$	$MAV=\frac{1}{N}\sum_{i=1}^{N}\lvert x_i\rvert$
$MAV1=\frac{1}{N}\sum_{i=1}^{N}w_i\lvert x_i\rvert \;,\; w_i=\begin{cases}1 & 0.25N\le i\le 0.75N\\ 0.5 & else\end{cases}$	$VAR=\frac{1}{N}\sum_{i=1}^{N}(x_i-\bar{x})^2$	$RMS=\frac{1}{N}\sum_{i=1}^{N}x_i^{\,2}$
$SSC=\frac{1}{N}\sum_{i=1}^{N}f_i \;,\; f_i=\begin{cases}1 & x_i>x_{i-1}, x_i>x_{i+1} or x_i<x_{i-1}, x_i<x_{i+1},\\ & \lvert x_i-x_{i+1}\rvert>x_{th} or \lvert x_i-x_{i-1}\rvert>x_{th}\\ 0 & otherwise\end{cases}$	$FMN=\sum_{j=1}^{M}f_jP_j\Big/\sum_{j=1}^{M}P_j$	$PSD=\frac{1}{M}\sum_{j=1}^{M}p_j$
$WAMP=\frac{1}{N}\sum_{i=1}^{N-1}f(\lvert x_k-x_{k+1}\rvert) \;,\; f(x)=\begin{cases}1 & x>x_{th}\\ 0 & otherwise\end{cases}$	$\sum_{j=1}^{f_{md}}P_j=\sum_{j=f_{md}}^{M}P_j=\frac{1}{2}\sum_{j=1}^{M}P_j$	
$ZC=\frac{1}{N}\sum_{i=1}^{N}f_i \;,\; f_i=\begin{cases}1 & x_ix_{i+1}<0. \lvert x_i-x_{i+1}\rvert>x_{th}\\ 0 & else\end{cases}$	$x_i=\sum_{j=1}^{n}a_jx_{i-j}$, n^{th} order AR model	

was approximately 0.13, with a confidence interval between 0.07 and 0.28. If considering identical probability for unselected classes, the entropy 0.13 implies a probability of 99.35% for the selected class. For noncorrect classifications, the entropy was approximately 0.74 with a confidence interval between 0.49 and 0.94. This produced a probability of less than 93.8% for the selected class.

V. CONCLUSION

This paper introduced and evaluated the application of an SVM to classify upper limb motions using MESs. It suggested the most advantageous data segmentation technique, feature set, parameter adjustment approach for an SVM, and postprocessing method. Although experiments were conducted on healthy subjects with intact limbs, rather than on limb deficient subjects experience in [3] and [13] shows that this does not invalidate the achieved conclusions in the design of pattern-recognition-based myoelectric control. The observations and conclusions are not definitive solutions; however, they contain some notable contributions to advance knowledge in the field.

An SVM offers classification performance that matches or exceeds other classifiers, and does so in a computationally efficient manner. A TD multifeature set (i.e., MAV + WL + ZC + SSC) outperforms other features, because of its relatively high rate of accuracy, stability against changes in segment length, low discrepancy over several sessions, and computational simplicity. WL outperforms single features because of its high rate of accuracy and stability to changes in segmentation method. A disjoint segmentation with a length of 200 ms provides high performance during MES classification and a reasonable response time to allow real-time application. Overlapped segmentation with a length of 200 ms and an increment of 50 ms shortens the response time without a noticeable degradation in accuracy. MV provides no remarkable improvement in accuracy, but more stable output. The entropies of correct and noncorrect outputs are significantly distinct.

SVM has a high potential for application as a core for classification in myoelectric control systems, and appear to be capable of recognizing patterns that are more complex. This could lead to an expansion of the functionality of myoelectric control, as well as the development of advanced online training schemes to support long-term operation.

REFERENCES

[1] K. Englehart, B. Hudgins, and P. Parker, *Intelligent Systems and Technologies in Rehabilitation Engineering*. Boca Raton, FL: CRC Press, 2001, ch. 5.

[2] M. A. Oskoei and H. Hu, "Myoelectric control systems-a survey," *Biomed. Signal Process. Control (Elsevier)*, vol. 2, no. 4, pp. 275–294, Oct. 2007.

[3] B. Hudgins, P. Parker, and R. Scott, "A new strategy for multifunction myoelectric control," *IEEE Trans. Biomed. Eng.*, vol. 40, no. 1, pp. 82–94, Jan. 1993.

[4] K. Englehart, B. Hudgins, and P. A. Parker, "A wavelet-based continuous classification scheme for multifunction myoelectric control," *IEEE Trans. Biomed. Eng.*, vol. 48, no. 3, pp. 302–310, Mar. 2001.

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

[6] N. Chaiyaratana, A. M. S. Zalzala, and D. Datta, "Myoelectric signals pattern recognition for intelligent functional operation of upper-limb prosthesis," in *Proc. 1st Euro Conf. Disabil., Virtual Real. Assoc. Technol.*, U.K., 1996, pp. 151–160.

[7] A. T. C. Au and R. F. Kirsch, "EMG-based prediction of shoulder and elbow kinematics in able-bodied and spinal cord injured individuals," *IEEE Trans. Rehabil. Eng.*, vol. 8, no. 4, pp. 471–480, Dec. 2000.

[8] A. B. Ajiboye and R. F. Weir, "A heuristic fuzzy logic approach to EMG pattern recognition for multifunctional prosthesis control," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 13, no. 3, pp. 280–291, Sep. 2005.

[9] F. H. Y. Chan, Y. S. Yang, F. K. Lam, Y. T. Zhang, and P. A. Parker, "Fuzzy EMG classification for prosthesis control," *IEEE Trans. Rehabil. Eng.*, vol. 8, no. 3, pp. 305–311, Sep. 2000.

[10] K. Kiguchi, T. Tanaka, K. Watanabe, and T. Fukuda, "Design and control of an exoskeleton system for human upper-limb motion assist," in *Proc. IEEE/ASME Int. Conf. Adv. Intell. Mechatron. (AIM)*, 2003, pp. 926–931.

[11] M. Vuskovic and S. J. Du, "Classification of prehensile EMG patterns with simplified fuzzy ARTMAP networks," in *Proc. Int. Joint Conf. Neural Netw.*, 2002, vol. 3, pp. 2539–2544.

[12] J. S. Han, Z. Z. Bien, D. J. Kim, H. E. Lee, and J. S. Kim, "Human–machine interface for wheelchair control with EMG and its evaluation," in *Proc. 25th IEEE Int. Conf. Eng. Med. Biol. Soc.*, Mexico, Sep. 2003, pp. 1602–1605.

[13] Y. Huang, K. Englehart, B. Hudgins, and A. D. C. Chan, "A Gaussian mixture model based classification scheme for myoelectric control of powered upper limb prostheses," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 11, pp. 1801–1811, Nov. 2005.

[14] O. Fukuda, T. Tsuji, M. Kaneko, and A. Otsuka, "A human-assisting manipulator teleoperated by EMG signals and arm motions," *IEEE Trans. Robot. Autom.*, vol. 19, no. 2, pp. 210–222, Apr. 2003.

[15] A. D. C. Chan and K. Englehart, "Continuous classification of myoelectric signals for powered prostheses using Gaussian mixture models," in *Proc.*

25th Ann. Int. Conf. Eng. Med. Biol. Soc., Mexico, Sep. 2003, vol. 3, pp. 2841–2844.
[16] A. D. C. Chan and K. Englehart, "Continuous myoelectric control for powered prostheses using hidden markov models," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 1, pp. 121–124, Jan. 2005.
[17] E. Osuna, "Applying SVMs to face detection," *IEEE Intell. Syst.*, vol. 13, pp. 23–26, Jul./Aug. 1998.
[18] S. Dumais, "Using SVMs for text categorization," *IEEE Intell. Syst.*, vol. 13, no. 4, pp. 21–23, Jul./Aug. 1998.
[19] G. Dror, R. Sorek, and S. Shamir, "Accurate identification of alternatively spliced exons using support vector machine," *Bioinformatics*, vol. 21, no. 7, pp. 897–901, Apr. 2005.
[20] C.-C. Chang and C.-J. Lin (2001). LIBSVM: A library for support vector machines [Online]. Available: http://www.csie.ntu.edu.tw/∼cjlin/libsvm.
[21] P.-H. Chen, C.-J. Lin, and B. Schölkopf, "A tutorial on nu-support vector machines," *Appl. Stochastic Models Bus. Ind.*, vol. 21, pp. 111–136, 2005.
[22] R. M. Rifkin, "Everything old is new again: A fresh look at historical approaches in machine learning," Ph.D. dissertation, MIT Cambridge, Cambridge, Sep. 2002.
[23] C.-W. Hsu and C.-J. Lin, "A comparison of methods for multiclass support vector machines," *IEEE Trans. Neural Netw.*, vol. 13, no. 2, pp. 415–425, Mar. 2002.
[24] T. F. Wu, C. J. Lin, and R. C. Weng, "Probability estimates for multi-class classification by pairwise coupling," *J. Mach. Learn. Res.*, vol. 5, pp. 975–1005, 2004.
[25] M. A. Oskoei and H. Hu, "GA-based feature subset selection for myoelectric classification," in *Proc. Int. Conf. Robot. Biomimetics*, China, Dec. 2006, pp. 1465–1470.
[26] M. H. DeGroot and M. J. Schervish, *Probability and Statistics*. Reading, MA: Addison-Wesley, 2002, pp. 589–694.
[27] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, Sep. 1995.
[28] C. Lake and J. M. Miguelez, "Comparative analysis of microprocessors in upper limb prosthetics," *J. Prosthet. Orthot.*, vol. 15, pp. 48–65, 2003.
[29] J. VallsSole, J. C. Rothwell, F. Goulart, G. Cossu, and E. Muñnoz, "Patterned ballistic movements triggered by a startle in healthy humans," *J. Physiol.*, vol. 516, no. 3, pp. 931–938, 1999.
[30] L. Stark, *Neurological Control Systems Studies in Bioengineering*. New York: Plenum, 1968, pp. 389–397.

Mohammadreza Asghari Oskoei (S'08) received the B.Sc. and M.Sc. degrees in electrical engineering from the University of Tehran, Tehran, Iran, in 1991 and 1993, respectively. He is currently working toward the Ph.D. degree in the Department of Computing and Electronic Systems, University of Essex, Colchester, U.K.

His current research interests include adaptive schemes in myoelectric control, machine learning, support vector machines with applications to biomedical signal processing, and neuromuscular rehabilitation.

Huosheng Hu (M'94–SM'01) received the M.Sc. degree in industrial automation from the Central South University, Changsha, China, in 1982, and the Ph.D. degree in robotics from the University of Oxford, Oxford, U.K., in 1993.

He is currently a Professor in the Department of Computing and Electronic Systems, University of Essex, Colchester, U.K., where he is leading the Human Centred Robotics Group. Since 2000, he has been a Visiting Professor at six universities in China—the Central South University, Shanghai University, Wuhan University of Science and Engineering, Kunming University of Science and Technology, Chongqing University of Post and Telecommunication, and the Northeast Normal University. He is the author or coauthor of more than 250 papers published in various journals, books, and conferences. His current research interests include autonomous mobile robots, human–robot interaction, evolutionary robotics, multirobot collaboration, embedded systems, pervasive computing, sensor integration, RoboCup, intelligent control, and networked robotics.

Prof. Hu was the General Co-Chair of the IEEE International Conference on Mechatronics and Automation, Harbin, China, 2007, the Publication Chair of the IEEE International Conference on Networking, Sensing, and Control, London, 2007, the Co-Chair of the Special and Organized Sessions of the IEEE International Conference on Robotics and Biomimetics, Sanya, China, 2007, the Co-Chair of the International Program Committee, the First International Conference on Human Body Simulation and Modeling, Shanghai University, China, October 28–30, 2004, etc. He is the Founder Member of the Networked Robots of the IEEE Society of Robotics and Automation Technical Committee since 2001 and a member of the International Association of Science and Technology for Development (IASTED) Technical Committee on "Robotics" from 2001 to 2004. From 1997 to 2000, he was a Member of the Editorial Advisory Board for the *International Journal of Industrial Robots*, and is currently the Editor-in-Chief for the *International Journal of Automation and Computing*. He is a Reviewer for a number of international journals such as the IEEE TRANSACTIONS ON ROBOTICS, AUTOMATIC CONTROL, NEURAL NETWORKS and the *International Journal of Robotics Research*. He is a Chartered Engineer, a Senior Member of the Association for Computing Machinery (ACM), and a member of the Institution of Engineering and Technology (IET), the Association for the Advancement Artificial Intelligence (AAAI), IASTED, and the Institute of Advanced Study (IAS).