Best Feature Selection and Classification For Elbow Flexion And Extension Movements

A. Rubiano[1,3], J. L. Ramirez[1,3], M. N. El Korso[1,3], L. Gallimard[1,3], N. Jouandeau[2,3], O. Polit[1,3]

[1]LEME, Université Paris Ouest Nanterre La Défense,
[2]LIASD, Université Paris 8,
[3]Université Paris Lumière,
50, rue de Sèvres, 92410 Ville d'Avray, France.

E-mail: a.rubiano_fonseca@u-paris10.fr

Abstract. In this paper, a novel identification method for elbow flexion and extension movements based on the analysis of the surface electromyography (EMG) signals is presented. Thanks to 3D VICON cameras, we analyze the elbow angle and angular velocity in order to extract in an accurate way the EMG signals of interest from the biceps and the triceps muscles. Then, a feature selection criterion is proposed in order to identify the best features relevant to elbow flexion and extension movements classification. These features are used as input for a non-linear SVM algorithm which is optimized w.r.t. the so-called overfitting and radius coefficients. Finally, numerical simulation assess the accuracy of the classification, as well as the robustness of the proposed approach in non idealistic conditions.

1. INTRODUCTION

EMG signals analysis is of a great importance for hand prosthesis design since they contain pertinent information about the electrical activity produced by the neuromuscular process during muscle's relaxation and contraction [5]. Such analysis requires characteristic extraction in order to identify the produced and the expected movements. Nevertheless, the nature of the observed EMG signals makes the detection and identification a tricky task [1]. One can find in the literature some relevant studies, namely, Oskoei and Huosheng, focused on the classification of five hand movements in which the obtained accuracy range of 94% [4]. Whereas, in [7], the authors proposed the identification of the wrist and ring finger movements, and earned an accuracy of 87.3%. More recently, Wang *et al.* put forward the recognition of eight grasping gestures, and achieved an accuracy between 96.9% and 99.65% [6].

Nevertheless, none of them have considered the synchronization of the upper limb kinematic with the EMG signal. This will allow us to extract in an accurate way the signal fragment that matches exactly with the appropriate movement in order to enhance the rate of successful classification. First, to fill this lake, we propose to design a real time synchronized process w.r.t. both kinematic and EMG signal, which guarantees the correct extraction of the EMG signals (this open source software will be soon available in our webpage). Based on quaternion approach, the proposed design fills automatically data gaps and outliers. Thus, we consider variations in kinematical parameters (namely, speed and angle range) that affects considerably the EMG signals, and consequently, the classification accuracy. Secondly, a feature selection is performed by choosing the best couple of feature describing flexion and extension movements. By best, we mean the couple of feature that maximizes the distance between the nearest values of each couple of feature for different labeling. Finally, a non-linear SVM algorithm (which is optimized w.r.t. the so-called overfitting and radius coefficients.) is applied for the classification purpose.

In the proposed numerical simulations we consider a more realistic scenario which can take into account the sweat and fatigue of the subject, sensors' thermal disturbance etc. Some of such non idealistic conditions can be modeled as an additive white Gaussian noise. Numerical simulations reveal that we achieve an accuracy of 99.79% using corrupted noisy measurements as input unlike the previous results in the literature which used only noise-free measurements (i.e., observed EMG signal without any contamination). Furthermore, we note that the presented results are obtained by testing the EMG signal from a new subject (namely, the proposed scheme has both accurate accuracy and robustness to new subjects which are not included in the learning process).

2. EXPERIMENT AND PRE-PROCESSING OF DATA

2.1. Materials presentation

In the proposed study we consider 3 healthy subjects which executed flexion and extension movements[1]. For each movement, we observed EMG signals (using two EMG sensors with acquisitions sampled at 1 kHz) from the triceps and the biceps as well the tracking of the upper limb motion (using seven 3D VICON cameras with acquisitions sampled at 0.2 kHz). The triceps and the biceps have been selected since they generate a higher activity during the flexion and the extension movement compared to others muscles [9]. Furthermore, for the tracking of the upper limb motion (TULM) we considered 30 retro-reflective markers with the distribution shown in Fig. 1.

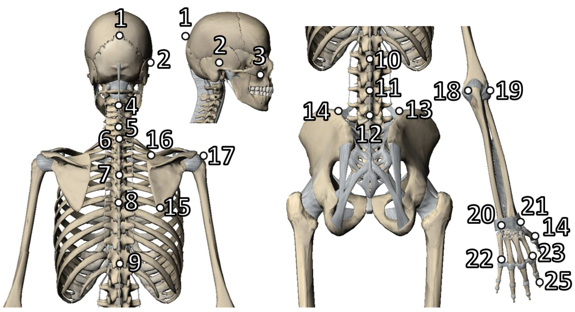

Figure 1. Markers positions for the upper limb kinematic chaine.

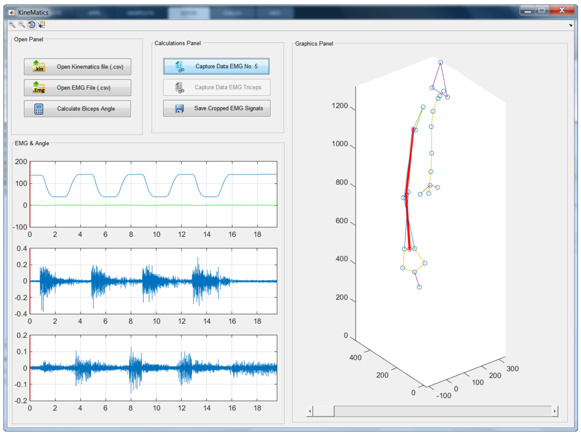

Figure 2. The interface of the proposed open source software.

2.2. Dealing with the missed reflectors

In order to pre-process the observation in an accurate manner we develop an open source software[2] (c.f. Fig. 2) that allows us to display the observation (both TULM and EMG signals), calculate the angular velocity of the elbow during the movement and finally, it synchronizes the EMG and TULM data in order to perform an accurate extraction even in the presence of different sampling frequencies.

Nevertheless, it is common that some markers disappear. In such situation, the miss detection of the reflector results annulling the coordinates of the missed markers (e.g., c.f. Fig. 3). Such annulling should be corrected since it has a dramatic effect on the accuracy of the classification procedure.

In order to overcome this disadvantage, we develop a mathematic model, which allows us to predict the coordinates of the missing markers. First, we cluster the set of markers by regions

[1] More precisely, the first, second and third subject executed, respectively, one, four and six trials

[2] The open source software will be available soon at `http://leme.u-paris10.fr/promain-565200.kjsp?RH=1415865400449`

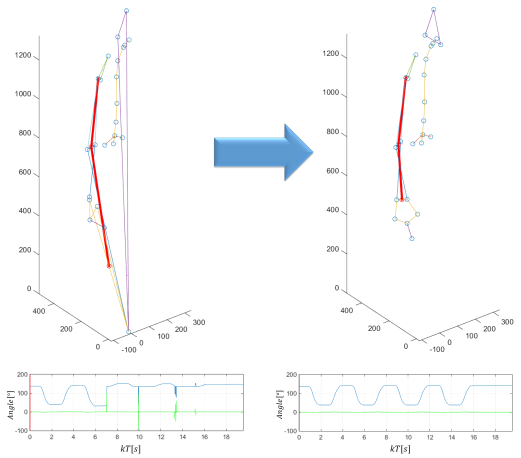

Figure 3. Examples of missing TULM Markers.

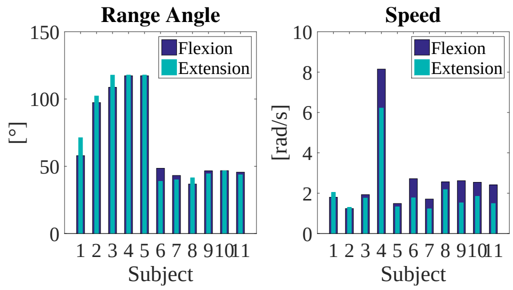

Figure 4. Extracted kinematical parameters after the pre-processing step.

(more precisely, we consider 7 regions : head, spinal column, sacrum, scapula, arm, forearm and hand). Each marker from each region is connected to its nearest marker by a straight line with a given direction representing a vector. Secondly, we use a quaternion based approach in order to formulate the spatial rotation [8]. More precisely, we define a group of quaternions expressing each rotation between the created vectors which allows us to predict the position each marker. Then, we check the position of each marker. If the expected marker's coordinate is above a certain threshold w.r.t. to the observed one, thus, the coordinate of the observed vector is replaced by the expected one. This procedure allows us to automatically correct the missed and miss-observed markers.

2.3. Evaluation of the elbow angle

After the correction of the visualized 3D model, we can easily evaluate the elbow angle, θ, and angular velocity η (c.f. Fig. 4). For example, we can use the vector linking the elbow to the wrist (arm vector) and the vector linking the elbow to the shoulder (forearm vector). The arm vector, $\boldsymbol{x}_a$, is defined from the point p_1 (which is defined as the mean of markers' 19 and 20 coordinates which are placed over the coronoid process ulna and radius) to p_3 (marker 18 which is placed over the epicondyle of humerus). Whereas, the forearm vector, $\boldsymbol{x}_{fa}$, is defined from p_1 to p_2 (which is defined as the mean of markers' 21 and 22 coordinates which are placed over the styloid process of ulna and radius). Thus, the elbow angle is evaluated by $\theta = \arccos\left(\frac{\langle \boldsymbol{x}_a, \boldsymbol{x}_{fa}\rangle}{\|\boldsymbol{x}_a\|\|\boldsymbol{x}_{fa}\|}\right)$. The behavior of the elbow angle (equivalently, the sign of the angular velocity $\eta = {}^{d\theta}/_{dt}$) determines the movement that has been performed. This is of importance, since the sign of the angular velocity and the the synchronisation of the EMG signal and the TUML, allow us an accurate extraction of EMG labelling the training data.

3. FEATURE SELECTION

Before performing the classification step, we focus on the feature selection in the context of the EMG signal applied to the identification of the elbow movements.

In the literature, several feature have been proposed for analyzing EMG signals [4]. Among the commonly used one:

- Mean Absolute Value : $\mathrm{MAV}(\boldsymbol{y}) = n^{-1}\sum_{i=1}^{n}|[\boldsymbol{y}]_i|$, where $[\boldsymbol{y}]_i$ represents the i-th sample of the signal $\boldsymbol{y}$, and n is the number of samples.
- Mean value: $\mathrm{M}(\boldsymbol{y}) = n^{-1}\sum_{i=1}^{n}[\boldsymbol{y}]_i$.
- Entropy : $\mathrm{Ent}(\boldsymbol{y}) = -\sum_i e_i^2 \log_2(e_i^2)$ where e_i represents the projection coefficients of the signal $\boldsymbol{y}$ in an orthonormal basis [2].
- Harmonic mean: $\mathrm{HM}(\boldsymbol{y}) = n(\sum_{i=1}^{n}[\boldsymbol{y}]_i^{-1})^{-1}$.

- Mean frequency: $\mathrm{MF}(\boldsymbol{y}) = (\sum_{j=1}^{N} I_j)^{-1} \sum_{j=1}^{N} I_j f_j$ where N denotes the number of harmonics in the spectrum, I_j represents the magnitude of the j-th harmonic, and f_j is the frequency of the j-th harmonic.

In the following, we aim to select the *best couple* of feature describing the two movements of our interest which are the flexion and the extension movement of the elbow (in which all the extracted features are presented in Table 1). This selection allows us to create two sets of features; one for extension and the other one for flexion. Thus, the meaning of *best couple* feature is the couple that maximize the distance between the nearest values of each couple of feature for different movements[3]. More precisely, using the training data, we evaluate the best couple of feature as $(\hat{f}_1, \hat{f}_2)$, for $\hat{f}_1 \neq \hat{f}_2$, which satisfies

$$(\hat{f}_1, \hat{f}_2) = \arg\max_{f_1, f_2} \left(\min \left(\sum_{i=1}^{2} \left(\mathcal{F}_{f_i}^{\mathrm{ext}} - \mathcal{F}_{f_i}^{\mathrm{fle}} \right)^2 \right) \right) \tag{1}$$

where $\mathcal{F}_{f_i}^{\mathrm{ext}}$ and $\mathcal{F}_{f_i}^{\mathrm{fle}}$ denote the value of the f_i-th features for extension and flexion, respectively. Using the pre-processed training data, the couple (Ent,MF) is shown to be the best couple in the sense of the proposed criterion (1) among all the features introduced above.

4. CLASSIFICATION SCHEME AND NUMERICAL RESULTS

4.1. Classification using the pre-processed data

After the pre-processing step and the feature selection and extraction, we are now ready to present the classification step.

First, let us gather the value of the selected couple of feature for each trial into the matrix $\boldsymbol{F}$ defined as $[\boldsymbol{F}]_l = \boldsymbol{f}_l = [\mathrm{Ent}(\boldsymbol{y}_{bi}), \mathrm{MF}(\boldsymbol{y}_{bi}), \mathrm{Ent}(\boldsymbol{y}_{tr}), \mathrm{MF}(\boldsymbol{y}_{tr})]$ in which the l denotes the l-th trial, $\boldsymbol{y}_{bi}$ and $\boldsymbol{y}_{tr}$ are the pre-processed observed biceps and triceps EMG signal.

Each row of $\boldsymbol{F}$ is a four element vector, denoted by $\boldsymbol{f}_l$, which can be represented as a point in a 4D features space. We assign an etiquette $z_l \in \{1, -1\}$ to each point. The binary etiquette classifies this point into two categories: extension (for $z = 1$) or flexion (for $z = -1$). Using a proper support vector machine (SVM) learning methodology, we aim to separate as well as possible the two sets of points in the feature space extracted from $\boldsymbol{F}$ having different labelling. More precisely, we can determine an optimal hyperplane between the two categories using a matrix $\boldsymbol{F}'$ composed by q rows, in which each row $\boldsymbol{f}'_q$ is a support vector.

Since a linear separation is not possible as shown in Fig. 5, we propose the use of a non-linear SVM classifier, in which the optimal hyperplane is obtained by solving the following quadratic programming problem [3]

$$\min_{\boldsymbol{h}, b, \boldsymbol{\kappa}} \frac{1}{2} \boldsymbol{h}^T \boldsymbol{h} + C \sum_{q=1}^{m} [\boldsymbol{\kappa}]_q \quad \text{s.t.} \quad z_q \left(\boldsymbol{h}^T \theta(\boldsymbol{f}'_q) + \gamma \right) \geq 1 - [\boldsymbol{\kappa}]_q \quad \text{and} \quad [\boldsymbol{\kappa}]_q \geq 0 \;\; \text{for} \;\; q = 1, \ldots, m \tag{2}$$

in which $[\boldsymbol{\kappa}]_q$ is the error soft margin, $\boldsymbol{h}$ and γ determine the hyperplane in feature space, C denotes a term to control the overfitting, m represents the amount of support vectors inside $\boldsymbol{F}'$, and θ maps $\boldsymbol{f}'_q$ into a higher-dimensional space. The following decision function represents the solution of the above problem:

$$\Psi(\boldsymbol{f}_l) = sign \left(\sum_{q=1}^{m} \beta_q z_q \mathcal{K}(\boldsymbol{f}'_q, \boldsymbol{f}_l) + \gamma \right)$$

in which β_q denotes the proper lagrange coefficients (see [2] for more details), and $\mathcal{K}(\boldsymbol{f}'_q, \boldsymbol{f}_l)$ is the radial basis kernel function[4]

[3] Note that the kinematical parameters variation (especially, the speed and the angle range) affects the set features' values.

[4] This choice is motived due to its capacity to maximize the distance between the hyperplane and the points.

Flexion									
Biceps					**Triceps**				
Ent	Mav	M	HM	MF	Ent	Mav	M	HM	MF
3.3E+00	$4.4E-02$	$-5.9E-04$	$0.0E+00$	$2.7E+02$	$2.3E+00$	$1.1E-02$	$1.0E-03$	$-7.2E-02$	$2.6E+02$
$3.4E+00$	$4.8E-02$	$-1.8E-03$	$-1.1E-05$	$2.7E+02$	$2.2E+00$	$1.2E-02$	$-1.8E-03$	$-8.3E-05$	$2.6E+02$
$3.5E+00$	$5.5E-02$	$-1.6E-03$	$-8.5E-06$	$2.7E+02$	$2.2E+00$	$1.3E-02$	$-1.8E-03$	$4.3E-05$	$2.6E+02$
$3.2E+00$	$7.1E-02$	$-4.5E-03$	$-2.5E-05$	$2.7E+02$	$2.5E+00$	$2.1E-02$	$8.7E-04$	$-3.3E-02$	$2.6E+02$
$3.5E+00$	$5.1E-02$	$-1.6E-03$	$1.2E-01$	$2.7E+02$	$2.4E+00$	$1.5E-02$	$-1.2E-03$	$1.6E-05$	$2.6E+02$
$4.4E+00$	$1.8E-01$	$4.2E-04$	$-7.0E-07$	$2.8E+02$	$2.9E+00$	$3.2E-02$	$-2.5E-03$	$2.6E-06$	$2.7E+02$
$3.0E+00$	$2.8E-02$	$2.1E-04$	$0.0E+00$	$2.7E+02$	$2.1E+00$	$1.9E-02$	$-2.9E-03$	$7.2E-05$	$2.7E+02$
$3.2E+00$	$7.1E-02$	$-4.5E-03$	$-2.5E-05$	$2.7E+02$	$2.5E+00$	$2.1E-02$	$8.7E-04$	$-3.3E-02$	$2.6E+02$
$3.4E+00$	$4.5E-02$	$2.9E-03$	$1.6E-06$	$2.7E+02$	$1.7E+00$	$1.3E-02$	$-3.4E-03$	$1.9E-04$	$2.6E+02$
$4.3E+00$	$1.6E-01$	$2.0E-03$	$7.1E-02$	$2.7E+02$	$3.0E+00$	$4.1E-02$	$-5.3E-03$	$1.1E-06$	$2.7E+02$
$3.3E+00$	$4.4E-02$	$2.3E-03$	$-4.0E-02$	$2.7E+02$	$2.2E+00$	$2.2E-02$	$-4.2E-03$	$-6.1E-07$	$2.7E+02$
Extension									
Biceps					**Triceps**				
Ent	Mav	M	HM	MF	Ent	Mav	M	HM	MF
$3.4E+00$	$6.1E-02$	$-1.1E-03$	$4.8E+00$	$2.7E+02$	$3.7E+00$	$6.3E-02$	$1.9E-04$	$0.0E+00$	$2.7E+02$
$2.0E+00$	$8.0E-03$	$3.9E-04$	$0.0E+00$	$2.6E+02$	$2.0E+00$	$9.0E-03$	$2.6E-04$	$0.0E+00$	$2.6E+02$
$2.0E+00$	$8.8E-03$	$3.6E-04$	$3.7E-06$	$2.6E+02$	$2.0E+00$	$1.0E-02$	$1.4E-04$	$0.0E+00$	$2.6E+02$
$2.5E+00$	$1.8E-02$	$-7.2E-04$	$-2.1E-06$	$2.6E+02$	$2.5E+00$	$1.8E-02$	$2.8E-04$	$-1.1E-01$	$2.6E+02$
$2.1E+00$	$9.1E-03$	$5.5E-04$	$0.0E+00$	$2.6E+02$	$1.8E+00$	$6.1E-03$	$4.2E-04$	$7.0E-05$	$2.6E+02$
$2.8E+00$	$2.3E-02$	$-5.5E-04$	$-7.5E-06$	$2.7E+02$	$2.6E+00$	$1.8E-02$	$1.4E-03$	$9.7E-06$	$2.6E+02$
$2.2E+00$	$1.2E-02$	$5.9E-04$	$-1.5E-05$	$2.6E+02$	$2.3E+00$	$1.6E-02$	$4.6E-04$	$-8.4E-06$	$2.7E+02$
$2.5E+00$	$1.8E-02$	$-7.2E-04$	$-2.1E-06$	$2.6E+02$	$2.5E+00$	$1.8E-02$	$2.8E-04$	$-1.1E-01$	$2.6E+02$
$2.1E+00$	$9.3E-03$	$3.4E-04$	$8.4E-07$	$2.6E+02$	$2.3E+00$	$1.4E-02$	$5.1E-04$	$-4.8E-05$	$2.6E+02$
$2.7E+00$	$2.1E-02$	$-2.7E-05$	$4.6E-05$	$2.6E+02$	$2.6E+00$	$1.8E-02$	$6.1E-04$	$-2.2E-04$	$2.6E+02$
$2.3E+00$	$1.0E-02$	$1.2E-03$	$-2.3E-06$	$2.6E+02$	$2.1E+00$	$1.3E-02$	$-5.6E-05$	$3.9E-05$	$2.6E+02$

Table 1. Values of extracted features.

$$\mathcal{K}(\boldsymbol{f}'_q, \boldsymbol{f}_l) = \exp^{-\frac{(\boldsymbol{f}'_q - \boldsymbol{f}_l)(\boldsymbol{f}'_q - \boldsymbol{f}_l)^T}{2\sigma^2}}$$

in which σ is a positive parameter for controlling the radius.

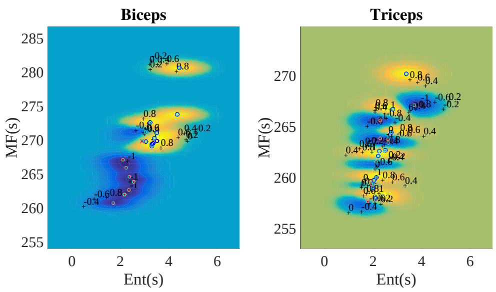

Figure 5. Flexion and extension features' distribution (blue and yellow circles denote flexion and extension features, respectively).

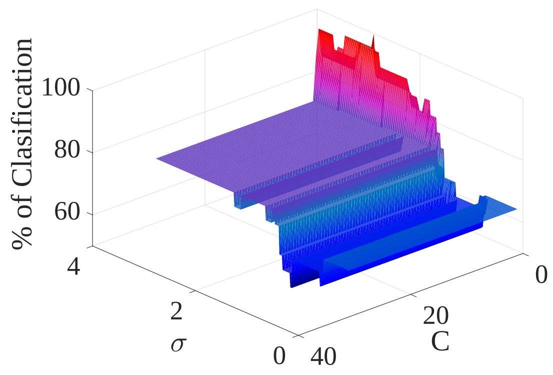

Figure 6. Percentage of correct classification as function of σ and C.

The above training was performed with only 13.63% of the whole trials. In the following, the decision function will be applied using the remaining 86.37% trials, in which the outcome of $\Psi(\boldsymbol{f}_l)$ is compared with the true known labelling z_l. We note that a randomly chosen value of σ and C may have a dramatic and poor performances which affects the successful classification's rate. Aiming to improve this successful classification's rate, we identified the optimal σ_{opt} and C_{opt} using the following criterion

$$(C_{\text{opt}}, \sigma_{\text{opt}}) = \underset{C,\sigma}{\arg\max} \sum_{l=1}^{M} \delta(\Psi(\boldsymbol{f}_l) - z_l)$$

in which M denotes the number of the test data and $\delta(.)$ is the kronecker function (which equals to 1 if its argument is equal to zero, and the function is equal to zero elsewhere). Numerical simulations reveal that the optimal value is not unique (as shown in Fig. 6). For example, for the couple $(C_{\text{opt}}, \sigma_{\text{opt}}) = (2.5, 1.1)$ we obtained $\sum_{l=1}^{M} \delta(\Psi(\boldsymbol{f}_1) - z_l) = M$, which mean that we reach 100% of successful classification's rate (the exact 100% rate is due to the fact that trials number is finite and relatively not large; 44 trials). We should precise, that this result is obtained using the first subject which represents a new pattern since the first subject 's trial was not considered for the training step.

4.2. Classification robustness under non idealistic conditions

In a more realistic scenario one has to take into account the sweat and fatigue of the subject, displacement of the recording electrodes, sensors' thermal disturbance etc. Some of such non idealistic conditions can be modeled as an additive white Gaussian noise [5].

We corrupted the testing data by an additive white Gaussian noise for three subjects (EMG signal measurements from both biceps and triceps). The numerical analysis is performed using 1000 Monte Carlo trials for different values of signal to noise ratio. Using the proposed scheme described below we achieve a successful rate of classification of 99.79% for a contamination of 2% of the power of the observed signals, which converges to 100% by decreasing the contamination rate. Then, we conclude that the proposed scheme is robust under medium noise contamination (due space limitation, the results of these simulations are not presented here).

5. CONCLUSION

We proposed a novel method for elbow flexion and extension classification. First, a quaternion based scheme was employed in order to automatically correct the values of the missing kinematic markers. Then, the elbow angle and angular velocity have been calculated in order to extract in a accurate way the signal of interest. Secondly, a feature selection procedure was performed based on a proposed criterion. Finally, a non-linear SVM algorithm was applied for the classification purpose. Using only 13.63% of the whole data as training data, our algorithm presented 100% successful rate of classification even in the presence of a new subject which was not included in the training procedure. Furthermore, numerical simulations show that in a non idealistic conditions (moderate contaminated measurement), the proposed scheme performs, also, almost perfectly (99,79%).

References

[1] Alkan, A. and Gnay, M., *Identification of EMG signals using discriminant analysis and SVM classifier*, Expert Systems with Applications, Vol. 39, 2012, pp. 44-47.

[2] Liu, Q. and Zhao, X. and Hou, Z., *Metric learning for event-related potential component classification in EEG signals*, Proceedings of the 22nd European Signal Processing Conference (EUSIPCO), 2014, Lisbon, Portugal, pp. 2005-2009.

[3] Chamasemani, F. F. and Singh, Y. P., *Multi-class Support Vector Machine (SVM) Classifiers - An Application in Hypothyroid Detection and Classification*, Sixth conference on Bio-Inspired Computing: Theroies and applications, IEEE, 2011, Penang, Malaysia, pp. 351-356.

[4] Oskoei, M.A. and Huosheng, H. *Support Vector Machine-Based Classification Scheme for Myoelectric Control Applied to Upper Limb*, IEEE Transactions on Biomedical Engineering, Vol. 55 2008, pp. 1956-1965.

[5] Reaz, M. B. I. and Hussain, M. S. and Mohd-Yasin, F., *Techniques of EMG signal analysis: detection, processing, classification and applications,* Biological Procedures Online, Vol. 8, 2006, pp. 11-35.

[6] Wang, N. and Chen, Y. and Zhang, X., *The recognition of multi-finger prehensile postures using LDA*, Biomedical Signal Processing and Control, Vol. 8 2013, pp. 706-712.

[7] Naik, G. R. and Kumar, D.K. and Jayadeva, J., *Twin SVM for Gesture Classification Using the Surface Electromyogram,* IEEE Transactions on Information Technology in Biomedicine, Vol. 14, 2010, pp. 301-308.

[8] Bell, P. and Potapov, I., *Reachability problems in quaternion matrix and rotation semigroups,* Information and Computation, Vol. 206, 2008, pp. 1353-1361.

[9] Ali, A. and Sundaraj, K. and Ahmad, R. B. and Ahamed, N. U. and Islam, A., *Surface electromyography for assessing triceps brachii muscle activities: A literature review,* Biocybernetics and Biomedical Engineering, Vol. 33, 2013, pp. 187-195.