# Features selection and classification to estimate elbow movements

A. Rubiano, J. L. Ramírez, Mohammed Nabil El Korso, N. Jouandeau, L. Gallimard, Olivier Polit

# Features selection and classification to estimate elbow movements

**A Rubiano[1,3], J L Ramírez[1], M N El Korso[1], N Jouandeau[2], L Gallimard[1] and O Polit[1]**

[1]LEME, Université Paris Ouest Nanterre la Défense, 50 rue de Sèvres, 92410 Ville d'Avray, FR
[2]LIASD, Université Paris 8,2 Rue de la Libert, 93526 Saint-Denis, FR
[3]Universidad Militar Nueva Granada, Cr 11 101-80, 110111 Bogotá, COL

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

**Abstract.** In this paper, we propose a novel method to estimate the elbow motion, through the features extracted from electromyography (EMG) signals. The features values are normalized and then compared to identify potential relationships between the EMG signal and the kinematic information as angle and angular velocity. We propose and implement a method to select the best set of features, maximizing the distance between the features that correspond to flexion and extension movements. Finally, we test the selected features as inputs to a non-linear support vector machine in the presence of non-idealistic conditions, obtaining an accuracy of 99.79% in the motion estimation results.

## 1. Introduction

The Electromyography (EMG) signal is a measure of the electrical activity produced by muscles during the movement. Therefore, it is possible to estimate the motion through the analysis of these signals [1]. Even that, due to the nature of the EMG the estimation is not straightforward [2].

In the literature, several relevant studies have been carried out. Some of them are focused on the relation of the EMG signals and the movement, and some others in the classification of signals to estimate movements. Regarding the approach to the classification of EMG signals, Oskoei and Huosheng [3] performed the classification of five movements of the hand, they reached an accuracy of 94%. Likewise, in [4], the authors proposed the identification of the wrist and the ring finger movements, and they earned an accuracy of 87.3%. In the same way, Wang *et al.* [5] put forward the recognition of eight grasping gestures and achieved an accuracy between 96.9% and 99.65%. Concerning the analysis of the relationship between EMG and motion, Kamavuako *et al.* [6] carried out a study to find the relationship between the grasping force and the EMG signals.

Nevertheless, these important studies have not established a clear relationship, and so, the classification process could be improved. Consequently, we introduce a new method to analyse the EMG signals based on features. A feature is a scalar value extracted from the EMG signal in time or frequency domains that contains relevant information that could be linked with motion. Our method, analyse the kinematic parameters as angular velocity an angular position, to extract in an accurate way the signal fragment that matches exactly with the appropriate

movement. Therefore, we enhance the possibility of establishing a relationship EMG-motion to improve the process of motion estimation.

Afterwards, we perform a correlation between the kinematic data and the EMG signals. This relationship allows to propose a criterion for selecting the couple of feature that best separates the flexion and extension movements.

Finally, the results of our criterion (best features) are validated using a non-linear SVM algorithm (which is optimized with respect to the so-called overfitting and radius coefficients). We add simulated white zero-mean Gaussian noise to the signals for assessing our method in non-idealistic condition. As a result, we achieved a classification percentage of 99.79% which is higher than the classification achieved in the state of the art.

## 2. Experimental Set-up

In this study, we consider three healthy subjects that execute flexion and extension movements, the first, second and third subject execute one, four and six trials, respectively. For each movement of flexion and extension, we measured EMG signals of the triceps and the biceps (due to their high electrical activity during these movements [7]) and tracked the upper limb motion.

The motion was tracked using seven 3D VICON cameras with a sampling frequency of 200 Hz, and 30 retro-reflective markers distributed as shown in Fig. 1. The EMG signals were captured using wireless sensors with a sampling frequency of 1 kHz.
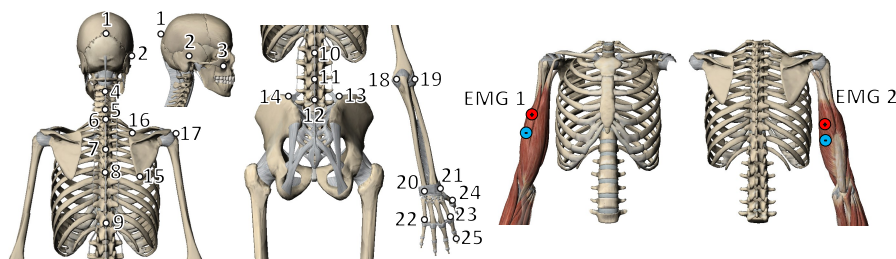


**Figure 1.** The Position of retro-reflective markers and EMG sensors.

To preprocess the experimental data in an accurate way, we develop an open source software[1] that allows us to display the measures (motion and EMG) and calculate kinematic parameters as rotation and angular velocity. Our software synchronizes motion capture and EMG signals (measured in real time) to perform an accurate analysis of the relation between motion and EMG.

## 3. Kinematic analysis and features selection

In this section, we present the kinematic analysis of the upper limb, and we introduce the relation between the EMG signals and kinematic parameters. afterwards, we perform a feature extraction and we propose a new strategy to select the best couple of features.

### 3.1. Kinematic analysis of the movement

The synchronization of the 3D motion and EMG signals, allows to evaluate the kinematics data i.e. the elbow angle, $\theta$, and angular velocity $\eta$. The Fig. 2 shows these values for the eleven trials. To calculate the angle, we introduce a vector linking the elbow to the wrist (forearm's vector) and a vector linking the elbow to the shoulder (arm's vector). The arm's vector $\boldsymbol{x}_a$, is defined from the point $p_1$ (located at the middle point between markers 18 and 19 placed over

---

[1] The open source software is available at `http://leme.u-paris10.fr/promain-565200.kjsp`

the coronoid process of the ulna and the radius) to $p_3$ (marker 17 placed over the epicondyle of the humerus). The forearm vector, $\boldsymbol{x}_{fa}$, is defined from $p_1$ to $p_2$ (located at the middle point between markers 20 and 21 placed over the styloid process of the ulna and the radius). Consequently, the elbow angle is evaluated by $\theta = \arccos\left(\langle\boldsymbol{x}_a, \boldsymbol{x}_{fa}\rangle / \|\boldsymbol{x}_a\|\|\boldsymbol{x}_{fa}\|\right)$. The sign of the angular velocity, calculated as $\eta = d\theta/dt$, defines if the movement corresponds to an extension (positive) or a flexion (negative). Considering that the motion capture and the EMG signals are synchronized, we use the sign of the angular velocity $\eta$ to extract the part of the EMG signal that corresponds to a particular movement (flexion or extension).
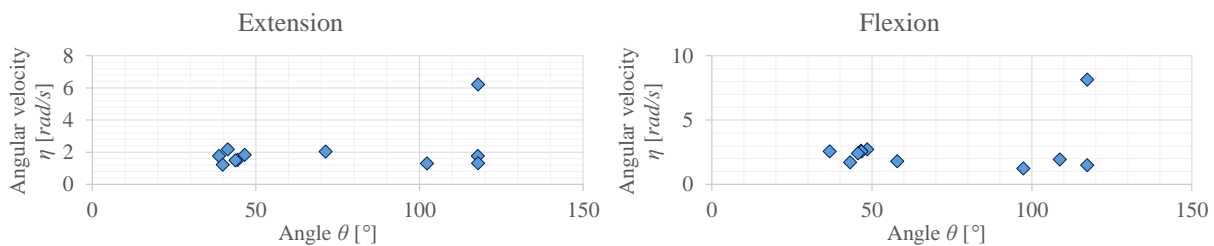


**Figure 2.** Extracted kinematical parameters after the pre-processing step.

*3.2. Relationship between EMG and Kinematic*
The EMG signal is a measure of the electrical activity produced by muscle contractions during motion. From these signals, it is possible to extract scalar values that could contain relevant information of the movement, and so on, establish a relationship between EMG and motion. Consequently, we choose the following set of features in time and frequency domain:

- 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 [8].
- 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.

Once we apply the feature extraction models to all trials, we get a set of scalar values that correspond to the flexion and the extension. Considering that the values have different scales, we apply a normalisation. The table 1 shows the normalised values of the features; the row one contains the values for trial one, the row two represents trial two, and so on.

Comparing the normalised values of the features during the different movements, we find that M and HM have significant variations and few local minimal and maximal values (see figure 3 left side). These long fluctuations could be useful to model small angle or velocity changes. But in the opposite, the identification of the complete range will be more difficult. Moreover, Ent, MAV and MF show a similar pattern, a more regular behaviour and several local minimal and maximal values (see figure 3 right). These particularities make possible to model and consequently describes a long range movement. It shows that there is a relation between the features extracted from the EMG signals and the motion. Thus we propose a strategy to choose the best features to identify the flexion and extension movements.

**Table 1.** The normalised values of the features.

| Flexion | | | | | | | | | | Extension | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Biceps | | | | | Triceps | | | | | Biceps | | | | | Triceps | | | | |
| Ent | Mav | M | HM | MF | Ent | Mav | M | HM | MF | Ent | Mav | M | HM | MF | Ent | Mav | M | HM | MF |
| −0.47 | −0.57 | 0.01 | −0.31 | −0.69 | −0.07 | −0.94 | 1.36 | −2.50 | −0.66 | 2.25 | 2.81 | −1.60 | 3.02 | 2.14 | 2.55 | 2.90 | −0.57 | 0.45 | 1.65 |
| −0.26 | −0.49 | −0.46 | −0.31 | −0.42 | −0.57 | −0.82 | 0.03 | 0.52 | −0.88 | −1.05 | −0.66 | 0.52 | −0.30 | −1.00 | −0.86 | −0.62 | −0.40 | 0.45 | −1.10 |
| −0.03 | −0.34 | −0.37 | −0.31 | −0.19 | −0.52 | −0.80 | 0.03 | 0.53 | −0.86 | −0.92 | −0.61 | 0.47 | −0.30 | −0.79 | −0.74 | −0.56 | −0.73 | 0.45 | −1.09 |
| −0.63 | −0.03 | −1.55 | −0.32 | 0.28 | 0.47 | 0.07 | 1.28 | −0.87 | −0.31 | 0.19 | −0.01 | −1.06 | −0.30 | 0.04 | 0.16 | −0.04 | −0.34 | −2.02 | 0.07 |
| 0.00 | −0.42 | −0.41 | 2.43 | −0.55 | −0.01 | −0.59 | 0.30 | 0.53 | −0.09 | −0.76 | −0.59 | 0.74 | −0.30 | −1.28 | −1.20 | −0.81 | 0.04 | 0.45 | −1.71 |
| 2.03 | 2.20 | 0.41 | −0.31 | 2.67 | 1.38 | 1.32 | −0.33 | 0.53 | 1.49 | 0.91 | 0.34 | −0.81 | −0.30 | 1.16 | 0.45 | −0.03 | 2.61 | 0.45 | 0.34 |
| −1.02 | −0.89 | 0.33 | −0.31 | −0.57 | −0.74 | −0.11 | −0.48 | 0.53 | 0.66 | −0.51 | −0.42 | 0.80 | −0.30 | −0.12 | −0.19 | −0.14 | 0.15 | 0.45 | 1.12 |
| −0.63 | −0.03 | −1.55 | −0.32 | 0.28 | 0.47 | 0.07 | 1.28 | −0.87 | −0.31 | 0.19 | −0.01 | −1.06 | −0.30 | 0.04 | 0.16 | −0.04 | −0.34 | −2.02 | 0.07 |
| −0.35 | −0.56 | 1.40 | −0.31 | −0.64 | −1.80 | −0.71 | −0.73 | 0.54 | −1.25 | −0.68 | −0.57 | 0.44 | −0.30 | −0.71 | −0.19 | −0.27 | 0.27 | 0.45 | 0.50 |
| 1.84 | 1.69 | 1.03 | 1.32 | 0.62 | 1.77 | 2.27 | −1.62 | 0.53 | 1.89 | 0.77 | 0.23 | −0.07 | −0.30 | 0.52 | 0.42 | −0.02 | 0.56 | 0.44 | 0.47 |
| −0.48 | −0.58 | 1.17 | −1.24 | −0.77 | −0.38 | 0.22 | −1.12 | 0.53 | 0.31 | −0.39 | −0.52 | 1.64 | −0.30 | 0.01 | −0.54 | −0.35 | −1.25 | 0.45 | −0.32 |
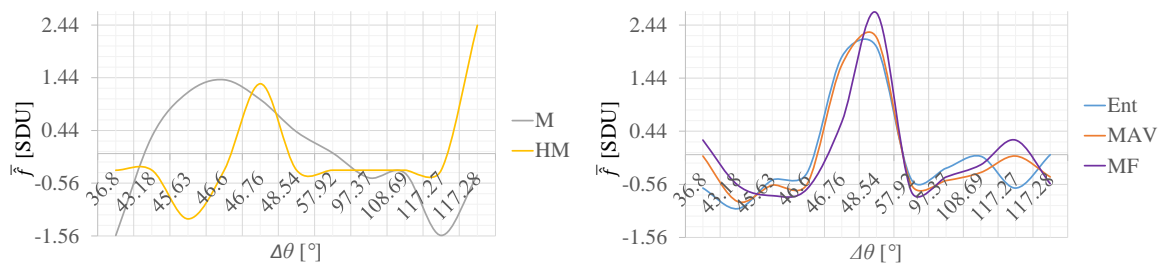


**Figure 3.** The behaviour of the features with respect to the movement of flexion.

### 3.3. Features selection

For this purpose, each movement is represented in an n-dimensional space $\Upsilon$. For two muscles, we extract two features from the biceps signal and one for the triceps, and the space $\Upsilon$ will have four dimensions. One point in this space represents one movement, flexion or extension.

The strategy for selecting the best couple of features is based on the maximization of the Euclidean distance between flexion and extension points. For a couple of features $(\hat{f}_1, \hat{f}_2)$, for $\hat{f}_1 \neq \hat{f}_2$, this function satisfies:

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

where $\mathcal{F}_{f_i}^{\text{ext}}$ and $\mathcal{F}_{f_i}^{\text{fle}}$ denote the values of the $i^{th}$ feature for extension and flexion, respectively. As a result of the proposed criterion (1), the couple (Ent, MF) shows the best fitness among all the features introduced above.

## 4. Supervised Learning Classification

The supervised learning method is introduce in this section. It is used to assess the performance of the selected features for motion classification. Furthermore, external perturbations [2] of the EMG signal is simulated with additive white zero-mean Gaussian noise (AWZMGN) [1].

### 4.1. Support Vector Machine

Support vector machine (SVM) is a supervised learning method that can separate a set of samples $\boldsymbol{F}$ into two categories $z_l \in \{-1, 1\}$. The method separates the categories using a subset $\boldsymbol{F'}$ as

---

[2]  Sweat and fatigue of the subject, displacement of the recording electrodes, thermal disturbances, among others.

training data to calculate an optimal hyperplane between them. The subset $\boldsymbol{F}'$ is composed by $q$ samples denominated support vectors $\boldsymbol{f}'_q$.

In this study, the set of samples $\boldsymbol{F}$ is composed of the selected features (MF and Ent) extracted from the biceps' and the triceps' EMG signals. Consequently, each member of $\boldsymbol{F}$ is defined as $[\boldsymbol{F}]_l = \boldsymbol{f}_l = [\mathrm{Ent}(\boldsymbol{y}_{bi}), \mathrm{MF}(\boldsymbol{y}_{bi}), \mathrm{Ent}(\boldsymbol{y}_{tr}), \mathrm{MF}(\boldsymbol{y}_{tr})]$ where index $l$ is the trial number. Furthermore, the $\boldsymbol{F}'$ is composed by three elements of $\boldsymbol{F}$. Considering that a linear separation is not possible, we propose the use of a non-linear SVM classifier, in which the optimal hyperplane is obtained by solving the following quadratic programming problem [9]:

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

where $[\boldsymbol{\kappa}]_q$ is the error soft margin, $\boldsymbol{h}$ and $\gamma$ 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 $\Theta$ maps $\boldsymbol{f}'_q$ into a higher-dimensional space. The following decision function represents the solution of the following problem:

$$\Psi(\boldsymbol{f}_l) = sign\left(\sum_{q=1}^{m}\beta_q z_q \mathcal{K}(\boldsymbol{f}'_q, \boldsymbol{f}_l)+\gamma\right) \quad \text{for} \quad \mathcal{K}(\boldsymbol{f}'_q, \boldsymbol{f}_l) = \exp^{-\frac{(\boldsymbol{f}'_q-\boldsymbol{f}_l)(\boldsymbol{f}'_q-\boldsymbol{f}_l)^T}{2\sigma^2}} \quad (3)$$

where $\beta_q$ denotes the Lagrange coefficients [8], $\mathcal{K}(\boldsymbol{f}'_q, \boldsymbol{f}_l)$ is the radial basis kernel function[3] and $\sigma$ is a positive parameter associated with the radius.
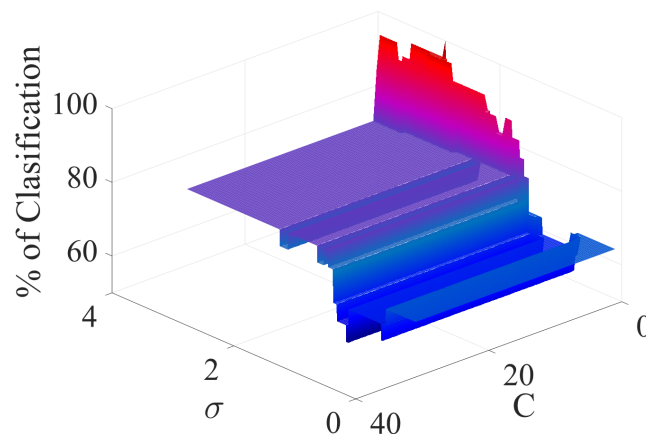


**Figure 4.** Percentage of correct classification as function of $\sigma$ and $C$.

Subsequently, we apply the decision function $\Psi(\boldsymbol{f}_l)$ to the elements of $\boldsymbol{F}$ that have not been used as support vectors, and we compare the result with the corresponding $z_l$ which are known. If the result of the decision function match with $z_l$, then the classification is correct.

*4.2. Classifier Implementation*
It is necessary to take into account the influence of the parameters $\sigma$ and $C$ for the classification performance. Therefore, we propose to find the optimal values $\sigma_{\mathrm{opt}}$ and $C_{\mathrm{opt}}$, following the

---

[3] This function allows classify using a Euclidean distance.

criterion:

$$(C_{\text{opt}}, \sigma_{\text{opt}}) = \arg\max_{C,\sigma} \sum_{l=1}^{l} \delta(\Psi(\boldsymbol{f}_l) - z_l) \tag{4}$$

where $\delta(.)$ is the Dirac function (equals to one if its argument is zero and equal to zero elsewhere).

We implement the SVM defined by equations (2) and (3), using three support vectors, then the criterion (4) is applied. We find that the optimal value is not unique. Fig. 4 shows the relation between $C$, $\sigma$ and the percentage of classification; the red zone is the optimal zone where the classification percentage reach 100%. Therefore, it must be denoted that the classifier identifies the movements of the first subject although the support vectors is based on features extracted from the signal of subjects two and three; the test is performed with all subjects.

Finally, we add AWZMGN to the original set of EMG signals using 1000 Monte Carlo trials for different values of signal to noise ratio (SNR). Afterwards, we use the noised signals as the new set of samples $\boldsymbol{F}$ and we perform the SVM classification described in section 4.1. We achieve a classification percentage of 99.79% for an SNR of $37.5 dB$.

## 5. Conclusions

In this paper, the relationship between EMG signals from the biceps and triceps for flexion and extension movements of the elbow is analysed. A novel method to assess and select the best features is presented for the classification of the movements. A criterion is defined to select the best features associated with the motion.

The test performed with the non-linear SVM algorithm, using only 13.63% of the whole data as training data, allow us to achieve a 100% successful classification. For a new subject and after adding simulated AWZMGN, the classification reach a percentage of 99.79%. Therefore, the proposed method to analyse EMG signals is efficient and robust.

## References

[1] 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.

[2] 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.

[3] 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.

[4] 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.

[5] 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.

[6] Kamavuako, E. N. and Farina, D. and Yoshida, K. and Jensen, W., *Relationship between grasping force and features of single-channel intramuscular EMG signals*, Journal of Neuroscience Methods, Vol. 185, 2009, pp 143-150.

[7] 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.

[8] 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.

[9] 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.