

# Gait Recognition Based on EMG Information with Multiple Features

Yueying Cao<sup>1</sup>, Farong Gao<sup>1,\*</sup>, Liling Yu<sup>2</sup>, and Qingshan She<sup>1</sup>

<sup>1</sup>School of Automation, Hangzhou Dianzi University, Hangzhou 310018, China

<sup>2</sup>Department of Rehabilitation, Zhejiang Hospital, Hangzhou 310013, China  
E-mail: frgao@hdu.edu.cn

**Abstract.** In order to evaluate the effects of time domain (TD) and frequency domain (FD) features as well as muscle number on gait classification recognition, eight channels of electromyography (EMG) signals were collected from four thigh and four lower leg muscles, and two TD features and two FD features were extracted in this study. The method of support vector machine (SVM) was presented to investigate the classification property. For the classification stability and accuracy, 3-fold cross validation was verified and selected to classify the lower limb gait. The results show that the FD features can obtain higher accuracy than TD features. In addition, accuracy of gait recognition increased with the augment of muscle number.

**Keywords:** Gait recognition, EMG, Time domain, Frequency domain,  $k$ -fold, Support vector machine

## 1 Introduction

Gait is a dynamic and successive activity achieved through the movement of limbs while daily walking. Normal human gait has the characteristics of coordination, proportionality and periodicity. According to the toe-off and heel-strike, an integrated gait cycle can be basically divided into two gait phases, the swing phase and the stand phase [1]. Gait analysis is the powerful tool of walking function in the areas of sports science, rehabilitation therapy and clinical assessment. Santilli *et al.* examined fourteen athletes with functional ankle instability during gait cycle [2]. Bogataj presented an aggressive approach to gait relearning in patients who were unable to walk independently with hemiplegia by means of multichannel functional electrical stimulation combined with conventional therapy [3]. Fish *et al.* provided a clinical assessment of the gait which may be used to monitor the value of physiotherapy treatment for people who have neurological deficits [4].

It was mentioned in [5] that gait analysis has used different types of motion sensors, such as the accelerometers, gyroscopes, force sensors and electromyography sensors. In contrast with other sensors, the surface electromyography (sEMG) sensors have an inherent advantage in reflecting the internal activities of the muscle and differentiating various conditions. The sEMG had developed to perform an indirect measurement of muscle activity using surface electrodes which can detect voltage change to provide critical information during the onsets of muscle contraction and relaxation. And these electrodes are portable which is perfectly suitable for continuous motion. The sEMG signal contains abundant information. Its acquisition technology is mature and reflects human movement intention. It has shown that sEMG has been successfully used in both academic study and clinical applications in several different fields such as sports science[2], neurophysiology[6] and rehabilitation[7]. Three main procedures that consist of data preprocessing, feature extraction and classification algorithm should be carefully considered in recognition of sEMG signals of human gait cycles.

Feature extraction is an important step to extract the useful information that is hidden in surface EMG signal [8]. Appropriate features should be capable of presenting the characteristics for different limb motions, contribute to preferable performance in classification accuracy and precise control of mputee's prosthesis. How to select effective features is the fundamental problem of EMG gait pattern recognition.

Generally, there are three major categories of features extracted from sEMG signals, which are time domain(TD), frequency domain(FD), and time-frequency or time-scale representation [9]. Several surveys [10, 11] have presented possible TD and FD features and detailed introductions to sEMG feature extraction. The study on these features has achieved certain progress.

Linear Discriminant Analysis was studied by performed six hand movement's task to investigate the performance of TD and FD features from EMG signal [12], it is found that FD features in discriminating the hand movements were more accurate. As reported in [13], the result indicated the selected TD single features provide more discriminative information than FD single features, and FD features imposed relatively high load of computation using support vector machine (SVM) to classify five limb motions. Altın *et al.* [14] obtained EMG signals for different elbow gestures and discussed the effects of the time domain and frequency domain. Ref. [15] proposed the behavior of fifty TD and FD features to classify ten upper limb motions and gained certain systematic conclusions, which signified that TD features show a better performance for pattern classification of upper limb. It's interesting to note that the previously contributions were focused on upper limb, and they shared different conclusions under similar conditions. Yet it is of vital necessity to explore the rules for the lower limb recognition.

## **2 Methods**

### **2.1 Data Acquisition**

During the human walking, many muscles participate in the motion process. Considering the role of different gait stage, four upper leg (thigh) muscles, i.e., Vastus

Medialis (VM), Adductor Longus(AL), Tensor Fascia Lata(TF), Semitendinosus(ST), and four lower leg muscles, i.e., Rectus Femoris(RF), Tibialis Anterior(TA), Gastrocnemius(GM) and Soleus(SO) Muscles, were selected to acquire EMG signals.



**Figure 1.** The experimental equipment and process

As shown in Fig. 1, the EMG signals were collected by the portable surface electromyography (DataLINK SX230-1000, UK Biometrics Ltd Company), the three-dimensional positional signals were collected by the motion capture system (Vicon, UK Oxford Metrics Limited Company). The voluntary testers were three healthy 25-year-olds, which were arranged to walk at the speed of 5km/h on the treadmill (Track Master TMX428CP, US Trackmaster Company). To reduce the interference of random factors, the volunteers were asked to walk 70 steps continuously[16]. Sampling frequencies of EMG signals was 1000 Hz, while the frequency of motion capture signals was 100 Hz.

## 2.2 Data Processing

The raw EMG signals contain some noises such as surrounding environmental noise, device noise and physical noise, which caused a certain distortion of the EMG signal acquired, and cause to a low signal-to-noise ratio. In this paper, for de-noising processing, a modulus maxim method is employed.

In this paper, TD features are extracted including mean absolute value *MAV* and variance *VAR*. They are proposed to represent the levels of muscle contraction, which can be expressed as follows[17],

$$MAV = \frac{1}{N} \sum_{i=1}^N |x_i| \quad (1)$$

$$VAR = \frac{1}{N-1} \sum_{i=1}^N x_i^2 \quad (2)$$

where  $x_i$  is the  $i$ -th EMG sample sequence,  $N$  is the length of EMG sample segment.

Subsequently two TD features, mean power frequency *MPF* and median frequency *MDF* were selected, for each feature sequence which are described as follows,

$$MPF = \frac{\int_0^{+\infty} fP(f)df}{\int_0^{+\infty} P(f)df} \quad (3)$$

$$\int_0^{MDF} P(f)df = \int_{MDF}^{+\infty} P(f)df = \frac{1}{2} \int_0^{+\infty} P(f)df \quad (4)$$

in which  $P$  denotes power spectral density (PSD) function of sEMG signal and  $f$  represents frequency. Thus, the feature vectors can constitute a feature matrix  $\mathbf{Z}$ ,

$$\mathbf{Z} = \begin{bmatrix} Z_{qg} \end{bmatrix} \quad (5)$$

where  $q = 1, 2, \dots, Q$  is the sequence of feature vectors, and  $g = 1, 2, \dots, G$  with  $G = uv$ , in which  $u$  is the number of selected muscles,  $v$  the number of feature vectors respectively. Thus, the samples can be transformed from the low-dimensional space to the high-dimensional space by SVM.

### 2.3 Gait phase

Human gait is a dynamic and periodic process. A gait cycle is defined as the time period between two consecutive heel strikes of the same foot. Through the calculation at the occurrence time for the heel strike and toe off, a gait cycle can be divided into the stance phase and swing phase, which can also be subdivided into many different phases. In the current study, five phases were applied according to proportion [1], i.e., the pre-stance phase, mid-stance phase, terminal stance phase, pre-swing phase and the terminal swing phase, respectively.

After de-noising processing and gait division, eight-dimension feature vectors were extracted for four-way EMG signals, which were abstracted by the sliding window method in five gait phases. In this paper, EMG data were divided into 100ms windows with 30ms increment. Then the feature vectors  $\mathbf{Z}$  were calculated according to Eq. (5).

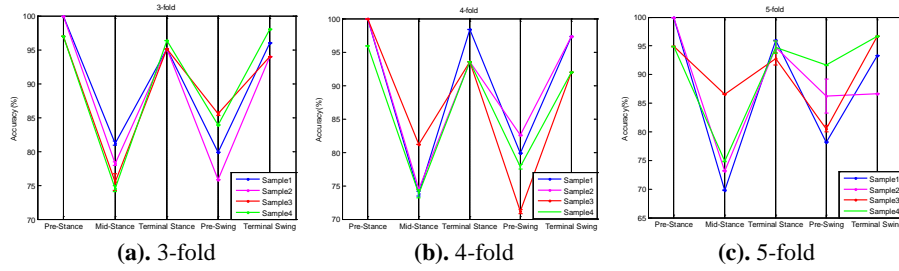
## 3 Results and Discussion

### 3.1 Cross Validation

In following analysis,  $k$ -fold cross validation is used to determine the validity and reliability of our results. It can effectively avoid over-learning and therefore the final result has to persuade nature even more.

The specific principle is to randomly divide all samples into  $k$  exclusive subsets of approximately equal size.  $k-1$  out of  $k$  subsets are used for training and the remaining one subset is used for testing. Continue to take a copy as a new test sample, the remaining  $k-1$  copies as new training samples. Proceed this way until all  $k$  samples are identified as test samples, and then average all recognition results as the estimate of the true generalized performance of the classifier [18].

According to the characteristics of the EMG signal, 3-fold, 4-fold and 5-fold were calculated, respectively. The recognition rates with cross validation were presented in Figure 2.



**Figure 2.** Results of  $k$ -fold cross validation

It's interesting to observe that the accuracy of five gait phases for all  $k$ -fold cross validation shows significant distinction. We can find that the recognition rates of the mid-stance and the pre-swing period are the lowest among the five periods and the variance is large, indicating that the recognition rate of these periods fluctuates greatly. In addition, in terminal-stance period, the variance of the 5-fold cross-validation is larger than other periods resulting in an obvious fluctuation. Conversely the results of 3-fold cross validation exhibit more stable and reliable. Therefore, 3-fold cross validation method is more applicable to be proposed for identification in this paper.

### 3.2 Time Domain (TD) Features

Time domain analysis method is the earliest traditional feature extraction method. Features in time domain are measured as a function of time. The data are assumed as stationary signals and do not need any transformation [19]. As a result, with the characteristics of their lower computational complexity and reduction in complexity that related with the feature extraction, low noise in environment, and easy to implementation, this kind of method has been extensively used in biological systems, engineering applications and medical researches [17].

In this paper, four thigh muscles (VM, AL, TF, ST), and two TD features (MAV, VAR) were selected. 8-dimensional feature vectors were obtained with the usage of the sliding window method. Take this feature vectors as input to the support vector machine with appropriate parameters [20], and the recognition rates and the averaged accuracies of five gait phases were obtained, as shown in Table 1.

**Table 1.** Classification accuracy in five gait phases (TD)

|                             | Pre-Stance (%) |      | Mid-Stance (%) |      | Terminal Stance (%) |      | Pre-Swing (%) |      | Terminal Swing (%) |      |      |      |      |      |      |
|-----------------------------|----------------|------|----------------|------|---------------------|------|---------------|------|--------------------|------|------|------|------|------|------|
| Classification Accuracy (%) | 93.4           | 95.7 | 96.4           | 72.3 | 79.1                | 69.9 | 92.2          | 93.1 | 93.3               | 76.4 | 85.3 | 80.2 | 97.9 | 97.3 | 96.3 |
| Average Accuracy (%)        | 95.17          |      | 73.77          |      | 92.87               |      | 80.63         |      | 97.17              |      |      |      |      |      |      |
| Variance                    | 2.46           |      | 22.77          |      | 0.34                |      | 19.94         |      | 0.65               |      |      |      |      |      |      |

It's illustrated that the averaged classification accuracy obtained from SVM of all test trials during terminal swing period is up to 97.17% instead mid-stance is only in 73.77%. At the same time, during the period of pre-stance, terminal stance and terminal swing, the recognition rates are higher than that of mid-stance and pre-swing.

The variance of the recognition rate reflects the discrete degree of random variables and describes the fluctuation of accurate data. From the perspective of variance, the periods of mid-stance and pre-swing were larger than that of terminal swing, which proved that the recognition rates in these periods are not stable.

### 3.3 Frequency Domain (FD) Features

Frequency or spectral domain is one of the methods used to analysis the signal data obtained in frequency. The feature extracted by this method are stable and its characteristic is to directly observe the frequency distribution and changes of surface electromyogram signals. The frequency domain analysis is mainly based on power spectrum analysis (PSD) [15]. Main feature parameters of PSD include Mean Power Frequency (MPF) and Median Frequency (MDF). MPF is a common feature. MF has been proved to possess the characteristics of sensitivity to physiological parameter change, strong anti-noise and signal aliasing ability.

Four thigh muscles (VM, AL, TF, ST), two FD features (MPF, MDF) were selected subsequently. With the usage of the same method, the recognition rates of five gait phases were obtained, as shown in Table 2.

**Table 2.** Classification accuracy of five gait phases (FD)

|                             | Pre-Stance (%) |      |      | Mid-Stance (%) |      |      | Terminal Stance (%) |      |      | Pre-Swing (%) |      | Terminal Swing (%) |      |      |      |
|-----------------------------|----------------|------|------|----------------|------|------|---------------------|------|------|---------------|------|--------------------|------|------|------|
| Classification Accuracy (%) | 94.6           | 99.7 | 98.8 | 83.0           | 78.3 | 75.3 | 90.0                | 93.9 | 91.6 | 91.9          | 85.1 | 91.5               | 97.2 | 99.1 | 99.5 |
| Average Accuracy (%)        | 97.69          |      |      | 78.87          |      |      | 91.83               |      |      | 89.50         |      | 98.60              |      |      |      |
| Variance                    | 6.74           |      |      | 15.06          |      |      | 3.84                |      |      | 14.56         |      | 1.51               |      |      |      |

The result of the recognition rate obtained by adding the FD features is improved to compare with that of the TD features, which indicates the FD features are notably effective. There was an obvious increase in pre-swing period than any others, and its variance of the frequency domain features was balanced. However, in mid-stance and pre-swing periods, the variances were still large and stable.

### 3.4 Time Domain and Frequency Domain Features

Four thigh muscles (VM, AL, TF, ST), two TD features (MAV, VAR) and two FD features (MPF, MDF) were selected. Take the same approach to obtain 16-dimensional feature vectors. The results are shown in Table 3.

**Table 3.** Classification accuracy rate of five gait phases (TD+FD)

|                             | Pre-Stance (%) |      |      | Mid-Stance (%) |      |      | Terminal Stance (%) |      |      | Pre-Swing (%) |      |      | Terminal Swing (%) |      |      |
|-----------------------------|----------------|------|------|----------------|------|------|---------------------|------|------|---------------|------|------|--------------------|------|------|
| Classification Accuracy (%) | 97.2           | 95.8 | 94.6 | 73.6           | 82.4 | 79.6 | 94.7                | 92.3 | 91.0 | 88.5          | 89.9 | 90.2 | 99.7               | 99.3 | 99.6 |
| Average Accuracy (%)        | 95.87          |      |      | 78.53          |      |      | 92.67               |      |      | 89.53         |      |      | 99.53              |      |      |
| Variance                    | 1.69           |      |      | 20.21          |      |      | 3.52                |      |      | 0.82          |      |      | 0.04               |      |      |

After combined two TD features with two FD features, the recognition rates of mid-stance and pre-swing periods have dramatically increased, the recognition rate of pre-stance period has elevated 20.6% but that mid-stance period is still unstable. The recognition rate of the multiple features is similar to that of FD features only, but the variance of the former is smaller. The more features are extracted, the more stable result and the less fluctuation can we get. TD and FD features are combined into high-dimensional feature vectors, which can improve the average rates of gait pattern recognition, but face with the disadvantages of the prolonged computational time.

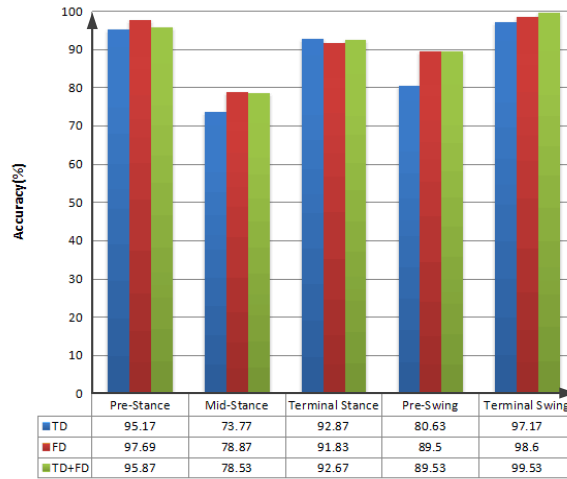
**Figure 3.** Combination of feature effect

Fig. 3 showed the results of the three groups: the recognition results of TD features are the lowest of all, while that of FD features and multiple features are similar, indicating that FD features have significantly increased classification accuracy as compared to the other two groups. Instead of improving the recognition rate greatly, combined features will spend more computational time. To further explore the effect of muscle number on recognition rate, we added four muscles as a reference.

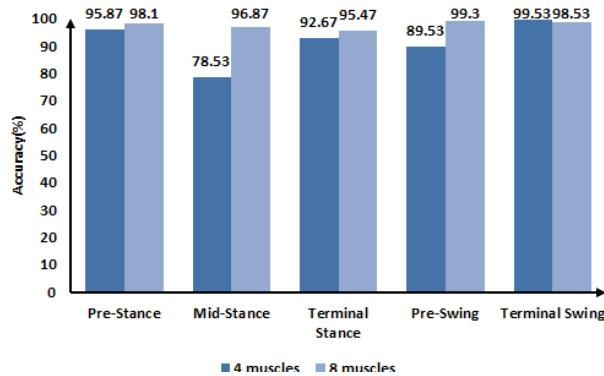
### 3.5 Muscle Number Effect

Eight muscles from Thigh and Calf (VM, AL, TF, ST, RF, TA, GM and SO Muscles), two TD features (MAV, VAR) and two FD features (MPF, MDF) were selected. Using the same method, the 32-dimension feature vectors were extracted, and the recognition rates obtained by 3-fold cross validation were shown in table 4.

**Table 4.** Classification accuracy rate of five gait phases.

|                             | Pre-Stance (%) |      |      | Mid-Stance (%) |      |      | Terminal Stance (%) |      |      | Pre-Swing (%) |      |      | Terminal Swing (%) |      |      |
|-----------------------------|----------------|------|------|----------------|------|------|---------------------|------|------|---------------|------|------|--------------------|------|------|
| Classification Accuracy (%) | 98.8           | 96.9 | 98.6 | 96.7           | 96.6 | 97.3 | 96.9                | 94.4 | 95.1 | 98.6          | 99.5 | 99.8 | 98.1               | 99.0 | 98.5 |
| Average Accuracy (%)        | 98.10          |      |      | 96.87          |      |      | 95.47               |      |      | 99.30         |      |      | 98.53              |      |      |
| Variance                    | 1.09           |      |      | 0.14           |      |      | 1.66                |      |      | 0.39          |      |      | 0.20               |      |      |

Compared with four muscles, it can be seen that the recognition rates of eight muscles have been improved greatly. The recognition accuracy of the Mid-Stance rises obviously which has reached from 78.53% to 96.87%. The stability of each period is also the best of the four experiments, and the variance is significantly reduced. Based on the results of four groups of experimental data, we found that the effects of muscle number augment on gait recognition rates are greater than that of feature number increasing. It is therefore suggested that the multi-channel signals can improve the accuracy of gait recognition.



**Figure 4.** Muscle number effect

Fig. 4 is a histogram of gait pattern recognition under different muscle groups. Before added four muscles, the average accuracy rate is 91.23%. The recognition rates are elevated as the muscle number augment. With the gradual increase of the recognition rate in each period, the differences in five periods had decreased, and each period is more uniform, reaching more than 90% of all, 97.65% on average.



### 3.6 Discussion

Based on normal human walking using four EMG channels, there are several interested issues that could be discussed as follows:

- (1) Time domain, frequency domain and time-frequency domain are three main groups in the analysis of sEMG feature. The reason why we consider TD and FD features is that feature extracted from time-frequency methods such as short-time Fourier transform and discrete wavelet transform should be reduced their high dimensions before sending them to a classifier, which require much more complicated processing and cannot be directly used by themselves. However, as mentioned in several studies, one of the major properties of surface EMG signal is nonlinearity. Nonlinear analysis techniques can extract the real hidden information from surface EMG data. As for further study, we shall use nonlinear features to detect the effective recognition.
- (2) From the results of our study and previously research, it can be found that the conclusion gained above differ from each other. But that doesn't mean the work we do is pointless. Here are some contrasts and introspection in summary. One of the issues that may matter the result is the difference in the total number of the feature selected. And different algorithms of previous studies were used, such as SVM and LDA. In Ref. [15], the classification results of multiple features can be distinctive between LDA and SVM. The optimal classifier needs to be further verified in gait analysis. On the other hand, each of the previous studies was conducted on different experimental settings and different upper limb motions. In the aspect of gait recognition, this research is far from maturity and still needs sustained investigation.

## 4 Conclusion

Gait recognition has application in the rehabilitation training and prosthesis control of lower limb amputees. We systematically analyse a continuous recording of human gait in order to obtain a deeper insight into the dynamics of the TD and FD features. In this paper, multiple features of EMG signals were utilized to calculate the recognition rate of five gait phases with SVM. The results show that FD features get the high accuracy of 91.29%, obviously elevated classification accuracy as compared to TD features and combination of TD and FD which demonstrate that FD features are better qualified and more appropriate to the lower limb prosthetic controlled system in rehabilitation. While TD and FD features were determined, the increase of muscle number can also improve the classification accuracy.

There are several limitations associated with this study that needs further development. While findings for the present study are only examined on healthy human, this work needs EMG data from disabled or the amputee as a contrast and other age as well.

## References

1. Vaughan, C.L., Davis, B.L., O'Connor, J.C.: Dynamics of Human Gait. Human Kinetics Publishers (1999)
2. Santilli, V., Frascarelli, M.A., Paoloni, M., Frascarelli, F., Camerota, F., De, L.N., De, F.S.: Peroneus Longus Muscle Activation Pattern During Gait Cycle in Athletes Affected by Functional Ankle Instability: A Surface Electromyographic Study. *American Journal of Sports Medicine* 33, 1183-1187 (2005)
3. Bogataj, U., Gros, N., Kljajić, M., Aćimović, R., Malezic, M.: The Rehabilitation of Gait in Patients with Hemiplegia: A Comparison between Conventional Therapy and Multichannel Functional Electrical Stimulation Therapy. *Physical Therapy* 75, 490-502 (1995)
4. Fish, D.J., Nielsen, J.-P.: Clinical Assessment of Human Gait. *Journal of Prosthetics & Orthotics* 5, 39-48 (1993)
5. Tao, W., Liu, T., Zheng, R., Feng, H.: Gait Analysis Using Wearable Sensors. *Sensors* 12, 2255-2283 (2012)
6. JY, H.: Clinical Applications of Surface Electromyography in Neuromuscular Disorders. *Clinical Neurophysiology* 35, 59-71 (2005)
7. Wang, P., Low, K.H.: Qualitative Evaluations of Gait Rehabilitation via EMG Muscle Activation Pattern: Repetition, Symmetry, and Smoothness. In: *IEEE International Conference on Robotics and Biomimetics*, pp. 215-220, Guilin, China (2009)
8. Boostani, R., Moradi, M.H.: Evaluation of the Forearm EMG Signal Features for the Control of A Prosthetic Hand. *Physiological Measurement* 24, 309-319 (2003)
9. Zecca, M., Micera, S., Carrozza, M.C., Dario, P.: Control of Multifunctional Prosthetic Hands by Processing the Electromyographic Signal. *Critical Reviews in Biomedical Engineering* 30, 459-485 (2002)
10. Burhan, N., Kasno, M.A., Ghazali, R.: Feature Extraction of Surface Electromyography (sEMG) and Signal Processing Technique in Wavelet Transform: A Review. In: *IEEE International Conference on Automatic Control and Intelligent Systems*, pp. 141-146 (2017)
11. Asghari Oskoei, M., Hu, H.: Myoelectric Control Systems-A Survey. *Biomedical Signal Processing and Control* 2, 275-294 (2007)
12. Too, J., Abdullah, A.R., Zawawi, T.N., Saad, N.M., Musa, H.: Classification of EMG Signal Based on Time Domain and Frequency Domain Features. *International Journal of Human and Technology Interaction* 1, 25-29 (2017)
13. Oskoei, M.A., Hu, H.: Support Vector Machine-based Classification Scheme for Myoelectric Control Applied to Upper Limb. *IEEE Transactions on Biomedical Engineering* 55, 1956-1965 (2008)
14. Altın, C., Er, O.: Comparison of Different Time and Frequency Domain Feature Extraction Methods on Elbow Gesture's EMG. *European Journal of Interdisciplinary Studies* 5, 35-44 (2017)
15. Phinyomark, A., Quaine, F., Charbonnier, S., Serviere, C., Tarpin-Bernard, F., Laurillau, Y.: EMG Feature Evaluation for Improving Myoelectric Pattern Recognition Robustness. *Expert Systems with Applications* 40, 4832-4840 (2013)
16. Wang, J., Gao, F., Yao, S., Luo, Z.: Non-Uniform Characteristics and Its Recognition Effects for Walking Gait Based on sEMG. *Chinese Journal of Sensors & Actuators* (2016)
17. Phinyomark, A., Phukpattaranont, P., Limsakul, C.: Feature Reduction and Selection for EMG Signal Classification. *Expert Systems with Applications* 38, 7420-7431 (2011)
18. Kohavi, R.: A Study of Cross-validation and Bootstrap for Accuracy Estimation and Model Selection. In: *International Joint Conference on Artificial Intelligence*, pp. 1137-1143 (1995)
19. Hudgins, B., Parker, P., Scott, R.N.: A New Strategy for Multifunction Myoelectric Control. *IEEE Transactions on Biomedical Engineering* 40, 82-94 (1993)
20. Gao, F.R., Wang, J.J., Xi, X.G., She, Q.S., Luo, Z.Z.: Gait Recognition for Lower Extremity Electromyographic Signals Based on PSO-SVM Method. *Journal of Electronics & Information Technology* 37, 1154-1159 (2015)