Article

Estimation of Knee Movement from Surface EMG Using Random Forest with Principal Component Analysis

Zhong Li , Xiaorong Guan *, Kaifan Zou and Cheng Xu

School of Mechanical Engineering, Nanjing University of Science and Technology, Nanjing 210094, China; zhong0814@njust.edu.cn (Z.L.); zoukaifan@njust.edu.cn (K.Z.); xucheng62@njust.edu.cn (C.X.)

* Correspondence: gxr@njust.edu.cn; Tel.: +86-186-5209-6966

Received: 28 November 2019; Accepted: 25 December 2019; Published: 28 December 2019

Abstract: To study the relationship between surface electromyography (sEMG) and joint movement, and to provide reliable reference information for the exoskeleton control, the sEMG and the corresponding movement of the knee during the normal walking of adults have been measured. After processing the experimental data, the estimation model for knee movement from sEMG was established using the novel method of random forest with principal component analysis (RFPCA). The influence of the sample size and the previous sEMG data on the prediction efficiency was analyzed. The estimation model was not sensitive to the sample size when samples increased to a certain value, and the results of different previous sEMG showed that the prediction accuracy of the estimation models did not always improve with the increasing features of input. By comparing the estimation model of back propagation neural network with principal component analysis (BPPCA), it was found that RFPCA was suitable for all participants in the experiment with less execution time, and the root mean square error was around 5° which was lower than BPPCA with errors varying from 7° to 25°. Therefore, it was concluded that the RFPCA method for the estimation of knee movement from sEMG is feasible and could be used for motion analysis and the control of exoskeleton.

Keywords: sEMG; knee; random forest; principal component analysis; back propagation; estimation model

1. Introduction

The exoskeleton robot could be a revolutionary technology in human limb rehabilitation [1] and power enhancement [2]. However, the motion intention of its wearer has limited the development of this technology, because traditional sensors for the exoskeleton are unable to detect motion tendency ahead of time. Since the surface electromyography (sEMG) signal is noninvasive, and has the potential to predict people's movement intentions 30–100 ms in advance [1], the sEMG has been favored by many researchers, and with the help of sEMG sensors, the performance of wearable devices would be improved [1,3–5]. Thus, in addition to the exoskeleton technology, there are a wide range of applications for using sEMG, such as wearable devices [6], prosthetic limbs [7], and other such myoelectric control systems [8]. The existing studies focus on how the sEMG signals relate to human movement for better control of the wearable devices and similar products.

Some of the existing research has investigated the relationship between sEMG and human biomechanics. Chen et al. [9] proposed a musculoskeletal biomechanical model connecting sEMG and knee joint torque, based on the underlying physiological mechanism facilitating the study of neural control. Tagliapietra et al. [10] used a subject-specific EMG-driven Neuro MusculoSkeletal (NMS) model to estimate ankle torque and muscle forces expressed by the subject. Zhuang et al. [11] proposed

an sEMG-based admittance controller that could enable a more synchronized human–robot interaction, as compared to the torque-sensing-based admittance controller.

However, to avoid building a complicated biomechanical model, some researchers have tried to use a data training method to do the job. Anwar et al. [12] proposed an adaptive neuro fuzzy inference system (ANFIS), such as a neuro-fuzzy type knowledge-based adaptive network that contained a non-parametric model, with an EMG signal of two muscles used as the input to estimate torque. Gui et al. [13] used radial basis function (RBF) neural networks to approximate the active joint torque of subjects during the swing phase.

Besides the biomechanical method, there is a new intuitive myoelectric control strategy for assistive devices, which relies on the sEMG-based intention estimation of human motion. These predictions can be broadly categorized as classification and regression models [14]. For the classification, Toledo-Pérez et al. [15] used a support vector machine (SVM) based on sEMG to classify the intention of right lower limb movement. Morbidoni et al. [16] proposed a deep learning (DL) approach for sEMG-based classification of stance/swing phases and the prediction of the foot–floor-contact signal in more natural walking conditions. Nazmi et al. [17] proposed a classification system for both stance and swing phases, by extracting the patterns of electromyography signals from time domain features and feeding them into an artificial neural network (ANN) classifier.

For the continuous estimation of the joint angle, there are various methods. Bao et al. [18] presented a single stream convolutional neural network (CNN) for mapping sEMG to wrist angles within three degrees-of-freedom. Xiao et al. [19] used the mean absolute value, waveform length, zero crossing, slope signs changes, and the difference in absolute standard deviation value of sEMG, in order to estimate continuous elbow motion by random forest (RF). Lei Z. [20] used the back propagation (BP) neural network to establish a model of the relationship between elbow angles and sEMG signals features, through which they estimated the angles of the elbow joint and achieved continuous motion control of the exoskeleton. Huang et al. [21] presented deep-recurrent neural networks (RNNs) for predicting the knee joint angle in real-time, based on a fusion of sEMG and kinematics signals.

It can be concluded that most of the existing studies are model-free approaches for the estimation of joint angles from sEMG, based on machine learning (ML). Furthermore, the majority of the existing research used a single method of ML, and most of them seldom considered the influence of the previous sEMG as the input on the accuracy of their methods, even though both the joint movement and sEMG signals are in a continuous time-sequence. Also, the size of the training sample for ML is also a debatable issue in terms of the accuracy of estimation, since a large training sample would lead to overfitting and be time consuming.

Thus, a novel double-ML-method, based on random forest combined with principal component analysis (RFPCA) has been proposed in this work to estimate the movement of the knee joint from sEMG, with the expectation of achieving high accuracy and efficiency. This method was also utilized to analyze how the input of the previous sEMG and the sample size for model validation affect the estimation of the knee joint movement. Moreover, a BPPCA was constructed to compare with the RFPCA, and the RFPCA presented a better performance in this work.

The remainder of the paper is organized as follows. Section 2 introduces the experiment and proposes the knee angle estimation method. Section 3 presents the estimation results using RFPCA and BPPCA, followed by the discussion in Section 4. Finally, the conclusions are drawn in Section 5.

2. Materials and Methods

2.1. Subjects

A total of six healthy subjects, who have never suffered from muscular atrophy or disorders, participated in this study, with an average height, age and weight of 181 ± 3.8 cm, 72.5 ± 6.9 kg and 24.2 ± 1.6 yrs, respectively. All of the subjects enrolled in this study knew the procedure of the experiment, and signed and agreed to participate in the experimental study as the test subjects.

The experiment was approved by Nanjing University of Science and Technology (Date granted: June 12, 2019) and performed in accordance with the Declaration of Helsinki.

2.2. Experimental Preparation and Protocol

The unilateral lower limb of the human body contains at least 30 muscles that drive 7 degrees of freedom of the lower limb. Since the knee is a vital joint in the lower limb, we have chosen it as the key aspect in our research. The sEMG signals related to the knee of the unilateral leg were obtained by Trigno wireless sEMG instrument. To obtain better sEMG signals, several muscles that are easily detected were selected, including the vastus lateralis (VL), rectus femoris (RF), vastus medialis (VM), gastrocnemius medialis (GM) and gastrocnemius lateralis (GL). The adhesive positions (shown in Figure 1) of the surface electrodes are the positions recommended by SENIAM [22]. Before the experiment, the skin surfaces, where the electrodes had to be placed, were shaved and then cleaned with alcohol. This was done to reduce the impedance between the measured skin and electrodes and also to improve the sensor–skin contact.

Figure 1. The locations of surface electromyography (sEMG) sensors.

On the lateral side of the leg with sEMG sensors, three markers were placed to obtain the kinematic data of the knee, through a 3D motion capture system called Codamotion. As shown in Figure 2, Marker B was placed at the approximate center of the knee joint, on the sagittal skin of the subject, while Marker A and Marker C were respectively placed on the projection line of the thigh femur and calf tibia on the sagittal skin of the subject. The approximate flexion–extension motion angle of the knee joint was obtained by collecting the spatial motion trajectory of Markers A, B and C.

Figure 2. The placement of the Codamotion markers.

During the experiment, the subject was asked to walk back and forth nearly 20 times along a straight line about 5 m in length, at his natural or free cadence, as shown in Figure 3. The two Codamotion CX1 units on the right of the subject collected the kinematic data of the markers at a sampling frequency of 100 Hz, while the wireless Trigno system picked up the raw sEMG signals from the five muscles at a frequency of 2000 Hz. The data were then synchronized and transferred to

the Codamotion hub, and finally, transmitted to the PC and stored. If the subject felt uncomfortable while the experiment was in progress, he would rest for 10 minutes to reduce the effects of such factors as fatigue. If he felt better after the rest, the experiment would continue. Otherwise, the experiment would be stopped. In this way, the data of one trail from as many as 8 strides could be collected, which was more than adequate to obtain 4 representative sEMG profiles for each muscle and gait cycle (GC) of the leg. Approximately 80 full GCs and their corresponding sEMG signals were recorded in all for each subject. Since each subject needed to be properly equipped with the sensors, and it required approximately one hour for each subject to undergo the experiment process, including preparation (15 min), resting (30 min) and walking (15 min), the entire experiment lasted 2 days, with one person in the morning and two in the afternoon.

Figure 3. The schematic of the experiment.

2.3. Signal Processing

The sEMG signals were focused on the time domain because of the time-sequence. The amplitude extracted from the raw sEMG signals was used for the training of the RFPCA, similar to the work done by the authors of [16], who directly used the envelopes of the EMG signal to train the network. This feature was carried out from digital filtering and using simple math. A Butterworth band-pass filter (band length of 10–500 Hz, 4th order) was used to filter interference signals and extract effective signals. After that, a full-wave rectification of the sEMG signal was conducted, and the signal was then filtered through a low-pass filter (Butterworth at 6 Hz, 2nd order) [23]. One of the representative examples of the sEMG signal processing of one trial is shown in Figure 4.

Figure 4. The signal processing of sEMG.

For processing the kinematic data of the knee, the vector $\vec{V}_{AB}$ consisting of Markers A and B represents the direction of the thigh, while vector $\vec{V}_{BC}$ consisting of Markers B and C represents the direction of the shank. The motion angle of knee could then be calculated approximatively as:

$$\theta_{knee} = \arccos \left(\frac{\vec{V}_{AB} \cdot \vec{V}_{BC}}{\|\vec{V}_{AB}\| \|\vec{V}_{BC}\|} \right). \quad (1)$$

In order to keep the data synchronized with the knee and facilitate subsequent knee movement prediction, the sEMG signal, after the original processing, was resampled at 100 Hz. Thus, a set of data $[E_{VL,t}, E_{RF,t}, E_{VM,t}, E_{GM,t}, E_{GL,t}, y_{knee,t}]$ at time t was obtained, where $E_{i,t}$ represents the envelope of i sEMG signal after processing, and $y_{knee,t}$ represents the knee angle θ_{knee} at time t .

For the study of the sEMG historical influence, the data from one trail of each subject was chosen for training, in order to establish the estimation models. The influence of previous sEMG signals was studied by the number of previous sEMG signals used as a feature vector in the learning course. A typical sample is a combination of sEMG and knee angle $[X_t, Y_t]$, where the sEMG X_t is the input, while the knee angle $Y_t = y_{knee,t}$ is the output. In this work, the sEMG input X_t at time t is defined as follows:

$$X_t = [x_{t-n\Delta t}, x_{t-(n-1)\Delta t}, \dots, x_{t-\Delta t}, x_t], \quad (2)$$

where, x_t is the vector of sEMG at time t after processing, namely $[E_{VL,t}, E_{RF,t}, E_{VM,t}, E_{GM,t}, E_{GL,t}]$; n is the number of previous sEMG signals used as the input, Δt is the sample interval which is 1/100 Hz. The input dimension of X_t increases as n increases.

For the sample size, n was zero, and all of the processed data of each subject was segmented into groups according to the GC. One group contains different sampling numbers for different subjects, which are listed in Table 1. For each subject, 61 complete groups of data (GDs, approximately 7808 samples) were selected for the following work.

Table 1. The average sampling numbers contained in a gait cycle (GC) of different subjects.

Subject	1#	2#	3#	4#	5#	6#	Mean
Number	129 ± 9	127 ± 6	126 ± 7	131 ± 6	134 ± 9	122 ± 7	128 ± 8

2.4. Estimation Model

Because of the complexity of the sEMG signal and the differences between subjects, it is hard to establish a general mathematical model to represent the mapping relationship from sEMG to the knee angle. Furthermore, the biomechanical model describing the relationship between the sEMG and joint angle is also complicated and difficult to construct for practical application. Therefore, to establish a universal sEMG–angle model of a human joint, with a learning function, this study adopted a novel model-free method using random forest (RF) in combination with principal component analysis (PCA), in order to set up the estimation model between sEMG signal and knee movement. It is expected that this coupled ML method will be able to handle the estimation issue for different participants with a parametric adaptive approach. The input of the model is the processed sEMG, and the output is the knee angle.

2.4.1. Random Forest

Random forest (RF), based on the theory of decision trees, was first proposed by Breiman [24]. RF is an effective tool in prediction, because with the right input, RF produces accurate classifiers and regressors [24]. In standard trees, each node is split using the best split among all variables. In a RF, each node is split using the best among a subset of predictors randomly chosen at that node. This somewhat counterintuitive strategy performs very well in comparison to other classifiers, including discriminant analysis, SVM and neural networks, and is robust against overfitting [25]. Additionally, the execution time of RF is far less than RBF and SVM when used to process high dimensional data, because the RF algorithm itself can select the important features automatically [19]. It is also relatively robust to outliers and noise. Due to these advantages, RF was chosen for application to the relationship between the sEMG and knee joint in this study.

As an ensemble learning method, RF achieves better generalization performance by establishing multiple decision trees. If RF has N decision trees, it is necessary to generate N sample sets to train each tree. Each tree is grown as follows [26]:

- If the number of cases in the training set is T_r , then T_r cases at random are sampled—but with replacement, from the original data. This sample will be the training set for growing the tree.
- If there are I_v input variables, a number $N_f \ll I_v$ is specified, such that at each node, N_f variables are selected at random out of the I_v and the best split on these, m , is used to split the node. The value of N_f is held constant during the forest growing.
- Each tree is grown to the largest extent possible. There is no pruning.

2.4.2. Principal Component Analysis

Principal component analysis (PCA) is a statistical technique that performs a linear transformation from an original set of values into a smaller one of uncorrelated variables [15]. The idea was conceived of by K. Pearson [27] and later developed by Hotelling [28].

PCA needs to find some encoding function that produces the code for an input, and a decoding function that produces the reconstructed input, given its code [29]. In general, PCA uses a covariance matrix to reduce the data dimension. By calculating the eigenvalue eigenvector of the covariance matrix of the data, and selecting the matrix composed of eigenvectors corresponding to k features with the largest eigenvalue (i.e., the largest variance), the data matrix can be converted into a new space to achieve dimensional reduction of data features. However, in our study, PCA is used to improve the estimation performance, rather than dimension reduction, which is the same PCA application described in [15].

2.4.3. Random Forest with Principal Component Analysis

The structural diagram of the RFPCA method used for the knee estimation from the sEMG is shown in Figure 5. In this figure, the blue lines represent the training process and the red lines describe the studying course. The time-domain amplitudes were extracted from 5 sEMG channels. After the process and PCA, one part of the data would be used for training to build the estimation model and the rest would be used for testing.

Figure 5. The structural diagram of the random forest with principal component analysis (RFPCA) used for knee estimation from sEMG.

In our work, the processed sEMG X_t forms the original data set X_{emg} , and X_{emg} which is then transformed through a matrix P by PCA. Next, the input X of the following RF is obtained, which is described as follows:

$$X = PX_{emg}. \quad (3)$$

with the combination of X and the knee angle y_{knee} , the data set after PCA is derived as (X, y_{knee}) . These data are divided into training set and testing set.

When the training sample $T_r = (X_{tr}, Y_{tr})$ is given, the goal is to use the T_r to establish a estimation model $E(X)$ and apply it to estimate the new knee angle from sEMG. The RF consists of a collection of N randomized regression trees $r(X, v_i)$, where v_i ($i = 1, 2, \dots, N$) are the independent random

variables. They are used to resample the training set and select the successive direction for splitting [19]. The estimation model $E(X)$ is an average of the regression trees in the RF, expressed as:

$$E(X) = \frac{\sum_{i=1}^N r(X, v_i)}{N}, \quad (4)$$

When the testing data set X_{test} is used to validate the estimation performance, the estimated knee angle can be calculated as follows:

$$\hat{y}_{knee} = E(X_{test}; v_1, v_2, \dots, v_N, T_r) = \frac{1}{N} \sum_{i=1}^N r(X_{test}; v_i, T_r). \quad (5)$$

2.5. Arguments Selection

For comparison with RFPCA in this study, another method, namely BPPCA, was also presented. This method is similar to RFPCA in Section 2.4.3, with a substitution of a conventional BP for RF. For different methods, there are different arguments regarding the effect they have on the performance of the estimation. In a model of RF, there are three main parameters to be determined, namely the number of trees in the forest (N), the number of features of the input (N_f) and the minimum size of the terminal nodes (N_m). The parameters for the BP are the number of epochs (N_e), the number of hidden layer nodes (N_h), the learning rate (L_r) and the learning goal (L_g).

During the test of the ML methods, the results showed that the performance of motion estimation was not sensitive to the difference of individuals using the same parameters, which was similar to the results found in [19]. Thus, it meant that universal values could be chosen from these parameters. Referring to the parameter settings in [19], and to reach a compromise between estimation accuracy and computational time, the parameters for our methods were set as recorded in Table 2.

Table 2. The values of parameters of the machine learning (ML) methods.

Argument	N	N_f	N_m	N_e	N_h	L_r	L_g
Value	50	5	5	100	40	0.1	0.00001

To study the historic effect of previous sEMG, the data from one trail with 4 GCs from each subject would be introduced. A total of 75% of the data was used for training and the rest was used for the testing set.

For the study of the sample size, as noted, 61 GDs were utilized for the estimation work of one subject, and the last GD, with approximately 128 samples, was always the testing set. Hereby, parameter S_s in the interval of [1, 60] was defined to represent the sample size. For example, when $S_s = 1$, the 60th GD would be the training set, when $S_s = 2$, the 60th and the 59th GDs were the training set, and by that logic, when S_s was 60, all of the 60 GDs would be used for training. The sample size was nearly proportional to the S_s .

2.6. Evaluation Metric

The performance of the estimation model was mainly evaluated by comparing the root mean square error (R) between the estimated value $\hat{y}_{knee}$ and the experiment value (EV) y_{knee} defined as follows:

$$R = \sqrt{\frac{1}{S} \sum_{i=1}^S (\hat{y}_{knee,i} - y_{knee,i})^2}, \quad (6)$$

where S is the number of test samples. In this work, R is an indicator for the judgement of the estimation models, and the accuracy of the estimation model increases with a decrease of R .

3. Results

In this section, different kinds of estimation methods for the estimation of the knee from the sEMG signals were compared to investigate whether the sample size and the previous sEMG signals would affect the performance of the estimation. All of the data processing was done on the software of MATLAB R2016b. In order to avoid the influence of the processor on the computation time, all the execution time in our study is relative.

3.1. The Results of Different Sample Size

Different sample sizes from 1 GD to 60 GDs were utilized for RFPCA and BPPCA to predict the knee joint angle.

Partial results of Subject 1 for estimation are shown in Figure 6. It can be seen that, as S_s increases, the estimations of BPPCA and RFPCA are close to the EV. However, the estimation results of RFPCA are always closer to the EV and more robust than BPPCA.

Figure 6. Partial results for estimation with different sample sizes of Subject 1.

The relative execution time (the execution time of each calculation processed against the longest computation among all of the sample size results), and the evaluation metric for the estimation with different methods and different subjects, are depicted in Figure 7. As the sample size increases, so does the execution time. Compared to the increase in time using BPPCA, RFPCA appears more efficient for the training, as it is less time consuming. Although the value of R decreases and the error decreases with the larger sample size from 1 to 10 GDs, both of the methods have no significant changes after $S_s = 20$. When $S_s \geq 30$, the R of BPPCA is slightly smaller than RFPCA, with a longer execution time.

Figure 7. The relative execution time and R for the estimation with different S_s .

3.2. The Results of Different Previous sEMG Input

In the process of training, different previous sEMG signals were used, and the result of relative execution time and the evaluation metric are shown in Figure 8. As the attributes of input X_t increase, the execution time of the different methods also increases, and this is more pronounced for BPPCA. RFPCA is much smaller, which is conducive to the application of myoelectric control. For the results of R , the estimation errors of different methods are various. The BPPCA results are much larger than RFPCA, and the method is insensitive to the input dimension when n is larger than approximately 7. The results of RFPCA seem to increase when n is bigger than 2. The results of R also show that BPPCA has larger standard deviations than RFPCA.

Figure 8. The relative execution time and R for the estimation with different previous input.

The estimation results of all six subjects using various methods when $n = 2$ are shown in Figure 9. From the figure, it is clear that the EV tracking errors of RFPCA are much smaller than BPPCA.

Figure 9. The estimation results of different models from different subjects when $n = 2$.

Furthermore, the evaluation metric R of different models for different subjects when $n = 2$ is depicted in Figure 10. The R of RFPCA is almost 5° , while the BPPCA has a poor prediction ability of the knee angle estimation, showing a large variation between different subjects, with errors ranging between 7° to 25° .

Figure 10. The evaluation metric of different models for different subjects when $n = 2$.

4. Discussion

In this paper, a novel estimation method of RFPCA was proposed to study the relationship between sEMG and knee movement. Compared with the results of BPPCA, the RFPCA performs better, both in terms of the root mean square error and the execution time. All of the estimation results using RFPCA are also generally in line with the EV. These results may be due to the strong regression ability of RF, which generates an internal unbiased estimate of the generalization error as the forest building progresses. PCA is able to generate a better input for RF from the original data, which also promotes the accuracy of the results of estimation.

As seen in Figure 7, with the increasing of input samples, R starts to decrease and eventually stabilizes, which means that the prediction accuracy increases at first, and then does not change significantly for both methods. As known, walking is a regular movement, the kinematic parameter of the gait is a cyclic process and the sEMG also appears as a periodic signal in different GCs. Thus, we believe that when the sample size increases to a certain value, the differences between the samples decrease, so that the prediction results show little change. However, the larger the data, the longer the time. With acceptable accuracy, choosing a better sample size can effectively reduce the time consumption caused by large samples, and this would contribute to the efficiency of online control using sEMG.

The historic effect of the input has a positive influence on motion estimation, according to the authors of [19]. In this work, as the input size increases, there is a tendency for the R of both RFPCA and BPPCA to increase, as shown in Figure 8, and the previous signals seem to have little to no effect on the estimation after more dimensional data is involved in the calculation. Generally, in the high-dimensional input case, the problem of sparse data samples and the difficulty of distance calculation are a common and serious obstacle for all machine learning methods, which is called the "curse of dimensionality". Thus, except for the PCA used in our work, further study of the input dimension of sEMG needs to be considered. Multichannel sensors of sEMG detecting are also worthy of research.

As seen in Figure 9, in terms of the physical and mental quality of the test subjects, the prediction results of different subjects also vary. In addition, the skin preparations of the subjects vary from one subject to another, which can also contribute to detection errors of in the raw sEMG. Moreover, different times for the experiment and other environmental factors will also cause diversity in the outcome when the raw sEMG is being collected. However, both Figures 9 and 10 show that the results of the BPPCA are more unstable for all participants, relatively, and that the results of the RFPCA have estimations similar to the EV. That is, RFPCA has a better error tolerance and is not adversely affected by variations in test subjects.

The root mean square error of RFPCA was small ($R \approx 5^\circ$) in the previous sEMG study when $n = 2$, as shown in Figure 10, and this result seems to be large in comparison to the motion angle of the knee. However, we believe that this is acceptable for the exoskeleton control, since the control precision of the exoskeleton joint is not in high demand as a machine tool, and the pilot of the exoskeleton is able to tolerate the small differences of several degrees due to the flexibility of the human body. Also, the estimation model using RFPCA has some hyperparameters in the structure of the forest. Therefore, a better parameter selection may lead to a more accurate estimation model of RFPCA, and this would also help the application of RFPCA for estimation of joint movement in myoelectric control. Furthermore, while the RFPCA performed better in our study, the BPPCA may have advantages over the RFPCA with a better parameter choice.

Since the aim of this work is using ML to build an estimation model which can be adjusted to suit the subject himself, the training and testing data are from the same subject for each validation, and a similar method can be found in [30]. The results of the subjects in this work using RFPCA was acceptable. However, sEMG is unstable, as mentioned above, and the RFPCA method may be unfeasible for subjects who are middle-aged and elderly, since the subjects in our study were under 30 years of age, and the data from different people for training and testing may be an issue through the method of RFPCA. Thus, more people will be invited to participate in the study to further validate the proposed method.

5. Conclusions

In this study, a new method of combining RF and PCA was utilized to establish an estimation model from sEMG to the knee. Better than BPPCA, the RFPCA method is able to predict the knee angle at a low root mean square error about 5° and the execution time is several times smaller, and these results indicate that RFPCA is suitable for knee movement estimation with high accuracy and requiring minimal time. Both the sample size and the input dimension of the RFPCA were investigated. RFPCA is insensitive to the estimation accuracy when the sample size increases to a certain value, but requires more time. Also, as the previous sEMG increases, the accuracy of the RFPCA first increases and then decreases, and the work would be beneficial to the estimation model construction using ML methods. Moreover, the results of the RFPCA are more stable in comparison to the BPPCA in terms of sample size and previous sEMG input, and RFPCA is also robust in terms of differences among test subjects. All in all, the estimation of RFPCA performs well in the estimation from sEMG to knee movement in this work, which is conducive to motion analysis and exoskeleton control.

Future work of improving the estimation accuracy will be done by extracting more signal features or fusing other physical sensors, as well as testing more people and other daily activities such as different speeds of walking, and ascending or descending stairs. Furthermore, a wider application for RFPCA may be utilized for estimation using different sEMG data from various people to find a general relationship between sEMG and joint motion. Finally, the RFPCA method will be applied to estimate human motion for exoskeleton control.

Author Contributions: Conceptualization, Z.L., X.G. and C.X.; methodology, Z.L. and X.G.; software, Z.L.; validation, Z.L., X.G. and C.X.; formal analysis, Z.L. and X.G.; investigation, Z.L. and K.Z.; resources, Z.L. and K.Z.; data curation, Z.L., X.G. and C.X.; writing—original draft preparation, all the authors; writing—review and editing, all the authors; visualization, Z.L. and X.G.; supervision, X.G. and C.X.; project administration, X.G.; funding acquisition, X.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by NATIONAL DEFENSE BASIC SCIENTIFIC RESEARCH PROGRAM OF CHINA, grant number B1020132012.

Acknowledgments: Thanks to the help of the participants in the experiment and the support of Laboratory for Individual Equipment Technology at School of Mechanical Engineering, Nanjing University of Science and Technology. We are also grateful for the assistance of the editors and reviewers.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Fan, Y.; Yin, Y. Active and Progressive Exoskeleton Rehabilitation Using Multisource Information Fusion from EMG and Force-Position EPP. *IEEE Trans. Biomed. Eng.* **2013**, *60*, 3314–3321. [CrossRef]
2. Chen, C.; Du, Z.; He, L.; Wang, J.; Wu, D.; Dong, W. Active Disturbance Rejection with Fast Terminal Sliding Mode Control for a Lower Limb Exoskeleton in Swing Phase. *IEEE Access* **2019**, *7*, 72343–72357. [CrossRef]
3. Chae, J.; Jin, Y.; Sung, Y.; Cho, K. Genetic Algorithm-Based Motion Estimation Method using Orientations and EMGs for Robot Controls. *Sensors* **2018**, *18*, 183. [CrossRef]
4. Kang, I.; Kunapuli, P.; Hsu, H.; Young, A.J. Electromyography (EMG) Signal Contributions in Speed and Slope Estimation Using Robotic Exoskeletons. In Proceedings of the 16th International Conference on Rehabilitation Robotics, Toronto, ON, Canada, 24–28 June 2019; pp. 548–553. [CrossRef]
5. Moon, D.H.; Kim, D.; Hong, Y.D. Intention Detection Using Physical Sensors and Electromyogram for a Single Leg Knee Exoskeleton. *Sensors* **2019**, *19*, 4447. [CrossRef]
6. Kartsch, V.; Benatti, S.; Mancini, M.; Magno, M.; Benini, L. Smart Wearable Wristband for EMG based Gesture Recognition Powered by Solar Energy Harvester. In Proceedings of the IEEE International Symposium on Circuits and Systems, Florence, Italy, 27–30 May 2018; pp. 1–5. [CrossRef]
7. Resnik, L.; Huang, H.; Winslow, A.; Crouch, D.L.; Zhang, F.; Wolk, N. Evaluation of EMG pattern recognition for upper limb prosthesis control: A case study in comparison with direct myoelectric control. *J. Neuroeng. Rehabil.* **2018**, *15*, 23:1–23:13. [CrossRef]
8. Jang, G.; Kim, J.; Lee, S.; Choi, Y. EMG-based continuous control scheme with simple classifier for electric-powered wheelchair. *IEEE Trans. Ind. Electron.* **2016**, *63*, 3695–3705. [CrossRef]
9. Chen, J.; Zhang, X.; Gu, L.; Nelson, C. Estimating Muscle Forces and Knee Joint Torque Using Surface Electromyography: A Musculoskeletal Biomechanical Model. *J. Mech. Med. Biol.* **2017**, *17*, 3:1–3:21. [CrossRef]
10. Tagliapietra, L.; Vivian, M.; Sartori, M.; Farina, D.; Reggiani, M. Estimating EMG signals to drive neuromusculoskeletal models in cyclic rehabilitation movements. In Proceedings of the 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Milan, Italy, 25–29 August 2015; pp. 3611–3614. [CrossRef]
11. Zhuang, Y.; Yao, S.; Ma, C.; Song, R. Admittance Control Based on EMG-Driven Musculoskeletal Model Improves the Human–Robot Synchronization. *IEEE Trans. Ind. Inf.* **2018**, *15*, 1211–1218. [CrossRef]
12. Anwar, T.; Al-Jumaily, A.; Watsford, M. Estimation of Torque Based on EMG using ANFIS. *Procedia Comput. Sci.* **2017**, *105*, 197–202. [CrossRef]
13. Gui, K.; Liu, H.; Zhang, D. A Practical and Adaptive Method to Achieve EMG-Based Torque Estimation for a Robotic Exoskeleton. *IEEE-ASME Trans. Mech.* **2019**, *24*, 483–494. [CrossRef]
14. Bi, L.; Feleke, A.G.; Guan, C. A review on EMG-based motor intention prediction of continuous human upper limb motion for human-robot collaboration. *Biomed. Signal Process.* **2019**, *51*, 113–127. [CrossRef]
15. Toledo-Pérez, D.; Martínez-Prado, M.; Gómez-Loenzo, R.; Paredes-García, W.; Rodríguez-Reséndiz, J. A Study of Movement Classification of the Lower Limb Based on up to 4-EMG Channels. *Electronics* **2019**, *8*, 259. [CrossRef]
16. Morbidoni, C.; Cucchiarelli, A.; Fioretti, S.; Di Nardo, F. A Deep Learning Approach to EMG-Based Classification of Gait Phases during Level Ground Walking. *Electronics* **2019**, *8*, 894. [CrossRef]
17. Nazmi, N.; Abdul Rahman, M.A.; Yamamoto, S.; Ahmad, S.A. Walking gait event detection based on electromyography signals using artificial neural network. *Biomed. Signal Process.* **2019**, *47*, 334–343. [CrossRef]
18. Bao, T.; Zaidi, A.; Xie, S.; Zhang, Z. Surface-EMG based Wrist Kinematics Estimation using Convolutional Neural Network. In Proceedings of the 16th International Conference on Wearable and Implantable Body Sensor Networks (BSN), Chicago, IL, USA, 19–22 May 2019; pp. 1–4. [CrossRef]
19. Xiao, F.; Wang, Y.; Gao, Y.; Zhu, Y.; Zhao, J. Continuous estimation of joint angle from electromyography using multiple time-delayed features and random forests. *Biomed. Signal Process.* **2018**, *39*, 303–311. [CrossRef]
20. Lei, Z. An upper limb movement estimation from electromyography by using BP neural network. *Biomed. Signal Process.* **2019**, *49*, 434–439. [CrossRef]
21. Huang, Y.; He, Z.; Liu, Y.; Yang, R.; Zhang, X.; Cheng, G.; Yi, J.; Ferreira, J.P.; Liu, T. Real-Time Intended Knee Joint Motion Prediction by Deep-Recurrent Neural Networks. *IEEE Sens. J.* **2019**, *19*, 11503–11509. [CrossRef]
22. Hermens, H.J.; Freriks, B. Welcome to SENIAM. Available online: <http://www.seniam.org/> (accessed on 5 June 2019).

23. Konrad, P. *The ABC of EMG: A Practical Introduction to Kinesiological Electromyography*; Noraxon USA Inc.: Scottsdale, AZ, USA, 2006; ISBN 0-9771622-1-4.
24. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
25. Liaw, A.; Wiener, M. Classification and Regression by RandomForest. *R News* **2002**, *2*, 18–22.
26. Breiman, L.; Cutler, A. Random Forests. Available online: https://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm (accessed on 21 September 2019).
27. Pearson, K. LIII. On lines and planes of closest fit to systems of points in space. *Lond. Edinburgh Dublin Philos. Mag. J. Sci.* **1901**, *2*, 559–572. [[CrossRef](#)]
28. Hotelling, H. Analysis of a complex of statistical variables into principal components. *J. Educ. Psychol.* **1933**, *24*, 417–441. [[CrossRef](#)]
29. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2016; Volume 2, pp. 45–46.
30. Morbidoni, C.; Principi, L.; Mascia, G.; Strazza, A.; Verdini, F.; Cucchiarelli, A.; Di Nardo, F. Gait Phase Classification from Surface EMG Signals Using Neural Networks. *IFMBE Proc. Springer Cham.* **2020**, *76*, 75–82. [[CrossRef](#)]

© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).