High dimensional feature data reduction of multichannel sEMG for gesture recognition based on double phases PSO

Jing Jie¹ · Kairui Liu¹ · Hui Zheng¹ · Beibei Wang¹ · Rui Dai¹

Received: 14 August 2020 / Accepted: 3 November 2020 / Published online: 30 March 2021
© The Author(s) 2021

Abstract

Surface electromyography (sEMG) is a kind of valuable bioelectric signal and very potential in the field of human–machine interaction. Ideal interactions require sEMG based patterns recognition not only with high accuracy but also with good rapidity. However, too much real-time feature-related computation will greatly slow down the interaction, especially for multichannel sEMG. To decrease the feature-related time consumption, the paper formulates the feature reduction as an optimization problem, and develops a double-phases particle swarm optimization (PSO) with hybrid coding to solve the problem. In the research, the initial feature data set with 31 kinds of feature is built firstly based on eight subjects' 16 channels forearm sEMG signals, then PSO is introduced to conduct the feature reduction of 31×16 dimensions through the feature and channel optimization in double phases. During the optimization, two improved k -nearest neighbor (KNN) methods such as weighted representation based KNN (WRKNN) and weighted local mean representation based KNN (WLMRKNN) are introduced to classify the gestures, and the classification accuracy is used to evaluate the particles of PSO. Experimental results and comparison analysis show that PSO based feature reduction methods outperform genetic algorithm (GA), ant colony optimization (ACO) and principal component analysis (PCA) based feature reduction methods. With the optimized feature data subset by PSO, WRKNN and WLMRKNN are superior to KNN, quadratic discriminant analysis (QDA), and naive bayes (NB) greatly. The proposed method can be applied in the pattern recognition of high dimensional sEMG with multichannel or high-density channels for the purpose of rapidity and without a decline of accuracy in real-time control. Further, it can be used to reduce the economic cost of the personalized customization equipment through the optimal channels for any subjects in the future.

Keywords Surface electromyography · Particle swarm optimization · Feature reduction · Gesture classification · Pattern recognition

Introduction

Surface electromyography (sEMG) is a kind of weak bio-electrical signal, which contains rich information to reflect human's neuron-muscular activity and moving intentions. The sEMG signals can be obtained in a simple and non-invasive way by placing electrodes on the skin surface of the experimental subjects, which is low-cost and does not harm the human body [1]. Therefore, the sEMG signals are more

and more popular in many fields such as clinical diagnosis, rehabilitation, human–computer interaction, etc. Besides, it is also widely applied in the commercial wearable devices such as prosthetic hand or limb [2] and provides a natural human–machine interaction method for users. However, the control accuracy, robustness and rapidity of the sEMG based pattern recognition directly influence the user's comforting satisfaction in the real-time interaction. To improve the real-time control performance of sEMG based human–machine interaction, feature extraction and classification methods should be designed in details during the pattern recognition of sEMG [3].

Feature extraction can effectively extract hidden information from surface EMG signals and reduce the impact of noise, so researchers have been studying the features of surface EMG signals for decades. Numerous features have

✉ Hui Zheng
111003@zust.edu.cn

Jing Jie
jingjie@zust.edu.cn

¹ Zhejiang University of Science and Technology,
Hangzhou 310023, China

been proposed in recent years. Graupe and Cine introduced the autoregressive (AR) model to classify EMG signal [4], Meek and Fetherston used signal-to-noise ratio to measure the quality of EMG signal [5]. Zardoshti et al. evaluated a variety of EMG features including integral of average value, the variance, the number of zero crossing, Willison amplitude, v-order, log detector, histogram and autoregressive model parameters [6]. The approximate entropy (ApEn), introduced by Pincus, has proved to be an efficient measure of the regularity of a time series, especially in the field of physiology and medicine [7, 8]. Later, Richman and Moorman developed ApEn into Sample entropy (SampEn) to reduce the bias induced by self-matching [9]. SampEn and ApEn can be used for sEMG pattern recognition, but they usually consume a large amount of computation which significantly reduce the speed of feature recognition in real-time control. Recently, several fractal analysis methods have been proposed as sEMG features. For example, Arjunan introduced a novel feature, Maximum Fractal Length (MFL), to measure the strength of contraction of the associated muscle, after that he presented another new fractal method named Higuchi's fractal dimension (HFD), which has shown better performance than other fractal methods in the study of [10–12].

To ensure the performance of sEMG based pattern recognition, especial in real-time control, feature reduction or feature selection becomes an important problem in the feature extraction process. Principal component analysis (PCA) is commonly used to deal with feature reduction problem [13, 14], but it performs weakly when solving high dimensional feature reduction problems. Some researchers try to find out some better feature combination to describe the sEMG signal. Phinyomark et al. summarized fifty features including time domain, frequency domain and fractal method, and compared the possible combinations of 2, 3 and 4 out of 50 features to find the best feature combination [15, 16]. Bai manually compared several time-domain feature combinations and finally choose MAV, ZC and RMS as the best feature combination [17]. However, those fixed feature combinations are less robust and cannot be suitable for subjects with different physiological characteristics simultaneously.

Classifiers also have a significant impact on the pattern recognition accuracy. Numerous researches had explored the appropriate classifiers for sEMG signal pattern recognition. The popular classification algorithms, including *k*-nearest neighbor (KNN), artificial neural network (ANN), linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), support vector machine (SVM) are introduced as the classifiers for sEMG recognition [18, 19]. Phinyomark adopted LDA to classify sEMG based on the elaborated chosen feature combination of 2, 3 and 4 out of 50 features [16]. Kim observed the classification accuracy of KNN, QDA and LDA based on thirty features

of two-channel sEMG, and found out that KNN performs better than QDA and LDA [20]. Pan [21] introduced two improved KNN, named as the weighted representation based *k*-nearest neighbor (WRKNN) and weighted local mean representation based KNN (WLMRKNN) [22], to classify the sEMG, and the results showed that WRKNN and WLMRKNN are superior to KNN and ANN. In fact, it is difficult to judge which feature or which feature combination are superior or inferior for one classifier because of the physical difference between people.

Particle swarm optimization (PSO) is a random search algorithm based on group collaboration, firstly proposed by Kennedy and Eberhart [23]. As a robust swarm intelligence technology, PSO has been widely used to solve complex optimization problems in real engineering area [24–28], it also has been introduced into machine learning to solve pattern recognition problems. Huang [29] proposed a PSO-SVM model to improve the performance of support vector machine (SVM) classifier and select the discriminating input features in the field of data mining. To find an optimal feature set, Khushaba [30] proposed a feature selection method based on modifying the binary PSO and mutual information (BPSOMI). All the researches show PSO based method can be applied to feature reduction.

To decrease the feature-related time consumption, the paper describes the feature reduction as an optimization problem, and develops a double-phases PSO with hybrid coding to solve the problem. PSO is introduced to sEMG pattern recognition procedure and to conduct the feature reduction through the feature and channel optimization in double phases. During the optimization, WRKNN or WLMRKNN is introduced to classify the gestures, and the classification accuracy is used to evaluate the particles of PSO.

The remainder of the paper is organized as follows. Section 2 describes the detailed process about the pattern recognition of sEMG, including signal collection, signal preprocessing, feature extraction and classifier. Section 3 describes the formulation of the optimization problem and the methodology of double-phases PSO based feature optimization and channel optimization. Section 4 shows the experiment results and analysis. Finally, Sect. 5 provides the conclusion and outlook.

Signal acquisition and processing

This section introduces the process of sEMG pattern recognition which includes the signal collection, signal preprocessing, features used for feature extraction, and common classifiers.

Signal collection

In this research, a kind of sEMG signal collecting device with an electrode sleeve named ELONXI EMG 100-Ch-Y-RA is used to collect raw sEMG signal. The 18 dry electrodes on the electrode sleeve are arranged in a novel bipolar Montage distribution [1]. Each electrode is modified from the standard disposable Ag/AgCl ECG electrodes. 16 of the 18 dry electrodes are used for data acquisition, and the other two are electrical reference electrode and bias electrode. The distribution of the myoelectric electrodes and its wearing method is shown in Fig. 1.

Eight human subjects, seven males and one female, aged 21–29 years, participated in the study. They are all healthy and volunteered for the study. Each subject was required to collect sEMG data for three trials. For every trial, each subject was asked to make five gestures, including OK, Victory, Eight, Orchid Fingers, and Thumb Up, which are shown in the following figures (Fig. 2).

The data collection for each gesture lasted 20 s. The first 8 s is the rest time, and the volunteers' forearms and fingers were placed in a natural state of relaxation. The next 12 s is the time for gesture signal collection. At this time, subjects should make corresponding gestures and keep them until the end of gesture data collection. For every 12 s of gesture signals, to ensure the stability of

gesture data, the middle 10 s are selected as samples and then marked.

Signal preprocessing

After sEMG signal is collected, it will go through the pre-processing of filter, which can validly reduce the noise of the raw sEMG signal. Since the most useful information of the sEMG is between the frequency range of 20–500 Hz, the sampling frequency is set to 1000 Hz according to the Shannon Theory, thus the signal below 500 Hz can be collected without distortion. The raw sEMG signals are filtered with a 20 Hz Butter-worth high pass filter and a notch filter of 50 Hz. This filtering method can ensure that the acquired signal contains most of the muscle information under the premise of effective noise reduction, as shown in Figs. 3 and 4. Obviously, the filtered signal reduces the low-frequency noise existing in the entire signal collecting process.

Feature extraction

Feature extraction is an important process during pattern recognition of sEMG. Referring to the research of Fang [31], the sEMG signals were segmented with a 300 ms window and 100 ms window shift for feature extraction. A total of 31 features are collected to describe gestures,

Fig. 1 Distribution of the myoelectric electrodes and wearing method of myoelectric sleeve

Fig. 2 Gestures collection

Fig. 3 Power spectrum of single channel in a subject's signal before filtering

Fig. 4 Power spectrum of single channel in a subject's signal after filtering

including 27 time-domain features and 4 frequency-domain features. The detailed information on those features is shown below (Table 1):

Integrated EMG Integrated EMG (IEMG) was presented by Merletti [32]. It is expressed as an integral value of the absolute value of the sEMG signal and can be used to detect muscle activity. It is widely used in the field of non-pattern recognition and clinical medicine. IEMG is usually used together with the sliding window method.

Mean Absolute Value Mean absolute value (MAV) feature is one of the most popular EMG features. Like the IEMG feature, it is often used as an indicator of the disease, especially in the detection of the surface EMG signal for the prosthetic limb control [6, 33]. The MAV feature is the mean value of the absolute amplitude of the signal processed by the time window method.

Modified MAV1 and MAV2 Modified mean absolute value type1 (MAV1) is an extension of the MAV feature and the weight coefficient w_i is added to improve the robustness of the MAV feature [6]. Modified mean absolute value type2 (MAV2) is also an extension of the MAV feature, and it is similar to the MAV1 feature, but the difference is that the weight value w_i of the window in the formula is assigned by a continuous function. So MAV2 is smoother than MAV1 [34].

Simple Square Integral Simple square integral (SSI) or integral square represents the energy of the EMG signal. It is the sum of the squares of the amplitudes of the EMG signals. In general, this feature is defined as the energy index of the EMG signal [35].

Variance of EMG Variance of EMG (VAR) reflects the concentration and dispersion of the signal data value. It also reflects the index of signal energy [6], which is generally defined as the mean value of the sum of squares of the EMG signal.

Absolute Value of the 3rd, 4th, and 5th Temporal Moment The 3rd, 4th, and 5th temporal moment (TM3, TM4, and TM5) feature were proposed by Saridis and Gootee in 1982 for the control of prostheses [36].

Root Mean Square Root mean square (RMS) is a widely used EMG feature [20], which is related to the contraction force of the muscles and the state of muscle fatigue.

V-Order V-Order is also a feature that has a certain relationship with muscle contraction force. According to the study [6], the optimal value of the variable v was set to 2, so that the v-order feature is the same as the RMS feature.

Log Detector Log detector (LOG) feature is generally used to estimate the contractility of muscle.

Waveform Length The waveform length (WL) feature, also known as waveform features, describes the complexity of EMG signals [33]. The WL feature is the cumulative length of the entire waveform, and it usually used together with the sliding window method.

Difference Absolute Standard Deviation Value The difference absolute standard deviation value. (DASDV) feature looks like a combination of WL features and RMS features, which can be seen as the standard deviation of the waveform length [20].

Zero Crossing Zero crossing (ZC) is defined as the number of times the signal passes through zero, which can reflect the fluctuation degree of point data and is an important feature of signal recognition. ZC is also a method for obtaining frequency information of myoelectric signals from the time-domain [37]. By setting a threshold, low voltage fluctuations can be avoided and noise eliminated.

Myopulse Percentage Rate Myopulse percentage rate (MYOP) is the average of a series of myopulse outputs, and the myopulse output is 1 if the myoelectric signal is greater than a pre-defined threshold [37].

Willison Amplitude Willison amplitude (WAMP), like the ZC feature, can be used to obtain frequency information of the myoelectric signal from the time domain [37]. It is determined by the difference between the amplitudes of adjacent EMG signals and a predetermined threshold. It is related to the firing of motor unit action potentials (MUAP) and muscle contraction force [35].

Slope Sign Change Slope sign change (SSC) is similar to the ZC, MYOP and WAMP features and can also be used to obtain frequency domain information from the time domain of the signal. The threshold is preferably 50–100 mV [34]. The value of threshold needs to be determined based on the actual equipment and noise.

Auto-Regressive Coefficients Previous studies have shown that the changes in muscle state will cause changes in AR coefficients [38], which describe each segment of the signal as a linear combination of signal before the p th segment and noise [39]. In its equation, a_p is AR coefficient, and P is the order of AR model. Several researches showed that a model order of 4 is adequate for AR time series modeling of sEMG signal.

Cepstral Coefficients Cepstral coefficients (CC) is defined as the inverse Fourier transform of the logarithm of the magnitude of the signal power spectrum [40].

Mean Frequency The mean frequency (MNF) of a spectrum is calculated as the sum of the product of the spectrogram intensity (in dB) and the frequency, divided by the total sum of spectrogram intensity [41]. In its equation, M is the number of frequency bins in the spectrum, f_j is the frequency of spectrum at bin j of M , and P_j is Intensity (dB scale) of the spectrum at bin j of M .

Peak Frequency Peak frequency (PKF) is the maximum frequency within a window of the EMG timing signal [42].

Mean Power Mean power (MNP) is the average power of the sEMG energy spectrum [39].

Total Power The total power (TTP) is the sum of all the frequencies in the power spectrum.

Maximum Fractal Length Maximum fractal length (MFL) is related to the strength of contraction of the associated muscle [10].

Classifiers

In general, the most popular classifiers for sEMG pattern recognition are LDA, QDA, KNN, SVM, ANN, NB, etc. LDA, QDA and KNN are characterized by fast computing speed, but their classification accuracy is not well enough to reach the application level; SVM, ANN, NB can achieve higher classification accuracy, but they require huge computational costs. To be applied in clinical or industrial environments, classification algorithms for sEMG need to have some necessary characteristics, such as high classification accuracy, strong robustness and good rapidity. WRKNN and

WLMRKNN are firstly proposed by Gou [22] for the pattern recognition of UCI and UCR data sets as well as face databases. Comparing to KNN, these two algorithms increase the weight of nearest neighbors. In 2019, we introduced WRKNN and WLMRKNN for sEMG pattern recognition, the results showed their superiority. So, these two modified KNN are involved in the following experiments.

Weighted representation-based K-nearest neighbor algorithm

WRKNN is a weighted extension of the KNN. The main process of WRKNN is as follows. Firstly, choose k nearest neighbors of test sample y from each class j based on Euclidean distance as follows:

$$d(y, x_i^j) = \sqrt{(y - x_i^j)^T (y - x_i^j)} \quad (1)$$

where x_i^j indicates the i th training sample from class j and denotes the k -nearest neighbor as $X_{kN}^j = [x_{1N}^j, x_{2N}^j, \dots, x_{kN}^j]$. Secondly, represent the test sample y as the linear combination of categorical k nearest neighbors which can be defined as:

$$y = x_{1N}^j \eta_1^j + x_{2N}^j \eta_2^j + \dots + x_{kN}^j \eta_k^j = X_{kN}^j \eta^j \quad (2)$$

where η_i^j indicates the i th coefficient of i th nearest neighbor from class j . Thirdly, solve the optimal representation coefficient η^{j*} which can be defined as:

$$\eta^{j*} = \underset{\eta^j}{\operatorname{argmin}} \left\{ \|y - X_{kN}^j \eta^j\|_2^2 + \lambda \|T^j \eta^j\|_2^2 \right\} \quad (3)$$

$$= ((X_{kN}^j)^T X_{kN}^j + \lambda (T^j)^T T^j)^{-1} (X_{kN}^j)^T y$$

where λ is the regularization coefficient, T^j is the distance matrix between the test sample and each nearest neighbor, W^j is defined as:

$$T^j = \begin{Bmatrix} \|y - x_{1N}^j\|_2 & & 0 \\ & \ddots & \\ 0 & & \|y - x_{kN}^j\|_2 \end{Bmatrix} \quad (4)$$

Then calculate the categorical representation-based distance between the test sample and k -local mean vectors in class j as:

$$r_k^j(y) = \|y - X_{kN}^j \eta^{j*}\|_2^2 \quad (5)$$

Finally, classify the test sample y into the class with the minimum categorical representation-based distance among all the classes, the definition can be expressed as:

$$l_y = \operatorname{argmin}_{l_j} (r_k^j(y)), j = 1, 2, \dots, C \quad (6)$$

Weighted local mean representation-based K -nearest neighbor

WLMRKN can be regarded as an improved one of WRKNN. Different from WRKNN, WLMRKN uses k local mean vectors to represent the test sample y . Supposed there are k nearest neighbors of test sample y as the following, which are chosen from class j based on Euclidean distance:

$$X_{kN}^j = [x_{1N}^j, x_{2N}^j, \dots, x_{kN}^j] \quad (7)$$

Here x_i^j indicates the i th training sample from class j as a nearest neighbor. Then the mean vectors of k nearest neighbors from each class can be expressed as:

$$\bar{x}_{iN}^j = \frac{1}{i} \sum_{h=1}^i x_{hN}^j (i = 1, 2, \dots, k) \quad (8)$$

where $\bar{x}_{iN}^j$ indicates the i th local mean vector from class j , and represent the test sample y as the linear combination of categorical k local mean vectors which can be defined as:

$$y = \bar{x}_{1N}^j s_1^j + \bar{x}_{2N}^j s_2^j + \dots + \bar{x}_{kN}^j s_k^j = \bar{X}_{kN}^j s^j \quad (9)$$

where s_i^j indicates the i th coefficient of i th local mean vector from class j . The optimal coefficient s^{j*} can be computed by the following:

$$s^{j*} = \operatorname{argmin}_{s^j} \left\{ \|y - \bar{X}_{kN}^j s^j\|_2^2 + \gamma \|W^j s^j\|_2^2 \right\} \\ = ((\bar{X}_{kN}^j)^T \bar{X}_{kN}^j + \gamma (W^j)^T W^j)^{-1} (\bar{X}_{kN}^j)^T y \quad (10)$$

Here γ is the regularization coefficient, W^j is the distance matrix between the test sample and each local mean vector which is defined as:

$$W^j = \begin{Bmatrix} \|y - \bar{x}_{1N}^j\|_2 & & 0 \\ & \ddots & \\ 0 & & \|y - \bar{x}_{kN}^j\|_2 \end{Bmatrix} \quad (11)$$

So, the categorical representation-based distance between the test sample and k -local mean vectors in class j is computed as:

$$r_k^j(y) = \|y - \bar{X}_{kN}^j s^{j*}\|_2^2 \quad (12)$$

Finally, classify the test sample y into the class with the minimum categorical representation-based distance among all the classes, the definition can be expressed as:

$$l_y = \operatorname{argmin}_{l_j} (r_k^j(y)), j = 1, 2, \dots, C \quad (13)$$

Feature data reduction based on double phases PSO

Theoretically, the more characteristic information there is, the higher the classification accuracy will be. However, high dimensional features set of sEMG will consume great computation, especially for sEMG with high-density channels. As it's well known, the high cost will directly influence the application of sEMG in real-time man-machine interaction. So, it's necessary to conduct the feature reduction before pattern classification for high dimensional features set with multiple channels, shown in Fig. 5 sEMG pattern recognition flowchart. To achieve better results, we introduce PSO to reduce the feature dimensions in double phases.

Double-phases PSO is divided into two parts. In the first phase, PSO is used to obtain an optimal feature subset in the

Fig. 5 sEMG pattern recognition flowchart

whole channel data set. In the second phase, PSO is adopted to go on the channel optimization with the optimal feature subset. The methods will be described in detail as follows, including problem description, PSO, feature and channel optimization.

Problem description

Suppose there are M hand gestures need to be recognized, so the data set Q for those gestures can be recorded as the following:

$$Q = \{(D^1, y_1), (D^2, y_2), \dots, (D^m, y_m), \dots, (D^M, y_M)\} \quad (14)$$

Here, D^m stands for the feature data vector of the m th hand gesture, while y_m stands for the real label of D^m . If the sEMG signal of each hand gesture is acquired by an equipment with C channels over time, and F features are adopted to describe each gesture, D^m can be recorded as the following equation:

$$D^m = (d_{11}^m, d_{12}^m, \dots, d_{1F}^m, \dots, d_{c1}^m, d_{c2}^m, \dots, d_{cF}^m, \dots, d_{C1}^m, d_{C2}^m, \dots, d_{CF}^m, \dots, d_{CF}^m) \quad (15)$$

where d_{cf}^m means the f th feature data of the c th channel for the m th hand gesture.

Suppose that classifier Z is used to recognize M gestures for any one subject, and each gesture has been collected K samples over time in data set Q , the classification accuracy of the classifier Z can be computed as the following:

$$\text{Accu}(Z; Q) = \frac{1}{MK} \sum_{m=1}^M \sum_{k=1}^K 1(Z(D^m) = y_m) \quad (16)$$

Obviously, there are $C \times F$ dimensions of feature data for each gesture in the data set Q . The larger the C and F , the higher the dimension of feature data, and the higher the computational cost of gesture recognition. To reduce the computing cost in real-time tasks, we use PSO to find the optimal feature subset from the data set Q to provide the optimal accuracy for gesture recognition.

Suppose the optimal feature data subset is recorded as Q_{OS} :

$$Q_{OS} = \{(D_o^1, y_1), (D_o^2, y_2), \dots, (D_o^m, y_m), \dots, (D_o^M, y_M)\} \quad (17)$$

Here, D_o^m stands for the optimal feature data of the m th gesture chosen from the data set Q by PSO in the Q_{OS} . According to the Eq. (16), the feature optimization problem can be described as follows:

$$\text{Max}[\text{Accu}(Z; Q_{OS})] = \text{Max} \left[\frac{1}{MK} \sum_{m=1}^M \sum_{k=1}^K 1(Z(D_o^m) = y_m) \right] \quad (18)$$

Particle swarm optimization

PSO originated from the study of the behavior of preying on birds, its basic idea is that the whole swarm of birds will tend to follow the bird which found the best path to food. To search an optimum, PSO defines a swarm of particles to represent the potential solutions to an optimization problem. Each particle begins with an initial position randomly and flies through the D -dimensional solution space. The flying behavior of each particle can be described by its velocity and position in standard particle swarm optimization (SPSO) [43] as the following.

$$v_{id}(t+1) = wv_{id}(t) + c_1r_1(p_{id}(t) - x_{id}(t)) + c_2r_2(p_{gd}(t) - x_{id}(t)) \quad (19)$$

$$x_{id}(t+1) = x_{id}(t) + v_{id}(t+1) \quad (20)$$

where $V_i = (v_{i1}, v_{i2}, \dots, v_{id}, \dots, v_{iD})$ is the velocity vector of the i th particle; $X_i = (x_{i1}, x_{i2}, \dots, x_{id}, \dots, x_{iD})$ is the position vector of the i th particle; $P_i = (p_{i1}, p_{i2}, \dots, p_{id}, \dots, p_{iD})$ is the best

position found by the i th particle; $P_g = (p_{g1}, p_{g2}, \dots, p_{gd}, \dots, p_{gD})$ is the global best position found by the whole swarm; c_1, c_2 are two learning factors, usually $c_1 = c_2 = 2$; r_1, r_2 are random numbers between (0, 1); w is the inertia weight to control the velocity.

After some comparative analysis, Shi and Eberhart found out that the inertia weight has a great influence on the optimization performance [43]. To get suitable search step, they developed a modified PSO, in which a strategy of linearly decreased inertial weight over time (LDIW) is introduced to keep the trade-off between the exploration and the exploitation, the modified equation of inertial weight can be found as the following:

$$w = w_1 - (w_1 - w_T) \cdot \frac{t}{T_{\max}} \quad (21)$$

where w_1 and w_T stand for the initial and the terminal inertial weights respectively, while $T_{\max}$ is the terminal iteration.

Double phases PSO for feature and channel optimization

In our work, we choose 31 features to describe the sEMG of 16 channels, that is, there are $31 \times 16 = 496$ dimension feature data in each gesture feature data set Q . It's no doubt that such a data set is a super-high dimensional problem. To reduce the data computation cost and make clear how the features and the

channels influence the recognition accuracy for each subject, we develop double phases PSO for feature data optimization.

Feature optimization

In the first phase of optimization, PSO is anticipated to find the best feature combination among 31 features to reduce the feature data. Considering the problem and the optimization algorithm as a whole, a hybrid coding method with real and binary coding is adopted to describe the particle information in the search and the decision space. For example, the position vector in the search space of arbitrary particle i is encoded in the real string:

$$X_i^R = (x_{i1}^R, x_{i2}^R, \dots, x_{if}^R, \dots, x_{iF}^R) \quad (22)$$

At the same time, its velocity also is encoded in real value in the search space.

$$V_i^R = (v_{i1}^R, v_{i2}^R, \dots, v_{if}^R, \dots, v_{iF}^R) \quad (23)$$

where f stands for the feature index, $f \in [1, F]$, and $F = 31$. $x_{if}^R \in [0, 1]$, $v_{if}^R \in [-V_{\max}, V_{\max}]$ and $V_{\max}$ is set as 1.0.

In the decision space of the feature optimization problem, the position of the particle can switch to a binary code:

$$X_i^B = (x_{i1}^B, x_{i2}^B, \dots, x_{if}^B, \dots, x_{iF}^B) \quad (24)$$

where $x_{if}^B \in \{0, 1\}$. If the fraction of x_{if}^R is bigger than 0.5, $x_{if}^B = 1$, which indicates the feature f is selected into the feature subset during the classification process; else $x_{if}^B = 0$, which means the feature f is not selected into the feature subset. Binary coding-based particle i makes it easy to construct feature subsets.

To clarify the hybrid coding method, a coding example for particle i is presented in the following table:

According to the decision vector X_i^B in Table 2, the value of 1th, 3th and 7th sub-vectors are equal to 1, what means there are the 1th, 3th and 7th features are chosen to construct the feature subset {IEMG, MAV1, TM3}. Based on the feature subset, the corresponding data subset for particle i can be obtained:

$$D^{im} = (d_{11}^{im}, d_{13}^{im}, d_{17}^{im}, \dots, d_{c1}^{im}, d_{c3}^{im}, d_{c7}^{im}, \dots, d_{C1}^{im}, d_{C3}^{im}, d_{C7}^{im}) \quad (25)$$

Obviously, the dimension of the feature data has been reduced greatly through feature optimization. The reduced feature data subsets are directly used for pattern recognition, and the recognition accuracy is calculated as the fitness of particle i . After sEMG signal is preprocessed, the above hybrid coding method can be used for feature optimization by PSO. The main procedure is as the following:

Channel optimization

As it's well known, each gesture is determined by the movement of multiple muscle blocks, and each involved muscle block provides a different contribution to the output of sEMG signal. Therefore, it is necessary to identify the most significant muscle blocks involved in each gesture movement, which mainly determine the corresponding EMG signal strength of the gesture. During the acquisition of sEMG signal, each channel index just indicates the muscle block motion signal corresponding to the acquisition point. Therefore, after the feature optimization in the first phase, it is expected to use PSO for channel optimization in the second phase, further reducing the dimension of the data set.

As same as the feature optimization, the hybrid coding method with real and binary coding is also adopted to describe the particle information in the search and the decision space during the channel optimization.

Considering the coding of the arbitrary particle j , its position and velocity vector in the search space is coded in real strings:

$$X_j^R = (x_{j1}^R, x_{j2}^R, \dots, x_{jc}^R, \dots, x_{jC}^R) \quad (26)$$

$$V_j^R = (v_{j1}^R, v_{j2}^R, \dots, v_{jc}^R, \dots, v_{jC}^R) \quad (27)$$

where $C = 16$. c stands for the channel index and $c \in [1, C]$. $x_{jc}^R \in [0, 1]$, $v_{jc}^R \in [-V_{\max}, V_{\max}]$ and $V_{\max}$ is set as 1.0.

In the decision space of the channel optimization problem, the position of the particle can be switched to a binary code:

$$X_j^B = (x_{j1}^B, x_{j2}^B, \dots, x_{jc}^B, \dots, x_{jC}^B) \quad (28)$$

where $x_{jc}^B \in \{0, 1\}$. If the fraction of x_{jc}^R is bigger than 0.5, $x_{jc}^B = 1$, which indicates the channel c is selected into the channel subset during the classification process; else $x_{jc}^B = 0$, which means the channel c is not selected into the channel subset. So, a channel subset is easy to be decided based on the binary coding of the particle j .

Suppose the value of 1th, 2th, 5th, 8th sub-vectors in X_j^B are equal to 1, which means that the 1th, 2th, 5th and 8th channels provide a significant contribution to the sEMG signal of the gestures, and can be chosen to construct the channel subset. The data subset in the Eq. (25) will be further reduced as the following:

$$D^{im} = (d_{11}^{im}, d_{13}^{im}, d_{17}^{im}, d_{21}^{im}, d_{23}^{im}, d_{27}^{im}, d_{51}^{im}, d_{53}^{im}, d_{57}^{im}, d_{81}^{im}, d_{83}^{im}, d_{87}^{im}) \quad (29)$$

Table 1 Index and information of features

No	Domain	Feature	Equation
1	Time-domain	IEMG (integrated EMG)	$\text{IEMG} = \sum_{i=1}^N x_i $
2		MAV (mean absolute value)	$\text{MAV} = \frac{1}{N} \sum_{i=1}^N x_i $
3		MAV1 (modified mean absolute value type1)	$\text{MAV1} = \frac{1}{N} \sum_{i=1}^N w_i x_i \quad w_i = \begin{cases} 1, & \text{if } 0.25N \leq i \leq 0.75N \\ 0.5, & \text{otherwise} \end{cases}$
4		MAV2 (modified mean absolute value type 2)	$\text{MAV2} = \frac{1}{N} \sum_{i=1}^N w_i x_i \quad w_i = \begin{cases} \frac{1}{4(i-N)N}, & \text{if } 0.25N \leq i \leq 0.75N \\ \frac{4i}{N}, & \text{else if } i < 0.25N \\ 0, & \text{otherwise} \end{cases}$
5		SSI (simple square integral)	$\text{SSI} = \sum_{i=1}^N x_i^2$
6		VAR (variance of EMG)	$\text{VAR} = \frac{1}{N-1} \sum_{i=1}^N x_i^2$
7		TM3, TM4, TM5 (3rd, 4th, and 5th temporal moment)	$\text{TM3} = \frac{1}{N} \sum_{i=1}^N x_i^3$
8			$\text{TM4} = \frac{1}{N} \sum_{i=1}^N x_i^4$
9			$\text{TM5} = \frac{1}{N} \sum_{i=1}^N x_i^5$
10		RMS (root mean square)	$\text{RMS} = \sqrt{\frac{1}{N} \sum_{i=1}^N x_i^2}$
11		V-Order	$\text{V - Order} = \left(\frac{1}{N} \sum_{i=1}^N x_i^v \right)^{\frac{1}{v}}$
12		LOG (log detector)	$\text{LOG} = e^{\frac{1}{N} \sum_{i=1}^N \log(x_i)}$
13		WL (waveform length)	$\text{WL} = \sum_{i=1}^{N-1} x_{i+1} - x_i $
14		DASDV (difference absolute standard deviation value)	$\text{DASDV} = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N-1} (x_{i+1} - x_i)^2}$
15		ZC (zero crossing)	$\text{ZC} = \sum_{i=1}^N [\text{sgn}(x_i \times x_{i+1}) \cap x_i - x_{i+1} \geq \text{threshold}]$ $\text{sgn}(x) = \begin{cases} 1, & \text{if } x < 0 \\ 0, & \text{otherwise} \end{cases}$
16		MYOP (myopulse percentage rate)	$\text{MYOP} = \frac{1}{N} \sum_{i=1}^N [f(x_i)]$ $f(x) = \begin{cases} 1, & \text{if } x \geq \text{threshold} \\ 0, & \text{otherwise} \end{cases}$
17		WAMP (Willson amplitude)	$\text{WAMP} = \sum_{i=1}^{N-1} [f(x_n - x_{n+1})]$ $f(x) = \begin{cases} 1, & \text{if } x \geq \text{threshold} \\ 0, & \text{otherwise} \end{cases}$
18		SSC (slope sign change)	$\text{SSC} = \sum_{i=2}^{N-1} [f((x_i - x_{i-1}) \times (x_i - x_{i+1}))]$ $f(x) = \begin{cases} 1, & \text{if } x \geq \text{threshold} \\ 0, & \text{otherwise} \end{cases}$
19		AR (auto-regressive coefficients)	$x_i = \sum_{p=1}^P a_p x_{i-p} + w_i$

Table 1 (continued)

No	Domain	Feature	Equation
20			
21			
22			
23		CC (cepstral coefficients)	$x_i = \sum_{p=1}^P a_p x_{i-p} + w_i$ $c_1 = -a_1$ $c_p = -a_p - \sum_{l=1}^{p-1} \left(1 - \frac{l}{p}\right) a_p c_{p-l}$
24			
25			
26			
27		MFL (maximum fractal length)	$MFL = \log_{10} \sqrt{\left(\sum_{n=1}^{N-1} (x(n+1) - x(n))^2 \right)}$
28	Frequency-domain	MNF (mean frequency)	$MNF = \frac{\sum_{j=1}^M f_j P_j}{\sum_{j=1}^M P_j}$
29		PKF (peak frequency)	$PKF = \max(P_j), \quad j = 1, \dots, M.$
30		MNP (mean power)	$MNP = \frac{\sum_{j=1}^M P_j}{M}$
31		TTP (total power)	$TTP = \sum_{j=1}^M P_j = SM0$

Table 2 Hybrid coding for particle i

Feature index	1	2	3	4	5	6	7	8	9	10	...	F
Search vector X_i^R	0.65	0.13	0.72	0.25	0.1	0.3	0.8	0.21	0.31	0.2	...	0.3
Search vector V_i^R	0.25	0.31	0.11	0.22	0.3	0.05	0.15	0.22	0.13	0.25	...	0.15
Decision vector X_i^B	1	0	1	0	0	0	1	0	0	0	...	0

The feature index is same as Table 1

Based on the reduced data subset, the gesture patterns are recognized and the recognition accuracy will be computed as the fitness to evaluate the particle j . The main procedure of channel optimization by PSO is same as the procedure of feature optimization in the above section.

Algorithm procedure

The algorithm procedure is described below.

Double phases PSO Based Feature Data Reduction

- 1: Acquire and preprocessing of sEMG Signal with 16 channels.
- 2: Extract features data for the above 31 features with moving window method to construct the feature data space, and normalize those feature data.
- 3: Conduct first phase of feature optimization by PSO and get the optimal feature subset to reduce the feature data.
- 4: Conduct second phase of channel optimization by PSO and get the optimal channel subset to further reduce the feature data.
- 5: Classify gesture and evaluate the relative recognition accuracy based on the optimal feature data subset.

Experiments and discussions

In the experiments, three trials of sEMG signal data are acquired under the same sleeve wearing position from eight subjects firstly. After that, the first trial of data is selected as a training data set, and one of the rest trials of data is used as a test data set for each subject. The parameters of PSO are set as: $c_1 = c_2 = 2.0$, $w_I = 0.7$, $w_T = 0.2$, and swarm size $N_s = 20$; In GA, crossover probability is 0.8, and mutation probability is 0.1; in ACO, evaporation pheromone rate is 0.8, and P_0 is 0.2. All the programs such as filtering, feature extraction, classifier and optimization are completed on the MATLAB R2018b platform.

Experimental comparison analysis of feature optimization

In the first experiment, PSO, GA, ACO and PCA were combined with five classifiers (KNN, QDA, NB, WRKNN and WLMRKN) respectively to perform feature reduction and classification of sEMG signal, thus 20 different combinations were obtained. The classification accuracy of the combined methods for each subject and the average accuracy for all subjects are shown in Tables 3, 4, 5, 6 and 7 below.

According to the average accuracy of each classifier for all subjects in the above five tables, most of the classifiers combined with PCA-based feature reduction are below 60%, and only the average classification accuracy of PCA-QDA classifier was slightly higher (62.18%). The reason for this result

is that PCA inevitably weakens and loses information when reducing large dimensions to very small ones. The average results of ACO are a little better than PCA, but it performed unstable and sometimes may be trapped into local optimum. GA based feature optimization method performed well but it's still worse than PSO except with KNN classifier in Table 3. According to the result of PSO in five tables, whether combined with any classifier, the average classification accuracy of PSO based feature reduction has been significantly improved, which means that PSO based feature reduction is very valid for the high dimensional feature reduction in the gesture recognition, and can effectively select the optimal feature combination from the large feature set with less redundant information.

In addition, the results in Tables 3, 4, 5, 6 and 7 show that different classifiers combined with different feature extraction methods have different performance. It can be seen from Tables 6 and 7, PSO-WRKNN and PSO-WLMRKN are obviously superior to other methods. Compared with PSO-KNN, the average accuracy of these two methods are improved by about 4%. This is because WRKNN and WLMRKN increase the nearest neighbor weights and reduce the influence of data from the remote time window. Therefore, these two improved KNN can be regarded as a satisfactory choice for classifiers with PSO based feature optimization to improve the classification accuracy.

To further make clear the influence of different features, the distribution statistics about the optimal features in each trial are presented in the following Figs. 6, 7 and 8.

Figures 6, 7 and 8 shows the rate at which each feature is selected in the optimal feature set. Among them,

Table 3 Average classification accuracy for each subject through Feature optimization and KNN classifier with all channels sEMG data

Training data\test data	PSO-KNN (%)	GA-KNN (%)	ACO-KNN (%)	PCA-KNN (%)
CYB1\CYB2	94.88	79.84	77.52	47.43
CYB1\CYB3	95.70	64.79	72.89	56.86
CYX1\CYX2	79.83	94.38	40.17	45.95
CYX1\CYX3	69.59	94.38	40.83	35.2
SK1\SK2	93.55	78.84	42.81	43.3
SK1\SK3	96.86	64.30	85.79	51.07
TP1\TP2	93.3	91.57	57.52	52.56
TP1\TP3	93.39	96.03	47.60	55.04
YJX1\YJX2	88.26	92.73	57.03	60.33
YJX1\YJX3	75.21	93.72	25.46	59.5
YKJ1\YKJ2	92.73	87.44	75.87	60.33
YKJ1\YKJ3	76.20	95.04	54.55	52.07
ZJN1\ZJN2	80.00	91.07	68.93	58.34
ZJN1\ZJN3	75.04	76.69	67.93	35.54
CHF1\CHF2	80.83	79.01	47.77	44.96
CHF1\ CHF3	61.65	76.03	53.88	40.16
Average accuracy	84.19	84.74	57.28	49.92

The bold symbol indicates that the average classification accuracy of this method is the highest
CYB, CYX, SK, TP, YJX, YKJ, ZJN and CHF stand for the eight subjects

Table 4 Average classification accuracy of each subject through feature optimization and QDA classifier with all channels sEMG data

Training data\test data	PSO-QDA (%)	GA-QDA (%)	ACO-QDA (%)	PCA-QDA (%)
CYB1\CYB2	92.78	72.56	47.77	81.49
CYB1\CYB3	96.29	63.14	74.88	75.04
CYX1\CYX2	84.38	87.11	32.23	49.59
CYX1\CYX3	71.74	91.24	46.45	34.71
SK1\SK2	97.24	66.45	25.62	44.46
SK1\SK3	98.02	59.34	84.46	58.51
TP1\TP2	96.12	82.98	77.03	76.03
TP1\TP3	95.37	95.21	26.94	73.88
YJX1\YJX2	92.07	91.57	77.36	64.96
YJX1\YJX3	99.01	90.74	51.57	73.55
YKJ1\YKJ2	90.08	85.62	69.59	60.00
YKJ1\YKJ3	84.46	82.98	60.66	60.6
ZJN1\ZJN2	84.05	91.41	67.93	73.72
ZJN1\ZJN3	82.98	79.17	24.13	59.50
CHF1\CHF2	87.77	78.35	23.64	60.83
CHF1\CHF3	69.92	65.79	33.39	48.10
Average accuracy	88.89	80.23	51.48	62.18

The bold symbol indicates that the average classification accuracy of this method is the highest

Table 5 Average classification accuracy of each subject through Feature optimization and NB classifier with all channels sEMG data

Training data\test data	PSO-NB (%)	GA-NB (%)	ACO-NB (%)	PCA-NB (%)
CYB1\CYB2	97.77	76.53	71.57	83.80
CYB1\CYB3	94.96	63.47	75.54	68.76
CYX1\CYX2	93.14	87.60	25.29	47.77
CYX1\CYX3	77.28	89.09	37.03	38.84
SK1\SK2	93.72	65.95	34.88	37.52
SK1\SK3	97.02	62.31	63.14	64.96
TP1\TP2	98.02	79.34	70.74	71.90
TP1\TP3	98.84	93.55	80.17	73.22
YJX1\YJX2	88.51	85.79	64.96	55.2
YJX1\YJX3	96.20	89.75	77.03	60.50
YKJ1\YKJ2	93.80	82.31	73.88	63.80
YKJ1\YKJ3	80.67	90.08	49.26	54.88
ZJN1\ZJN2	80.17	85.12	60.50	55.54
ZJN1\ZJN3	78.60	76.20	51.74	44.30
CHF1\CHF2	83.80	76.20	64.13	63.47
CHF1\CHF3	63.64	59.67	42.98	48.10
Average accuracy	88.51	78.94	58.93	58.28

The bold symbol indicates that the average classification accuracy of this method is the highest

both MNF and MFL are most likely to be chosen, more than other characteristics, indicating these two features are suitable to describe the forearm sEMG signal for more subjects in gesture recognition. It is worth mentioning that, according to the results of multiple optimal feature combinations of eight subjects, each subject's optimal feature sub-set tends to select specific features, that is, each person has a set of optimal features that are most suitable for his or her physical condition and dressing condition.

Experimental comparison analysis of channel optimization

Through the above analysis, PSO is very effective for dimension reduction of high-dimensional features. On this basis, PSO is continued to be used for channel optimization. In the experiment, we observed five classifiers combined with PSO for double phases optimization respectively, and the statistical data are shown in Table 8.

Table 6 Average classification accuracy of each subject through Feature optimization and WRKNN classifier with all channels sEMG data

Training data\test data	PSO-WRKNN (%)	GA-WRKNN (%)	ACO-WRKNN (%)	PCA-WRKNN (%)
CYB1\CYB2	96.86	83.97	25.62	55.37
CYB1\CYB3	97.85	65.12	83.64	59.83
CYX1\CYX2	79.34	97.69	47.27	52.23
CYX1\CYX3	72.73	95.37	79.17	45.95
SK1\SK2	96.53	90.41	22.98	55.04
SK1\SK3	97.19	77.03	90.41	55.04
TP1\TP2	94.21	91.90	56.86	55.87
TP1\TP3	95.87	97.36	90.58	56.36
YJX1\YJX2	90.58	97.52	71.57	62.48
YJX1\YJX3	97.19	99.01	92.40	65.29
YKJ1\YKJ2	94.88	89.09	62.31	60.50
YKJ1\YKJ3	85.12	96.03	75.37	51.57
ZJN1\ZJN2	83.64	94.38	65.95	62.48
ZJN1\ZJN3	82.15	81.32	78.51	41.49
CHF1\CHF2	87.44	80.33	29.9	46.94
CHF1\CHF3	69.09	78.02	61.65	40.17
Average accuracy	88.79	88.41	64.64	54.16

The bold symbol indicates that the average classification accuracy of this method is the highest

Table 7 Average classification accuracy of each subject through Feature optimization and WLMRKNN classifier with all channels sEMG data

Training data\test data	PSO-WLMRKNN (%)	GA-WLMRKNN (%)	ACO-WLMRKNN (%)	PCA-WLMRKNN (%)
CYB1\CYB2	98.02	83.31	77.03	57.02
CYB1\CYB3	94.47	64.13	21.49	59.34
CYX1\CYX2	92.98	97.69	76.69	47.44
CYX1\CYX3	77.36	95.54	74.71	45.79
SK1\SK2	93.99	91.07	60.66	51.24
SK1\SK3	97.85	75.70	65.79	52.89
TP1\TP2	97.61	91.41	44.63	55.04
TP1\TP3	99.01	95.21	87.93	56.03
YJX1\YJX2	88.26	97.69	90.91	61.16
YJX1\YJX3	95.87	98.51	92.40	64.13
YKJ1\YKJ2	87.77	87.93	78.35	58.84
YKJ1\YKJ3	81.49	96.36	45.29	51.40
ZJN1\ZJN2	79.59	94.22	60.66	60.66
ZJN1\ZJN3	78.18	80.00	74.55	40.00
CHF1\CHF2	83.97	78.84	27.60	43.80
CHF1\CHF3	64.63	78.35	63.47	39.00
Average accuracy	88.19	87.87	65.13	52.73

The bold symbol indicates that the average classification accuracy of this method is the highest

Table 8 shows the gesture classification accuracy after feature and channel optimization of five different classifiers. The accuracy of PSO-KNN, PSO-QDA, and PSO-NB increased by 4.16%, 3.78%, and 3.53% respectively after channel optimization, meanwhile, the accuracy PSO-WRKNN and PSO-WLMRKNN is also improved slightly by

0.70% and 0.13%, respectively. Obviously, the channel optimization doesn't cause accuracy degradation but improves it more or less. The fact is that, in most situation, the sEMG signal acquisition device acquires a series of noises due to the crashed sEMG electrodes or bad contact, and this problem can be solved through channel optimization to a certain

Fig. 6 Distribution statistics of PSO optimal feature

Fig. 7 Distribution statistics of GA optimal feature

Fig. 8 Distribution statistics of ACO optimal feature

Table 8 Accuracy of gesture recognition for each subject after feature and channel optimization

Training data\test data	PSO-KNN (%)	PSO-QDA (%)	PSO-NB (%)	PSO-WRKNN (%)	PSO-WLMRKNN (%)
CYB1\CYB2	96.86	89.09	82.98	92.78	95.37
CYB1\CYB3	97.85	89.59	87.77	96.29	96.95
CYX1\CYX2	79.34	77.85	77.19	84.38	79.84
CYX1\CYX3	72.73	72.23	71.57	71.74	70.23
SK1\SK2	96.53	89.09	91.40	97.24	96.69
SK1\SK3	97.19	91.74	91.24	98.02	97.94
TP1\TP2	94.21	93.22	94.38	96.12	96.28
TP1\TP3	95.87	89.75	90.41	95.37	96.5
YJX1\YJX2	90.58	84.79	85.12	92.07	87.69
YJX1\YJX3	97.19	90.41	83.64	99.01	98.27
YKJ1\YKJ2	94.88	96.53	86.45	90.08	93.72
YKJ1\YKJ3	85.12	81.32	86.45	84.46	85.71
ZJN1\ZJN2	83.64	76.53	78.02	84.05	81.82
ZJN1\ZJN3	82.15	79.01	70.41	82.98	83.22
CHF1\CHF2	87.44	81.16	88.26	87.77	87.03
CHF1\CHF3	69.09	53.39	69.59	69.92	71.00
Average accuracy	88.79	83.48	83.43	88.89	88.64

The bold symbol indicates that the average classification accuracy of this method is the highest

extent. In other words, channel optimization can accurately remove the unimportant channels which provided less contribution to the sEMG of one gesture, so that eliminating the signal data of some channels after channel optimization not only can improve the operation speed in real-time control, but also ensure the classification accuracy in real-time control.

To further make clear the influence of the channels, the number of the optimal channels in each trial is observed and the distribution statistics of them are illustrated in the following Fig. 9.

It can be seen from Fig. 9, the optimal number of channels in this experiment is mostly distributed between 6 and 10, among which 9 channels are the most, and the experimental results of using nine channels as the optimal channel account for 35.24%. It should be noted that in this experiment, the

maximum number of channels is limited to 10, because the more channels, the more difficult it is to reduce the channel data. All the results further reveal that there are some individual differences among different subjects for the same gesture. Therefore, PSO based channel optimization not only helps to reduce the time consumption of pattern recognition but also provides some valuable ideas for reducing the economic cost of sEMG based human–computer interaction devices.

Conclusions

Improving the rapidity of real-time control while ensuring accuracy is a key problem for sEMG based human–computer interaction. Suitable feature extraction and classification play

Fig. 9 Distribution statistics of PSO based optimal channel number

a vital role in pattern recognition. To get high classifying accuracy with low time consumption in pattern recognition, the paper presents a double-phases PSO model combined with WRKNN and WLMRKNN classifiers for multichannel sEMG signal with high dimensional features. After the preprocessing of sEMG, PSO is introduced firstly to select the optimal feature subset from the total 31 feature set, and then to select the optimal channel. During the optimization, WRKNN and WLMRKNN classifiers are used to classify the gestures based on the optimal channels and the optimal features. The experimental results show that PSO based feature reduction method outperform GA, ACO and PCA based feature reduction methods in high dimensional feature space of sEMG. The combination of PSO and WRKNN or WLMRKNN classifier further improves the classification accuracy. The proposed method requires only a small amount of feature extraction computation and can still reach a significant accuracy for multichannel sEMG. In the future, we will further study the sEMG based applications in the human–machine interaction, such as the real control of the robot, or some rehabilitation treatments based on sEMG, and seek for better pattern recognition method [44, 45].

Author contributions Conceptualization, JJ; methodology, JJ and HZ; validation, KL, BW, RD; data curation, KL; writing—original draft preparation, KL and JJ; writing, review and editing, JJ and HZ; supervision, JJ and HZ.

Funding This work was supported in part by National Natural Science Foundation of China (Grant No. 62003306) and Natural Science Foundation of Zhejiang Province (LY19F030003, LQ21F030009).

Data availability All data generated or analyzed during this study are included in this published article.

Code availability If the code is necessary, please email 111003@zust.edu.cn for it.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Fang Y, Hettiarachchi N, Zhou D et al (2016) Multi-modal sensing techniques for interfacing hand prostheses: a review. *IEEE Sens J* 15(11):6065–6076
2. Deluca C (1979) Physiology and mathematics of myoelectric signals. *IEEE Trans Biomed Eng* 26(3):313–325
3. Phinyomark A, Phukpattaranont P, Limsakul C (2012a) Feature reduction and selection for emg signal classification. *Expert Syst Appl* 39(8):7420–7431
4. Graupe J, Cine K (1975) Functional separation of EMG signals via ARMA identification method for prosthesis control purpose. *IEEE Trans Syst Man Cybern* 5:252–259
5. Meek SG, Fetherston SJ (1992) Comparison of signal-to-noise ratio of myoelectric filters for prosthesis control. *J Rehab R&D* 29(4):9–20
6. Zardoshti-Kermani M, Wheeler BC, Badie K, Hashemi RM (1995) EMG feature evaluation for movement control of upper extremity prostheses. *IEEE Trans Rehabil Eng* 3(4):324–333
7. Pincus SM (1991) Approximate entropy as a measure of system complexity. *Proc Natl Acad Sci USA* 88:2297–2301
8. Pincus SM (1995) Approximate entropy (ApEn) as a complexity measure. *Chaos* 5:110–117
9. Richman JS, Moorman JR (2000) Physiological time-series analysis using approximate entropy and sample entropy. *Am J Physiol Heart Circ Physiol* 278(6):H2039–H2049
10. Arjunan SP (2008) Fractal features of surface electromyogram: a new measure for low level muscle activation. RMIT University, Melbourne
11. Arjunan SP, Kumar DK (2010) Decoding subtle forearm flexions using fractal features of surface electromyogram from single and multiple sensors. *J Neuro Eng Rehabil* 7:53
12. Esteller R, Vachtsevanos G, Echaz J, Litt B (2001) A comparison of waveform fractal dimension algorithms. *IEEE Trans Circ Syst I Fundam Theory Appl* 48(2):177–1831
13. Xu YB (2018) Teleoperation control of robot based on EMG signal. South China University of Technology, Guangzhou ((in Chinese))
14. Zhou L, Zheng J, Ge ZQ et al (2018) Multimode process monitoring based on switching autoregressive dynamic latent variable model. *IEEE Trans Industr Electron* 65(10):8184–8194
15. Phinyomark A, Phukpattaranont P, Limsakul C (2012b) Fractal analysis features for weak and single-channel upper-limb EMG signals. *Expert Syst Appl* 39(12):11156–11163
16. Phinyomark A, Quaine F, Charbonnier S, Serviere C, Tarpin-Bernard F, Laurillau Y (2013) EMG feature evaluation for improving myoelectric pattern recognition robustness. *Expert Syst Appl* 40(12):4832–4840
17. Bai Y (2018) Research on robot control method based on EMG signal. Shenyang Ligong University, Shenyang ((in Chinese))
18. Sang HP, Seok PL (1998) EMG pattern recognition based on artificial intelligence techniques. *IEEE Trans Rehabil Eng* 6:400–405
19. Reza B, Mohammad HM (2003) Evaluation of the forearm EMG signal features for the control of a prosthetic hand. *Physiol Meas* 24:309–319
20. Kim KS, Choi HH, Moon CS et al (2011) Comparison of k-nearest neighbor, quadratic discriminant and linear discriminant analysis in classification of electromyogram signals based on the wrist-motion directions. *Curr Appl Phys* 11(3):740–745
21. Pan S, Jie J, Liu K, et al. (2019) Classification methods of sEMG through weighted representation-based k-nearest neighbor. In: International conference on intelligent robotics and applications. Springer, Cham, pp. 456–466.

22. Gou JP, Qiu WM, Zhang Yi et al (2019) Locality constrained representation-based K-nearest neighbor classification. *Knowl Based Syst* 167(3):38–52
23. Kennedy J, Eberhart R. (1995) Particle swarm optimization. In: *Proceedings of ICNN'95-international conference on neural networks*. IEEE, vol 4, pp 1942–1948.
24. Sun C, Jin Y, Cheng R et al (2017) Surrogate-assisted cooperative swarm optimization of high-dimensional expensive problems. *IEEE Trans Evol Comput* 21(4):644–660
25. Qin S, Sun C, Zhang G et al (2020) A modified particle swarm optimization based on decomposition with different ideal points for many-objective optimization problems. *Complex Intell Syst* 6:263–274
26. Guo Y, Zhang X, Gong D et al (2020) Novel interactive preference-based multi-objective evolutionary optimization for bolt supporting networks. *IEEE Trans Evol Comput* 24(4):750–764
27. Guo Y, Yang H, Chen M et al (2019) Ensemble prediction-based dynamic robust multi-objective optimization methods. *Swarm Evol Comput* 48:156–171
28. Jie J, Zhang J, Zheng H et al (2016) Formalized model and analysis of mixed swarm based cooperative particle swarm optimization. *Neurocomputing* 174:542–552
29. Huang CL, Dun JF (2008) A distributed PSO-SVM hybrid system with feature selection and parameter optimization. *Appl Soft Comput* 8(4):1381–1391
30. Khushaba RN, AlSukker A, Al-Ani A et al (2009) A novel swarm based feature selection algorithm in multifunction myoelectric control. *J Intell Fuzzy Syst* 20(4):175–185
31. Fang Y (2015) Interacting with prosthetic hands via electromyography signals. University of Portsmouth, London, pp 62–65
32. Merletti R (1996) Standards for reporting EMG data. *J Electromyogr Kinesiol* 6(1):3–4
33. Hudgins B, Parker P, Scott R (1993) A new strategy for multifunction myoelectric control. *IEEE Trans Biomed Eng* 40(1):82–94
34. Phinyomark A, Limsakul C, Phukpattaranont P (2009) A novel feature extraction for robust EMG pattern recognition. *J Comput* 1:71–80
35. Du S, Vuskovic M. Temporal vs. spectral approach to feature extraction from prehensile EMG signals. In: *Proceedings of IEEE international conference on information reuse and integration*, pp. 344–350, 2004.
36. Saridis GN, Gootee TP (1982) EMG pattern analysis and classification for aprosthetic arm. *IEEE Trans Biomed Eng* 29(6):403–412
37. Philipson L (1987) The electromyographic signal used for control of upper extremity prostheses and for quantification of motor blockade during epidural an aesthesia. Linköping University, Linköping
38. Granpe D, Salahi J, Zhang D (1985) Stochastic analysis of myoelectric temporal signatures for mula-functional single-site activation of prostheses and orthoses. *J Biomed Eng* 7(1):18–28
39. Zecca M, Micera S, Carrozza MC, Dario P (2002) Control of multifunctional prosthetic hands by processing the electromyographic signal. *Crit Rev Biomed Eng* 30(4–6):459–485
40. Park SH, Lee SP (1998) EMG pattern recognition based on artificial intelligence techniques. *IEEE Trans Rehabil Eng* 6(4):400–405
41. Öberg T, Sandsjö L, Kadefors R (1994) EMG mean power frequency: obtaining a reference value. *Clin Biomech* 9(4):253–257
42. Winter DA (2005) *Biomechanics and motor control of human movement*, 3rd edn. Wiley, Hoboken
43. Shi Y, Eberhart RC. Empirical study of particle swarm optimization. In: *Proceedings of IEEE congress on evolutionary computation*, Washington, pp. 1945–1950, 1999.
44. Sharma D, Willy C, Bischoff J (2020) Optimal subset selection for causal inference using machine learning ensembles and particle swarm optimization. *Complex Intell Syst* 466:1–19
45. Zhang H, Zhou A, Lin X (2020) Interpretable policy derivation for reinforcement learning based on evolutionary feature synthesis. *Complex Intell Syst* 6(3):741–753

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.