Biomedical Signal Processing and Control 73 (2022) 103364

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

Gesture recognition using dual-stream CNN based on fusion of sEMG energy kernel phase portrait and IMU amplitude image

Liukai Xu [a,b], Keqin Zhang [c], Genke Yang [a,b,*], Jian Chu [a,b]

[a] Ningbo Artificial Intelligence Institute of Shanghai Jiao Tong University, Zhejiang 315000, PR China
[b] Department of Automation, Shanghai Jiao Tong University, Shanghai 200240, PR China
[c] Ningbo Industrial Internet Institute, Zhejiang 315000, PR China

ARTICLE INFO

Keywords:
sEMG
Gesture recognition
Energy kernel
CNN
IMU

ABSTRACT

Since the surface electromyography signal (sEMG) is a weak and unstable signal, human action recognition based on sEMG time domain or frequency domain feature extraction has poor stability and low recognition rate. Based on modeling analysis of sEMG, the dual-stream convolutional neural network (CNN) is built to perform feature fusion of sEMG energy kernel phase portrait and inertial measurement unit(IMU) data for gesture recognition in this paper. Firstly, methods of matrix counting method and time window amplitude are utilized to represent the sEMG and IMU data as images. Secondly, the moving average method is presented to filter the images. Thirdly, the dual-stream CNN is used to perform feature extraction, fusion and classification of sEMG and IMU images respectively. Finally, the experiments yield an average recognition accuracy of 95.78% for 6 gestures performed by 5 subjects, which demonstrates that the proposed method has better performance for human gestures recognition.

1. Introduction

With the intensification of population aging, the number of elderly patients with movement disorders has also increased rapidly[1]. There is a great demand for robots that can assist in rehabilitation and mobility. The exoskeleton robot is a new type of mechatronics device proposed by imitating the exoskeleton of the biological world and combines sensing, control, information fusion, mobile computing and other technologies, which can not only provide support and protection for the wearer, but also assist the human limbs to perform moderate movements[2]. Obtaining sensor information and further identifying the movement intention of the wearer are important foundations for realizing the auxiliary movement function of exoskeleton robots, and have also been research hotspots in human–computer interaction[3]. According to the sensor types, the recognition of human movement intention is mainly based on human physiological signals or human–computer interaction information. Among human physiological signals, surface electromyography (sEMG) is an unstable, non-periodic and weak electrical signal generated during muscle contraction, which can express muscle movement and information about limb activities[4]. The sEMG signal contains a wealth of information and can non-invasively detect muscle activity, which is now widely used[5]. sEMG is generally generated 30 to 150 ms earlier than the movement of human limbs, which provides a basis for the real-time control of exoskeleton equipments[4]. Human intention recognition based on sEMG signals mainly includes two major issues: feature extraction and classifier design. Commonly used sEMG signal features include three categories: time domain features, frequency domain features, and time–frequency analysis[6-8]. Although the time-domain and frequency-domain features are easy to calculate, the instability of the sEMG signal results in poor stability of these features[9]. Even though the time–frequency feature analysis is stable, its calculations are complex and can not satisfy the real-time requirements of motion intention recognition[9]. Due to the instability of time domain and frequency domain features in sEMG signal recognition, some scholars have begun to model and analyze muscle biological mechanisms. [9] proposed the energy kernel method to estimate contractility and some inherent characteristics of muscles, and proved the robustness and strong generalization of this feature through experiments. Previous studies on the energy kernel of sEMG signals mainly focus on how to calculate the ellipse distribution area of the energy kernel phase portrait more efficiently[9-11]. Manual calculation of the area of the energy kernel ellipse not only lacks a unified parameter standard, but also leads to insufficient extraction of energy kernel features. Intention recognition based on human–computer

* Corresponding author at: Department of Automation, Shanghai Jiao Tong University, Shanghai 200240, PR China.
E-mail address: gkyang@sjtu.edu.cn (G. Yang).

https://doi.org/10.1016/j.bspc.2021.103364
Received 22 July 2021; Received in revised form 18 October 2021; Accepted 7 November 2021
Available online 20 December 2021

interaction information mainly uses information such as acceleration and angular velocity to recognize human activities or fall events[12-15], and the commonly used sensor is the inertial measurement unit (IMU), which consists of accelerometer, gyroscope and magnetometer. [12] proposed a deep convolutional neural network to perform human activity recognition by exploiting the inherent characteristics of activity and time-series signals. [14] presented a fall detection system utilizing the total sum vector magnitude of tri-axial accelerometer sensor data to detect fall events. In order to improve the recognition reliability of human actions, many current researches are based on information fusion of sEMG and IMU [16-19]. [16] presented a support vector machine (SVM) framework based on fusion information of wrist pitch and roll angle calculated from acceleration, gyro data acquired in IMU and the waveform length (WL) feature of sEMG, which yielded an accuracy of more than 98% for the recognition of human body's six wrist movements. [17] proposed a hierarchical recognition framework for hand movements by extracting the time domain features of arm sEMG data, angular velocity and acceleration information from IMU respectively. Then it employed SVM and decision tree to perform classification, and this arrangement had up to 95.6% accuracy in classifying 6 kinds of gestures. [18] employed the average absolute value, waveform length and other time-domain information extracted from sEMG and IMU sensors. It used linear discriminant analysis (LDA) algorithm to identify and classify signal features and the average recognition accuracy of 8 gestures reached 92.6%. [19] utilized the HMM model to recognize the frequency domain features of sEMG data and the posture angle information of the IMU, and the classification results were then applied to the control instructions of the mobile robot. Previous human intention recognition methods based on sEMG and IMU information fusion mainly perform feature fusion at the level of one-dimensional vector.

For the sEMG signal processing, manually extracting sEMG time domain or frequency domain features is widely used, which has the limitation of instability. For IMU signal processing, the calculation algorithms are often empolyed to solve the attitude angles, which has high computational complexity and leads to insufficient feature extraction. In order to overcome the limitations of manually extracting features, artificial neural network (ANN) algorithms are naturally introduced. Among them, convolutional neural network (CNN), multilayer perceptron (MLP), recurrent neural network (RNN) and long short-term memory (LSTM) are widely used [20-23]. [20] utilized the two-dimensional matrix to present the amplitude of the multi-channel sEMG signal, and used CNN to extract and classify the characteristics of the amplitude matrix. [21] employed multi-stream CNN to train the instantaneous amplitude map of high-density sEMG signals. It divides the multi-channel sEMG signal amplitude map into multi-stream CNN for fusion and recognition. [22] used MLP and LSTM to extract the features of the sEMG signal timing and amplitude matrix respectively, and then fused the features learned by the networks into new feature vectors for recognition. [23] utilized the LSTM network to extract the timing features of the sEMG signal, and then applied CNN to perform secondary feature extraction and recognition of the output features from LSTM.

According to the characteristics of sensor data and the preprocessing method of this article, the dual-stream convolutional neural network (CNN) is adopted to perform feature fusion of sEMG energy kernel phase portrait and IMU data for gesture recognition. The main contributions of this paper are summarized as follows:

1) The matrix counting method and sEMG energy kernel theory are utilized to represent the sEMG signals in form of two-dimensional images, which eliminates area calculation steps and reduces calculation errors.
2) The time window amplitude method is used to represent IMU data as images, which avoids the time complexity of the angle calculation.
3) Automatic feature extraction and fusion are performed through the dual-stream CNN training, which avoids the tedious process of manually extracting features.
4) Signals of sEMG and IMUs are further tested to verify the effectiveness and practicability of the recognition algorithm.

2. Energy kernel theory of sEMG

2.1. Energy kernel phase portrait of sEMG

Different from the features of time domain, frequency domain and time–frequency analysis, the energy kernel is a new method by using sEMG signal modeling. sEMG is a zero-averaged stochastic wave signal, whose amplitude is featured by the reciprocating motions accompanied by noise [24]. From the perspective of mathematics, such motion patterns can be abstracted as an "oscillator" [9-11], which are described by differential equations[25]. In the x-y phase plane, the amplitude of sEMG signal is defined by the variable x, and the derivative of amplitude over time is defined by the variable y, then the phase portrait of a sEMG segment can be drawn. Fig. 1 shows a segment of two-channel sEMG signals (sampling rate, 1000 Hz) and the corresponding energy kernel phase portrait. In Fig. 1(a) the blue line in the two-channel sEMG signal represents the channel 1 signal, and the orange curve represents the channel 2 signal. The phase portrait of the sEMG data enclosed by the green dashed box (200 ms) is drawn as scatter points with the dotted line indicating the distribution boundary of the corresponding phase portrait points. As Fig. 1(b) shows, the scatter points of phase portrait diagram approximately present the elliptical distribution in the coordinate plane [9-11]. [11] applied the linear fencing (LF) method to locate the ellipse boundary and found that the long axis of the ellipse was perpendicular or parallel to the abscissa axis most of the time.

2.2. Oscillator model of sEMG

The sEMG phase portrait is based on the oscillator model of sEMG. Suppose there is a simple harmonic spring oscillator with mass m and spring stiffness k, where x and y respectively represent the magnitude and the velocity of the oscillator. According to the Newtonian equation of motion, the differential equations of "oscillator" can be obtained as follows:

$$\dot{x} = y,\ \dot{y} = -\frac{k}{m}x \tag{1}$$

where $\dot{x}$ and $\dot{y}$ represent the first derivative of x and y respectively. In a short time, the sEMG signal can be approximately modeled as oscillator. For the spring oscillator model, its energy is the accumulation of kinetic energy and elastic potential energy, thus the energy E of the sEMG signal oscillator model can be expressed as follows:

$$E = \frac{1}{2}kx^2 + \frac{1}{2}my^2 \tag{2}$$

where k and m are only related to the internal characteristics of the sEMG signal. The Eqs.(2) can also be expressed as elliptic equation, as follows:

$$\frac{x^2}{2E/k} + \frac{y^2}{2E/m} = 1 \tag{3}$$

From Eqs.(3), the proportional relationship between the ellipse area S enclosed by the scatter point distribution boundary of the phase portrait and the energy E of the sEMG signal oscillator model is shown as follows:

$$S = \frac{2\pi}{\sqrt{km}}E \tag{4}$$

thus, the planar ellipse on the phase portrait(Fig. 1(b)) is named as "energy kernel".

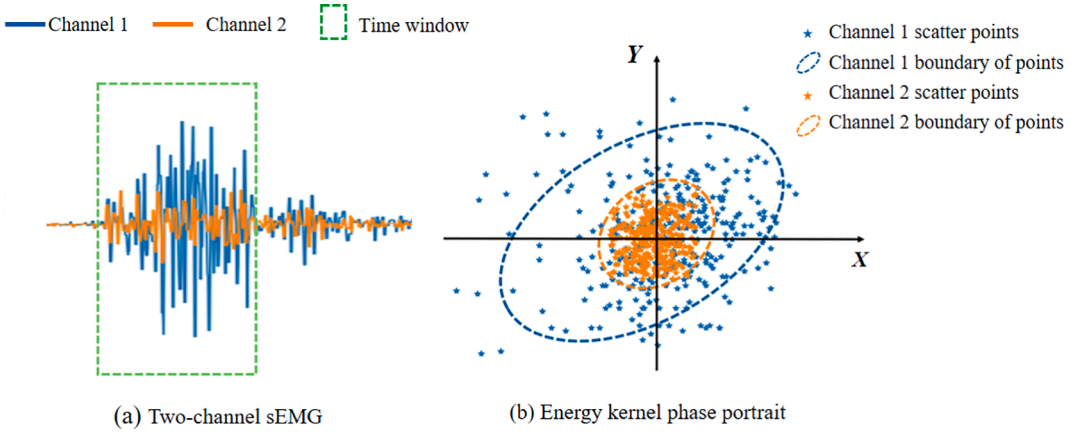

Fig. 1. Two-channel sEMG signal and corresponding energy kernel phase portrait.

sEMG signal is formed by the superposition of the motor unit action potential train (MUAPT) generated from all active motor units during muscle contraction, which can be considered as propagating waves spreading on muscle fibers[26]. The energy of the sEMG signal oscillator is actually the energy summation of all the harmonic waves formed by MUAPT, and the average energy density $\overline{E}$ of the harmonics can be calculated as follows:

$$\overline{E} \cong \frac{1}{2}\rho\left(\sum_j A_j^2\right)\omega^2 \tag{5}$$

where A_j denotes the the amplitude of the jth harmonic wave. ρ represents the mass density. ω denotes the dominating frequency component of the vibration source. Previous researches have recognized that the root mean square (RMS) of the sEMG signal has a certain linear relationship with the isometric contraction force of the muscle in a short time[27]. Therefore, the ellipse area can be used to characterize the muscle force features inside the sEMG signal.

We can summarize that the ellipse area of the energy kernel phase portrait of the sEMG signal can characterize muscle force. [9] proved that the energy kernel method has the characteristic of high signal-to-noise ratio and is very robust. Researchers have proposed many recognition algorithms for calculating the ellipse area of the energy kernel phase portrait and applied them in the recognition of sEMG signals. The energy kernel phase portrait contains rich information such as the area of the ellipse and the distribution characteristics of sample points. The recognition algorithm that simply extracts the area ignores the features of the signal sampling point distribution trend and sparsity in the phase diagram, which is the limitation of manually extracting features. In this paper, we propose the CNN algorithm for feature learning and recognition of the sEMG signal energy kernel phase portrait, which eliminates the time complexity caused by the ellipse area calculation, and employs the network to automatically learn more distribution features of the phase portrait.

3. Material and methods

3.1. Image processing of sEMG and IMU data

3.1.1. Image processing of sEMG

According to the energy kernel model, the sEMG signal of each acquisition channel corresponds to a scatter point distribution map containing N sampling points. Filter out the interference points with a large difference from the coordinate origin, and then the grid with the size of $N_1 \times N_2$ is used to cover the scattered points on the phase portrait. By counting the number of scattered points inside each grid, the grid is then converted into a $N_1 \times N_2$ matrix. This two-dimensional digital matrix $\boldsymbol{M}$ constitutes a grayscale digital image of $N_1 \times N_2$, and each element of $\boldsymbol{M}$ corresponds to the image pixel. The calculation of the length l and width w of each grid cell is shown as follows:

$$w = (x_{\max} - x_{\min})/N_1, l = (y_{\max} - y_{\min})/N_2 \tag{6}$$

where $x_{\max}$, $y_{\max}$, $x_{\min}$, and $y_{\min}$ are the maximum and minimum values of coordinates of the energy kernel scattered points, and N_1 and N_2 represent the size of the counting matrix. Due to the differences in the cell grid area of different counting grids, the weighting coefficient of cell area is added to the counting matrix, as follows:

$$\boldsymbol{I} = \boldsymbol{M} \times \alpha, \tag{7}$$

where $\boldsymbol{I}$ represents the processed grayscale image, $\boldsymbol{M}$ denotes the counting matrix corresponding to the counting grid, and α represents the area of each rectangular grid cell. In addition, according to [9], scatter points in the energy kernel phase portrait are elliptically distributed around the origin of the coordinate, and the distribution density gradually decreases from the inside to the outside. Due to drawing error factors, the grid count result close to the origin of the coordinate may be zero, which would result the phenomenon of pixel depression in the image and affect effectiveness of the energy kernel phase portrait. Therefore, the moving average method is applied to perform sliding filtering on the matrix image. The filtering formula is

$$\overline{p}_{i,j} = \frac{1}{9}(p_{i-1,j-1} + p_{i-1,j} + p_{i-1,j+1} + p_{i,j-1} + p_{i,j} + p_{i,j+1} + p_{i+1,j-1} + p_{i+1,j} + p_{i+1,j+1}) \tag{8}$$

where $\overline{p}_{i,j}$ ($i = 1, \ldots, N_1$, $j = 1, \ldots, N_2$) represents the averaged statistical value of each cell in the counting grid. It is more reasonable that the filtering method can effectively avoid the phenomenon of pixel depression and remove some interference noise on the image. The single-channel sEMG signal energy kernel phase portrait obtained by the above preprocessing forms a grayscale image with the size of $N_1 \times N_2$. Here, according to multiple experiments and the spatial relative positions of the sEMG signal acquisition electrodes, the grayscale images are image-spliced in the order of the channels. Thus, the stitched images contain the spatial position information of the electrodes and are finally processed into grayscale images with size of $2N_1 \times 4N_2$. In order to ensure the continuity of features, the method of "time-incremental window" as shown in Fig. 2 is applied to extract the features of the sEMG signal segments. *WT* and *ST* respectively represent the length of the time window and the length of the incremental window at time *K*.

3.1.2. Image processing of IMU data

The IMU data consists of 3-axis acceleration signals, 3-axis angular velocity signals, and 3-axis magnetometer signals. Considering that the magnetometer is greatly interfered by the outside world, the IMU data in

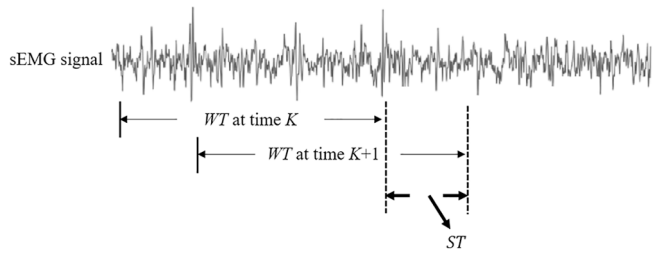

Fig. 2. Sliding time window method to extract signal segments.

this paper only contains acceleration and angular velocity data. In order to fully retain the characteristic information of the IMU data, this paper arranges the IMU's 6-axis signals in the order of X, Y, and Z axes, and applies the "time-incremental window" method to extract the IMU signal segments. The IMU signals are shown in Fig. 3(a), where Acc_X1 and Gyro_X2 represent the acceleration signal and angular velocity in X axis of two IMUs respectively and the remaining symbols have similar meanings. The amplitude images of the IMU data enclosed by the black dashed boxes are shown in Fig. 3(b). The length of time window and the moving window are respectively defined by variables of *WTimu* and *STimu*, the signals intercepted in each time window form a two-dimensional matrix with the size of $WTimu \times 12$, where each element of the matrix represents the signal amplitude at the corresponding position of the two IMUs. Then the amplitude matrix is normalized to form a grayscale image format, thereby completing the image processing of the IMU data.

3.2. Dual-stream CNN recognition framework

CNN is a feed-forward neural network that includes convolution calculations, which can automatically learn effective features and free researchers from the tedious process of manually constructing features [21]. Inspired by the excellent performance of Alexnet on classification problems[28], a dual-stream CNN recognition network is built and shown in Fig. 4, which can perform feature extraction and network learning on sEMG and IMU images respectively. Features of sEMG and IMU signals learned by the network are merged in the concatenation layer. The dual-stream CNN is mainly composed of convolutional layers, pooling layers, and fully connected layers. The convolution kernels in the convolutional layers extract features by performing convolution operations on the input samples. The pooling layers use spatial operations and pooling rules to reduce the size of the feature maps generated by convolutional layers, and the fully connected layers are applied to collect the feature space diffusion information of the features after convolution and pooling. As an end-to-end network, CNN has a strong learning ability in terms of convolutional feature extraction and classification of fully connected layers[29].

The dual-stream CNN recognition algorithm based on sEMG energy kernel phase portrait and IMU amplitude image proposed in this paper has the following advantages:

1) In the aspect of sEMG feature extraction, the energy kernel phase portrait is applied to characterize sEMG to overcome the limitations of time domain and frequency domain features, and at the same time, the step of calculating the ellipse area of the energy kernel scatter points distribution is omitted.
2) For IMU signal processing, the two-dimensional amplitude matrix of the multi-axis IMU data is applied to preserve the information characteristics of the IMU signals to the greatest extent, and the posture settlement step is omitted.
3) In terms of classifier design, the dual-stream CNN network with excellent image processing performance is used to automatically extract and learn features of sEMG and IMU grayscale images, avoiding the complexity and limitations of manual extraction of signal features.
4) The sEMG and IMU features learned and merged from the network can improve the performance of the recognition.

4. Gesture recognition experiments

4.1. Experimental configurations

In order to verify the effectiveness of the recognition algorithm based on sEMG and IMU signals, this study uses the sEMG sensor named "gForce Pro Wristband" to acquire the sEMG signals of human forearm with the sampling frequency of 1000 Hz. Two IMU sensors named "MPU9250" are placed on the finger and wrist to acquire the hand IMU data, whose sampling frequency is about 200 Hz. For 5 male subjects aged 22 to 25, sEMG and IMU signals are acquired from six gestures: fist, OK, inward wave, outward wave, pistol, and outstretched hand. The specific recognition framework is shown as Fig. 5. After filtering and denoising the signals, training and test image samples are divided. According to parameters of sEMG processing in [10,11] and experimental results, on the sEMG image processing in this paper, the counting grid size is $N_1 = N_2 = 20$. the length of time window (*WT*) and the moving window (*ST*) are 200 ms and 50 ms. Thus the energy kernel phase portrait extracted from the sEMG signals forms a training set with the image size of 40×80. In the IMU image processing, *WTimu*, *STimu*, and the sample size are 40, 20, and 40×12 for two IMUs. As the sampling frequency of sEMG and IMU is different, there will be a difference in the

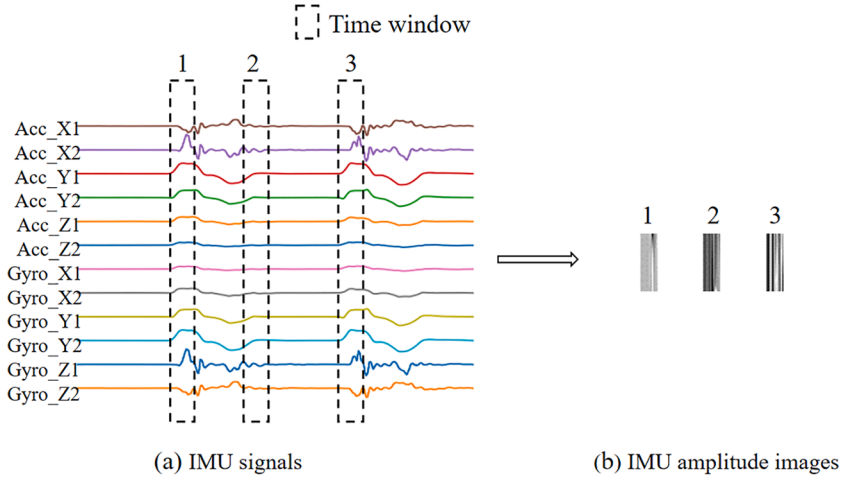

Fig. 3. Signals arrangement from two IMUs and the amplitude images of the signals in the time windows.

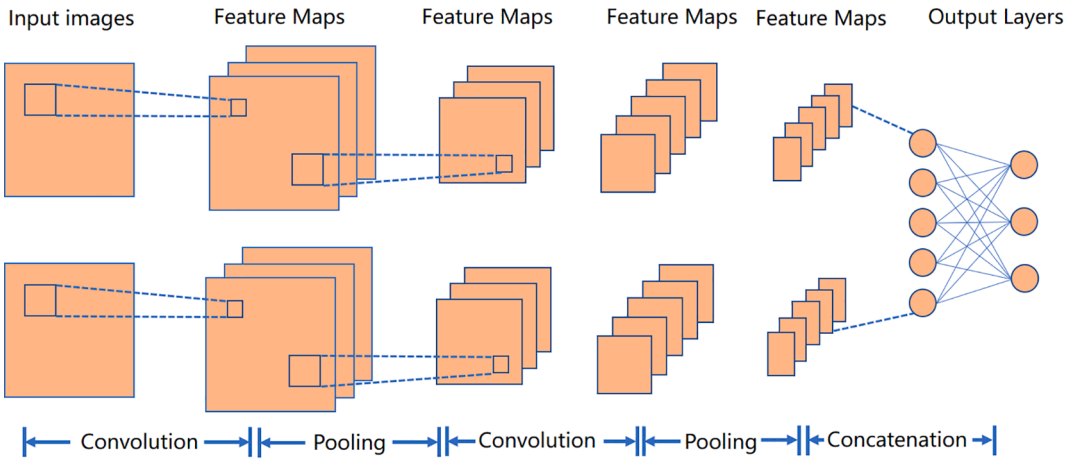

Fig. 4. Main structure of dual-stream CNN.

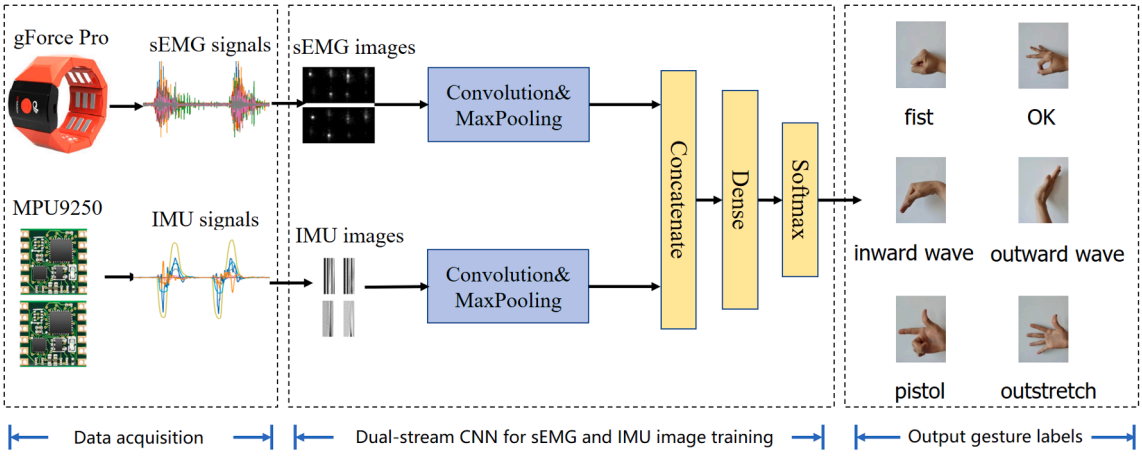

Fig. 5. Hardware setup and recognition framework of dual-stream CNN.

number of samples of two types. Thus, in the preparation of the training set, the IMU samples are replicated in the same proportion to achieve the same number of sEMG samples. In terms of dual-stream CNN network design, the training network for the phase portrait image of the sEMG energy kernel uses two convolutional layers and two pooling layers. The specific structure is shown in Table 1, where s and c respectively represent the stride of the convolution and the number of output channels. The convolutional layers utilize four and eight convolution kernels with size of 3×3 respectively. The training network for the IMU amplitude image has two convolution layers and two pooling layers, which includes 8 convolution kernels whose size is 3×3 and 16 convolution kernels with the size of 3×3 respectively. The specific structure is shown in Table 2. The sEMG and IMU features learned after convolution, pooling and other layers are output as feature vectors through the flatten layer, then the fully connected layers are applied for feature vectors splicing, fusion and classification. The fully connected classification module mainly includes a hidden layer composed of 16 neural units and an output layer using a normalized exponential function (softmax). The specific parameters are shown in Table 3. This dual-stream CNN employs cross-entropy function, the adaptive optimizer "Adam" and rectified linear unit(ReLU) for the network training. By performing 5-fold cross-validation on the training set and analyzing the trend of loss function value to judge whether overfitting occurs, the CNN iterations are then determined. Furthermore, the "dropout" method is

Table 2
Specific CNN layers and parameters for IMU training

CNN layers	Output dimension
Conv 8@3 × 3, $s = 3$, $c = 8$	(None, 14, 12, 8)
MaxPooling, $s = 2$	(None, 7, 6, 8)
Conv 16@3 × 3, $s = 1$, $c = 16$	(None, 7, 6, 16)
MaxPooling, $s = 2$	(None, 3, 3, 16)
Flatten	(None, 144)

Table 1
Specific CNN layers and parameters for sEMG training.

CNN layers	Output dimension
Conv 4@3 × 5, $s = 3$, $c = 4$	(None, 13, 26, 4)
MaxPooling, $s = 2$	(None, 6, 13, 4)
Conv 8@3 × 3, $s = 1$, $c = 8$	(None, 6, 13, 8)
MaxPooling, $s = 2$	(None, 3, 6, 8)
Flatten	(None, 144)

Table 3
Fully connected layers and parameters for feature fusion of sEMG and IMU

Feature fusion layers	Output dimension
Concatenate	[(None, 144), (None, 144)]
Dropout(0.5)	(None, 288)
Dense(16)	(None, 16)
Dropout(0.5)	(None, 16)
Dense(6)	(None, 6)

applied to randomly discard 50% of the output features in each training to reduce the degree of overfitting.

4.2. Experimental results

4.2.1. Single signal and fusion signals comparison

In order to validate the performance of the recognition algorithm based on sEMG and IMU information fusion, comparative experiments are carried out. The classification network is CNN, and the recognition accuracy of 6 gestures from 5 subjects is shown in Fig. 6. The average recognition accuracy of the algorithm based on sEMG and IMU information fusion exceeds 93% and is improved by more than 5% on average comparing with the recognition algorithm relying solely on sEMG or IMU information, which demonstrates that the feature fusion method of sEMG and IMU can improve the recognition accuracy and reliability of human gestures.

4.2.2. Parameter selection of time window WT and WTimu

For the purpose of determining the appropriate time window length in image processing of sEMG and IMU signals, experiments with different time window lengths are carried out. According to [4,30], the period of human intent recognition and control should be less than 300 ms, so that the wearers would not notice the operation delay. Considering that the time of signal preprocessing and online recognition of the model is about 100 ms, thus the length of the time window should not exceed 200 ms. Based on this, the length of the time window in the experiment varies from 50 ms to 200 ms, and the average experimental results are shown as Fig. 7. It can be seen that when the time window length increases, the recognition effect is better. Considering the accuracy and real-time of recognition, *WT* and *WTimu* are both chosen as 200 ms.

4.2.3. Performance of energy kernel phase portrait for sEMG image processing

In order to verify the advantages of energy kernel phase portrait extracted from sEMG signal, we give comparative experiments based on amplitude images of sEMG signals. The amplitude image method is similar to the processing of IMUs, and the image size of 8-channel sEMG is 200×8. From the recognition results shown as Fig. 8, the energy kernel method has a higher accuracy than the amplitude method in gesture recognition of five subjects. The results demonstrate that the extraction of the energy kernel phase portrait has better recognition performance than the amplitude image without energy kernel feature extraction, which further illustrates the effectiveness of the energy

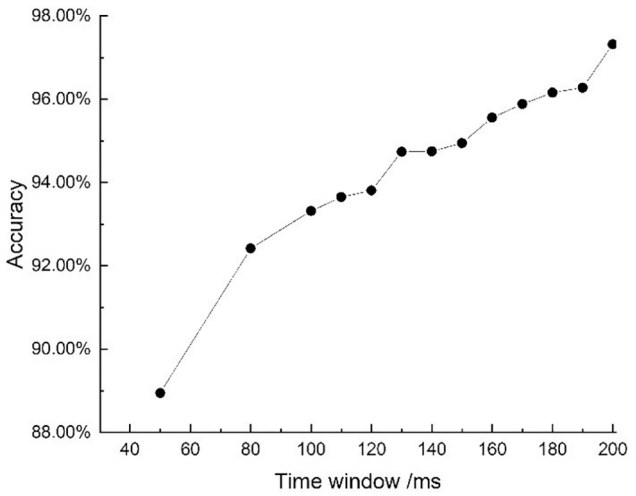

Fig. 7. Recognition rate varies with the length of time window.

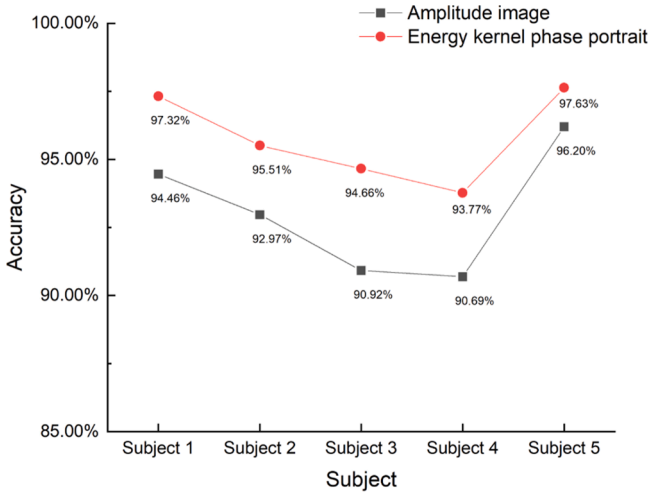

Fig. 8. Performance of energy kernel phase portrait for sEMG image processing.

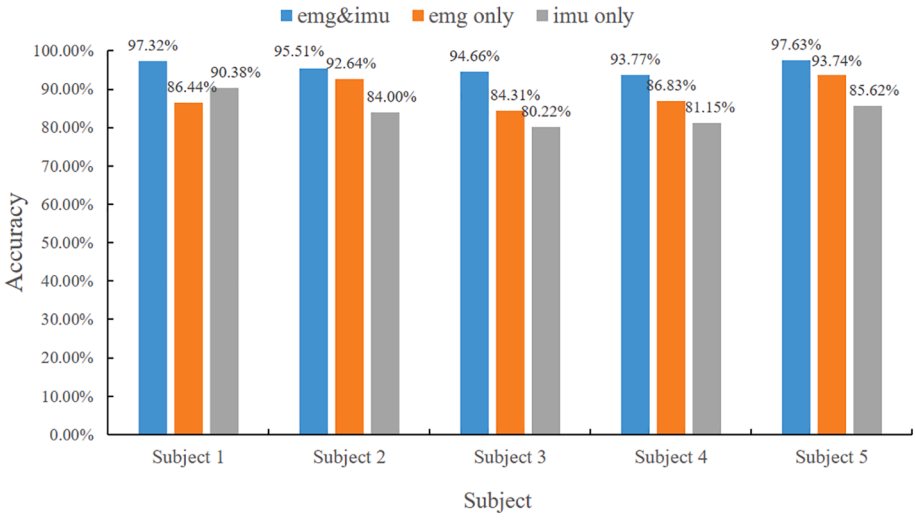

Fig. 6. Single signal and fusion signals comparison results on 5 subjects.

kernel phase portrait.

In order to verify the efficiency of the sEMG processing method in this paper, we compare the area calculation methods of linear fencing (LF) and discrete box counting (DBC) proposed in [9,10]. [11] used LF and DBC methods to calculate 100 sEMG data segments, which took 1.051 s and 0.789 s respectively. This paper uses the same data segment and the average calculation time is 0.298 s. It can be concluded that the calculation efficiency of the image processing for sEMG energy kernel model is higher than area calculation.

4.2.4. Feature extraction comparison experiment

In order to validate the performance of the two-dimensional image processing of sEMG and IMU signals proposed in this paper, we give the recognition experiment in light of one-dimensional feature extraction, such as the characteristics used in [17,18], from fusion of sEMG and IMU signals. For the sEMG signal, features of root mean square(RMS), mean absolute value(MAV), waveform length(WL), zero crossing(ZC) and slope sign changes(SSC) are extracted. For the IMU signals, RMS, MAV and WL features are extracted. Features from 8-channel sEMG sensor and two IMU sensors are then spliced to form feature vectors with the length of 76. Classifiers of naive bayes(NB), *K*-nearest neighbor(KNN), logistic regression(LR), decision tree(DT) and support vector machine (SVM) are applied to recognize the feature vectors. The average classification results of four subjects are shown in Table 4. It shows that the recognition algorithm which uses the dual-stream CNN to recognize the two-dimensional image features of sEMG and IMU has higher accuracy than the one-dimensional feature vector recognition algorithm. This experiment demonstrates that the feature fusion in form of two-dimension has higher accuracy rate for gesture recognition.

4.2.5. Comparison of amplitude image processing and angle calculation for IMU signals

In order to compare the IMU image processing with the recognition method of angle calculation for IMU signals proposed in [16], this paper performs the pitch and roll angles extraction on the IMU signals. Referring to the sEMG and IMU signal processing in [16], the WL features of the sEMG signals, the RMS and MAV features of the IMU signal attitude angles are extracted respectively. The experimental results are shown in the Table 5, which can be seen that the image processing of IMU can not only reduce the time complexity of posture solution, but also improve the recognition accuracy rate.

4.2.6. Comparison of different artificial neural networks

In order to validate the performance of the proposed CNN framework, we give experiments based on classic time series training networks such as dual stream recurrent neural network(Dual-RNN), dual stream long-short term memory(Dual-LSTM), and dual stream gate recurrent unit(Dual-GRU). The network inputs are all two-dimensional images of sEMG and IMU signals, and the recognition results are shown in the Table 6. The experimental results show that the recognition accuracy of the CNN is higher than that of the other three networks, indicating that CNN has better performance and advantages for two-dimensional image processing.

5. Conclusion

This paper proposes a dual-stream CNN gesture recognition algorithm based on the fusion of the sEMG signal energy kernel phase portrait and the IMU amplitude image. In order to verify the effectiveness of the algorithm, the paper designs a CNN recognition framework and conducts signal acquisition and recognition experiments for 6 gesture actions from 5 subjects. The recognition accuracy rate of all experiments exceeds 93%, which satisfy the requirements of exoskeleton robots. In order to compare the performance of different classification algorithms, this paper mainly gives five types of comparison experiments. The experimental results show the feasibility and effectiveness of the recognition algorithm in this paper. The recognition framework also provides a new idea for the current research on human–computer interaction based on sEMG signals. This study only gives experiments on gesture sEMG and IMU signals, and we next will focus on the research of lower limbs assisting exoskeleton robots. The follow-up researches would acquire the sEMG signals of the leg muscles and integrate more sensors such as force and acceleration to study a more reliable method of human movement intention recognition.

Table 4
Average classification results of one-dimensional feature extraction

Classifier	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5
NB	74.68%	69.65%	63.89%	68.31%	65.69%
KNN	93.51%	90.07%	83.84%	79.14%	88.29%
LR	93.64%	90.75%	84.51%	83.48%	88.89%
DT	95.35%	93.37%	91.11%	87.14%	93.09%
SVM	88.76%	88.88%	77.53%	72.82%	84.61%
Our method	97.32%	95.51%	94.66%	93.77%	97.63%

Table 5
Experimental results of angle calculation for IMU signals

Classifier	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5
SVM	74.11%	81.24%	71.66%	46.31%	79.08%
LDA	81.24%	74.32%	72.80%	77.41%	81.28%
KNN	86.07%	87.35%	85.57%	64.98%	87.06%
LR	71.69%	74.54%	72.33%	76.35%	74.63%
DT	87.54%	86.80%	86.86%	83.12%	94.05%
Our method	97.32%	95.51%	94.66%	93.77%	97.63%

Table 6
Comparison results of different artificial neural networks

ANNs	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5
Dual-RNN	82.86%	75.48%	78.79%	73.51%	85.99%
Dual-LSTM	94.53%	89.64%	88.71%	91.08%	94.50%
Dual-GRU	90.36%	90.14%	90.00%	89.09%	94.19%
Dual-CNN	97.32%	95.51%	94.66%	93.77%	97.63%

Funding

This work is supported by the China National R&D Key Research Program(2020 YFB1711200).

CRediT authorship contribution statement

Liukai Xu: Conceptualization, Investigation, Software, Writing – original draft. **Keqin Zhang:** Validation, Project administration. **Genke Yang:** Writing – review & editing, Methodology. **Jian Chu:** Project administration.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Z. Zhang, B.A. Muthu, C.B. Sivaparthipan, The necessary of constructing preventive health intervention policy under the trend of deep aging in China[J], J. Ambient Intell. Hum. Comput. 12 (3) (2021) 3539–3547, https://doi.org/10.1007/s12652-020-02594-8.

[2] Z. Lovrenovic, M. Doumit, Review and analysis of recent development of lower extremity exoskeletons for walking assist[C]//2016 IEEE EMBS International Student Conference (ISC), IEEE, Ottawa, ON, Canada, 2016, pp. 1–4.

[3] J.G. Ngeo, T. Tamei, T. Shibata, Continuous and simultaneous estimation of finger kinematics using inputs from an EMG-to-muscle activation model[J], J. NeuroEng. Rehabil. 11 (1) (2014) 1–14, https://doi.org/10.1186/1743-0003-11-122.
[4] J.-U. Chu, I. Moon, Y.-J. Lee, S.-K. Kim, M.-S. Mun, A supervised feature-projection-based real-time EMG pattern recognition for multifunction myoelectric hand control[J], IEEE/ASME Trans. Mechatron. 12 (3) (2007) 282–290, https://doi.org/10.1109/TMECH.2007.897262.
[5] P. Artemiadis, EMG-based robot control interfaces: past, present and future[J], Advances in Robotics & Automation 1 (2) (2012) 1–3, https://doi.org/10.4172/ara.1000e107.
[6] J.M. Fajardo, O. Gomez, F. Prieto, EMG hand gesture classification using handcrafted and deep features[J], Biomed. Signal Process. Control 63 (2021) 102210, https://doi.org/10.1016/j.bspc.2020.102210.
[7] Z. Lei, An upper limb movement estimation from electromyography by using BP neural network[J], Biomed. Signal Process. Control 49 (2019) 434–439, https://doi.org/10.1016/j.bspc.2018.12.020.
[8] X. Zhai, B. Jelfs, R.H.M. Chan, et al., Self-recalibrating surface EMG pattern recognition for neuroprosthesis control based on convolutional neural network[J], Front. Neurosci. 11 (2017) 379, https://doi.org/10.3389/fnins.2017.00379.
[9] X. Chen, Y. Yin, Y. Fan, EMG oscillator model-based energy kernel method for characterizing muscle intrinsic property under isometric contraction[J], Chin. Sci. Bull. 59 (14) (2014) 1556–1567, https://doi.org/10.1007/s11434-014-0147-3.
[10] X. Chen, Y. Zeng, Y. Yin, Improving the transparency of an exoskeleton knee joint based on the understanding of motor intent using energy kernel method of EMG[J], IEEE Trans. Neural Syst. Rehabil. Eng. 25 (6) (2017) 577–588, https://doi.org/10.1109/TNSRE.733310.1109/TNSRE.2016.2582321.
[11] Y. Zeng, JianTao Yang, C. Peng, Y. Yin, Evolving Gaussian process auto regression based learning of human motion intent using improved energy kernel method of EMG[J], IEEE Trans. Biomed. Eng. 66 (9) (2019) 2556–2565, https://doi.org/10.1109/TBME.1010.1109/TBME.2019.2892084.
[12] A. Ignatov, Real-time human activity recognition from accelerometer data using Convolutional Neural Networks[J], Appl. Soft Comput. 62 (2018) 915–922.
[13] C.A. Ronao, S.B. Cho, Human activity recognition with smartphone sensors using deep learning neural networks[J], Expert Syst. Appl. 59 (2016) 235–244, https://doi.org/10.1016/j.eswa.2016.04.032.
[14] Saadeh W, Altaf M A B, Altaf M S B. A high accuracy and low latency patient-specific wearable fall detection system[C]//2017 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI). IEEE, 2017: 441-444. https://doi.org/10.1109/BHI.2017.7897300.
[15] W. Saadeh, S.A. Butt, M.A.B. Altaf, A patient-specific single sensor IoT-based wearable fall prediction and detection system[J], IEEE Trans. Neural Syst. Rehabil. Eng. 27 (5) (2019) 995–1003, https://doi.org/10.1109/TNSRE.733310.1109/TNSRE.2019.2911602.
[16] H. Ali, W. Yanen, in: SVM Classification for Novel Time Domain IMU and EMG fused features for control of 6-DOF industrial robot[C]//2020, IEEE, Beijing, China, 2020, pp. 18–22, https://doi.org/10.1109/ICMA49215.2020.9233536.
[17] W. Chang, L. Dai, S. Sheng, et al., in: A hierarchical hand motions recognition method based on IMU and sEMG sensors[C]//2015, IEEE, Zhuhai, China, 2015, pp. 1024–1029, https://doi.org/10.1109/ROBIO.2015.7418906.
[18] S. Jiang, B. Lv, X. Sheng, et al., in: Development of a real-time hand gesture recognition wristband based on sEMG, IEEE, Qingdao, China, 2016, pp. 1256–1261, https://doi.org/10.1109/ROBIO.2016.7866498.
[19] Shin S O, Kim D, Seo Y H. Controlling mobile robot using imu and emg sensor-based gesture recognition[C]//2014 Ninth International Conference on Broadband and Wireless Computing, Communication and Applications. Guangdong, China: IEEE, 2014: 554-557. https://doi.org/10.1109/BWCCA.2014.145.
[20] W. Geng, Y.u. Du, W. Jin, W. Wei, Y.u. Hu, J. Li, Gesture recognition by instantaneous surface EMG images[J], Sci. Rep. 6 (1) (2016), https://doi.org/10.1038/srep36571.
[21] W. Wei, Y. Wong, Y.u. Du, Y.u. Hu, M. Kankanhalli, W. Geng, A multi-stream convolutional neural network for sEMG-based gesture recognition in muscle-computer interface[J], Pattern Recogn. Lett. 119 (2019) 131–138, https://doi.org/10.1016/j.patrec.2017.12.005.
[22] Y. He, O. Fukuda, N. Bu, et al., Surface emg pattern recognition using long short-term memory combined with multilayer perceptron[C]//2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE (2018) 5636–5639, https://doi.org/10.1109/EMBC.2018.8513595.
[23] Y. Wu, B. Zheng, Y. Zhao, Dynamic gesture recognition based on LSTM-CNN[C]//2018 Chinese Automation Congress (CAC), IEEE (2018) 2446–2450, https://doi.org/10.1109/CAC.2018.8623035.
[24] D. Staudenmann, K. Roeleveld, D.F. Stegeman, J.H. van Dieën, Methodological aspects of SEMG recordings for force estimation–a tutorial and review[J], J. Electromyogr. Kinesiol. 20 (3) (2010) 375–387, https://doi.org/10.1016/j.jelekin.2009.08.005.
[25] Y.-C. Du, C.-H. Lin, L.-Y. Shyu, T. Chen, Portable hand motion classifier for multi-channel surface electromyography recognition using grey relational analysis[J], Expert Syst. Appl. 37 (6) (2010) 4283–4291, https://doi.org/10.1016/j.eswa.2009.11.072.
[26] A.J. Mccomas K. Mrozek The electrical properties of muscle fiber membranes in dystrophia myotonica and myotonia congenita[J] Journal of Neurology, Neurosurgery, and Psychiatry 31 5 1968 441 https://dx.doi.org/10.1136%2Fjnnp.31.5.441.
[27] G.S. Trajano, K. Nosaka, A.J. Blazevich, Neurophysiological mechanisms underpinning stretch-induced force loss[J], Sports Medicine 47 (8) (2017) 1531–1541, https://doi.org/10.1007/s40279-017-0682-6.
[28] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks[J], Advances in Neural Information Processing Systems 60 (6) (2017) 84–90, https://doi.org/10.1145/3065386.
[29] J.O. Pinzón-arenas, R. Jiménez-moreno, J.E. Herrera-benavides, Convolutional neural network for hand gesture recognition using 8 different emg signals[C]//2019 XXII Symposium on Image, Signal Processing and Artificial Vision (STSIVA), IEEE, Bucaramanga, Colombia, 2019, pp. 1–5.
[30] K. Englehart, B. Hudgin, P.A. Parker, A wavelet-based continuous classification scheme for multifunction myoelectric control[J], IEEE Trans. Biomed. Eng. 48 (3) (2001) 302–311, https://doi.org/10.1109/10.914793.