**sEMG Motion Intention Recognition Based On Wavelet Time-Frequency Spectrum And ConvLSTM**

**Yixin Liu, Jin Hua, Daxin Xin and Mengzhao Liu**  
Xi'an Technological University, College of Electronic Information Engineering,  
Xi'an 710021, China

Email: 974795316@qq.com

**Abstract.** In this paper, a recognition method based on convLSTM-CNN is presented for time - frequency spectrum classification method of sEMG signals. ConvLSTM had the ability of global time-frequency spectrum feature extraction, and was designed to extract the time-frequency spectrum features of each channel. Due to its strong local feature extraction ability, CNN was designed to further extract the fused feature information and realize end-to-end classification. Experimental results show that this algorithm has better classification performance than the existing methods.

**1．Introduction**

The electromyographic signal applications, due to the electromyographic signal acquisition process is easily affected by noise, and electromyographic signal itself is a kind of weak electrical signals of the acquisition to the poor quality of the electromyographic signal of itself, is difficult to distinguish, these features for the follow-up action recognition based on electromyographic signal has brought great challenges, become the research in this field.

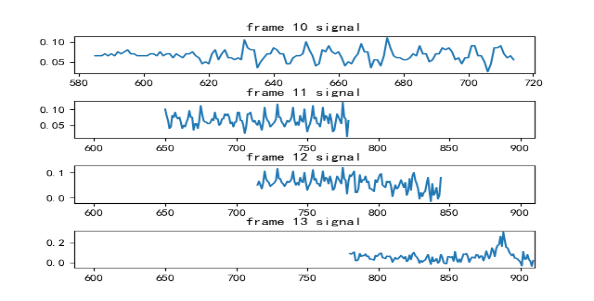
Because the signal is complex and has non-stationary randomness, it is difficult to obtain satisfactory classification and recognition effect simply by taking sEMG signal as the input of a classifier [1]. Therefore, the signal feature extraction is carried out to represent the signal type. In this paper, features of sEMG signals are extracted by two-dimensional time-frequency graph conversion method and the subsequent actions represented by sEMG signals are identified by deep learning method, so as to judge the user's motion intention.

**2．Data processing method**

2.1. Pretreatment

In the process of motion intention recognition, in order to achieve the purpose of stable recognition, the data is divided into data frames for processing. Sliding window segmentation is the most common and simple method. In this way, the data is taken as the frame data (window length) by a certain time length (e.g. 100ms), and then moved on the original data by a certain length such as half of the window length) as the next frame data. There is certain overlap between such window and window. This method not only increases the stability of each frame, but also increases the number of frames divided into individual segments. It is also simple and fast, so it is widely used. But this way to the window length and overlap length is difficult to determine. For sEMG signals, window length is usually selected between 100 and 300ms, and window overlap length is half of window length.

When the frame length is 256 and the frame is moved to 128, the sEMG signal from frame 10 to 13 is shown in Figure 1.



**Figure 1.** Frame 10 to 13 sEMG signal

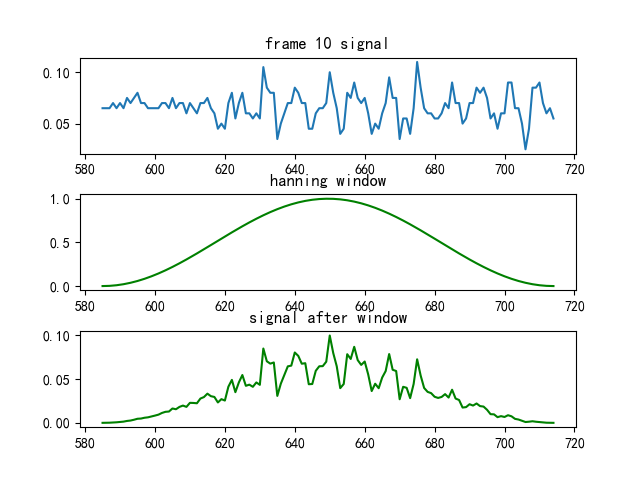
In general, it is necessary to add Windows for signal truncation and framing, because all truncations have energy leakage in the frequency domain, and window function can reduce the impact of truncation. The Hanning window, also known as the ascending cosine window, can be seen as the sum of the spectrums of the three rectangular time Windows, or the sum of the three functions, and the two terms in parentheses move to the left and right relative to the first one, thereby canceling out the side-lobes and eliminating the high frequency interference and energy leakage. It can be seen that the main lobe of the hanning window is widened and reduced, while the side lobe is significantly reduced. From the point of view of reducing leakage, the hanning window is superior to the rectangular window. However, the widened main lobe of hanning window is equivalent to the widened analysis bandwidth and the reduced frequency resolution.

Hanning window:

(1)

Where the window length is L.

In this project, hanning window is used to add window processing to sEMG signals. The hanning window function is shown in the second figure in Figure 2. After window processing to sEMG signals, signals at both ends attenuate.



**Figure 2.** Frame and window sEMG signal

2.2. Deal with the noise

Based on the difference between the real signal and the noise signal, wavelet transform separates the real signal and the noise signal through layer by layer decomposition, removes the noise signal through appropriate threshold selection, and finally reconstructs the original state of the pure sEMG signal through reconstruction algorithm.

Traditional data processing methods include filter, Kalman and empirical formula filtering. The analysis process of these data processing methods is relatively complicated, and some of them need some basic mathematical tools and empirical knowledge, so the denoising effect is not ideal. And discrete wavelet transform multi-scale refinement, ultimately achieve high frequency time segment, the low frequency in the frequency segment, can automatically adapt to the requirement of time-frequency signal analysis, which can be on any details of the signal, the operation is simple, easy to understand, coupled with the use of digital filter and a trap, incorporates the advantages of three kinds of denoising tool as a whole, the signal is pure and with the original signal with high similarity, therefore very suitable for sEMG this nonlinear signal processing.

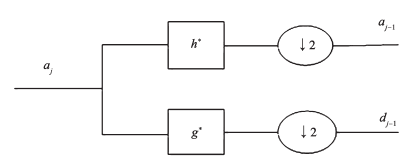
According to the property of discrete wavelet transform, any signal (function) can be decomposed into:

(2)

Where,is the scale coefficient, is the wavelet coefficient:

(3)

According to the above two equations, and of the J +1 layer can be decomposed to the scale coefficient of the J layer by filtering coefficient and .The decomposition is shown in Figure 3.



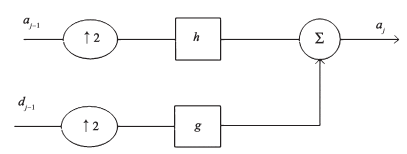
**Figure 3.** Discrete wavelet decomposition

Note: and are conjugates of h and g.

Reconstructing the algorithm formula of scale coefficient:

(4)

The reconstruction diagram is shown in Figure 4.



**Figure 4.** Wavelet reconstruction

In this paper, discrete wavelet is used to denoise the sEMG signals of the upper limbs. The signals of one movement and one channel are decomposed and compared. The comparison diagram of decomposition before and after denoising is shown in Figure 5.

图片包含 游戏机

描述已自动生成

**Figure 5.** Contrast of waveform before and after denoising

In the process of data collection, there is a small shift between the muscle contraction and the electrode, which is called baseline drift, and a linear filter is needed to remove the baseline drift.The processing results are shown in Figure 6.

图片包含 游戏机

描述已自动生成

**Figure 6.** De-baselining drift

2.3. Time-frequency spectrum

The basic idea of wavelet change is similar to the Fourier transform, which also USES a family of functions to represent signals, which is called the wavelet function system. Set function, and , the of the gens function expansion and translation.

（5）

Type: for the analysis of the wavelet or continuous wavelet; is the mother wavelet or basic wavelet. is the expansion factor that changes the shape of the wavelet; is the translation factor of the wavelet shift.

For an arbitrary function, the continuous wavelet transform is:

（6）

Type: is the complex conjugate for ; is inner product of them, the invisible is represented in coefficient of wavelet function when scale for position offset for , characterization of the similarity of wavelet function and the original signal. Where, and are continuous variables, so it is called continuous wavelet transform.

In this paper, using continuous wavelet transform, the algorithm steps of generating time-frequency graph by sEMG transformation are as follows:

Step 1: Set as the scale (stretching factor), as the sampling frequency, and as the wavelet center frequency, then the actual frequency of is:

（7）

Step 2: According to Equation (7), in order to the converted frequency sequence to be an isometric sequence, the scale sequence must take the following form:

（8）

Where: is the length of scale sequence used in wavelet transform of signal (preset to 256 in this paper), and C is a constant.

Step 3: It can be seen from Equation (7) that the actual frequency corresponding to scale should be fs/2, so it can be obtained:

（9）

Substitute Equation (9) into Equation (8) to obtain the required scale sequence t.

Step 4: After determining the wavelet base and scale, the wavelet coefficient is obtained by using the principle of continuous wavelet transform (Equation (6)). Then the scale sequence is converted into the actual frequency sequence F according to the principle of Equation (7). Finally, combining with the time series t, the wavelet time-frequency spectrum can be drawn to obtain the characteristic information.

图片包含 游戏机, 电脑

描述已自动生成图片包含 游戏机, 电脑

描述已自动生成

1. (b)

图片包含 游戏机, 电脑

描述已自动生成图片包含 游戏机, 电脑

描述已自动生成

(c) (d)

**Figure 7.** Time-Frequency spectrum. (a) time-frequency spectrum of make a fist in channel 1.(b) time-frequency spectrum of make a fist in channel 2.(c) time-frequency spectrum of Palm down in channel 1.(d) time-frequency spectrum of Palm down in channel 2

The x-coordinate of the spectrum is time, the y-coordinate is frequency, and the coordinate point value is the myoelectric signal data energy. Since the two-dimensional plane is used to express three-dimensional information, the energy value is expressed by color. The upper end of the color band indicates that the energy of the point is stronger, while the lower end of the color band indicates that the energy of the point is weaker.

The time-frequency map obtained from the sEMG signal in Figure 7 clearly shows that the sEMG time-frequency map from different channels of the same action is different and contains rich biological information about different muscles of the action. Therefore, the time-frequency graph can be made into a data set to collect the characteristics of sEMG signals and complete the upper limb motion recognition based on sEMG signals.

**3. Network structure**

Electromyographic signals are bioelectrical signals produced by muscle movements. Different muscles produce different electromyographic signals under the same movement. In this paper, eight-channel sEMG sensors are used to collect sEMG signals to obtain more biological information representing this action. Since there is a strong correlation between the sEMG signals of different channels and they represent the same action, the sEMG signals of different channels can be regarded as sequence data.

The sEMG signals of different channels contain different information, and the more information, the better the recognition performance. Therefore, using eight-channel sEMG signal for motion recognition can improve the recognition performance. The temporal feature extraction mechanism is also suitable for sEMG signal based motion recognition. At the same time, convolution operation can extract effective spatial features. ConvLSTM has the ability to extract structural features and to mine dependencies and correlations. Thus, in the course of motion classification and recognition of multichannel sEMG signals, the spectrum of each ConvLSTM was treated as input at different moments. ConvLSTM cells were used to extract correlations between different channels and structural features. On this basis, the corresponding relationship between the sEMG signals of different channels is used to improve the recognition performance.

The motion recognition structure of ConvLSTM based sEMG signals time-frequency spectrum is shown in Figure 8.The ConvLSTM consisted of three ConvLSTM layers, a CNN layer, a full ConvLSTM layer, and a Softmax layer. After pretreatment, the sEMG time-frequency spectrum of the different channels are fed into ConvLSTM as sequence data. In ConvLSTM layer, convolution kernel is introduced to extract structural features, thus reducing the dimension of the structure.The spatial information of sEMG signals is extracted from different channels by convolution kernel sliding data. At the same time, the output of the previous cell and the input of the current cell are merged into the input of the current cell.The features extracted by convolution operation are transferred to the next layer. In the last layer, a Softmax layer is introduced as a classifier to identify targets. ConvLSTM uses convolution to extract local spatial features. Meanwhile, the structure of the LSTM enable ConvLSTM to transmit previous information. Thus, ConvLSTM can use not only the time series features of the sEMG signals, but also the convolution operation to extract the spatial features.



**Figure 8.** Architecture of identification network

**4.Experiment and analysis**

The surface electromyogram signal (sEMG) is collected by zTEMG-8000 eight-channel EMG signal acquisition equipment of Qingdao Zhituo Company. The sampling resolution is 10 bit and the sampling frequency was 500 Hz. Eight surface sticking electrodes are used for the acquisition electrode. Eight upper limb movements including extending palm, clenching, palm up, palm down, radial bending, ruler bending, arm up and arm down are collected, as shown in Figure 9.The subjects are 5 healthy subjects aged 23-25 years. Each subject is asked to make approximately 3s for each gesture, followed by stretching five fingers to relax and rest for 3s. Each gesture intention is collected for 20 times, and the collected signals are stored in the form of a table. A total of 800 sets of data are obtained. The experimental data set is composed of eight-channel time-frequency sEMG atlas of eight movements. Eighty percent of the samples for each movement are training data, and twenty percent of the samples for each movement are test data.

图片包含 照片, 男人, 桌子, 房间

描述已自动生成

**Figure 9.** Schematic diagram of eight gestures

The model in this paper is implemented on TensorFlow and Keras frameworks. Each sEMG signal is divided into 20 frames. The input space size is adjusted to 256×256.Convolution kernel sizes were 5×5×5, 3×3×3 and 1×1×1, and three ConvLSTM networks with 256 hidden units each were used. The model was trained using the stochastic gradient descent method, and the Objective loss function was optimized using the Adam network optimizer. The initial learning rate was 0.001, the momentum was set at 0.9, and the batch size was set at 32.

Table 1 compares the recognition effects of the eight actions. The experimental results show that the network can effectively extract the features of the eight sEMG time-frequency spectrums and complete the feature classification, without the need to manually design feature vectors and classifiers according to the characteristics of the actions to be recognized.

|  |  |
| --- | --- |
| **Table 1.** Comparison of the eight actions on our network architectures. | |
| Motion | Accuracy |
| Extend palm | 98.2% |
| Make fist | 97.3% |
| Palm up | 92% |
| Palm down | 93.2% |
| Radial bend | 96% |
| Ruler bend | 96.3% |
| Arm up | 87.3% |
| Arm down | 84.8% |
| All motion | 93.13% |

Table 2 shows the recognition rates of different methods. The results show that the test results of this method in the existing data set are superior to other identification and classification methods. LSTM can only extract the temporal features but not the spatial features, thus obtaining the lowest recognition rate. CNN can extract the spatial details of myoelectric time-frequency spectrum, and its recognition rate is good. Since both convolution operation and feedback mechanism could be used to extract temporal features, ConvLSTM performed better than LSTM which could only extract temporal features and CNN which could only extract spatial features, with an recognition rate of 92.5%.ConvLSTM not only had LSTM timing modeling capability, but also could extract global features, so to speak, spatio-temporal features. Finally, CNN was used for further local feature extraction, and the experimental results showed that the performance of our method was better than other methods.

|  |  |
| --- | --- |
| **Table 2.** Comparison of the four network architectures on our dataset. | |
| Methods | Accuracy |
| LSTM | 90.8% |
| CNN | 91.22% |
| ConvLSTM | 92.5% |
| Ours | 93.13% |

**5. Conclusion**

In this paper, eight kinds of motions were selected for sEMG signal collection. In combination with the characteristics of time-frequency spectra, the relevant theories and methods of convolutional neural network and convolutional short-time memory network were studied, and the sEMG signal ConvLSTM-CNN identification model was established to classify the eight motions. In the preprocessing, wavelet transform is used to denoise sEMG signals. Existing algorithms cannot combine the sEMG characteristics of different channels. LSTM can well extract the correlation characteristics of sEMG signals from different channels. Meanwhile, CNN can effectively extract local spatial features through convolution operation. Therefore, combining the advantages of CNN and LSTM, this paper introduced a ConvLSTM network which could simultaneously extract the space-time characteristics. The sEMG signals of the different channels were thought to be sequence inputs of ConvLSTM. Compared with the four networks, the experimental results show that ConvLSTM combined with CNN algorithm is a feasible sEMG signal recognition method.

**REFERENCES**

[1] Lin K, Wu C, Huang X, Ding Q and Gao X. A robust gesture recognition algorithm based on surface EMG[C]. Advanced Computational Intelligence (ICACI 2015). IEEE, 2015.

[2] PukhovaV, Gorelova E, Ferrini G, et al. Time-frequency representation of signals by wavelet transform[C]. Young Researchers in Electrical & Electronic Engineering. IEEE, 2017:715-718.

[3] Luo K, Du K, Cai Z, et al. A modified frequency slice wavelet transform for physiological signal time-frequency analysis[C]. 2017 Chinese Automation Congress (CAC). IEEE, 2017:3441-3444.

[4] Chao H, He Y, Zhang J, et al. GaitSet: Regarding Gait as a Set for Cross-View Gait Recognition[J]. arXiv:1811.06186v4, 2018.

[5] Feng J, Wu X, Chen J, et al. Joint Multilayer Spatial-Spectral Classification of Hyperspectral Images Based on CNN and Convlstm[C]. IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium. IEEE, 2019: 588-591.

[6] [Manfredo Atzori](https://www.nature.com/articles/sdata201453/#auth-1), [Arjan Gijsberts](https://www.nature.com/articles/sdata201453/#auth-2), [Claudio Castellini](https://www.nature.com/articles/sdata201453/#auth-3), et al. Electromyography data for non-invasive naturally-controlled robotic hand prostheses[J]. Nature, Scientific data1, 140053(2014).

[7] Hu W S, Li H C, Pan L, et al. Spatial–Spectral Feature Extraction via Deep ConvLSTM Neural Networks for Hyperspectral Image Classification[J]. IEEE Transactions on Geoence and Remote Sensing, 2020, 58(6):4237-4250.

[8] Chen W, Zhang L, Xi Y, et al. Polarimetric HRRP Target Recognition Based on Convlstm[C]. IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium. IEEE, 2019:334-337.

[9] Dawei Huang, Badong Chen, Surface EMG Decoding for Hand Gestures Based on Spectrogram and CNN-LSTM[C]. the 2nd China Symposium on Cognitive Computing and Hybrid Intelligence (CCHI 2019). IEEE,2019:123-126.

**Acknowledgement**National Fund Youth Project (11804263).