

A Teleoperated Manipulator Control System Based on Surface Electromyography Using Deep Learning

Zhiping Lai¹, Xueze Zhang¹, Junkongshuai Wang¹, Wei Mu¹, Aiping Wang¹, Lan Niu², Lihua Zhang^{1,2}, Hongbo Wang¹, Xiaoyang Kang^{1,2,3,4}

¹Laboratory for Neural Interface and Brain Computer Interface, Engineering Research Center of AI & Robotics, Ministry of Education, Shanghai Engineering Research Center of AI & Robotics, MOE Frontiers Center for Brain Science, State Key Laboratory of Medical Neurobiology, Institute of AI & Robotics, Academy for Engineering & Technology, Fudan University, Shanghai, China

²Ji Hua Laboratory, Foshan, Guangdong Province, China

³Yiwu Research Institute of Fudan University, Chengbei Road, Yiwu City, 322000 Zhejiang, China.

⁴Research Center for Intelligent Sensing, Zhejiang Lab, Hangzhou 311100, China
e-mail: xiaoyang_kang@fudan.edu.cn

Abstract— Gesture recognition based on surface electromyography (sEMG) is a popular topic in teleoperation field, which has the characteristics of convenience, maneuverability, flexibility and adaptability. With the continuous development of deep learning technology, the variety and accuracy of gesture recognition are constantly improving. In our previous research, a new neural network based on graph convolutional network and temporal convolutional network (STCN-GR) was proposed and achieved good results in gesture recognition task. In this paper, we develop a teleoperated manipulator control system based on STCN-GR combined with majority voting. The experimental results show that the performance of proposed system can achieve relatively excellent performance in recognition accuracy and computation speed. In this paper, we choose a mechanism hand as a peripheral to test the performance of the STCN-GR based control system, and the results show that this system can be applied in teleoperation in feature work to provide convenience for human life.

I. INTRODUCTION

Human-computer interaction is an advanced technology that interacts with computers through various media. Since the invention of computers, people have begun to study various human-computer interaction technologies and use those technologies for the peripheral teleoperation. Early human-computer interaction technology mainly uses mouse, keyboard and other external devices, through the input of instructions to complete the interaction with the computer. After entering the 21st century, with the continuous development of software and hardware technology, peripheral-based interaction has been unable to meet people's daily needs, researchers began to study the use of more intelligent teleoperation technology.

An important way to realize human-centered human-computer interaction teleoperation system is perceptual computing, which uses visual, auditory, tactile and other

information to sense human intention and then converts it into corresponding instructions to complete interaction with the computer and complete the teleoperation [1]. Hand is the most flexible limb of human body, gesture contains abundant information, gesture recognition has become an important research content of perceptual computing. Gesture recognition is essentially a pattern recognition problem, which needs to recognize the corresponding gesture pattern from the signals obtained by cameras, sensors and other devices. Surface electro-myography (sEMG) signal is a kind of physiological signal collected from the skin surface. It has the characteristics of non-invasive and high precision, and is one of the main signal sources for gesture recognition at present.

Gesture recognition, as an important way to complete the teleoperation, has broad application prospects in medical rehabilitation, prosthesis control, virtual reality and other fields. Among many gesture recognition methods, gesture recognition technology based on sEMG signals is not easily affected by occlusion, illumination and other environmental factors, and can recognize complex and fine gestures. In addition, patients with hand dysfunction will also produce weak sEMG signals on their arms when attempting hand movements [2], which makes it possible for them to control peripherals, such as wheelchairs, exoskeletons and mechanical arms, etc.

In our previous work [3], we proposed a sEMG-based gesture recognition algorithm, which uses spatial-temporal convolutional networks comprising of graph convolutional network and temporal convolutional network to extract spatial-temporal features. The excellent performance on CapgMyo [4] and BandMyo [5] show its great advantage. In this paper, we build a manipulator control system based on sEMG signal which uses the gesture recognition algorithm proposed before.

This work was supported in part by the National Natural Science Foundation of China, grant no. 61904038 and no. U1913216; National Key R&D Program of China, grant no. 2021YFC0122702 and no. 2018YFC1705800; Shanghai Sailing Program, grant no. 19YF1403600; Shanghai Municipal Science and Technology Commission, grant no. 19441907600, no.19441908200, and no. 19511132000; Opening Project of Zhejiang Lab, grant no. 2021MC0AB01; Fudan University-CIOMP

Joint Fund, grant no.FC2019-002; Opening Project of Shanghai Robot R&D and Transformation Functional Platform, grant no. KEH2310024; Ji Hua Laboratory, grant no. X190021TB190 and no. X190021TB193; Shanghai Municipal Science and Technology Major Project, grant no. 2021SHZDZX0103 and no. 2018SHZDZX01. ZJ Lab, and Shanghai Center for Brain Science and Brain-Inspired Technology.

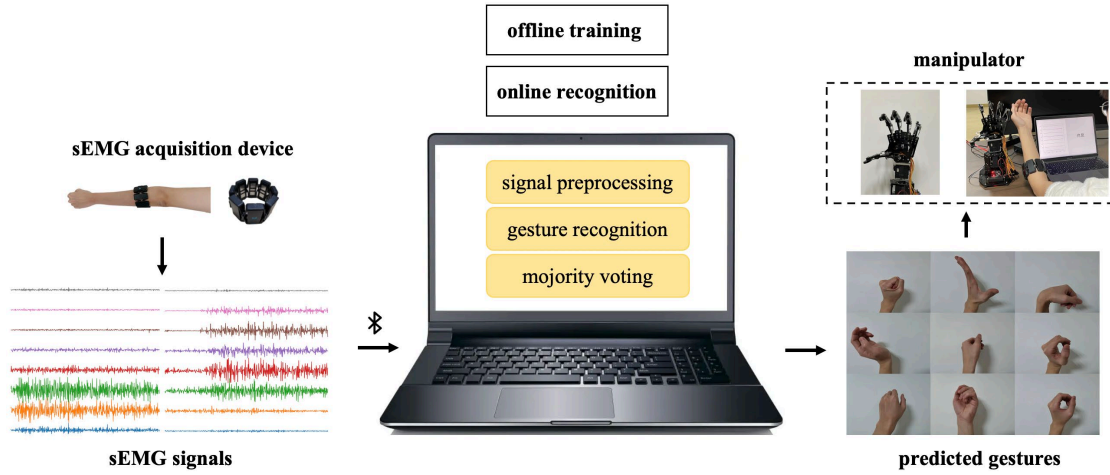


Figure 1 The overview of teleoperated manipulator control system. During offline training, the gesture recognition model based on deep learning is trained using sEMG signals that sensed from arm. During online recognition, acquired sEMG signals are sent to computing device via Bluetooth, and predicted gestures can be obtained using preprocessed signals after gesture recognition and majority voting, which will convert to motion instructions for manipulator.

II. RELATED WORK

In the existing hand gesture recognition methods based on sEMG signal, it can be divided into traditional machine learning method and deep learning method. The traditional machine learning gesture recognition technology mainly uses some classification algorithms to classify gestures based on feature extraction. Ju et al. [6] used root mean square (RMS) as the basic feature and adopted gaussian mixture model and support vector machine (SVM) to classify 10 hand movements, achieving classification accuracy of 88.5% and 88.75% respectively. Timemy et al. [7] combined the feature extraction method of time-domain autoregression with 15 kinds of gestures of normal people and 12 kinds of amputees. The accuracy of gesture recognition was 98% and 90% respectively. Sui et al. [8] optimized SVM with particle swarm optimization algorithm, and achieved 90.66% recognition accuracy in six kinds of upper limb motion classification, and the training time was shorter than that of traditional SVM method. It is not difficult to see from the above research that the traditional machine learning method can achieve good recognition effect, has strong robustness, and the calculation is relatively not very complicated. But on the other hand, the selection and construction of features in the time domain and frequency domain, as well as the selection of classification algorithms, need to rely on specific expert knowledge, and the recognition effect is limited, which is also a major drawback of traditional machine learning recognition methods.

Deep learning technology can automatically complete feature construction and extraction, the design of algorithms can be more diversified, and the ability to use information is more prominent. Based on the development of deep learning methods, researchers began to use deep learning methods to solve gesture recognition tasks based on sEMG signals. Amor et al. [9] used MYO to collect sEMG signals for sign language recognition, and used RNN to extract features from sequence data to analyze sign language gestures. Shin et al. [10] used RNN network containing three long-short term memory (LSTM) layers to extract features from sEMG

signals and IMU signals to realize Korean symbol recognition. Geng et al. [11] construct instantaneous sEMG images by spatial arrangement of channels of single frame sEMG signals, and identify corresponding gesture patterns from the instantaneous sEMG images, combining with majority voting method to achieve better recognition effect. Rahimian et al. [12] proposed a new network module structure, XceptionTime, and used this module to build an recognition network. Compared with the traditional model, this network model has fewer parameters and is insensitive to window size, and its recognition accuracy is as high as 5.43% on NinaPro DB1. Tsinganos et al. [13] used sequential convolutional neural network and attention mechanism to capture sequential and task-related information, and finally achieved a good recognition effect with a simple structure and fewer parameters.

As gesture recognition technology becomes more and more mature, many scholars begin to study peripheral control based on gesture recognition technology. Leonardis et al. developed the bilateral hand functional rehabilitation training equipment to detect the sEMG signals of the opposite hand with hemiplegia and classify them into patterns [14]. Then, the classification results were used as driving signals to the wearable training equipment, which drove the hemiplegic hand to complete the movement training consistent with the contralateral hand. Zainab et al. realized the teleoperated control of wheelchair by sEMG signal, and convert the ten gestures into ten modes of wheelchair control instruction by KNN and calculation of minimum Euclidean distance [15]. Ana et al. designed a hand exoskeleton driven by sEMG signals. The system provides real-time control of the hand exoskeleton by calculating muscle activation [16]. Farooq et al. realized 3D printed prosthetic hand control by using the method of traditional feature learning and SVM classifier. [17]. These results show that sEMG signals can be well applied to the control of peripherals, but to achieve more refined intelligent control, the reliability of the algorithm needs to be further improved.

III. METHOD

A. Overall design of system

We develop a manipulator control system in this paper, which is illustrated in Figure 1. The MYO armband with 200Hz sampling rate is used for sEMG signals recording, which can record 8 channels sEMG signals simultaneously. The peripheral adopts mechanism hand, which has 7 degrees of freedom and can perform forearm internal/external rotation, wrist flexion/extension, and each single finger movement. The performance of the ulnar/radius deviation completed by the combination of the internal/external rotation and wrist flexion/extension. The sEMG signals sensed from skin using MYO are sent to computer via Bluetooth. The computer will finish the preprocessing of signals before training and reasoning, such as signal segmentation. At the stage of offline training, segmented signals and corresponding gesture labels will compose samples and be fed into a neural network to train the gesture recognition model. At the stage of online recognition, the samples from the part signals in a trial will be loaded to the gesture recognition model and the decoded gesture is transferred to the mechanism hand movement instruction. In the both offline and online stages, the 1-200 Hz band pass filter and a 50-100-150Hz trap filter were used to clean sEMG signals.

B. sEMG-based gesture recognition

Recognizing gestures using sEMG is a challenging task for its non-stationarity and nonlinearity. For a system driven by sEMG-based gesture recognition, the gesture recognition algorithm is extremely important, which can greatly improve the performance of system. With the prosperity and

development of deep learning, many excellent gesture recognition algorithms have emerged. In our previous work, we propose a spatial-temporal convolutional networks for gesture recognition named STCN-GR [3], which shows great advantages in the task of sEMG-based gesture recognition. The structure of STCN-GR is illustrated in Figure 2, which is composed of 4 spatial-temporal convolution blocks (STCBs) followed by a global average pooling (GAP) layer and an optional dropout layer.

After acquiring abundant sEMG signals, sEMG samples can be obtained using sliding window strategy, which can be used to train gesture recognition model. Single sample and its corresponding gesture label are just required when training. But in the online reasoning scenario, e.g., gesture-recognition-driven system in this paper, the robustness of recognition results must be guaranteed for security and user experience. For that, majority voting is used in this paper, which can be formulated as Eq.1.

$$y = \underset{y_i \in \{1,2,3,\dots,G\}}{\operatorname{argmax}} \sum_{t=1}^T \mathbb{I}_{y_i}(\hat{y}_t) \quad (1)$$

where $\hat{y}_t$ is the t^{th} predicted gesture label, T is number of labels for majority voting, y_i is the ground truth, G is the number of gestures. $\mathbb{I}_{y_i}(\cdot)$ is an indicator function and can be formulated as Eq.2.

$$\mathbb{I}_{y_i}(\hat{y}_t) = \begin{cases} 1, & \text{if } y_i = \hat{y}_t \\ 0, & \text{if } y_i \neq \hat{y}_t \end{cases} \quad (2)$$

As an ensemble learning algorithm, majority voting can ensure the robustness of the recognition results as much as possible. After determining the final gesture recognition result y , the motion instruction of the manipulator can be obtained, which can control manipulator to complete the corresponding gestures.

IV. EXPERIMENTS

To evaluate the performance of proposed manipulator control system based on gesture recognition using sEMG, we recruited subjects and conducted experiments in online recognition and control scenario.

After loading the offline-trained model, one of the 9 gestures was randomly prompted on the screen, and a buffer time of 1s was provided for subject to make corresponding gesture. The subject completed and kept the gesture within the specified time. Once the gesture is completed, the predicted gesture can be determined by recognition model combined with majority voting. The predicted gesture was converted into the designed control instruction to control the manipulator to perform the corresponding gesture. At the same time, the cumulative recognition accuracy is calculated,

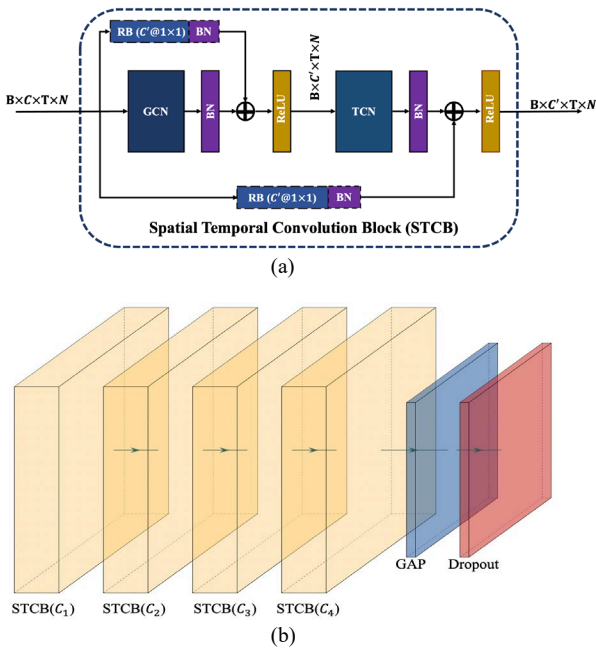


Figure 2. (a) Structure of spatial temporal convolution block (STCB). STCB is composed of graph neural network (GCN), temporal neural network (TCN), batch normalization (BN), residual block (RB) and activation function ReLU. (b) Structure of STCN-GR. STCN-GR is composed of 4 spatial-temporal convolution blocks (STCBs) followed by a global average pooling (GAP) layer and an optional dropout layer.

TABLE I. RECOGNITION ACCURACY BASED ON DIFFERENT VOTING WINDOW

Subject	Accuracy (%)		
	1s	3s	5s
1	83.00	89.00	87.00
2	91.00	94.00	92.00

and the recognition result is fed back it to subjects through the whole, for subject 1, open, rest and external rotation were

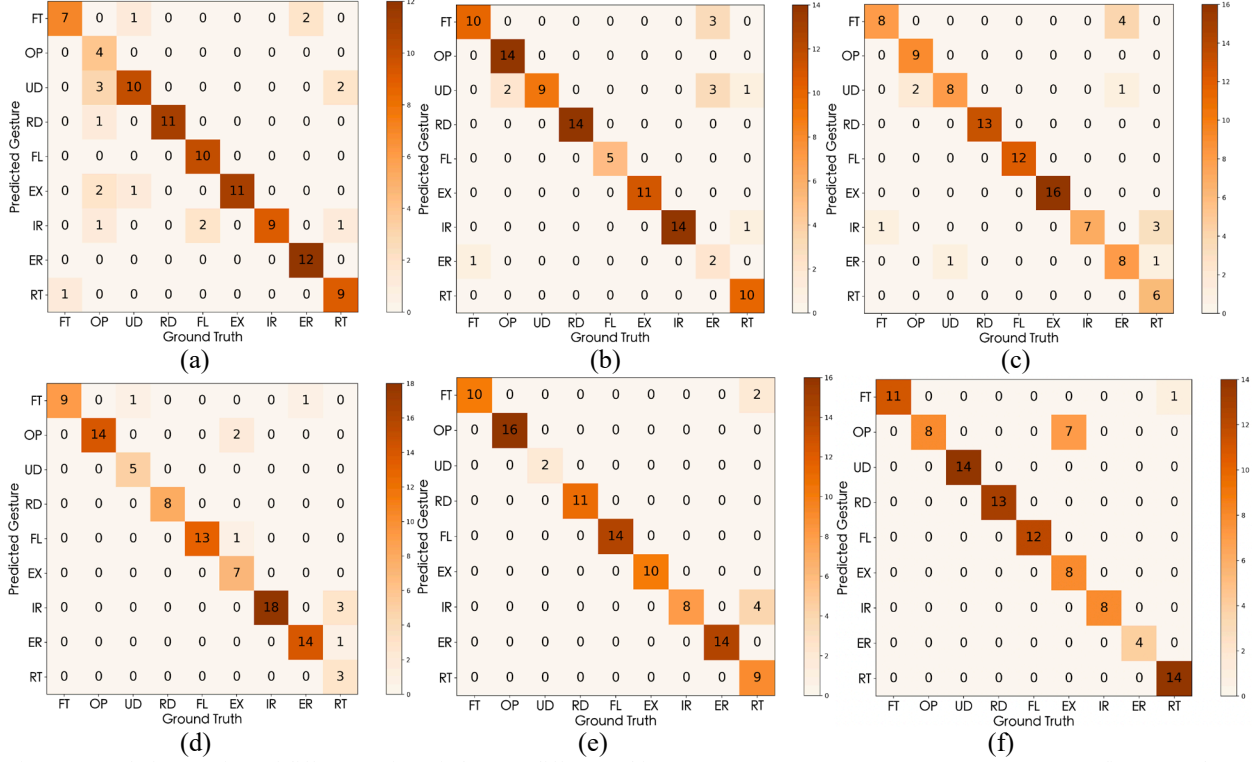


Figure 3. Confusion matrices of different voting windows on different subjects. FT, OP, UD, RD, FL, EX, IR, ER, RT represent fist, open, ulnar deviation, radial deviation, flexion, extension, internal rotation, external rotation and rest, respectively. (a)~(c) are the confusion matrices of subject 1 based on voting windows of 1s, 3s and 5s, respectively. (d)~(f) are the confusion matrices of subject 2 based on voting windows of 1s, 3s and 5s, respectively.

the screen. There will be 5s for rest after each trial. Repeat 100 times to get the final online recognition accuracy. We recruited two subjects for our experiments, and the online recognition result is showed in TABLE I.

It can be seen from TABLE I that the recognition accuracy on two subjects is above 80%, and the highest is 94%, which shows the excellent performance in the actual control environment. The performance of different voting window indicates that 3s is the most suitable voting window.

For further analysis, confusion matrices of the recognition results are drawn. The confusion matrices for different voting window are illustrated in Figure 3. For the voting window of 1s, the gestures that are hardly identified are open, flexion, external rotation and rest for subject 1, and for subject 2 are extension and rest; For the voting window of 3s, the incorrectly recognized gestures on subject 1 are mainly open and external rotation, and the incorrectly recognized gesture on subject 2 is rest; For the voting window of 5s, the incorrectly recognized gestures on subject 1 are mainly external rotation and rest, and the wrong gestures identified on subject 4 were mainly extension. On

the most easily identified wrong gestures, while for subject 4, the gestures most easily identified wrong were extension and rest. For the different habits and physiological conditions, the situation of gesture errors is not the same. Due to the data distributions are variable, it is difficult for the model to learn this part of knowledge, but it is easily found that there is a large degree of similarity between correct/wrong gesture. In addition, rest is the gesture that both subjects are prone to recognize wrong, the reason is that the amplitude and force of some gestures are small, which leads to the model overfitting the noises.

Besides recognition accuracy, the real-time performance affect significantly for a sEMG-based gesture-recognition-driven control system. The overall delay of manipulator control system is mainly composed of six parts: signal acquisition delay, signal processing delay, signal transmission delay, instruction recognition delay, instruction transmission delay and manipulator motion delay, which is illustrated in Figure 4. Signal acquisition and processing is mainly generated from acquisition device, and relatively small. The signal transmission delay is mainly related to the

TABLE II. RECOGNITION SPEED BASED ON DIFFERENT VOTING WINDOW

	100ms	200ms	250ms	300ms
Time (ms)	5.56×10^{-3}	6.43×10^{-3}	7.19×10^{-3}	7.31×10^{-3}
Frame Rate (fps)	1.80×10^5	1.55×10^5	1.39×10^5	1.37×10^5

transmission protocol, transmission method and distance. In this paper the Bluetooth wireless transmission method is adopted and the transmission distance is less than 1m, the delay of this part can be ignored. The instruction recognition delay is the main component of delay, which includes multi-parts. The instruction transmission delay comes from the delay of transmitting the instruction to the microcontroller, which can also be ignored. the manipulator control delay is the delay from receiving the control instruction to drive the manipulator to perform gestures, which is also an important part that affects the real-time performance of the system, but this delay mainly depends on the manipulator. Therefore, in the control system based on gesture recognition using sEMG, the delay caused is mainly the delay of the instruction recognition.

There are five parts in the instruction recognition delay: signal cache delay, signal segmentation delay, gesture prediction delay, majority voting delay and instruction output delay. Since the recognition accuracy is the most important factor affecting the performance of the control system, a certain amount of sEMG signals need to be cached, which will bring a large time delay. After getting a certain amount of signals, the sliding window method is used to obtain sEMG signal samples. This process needs to traverse all the data, and the delay is also important. After obtaining the sEMG sample, it needs to convert the format and input it into the gesture recognition model to obtain the predicted gesture, which is the main source of delay in the instruction recognition except for the data cache and depends on the gesture recognition algorithm, processing speed and computing power of device.

To clarify the delay of gesture recognition. We carried out experiment on a device with a CPU model of Intel i7-8700 and a frequency of 3.20GHz, and the average value of

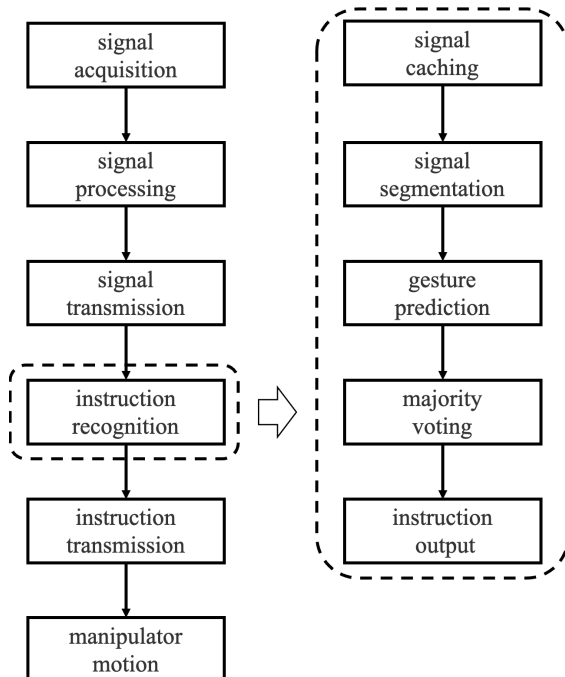


Figure 4. The component of control delay. Delay comprises of six parts, and instruction recognition delay determines the performance of whole system, especially signal caching and gesture prediction.

1,000,000 samples is calculated serially. The recognition speed of the model is shown in TABLE II.

We can see that the frame rate reaches 1.37×10^5 fps, that is, 1s can process 1.37×10^5 samples, and the operation speed is very impressive. In terms of processing time, the processing time of a single sample can be controlled at the microsecond, and the delay is extremely small. In addition, since a certain number of samples are cached, serial computing can be performed with the support of GPU, which further reduces the computing delay. The majority voting stage mainly votes on the output results of the model, and the operation delay is negligible compared with the model calculation. Similarly, the latency of instruction conversion and output is negligible.

V. CONCLUSION

In this paper, we build a teleoperated manipulator control system based on sEMG using deep learning algorithm proposed in our previous work. Combined with majority voting strategy, proposed system can realize reliable recognition results, which show the possibility of teleoperation. For a teleoperation system, real time is an important content to be considered. The analysis of delay and speed shows that the sEMG-based teleoperated control system can fully meet the real-time requirements of teleoperation. In our future work, we will concentrate on the optimization of engineering and further improve the performance of whole system.

ACKNOWLEDGMENT

This work was supported in part by the National Natural Science Foundation of China, grant no. 61904038 and no. U1913216; National Key R&D Program of China, grant no. 2021YFC0122702 and no. 2018YFC1705800; Shanghai Sailing Program, grant no. 19YF1403600; Shanghai Municipal Science and Technology Commission, grant no. 19441907600, no.19441908200, and no. 19511132000; Opening Project of Zhejiang Lab, grant no. 2021MC0AB01; Fudan University-CIOMP Joint Fund, grant no.FC2019-002; Opening Project of Shanghai Robot R&D and Transformation Functional Platform, grant no. KEH2310024; Ji Hua Laboratory, grant no. X190021TB190 and no. X190021TB193; Shanghai Municipal Science and Technology Major Project, grant no. 2021SHZDZX0103 and no. 2018SHZDZX01. ZJ Lab, and Shanghai Center for Brain Science and Brain-Inspired Technology.

REFERENCES

- [1] Harper, Lisa D. et al. "Perceptive assistive agents in team spaces." IUI '04 (2004).
- [2] Ting, Jordyn E. et al. "Sensing and decoding the neural drive to paralyzed muscles during attempted movements of a person with tetraplegia using a sleeve array." medRxiv (2021): n. pag.
- [3] Lai, Zhiping et al. "STCN-GR: Spatial-Temporal Convolutional Networks for Surface-Electromyography-Based Gesture Recognition." ICONIP (2021).
- [4] Yu D, Jin W, Wei W, et al. Surface EMG-Based Inter-Session Gesture Recognition Enhanced by Deep Domain Adaptation[J]. Sensors, 2017, 17(3):458.

- [5] Zhang, Yingwei et al. "Learning Effective Spatial–Temporal Features for sEMG Armband-Based Gesture Recognition." *IEEE Internet of Things Journal* 7 (2020): 6979-6992.
- [6] Ju, Zhaojie and Honghai Liu. "A generalised framework for analysing human hand motions based on multisensor information." 2012 IEEE International Conference on Fuzzy Systems (2012): 1-6.
- [7] Al-Timemy A H, Bugmann G, Escudero J, et al. Classification of Finger Movements for the Dexterous Hand Prosthesis Control With Surface Electromyography[J]. *IEEE Journal of Biomedical and Health Informatics*, 2013, 17(3):608-618.
- [8] Sui X, Wan K, Zhang Y. Pattern recognition of SEMG based on wavelet packet transform and improved SVM[J]. *Optik*, 2019, 176: 2
- [9] Amor, Amina Ben Haj et al. "Toward sign language handshapes recognition using Myo armband." 2017 6th International Conference on Information and Communication Technology and Accessibility (ICTA) (2017): 1-6.
- [10] Shin, Seongjoo et al. "Korean sign language recognition using EMG and IMU sensors based on group-dependent NN models." 2017 IEEE Symposium Series on Computational Intelligence (SSCI) (2017): 1-7.
- [11] Geng, Wei-dong et al. "Gesture recognition by instantaneous surface EMG images." *Scientific Reports* 6 (2016): n. pag.
- [12] Rahimian, Elahe et al. "XceptionTime: Independent Time-Window Xceptiontime Architecture for Hand Gesture Classification." *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (2020): 1304-1308.
- [13] Tsinganos, Panagiotis et al. "Improved Gesture Recognition Based on sEMG Signals and TCN." *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (2019): 1169-1173.
- [14] Leonardis D, Barsotti M, et al. An EMG-controlled robotic hand exoskeleton for bilateral rehabilitation[J]. *IEEE Transactions on Haptics*, 2015, 8(2): 140-151.
- [15] Alibhai, Zainab et al. "A Human-Computer Interface For Smart Wheelchair Control Using Forearm EMG Signals." 2020 3rd International Conference on Data Intelligence and Security (ICDIS) (2020): 34-39.
- [16] Cignal, Ana et al. "RobHand: A Hand Exoskeleton With Real-Time EMG-Driven Embedded Control. Quantifying Hand Gesture Recognition Delays for Bilateral Rehabilitation." *IEEE Access* 9 (2021): 137809-137823.
- [17] Farooq, Umar et al. "EMG control of a 3D printed myo electric prosthetic hand." *IOP Conference Series: Materials Science and Engineering* (2019): n. pag.