Brain–Computer Interface Integrated With Augmented Reality for Human–Robot Interaction

Bin Fang, Wenlong Ding, Fuchun Sun, *Fellow, IEEE*, Jianhua Shan, Xiaojia Wang, Chengyin Wang, and Xinyu Zhang, *Member, IEEE*

***Abstract*—Brain–computer interface (BCI) has been gradually used in human–robot interaction systems. Steady-state visual evoked potential (SSVEP) as a paradigm of electroencephalography (EEG) has attracted more attention in the BCI system research due to its stability and efficiency. However, an independent monitor is needed in the traditional SSVEP-BCI system to display stimulus targets, and the stimulus targets map fixedly to some preset commands. These limit the development of the SSVEP-BCI application system in complex and changeable scenarios. In this study, the SSVEP-BCI system integrated with augmented reality (AR) is proposed. Furthermore, a stimulation interface is made by merging the visual information of the objects with stimulus targets, which can update the mapping relationship between stimulus targets and objects automatically to adapt to the change of the objects in the workspace. During the online experiment of the AR-based SSVEP-BCI cue-guided task with the robotic arm, the success rate of grasping is 87.50 ± 3.10% with the SSVEP-EEG data recognition time of 0.5 s based on FB-tCNN. The proposed AR-based SSVEP-BCI system enables the users to select intention targets more ecologically and can grasp more kinds of different objects with a limited number of stimulus targets, resulting in the potential to be used in complex and changeable scenarios.**

***Index Terms*—Augmented reality (AR), brain–computer interface (BCI) system, FB-tCNN, human–robot interaction, steady-state visual evoked potential (SSVEP), stimulation interface, visual information.**

I. Introduction

SOME patients have difficulty in pronunciation and physical movement. However, these patients can still generate electrical signals from their brains that can be detected by a brain–computer interface (BCI). BCI can realize information communication between the human brain and external devices [1], [2]. It can help patients communicate with the outside world and improve the quality of their lives, thus reducing the burden on patients' families and nursing staff.

Although the signal-to-noise ratio (SNR) of noninvasive BCI is lower than invasive BCI, noninvasive BCI has a greater application value because of its simple operation and more secure and reliable [3]–[7]. Electroencephalography (EEG) is a type of brain electrical signal used commonly in noninvasive BCI [8]–[12]. As a paradigm of EEG, steady-state visual evoked potential (SSVEP) is relatively stable and reliable since its frequency-domain characteristics are evident and easy to identify. Several stimulus target flicker at different fixed frequencies to induce specific SSVEP signals. By analyzing the induced SSVEP signals, which stimulus target the users are focusing on can be determined. SSVEP has proved to be the most promising brain model among the BCI application system [13]. This article mainly discusses the SSVEP-BCI application scenario in which the users control the robotic arm to grasp the target object.

An independent monitor is used to display stimulus targets in the traditional SSVEP-based BCI application system. However, the independent monitor is not only inconvenient to carry but also difficult to adapt to complex realistic scenarios. In addition, stimulus targets are difficult to present well on a small monitor [19]. Furthermore, in the first perspective of users, the stimulus targets on the independent monitor are always not in the same view field as the target objects in the workspace. On the one hand, users need to stare at stimulus targets on the independent monitor to generate intention commands. On the other hand, they need to look at the robotic arm to make sure whether it grasps the intended object. The users had to switch their eyes frequently between the independent monitor and the robotic arm, which might make the users feel tired. Therefore, the independent monitor is not conducive to the SSVEP-BCI application in human–robot interaction.

Augmented reality (AR) device used to display stimulus targets has the potential to alleviate the problem of attention shift. AR can be divided into video see-through (VST) AR and optical see-through (OST) AR [14], [15]. The VST-AR

Manuscript received 30 November 2021; revised 7 June 2022; accepted 21 July 2022. Date of publication 28 July 2022; date of current version 11 December 2023. This work was supported in part by the National Natural Science Foundation of China under Grant 91848206 and Grant 62173197, and in part by the Natural Science Foundation of Anhui Province under Grant 2108085MF224. *(Bin Fang and Wenlong Ding contributed equally to this work.) (Corresponding authors: Fuchun Sun; Jianhua Shan.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Local Ethics Committee at the Department of Psychology, Tsinghua University under Application No. IRB202138.

Bin Fang and Fuchun Sun are with the Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China (e-mail: fcsun@tsinghua.edu.cn).

Wenlong Ding, Jianhua Shan, and Chengyin Wang are with the Department of Mechanical Engineering, Anhui University of Technology, Ma'anshan 243032, Anhui, China (e-mail: shanjianhua.vip@qq.com).

Xiaojia Wang is with the Department of Electrical and Computer Engineering, Clemson University, Clemson, SC 29634 USA.

Xinyu Zhang is with the State Key Laboratory of Automotive Safety and Energy, Tsinghua University, Beijing 100084, China.

This article has supplementary material provided by the authors and color versions of one or more figures available at https://doi.org/10.1109/TCDS.2022.3194603.

Digital Object Identifier 10.1109/TCDS.2022.3194603

system is a method to acquire real images via a camera or computer and combine them with virtual images for real-time display on the monitor. Head-mounted display (HMD) is the preferred approach for OST-AR technology, which merges the real with the virtual by creating a scene generator that displays translucent virtual video or images. The working principle of OST-HMD is to place the optical combinator in front of the users' eyes [16] so that the users can see the real world and translucent virtual video or image at the same time in the first perspective. OST-AR technology can overcome the limitation of the independent monitor and improve the flexibility of the SSVEP-BCI application. Using AR to replace the independent monitor not only improves the portability and adaptability of the SSVEP-BCI application system in complex environments but also avoids the distraction and fatigue caused by attention shift. Therefore, we propose an SSVEP-BCI online control system based on OST-HMD-AR.

In recent years, some researchers have tried to combine OST-AR with the SSVEP-BCI system [17]–[22]. Si-Mohammed *et al.* [17] mainly discussed the feasibility of the AR-based SSVEP-BCI in the application. The effect of stimulus target layout in the AR-SSVEP is discussed in [18]. The application of the AR-based SSVEP-BCI system is studied, respectively, in [19]–[22]. Ke *et al.* [19] proposed a BCI system based on high-speed online SSVEP-recognition in an OST-AR environment. The data length of the recognition time window is 1 s in the online experiment. Their results show that the AR-SSVEP has the potential to provide a high-performance BCI system. Park *et al.* [20] implemented a new control system for home appliances by AR-based SSVEP-BCI. A four-target SSVEP-BCI system is constructed to turn on or off the home appliances, and a 2.6-s data length of the recognition time window is used. Chen *et al.* [21] designed a robotic arm control system by merging AR with SVEP-BCI. In their study, a four-target SSVEP-BCI system is constructed to enable users to control the robotic arm to grasp specific objects. The data length of the recognition time window in the online experiment is 2.6 s. In the study of Wang *et al.* [22], AR glasses are used as visual stimulators in the BCI system to control the humanoid robot to grasp objects.

In the above studies, the stimulus targets map fixedly to some preset target commands. For example, in the study of Chen *et al.* [21], three stimulus targets map fixedly to three target objects so that they could only grasp the three objects but no other objects. Therefore, we propose a simulation interface that merges the visual information of the objects with the stimulus targets. The mapping relationship between stimulus targets and objects can be updated automatically to adapt to the changes of objects in the workspace, which allows the user to control the robotic arm to grasp more kinds of objects with a limited number of stimulus targets. In addition, the proposed stimulation interface makes the mapping relationship between stimulus targets and objects more clear, which can also make the users select the intended objects more ecologically.

It will undoubtedly cause visual fatigue when the stimulus targets displayed by AR glasses flicker all the time. Therefore, the stimulus targets will stop flickering after the intention is predicted by the short time-window classification algorithm FB-tCNN [23]. Besides, the stimulation interface will be turned off after the prediction check stage. These can reduce the time the stimulus targets are presented in the users' view field to relieve the fatigue.

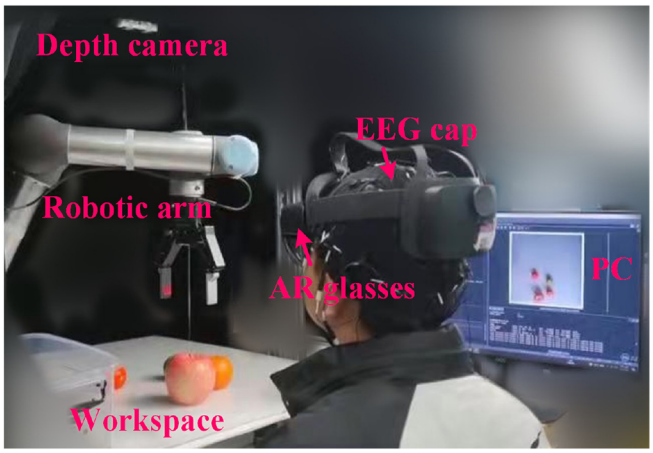

Fig. 1. Experimental setup of the proposed AR-based SSVEP-BCI-controlled robotic arm.

The main contributions of this article can be summarized as follows.

1) We propose an online SSVEP-BCI control system with a robotic arm based on AR, which improves the portability of the SSVEP-BCI application system and avoids attention shift. It has made contributions to the AR-based SSVEP-BCI field.
2) We propose a stimulation interface that merges the visual information of the objects with stimulus targets. Users can choose the target object more ecologically and control the robotic arm to grasp more kinds of objects with a limited number of stimulus targets.
3) Reducing the time the users stare at the stimulation interface to relieve users' fatigue. The FB-tCNN algorithm used in the proposed AR-based SSVEP-BCI system can realize an effective recognition for SSVEP-EEG data in a short time window (such as 0.5 s). Moreover, the stimulation interface displayed by AR will be turned off after the prediction check stage.

The structure of this article is as follows. Section I gives a brief introduction to the AR-based SSVEP-BCI system. Section II presents the proposed methods. Section III shows the results. In Section IV, we discuss the potential and advantage of the proposed AR-based SSVEP-BCI system and put forward future work. In Section V, we make a summary.

II. Methods

A. System Description

The proposed AR-based SSVEP-BCI control system consists of AR glasses (Hololens 2), EEG acquisition equipment (Neuroscan), a depth camera (Realsense D455), a robotic arm (UR 5 and Robotiq), and a personal computer (PC). The experimental setup is shown in Fig. 1. The AR glasses display the stimulation interface, the EEG acquisition equipment acquires the EEG data, the depth camera acquires the visual information

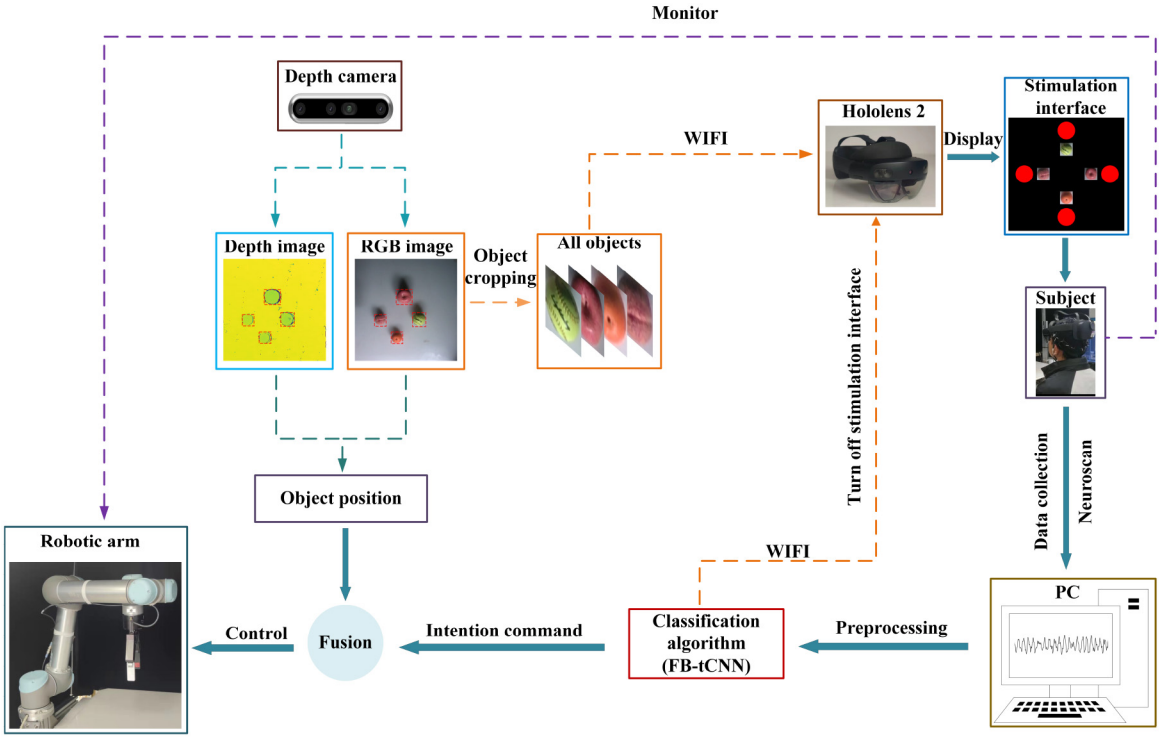

Fig. 2. System architecture of the proposed AR-based SSVEP-BCI-controlled robotic arm.

of objects in the workspace (depth image and RGB image), the robotic arm is used to grasp and transfer the target object, and the PC is used for processing the acquired EEG data and image data. The system architecture is shown in Fig. 2.

First, the PC obtains the visual information of the objects through the depth camera, then crops out all objects in the RGB image and sends them to AR glasses through WIFI (TCP/IP). After these cropped images are sent, a "start" command is sent to AR glasses via WIFI to make AR glasses turn on the stimulation interface. These cropped images are placed at a specific position in a specific order next to the stimulus target on the stimulation interface. Each stimulus target flickers at a fixed frequency, which is different from the other. The SSVEP-EEG data are induced by staring at a stimulus target next to the target object image. The EEG data are collected by the EEG acquisition equipment and processed by the PC. A deep learning-based classification algorithm FB-tCNN is used for the collected EEG data and to predict the intention of the users. Then, the predicted intention command is sent to AR glasses, and the object image corresponding to the intention command would be placed at the center of the stimulation interface to make users check whether the predicted intention is consistent with the real intention. After that, a "stop" command is sent to AR glasses through WIFI to make AR glasses turn off the stimulation interface. Which object is the users' target can be known by the predicted intention, and the position of the target object in the real world can be obtained by the depth image, RGB image, internal parameter, and external parameter (the specific calculation process will not be presented in this article). Finally, the position of the target object is transmitted to the robot controller through wire transmission (TCP/IP) communication technology, which can control the robotic arm to grasp the target object and transfer it to a specific area. Users can monitor the robotic arm to prevent an unsafe trajectory of the robotic arm and check whether the robotic arm grasps the target object.

In the proposed AR-based SSVEP-BCI system, the users need only to stare at the stimulus target next to the cropped image of the target object to grasp the target object. Users can choose the objects they intend to grasp more ecologically. The intention command represented by the stimulus target corresponds to the cropped object image beside it. Instead of a fixed mapping relationship between the stimulus target and the predetermined limited target object, the proposed stimulation interface can update the mapping relationship between stimulus targets and objects automatically to adapt to the change of the objects in the workspace. Therefore, the robotic arm can grasp more kinds of objects (such as six-class objects) with fewer stimulus targets (such as four stimulus targets) in the proposed system. To reduce users' fatigue and create a more user-friendly BCI experience, the stimulation interface displayed by AR glasses is not always present in the users' first perspective. The stimulation interface will be turned off when the predicted intention check is completed. It can reduce the time the users spend staring at the flickering stimulus targets.

B. Merge Process of Visual Information and Stimulus Targets

The RGB and depth images are captured by the depth camera located above the workspace. The pixel size of the RGB and depth images is 720 × 480. These two images are cropped to make only the workspace emerge in the two

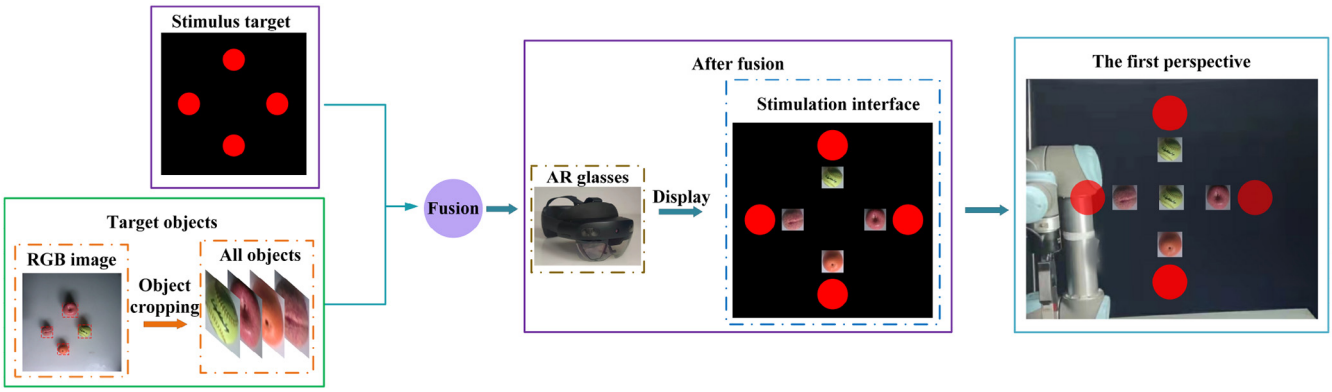

Fig. 3. Merge process of visual information and stimulus targets.

images to prevent other irrelevant objects out of the workspace from interfering with the object recognition. The pixel size of cropped RGB and depth images is 360 × 360. Mask-RCNN [32] model is used for target recognition of the objects in the RGB image. Each object is recognized by Mask-RCNN and boxed with a rectangle, and then cropped out from the RGB image along with the rectangular box. After that, all cropped object images are sent to AR glasses through WIFI and then placed in the specified position (up, right, down, and left) in a predetermined sequence on the stimulation interface. The merge process of cropped object images and stimulus targets is shown in Fig. 3, the first perspective is in the prediction check stage. The merging stimulation interface contained the visual information of the target objects and can make users choose target objects more ecologically.

C. Subjects

Six healthy subjects (ages 22–28, all male) participated in the experiment. All subjects participated in both offline and online sessions of the experiment. Four subjects had never used BCI, while the others had experience with BCI experiments. The study was approved by the local Ethics Committee at the Department of Psychology, Tsinghua University (IRB202138).

D. Paradigm of Stimulus Target

Taking into account the contrast with the environment, the color of the stimulus targets is chosen as red [19]. Referring to [24]–[26], we adopt a sampled sinusoidal stimulation method to present flickering stimulus targets. The red-color stimulus target brightness $s(n, f_i, \varphi)$ is expressed as

$$s(n, f_i, \varphi) = \left\{1 + \sin\left[2\pi f_i (n/R_s) + \varphi\right]\right\}/2 \tag{1}$$

where sin() generates a sine wave, n indicates the frame index in the sequence, R_s indicates the screen refresh rate, φ indicates the phase, and f_i indicates the stimulus frequency of the ith target, $i = 1, 2, 3, 4$.

The frequencies of the four SSVEP targets are designed to flicker at 9, 11, 13, and 15 Hz. The phases of the four SSVEP targets are designed to flicker at 0, 0.5π, π, and 1.5π. The four SSVEP stimulus targets are placed at four positions (up, right, down, and left) on the stimulation interface.

E. Experiments of AR-Based SSVEP-BCI Robotic Arm Control System

The experiments are divided into offline session and online session. The two sessions are the two experiments on different days for the same person.

1) Offline Session: The offline session is mainly designed to collect the AR-based and PC-based SSVEP-EEG data. The purposes of this experiment are: 1) to optimize the parameters of the AR-based SSVEP-BCI control system for the online session, such as the data length of the recognition time window; 2) to make the subjects familiar with the AR-SSVEP experiment (especially familiar with the experimental paradigm and the usage of AR glasses); and 3) to compare AR-SSVEP with PC-SSVEP.

In the offline session, the stimulation interface is displayed by AR glasses (developed in C# under Unity 3-D) and PC (developed in Python) in the AR-SSVEP experiment and PC-SSVEP experiment, respectively. To better explain the experimental design of AR-SSVEP, the experimental design of PC-SSVEP is introduced first.

During the PC-SSVEP experiment, the subjects are asked to sit in front of the independent monitor. The experimental design of PC-SSVEP is shown in Fig. 4. Each trial began with a visual cue for 2 s (preparation), prompting the location of the stimulus target that the subjects need to focus on. Then, each red-color stimulus target flickers for 4 s (stimulation), followed by a rest period lasting for 4 s. At the visual cue stage, the stimulus target at the specified location would present in white to facilitate the subjects to find it. Each stimulus frequency presents five times randomly in one block, there is a total of six blocks for the experiment. Therefore, the collected EEG data has 120 trials (4 classes × 5 times × 6 blocks). The subject would get enough time for rest after each block.

The experiment design of AR-SSVEP is the almost same as PC-SSVEP as shown in Fig. 4. The difference is the flickering stimulus targets are displayed by AR glasses and the cues of the preparation and rest stage are still displayed by the independent monitor. Subjects need to adjust the AR glasses to ensure the four stimulus targets in the first perspective are in the independent monitor and should focus on the specified stimulus target displayed by AR glasses in

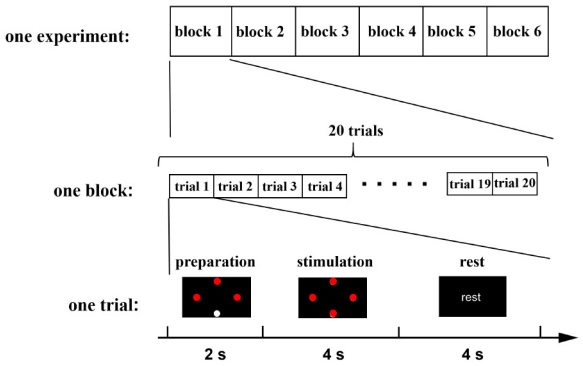

Fig. 4. Experimental design of the offline session.

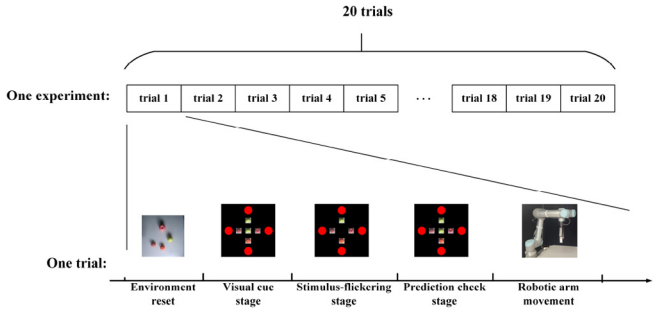

Fig. 5. Experimental design of the evaluation phase in the online session.

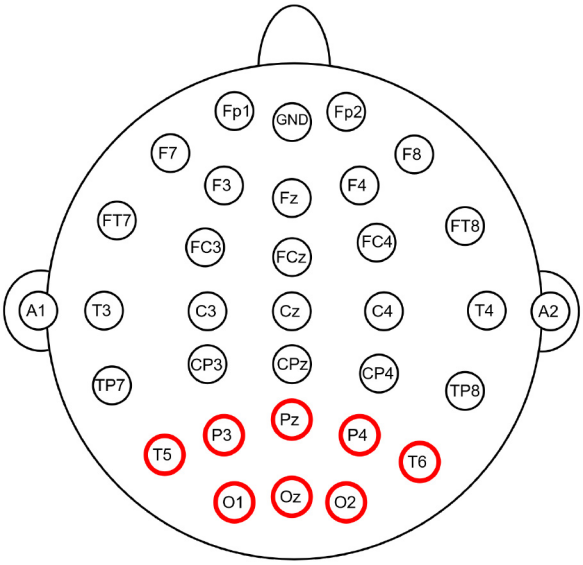

Fig. 6. Position distribution of the selected electrodes on the EEG cap.

the stimulation stage according to the cue displayed by the independent monitor in the preparation stage.

2) Online Session: In the online session, the subjects need to control the robotic arm to grasp the target object via the proposed AR-based SSVEP-BCI system during a cue-guided task. The purpose of the online session is to evaluate the feasibility of the proposed AR-based SSVEP-BCI system. The online session is divided into the training phase and the evaluation phase. According to the results analysis of the offline session, the data length of the recognition time window is determined to be 0.5 s.

Training Phase: The experimental design of the training phase is the same as the AR-SSVEP experiment of the offline session (Fig. 4), which mainly provides training data for the classification algorithm FB-tCNN.

Evaluation Phase: To facilitate the evaluation of system performance, a visual cue is added in the evaluation phase. One of these objects on the workspace is randomly selected as the target object, and its cropped image is placed at the center of the stimulation interface as a visual cue. The subjects are required to find the cropped image, which is the same as the visual cue, and then stare at the stimulus target next to it.

The experimental design of the evaluation phase is shown in Fig. 5. First, four different objects are randomly placed in the workspace. The depth camera then acquires the visual information about the workspace, the images of these objects in the workspace are cropped out from the RGB image and sent to the AR glasses. Then, the AR glasses display the stimulation interface that merges the stimulus targets with the cropped object images, the stage of the visual cue is started and lasted for 4 s. The subjects should focus on the stimulus target next to the peripheral cropped image, which is consistent with the cue (the object image at the center). Then, the stimulus-flickering stage is started, and each stimulus target begins to flicker. After the intention of the subjects is predicted, the stimulus targets stop flickering. The predicted intention would be presented as an object image at the center of the stimulation interface to make the subjects check the predicted intention in the prediction check stage which lasted for 2 s. After the predicted intention is checked, the stimulation interface will be turned off. Finally, the robotic arm performs the corresponding grasping operation according to the predicted intention command in the robotic arm movement stage.

After each trial of grasp, a new object needs to be randomly added to the workspace within 10 s to ensure there are four objects in the workspace. A total of 20 trials of grasping tasks in the evaluation phase. There are six different objects in the experiment. To reduce unnecessary computation, the EEG data are only processed for classification recognition while the stimulation interface is turned on.

F. EEG Data Acquisition

The EEG acquisition device is used to record EEG data at a sampling rate of 1000 Hz. Considering simplicity and SNR, the EEG data of only eight electrodes (O1, O2, Oz, P3, P4, Pz, T5, and T6) close to the occipital region are recorded [27]–[29]. The reference electrode is Cz. Electrode impedances are kept below 10 kΩ during recording. The channel configuration of the International 10–20 system. The distribution of these electrodes on the EEG cap is shown in Fig. 6, the selected electrodes are marked with red.

G. Preprocessing of EEG Data

To reduce the calculation, the EEG data of these electrode channels are downsampled to 250 Hz. Referring to our earlier work on the algorithm FB-tCNN [23], three subfilters with different bandpass ranges are designed as 7–17, 16–32, and

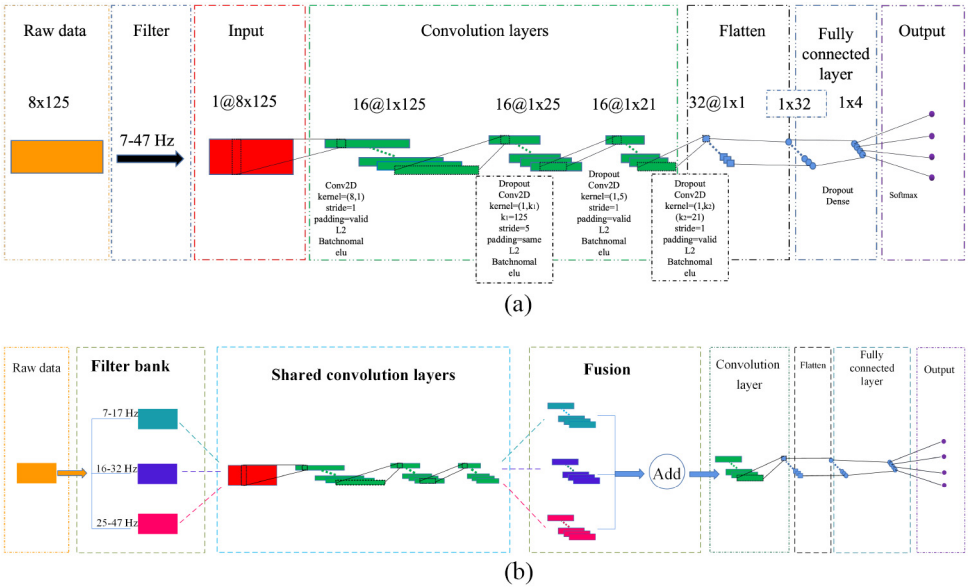

Fig. 7. (a) Architecture framework of tCNN, "Conv2D" indicates 2-D convolution, "kernel" indicates convolution kernel size, "stride" indicates convolution operation step size, "Batchnormal" indicates batch normalization, "elu" indicates "ELU" activation function, and "L2" indicates "L2 regularization." (b) Architecture framework of FB-tCNN, the network hyperparameters are the same as tCNN, which will not be described here.

25–47 Hz, respectively. These three subfilters make up the filter bank. Considering a latency in the visual system [30], the data range used in each trial of the offline session and the training phase of the online session is $[0.14, 0.14 + L]$ s, which starts 0.14 s after the onset time of the stimulus, and L s is the lasted time of the stimulus.

H. SSVEP Classification by FB-tCNN

To better explain the network architecture of FB-tCNN, the network architecture of tCNN is introduced first. The 0.5-s time window is taken as an example of illustration.

1) Network Architecture of tCNN: The network architecture of the tCNN is shown in Fig. 7(a). The input data are the 0.5-s time window of eight electrode channels, and the sampling frequency is moderated to 250 Hz by downsampling. Therefore, the input size is 8×125 (be reshaped into $1@8 \times 125$). The convolution kernel size of the first convolution is 8×1, the step size is 1 without padding, the number of output channels is 16, and the output is $16@1 \times 125$. The convolution kernel of the second convolution is $1 \times k_1$, k_1 is the time-frame length of the input (here k_1 is 125), the step size is 5 with padding, the number of output channels is 16, and the output is $16@1 \times 25$. The convolution kernel of the third convolution is 1×5, the step size is 1 without padding, the number of output channels is 16, and the output is $16@1 \times 21$. The convolution kernel of the fourth convolution is $1 \times k_2$ ($1 \times k_2$ is the size of a channel of the front layer network, here is 1×21), the step size is 1 with padding, the number of output channels is 32, and the output is $32@1 \times 1$. Finally, through the flatten and fully connected layers, we can obtain the scores of four classes through softmax and the class with the highest score is considered the prediction class.

The L2 regularization and dropout parameters used in this study are set to 0.01 and 0.4, respectively.

2) Network Architecture of FB-tCNN: The network architecture of the FB-tCNN is shown in Fig. 7(b). For the network hyperparameters of FB-tCNN are the same as tCNN, here we only illustrate the network architecture of FB-tCNN simply. The raw data are filtered by three subfilters with different filtering ranges in the filter bank module to obtain three subinputs. These three subinputs are, respectively, passed through three convolution layers to obtain three subfeatures. These three subinputs' three convolution layers share weights. Next, these three subfeatures are fused ([]added) in the fusion module. Then, pass through the fourth convolution layer. After that, the feature dimension is reduced by flattening. Finally, through a fully connected layer, the output is obtained via softmax. For more details, refer to [23].

I. Performance Evaluation

We analyze the amplitude spectrum and SNR in the frequency domain of SSVEP-EEG data in the offline analysis. The fast Fourier transform (FFT) is used to calculate amplitude spectrum $y(f)$, where f is a certain frequency in the frequency domain. The data length used in the offline analysis is 4 s. Thus, the frequency resolution of FFT is 0.25 Hz. Referring to Chen *et al.* [21], the SNR can be expressed as

$$\mathrm{SNR}(f) = 20\log_{10}\frac{10 \times y(f)}{\sum_{q=1}^{5}\left[y(f - 0.25 \times q) + y(f + 0.25 \times q)\right]}. \tag{2}$$

The other two parameters used to evaluate the performance of the SSVEP-BCI system are accuracy and information transfer rate (ITR) [31].

In most studies, accuracy has been defined as the ratio of the number of correct samples to the number of total samples, and the accuracy P is expressed as

$$P = m_1/m \tag{3}$$

where m_1 represents the number of correct samples, and m represents the number of total samples. The accuracy can be used to evaluate the performance of the classification algorithm in the SSVEP-BCI.

In the offline session, leave-one-out cross-validation is used. Using each of the six blocks as the testing set, and the remaining five blocks as training data. The training set and validation set are randomly divided by the ratio of 9:1 in the training data. During the test process, P indicates the test accuracy, m_1 indicates the number of correctly predicted samples, and m indicates the total number of samples (1000 samples). The test process repeats six times, and the classification accuracy is taken as the average of the six test accuracies. In the online session, all the collected data in the training phase is set as the training data to train the FB-tCNN model which is used in the evaluation phase. P indicates the grasping success rate, m_1 indicates the number of trials that the object is grasped successfully, and m indicates the total number of trials (20 trials).

Although accuracy is an important index, the ITR should be also considered in the BCI application. ITR considers the tradeoff between accuracy and data length. The ITR (bits/min) can be written as follows:

$$\mathrm{ITR} = 60 \times \left[\log_2 K + P\log_2 P + (1-P)\log_2 \frac{1-P}{K-1}\right]/d \quad (4)$$

where K indicates the number of stimulus targets, P indicates the accuracy, and d indicates the data length of the recognition time window both in the offline and online analyses. In the online system, d is generally set to the sum of stimulus time and task execution time. However, the robotic arm moves slowly in the evaluation phase of the online session due to the safety factor, and it takes about 25 s for a complete trial in the evaluation phase. In addition, the AR glasses need a random delay time (1–2 s) to turn on the stimulation interface due to the limitation of AR glasses equipment performance. Therefore, only the data length of the recognition time window is taken as the d in the evaluation phase, regardless of the time of AR glasses delay and robotic arm movement.

III. Results

A. Results of Offline Session

In this section, offline analysis is conducted to study the difference between AR-based and PC-based SSVEP-EEG data.

1) Amplitude Spectrum and SNR in Frequency Domain: The average amplitude spectrum and SNR of each stimulus target frequency across all trials, channels, and subjects in the offline session are shown in Figs. 8 and 9, respectively. Amplitudes spectrum and SNR spikes in the frequency domain at all stimulus target frequencies of PC-SSVEP and AR-SSVEP are visible. The amplitude spectrum and SNR spikes of AR-SSVEP at the target frequency are smaller than those of PC-SSVEP, indicating that the task-related features in the frequency domain of AR-SSVEP are weaker than PC-SSVEP.

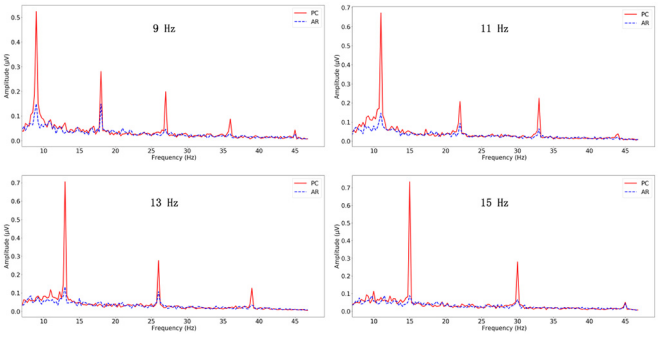

Fig. 8. Amplitude spectrum of AR-based and PC-based SSVEP-EEG data averaged across all subjects for each stimulus target.

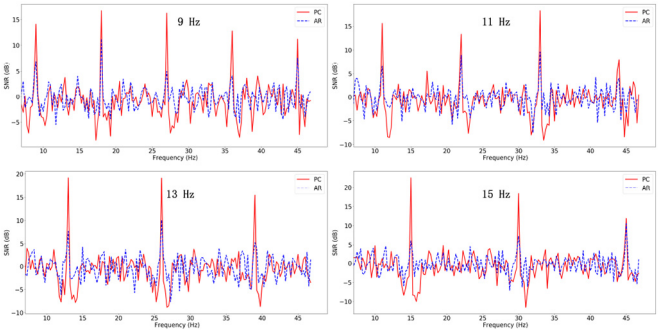

Fig. 9. SNR of AR-based and PC-based SSVEP-EEG data averaged across all subjects for each stimulus target.

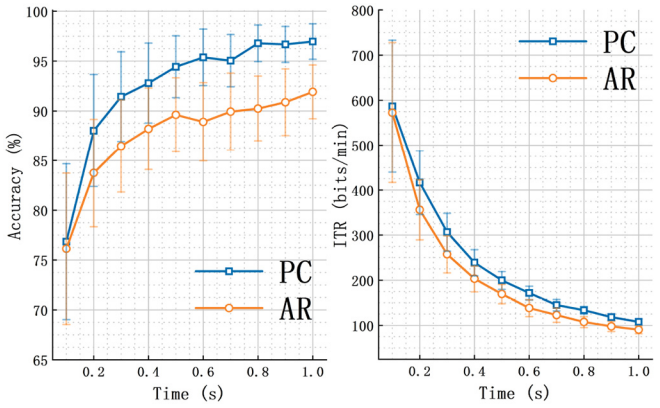

Fig. 10. Average classification accuracy and ITR of AR-SSVEP and PC-SSVEP at different data lengths. The error bars denote SEM (the standard error of the mean).

2) Average Classification Accuracy and ITR: Fig. 10 illustrates the average classification accuracy and ITR of PC-SSVEP and AR-SSVEP across all subjects in the offline session yielded by FB-tCNN at ten data lengths, which ranged from 0.1 to 1.0 s with an interval of 0.1 s. The accuracy of PC-SSVEP and AR-SSVEP is generally positively correlated with the data length. However, the improvement in accuracy is not obvious after 0.5 s. The paired t-test (tails = 2) is used to evaluate the significant difference between PC-SSVEP and AR-SSVEP, showing that there is no significant difference in average classification accuracy and ITR between PC-SSVEP and AR-SSVEP at all data lengths (all $p > 0.05$).

TABLE I
SUCCESS RATE (%) AND ITR CROSS ALL SUBJECTS AT 0.5 S IN THE EVALUATION PHASE

Subject	Success rate	ITR
1	100	240
2	85	138.29
3	80	115.33
4	80	115.33
5	90	164.70
6	90	164.70
Mean ± SEM	**87.50 ± 3.10**	**159.40 ± 19.00**

B. Results of Online Session

Considering the tradeoff between the classification accuracy and the time the subjects stare at the flickering stimulus target, 0.5 s is selected as the data length in the evaluation phase of the online experiment. Table I shows the results of the evaluation phase. In the AR-based SSVEP-BCI cue-guided task with the robotic arm, the grasping success rate across all subjects is $87.50 \pm 3.10\%$, and the ITR reaches 159.40 ± 19.00 bits/min.

IV. DISCUSSION

A. Combination of AR and SSVEP-BCI

In this section, we mainly discuss the feasibility of the combination of AR devices and SSVEP-BCI. We first analyze the results of the offline session to show the application potential of AR-based SSVEP-EEG data and then analyze the feasibility of the AR devices used in the SSVEP-BCI system.

1) Application Potential of the AR-Based SSVEP-EEG Data: As shown in Figs. 8 and 9, both the amplitude spectrum and SNR of AR-based SSVEP-EEG data for each stimulus target produce obvious spikes at the fundamental and harmonic frequencies in the frequency domain, which indicates that AR-based stimulus targets can also induce SSVEP components with obvious fundamental and harmonic signals. As shown in Fig. 10, the classification accuracy of AR-SSVEP is lower than that of PC-SSVEP, but the difference is not large (the gap is within 7%). In fact, AR-SSVEP can also achieve high classification accuracy (more than 90% after 0.7 s). It can be concluded that AR-based SSVEP-EEG data have application potential. Similar conclusions are presented in other studies [18], [19], there are more subjects in their experiments to verify the conclusions.

2) Advantage and Feasibility of AR Devices: The stimulation interface displayed by AR glasses enables the subjects to observe both the workspace and the stimulation interface in the same field of view (see the first perspective in Fig. 3), which makes the users rarely switch their attention between the workspace and the stimulation interface during the task execution. This can relieve the fatigue caused by the subjects' frequent attention switching to a certain extent and make the subjects concentrate more. During the evaluation phase of the online session, all the subjects completed the experiment well, and the proposed AR-based system exhibited relative stability. Therefore, the usage of AR devices in the SSVEP-BCI application system is feasible.

B. Combination of Visual Information and Stimulus Targets

In the traditional SSVEP-BCI system, each stimulus target maps fixedly to a preset intention command. Users can generate only a fixed number of certain preset intention commands. Taking the grasping experiment as an example, different stimulus targets in the traditional SSVEP-BCI system correspond to some specific objects in a fixed mapping way. Such a system has the limitation of object category and quantity, and the process of object selection is not ecological.

A solution is to add the stimulus targets with different flickering frequencies to all objects after the objects are recognized by the image target recognition algorithm (such as Mask-RCNN), which allows users to choose the target object more ecologically in the SSVEP-BCI system. However, it is difficult to realize in the OST-AR-based SSVEP-BCI system. Therefore, the opposite approach is taken. Instead of placing the flickering stimulus targets on the objects in the image, the images of all recognized objects are cropped and placed next to the stimulus targets on the stimulation interface. To grasp a target object, users just need to stare at the flickering stimulus target next to the cropped target object image. The stimulation interface that merges the visual information of objects with stimulus targets can make the users choose the target object more ecologically than the traditional stimulation interface in the SSVEP-BCI system. The proposed SSVEP-BCI system can be applied in complex and changeable scenarios and is not limited to the fixed object category and quantity because the merging stimulation interface can update the mapping relationship between the objects and stimulus targets automatically. The proposed merging stimulation interface contributes to the SSVEP-BCI application in life.

C. Future Work

1) Study on Multimodal BCI Control System: There are some other bioelectrical signals that can be used in the human–robot control system, such as electrooculogram (EOG) [33], [34] and surface electromyogram (sEMG) [35], [36]. Introducing other bioelectrical signals to realize a multimodal BCI control system can further facilitate the application of the SSVEP-BCI in life. We intend to introduce the EOG signal as the starting signal of the SSVEP-BCI control system. When the users observe the objects they need in the workspace, the proposed AR-based SSVEP-BCI control system can be started through EOG. During the predicted intention check stage, when the users note that the predicted intention command is inconsistent with their real intention, they can go back to the stimulus-flickering stage through EOG to reselect the target object.

2) Study on High-Frequency AR-SSVEP: Some methods are adopted in this study to reduce the time for the subjects to be stimulated by the stimulus targets, but the low-frequency stimulus target in the long-term experiment still inevitably causes the subjects to feel tired. Furthermore, the low-frequency stimulus target has the risk of inducing light-sensitive epilepsy. The problem can be alleviated by making the stimulus target flicker with high frequency. However, SSVEP signal intensity induced by a high-frequency stimulus

target is weaker than that induced by a low-frequency stimulus target. Thus, the classification accuracy is lower for high-frequency SSVEP-EEG data. Therefore, developing a new method to recognize high-frequency SSVEP is necessary.

V. Conclusion

In this study, the novel stimulation interface of the AR-based SSVEP-BCI system for the human–robot interaction was proposed, which merges the visual information of the objects with stimulus targets. We first conducted a comparative experiment between PC-SSVEP and AR-SSVEP in the offline session. The results of the offline session showed that the AR-SSVEP is potential in the SSVEP-BCI application system. Then, a cue-guided task with a robotic arm based on the proposed AR-based SSVEP-BCI system was conducted in the online session, all subjects completed the task well. Next, we discussed the feasibility of the proposed AR-based SSVEP-BCI control system. The proposed AR-based merging stimulation interface can make users select objects more ecologically and adapt to complex and changeable scenarios. Finally, we put forward future work.

References

[1] J. J. Vidal, "Toward direct brain-computer communication," *Annu. Rev. Biophys. Bioeng.*, vol. 2, no. 1, pp. 157–180, 1973.

[2] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain–computer interfaces for communication and control," *Clin. Neurophysiol.*, vol. 113, no. 6, pp. 767–791, 2002.

[3] G. Dornhege, J. R. Millán, T. Hinterberger, D. J. McFarland, and K.-R. Müller, *Toward Brain-Computer Interfacing*. Cambridge, MA, USA: MIT Press, 2007.

[4] D. J. Krusienski and J. R. Wolpaw, "Brain–computer interface research at the wadsworth center developments in noninvasive communication and control," *Int. Rev. Neurobiol.*, vol. 86, pp. 147–157, Jan. 2009.

[5] F. Vogel, "The genetic basis of the normal human electroencephalogram (EEG)," *Humangenetik*, vol. 10, no. 2, pp. 91–114, 1970.

[6] F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, and B. Arnaldi, "A review of classification algorithms for EEG-based brain-computer interfaces," *J. Neural Eng.*, vol. 4, no. 2, p. R1, 2007.

[7] J. P. Cunningham, P. Nuyujukian, V. Gilja, C. A. Chestek, S. I. Ryu, and K. V. Shenoy, "A closed-loop human simulator for investigating the role of feedback control in brain-machine interfaces," *J. Neurophysiol.*, vol. 105, no. 4, pp. 1932–1949, 2011.

[8] T. Wilaiprasitporn, A. Ditthapron, K. Matchaparn, T. Tongbuasirilai, N. Banluesombatkul, and E. Chuangsuwanich, "Affective EEG-based person identification using the deep learning approach," *IEEE Trans. Cogn. Develop. Syst.*, vol. 12, no. 3, pp. 486–496, Sep. 2020.

[9] J. Li, S. Qiu, C. Du, Y. Wang, and H. He, "Domain adaptation for EEG emotion recognition based on latent representation similarity," *IEEE Trans. Cogn. Develop. Syst.*, vol. 12, no. 2, pp. 344–353, Jun. 2020.

[10] B. J. Edelman, B. Baxter, and B. He, "EEG source imaging enhances the decoding of complex right-hand motor imagery tasks," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 1, pp. 4–14, Jan. 2016.

[11] J. Cao, J. Zhu, W. Hu, and A. Kummert, "Epileptic signal classification with deep EEG features by stacked CNNs," *IEEE Trans. Cogn. Develop. Syst.*, vol. 12, no. 4, pp. 709–722, Dec. 2020.

[12] Y. Li *et al.*, "A novel bi-hemispheric discrepancy model for eeg emotion recognition," *IEEE Trans. Cogn. Develop. Syst.*, vol. 13, no. 2, pp. 354–367, Jun. 2021.

[13] Y. Zhang, P. Xu, T. Liu, J. Hu, R. Zhang, and D. Yao, "Multiple frequencies sequential coding for SSVEP-based brain-computer interface," *PLoS ONE*, vol. 7, no. 3, 2012, Art. no. e29519.

[14] P. Gang *et al.*, "User-driven intelligent interface on the basis of multimodal augmented reality and brain-computer interaction for people with functional disabilities," in *Proc. Future Inf. Commun. Conf.*, 2018, pp. 612–631.

[15] K. Takano, N. Hata, and K. Kansaku, "Towards intelligent environments: An augmented reality–brain–machine interface operated with a see-through head-mount display," *Front. Neurosci.*, vol. 5, p. 60, Apr. 2011.

[16] A. Tang, J. Zhou, and C. Owen, "Evaluation of calibration procedures for optical see-through head-mounted displays," in *Proc. 2nd IEEE ACM Int. Symp. Mixed Augmented Reality*, 2003, pp. 161–168.

[17] H. Si-Mohammed *et al.*, "Towards BCI-based interfaces for augmented reality: Feasibility, design and evaluation," *IEEE Trans. Vis. Comput. Graphics*, vol. 26, no. 3, pp. 1608–1621, Mar. 2020.

[18] X. Zhao, C. Liu, Z. Xu, L. Zhang, and R. Zhang, "SSVEP stimulus layout effect on accuracy of brain-computer interfaces in augmented reality glasses," *IEEE Access*, vol. 8, pp. 5990–5998, 2020.

[19] Y. Ke, P. Liu, X. An, X. Song, and D. Ming, "An online SSVEP-BCI system in an optical see-through augmented reality environment," *J. Neural Eng.*, vol. 17, no. 1, 2020, Art. no. 16066.

[20] S. Park, H.-S. Cha, J. Kwon, H. Kim, and C.-H. Im, "Development of an online home appliance control system using augmented reality and an SSVEP-based brain-computer interface," in *Proc. 8th Int. Winter Conf. Brain-Comput. Interface (BCI)*, 2020, pp. 1–2.

[21] X. Chen, X. Huang, Y. Wang, and X. Gao, "Combination of augmented reality based brain- computer interface and computer vision for high-level control of a robotic arm," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, pp. 3140–3147, 2020.

[22] Y. Wang, X. Zhang, K. Li, J. Wang, and X. Chen, "Humanoid robot control system based on AR-SSVEP," in *Proc. 6th Int. Conf. Comput. Artif. Intell.*, 2020, pp. 529–533.

[23] W. Ding, J. Shan, B. Fang, C. Wang, F. Sun, and X. Li, "Filter bank convolutional neural network for short time-window steady-state visual evoked potential classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 2615–2624, 2021.

[24] N. V. Manyakov, N. Chumerin, A. Robben, A. Combaz, M. van Vliet, and M. M. Van Hulle, "Sampled sinusoidal stimulation profile and multichannel fuzzy logic classification for monitor-based phase-coded SSVEP brain-computer interfacing," *J. Neural Eng.*, vol. 10, no. 3, 2013, Art. no. 36011.

[25] X. Chen, Z. Chen, S. Gao, and X. Gao, "A high-ITR SSVEP-based BCI speller," *Brain-Comput. Interfaces*, vol. 1, nos. 3–4, pp. 181–191, 2014.

[26] Y. Wang, X. Chen, X. Gao, and S. Gao, "A benchmark dataset for SSVEP-based brain–computer interfaces," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, pp. 1746–1752, 2017.

[27] Z. Lin, C. Zhang, W. Wu, and X. Gao, "Frequency recognition based on canonical correlation analysis for SSVEP-based BCIs," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 6, pp. 1172–1176, Jun. 2007.

[28] G. Bin, X. Gao, Y. Wang, Y. Li, B. Hong, and S. Gao, "A high-speed BCI based on code modulation VEP," *J. Neural Eng.*, vol. 8, no. 2, 2011, Art. no. 25015.

[29] Q. Wei, S. Zhu, Y. Wang, X. Gao, H. Guo, and X. Wu, "A training data-driven canonical correlation analysis algorithm for designing spatial filters to enhance performance of SSVEP-based BCIs," *Int. J. Neural Syst.*, vol. 30, no. 5, 2020, Art. no. 2050020.

[30] F. Di Russo and D. Spinelli, "Electrophysiological evidence for an early attentional mechanism in visual processing in humans," *Vis. Res.*, vol. 39, no. 18, pp. 2975–2985, 1999.

[31] J. R. Wolpaw, H. Ramoser, D. J. McFarland, and G. Pfurtscheller, "EEG-based communication: Improved accuracy by response verification," *IEEE Trans. Rehabil. Eng.*, vol. 6, no. 3, pp. 326–333, Sep. 1998.

[32] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2980–2988.

[33] M. Minamoto, Y. Suzuki, T. Kanno, and K. Kawashima, "Effect of robot operation by a camera with the eye tracking control," in *Proc. IEEE Int. Conf. Mechatronics Autom. (ICMA)*, 2017, pp. 1983–1988.

[34] Y. Cao, S. Miura, Y. Kobayashi, K. Kawamura, S. Sugano, and M. G. Fujie, "Pupil variation applied to the eye tracking control of an endoscopic manipulator," *IEEE Robot. Autom. Lett.*, vol. 1, no. 1, pp. 531–538, Jan. 2016.

[35] H. Zeng, K. Li, N. Wei, R. Song, and X. Tian, "A sEMG-controlled robotic hand exoskeleton for rehabilitation in post-stroke individuals," in *Proc. IEEE Int. Conf. Cyborg Bionic Syst. (CBS)*, 2018, pp. 652–655.

[36] N. Feng, H. Wang, F. Hu, and J. Gong, "Humanoid soft hand design based on sEMG Control," in *Proc. 9th Int. Conf. Inf. Technol. Med. Educ. (ITME)*, 2018, pp. 187–191.

Bin Fang received the Doctoral degree from Beihang University, Beijing, China, in 2014.

He is a Research Assistant with the Department of Computer Science and Technology, Tsinghua University, Beijing. His research interests include sensor fusion, wearable devices, robotics, and human–robot interaction.

Wenlong Ding received the B.S. degree from Anhui University of Technology, Ma'anshan, Anhui, China, in 2019, where he is currently pursuing the master's degree.

His current research interests include brain–computer interface and deep learning.

Fuchun Sun (Fellow, IEEE) received the Doctoral degree from Tsinghua University, Beijing, China,ăin 1997.

He is a Full Professor with the Department of Computer Science and Technology, Tsinghua University. His current research interests include robotic perception and cognition.

Prof. Sun serves as an Associate Editor for a series of international journals, including the IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS: SYSTEMS, IEEE TRANSACTIONS ON FUZZY SYSTEMS, and *Robotics and Autonomous Systems*.

Jianhua Shan received the Doctoral degree from the University of Science and Technology of China, Hefei, China, in 2007.

He is a Full Professor with the Department of Mechanical Engineering, Anhui University of Technology, Ma'anshan, Anhui, China. His research interests include sensor fusion, wearable devices, robotics, and HMI.

Xiaojia Wang received the bachelor's degree in automation from the North China University of Technology, Beijing, China, in 2014. He is currently pursuing the Doctorate degree with Clemson University, Clemson, SC, USA.

His current research interests include embedded systems, VR/AR, human–machine interaction, and stochastic computing.

Chengyin Wang received the B.S. degree from Anhui University of Technology, Ma'anshan, Anhui, China, in 2019, where he is currently pursuing the master's degree.

His current research interests include human–robot interaction and deep learning.

Xinyu Zhang (Member, IEEE) received the B.E. degree from the School of Vehicle and Mobility, Tsinghua University, Beijing, China, in 2001.

He was a Visiting Scholar with the University of Cambridge, Cambridge, U.K. He is currently a Researcher with the School of Vehicle and Mobility, and the Head of the Mengshi Intelligent Vehicle Team, Tsinghua University. His research interests include intelligent driving and multimodal information fusion.