The SSHVEP Paradigm-Based Brain Controlled Method for Grasping Robot Using MVMD Combined CNN Model

Rui Li, Duanyang Bai, Zhijun Li, Shiqiang Yang, Weiping Liu, Yichi Zhang, Jincao Zhou, Jing Luo, and Wen Wang

***Abstract*— In recent years, the steady-state visual evoked potentials (SSVEP) based brain control method has been employed to help people with disabilities because of its advantages of high information transmission rate and low training time. However, the existing SSVEP brain control methods cannot adapt to dynamic or unstructured environments. Moreover, the recognition accuracy from the conventional decoding algorithm still needs to improve. To address the above problems, this study proposed a steady-state hybrid visual evoked potentials (SSHVEP) paradigm using the grasping targets in their environment to improve the connection between the subjects' and their dynamic environments. Moreover, a novel EEG decoding method, using the multivariate variational mode decomposition (MVMD) algorithm for adaptive sub-band division and convolutional neural network (CNN) for target recognition, was applied to improve the decoding accuracy of the SSHVEPs. 18 subjects participated in the offline and online experiments. The offline accuracy across 18 subjects by the 9-target SSHVEP paradigm was up to 95.41 ± 2.70%, which is a 5.80% improvement compared to the conventional algorithm. To further validate the performance of the proposed method, the brain-controlled grasping robot system using the SSHVEP paradigm was built. The average accuracy reached 93.21 ± 10.18% for the online experiment. All the experimental results demonstrated the effectiveness of the brain-computer interaction method based on the SSHVEP paradigm and the MVMD combined CNN algorithm studied in this paper.**

Manuscript received 24 July 2023; revised 2 December 2023 and 1 June 2024; accepted 1 July 2024. Date of publication 9 July 2024; date of current version 19 July 2024. This work was supported in part by the National Natural Science Foundation of China under Grant 52305035 and in part by the Northwestern Polytechnical University Unmanned Aerial Vehicle (UAV) Key Laboratory Fund under Grant 2022-JCJQ-LB-719. *(Corresponding author: Rui Li.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Ethics Review Committee at Xi'an University of Technology under Application No. 2023066.

Rui Li, Duanyang Bai, Zhijun Li, Shiqiang Yang, Yichi Zhang, Jincao Zhou, and Wen Wang are with the School of Mechanical and Instrumental Engineering, Xi'an University of Technology, Xi'an 710048, China (e-mail: liruizixing@163.com; baiduanyang2022@163.com; 15148812969@163.com; yangsq@126.com; zhangyichigz@163.com; jzhou324@xaut.edu.cn; wangwen@xaut.edu.cn).

Weiping Liu is with Xi'an People's Hospital, Xi'an 710004, China (e-mail: neurosurgeonliu@163.com).

Jing Luo is with Shaanxi Key Laboratory for Network Computing and Security Technology, School of Computer Science and Engineering, and the Human-Machine Integration Intelligent Robot Shaanxi University Engineering Research Center, Xi'an University of Technology, Xi'an, Shaanxi 710048, China (e-mail: luojing@xaut.edu.cn).

Digital Object Identifier 10.1109/TNSRE.2024.3425636

***Index Terms*— Steady-state hybrid visual evoked potentials (SSHVEP), convolutional neural network (CNN), grasping robot, human–machine interaction, multivariate variational mode decomposition (MVMD).**

I. Introduction

With the number of ageing people growing, it has become an essential social problem to help disabled patients, especially those with ageing disabilities to restore their essential life ability. Brain-computer interface (BCI) is a new type of human-computer interaction technology that does not rely on peripheral neuromuscular tissue and establishes a direct communication channel between the brain and external devices [1], [2]. This novel technology shows significant innovative and great application value compared with traditional human-computer interaction methods (such as keyboard, joystick). With the development of brain science and technology, brain-computer fusion technology can control the robot to complete complicated control by analyzing the patient's intention. This technology has been widely used in various fields, such as disability-assisted robots and rehabilitation robots, which can not only help patients restore their daily life abilities but also enhance the patient's confidence in life [3]. Since the non-implantable BCI has the advantages of low cost, low risk, and easy operation, it become the mainstream research in brain control systems.

According to the mechanism of EEG, brain control systems can be divided into spontaneous and evoked systems [4]. Spontaneous brain control systems that could autonomously regulate EEG rhythms or potentials through specific mental activities without external stimulation, such as ERD/ERS, slow cortical potentials, etc. One of the most typical systems is the motor imagery (MI) based brain control method. A spontaneous brain control system's advantage is that it can naturally control neural activity without relying on external stimuli. However, it requires a long training time and few control commands [5].

In recent years, affective science-based brain-computer interface system as a new BCI paradigm has received extensive attention. Prof Lu's team proposed a computational model of optimal graph-coupled semi-supervised learning to discriminate different human emotions [6]. Our team has researched facial expression-based BCI systems since 2018 and successively realized prosthesis control using four

types of EEG resulting from facial expressions [7], [8]. Unfortunately, the major drawback of this kind of brain control paradigm is EMG artifact interference.

Other standard modalities used in brain control systems include SSVEPs and P300-based brain control systems, which induce the EEG by external physical stimuli [9]. Kim et al designed a P300-based unmanned aerial vehicle (UAV) control method which successfully operated UAV up, down, left, right, and rotation [10]. Although P300 potentials have the advantage of fast response and high recognition accuracy, the responses of these potentials rely on novel stimuli, which causes the communication rate is not satisfied for patients [11].

Another evoked BCI system based on SSVEP paradigm achieved more attention due to the possibility of achieving higher recognition rates, larger sets of instructions, and shorter training times [12]. For visually evoked BCI system, designing a suitable SSVEP paradigm plays an important role in establishing such a BCI system [13]. Traditional paradigms have been employed in SSVEP-based brain-controlled systems, including flashing LED lights, black and white squares, and the checkerboard grid flipping [9]. Despite these paradigms having achieved relatively good results, their development is in its infancy, such as long-term adaptation by paradigm stimulation, the limitation of usable stimulus frequencies, and difficulty focusing on attention. Therefore, most research teams focus on optimizing the design of SSVEP paradigms.

To improve the BCI performance, Prof. Xu and his colleagues reported a novel Newton's ring paradigm with periodic motion, which enhanced the activation of SSVEPs through periodic motion stimulation [14]. This group also designed four novel stimulation paradigms based on motion pattern [15]. To simplify the paradigm layout and alleviate user fatigue, the group at Zhejiang University designed a new SSVEP paradigm based on the multi-target overlap, which makes the target stimulus unit share the exact physical space location with other stimulus units [16]. Most of the past research on the SSVEP paradigm focused on reducing user fatigue and enhancing the number of usable stimuli frequencies. However, the need for more adaptability to dynamic, unstructured environments for the SSVEP paradigm was ignored, which caused it only be used in pre-defined scenarios.

In addition, improving the accuracy of SSVEPs is challenging because these characteristics were not fully utilized [17], [18]. Chongqing University reported an SSVEP decoding algorithm based on a spatial dimensional fusion strategy, which used the information from all spatial filters in the standard maximized signal fraction analysis (MSFA) to identify the frequencies of SSVEP signals [19]. The University of Macau achieves multi-target SSVEP classification by considering the characteristics of the target and the adjacent non-target frequencies [20]. Based on a similar motivation, a new deep neural network (DNN) architecture was proposed by performing a commonality training model first and then adjusting the model parameters according to individual differences [21]. Although most of the past research improved the SSVEPs accuracy, it remains to explore optimizing the decoding algorithm by efficiently utilizing the harmonic components of SSVEP.

Besides the decoding algorithms, combining biological and machine intelligence is another critical role in SSVEP-based BCI applications. Up to now, remarkable progress on brain-controlled systems has been made using the SSVEP method. One of the pioneering works integrated machine vision with visual stimuli and used it for robot arm localization and grasping in dynamic environments [22]. Another report designed a hybrid brain control system for humanoid robots, combining the SSVEP mechanism with AR technology to control the robot walking in a maze through AR glasses [23].

Although the SSVEP-based brain control system has been significantly enhanced in robotics applications, it is still far away for clinical applications [24]. Overall, the shortcomings of SSVEP based control method for robot application can be summarized as follows:

1). The traditional SSVEP paradigm primarily uses black-and-white stimuli, poorly connected to the natural environment.

2). The decoding algorithms of SSVEP fail to combine harmonic components adaptively, and there is still a need to find suitable algorithms to improve the utilization of harmonic components.

3). The traditional SSVEP-based control methods cannot cope with dynamic, unstructured environments.

To address the above problems, this study proposed a novel SSHVEP paradigm combining the Yolo algorithm and SSVEP mechanism to adopt a dynamic, unstructured environment and a MVMD combined with the CNN model was implemented to decode SSHVEP. The proposed SSHVEP-based controlled method precisely controls a seven degree-of-freedom (DOF) robotic arm to grasp living objects under a dynamic non-structural environment. From our experimental validation, the main contributions of this work can be summarized as follows:

First, to address the shortcomings of poor adaptability to dynamic, unstructured environments for traditional SSVEP paradigms, a novel SSHVEP stimulus paradigm was designed by combining targets in dynamic, unstructured environments with a radial grid. It provides additional options to improve the applicability of the visually evoked brain control system.

Second, to solve the problem of underutilizing the harmonic components of EEG, a decoding algorithm combining MVMD and CNN model is designed to extract the features in the SSHVEP signals automatically.

Finally, to verify the performance of the proposed method, a brain-controlled robotic arm system using the SSHVEP paradigm was built, and the robotic arm achieves autonomous and precise grasping using the proposed method.

The remainder of this paper is organized as follows. Section II. Experimental Protocol presents the details of the proposed paradigm and its experimental design. The section III. Data Processing Method introduced the selected MVMD combined CNN for SSHVEP decoding. The experimental results are analyzed and discussed in the IV and V section, respectively. The final section concludes this paper.

II. Experimental Protocol

To better understand the working mechanism of our proposed method, we described the construction of the brain-controlled grasping robot system and the design of the SSHVEP paradigm. Moreover, the experimental setup was also systematically investigated.

A. Experimental Equipment

Based on previous experiences and the performance criteria for grasping robot control [25], [26], the brain-controlled grasping robot system used in this study consisted of four modules: a visual stimulator, EEG acquisition equipment, EEG processing equipment, and a seven DOF grasping robot.

A 64-channel wireless NeuSen-W64, manufactured by Neuracle Technology Co. Ltd, was selected as the EEG acquisition module (Fig. 1C). All electrode distributions of the NeuSen-W64 follow the International Standard 10-20 Electrode Location System.

A Lenovo Saver Y9000 microprocessor with AMD Ryzen 9-5900HX with Radeon Graphics CPU and a screen resolution of 2048∗1080 was employed as EEG visual stimulator and EEG processing equipment.

The grasping robot module was Risenko's seven DOF Sawyer robot. The main components of the grasping robot include a vision sub-module composed by depth camera and seven joints with four joints to rotation and the remaining three to pitch. The details of grasping robot parameters are shown in Table I.

The overall control strategy of the grasping robot system is illustrated in Fig. 1A. When the subject has the intention to grasp, the vision sub-module of the Sawyer robot identifies the living objects within its range and presents them to the visual stimulator after processing with different frequency modulation; the subject was gazing at the stimulus according to the demand; the EEG acquisition module recorded EEG simultaneously and transmitted it to the EEG processing module. Then, the computer processed the EEG data and recognized one result. The result was translated into a control command to control the robot arm for grasping according to the subject's intention.

B. Subjects and Data Acquisition

Eighteen healthy subjects from 22 to 26 years old (12 males and 6 females) volunteered to participate in the experiment and none of them has relevant BCI experiment experience. All participants were right-handed, had normal or corrected-to-normal vision, and had no neurological or psychiatric history. Before participating in the experiment, all subjects signed an informed consent form and received compensation. The Ethics Review Committee at Xi'an University of Technology approved the proposed experiment, and all experiments followed the Declaration of Helsinki. More details of sample size estimation can be found in section III-E Statistical analysis.

EEG data were recorded through NeuSen-W64 EEG amplifier with a sampling rate of 1000Hz. In the experiment, eight electrodes of POz, PO3, PO4, PO5, PO6, Oz, O1, and O2 from the occipital cortex were selected to record EEG. CPz and AFz were employed as the reference and grounding electrodes, respectively. The electrode position of the selected channels was shown in Fig. 1B. During the experiment, the impedance of each EEG recording electrode is lower than 5kΩ.

C. The SSHVEP Paradigm Design

Beyond system construction, several additional factors affect the full system's performance. One of the most important factors is the SSHVEP paradigm design. Considering the drawbacks of traditional SSVEP paradigm, in this section, the SSHVEP paradigm based on a dynamic environment target and radial checkerboard grid was designed to solve the problem of un-effective interaction between visually evoked systems and their natural environments. The detailed paradigm design procedure is as follows.

- Firstly, You Only Look Once v5 (YOLOv5) recognizes the target of daily living equipment in the dynamic environment where the grasping robot is located.
- Secondly, the stimulus combined the graspable target in a dynamic environment identified by YOLOv5 and a radial checkerboard grid using a nested format.
- Thirdly, the paradigm is sinusoidally modulated with different frequencies.

More detailed information about SSHVEP paradigm design can be found below.

1) Target Recognition Based on YOLOv5 Algorithm: One critical role in deciding the performance of visually evoked paradigm is the stimulus unit selection. The traditional stimulus has the severe drawback of needing more relevance to the application scenario. Considering our previous study of Scene Graph-SSVEP [25], the stimulus in the SSHVEP paradigm consists of the graspable daily living equipment in the dynamic environment where the robot is located. Specifically, the dynamic environment described in the paper means the current environment in which the grasping robot is located now. The stimuli in the SSHVEP paradigm are the possible targets that can be grasped in the above environment, which the YOLOv5 algorithm can recognize.

The current target detection methods used in dynamic environments are mainly divided into traditional machine learning and deep learning-based methods [27]. A representative deep learning-based algorithm of YOLOv5 is used in this study for possible graspable target recognition due to its superiority in real-time and accuracy in target recognition [28].

The YOLOv5 model mainly consists of four parts: the input, the backbone network, the neck network and the output, the procedure of basic YOLOv5 is shown in Fig. 2 [29]. The YOLOv5 network model is first trained using the coco dataset, a large and diverse dataset for object detection and segmentation that includes daily necessities, fruits, etc [30]. Then the pre-trained YOLOv5 network model is used to recognize the actual target to be grasped in the dynamic environment where the brain-controlled grasping robot is located. After that, the recognized target is segmented according to the coordinates to obtain the target picture.

After segmentation, the recognized target image must be pre-processed and sent to the SSHVEP paradigm presentation. Fig. 3A shows an example of an experiment scene where the

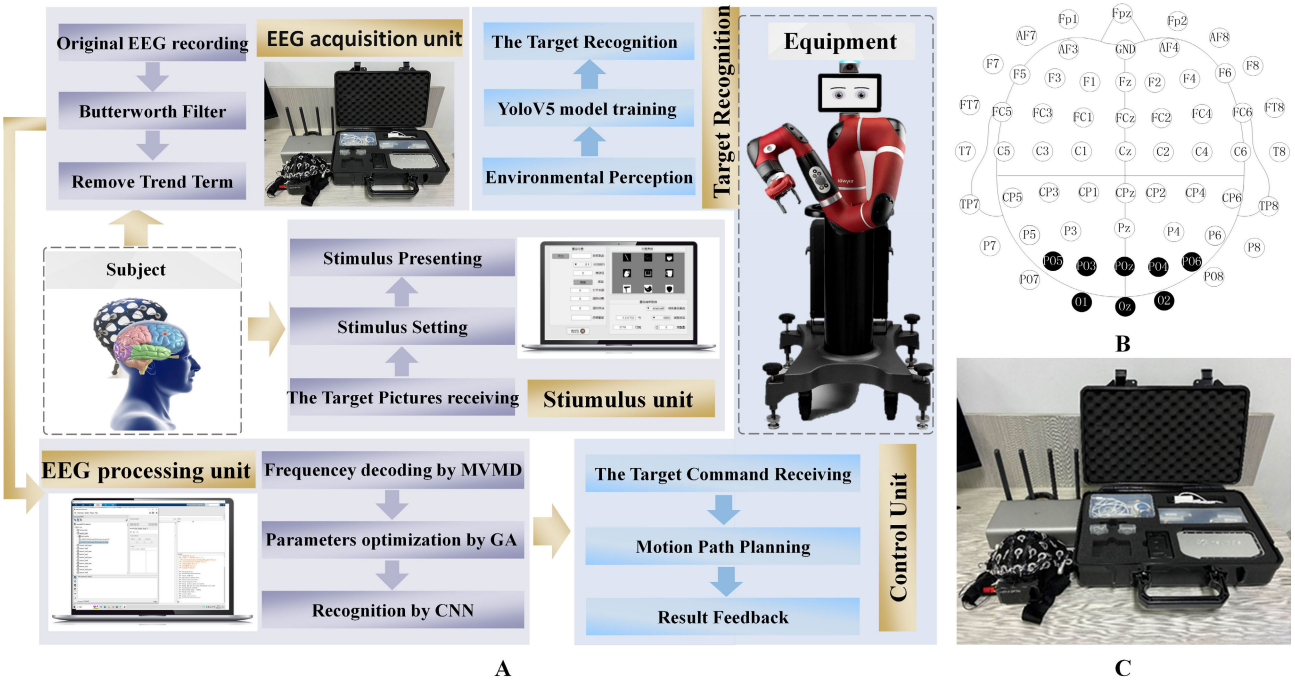

Fig. 1. (A) Overall system structure diagram. (B) Electrode arrangement diagram. (C) NeuSen-W64 EEG acquisition device.

TABLE I
SAWYER ROBOT PARAMETERS

Basic structure		Performance		Head camera		Wrist camera	
Parameter	value	Parameter	value	Parameter	value	Parameter	value
DOF	7	Arm Span	1260mm	Resolution	1280×800	Resolution	752×480
Weight	19KG	Precision	±0.1mm	Lens Type	Wide-angle	Lens Type	Wide-angle
Payload	4KG	Protection	IP54	Color	RGB	Color	Grayscale

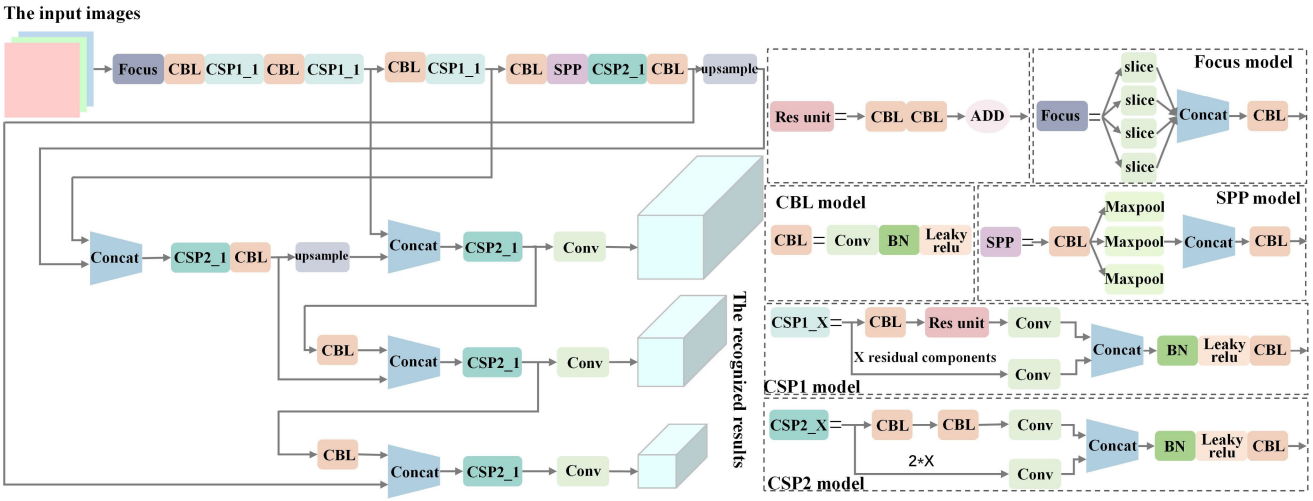

Fig. 2. YOLOv5 network structure diagram.

grasping robot is located. After the YOLOv5 model was recognized, nine possible grasping targets were recognized. Fig. 3B takes the possible grabbing target banana as an example and thoroughly describes its preprocessing process. In detail, after the YOLOv5 model recognizes the banana target, it first performs image segmentation on the banana target. Then, the banana target image is processed as follows: background removal, binarization, and edge extraction on the banana target image. In the above process, the pre-trained Mask R-CNN network model performs the background removal, the binarization is done by the OTSU algorithm [31] and edge processing is implemented by the Roberts operator [32].

2) The SSHVEP Paradigm: In this section, we nested the target image with a radial motion checkerboard to form the SSHVEP paradigm. It was developed by MATLAB (MathWorks) using the Psychophysics Toolbox [33].

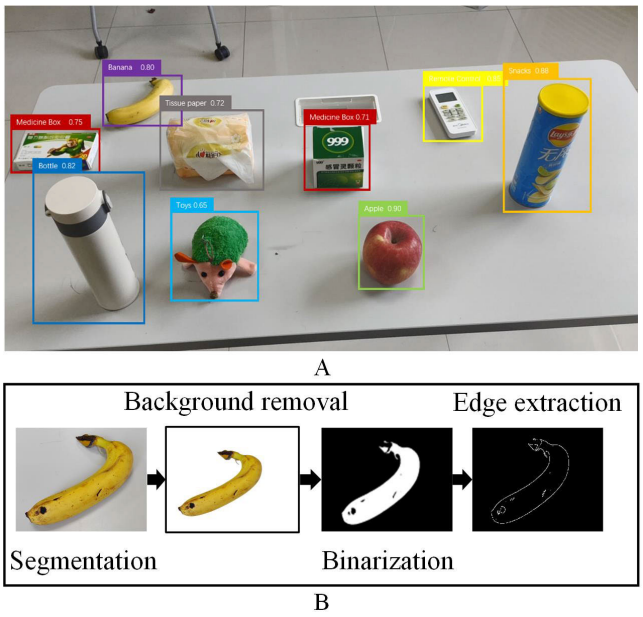

Fig. 3. (A) Experiment scene. (B) Target image processing.

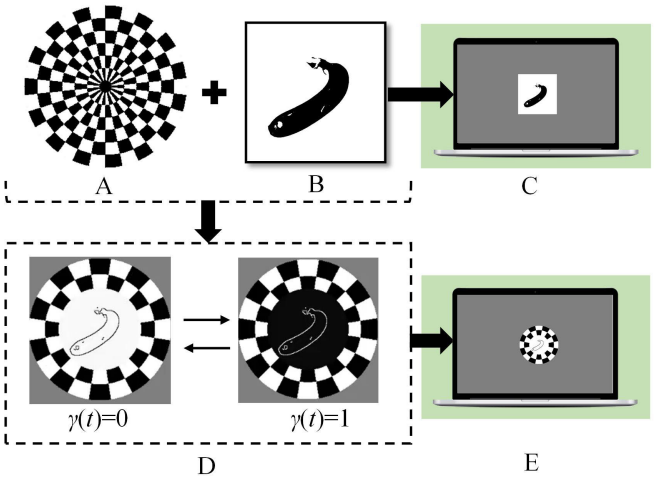

Fig. 4. (A) Radial checkerboard grid. (B) Scene target picture (take the banana target as an example). (C) The interface of SG-SSVEP paradigm. (D) Mixed checkerboard grid. (E) The interface of SSHVEP paradigm.

The radial motion checkerboard in Fig. 4A consists of multiple concentric rings. In detail, each circle is divided into white and black grids, and the white and black areas are the same size. The mechanism of visual response is performing contraction-dilation changes with different frequencies [34]. The formula of the radial motion checkerboard could be represented as follows:

$$stim = sign(\cos(\pi \cdot \frac{r(x,y)}{H} + \varphi(t) \cdot \frac{S}{H}) \cdot \cos(ang(x,y) \cdot N)) \quad (1)$$

where $\varphi(\ )$ is the phase value, sign represents a symbolic function, $r(x, y)$ and $ang(x,y)$ is the radius and angle of the radial motion checkerboard, respectively. N is the number of cells for a single circular division, and S and H denote the amplitude and the width of the checkerboard grid, respectively.

The formula of phase change can be described as follow:

$$\varphi(t) = \frac{\pi}{2} \cdot (1 + \sin(2\pi f t - \frac{\pi}{2})) \quad (2)$$

where f is the frequency.

In our research, the target images obtained from the current environment are nested to the radial motion checkerboard, which consists of the hybrid visual evoked stimulus named SSHVEP stimulus. In this way, a proposed stimulus was manipulated to evoke SSHVEPs via sinusoidal stimulation modulation, which is given by the following equation:

$$I = \begin{cases} |pic - 255 \cdot \varphi(t)|, & r < R_1 \\ sign(\cos(\pi \cdot \frac{r(x,y)}{H} + \gamma(t) \cdot \frac{S}{H}) \cdot \\ \cos(ang(x,y) \cdot N)), & R_1 < r < R_2 \\ I_0, & r > R_2 \end{cases} \quad (3)$$

where pic is the processed target images; I_0 denotes the background brightness; R_1 and R_2 are the inner and outer diameters of the checkerboard grid, respectively.

In each SSHVEP stimulus, the brightness is modulated by phase change $\gamma(t)$, which is computed by the following formula:

$$\gamma(t) = \frac{1}{2} \cdot (1 + \sin(2 \cdot \pi \cdot f \cdot \frac{n}{fr} - \frac{\pi}{2})) \quad (4)$$

where fr is the screen refresh rate; f is the frequency of change. When the phase function γ(t) goes from 0 to 1, the checkerboard grid expands, and the target picture becomes darker; when the phase function γ(t) goes from 1 to 0, the checkerboard grid contracts and the target picture becomes lighter. Taking the banana target as an example, the SSHVEP paradigm display is shown in Fig. 4D when the phase function γ(t) = 0 or 1.

The proposed SSHVEP paradigm can be used in various scenes and combinations in practical applications. The SSHVEP paradigm can be adjusted automatically for different working environments where the grasping robot is positioned.

To evaluate the superiority of the SSHVEP paradigm, we selected our previous designed SG-SSVEP paradigm as a comparison paradigm, which is shown in Figure 4C [25]. It realized the stimulus paradigm based on a predefined picture of the scene, such as the steps for drinking water.

D. Offline and Online Experiments

To verify the feasibility and the practicality of the SSHVEP-based brain-controlled method, all subjects performed both offline and online experiments. Each subject conducted the experiments in a quiet room on two separate days. In the experiment, the subjects were instructed to sit on a 60-100 cm chair away from the screen. The purpose of offline experiments was to evaluate the SSHVEP response of the proposed paradigm, and online experiments verified the performance of the SSHVEP-based brain-controlled method for grasping robots.

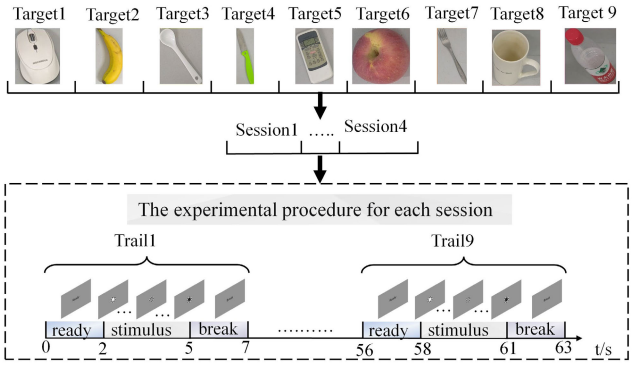

Fig. 5. Offline experiment process.

1) Offline Experiment: The offline experiment contained both the traditional SG-SSVEP paradigm (Fig. 4C) and the proposed SSHVEP paradigm (Fig. 4E). We set 7Hz, 8Hz, 9Hz, 10Hz, 11Hz, 12Hz, 8.5Hz, 9.5Hz, 10.5Hz frequency for all stimulus paradigms. The display and procedure of the traditional SG-SSVEP experiment were identical to those of the SSHVEP experiment.

Fig. 5 describes the time series of the offline experiment. During the offline experiment, each subject participated in 4 sessions of each stimulus in each paradigm. There were 9 trials in one session, and one trial lasted 7s with 2s preparation, 3s visual stimulation and 2s rest. To avoid visual fatigue, 3∼5min rest time between each session and 10∼15min rest between each paradigm was provided to the subject.

2) Online Experiment: The online experiment will use the SSHVEP paradigm to control the robot to grasp.

At the beginning of the online experiment, the visual sensors on the robot perceive the object information of the surrounding environment. These information is then presented in the visual stimulus paradigm.

The experiment scene is shown in Fig. 6A and the interface of the SSHVEP paradigm display is shown in Fig. 6B. Each subject was required to complete 9 experiments. The online experiment series consisted of 1s preparation, 1s cue, 2s visual stimulation, 1s feedback and 1s rest. After the visual stimulation, the EEG decoding algorithm recognizes the EEG during the visual stimulation, and the grasping robot maintains the initial position until the recognition result comes out. At the same time, the recognition result is also fed back to the SSHVEP interface in the form of the red box in real-time and will last 1s, which is shown in Fig. 6B. When the subject receives the feedback, the robot simultaneously grabs the target object chosen by the subject and moves it to their side.

During the rest time and the period while the robot grasps the target item, the subject can observe any changes in the information of the actual surrounding environment.

When the subject receives the item grabbed by the robot, the next experiment begins. The online experiment continues until the subject has completed all the experiments.

III. Data Processing Method

The decoding accuracy is determined by the signal quality of SSHVEP and the performance of the selected decoding algorithm, which further influences the performance of the

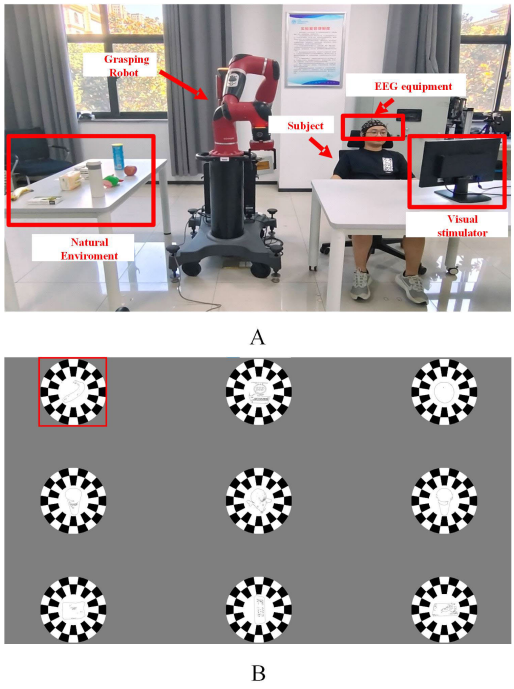

Fig. 6. (A) Online experiment scene. (B) The SSHVEP paradigm arrangement.

brain-controlled system. Hence, this section describes the SSHVEP decoding algorithm in detail. Firstly, to reduce noise contamination, SSHVEP data from online and offline experiments were extracted using the beginning and ending markers. Then, a Butterworth bandpass filter was used to filter EEG data into 4–45 Hz, and the trend of each channel was isolated. Lastly, the EEG was decoded by the MVMD-CNN algorithm.

A. Multivariate Variational Modal Decomposition(MVMD)

To fully detect the features from the SSHVEP, we selected the modal decomposition method of MVMD to divide the subbands of SSHVEP. MVMD, as a frequency domain separation method, has the advantage of adaptively adjusting the central frequency and bandwidth of each modal component to decompose a fixed number of sub-bands. Compared with the traditional filter bank canonical correlation analysis (FBCCA), MVMD can improve the efficiency of correlated subbands from visual evoked potentials [22].

MVMD decomposes nonlinear and non-smooth signals into multiple multivariate modulated oscillatory signals, which can effectively exploit the different characteristics of each harmonic [35]. Due to the diversity, complexity, and nonlinear and nonstationary characteristics of EEG, we selected MVMD to decompose the SSHVEPs and further analysis.

In the MVMD algorithm of this study, the SSHVEPs to decompose to k modulation components adaptively. The algorithm step includes:

1). Suppose the selected channel of SSHVEPs $X(t) = [x_1(t), x_2(t) \ldots\ldots x_c(t)]$ include k decomposition result from the MVMD, which is defined as:

$$x(t) = \sum_{k=1}^{K} u_k(t) \tag{5}$$

where k is the number of multivariate components, in this study $k = 3$ [22].

2). The analytic signal $u_+^k(t)$ calculated by Hilbert transform from the $u_k(t)$ is:

$$u_+^k(t) = \begin{bmatrix} u_+^{k,1}(t) \\ u_+^{k,2}(t) \\ \cdots \\ u_+^{k,C}(t) \end{bmatrix} = \begin{bmatrix} (\delta(t) + \frac{j}{\Pi t}) \times u_{k,1}(t) \\ (\delta(t) + \frac{j}{\Pi t}) \times u_{k,2}(t) \\ \cdots \\ (\delta(t) + \frac{j}{\Pi t}) \times u_{k,C}(t) \end{bmatrix} \quad (6)$$

where $u_+^{k,1}(t) \sim u_+^{k,c}(t)$ is the analytic signal from the corresponding SSHVEP channels 1 to C; $u_{k,1}(t) \sim u_{k,c}(t)$ is the of multivariate components from channels 1 to C of $u_k(t)$.

3). There is necessary to find the center frequency w_k in multiple channels from the $u_k(t)$, the spectrum of the analyzed signal is shifted to the baseband using a frequency shift operation:

$$u_+^k(t)e^{-jw_k t} = \begin{bmatrix} [(\delta(t) + \frac{j}{\Pi t}) \times u_{k,1}(t)]e^{-jw_k t} \\ [(\delta(t) + \frac{j}{\Pi t}) \times u_{k,2}(t)]e^{-jw_k t} \\ \cdots \\ [(\delta(t) + \frac{j}{\Pi t}) \times u_{k,C}(t)]e^{-jw_k t} \end{bmatrix} \quad (7)$$

4). Constructing constrained variational problems using the L^2-parametrization of gradients:

$$\begin{cases} \min\limits_{\{u_k(t)\}\{w_k\}} \left\{ \sum\limits_k \sum\limits_c \left\| \partial_t u_+^{k,c}(t) e^{-jw_k t} \right\|_2^2 \right\} \\ s.t \sum\limits_k u_{k,c}(t) = x_c(t), c = 1, 2, 3, \ldots, C \end{cases} \quad (8)$$

where $x_c(t)$ is the data of the cthchannel of the identified EEG.

5). Introducing a quadratic penalty term α and a Lagrange multiplier λ transformed the stated constrained variational problem into an unconstrained optimization problem. In this study, we applied alternate direction method of multipliers to update the decomposed components obtained by u_k. The update equation of u_k is expressed as:

$$u_{k,c}^{n+1} = x_c(w) - \sum_{u_i \neq k} u_{i,c}(w) + \lambda_c(w)/1 + 2\alpha\,(w - w_k)^2 \quad (9)$$

6). The update of the central frequency w_k is expressed as:

$$w_k^{n+1} = \int_0^\infty w \left| u_{k,c}(w) \right|^2 dw \Big/ \int_0^\infty \left| u_{k,c}(w) \right|^2 dw \quad (10)$$

After iterative optimization to obtain the final decomposed modes, the next step is to detect the target frequeny implicit in SSHVEP.

B. Convolutional Neural Networks

It is widely known that the accuracy of SSHVEP significantly relies on the performance of selected machine learning algorithms. CNNs are one of representative deep learning algorithms, as they provide faster training, superior information preservation across hierarchical processes, and a lower risk of overfitting. These advantages allow the CNN classifier to automatically learn the appropriate features from the EEG data while maintaining its translation invariance and data hierarchy [36]. Therefore, this paper selects CNN model for feature extraction and classification of SSHVEPs.

In this study, a five-layer CNN model was constructed considering the characteristics of the multi subbands of SSHVEPs from MVMD. The architecture of the proposed CNN model is illustrated in Fig. 7. In our study, the input data fed into the CNN model was a $C_{chan} \times N_t \times C$, where C_{chan} corresponds to the 8 selected EEG channel, N_t indicates the sampling point of each trail, C is the number of variational modes computed by MVMD. In this study, the input matrix is 8∗2000∗3. Due to the CNN model was used to discriminate the 9 objects used in daily life, the output layer was designed to have 9 outputs.

To avoid overfitting the network, the Dropout layer and L2 regularization are added to the network to reduce the complexity and improve the generalization of the network. The probability of a neuron stopping working in the Dropout layer is set to 0.5. In the regularization, $\beta 1$ is set as 0.99, and $\beta 2$ set 0.999. The remaining parameters in the CNN model are as follows: Batch is set to 16, Epoch is set to 64, and the number of iterations is 200. The number of convolutional kernels in the convolutional layer and the number of neurons in the fully connected layer are searched for using a genetic algorithm. The specific search procedure has been presented in our previous study [8]. Moreover, the cross-entropy loss function is used in the training process of the network to calculate the magnitude of the difference between the training results and the labels, and the Adam optimizer is used to update the network parameters to obtain optimal results.

To detect the robustness of the proposed MVMD-CNN and prevent an over-fitting problem, 5-fold cross-validation was used to investigate the classification accuracy, and each subject's data was used to train his/her own classifier. The 5-fold cross-validation randomly divided the offline SSHVEP dataset into five equal-sized subsets, four cross-validation were performed. During each validation, four subsets of data were used for training, and one was used for testing.

To evaluate the superiority of MVMD-CNN in recognizing SSHVEPs, the FBCCA and CNN algorithms were selected as comparison algorithms. FBCCA applied a filter bank analysis to integrate stimulus and harmonic frequency components [37]. The reference signals of FBCCA were associated with the nine stimulation frequencies, and the number of sub-band decomposition was set to 3, considering the MVMD decomposition results. The structure of the comparison CNN model and its parameter values were consistent with those of the MVMD-CNN model.

C. The Motion Planning of Grasping Robot

In addition to the above, exploration the motion planning method of grasping robots is also important. In this study, the grasping robot reaches the target position through an inverse kinematics solution. The motion information of each joint was computed by the Newton-Raphson method [38]. The whole process of the Sawyer robot in the given task, it consists of reaching the target position, grasping the target and delivering the target to the specified location. The whole process for Sawyer robot grasping is described as follows:

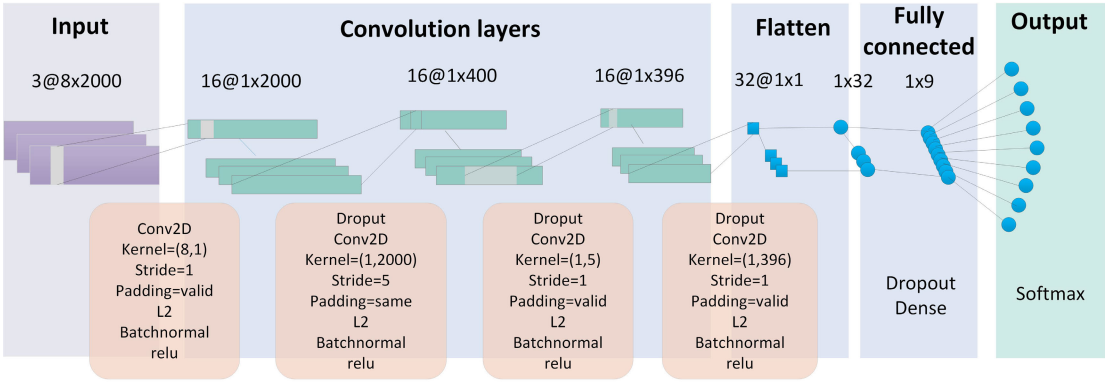

Fig. 7. CNN network structure diagram.

1). Set the current joint angle as θ_0 the target end position X_d, the error accuracy ε, the damping coefficient λ_0, and the minimum singularity σ_0;

2). Positive kinematics calculates the current end position X_c corresponding to the current joint angle θ_c;

3). Calculate the difference e between the target end position X_d and the current position;

4). If $\| e \| \leq \varepsilon$, output the joint angle θ_c and exit the program, otherwise, go to step 5;

5). Calculate the singular value σ of the pseudo-inverse matrix of the robot and select the appropriate damping coefficient λ by the singular value and the minimum singular value σ_0;

6). Calculate the amount of joint change $\Delta\theta_c$ according to $J^+ = J^T(JJ^T + \lambda I)^{-1}, \theta_c = J^+(\theta_c)e$;

7). Let $\theta_c = \theta_c + d\theta_c$ update θ_c according to the amount of change of the robot arm's motion joints, return to the second step, and loop again.

Until the robot delivers the target to the specified position, the cycle ends.

D. Performance Evaluation

To evaluate the performance of the proposed method, metrics of signal-to-noise ratio (SNR) and information transfer rate (ITR) were calculated. The SNR was used to evaluate the strength of the SSHVEP responses. For calculating power values, we normalized the power at a fundamental frequency *fi* to the power of the surrounding frequencies in the narrow band. The SNR can be computed as follows:

$$SNR(f_i) = 20\log_{10}\frac{mF(f_i)}{\sum_{k=1}^{m}[F(f_i + k\Delta f) + F(f_i - k\Delta f)]} \quad (11)$$

where *fi* denotes the response frequency, Δf is the frequency resolution and the value is 0.25Hz, k is the number of surrounding frequencies, defined as 8.

The ITR is used to evaluate the system transmission rate. In this study, it was calculated as follows:

$$V_{ITR} = \frac{60}{T}[\log_2 N + p\log_2 p + (1-p)\log_2(\frac{1-p}{N-1})] \quad (12)$$

where T denotes the time interval of each section, N is the number of targets, and P represents the correct recognition accuracy.

E. Statistical Analysis

The One-way analysis of variance (ANOVA) was used to test three different classification accuracy of FBCCA, CNN and MVMD- CNN. The paired t-test was conducted to assess the corresponding ITR value differences between the two paradigms. Moreover, the Greenhouse-Geisser correction was applied for p-value adjustment. It is well known that the sample size was determined by three parameters: the significance level (α), the desired statistical power (1)-β) and the effect size [39]. In this study, the desired statistical power was set as 0.8 (1-β = 0.8), the level of significance as 0.05 (α = 0.05) and the desired effect size as (f = 0.85). Under this given condition, the estimated sample size was 18 subjects using the statistical software G∗ Power.

After the experiments, each subject finished a simple questionnaire about two paradigms. There were three questions as follows, and each question was scored on a scale of 1-10, representing disagreement (1 point) to strong agreement (10 points).

1). Is the paradigm easy to focus on?

2). Does the paradigm cause visual fatigue?

3). Are you satisfied with the paradigm?

IV. Experimental Results and Analysis

In this section, both offline and online experiments were conducted. The offline experiment was used to evaluate the feasibility and superiorities of the proposed SSHVEP paradigm. Moreover, the effectiveness of the MVMD-CNN method was also observed in the offline section.

During the online experiment, the practicality of the proposed brain-controlled system was used in the grasping robot system.

A. The Experimental Results of Offline Experimental

1) The Effective of SSHVEP Paradigm: Source localization plays a crucial role in subject intention mapping. Hence, the standardized low-resolution brain electromagnetic tomography (sLORETA) method was firstly utilized to decode the brain source localization of the SSHVEP paradigm for subject intention. The EEG of each trail was averaged across, and the source localization result of 7 Hz from one representative subject S9 is plotted in Fig. 8. In this figure, the time stamps were set as 25ms, so there are a total of 8 cortical

surface activation sliced temporally. It was observed that the strong neural activation of SSHVEP detected in both occipital regions and several frontal-related regions (frontal pole (FP), pars orbitalis (PORB), lateral orbitofrontal cortex (LOF), and medial orbitofrontal cortex (MOF)). As time passed, the stable and obvious SSHVEP existed in occipital regions during the 2s. Other subjects show similar results. Hence, the result of EEG source localization can be concluded that the effectivity of the proposed novel paradigm can evoke SSHVEP successfully. During the time period, no obvious response decreases were found. This performance also confirmed that the performance of the proposed SSHVEP paradigm can evoke effective SSHVEP stability. In other words, the results of the sLORETA analysis proved that the proposed paradigm is satisfying the visual evoked mechanism.

To reveal the connection patterns among the different neural regions, the brain network properties under stimulus states among different frequencies were further analyzed. One of the crucial parameters of the Pearson correlation coefficient (PCC) among different regions was computed. The functional connectivity dynamics of weights between pairs of all channels from 9Hz was displayed in Fig. 9A, and its Adjacency matrix from all channels was shown in Fig. 9B.

This study set the sparsity value as 0.65 to reduce the spurious edges of all averaged networks before computing the topological properties. The sparsity value was generated based on all original networks of two frequencies across 18 subjects. It was found that the enhanced connection link mostly appeared among the parietal-occipital and frontal regions in the stimulus-evoked state. Conversely, the weak connections link among other regions. That is the reason why only the occipital region can cause strong SSHVEP. It also demonstrated that the PCC could be supposed a result of the enhanced synchronous neural activity of SSHVEP among brain regions. The analyzed result of brain network performance further demonstrated that the SSHVEP paradigm could induce the response from the occipital regions. That's the reason why eight channels of POz, PO3, PO4, PO5, PO6, Oz, O1, O2 were selected for further study. All the above results confirmed the feasibility of the proposed SSHVEP paradigm and gave the basis for selecting the optimal channel for a brain-controlled robot system.

To further compare the performance between SSHVEP and the conventional paradigm, the time-frequency and frequency spectrogram from two paradigms of S9 were shown in Fig. 10. Eight selected channels were fused to single channel data, which further used to a frequency domain and time-frequency domain analysis. The experiment results show the FFT power spectrum for 9 stimulus frequencies were significant at fundamental frequency and its harmonic components. It is worth noting that each time-frequency spectrogram observed remarkable and robust power increases in its fundamental frequency and harmonic frequencies during the whole period. Only a few interference peaks could be found in SSHVEP; the amplitude of background EEG activities also decreased when the frequency response increased. This performance indicated that the subjects with strong SSHVEP will show good accuracy within a short window. It also indicated that the designed new paradigm could effectively elicit SSHVEPs and satisfy the mechanism of visually evoked potentials.

The averaged SNRs of SSVEP and SSHVEPs are shown in Fig. 11. Although no apparent differences were found between the two paradigms, the SSHVEP paradigm still had some notable merits. It is found that the harmonics exhibited precise peak frequencies of SNR. The mean SNR value of the fundamental frequency was 8.87±0.83dB of SSHVEP and 7.27±1.16dB of SSVEP, respectively. The averaged SNR value of the second harmonic frequency was 7.73±1.41dB and 5.49±0.77 dB, respectively. There were significant differences between the SG-SSVEP paradigm and the proposed SSHVEP paradigm in the SNR values ($p < 0.05$, paired sample t-test). Overall, the SNR from the fundamental frequency and its harmonic components was improved 22.01% by SSHVEP. Since the SNRs were significantly positively correlated with classification accuracy, which means the SSHVEP components can be detected more easily. In other words, classification performance highly correlates with the SNR. The SSHVEP contains richer and more robust information that benefits different target recognitions. Overall, all the above results confirmed the proposed SSHVEP paradigm's feasibility and superiority.

2) The Recognition Performance of the Proposed MVMD-CNN: Fig. 12A and Fig. 12B are the visualization of the sub-band decomposition result from one representative subject S7, which uses Filter Banks and MVMD, respectively. The 1st subband from the two methods are close to each other. However, the other two sub-bands show obvious differences between the two methods. In detail, the frequency peak in the second subband from Filter Banks is 14Hz, 21HZ compared with 7Hz, 21Hz, 28Hz, 35 Hz, 42Hz by MVMD. The frequency and magnitude distributions of the second subband of MVMD and the filter bank are significantly different. The magnitude of the filter bank is significantly higher than that of MVMD. The higher harmonics have a larger magnitude than the lower harmonics, which is consistent with the mechanism of visual evoked potentials. To further demonstrate the advantages of MVMD, we also calculated the SNR of the subbands spectrum divided by the filter bank and MVMD [40], the results were shown in Fig. 13. Combining Fig. 12 and Fig. 13, it can be observed that the subbands decomposed by MVMD contain more efficient information from the harmonics, and it also limits the noises in background EEG. This finding also implied that the overlap of the frequency range in the sub-bands has yet to be fully solved by FBCCA, which will cause an insignificant improvement to the overall classification performance for visual evoked. The advantage of adaptively adjusting the central frequency and bandwidth of each modal component MVMD is that it will further improve the accuracy detection of SSHVEPs.

In more detail, it is evident that the subbands decomposed by MVMD enhanced the SSHVEP responses effect and reduced the influence of background EEG activities. More effective information embedded in the harmonic components can be helpful for target identification. Hence, the subbands extracted by MVMD are more suitable in the proposed system.

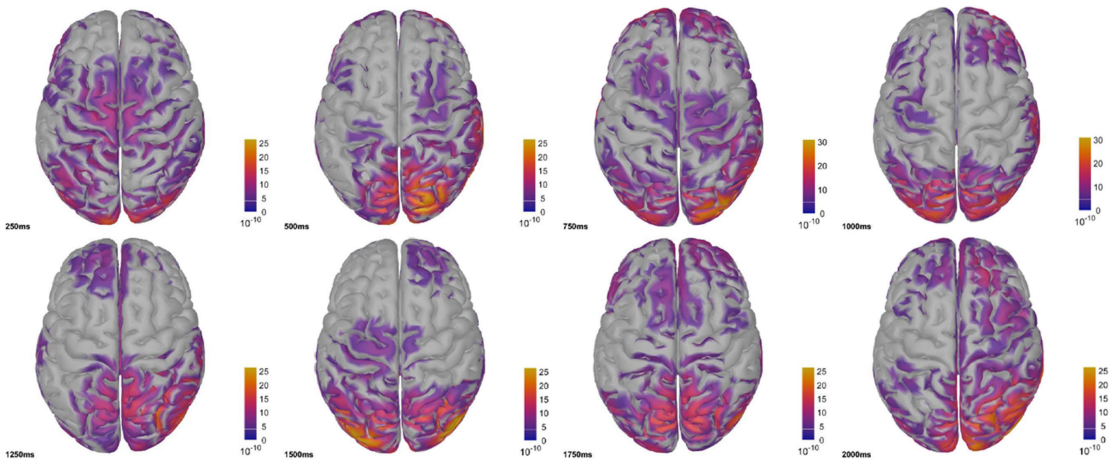

Fig. 8. The brain source localization analysis of one representative subject S9.

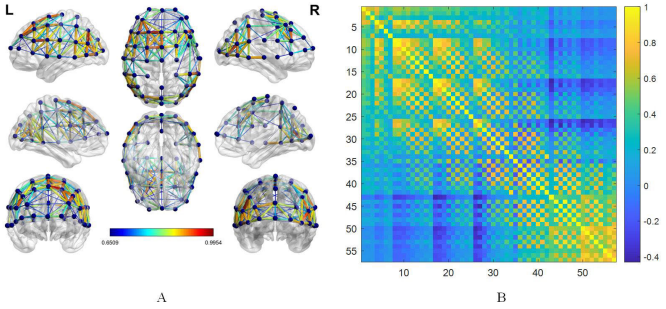

Fig. 9. Brain network analysis. (A) Brain network structure diagram. (B) Adjacency matrix.

TABLE II
ACCURACY COEFFICIENT OF MVMD-CNN, FBCCA AND CNN

Subject	MVMD-CNN	FBCCA	CNN
S1	97.02%	91.05%	88.60%
S2	99.64%	93.52%	86.15%
S3	95.61%	90.12%	85.43%
S4	92.43%	96.60%	86.03%
S5	97.26%	95.06%	95.08%
S6	98.21%	76.85%	86.99%
S7	94.62%	87.96%	93.56%
S8	98.21%	78.40%	89.65%
S9	96.42%	95.22%	87.25%
S10	95.83%	94.44%	86.32%
S11	98.85%	88.89%	91.12%
S12	91.07%	87.49%	84.78%
S13	97.19%	96.91%	88.58%
S14	92.78%	93.21%	86.67%
S15	93.21%	85.19%	87.36%
S16	91.13%	88.97%	87.19%
S17	92.25%	89.81%	89.58%
S18	95.58%	90.58%	91.61%
Avg±Std	95.41±2.70%	89.61±5.48%	88.44±2.85%

Table II lists the offline classification accuracies of all subjects of the three algorithms MVMD-CNN, FBCCA and CNN. The time window was set as 3s. The global accuracy of the MVMD-CNN was 95.41±2.70%, with an average accuracy of 89.61±5.48% for conventional FBCCA and an average accuracy of 88.44±2.85% for CNN. Compared with the conventional algorithm, the performance of the proposed MVMD-CNN has improved by 5.80%(FBCCA) and 6.97%(CNN). The One-way ANOVA analysis was conducted, and its result revealed a significant difference exists among the three algorithms ($p < 0.05$).

In more detail, the results from S6 and S8 were significantly improved using the proposed algorithm, which is 21.36% and 19.81%, respectively. To be noted, the MVMD-CNN methods can also reduce individual differences in the BCI system. In a word, the algorithm analysis suggested that the performance was significantly improved in terms of the accuracy when the proposed MVMD-CNN method was applied. It can be applied more effectively than FBCCA as well as CNN.

To further study the performance of the online experiment, the accuracy and corresponding ITR with time changes under the MVMD- CNN algorithm was summarized in Table III. As time increased, the accuracy rate also continued to increase. The averaged accuracy from 4-time windows is 80.62±7.53%, 88.80±3.78%, 90.39±3.60% and 95.41±2.70%, respectively. The corresponding averaged ITR of 4-time windows are also calculated using the proposed algorithm. The averaged ITR under different time windows are 45.71±9.15, 46.76±4.55%, 41.75±3.83% and 41.63±3.03, respectively. The best result was reached at 2s. Considering the practicality of the online experiment, the optimal window lines were set as 2s for online experiment.

3) Subjective Reports: As one might anticipate, the performance of comfort from SSHVEP is better than the traditional paradigm. The Statistical results are shown in Fig. 14. The averaged statistical results from the proposed paradigm were 7.50±2.07, 4.78±2.16, 7.28±1.60 and the corresponding results from the conventional paradigm were 4.89±1.28, 5.56±1.98, 5.83±1.76. Most subjects reported that the SSHVEP paradigm was more comfortable and easily acceptable than the traditional SSVEP paradigm. The paired t-test revealed a significant difference between the two paradigms ($p < 0.05$).

It will know that the performance of SSVEP-based BCI is predominantly affected by stimulation paradigm designs and recognition methods. All the above results further demonstrated the superiority of the proposed SSHVEP paradigm and its decoding algorithm of MVMD-CNN.

B. The Experimental Results of Online Experimental

Offline experiments verify the effectiveness of the proposed paradigm and its corresponding algorithm, and the practicality

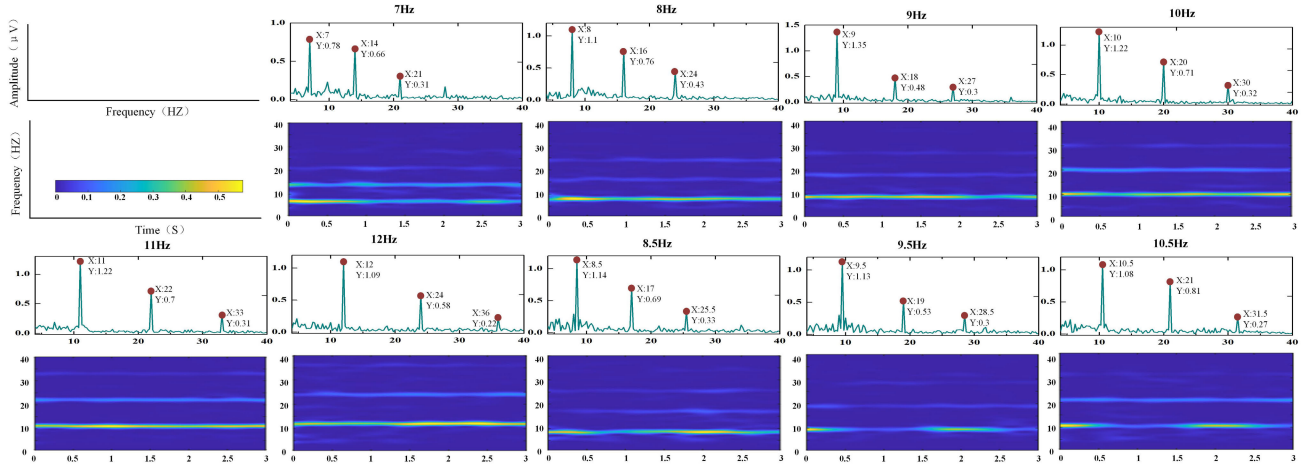

Fig. 10. Spectrograms at different stimulation frequencies and time-frequency graphs.

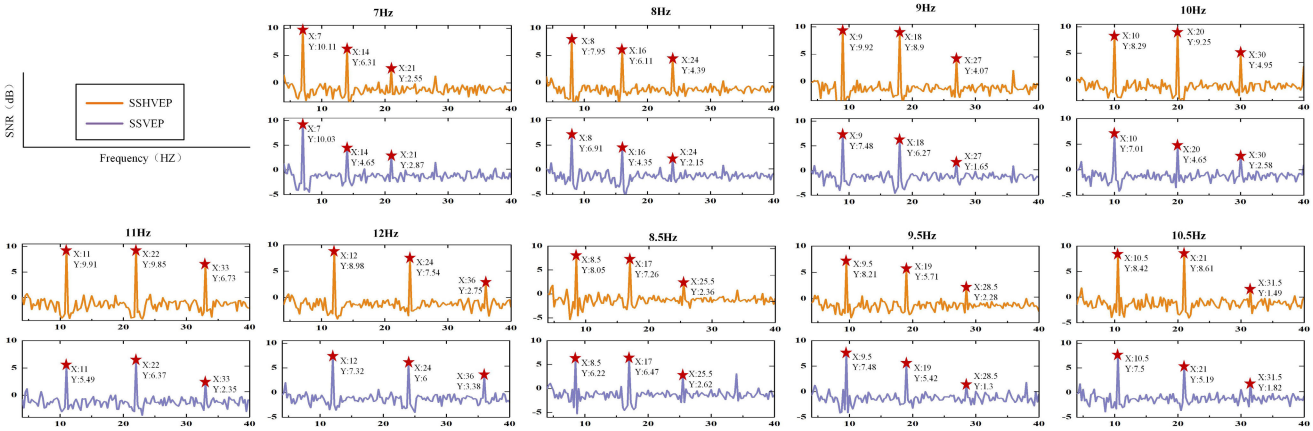

Fig. 11. The averaged SNRs of SSVEP and SSHVEPs.

TABLE III
CORRECT RATE AND INFORMATION TRANSMISSION RATE UNDER DIFFERENT TIME WINDOWS

Subject	1.5s		2.0s		2.5s		3.0s	
	Acc	ITR	Acc	ITR	Acc	ITR	Acc	ITR
S1	82.01%	46.81	90.38%	48.49	91.47%	42.74	97.02%	43.31
S2	89.95%	57.55	95.87%	55.96	94.83%	46.65	99.64%	46.87
S3	85.79%	51.69	86.05%	43.37	88.10%	39.20	95.61%	41.67
S4	73.54%	37.02	87.08%	44.54	90.74%	41.95	92.43%	38.34
S5	87.83%	54.49	93.04%	51.93	95.39%	47.35	97.26%	43.60
S6	75.19%	38.82	81.34%	38.32	84.16%	35.39	98.21%	44.80
S7	85.19%	50.89	87.21%	44.69	92.76%	44.19	94.62%	40.59
S8	70.72%	34.06	89.85%	47.83	91.04%	42.27	98.21%	44.80
S9	91.21%	59.44	94.06%	53.33	95.83%	47.91	96.42%	42.60
S10	85.71%	51.59	91.21%	49.54	93.28%	44.79	95.83%	41.92
S11	78.31%	42.35	82.11%	39.11	90.56%	41.76	98.85%	45.67
S12	70.01%	33.34	87.68%	45.24	84.21%	35.43	91.07%	37.02
S13	73.02%	36.46	90.91%	49.16	92.47%	43.86	97.19%	43.51
S14	83.60%	48.82	85.29%	42.52	88.49%	39.59	92.78%	38.69
S15	89.95%	57.55	89.72%	47.67	90.58%	41.78	93.21%	39.12
S16	70.15%	33.48	89.85%	47.83	84.54%	35.74	91.13%	37.08
S17	73.32%	36.78	86.86%	44.29	90.65%	41.85	92.25%	38.16
S18	85.69%	51.56	89.91%	47.91	87.98%	39.08	95.58%	41.64
Avg±Std	80.62±7.53%	45.71±9.15	88.80±3.78%	46.76±4.55%	90.39±3.60%	41.75±3.83%	95.41±2.70%	41.63±3.03

of the constructed brain-controlled robot grasping system is analyzed by online experiments. To verify the feasibility of the constructed robot grasping system, the simulation environment was built in Gazebo under Robot Operating System (ROS), and the autonomous grasping code was written to realize the robot's autonomous grasping and placing. The simulation experiment provided the foundation for the real environment. Fig. 15 describes an example of grasping banana. Fig. 15A is the robot arm in the initial state, Fig. 15B is to grasp the target object, Fig. 15C is to place the object at the specified

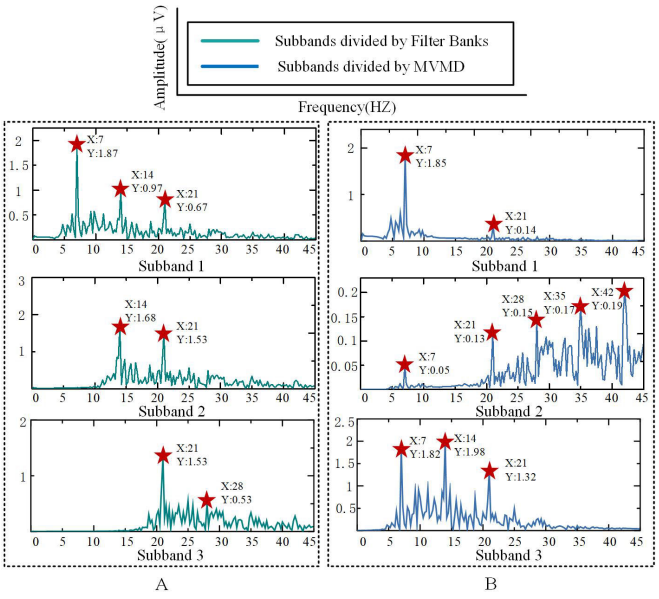

Fig. 12. (A) Spectrogram of subbands divided by the Filter Banks. (B) Spectrogram of subbands divided by the MVMD.

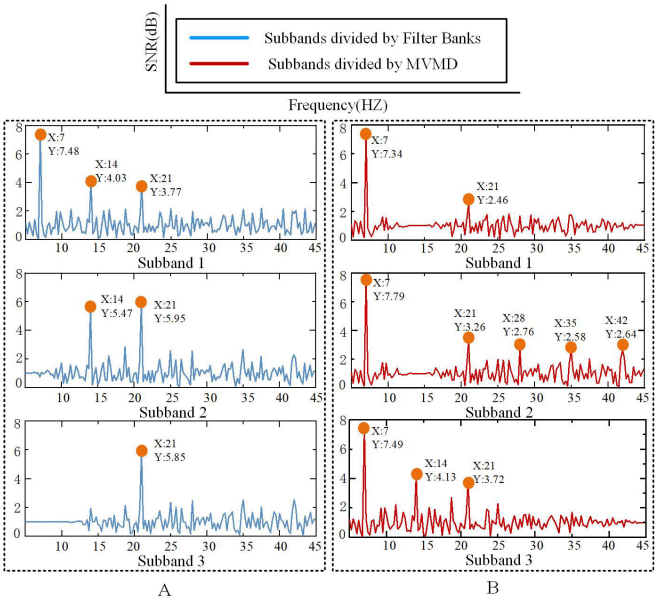

Fig. 13. (A) SNR of the subbands spectrogram divided by the Filter Banks. (B) SNR of the subbands spectrogram divided by the MVMD.

position, Fig. 15D is the robot arm back to the initial state, and the remaining steps are the transition state of the grasping process.

All the subjects performed grasping experiments on nine targets, and the experimental results are shown in Table IV. The correct average rate was 93.21 ± 10.18%, and the highest correct rate reached 100% from 11 subjects, which proved the feasibility of the proposed method and the practicality of the constructed brain-controlled robot grasping system. This performance further demonstrated the generalization ability of the SSHVEP paradigm and its corresponding MVMD-CNN algorithm. All the results of the online experiment demonstrated the effectiveness and feasibility of the SSHVEP-based BCI system.

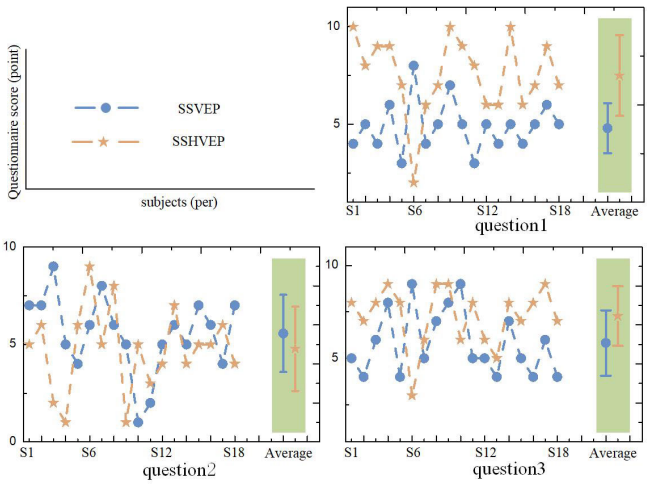

Fig. 14. Statistical results of the questionnaire survey.

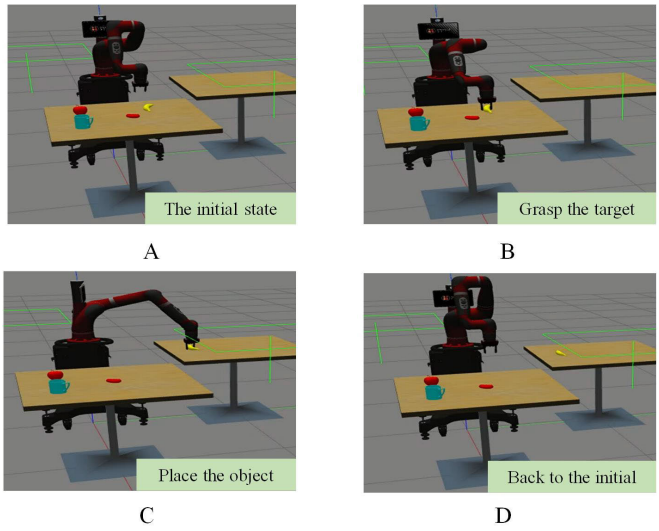

Fig. 15. Diagram of robot grasping process under simulation environment. (A) The robot arm in the initial state. (B) The robot arm grasping the target object. (C) The robot arm placing the object at the specified position. (D) The robot arm backing to the initial state.

V. Discussion

The result in this work demonstrated the effectiveness of a novel SSHVEP brain-controlled method for grasping robots. Three crucial factors of the SSHVEP paradigm, corresponding decoding algorithm and its control strategy for grasping robot were studied. The proposed paradigm can improve the relationship between natural sense and its visual paradigm. To fully use the characteristic from SSHVEPs, a modified algorithm combining MVMD and CNN model was designed. For robot grasping, we use inverse kinematics solution control mode to increase efficiency. Both the offline and online experiments verified the effectiveness of our proposed SSHVEP-based system.

A. The Significance of Proposed SSHVEP Paradigm

Recent studies have paid more attention to augmenting the number of targets and improving its ITR value for the visually evoked paradigm. Previous studies have reported some novel paradigms for evoking SSVEP. The study of the SSVEP

TABLE IV
SUBJECTS GRASP CORRECT ACCURACY

Sub	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
acc	100%	100%	88.89%	88.89%	100%	66.67%	100%	100%	100%	100%
Sub	S11	S12	S13	S14	S15	S16	S17	S18	Avg±Std	
acc	77.78%	100%	100%	88.89%	100%	88.89%	100%	77.78	93.21±10.18%	

paradigm in reported a novel motion-based Newton's Circle to evoke SSVEP [15]. Another novel SSVEP paradigm in uses a visual stimulus paradigm based on the multi-target overlap [16]. The traditional stimulus mode of SSVEP is light flashing or graphic flipping [41], using luminance as a stimulus paradigm. The advantage of these existing researches is that they can induce SSVEP successfully by different stimulus paradigms. However, this type of stimulation is prone to visual fatigue and even has the risk of inducing seizures in the user, which ultimately affects the performance of the BCI system [42]. One major drawback of the traditional visually evoked paradigm is that ignores the relationship between the paradigm and the natural environment. In other words, the human visual system is sensitive not only to the color of an object but also to its shape and other features.

To address the weak connection between the current visually evoked paradigm and its dynamic environment, our previous study suggested that the visually evoked paradigm regarding the Scene Graph of a subject's intention decreased the degradation of SSVEP and reduced the visual fatigue in comparison with the traditional SSVEP-BCI system [25]. In our study, we further optimized the design of a visually evoked paradigm SSHVEP regarding a subject's intention. In detail, a hybrid visually evoked paradigm based on a combination of target pictures from the dynamic environment and the radial checkerboard grid was designed. The analysis result of the responses property in Fig. 10 and corresponding functional brain networks in Fig. 9 demonstrated the effectiveness of the proposed paradigm. According to the comparison results of SNR recognition accuracy, the improved paradigm further optimizes the system's performance regarding the subject's intention. There is no denying that increasing the interaction between the BCI paradigm and the environments for the brain-controlled method is the main objective for BCI application. Moreover, most subjects reported that they experienced less visual fatigue and minor annoyance compared with the SG-SSVEP paradigm because it was easier to focus their attention on the stimulus target.

These advantages are expected to improve usability and practicality in SSVEP-based BCI systems, such as brain-controlled robots or wheels.

B. The Effectively of the Proposed MVMD-CNN Algorithm

Another essential factor that influences the performance of a brain-controlled system is its decoding accuracy. The performance of EEG processing algorithms can influence the accuracy and robustness of decoding. Most SSVEP decoding methods use the power spectrum features by their experience, such as minimum energy combination or Extended-CCA [18]. However, EEG is highly complex, non-stationary and nonlinear signals. The existing traditional methods need the assumption of signal stationarity and predefined basis functions. It is more difficult to identify each target correctly with the limited information. One effective way to improve detection accuracy is enhancing the feature quality of EEG. From the experimental results in Fig. 12 and Fig. 13. it is seen that the subbands decomposed by MVMD contain more efficient information from the harmonics, and it also limits the noises in background EEG. The SSHVEP decomposition analysis on amplitude and SNR (Fig. 10, Fig. 11) also shows that it contains rich and robust information in its harmonics.

For the automatic extract more valuable features, deep learning-based algorithms can effectively improve the classification accuracy of SSVEP. A conventional CNN model designed is used to address the SSVEP target identification [43]. Unfortunately, only using the CNN model can not comprehensively select the harmonic characteristic of the target frequency. Hence, learning more hidden features and eliminating redundant information from SSHVEP could enhance the overall capability of the BCI system. MVMD, as a frequency domain separation method, has the advantage of adaptively adjusting the center frequency and bandwidth of each mode component. Meanwhile, CNN can automatically extract valuable features and eliminate redundant information in each subband from SSHVEPs, which can effectively improve the recognition accuracy of the response EEG. The selected feature associated closely with the task will improve the performance of classification. Hence, to fully extract the effective feature information from the SSHVEP, this study chose a decoding algorithm that combined the MVMD and CNN model. The decoding results showed that the MVMD combined CNN model had a higher average classification accuracy compared with the traditional model. One of the important reasons is that MVMD extracts more deep frequencies characteristic of SSHVEP. The analyzed result may also provide possible evidence to demonstrate the whole performance of the SSHVEP-based brain-controlled system.

Even though our study achieved higher accuracy than previous work, only a basic CNN model was used for classification. To optimize system performance and reduce individual differences, future work should consider more efficient EEG decoding methods, such as Transformer, EEGNet, and DeepConvNet models, which will further improve the system's generalization capability and performance [44].

With recent advances in computer vision, it is possible to employ computer vision mechanisms to process EEG signals. After realizing the visual representation of EEG, we employed a multi-head self-attention mechanism to detect hidden characteristics in EEG data [45]. By incorporating graph convolutional modules, attention modules, and residual

modules into the CNN model, we can create a hybrid decoding method that combines SSHVEP with computer vision. This may be an effective way to extract high-quality features from raw EEG and further improve accuracy. However, this approach will increase the complexity of the entire system.

C. Applications

The main application for SSVEP-based brain-controlled system is the rehabilitation of people with disabilities. Quiles E reported a low-cost robotic arm control system using by conventional SSVEP paradigm. SSVEP paradigm not only controls the robot arm but also controls the prosthesis hand [26]. The group from Prof Zhang developed a brain-controlled prosthesis using the scene graph paradigm [25]. However, these applications ignore the robot's intelligence and have a low connection with their intention, resulting in low execution efficiency and weak fault tolerance in actual operation. This paper's ultimate objective is not only to help people with disabilities but also to expand to a broader user base on the SSHVEP-based brain-controlled method. One efficient brain-computer lies at the core of the research and development of BCI technology. And we combine human calculations with machine intelligence, using task-level control, where the user makes target selection and the gripping robot performs precision motion planning as well as gripping and delivery work, which can greatly improve work efficiency. In addition to its application in rehabilitation, our approach can be further used for the control of complex environments that are not suitable for humans.

D. The Potential Discrepancies of Proposed System

Although our proposed brain-controlled system demonstrated superiority compared to the previous system, potential discrepancies remain in conceptual, methodological and practical aspects.

In this study, the main target subjects for the proposed system are disabled people. However, with the number of elderly people growing, caring for the elderly is becoming a serious social problem [46]. Hence, it may be expected that our system can also be applied to some elderly people with reduced motor function. Moreover, the system could further evaluate the fatigue caused by the proposed system. These reasons may cause a discrepancy among conceptual aspects.

In the methodological aspects, the proposed system neglects the cognitive workload during EEG recording. Previous studies reported that increased workload and mental fatigue will cause degradation of EEG quality [47]. This may cause a potential discrepancy in our proposed methodology. Investigating the quality of EEG before its recording should be further considered.

In practice, the BCI performance might be affected by environmental factors. Moreover, the usability and user acceptance among disabled patients may promote the system utilization in practical scenarios. Hence, there is another potential factor causing discrepancies in the practical aspects of the proposed system.

E. The Limitations of the Study and Further Work

Even though our study improved the performance of the SSVEP-based brain-controlled system, some challenges still need to be solved in the future. Firstly, we need more subjects to further improve our paradigm, especially the disabled and older people. The proposed system needs a large sample size and a balance of male and female subjects to be thoroughly evaluated.

Although SSHVEP is an evoked potential highly dependent on stimulus performance and less susceptible to individual differences and noise contamination, the SSHVEP decoding CNN algorithm is nonlinear and still requires multiple samples for comprehensive training. Increasing the training sample size of the CNN can further optimize its performance. Additionally, individual differences in SSHVEP potentials, such as cognitive ability and attention level, should be further explored.

Since this experiment was conducted in a quiet indoor environment, it differs from real application scenarios. Future research should explore the impact of environmental factors such as noise, interference, light intensity, and task design on the stability and reliability of BCI systems.

A better system that allows for different numbers of stimuli should be constructed, and the improvement of the corresponding decoding algorithm should also be expected to further study. Meanwhile, more presentation ways of the SSHVEP paradigm should be considered, such as placing these symbolic images alongside the targets. Furthermore, it remains a significant challenge to establish a closed-loop system for real-time interaction between the subject and external devices.

VI. Conclusion

This paper proposes a novel brain-controlled method for grasping robots using the SSHVEP paradigm and a decoding algorithm combined with the MVMD and CNN model. To verify the effectiveness of the proposed method, 18 subjects participated in both offline and online experiments. The average accuracy of the proposed method is up to 95.41 $\pm$ 2.70%, which is 5.80% higher than that of the traditional method. The feasibility of the proposed system in online applications was also verified through a robot grasping task. The proposed system obtained a mean accuracy of 93.21$\pm$10.18%; the highest accuracy was achieved at 100%. All experimental results proved the practicality of the proposed SSHVEP-based controlled method and further provided a basis for the practical application of brain control technology. Future work aims to introduce shared control strategies to improve the control performance of brain-controlled grasping systems.

References

[1] X. Gao, Y. Wang, X. Chen, and S. Gao, "Interface, interaction, and intelligence in generalized brain–computer interfaces," *Trends Cognit. Sci.*, vol. 25, no. 8, pp. 671–684, Aug. 2021.

[2] M. Zabcikova, Z. Koudelkova, and R. Jasek, "Recent advances and currenttrends in brain-computer interface research and their applications," *Int. J. Develop. Neurosci.*, vol. 82, no. 2, pp. 107–123, 2022.

[3] G. Cheng, S. K. Ehrlich, M. Lebedev, and M. A. L. Nicolelis, "Neuroengineering challenges of fusing robotics and neuroscience," *Sci. Robot.*, vol. 5, no. 49, p. 1911, Dec. 2020.

[4] A. Singh, A. A. Hussain, S. Lal, and H. W. Guesgen, "A comprehensive review on critical issues and possible solutions of motor imagery based electroencephalography brain-computer interface," *Sensors*, vol. 21, no. 6, p. 2173, Mar. 2021.

[5] M. A. Khan, R. Das, H. K. Iversen, and S. Puthusserypady, "Review on motor imagery based BCI systems for upper limb post-stroke neurorehabilitation: From designing to application," *Comput. Biol. Med.*, vol. 123, Aug. 2020, Art. no. 103843.

[6] Y. Peng, F. Jin, W. Kong, F. Nie, B.-L. Lu, and A. Cichocki, "OGSSL: A semi-supervised classification model coupled with optimal graph learning for EEG emotion recognition," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 1288–1297, 2022.

[7] R. Li et al., "A novel EEG decoding method for a facial-expression-based BCI system using the combined convolutional neural network and genetic algorithm," *Frontiers Neurosci.*, vol. 16, Sep. 2022, Art. no. 988535.

[8] R. Li et al., "An approach for brain-controlled prostheses based on a facial expression paradigm," *Frontiers Neurosci.*, vol. 12, p. 943, Dec. 2018.

[9] F.-B. Vialatte, M. Maurice, J. Dauwels, and A. Cichocki, "Steady-state visually evoked potentials: Focus on essential paradigms and future perspectives," *Prog. Neurobiol.*, vol. 90, no. 4, pp. 418–438, Apr. 2010.

[10] S. Kim, S. Lee, H. Kang, S. Kim, and M. Ahn, "P300 brain-computer interface-based drone control in virtual and augmented reality," *Sensors*, vol. 21, no. 17, p. 5765, Aug. 2021.

[11] B. Z. Allison, A. Kübler, and J. Jin, "30+ years of P300 brain–computer interfaces," *Psychophysiology*, vol. 57, no. 7, p. 13569, Jul. 2020.

[12] X. Chi, H. Cui, and X. Chen, "Advances in multi-modal brain-computer interface combined with steady-state visual evoked potential," *Chin. J. Biomed. Eng.*, vol. 41, no. 2, pp. 204–213, 2022.

[13] M. Li, D. He, C. Li, and S. Qi, "Brain–computer interface speller based on steady-state visual evoked potential: A review focusing on the stimulus paradigm and performance," *Brain Sci.*, vol. 11, no. 4, p. 450, Apr. 2021.

[14] J. Xie, G. Xu, J. Wang, F. Zhang, and Y. Zhang, "Steady-state motion visual evoked potentials produced by oscillating Newton's rings: Implications for brain-computer interfaces," *PLoS ONE*, vol. 7, no. 6, Jun. 2012, Art. no. e39707.

[15] W. Yan, G. Xu, J. Xie, M. Li, and Z. Dan, "Four novel motion paradigms based on steady-state motion visual evoked potential," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 8, pp. 1696–1704, Aug. 2018.

[16] L. Wang, Z. Zhang, and D. Han, "Single stimulus location for two inputs: A combined brain-computer interface based on Steady-State Visual Evoked Potential (SSVEP)," *Eur. J. Neurosci.*, vol. 53, no. 3, pp. 861–875, 2021.

[17] Y. Zhang, S. Q. Xie, H. Wang, and Z. Zhang, "Data analytics in steady-state visual evoked potential-based brain–computer interface: A review," *IEEE Sensors J.*, vol. 21, no. 2, pp. 1124–1138, Jan. 2021.

[18] Y. Zhang, M. Xia, and K. Chen, "Progresses and prospects on frequency recognition methods for steady-state visual evoked potential," *J. Biomed. Eng.*, vol. 39, no. 1, pp. 192–197, 2022.

[19] Z. Li, K. Liu, X. Deng, and G. Wang, "Spatial fusion of maximum signal fraction analysis for frequency recognition in SSVEP-based BCI," *Biomed. Signal Process. Control*, vol. 61, Aug. 2020, Art. no. 102042.

[20] C. M. Wong et al., "Learning across multi-stimulus enhances target recognition methods in SSVEP-based BCIs," *J. Neural Eng.*, vol. 17, no. 1, Jan. 2020, Art. no. 016026.

[21] O. B. Guney, M. Oblokulov, and H. Ozkan, "A deep neural network for SSVEP-based brain-computer interfaces," *IEEE Trans. Biomed. Eng.*, vol. 69, no. 2, pp. 932–944, Feb. 2022.

[22] K. Wang, D.-H. Zhai, Y. Xiong, L. Hu, and Y. Xia, "An MVMD-CCA recognition algorithm in SSVEP-based BCI and its application in robot control," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 5, pp. 2159–2167, May 2022.

[23] S. Zhang, S. Gao, and X. Chen, "Humanoid robot walking in maze controlled by SSVEP-BCI based on augmented reality stimulus," *Frontiers Hum. Neurosci.*, vol. 16, Jul. 2022, Art. no. 908050.

[24] K. Sharma and S. K. Maharaj, "Continuous and spontaneous speed control of a robotic arm using SSVEP," in *Proc. 9th Int. Winter Conf. Brain-Comput. Interface (BCI)*, Feb. 2021, pp. 1–5.

[25] R. Li, X. Zhang, H. Li, L. Zhang, Z. Lu, and J. Chen, "An approach for brain-controlled prostheses based on scene graph steady-state visual evoked potentials," *Brain Res.*, vol. 1692, pp. 142–153, Aug. 2018.

[26] E. Quiles, J. Dadone, N. Chio, and E. García, "Cross-platform implementation of an SSVEP-based BCI for the control of a 6-DOF robotic arm," *Sensors*, vol. 22, no. 13, p. 5000, Jul. 2022.

[27] J. Zhang, C. Nong, and Z. Yang, "Review of object detection algorithms based on convolutional neural network," *J. Ordnance Equip. Eng.*, vol. 43, no. 6, pp. 37–47, 2022.

[28] H.-K. Jung and G.-S. Choi, "Improved YOLOv5: Efficient object detection using drone images under various conditions," *Appl. Sci.*, vol. 12, no. 14, p. 7255, Jul. 2022.

[29] B. Yan, P. Fan, X. Lei, Z. Liu, and F. Yang, "A real-time apple targets detection method for picking robot based on improved YOLOv5," *Remote Sens.*, vol. 13, no. 9, p. 1619, Apr. 2021.

[30] K. Tong and Y. Wu, "Rethinking Pascal-VOC and MS-COCO dataset for small object detection," *J. Vis. Commun. Image Represent.*, vol. 93, May 2023, Art. no. 103830.

[31] P. Yang, W. Song, X. Zhao, R. Zheng, and L. Qingge, "An improved Otsu threshold segmentation algorithm," *Int. J. Comput. Sci. Eng.*, vol. 22, no. 1, p. 146, 2020.

[32] J. Owotogbe, T. S. Ibiyemi, and B. Adu, "Edge detection techniques on digital images—A review," *Int. J. Sci. Res.*, vol. 4, no. 11, pp. 329–332, 2019.

[33] D. H. Brainard, "The psychophysics toolbox," *Spatial Vis.*, vol. 10, no. 4, pp. 433–436, 1997.

[34] C. Han, G. Xu, J. Xie, C. Chen, and S. Zhang, "Highly interactive brain–computer interface based on flicker-free steady-state motion visual evoked potential," *Sci. Rep.*, vol. 8, no. 1, p. 5835, Apr. 2018.

[35] N. u. Rehman and H. Aftab, "Multivariate variational mode decomposition," *IEEE Trans. Signal Process.*, vol. 67, no. 23, pp. 6039–6052, Dec. 2019.

[36] G.-A. Vargas-Hákim, E. Mezura-Montes, and H.-G. Acosta-Mesa, "A review on convolutional neural network encodings for neuroevolution," *IEEE Trans. Evol. Comput.*, vol. 26, no. 1, pp. 12–27, Feb. 2022.

[37] T. Lee, S. Nam, and D. J. Hyun, "Adaptive window method based on FBCCA for optimal SSVEP recognition," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 78–86, 2023.

[38] C.-F. Yang, S.-T. Zheng, J. Jin, S.-B. Zhu, and J.-W. Han, "Forward kinematics analysis of parallel manipulator using modified global Newton–Raphson method," *J. Central South Univ. Technol.*, vol. 17, no. 6, pp. 1264–1270, Dec. 2010.

[39] D. Lakens, "Sample size justification," *Collabra, Psychol.*, vol. 8, no. 1, 2021, Art. no. 33267.

[40] Y. Zhang, P. Xu, Y. Huang, K. Cheng, and D. Yao, "SSVEP response is related to functional brain network topology entrained by the flickering stimulus," *PLoS ONE*, vol. 8, no. 9, Sep. 2013, Art. no. e72654.

[41] 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.

[42] R. J. Snowden and T. C. A. Freeman, "The visual perception of motion," *Current Biol.*, vol. 14, no. 19, pp. 828–831, 2004.

[43] J. J. Podmore, T. P. Breckon, N. K. N. Aznan, and J. D. Connolly, "On the relative contribution of deep convolutional neural networks for SSVEP-based bio-signal decoding in BCI speller applications," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 4, pp. 611–618, Apr. 2019.

[44] Y. He et al., "Classification of attention deficit/hyperactivity disorder based on EEG signals using a EEG-transformer model," *J. Neural Eng.*, vol. 20, no. 5, Oct. 2023, Art. no. 056013.

[45] J. Luo et al., "A shallow mirror transformer for subject-independent motor imagery BCI," *Comput. Biol. Med.*, vol. 164, Sep. 2023, Art. no. 107254.

[46] S. Park, J. Ha, J. Park, K. Lee, and C.-H. Im, "Brain-controlled, AR-based home automation system using SSVEP-based brain-computer interface and EOG-based eye tracker: A feasibility study for the elderly end user," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 544–553, 2023.

[47] Y. Zhao et al., "Effects of distracting task with different mental workload on steady-state visual evoked potential based brain computer interfaces—An offline study," *Frontiers Neurosci.*, vol. 12, p. 79, Feb. 2018.