An SSVEP Stimuli Design using Real-time Camera View with Object Recognition

Shih-Kang Chen, Student Member, IEEE
Graduate Institute of Automation Technology
National Taipei University of Technology,
Taipei, Taiwan
jimmy4920@gmail.com

Chin-Sheng Chen, Member, IEEE
Graduate Institute of Automation Technology
National Taipei University of Technology,
Taipei, Taiwan
saint@ntut.edu.tw

Yu-Kai Wang, Member, IEEE
CIBCI Lab, Centre for AI, FEIT
University of Technology Sydney
Sydney, Australia
yukai.wang@uts.edu.au

Chin-Teng Lin, Fellow, IEEE
CIBCI Lab, Centre for AI, FEIT
University of Technology Sydney
Sydney, Australia
Chin-Teng.Lin@uts.edu.au

***Abstract*—Most SSVEP-based stimuli BCIs are pre-defined using the white blocks. This kind of scenario lead less flexibility in the real life. To represent the flickers with the location, types and configurations of the objects in real world, this paper proposes an SSVEP-based BCI using real-time camera view with object recognition algorithm to provide intuitive BCI for users. A deep learning-based object recognition algorithm is used to calculate the location of the objects on the online camera view from a depth camera. After the bounding box of the objects is estimated, the location of the SSVEP flickers are designed to overlap on the object locations. An overlapping FFT and SVM is used to recognize the EEG signals into corresponding classes. In experimental results, the classification rate for camera view scenario is more than 94.1%. The results show that proposed SSVEP stimuli design is available to create an intuitive and reliable human machine interaction. The proposed results can be used for the users who have motor disabilities to further used to interact with assistive devices, such as: robotic arm and wheelchairs.**

Keywords—Object recognition, steady state visually evoked potential (SSVEP), brain-computer interface (BCI).

I. Introduction

Robotic technologies are widely and important research field in recent years. On the other hand, multi-axis robotic arm is widely used as an assistive device for the physically challenged people. However, learning how to operate a robotic arm is big a challenge for an engineer, not to mention for the elderly [1]. Therefore, combining with robotic arm with an intuitive teach or operate technologies are presented to aid human operating a robot in order to reduce the difficulty of robot programming [2]. Brain-computer interface (BCI) systems acquire electroencephalography (EEG) signals from human brain and translate into digital command can be recognized on a computer using advanced algorithm. BCI can provide a new interface for users who have motor disabilities to control assistive devices [3, 4].

The BCIs have already been studied for several decades, the most commonly applied brain patterns include event-related potential, steady state visually evoked potential (SSVEP) and motor imagery [5, 6]. The SSVEP has the advantage of good signal-to-noise ratio (SNR), high information transfer rate (ITR) and the ability of large number of classes. In this research, SSVEP is adopt to present the brain patterns. However, the traditional SSVEP scenario was usually used white block and black background to present the flickers with different frequencies to maintain avoid the uncertainties and maintain the SNR [7-9]. Most interfaces of SSVEP-based BCIs are pre-defined. The kind of design lead less flexibility to the dynamic changes in real life, such as number, type or location of objects. In this research, we present new design for combining with the SSVEP scenario with camera view and using the object recognition algorithm to estimate the object position in the camera view to locate the flickers. The proposed scenario can provide a more interactive and intuitive human machine interaction. We even plan to test the combined scenario to various robotic systems and assess the changes of system performance in the following research.

Object recognition is an important topic in machine vision field. Since the performance of graphics processing units (GPU) grow up faster in resent years, the deep learning algorithm can be easily applied on the machine vision. The features extractors (e.g. SSD, R-CNN and R-FCN) and network architecture (e.g. VGG-16, MobileNet and Resnet) were applied and lead the significant improve for object recognition [10-13]. Combining with the SSVEP scenario, an efficient object recognition algorithm for real-time image is required. To approach the goal, was a Convolutional Neural Network (CNN)-based real-time object detection system, the You only look once (YOLO), proposed by [14] is applied in this research.

Finally, recognizing the EEG signals which is collected from the proposed camera view scenario, the methods such as power spectral density analysis, canonical correlation analysis (CCA), support vector machine (SVM), *k*-nearest neighbor (*k*-NN) and linear discriminant analysis (LDA) are frequently adopt for feature extraction and signal recognition in SSVEP signal processing [15, 16]. In this paper, a power value derived by overlapping fast Fourier transform (FFT) is used as the feature for SVM to recognize the SSVEP events.

An SSVEP-based BCI using real-time camera view with object recognition is proposed in this paper, the system

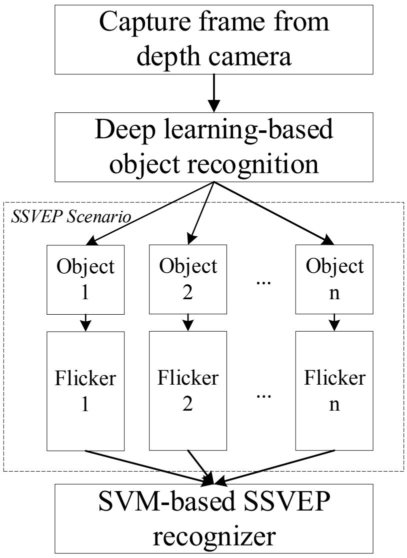

Fig. 1 The architecture of the full system. The current works in this paper was applied online camera view with object recognition to design the SSVEP flickers. After that, an SSVEP recognizer was used to classifier the EEG signals into corresponding classes then determine the selected object from human.

RGB sensor	1920x1080, 30FPS
Depth sensor	1280x720, 90FPS
Features	Global Shutter, 3μm x 3μm pixel size
Physical	90 mm x 25 mm x 25 mm
Interface	USB-C 3.1

Fig. 2 The specifications of the Intel® RealSense D435i depth camera.

architecture, material, methods and experimental results are demonstrated. Furthermore, this research is working toward to integrate with robotic system, creating an intuitive human machine interaction.

II. Architecture

The full system architecture of this research is shown in Fig. 1. An RGB-D camera was used to capture the color image with depth information to generate point cloud data. After the frame was captured, a deep learning-based multi object detection algorithm was used to calculate the bounding box of region of interest (ROI) image on color image. With the bounding box, the SSVEP flicker can be directly overlapping on the object location on the color image, creating an intuitive BCI scenario for the participants. Following the EEG signals were captured, an SVM-based SSVEP recognition algorithm was used to classify the ROI image that participant is in focused. Finally, the point cloud data of chosen object was used to estimate the position and orientation of the object in Cartesian space than command the robotic arm. In this paper, the SSVEP BCI with object recognition on online camera view was implemented and discussed, this research is still working on further integrating with the robotic systems.

III. Object Recognition on Real-Time Camera View

A. RGB-D camera

In order to recognize the 6 degree-of-freedom (DOF) object information in Cartesian space, an Intel® RealSense

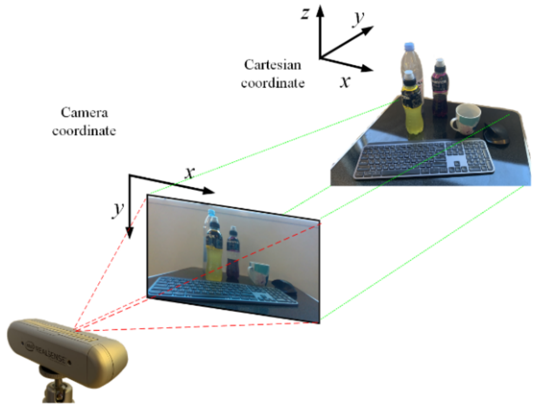

Fig. 3 The relationships between camera coordinate and Cartesian coordinate.

D435i RGB-D camera was used to capture the color and depth image and the point cloud in this research, shown in Fig. 2. To improve the performance of the object recognition, this research applied the deep learning-based object recognition algorithm on color image to locate 2 DOF object position.

B. Object recognition algorithm on color image

Real-time object detector operation on conventional GPU allows their mass usage and proposed an acceptable performance. A YOLOv4 was applied in this research to calculate the bounding box of the ROI image [17]. Meanwhile, the x-axis and z-axis location of the object in Cartesian space can be determined as the relationships shown in Fig. 3.

IV. Brain Computer Interface for Object Selection

The object locations was estimated in previous section. Therefore, the ROIs information are used to determine the location of the SSVEP flickers. In order to design a more intuitive scenario for SSVEP-based BCI for operators, the flickers directly overlaying on the camera view can create a scenario for operators easily and determine the object they want.

A. Scenarios configuration

There are two different SSVEP scenarios were designed in the research to collect the EEG datasets. To collect the dataset, a simple scenario (Scenario 1) using black background with white block flickers located in the four corners. Flickers with ROI image (Scenario 2) is used to collect and enclose the datasets. Even the background is not the camera view in Scenario 1 and 2, considering that the affiect from the layouts, the layouts of flickers is accorded to the real dimension from the detected objects, then displayed at the four corners of the display screen. Finally, a camera view with flickers, overlapping on the object location (Scenario 3), was collected for verifying the SVM-based SSVEP recognizer. In each scenario, most four objects are detected and displayed with flickering frequencies as 7, 9, 11 and 13Hz according to the band-pass filter region to make sure the harmonic frequencies can be retained after the filtering process. The reason that used relatively prime number to design flickers in this research is to avoid the overlapping issue for the harmonic frequencies with the main frequencies. The configurations in this research is shown in Fig. 4.

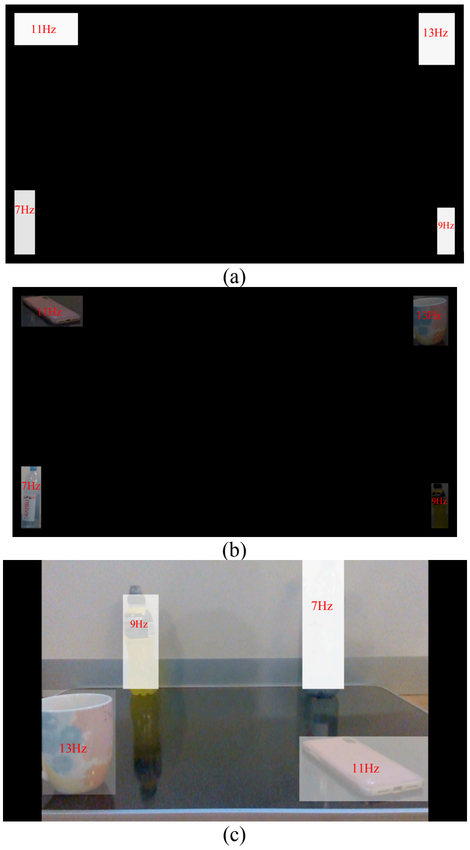

Fig. 4 The three scenario configurations used in this paper: (a) Scenario 1, black background and white block flickers with object layout at four corners, (b) Scenario 2, black background and ROI image flickers at four corners, and (c) Scenario 3, camera view background and overlapping the flicker on the objects.

B. *Feature extraction*

In this paper, the power spectrum derived by overlapping FFT is used as feature for SSVEP recognizer. A 5-s trial using a 2-s window with moving shift for 0.5-s each iteration, shown in Fig. 5, is used to calculate the overlapping FFT power for generating the feature for further classifier. Considering that the EEG signals contain 50Hz AC noise and segment the suitable frequency region for classifier, a four-order Butterworth band-pass filter is applied to filter the noise and remain the power bounded in 3 to 30Hz before FFT.

C. *SVM-based recognizer for SSVEP BCI with error detection*

In order to recognize the EEG signal via SSVEP, a SVM-based recognizer is used in this paper. The SVM was supervised learning model with associated learning algorithms that analyzed data used for classification and regression analysis [18]. In this paper, the SVM classifier was used to recognize the EEG signals into different classes of SSVEP event. During a human operating the BCI robotic arm system with camera view, many reasons may cause that the expected object may not be displayed or detected in the SSVEP scenario. Also, if the human is not focused on the flicker during the data recording period, the recorded EEG signals will not be able to classify into correct classes. However, in the previous research, the input EEG signal is classified forcibly into one most similar defined class, even

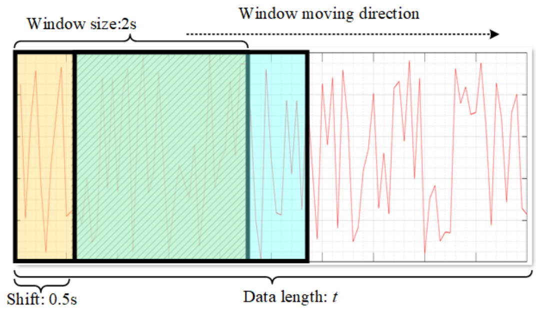

Fig. 5 The overlapping FFT schematic diagram. A 2 seconds window with 0.5 seconds moving shift for *t* seconds data.

Fig. 6 The system setup for the SSVEP BCI experiment.

the data is not belong to any class of them. Therefore, the paper proposed a SVM-based recognizer with radial basis function (RBF) kernel function to solve the above problems. Here, the MATLAB Optimization toolbox is used to calculate the optimized results of the corresponding parameters.

V. EXPERIMENTS RESULTS

A. *Experimental Setup*

In this paper, the EEG signals were acquired using Compumedics NeuroScan SynAmps RT 64-channel amplifier via Curry 8 software. For BCI experiment, a high level stimulating monitor with 144Hz refresh rate, BenQ XL2430T, was used to present stimuli. The subject was requested to focus on the specified object follow the instructions displayed on the screen to collect the data for 7Hz, 9Hz, 11Hz and 13Hz frequencies of the flickers. After the data acquisition, the recorded EEG data from scenario 1and 2 are randomly separated into training and testing dataset by 53% and 47%. The number of collected EEG dataset, including three configurations(scattered, without occluded and occluded) shown in Fig. 7, for training and verifying the SSVEP recognizer is shown in TABLE I. Considering that the Scenario 1 and Scenario 2 are both designed to the black background, the differences between these two scenarios are the flickers use the white block or the ROI image. The EEG data which was collected via Scenario 1 and 2 were used to train and verify the SSVEP recognizer. The EEG data which is collected via Scenario 3 was used to verify proposed recognizer and analyze the effect in this paper. During the training process, 600 iterations were set to train the SVM and minimize the errors.

To simulate and test the recognition performance for the SVM-based SSVEP recognizer in the real scene, there are three configurations, scattered, normal placed without occluded and occluded, are used in Scenario 3.

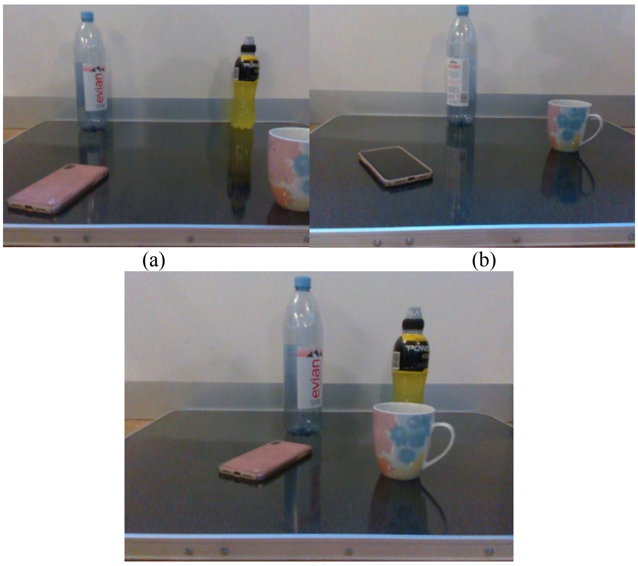

Fig. 7 The three configuration in Scenario 3 to simulate the real scene effect for SSVEP recognizer: (a) objects are scattered on the table, (b) objects are configured with normal placed without occluded between each other on the table and (c) some objects are occluded on the table.

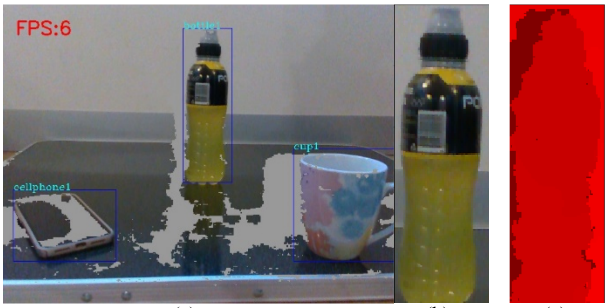

Fig. 8 The captured image: (a) mapped color frame with depth frame and bounding box for objects, (b) cropped ROI color image of bottle and (c) cropped ROI depth image of bottle.

B. *Object Recognition and position estimation results*

In this paper, a laptop configured with an Intel i7-9750H CPU, 32GB RAM and RTX 2060 6GB GPU was used for real-time object recognition from Intel® RealSense D435i RGB-D camera. A result for the captured image from RGB-D camera is shown in Fig. 8. The captured image for color frame, which is mapped with depth frame, and the bounding box for objects is shown in Fig. 8(a). The cropped ROI color and depth image of bottle using the information from YOLOv4 are shown in Fig. 8(b) and Fig. 8(c).

C. *SSVEP experimental results*

The acquired EEG data from the amplifier were 64 channels. In this paper, the O1 and O2 channels are adopted to be the input data in feature extraction, introduced in section IV.B. After that, the features from O1 and O2 channels are fed into the SVM-based SSVEP recognizer.

The of confusion matrix of the SSVEP recognition results for black background test (Scenario 1 and 2) and camera view scenario (Scenario 3) are shown in TABLE II and TABLE III, respectively. There are four terms in the confusion matrix, true positive (TP), false positive (FP), true negative (TN) and false negative (FN). Two factors are used

TABLE I The number of EEG data for training and testing in three scenarios.

Frequency(Hz)	Train	Test	
	Scenario 1&2	*Scenario 1&2*	*Scenario 3*
7	150	129	90
9	155	134	90
11	102	87	60
13	51	44	30
Total	458	394	270

TABLE II The confusion matrix of the SSVEP recognition results for black background test (Scenario 1 and 2).

		Predicted Class					ACC (%)
		7Hz	*9Hz*	*11Hz*	*13Hz*	*Unknown*	
True Class	*7Hz*	119	1	1	2	6	92.2
	9Hz	2	124			8	92.5
	11Hz	4	2	75	1	5	86.2
	13Hz	4		1	38	1	86.4
Precision(%)		92.2	97.6	97.4	92.7	NA	

TABLE III The confusion matrix of the SSVEP recognition results for camera view test (Scenario 3)

		Predicted Class					ACC (%)
		7Hz	*9Hz*	*11Hz*	*13Hz*	*Unknown*	
True Class	*7Hz*	87				3	96.7
	9Hz	1	87			2	96.7
	11Hz			55	1	4	91.7
	13Hz	1	1		25	3	83.3
Precision(%)		97.8	98.9	100	96.2	NA	

to analysis the recognition results for different scenarios, accuracy (ACC) and precision as:

$$\text{ACC}(\%)=\frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

$$\text{Precision}(\%)=\frac{TP}{TP+FP} \tag{2}$$

In TABLE II, the ACC of the proposed SVM-based SSVEP recognizer for Scenario 1 and 2 in 7Hz, 9Hz, 11Hz and 13Hz are 92.2%, 92.5%, 86.2% and 86.4%, respectively. The average ACC for all frequencies is 90.4%. The precision of the proposed recognizer for Scenario 1 and 2 in 7Hz, 9Hz, 11Hz and 13Hz are 92.2%, 97.6%, 97.4% and 92.7%, respectively. As a result, the proposed SSVEP recognizer can effectively classify the EEG signals. The precisions of the proposed recognizer are higher than 92%, the results show that even the quality of the EEG data is not enough to classify into correct classes, the proposed recognizer still can detect the input and classify it into unknown class to avoid the false positive detection.

In TABLE III, the ACC of the proposed SVM-based SSVEP recognizer for camera view scenario in 7Hz, 9Hz, 11Hz and 13Hz are 96.7%, 96.7%, 91.% and 83.3%, respectively. The average ACC for all frequencies is 94.1%. The precision of the proposed recognizer for Scenario 1 and 2 in 7Hz, 9Hz, 11Hz and 13Hz are 97.8%, 98.9%, 100% and

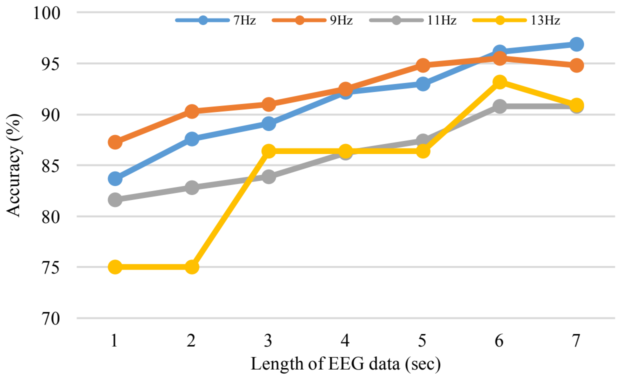

Fig. 9 The accuracy for four classes SSVEP flickers on black background scenario for different input length of EEG data.

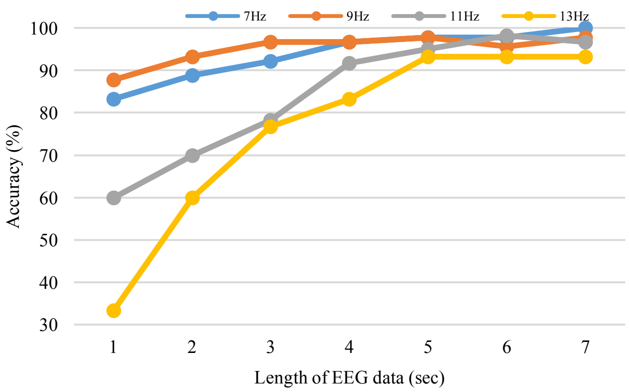

Fig. 10 The accuracy for four classes SSVEP flickers on camera view scenario for different input length of EEG data.

96.2%, respectively. As a result, the proposed SSVEP recognizer can effectively classify the EEG signals. The precisions of the proposed recognizer are higher than 97%, the results show that the proposed recognizer still can detect the input and classify it into unknown class in the camera view scenario, maintaining the performance and avoiding the damage for further combining with the robotic arm.

To discuss about the effect from the length of the EEG data sequence, there are seven different sequence lengths were tested in this paper. The accuracy for four classes of SSVEP flickers on black background scenarios and camera view scenario are shown in Fig. 9 and Fig. 10, respectively. The results show that the accuracy and the length of the EEG data are positive correlation in all scenarios in this paper. However, the participants will feel uncomfortable undergoing the long EEG data acquisition. A balance must be taken a count in the data collection stage for the participants. Therefore, a five second length for each flicker of the data acquisition period was used in this paper.

VI. Conclusions

In this paper, an SSVEP-based BCI with camera view flicker design is implemented. First, a deep learning-based object recognition algorithm YOLOv4 is used to calculate ROI on the color image from Intel® RealSense D435i RGB-D camera. Finally, the object location is used to design the SSVEP flicker in this paper. There are three objects configurations are used in this paper to evaluate real scene effect for SSVEP recognizer. In the experimental results, the moving window is used to calculate the overlapping FFT in feature extraction stage. The average recognition accuracy for camera view scenarios is 94.1%. The results show that the proposed SVM-based SSVEP can effectively recognize the EEG signals into corresponding class. In conclusions, the proposed SSVEP-based BCI can be used in real world application, combing with the flickers and actual objects, based on the upon the scenarios presented. This works is continuous toward to integrate with robotic arm to provide a novel intuitive operation interface.

Acknowledgment

This study was financial sponsored by the Ministry of Science and Technology under the project No. MOST 108-2221-E-027-094-

References

[1] M. Umar Suleman, and M. M. Awais, “Learning from demonstration in robots: Experimental comparison of neural architectures,” *Robotics and Computer-Integrated Manufacturing,* vol. 27, no. 4, pp. 794-801, 2011/08/01/, 2011.

[2] Q. N. Jiang, M. Z. Liu, X. Q. Wang, M. G. Ge, and L. Lin, “Human motion segmentation and recognition using machine vision for mechanical assembly operation,” *Springerplus,* vol. 5, pp. 18, 2016.

[3] A. H. Do, P. T. Wang, C. E. King, A. Abiri, and Z. Nenadic, “Brain-Computer Interface Controlled Functional Electrical Stimulation System for Ankle Movement,” *Journal of NeuroEngineering and Rehabilitation,* vol. 8, no. 1, pp. 49, 2011/08/26, 2011.

[4] X. Chen, Y. Wang, S. Gao, T.-P. Jung, and X. Gao, “Filter bank canonical correlation analysis for implementing a high-speed SSVEP-based brain–computer interface,” *Journal of Neural Engineering,* vol. 12, no. 4, pp. 046008, 2015/06/02, 2015.

[5] S. F. Anindya, H. H. Rachmat, and E. Sutjiredjeki, "A prototype of SSVEP-based BCI for home appliances control." pp. 1-6.

[6] C. Lin, C. Chiu, A. K. Singh, J. King, L. Ko, Y. Lu, and Y. Wang, “A Wireless Multifunctional SSVEP-Based Brain–Computer Interface Assistive System,” *IEEE Transactions on Cognitive and Developmental Systems,* vol. 11, no. 3, pp. 375-383, 2019.

[7] X. Chen, B. Zhao, Y. Wang, and X. Gao, “Combination of high-frequency SSVEP-based BCI and computer vision for controlling a robotic arm,” *Journal of Neural Engineering,* vol. 16, no. 2, pp. 026012, 2019/01/29, 2019.

[8] Y.-T. Wang, Y. Wang, and T.-P. Jung, "A Cell-Phone Based Brain-Computer Interface for Communication in Daily Life." pp. 233-240.

[9] R. Singla, A. Khosla, and R. Jha, “Influence of stimuli colour in SSVEP-based BCI wheelchair control using support vector machines,” *Journal of Medical Engineering & Technology,* vol. 38, no. 3, pp. 125-134, 2014/04/01, 2014.

[10] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single Shot MultiBox Detector," *Computer Vision – ECCV 2016.* pp. 21-37.
[11] R. Girshick, "Fast r-cnn." pp. 1440-1448.
[12] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition." pp. 770-778.
[13] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, “Mobilenets: Efficient convolutional neural networks for mobile vision applications,” *arXiv preprint arXiv:1704.04861*, 2017.
[14] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection." pp. 779-788.
[15] R. Singla, and B. Haseena, “Comparison of ssvep signal classification techniques using svm and ann models for bci applications,” *International Journal of Information and Electronics Engineering,* vol. 4, no. 1, pp. 6, 2014.
[16] N. Kobayashi, and K. Ishizuka, "LSTM-based Classification of Multiflicker-SSVEP in Single Channel Dry-EEG for Low-power/High-accuracy Quadcopter-BMI System." pp. 2160-2165.
[17] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, “YOLOv4: Optimal Speed and Accuracy of Object Detection,” *arXiv preprint arXiv:2004.10934*, 2020.
[18] K. Polat, S. Güneş, and A. Arslan, “A cascade learning system for classification of diabetes disease: Generalized Discriminant Analysis and Least Square Support Vector Machine,” *Expert Systems with Applications,* vol. 34, no. 1, pp. 482-487, 2008/01/01/, 2008.