A Generalized Encoding System for Alpha Oscillations Through Visual Saliency Analysis

Zhen Liang, Fangchao Li, Wanrou Hu, Gan Huang, Shigeyuki Oba, Zhiguo Zhang, and Shin Ishii

***Abstract*— By learning how the brain reacts to external visual stimuli and examining possible triggered brain statuses, we conduct a systematic study on an encoding problem that estimates ongoing EEG dynamics from visual information. A novel generalized system is proposed to encode the alpha oscillations modulated during video viewing by employing the visual saliency involved in the presented natural video stimuli. Focusing on the parietal and occipital lobes, the encoding effects at different alpha frequency bins and brain locations are examined by a real-valued genetic algorithm (GA), and possible links between alpha features and saliency patterns are constructed. The robustness and reliability of the proposed system are demonstrated in a 10-fold cross-validation. The results show that stimuli with different saliency levels can induce significant changes in occipito-parietal alpha oscillations and that alpha at higher frequency bins responded the most in involuntary attention related to bottom-up-based visual processing. This study provides a novel approach to understand the processing of involuntary attention in the brain dynamics and would further be beneficial to the development of brain-computer interfaces and visual design.**

***Index Terms*— EEG, alpha oscillations, brain encoding, visual saliency, involuntary attention.**

I. Introduction

ALPHA oscillations are one of the most prominent indicators in electroencephalogram (EEG) studies [1]–[3]. These oscillations are dominant in the human brain [4] and play a critical role in understanding the patterns in acquired raw signals. In Klimesch's research [5], the significance of alpha oscillations was delineated in cognitive and memory performance. Their study demonstrated that a change in alpha bandpower was positively correlated with brain cognitive functions, and that an increase in alpha bandpower could reflect the maturity process of the brain. Furthermore, alpha oscillations in the frequency range of 8-12 Hz were demonstrated to be critical in the ability to designate attentional processes [6], [7] and were also shown to be associated with cognitive functioning and impairment [8], [9], working memory [10], human emotion [11] and so forth. Previous studies of visual attention also revealed that the alpha oscillations are highly associated with visual processing [4] and selection [12], [13]. For example, an alpha suppression could be considered a significant marker to index the increase in attention level when observing stimuli [14]. To further broaden the current understanding of how alpha oscillations work with visual selective processing, this paper presents a generalized encoding study of alpha oscillations through visual saliency analysis.

In the field of visual attention, the **human visual system** (HVS) is composed of two forms of processing: **bottom-up** and **top-down** processing. In bottom-up processing, **involuntary attention** is allocated to potentially important items in visual information that have discriminant visual features, such as a red spot against a green background. Top-down processing refers to goal-directed visual perception processing, with **voluntary attention** allocated to certain items in visual input that are related to a given task, such as finding a blue balloon from a number of colorful balloons. In other words, the bottom-up processing is a fast and simple procedure triggered by the low-level visual features in the visual content only, whereas the top-down processing is a slow and complex procedure triggered by the prior knowledge and given task. The allocation of attention in these two processing forms is quite different, uncorrelated, and independent [15]. For complex visual observation tasks, the triggered top-down processing could be quite different from person to person, due to differences in education background, experience, religion, and so on. To simplify the question raised in this encoding study, the present work will focus on the simple and fast visual selective processing without any high-level visual understanding involved, namely, bottom-up processing. Meanwhile, inspired by the processing of bottom-up-based visual attention, Koch and Ullman [16] proposed the concept of a **saliency map** to highlight the conspicuousness of pixels in a scene. Later, Itti *et al.* [17] realized the generation

Manuscript received April 16, 2020; revised September 30, 2020; accepted October 29, 2020. Date of publication November 17, 2020; date of current version January 29, 2021. This work was supported in part by the National Natural Science Foundation of China under Grant 61906122, in part by the New Energy and Industrial Technology Development Organization (NEDO), in part by the Post-K Project from Ministry of Education, Sports, Science and Technology (MEXT), and in part by the Japan Society for the Promotion of Science, Grants-in-Aid for Scientific Research (JSPS KAKENHI) under Grant JP19H04180 and Grant 17H06310. *(Corresponding author: Zhiguo Zhang.)*

Zhen Liang is with the Health Science Center, School of Biomedical Engineering, Shenzhen University, Shenzhen 518060, China, and also with the Graduate School of Informatics, Kyoto University, Kyoto 606-8501, Japan (e-mail: janezliang@szu.edu.cn).

Fangchao Li, Wanrou Hu, Gan Huang, and Zhiguo Zhang are with the Health Science Center, School of Biomedical Engineering, Shenzhen University, Shenzhen 518060, China (e-mail: lifangchao2018@email.szu.edu.cn; huwanrou2019@email.szu.edu.cn; huanggan@szu.edu.cn; zgzhang@szu.edu.cn).

Shigeyuki Oba is with the Graduate School of Informatics, Kyoto University, Kyoto 606-8501, Japan (e-mail: oba@i.kyoto-u.ac.jp).

Shin Ishii is with the Graduate School of Informatics, Kyoto University, Kyoto 606-8501, Japan, and also with ATR Neural Information Analysis Laboratories, Kyoto 619-0288, Japan (e-mail: ishii@i.kyoto-u.ac.jp).

This article has supplementary material provided by the authors and color versions of one or more figures available at https://doi.org/10.1109/TNSRE.2020.3038789.

Digital Object Identifier 10.1109/TNSRE.2020.3038789

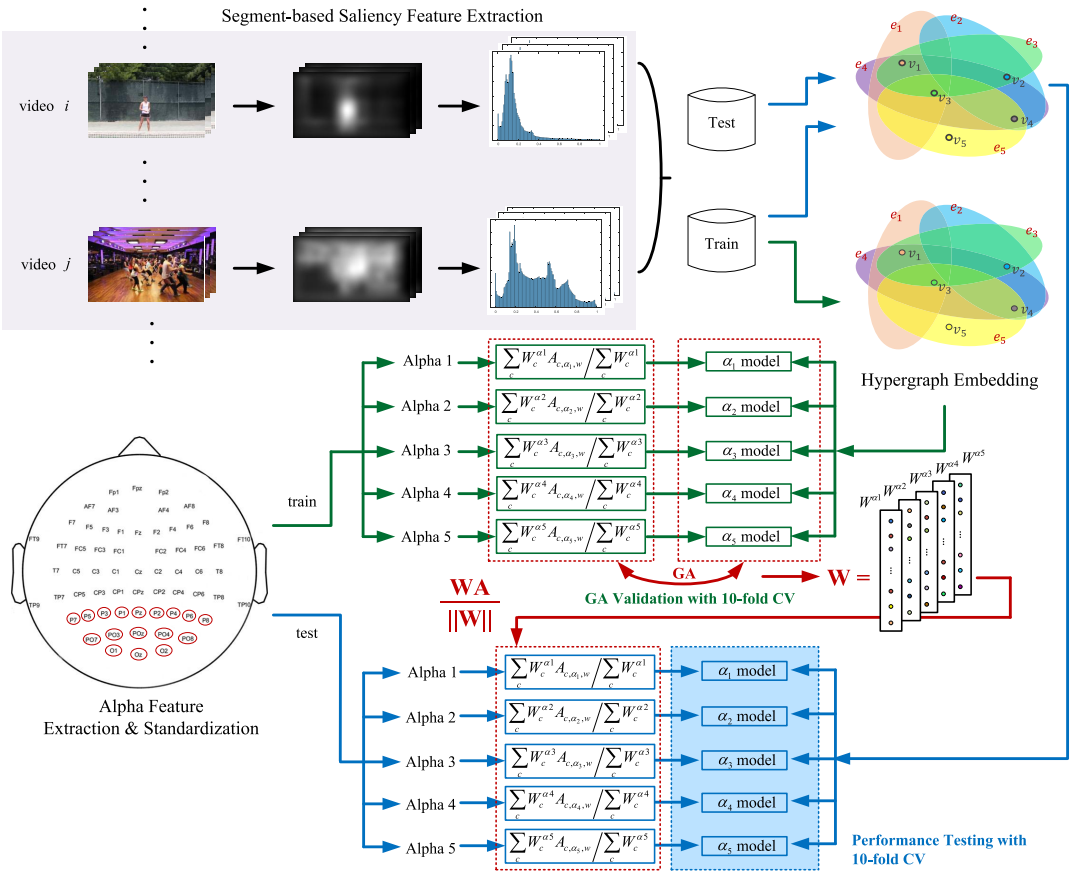

Fig. 1. The proposed generalized encoding system. Saliency information were extracted in terms of frame-based histogram representation and reduced to a lower feature dimensionality by using a hypergraph embedding approach. Then, the extracted saliency information were used to encode the alpha oscillations recorded from the parieto-occipital sites, with a real-valued genetic algorithm based approach and a support vector regression method.

of a saliency map in a computational manner, where pixel conspicuity was evaluated in a topographic structure under the consideration of three primitive visual properties: color, orientation and intensity.

Alpha oscillations are well known to be associated with the processing of visual information and are treated as a marker in many event-related desynchronization (ERD) or event-related synchronization (ERS) studies [18]–[21]. Alpha desynchronization and synchronization reflect certain types of processing, including information receiving, retrieval and storage. In addition, modulations in the alpha rhythm were found to be significantly correlated with the intensity of visual sources [22] and to be influenced by visual selection [13]. In McDermott *et al.* 's flanker task study (including both congruent and incongruent conditions), the results showed that the alpha oscillations played an important role in the task performance, where a significant and robust decrease of the underlying alpha dynamics was observed in the occipital and parietal regions [23]. Instead of using simple visual stimuli as the flanker task, Noah *et al.* extended the neural dynamic study of attentional selective mechanism to natural scene processing with specific targets [24]. Similar observations of neural dynamics in alpha changes were reported, as it was found in spatial attention [25] and feature attention [26]. These results proved that the EEG alpha oscillations (8-12 Hz) are closed related to the attentional selection mechanisms in visual cortex. A decrease of alpha bandpower refers to an enhancement of focal cortical and vice versa. Thus, the alpha oscillations in ongoing brain activity could be considered as the fundamental component of the neural mechanism of attention [24]. On the other hand, how the human brain reacts to the involuntary processing related to a general natural and dynamic visual content, especially the corresponding performance in terms of EEG signals, is still incompletely understood. For a better understanding of the visual attention mechanism, a hypothesis was made that alpha oscillations can also be an indicator of the saliency information in displayed visual stimuli, and the possible associations between these two factors in a generalized encoding system were investigated. The effect of involuntary attention study to the changing, complex, and meaningful real-world is a critical aspect of attentional selection.

In this study, we conducted a visual experiment with 20 healthy subjects and 18 natural videos, and the raw EEG signals were simultaneously acquired. As illustrated in Fig. 1, saliency maps were computed for every video frame in the videos. The corresponding saliency information was extracted in a histogram representation and further reduced to a lower feature dimensionality by hypergraph embedding. Meanwhile,

alpha oscillations in the acquired raw EEG signals were characterized at five alpha frequency bins measured at parieto-occipital sites. Because this encoding study was based on the visual saliency computed in a bottom-up-based computational approach, we only considered the brain areas involved in low-level visual processing. In previous studies, the occipital lobe was recognized as the center of visual processing (e.g., V1, V2, V3, V4, and V5) [27]; the brain responses in the parietal area are related to visual attention [28] and the processing of sensory information, such as spatial awareness, perception and integration [29]–[32]. The analysis of the alpha bandpowers at different frequency bins may yield more accurate information than simply averaging the alpha bandpower in a broad band and may reveal different functional processes in the reflection of saliency information. Furthermore, to determine the contribution of the channels on the alpha reactivities related to visual saliency, a real-valued genetic algorithm (GA)-based approach for optimal channel-weight searching was adopted in which the fitness function was designed under the consideration of both time and space distributions. Then, an alpha index was generated by a weighted summation based on the discovered optimal weights. Subsequently, support vector machine (SVM) regressions were applied to encode the extracted saliency information to the generated alpha indices at each frequency bin in a cross-validation (CV) manner. To the best of our knowledge, no study has examined whether visual saliency can be used to encode alpha dynamics. Through this study, the relationship between brain responses and visual content will be thoroughly examined, and the possibility of using visual inputs to encode brain dynamics in EEG signals will be explored. The main contribution of this study is to explore involuntary attentional selection to the dynamic scenes under naturalistic conditions, instead of using the simple visual tasks or static scenes. The low-level based visual saliency map was introduced to represent the content-specific attentional information, and a neural signature related to the involuntary attention was examined. The proposed generalized encoding model can provide insights into the processing of involuntary attention and offer a novel view regarding the selection of the visual stimuli in the brain-computer interface applications.

II. Experiment

In this study, we conducted a covert brain-visual experiment, where the conductor kept the true intention of this experiment concealed from the participants. This approach offers us a better and valid means to objectively study bottom-up visual processing. The participants were invited to watch different natural videos in a free-watching mode, and the corresponding EEG signals were simultaneously recorded. The data collection pipeline is provided in Appendix A of the Supplementary Materials.[1] This experiment was approved by the Ethics Committee of the Health Science Center, Shenzhen University.

A. Participants and Stimuli

A total of 20 healthy participants (male/female: 11/9; age: 19 to 25 years old, 20.80±1.96; with normal or corrected-to-normal vision) from Shenzhen University participated in the study. All participants had no prior knowledge of the experiment, provided written consent to participate, and had no history of neurological/eye diseases or disorders. Total 18 natural-colored videos (960 pixels width and 540 pixels height) with a length of 58 s were selected from Activity Net.[2] All the videos are color videos and cover different kinds of complex human activities in daily life, such as dancing, playing sports, caring for animals and so on. In the video selection, two criteria were considered. One is the selected videos should cover a certain amount of visually salient parts to guarantee involuntary attention triggered; another is the video content do not include too many unexpected or drastic scene changes to eliminate any surprise factor during the data collection [33]. During data collection, the videos were displayed on a 24.5-inch Dell Alienware Aw2518H monitor. The display resolution was 1920 × 1080, and the refresh rate was 240 Hz. At a fixed viewing distance of 60 cm, it provided a subtended visual angle of 48.75 degree × 28.31 degree. The average frame rate of the video stimuli was 28.81 fps. To dissociate the auditory effect in the recorded EEG signals, all the videos were played with the sound muted.

B. Experimental Design and Data Collection

For each participant, the brain-visual experiment included a total of 18 trials (corresponding to the 18 selected videos). To exclude other factors that would affect fluctuations in alpha waves, we introduced a baseline collection before playing the videos. Thus, one trial consisted of a 4 s baseline and a 58 s video task:

- *baseline:* A white fixation point was shown at the center of a black screen. The participants were instructed to relax and maintain a clear mind; the collected EEG data are referred to as **baseline data** below.
- *video playing:* The videos were displayed in a randomized sequence. The participants were instructed to watch the videos in a free-viewing mode, without thinking about anything else; in other words, the participants should be fully engaged in watching the videos. The collected EEG data are referred to as **video data** below.

During data collection, the participants were allowed to freely blink and move their eyes. To encode the brain activities from the visual information in a real visual processing situation, there was no specific task assigned to the participants during video watching. To minimize possible artifacts in the recorded EEG signals induced by head or body movements, a chin-rest was used to immobilize the participant's head. Here, EEG signals were recorded by a BrainAmp system (Brain Products GmbH, Germany) using 64 Ag/AgCI electrode channels. In the BrainAmp system, the online filter was a fifth-order butterworth filter with 30 dB/octave. The corresponding lower and high cutoffs were 0.016 Hz/10s and 1000 Hz, respectively. More technical specifications are available here.[3] The placement of the electrode channels followed the standard international 10-20 electrode system,

[1] https://drive.google.com/file/d/1gTVhnEdePN45FAAFKlHugI7iw8Gs-oPn/view?usp=sharing

[2] http://activity-net.org

[3] https://www.brainproducts.com/productdetails.php?id=1

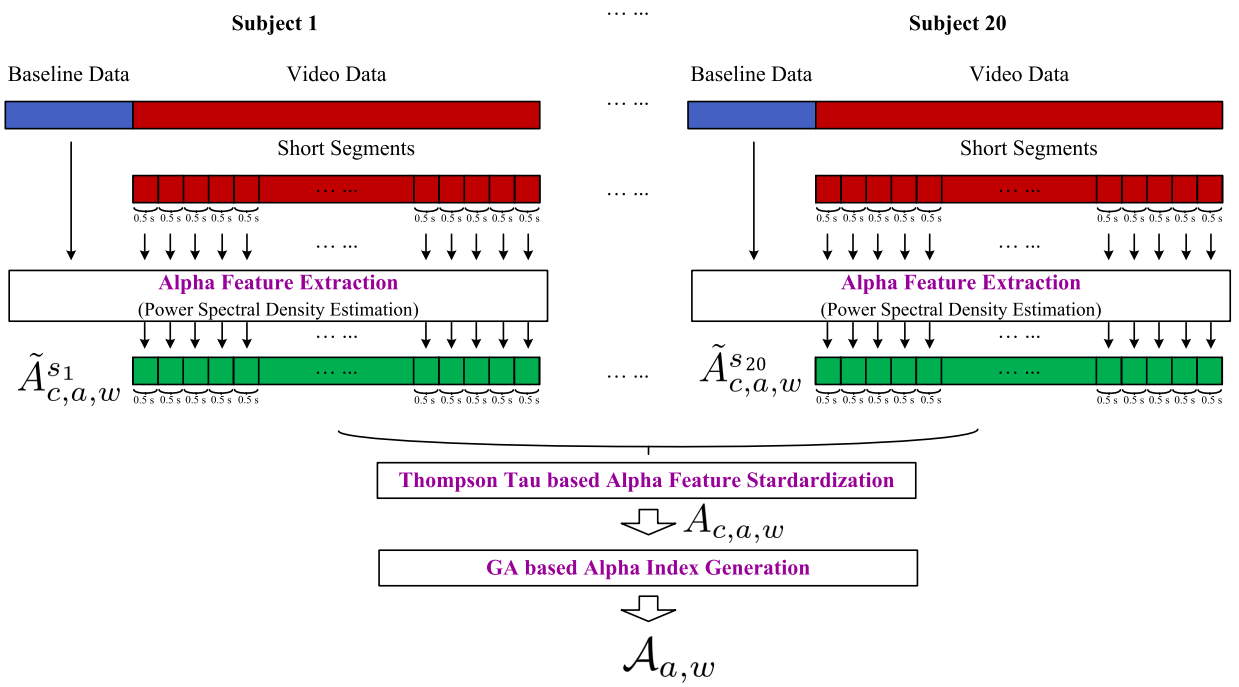

Fig. 2. The feature extraction pipeline of the alpha indices. First, alpha spectral powers were extracted from baseline and video data, respectively; Second, the bandpower changes during video watching were measured by normalizing the video feature to the baseline feature; Third, an alpha feature standardization was conducted to aggregate the features from participants with a Thompson Tau method; Finally, a real-valued genetic algorithm based approach was introduced to fuse the extracted features from different channels into an alpha index.

where the placement of electrodes were based on the relative interelectrode distances of either 10% or 20% of the total front-back or left-right distance of the skull [34]. The electrode placements with the GND and reference locations are shown in Appendix A of the Supplementary Materials. The recorded signals from each channel were digitized at a sampling rate (*fs*) of 1000 Hz. The impedance for all channels were kept lower than 5 kΩ in the experiments. Note that the participants kept their eyes open throughout the entire experiment.

III. Methodology

A. EEG Data Analysis

1) Preprocessing: To remove the common artifacts from the acquired raw EEG, such as physiological artifacts (e.g. ocular activity, cardiac activity, muscle activity, respiration) and non-physiological artifacts (e.g. body movement, AC electrical, eletromagnetic inferences), preprocessing was first conducted in a standard procedure. Here, after IO channel removal and FCz channel interpolation, we re-referenced the signals using a common average reference by calculating the mean signal from all the EEG channels and this mean was subsequently extracted from the signals of each channel. Then, we filtered the raw signals by a butterworth filter of [1 Hz, 45 Hz] and conducted a notch filter at 50 Hz, where the DC components at the low frequencies, high frequency artifacts and 50 Hz power-line interference from the recording signals were removed. It was proven that blind source separation using an independent component analysis (ICA) is an efficient approach to remove the artifacts [35]. We computed ICA components by using runica algorithm as implemented in EEGLAB [36] and removed the independent components that capture the artifacts such as eye blinks, eye movements, muscle activity and so on. Finally, clean EEG signals were reconstructed based on the remaining ICA components. More details about every step in the preprocessing procedure were clearly clarified in Appendix B of the Supplementary Materials. As proven in the existing studies, the occipital lobe is the visual processing center, and the parietal lobe plays a critical role in the integration of sensory information in the human vision. To focus on the EEG dynamics involved in low-level visual processing and to eliminate the influence of high-level visual processing, we selected only the channels in the parietal and occipital lobes (a total of 17 channels were selected, which are highlighted by red circles in Fig. 1). The corresponding alpha indices from these selected channels were further extracted, as illustrated in Fig. 2.

2) Alpha Feature Extraction: The commonly used representative features from the acquired EEG signals are the spectral power distributions at different frequency bins, which are characterized by power spectral density (PSD) estimation in the frequency domain. In this study, we focused on exploring the visual saliency effect on the spectral power distributions in a possible alpha-located frequency range ([8 Hz, 12 Hz]). Rather than directly averaging all the power distributions on a broad alpha frequency range, we computed the alpha spectral power and evaluated the alpha at each frequency bin, as given in Table I. This approach allowed us to examine the encoding study of each individual alpha bandpower separately and determine which alpha frequency corresponded the most to the saliency information in the presented stimuli.

TABLE I
THE ALPHA FREQUENCY BINS EVALUATED IN THIS STUDY

	$\alpha1$	$\alpha2$	$\alpha3$	$\alpha4$	$\alpha5$
frequency bin	8 Hz	9 Hz	10 Hz	11 Hz	12 Hz

In a single trial, the recorded raw EEG signals in the baseline and video playing sessions were denoted E_c^b and E_c^v, respectively, where c indicates the selected EEG channels ($c = 1, \ldots, c_n$). To thoroughly examine the changing trend of alpha oscillations in the process of video watching, we first measured the alpha powers in E_c^b and then used them as references to measure the bandpower changes in E_c^v. Specifically, the alpha bandpowers were extracted from the 4 s baseline data by the power spectral estimation algorithm [37] with a 50% overlap and a Hamming window, and these bandpowers were denoted $\dot{A}_{c,a}^b, a = \alpha1, \ldots, \alpha5$. To better assess alpha dynamics over time, rather than extracting bandpower information from the entire 58 s of video data, we first divided E_c^v into a number of short segments with a time length of 0.5 s ($E_{c,w}^v$, where $w \in [1, w_n]$ refers to the EEG-based segment index). Note that no content overlap occurred between any of the adjacent short segments ($\{E_{c,w}^v\}_{w=1}^{w_n}$). The alpha bandpowers were also extracted from each segment as the baseline data, which were denoted $\dot{A}_{c,a,w}^v, a = \alpha1, \ldots, \alpha5$ and $w = 1, \ldots, w_n$. Subsequently, to establish the elicited brain activities in the process of video viewing, the extracted $\dot{A}_{c,a,w}^v$ were normalized by $\dot{A}_{c,a}^b$ as

$$\tilde{A}_{c,a,w} = \dot{A}_{c,a,w}^v - \dot{A}_{c,a}^b, \tag{1}$$

where $\tilde{A}_{c,a,w}$ can be considered the alpha dynamic that was caused only by the displayed video content.

3) Alpha Feature Standardization: In the previous section III-A.2, the alpha features were extracted from a single participant's individual trial data. Because the focus of the present paper is on a cross-subject study and a generalized encoding system, after feature processing and feature extraction, the features extracted from the participants were aggregated. In the past studies, the commonly used fusion approach is to simply average the extracted features from all the participants. However, this average based method would include the contaminated feature(s) (also called "noise") in the feature fusion processing and further affect the following encoding model proposal. To eliminate the effect of the "noise" in the features, we suggested to check the extracted feature quality first and cancel out the "noise" before feature fusion / standardization. In other words, only the shared common patterns across participants should be remained and aggregated. Suppose that the extracted alpha features from different participants while watching the same stimuli were denoted $\tilde{A}_{c,a,w}^s$, where s is ranging from 1 to s_n (here, the feature size was equal to 20, which was the same as the subject number s_n). Furthermore, $\{\tilde{A}_{c,a,w}^s\}_{s=1}^{s_n}$ were integrated by means of feature standardization and formed into a single participant-independent feature vector to capture the common EEG dynamics evoked by the presented video. In this paper, a statistical-based feature standardized approach called the Thompson Tau method was adopted to adaptively filter the "noises" from $\{\tilde{A}_{c,a,w}^s\}_{s=1}^{s_n}$ as described below. Compared to the other popular outlier detection methods such as Z-score and K-means, the Thompson Tau method has been proven as a credible, useful and high-efficient method in various studies [38]–[41], which performed reliably and provided an objective approach to determine whether the sample in the data distribution is an anomaly point.

1) Calculate the respective mean and standard deviation as

$$\mu_{\tilde{A}_{c,a,w}} = \frac{\sum_{s=1}^{s_n} \tilde{A}_{c,a,w}^s}{s_n}, \tag{2}$$

$$\sigma_{\tilde{A}_{c,a,w}} = \sqrt{\frac{\sum_{s=1}^{s_n} (\tilde{A}_{c,a,w}^s - \mu_{\tilde{A}_{c,a,w}})^2}{s_n - 1}}. \tag{3}$$

2) Compute the Thompson Tau value, as

$$tau = \frac{t_{\alpha/2} \times (s_n - 1)}{\sqrt{s_n} \times \sqrt{s_n - 2 + t_{\alpha/2}^2}}, \tag{4}$$

where $t_{\alpha/2}$ is the $\alpha/2\%$ point of Student's t-distribution ($\alpha = 0.05$).

3) Measure the value of TS as

$$TS = tau \times \sigma_{\tilde{A}_{c,a,w}}. \tag{5}$$

4) Remove the outliers ("noises") as presented in Algorithm 1. For clarity, we denoted the returned $\{\tilde{A}_{c,a,w}^s\}_{s=1}^{s_n}$ as $\{\hat{A}_{c,a,w}^s\}_{s=1}^{\hat{s}_n}$, where the remaining number of features was $\hat{s}_n$ ($\hat{s}_n \leq 20$).

Algorithm 1 Outlier Removal based on the Thompson Tau Method

```
s = 1;
while do
    if |Ã^s_{c,a,w} − μ_{Ã_{c,a,w}}| > TS then
        remove Ã^s_{c,a,w} from {Ã^s_{c,a,w}}_{s=1}^{s_n}, and update s_n =
            s_n − 1 and {Ã^s_{c,a,w}}_{s=1}^{s_n}
    else
        s = s + 1
    end if
    if s > s_n then
        break
    end if
end while
return the final obtained {Ã^s_{c,a,w}}_{s=1}^{s_n}
```

5) Form the generalized alpha feature across the remaining clean features as

$$A_{c,a,w} = \frac{\sum_{s=1}^{\hat{s}_n} \hat{A}_{c,a,w}^s}{\hat{s}_n}, \tag{6}$$

where c, a and w refer to the EEG channel, alpha frequency bin and short segment, respectively. In practice, feature standardization was conducted individually for every combination of alpha frequency bins (5), segments (116), channels (17) and videos (18). For each combination, the feature size of the inputs to the Thompson Tau outlier detection method were 20 (equal to the total number of participants). The total

number of the combinations was 177480 ($5 \times 116 \times 17 \times 18 = 177480$). We checked the outlier removal results in the total 177480 combinations, and found the average and standard deviation of the removed feature samples across all the combinations were 1.35 and 0.56, respectively. In other words, the average number of samples for each of the remaining clean features in step 5) for further feature standardization across all the combinations was equal to 18.65 ($20 - 1.35 = 18.65$).

4) GA-Based Alpha Index Generation: A key question is how to combine the extracted features from different channels. Based on the principle of natural genetic selection operations, the genetic algorithm (GA) [42], a probabilistic and heuristic optimization technique, was proposed to learn and discover optimal solutions that can produce promising performances and are computationally practical. The conventional GA is a binary coding method based on the strings of 0 and 1, which would cause inefficient occupation of computer memory and are not sufficiently flexible to solve complex problems. To overcome these limitations, we adopted a real-valued GA that utilizes floating point codings in its solutions [43]. The real-valued GA performed in a more straightforward and efficient manner than the binary GA. The benefits of the real-valued GA technique have been widely demonstrated in various search problems, such as feature weight learning [44]–[46].

In the real-valued GA technique, the possible solutions are named chromosomes. After initialization, a set of chromosomes form the very first population. The merit of each chromosome is evaluated by the predefined **fitness function**, and only the chromosomes with high fitness values are selected and remained in the next generation of the population. Additionally, to increase the population diversity, the selected chromosomes are used to generate new chromosomes through two manners: mutation and crossover. Mutation is conducted based on the existing chromosomes to increase the randomness of the chromosomes and decrease the possibility of reaching a local optimum, whereas crossover is based on two chromosomes to generate a new offspring. After repeated alternations of the population selection and generation steps, the searching process terminates once the maximum generation number (GA_M) is reached. In this study, we applied the real-valued GA technique to explore the optimum weights $\mathbf{W} = \{W_c^a\}_{c=1}^{c_n}$ of the selected EEG channels $\{c\}_{c=1}^{c_n}$ for each individual alpha frequency bin $\{a\}_{a=\alpha 1}^{\alpha 5}$, where one set of weights is one chromosome. The obtained optimum weights $\{W_c^a\}_{c=1}^{c_n}$ can be interpreted as indicators of the importance of the EEG channels that contribute to the alpha dynamics related to visual saliency during video viewing. Then, the extracted alpha features $\{A_{c,a,w}\}_{c=1}^{c_n}$ were weighted summed into an alpha index, as

$$\mathcal{A}_{a,w} = \sum_{c=1}^{c_n} W_c^a \times A_{c,a,w}. \tag{7}$$

To quantify the quality of a chromosome and guide population selection and generation toward globally optimal solutions, we designed the fitness function as

$$f_{GA} = |eval_1| \times \exp(eval_2 + 1), \tag{8}$$

where $eval_1$ and $eval_2$ are two evaluation standards of regressions between the prediction and target data: the Pearson correlation coefficient and NMSE. A high correlation coefficient indicates that the prediction and target data share a similar trend, while a high value for NMSE indicates the deviation between the prediction and target data is small in both space and time. Thus, the global optimum in this search problem should be the one that can maximize both the correlation coefficient and NMSE, namely, the optimum that can maximize the given fitness function. The operation process for GA-based weight searching was independently performed for each individual alpha frequency bin. In the implementation, each chromosome was a vector $\{W_c^a\}_{c=1}^{c_n}$, which was initialized by random numbers in the range of $[0, 1]$. The size of the initial population GA_N and the maximum generation number GA_M were varied from 10 to 1000, and the corresponding performances are compared in Section IV. Moreover, the non-uniform mutation and arithmetic crossover approaches were utilized to generate new chromosomes. The normalized geometric selection probability was equal to 0.08, which was demonstrated as an optimal setting in the GA based function optimization applications [47].

B. Video Data Analysis

In this study, we introduced video saliency to encode EEG dynamics via time-series regressions. As illustrated in Fig. 4, the video-based saliency maps were first computed at every video frame and were then characterized and embedded in the form of short segment-based structures. More details about the computation process are reported in Appendix C of the Supplementary Materials.

1) Frame-Based Saliency Map Generation: The model of Itti *et al.* is the most known bottom-up-based computational visual attention model [17], and it was applied to compute the visual saliency of the video frames in this study. Figure 3 shows an example result produced by Itti's model, where the contrast difference in the input was first examined according to three primitive visual features, namely, color, intensity, and orientation (denoted S_f^C, S_f^I and S_f^O, respectively), and then formed into one final saliency map (denoted S_f). In the example shown in Fig. 3, the final saliency map revealed that the woman in the center was assigned the highest saliency values close to 1 (the bright areas), while the background was close to 0 (the dark areas). To avoid confusion, the primitive feature-based saliency results are referred to as feature maps below. In the encoding studies, in addition to evaluating the encoding performance using S_f, we also investigated the encoding effects with each type of feature map.

2) Segment-Based Saliency Map Generation: For each video frame V_f in one video V, the corresponding feature maps and the final saliency map were computed as presented in Section III-B.1 and were then downsampled to a lower spatial resolution (18×32). Notably, the values in the obtained feature/saliency maps reflected the conspicuousness of pixels in terms of each feature or the overall conspicuousness of pixels in the video frame across all the features, which were all in the range of $[0, 1]$. Similar to the feature extraction

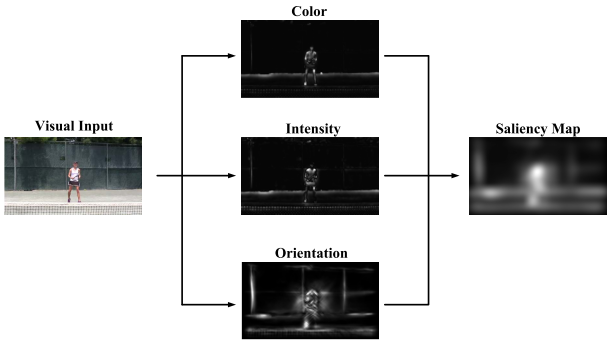

Fig. 3. An example result of Itti *et al.* model. The feature map of the input image was first evaluated in three primitive visual features (color, intensity and orientation) and then formed into one saliency map.

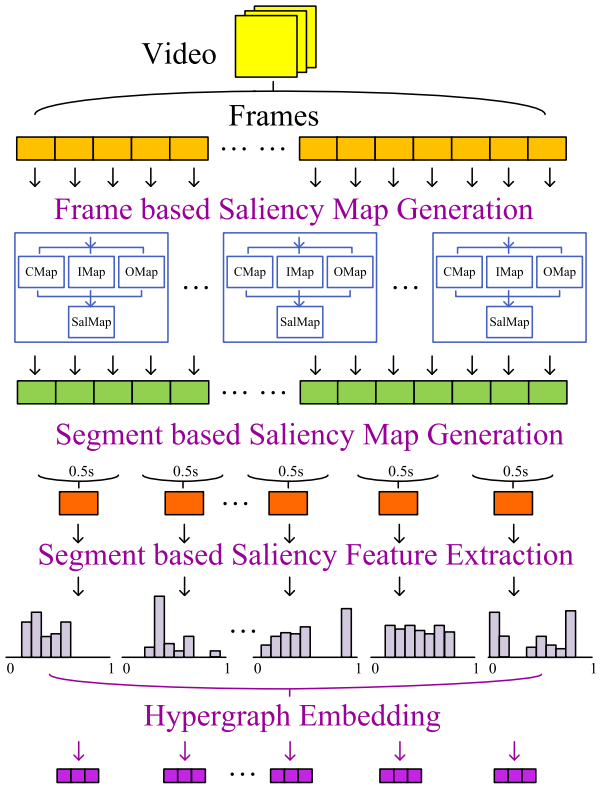

Fig. 4. The flowchart of saliency feature extraction. The saliency maps of video frames were first computed. The computed frame-based saliency maps were then integrated into segment-based saliency maps and a 128-bin histogram representation of each segment-based saliency map was characterized. To lower computation complexity and remove redundant features / noise, a hypergraph embedding method was introduced to capture the essence of the data and reduce the saliency representation to a lower feature space.

process in the raw EEG signals, the saliency feature extraction from videos was also implemented in terms of short segments with a length of 0.5 s. For one video, we integrated the obtained frame-based feature/saliency maps to segment-based feature/saliency maps, denoted S_w^C, S_w^I, S_w^O and S_w.

3) Segment-Based Saliency Feature Extraction: A histogram-based feature representation is a simple and direct method that shows good results in various applications [48], [49]. Compared to a single feature index (e.g., average or medium) to describe the distribution, a histogram-based feature representation can provide a more accurate description to show the probability estimation of the underlying distribution of the input. In the present study, we adopted the histogram-based feature representation to characterize the obtained segment-based feature/saliency maps as follows. (1) define the number of bins: the number of bins b_n was set to 128 here, as the other histogram-based feature representation studies; (2) calculate the bin width and edges: as the defined bins ($b_p = 1, \ldots, b_n$) were a series of consecutive and non-overlapping intervals with equal size, the bin width could be calculated as the entire value range in feature/saliency maps divided into b_n. In this paper, as the range of saliency map was normalized to the range of [0,1], the width of each histogram bin was equal to $1/128$; (3) count the number of values in each bin: counted how many values in feature/saliency maps fall into each bin according to the defined bin width and edges; (4) normalization: to render the histogram-based feature representations comparable among different segments and videos, normalization was conducted by diving the number of pixels in the saliency map. More details are reported in Appendix C of the Supplementary Materials. The obtained normalized features for feature and saliency maps were denoted $\tilde{C}^C_{b_p,w}$, $\tilde{C}^I_{b_p,w}$, $\tilde{C}^O_{b_p,w}$ and $\tilde{C}_{b_p,w}$, respectively.

4) Hypergraph Embedding: The size of the extracted segment-based saliency feature vector $\{\tilde{C}_{b_p,w}\}_{b_p=1}^{b_n}$ was 1×128 ($M = 128$). In practice, reducing the feature dimensionality to a lower dimension is necessary and beneficial, e.g. $M' < M$. For dimensionality reduction with graph embedding techniques, the common approach is to construct an undirected weighted graph and identify an optimal surface to represent the original feature space. However, traditional graph embedding approaches may cause substantial loss of information because only simple pairwise relations are considered. To avoid misleading representation and information loss, hypergraph embedding provides a more natural method to identify the optimal subspace, where multiple or higher-order relationships among vertices are captured. In this study, we treated each segment-based saliency feature vector as a vertex and formed a corresponding hypergraph. A hypergraph, G_h, is composed of a number of vertices V_h ($v_{h,i} \in V_h, i = 1, \ldots, |V_h|$) and hyperedges E_h ($e_{h,j} \in E_h, j = 1, \ldots, |E_h|$), where a hyperedge can connect more than two vertices. $v_{h,i}$ and $e_{h,j}$ are a vertex and a hyperedge on the hypergraph, respectively. In this study, a hyperedge was formed by a centroid vertex (the centroid of the hyperedge) and its nearest neighbors. Each vertex in the database was used as the centroid once. Thus, the formed number of hyperedges was the same as the number of vertices. More details about hypergraph construction are provided below. (1) Set the hyperedge size to τ_h. (2) Set a vertex as a centroid and measure the pairwise similarities between this centroid vertex and other available vertices. The similarity between any two vertices was defined as

$$\zeta(v_{h,i}, v_{h,i'}) = \frac{1}{1 + dist_{i,i'}}, \tag{9}$$

where $dist_{i,i'}$ is given as

$$dist_{i,i'} = \sum_{b_p=1,\ldots,b_n} \frac{(\tilde{C}_{v_{h,i}}(b_p) - \tilde{C}_{v_{h,i'}}(b_p))^2}{\tilde{C}_{v_{h,i}}(b_p) + \tilde{C}_{v_{h,i'}}(b_p)}, \tag{10}$$

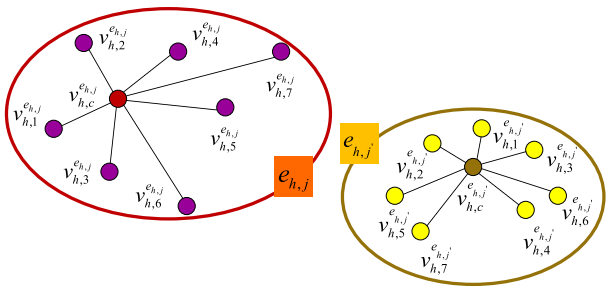

Fig. 5. An illustration of hyperedge formation. A centroid vertex with the nearest vertices were formed a hyperedge.

where $\tilde{C}_{v_{h,i}}$ and $\tilde{C}_{v_{h,i'}}$ are the extracted histogram representations of vertices $v_{h,i}$ and $v_{h,i'}$. (3) Rank the similarities in descending order. (4) Form a hyperedge by connecting the centroid vertex with the top $\tau_h - 1$ nearest vertices. Examples of the formed hyperedges are shown in Fig. 5. (5) Repeat Steps (2) to (4) until each vertex is treated as a centroid once and the corresponding hyperedge is formed. Consequently, the number of the constructed hyperedges was the same as the number of vertices. The corresponding incidence matrix H_h was a $|V_h|$-by-$|E_h|$ matrix defined as

$$h_h(v_{h,i}, e_{h,j}) = \begin{cases} 1 & \text{if } v_{h,i} \in e_{h,j} \\ 0 & \text{if } v_{h,i} \notin e_{h,j}. \end{cases} \tag{11}$$

The weight of one hyperedge was a summation of the similarities between the centroid vertex and the top $\tau_h - 1$ nearest vertices,

$$w_h(e_{h,j}) = 1 + \sum_{i_n=1,\ldots,\tau_h-1} \zeta(v_{h,c}^{e_{h,j}}, v_{h,i_n}^{e_{h,j}}), \tag{12}$$

where $v_{h,c}^{e_{h,j}}$ is the centroid vertex of the hyperedge $e_{h,j}$ and $\{v_{h,i_n}^{e_{h,j}}\}_{i_n=1}^{\tau_h-1}$ are the connected top nearest vertices. The hyperedge weight matrix W_h was a $|E_h|$-by-$|E_h|$ diagonal matrix in which the diagonal elements were the hyperedges' weights calculated by Eq. 12. The degree of a vertex $v_{h,i}$ was a summation of all hyperedge weights of all the hyperedges belonging to the matrix, as

$$d_v(v_{h,i}) = \sum_{v_{h,i} \in e_h^{v_{h,i}}} w_h(e_h^{v_{h,i}}), \tag{13}$$

where $e_h^{v_{h,i}}$ are all the hyperedges that were connected to the vertex $v_{h,i}$. The degree of a hyperedge $e_{h,j}$ was a summation of all the connected vertices, as

$$d_e(e_{h,j}) = \sum_{v_h^{e_{h,j}} \in e_{h,j}} h(v_h^{e_{h,j}}, e_{h,j}), \tag{14}$$

where $v_h^{e_{h,j}}$ refers to all the connected vertices in the hyperedge $e_{h,j}$. Therefore, the vertex degree matrix D_v and the hyperedge degree matrix D_e were constructed accordingly, which were used to compute the diagonal elements $d_v(v_{h,i})$ and $d_e(e_{h,j})$, respectively. The optimal subspace of the constructed hypergraph was the eigenvectors with the smallest nonzero eigenvalues of the computed hypergraph Laplacian, established as

$$\Delta = I - D_v^{-(1/2)} H_h W_h D_e^{-1} H_h^T D_v^{-(1/2)}. \tag{15}$$

In practice, we calculated the hypergraph Laplacian Δ, measured the corresponding eigenvalues, sorted them in ascending order, retained only the eigenvectors whose eigenvalues were the top M' smallest eigenvalues, and formed a new and optimal surface of feature representations. The new feature surface with the lower dimension was denoted C_w. In a similar manner, the feature representations of the feature maps $\tilde{C}_{b_p,w}^C$, $\tilde{C}_{b_p,w}^I$, and $\tilde{C}_{b_p,w}^O$ were also reduced to a lower dimension, which were denoted C_w^C, C_w^I, and C_w^O, respectively. Later, the new feature representations were used to estimate the extracted alpha index $\mathcal{A}_{a,w}$ in a supervised learning approach.

IV. Experimental Results

A. Cross-Validation Based Encoding Results

In this study, we conducted continuous encoding from visual saliency to alpha powers at each frequency bin, with a time length of 0.5 s. In total, we had 2088 samples in the database for modeling (18 videos $\times$ 116 short time segments). The proposed generalized encoding system was evaluated through a 10-fold cross validation (CV) to assess the robustness of performance on an independent / new dataset in a statistical analysis approach. To measure the encoding performance in both the space and time domains, the encoding performance was evaluated with both Pearson correlation coefficient and normalized mean square error (NMSE) (definitions given in Appendix D of the Supplementary Materials). In this study, all the presented correlation coefficient results had a significance level that was much smaller than 0.05. The value of hyperedge size τ_h and the feature dimensionality M' were set to 30 and 110 empirically. A discussion on the two value selections were reported in section IV-D.

We constructed a series of support vector regression (SVR) models to predict the extracted alpha indices $\mathcal{A}_{a,w}$ by using the extracted saliency representation C_w in the final saliency map S_f, and we evaluated the results in 10-fold CV in terms of the correlation coefficient, NMSE and fitness function (the model construction and evaluation process refer to Appendix E of the Supplementary Materials). The overall results across all alpha frequency bins are presented in Table II, where the initial population (GA_N) and the maximum generation number (GA_M) were adjusted from 10 to 1000. In accordance with our predictions, the results showed that an increase in GA_N and GA_M values led to better encoding performance. The best results were obtained when $GA_N = 1000$ and $GA_M = 1000$, where the correlation coefficient and NMSE were equal to 0.6207 and -2.4625, respectively. The best calculated fitness value (f_{GA}) was 0.6544. The results showed an increase of population size and generation number could lead to an improvement of the encoding performance in terms of correlation coefficient, NMSE and f_{GA}. This is consistent with the idea of genetic algorithm: the larger the population size and the generation number are, the greater chance to reach to an optimal solutions after iterative mutation and crossover [50]. Moreover, to better illustrate the encoding behavior at each alpha frequency bin, we also examined the corresponding encoding results in Table III. We found that the encoding performances in higher frequency bins, such as $\alpha 3$, $\alpha 4$

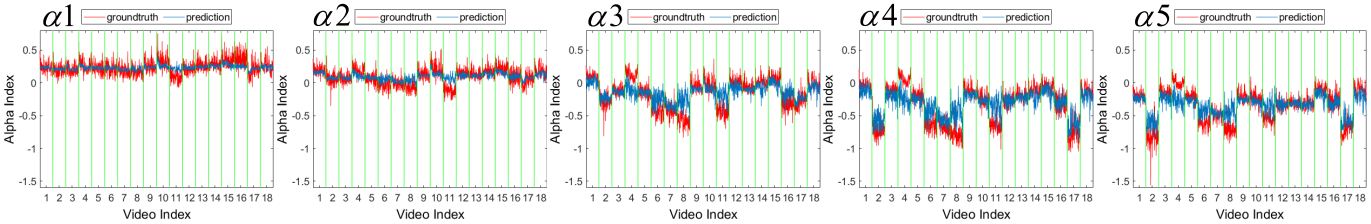

Fig. 6. The 10-fold CV encoding results of $\mathcal{A}_{a,w}$ by using C_w in an alignment of video sequences. Red and blue lines indicate the true value and predicted value of $\mathcal{A}_{a,w}$, respectively. Compared to $\alpha 1$ and $\alpha 2$, the estimated alpha dynamics in $\alpha 3$, $\alpha 4$, and $\alpha 5$ showed better encoding performance.

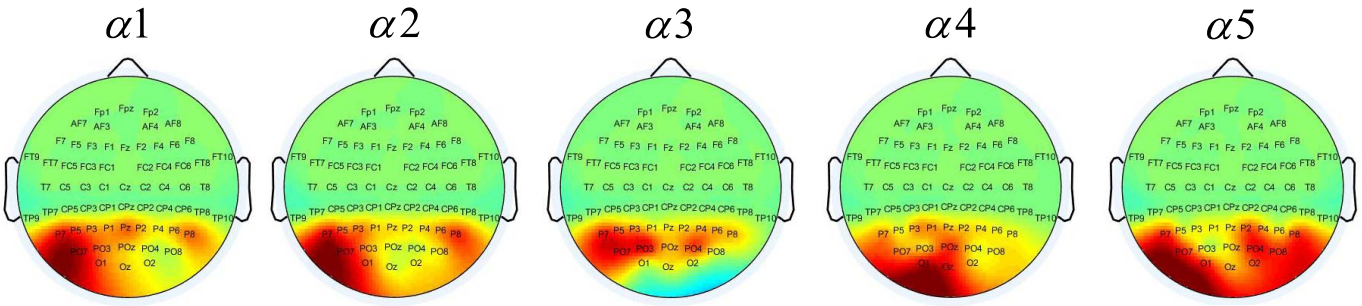

Fig. 7. The final discovered optimum channel weights at each alpha frequency bin. Here, the map colors were all scaled to the range of [−1,1]. It shows O1 in $\alpha 4$ responses contributed the most to visual saliency, with a channel weight of 0.8973.

TABLE II
THE OVERALL 10-FOLD CV PERFORMANCE OF $\mathcal{A}_{a,w}$ BY USING C_w, C_w^C, C_w^I, AND C_w^O, RESPECTIVELY

	GA_N	GA_M	corr	NMSE	f_{GA}
C_w	10	10	0.6083	-2.8459	0.5587
	100	100	0.6154	-2.6181	0.6270
	1000	1000	0.6207	-2.4625	0.6544
C_w^C	10	10	0.5959	-3.0863	0.5459
	100	100	0.6085	-2.6696	0.6183
	1000	1000	0.6093	-2.6838	0.6427
C_w^I	10	10	0.6444	-2.1077	0.7166
	100	100	0.6561	-1.9043	0.8140
	1000	1000	0.6583	-1.9211	0.8400
C_w^O	10	10	0.5388	-4.3289	0.3291
	100	100	0.5498	-3.9541	0.3850
	1000	1000	0.5590	-3.6792	0.4152

and $\alpha 5$, were significantly better than those in lower frequency bins ($\alpha 1$ and $\alpha 2$). The best encoding performance occurred in $\alpha 4$ with a significant correlation coefficient of **0.74** and an NMSE of $-$**0.30**, where the fitness value was equal to **1.48**. These results suggest that, comparing to the frequency range of 8 to 9 Hz, the alpha oscillations in the frequency range of 10 to 12 Hz were more sensitive to saliency changes in the visual content (better encoding performance achieved). The results are consistent with the general agreement that upper alpha could reflect the processing of sensory information [51]–[53]. We also assessed the true and predicted values of C_w in an alignment of video sequences from 1 to 18, as shown in Fig. 6. The line distribution of the true value of C_w (red line) further verified our assumption that the alpha dynamics in $\alpha 1$ and $\alpha 2$ did not vary significantly in their responses to a change in the video content. Rather, the estimated results (blue line) in $\alpha 3$, $\alpha 4$, and $\alpha 5$ showed promising coincidences with the true alpha responses while watching videos.

B. Results of the Optimal Channel Weights

In the exploration of the contributions of all the 17 selected channels, we discovered the optimal channel weights ($\mathbf{W}$) using the real-valued GA-based searching approach. The visualization results of the discovered $\mathbf{W}$ are shown in Fig. 7. In the heatmap, higher weights tend to exhibit hotter orange and red colors, whereas lower weights tend to exhibit a green color. The results show that the contributions of the channels at each alpha frequency bin were slightly different. For example, left occipital areas (O1 and Oz) contributed the most in the $\alpha 4$ responses to visual saliency, with $W_{O1}^{\alpha 4} = 0.8973$ and $W_{Oz}^{\alpha 4} = 0.7628$. PO4 in the parietal lobe obtained the highest weight ($W_{PO4}^{\alpha 5} = 0.8135$) in the response of $\alpha 5$ to visual saliency. For the alpha dynamics in $\alpha 3$, the site with the greatest contribution was PO3 ($W_{PO3}^{\alpha 4} = 0.8354$). The obtained optimal weights in every loop of CV-based GA multiobjective searching are reported in Appendix F of the Supplementary Materials. These results showed that involuntary attention processing of sensory information at different frequency bins would exhibit distinct changes and depend on the electrode locations, as found in other mental activities [51], [54]–[57].

C. Encoding Performance Using Feature Maps

We also evaluated the encoding performance by only using the feature maps of S_w^C, S_w^I, and S_w^O on the basis of the extracted features C_w^C, C_w^I, and C_w^O. The overall performances

TABLE III
THE 10-FOLD CV PERFORMANCE OF $\mathcal{A}_{a,w}$ BY USING C_w, C_w^C, C_w^I, AND C_w^O

	GA_N	GA_M	$\alpha1$ corr	$\alpha1$ NMSE	f_{GA}	$\alpha2$ corr	$\alpha2$ NMSE	f_{GA}	$\alpha3$ corr	$\alpha3$ NMSE	f_{GA}	$\alpha4$ corr	$\alpha4$ NMSE	f_{GA}	$\alpha5$ corr	$\alpha5$ NMSE	f_{GA}
C_w	10	10	0.42	-7.76	0.00	0.53	-3.99	0.03	0.63	-1.43	0.41	0.72	-0.44	1.27	0.73	-0.60	1.09
	100	100	0.42	-7.52	0.00	0.55	-3.48	0.05	0.65	-1.20	0.53	0.73	-0.35	1.40	0.73	-0.54	1.16
	1000	1000	0.43	-7.08	0.00	0.56	-3.24	0.06	0.65	-1.17	0.55	0.74	-0.30	1.48	0.73	-0.52	1.19
C_w^C	10	10	0.38	-8.71	0.00	0.52	-4.29	0.02	0.64	-1.32	0.46	0.71	-0.48	1.19	0.72	-0.62	1.06
	100	100	0.40	-7.57	0.00	0.54	-3.72	0.04	0.66	-1.12	0.59	0.72	-0.40	1.31	0.73	-0.54	1.16
	1000	1000	0.38	-8.10	0.00	0.56	-3.39	0.05	0.67	-1.04	0.64	0.72	-0.36	1.36	0.73	-0.53	1.17
C_w^I	10	10	0.48	-5.87	0.00	0.58	-2.97	0.08	0.67	-0.96	0.70	0.74	-0.29	1.50	0.75	-0.45	1.30
	100	100	0.48	-5.53	0.01	0.60	-2.63	0.12	0.68	-0.80	0.84	0.75	-0.20	1.67	0.76	-0.36	1.45
	1000	1000	0.48	-5.86	0.00	0.61	-2.47	0.14	0.69	-0.76	0.88	0.75	-0.16	1.75	0.76	-0.37	1.43
C_w^O	10	10	0.33	-11.52	0.00	0.44	-6.16	0.00	0.58	-2.07	0.20	0.67	-0.75	0.86	0.67	-1.15	0.58
	100	100	0.33	-10.85	0.00	0.45	-5.55	0.00	0.61	-1.63	0.33	0.68	-0.60	1.02	0.67	-1.15	0.57
	1000	1000	0.35	-10.16	0.00	0.48	-5.03	0.01	0.61	-1.62	0.33	0.69	-0.54	1.09	0.67	-1.03	0.65

TABLE IV
THE 10-FOLD CV PERFORMANCE OF $\mathcal{A}_{a,w}$ BY USING C_w AT EACH FREQUENCY BIN WHEN A SIMPLE GRAPH WAS ADOPTED

GA_N	GA_M	$\alpha1$ corr	$\alpha1$ NMSE	f_{GA}	$\alpha2$ corr	$\alpha2$ NMSE	f_{GA}	$\alpha3$ corr	$\alpha3$ NMSE	f_{GA}	$\alpha4$ corr	$\alpha4$ NMSE	f_{GA}	$\alpha5$ corr	$\alpha5$ NMSE	f_{GA}
10	10	0.12	-31.30	0.00	0.21	-21.37	0.00	0.27	-13.37	0.00	0.35	-7.64	0.00	0.38	-7.28	0.00

across all frequency bins are reported in Table II, and the detailed results at each frequency bin are shown in Table III. The results indicated that, except intensity map, poor encoding performance was achieved if only a single color or orientation feature map was applied. This observation is consistent with previous psychophysical studies in HVS [16], [17], [58], [59]. The same as proven in [22], the intensity feature showed the most stable encoding performance than the others. While, orientation-based encoding for alpha oscillations was the weakest among the three. For different frequency bins, the encoding at $\alpha4$ was consistently more sensitive to all the single feature maps, which was similar to the encoding results to saliency map. Furthermore, compared to the encoding results of $\alpha3$, $\alpha4$ and $\alpha5$, the performances of $\alpha1$ and $\alpha2$ were still the worst in the feature-map-based encoding.

D. Effects of Parameter Settings

We verified the effect of hyperedge size τ_h and the reduced feature dimensionality M' on encoding performance, which were used in hypergraph embedding for dimensionality reduction. This testing was conducted when both GA_N and GA_M were equal to 10. We quantified and compared the influence of hyperedge size by varying τ_h from 3 to 700 when M' was set to 110. The overall encoding results across all frequency bins in terms of the calculated correlation coefficient values are compared in Fig. 8 (a). An increase in the hyperedge size resulted in an increase in correlation coefficient until the highest encoding performance was achieved when τ_h was 30. Subsequently, the encoding performance notably decreased. On the other hand, when the hyperedge size was equal to 2, it was the same as a simple graph. The corresponding encoding performance is reported in Table IV, where the encoding performance was quite unsatisfactory. This result

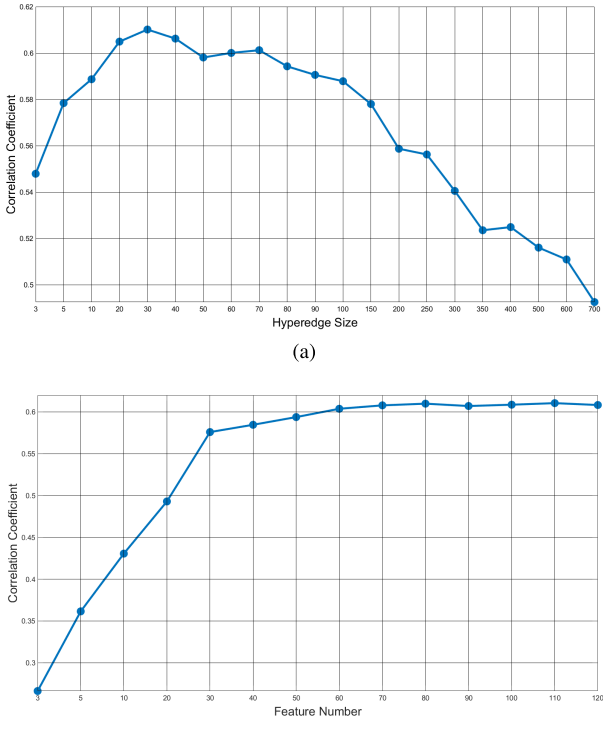

Fig. 8. The obtained correlation coefficient values when (a) the hyperedge size and (b) the feature number were set to different values. The best encoding performance was achieved when τ_h was 30 and M' was equal to or larger than 60.

also indicated the benefit of the hypergraph structure in feature dimensionality reduction. Furthermore, the feature size M' was varied from 3 to 120 when $\tau_h = 30$. The performance is presented in Fig. 8 (b). The curve showed that similar

good encoding performance was achieved when M' was equal to or larger than 60. The results demonstrated that the encoding framework with the proper parameter settings could benefit the estimation of brain dynamics from visual saliency.

V. Discussion and Conclusion

The human visual system has a remarkable ability to automatically select the salient content from the visual inputs, which enables people to allocate limited resources to process relevant and important information [60]–[62]. In the human visual system, involuntary attention is a fundamental component of visual attention mechanism, in which the attention is navigated according to the pure low-level visual features and will be used to further support and interact with the voluntary attention. Current attentional selection has typically been studied in visual tasks with simplified or artificial displays [23], [25], [26]. Although these studies are very important to reflect the basic principles of attention, they do not explain how the attention reacts to the complicated real-world. Few attention studies were conducted with natural visual processing, but they only focused on the static scenes [24], [62]. To further investigate how the brain efficiently selects salient information in dynamic and natural scenes such as natural-colored videos, in this paper we proposed a generalized encoding framework for real-world visual processing that connects the relationships between involuntary attention modulation (in terms of alpha oscillations) and visual information (in terms of saliency maps generated from the natural-colored videos).

The past physiological studies have demonstrated that the performance of the selective visual attention are highly correlated with the neural dynamics in perceptual systems [63]. We chose the generated saliency map by a bottom-up based computational model [17] as the topographic representation of attentional information in the visual inputs, because there is converging evidence that the generated saliency maps have been closely linked to the selection process in primary visual cortex [4], [64], [65]. We could use the saliency maps to estimate which are more likely to capture the attention of the stimuli, and further extend to evaluate whether different saliency levels could induce the changes in neural oscillations. Our results presented in this paper are in agreement with the existing findings that we found a highly significant correspondences between computational model of attention and neural oscillations, where the neural oscillation at a frequency of 11 Hz reflects the most to the bottom-up, stimulus-driven attributes. The results also suggested that the neural oscillation in the frequency range of 10 to 12 Hz played a significant role in the involuntary attention (early attentional mechanism). Besides, we conducted a t-test analysis to investigate whether there exists a statistical significance of the differences of the extracted saliency features between any two videos. The results showed the extracted features from different videos are statistically different with a confidence level of 95%. In other words, the present videos could be considered to have different involuntary attention effects. In the collected EEG data, the attentional influences on neural oscillations in the frequency range of 10 to 12 Hz during watching different videos were also observed, that suggests the upper alpha band could be considered as the fundamental component of the neural mechanism of involuntary attention.

In addition to looking at the relationship between involuntary attention and the frequency of neural oscillations, we also examined the attentional effects to the brain locations. The existing studies found the attention-related neural activation was mainly located in primary visual cortex, and the involuntary attentional effects are most correlated to the parieto-occipital areas [27], [28], [66]. In response to visual attention associated tasks, a signification neural activation in the parieto-occipital region was observed [67]–[69]. To further investigate finer-scale brain regions to involuntary attention, we evaluated the contribution of every single EEG channel in the parieto-occipital areas by building a genetic algorithm based regression model. The obtained weight of each channel in the regression process could be considered as an indicator to show how important of the channel location to the selective processing and measure the brain source of the interaction to the involuntary attention. Our results showed that the most related channels to involuntary attention were distributed on the O1 and PO3 sites, with a great channel weight obtained (0.8973 and 0.8354). These locations may be responsible for the sensory information processing in the involuntary attention.

Investigating the possibility of encoding brain dynamics by visual content is a critical aspect in the cognitive neuroscience and biomedical engineering. This current study demonstrated the effects of low-level visual attributes on the neural mechanisms involved in the involuntary attention via the changes in EEG oscillations and proposed a generalized encoding pipeline for alpha oscillations from visual saliency features. Our work outlined possible processing steps to solve an EEG encoding problem, supporting the future development of brain analysis. The results indicated that visual saliency information can be indeed reflected by different alpha frequency bins and different locations. We found that $\alpha 4$ responded the most to involuntary attention related to bottom-up-based visual information processing and that the extracted saliency features were reliable predictors of alpha dynamics. In light of neuroadaptive technologies, the obtained results in the present work could be applied to the field of brain-computer interfaces where suitable visual stimuli will be interactively and automatically selected to elicit certain mental statuses and further improve the interaction experience in terms of usability, accessibility, and pleasure. The current work successfully investigated involuntary attention in a set of natural-colored videos. Instead of considering the low-level visual features (such as color, texture, orientation) in the dynamic scenes, we introduced saliency map as a reflection of attention in the present variable and cluttered videos and evaluated the performance of alpha oscillations at different frequency bins to the involuntary attentional selection under naturalistic conditions. An intelligent-based generalized encoding pipeline was built to help a mapping from involuntary attention reflection in the visual content to the involuntary attention modulation with alpha oscillations.

The present study suggests the feasibility of estimating brain dynamics by monitoring the saliency changes in the present

visual stimuli and offers an opportunity to regulate neural oscillations through multimedia information. Limitations of this study include the relatively simple experimental design and the limited number of participants and video clips that were examined at the current stage. In the future extend experiment, the number of participants will be increased, and visual stimuli with various saliency distributions will be examined. Based on the study proposed in this paper, the design of visual content could be improved by interacting with the estimated brain responses while watching them, and more intelligent-inspired multimedia could be produced.

VI. Conflicts of Interest

The authors declare that they have no conflicts of interest to disclose.

Acknowledgment

The authors sincerely appreciate the anonymous reviewers for their detailed comments.

References

[1] E. Başar, M. Schürmann, C. Başar-Eroglu, and S. Karakaş, "Alpha oscillations in brain functioning: An integrative theory," *Int. J. Psychophysiol.*, vol. 26, nos. 1–3, pp. 5–29, Jun. 1997.

[2] H. Laufs *et al.*, "Electroencephalographic signatures of attentional and cognitive default modes in spontaneous brain activity fluctuations at rest," *Proc. Nat. Acad. Sci. USA*, vol. 100, no. 19, pp. 11053–11058, Sep. 2003.

[3] O. M. Bazanova and D. Vernon, "Interpreting EEG alpha activity," *Neurosci. Biobehav. Rev.*, vol. 44, pp. 94–110, Jul. 2014.

[4] W. Klimesch, "Alpha-band oscillations, attention, and controlled access to stored information," *Trends Cognit. Sci.*, vol. 16, no. 12, pp. 606–617, Dec. 2012.

[5] W. Klimesch, "EEG alpha and theta oscillations reflect cognitive and memory performance: A review and analysis," *Brain Res. Rev.*, vol. 29, nos. 2–3, pp. 169–195, Apr. 1999.

[6] W. Ray and H. Cole, "EEG alpha activity reflects attentional demands, and beta activity reflects emotional and cognitive processes," *Science*, vol. 228, no. 4700, pp. 750–752, May 1985.

[7] P. Sauseng *et al.*, "A shift of visual spatial attention is selectively associated with human EEG alpha activity," *Eur. J. Neurosci.*, vol. 22, no. 11, pp. 2917–2926, Dec. 2005.

[8] K. Abuhassan, D. Coyle, A. Belatreche, and L. Maguire, "Compensating for synaptic loss in Alzheimer's disease," *J. Comput. Neurosci.*, vol. 36, no. 1, pp. 19–37, 2014.

[9] A. Fink and M. Benedek, "EEG alpha power and creative ideation," *Neurosci. Biobehav. Rev.*, vol. 44, pp. 111–123, Jul. 2014.

[10] Z.-X. Liu, D. Glizer, R. Tannock, and S. Woltering, "EEG alpha power during maintenance of information in working memory in adults with ADHD and its plasticity due to working memory training: A randomized controlled trial," *Clin. Neurophysiol.*, vol. 127, no. 2, pp. 1307–1320, Feb. 2016.

[11] A. Brzezicka, J. Kamiński, O. K. Kamińska, D. Wołyńczyk-Gmaj, and G. Sedek, "Frontal EEG alpha band asymmetry as a predictor of reasoning deficiency in depressed people," *Cognition Emotion*, vol. 31, no. 5, pp. 868–878, Jul. 2017.

[12] J. J. Foxe and A. C. Snyder, "The role of alpha-band brain oscillations as a sensory suppression mechanism during selective attention," *Frontiers Psychol.*, vol. 2, p. 154, Jul. 2011.

[13] B. Zoefel and R. VanRullen, "Oscillatory mechanisms of stimulus processing and selection in the visual and auditory systems: State-of-the-art, speculations and suggestions," *Frontiers Neurosci.*, vol. 11, p. 296, May 2017.

[14] B. F. Händel, T. Haarmeier, and O. Jensen, "Alpha oscillations correlate with the successful inhibition of unattended stimuli," *J. Cognit. Neurosci.*, vol. 23, no. 9, pp. 2494–2502, Sep. 2011.

[15] Y. Pinto, A. R. van der Leij, I. G. Sligte, V. A. Lamme, and H. S. Scholte, "Bottom-up and top-down attention are independent," *J. Vis.*, vol. 13, no. 3, p. 16, 2013.

[16] C. Koch and S. Ullman, "Shifts in selective visual attention: Towards the underlying neural circuitry," *Human Neurobiol.*, vol. 4, no. 4, pp. 219–227, 1985.

[17] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 11, pp. 1254–1259, Nov. 1998.

[18] T.-P. Jung, S. Makeig, M. Stensmo, and T. J. Sejnowski, "Estimating alertness from the EEG power spectrum," *IEEE Trans. Biomed. Eng.*, vol. 44, no. 1, pp. 60–69, Jan. 1997.

[19] S. Hanslmayr *et al.*, "Visual discrimination performance is related to decreased alpha amplitude but increased phase locking," *Neurosci. Lett.*, vol. 375, no. 1, pp. 64–68, Feb. 2005.

[20] W. Klimesch, R. Fellinger, and R. Freunberger, "Alpha oscillations and early stages of visual encoding," *Frontiers Psychol.*, vol. 2, p. 118, May 2011.

[21] S. D. Mayhew, D. Ostwald, C. Porcaro, and A. P. Bagshaw, "Spontaneous EEG alpha oscillation interacts with positive and negative BOLD responses in the visual-auditory cortices and default-mode network," *NeuroImage*, vol. 76, pp. 362–372, Aug. 2013.

[22] T. Ergenoglu, T. Demiralp, Z. Bayraktaroglu, M. Ergen, H. Beydagi, and Y. Uresin, "Alpha rhythm of the EEG modulates visual detection performance in humans," *Cognit. Brain Res.*, vol. 20, no. 3, pp. 376–383, Aug. 2004.

[23] T. J. McDermott, A. I. Wiesman, A. L. Proskovec, E. Heinrichs-Graham, and T. W. Wilson, "Spatiotemporal oscillatory dynamics of visual selective attention during a flanker task," *NeuroImage*, vol. 156, pp. 277–285, Aug. 2017.

[24] S. Noah, T. Powell, N. Khodayari, D. Olivan, M. Ding, and G. R. Mangun, "Neural mechanisms of attentional control for objects: Decoding EEG alpha when anticipating faces, scenes, and tools," *J. Neurosci.*, vol. 40, no. 25, pp. 4913–4924, 2020.

[25] O. Jensen and A. Mazaheri, "Shaping functional architecture by oscillatory alpha activity: Gating by inhibition," *Frontiers Hum. Neurosci.*, vol. 4, p. 186, Nov. 2010.

[26] A. C. Snyder and J. J. Foxe, "Anticipatory attentional suppression of visual features indexed by oscillatory alpha-band power increases: A high-density electrical mapping study," *J. Neurosci.*, vol. 30, no. 11, pp. 4024–4032, 2010.

[27] Z. Li, "A saliency map in primary visual cortex," *Trends Cognit. Sci.*, vol. 6, no. 1, pp. 9–16, Jan. 2002.

[28] C. L. Colby and M. E. Goldberg, "Space and attention in parietal cortex," *Annu. Rev. Neurosci.*, vol. 22, no. 1, pp. 319–349, Mar. 1999.

[29] M. Critchley, *The Parietal Lobes*. Baltimore, MD, USA: Williams & Wilkins, 1953.

[30] M.-M. Mesulam, "A cortical network for directed attention and unilateral neglect," *Ann. Neurol.*, vol. 10, no. 4, pp. 309–325, Oct. 1981.

[31] G. Rizzolatti, L. Fogassi, and V. Gallese, "Parietal cortex: From sight to action," *Current Opinion Neurobiol.*, vol. 7, no. 4, pp. 562–567, Aug. 1997.

[32] G. Coricelli, H. D. Critchley, M. Joffily, J. P. O'Doherty, A. Sirigu, and R. J. Dolan, "Regret and its avoidance: A neuroimaging study of choice behavior," *Nature Neurosci.*, vol. 8, no. 9, p. 1255–1262, 2005.

[33] C. Summerfield and T. Egner, "Expectation (and attention) in visual cognition," *Trends Cognit. Sci.*, vol. 13, no. 9, pp. 403–409, Sep. 2009.

[34] R. Oostenveld and P. Praamstra, "The five percent electrode system for high-resolution EEG and ERP measurements," *Clin. Neurophysiol.*, vol. 112, no. 4, pp. 713–719, Apr. 2001.

[35] T.-P. Jung *et al.*, "Removing electroencephalographic artifacts by blind source separation," *Psychophysiology*, vol. 37, no. 2, pp. 163–178, Mar. 2000.

[36] A. Delorme and S. Makeig, "EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *J. Neurosci. Methods*, vol. 134, no. 1, pp. 9–21, Mar. 2004.

[37] P. Welch, "The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms," *IEEE Trans. Audio Electroacoustics*, vol. 15, no. 2, pp. 70–73, Jun. 1967.

[38] J. Griffin, T. Schultz, R. Holman, L. S. Ukeiley, and L. N. Cattafesta, "Application of multivariate outlier detection to fluid velocity measurements," *Exp. Fluids*, vol. 49, no. 1, pp. 305–317, Jul. 2010.

[39] S. J. Klawikowski, C. Zeringue, L. S. Wootton, G. S. Ibbott, and S. Beddar, "Preliminary evaluation of the dosimetric accuracy of the in vivo plastic scintillation detector OARtrac system for prostate cancer treatments," *Phys. Med. Biol.*, vol. 59, no. 9, p. N27, 2014.

[40] S. M. Mahmoud, H. A. Alabbasi, and T. E. Abdulabbas, "Monitoring and detecting outliers for elder's life activities in a smart home: A case study," in *Proc. E-Health Bioeng. Conf. (EHB)*, Jun. 2017, pp. 458–461.
[41] A. Megahed, S. M. Fadl, Q. Han, and Q. Li, "Handwriting forgery detection based on ink colour features," in *Proc. 8th IEEE Int. Conf. Softw. Eng. Service Sci. (ICSESS)*, Nov. 2017, pp. 141–144.
[42] D. E. Goldberg and J. H. Holland, "Genetic algorithms and machine learning," *Mach. Learn.*, vol. 3, nos. 2–3, pp. 95–99, 1988.
[43] Y.-P. Huang and C.-H. Huang, "Real-valued genetic algorithms for fuzzy grey prediction system," *Fuzzy Sets Syst.*, vol. 87, no. 3, pp. 265–276, May 1997.
[44] Z. Liang, H. Fu, Z. Chi, and D. Feng, "Refining a region based attention model using eye tracking data," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2010, pp. 1105–1108.
[45] R. Irani and R. Nasimi, "Evolving neural network using real coded genetic algorithm for permeability estimation of the reservoir," *Expert Syst. Appl.*, vol. 38, no. 8, pp. 9862–9866, Aug. 2011.
[46] G. Chandrashekar and F. Sahin, "A survey on feature selection methods," *Comput. Electr. Eng.*, vol. 40, no. 1, pp. 16–28, Jan. 2014.
[47] C. R. Houck, J. Joines, and M. G. Kay, "A genetic algorithm for function optimization: A MATLAB implementation," *Ncsu-ie tr*, vol. 95, no. 09, pp. 1–10, 1995.
[48] A. Ion-Margineanu *et al.*, "Comparison of manual and semi-manual delineations for classifying glioblastoma multiforme patients based on histogram and texture MRI features," in *Proc. 25th Eur. Symp. Artif. Neural Netw., Comput. Intell. Mach. Learn.*, 2017, pp. 501–506.
[49] F. Lee, J. Zhao, K. Kotani, and Q. Chen, "Video copy detection using histogram based spatio-temporal features," in *Proc. 10th Int. Congr. Image Signal Process., Biomed. Eng. Informat. (CISP-BMEI)*, Oct. 2017, pp. 1–5.
[50] S. G. B. Rylander and B. Gotshall, "Optimal population size and the genetic algorithm," *Population*, vol. 100, no. 400, p. 900, 2002.
[51] W. Klimesch, M. Doppelmayr, T. Pachinger, and H. Russegger, "Event-related desynchronization in the alpha band and the processing of semantic information," *Cognit. Brain Res.*, vol. 6, no. 2, pp. 83–94, Oct. 1997.
[52] W. Klimesch, M. Doppelmayr, H. Russegger, T. Pachinger, and J. Schwaiger, "Induced alpha band power changes in the human EEG and attention," *Neurosci. Lett.*, vol. 244, no. 2, pp. 73–76, Mar. 1998.
[53] I. Babu Henry Samuel, C. Wang, Z. Hu, and M. Ding, "The frequency of alpha oscillations: Task-dependent modulation and its functional significance," *NeuroImage*, vol. 183, pp. 897–906, Dec. 2018.
[54] A. J. Shackman, B. W. Mcmenamin, J. S. Maxwell, L. L. Greischar, and R. J. Davidson, "Identifying robust and sensitive frequency bands for interrogating neural oscillations," *NeuroImage*, vol. 51, no. 4, pp. 1319–1333, Jul. 2010.
[55] P. Capotosto, C. Babiloni, G. L. Romani, and M. Corbetta, "Resting-state modulation of alpha rhythms by interference with angular gyrus activity," *J. Cognit. Neurosci.*, vol. 26, no. 1, pp. 107–119, Jan. 2014.
[56] T. Chow, T. Javan, T. Ros, and P. Frewen, "EEG dynamics of mindfulness meditation versus alpha neurofeedback: A sham-controlled study," *Mindfulness*, vol. 8, no. 3, pp. 572–584, Jun. 2017.
[57] Y. Kitaura *et al.*, "Functional localization and effective connectivity of cortical theta and alpha oscillatory activity during an attention task," *Clin. Neurophysiol. Pract.*, vol. 2, pp. 193–200, 2017.
[58] R. Desimone and J. Duncan, "Neural mechanisms of selective visual attention," *Annu. Rev. Neurosci.*, vol. 18, no. 1, pp. 193–222, Mar. 1995.
[59] S. Kollmorgen, N. Nortmann, S. Schröder, and P. König, "Influence of low-level stimulus features, task dependent factors, and spatial biases on overt visual attention," *PLoS Comput. Biol.*, vol. 6, no. 5, May 2010, Art. no. e1000791.
[60] M. V. Peelen and S. Kastner, "Attention in the real world: Toward understanding its neural basis," *Trends Cognit. Sci.*, vol. 18, no. 5, pp. 242–250, May 2014.
[61] E. H. Cohen and F. Tong, "Neural mechanisms of object-based attention," *Cerebral Cortex*, vol. 25, no. 4, pp. 1080–1092, Apr. 2015.
[62] D. Kaiser, N. N. Oosterhof, and M. V. Peelen, "The neural dynamics of attentional selection in natural scenes," *J. Neurosci.*, vol. 36, no. 41, pp. 10522–10528, Oct. 2016.
[63] T. Moore and M. Zirnsak, "Neural mechanisms of selective visual attention," *Annu. Rev. Psychol.*, vol. 68, no. 1, pp. 47–72, Jan. 2017.
[64] S. Treue, "Visual attention: The where, what, how and why of saliency," *Current Opinion Neurobiol.*, vol. 13, no. 4, pp. 428–432, Aug. 2003.
[65] O. Jensen, M. Bonnefond, and R. VanRullen, "An oscillatory mechanism for prioritizing salient unattended stimuli," *Trends Cognit. Sci.*, vol. 16, no. 4, pp. 200–206, Apr. 2012.
[66] M. Kawasaki and Y. Yamaguchi, "Effects of subjective preference of colors on attention-related occipital theta oscillations," *NeuroImage*, vol. 59, no. 1, pp. 808–814, Jan. 2012.
[67] K.-M. G. Fu, J. J. Foxe, M. M. Murray, B. A. Higgins, D. C. Javitt, and C. E. Schroeder, "Attention-dependent suppression of distracter visual input can be cross-modally cued as indexed by anticipatory parieto–occipital alpha-band oscillations," *Cogn. Brain Res.*, vol. 12, no. 1, pp. 145–152, 2001.
[68] H. Kojima and T. Suzuki, "Hemodynamic change in occipital lobe during visual search: Visual attention allocation measured with NIRS," *Neuropsychologia*, vol. 48, no. 1, pp. 349–352, Jan. 2010.
[69] C. Liang, C.-T. Lin, S.-N. Yao, W.-S. Chang, Y.-C. Liu, and S.-A. Chen, "Visual attention and association: An electroencephalography study in expert designers," *Des. Stud.*, vol. 48, pp. 76–95, Jan. 2017.