Decoding Emotional Valence from EEG in Immersive Virtual Reality

Guanxiong Pei*§, Bingjie Li†§, Taihao Li*ʃ, Ruohao Xu*, Jianmin Dong*, Jia Jin‡ʃ
*Zhejiang Lab, Hangzhou, China
E-mail: lith@zhejianglab.com Tel: +86-0571-58005295
†National University of Singapore, Singapore, Singapore
email:bjlistat@nus.edu.sg Tel: +65-93978342
‡Shanghai International Studies University, Shanghai, China
E-mail: jinjia.163@163.com Tel: +86-021-35372000
§Guanxiong Pei and Bingjie Li contribute equally to the article.
ʃCorrespondence: lith@zhejianglab.com, jinjia.163@163.com

Abstract— **Due to the strong sense of reality, immersion, and interaction, virtual reality technology has been widely used in emotional induction, psychological assistance, and the diagnosis of emotion-related disorders. It is a challenging problem that involves evaluating the interventional effect of virtual reality in the above applications and objectively judging an individual's emotional state. The primary purpose of this paper is to introduce innovative methods for obtaining reliable distinguishing features and improve the classification accuracy of emotional valence from EEG signals in immersive virtual reality. Firstly, we established a relatively standard emotion-induced virtual reality video library. Participants' EEG data were collected synchronously while they were watching virtual reality clips. Then, EEG features of the energy spectrum, differential entropy, differential asymmetry, and rational asymmetry were extracted to represent the characteristics associated with emotional valence. The results show that the random forest (RF) generally performed better than the backpropagation neural network (BPNN). By combining the dimensionality reduction method (the F-test or PCA) and the RF classifier, it is possible to achieve encouraging classification results and increase computation speed and stability. PCA-RF achieved the highest average classification accuracy of 95.6%. In addition, it is demonstrated that features extracted from the theta band were superior to features from other frequency bands for emotional valence decoding. This may facilitate the application of EEG-based affective computing technology in the virtual reality brain-computer interface field.**

I. Introduction

Virtual reality (VR) is a computer-generated environment in which scenarios appear realistic and allow users to feel immersed [1]. According to a recent literature review [2], immersive VR is becoming more and more popular in affective computing. On the one hand, VR allows individuals to experience lively, naturalistic, and interactive situations, allowing researchers to study emotions in more realistic settings within controlled laboratory conditions. It is beneficial to the ecological validity of research conclusions. On the other hand, VR can be widely used in emotional induction, mental health care, and the diagnosis of emotion-related disorders [3] to make scientific research and real-life applications closely integrated to benefit human life.

According to literature review, one evolutionary trend of current research is the combination of immersive VR and implicit measurements to monitor physical and psychological states. One example is electroencephalogram (EEG) technology [2]. As one of the most mobile neurophysiological techniques, EEG has emerged as a powerful tool for quantitatively studying brain activities combined with the VR head-mounted display (HMD) [4]. Compared with other behavioral signals for affective computing, such as facial expressions and sound recordings, EEG signals directly reflect changes in the central nervous system, which provide more reliable information for emotions in contrast with visual and audio cues [5]. In addition, the use of EEG is noninvasive and inexpensive, which has attracted many researchers to unravel the evocation of emotional responses in the brain [6]. To summarize, the combination of immersive VR and EEG enables naturalistic neuroscientific research while maintaining experimental control.

As a rapidly growing field of research, EEG-based affective computing in immersive VR is gradually transitioning from traditional statistical analysis to supervised machine learning-based analysis. Previous studies focused more on the neural mechanisms of emotions underlying the VR environment or the frequency oscillations produced during various affective states by using hypothesis testing and correlation analysis [7-9]. For example, ref. [10] demonstrated that the brain's lower beta-band desynchronization could be used to predict emotional arousal when participants experienced a virtual roller coaster game. Another study showed that the VR sequence with entertaining content was closely related to beta high (β_H) bands, while the VR sequence with horrific content was closely associated with theta bands [11].

To our knowledge, ref. [12] was the first study that used EEG in an immersive scenario combined with machine learning algorithms to recognize various emotional states. An extensive set of EEG (frequency band power and phase coherency) and heart rate variability (HRV) features were put

978-616-590-477-3

into a nonlinear SVM classifier. The model realized an accuracy of 71.21% at the two levels of emotional valence. Another study used SVM in combination with EEG and HRV, and achieved recognition accuracies of 75% (arousal) and 71.08% (valence) in a realistic 3D virtual museum [13]. Later, in 2022, a live automatic emotion recognition system was established to decode the real-time emotional states using the low-cost wearable EEG headset with only four electrodes to detect the four distinct emotion classes, which obtained an 85.01% classification accuracy [14]. Limitations of the studies mentioned above can be summarized in two perspectives: (i) Existing research does not extract features from the multi-dimensional perspective of EEG itself but carries out emotion recognition by combining multi-modal features, such as HRV, which increases experimental complexity and cost; and (ii) The accuracy of emotion classification and recognition is generally low and cannot support a wide range of applications.

To overcome these limitations, we try to propose a reliable method to recognize the emotional valence in immersive VR. Inspired by previous affective computing studies, we have sorted out the EEG features commonly used in VR research and the features generally considered to perform well in emotional valence recognition in 2D scenarios [12-15]. Finally, we extracted energy spectrum (ES), differential entropy (DE), differential asymmetry (DASM), and rational asymmetry (RASM) as features. The F-test and the principal components analysis (PCA) algorithm were used as dimensionality reduction methods to reduce the computational cost of modeling. The classification performance of the random forest (RF) and the backpropagation neural network (BPNN) were compared.

II. MATERIALS AND METHODS

A. Participants

A total of 28 college students participated in this experiment, with an average age of 21.44 years and a standard deviation of 2.87. The data of one participant was deleted because of extensive artifacts. Finally, the data of 27 participants were used for further analysis. All participants were right-handed and natively Chinese, with normal or corrected-to-normal vision, and free of achromatopsia and neurological disorder. Participants were requested not to drink coffee or tea 3 hours before coming to the lab and less than 5 hours of experience with VR. All of them signed informed consent before participation. The Ethics Committee approved the study. Each participant received ¥ 45 (about $ 7.0) as compensation for their time.

In compliance with ethical guidelines, participants completed the 21-item Chinese version of the Depression Anxiety Stress Scales (DASS-21) before the start of the study [16]. As a short version of a 42-item self-report scale designed to measure three related negative emotional states: depression, anxiety, and tension/stress, this questionnaire has high validity and reliability. Participants read each statement and circled a number 0 (not applicable to me at all), 1 (applicable to me to some degree, or some of the time), 2 (applicable to me to a considerable degree, or a good part of the time) or 3 (applicable to me very much, or most of the time) that indicates to what extent the statement applies to them over the previous week. The DASS-21 results were obtained immediately so that unsuitable participants could be excluded (subjects with a depression score above 9, an anxiety score above 7, or a stress score above 14). None of the participants were excluded from the study (depression scores: Mean = 3.63, SD = 3.16; anxiety scores: Mean = 3.26, SD =2.54; stress scores: Mean = 5.37, SD = 4.43).

B. Materials

We spent over one month searching for clips of immersive VR with the potential function of emotion induction. In total, more than 100 VR clips were viewed and assessed. The subsequent selection round was conducted based on the criteria employed by ref. [17]. Firstly, the clips had to be of relatively short length. This is especially important as longer clips might induce fatigue and dizziness among participants. Secondly, the clips had to be comprehensible on their own without the need for further description. Thirdly, the clips were likely to induce emotions. Three psychologists judged this from Shanghai International Studies University. Finally, 15 high-resolution clips were selected for the study, each with a duration of 30 seconds. Clips 1-5 were used for negative emotion induction, mainly with horrible and dark scenarios; Clips 6-10 were used for positive emotion induction, mainly with contents of cute pets or dance performances; and Clips 11-15 clips were used for neutral emotion induction mainly with natural or urban street scenarios. Fig.1 shows the procedure of the experiment and the screenshot of each clip.

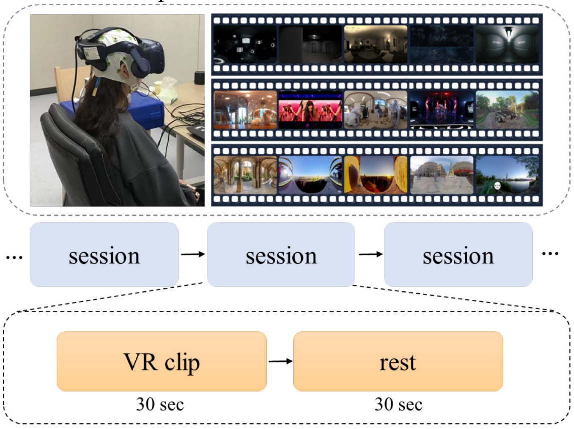

Fig. 1 Procedure of the experiment and the screenshot of each clip

C. Experimental Equipment

An HTC Vive HMD was placed on the EEG cap using custom-made cushions to avoid pressure artifacts and enhance wearing comfort. The HMD can provide stereoscopy with two 1440×1600-pixel OLED displays and a refresh rate of 90 Hz. VR clips were loaded into Viveport Video software (Ver. 3.0.5), which ran on a 3.6 GHz Intel i9 computer with the Nvidia GTX 1080 graphics card. The equipment used for EEG recording was a 32-channel (international standard 10-20 system distribution) BP EEG measurement system (Brain Products GmbH, Gilching, Germany), including signal amplifiers, photoelectric converters, frequency division switches, and the EEG caps (ranged from 54 cm to 58 cm, each was selected according to the participant's head size). The VR-EEG time synchronization interface was developed by Shanghai Qingyan Technology Co., Ltd., with a maximum synchronization frequency of 100 Hz.

D. *Experimental Design*

When participants arrived, they were led to wash and blow dry their hair to ensure that the scalp was free of oil. Then, they sat in a magnetic-insulated and sound-insulated room. They were told that they would wear an HMD and an EEG Electrode cap at the same time. They were allowed to request to quit anytime should they feel uncomfortable or dizzy. All of the participants were informed of the harmlessness of the equipment. After the electrode cap was put on, the conductive gel was injected into the electrodes. We conducted experiments after confirming that the impedance of each electrode point dropped to a reasonable range (below 5 kΩ). Afterward, the HMD was put on, and customized cushions were placed below the straps to reduce the interference with the EEG electrodes. Participants were also required to reduce body and eye movement to decrease electromyography (EMG) and electrooculography (EOG) noise. After the above steps were completed, participants entered the formal experiment.

In the formal experiment, EEG data were continuously acquired when participants watched the 15 VR clips, with a sampling rate of 500 Hz and a frequency band of direct current (DC) to 100 Hz. Between every two clips, a 30-second blank screen was placed as an interval to reduce interference between evoked emotions while participants were allowed to have a rest to minimize the chances of fatigue or dizziness. Before the VR exposure, 30 seconds of resting EEG activity were recorded. The total duration of the formal experiment was about 15 minutes. Because participants wore the EEG cap and the HMD simultaneously, their heads were under a certain amount of pressure. Participants could moderately relax and adjust their posture in the blank screen stage if they felt uncomfortable. The behavioral data were collected by the Self-assessment manikin (SAM) scale, which was a nonverbal image-oriented assessment scale, as shown in Fig.2 [18]. Participants were asked to report their emotional arousal and valence to various clips.

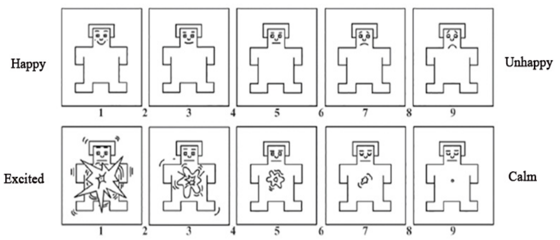

Fig. 2 The SAM scale used in this experiment

E. *EEG Data Preprocessing and Feature Extraction*

The EEG data were further processed offline with the Matlab EEGLab toolbox and several plugins. The EEG data were band-pass filtered at 0.5-50 Hz with a zero-phase shift FIR filter. The Notch filter removed electrical interference from the 50 Hz-line noise. High-density EEG activities referenced the average of both mastoids (TP9 & TP10). The EEG data were downsampled to speed up computation with a sampling frequency of 256 Hz. Combining EEG with VR provokes additional challenges: the signal-to-noise ratio (SNR) may be decreased due to mechanical interference of the VR headset with the EEG cap. To ensure high data quality, we applied multiple measures to prevent, recognize, discard, or correct artifacts in the EEG signal. Firstly, the EEG data were manually checked for contamination by muscles or other artifacts, and bad EEG epochs were rejected. Secondly, the location of removed bad electrodes was interpolated using spherical interpolation. Thirdly, the independent component analysis approach removed nonbrain-related artifacts such as significant muscle activity and eye movements.

Through the above steps, cleaned EEG data were produced. Each participant's data were segmented into 2-s epochs. For a better understanding of human brain activity, the EEG signal waves were divided into five significant sub-bands, which were separated from low to high frequencies known as delta (1-3 Hz), theta (4-7 Hz), alpha (8-13 Hz), beta (14-30 Hz), and gamma (31-49 Hz) bands. Frequency domain features were employed in this study. A 512-point short-time Fourier transform (STFT) with a non-overlapped Hanning window of 1s was used to convert EEG raw data from the time domain to the frequency domain, dividing the data into different frequency bands as required.

Four features (ES, DE, DASM, and RASM) were extracted. ES is the average energy of EEG signals in the five frequency bands mentioned above. DE is equivalent to the logarithm ES and calculated according to ref. [15], The time series X obeys the Gauss distribution N(μ, δ^2).

$$\mathrm{h(X)} = -\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \log\left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\right) dx = \frac{1}{2}\log(2\pi e\sigma^2) \quad (1)$$

DASM and RASM are the differences and ratios between DE of 13 pairs of hemispheric asymmetry electrodes (left-right: FP1-FP2; F3-F4; C3-C4; P3-P4; O1-O2; F7-F8; T7-T8; P7-P8; FC1-FC2; CP1-CP2; FC5-FC6; CP5-CP6; TP9-TP10), which can be calculated as

$$\mathrm{DASM} = \mathrm{h}\left(X_i^{left}\right) - \mathrm{h}\left(X_i^{right}\right) \quad (2)$$

$$\mathrm{RASM} = \mathrm{h}\left(X_i^{left}\right)/\mathrm{h}\left(X_i^{right}\right) \quad (3)$$

where h(X) is defined in Equation 1, and *i* is the pair number.

F. *Dimensionality Reduction and Emotion Classification*

We extracted the features (ES, DE, DASM, and RASM) of 27 participants within 5 frequency bands in different electrodes. We confronted with data that were of a high dimensionality. Dimensionality reduction methods benefit emotional valence recognition to reduce the computational cost of modeling. This study employed an F-test (p-value < 0.05) and PCA (cumulative proportion of variance more than 80%) to get the optimal feature sets. F-Test is helpful in feature selection as we know each feature's significance in improving the model. PCA is useful for selecting a subset of variables that preserves as much information as possible in the complete features. After feature selection, the data were normalized to avoid dependence on the choice of measurement units. In this study, we employed the min-max normalization method. It is commonly used when features are on drastically different scales. For every feature, the minimum value of that feature was converted into 0, the maximum value was transformed into 1, and every other value was converted into a decimal between 0 and 1.

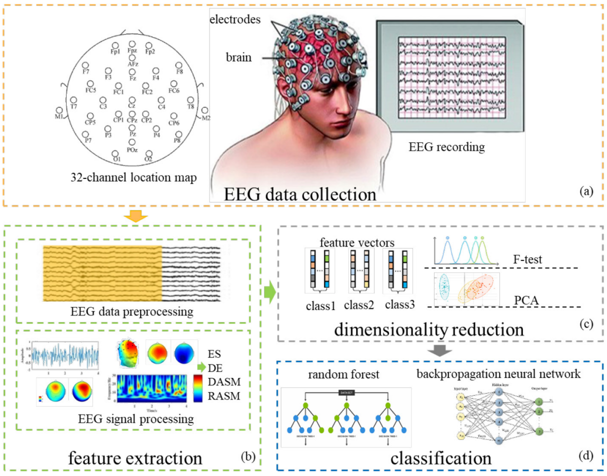

Fig. 3 Flow chart of the EEG data processing

The processed features were further applied to emotion classification using RF and BPNN. RF is an ensemble learning method that combines the outputs of multiple decision trees to get a single result, usually trained with bagging or bootstrap aggregating. RF is user-friendly with fewer parameters, and its performance is not sensitive to parameter value. Ease of use and flexibility have boosted the adoption of RF. More importantly, it is immune to irrelevant outliers and robust to overfitting. In our model, the number of trees in the forest is 100, the maximum depth of the trees is 81, and the minimum number of samples required to split an internal node is 3. Different from RF, BPNN builds upon the human nervous system. The topology structure of BPNN includes one input layer, multi-hidden layers and one output layer. Compared with RF, more factors are needed in constructing the BPNN, including the initialization of the network, the transfer functions, and so on, which require experience and prior knowledge to determine many parameters. In fitting a neural network, backpropagation calculates the gradient of a loss function concerning all the weights in the network. It should be noted that the 10-fold cross-validation method was used to evaluate the model's generalization performance. We first divided our dataset into ten equally sized subsets. Then, we repeated the train-test method ten times such that each time one of the ten subsets was used as a test set and the rest nine subsets were used together as a training set. We have selected several performance indicators of the classification model, including accuracy, precision, recall, and macro-F1 (macro-averaged F1-score) based on the confusion matrix. The calculation method for each index is as follows.

$$Accuracy=(TP+TN)/(TP+FN+FP+TN) \quad (4)$$

$$Precision=TP/(TP+FP) \quad (5)$$

$$Recall=TP/(TP+FN) \quad (6)$$

$$F1\text{-}score=2\times Recall\times Precision/(Recall+Precision) \quad (7)$$

where TP = Number of true positives among the total predictions; FN = Number of false negatives among the total predictions; FP = Number of false positives among the total predictions; TN = Number of true negatives among the total predictions.

III. RESULTS

A. Self-reported Results

The self-reported results show that 48.15% of the participants had experienced VR equipment before, and 51.85% wore VR equipment for the first time. In all, 18.23% of the participants had a slight sense of dizziness during the experiment. Statistical comparisons (ANOVA with Fisher's LSD post hoc test) of SAM (9-point Likert scale) were performed to detect significant differences in self-reported emotional valence and arousal with positive, negative, and neutral VR clips priming. The ANOVA analysis of the self-reported emotional valence was significant [$F(2,78)= 40.50, p<0.001$]. The LSD Post Hoc test indicated that positive VR clips induced significantly higher valence ratings than negative and neutral VR clips ($p<0.001$). Neutral VR clips generated significantly higher valence ratings than negative VR clips ($p<0.001$). The ANOVA analysis of the self-reported emotional arousal was significant [$F(2,78)= 10.325, p<0.001$]. The LSD Post Hoc test indicated that neutral VR clips induced significantly lower arousal ratings than negative and positive VR clips ($p<0.001$). There were no statistically significant differences in arousal between negative VR clips and positive VR clips ($p>0.05$), as shown in Table I.

TABLE I

THE RESULTS (MEAN ± S.D.) OF EMOTIONAL VALENCE AND AROUSAL FROM SAM

dimension	positive	negative	neutral
valence	6.35±0.89	4.12±1.19	5.43±0.54
arousal	5.07±1.69	5.18±1.57	3.38±1.66

B. EEG Results

The performance of different method combinations trained with multivariate features (ES, DE, DASM, and RASM) on five frequency bands are shown in Fig.4, Table II and Table III. The RF classifier performed better than the BPNN classifier reflected by the four indicators. The RF classifier offers a more accurate and stable recognition performance regardless of the frequency band.

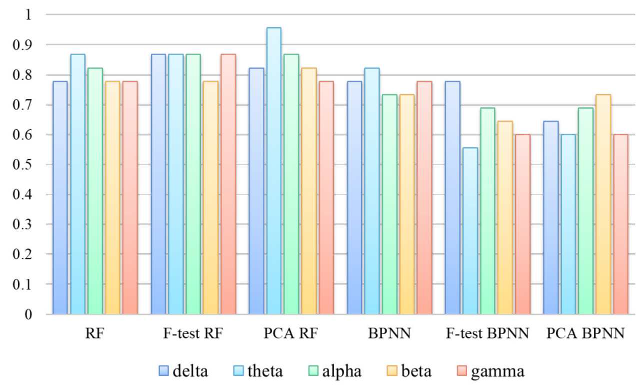

Fig.4 Classification accuracy of methods in each frequency band

As we can see from the results, theta frequency bands generally performed better than other frequency bands, which achieved the highest average classification accuracy of 95.6% with the PCA-RF method and 82.2% with the BPNN method. In addition, theta frequency bands achieved the highest average recall rate of 93.3% with the PCA-RF method and 73.3% with the BPNN method.

TABLE II
PERFORMANCE EVALUATION OF THE RANDOM FOREST

method	band	accuracy	precision	recall	F1
	delta	0.778	0.724	0.667	0.672
	theta	0.867	0.794	0.800	0.792
RF	alpha	0.822	0.739	0.733	0.731
	beta	0.778	0.683	0.667	0.671
	gamma	0.778	0.672	0.667	0.665
	delta	0.867	0.794	0.800	0.792
	theta	0.867	0.806	0.800	0.798
F-test RF	alpha	0.867	0.800	0.800	0.800
	beta	0.778	0.667	0.667	0.667
	gamma	0.867	0.800	0.800	0.800
	delta	0.822	0.756	0.733	0.739
	theta	0.956	0.944	0.933	0.933
PCA RF	alpha	0.867	0.838	0.800	0.794
	beta	0.822	0.790	0.733	0.739
	gamma	0.778	0.672	0.667	0.665

For the RF classifier, the PCA can significantly improve its performance. The average classification accuracies of PCA and RF methods were improved by 4.4%, 8.9%, 4.4%, and 4.4% in delta, theta, alpha, and beta bands, respectively, compared with the RF method alone. The average classification precisions of the combination of PCA and RF methods were improved by 3.2%, 15%, 9.9%, and 10.7% in delta, theta, alpha, and beta bands, respectively, compared with the RF method alone. The average classification recall rates of PCA and RF methods were improved by 6.7%, 13.3%, 6.7%, and 6.7% in delta, theta, alpha, and beta bands, respectively, compared with the RF method alone. The average classification F1-scores of PCA and RF methods were improved by 6.6%, 14.1%, 6.3%, and 6.8% in delta, theta, alpha, and beta bands, respectively, compared with the RF method alone.

TABLE III
PERFORMANCE EVALUATION OF THE BACKPROPAGATION NEURAL NETWORK

method	band	accuracy	precision	recall	F1
	delta	0.778	0.683	0.667	0.671
	theta	0.822	0.756	0.733	0.739
BPNN	alpha	0.733	0.617	0.600	0.604
	beta	0.733	0.568	0.600	0.572
	gamma	0.778	0.672	0.667	0.665
	delta	0.778	0.679	0.667	0.656
	theta	0.556	0.436	0.333	0.389
F-test BPNN	alpha	0.689	0.556	0.533	0.484
	beta	0.644	0.444	0.467	0.401
	gamma	0.600	0.431	0.400	0.395
	delta	0.644	0.692	0.467	0.598
	theta	0.600	0.442	0.400	0.421
PCA BPNN	alpha	0.689	0.605	0.533	0.646
	beta	0.733	0.728	0.600	0.757
	gamma	0.600	0.443	0.400	0.421

For the RF classifier, the F-test can also significantly improve its performance. The average classification accuracies of the F-test and RF methods were improved by 8.9%, 4.4%, and 8.9% in delta, alpha, and gamma bands, respectively, compared with the RF method alone. The average classification precisions of the F-test and RF methods were improved by 7.1%, 1.1%, 6.1%, and 12.8% in delta, theta, alpha, and gamma bands, respectively, compared with the RF method alone. The average classification recall rates of the F-test and RF methods were improved by 13.3%, 6.7%, and 13.3% in delta, alpha, and gamma bands, respectively, compared with the RF method alone. The average classification F1-scores of F-test and RF methods were improved by 12%, 0.6%, 6.9%, and 13.5% in delta, theta, alpha, and gamma bands, respectively, compared with the RF method alone.

IV. Discussion

Affective computing based on EEG signals in immersive VR is a rapidly growing field that enables naturalistic neuroscientific research while maintaining experimental control. It has a wide range of applications in emotional induction, psychological assistance, and the diagnosis of emotion-related disorders. Since 2018, researchers have proposed many emotional valence recognition methods based on supervised machine learning algorithms to automatically identify the individual's emotional states. However, these methods have high time complexity, high cost, and insufficient accuracy. This study has proposed an effective and reliable way to extract EEG features and encode emotional valence with encouraging performance.

As EEG signals involve a considerable amount of data, determining how to extract valuable features effectively is still the focus of much research. A recent review article shows that the most frequently used EEG features for affective computing in VR scenarios are frequency domain features [19]. Inspired by previous affective computing studies, we have sorted out the EEG frequency domain features commonly used in VR research and the features generally considered to perform well in emotional valence recognition in 2D scenarios [12-15]. Finally, we extracted ES, DE, DASM, and RASM as features. We can get a more comprehensive mapping and representation of emotions from various aspects, including the power attribute, logarithmic attribute, and cerebral asymmetry attribute of frequency-domain data. The results demonstrate that each band's joint presentation of multivariate features was complementary. These distinguishing features can be used to characterize emotional changes in EEG signals and achieve good classification results.

To reduce the computational cost of modeling, we applied PCA for dimensionality reduction. As an unsupervised linear transformation technique, PCA aims to find the directions of maximum variance in high-dimensional data and project the data onto a new subspace with equal or fewer dimensions than the original one. The PCA's validity is verified when combined with the RF method. It saves computing resources and significantly improves the performance of the model. The average classification accuracies of PCA and RF methods were improved by 4.4%, 8.9%, 4.4%, and 4.4% in delta, theta, alpha, and beta bands, respectively, compared with the RF method alone. PCA and RF methods preserve valid feature information while reducing the data dimensionality.

As a comparison, we proposed to combine the one-way ANOVA F-test statistics scheme to determine the most important features contributing to emotional valence recognition. This method was used to reduce the high data dimensionality of the feature space before the classification process. The F-test's validity is verified when combined with the RF method. It reduces the amount of calculation and significantly improves the model's performance. The average classification accuracies of the F-test and RF methods were improved by 8.9%, 4.4%, and 8.9% in delta, alpha, and gamma bands, respectively, compared with the RF method alone. In summary, combining the dimensionality reduction method and the RF classifier helps increase the model's accuracy and stability. Some of the feature values are shown to be irrelevant to emotional valence recognition, and some are redundant in our task. This discovery helps us reduce the computations of features and the complexity of the classification models. However, it is worth mentioning that combining the dimensionality reduction method with some classifiers, such as the BPNN in this paper, may play a negative role in the model performance. Therefore, choosing the appropriate combination according to the specific situation is necessary. Researchers should consider not only the computational cost and time complexity of the model, but also the classification model's performance.

It can be seen that the RF method performs better than the BPNN method, and there may be three reasons for this. (i) BPNN learns a non-linear mapping through a multi-layer network, which generally requires a large amount of training data. In our experiment, constrained by the number of experimenters, we adopt a small data set, which is detrimental to BPNN but has little impact on the RF algorithm [20]. (ii) RF requires far fewer hyperparameters than BPNN, which is more conducive to better classification results by adjusting the hyperparameters of the former. (iii) BPNN trains the parameters by gradient methods, so local optimum is usually obtained. This phenomenon restricts the stability of BPNN [21]. In contrast, RF can effectively avoid falling into local optima by randomly selecting samples and features.

The results demonstrate that the classification effect of the RF method with features in the theta band is superior to features in other frequency bands. The best performance was achieved by RF when it was combined with PCA, with the results of 95.6% (accuracy), 94.4% (precision), 93.3% (recall rate), and 93.3% (F1-score), showing the theta band's power in characterizing and separating emotional valence. It was found that theta asymmetry was one of the best indices for emotion recognition, which was quite robust to individual differences [22]. Theta power over the left and suitable frontal cerebral regions responded differentially to emotional valence. Positive valence emotions elicited larger theta power in the left hemisphere, whereas negative valence emotions elicited greater theta power in the right hemisphere [23]. As a biomarker, the power activation of the theta band may play a more critical role in emotion recognition [24]. Several recent studies have introduced different methods combined with frequency band features for automatically monitoring emotional valence in VR scenarios [12-14]. Compared with these studies, this research demonstrates higher discriminative performance.

The combination of EEG and VR has broad application prospects. Due to the fast pace of modern life and the aggravation of work pressure, emotional regulation and management have become more important. Previous studies have demonstrated that HMD made viewers perceive more reality than 2D environments, and positive emotions were triggered at a higher level of strength [25]. Thus, the emotion regulation system based on VR technology can help us stay positive and control negative urges during emotional distress. Based on the EEG analysis method of this study, the validity and reliability of the VR emotion regulation system can be objectively evaluated to guide the design of VR scenarios, which can avoid the social desirability bias brought by self-assessment questionnaires. Moreover, the initial screening and diagnosis of people with emotional disorders is an essential issue in clinical psychology. Taking advantage of the emotion induction by VR, objective electrophysiological neural markers can be extracted under different valence conditions to distinguish between normal and pathological populations. In addition, EEG signal-driven VR interaction technology is a new direction in brain-computer interfaces. This further expands the handle, button, and eye movement interaction, which can achieve more direct and naturalistic human-computer interaction. Future research could conduct real-time emotion monitoring based on EEG signals. Thus, the user's emotional state can control the virtual environment.

It should be noted that though this study has achieved encouraging results for automatically identifying emotional valence, several challenges still require additional effort. Firstly, future research should compare VR and ordinary imaging scenarios, which help determine the characteristics of EEG signals in VR scenarios. Secondly, although we have achieved high classification and recognition accuracy through machine learning methods, deep learning methods are also a path worth trying. In recent years, most studies have extracted hand-crafted features and used supervised machine learning methods to identify different emotional states in VR scenarios. However, hand-crafted feature engineering needs some background knowledge of neuroscience and more human intervention, which is complex, time-consuming, and experience-driven. Deep learning methods have demonstrated great promise in helping make sense of signals because of their capacity to automatically learn excellent feature representations from raw EEG data [26], which should be introduced to infer subjects' emotional states in VR scenarios.

V. Conclusions

This paper has established a standard emotion-induced VR video library presented through the portable HMD. We used the EEG acquisition device to obtain EEG signals generated by the participants while they were watching VR videos designed to elicit positive, negative, and neutral emotional states. We extracted EEG features of ES, DE, DASM, and RASM to represent the characteristics associated with emotional valence and compared their classification accuracy in five frequency bands. The results demonstrated that features extracted from the theta band were superior to features from other frequency bands for emotional valence decoding. The RF generally performed better than the BPNN. By combining the dimensionality reduction method (the F-test or PCA) and the RF classifier, it is possible to achieve encouraging classification results and increase computation speed and stability, which may facilitate the application of EEG-based affective computing technology in the virtual reality brain-computer interface field.

Acknowledgment

This work was supported by Grant No. LQ22C090007 from the Zhejiang Provincial Natural Science Foundation of China, No. 71942002 from the National Natural Science Foundation of China, No. 2021ZD0114303 from the National Science and Technology Major Project of the Ministry of Science and Technology of China, and No. 21YJAZH035 from the Ministry of Education of China. This work was also supported by Grant No. 111011-AC2202 and No. 2020KB0AC01 from the Scientific Project of Zhejiang Lab.

References

[1] J. Diemer, G. W. Alpers, H. M. Peperkorn, Y. Shiban and A. Mühlberger, "The impact of perception and presence on emotional reactions: a review of research in virtual reality," *Frontiers in Psychology*, vol. 6, 2015, pp.26. DOI: 10.3389/fpsyg.2015.00026

[2] N. Döllinger, C. Wienrich and M. E. Latoschik, "Challenges and opportunities of immersive technologies for mindfulness meditation: a systematic review," *Frontiers in Virtual Reality*, vol. 2, 2021, pp.644683. DOI: 10.3389/frvir.2021.644683

[3] D. Colombo, A. Díaz García, J. Fernandez Álvarez and C. Botella, "Virtual reality for the enhancement of emotion regulation," *Clinical Psychology & Psychotherapy*, vol. 28, no. 3, 2021, pp.519-537. DOI: 10.1002/cpp.2618

[4] J. G. Cruz-Garza, J. A. Brantley and S. Nakagome et al., "Deployment of mobile EEG technology in an art museum setting: Evaluation of signal quality and usability," *Frontiers in Human Neuroscience*, vol. 11, 2017, pp.527. DOI: 10.3389/fnhum.2017.00527

[5] S. M. Alarcao and M. J. Fonseca, "Emotions recognition using EEG signals: A survey," *IEEE Transactions on Affective Computing*, vol. 10, no. 3, 2017, pp.374-393. DOI: 10.1109/TAFFC.2017.2714671

[6] A. Shalbaf, S. Bagherzadeh and A. Maghsoudi, "Transfer learning with deep convolutional neural network for automated detection of schizophrenia from EEG signals," *Physical and Engineering Sciences in Medicine*, vol. 43, no. 4, 2020, pp.1229-1239. DOI: 10.1007/s13246-020-00925-9

[7] M. Banaei, A. Ahmadi, K. Gramann and J. Hatami, "Emotional evaluation of architectural interior forms based on personality differences using virtual reality," *Frontiers of Architectural Research*, vol. 9, no. 1, 2020, pp.138-147. DOI: 10.1016/j.foar.2019.07.005

[8] F. Tian, M. Hua, W. Zhang, Y. Li and X. Yang, "Emotional arousal in 2D versus 3D virtual reality environments," *Plos One*, vol. 16, no. 9, 2021, pp.e256211. DOI: 10.1371/journal.pone.0256211

[9] C. Stolz, D. Endres and E. M. Mueller, "Threat-conditioned contexts modulate the late positive potential to faces—A mobile EEG/virtual reality study," *Psychophysiology*, vol. 56, no. 4, 2019, pp.e13308. DOI: 10.1111/psyp.13308

[10] S. M. Hofmann, F. Klotzscheand A. Mariola et al., "Decoding subjective emotional arousal from EEG during an immersive Virtual Reality experience," *eLife,* vol. 10, 2021, pp. e64812. DOI: 10.7554/eLife.64812

[11] M. Horvat, M. Dobrinić, M. Novosel and P. Jerčić. "Assessing emotional responses induced in virtual reality using a consumer EEG headset: A preliminary report," In 2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO). IEEE, 2018, pp. 1006-1010. DOI: 10.23919/MIPRO.2018.8400184

[12] Marín-Morales, J. L. Higuera-Trujilloand A. Greco et al., "Affective computing in virtual reality: emotion recognition from brain and heartbeat dynamics using wearable sensors," *Scientific Reports*, vol. 8, no. 1, 2018, pp.13657. DOI:10.1038/s41598-018-32063-4

[13] J. Marín-Morales, J. L. Higuera-Trujilloand A. Greco et al., "Real vs. immersive-virtual emotional experience: analysis of psycho-physiological patterns in a free exploration of an art museum," *Plos One*, vol. 14, no. 10, 2019, pp. e223881. DOI: 10.1371/journal.pone.0223881

[14] N. S. Suhaimi, J. Mountstephens and J. Teo, "A Dataset for Emotion Recognition Using Virtual Reality and EEG (DER-VREEG): Emotional State Classification Using Low-Cost Wearable VR-EEG Headsets," *Big Data and Cognitive Computing*, vol. 6, no. 1, 2022, pp.16. DOI: 10.3390/bdcc6010016

[15] R. Duan, J. Zhu and B. Lu. "Differential entropy feature for EEG-based emotion classification", *in 2013 6th International IEEE/EMBS Conference on Neural Engineering (NER)*. IEEE, 2013: pp. 81-84. DOI: 10.1109/NER.2013.6695876

[16] M. Taouk, P. F. Lovibond and R. Laube, "Psychometric properties of a Chinese version of the short Depression Anxiety Stress Scales (DASS21)," *Report for new South Wales Transcultural Mental Health Centre, Cumberland Hospital, Sydney*, 2001.

[17] B. J. Li, J. N. Bailenson, A. Pines, W. J. Greenleaf and L. M. Williams, "A public database of immersive VR videos with corresponding ratings of arousal, valence, and correlations between head movements and self report measures," *Frontiers in Psychology*, vol. 8, 2017, pp.2116. DOI: 10.3389/fpsyg.2017.02116

[18] M. M. Bradley and P. J. Lang, "Measuring emotion: the self-assessment manikin and the semantic differential," Journal of Behavior Therapy and Experimental Psychiatry, vol. 25, no. 1, 1994, pp.49-59. DOI: 10.1016/0005-7916(94)90063-9

[19] J. Marín-Morales, C. Llinares, J. Guixeres and M. Alcañiz, "Emotion recognition in immersive virtual reality: From statistics to affective computing," *Sensors*, vol. 20, no. 18, 2020, pp.5163. DOI: 10.3390/s20185163

[20] A. Liaw and M. Wiener, "Classification and regression by random forest," *R news*, vol. 2, no. 3, 2002, pp.18-22.

[21] M. Liu, M. Wang, J. Wang and D. Li, "Comparison of random forest, support vector machine and back propagation neural network for electronic tongue data classification: Application to the recognition of orange beverage and Chinese vinegar," *Sensors and Actuators B: Chemical*, vol. 177, 2013, pp.970-980. DOI: 10.1016/j.snb.2012.11.071

[22] J. A. Coanand J. J. B. Allen, "Frontal EEG asymmetry as a moderator and mediator of emotion," *Biological psychology*, vol.67, no.1, 2004, pp.7-50. DOI:10.1016/j.biopsycho.2004.03.002

[23] L. Santamaria, V. Noreikaand S. Georgieva et al., "Emotional valence modulates the topology of the parent-infant inter-brain network," *bioRxiv*, 2019, pp.623355. DOI: 10.1016/j.neuroimage.2019.116341

[24] R. Du, R. M. Mehmood and H. J. Lee, "Alpha activity during emotional experience revealed by ERSP," *Journal of Internet Technology*, vol. 15, 2014, pp.775-782. DOI: 10.6138/JIT.2014.15.5.07

[25] M. Magdin, Z. Balogh and J. Reichel et al., "Automatic detection and classification of emotional states in virtual reality and standard environments (LCD): Comparing valence and arousal of induced emotions," *Virtual Reality*, vol. 25, no. 4, 2021, pp.1029-1041. DOI: 10.1007/s10055-021-00506-5

[26] D. Merlin Praveena, D. Angelin Sarahand S. Thomas George, "Deep learning techniques for EEG signal applications–a review," *IETE Journal of Research*, 2020, pp.1-8. DOI: 10.1080/03772063.2020.1749143