MERD-360VR: A Multimodal Emotional Response Dataset from 360° VR Videos Across Different Age Groups

Qiang Chen
Jiangxi University of Finance and Economics
Nanchang, China
qiangchen@jxufe.edu.cn

Shikun Zhou
Jiangxi University of Finance and Economics
Nanchang, China
zhoushikun1015@163.com

Yuming Fang
Jiangxi University of Finance and Economics
Nanchang, China
fa0001ng@e.ntu.edu.sg

Dan Luo
Jiangxi University of Finance and Economics
Nanchang, China
luodan678@163.com

Tingsong Lu
Jiangxi University of Finance and Economics
Nanchang, China
lutingsong@foxmail.com

Abstract

Virtual Reality (VR) technologies provide interactive experiences capable of evoking a wide spectrum of emotional responses from users. However, there is a notable scarcity of VR-based multimodal emotional response datasets designed to enhance the accuracy of emotionally immersive film and television productions. To address this gap, we conducted a study to develop a comprehensive, multimodal, annotated dataset capturing users' affective, physiological, and behavioral responses to 360° panoramic videos in immersive VR environment. Our dataset specifically focuses on two age groups: adults and minors. The dataset collection process involved gathering participants' self-reported emotional responses alongside objective measures, including behavioral data (e.g., head movements and gaze patterns) and physiological data (e.g., heart rate and skin conductance signals). To analyze the data, we employed subject-independent baseline classification algorithms to evaluate the usefulness of the dataset for emotion analysis. Furthermore, we assessed the consistency of participants' interactions with specific regions of the 360° panoramic videos across experimenters, and examined the correlation between physiological data and self-reported emotional responses. This publicly available multimodal dataset provides a valuable resource to facilitate numerous future efforts on VR-based affective computing research and on tailoring VR content to diverse audiences based on emotional and demographic profiles.

Keywords

Emotional arousal, user perception, affective computing, physiological signals, multimodal datasets, 360° panoramic video, virtual reality experience.

ACM Reference Format:
Qiang Chen, Shikun Zhou, Yuming Fang, Dan Luo, and Tingsong Lu. 2025. MERD-360VR: A Multimodal Emotional Response Dataset from 360° VR Videos Across Different Age Groups. In *Proceedings of the 27th International Conference on Multimodal Interaction (ICMI '25), October 13–17, 2025, Canberra, ACT, Australia.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3716553.3750777

ICMI '25, October 13–17, 2025, Canberra, ACT, Australia

ACM ISBN 979-8-4007-1499-3/2025/10
https://doi.org/10.1145/3716553.3750777

1 Introduction

In recent years VR technology has seen extensive applications, particularly through the widespread adoption of 360° VR content viewing via head-mounted displays (HMDs). This technology has demonstrated significant potential across various sectors, including gaming [21], film [16], tourism [15], and streaming platforms. As a new form of multimedia, 360° VR allows the viewer to freely rotate his/her head and eyes, thus providing more immersive experience. Such immersive interactions not only increase the engagement and immersion of the content, but also trigger strong emotional resonance.

The expression of human emotions is heavily dependent on ocular features, head movements, vocal qualities, and body gestures [20]. Notably, the eyes, often called the *windows to the soul*, play a crucial role in conveying emotions. Recently researchers have begun collecting physiological signals to analyze human emotional states using sensors. For example, eye and head movements can be used to infer emotional changes [12, 49]. Galvanic skin responses (GSR) [11, 27] reflect an individual's mental state by measuring fluctuations in skin conductance, while electrocardiograms (ECGs) [17, 25] reveal emotional variations through the change of heart rates. These techniques offer new tools for investigating human emotions. However, emotions are essentially subjective, and the emotional responses of individuals may not always manifest outwardly. In addition, significant age-related differences exist in emotional responses to the same stimuli. As highlighted in prior research [44], current 360° video emotion datasets do not adequately represent various age groups, leading to incomplete understanding of emotional responses across ages.

To address this gap, our study focuses on collecting a comprehensive dataset (called the MERD-360VR dataset) comprising eye and head movements, physiological data, and questionnaire responses from both minors and adults as they engage with immersive 360° VR content. This dataset enables us to investigate emotional and behavioral differences across age groups while uncovering the underlying mechanisms of emotional responses in VR environments. Given the rapid fluctuations in emotions [23], we have accounted for the potential comprehension challenges faced by minors and prioritized improving the efficiency of data collection [14]. To achieve this, we used a method that improves the accuracy of capturing user emotional responses, thus increasing the reliability and practical value of the collected data. This dataset can serve as a valuable resource for advancing VR-based affective computing research.

The main contributions can be summarized as follows.

- We introduced a new, publicly available, multimodal dataset (MERD-360VR) containing both the subjective emotional responses and objective behaviors of users who immersively view 360° videos in VR environment, including self-reported emotional responses, head movement and gaze, ECG data, and GSR data.
- We conducted statistical analysis on self-reported emotion responses, physiological data, and behavioral data, for two different age groups.

Our dataset is a valuable resource for affective computing, enabling studies of transient behavioral/physiological responses and emotional states in 360° panoramic video applications on HMDs. It also supports in-depth analysis of 360° visual attention modeling and exploration of the relationship between behavioral/physiological signals and emotions across age groups [43], enhancing understanding of human emotion elicitation and analysis. The full MERD-360VR dataset and comprehensive documentation are available on GitHub (https://qiangchen-cg.github.io/ or https://drive.google.com /file/d/1aI9KqXMiDMDnbweoKTGnKSsx6FGru0tb/view).

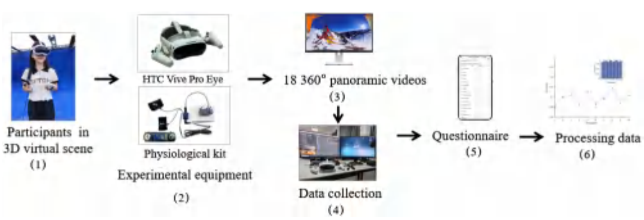

Figure 1: Pipeline of our dataset collection and processing

2 RELATED WORK

Emotions are subjective experiences arising from physiological and psychological responses [26]. They are integral to human life, influencing learning, expression, cognition, social communication, decision-making, and memory formation [39, 46]. The generation and recognition of emotions have been central to research in human-computer interaction (HCI) and VR, with emotion recognition playing a critical role in affective computing applications such as facial recognition, mental health assessment, and interpretation of human behavior [45].

To study and elicit emotions, researchers have developed various models to quantify and characterize emotional states. Notable models include Plutchik's Wheel of Emotions [22], the Tree of Emotions [1], and the Circumplex Model of Affect (CMA) [24]. The CMA evaluates emotions using three dimensions: *valence* (positive or negative emotions), *arousal* (emotional intensity), and *dominance* (the sense of control over emotions). Emotion induction in many studies typically involves presenting participants with emotional stimuli, such as audio [32] or video clips [40]. However, these traditional methods are non-immersive and limited in their ability to fully engage participants.

VR has become a powerful tool for emotional engagement, providing immersive 3D environments where users achieve full sensory immersion via HMDs [2]—environments especially effective at inducing and amplifying emotional responses [44]. Despite rising interest in VR-based emotion studies, there is a notable lack of public datasets tailored to them; many existing VR datasets focus mainly on visual saliency detection in computer vision. While researchers have recently recognized the value of integrating eye movement trajectories, head movement data, and physiological signals for understanding emotions, existing datasets are broadly divided into two categories: *360° panoramic static datasets* and *360° panoramic dynamic datasets* (the latter including immersive 2D flat video and 360° panoramic video datasets). This gap in comprehensive VR-based datasets underscores the need for more holistic resources that combine emotional stimuli, behavioral data, and physiological measurements to advance the field.

360° Panoramic Static Datasets. Tang et al. [37] conducted an experiment with 19 participants who viewed 360° panoramic images across various VR scenarios. They collected both eye movement trajectories and self-reported emotional responses to the stimuli. This dataset provides detailed emotional ratings for each image, making it a valuable resource for emotional and psychological research. Similarly, Riva et al. [31] explored users' emotional responses —specifically anxiety, relaxation, and neutrality — within three different VR park scenarios. Additionally, Lo et al. [19] presented a series of studies on emotion recognition, where physiological data, including sensor readings, were collected from 50 participants after viewing panoramic images and other stimuli. These datasets contribute valuable insights into emotion induction in static VR environments.

360° Panoramic Motion Datasets. Li et al. [18] used 360° panoramic videos and the Self-Assessment Manikin (SAM) scale to evaluate emotions, recording users' head movements, emotional arousal, and valence while they watched 73 videos. Similarly, Xue et al. [44] had participants wear a Vive Pro Eye headset to view 8 validated emotional 360° video clips, continuously annotating their valence and arousal while collecting physiological responses (e.g., heart rates, skin temperatures, electrodermal activities); they also gathered self-reports on momentary emotions, motion sickness, presence, and workload from 32 participants aged 18–33. Suhaimi et al. [35] developed a VR-based emotion recognition dataset using EEG, with low-cost wearable VR-EEG headsets to classify emotional states via brainwave patterns, involving 32 healthy participants (7 females, 25 males) aged 23–45. Further, Tabbaa et al. [36] introduced the *VREED* dataset, which combined behavioral signals (head movements, eye movements, pupillometry) with physiological responses (heart rates, skin temperatures, electrodermal activities) and EEG brainwave data, serving as a useful resource for emotion recognition research.

Existing 360° panoramic video-based emotion induction methods have limitations: small samples, poor scalability, insufficient diverse age representation, lack of behavioral/physiological data, and absence of post-experiment emotional perception questionnaires in many datasets. To address these, this work presents a multimodal dataset for age-specific emotional responses, featuring 18 partially validated 360° VR videos (six basic emotions) and collecting physiological signals, behavioral data, self-reported emotion ratings, and post-experiment questionnaires from participants across age groups.

3 MERD-360VR Dataset Collection

In this section, we describe the experimental protocol employed to view immersive 360° panoramic videos while simultaneously tracking and collecting multimodal physiological and behavioral data to analyze viewers' emotional responses in VR environment. Our experimental procedure is described below and illustrated in Figure 1.

We developed a custom scene in the Unity Engine to display 360° panoramic videos. Participants viewed a total of 18 emotionally evocative 360° panoramic videos using an HTC Vive Pro Eye headset. The headset includes an integrated Tobii Pro eye tracker with a sampling frequency of 120 Hz and an accuracy of 0.5°. It offers a 2K resolution, a refresh rate of 90 Hz, and a 120° field of view. The participants comfortably sat while viewing the stimuli, and the headset passively recorded their head and eye movement data.

In addition to the headset, participants wore BMD101 ECG sensors on their chests and GSR sensors on their fingers to record physiological signals. These sensors are illustrated in Figure 1. The BMD101 ECG sensor features a 16-bit analog-to-digital converter, enabling real-time measurement of heart rate, heart rate variability, respiration rate, relaxation levels, and average heart rate. The GSR sensors, operating at a voltage of 3.3V/5V with a sampling frequency of 193 Hz, recorded electrodermal activity (EDA) data.

Participants used a VR controller to randomly select and play 360° panoramic videos in the VR environment. After viewing each video, they were prompted with questions to assess whether an emotion was induced, the specific emotion, and its intensity. Specifically, participants responded to three groups of questions:

- **Emotion Induction**: Was an emotion induced? The options included positive emotion, negative emotion, or none.
- **Emotion Type**: If an emotion was induced, which emotion was it? The options included anger, sadness, fear, disgust, happiness, and surprise.
- **Emotion Intensity**: To what extent was the emotion experienced? This was recorded using a 5-point Likert scale, where 1 indicates the lowest intensity and 5 indicates the highest intensity.

Participants were allowed a short break between videos to calm down before viewing the next one. This process resulted in the collection of both annotated emotional data and corresponding physiological signals.

After completing the experiment, participants were immediately asked to fill out a questionnaire designed to evaluate their experiences during the study. This included assessing symptoms of motion sickness, describing their overall experience, and sharing their perspectives on any emotional challenges encountered. These responses were used as subjective measures in our data analysis. Motion sickness was specifically assessed using the widely used Simulator Sickness Questionnaire (SSQ) [6].

3.1 Stimuli

To collect data, we prepared 360° panoramic videos to induce mood swings—each with unique visual elements for easy immersion. For stimuli, we classified 360° videos by six basic emotions (happy, surprised, sad, angry, fearful, disgusted), selecting representative samples from public databases [36, 44] and 1080p clips with agreed emotion labels from YouTube. We chose 3 videos per emotion, totaling 18 clips. All included background music; 17 lasted 1 minute, and 1 lasted 3 minutes.

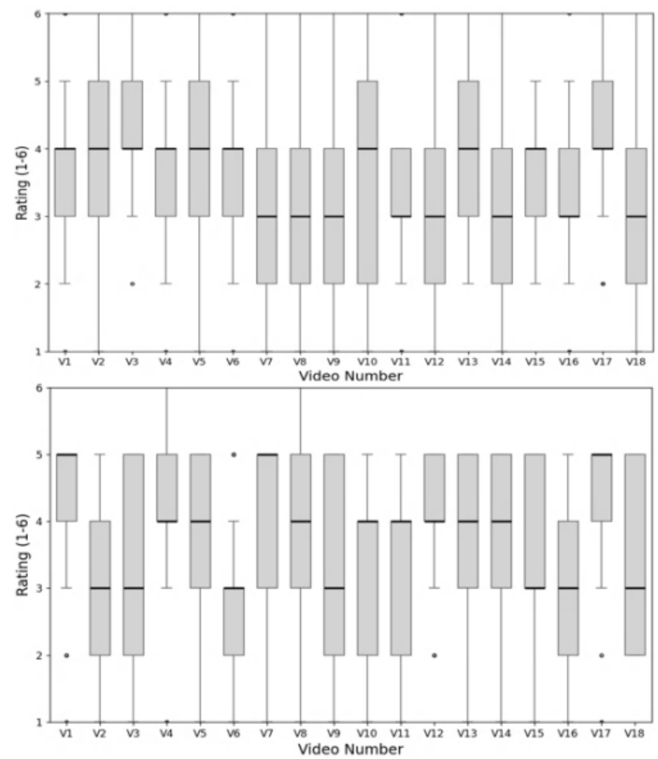

Figure 2: Boxplots of the emotion ratings by the adult group (top) and the minor group (bottom) ratings for the 18 videos in the study. The vertical coordinate Y denotes the level of emotional arousal while watching each video, and the horizontal coordinate X denotes the 18 used videos. The Length is determined by the distribution of the upper and lower quartiles. The top, bottom and middle lines of the box plot represent the top and bottom edges of the data and the median, respectively.

3.2 Participants

A total of 62 participants, including 31 females and 31 males, participated in our study. The group comprised 21 elementary school students aged 9 to 11 years (11 females and 10 males) and 41 college students aged 18 to 30 years (20 females and 21 males). None of them had prior experience with computer animation or simulation. They were monetarily compensated for their participation. All of them had normal corrected vision and were not color blind, and more than 80% of them had fewer than three VR viewing experiences previously.

4 Dataset Analysis

4.1 Comparative Analysis of Emotion Ratings

We conducted comparative analyses by dividing participants into adults and minors. Both groups were assessed using two-way randomization, absolute consistency, and mean Intraclass Correlation Coefficients (ICC) [3]. 41 adults were used for mood rating assessment. The mean ICC results across the 18 videos demonstrated that adults exhibited very good reliability in potency ratings, with an overall mean ICC of 0.75 ($p < 0.05$). Similarly, 21 minors were used for mood ratings, and their mean ICC results for the same 18 videos indicated good reliability, with an overall mean ICC of 0.60 ($p < 0.05$).

Figure 2 presents box plots illustrating the video mood ratings. Upon analysis of the box plots, we observe that adults were relatively conservative in their ratings compared to minors. This suggests that emotional fluctuations experienced by adults while watching the videos were less pronounced than those experienced by minors. For neutral emotions, adults were more likely to report feeling no emotion, which may be linked to individual preferences and levels of acceptance. In contrast, minors rarely reported a lack of emotional response during the experiment.

These findings collectively suggest the following: First, the methodology for video emotion ratings demonstrates a high degree of consistency and reliability. Second, emotional perception and fluctuations are more pronounced in minors compared to adults, highlighting age-related differences in emotional responses.

4.2 Head Movement Data Analysis

When viewing 360° videos in VR environment, head movement (HM) reflects the dynamic viewports of the participant [41]. As shown in Figure 3, we generated HM saliency maps for various videos based on the combined gaze sample points of 62 participants in our study. Six images, representing the six basic human emotions, are selected for display. All HM saliency maps are presented in an equirectangular format, with the Y-axis indicating the pitch angle of the head and the X-axis representing the yaw angle of the head.

When viewing 360° panoramic videos, existing studies [10, 42] reported that human attention tends to be strongly front-biased and equator-biased. To address the biases, some previous studies [44] used calibration of equipment during initialization to mitigate its effects. In our study, we instructed the participants to look straight ahead and adjust their viewing angles at the start of the experiment.

From Figure 3, we observe that most participants' attention remained concentrated near the equator while watching the videos. However, certain stimuli caused notable deviations. For example, in the case of a Disgust video featuring a cockroach, the attention of the participants shifted toward the top and bottom regions, indicating avoidance behavior. In contrast, the Fear video, which features a shark, captured the participants' attention predominantly on the shark itself.

Positive emotion videos such as Happy and Surprise exhibited more pronounced shifts in attention. For example, in a Happy video featuring SpongeBob SquarePants, the participants' attention frequently shifted to the right, likely due to the character's continuous forward motion. Similarly, in the Surprise video involving prisoners, the participants focused on dynamic elements, resulting in noticeable shifts in their attentional perspective. In comparison, videos depicting Anger and Sadness prompted less dramatic attention shifts. The participants mainly focused on the movements of the characters and the narrative content within these videos.

These findings highlight that the combination of emotional content and video stimuli significantly influences participants' viewing angles and attentional focus. Different emotional responses elicited by the videos result in heterogeneous patterns of attention distribution.

4.3 EDA Data Analysis

We also performed a quantitative analysis of the recorded EDA data. First, we followed the methodology outlined in previous work [48] to process the data for each participant who watched the videos. This included denoising, adjusting, and improving the data, as well as normalizing the physiological signals for each psychological state.

Previous studies [13, 47], heart rate (HR) data effectively reflects time-domain changes, so we used mean HR as the key analysis feature [48]. As shown in Figure 4, we processed and compared minors' and adults' HR data; violin plots show mean HR distribution across 18 videos for each age group, with a horizontal center line at 0 for easy comparison. Consistent with [44], we found no significant physiological differences across video types, possibly because most emotionally stimulating videos (around 1 minute) were too short to induce strong emotional effects. Additionally, standardized analysis of videos with inconsistent lengths is challenging, and prolonged VR viewing may increase participant stress, workload, or 3D vertigo [5].

Despite these limitations, Figure 4 shows that psychological traits can effectively characterize specific types of emotional videos. Our study revealed instances where higher rated videos (associated with higher arousal emotions) corresponded to greater HR changes compared to lower rated videos. We use a "V" followed by a number to represent the NTH video, and the number in parentheses represents the degree of emotional arousal collected using the 5-point Likert scale. For example, in the adult panorama V5 (V=3.93) and the minor panorama V12 (V=4.19) — the highest rated videos in their respective age groups — these videos also exhibited the highest mean HR values. Furthermore, we observed that minors generally had higher mean HR values compared to adults. This suggests that minors may be more sensitive to emotional stimuli.

4.4 GSR Data Analysis

As shown in Figure 5, we plot the frequency histograms of the distribution of the participants' mean GSR values in different emotional states. We also plot the corresponding kernel density curves. Also, as shown in Figure 6, we plot the mean GSR values against the emotional arousal levels and calculate the Spearman correlation coefficient for the entire dataset. According to the previous study

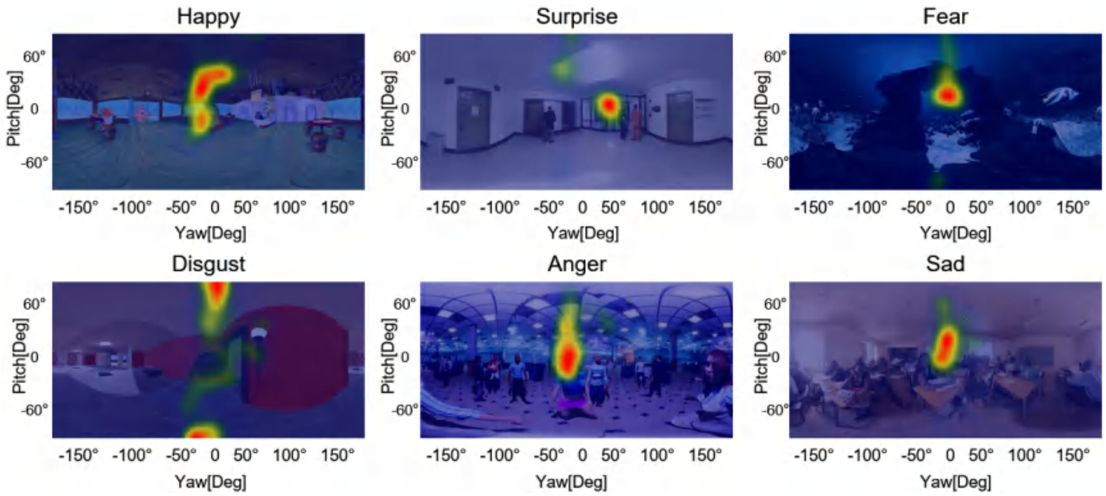

Figure 3: Six basic emotion sample frames with the saliency maps from 62 participants, where the Y-axis and X-axis denote the yaw and pitch angles of the head motion, respectively.

in [38], the ranking of videos based on averaged arousal ratings aligns closely with their ranking based on averaged GSR values. We verify this observation using our selected six videos (Figure 6). Additionally, we calculate the Spearman correlation coefficient (CC) between the mean GSR values and their corresponding emotional arousal ratings. The Spearman correlation coefficient ρ is found to be 0.94, with a p-value of 0.004, indicating a strong and statistically significant correlation.

Table 1: Questions included in the post-study questionnaire

Question	Answer
1. How influence do you think background music has on emotional arousal?	1-5 points
2. Do you think age affects the expression of one's emotions?	Yes/No
3. Do you think negative or positive emotions affect you more?	Yes/No
4. Do you think gender can affect a person's mood?	Yes/No
5. For the six emotions you experience, which one(s) do you think have the greatest impact on you?	Happy/Surprise/Sad/ Fear/Anger/Disugst

4.5 Questionnaire Response Analysis

We collected post-study questionnaire responses from a total of 62 participants, all of whom reported experiencing emotional changes during the experiment. The majority also indicated that they did not feel any discomfort during participation. The questionnaire contains five questions, as shown in Table 1. Figure 7 presents an analysis of the questionnaire responses across different age groups.

In general, 70.71% of the participants stated that negative emotions had the most significant impact on their moods, while more than 58.20% identified happiness or fear as the most influential emotions. Among minors, 61.91% believed that positive emotions had the greatest impact, with more than 56.12% specifically highlighting happiness or fear as the most influential. In contrast, 75.11% of adults reported that negative emotions had the strongest effect, with more than 59. 24% also emphasizing happiness or fear as the most impactful.

Regarding the influence of background music on emotions, both minors and adults exhibited similar responses. 61.26% of the participants believed that background music had a strong emotional effect, while 25.13% considered its impact significant. In particular, none of the participants believed that the background music had no effect on emotions.

When examining the effects of age and gender on emotional expression, 85.51% of the participants believed that age played a significant role, while 68.42% felt that gender also had a strong influence.

During follow-up discussion with the participants, minors generally reported that animated cartoon videos elicited stronger emotional responses, whereas adults found that more realistic videos had a greater impact on their emotions.

5 Experiments

To enable further analysis of the validity and reliability of our dataset, in this section we provide baseline classification experiments based on state-of-the-art machine learning algorithms, as well as ablation study experiments.

5.1 Classification Experiments

5.1.1 Classification Tasks. Our baseline experiments are inspired by previous work [29]. Two main tasks were tested on our dataset:

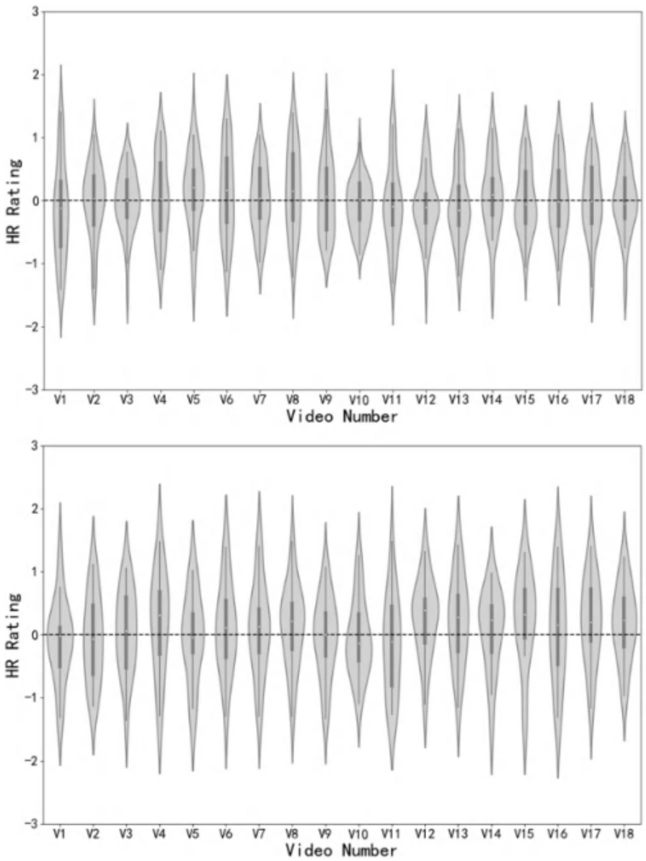

Figure 4: Heart rate characteristic distributions of the adult (top) and minor (bottom) groups across the 18 panorama videos, where the Y-axis represents the distribution of the experimenters' mean heart rates mapped to the interval from -3 to 3, and V1 to V18 on the X-axis represent the 18 videos used in our study. The dotted line in the middle is used as a marker for comparison between the different graphs, the width of the graphs represents the density of the mean heart rate distribution.

(a) (*3-class classification*) Ternary classification of positive, negative, and neutral, and (b) (*7-class classification*) a heptadic categorization based on the six basic human emotions plus neutral: happy, surprised, fear, disgust, sadness, anger, and neutral.

Table 2: Classification performance comparisons among traditional machine learning methods. "DT" denotes the decision tree method and "RF" denotes the random forest method. The best result for each metric is highlighted in bold.

	3-Class			7-Class		
ML-Classifier	ACC	PRE	REC	ACC	PRE	REC
KNN	60.27%	**50.60%**	0.40	50.23%	33.13%	0.32
DT	**68.95%**	47.51%	**0.48**	**53.88%**	44.87%	**0.47**
RF	66.67%	45.10%	0.46	48.40%	**46.81%**	0.40

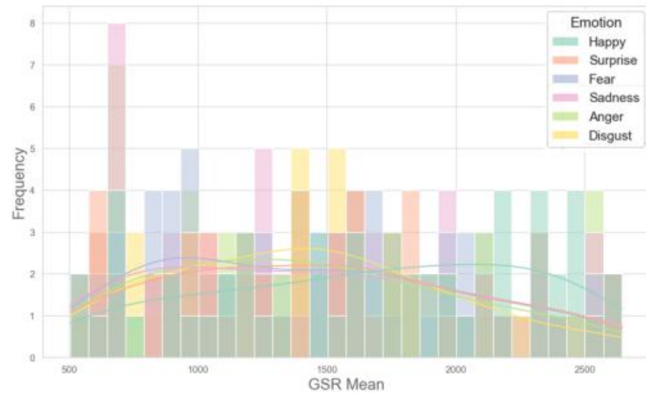

Figure 5: The frequencies of different mean GSR values when participants had different emotional states. The color kernel density curves represent the fluctuations of the distributions for different emotions.

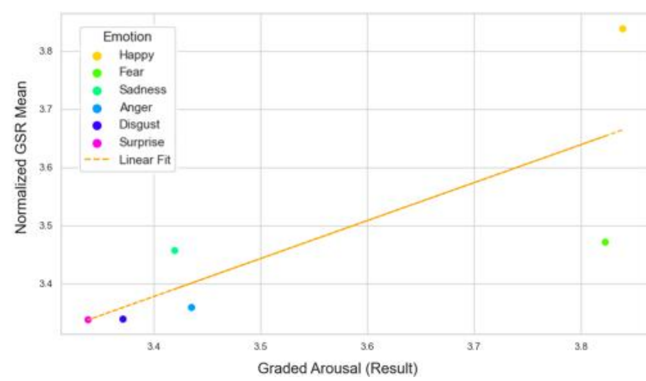

Figure 6: Plotting of the relation between the mean GSR values and emotional arousal levels for the full range dataset. The X-axis represents the subjective emotional arousal ratings of the experimenters, and the Y-axis represents the normalized mean GSR values.

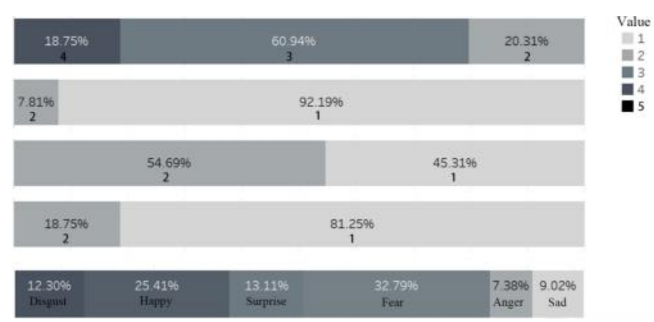

Figure 7: Visualization of the post-study questionnaire responses. The visualizations from top to bottom correspond to the responses to the questions 1 to 5 in Table 1.

5.1.2 Classification and Evaluation Methods. Classical machine learning methods (including deep learning algorithms) have been

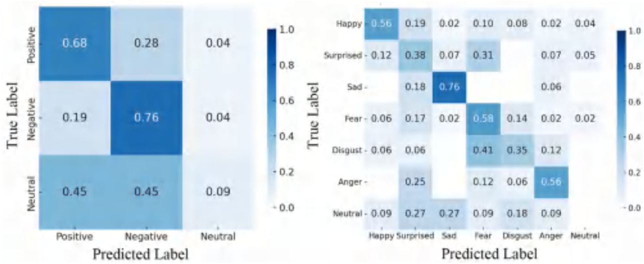

Figure 8: The confusion matrices of the classification results (left: 3-class classification; right: 7-class classification) by the decision tree algorithm.

Table 3: Classification performance comparisons of the CNN and LSTM methods. The best result for each metric is highlighted in bold.

	3-Class			7-Class		
DL-Classifier	ACC	PRE	REC	ACC	PRE	REC
LSTM	60.27%	**41.86%**	0.41	31.51%	13.54%	0.17
CNN	**61.19%**	41.39%	**0.42**	**37.44%**	**29.21%**	**0.26**

Table 4: Ablation study results. "Physio" denotes physiological signals, "AS" denotes age and sex features, and "HM/EM" denotes behavioral signals. The best result for each metric is highlighted in bold.

	3-Class			7-Class		
	ACC	PRE	REC	ACC	PRE	REC
Physio+AS	54.71%	36.36%	0.36	26.01%	25.46%	0.16
HM/EM	59.36%	40.40%	0.40	29.22%	12.55%	0.19
HM/EM+Ph+AS	**60.27%**	**50.60%**	**0.40**	**50.23%**	**33.13%**	**0.32**

used to classify emotions based on our dataset. Specifically, we tested the following methods: Nearest Neighbor [9], Decision Tree [33], and Random Forest (RF) [34], Convolutional Neural Networks (CNN), and LSTM Neural Networks.

We first preprocessed the data to eliminate missing and problematic entries. Then, we extracted some features from the data. Inspired by previous studies [43], we extracted the following features from behavioral and physiological signals, including the mean, median, standard deviations, first order and second order differentials of the pitch/yaw angles of head movement. We also extracted features from the collected physiological GSR and EDA signals, including the raw value, mean, standard deviations, and first order and second order differentials. Also, the extracted features from heart rate data include maximum, mean, and acceleration mean values.

The specific network architectures of the two deep learning methods (CNN-based and LSTM-based) are described below. Both of them are two-layer convolutional neural networks consisting of a convolutional layer, an activation layer, and a pooling layer. The number of neurons is 16, the maximum number of training rounds is 70, and the batch size is set to 50. During the training process, the feature selection of all the methods was carried out by adopting an L1 paradigm for the selection of the ten most significant features. The ratio among the training set, the validation set, and the test set is 7:1:2. In order to ensure the robustness of the experimental results, we conducted a ten-fold cross-validation for all the methods.

Finally, in order to evaluate the performance of these methods, we chose three widely-used metrics to assess their quantitative performances: (1) accuracy (ACC), which is used to measure the proportion of correctly predicted samples to the total samples; (2) precision (PRE), which is used to measure the proportion of samples predicted to be a positive class that are actually positive; and (3) recall (REC), which is used to measure the proportion of samples actually predicted to be a positive class that are actually positive.

5.1.3 Classification Results. The classification results by selected traditional machine learning methods are shown in Table 2. In the three-class classification case, the decision tree method achieved the highest accuracy (ACC) and recall (REC) rate, and the KNN method has the highest precision (PRE). In the seven-class classification case, the decision tree method achieved the highest accuracy and the recall rate. Figure 8 shows the confusion matrices of the classification results by the decision tree algorithms. Table 3 shows the classification results by the two selected deep learning methods. In general, the CNN method performed better, with a slightly lower accuracy than LSTM in the three-class classification task, while having other higher metrics than LSTM in all other cases.

5.2 Ablation Study

In order to analyze the significance of different data modalities for classifications, we performed an ablation study using the nearest neighbor method to assess the impact of physiological data (Physio), behavioural data (HM/EM), and age/sex information (AS) in both the tri- and hepta-classification tasks.

As shown in Table 4, when physiological or behavioural data were used alone, both performed relatively close to each other in terms of classification performance, and no significant differences was presented in either the three- or seven-classification tasks. This suggests that it is difficult to achieve a significant advantage in the classification tasks by relying only on information from a single modality. However, when all the modalities are fused, the classification performance is significantly improved, especially in the seven-classification task, where all of ACC, PRE, and REC are improved substantially. This result suggests that although there is no significant difference between the individual performances of physiological signals and behavioral data, the combination of the two can effectively improve the classification accuracy. Meanwhile, this also further validates the advantages of multimodal data fusion in emotion classification tasks [7].

5.3 Discussion

First, from the quantitative analysis of our dataset, we demonstrate that the emotion-level rating method used in our study is highly consistent and reliable with the expected attributes of the selected videos. Additionally, we observed that minors exhibit more pronounced emotional perception and fluctuations compared to individuals of different age groups.

Second, saliency maps based on head movement among viewers reveal distinct patterns, with variations in emotional responses and viewing behavior influenced by the content of the videos. Different emotional triggers lead to heterogeneous viewing perspectives and attentional focus. Our findings align with previous studies [10, 42, 44], confirming a similar front-and-center bias in saliency maps.

Third, an analysis of physiological data indicates a strong correlation between emotional ratings and heart rates. Specifically, when a video is reported as highly emotionally arousing, heart rate values tend to increase for most participants [44]. In addition, minors generally have higher mean heart rate values compared to adults. Consistent with a previous finding [38], we found that the ordering of mean graded arousal in videos closely aligns with the ordering of mean GSR values. In addition, emotional arousal ratings and GSR signals show a positive and statistically significant correlation.

Fourth, the responses to the post-experiment questionnaire highlight key differences in emotional sensitivity between age groups. Minors tend to be more responsive to positive emotions, while adults are more sensitive to negative emotions. Fear, however, has a strong emotional impact across both age groups. The participants agreed that background music, age, and gender influence emotional responses. Furthermore, minors perceived animated cartoon videos as more emotionally stimulating, whereas adults felt that videos with a more realistic aesthetic had a stronger effect on mood.

6 Potential Research and Applications

Many potential research and applications can be developed based on our MERD dataset. In the following, we describe some potential research directions.

- *Emotion-labelled Attention and Saliency Models.* This dataset can be used for in-depth study of attentional mechanisms and emotion saliency models, and to analyse the relationship between an individual's attentional allocation and their emotional state when viewing 360° panoramic videos. Combining eye movement trajectories, head movements and physiological signals can optimise emotion-driven visual saliency prediction models and enhance emotion perception in VR and human-computer interaction systems [30].
- *Multimodal Sentiment Analysis.* Combining physiological signals, behavioural data and demographic information, multimodal sentiment analysis enables more accurate emotion recognition. This dataset can be used to build cross-modal emotion understanding systems, improve the robustness of affective computing, and play a role in areas such as mental health monitoring and intelligent recommendation.
- *Reverse Mood Analysis.* By analysing the patterns of physiological signals over time, the patterns and triggers of mood fluctuations can be studied. Using time series modelling, we could predict emotional transitions and apply them to personalised emotional intervention, emotionally adaptive VR experiences, and intelligent feedback systems to improve emotional prediction and regulation.
- *Demographic-based Emotion Research.* To study the emotional, physiological and behavioural changes of different age and gender groups in 360° panoramic videos, and to explore the impact of demographic factors on emotional responses. The dataset can be used to construct personalised emotion prediction models to facilitate accurate emotion calculation and personalised interaction systems.
- *Advanced Feature Fusion and Model Optimisation.* Further explore multimodal feature fusion strategies to optimise the generalisation capability of emotion recognition using deep learning. Combined with self-supervised learning, multi-task learning and other methods, the cross-scene adaptability of the model could be improved to support the application of AI in affective computing, mental health assessment, and intelligent interaction.

7 Safe and Responsible Innovation Statement

The MERD-360VR dataset will be made publicly available on GitHub upon publication to support reproducibility and encourage research in affective modeling, sentiment analysis, and VR content customization. This study was conducted in accordance with institutional policies, and per the lead author's institution, it qualifies for exemption from IRB review. The dataset will be released with accompanying documentation outlining consent procedures, privacy protections, and usage limitations. We acknowledge the potential for misuse in emotionally manipulative applications and encourage future users to consider ethical deployment, inclusivity, and bias mitigation when leveraging this resource in multimodal interaction systems.

8 Conclusion

We introduce a multimodal dataset for age-specific emotional perception in panoramic virtual videos, collecting objective (physiological, behavioral) and subjective data from participants, with accompanying data analysis and classification experiments. However, the dataset has limitations: first, 360° video selection is constrained by scarce public databases with validated emotion labels, and varying video emotional content may weaken uniform participant responses; second, the experimental setup—requiring VR headsets plus GSR/EDA sensors—may cause discomfort and limit non-laboratory use [8], though sensor advances could expand applications beyond clinical settings; third, EEG data was not collected due to challenges in obtaining stable, high-quality signals with HMDs [17]; finally, certain groups (e.g., individuals with depression [28]) may struggle to report emotions, risking mismatches between self-assessments and physiological responses. In sum, the MERD dataset fills gaps in emotion computation research on age-related physiological/behavioral response differences and supports future VR-based emotion prediction and perception studies [4].

Acknowledgments

Qiang Chen was supported in part by the National Natural Science Foundation of China (Grant No. 62262024), and Science Foundation of the Jiangxi Province of China (Grant No. 20232BAB202023). Yuming Fang was supported in part by the National Natural Science Foundation of China (Grant No. U24A20220).

References

[1] Mohammed Isam Mohammed Abdel-Magid. 2017. The tree of emotions: Exploring the relationships of basic human emotions. *International Journal of Indian Psychology* 5, 1 (2017), 22–37.

[2] Vladislav Angelov, Emiliyan Petkov, Georgi Shipkovenski, and Teodor Kalushkov. 2020. Modern virtual reality headsets. In *2020 International congress on human-computer interaction, optimization and robotic applications (HORA)*. IEEE, 1–5.

[3] John J Bartko. 1966. The intraclass correlation coefficient as a measure of reliability. *Psychological reports* 19, 1 (1966), 3–11.

[4] Javad Birjandtalab, Diana Cogan, Maziyar Baran Pouyan, and Mehrdad Nourani. 2016. A non-EEG biosignals dataset for assessment and visualization of neurological status. In *2016 IEEE International Workshop on Signal Processing Systems (SiPS)*. IEEE, 110–114.

[5] Marc Van den Broeck, Fahim Kawsar, and Johannes Schöning. 2017. It's all around you: Exploring 360 video viewing experiences on mobile devices. In *Proceedings of the 25th ACM international conference on Multimedia*. 762–768.

[6] Phillip Brown, Pieter Spronck, and Wendy Powell. 2022. The simulator sickness questionnaire, and the erroneous zero baseline assumption. *Frontiers in Virtual Reality* 3 (2022), 945800.

[7] Carlos Busso, Zhigang Deng, Serdar Yildirim, Murtaza Bulut, Chul Min Lee, Abe Kazemzadeh, Sungbok Lee, Ulrich Neumann, and Shrikanth Narayanan. 2004. Analysis of emotion recognition using facial expressions, speech and multimodal information. In *Proceedings of the 6th international conference on Multimodal interfaces*. 205–211.

[8] Hui-Ling Chan, Po-Chih Kuo, Chia-Yi Cheng, and Yong-Sheng Chen. 2018. Challenges and future perspectives on electroencephalogram-based biometrics in person recognition. *Frontiers in neuroinformatics* 12 (2018), 66.

[9] Padraig Cunningham and Sarah Jane Delany. 2021. K-nearest neighbour classifiers-a tutorial. *ACM computing surveys (CSUR)* 54, 6 (2021), 1–25.

[10] Erwan J David, Jesús Gutiérrez, Antoine Coutrot, Matthieu Perreira Da Silva, and Patrick Le Callet. 2018. A dataset of head and eye movements for 360 videos. In *Proceedings of the 9th ACM multimedia systems conference*. 432–437.

[11] Amita Dessai and Hassanali Virani. 2023. Emotion Classification Based on CWT of ECG and GSR Signals Using Various CNN Models. *Electronics* 12, 13 (2023), 2795.

[12] Yu Ding, Lei Shi, and Zhigang Deng. 2017. Perceptual enhancement of emotional mocap head motion: An experimental study. In *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, 242–247.

[13] Julien Fleureau, Philippe Guillotel, and Izabela Orlac. 2013. Affective benchmarking of movies based on the physiological responses of a real audience. In *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*. IEEE, 73–78.

[14] Shkurta Gashi, Elena Di Lascio, Bianca Stancu, Vedant Das Swain, Varun Mishra, Martin Gjoreski, and Silvia Santini. 2020. Detection of artifacts in ambulatory electrodermal activity data. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 2 (2020), 1–31.

[15] Hanyoung Go and Myunghwa Kang. 2023. Metaverse tourism for sustainable tourism development: Tourism agenda 2030. *Tourism Review* 78, 2 (2023), 381–394.

[16] Michael Gödde, Frank Gabler, Dirk Siegmund, and Andreas Braun. 2018. Cinematic narration in VR–rethinking film conventions for 360 degrees. In *Virtual, Augmented and Mixed Reality: Applications in Health, Cultural Heritage, and Industry: 10th International Conference, VAMR 2018, Held as Part of HCI International 2018, Las Vegas, NV, USA, July 15-20, 2018, Proceedings, Part II 10*. Springer, 184–201.

[17] Muhammad Anas Hasnul, Nor Azlina Ab. Aziz, and Azlan Abd. Aziz. 2023. Augmenting ECG data with multiple filters for a better emotion recognition system. *Arabian Journal for Science and Engineering* 48, 8 (2023), 10313–10334.

[18] Benjamin J Li, Jeremy N Bailenson, Adam Pines, Walter J Greenleaf, and Leanne M Williams. 2017. 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* 8 (2017), 2116.

[19] Wen-Chih Lo, Ching-Ling Fan, Jean Lee, Chun-Ying Huang, Kuan-Ta Chen, and Cheng-Hsin Hsu. 2017. 360 video viewing dataset in head-mounted virtual reality. In *Proceedings of the 8th ACM on Multimedia Systems Conference*. 211–216.

[20] Yifei Lu, Wei-Long Zheng, Binbin Li, and Bao-Liang Lu. 2015. Combining Eye Movements and EEG to Enhance Emotion Recognition.. In *IJCAI*, Vol. 15. Buenos Aires, 1170–1176.

[21] Aline Menin, Rafael Torchelsen, and Luciana Nedel. 2018. An analysis of VR technology used in immersive simulations with a serious game perspective. *IEEE computer graphics and applications* 38, 2 (2018), 57–73.

[22] Manshad Abbasi Mohsin and Anatoly Beltiukov. 2019. Summarizing emotions from text using Plutchik's wheel of emotions. In *7th scientific conference on information technologies for intelligent decision making support (ITIDS 2019)*. Atlantis Press, 291–294.

[23] Frederik Nagel, Reinhard Kopiez, Oliver Grewe, and Eckart Altenmüller. 2007. EMuJoy: Software for continuous measurement of perceived emotions in music. *Behavior Research Methods* 39, 2 (2007), 283–290.

[24] Labake Odushegun. 2023. Aesthetic semantics: Affect rating of atomic visual web aesthetics for use in affective user experience design. *International Journal of Human-Computer Studies* 171 (2023), 102978.

[25] Faruk Enes Oğuz, Ahmet Alkan, and Thorsten Schöler. 2023. Emotion detection from ECG signals with different learning algorithms and automated feature engineering. *Signal, Image and Video Processing* 17, 7 (2023), 3783–3791.

[26] Edward F Pace-Schott, Marlissa C Amole, Tatjana Aue, Michela Balconi, Lauren M Bylsma, Hugo Critchley, Heath A Demaree, Bruce H Friedman, Anne Elizabeth Kotynski Gooding, Olivia Gosseries, et al. 2019. Physiological feelings. *Neuroscience & Biobehavioral Reviews* 103 (2019), 267–304.

[27] Varsha Kiran Patil, Vijaya R Pawar, Shreiya Randive, Rutika Rajesh Bankar, Dhanashree Yende, and Aditya Kiran Patil. 2023. From face detection to emotion recognition on the framework of Raspberry pi and galvanic skin response sensor for visual and physiological biosignals. *Journal of Electrical Systems and Information Technology* 10, 1 (2023), 24.

[28] Anders Petersen and Ole Jacob Madsen. 2017. Depression: Diagnosis and suffering as process. *Nordic Psychology* 69, 1 (2017), 19–32.

[29] Bastian Pfleging, Drea K Fekety, Albrecht Schmidt, and Andrew L Kun. 2016. A model relating pupil diameter to mental workload and lighting conditions. In *Proceedings of the 2016 CHI conference on human factors in computing systems*. 5776–5788.

[30] Minglang Qiao, Mai Xu, Zulin Wang, and Ali Borji. 2020. Viewport-dependent saliency prediction in 360 video. *IEEE Transactions on Multimedia* 23 (2020), 748–760.

[31] Giuseppe Riva, Fabrizia Mantovani, Claret Samantha Capideville, Alessandra Preziosa, Francesca Morganti, Daniela Villani, Andrea Gaggioli, Cristina Botella, and Mariano Alcañiz. 2007. Affective interactions using virtual reality: the link between presence and emotions. *Cyberpsychology & behavior* 10, 1 (2007), 45–56.

[32] Mehmet Cenk Sezgin, Bilge Gunsel, and Gunes Karabulut Kurt. 2012. Perceptual audio features for emotion detection. *EURASIP Journal on Audio, Speech, and Music Processing* 2012 (2012), 1–21.

[33] Yan-Yan Song and LU Ying. 2015. Decision tree methods: applications for classification and prediction. *Shanghai archives of psychiatry* 27, 2 (2015), 130.

[34] Jaime Lynn Speiser, Michael E Miller, Janet Tooze, and Edward Ip. 2019. A comparison of random forest variable selection methods for classification prediction modeling. *Expert systems with applications* 134 (2019), 93–101.

[35] Nazmi Sofian Suhaimi, James Mountstephens, and Jason Teo. 2022. 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* 6, 1 (2022), 16.

[36] Luma Tabbaa, Ryan Searle, Saber Mirzaee Bafti, Md Moinul Hossain, Jittrapol Intarasisrisawat, Maxine Glancy, and Chee Siang Ang. 2021. VREED: Virtual reality emotion recognition dataset using eye tracking & physiological measures. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies* 5, 4 (2021), 1–20.

[37] Wei Tang, Shiyi Wu, Toinon Vigier, and Matthieu Perreira Da Silva. 2020. Influence of emotions on eye behavior in omnidirectional content. In *2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, 1–6.

[38] Alexander Toet, Fabienne Heijn, Anne-Marie Brouwer, Tina Mioch, and Jan BF van Erp. 2020. An immersive self-report tool for the affective appraisal of 360 VR videos. *Frontiers in Virtual Reality* 1 (2020), 552587.

[39] Samarth Tripathi, Shrinivas Acharya, Ranti Sharma, Sudhanshi Mittal, and Samit Bhattacharya. 2017. Using deep and convolutional neural networks for accurate emotion classification on DEAP data. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 31. 4746–4752.

[40] Minchao Wu, Wei Teng, Cunhang Fan, Shengbing Pei, Ping Li, and Zhao Lv. 2023. An investigation of olfactory-enhanced video on EEG-based emotion recognition. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 31 (2023), 1602–1613.

[41] Mai Xu, Chen Li, Shanyi Zhang, and Patrick Le Callet. 2020. State-of-the-art in 360 video/image processing: Perception, assessment and compression. *IEEE Journal of Selected Topics in Signal Processing* 14, 1 (2020), 5–26.

[42] Mai Xu, Yuhang Song, Jianyi Wang, MingLang Qiao, Liangyu Huo, and Zulin Wang. 2018. Predicting head movement in panoramic video: A deep reinforcement learning approach. *IEEE transactions on pattern analysis and machine intelligence* 41, 11 (2018), 2693–2708.

[43] Tong Xue, Abdallah El Ali, Gangyi Ding, and Pablo Cesar. 2021. Investigating the relationship between momentary emotion self-reports and head and eye movements in HMD-based 360 VR video watching. In *Extended abstracts of the 2021 CHI conference on human factors in computing systems*. 1–8.

[44] Tong Xue, Abdallah El Ali, Tianyi Zhang, Gangyi Ding, and Pablo Cesar. 2021. CEAP-360VR: A Continuous Physiological and Behavioral Emotion Annotation Dataset for 360° VR Videos. *IEEE Transactions on Multimedia* 25 (2021), 243–255.

[45] Dongri Yang, Abeer Alsadoon, PW Chandana Prasad, Ashutosh Kumar Singh, and Amr Elchouemi. 2018. An emotion recognition model based on facial recognition in virtual learning environment. *Procedia Computer Science* 125 (2018), 2–10.

[46] Shanbin Zhang, Guangyuan Liu, and Xiangwei Lai. 2015. Classification of evoked emotions using an artificial neural network based on single, short-term physiological signals. *Journal of Advanced Computational Intelligence and Intelligent Informatics* 19, 1 (2015), 118–126.

[47] Tianyi Zhang, Abdallah El Ali, Chen Wang, Alan Hanjalic, and Pablo Cesar. 2020. Rcea: Real-time, continuous emotion annotation for collecting precise mobile video ground truth labels. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–15.

[48] Bobo Zhao, Zhu Wang, Zhiwen Yu, and Bin Guo. 2018. EmotionSense: Emotion recognition based on wearable wristband. In *2018 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*. IEEE, 346–355.

[49] Ling Zhu and Jiufang Lv. 2023. Review of studies on user research based on EEG and eye tracking. *Applied Sciences* 13, 11 (2023), 6502.