Check for updates

Facial expression recognition in virtual reality environments: challenges and opportunities

Zhihui Zhang*, Josep M. Fort and Lluis Giménez Mateu

Escola Tècnica Superior d'Arquitectura de Barcelona, Universitat Politècnica de Catalunya, Barcelona, Spain

This study delved into the realm of facial emotion recognition within virtual reality (VR) environments. Using a novel system with MobileNet V2, a lightweight convolutional neural network, we tested emotion detection on 15 university students. High recognition rates were observed for emotions like "Neutral", "Happiness", "Sadness", and "Surprise". However, the model struggled with 'Anger' and 'Fear', often confusing them with "neutral". These discrepancies might be attributed to overlapping facial indicators, limited training samples, and the precision of the devices used. Nonetheless, our research underscores the viability of using facial emotion recognition technology in VR and recommends model improvements, the adoption of advanced devices, and a more holistic approach to foster the future development of VR emotion recognition.

1. Introduction

The intersections of technology, emotion, and human experience have led to groundbreaking advancements and insights in numerous fields. Two technological paradigms that have witnessed profound evolution over the decades and remain at the forefront of such intersections are Virtual Reality (VR) and Emotion Recognition.

VR technology, from its inception, has promised to revolutionize our perception of reality. Emerging as a game-changer in the world of interactive entertainment, VR offered players not just games but lived experiences. The watershed moment for this was marked by the release of devices like the Oculus Rift, signaling a surge of interest and investment in the VR domain (Tan et al., 2015). However, the realm of VR isn't confined solely to gaming. It spans diverse genres, from adventurous explorations to competitive sports, and with the continuous strides in AI and interactivity, the horizon looks even more promising for VR gaming.

Beyond entertainment, VR's tentacles have reached out into transformative areas such as education, where it replicates real-world scenarios, affording students hands-on opportunities (Merchant et al., 2014). The medical sector has not been untouched either. Breakthroughs in patient rehabilitation and surgical simulations have been made possible, offering new avenues for practitioners (Rose et al., 2005). The immersive nature of VR has found resonance in architectural design, letting clients and designers visualize spaces before they materialize (Wang et al., 2013). Furthermore, in a world challenged by a pandemic, virtual tourism stood out as a beacon for the travel industry (Guttentag, 2010).

VR's influence on psychology is particularly compelling. By generating controlled environments, clinicians have found success in treating conditions ranging from anxiety and phobias to pain management, assisting countless individuals to overcome their fears and traumas (Hoffman et al., 2000; Gorini et al., 2010). Moreover, VR environments have proven especially useful for children with autism, enabling them to practice social scenarios and hone their interpersonal skills (Parsons and Mitchell, 2002).

Parallelly, Emotion Recognition technology has undergone transformative phases, evolving from rudimentary self-reporting methods, which were often influenced by personal biases and cultural nuances (Mesquita and Frijda, 1992; Russell, 1994), to sophisticated deep learning models. The journey has seen the incorporation of biopsychological insights, with physiological signals like heart rate and skin conductivity becoming tools for more precise emotion detection (Picard et al., 2001; Koelstra et al., 2012). The theories of eminent scholars like Ekman gave impetus to recognizing emotions through facial expressions (Ekman and Friesen, 1978; Pantic and Rothkrantz, 2000), while the tonal nuances of voice also emerged as vital emotional indicators (Corive et al., 2001; Scherer, 2003).

The digital age supercharged the field of Emotion Recognition. With machine learning techniques, textual data became a source of sentiment analysis, widely used for gauging public opinions and gathering customer feedback (Liu et al., 2005; Pang and Lee, 2008). The complex architectures of neural networks, such as CNNs and RNNs, provided unprecedented precision in processing multi-modal emotional data (Tzirakis et al., 2017).

Where the worlds of VR and Emotion Recognition converge, we witness a magnified potential for insights. This convergence has enabled researchers to understand and assess "presence" or "immersion" within VR platforms, thereby amplifying the intensity of emotional responses (Riva, 2005; Slater, 2018). Combining VR with traditional self-reporting methodologies has further refined our understanding of emotions in simulated real-world scenarios (MacCann et al., 2003; Felnhofer et al., 2015). The incorporation of biometrics such as EEG and ECG into VR frameworks offers a richer, multi-dimensional view of emotional responses (Mingyu et al., 2006; Ito et al., 2017; Tauscher et al., 2019). Moreover, recent endeavors in merging eye-tracking technology with VR herald new pathways for emotion research, although they come with their sets of challenges (Geraets et al., 2021; Gori et al., 2021; Tabbaa et al., 2021).

In synthesizing the above, it becomes evident that the symbiotic evolution of VR and Emotion Recognition technologies offers a tantalizing promise - a deeper, more nuanced understanding of human emotions, and myriad ways to explore, interact, and harness these emotions in previously uncharted terrains.

2. VR emotion detection system

2.1. Design of the System Architecture

In constructing a VR-based emotion recognition system (see Figure 1), we adopted a comprehensive system architecture and suite of software tools to ensure precise, efficient emotion detection alongside an immersive virtual reality experience. Firstly,

we chose Meta Quest Pro as the principal input device. This decision was rooted in Quest Pro's capabilities, as it not only supports conventional VR inputs but also incorporates natural facial expression recognition sensors. While similar devices, such as the Vive, also feature facial tracking capabilities that can be paired with VIVE Focus 3 to track facial expressions (Hu et al., 2023), Vive's software programming isn't based on Unity. This makes it inconvenient for potential transitions to Apple's forthcoming Vision Pro (Waisberg et al., 2023), which boasts superior facial recognition capabilities. Thus, our choice veered towards Quest Pro. By capturing users' facial movements, Quest Pro can map their real-world expressions to virtual characters within the software. Even though we analyzed facial imagery and eye-tracking capabilities, the use of virtual avatars ensured that no personal information was retained, thereby guaranteeing privacy and security.

For rendering the virtual environment, we opted for Unity as the rendering engine. Its robust compatibility with Meta Quest Pro and extensive SDK support allowed us to rapidly design and deploy realistic virtual scenarios. Opting for Unity not only streamlined the developmental process but also provided a powerful graphics and physics engine (Meta, 2023). This support underpins a myriad of intricate virtual interactions and visual effects. In terms of environmental inputs, our software was designed to allow users to choose their scenes, offering flexibility for other researchers to set the scenarios they wish to study.

For facial emotion analysis, we decided on PyTorch for deep learning. We combined virtual character data for seven fundamental emotions (surprise, sadness, neutral, happiness, disgust, fear, and anger) with the FER2013 database to create our dataset (Giannopoulos et al., 2018). Machine learning training yielded a pth model, which we subsequently converted to the ONNX model format using PyTorch. Unity's Barracuda package enabled us to use the Onnx model for inference and integrate it within the Quest Pro application. This accomplished the machine learning inference component within Unity, providing us with the desired emotion values.

Regarding data storage and management, the identified emotion data–collected at a frequency of 10 times per second–and the visitor's scene information were stored in an Excel spreadsheet. This method ensured timely and orderly data storage, making subsequent analysis and processing more convenient. Our focus on data storage underlines our holistic consideration of the entire system workflow, from emotion capture and analysis to virtual environment rendering and data storage.

2.2. Technical Details of Facial Recognition in Unity

Within the Unity environment (see Figure 2), we initially position the 3D Avatar into the desired scene. Researchers can select their own scenes using panoramic images or FBX models. To ensure real-time capture of the avatar's facial expressions, we placed a camera directly in front of the Avatar. This camera is programmatically set to always track and aim at the Avatar's face, capturing its frontal facial features from the optimal angle.

FIGURE 1
Design of the system architecture.
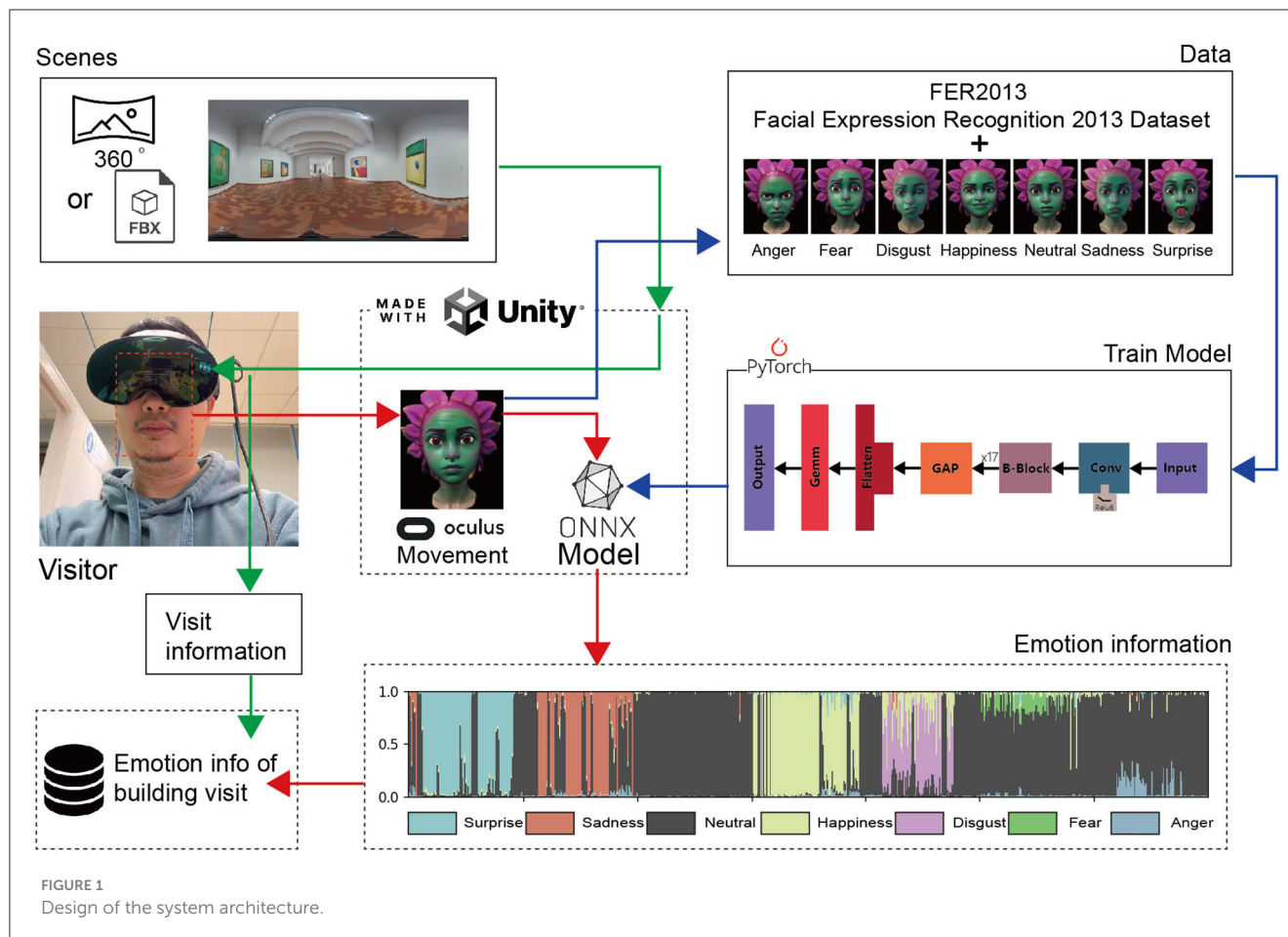
For real-time monitoring of the Avatar's facial expressions, we set the camera to render at a rate of 10 frames per second. This provides us with 10 2D texture images within a second, recording the Avatar's current facial state. Each generated texture is first fed into the retinaface-mnet model (Li et al., 2020). We chose this model because it's optimized specifically for portable devices. After being processed by the retinaface-mnet model, we can quickly and precisely extract the facial part from the original 2D texture image.

The retrieved facial texture then goes through a series of preprocessing steps, ensuring it's ready for further machine learning handling. These steps include cropping the image from its center to produce a 224x224 square image, scaling the image using bilinear interpolation, and converting the RGB image into grayscale. We opted for single-channel processing to avoid inconsistencies across different software and to improve efficiency. The processed R8 single-channel image is then sent to our pre-trained MobileNet V2 model (Sandler et al., 2018). This model performs real-time inferences on the image, identifying facial emotional characteristics accurately. Ultimately, this emotional data is rendered in real-time on a Canvas, allowing researchers to view and analyze the results directly on a monitor.

Our decision to adopt the above workflow was based on two main considerations. First, although we could directly process the raw data from the Quest Pro sensors, this would necessitate handling intricate micro-expression restorations. Given

Meta's expertise in this domain, we decided to utilize their provided SDK as it is more reliable. Secondly, we wanted to ensure our solution is feasible across a variety of devices, such as the upcoming Apple Vision Pro. With potential differences in sensor technologies across devices, we opted to perform facial recognition directly on 2D images, bypassing the handling of raw sensor data. This choice ensures the versatility of our solution, allowing it to easily adapt to various devices.

2.3. Model selection: mobileNet V2

Choosing the right model is a crucial task in our facial expression recognition project. Initially, we tried deep learning models like ResNet 50 and VGG 19. However, due to compatibility issues with Barracuda, we ultimately settled on MobileNet V2 (see Figure 3) as our solution. MobileNet V2 is a lightweight convolutional neural network, especially suitable for running on resource-constrained devices like VR headsets. MobileNet V2 introduces what's called Inverted Residual Blocks, which utilize lightweight depthwise convolutions and pointwise convolutions. These inverted residual structures effectively reduce the computational complexity and the number of parameters of the model. Each inverted residual block in the model ends with a Linear

FIGURE 2
Technical details of facial recognition in unity.

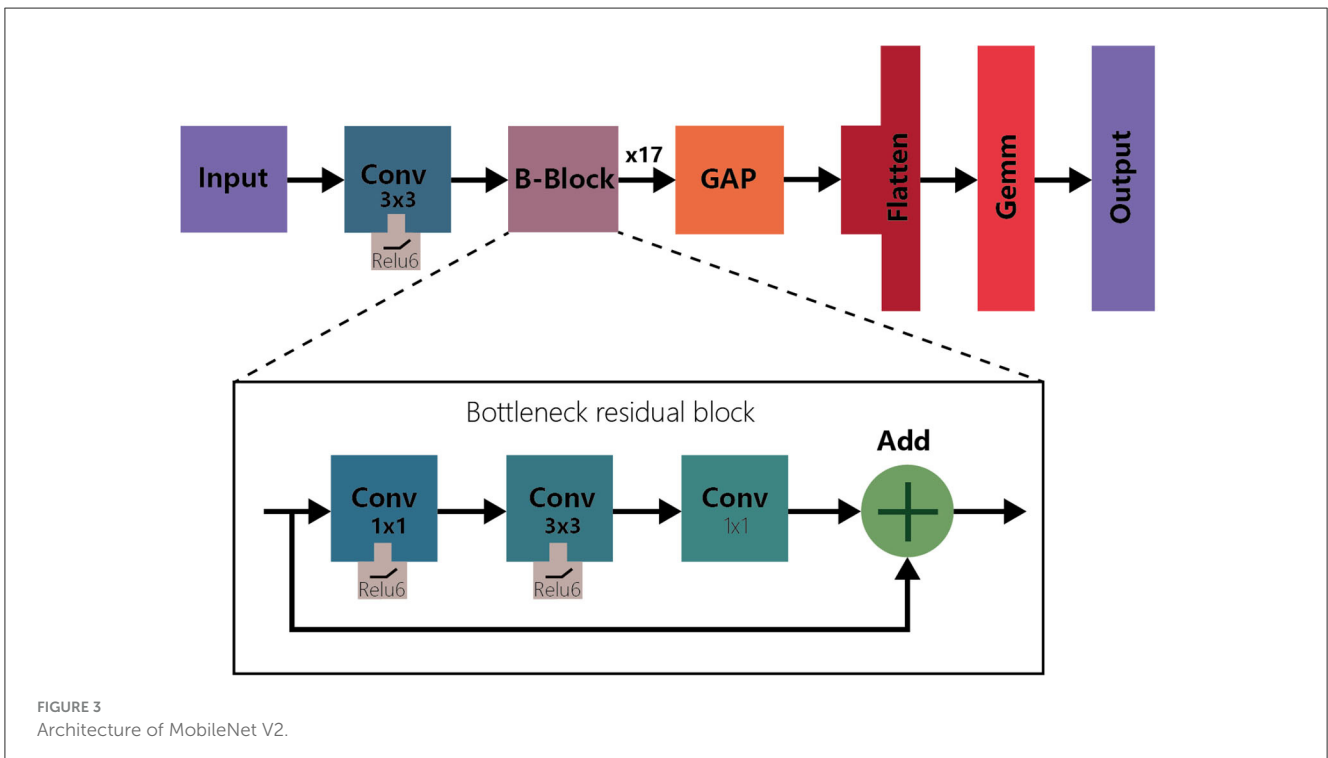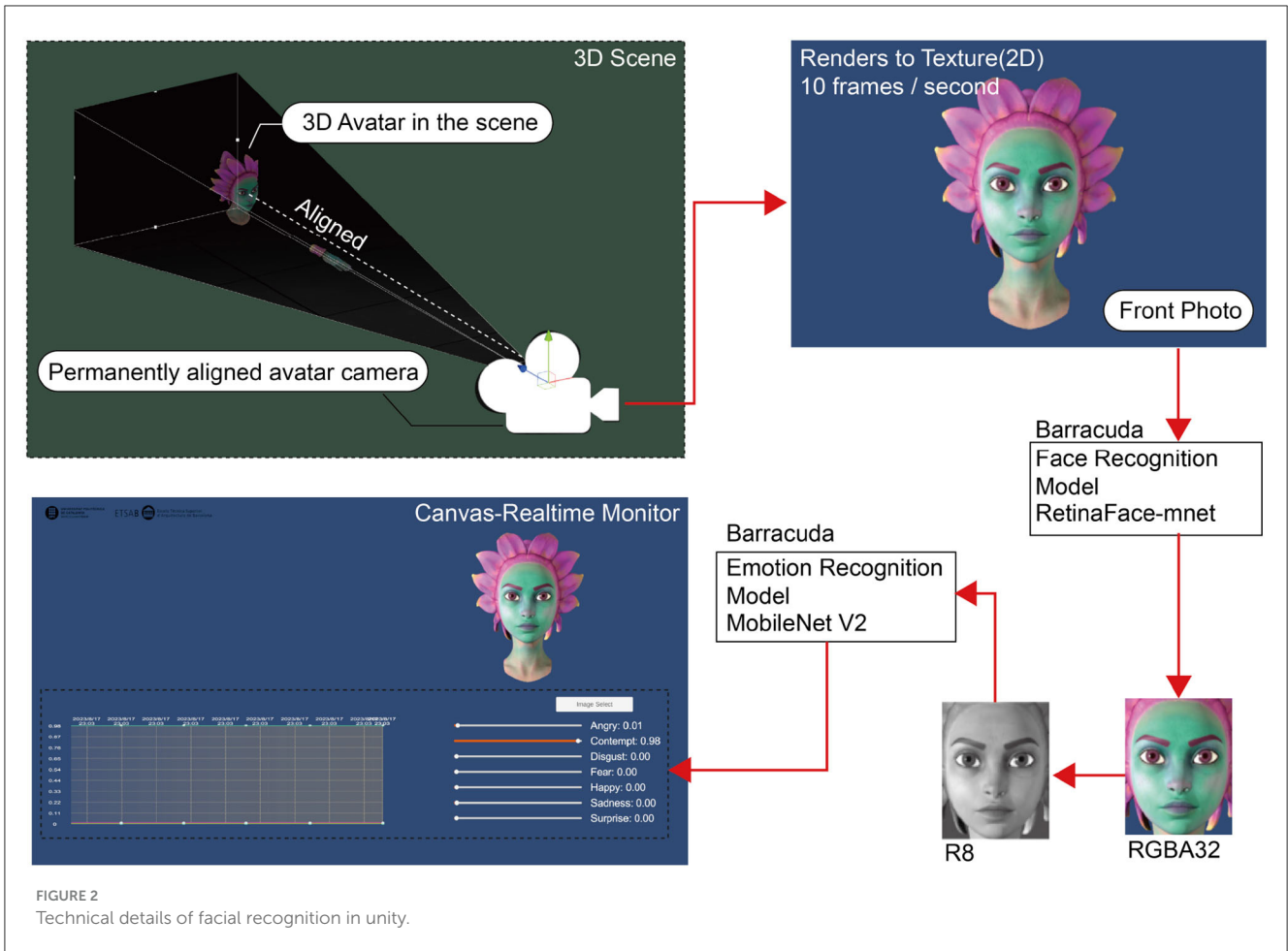**FIGURE 3**
Architecture of MobileNet V2.

TABLE 1 Accuracy percentages of the VR facial expression recognition system. The table shows the system's detection rate for each emotion, in comparison to the emotion that participants were instructed to express.Each column sums up to 100% indicating the recognition rate of predicted emotions, while the sum of values in rows may vary.

Emotion	Emotion requested of participant						
	Surprise	Sadness	Neutral	Happiness	Disgust	Fear	Anger
Surprise	88.66%	0.00%	0.20%	0.14%	0.25%	1.46%	1.02%
Sadness	4.98%	93.11%	0.02%	0.00%	0.87%	0.06%	1.17%
Neutral	2.78%	0.49%	97.52%	3.87%	18.83%	68.41%	70.50%
Happiness	2.72%	5.20%	0.23%	94.15%	8.21%	6.35%	2.34%
Disgust	0.00%	0.15%	0.00%	0.00%	71.24%	4.75%	0.00%
Fear	0.00%	0.00%	0.00%	1.17%	0.00%	16.48%	0.00%

Bottleneck structure, which helps minimize the model size and complexity while retaining essential information. This structural optimization allows MobileNet V2 to achieve optimizations in speed and performance while maintaining a high sensitivity to complex facial expression features. Compared to other larger and more intricate models, MobileNet V2 boasts a smaller footprint and faster computation speed, making it an ideal choice for real-time analysis and tracking of user facial expressions in virtual environments. Furthermore, due to the efficient performance of the MobileNet V2 model, it also allows us to achieve seamless real-time emotional mapping on VR devices, providing users with a more natural and immersive virtual reality experience.

3. Method

3.1. Participants

In our facial expression recognition experiment, we had a total of 15 participants, all of whom came from a university in a major city, with ages ranging from 19 to 42. All participants were university students, covering a variety of academic backgrounds. We did not record the gender information of the participants because the focus of the experiment was on facial expression recognition, and all participants' facial expressions were mapped onto the same avatar. Given that the aim of the experiment was to test the feasibility of VR facial expression recognition, gender was not a primary consideration. Furthermore, not recording gender also adhered to some ethical principles of global health research (Heidari et al., 2016). Participants with myopia were allowed to wear corrective glasses during the experiment. All participants were informed and signed a consent form, ensuring the experiment met ethical standards.

3.2. Experiment procedure

Once the participants understood the study's objectives and gave their consent, they were assisted in equipping and calibrating the VR tools. This ensured both comfort for the participant and accurate detection for the study. The procedure was primarily categorized into two phases: the emotion demonstration and the experiment execution. During the emotion demonstration phase,

participants were introduced to standardized expressions for seven primary emotions: surprise, sadness, neutral, happiness, disgust, fear, and anger, presented in this specific sequence. Participants were then guided to mimic these emotions in the same order.
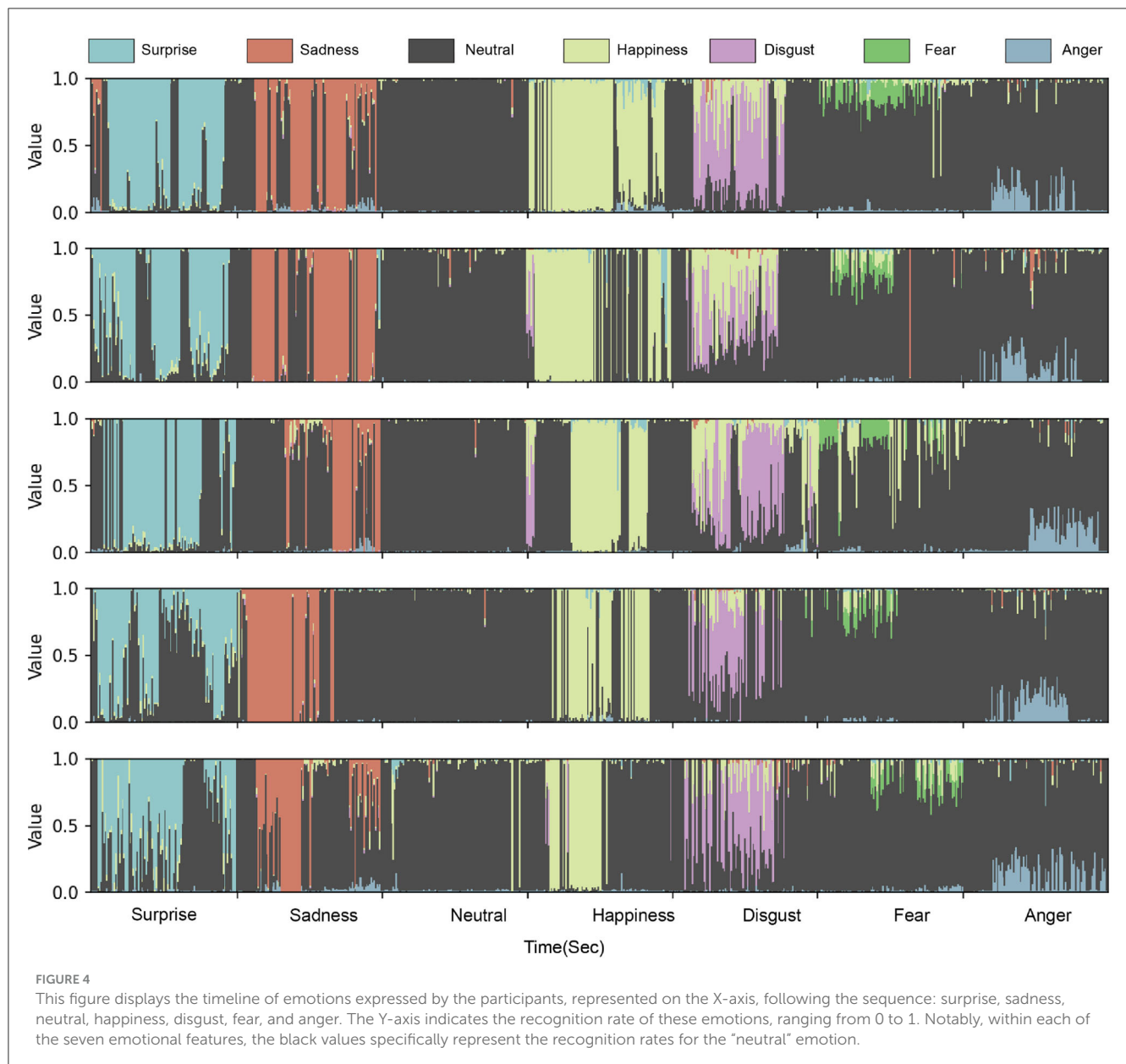
In the execution phase, at every minute interval, participants were prompted to emulate the specified emotions within the VR setting. Concurrently, computer monitors displayed and chronicled their avatar's facial expressions, adhering to the previously mentioned emotion sequence. After the experiment, thorough disinfection of the VR equipment was carried out. This ensured both the hygiene of the next participant and the undisturbed functioning of subsequent sessions. As a precautionary measure against interfering with the equipment's sensors, participants were advised against wearing disposable eye masks.

3.3. Data analysis

For every emotion presented in the sequence – surprise, sadness, neutral, happiness, disgust, fear, and anger – we extracted a 10-second data segment from each participant's feedback. Although participants were instructed to express each emotion within a minute, inconsistencies in their response time and the subtle nuances of facial expressions necessitated that we focus on a 10-second window, deeming it sufficient to capture the core of each emotion.

Rather than directly averaging the values within this time frame, we honed in on peak emotion intensity. Given our system's capability of capturing emotion at 10 frames per second, this 10-second segment provided 100 frames of data. From these, we pinpointed the top 10 emotion values, effectively representing the most intense emotion readings in a one-second span, though not necessarily from a consecutive second. This approach ensured our evaluations were centered on the most pronounced expression of each emotion, minimizing the influence of subdued or transitional values.

Once the top 10 values for each emotion from every participant were aggregated, we derived an average across all 15 participants. This methodology resulted in our confusion matrix, a detailed portrayal of the system's efficacy in discerning each emotion. In Table 1, the horizontal 'Emotion Requested of Participant' row indicates the emotions participants were instructed to show.

FIGURE 4
This figure displays the timeline of emotions expressed by the participants, represented on the X-axis, following the sequence: surprise, sadness, neutral, happiness, disgust, fear, and anger. The Y-axis indicates the recognition rate of these emotions, ranging from 0 to 1. Notably, within each of the seven emotional features, the black values specifically represent the recognition rates for the "neutral" emotion.

The vertical 'Emotion' column represents the system's detected expressions. For instance, the 88.66% under 'Surprise' against the 'Surprise' row means that our system accurately identified the 'Surprise' expression 88.66% of the time when it was presented.

4. Result

In our VR emotion recognition experiment, we utilized a confusion matrix to evaluate the system's ability to identify various emotions(see Table 1). The results indicate that specific emotions such as "Neutral", "Happiness", "Sadness", and "Surprise" achieved relatively high accurate identification rates. Particularly, the identification rate for "Neutral" reached a remarkable 97.52%, suggesting that the model can distinguish this emotion very accurately. The recognition rate for the "Happiness" emotion

reached 94.15%, while "Sadness" and "Surprise" were at 93.11% and 88.66%, respectively. However, for emotions like "Anger" and "Fear", the model's performance was relatively unsatisfactory. Especially for the "Anger" emotion, its accurate rate was only 24.84%, and it was frequently misidentified as "Neutral" with an error rate as high as 70.50%. Similarly, the "Fear" emotion was often misclassified as "Neutral", with an error rate reaching 68.41

Regarding other emotions, such as "Disgust", although its recognition rate was 71.24%, it was sometimes confused with 'Neutral', having an error rate of 18.83%. For the "Happiness" emotion, despite its high accuracy rate, there was still confusion with "Disgust" and "Sadness", with error rates of 8.21% and 5.20% respectively. Considering the results as a whole, it's evident that the model displayed clear strengths and weaknesses in recognizing certain emotions. Particularly concerning is the model's difficulty in identifying the "Anger" and "Fear" emotions. As observed from our

matrix (Table 1), certain rows have notably high recognition rates. This indicates that other true emotions are frequently misclassified into the respective predicted emotion categories. Such insights shed light on areas where the model might require refinement

5. Discussion

In this research, we delved deeply into VR emotional recognition, aiming to uncover the potential capabilities and limitations of the system. Our findings underscore the complexity of emotion recognition and highlight the challenges of current technology and models.

5.1. Model recognition robustness and limitations

The high recognition rates of specific emotions, such as "Neutral", "Happiness", "Sadness", and "Surprise", point out the model's superiority in certain aspects, particularly those emotions that may have clear physiological or facial markers. However, for emotions like "Anger" and "Fear", the model's performance declined (see Figure 4). This could be due to the overlapping physiological and facial features of these emotions with others, combined with a lack of samples for these emotions during model training, leading to its deficiencies. As observed from our matrix (Table 1), certain rows (specifically, 'Neutral') have notably high recognition rates. This indicates that other true emotions are frequently misclassified into the respective predicted emotion categories. Such insights shed light on areas where the model might require refinement.

5.2. Impact of device choice

Results obtained from the Quest Pro device indicate that certain emotions, especially "Anger" and "Fear", might be challenging to capture accurately due to the precision of VR devices. In contrast, the Vision Pro, with its higher precision and additional sensors, could potentially offer more accurate emotion recognition outcomes. Nonetheless, our current research has established a solid foundation for future endeavors.

5.3. Comparing eye-tracking technology

In recent years, although VR eye-tracking technology has been explored in emotion recognition, solely relying on it has inherent limitations. The complexity and diversity of emotions imply that single eye-tracking might be insufficient to capture all emotional nuances (Geraets et al., 2021; Gori et al., 2021; Tabbaa et al., 2021). For instance, different emotions could lead to similar eye behaviors; anger and anxiety might both manifest the same eye fixations, yet they are physiologically and psychologically distinct. Moreover, emotions are not solely reflected in facial expressions; subtle changes in head movements and body trajectories could also reveal nuanced emotional differences. Therefore, a more holistic approach, integrating eye, face, and motion trajectory recognition, is necessary.

5.4. Future directions

We recommend using more advanced devices and delving deeper into the physiological responses of emotions like "Anger" and "Fear", considering the integration of VR device motion trajectories with facial recognition. By combining more sophisticated technologies and larger datasets, there's immense potential and room for improvement in this field.

Overall, this study highlights the challenges and opportunities in the domain of VR emotion recognition. While current models and technologies still have their limitations, with the continual advancement of technology, VR facial emotion recognition is bound to find broad applications in the field of psychological research.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by Ethics Committee of the Universitat Politcnica de Catalunya. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any identifiable images or data included in this article.

Author contributions

ZZ: Conceptualization, Data curation, Methodology, Software, Visualization, Writing—original draft. JF: Methodology, Supervision, Writing—review and editing. LG: Methodology, Supervision, Writing—review and editing.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Corive, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., et al. (2001). Emotion recognition in human-computer interaction. IEEE Signal Process. Mag. 18, 3280. doi: 10.1109/79.911197

Ekman, P., and Friesen, W. V. (1978). Facial action coding system. Environ. *Psychol. Nonverb. Behav.* doi: 10.1037/t27734-000

Felnhofer, A., Kothgassner, O. D., Schmidt, M., Heinzle, A. K., Beutl, L., Hlavacs, H., et al. (2015). Is virtual reality emotionally arousing? investigating five emotion inducing virtual park scenarios. Int. J. Hum. Comput. Stud. 82, 4856. doi: 10.1016/j.ijhcs.2015.05.004

Geraets, C. N., Tuente, S. K., Lestestuiver, B. P., van Beilen, M., Nijman, S. A., Marsman, J. B., et al. (2021). Virtual reality facial emotion recognition in social environments: an eye-tracking study. Inter. Intervent. 25, 100432. doi: 10.1016/j.invent.2021.100432

Giannopoulos, P., Perikos, I., and Hatzilygeroudis, I. (2018). Deep learning approaches for facial emotion recognition: a case study on fer-2013. Smart Innovat, Syst Technol. 85, 116. doi: 10.1007/978-3-319-66790-4_1

Gori, M., Schiatti, L., and Amadeo, M. B. (2021). Masking emotions: face masks impair how we read emotions. *Front. Psychol.* 12:669432. doi: 10.3389/fpsyg.2021.669432

Gorini, A., Pallavicini, F., Algeri, D., Repetto, C., Gaggioli, A., and Riva, G. (2010). Virtual reality in the treatment of generalized anxiety disorders. Stud. Health Technol. Inform. 154, 3943.

Guttentag, D. A. (2010). Virtual reality: applications and implications for tourism. Tour. Managem. 31, 637651. doi: 10.1016/j.tourman.2009.07.003

Heidari, S., Babor, T. F., Castro, P. D., Tort, S., and Curno, M. (2016). Sex and gender equity in research: rationale for the sager guidelines and recommended use. Res. Integ. Peer Rev. 1, 19. doi: 10.1186/s41073-016-0007-6

Hoffman, H. G., Doctor, J. N., Patterson, D. R., Carrougher, G. J., and Furness, T. A. (2000). Virtual reality as an adjunctive pain control during burn wound care in adolescent patients. Pain 85, 305309. doi: 10.1016/S0304-3959(99)00275-4

Hu, R., Hui, Z., Li, Y., and Guan, J. (2023). Research on learning concentration recognition with multi-modal features in virtual reality environments. Sustainability 15, 11606. doi: 10.3390/su151511606

Ito, K., Usuda, S., Yasunaga, K., and Ohkura, M. (2017). Evaluation of "feelings of excitement" caused by a VR interactive system with unknown experience using ECG. Adv. Intell. Syst. Comput. 585, 292302. doi: 10.1007/978-3-319-60495-4_32

Koelstra, S., Mhl, C., Soleymani, M., Lee, J. S., Yazdani, A., Ebrahimi, T., et al. (2012). Deap: A database for emotion analysis; using physiological signals. IEEE Trans. Affect. Comput. 3, 1831. doi: 10.1109/T-AFFC.2011.15

Li, Q., Guo, N., Ye, X., Fan, D., and Tang, Z. (2020). Video face recognition system: Retinaface-mnet-faster and secondary search. Adv. Intell. Syst. Comput. 1364, 625636. doi: 10.1007/978-3-030-73103-8_44

Liu, M., Wang, J., Yan, N., and Yang, Q. (2005). "Development of eeg biofeedback system based on virtual reality environment," in IEEE Engineering in Medicine and Biology Society. Annual Conference. Shanghai: IEEE, 53625364.

MacCann, C., Matthews, G., Zeidner, M., and Roberts, R. D. (2003). Psychological assessment of emotional intelligence: a review of selfâĂŘreport and performanceâĂŘbased testing. Int. J. Organ. Anal. 11:247274. doi: 10.1108/eb028975

Merchant, Z., Goetz, E. T., Cifuentes, L., Keeney-Kennicutt, W., and Davis, T. J. (2014). Effectiveness of virtual reality-based instruction on students' learning outcomes in k-12 and higher education: a meta-analysis. Comput. Educ. 70:2940. doi: 10.1016/j.compedu.2013.07.033

Mesquita, B., and Frijda, N. H. (1992). Cultural variations in emotions: a review. Psychol. Bull. 112, 179204. doi: 10.1037/0033-2909.112.2.179

Meta (2023). *Unity-Movement.* Available online at: https://github.com/oculus-samples/Unity-Movement

Mingyu, L., Jue, W., Nan, Y., and Qin, Y. (2006). "Development of EEG biofeedback system based on virtual reality environment," in *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference* (Piscataway, NJ: IEEE), 5362–5364. doi: 10.1109/IEMBS.2005.1615693

Pang, B., and Lee, L. (2008). Opinion mining and sentiment analysis. Trends Inf. Retr. 2, 1135. doi: 10.1561/9781601981516

Pantic, M., and Rothkrantz, L. M. (2000). Automatic analysis of facial expressions: the state of the art. IEEE Trans. Pattern Anal. Mach. Intell. 22, 14241445. doi: 10.1109/34.895976

Parsons, S., and Mitchell, P. (2002). The potential of virtual reality in social skills training for people with autistic spectrum disorders. J. Intellect. Disabil. Res. 46, 430443. doi: 10.1046/j.1365-2788.2002.00425.x

Picard, R. W., Vyzas, E., and Healey, J. (2001). Toward machine emotional intelligence: Analysis of affective physiological state. IEEE Trans. Pattern Anal. Mach. Intell. 23, 11751191. doi: 10.1109/34.954607

Riva, G. (2005). Virtual reality in psychotherapy: review. CyberPsychol. Behav. 8, 220230. doi: 10.1089/cpb.2005.8.220

Rose, F. D., Brooks, B. M., and Rizzo, A. A. (2005). Virtual reality in brain damage rehabilitation: review. Cyberpsychol. Behav 8, 241262. doi: 10.1089/cpb.2005.8.241

Russell, J. A. (1994). Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. Psychol. Bull. 115, 102141. doi: 10.1037/0033-2909.115.1.102

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., and Chen, L. C. (2018). "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Salt Lake City, UT), 4510–4520.

Scherer, K. R. (2003). Vocal communication of emotion: a review of research paradigms. Speech Commun. 40, 227256. doi: 10.1016/S0167-6393(02)00084-5

Slater, M. (2018). Immersion and the illusion of presence in virtual reality. Br. J. Psychol. 109, 431433. doi: 10.1111/bjop.12305

Tabbaa, L., Searle, R., Bafti, S. M., Hossain, M. M., Intarasirisawat, J., Glancy, M., et al. (2021). Vreed: Virtual reality emotion recognition dataset using eye tracking & physiological measures. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 5, 2. doi: 10.1145/3495002

Tan, C. T., Leong, T. W., Shen, S., Dubravs, C., and Si, C. (2015). "Exploring gameplay experiences on the oculus rift," in *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play*, 253–263. doi: 10.1145/2793107.2793117

Tauscher, J.-P., Schottky, F. W., Grogorick, S., Bittner, P. M., Mustafa, M., and Magnor, M. (2019). Immersive EEG: Evaluating Electroencephalography in Virtual Reality. Osaka, Japan: IEEE, 17941800

Tzirakis, P., Trigeorgis, G., Nicolaou, M. A., Schuller, B. W., and Zafeiriou, S. (2017). End-to-end multimodal emotion recognition using deep neural networks. IEEE J. Sel. Top. Signal Process. 11, 13011309. doi: 10.1109/JSTSP.2017.2764438

Waisberg, E., Ong, J., Masalkhi, M., Zaman, N., Sarker, P., Lee, A. G., et al. (2023). The future of ophthalmology and vision science with the apple vision pro. Eye 2023, 12. doi: 10.1038/s41433-023-02688-5

Wang, X., Love, P. E., Kim, M. J., Park, C. S., Sing, C. P., and Hou, L. (2013). A conceptual framework for integrating building information modeling with augmented reality. Automat. Construct. 34, 3744. doi: 10.1016/j.autcon.2012.10.012