

Central Lancashire Online Knowledge (CLoK)

Title	Evaluating Emotional Responses to Experimental Video Art
Type	Article
URL	https://knowledge.lancashire.ac.uk/id/eprint/56361/
DOI	https://doi.org/10.1109/NTMS65597.2025.11076746
Date	2025
Citation	Tsioutas, Konstantinos, Doumanis, Ioannis, Dalezios, Konstantinos, Siafla, Dora and Tiligadis, Konstantinos (2025) Evaluating Emotional Responses to Experimental Video Art. 2025 12th IFIP International Conference on New Technologies, Mobility and Security (NTMS). pp. 265-271. ISSN 2157-4952
Creators	Tsioutas, Konstantinos, Doumanis, Ioannis, Dalezios, Konstantinos, Siafla, Dora and Tiligadis, Konstantinos

It is advisable to refer to the publisher's version if you intend to cite from the work.
<https://doi.org/10.1109/NTMS65597.2025.11076746>

For information about Research at UCLan please go to <http://www.uclan.ac.uk/research/>

All outputs in CLoK are protected by Intellectual Property Rights law, including Copyright law. Copyright, IPR and Moral Rights for the works on this site are retained by the individual authors and/or other copyright owners. Terms and conditions for use of this material are defined in the <http://clock.uclan.ac.uk/policies/>

Evaluating Emotional Responses to Experimental Video Art

Konstantinos Tsioutas,^{*} Ioannis Doumanis,[†] Konstantinos Dalezios,[‡] Dora Siafla,[‡] Konstantinos Tiligadis,[‡]

^{*} Mobile Multimedia Laboratory, Department of Informatics,

School of Information Sciences and Technology, Athens University of Economics and Business, Greece

[†] School of Psychology and Computer Science University of Central Lancashire Preston United Kingdom

[‡] Audio and Visual Arts Department, Ionian University, Corfu, Greece

Abstract—Video is a medium that conveys information capable of reflecting multiple emotions to the viewers. Experimental video, a new contemporary art form, is a complex stimulus that can induce diverse emotions in its audience. This study exposed 22 individuals to two experimental videos, each comprising multiple sequences of experimental images and sounds. We asked participants to record their facial expressions while watching the videos using web cameras. We analysed the recorded footage using DeepFace and Facial Emotion Recognition (FER) algorithms to estimate the seven universal emotions. The Deepface algorithm demonstrated higher sensitivity than FER. We performed time analysis to detect possible visual events that might trigger emotional alterations and Principal Component Analysis (PCA) to search for similarities in participants’ emotional profiles. The sudden voice of a child and the sudden appearance of a human figure, possibly triggered the increase of happiness as the time analysis exposed. PCA revealed high variability in neutral emotion across clusters, with notable increases in anger in two of the three clusters. The first cluster showed the most variability in sadness, while happiness slightly increased only in this cluster.

Index Terms—Experimental video art, Emotions, Facial Expressions

I. INTRODUCTION

Experimental video, or video art, is a genre of contemporary art that emphasises experimentation with elements such as movie scenes, moving images, audio, text, and various visual and auditory content [1]. Multiple techniques are employed in creating such works, including timeline montage, visual and audio effects, sound design, mixing and mastering, animation, and more [2]. These projects are often showcased in art galleries and exhibitions, typically displayed on digital screens or projected onto walls [3]. Video art is a relatively recent visual art medium, with many artists producing works featured in thousands of exhibitions worldwide. As a form of contemporary art, it seeks to provoke questions and evoke emotions in viewers, as all forms of art aim to do [3]. Videos, in general, act as mediums that elicit various emotional responses from viewers [4]. For instance, a humorous video may evoke joy, while a sombre one might induce sadness. However, experimental video is a particularly complex art form, as artists often employ subtle and unconventional methods to express their thoughts and convey messages to

the audience [5]. In video art, audiovisual materials become intricate stimuli, eliciting unique and diverse emotional reactions from each viewer [6]. Given the complexity and depth of emotions associated with experimental video art, machine learning techniques, such as facial analysis, are increasingly utilised to estimate and interpret these emotional responses [7]. Standardised models categorise emotions into seven universal types: happiness, anger, sadness, neutrality, disgust, surprise, and fear [8]. This paper investigates the emotions induced in 22 viewers who watched two specific experimental videos created by our team. The contributions of this paper are as follows:

- 1) An investigation of the emotional responses in experimental video art.
- 2) A comparison of the DeepFace and Facial Emotion Recognition algorithms.
- 3) A principal component analysis of the emotional responses of the participants to the experimental videos.

The rest of the paper is organised as follows: In Section II, we discuss the concepts of video art and its emotional impact and reviews related studies. Section III details the techniques and content of the two videos. Section IV describes the methodology used, and Section V presents the results. Finally, Section VI analyses the findings.

II. BACKGROUND AND RELATED WORK

Experimental video art engages visual and auditory senses in ways that are often intense, abstract, and non-linear [2]. Such multi-modal stimuli can evoke strong emotional responses [9]. Unlike traditional narratives, video art invites diverse interpretations from its audience [10]. The use of materials such as images, colours, animations, and soundscapes can evoke a range of emotions, including anxiety, wonder, disorientation [11], as well as fear, desire, and nostalgia [12].

Immersive video art can elicit a sense of presence and participation, making viewers active participants rather than passive observers [13]. The immersive quality and video art’s unfamiliarity and abstract nature often challenge viewers’ perceptions and norms, leading to powerful emotional reactions. Such deviations from conventional expectations

disrupt viewers' anticipations, creating complex emotional experiences [14].

Video art's temporal visual and auditory elements, such as slow-motion sequences or rapid changes in visual content, can influence emotions ranging from melancholy to stress [15]. The abstract nature of video art frequently prompts personal reflections, raising existential questions and considerations of identity [16].

The auditory elements in video art—ranging from noise and ambient soundscapes to abstract audio forms—can independently influence mood and emotion, separate from visual cues [17]. The presence or absence of sound (e.g., silence versus music or soundscapes) significantly affects emotions: silence often induces tension, while music and soundscapes can evoke nostalgia or sadness [18].

Studies have demonstrated that automated facial expression recognition systems are reliable for detecting clear, universal emotions (e.g., happiness, anger) [19] [20]. However, these systems generally struggle to accurately detect subtle or ambiguous emotions [20]. For instance, emotions like contempt or boredom often exhibit minimal facial changes, and mixed emotions (e.g., a combination of happiness and sadness) may involve overlapping facial features. Since experimental video art can induce such complex emotions, additional biometric methods such as electroencephalography (EEG) and electromyography (EMG) are necessary. Several studies have utilised multimodal biometrics to measure viewers' emotions watching videos [21] [22]. Nevertheless, these studies typically require controlled laboratory conditions, which hinder capturing authentic engagement and contextual responses inherent in the viewer experience with experimental video art. Although wearable and portable biometric technologies are expected to facilitate running such studies in real-world settings eventually, automated facial expression recognition remains the cheapest and most reliable method available today.

III. EXPERIMENTAL VIDEO COMPOSITION

In this section, we describe the techniques for the composition of the two videos.

A. Techniques used in Video 1

In Video 1, which is available for watching here (link for video 1), we explore the multichannel video technique and sound. Figure 1 shows a snapshot of this video. We composed the film using six video recordings that capture a sea view from the vantage point of a vessel. Starting from a single recording, we edited the footage by cutting, looping, and flipping segments horizontally or vertically. We overlaid a frame from empty scanned film stock onto the footage to enhance the visual aesthetic.

We applied additional manipulations by using a mask to invert the colours of the underlying footage and introducing a sepia tone to evoke a sense of nostalgia. We also used image morphing filters to subtly alter the main video recording,

creating a unique portrayal of the sea that engages viewers visually and emotionally.

We paid significant attention to sound design, combining the visuals with characteristic recordings from familiar sources. By juxtaposing train station sounds with serene imagery of a vessel adrift at sea, we created a compelling contrast that enhances the immersive quality of the experience.

To further enrich the nostalgic visuals, we added a piano music track. We seamlessly sampled and mixed this music with a moving train and ambient passenger noises, adding emotional depth to the composition. Midway through the audio, we included the voice of a child recounting everyday moments from the past. This poignant auditory layer guides viewers through an imaginary voyage, leading them to the film's conclusion.

TABLE I
SUMMARY OF TECHNIQUES USED FOR THE TWO VIDEOS PRODUCTION

	Video 1	Video 2
Sound	Yes	No / Yes
Montage	No	Yes
Video Channels	Multiple	Single
Duration	7min	4min
Video Overlays	Yes	No
Color Filters	Yes	Yes
Human Forms	No	Yes
Aspect Ratio	16:9	16:9

B. Techniques used in Video 2

In Video 2, which is available for watching here (link for video 2), we created a single-channel film using fragments of footage we shot during a boat voyage in the COVID-19 pandemic. A snapshot of this video is provided in Figure 2. We recorded moving images inside and outside the vessel, primarily in a vertical aspect ratio. We chose this format deliberately to challenge conventional norms in film and video. By doing this, we treated the recordings as visual elements for composing an image rather than simply documenting events.

We applied extensive colour channel filters to the footage and juxtaposed these filtered clips with videos featuring human forms. Unlike Video 1, we included human elements in Video 2, int added a new dynamic to the composition. We made two versions of this video: one with sound design and music and another without audio.

By placing human figures, we experimented with the interaction of the human element with other film components through editing. This approach has been applied in the experiments of Lev Kuleshov conducted in the 1920s. Kuleshov showed that viewers could derive deeper meaning from a sequence of shots when human figures are present, compared to isolated ones, highlighting how context and juxtaposition shape visual storytelling.

IV. METHODOLOGY

We uploaded the videos to YouTube and shared the links with 22 participants. We divided the participants into four

groups, as shown in Table II. Group 1 included five students enrolled in a master's program in the arts. Group 2 consisted of six young males with no prior experience in experimental video art. Group 3 comprised six participants from another master's art class. Finally, Group 4 included five females with limited experience in video art. Group 1 and 2 watched the second video with sound while the Groups 3 and 4 watched the same video without sound.

We asked each participant to watch the videos and simultaneously record their facial expressions using their webcams. Since participants viewed the videos in uncontrolled environments, the experimental conditions varied. After collecting the recorded facial expressions, we analyzed the videos using the DeepFace algorithm [23] and the Facial Expression Recognition (FER) algorithm [24]. We conducted an agreement analysis to compare the results of the two algorithms.

As a next step, we applied temporal analysis, based on the results from DeepFace, to reveal possible time spots that triggered certain emotional responses to the viewers.

Finally, we performed principal component analysis (PCA) and clustering to explore the emotional profiles of the 22 participants.

TABLE II
GROUPS, VIDEOS AND EXPERIENCE

Group	User	Video 1 sound	Video 2 sound	Art experience
1	1-5	yes	yes	Extended
2	6-11	yes	yes	low
3	12-17	yes	no	Extended
4	18-22	yes	no	low

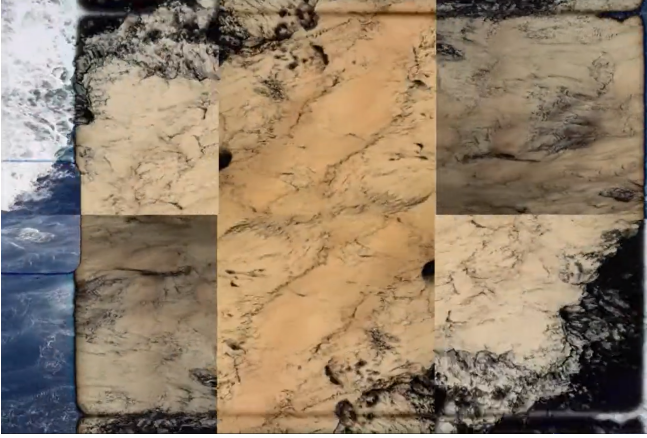


Fig. 1. A snapshot from the video 1 with multiple channels of the same content.

V. EVALUATION RESULTS

A. Deepface vs FER algorithms.

The DeepFace algorithm, built on Keras and TensorFlow, analyses facial expressions in images or videos to extract emotional data. It starts by detecting and isolating the face



Fig. 2. A snapshot from the video 2 which contains human figures.

from the background and non-face regions, a process known as face alignment. During this step, the face recognition module outputs a bounding box (a 4-element vector) for the face and a 10-element vector that localises five key facial landmarks: two for the eyes, two for the mouth, and one for the nose. In the final stage, DeepFace classifies the detected face into one of the basic emotion categories using a fully connected convolutional neural network (CNN) with three convolutional layers to extract the necessary features for classification.

The FER (Facial Expression Recognition) algorithm uses Multi-task Cascaded Convolutional Networks (MTCNN) [25] for face detection and alignment, which are essential for accurate emotion recognition. FER begins by detecting and isolating the face from the background by identifying facial landmarks, including the eyes, nose, and mouth. Once the face is aligned, it analyses the facial features and expressions using a CNN. The extracted features are then classified into one of the basic emotion categories: angry, disgust, fear, happy, sad, surprise, or neutral. This multi-step process allows FER to detect emotions from facial expressions in images and video frames efficiently.

The analysis of the datasets extracted from both algorithms revealed that DeepFace performed with more sensitivity compared to FER, providing estimations with up to five decimal places for each emotion per frame. In contrast, FER was less sensitive, frequently failing to detect a face and outputting zeros for many frames. A zero result indicates that the algorithm failed to identify a face in the frame [26].

We conducted a percentage agreement analysis to test the agreement between the two algorithms. We calculated the percentage agreement using a straightforward formula [27]: the ratio of the number of agreements to the total number of instances multiplied by 100. The results are shown in table III. We acknowledge that a comprehensive comparison of the two algorithms is a more complex process; however, it falls outside the scope of this study.

Overall, neither algorithm performed exceptionally well across all emotions. However, DeepFace was a better choice

for this study. First, the algorithm is more consistent in detecting faces and classifying emotions, as shown in its more precise estimations (to five decimal places). For these reasons, we based the analysis presented in this paper on the results from DeepFace.

TABLE III
AGREEMENT PER EMOTION BETWEEN DEEPFACE AND FER

Emotion	Agreement %
Angry	11.92
Disgust	90.26
Fear	7.07
Happy	30.71
Sad	6.03
Surprise	67.56
Neutral	8.83

B. Temporal analysis

In order to detect possible time spots of the two videos that triggered certain emotional responses to the participants, we performed a time analysis of the results from DeepFace. Considering that among the participants shown in table II, groups 3 and 4 watched the second video without sound, we applied this grouping to the results from DeepFace algorithm. We then plotted the average values of emotions per frame and per group of the 22 participants. An interesting observation can be found in Figure 4 where the average value of happiness per frame for groups 3 and 4 regarding video 1 is shown. The increase of the average value of happiness in around frame 9000, is possibly triggered by the voice of a child talking suddenly at that time of video 1. The same observation was found, but with lower increase of mean value for groups 1 and 2 who watched the same video. A similar event can be observed in Figure 5 where around frame 1400, there is a discrete increase of average value of happiness. This increase was possibly triggered by the sudden appearance of a human figure wearing a COVID protection mask similar to Figure 2. The same human figure exists at a later time spot of video 2 but it did not trigger any detectable alterations of the emotional responses.

The time analysis of the average scores exposed that Groups 3 and 4 that watched Video 2 without sound, reacted with lower sadness and higher levels of neutrality compared to Groups 1 and 2. This evidence may be an indication that the presence of sound may trigger negative or positive emotions compared to silence.

Finally percentage of users per dominant emotions and per video are shown in figure 3. Sharp alterations are observed for happiness and neutrality between the two videos suggesting a general decrease in happiness in Video 2 compared to Video 1.

C. Principal Component Analysis

We applied Principal Component Analysis (PCA) to identify similarities between the time series signals. PCA reduces

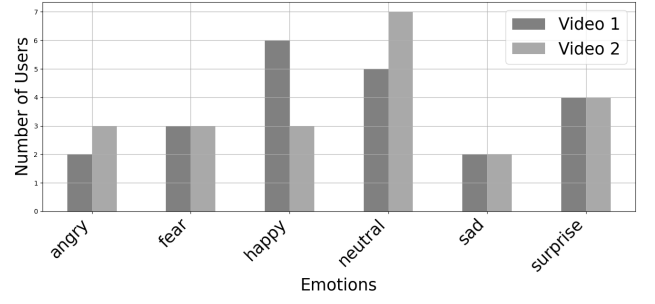


Fig. 3. Dominant emotions per video for 22 users.

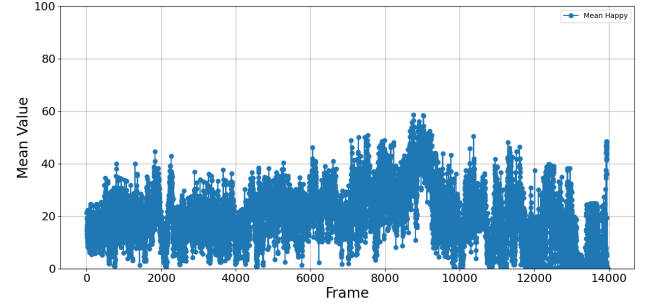


Fig. 4. Average value of happiness for groups 3 and 4 for video 1

the dimensionality of the data, transforming it into a lower-dimensional space to simplify complexity and reveal the most significant hidden features. To determine the optimal number of clusters, we used the elbow method [28], which identified three clusters.

Using this result, we ran PCA and generated the graph shown in Figure 6. Based on the three clusters computed from the results of Video 1, we plotted users' emotional profiles computed from Video 2 to compare possible changes in the two dimension space. The result is shown in 7.

We provide a possible interpretation of the position of each point in the plots in Table IV. From the first graph it is observed that participants in Cluster 3 (black points) experienced higher arousal and more positive emotions while watching Video 1. In contrast, Cluster 1 (light grey points) exhibited the opposite trend, with lower arousal and less positive emotions for Video 1. Finally, Cluster 2 (grey points in the middle of the plot) displayed neutral responses.

The second graph, as mentioned above, applies the grouping computed from Video 1 to generate the profiles for Video 2. Using this method, we can compare potential alterations in emotional responses between the two videos. As observed, most users did not exhibit detectable changes in their responses. In other words, the comparison of the two graphs suggests that Video 2 did not significantly affect the participants' emotional profiles compared to Video 1. The only clear change observed is related to users 12 and 9, whose profiles shifted towards the center of the graph, suggesting that their responses in Video 2 were more neutral.

Table V lists the user ids for each video and cluster. The

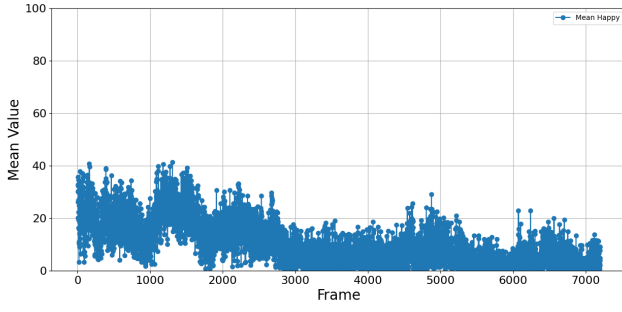


Fig. 5. Average value of happiness for groups 1 and 2 regarding video 2

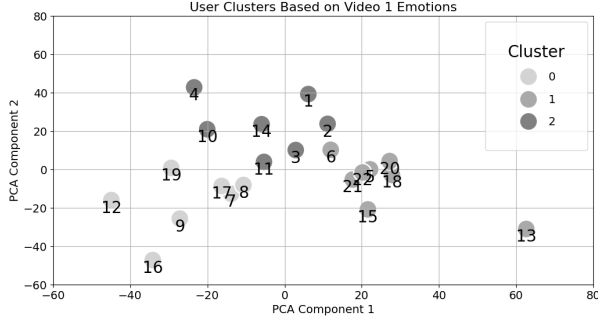


Fig. 6. PCA: 3 clusters of users for video 1 (DeepFace)

three clusters do not generally correspond to the initial four groups described in Section IV and presented in Table II. Additionally, users from different initial groups exhibit similar emotional profiles. For example, users in initial Group 2 (users 6, 7, 8, 9, 10, 11), consisting of six male undergraduates with minimal experience in experimental video art, displayed emotional profiles comparable to those of groups with extensive video art experience (cluster 2, video 2). This observation suggests that experience in video art may not influence the participants' emotional responses.

TABLE IV
POSSIBLE INTERPRETATION OF THE POSITION OF EACH SPOT IN THE PLOT

Position	Possible Interpretation
Middle area	Neutral Typical
Top left area	High arousal Low positivity
Top right area	High arousal High positivity
Bottom left area	Low arousal Low positivity
Bottom Right area	Low arousal High positivity

Based on the three clusters identified through PCA analysis, we plotted the variability of emotions for each cluster, as shown in Figure 8. We omitted the emotions of disgust and surprise because they were rarely detected.

The figure reveals that the neutral emotion exhibits high

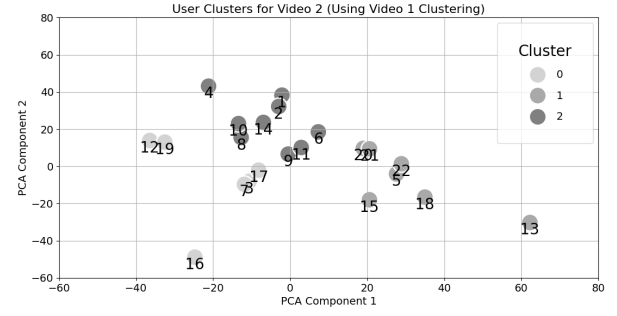


Fig. 7. PCA: placing users' profiles for video 2 in clusters computed for video 1(DeepFace)

variability across all clusters and both videos. In Cluster 2, anger increases by almost 15% compared to the two videos, while happiness in the same cluster decreases significantly. Cluster 3 shows higher levels of anger compared to the other clusters and lower levels of fear relative to the other subjects. Additionally, Cluster 1 displays a slight increase in happiness between the two videos and exhibits the greatest variability and higher median values in sadness compared to the other clusters.

TABLE V
USERS PER CLUSTER PER VIDEO

Cluster	Video 1
0	Users 7, 8, 9, 12, 16, 17, 19
1	Users 5, 6, 13, 15, 18, 20, 21, 22
2	Users 1, 2, 3, 4, 10, 11, 14
Cluster	Video 2
0	Users 3, 7, 12, 16, 17, 19
1	Users 5, 13, 15, 18, 20, 21, 22
2	Users 1, 2, 4, 6, 8, 9, 10, 11, 14

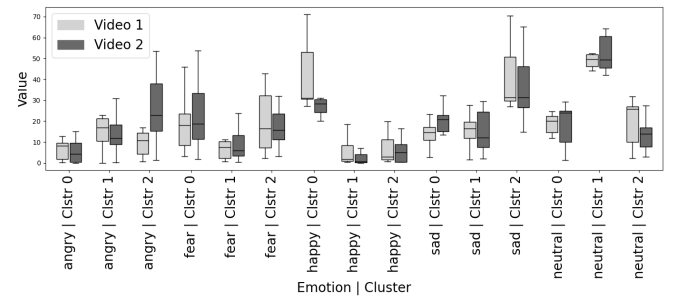


Fig. 8. Variance of emotions per video for the 3 clusters computed by the PCA analysis.

D. Statistical analysis

To examine the nature of the data and assess whether the assumptions for an ANOVA analysis were met, we performed the Shapiro-Wilk test. The results indicated that ANOVA could not be applied, as the data did not exhibit homogeneity of variance and did not follow a normal distribution. Overall,

the statistical tests showed that the two videos did not significantly affect the participants' emotional responses, whereas substantial emotional differences were observed among users.

VI. DISCUSSION

As shown in Figure 3 participants' facial expressions were generally neutral when watching videos and did not evoke other specific emotions. The absence of statistical findings suggests that the embedded emotional cue may not have been powerful or explicit enough to elicit distinct emotional patterns across all viewers. However, the PCA analysis revealed three distinct clusters of emotional profiles. This may be due to the nature of the experimental art videos, which involved ambiguity and abstraction, likely leading to diverse participant interpretations and emotional responses [29]. Therefore, these clusters are more likely driven by individual interpretation styles, aligning with the literature that ambiguity and abstraction in experimental art create space for subjective emotional reactions.

Additionally, we found that DeepFace is better suited for analysing emotions in experimental video art. As also reported in [30] an analysis based on facial expressions is not a conclusive procedure but it can strengthen the results of such studies. The abstract and ambiguous elements in the videos can elicit a wide range of subtle emotional responses. DeepFace's ability to calculate emotion scores to five decimal places allows for a more granular tracking of emotions than the FER algorithm. This precision facilitates detailed tracking of subtle emotional variation on a frame-by-frame basis, which is valuable in analysing the detailed emotional responses that experimental video art can evoke. Finally events like a sudden voice of a child and the sudden appearance of a human figure may trigger emotional alterations as the time analysis revealed.

VII. CONCLUSIONS AND FUTURE WORK

We experimented with 22 participants who watched two experimental art videos. We analysed the recorded videos using the DeepFace algorithm to estimate seven emotions and processed the results using statistical tools.

Experimental video art demands granular emotional tracking over time to gain deeper insights into participants' emotional engagement with the artwork throughout their experience. To address this, we propose combining DeepFace, FER, and a standard Transformer Encoder [31] to capture emotional transitions dynamically. This integrated pipeline aims to enhance emotion detection accuracy and provide insights beyond what is achievable with existing open-source tools. Specifically, the pipeline leverages FER for pre-processing (utilizing MTCNN for face detection and alignment) [32], DeepFace (CNN) for emotion classification [33], and a standard Transformer Encoder [31] to model emotional shifts over time. We are currently evaluating the tool's performance by analyzing the emotional responses of 22 participants.

REFERENCES

- [1] M. Yur, "Art experiment asan implementation of innovations." *Artistic Culture. Topical Issues*, vol. 18, no. 1, 2022.
- [2] A. L. Rees, *A history of experimental film and video*. Bloomsbury Publishing, 2019.
- [3] H. Kukka, J. Ylipulli, J. Goncalves, T. Ojala, M. Kukka, and M. Syrjälä, "Creator-centric study of digital art exhibitions on interactive public displays," in *Proceedings of the 16th International Conference on Mobile and Ubiquitous Multimedia*, 2017, pp. 37–48.
- [4] M. H. Järvenpää and K. O. Korhonen, "System and method for measuring and analyzing audience reactions to video," May 15 2014, uS Patent App. 13/674,683.
- [5] T. Barker, "Experiments with time: the technical image in video art, new media and the digital humanities," *Visual Communication*, vol. 16, no. 4, pp. 375–394, 2017. [Online]. Available: <https://doi.org/10.1177/1470357217702360>
- [6] R. Verleur, A. Heuvelman, and P. W. Verhagen, "Trigger videos on the web: Impact of audiovisual design," *British Journal of Educational Technology*, vol. 42, no. 4, pp. 573–582, 2011.
- [7] A. Sharma, V. Bajaj, and J. Arora, "Machine learning techniques for real-time emotion detection from facial expressions," in *2023 2nd Edition of IEEE Delhi Section Flagship Conference (DELCON)*, 2023, pp. 1–6.
- [8] P. Ekman, E. R. Sorenson, and W. V. Friesen, "Pan-cultural elements in facial displays of emotion," *Science*, vol. 164, no. 3875, pp. 86–88, 1969.
- [9] Ž. Eglīte, "The role of senses, emotions and the principles of the experience economy in the creative industries," *Culture Crossroads*, vol. 19, no. 1, pp. 230–242, 2021.
- [10] V. Berry and V. Shelton, "Watching music: Interpretations of visual music performance," *Journal of Communication Inquiry*, vol. 23, pp. 132–151, 04 1999.
- [11] B. Pons, I. Alvarez, and J. Thompson, "Aesthetic quality experiences (aqe) as a means of improving reading and writing: A case study," *Universal Journal of Educational Research*, vol. 8, no. 11, pp. 5376–5387, 2020.
- [12] C. Hales, "Weird and wonderful: how experimental film narratives can inform interactive digital narratives," in *Interactive Storytelling: 13th International Conference on Interactive Digital Storytelling, ICIDS 2020, Bournemouth, UK, November 3–6, 2020, Proceedings 13*. Springer, 2020, pp. 149–163.
- [13] A. Felnhöfer, O. Kothgassner *et al.*, "Presence and immersion: a tale of two cities," *Digital Psychology*, vol. 3, no. 2, pp. 3–6, 2022.
- [14] A. Daneli, P. Koulouri, C. Tzima, I. Chatzichristodoulou, and L.-V. Andreou, "Introducing video art in stem education alleviates negative academic emotion," in *2023 IEEE Global Engineering Education Conference (EDUCON)*. IEEE, 2023, pp. 1–5.
- [15] N. Marhamati and S. C. Creston, "Visual response to emotional state of user interaction," *arXiv preprint arXiv:2303.17608*, 2023.
- [16] S. Simmons, "Living the questions: existential intelligence in the context of holistic art education," 2006. [Online]. Available: <https://api.semanticscholar.org/CorpusID:190314278>
- [17] B. Riyanto, N. N. Alimin *et al.*, "Video art as environmental communication media for the problem of plastic waste in solo," in *IOP Conference Series: Earth and Environmental Science*, vol. 1114. IOP Publishing, 2022, p. 012090.
- [18] N. Bachman, Y. Palgi, and E. Bodner, "Emotion regulation through music and mindfulness are associated with positive solitude differently at the second half of life," *International Journal of Behavioral Development*, vol. 46, no. 6, pp. 520–527, 2022.
- [19] L. Kulke, D. Feyerabend, and A. Schacht, "A comparison of the affectiva imotions facial expression analysis software with emg for identifying facial expressions of emotion," *Frontiers in Psychology*, vol. 11, 2020. [Online]. Available: <https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2020.00329>
- [20] T. Küntzler, T. T. A. Höfling, and G. W. Alpers, "Automatic facial expression recognition in standardized and non-standardized emotional expressions," *Frontiers in Psychology*, vol. 12, 2021. [Online]. Available: <https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2021.627561>
- [21] T. Zhang, A. El Ali, C. Wang, A. Hanjalic, and P. Cesar, "Cornnet: Fine-grained emotion recognition for video watching using wearable

- physiological sensors,” *Sensors*, vol. 21, no. 1, 2021. [Online]. Available: <https://www.mdpi.com/1424-8220/21/1/52>
- [22] Y. Liu, T. Gedeon, S. Caldwell, S. Lin, and Z. Jin, “Emotion recognition through observer’s physiological signals,” 2020. [Online]. Available: <https://arxiv.org/abs/2002.08034>
- [23] S. Li and W. Deng, “Deep facial expression recognition: A survey,” *IEEE transactions on affective computing*, vol. 13, no. 3, pp. 1195–1215, 2020.
- [24] J. Shenk, “Fer: Facial expression recognition library,” <https://github.com/justinshenk/fer>, n.d., accessed: 1/10/2024.
- [25] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, “Joint face detection and alignment using multitask cascaded convolutional networks,” *IEEE Signal Processing Letters*, vol. 23, no. 10, pp. 1499–1503, 2016.
- [26] E. Owusu, E. Gavua, and Y.-Z. Zhan, “Facial expression recognition-a comprehensive review 1,” *International Journal of Technology and Management Research*, vol. 2, pp. 29–46, 01 2015.
- [27] J. Cohen, “A coefficient of agreement for nominal scales,” *Educational and Psychological Measurement*, vol. 20, no. 1, pp. 37–46, 1960. [Online]. Available: <https://doi.org/10.1177/001316446002000104>
- [28] M. A. Syakur, B. K. Khotimah, E. M. S. Rochman, and B. D. Satoto, “Integration k-means clustering method and elbow method for identification of the best customer profile cluster,” *IOP Conference Series: Materials Science and Engineering*, vol. 336, no. 1, p. 012017, apr 2018. [Online]. Available: <https://dx.doi.org/10.1088/1757-899X/336/1/012017>
- [29] C. Muth, V. M. Hesslinger, and C.-C. Carbon, “The appeal of challenge in the perception of art: How ambiguity, solvability of ambiguity, and the opportunity for insight affect appreciation,” *Psychology of Aesthetics, Creativity, and the Arts*, vol. 9, no. 3, p. 206, 2015.
- [30] Anonymous, “Multimodal assessment of network music performance,” in *Companion Publication of the 2021 International Conference on Multimodal Interaction*, ser. ICMI ’21 Companion. New York, NY, USA: Association for Computing Machinery, 2021, p. 284–290. [Online]. Available: <https://doi.org/10.1145/3461615.3485418>
- [31] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, “Attention is all you need,” 2023. [Online]. Available: <https://arxiv.org/abs/1706.03762>
- [32] J. Xiang and G. Zhu, “Joint face detection and facial expression recognition with mtcnn,” in *2017 4th International Conference on Information Science and Control Engineering (ICISCE)*, 2017, pp. 424–427.
- [33] H. Boughanem, H. Ghazouani, and W. Barhoumi, “Facial emotion recognition in-the-wild using deep neural networks: A comprehensive review,” *SN Computer Science*, vol. 5, no. 1, p. 96, 2023.