Emotion Recognition of Museum Visitors during Art Appreciation Based on Facial and Behavioral Features

Yohei Torigaki[1] and Yuki Matsuda[1,2][0000-0002-3135-4915]

[1] Okayama University, Okayama 700-0082, Japan
yohei.torigaki@cocolab.jp, yukimat@okayama-u.ac.jp
[2] RIKEN Center for Advanced Intelligence Project (AIP), Tokyo 103-0027 Japan

Abstract. In recent years, museums have been increasingly expected to facilitate communication in order to meet the intellectual needs of visitors. Although museum educators—experts in learning within museums and art galleries—have been employed to fulfill this role, there remains a persistent shortage of personnel. While systems such as audio guides are anticipated to serve as substitutes, a significant gap still exists between these systems and human guidance. Therefore, this study proposes a method to estimate the emotional state of museum visitors, aiming to provide dynamic guidance tailored to individual interests and concerns. To establish a method for emotional state estimation, we conducted a data collection experiment using a newly developed guidance system equipped with sensing capabilities at the Ohara Museum of Art, involving 30 participants. Based on the collected data of facial expressions and gestures, we constructed and evaluated an emotional state estimation model. As a result, the binary emotion classification model for positive and negative emotions achieved an accuracy of 64%. In contrast, regression models for satisfaction, comprehension, and perplexity did not yield effective results; however, some degree of correlation was observed, providing insights for future improvements.

Keywords: Emotion Recognition, Wearable Devices, Art Appreciation, Affective Computing, Audio Guidance, Museum, Art

1 Introduction

In museums and art galleries, it is important for visitors to feel, learn, and understand through the appreciation of exhibits. In recent years, museum educators—experts in learning within museums and art galleries—have played an important role in supporting visitors by providing explanations tailored to their interests and emotional states. However, there remains a persistent shortage of such personnel. In Japan, museum education is often supported primarily by curators, whose numbers remain limited. While the number of curators was approximately 3,000 in 2008, it increased by only about 20% to around 3,600 in 2018, and the average number of curators per museum remains fewer than four.

This situation indicates that many museums continue to face staff shortages, making it difficult to provide sufficient individualized guidance to all visitors [1].

In recent years, the use of audio guidance systems through rental devices and similar means has been advancing. However, most of these systems are limited to providing one-way explanations of artworks, essentially no more than digitized versions of in-gallery guides. Given this situation, many challenges remain for audio guidance systems and the like to serve as substitutes for museum educators.

In this study, we considered that this issue could be addressed by presenting content (compiling guidance content and order) interactively and dynamically, according to each visitor, based on estimates of the emotional states of people viewing exhibits. To this end, we newly developed an audio guidance system with the capability to sense diverse gestures and facial expressions, and conducted a data collection experiment on visitors' emotional states during art appreciation with 30 participants at the Ohara Museum of Art[3].

Using the constructed dataset, which consists of video data, acceleration and angular velocity data, and emotional state label data, we built and evaluated machine learning models to estimate emotional states. As a result of the evaluation, the binary emotion classification model (positive vs. negative) achieved an accuracy of 64%. On the other hand, regression models for satisfaction, comprehension, and perplexity did not yield effective results, though partial correlations were observed, providing insights for future improvements.

2 Related Work

Various methods have been proposed for emotional state estimation, particularly in the field of emotion recognition. For example, emotion estimation based on speech data collected through dialogue systems has been actively studied [8, 14]. In recent years, voice interaction for operating devices has become more realistic, and the use of voice is expected in a wider range of contexts, including tourism. To further improve recognition performance, multimodal emotion recognition methods that combine multiple sensor data have also been studied. For instance, Tzirakis et al. and Ghaleb et al. proposed approaches that combine audio with video data collected through dialogue [19, 5]. Other research has proposed emotion estimation based on eye gaze sensing and body movement sensing [21, 20]. In addition to emotions, Rach et al. proposed research on handling different emotional state dimensions such as interest, persuasiveness, comprehension, and relevance in argumentative dialogue systems [16, 15]. Prior studies have proposed methods for emotion recognition during sightseeing based on multimodal sensing approaches [12, 13].

As tools to support visitors' understanding in museums, navigation systems (audio guides) using rental devices or smartphone applications have been widely used both in Japan and abroad in recent years. Sun et al. reported that changing the way museum guide information is presented according to an individual's

[3] https://www.ohara.or.jp/

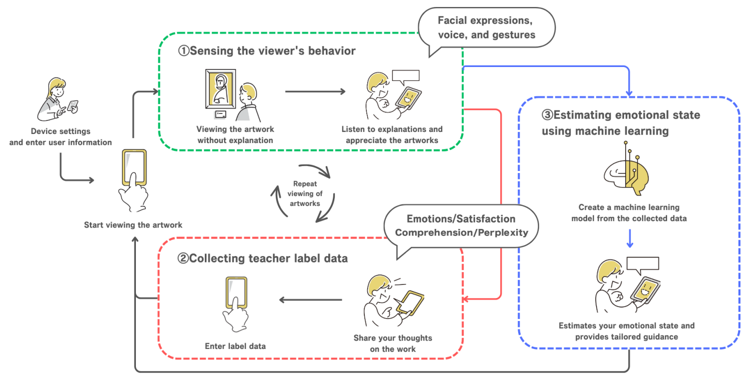

Fig. 1. Flow of Data Collection and Emotion Recognition

learning style can reduce cognitive load and enhance learning effects, highlighting the importance of such personalization [18]. Liu et al. conducted research to establish evaluation criteria for museums and, using those criteria, derived the relationship between the perceived attractiveness of museums and visitors' satisfaction [10].

Recent studies have applied affective computing techniques to museum and cultural heritage contexts. For example, emotion recognition based on facial expressions has been used to evaluate museum visitors' satisfaction [9], and psychophysiological analyses have been conducted to investigate emotional experiences during free exploration of art museums [11]. Other works have examined sensory museum design and conceptual approaches to understanding visitors' emotions during museum learning experiences [7, 4]. In addition, large-scale studies on facial expression recognition in unconstrained environments have provided important foundations for in-the-wild emotion analysis [3], and affective-computing-driven personalization has been explored in cultural and heritage settings [6].

In contrast, our work focuses on in-situ emotion estimation during real museum visits by combining facial expressions with behavioral and motion features through a sensing-enabled audio guidance system.

3 Proposed method

Various emotional states experienced when appreciating artworks are considered to be expressed through the visitor's facial expressions and gestures during appreciation. Therefore, we propose an emotional state estimation method based on sensing visitors during art appreciation. The workflow is shown in Figure 1, and each step is described below.

Step 0: Preparation At the beginning, the visitor inputs information such as prior experience with museums and guidance systems, as well as personality traits, and then starts the appreciation process.

Step 1: Artwork Appreciation The visitor selects and moves to an artwork of choice and begins appreciation. It is assumed that there are two possible modes of appreciation: appreciating the artwork without listening to any explanations, or appreciating it while listening to audio guidance of the artwork. During appreciation, behavioral data such as the visitor's facial expressions, voice, and gestures are sensed by a smartphone application providing guidance or by wearable devices.

Step 2: Label Input (for Model Construction Phase) After appreciating an artwork, the visitor verbally expresses impressions of the piece and inputs label data (ground-truth data) for the emotional state estimation model via the smartphone application.

Step 3: Emotional State Estimation (for Model Operation Phase) After appreciating an artwork, the emotional state estimation model estimates the visitor's recent emotional state, and the guidance is dynamically adapted and provided based on the estimation results. Steps 1–3 are then repeated.

3.1 Features and Ground Truth Labels Used for Model Construction

Emotional state estimation is realized by extracting features from collected data, using them as input values, and constructing a model that estimates the visitor's emotional states with label data as the output values.

The data obtained by sensing visitors during artwork appreciation include: video data captured by the smartphone's front camera (facial expressions during appreciation) and rear camera (viewing direction during appreciation), and acceleration/angular velocity data acquired from the smartphone and head-mounted motion sensors. The feature extraction methods for each sensor data are described below.

Video Data: For each frame of the video, we use OpenFace [2], a library capable of performing facial landmark detection, gaze estimation, head pose estimation, and Action Units (AUs) estimation from image data, to obtain frame-level features. From the resulting time-series features, the following are extracted and used:

- **Viewing Time (seconds):** The total time spent appreciating an artwork.
- **Counts of Continuous Viewing Durations (per time window) and App Viewing Counts:** Using gaze estimation data (`gaze_angle_x, _y`), we calculate the number of times the gaze continuously remained directed toward the artwork for longer than a predefined time window W_g (= 5, 10, 20, 30, 40, 50, 60, 90, 120 seconds), as well as the number of times the gaze shifted toward the smartphone during these intervals. Similarly, using head pose estimation data (`pose_Rx, Ry`), we calculate equivalent features with time windows W_h (= 0.5, 1, 2, 3 seconds).
- **Number of Blinks and Counts of Inter-Blink Intervals (per time window):** Using blink events (`AU45_c`) included in the Action Units, we

Table 1. Emotional state categories and their questions

Label	Question
Emotion	What kind of emotions did you feel when viewing the artwork?
Satisfaction	How satisfied were you with the artwork?
Comprehension	Do you feel that you understood the content or meaning of the artwork?
Perplexity	Were there any moments when you did not know what to focus on in the artwork?

Table 2. Options for user experience labeling

Emotion	Satisfaction	Comprehension	Perplexity
0: Excited / Surprised	0: Not satisfied at all	0: Did not understand at all	0: Never
1: Happy / Joyful	1: Dissatisfied	1: Did not understand	1: Not at all
2: Calm / Peaceful	2: Slightly dissatisfied	2: Hardly understood	2: Rarely
3: No particular change	3: Neutral	3: Neutral	3: Neutral
4: Sleepy / Tired	4: Somewhat satisfied	4: Somewhat understood	4: Somewhat infrequent
5: Miserable / Sad	5: Satisfied	5: Understood	5: Frequent
6: Bored	6: Very satisfied	6: Understood very well	6: Very frequent
7: Confused / Frustrated	–	–	–
8: Scared / Anxious / Angry	–	–	–

calculate the number of blinks and the number of inter-blink intervals (non-blinking durations) exceeding predefined time windows W_b (= 5, 10, 20, 30, 40, 50 seconds).

- **Statistical Features of Action Units:** For the AU intensity data `AU`$\boldsymbol{x}$`_r` (where $\boldsymbol{x}$ denotes the AU action ID), we compute summary statistics (mean, standard deviation, minimum, maximum, median, first quartile, third quartile, and mode).

Acceleration and Angular Velocity Data: From the six-axis time-series motion data (acceleration and angular velocity), the following features are extracted and used:

- **Number of Outlier Events:** An outlier detection algorithm is applied (a data point is considered an outlier if the absolute difference between the target data point and the mean of the past N samples exceeds twice the standard deviation of the past N samples), and the number of outlier events is counted.
- **Statistical Features of Acceleration and Angular Velocity:** For the acceleration and angular velocity data, summary statistics (mean, standard deviation, minimum, maximum, median, first quartile, third quartile, and mode) are calculated.

The emotional state labels (emotion, satisfaction, comprehension, perplexity) are obtained by presenting the questions in Table 1, and selecting one option from Table 2. The emotion labels are based on the two-dimensional circumplex model of affect defined by Russell et al. [17], and are divided into nine categories as shown in Table 2 (Emotion). For satisfaction, a 7-point Likert scale used in the Ministry of Land, Infrastructure, Transport and Tourism's tourism satisfaction survey is adopted. The same scale is also applied to comprehension and perplexity. For constructing the emotional state estimation model, a classification model is used for emotion labels, while regression models are constructed for the other labels.

In the following section, we describe the data collection and the construction of an emotional state estimation model using facial expression and gesture data.

4 Experiment and Evaluation at the Ohara Museum of Art

In this study, we newly developed an audio guidance system with sensing capabilities and conducted a data collection experiment at an actual art museum. The experiment was conducted on December 23, 2024. We collected data by sensing the viewing behavior of 30 participants for 30 artworks exhibited in the Ohara Museum of Art. This study was conducted with the approval of the "Research Ethics Committee for Studies Involving Human Subjects" at Okayama University (approval No.: Shizen 2024-13).

In the following, we describe the devices used and data collected, and the experimental procedure for data collection.

4.1 Devices Used and Data Collected

The devices used and the method of data collection for this experiment are detailed below. The experimental setup utilized three main devices: a smartphone, a headset, and a motion sensor. For the smartphone, we used five iPhone 12 Pro devices and five iPhone 14 Pro devices (all manufactured by Apple Inc.). We developed an application that serves multiple purposes: it provides audio guidance to the participants during viewing, delivered by an avatar[4]; it collects video/audio data during the viewing session; it records acceleration and gyroscope data from the smartphone itself; and it gathers post-viewing questionnaire data. To ensure the continuous recording of the participant's facial expressions and the artwork being viewed, the smartphone is secured to the participant's body using a neck-mounted smartphone holder. For the headset, we used the OpenRun[5] (manufactured by Shokz). This device connects to the smartphone and provides the artwork guidance audio to the participant viewing the application. Finally, the MetaMotion S[6] (manufactured by MBIENTLAB Inc.) was used as the motion sensor. It is attached to the left side of the headset to acquire acceleration and gyroscope data from the participant's head. The method for attaching this motion sensor is shown in Figure 3.

The data acquired by the devices were categorized into three main types: video data during artwork viewing, acceleration and gyroscope data, and emotional state labels. The video data during artwork viewing includes two streams

[4] CG-CA Gene (c) 2023 by Nagoya Institute of Technology, Moonshot R&D Goal 1 Avatar Symbiotic Society, `https://github.com/mmdagent-ex/gene/`

[5] `https://jp.shokz.com/products/openrun`

[6] `https://mbientlab.com/metamotions/`

recorded at a frame rate of 30 frames per second (30 fps). The participant's facial expressions during viewing were recorded using the smartphone's front camera. Concurrently, the appearance of the artwork being viewed was recorded using the smartphone's rear camera. The acceleration and gyroscope data were acquired from two sources: the motion sensor attached to the participant's head (at a frequency of 100 Hz) and the motion sensor embedded in the smartphone worn during viewing (at a frequency of 20 Hz). As the ground truth data, the emotional state labels were collected after the participant finished viewing each artwork. Participants responded to four questions presented via the application, as detailed in Table 1. These labels were categorized into four types: emotion, satisfaction, comprehension, and perplexity. The emotion was selected from nine distinct options, whereas satisfaction, comprehension, and perplexity were quantified using a 7-point scale.

4.2 Data Collection Procedure

Here, we explain the data collection flow, following the artwork viewing process using the audio guidance application.

Step 1: Explanation of the Data Collection Experiment
First, participants receive explanations regarding the purpose of the experiment, the experimental flow, and the method of using the application, and then they answer a pre-questionnaire. The pre-questionnaire collected demographic and background information (e.g., museum visit frequency); however, these data were not used in the current analysis and are reserved for future work. After completing the pre-questionnaire, participants secure the smartphone to their bodies and wear the bone conduction speaker, which is integrated with the motion sensor.

Step 2: Viewing Artworks Using the Application
Next, participants view the artworks and perform emotional state labeling using the guidance application shown in Figure 4. To begin viewing, participants stand in front of the artwork they wish to view and enter the artwork number, which is displayed near the artwork, into the application. After entering the artwork number and transitioning to the artwork viewing screen, participants first view the artwork without commentary (Figure 4 (a)). After viewing the artwork for several tens of seconds, they view it while listening to the audio commentary (Figure 4 (b)). After that, participants speak their impressions of the artwork for about 30 seconds and answer the questions that serve as emotional state labels on screens like those in Figure 4 (c) and (d). This completes one artwork viewing session, and this entire flow is repeated when viewing other artworks. The experiment in progress is shown in Figure 2. The artwork viewing time is 45 minutes, during which participants view the target artworks in the museum.

Step 3: Post-Questionnaire
After the artwork viewing is finished, participants answer a post-questionnaire.

Fig. 2. The experimental setup with the subjects

Fig. 3. Motion sensor attached to the bone conduction speaker

(a) Viewing start (b) Audio guide (c) Emotion input (d) Satisfaction input

Fig. 4. Screens of the guidance application

4.3 Building the Emotional State Estimation Model

Using the collected dataset, we built an emotional state estimation model as follows. First, we extracted the features shown in Section 3.1 from the collected data. Note that the audio data could not be used because it was not acquired due to device connection issues. Next, we built and evaluated the estimation models. For emotion label estimation, we built a 9-class emotional classification model to classify the nine labels shown in Table 2 (Emotion), and a 2-class classification model that groups them and classifies positive emotions (Positive: "Excited / Surprised," "Happy / Joyful," "Calm / Peaceful") and negative emotions (Negative: "Sleepy / Tired," "Miserable / Sad," "Bored," "Confused / Frustrated," "Scared / Anxious/Angry"). As the machine learning algorithm, we used the LightGBM Classifier[7] and performed hyperparameter tuning using Optuna[8], an automatic hyperparameter optimization framework. For sat-

[7] https://lightgbm.readthedocs.io/en/latest/pythonapi/lightgbm.LGBMClassifier.html
[8] https://optuna.org/

Table 3. Estimation results for emotion labels (9- and 2-class)

	9-class	2-class
Accuracy	0.14 ± 0.08	0.64 ± 0.09
Recall	0.19 ± 0.09	0.64 ± 0.09
Precision	0.19 ± 0.09	0.64 ± 0.09
F1	0.19 ± 0.09	0.64 ± 0.09

Table 4. Estimation results of satisfaction, comprehension, and perplexity labels

	Satisfaction	Comprehension	Perplexity
R2	-0.21 ± 0.12	-0.21 ± 0.15	0.05 ± 0.22
MAE	1.05 ± 0.13	1.24 ± 0.14	1.56 ± 0.20
RMSE	1.32 ± 0.14	1.55 ± 0.14	1.82 ± 0.21

isfaction, comprehension, and perplexity label estimation, we built regression models. As the machine learning algorithm, we used the LightGBM Regressor[9] and similarly performed hyperparameter tuning using Optuna. Furthermore, in evaluating the built estimation models, we employed stratified 10-fold cross-validation, taking into account the imbalanced number of labels. The current evaluation does not enforce strict participant-independent splits, and therefore primarily reflects within-subject generalization. Subject-wise generalization is left for future work.

4.4 Evaluation Results and Discussion

The evaluation results of the classification models for emotion label estimation are shown in Table 3.

Regarding the emotion labels, the results of the 9-class classification are shown in Figure 5, and the results of the 2-class classification are shown in Figure 6. The numerical values (0–8) on each axis in Figure 5 correspond to the IDs in Table 2 (Emotion), while in Figure 6, 0 corresponds to positive and 1 corresponds to negative. Although the accuracy for the 9-class classification was low at 14%, it was 64% for the 2-class classification, indicating that classification could be performed with a certain degree of accuracy.

While high estimation accuracy was not obtained in the 9-class classification, a certain degree of accuracy was obtained in the 2-class classification of positive and negative. A possible reason for this is that the amount of data secured per class was larger in the 2-class classification compared to the 9-class classification. Additionally, in this experiment, participants were asked to report their emotional response to the entire viewing of a single artwork. Therefore, it is possible that we could not distinguish whether that emotional state was elicited by the artwork, the guidance, or other factors. In the future, it will be necessary to take measures to minimize variation in responses, such as indicating the response criteria more clearly during the pre-experiment briefing.

The detailed evaluation results of the regression models for estimating satisfaction, comprehension, and perplexity are shown in Table 4. Regarding the estimation of satisfaction labels and comprehension labels, no trends were observed for either. Therefore, it became clear that estimating these using the features and methods employed this time is difficult. Regarding the estimation

[9] `https://lightgbm.readthedocs.io/en/latest/pythonapi/lightgbm.LGBMRegressor.html`

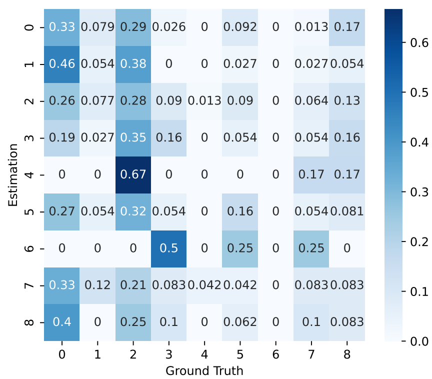

Fig. 5. Confusion matrix of 9-class emotion label estimation

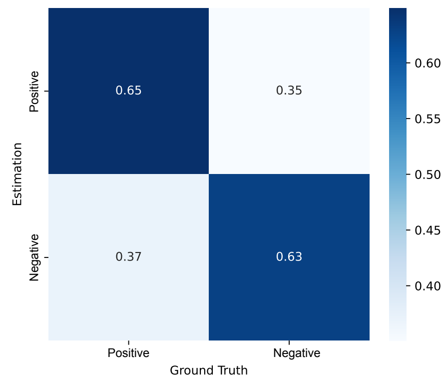

Fig. 6. Confusion matrix of 2-class emotion label estimation

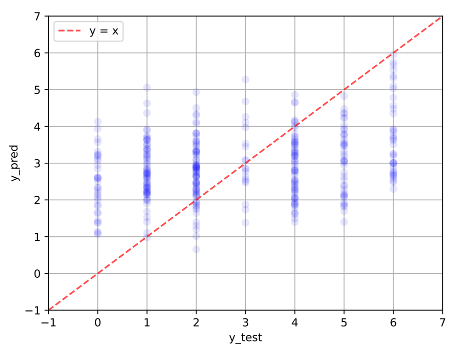

Fig. 7. Result of perplexity estimation (overall)

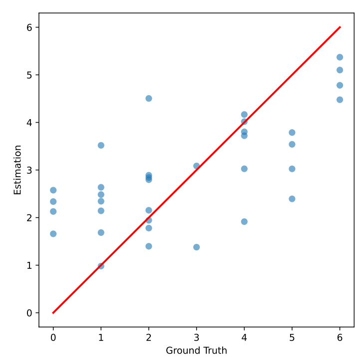

Fig. 8. Result of perplexity estimation (partial)

of perplexity labels, no correlation was observed overall, as shown in Figure 7. However, when focusing on data from one iteration of the cross-validation, results such as those in Figure 7 were obtained in some cases; for this data, the R2 score was 0.45, indicating a result with a certain degree of correlation. A possible reason why trends were observed in some parts for the perplexity labels, unlike the satisfaction and comprehension labels, is that the perplexity labels had less bias due to class balance compared to the other labels. Additionally, a possible reason why trends were observed only in some parts is that there were significant individual differences in the acquired data and features. Therefore, to improve the estimation accuracy for other labels, it can be considered necessary to: 1) equalize the class balance as much as possible, and 2) perform feature extraction that eliminates individual differences. Furthermore, it is necessary to explore methods to mitigate individual differences, such as adding features that account for individual variation.

5 Conclusion

In this study, we sensed the behavior of participants viewing artworks at the Ohara Museum of Art and conducted an experiment to collect multimodal sensing data and questionnaire data. Furthermore, using the data obtained from the experiment, we built models to estimate the emotional state during artwork viewing and analyzed the results. Regarding the estimation method using video data and acceleration/angular velocity (gyroscope) data, we achieved an accuracy of 64% in estimating emotion labels. However, no correlations were observed for satisfaction, comprehension, and perplexity. In the estimation of perplexity, a certain trend was observed in some of the estimation results. Future prospects for this study include: 1) exploring data and feature extraction methods useful for improving estimation accuracy, and 2) considering the construction of estimation models using other machine learning methods. By addressing these issues, we aim to improve estimation accuracy toward practical application.

Acknowledgments. This work was supported in part by the SCAT Research Grant and JSPS KAKENHI Grant Numbers JP22H03648 and JP24K20763. This research was also conducted with the cooperation of the Ohara Museum of Art. We would like to express our sincere appreciation.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Agency for Cultural Affairs: Changes in the number of museums, visitors, and curators. https://www.bunka.go.jp/seisaku/bijutsukan_hakubutsukan/shinko/suii/, accessed on 2 Feb., 2026 (in Japanese)
2. Baltrušaitis, T., Robinson, P., Morency, L.P.: OpenFace: An open source facial behavior analysis toolkit. In: IEEE Winter Conference on Applications of Computer Vision. pp. 1–10. WACV '16 (2016). https://doi.org/10.1109/WACV.2016.7477553
3. Benitez-Quiroz, C.F., Srinivasan, R., Feng, Q., Wang, Y., Martinez, A.M.: Emotionet Challenge: Recognition of Facial Expressions of Emotion in the Wild. arXiv preprint arXiv:1703.01210 (2017)
4. Fominska, I., Di Tore, S., Nappi, M., Iovane, G., Sibilio, M., Gelo, O.C.G.: Approaches to Identifying Emotions and Affections during the Museum Learning Experience in the Context of the Future Internet. Future Internet **16**(11), 417 (2024)
5. Ghaleb, E., Popa, M., Asteriadis, S.: Multimodal and temporal perception of audio-visual cues for emotion recognition. In: 8th International Conference on Affective Computing and Intelligent Interaction. pp. 552–558. ACII '19 (2019). https://doi.org/10.1109/ACII.2019.8925444
6. Hu, H., Wan, Y., Tang, K.Y., Li, Q., Wang, X.: Affective-Computing-Driven Personalized Display of Cultural Information for Commercial Heritage Architecture. Applied Sciences **15**(7) (2025)

7. Jelinčić, D.A., Šveb, M., Stewart, A.E.: Designing sensory museum experiences for visitors' emotional responses. Museum Management and Curatorship **37**(5), 513–530 (2022)
8. Kaya, H., Karpov, A.A., Salah, A.A.: Robust Acoustic Emotion Recognition Based on Cascaded Normalization and Extreme Learning Machines. In: Advances in Neural Networks. pp. 115–123. ISNN '16 (2016)
9. Kwon, D.H., Yu, J.M.: Real-Time Multi-CNN-Based Emotion Recognition System for Evaluating Museum Visitors' Satisfaction. ACM Journal on Computing and Cultural Heritage **17**(1), 1–18 (2024)
10. Liu, S., Guo, J.: Smart museum and visitor satisfaction. Journal of Autonomous Intelligence **7**(3), 1–13 (2024). https://doi.org/10.32629/jai.v7i3.1242
11. Marín-Morales, J., Higuera-Trujillo, J.L., Greco, A., Guixeres, J., Llinares, C., Gentili, C., Valenza, G.: Real vs. immersive-virtual emotional experience: Analysis of psycho-physiological patterns in a free exploration of an art museum. PLOS ONE **14**(10), e0223881 (2019)
12. Matsuda, Y., Fedotov, D., Takahashi, Y., Arakawa, Y., Yasumoto, K., Minker, W.: EmoTour: Estimating Emotion and Satisfaction of Users Based on Behavioral Cues and Audiovisual Data. Sensors **18**(11), 1–19 (2018). https://doi.org/10.3390/s18113978
13. Matsuda, Y., Fedotov, D., Takahashi, Y., Arakawa, Y., Yasumoto, K., Minker, W.: EmoTour: Multimodal Emotion Recognition using Physiological and Audio-Visual Features. In: Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers. pp. 946–951. UbiComp '18 (2018). https://doi.org/10.1145/3267305.3267687
14. Quck, W.Y., Huang, D.Y., Lin, W., Li, H., Dong, M.: Mobile Acoustic Emotion Recognition. In: 2016 IEEE Region 10 Conference. pp. 170–174. TENCON '16 (2016)
15. Rach, N., Matsuda, Y., Daxenberger, J., Ultes, S., Yasumoto, K., Minker, W.: Evaluation of Argument Search Approaches in the Context of Argumentative Dialogue Systems. In: The 12th International Conference on Language Resources and Evaluation. pp. 513–522. LREC '20 (2020)
16. Rach, N., Matsuda, Y., Ultes, S., Minker, W., Yasumoto, K.: Estimating Subjective Argument Quality Aspects from Social Signals in Argumentative Dialogue Systems. IEEE Access **9**, 11610–11621 (2021). https://doi.org/10.1109/ACCESS.2021.3051526
17. Russell, J.A.: A Circumplex Model of Affect. In: Journal of Personality and Social Psychology. pp. 1161–1178 (1980)
18. Sun, J.C.Y., Yu, S.J.: Personalized Wearable Guides or Audio Guides: An Evaluation of Personalized Museum Guides for Improving Learning Achievement and Cognitive Load. International Journal of Human–Computer Interaction **35**(4-5), 404–414 (2019). https://doi.org/10.1080/10447318.2018.1543078
19. Tzirakis, P., Trigeorgis, G., Nicolaou, M.A., Schuller, B.W., Zafeiriou, S.: End-to-end multimodal emotion recognition using deep neural networks. IEEE Journal of Selected Topics in Signal Processing **11**(8), 1301–1309 (2017)
20. Zhang, Z., Song, Y., Cui, L., Liu, X., Zhu, T.: Emotion recognition based on customized smart bracelet with built-in accelerometer. PeerJ **4**, e2258 (2016)
21. Zheng, W.L., Dong, B.N., Lu, B.L.: Multimodal emotion recognition using EEG and eye tracking data. In: 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. pp. 5040–5043 (2014). https://doi.org/10.1109/EMBC.2014.6944757