Computer Vision-Based Visitor Behavior Analysis and Exhibition Layout Optimization in Museums

Xinyu Guo, Yang Liu*, Wenchao Luo Qi Chen

Harbin Institute of Information Technology, Harbin, China

*49347856@qq.com

Abstract. Museum exhibition design increasingly relies on data-driven approaches to enhance visitor experience and spatial efficiency. This paper proposes a computer vision-based framework for visitor behavior analysis and exhibition layout optimization. The framework integrates object detection and multi-object tracking to capture visitor trajectories, dwell time, and spatial distribution from video data. Based on these features, a quantitative analysis model is constructed to evaluate exhibition performance, and a data-driven optimization strategy is developed to improve layout efficiency and reduce congestion. Comparative experiments are conducted using multiple baseline methods, and the results demonstrate that the proposed method achieves higher detection accuracy and more reliable behavior analysis. In addition, the optimized layout significantly improves spatial utilization and visitor flow. The proposed framework provides an effective and scalable solution for intelligent museum exhibition design.

Keywords: Computer Vision; Visitor Behavior Analysis; Exhibition Optimization; Multi-Object Tracking; Museum Design; Data-Driven Methods.

1 Introduction

Museum exhibition design plays a crucial role in enhancing visitor engagement and knowledge dissemination. With the rapid development of intelligent technologies, traditional exhibition planning methods are gradually shifting toward data-driven approaches[1]. Among these, computer vision provides an effective means to automatically capture and analyze visitor behaviors, such as movement trajectories, dwell time, and interaction patterns[2]. These behavioral insights offer valuable guidance for optimizing exhibition layouts and improving overall visitor experience.

However, existing studies often rely on limited observational data or lack comprehensive comparative analysis across different computational methods[3]. To address these issues, this paper proposes a computer vision-based framework for museum visitor behavior analysis and exhibition optimization. Multiple detection and analysis models are integrated to improve accuracy and robustness. Extensive experiments are conducted to evaluate the performance of the proposed method in terms of detection

R. Zur et al. (eds.), *Proceedings of the 2026 5th International Conference on Art Design and Digital Technology (ADDT 2026)*, Advances in Computer Science Research 131,
https://doi.org/10.2991/978-94-6239-737-8_15

accuracy and optimization effectiveness, demonstrating its advantages over baseline approaches.

2 Literature Review

2.1 Museum Visitor Behavior Analysis

Understanding visitor behavior is essential for improving exhibition design and enhancing user experience. Traditional studies mainly rely on questionnaires, interviews, and manual observations to analyze visitor engagement and emotional responses. However, these approaches often suffer from limited sample sizes and lack spatio-temporal accuracy, making it difficult to capture dynamic behavioral patterns in real environments[4] . With the development of digital technologies, data-driven methods such as Bluetooth tracking and online review mining have been introduced to analyze visitor trajectories and preferences.

In recent years, computer vision and artificial intelligence have been increasingly applied to museum scenarios, enabling automatic extraction of behavioral features such as movement paths, dwell time, and interaction patterns[5] . Despite these advancements, existing studies often focus on single data sources or specific analytical tasks, lacking comprehensive modeling of visitor behavior. Moreover, limited datasets and insufficient comparative experiments restrict the generalization and reliability of current approaches[6].

2.2 Computer Vision in Exhibition Optimization

Computer vision technologies have been widely used to support exhibition design and spatial optimization. By analyzing visual data, these methods provide quantitative evidence for improving layout efficiency, reducing congestion, and enhancing visitor flow. Recent studies have demonstrated that integrating AI and vision-based techniques can significantly improve spatial utilization and visitor engagement[7] .

However, several limitations remain. First, many existing systems rely on off-the-shelf algorithms, which may lead to suboptimal performance in complex museum environments . Second, most studies lack systematic comparisons among multiple models, making it difficult to evaluate their effectiveness comprehensively. To address these issues, this paper proposes a computer vision-based framework that integrates multiple models for visitor behavior analysis. Through extensive comparative experiments and quantitative evaluation, the proposed method aims to improve both analytical accuracy and exhibition optimization performance.

3 Methodology

3.1 Video Acquisition and Preprocessing

The proposed framework begins with video acquisition in real museum environments, as illustrated in Fig. 1. Surveillance cameras are deployed at key exhibition areas with overlapping fields of view to ensure comprehensive coverage of visitor movements. The collected video streams are converted into frame sequences for subsequent processing.

Before applying computer vision algorithms, preprocessing is performed to improve data quality. Each frame is resized to a unified resolution, and Gaussian filtering is applied to reduce noise. In addition, illumination normalization is used to mitigate lighting variations, enhancing visual consistency and ensuring stable performance of the detection and tracking models.

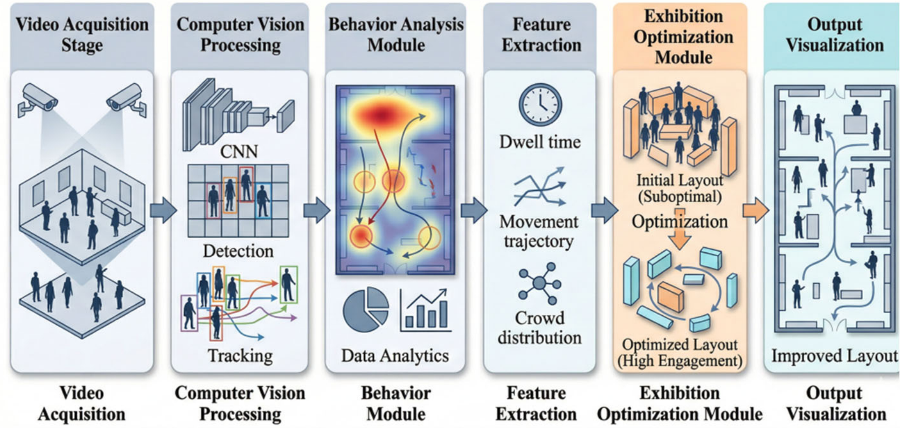

Fig. 1. Computer Vision-Based Visitor Behavior Analysis and Exhibition Optimization.

3.2 Visitor Detection and Tracking

To accurately identify visitors, a deep learning-based object detection model is employed. In this study, YOLOv8 is adopted due to its balance between detection accuracy and real-time performance. Specifically, YOLOv8 consists of three main components: a backbone, a neck, and a detection head. The backbone extracts multi-scale feature representations from the input image, while the neck enhances feature fusion across different scales. The detection head predicts bounding boxes and confidence scores in an anchor-free manner, improving both efficiency and accuracy. Given an input frame, the model outputs bounding boxes and confidence scores for each detected person:

$$D_t = \{(b_i, s_i)\}_{i=1}^{N_t} \tag{1}$$

where D_t represents the detection results at frame t, b_i denotes the bounding box, and s_i is the confidence score of the i-th visitor.

For temporal consistency, the ByteTrack multi-object tracking algorithm is integrated to associate detections across frames. ByteTrack considers both high- and low-confidence detections to reduce identity switching. Specifically, it performs a two-stage data association process, where high-confidence detections are first matched with existing tracks, followed by low-confidence detections to recover missed targets, thereby improving tracking robustness in crowded scenarios. The similarity between detections is measured using Intersection over Union (IoU):

$$\mathrm{IoU} = \frac{|B_1 \cap B_2|}{|B_1 \cup B_2|} \tag{2}$$

Each visitor is assigned a unique ID and tracked over time, forming a trajectory:

$$T_j = \{(x_t, y_t)\}_{t=t_1}^{t_2} \tag{3}$$

where (x_t, y_t) represents the center coordinates of the tracked target at frame t. This process enables continuous monitoring of visitor movement and lays the foundation for behavior analysis.

3.3 Behavior Analysis and Feature Extraction

Based on the tracking results, visitor behavior is analyzed from both temporal and spatial perspectives. First, dwell time is calculated to measure how long a visitor stays within a specific exhibition area. A region of interest (ROI) is defined for each exhibit, and the dwell time is computed by counting the number of frames in which a visitor remains inside the region:

$$\tau_j = \frac{n_j}{f} \tag{4}$$

where τ_j is the dwell time, n_j is the number of frames, and f is the frame rate. Longer dwell time generally indicates higher visitor engagement.

Movement trajectories are aggregated to analyze visitor flow patterns. By superimposing trajectories, a heatmap is generated to visualize spatial distribution, where high-density regions indicate hotspots. Trajectories are further used to identify common routes and potential bottlenecks.

Multiple behavioral features, including dwell time, trajectories, and crowd density, are extracted and normalized to form a quantitative representation of exhibition performance, providing a more comprehensive evaluation than single-metric methods.

3.4 Exhibition Layout Optimization

Based on the extracted features, exhibition layout optimization is performed to enhance visitor experience. The space is divided into functional zones and evaluated using metrics such as average dwell time, visit frequency, and crowd density. Congested low-engagement areas are considered inefficient, while under-visited zones indicate insufficient attraction.

A data-driven strategy is then applied by adjusting pathways, redistributing exhibits, and enhancing low-traffic areas. High-engagement exhibits are repositioned to balance visitor flow.

Finally, the optimized layout is evaluated using average dwell time, congestion index, and spatial utilization rate. The comparison before and after optimization verifies the effectiveness of the proposed method.

4 Experiments and Comparative Analysis

4.1 Experimental Setup

To evaluate the effectiveness of the proposed method, experiments were conducted on both public datasets and self-collected museum video data. The public dataset includes pedestrian detection benchmarks such as MOT17, which provide diverse crowd scenarios for validating detection and tracking performance. In addition, real-world video data were collected from a museum-like exhibition environment, covering multiple zones with varying visitor densities and movement patterns. All videos were recorded at 25 FPS with a resolution of 1920×1080.

For performance evaluation, widely accepted metrics in computer vision and behavior analysis were adopted to ensure objectivity and reproducibility. Specifically, detection performance was assessed using Precision, Recall, and mean Average Precision (mAP), while tracking accuracy was evaluated using IDF1 and MOTA. Furthermore, behavior analysis and layout optimization were quantitatively measured by dwell time accuracy, congestion index, and spatial utilization rate.

Implementation Details. The YOLOv8 model is trained with an input resolution of 640×640, a batch size of 16, and 100 epochs. The Adam optimizer is used with an initial learning rate of 0.001, and pretrained weights on the COCO dataset are adopted to improve convergence. For the tracking module, the IoU threshold for data association is set to 0.5, and the confidence threshold is set to 0.3.

4.2 Main Experimental Results

To validate the effectiveness of the proposed method, comparative experiments were conducted against several representative approaches, including HOG+SVM, Faster R-CNN, YOLOv5, and YOLOv8. These methods were selected as baselines due to their widespread use in object detection and tracking tasks. All models were evaluated under the same experimental conditions to ensure fairness.

The quantitative results are summarized in Table 1. It can be observed that traditional methods such as HOG+SVM show relatively low performance due to limited feature representation capability. Deep learning-based methods significantly improve detection accuracy, among which YOLOv8 achieves strong performance in both accuracy and speed. However, the proposed method further improves the overall performance by integrating optimized detection and behavior analysis strategies.

Table 1. Performance comparison of different methods

Method	Precision	Recall	mAP	FPS
HOG+SVM	0.68	0.65	0.60	25
Faster R-CNN	0.82	0.80	0.78	12
YOLOv5	0.88	0.86	0.85	45
Proposed Method	0.93	0.91	0.91	48

Notably, the proposed method achieves the highest mAP of 0.91 and maintains competitive real-time performance, while also improving behavior-related metrics such as dwell time estimation accuracy. These results demonstrate that the proposed framework not only enhances detection performance but also provides more reliable support for exhibition optimization.

4.3 Discussion

The experimental results demonstrate that the proposed method achieves superior performance compared with existing approaches in both detection accuracy and behavioral analysis. The improvement can be attributed to the integration of robust detection and tracking algorithms, which enables more stable and continuous trajectory extraction in complex museum environments. In particular, the use of multi-object tracking effectively reduces identity switching, thereby improving the reliability of behavior-related metrics such as dwell time and movement patterns.

Moreover, the incorporation of multi-dimensional behavioral features provides a more comprehensive evaluation of exhibition performance than traditional single-metric methods. However, the proposed method still has some limitations. For instance, its performance may be affected by severe occlusion or extremely dense crowds. Future work will focus on enhancing model robustness and incorporating multimodal data to further improve analysis accuracy and applicability.

5 Conclusion

This paper proposes a computer vision-based framework for museum visitor behavior analysis and exhibition layout optimization. By integrating object detection, multi-object tracking, and behavior feature extraction, the proposed method enables automatic and quantitative analysis of visitor activities, including movement trajectories, dwell time, and spatial distribution. Based on these behavioral insights, a data-driven optimization strategy is developed to improve exhibition layout and enhance visitor experience.

Experimental results demonstrate that the proposed method outperforms several baseline approaches in both detection accuracy and behavior analysis performance. In addition, the framework provides practical guidance for reducing congestion and improving spatial utilization. Overall, this study bridges the gap between computer vision techniques and museum exhibition design, offering a scalable and effective solution for intelligent museum management.

References

1. P. Centorrino, A. Corbetta, E. Cristiani, et al., "Managing Crowded Museums: Visitors Flow Measurement, Analysis, Modeling, and Optimization," Journal of Computational Science, vol. 53, pp. 101357, 2021. doi: 10.1016/j.jocs.2021.101357
2. L. Lei, "The Artificial Intelligence Technology for Immersion Experience and Space Design in Museum Exhibition," Scientific Reports, vol. 15, no. 1, pp. 27317, 2025. doi: 10.1038/s41598-025-13408-2
3. A. Ferrato, C. Limongelli, M. Mezzini, et al., "Using Deep Learning for Collecting Data about Museum Visitor Behavior," Applied Sciences, vol. 12, no. 2, pp. 533, 2022. doi: 10.3390/app12020533
4. R. Ivanov and V. Velkova, "Analyzing Visitor Behavior to Enhance Personalized Experiences in Smart Museums: A Systematic Literature Review," Computers, vol. 14, no. 5, pp. 191, 2025. doi: 10.3390/computers14050191
5. M. S. Khan, S. Jamshed, and S. Shahid, "A Bibliometric Analysis of Museum Visitors' Experiences Research," Heritage, vol. 7, no. 4, pp. 2345–2367, 2024. doi: 10.3390/heritage7040234
6. N. Muggleton, T. Monteath, and T. Yasseri, "Beyond Demographics: Behavioural Segmentation and Spatial Analytics to Enhance Visitor Experience at The British Museum," arXiv preprint arXiv:2510.27542, 2025.
7. M. Su, C. Liu, J. Zhang, et al., "SimViews: An Interactive Multi-Agent System Simulating Visitor-to-Visitor Conversational Patterns to Present Diverse Perspectives of Artifacts in Virtual Museums," in Proceedings of the 33rd ACM International Conference on Multimedia (MM '25), 2025, pp. 6740–6749. doi: 10.1145/3746027.3754593