YOLOv5를 활용한 전시회 관람객 동선 분석 및 시각화 시스템*

서도원[1], 남재홍[1], 김예경[1], 유성현[2], 정지훈[1*]
[1]충북대학교 소프트웨어학부
[2]충북대학교 컴퓨터과학전공
dsa3182@cbnu.ac.kr, {nam3762, yk_kim, sh.yu, jh.jeong}@chungbuk.ac.kr

Exhibition Visitor Trajectory Analysis and Visualization System Using YOLOv5

Do-Won Seo[1], Jae-Hong Nam[1], Ye-Kyung Kim[1], Seong-Hyun Yu[2], Ji-Hoon Jeong[1*]
[1]School of Computer Science, Chungbuk National University
[2]Department of Computer Science, Chungbuk National University

Abstract

This study focuses on the application of advanced human recognition and tracking technologies to analyze visitor movements and density in public spaces. It specifically examines the use of YOLOv5 and ByteTrack for tracking individuals, employing OpenCV for perspective transformation, and segmenting staff and visitors through k-means clustering. The system implemented in this study harnesses cutting-edge technological advancements to visualize visitor distribution and pinpoint popular exhibits, with the goal of improving visitor management and refining the layout of exhibition spaces. This research provides a technological framework that could significantly enhance the visitor experience in public venues.

1. Introduction

With the recent increase in large-scale events such as exhibitions and fairs, the importance of efficient visitor management and optimization of movement paths has come to the forefront [1]. Analyzing visitor movement paths is now crucial, not only for streamlining exhibition space operations but also for enhancing the overall visitor experience. Previous studies have predominantly utilized conventional sensor-based tracking systems or static statistical models, which, although effective to some extent, do not adequately meet the needs of environments requiring instantaneous data processing, offering greater versatility and accuracy. In this context, this study introduces an "Exhibition Visitor Trajectory Analysis and Visualization System" utilizing the latest computer vision technology, You Only Look Once (YOLO), to analyze and visualize the movement paths of exhibition visitors [2].

The system records video from a top-down view within the exhibition hall, tracking and analyzing visitors' movement paths using the data extracted from these recordings. YOLOv5, a real-time object detection technology, boasts high precision and fast processing speeds [3]. Through this technology, this study captures the movements of visitors, analyzing the data to derive various information such as popular areas within the exhibition space, congestion levels, and visitor routes. These insights empower exhibition organizers and managers to optimize exhibition space layouts, visitor movement pathways, and develop effective visitor response strategies

This paper introduces the development process of the Exhibition Visitor Movement Analysis System, discusses in detail the technical approaches used to implement the system, and the results obtained from it. Furthermore, it evaluates the potential impact of this system on the management and operation of exhibitions.

This research was supported by the MISP(Ministry of Science, ICT & Future Planning), Korea, under the National Program for Excellence in SW (2019-0-01183) supervised by the IITP(Institute of Information & communications Technology Planing & Evaluation) (2019-0-01183) and supported by the Institute of Information & communications Technology Planning & Evaluation (IITP) grant, funded by the Korea government (MSIT) (RS-2021-II212068, Artificial Intelligence Innovation Hub) and supported by Korea Institute of Planning and Evaluation for Technology in Food, Agriculture and Forestry(IPET) through Agriculture and Food Convergence Technologies Program for Research Manpower Development Program funded by Ministry of Agriculture, Food and Rural Affairs(MAFRA)(RS-2024-00398561).

2. Materials and methods

2.1. Object Detection and Tracking

YOLOv5 is a state-of-the-art deep learning algorithm, is deployed in the sutdy for its real-time object detection capabilities. Specifically, YOLOv5 swiftly detects individuals in CCTV footage of exhibitions, pinpointing their locations and bounding boxes within the video frames. This high precision in initial detection provides an accurate starting point for tracking visitor movements, which is essential for reliable monitoring throughout the exhibition space. Following the initial detection, ByteTrack is utilized to track the detected individuals across subsequent video frames [4]. This tracking technology assigns unique IDs to each visitor, allowing for continuous and accurate tracking of individual movements by linking detection data from frame to frame. This method ensures that each visitor’s trajectory is consistently followed, enhancing the analysis of movement patterns within the exhibition.

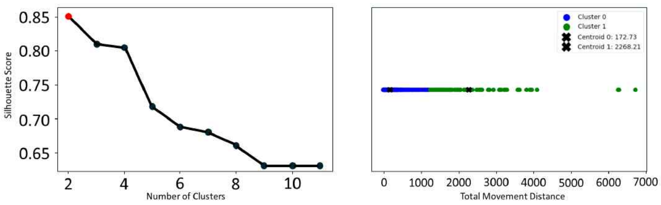
Additionally, the study integrates OpenCV for perspective transformation, adjusting the video data to accurately represent the actual spatial layout [5]. This adjustment is crucial for aligning the detected positions with the physical space, thus improving the accuracy of the data analysis.

2.2. Pre-processing

2.2.1. Removal of Outliers Using K-means Clustering

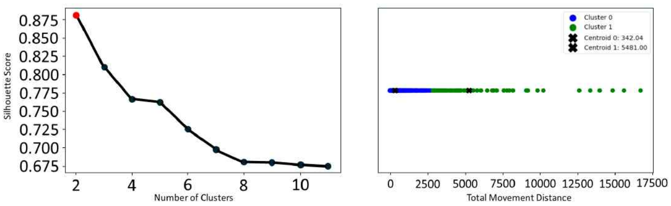
To analyze the pattern of the dataset based on the calculated moving distance, K-means Clustering is applied [6],[7]. Employing the Scikit-learn library, the optimal number of clusters is determined through the calculation of the Silhouette score, which gauges the cohesion and separation of the resulting groups [8]. The clustering result with the highest Silhouette score is chosen to identify and exclude datas with abnormally low moving distances (outliers). This step enhances the analysis's accuracy by distinguishing and removing outliers who exhibit abnormally short travel distances, ensuring they are not mistakenly considered part of the audience's typical movement patterns.

As can be seen in Fig. 1 and Fig. 2, it became clear that for both the 10-minute and 30-minute videos, the silhouette scores were maximized when the cluster count was set to 2. As a result, the individuals were segmented into two groups. The first group was identified as outliers with anomalously short movement distances, leading to the removal of their ID values. The second group was recognized as representing the normal movement patterns of the average visitors.



(A) (B)

Fig. 1. (A), Silhouette score analysis for a 10-minute video to determine the optimal number of clusters for visitor movement data. (B), K-means clustering of total movement distances from 10-minute video, with centroids indicating average distances within each cluster.



(A) (B)

Fig. 2. (A), Silhouette score analysis for a 30-minutes video to determine the optimal number of clusters for visitor movement data. (B), K-means clustering of total movement distances from 30-minutes video, with centroids indicating average distances within each cluster.

2.2.2. Perspective Transformation

After outliers are removed, the coordinates of the remaining data are adjusted using perspective transformation with the OpenCV library. This step adjusts the coordinates to better match the actual layout of the exhibition hall, taking into account the viewpoint of the CCTV cameras. These transformed coordinates are then analyzed on a top-down planar representation, offering a bird's-eye view of the exhibition space visitor traffic.

Fig. 4 illustrates the analysis of the spatial distribution of visitors using perspective-transformed coordinate data. This analysis employs kernel density estimation (KDE), a feature of Python's Matplotlib library, to render the visitors' distribution within the exhibition space in the form of a 3D heatmap.

(A) (B) (C)

Fig. 3. (A), An image from a scene of a video taken at an exhibition. (B), This image shows the result of object detection specifically humans within polygons using YOLOv5, represented with bounding boxes. (C), An image after the Perspective Transformation process.

3. Results and Discussion

3.1. Visitor Distribution Analysis Using 3D Heatmap

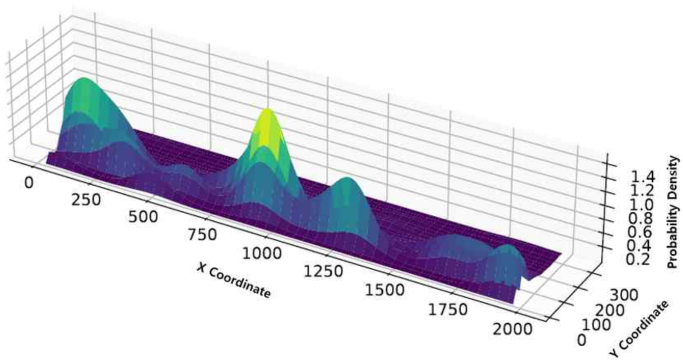


Fig. 4. 3D visualization of visitor density from a 5-minute video using Kernel Density Estimation (KDE) on exhibition space coordinates, highlighting areas of high visitor concentration.

KDE, a non-parametric method for estimating the probability density function of a random variable, applies a Gaussian kernel to each coordinate, effectively smoothing discrete data points into a continuous probability density surface. This technique is adept at revealing patterns and concentrations in spatial data, such as identifying areas within an exhibition hall where visitors are more concentrated, by creating a smooth density estimate that allows for visual identification of these regions [9]. Following the smooth density estimate created by the KDE process, this heatmap further assists in identifying popular areas within the exhibition. During the heatmap generation, the height of each coordinate indicates the density of visitors at that point, while the intensity of the color visually represents the concentration of visitors, making it easier to spot where people gather the most.

3.2. Visualization of Movement Paths Using 2D Graphs

As shown in Fig. 5, each visitor's movement through the exhibition is meticulously plotted on a two-dimensional layout using perspective-transformed moving coordinates. The vibrant spectrum of lines, each assigned a unique color by the Matplotlib library, traces the distinctive routes traversed by the attendees. Each visitor's path is distinctly marked with a unique color, creating a clear and easy-to-follow map of movement across the exhibition space. This approach not only differentiates each person's movements but also assembles a clear visual summary of how visitors are navigating the area. The varied colors work together to provide insights into the most frequented paths and highlight patterns of visitor engagement within the exhibition.

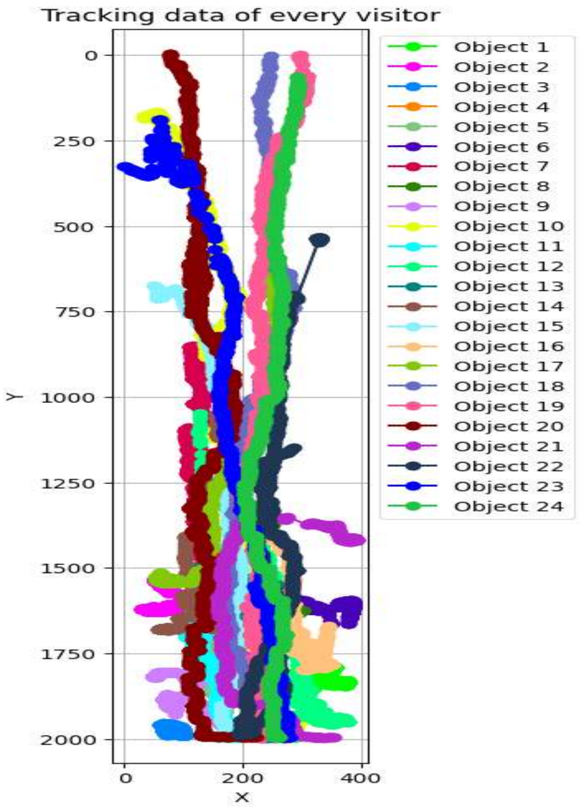


Fig. 5. Tracking data from 5-min video, with each visitor’s path uniquely color-coded, illustrates individual movements within the space, providing insights into visitor flow and for enhanced exhibition management.

3.3. Analysis of Visitor Distribution around Installations Using Fuzzy C-Means Clustering

The Fuzzy C-Means clustering method interprets the membership of data points in degrees of belonging, instead of as a binary distinction [10]. By applying this method to analyze the distribution of visitors around the installations within the exhibition space, it allows for a more nuanced understanding of how much interest each visitor has in specific installations. In this process, coordinates are established based on the central point of each installation, and these coordinates serve as the centers of clusters. Visitors are then classified based on their degree of membership to these cluster centers, quantifying their level of interest in each installation. Notably, visitors who do not reach a certain level of membership are considered to be wandering the exhibition space without a clear interest in the main installation. This allows for a more detailed analysis of visitor behavior patterns and interests.

4. Conclusion and future works

Through this study, the potential of applying advanced technologies for analyzing visitor movement and density in public spaces has been demonstrated. Employing YOLOv5, ByteTrack, OpenCV's perspective transformation, and k-means clustering has enabled a nuanced analysis, effectively differentiating staff from visitors and precisely tracking their movements. visualizing visitor distribution through Kernel Density Estimation has yielded essential insights into identifying high-traffic exhibits and has informed the optimization of visitor management and space planning. These technological approaches present a significant methodology for improving management and operations in public spaces and enhancing visitor experiences.

As future work, the system could be expanded to operate within multi-camera environments, a step that would involve coordinating and analyzing data from various vantage points. This approach would improve coverage and detail in mapping visitor movements, particularly in more extensive and complex venue layouts.

References

[1] Centorrino, P., Corbetta, A., Cristiani, E., Onofri, E, “Managing crowded museums: visitors flow measurement, analysis, modeling, and optimization.” Journal of Computational Science, 53, 101357, 2021.

[2] Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi, "You only look once: Unified, real-time object detection," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 779-788, 2016.

[3] N. M. Krishna, R. Y. Reddy, M. S. C. Reddy, K. P. Madhav and G. Sudham, "Object Detection and Tracking Using Yolo," 2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA), pp. 1-7, 2021.

[4] Yifu Zhang, Peize Sun, Yi Jiang, Dongdong Yu, Fucheng Weng, Zehuan Yuan, Ping Luo, Wenyu Liu, Xinggang Wang, "Bytetrack: Multi-object tracking by associating every detection box," European Conference on Computer Vision, Cham: Springer Nature Switzerland, pp 1-21, 2022.

[5] Rawat, Kuldeep S. "The pseudo-perspective transformation for motion compensation in digital video compression." Ph.D. dissertation, University of Louisiana at Lafayette, pp. 35-39, 2005.

[6] Taesik Yoon, Kyuseok Shim, "Efficient K-means Clustering for High-dimensional Large Data," in Proceedings of the Korean Institute of Information Scientists and Engineers Conference, vol. 38, no. 1(A), pp. 33-36, 2011.

[7] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman and A. Y. Wu, "An efficient k-means clustering algorithm: analysis and implementation," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 7, pp. 881-892, 2002.

[8] Peter J. Rousseeuw, "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis," Journal of Computational and Applied Mathematics, 20, pp 53-65, 1987.

[9] Soonhwan Kim, "Kernel Probability Density Function Estimation and Its Case Study," Master's Thesis, Graduate School of Duk Sung Women's University, Department of Mathematical Statistics, Statistics Major, Seoul, pp. 1-6, 2004.

[10] Bezdek, James C., Robert Ehrlich, and William Full, "FCM: The fuzzy c-means clustering algorithm," Computers & Geosciences, 10.2-3, pp 191-203, 1984.