REVISITING WILLIAM H. WHYTE'S DIAGRAM

An interactive video-tracking method for modelling spatiotemporal human behaviours within dynamic zoning geometries

CHAO YAN[1*], TE LI[2], SIYAN LIU[3], SHANSHU HE[4], LANMING ZHANG[5], SHU WEN[6], YUTONG AN[7], WENJIE WANG[8] and YINI LONG[9]
[1,2,3,4,5,6,7,8,9] *Tongji University, College of Architecture and Urban Planning.*
[1]*yanchao@tongji.edu.cn, 0000-0003-1446-4861*
[2]*2330265@tongji.edu.cn*
[3]*978939929@qq.com*
[4]*2252279@tongji.edu.cn*
[5]*2254149@tongji.edu.cn*
[6]*3166653915@qq.com*
[7]*281582047@qq.com*
[8]*2488265427@qq.com*
[9]*2151226@qq.com*

Abstract. This article explores the intersection of architectural forms and human behaviour through a video-tracking analysis method. Focusing on the impact of geometrical subdivision of public zones on social behaviour, it addresses limitations of current video-tracking tools. By introducing the customized BodyCAM software, architects can interactively delineate zoning boundaries on video footage, enabling detailed analysis of spatiotemporal patterns. The software tracks pedestrians, identifies geometric boundaries, and correlates behaviour patterns with spatial configurations, offering a novel approach to understanding human behaviour in complex architectural environments. This synthesis enhances both conceptual and operational aspects of architectural design in association with environmental behaviour, emphasizing the intricate interplay between architecture and the human body.

Keywords. Behaviour Study, Video-Tracking, Public Space, Zoning Geometry, Spatiotemporal Diagram.

1. Introduction

Video tracking has emerged as a pivotal tool in the realm of behavioural research, offering unprecedented insights into the dynamic behaviours of human occupants within their built environments. The ability to monitor and analyse the movements and gestures of individuals over extended periods has provided architects with a wealth of data that transcends the limitations of traditional observational methods, not only enhancing the precision and efficiency, but has also opening new avenues for understanding the intricate social interaction between individuals.

Meanwhile, video tracking as an observational method holds a distinct advantage over experimental approaches involving the wearing of tracking devices. The primary merit lies in the unobtrusive nature of video tracking, which allows for a more objective and naturalistic study of behavior. Unlike experimental setups where subjects may be influenced or cognizant of the devices they wear, video tracking allows researchers to observe individuals without introducing external interference, thereby capturing a more authentic representation of behavior patterns. This objectivity becomes particularly pronounced in complex environments, especially when studying social interactions. Participants can move freely and naturally in scenarios where individuals engage in intricate social behaviors within intricate spatial contexts.

In the realm of behavioural research in public spaces, various dynamic and static zoning contours shape the environment, encompassing physical space divisions, distribution of facilities, changes in shadowed areas, and the formation of social clusters. These elements wield a substantial influence on social behaviour. However, the limitations of video tracking become evident in its inability to establish a direct correlation with the intricate geometries of these zoning configurations. Video tracking excels in globally revealing spatiotemporal behavioural patterns but falls short in connecting with the detailed aspects of zoning geometry.

Furthermore, some zones are objectively defined, while others are latent or subject to the architect's interpretation. Regrettably, current video tracking methods do not empower architects with interactive opportunities to delineate and analyse these zones. This lack of interactivity restricts the exploration of potential spatial influences on social behaviour, leaving a gap in understanding the complex interplay between zoning geometries and human movements. Thus, there exists a need for innovative methodologies that not only capture global behavioural patterns but also allow architects to interactively engage with and manipulate zoning configurations for a more detailed analysis of social behaviours within public spaces.

This article explores the intricate interplay between behavior and geometry, introducing a customized tool and innovative methodology to augment architects' capacity for elucidating spatiotemporal patterns of social behavior in conjunction with their unique spatial interpretations. The core objective of this study is to discern the profound influence of geometrical subdivision within public spaces on complex social interactions within the context of our rapidly accelerating society.

2. Literature Review

Traditional observational techniques often struggle to record dynamic behavioural patterns, especially in complex environments. Video tracking, however, overcomes these challenges by enabling continuous, non-intrusive monitoring of subjects, allowing architects to gain comprehensive insights into their responses to environmental stimuli.

One key advantage of video tracking is its ability to facilitate the analysis of spatiotemporal patterns in behaviour. And it has roots in the seminal work of William

H. Whyte, a pioneer whose extensive research fundamentally shaped the understanding of human behaviour in urban spaces. Whyte's ground-breaking studies, notably his investigation of the Seagram Building's forecourt, offered invaluable insights into the temporal aspects of individuals interacting with their urban surroundings. Whyte introduced spatiotemporal diagrams, depicting the ebb and flow of pedestrian movements. In the iconic visualization 'A day in the life of the north front ledge at Seagram's', Whyte captured the patterns of social behaviours on the ledge. He constructed a chronological chart, resembling a player piano roll. Each line on the chart represented a sitter, with the length indicating their duration of occupation and the columns on the right specifying the total number of the occupants of the day. The continuous line at the chart's bottom, tracking the total number of sitters over time, added an additional layer of insight into the collective patterns of human engagement within the urban space (Whyte, 1980). Whyte's innovative and pioneering methodologies, exemplified in this study, laid the foundational principles for contemporary video-tracking techniques. Furthermore, his spatiotemporal diagrams intricately dissected and correlated behavioural patterns with the facility distribution on the Seagram Building's forecourt. This analysis contributed significantly to urban design by revealing the relationship between human activities and the geometrical placement of amenities within public squares, thereby informing more effective and user-centric architectural interventions.

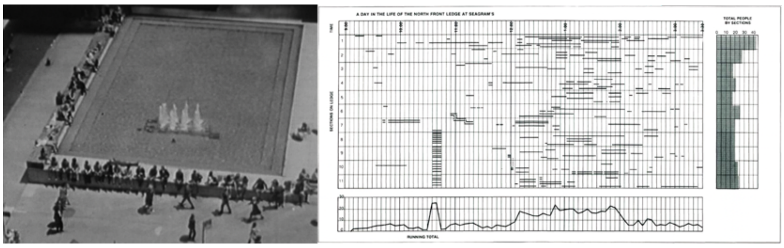

Figure 1. William H. Whyte's Spatiotemporal Diagram

Today, the advent of digital technologies, such as AI recognition and segmentation algorithms, has revolutionized video-tracking behaviour research, enhancing precision and automation, opening new frontiers for researchers to delve deeper into the intricacies of spatiotemporal dynamics. In architecture discipline, Gonzalez Rojas addresses the challenge of incorporating people's motion into architectural design through a data-driven model capturing 4D pedestrian behaviour (Gonzalez Rojas, 2017). In more detailed level, Ambrose introduces a novel digital investigation approach focusing on the geometric translations of the body in motion. This work explores the spatial experience generated by the human body, offering a reinterpretation of the body and its dynamic motion as a set of event spaces (Ambrose, 2007). For larger scale video-tracking, Guo, Wang, and Yuan utilize drones photography in human behaviour sensing within outdoor built environments. This research introduces a system integrating thermal-infrared images acquisition and processing, capturing city patterns and human behaviour characteristics (Guo et al., 2019). While some research focusing on developing video-tracking techniques, others explores the integration of both physical observation and virtual experiments. Yan proposes an integrated system utilizing automatic video tracking and video-quality virtual reality to enhance environmental behaviour study. This innovative approach not only provides substantial statistical measurements of behaviour patterns but also allows architects to interactively observe details of behaviours from various viewpoints in virtual reality (Yan, 2006). Pechlivanidou-Liakata et al. also explore spatial navigational patterns induced by real and virtual architectural environments. Their research focuses on identifying elements that trigger human behaviour in both contexts, whether material or immaterial. Through a series of experiments in real and virtual environments, the study confirms the primary role of spatio-exploratory perception in virtual environments, emphasizing the impact of specific architectural elements on subjects' movement and exploration (Pechlivanidou-Liakata et al., 2010).

The evolution of video tracking methods extends far beyond the realm of architecture, showcasing diverse applications across various domains. While architectural implementations predominantly focus on global mapping of behaviour, other domains are customizing video tracking methods for more diversified analyses. In the field of neuroscience, Arac et al. introduce DeepBehavior, a deep learning toolbox facilitating detailed behavioural analysis in animals and humans, for neuroscience and clinical motor function assessment (Arac et al., 2019). Regarding to specific behaviours, Auguste et al. contribute a behavioural analysis method for crowd movement, utilizing parameters derived from image analysis algorithms to detect abnormal phenomena in urban environments, offering potential applications in enhancing video surveillance for dangerous behaviours (Auguste et al., 2021). Intelligent surveillance, as proposed by Albusac et al., involves AI techniques and normality analysis to detect abnormal behaviours. This conceptual framework, based on normality concepts, provides realistic solutions for designing surveillance systems (Albusac et al., 2009). Wu et al. address long-term trajectory prediction in pedestrian movements, proposing a space-time tree search method that considers both spatial and temporal dimensions, showcasing its potential for autonomous driving and intelligent navigation (Wu et al., 2022). Torrens and Gu present inverse augmentation, an innovative approach that immerses real human users in urban simulations, transposing their embodied behaviour into models (Torrens and Gu, 2023). Lastly, Rezaei and Azarmi present DeepSOCIAL, a hybrid Computer Vision and deep neural network model for automated people detection and social distancing monitoring, demonstrating its superior performance and infection risk assessment capabilities in the context of the COVID-19 pandemic (Rezaei and Azarmi, 2020). These studies collectively underscore the versatility and customization of video tracking methodologies beyond architecture, providing detailed insights and applications in neuroscience, surveillance, crowd analysis, and public health.

In summary, in the realm of video-tracking behaviour research in architectural discipline, technological advancements focus more on addressing complex scenarios, enhancing tracking accuracy, and managing large-scale data. However, predominant approaches still prioritize macroscopic behaviour mapping through heatmap and trajectory drawing, often neglecting reconstruction of analytical dimensions and contents. Conversely, in diverse fields, innovative exploration of fundamental video-tracking mechanisms and analytical methods reveals diverse behavioural patterns. This not only opens avenues for understanding the intricate relationship between behaviour and space but also provides inspiration and technological foundations for redefining research paths in this study.

3. BodyCAM: A Customized Video-Tracking Tool

In architecture, existing tools and methodologies for behavioural tracking primarily address geometric challenges by simply mapping behavioural data onto spatial layouts. This approach, however, falls short in modelling the intricate spatiotemporal correlations between behaviours and forms within highly detailed small-scale public spaces. To overcome this limitation, this paper proposes an innovative analytical approach based on a custom video-tracking software named as BodyCAM.

This customized software is tailored for recognizing dynamic interactions among multiple individuals in complex urban and architectural spaces. The underlying computer vision technology of BodyCAM is built upon the YOLOX and ByteTrack open-source algorithms, both renowned for their exceptional performance in object detection and multi-object tracking. YOLOX, an advanced evolution of the YOLO series, adopts an anchor-free approach and introduces sophisticated detection techniques. This leads to various models, achieving superior performance in terms of Average Precision (AP) on the COCO dataset. ByteTrack, on the other hand, focuses on multi-object tracking, aiming to recover true objects and filter out background detections, ultimately achieving outstanding results in multiple tracking benchmarks.

To enhance its analytical capabilities, BodyCAM incorporates a simulated perspective distortion feature. Users can establish a transformation matrix by drawing reference points on the video, facilitating the conversion of the ground layout into a top-view orthographic projection. This transformation is crucial for accurately projecting behavioural trajectories and analysis visuals onto a planar map, eliminating dimension errors introduced by perspective distortion. This correction proves particularly beneficial when calculating relative metrics such as pedestrian distance and velocity, ensuring precision in spatial analyses.

BodyCAM extends its functionality to encompass a range of fundamental analyses for multiple target individuals. Firstly, the software facilitates social distancing calculations, presenting results through visualizations of heatmaps. Meanwhile, the software offers occupation analysis in a grid-based map representing the accumulation of different targets, color-coded to signify varying densities. It also provides calculations into the accumulated occupation time of targets in each grid, offering a nuanced understanding of temporal dynamics. Additionally, users can delve into further analyses, exploring pedestrian trajectories, velocities, and other behavioural attributes.

Most importantly, BodyCAM introduces a novel feature that empowers architects through an interactive interface. Architects can delineate zoning boundaries directly on video footage, allowing for the analysis of spatiotemporal patterns within these defined zones. The software not only enables the designation of zoning boundaries based on the physical layout of spaces but also allows architects to draw these zones with intricate geometries that align precisely with their understanding of the space. This interactive and customizable feature provides architects with a sophisticated tool to unravel the intricacies of spatiotemporal patterns within diverse architectural contexts.

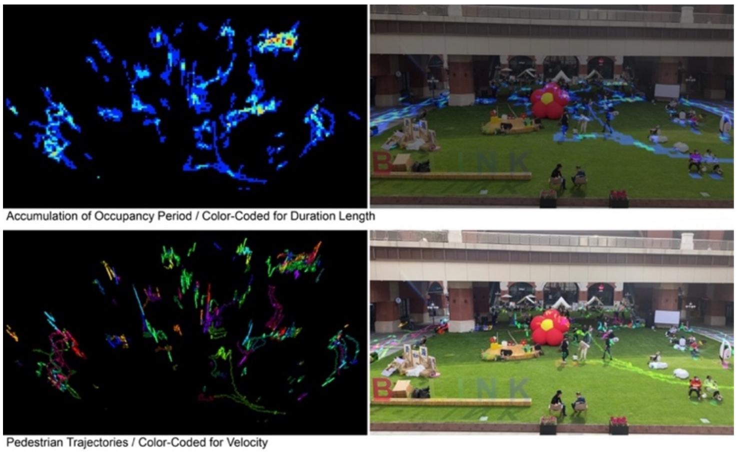

Figure 2. BodyCAM: Video Tracking and Analysis Tool

4. An Interactive Method of Behavioural Tracking for Design

The interactive zone delineation and spatiotemporal behaviour mapping functionality in BodyCAM offer architects a novel approach to complex behavioural analysis. Architects can now pre-analyse video footage or subjectively defining spatial layouts in various forms, revealing spatiotemporal patterns of behaviour distribution within different zoning geometries. This provides a unique method to establish correlations between spatial zone morphological elements and behavioural patterns, opening new possibilities for architects to gain insights into human dynamics within diverse architectural contexts.

4.1. CATEGORIES OF ZONING GEOMETRY

The foundation of this method involves a comprehensive exploration of various classifications of zoning geometry, encompassing both physical and virtual zoning geometries.

Physical zoning geometry refer to spatial divisions in video footage based on inherent physical elements.

- Spatial Element-Defined Regions: The first type of physical zoning geometry are defined areas by the designed spatial division elements like walls, furniture, green elements, pavement, and elevation differences that naturally segment the space.
- Attractor-Formed Fields: Another type of physical zoning geometry emerges from objects acting as attractors, forming fields that radiate outward. These objects can include trees, playground equipment, installations, seating facilities, and more.

Virtual zoning geometry, on the other hand, are potential spatial configurations requiring exploration by architects.

- Latent Visual Layers: One type involves visually judged zoning division, such as areas with shadows and those without.

- Architect's Subjective Interpretation: The second type relies on subjective judgments by architects, identifying implicit place-like regions within the physical space. These may include areas with a design sensibility that architects need to discern and then depict in the video.
- BodyCAM Pre-Analysis: The third type utilizes pre-analysis conducted through the basic functions of BodyCAM software. For example, tracking results in densely populated areas become references for zoning.
- External Software Pre-Analysis: The fourth type involves pre-analysis using external software, such as microclimate simulation tools that generate spatial heatmaps based on factors like thermal comfort and wind environment. These results yield diverse types of spatial configurations.

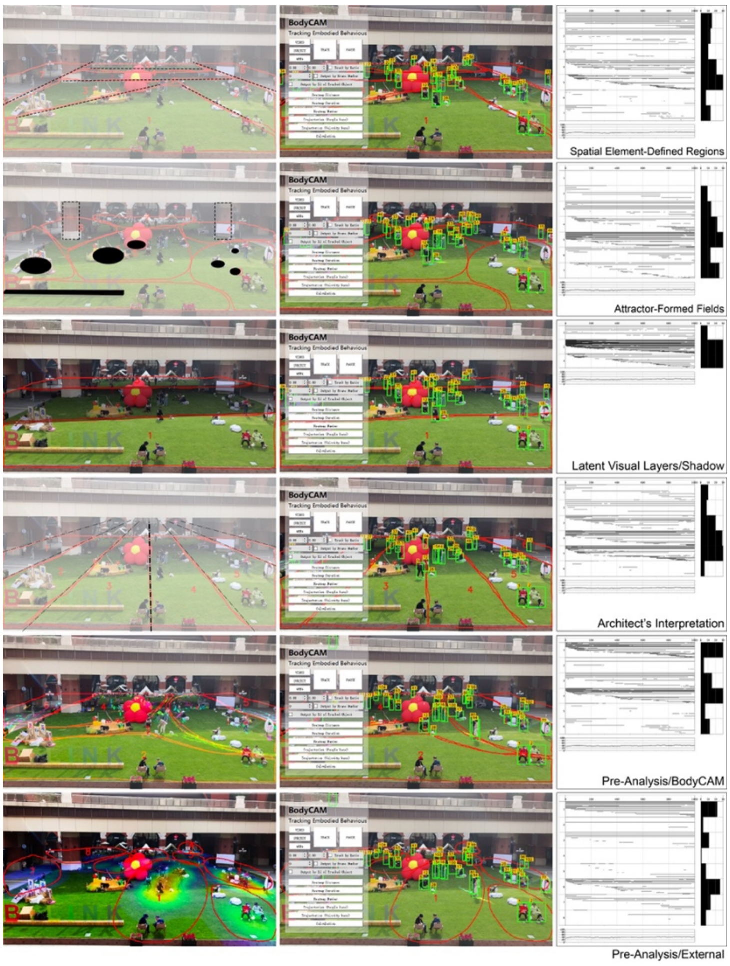

Figure 3. Categories of Zoning Geometry with its Corresponding Spatiotemporal Pattern

This typological research on zoning geometry in video space provides a detailed understanding of different spatial configurations, enhancing the ability to discern the impact of each type on behaviour. By employing the BodyCAM software, architects can create spatiotemporal distribution maps in responding to each category, elucidating the complex interplay between various spatial elements and behaviour patterns.

4.2. WORKFLOWS OF SPATIOTEMPORAL ANALYSIS

The spatiotemporal behaviour analysis process based on drawing zoning geometry involves three main stages. First, architects prepare zoning geometry by conducting pre-analyses using various software tools or sketching directly on video screenshots to conduct initial spatial assessments.

The subsequent step entails the use of the developed software interface to delineate various geometric boundaries within the spatial context. These boundaries can be the demarcation between different pavements, the outlines of urban furniture, and the curved contours of green areas, etc. By precisely tracing these boundaries, the space is subdivided into a series of zones, each characterized by distinct shapes. The software analysed the spatiotemporal occupancy of these zones, presenting the findings in the style of William H. Whyte's diagrams.

Third, the software tracked the movements of crowds within the space in real-time, mapping this data into the diagrams. Subsequently, through the interactive tracing of different geometric boundaries based on distinct spatial interpretations, the software unveiled diagrammatic correlations between behaviour patterns and the concealed layers of spatial geometry.

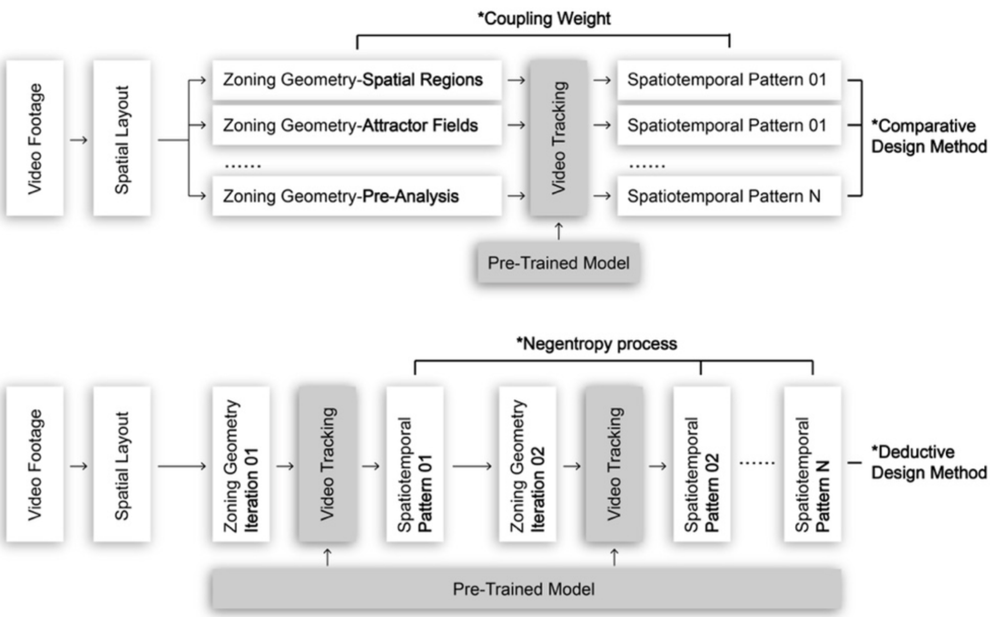

Figure 4. Workflows of the Video-Tracking Method for Design

From the methodological perspective of design, the spatiotemporal behaviour analysis based on drawing zoning geometry can be categorized into two main approaches. The first involves generating a set of zoning maps based on different features all at once, paired with their respective spatiotemporal diagrams of behaviour. This allows the analysis of the spatiotemporal patterns of behaviour within each feature, revealing the spatial influence mechanisms based on varied coupling weights. In the end, comparisons of spatiotemporal dense areas among different features reveal the varying degrees of influence of design elements on specific spatiotemporal behaviour patterns. The second method begins with a conceptual understanding of space, leading to the creation of one single zoning geometry, with its corresponding behaviour spatiotemporal diagrams. Then, through iterative adjustments based on certain rules and continuous observation of changes in spatiotemporal behaviour diagrams, this method progressively refines zoning forms to align with specific spatiotemporal behaviour patterns. The results and insights from both methods serve as valuable guidance for architects in spatial design.

5. Conclusion and Discussion

In conclusion, the advent of video tracking represents a paradigm shift in the field of environmental behavioural research. Its ability to capture, quantify, and analyse the intricate behaviours within their natural habitats has elevated our understanding of social dynamics to unprecedented levels. As technology continues to advance, video tracking itself also presents significant potential for further developments. However, unlocking its full capabilities requires researchers to engage in diverse explorations of its operational mechanisms, allowing for the unveiling of nuanced spatial-behavioural patterns across multiple dimensions.

The introduced analytical method, based on the customized tracking tool BodyCAM, opens new possibilities for modelling human behaviour in association with intricate zoning geometry of architectural spaces. This synthesis bridges the realms of advanced behavioural analysis and the classic formal analysis in architecture. As a result, this paper makes a significant contribution by extending both the conceptual and operational understanding of zoning morphology, rooted in the intricate interplay between architecture and behaviour.

The current paper still has limitations. It primarily focuses on the development of the tool and presents a preliminary exploration of the proposed method. It contributes to the theoretical understanding of zoning geometry types and outlines two potential analysis processes for architectural design at the methodological level. Future research endeavours should extend beyond tool development and delve into the practical validation of the method within diverse spatial contexts and design objectives. This validation process will be crucial for unlocking the full potential of the proposed method and ensuring its applicability across practical scenarios, further refining its theoretical foundations and methodological applications.

Acknowledgements

This paper is funded by the Industry-University Collaborative Education Program (202102560006).

References

Albusac, J., Vallejo, D., Jimenez-Linares, L., Castro-Schez, J. J., & Rodriguez-Benitez, L. (2009). Intelligent surveillance based on normality analysis to detect abnormal behaviors. *International Journal of Pattern Recognition and Artificial Intelligence*, 23(07), 1223-1244.

Ambrose, M. A. (2007). Body|Form|Space: Geometric translations of the body in motion. In *Em'body'ing Virtual Architecture: The Third International Conference of the Arab Society for Computer Aided Architectural Design (ASCAAD 2007)*, 28-30 November 2007, Alexandria, Egypt (pp. 431-438).

Arac, A., Zhao, P., Dobkin, B. H., Carmichael, S. T., & Golshani, P. (2019). DeepBehavior: A deep learning toolbox for automated analysis of animal and human behavior imaging data. *Frontiers in systems neuroscience*, 13, 20.

Auguste, A., Oudinet, G., Kaddah, W., Elbouz, M., & Alfalou, A. (2021, April). Implementation of a behavioral analysis method of crowd movement in the service of video surveillance. In *Pattern Recognition and Tracking XXXII* (Vol. 11735, pp. 126-133). SPIE.

Gonzalez Rojas, P. (2017). Space and Motion: Data-Driven Model of 4D Pedestrian Behavior. In *DISCIPLINES & DISRUPTION Proceedings of the 37th Annual Conference of the Association for Computer Aided Design in Architecture (ACADIA)* (pp. 266-273). Cambridge, MA: ACADIA. ISBN 978-0-692-96506-1.

Guo, Z., Wang, X., & Yuan, P. (2019). Sensing Human Behavior in the Built Environment. In *Hello, Culture! Proceedings18th International Conference, CAAD Futures 2019* (pp. 378-388).

Pechlivanidou-Liakata, A., Kerkidou, M. P., Zerefos, S. C., Stamenic, M., Mikrou, T., & Doulgerakis, A. (2010). Spatial Navigational Patterns Induced by Real and Virtual Architectural Environments. In *FUTURE CITIES 28th eCAADe Conference Proceedings* (pp. 755-763). ETH Zurich (Switzerland).

Rezaei, M., & Azarmi, M. (2020). Deepsocial: Social distancing monitoring and infection risk assessment in covid-19 pandemic. *Applied Sciences*, 10(21), 7514.

Torrens, P. M., & Gu, S. (2023). Inverse augmentation: Transposing real people into pedestrian models. *Computers, Environment and Urban Systems*, 100, 101923.

Whyte, W. H. (1980). *The social life of small urban spaces.* Conservation Foundation.

Wu, T., Lei, P., Li, F., & Chen, J. (2022). Space-time tree search for long-term trajectory prediction. *IEEE Access*, 10, 117745-117756.

Yan, W. (2006). Integrating Video Tracking and Virtual Reality in Environmental Behavior Study. In *Proceedings of the 25th Annual Conference of the Association for Computer-Aided Design in Architecture* (pp. 483-488).