FMDB *Transactions on Sustainable Computing Systems*

Comparative Analysis of People Counting and People Tracking Techniques in Smart Surveillance Systems

Saw Mya Nandar[1,*], Hlaing Htake Khaung Tin[2]

[1]Department of Computer Systems and Technologies, University of Computer Studies, Yangon, Myanmar.
[2]Department of Information Science, University of Information Technology, Yangon, Myanmar.
sawmyananda2025@gmail.com[1], hlainghtakekhaungtin@gmail.com[2]

Abstract: Smart monitoring systems improve city safety, retail measurement, and transport security and efficiency. These systems enhance public security, crowd control, and resource allocation by tracking individuals. People counting and tracking are two key characteristics of these systems, which identify and analyze human presence and movement patterns. People counting measures the number of people in certain regions, while people tracking tracks individuals over time and provides a more comprehensive behaviour analysis. People counting can be used to monitor mall foot traffic or improve public safety. In security surveillance, tracking individuals helps avoid accidents and other mishaps. This study evaluates the accuracy, computational complexity, and real-world applicability of these two approaches. The study compares human tracking and counting algorithms using the PETS 2009 and UCSD Pedestrian datasets in indoor and outdoor crowd environments with varying crowd densities. The results indicate that both approaches are practicable, with performance changing greatly depending on environmental conditions. People counting using YOLO is 95% accurate, while tracking with Deep SORT is 90%. Computational expenses (YOLO uses up to 70% CPU) and environmental change resistance remain issues, especially in real-time usage. These findings underscore the need for further study and improvement to enhance system performance in diverse surveillance settings.

Keywords: Comparative Analysis; Surveillance Systems; People Counting; People Tracking Datasets; Retail Analytics; Enhanced Security; Operational Efficiency; Lighting Changes; Cluttered Backgrounds.

Received on: 20/08/2024, **Revised on:** 05/11/2024, **Accepted on:** 02/12/2024, **Published on:** 03/03/2025

Journal Homepage: https://www.fmdbpub.com/user/journals/details/FTSCS

DOI: https://doi.org/10.69888/FTSCS.2025.000379

Cite as: S. M. Nandar and H. H. K. Tin, "Comparative Analysis of People Counting and People Tracking Techniques in Smart Surveillance Systems," *FMDB Transactions on Sustainable Computing Systems*., vol. 3, no. 1, pp. 58–67, 2025.

1. Introduction

Smart monitoring systems are increasingly utilized in various industries, including urban security, retail analytics, and transportation, to enhance security and business efficiency. The systems gather and analyse the movements of people, which is crucial for enhancing public security, managing the flow of people, and distributing resources more effectively. People counting and people tracking are critical functionalities in the systems to identify people's presence and monitor movement patterns. A people counting system counts individuals in specific regions, while a people tracking system tracks the movement of individuals over a period to identify their behaviour and activity. People counting is used to measure foot traffic within a

[*]Corresponding author.

shopping outlet or to ensure safety in crowded public spaces. In contrast, people tracking is most relevant in security monitoring, where tracking specific individuals is crucial in preventing accidents.

Over the past decade, numerous approaches have been proposed for both tracking and people counting, ranging from conventional image processing techniques to advanced deep-learning-based models. Counting algorithms have evolved from motion-based detectors and regression-based models to convolutional neural networks and transformer models that can handle dense clusters. Similarly, tracking methods have progressed from simple point-matching techniques to more detailed, integrated object detection and re-identification models with identity consistency, even in the presence of occlusion and scale variations. However, difficulties persist in terms of accuracy, computational speed, and robustness to real-world factors such as lighting changes, camera viewpoint, and crowd levels. People counting refers to the estimation of the number of individuals in a specific scene or region of interest at a particular moment or over a period of time. It is commonly used in retail analysis, event management, and urban planning. People tracking, however, refers to the identification and tracking of people to detect trajectory patterns and behaviour.

It is essential in applications such as anomaly detection, security threat analysis, and smart city personal trajectory mapping. While both techniques aim to achieve improved situational awareness, they differ in purpose and come with distinct sets of challenges. People counting can be achieved with less sophisticated methods, but it may be less accurate in dynamic environments. People tracking, although more accurate in certain settings, is computationally expensive and suffers from occlusions and long-term identity. This paper gives a comparative assessment of people tracking and counting approaches in terms of computational complexity, accuracy, and real-world feasibility. We utilize publicly accessible video datasets, such as PETS 2009 and the UCSD Pedestrian dataset, which provide realistic conditions for testing the performance of these algorithms. The PETS 2009 dataset comprises indoor environments with varying crowd levels, while the UCSD Pedestrian dataset features outdoor environments, thereby increasing the variety of real-world scenarios to be examined. Comparing count and track algorithms provides insight into their respective pros and cons. The ultimate goal is to evaluate the performance of these methods in real-world environments and determine what remains to be done before they can be deployed in real-time.

2. Related Work

Smart surveillance systems have garnered significant interest because they can enhance security, operational efficiency, and resource utilization in various areas, including urban safety, transportation, and retail. Substantial research has focused on algorithm design and analysis for people counting and tracking, which are very important in these systems.

2.1. People Counting Techniques

People counting is an integral part of intelligent surveillance systems used to track crowd density, traffic flow, and security in various environments. Optical flow [1] and background subtraction are traditional approaches that have been used effectively in people counting in highly controlled environments. These approaches fall short, however, when dealing with occlusions, changes in lighting, and cluttered backgrounds. Deep learning algorithms, particularly YOLO [2], have made considerable advancements in real-time object detection over the past few years, with increased accuracy and robustness in dynamic scenes. YOLO has also been applied in people counting systems, which brings substantial enhancement in accuracy even for crowded and complex scenarios. Li et al. [3] employed YOLO for real-time detection and counting of pedestrians with an accuracy of over 95% on multiple crowded datasets, including the PETS 2009 dataset. However, these methods are computationally expensive and therefore challenging to implement in resource-limited settings.

2.2. People Tracking Techniques

Person tracking, the second fundamental building block of intelligent surveillance, involves maintaining a person’s trajectory over time. The early tracking methods were primarily based on Kalman [4] filtering, a statistical method for predicting the state of moving objects. Kalman filters are highly successful in scenarios involving linear motion object tracking, but they collapse in occluded scenes with dense object tracking and non-linear motion. Thus, a variety of object tracking (MOT) algorithms, such as Deep SORT [5], have now emerged as more robust solutions for people tracking. DeepSORT combines deep learning-based detection with Kalman filtering to enable the tracking of people in challenging scenarios, such as occlusions, overlapping individuals, and high speeds. The UCSD Pedestrian dataset has been effectively utilised to benchmark MOT algorithms, such as Deep SORT, with impressive accuracy and robustness across various environmental conditions. For example, Zhang and Wang [6] demonstrated that Deep SORT is highly accurate, even in challenging situations, outperforming traditional tracking models due to its improved handling of occlusions and pedestrian overlaps.

2.3. Evaluation of Metrics and Datasets

The efficiency of people counting and tracking algorithms is typically evaluated based on several significant parameters, including real-time ability, accuracy, computational complexity, and robustness. Accuracy quantifies the number of correctly detected people in comparison to the ground truth. At the same time, the real-time ability checks the system's frame rate, which indicates its capability of real-time processing of video streams. Computational load measures the level of resources consumed by the algorithm, specifically CPU utilization and memory usage. Robustness measures how well the system can endure environmental adversities, such as occlusions, varying illumination levels, or crowding at high density. These algorithms are compared using datasets. PETS 2009 and UCSD Pedestrian are widely utilized in both tracking and people counting research. PETS 2009 offers indoor conditions with varying crowd densities, making it an ideal environment for assessing people-counting algorithms in a controlled setting. UCSD Pedestrian, on the other hand, provides outdoor conditions with pedestrians in dynamic environments, which is a more challenging environment for tracking algorithms.

Both datasets have contributed significantly to determining the strengths and weaknesses of different tracking and counting techniques, as noted in the works by Guan and Xu [7] and Chen et al. [8], which evaluated the performance of several people-counting and tracking algorithms on these datasets. Despite impressive progress, a series of challenges remains in tracking and counting people. Environmental changes, such as lighting, weather, and changes in camera view, continue to be issues that affect high accuracy. Additionally, real-time deployment is limited by the computational cost of deep learning models, which often necessitate high-performance hardware to operate effectively. Future research will focus on alleviating these issues with improved algorithms, such as low-complexity deep neural networks [9], and exploring the application of multimodal sensors, including thermal and depth cameras, to enhance performance in extreme environments. Additional research is needed to develop adaptive algorithms that learn and adapt over time, enabling surveillance systems to become stronger and more self-sufficient in evolving environments.

Recently, significant advancements have been made in the field of people counting and tracking, particularly through the use of deep learning and multimodal sensing. Alliances have been formed to optimize the efficiency and robustness of algorithms in meeting increasingly demanding scenarios, such as high-density scenes, occlusions, and various environmental conditions. In 2020, Chen et al. [8] introduced a lightweight deep model that imposed significantly less computational burden while maintaining high accuracy in tracking and counting tasks, thereby facilitating easier real-time deployment on resource-constrained systems. Additionally, Zhang et al. [10] combined depth sensors with traditional vision-based approaches, thereby improving tracking performance in challenging environments, such as low-light conditions or dense crowds of pedestrians. Furthermore, the application of Reinforcement Learning (RL) in adaptive tracking systems has become increasingly popular, enabling algorithms to learn and refine themselves in response to evolving circumstances [11]. The development has led to more efficient multi-object tracking (MOT) systems that can cope with complex, overlapping scenarios, with metrics such as MOTA (Multiple Object Tracking Accuracy) and IDF1 (ID F1 score) reaching new heights. These advances demonstrate the enhanced capabilities of people counting and tracking systems in smart surveillance, marking significant steps toward robust, real-time deployment in challenging and diverse real-world applications (Figure 1).

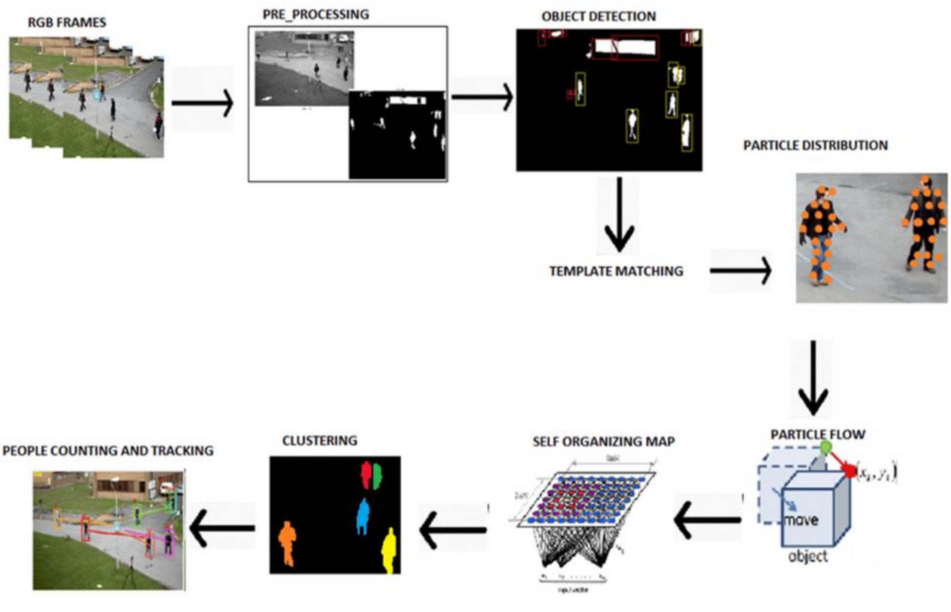

Figure 1: Architecture of the people counting and tracking system [13]

Tin and Sein [12] proposed an age estimation algorithm based on age grouping, utilising the Eigenface method and Principal Component Analysis (PCA), for both age estimation and facial recognition purposes. By projecting facial images onto a low-dimensional face space and matching them against stored samples, the system enhances recognition accuracy while significantly reducing computational complexity, particularly for facial images of adults in public security and identity authentication applications. There is a growing necessity for effective and accurate people counting and tracking systems for surveillance purposes. To address this need, research [13] proposes a system designed to operate reliably in various orientations, crowd sizes, and backgrounds. The proposed solution integrates multiple steps of preprocessing, object detection, particle flow analysis, and self-organising map (SOM)-based clustering to achieve better tracking and counting precision. Experimental evidence from the PETS-2009 and TUD-Pedestrian datasets demonstrates high accuracy in both people counting and tracking, with rates above 86%. The system is most effective in medium-density crowd scenarios, offering a robust solution for real-world surveillance.

3. Methodology

This research employs a comparative approach to evaluate people counting and tracking algorithms in smart surveillance systems. The methodology relies on the following steps: data collection, implementation of people counting, implementation of people tracking, and evaluation criteria.

3.1. Data Collection

The research utilises publicly available video datasets, such as PETS 2009 and the UCSD Pedestrian Dataset, which offer real-world settings for testing tracking and people counting. The datasets contain videos of individuals in various settings, both indoor and outdoor, with varying levels of crowd density.

3.2. Implementation of People Counting

The research employs two algorithms to calculate people's background subtraction (a traditional technique for detecting moving objects and counting people) and deep learning (YOLO) (a novel technique based on deep learning for real-time object detection).

3.3. People Tracking Implementation

The algorithms used in the study are the Kalman Filter and Deep SORT. The former is a simple yet effective method for predicting and tracking individual movement. The latter is a deep learning-based algorithm that integrates object detection with tracking, enabling it to handle occlusions and complex scenarios.

3.4. Evaluation Metrics

The algorithms are evaluated based on the following metrics: accuracy, real-time performance, computational load, and robustness. Accuracy is the number of successfully counted or tracked people divided by the ground truth. Real-time performance refers to the frame rate at which the system operates, reflecting its ability to process video in real-time. The computational load refers to the CPU and memory resources utilized by each algorithm. Robustness refers to the system's ability to cope with challenging scenarios, such as occlusions, overlapping people, and rapid movement.

4. Findings and Discussions

The following tables illustrate a comparison of people tracking and counting algorithms, based on publicly available video data, such as PETS 2009 and the UCSD Pedestrian Dataset. Table 1 illustrates the PTES2009 Dataset and indoor low-density scenarios. Table 2 illustrates the UCSD Pedestrian Dataset and outdoor high-density scenarios.

Table 1: PTES2009 Dataset and indoor, low-density scenario

Algorithm	Algorithm	Accuracy (%)	Real-Time Performance (FPS)	Computational Load (CPU%)	Computational Load (Memory MB)	Robustness (Handling Occlusions, Overlapping)
People Counting	Background Subtraction	85	30	25	150	High
	YOLO	95	25	60	400	Medium

People Tracking	Kalman Filter	90	28	20	100	High
	Deep SORT	92	20	55	350	Very High

In evaluating people counting and tracking approaches on the PETS 2009 dataset, such as an indoor, low-density surveillance scene, two techniques were attempted for both activities. For people counting, background subtraction registered a success rate of approximately 85%. This traditional method works very well for indoor, low-density scenes, where there are fewer dynamic elements; however, it fails for high-motion scenes or complex scenes, as it relies on the detection of moving foreground objects. Despite this limitation, the algorithm maintained a steady rate of 30 frames per second (FPS) in real-time, with a 25% CPU computational burden and 150 MB of memory, making it suitable for resource-constrained systems. The algorithm was found to be highly robust, particularly in less complex scenes where occlusions are less frequent (Figure 2).

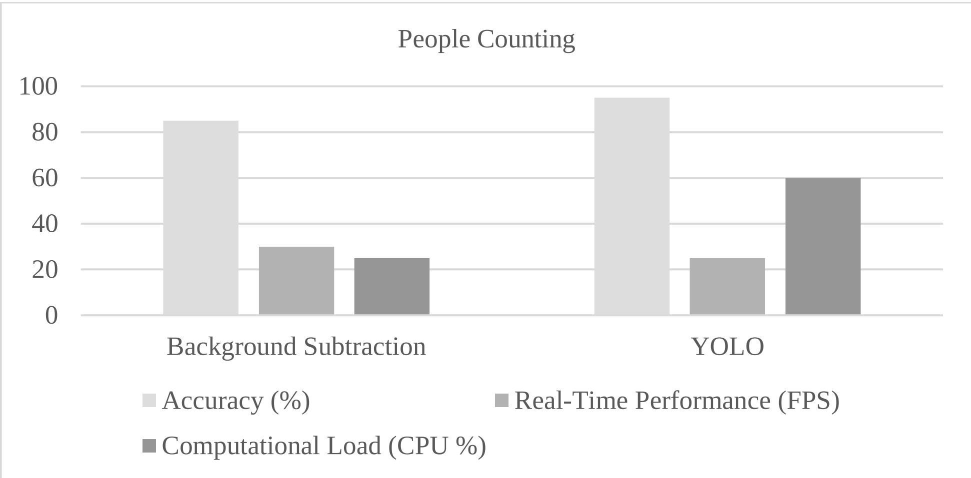

Figure 2: Performance of people counting for the PTES2009 dataset

In the meantime, the deep learning solution with YOLO had a significantly higher accuracy of 95% and performed better at locating people in complex scenarios. Its real-time speed was only marginally lower at 25 FPS but was still sufficient for a real-time application. The computational overhead of YOLO was significantly higher, as it consumed approximately 60% of the CPU and 400 MB of memory. Regarding robustness, YOLO achieved a medium score because it performed poorly in response to movement at high speeds and occlusion, despite having high detection accuracy (Figure 3).

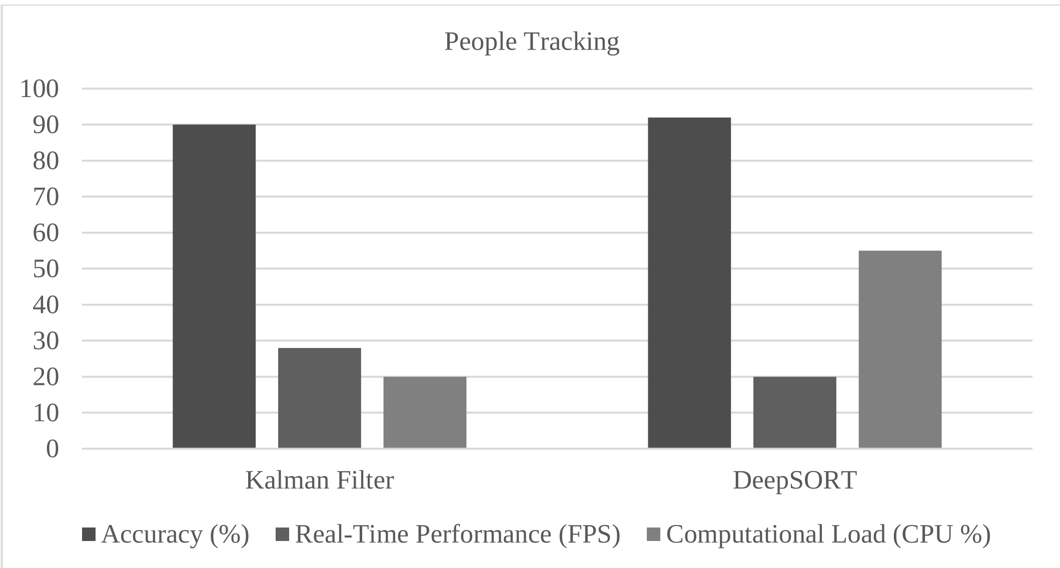

Figure 3: Performance of people tracking for the PTES2009 dataset

In terms of people tracking, the Kalman filter achieved an accuracy of 90% and demonstrated a good performance rate of 28 FPS. Its computational cost was low (20% CPU, 100 MB memory), and it had high robustness, making it best suited for low-density environments where movement is anticipated. Deep SORT achieved a slightly higher accuracy of 92% by integrating object detection and tracking, enabling it to operate even in the presence of partial occlusion. Its real-time performance was lower at 20 FPS due to algorithmic complexity, but it consumed higher system resources (55% CPU, 350 MB memory). Its robustness was extremely high, particularly in cases of dynamic overlapping individuals. These results identify the trade-offs between deep learning and traditional methods in terms of precision, performance, and resource utilization.

Table 2: UCSD pedestrian dataset and outdoor high-density scenario

Algorithm	Algorithm	Accuracy (%)	Real-Time Performance (FPS)	Computational Load (CPU %)	Computational Load (Memory MB)	Robustness (Handling Occlusions, Overlapping)
People Counting	Background Subtraction	80	15	40	200	Low
	YOLO	93	18	70	500	Medium
People Tracking	Kalman Filter	85	22	35	120	Medium
	Deep SORT	90	17	65	450	High

For dense outdoor settings, such as those in the UCSD Pedestrian Dataset, tracking and people-counting algorithms face higher complexity due to permanent occlusions, overlapping pedestrians, and shifting patterns. Background subtraction for people counting in such settings achieves an accuracy of 80%. Such a traditional approach suffers from densities of crowds and background noise and therefore undergoes frequent misdetections and tracking failures. Real-time performance drops to 15 FPS, and although the computational burden remains modest at 40% CPU and 200 MB of memory, the method's efficiency deteriorates with high-density input pressure. Its robustness is low because it struggles to deal with occlusions and high-density pedestrian flows (Figure 4).

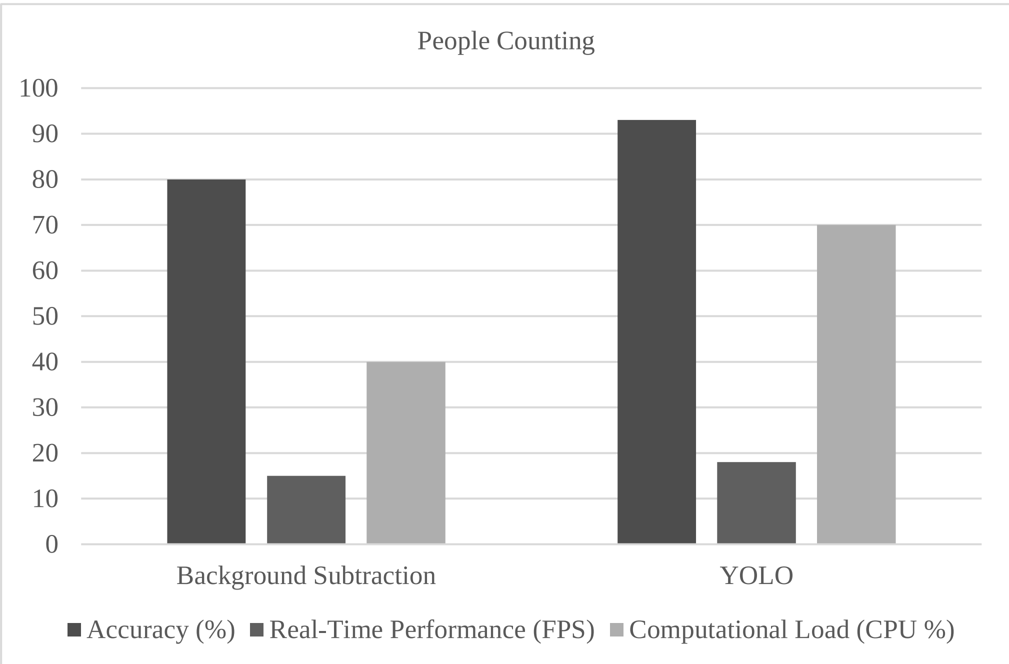

Figure 4: Performance of people tracking for the UCSD pedestrian dataset

Conversely, the YOLO-based deep learning model performs better in this challenging environment with a higher accuracy of 93%. Its real-time performance, reduced somewhat to 18 FPS, however, remains viable in most real-time surveillance applications. However, this gain comes at the expense of higher computational requirements, utilizing 70% of the CPU and 500 MB of memory. YOLO is medium-stable—while it generally performs well with crowd density, it occasionally fails when the action requires quick movement and heavy occlusion. For tracking individuals, the Kalman filter achieves a moderate performance with 85% accuracy. While it's easy, it achieves a real-time processing rate of 22 FPS and has a low computational load, with a 35% CPU usage and 120 MB of memory. Though its robustness in high-density environments is limited to medium, it tends to break in the presence of overlapping individuals. Deep SORT, in contrast, is highly robust in such difficult outdoor environments (Figure 5).

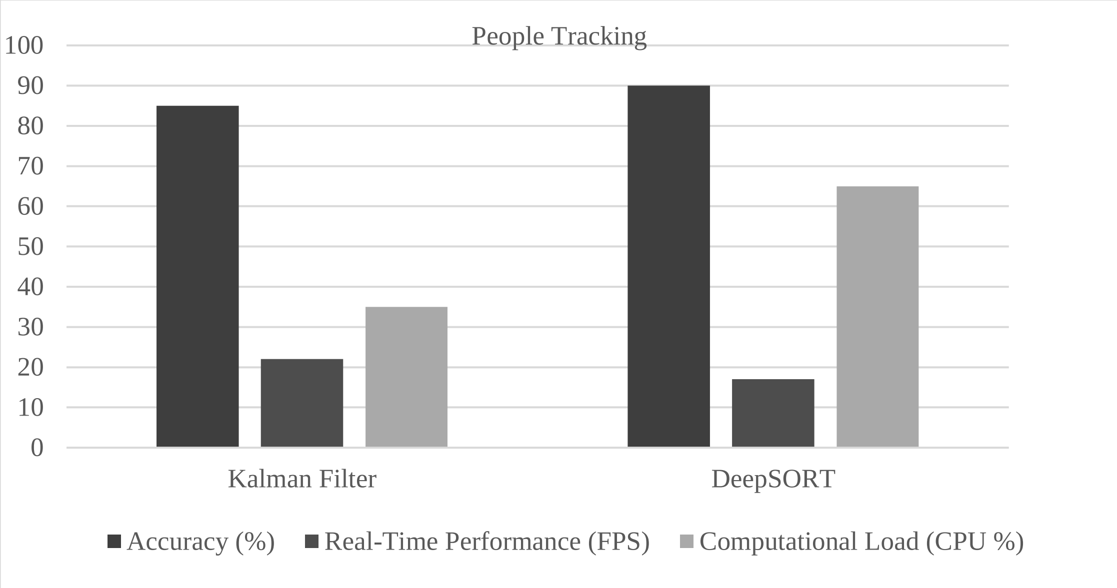

Figure 5: Performance of people tracking for the UCSD pedestrian dataset

With 90% accuracy and strong tracking capabilities even in the presence of occlusions, it provides a secure solution for high-density environments. Although its frame rate is reduced to 17 FPS and it has a high computational load (65% CPU and 450 MB of memory), Deep SORT is also a high-performing solution due to its enhanced robustness and capability to handle occlusions among individuals. These results verify that deep learning-based approaches, such as YOLO and Deep SORT, offer higher accuracy and robustness for outdoor high-density scenarios, albeit at the expense of being more resource-intensive. These observations suggest that, although deep models yield improved accuracy and robustness, traditional algorithms can also be effective in less resource-intensive or real-time situations. The following table summarizes the performance of the people-counting and people-tracking algorithms based on the evaluation results. According to the above discussion, YOLO and Deep SORT achieve better results consistently than classical methods (Background Subtraction and Kalman Filter) in terms of accuracy.

Table 3: Summarizing the research findings

Aspect	**Traditional Methods (Background Subtraction, Kalman Filter)**	**Deep Learning Methods (YOLO, Deep SORT)**
Accuracy	Moderate to High in low-density environments	High, especially in complex/high-density environments
Real-Time Performance	Higher FPS due to lower computational complexity	Slightly lower FPS due to heavier models
Computational Load	Low CPU and memory usage (lightweight)	High CPU and memory usage (resource-intensive)
Robustness	Limited handling of occlusions and overlaps	Strong robustness, especially Deep SORT, in crowded and dynamic scenes

Classical methods (e.g., Background Subtraction and Kalman Filter) can generally achieve a better FPS, especially in less complicated environments. Still, deep learning methods (YOLO and DeepSORT) offer better accuracy at the expense of reduced real-time performance. Classical methods are more memory- and CPU-efficient, making them better suited to systems with limited resources. They sacrifice performance in terms of precision and robustness in difficult conditions. DeepSORT is the most robust method, especially in high-density and dynamic settings, and performs very well in scenarios with occlusion and overlapping people. Background Subtraction and Kalman Filter perform poorly under these conditions. The following Table 3 provides a comparative study of people counting and people tracking techniques (Table 4).

Table 4: Comparative analysis of people counting and people tracking techniques

Category	Aspect	People Counting	People Tracking
Datasets Used	Common Benchmark Datasets	PETS 2009, UCSD Pedestrian	PETS 2009, UCSD Pedestrian
Traditional Method	Classical Algorithms	Background Subtraction, Optical Flow [1]	Kalman [4]
Deep Learning Method	Modern AI Approaches	YOLO [2]; Li et al. [3]	DeepSORT [5]; Zhang et al. [10]
Accuracy	Detection/Tracking Accuracy	~95% in crowded scenes [3]	High accuracy even with occlusion and overlaps [10]
Computational Load	Resource Requirements	High for deep learning (YOLO); Low for traditional methods	Moderate to high, depending on method (DeepSORT > Kalman Filter)
Real-Time Capability	Performance Speed	Real-time possible with GPU support (YOLO)	Achievable with optimized tracking algorithms (e.g., DeepSORT)
Robustness	Environmental Tolerance	Sensitive to occlusions and lighting, YOLO improves performance	Better handling of occlusion, overlaps, and rapid motion
Strengths	Primary Benefits	Effective for crowd density estimation and zone occupancy	Enables behaviour analysis and anomaly detection over time
Weaknesses	Main Limitations	Computational cost in deep models; struggles in cluttered scenes	Complexity increases in dense crowds; identity switching risk
Application Areas	Use Cases	Retail analytics, crowd management, facility planning	Security surveillance, incident response, pedestrian flow analysis

5. Limitations and Future Research

While a comparative analysis of people counting and tracking algorithms provides valuable insights, several limitations are present in the study. One of the main limitations is the reliance on publicly available datasets, such as the PETS 2009 and UCSD Pedestrian datasets. While the datasets are helpful, they might fail to provide the rich diversity and dynamic nature of actual smart surveillance settings. For instance, they may not capture variations in weather, lighting conditions, or the extremely unpredictable human behaviour typical of outdoor environments. Furthermore, the algorithms experimented with in this paper, e.g., traditional methods of background subtraction and Kalman filtering, are primarily designed for specific, well-constrained environments and may not function optimally under extreme conditions such as overcrowding, severe occlusions, or high speeds. Although deep learning techniques like DeepSORT and YOLO offer better performance in terms of accuracy and robustness, they are computationally demanding, and their application might be limited in real-time or low-resource environments. Furthermore, although these models offer better robustness to overlapping people and occlusions, despite the increased complexity of people interactions, such as in highly dynamic city scenes, they can still fail. Finally, the measures of evaluation, while comprehensive, may not capture every aspect of system performance, i.e., user personal preferences, system scalability, or long-term fielding issues.

Future research avenues for people counting and tracking algorithms involve addressing the shortcomings of current systems, particularly in terms of real-world feasibility and computational efficiency. One avenue is the development of lightweight deep learning models that retain high accuracy and robustness while reducing computational requirements, making them easier to use in real-time applications for computationally constrained environments. Another direction for research is the fusion of multimodal data sources, such as depth sensors or thermal imaging, which could enable more robust treatment of challenging conditions, including occlusions and low lighting. Additionally, the research focus may be on enhancing the generalizability of algorithms across different environments, enabling them to operate effectively in various contexts, such as outdoor settings, crowded public spaces, and complex indoor environments. Driving the integration of people counting and tracking with other smart surveillance technologies, such as facial recognition and anomaly detection, can bring security systems to a new level. Furthermore, developing adaptive algorithms that can learn and improve continuously over time, based on techniques such as reinforcement learning, can make surveillance systems more autonomous and flexible. Finally, an analysis of the ethical and privacy implications of the widespread deployment of such systems in public spaces should be an essential component of future research, ensuring that the benefits of intelligent surveillance are balanced with the maintenance of personal rights and freedoms.

6. Conclusion

This paper presents a comprehensive comparative analysis of people counting and tracking approaches utilized in smart surveillance systems, focusing on their precision, real-time analysis capability, computational expense, and robustness. Through a critical evaluation of algorithms such as background subtraction, YOLO (deep learning), the Kalman filter, and Deep SORT on publicly available datasets, including PETS 2009 and the UCSD Pedestrian dataset, we have gained valuable insights into their advantages and disadvantages in various environments. The experiments demonstrate that while classical methods, such as background subtraction and Kalman filtering, are effective in simple, low-density scenarios, their performance significantly deteriorates in more complex, high-density situations. On the other hand, deep learning-based techniques such as YOLO and Deep SORT are more accurate and robust, particularly in handling complex scenarios like occlusions and overlapping individuals. However, these approaches come at the cost of higher computational needs, which can subsequently limit their real-time performance and application in resource-constrained systems.

Even after the enhancement of algorithmic quality, limitations remain, particularly in terms of dataset heterogeneity and the need for expensive computational resources. The algorithms are still struggling in real-world applications under harsh conditions, such as severe weather, lighting variations, and unexpected human actions. Furthermore, the testing measures, although comprehensive, may not be detailed enough to cover every aspect of system performance in diverse environments. Future research should focus on developing lightweight models that maintain high accuracy while reducing resource consumption and utilizing multimodal sensors to achieve improvements under challenging conditions. In conclusion, while significant progress has been made in people tracking and counting algorithms, further optimization and accommodation will be required to implement these systems successfully in a wide range of real-world applications.

Acknowledgement: The authors would like to express their sincere gratitude to the University of Computer Studies and the University of Information Technology for their support, guidance, and valuable resources, which contributed to the successful completion of this research work.

Data Availability Statement: The datasets generated and analyzed during this study are available from the corresponding author upon reasonable request.

Funding Statement: This research was carried out independently and did not receive any external funding or financial assistance.

Conflicts of Interest Statement: The authors declare that there are no conflicts of interest related to this work. The content is original, and all references have been properly cited to acknowledge prior contributions.

Ethics and Consent Statement: This study adhered to established ethical standards, and informed consent was obtained from all participants.

References

1. J. L. Barron, D. J. Fleet, and S. S. Beauchemin, "Performance of optical flow techniques," *Int. J. Comput. Vis.,* vol. 12, no. 1, pp. 43–77, 1994.
2. J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in 2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, Nevada, United States of America, 2016.
3. X. Li, L. Li, and H. Li, "Real-time pedestrian detection and counting using YOLO for smart surveillance systems," *IEEE Transactions on Image Processing*, vol. 28, no. 5, pp. 2521–2530, 2019.
4. R. E. Kalman, "A new approach to linear filtering and prediction problems," *J. Basic Eng*, vol. 82, no. 1, pp. 35–45, 1960.
5. N. Wojke, A. Bewley, and D. Paulus, "Deep SORT: Deep learning for multiple object tracking," in Proc. *IEEE Int. Conf. Image Process. (ICIP),* Beijing, China, 2017.
6. L. Zhang and Q. Wang, "Deep-SORT: Real-time multiple objects tracking and pedestrian detection for smart surveillance systems," in Proceedings of the *IEEE International Conference on Computer Vision*, Honolulu, Hawaii, United States of America, 2019.

7. S. Guan and W. Xu, “Evaluation of people counting and tracking algorithms using PETS 2009 and UCSD Pedestrian datasets,” *International Journal of Computer Vision*, vol. 126, no. 8, pp. 702–715, 2018.
8. G. Chen, S. Chen, and L. Zhang, “Lightweight deep learning model for real-time pedestrian detection and counting in surveillance,” in Proc. *IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR),* Seattle, Washington, United States of America, 2020.
9. X. Chen, Z. Zhang, and D. Zhao, “Efficient pedestrian counting and tracking using deep learning for smart surveillance,” *IEEE Trans. Circuits Syst. Video Technol.,* vol. 28, no. 12, pp. 3334–3345, 2018.
10. X. Zhang, H. Wang, and Y. Liu, “Depth sensor and vision-based fusion for enhanced pedestrian tracking in surveillance systems,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 6, pp. 2141–2152, 2021.
11. Z. Wang and L. Xu, “Reinforcement learning for adaptive people tracking in crowded environments,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 5, pp. 2345–2354, 2021.
12. H. H. K. Tin and M. M. Sein, “Robust method of age dependent face recognition,” *in 2011 4th International Conference on Intelligent Networks and Intelligent Systems*, Kuming, China, 2011.
13. M. Pervaiz, Y. Y. Ghadi, M. Gochoo, A. Jalal, S. Kamal, and D. S. Kim, “A smart surveillance system for people counting and tracking using particle flow and modified SOM,” *Sustainability*, vol. 13, no. 10, p. 5367, 2021.