ISSN: 2073-9524
Pages:35-51

A Comprehensive Survey of Advanced Surveillance Techniques for Theft Detection in Retail Environments: Integrating Multi-Object Tracking and Human Pose Estimation

Ali Salah Hameed [1*], Taha Mohammed Hasan [2], Rokan Khaji [3]
[1,2,3] Computer Science Department, College of Science, University of Diyala, Baquba, Iraq, 32001.
scicomphd232402@uodiyala.edu.iq[1], dr.tahamh@sciences.uodiyala.iq[2], rkhaji2@gmail.com[3]

Abstract
This paper discusses recent advancements in theft surveillance techniques, including tracking multiple objects and estimating human movements. These technologies help with real-time monitoring, understanding behaviors, and ensuring public safety. The major computational challenges of the techniques have been looked at from different angles, including balancing accuracy with efficiency, occlusion management, and scalability in different environments. It also identifies areas that require further research and proposes potential future directions, such as incorporating diverse data types and developing compact and adaptable models. This work is meant to be a helpful tool for researchers and practitioners, giving them a look at the present status of these technologies and steering the creation of new smart systems.

Keywords: Human Pose estimation, Yolo, Object Tracking, Machine Learning

***Article history: Received: 04 Jan** 2025**; Accepted:** 21 Feb 2025**; Published:** 15 Mar 2025*

1. Introduction

Modern technologies have completely transformed the field of surveillance and safety, as these technologies have addressed major challenges in the field of behavioral monitoring, especially in the field of theft prevention. Surveillance technologies greatly help in monitoring locations with large crowds that are difficult to monitor using traditional methods [1].

Old surveillance methods, represented by the use of fixed cameras, are an essential element in combating suspicious behaviors, including theft in stores; thus, they play a major role in deterring potential shoplifters, enhancing safety among shoppers, and enhancing security operations. However, conventional surveillance mostly depends on human effort; therefore, this will result in a significant discrepancy between several cameras and will cause theft not to be caught in real time as humans would become fatigued and lose attention. Therefore, relying solely on human judgment may lead to erroneous conclusions when identifying suspicious activity [2].

As the number of objects to monitor in a video rise, it becomes increasingly challenging for humans to track many moving elements. Conversely, computer vision can identify and monitor numerous entities with exceptional precision. This is referred to as multi-object tracking. Tasks requiring overcoming several challenges are item detection and tracking. Apart from identifying the position and type of objects, detection systems have to deal with obstacles such as size variance, illumination changes, occlusions, and variations in form and appearance. It also has to

* **Corresponding Author**: scicomphd232402@uodiyala.edu.iq

 ISSN: 2073-9524
Pages:35-51

consider how to classify objects with different degrees of resemblance. Simultaneously, tracking algorithms must address emerging challenges such as monitoring rapid-moving entities, managing transient occlusions, and following objects [3].

A multitude of research studies have been conducted that utilize and suggest revolutionary designs and methodologies in computer vision, including Region-Convolutional Neural Network (R-CNN), Fast R-CNN, Faster R-CNN, and You Only Look Once (YOLO). These distinctive methodologies are often developed by major technology companies like Facebook, Microsoft, Google, and others. The current research landscape emphasizes the significance of computer vision. Numerous research studies are employed in object detection and tracking, which are inherently interconnected, as object tracking relies on object detection [4].

The Internet of Things (IoT) represents a significant technical advancement in security surveillance, facilitating the automatic communication and data exchange across diverse smart devices. These gadgets encompass security cameras, sensors, and alarm systems, augmenting businesses' capacity to monitor their surroundings with increased efficacy [5].

IoT technologies enhance incident response by providing immediate notifications upon the detection of anomalous activity, facilitating a swift reaction from security personnel. Furthermore, the data gathered from these devices allows the examination of individual activities, aiding in the identification of security patterns and trends. Automation and intelligent technologies contribute to lowering operating expenses by diminishing the necessity for a substantial workforce in surveillance activities [6].

Machine learning methods improve the precision of detecting suspicious actions, hence decreasing the incidence of false alarms. The implementation of IoT in security surveillance presents privacy and security issues since interconnected devices may be susceptible to hacking, necessitating stringent security protocols to safeguard sensitive data and information [7].

Human pose estimation is an important step in enhancing security surveillance by detecting key points and analyzing human behavior, especially when combined with other surveillance technologies; thus, it detects irregular activities and predicts threats. The use of surveillance methods has become crucial for retailers seeking to protect their assets, ensure client safety, and sustain operational efficiency in an increasingly competitive retail environment. Human pose estimation and motion tracking are important areas in computer science and computer vision [8].

They help us understand how people move and act in different situations. These approaches depend on the analysis of data derived from video clips, facilitating systems to reliably identify various human stances and monitor their movements [9].

The practical uses of these approaches span several industries, including healthcare, where they facilitate patient monitoring and movement analysis for treatment and rehabilitation. Moreover, the domain of security and surveillance uses these methodologies to identify suspicious activity and assess anomalous behavior [10]. The technology used includes machine learning and deep neural networks, which improve the accuracy of predictions by analyzing large amounts of data. Contemporary technology, including 3D cameras and sensors, enhances the efficacy of motion tracking systems. Research in this subject is now advancing breakthrough technologies that improve human-machine interaction and offer realistic experiences [11].

The study provides a comprehensive examination of these interconnected technologies, analyzing their individual functions and their collective contribution to the development of new surveillance systems. This effort aims to consolidate recent advancements and identify key concerns to guide the development of more intelligent, effective, and proactive safety solutions.

2. Related Work

Major steps forward in watching people, finding objects, and figuring out body positions have been made over time, driven by the need for alert and smart safety systems. The mix of IoT, deep learning methods, and sensors has been key in fixing the flaws of old watching methods. In the following subsections, we will review the most important technologies used in this field and explain their contribution briefly and clearly in Tables 1, 2, and 3. Using modern methods like Multi-Object Tracking (MOT) and Human Pose Estimation (HPE) greatly enhances the effectiveness of surveillance systems. Budget-friendly cameras can achieve this without sacrificing quality.

This affordable method uses computer technology skills to improve algorithms, achieving a balance between low cost and high performance while encouraging new ideas in surveillance technology. The tables below offer the techniques under examination, including MOT, HPE, and theft detection. Each article offers an in-depth précis of the methodology, packages, and accomplishments related to those technologies. The tables offer a comprehensive assessment, showcasing the strengths and trends of each method and illustrating their mixed impact on improving surveillance systems.

2.1 Surveillance Systems

The theft surveillance system has notably been boosted by the implementation of IoT-enabled systems and smart sensors, which can quite markedly enhance the ability to monitor [12]. Previously, conventional surveillance systems like CCTV cameras acted as passive tools in capturing footage of incidents after they had taken place. However, the present-day approach has shifted to proactivity. This includes human and object detection, motion sensors, and real-time alerts that can be sent via GSM or email, thereby making it possible to respond quickly to unauthorized access [13]. To top it all off, innovations like wireless cameras, night vision technology, and better storage options in DVRs and NVRs have made modern surveillance systems much more scalable, efficient, and reliable [14].

2.2 Multiple Object Tracking (MOT)

Multi-object tracking (MOT) has become an essential tool in computing, especially in the areas of passenger monitoring and security. Algorithms such as YOLO, Faster R-CNN, and Deep SORT provide efficient and effective ways to manage multiple objects in changing environments [15]. These techniques effectively handle challenges such as occlusion, flicker, and frame recognition, ensuring reliable performance in a variety of conditions. Furthermore, studies suggest that we can simultaneously improve tracking performance [16].

These improvements are essential for achieving effective real-time containment in large and complex environments, where traditional methods cannot always maintain the required accuracy and speed [17].

2.3 Human Pose Estimation (HPE)

Human pose evaluation (HPE) is useful for post-observation behavior analysis. The design of HPE algorithms aims to identify people's body parts and enhance the accuracy of pose determination in crowd management and violence detection [18]. This makes them useful for advanced surveillance packages. Advanced frameworks, such as HRNet and Pose Flow, use timestamps and active learning methods to handle problems such as motion, occlusion, and real-time needs very well [19].

2.4. Integrated Systems

The use of modern techniques such as Multi-Object Tracking (MOT) and Human Pose Estimation (HPE) makes surveillance systems much more effective, especially when low-cost cameras are used without sacrificing performance. Using computer science skills to improve algorithms is a cheap way to do this that strikes a balance between low cost and high functionality while also encouraging new developments in surveillance technologies [20].

In the work of Kukreja et al. [21], the authors proposed a model that can improve vehicle detection in public environments, thus enhancing security. The researchers also suggest improving the structure of the model to achieve better performance in terms of efficiency and accuracy, which will reduce the false positive rate and make it suitable for use in security applications. The study also discusses the possibility of integrating this modified model with deep learning techniques and discusses the development of a model based on YOLO to make it more efficient in overall performance. Whereas, the work of Manasa et al. [22] delineates the advancement of a real-time augmented system for identifying infractions that transpire whilst driving, encompassing various transgressions such as contravening traffic flow and exceeding speed limits.

This paper showed that the system is 96 percent accurate and can analyze data in less than 100 milliseconds per frame, making it suitable for monitoring traffic. The research article by Tang et al. [23] focuses on improving a model that tracks people and evaluates their body positions from different angles to build 3D environments. The researchers employed sophisticated ways to evaluate data from several cameras, enhancing the precision of object tracking in intricate situations. The researchers suggested combining human tracking with additional methods to attain precise outcomes in the creation of 3D sceneries. This technology achieved a 90 percent accuracy rate in tracking individuals. The paper presented by Li et al. [24] has developed an integrated approach that tracks more than one object using cues related to shape and appearance. The proposed approach achieved more than 80% accuracy in addition to reducing false positives. The contribution of this study was to improve the techniques related to tracking by combining motion information with appearance characteristics techniques.

This leads to the hybrid model performing well in difficult conditions. However, it faces difficulty in tracking objects in very crowded environments and requires high-quality input data. The research article provided by Larasati et al. [25] introduced a novel methodology in the domain of computational vision pertaining to quantum computing. The research article examined how quantum computing enhances algorithms for pattern recognition and image processing. In certain instances, this novel strategy demonstrated an approximate 50% increase in performance relative to conventional methods. Notwithstanding the obstacles confronting these technologies, especially quantum computing, this progress will facilitate continued advancement and enhancement in this domain. The research paper presented by Hassan et al. [26] recorded an accuracy of 70–90 percent for modern methods that focused on deep learning and motion analysis techniques. A research article by Chen et al. [27]. describes a system that alerts users when they stray from their designated lane. The system uses image analysis and a technique called semantic segmentation. The technology achieved a detection accuracy of up to 95 percent. This outcome will enhance the development of systems pertaining to road safety. The research article provided by Bullinger et al. [28] examined a system that monitors various items in real time. The accuracy of the provided model attained 85%, demonstrating the system's efficacy under various scenarios.

This approach contributed to the real-time analysis of things. This contribution demonstrates exceptional efficacy in monitoring overlapping objects. We can also use this approach for motion analysis and monitoring. Sakkos et al. [29] employed a technique that eliminates the background from films with 3D convolutional neural networks. This approach attained an accuracy of up to 90% in detecting moving objects against the backdrop. This precision illustrates the system's efficacy in real-time video clip analysis. The article proposes enhancing performance in challenging settings and intricate backdrops. Grant et al. [30] investigated cluster activity with a technique referred to as light flow in this study. The suggested study attained an accuracy of up to 85% in identifying and categorizing the behaviors of various clusters.

This work contributed to the knowledge of cluster motions and proposed the application of these strategies in security and transportation. A report by Mampilayil and Rahamathullah [31] revealed that deep learning methods were used to identify issues with three-wheeled vehicles, achieving an accuracy of 92%. This article enhances traffic control, offering valuable tools for relevant agencies to better traffic management. The research article provided by Suttiponpisarn et al. [32] focused on the creation of a model that identifies reverse driving using closed-circuit cameras. This methodology utilized image processing and machine learning methods, enabling the model to recognize cars travelling in the opposite direction in real time. This study's contribution improves road safety and decreases accidents caused by improper driving. All of this facilitates the advancement of traffic systems.

The research study authored by Li et al. [33] examined the application of a Kalman filter for multi-object tracking. This work introduced a beneficial paradigm for tracking the accuracy of these items across various situations. This work contributes by integrating motion and measurement data; hence, it enhances computer vision systems. Consequently, this enhances the domains of security and monitoring. Zhao et al. [34] authored a research article detailing their application of correlation filters to enhance object-tracking technology for autonomous vehicles, hence increasing asset tracking accuracy in congested environments.

This will subsequently improve the efficacy of the autonomous transportation system, particularly through the integration of artificial intelligence and computer vision technology. The study provided will be significant in enhancing the efficacy of smart automobiles. Liang et al. [35] conducted a research study employing a system for vehicle counting utilizing deep learning and multi-object tracking methodologies. The researchers illustrated the integration of machine learning and image processing methods to quantify the number of cars in various settings. This invention will enhance the comprehension of using artificial intelligence to attain sustainable transportation efficiency. The researchers Jean-Philippe Jodoin et al. [36] propose a cohesive methodology that combines machine learning with computer vision techniques. This work is significant because of its precise monitoring of items in urban settings, which enhances safety and transit efficiency. Zhang et al. [37] present an innovative method for tracking multiple objects by integrating motion data with attributes acquired from training deep neural networks. This amalgamation will enhance precision, particularly in intricate situations.

The tables below offer the techniques under examination, including MOT, HPE, and theft detection. Each desk offers an in-depth précis of the methodology, packages, and accomplishments related to those technologies. The tables offer a comprehensive assessment, showcasing the strengths and trends of each method and illustrating their mixed impact on improving surveillance systems.

Table 1 Techniques and Applications of Multi-Object Tracking (MOT).

Authors	Contribution	Performance
Kukreja et al. [21]	Proposed automation of traffic law enforcement using real-time video input and YOLO for object detection. Vehicles were categorized by lane, and violators' images were stored on Firebase.	Achieved over 95% accuracy with a fully automated system.
Manasa et al. [22]	Developed a real-time system for detecting wrong-way and clearway violations using video input and security lane detection. Violating vehicles were marked red.	Ensured reliable performance even in severe weather without using YOLO.

Tang et al. [23]	Proposed an improved tracking-by-detection strategy for multi-person tracking with a new appearance model, hierarchical trackers, and template-based classification of trackers.	Reduced false detections and improved overall tracking performance.
Li et al. [24]	Proposed a new MOT framework ("Flow-Tracker") using an IoU tracker, optical flow network, auxiliary tracker, and cascade matching strategy for missing detections.	Achieved an AP of 30.87, the highest accuracy across categories, and a trade-off between speed and precision.
Larasati et al. [25]	Developed a quantum computing-based approach using Yolov5, DeepSORT, Kalman filter, and Hungarian algorithm for optimization. Reduced mismatches and identity switching.	Improved MOTA by 16.03%, MOTP by 5.49%, and F1 score by 6.09%, outperforming similar models without quantum computing.
Hassan et al. [26]	Conducted a systematic literature review of multi-object tracking (MOT) in traffic environments, summarizing techniques, hardware, datasets, metrics, and research gaps.	Provided a comprehensive resource for MOT research until 2022, aiding in identifying new research directions.
Chen et al. [27]	Reviewed lane line departure warning systems, image processing algorithms, and semantic segmentation methods for lane detection. Evaluated systems using benchmarks and test cases.	Identified five major reliability problems and proposed techniques to improve accuracy and precision in LDWS, while outlining future research directions
Bullinger et al. [28]	Proposed an online multiple-object tracking approach using semantic instance segmentation and optical flow cues to track objects at the pixel level.	Demonstrated improved tracking of objects with high relative motions compared to SORT on the MOT 2D 2015 test dataset.
Sakkos et al. [29]	Proposed a temporal-aware, end-to-end background subtraction model using 3D convolutional neural networks (CNNs) for tracking foreground movement.	Effectively tracked the movement of the foreground.
Grant [30]	Developed an automated system for detecting unusual highway behaviors, such as sudden lane changes and wrong-way driving, using Lucas-Kanade optical flow.	Successfully detected vehicles moving in the wrong direction using dense optical flow.
Mampilayil and Rahamathullah [31]	Proposed a cost-effective, sensor-free system for detecting one-way traffic violations using vehicle trajectory points with OpenCV and TensorFlow.	Efficiently managed traffic in one-way areas and deployed easily without additional hardware.
Suttiponpisarn et al. [32]	Presented the LDVC framework for wrong-way detection using CCTV, incorporating road lane boundary detection and direction validation. Achieved high accuracy on both PC and embedded systems.	Achieved 95.23% accuracy on PCs and 94.66% on embedded systems, with potential for edge device deployment.
Li et al. [33]	Proposed an algorithm addressing real-time object tracking challenges (e.g., occlusion and splits) using centroid-based feature matching for continuous tracking.	Validated for efficient tracking of multiple moving objects in complex scenarios involving humans and vehicles.
Zhao et al. [34]	Proposed a multi-object tracking approach for autonomous vehicles, integrating a multi-scale object detection module and a compressed CNN-based correlation filtering module.	Outperformed state-of-the-art tracking methods on KITTI and MOT2015 benchmarks, achieving fewer false negatives and a low missing rate.

Liang et al. [35]	Developed a method combining YOLOv3 for vehicle detection with an Improved Kernel Correlation Filter (KCF) for vehicle tracking to ensure accurate vehicle counting and tracking.	Achieved high-precision vehicle detection and suitable tracking accuracy and maintained high tracking speed.
Jodoin et al. [36]	Presented the Urban Tracker algorithm for detecting and tracking objects in urban traffic using blob extraction, tracking, and object-tracking techniques.	Performed better than Track Initiation (TI) in terms of MOTA, particularly in handling occlusions and object size variations.
Zhang et al. [37]	Proposed a multimodal MOT method combining YOLOv3 for 2D detection and PointRCNN for 3D detection, integrating motion and appearance features.	Demonstrated superior tracking accuracy and reliability on KITTI Tracking Benchmark, reducing false detections.

Harshitha and Hussain [38] introduced a study article focused on the development of a surveillance robot utilizing the Internet of Things and Raspberry Pi technologies. The researchers developed a device that transmits data in real time while monitoring a designated region. The model had an accuracy rate of 90 percent. This research improved the application of the Internet of Things in security surveillance. The development of a system that monitors video using the web and PTZ (Pan-Tilt-Zoom) cameras was discussed by the study of Li et al. [39]. This development allowed for real-time monitoring, which allowed users to use remote cameras to monitor multiple areas. The accuracy of the results in this paper reached approximately 85%.

Therefore, this, in turn, will enhance the use of the web with artificial intelligence algorithms to enhance security and safety. In this research article of Arora et al. [40], the authors have devised a method that enhances data compression and storage economy while maintaining video quality. The suggested approach in this research study has decreased data storage space to around 60% while preserving video resolution at an acceptable standard. Consequently, this method will yield excellent resolutions to the storage issue for surveillance systems.

The study article by Lin and Tsao [41] employed a robot as a tour guide, utilizing RFID technology and a laser device for scanning places on maps. The researchers enhanced their method by integrating data from the robot and the geographic scanning laser apparatus. The algorithm demonstrated an accuracy of around 95% in locating sites across various conditions. Sanoob et al. [42] explored the application of smartphones in surveillance systems, utilizing embedded sensors to assess data and provide real-time monitoring. The detection accuracy of this system achieved roughly 90 percent in recognizing significant occurrences, hence enhancing surveillance systems.

Shao et al. [43] discussed the development of intelligent surveillance through sophisticated processing in the big data analysis of videos. Researchers concentrated on artificial intelligence to extract information from extensive recordings, achieving a result accuracy of 92% in scene analysis and item recognition. RSSl data, which indicate the intensity of the received signal, were utilized to enhance a model for tracking human movement [44].

The precision of the results in identifying particular places approached eighty-five percent, underscoring the significance of these signals. Ma et al. [45] devised a sensing system that identifies human postures utilizing infrared heat sensors. Researchers concentrated the system on a compressive measuring methodology to uphold precision and minimize data. The accuracy of identifying various postures was roughly 90 percent. This study greatly enhanced monitoring systems and underscored the relevance of using these sensors in diverse security applications. Yoon and Chun [46] employ CCTV cameras to create a system that monitors users, evaluates video footage, and identifies their positions via computer vision methodologies. The precision of object tracking attained 880%, and this method improved the efficacy of this investigation in monitoring.

ISSN: 2073-9524
Pages:35-51

Singh et al. [47] centered on the development of a monitoring system utilizing a robot compatible with Android smartphones that operates over Wi-Fi, delivering real-time video and audio streams. The system attained subpar outcomes in broadcasting; nonetheless, it has improved security and monitoring. Li and Li [48] discuss the architecture of a wireless network that integrates and supports a camera. The cameras employed sophisticated wireless communication technology, enabling remote monitoring due to their proficiency in transmitting data with little latency. This renders the system appropriate for real-time surveillance.

The prior work of Bensky [49] examines short-range wireless communication technologies, including Bluetooth, Zigbee, and Wi-Fi. It also examines the issues of signal interference and emphasizes the practical applications that depend on these technologies, such as smart home systems and others. Goyal et al. [50] proposed the integration of surveillance systems with encryption and verification systems to establish a multi-level model that minimizes unauthorized access and enhances result correctness. This system is deemed appropriate for sensitive settings, including governmental and commercial entities.

Pala and Inanc [51] demonstrated the application of RFID technology in developing a model that efficiently manages parking spots through automated tracking and identification of vacant areas. This study facilitated the efficient management of parking spots and enhanced user comfort. Gami [52] used the PIR sensor for the classification of movement and distance. They employed a sensor signal analysis technique to precisely estimate distance. These systems will deliver dependable information on individual movements, rendering them appropriate for security applications.

Table 2 Techniques and Applications of Theft Detection Systems**.**

Authors	Contribution	Performance
Harshitha and Hussain [38]	Notified the owner via a bot if an intruder was detected, based on collected data like face recognition.	Notifications sent as SMS are often overlooked in busy lifestyles.
Li et al. [39]	Implemented a concurrent monitoring system using pan-tilt-zoom (PTZ) cameras controlled remotely through a web server.	Coverage is limited to fixed camera positions, restricting observation to a defined area.
Arora et al. [40]	Reduced redundant CCTV frames using Mean Squared Error (MSE) to optimize storage space.	Processing time is required to compare adjacent frames for redundancy checks.
Lin and Tsao [41]	Adopted RFID tags and scanners for initializing robots and automated mapping using laser scanners.	Incorrect tags can lead to inaccurate results in location mapping.
Sanoob et al. [42]	Detected human presence using PIR sensors, sent alert messages, and uploaded captured video to cloud storage with links sent via email.	Alerts sent late at night through email or messages may not be noticed promptly.
Shao et al. [43]	Stored surveillance videos in big data environments using Multi-Point Association Analysis.	Data analysis and processing are complex and resource-intensive.
Booranawong et al.[44]	Developed a movement tracking device based on RSSI signal measurement.	Detection results are inaccurate if the RSSI signal measurements are incorrect.
Ma et al. [45]	Utilized PIR sensors to detect stationary human targets through servo control and arm movement.	PIR sensors can only detect humans when they are motionless.
Yoon and Chun [46]	Designed a CCTV-based tracking system for mobile users, enabling data transfer between mobile devices and servers for object recognition.	Monitoring is limited to the fixed location of the CCTV system.
Singh et al. [47]	Created a mini robot for live video streaming and remote control via a web application.	Requires two mobile devices for audio transfer and video streaming, increasing costs.
Li and Li [48]	Developed a Wi-Fi-enabled system with ARM and Linux OS for managing a wireless transceiver. Camera data is transmitted via a wireless router to a PC.	Requires higher receiver sensitivity when operating outdoors, particularly far from the router.

ISSN: 2073-9524
Pages:35-51

Bensky [49]	Used IR sensors for obstacle detection and mounted wireless cameras on roofs for monitoring.	Requires a clear line of sight between transmitter and receiver; supports only short distances.
Goyal et al. [50]	Designed a three-stage system with a digital keypad, OTP verification via Bluetooth, and RFID or face recognition for access control.	Alerts sent at night may not be seen promptly, reducing their effectiveness.
Pala and Inanc [51]	Introduced RFID technology in parking-lot systems for automated check-ins and check-outs controlled by RFID readers.	Fake tags complicate operations and increase security risks.
Gami [52]	Used a single PIR sensor with machine learning to detect human movement in surveillance zones.	Errors occur if incorrect data is provided to the machine learning model.

Cao et al. [53] proposed Part Affinity Fields (PAFs) for real-time multi-person 2D pose estimation, enabling efficient association of body parts. The model achieved state-of-the-art results on MPII and COCO datasets with high accuracy and real-time processing speed.

Sun et al. [54] developed HRNet, maintaining high-resolution representations for human pose estimation. The proposed model achieved superior performance on COCO and MPII datasets, providing more accurate and spatially precise key points.

Xiu et al. [55] proposed Pose Flow for efficient online pose tracking, combining cross-frame pose association with a novel NMS technique. The model improved multi-person pose tracking by 13 MAP and 25 MOTA on Pose Track datasets, enabling online tracking at 10 FPS.

Vakunov et al. [56] introduced Media Pipe Hands, an on-device hand-tracking solution with a two-stage pipeline for detecting and estimating hand poses. The proposed system enabled real-time tracking with high-quality predictions on mobile devices using only RGB input.

Table 3 Techniques of Human Pose Estimation (HPE).

Authors	Contribution	Key Focus
Cao et al. [53]	Proposed Part Affinity Fields (PAFs) for real-time multi-person 2D pose estimation, enabling efficient association of body parts.	Achieved state-of-the-art results on MPII and COCO datasets with high accuracy and real-time processing speed.
Sun et al. [54]	Developed HRNet, maintaining high-resolution representations for human pose estimation.	Achieved superior performance on COCO and MPII datasets, providing more accurate and spatially precise key points.
Xiu et al. [55]	Proposed Pose Flow for efficient online pose tracking, combining cross-frame pose association with a novel NMS technique.	Improved multi-person pose tracking by 13 MAP and 25 MOTA on Pose Track datasets, enabling online tracking at 10 FPS.
Vakunov et al. [56]	Introduced Media Pipe Hands, an on-device hand tracking solution with a two-stage pipeline for detecting and estimating hand poses.	Enabled real-time tracking with high-quality predictions on mobile devices using only RGB input.

3. Behavioral Analysis for Theft Prevention

The computational approaches discussed in the aforementioned papers highlight a diverse range of methodologies and challenges in theft detection. IoT systems utilizing Raspberry Pi and PIR sensors have been investigated for human presence detection, alarm activation, and image capture for real-time notifications [57]. Research has similarly highlighted the application of IoT for economical and efficient theft detection, utilizing systems that collect motion-activated photos and dispatch email notifications [58]. Another study proposed a home security framework utilizing PIR sensors and real-time monitoring via Raspberry Pi, focusing on minimal computational overhead [59].

Thief detection systems that use infrared sensors and cameras to spot unauthorized access have shown they can provide quick alerts and collect evidence effectively. These studies show that while sensor-based systems have improved, they still struggle with adapting and being accurate in complicated settings [60]. Recognizing suspicious behavior patterns plays a vital role in thwarting theft in retail environments. The advent of artificial intelligence (AI) has transformed traditional surveillance methods, enabling quick analysis of consumer behavior. Systems like Vision use smart algorithms to spot unusual behavior, like staying in one place for too long or moving in strange ways, which could suggest someone is trying to steal something [61]. These systems are intended to improve their accuracy over time by perpetually learning from the incoming data. AI-driven solutions can identify specific behavioral indicators linked to theft, including avoidance of scanning at self-checkout stations or excessive time spent in certain store sections.

These behaviors are seen as unusual compared to a standard, which allows for quick action to be taken when suspicious actions are noticed [62]. By utilizing object recognition technology, AI can monitor product interactions, identifying occurrences where goods are obscured or altered. This analysis helps find theft and allows businesses to create specific prevention strategies based on behavior patterns observed during risky times [63]. In stores using AI monitoring, understanding customer behavior helps create subtle ways to reduce theft without harming the shopping experience.

4. Methodologies and Approaches

The quick progress in monitoring technology has been fueled by new ideas and methods that tackle problems like theft detection, tracking multiple objects, and understanding human poses. This part describes the main methods and frameworks used across different fields, focusing on their unique contributions, advantages, and limitations.

4.1 Theft Detection Methodologies

Theft detection systems have improved from traditional CCTV surveillance to sophisticated, computerized answers powered with the aid of IoT and machine learning. IoT-enabled surveillance uses sensors, Wi-Fi communication, and edge computing to provide real-time alerts. It involves tools like motion sensors, RFID tags, and smart cameras. Rule-based detection uses set rules to identify unauthorized access or normal behavior. However, it struggles to adjust to changing situations. Machine learning and AI techniques, such as support vector machines and neural networks, examine surveillance data to spot unusual activities and behaviors, making detection much better [64].

Moreover, sensor fusion integrates records from numerous sources, which include cameras and movement detectors, to beautify precision and decrease fake positives, illustrating the transition in the direction of greater adaptable robbery detection structures [65]. Research on theft detection and monitoring devices reveals a notable disparity in performance parameters, including accuracy and computing efficiency. The YOLO method employed for vehicle detection demonstrates an accuracy of 95%, rendering it beneficial in traffic scenarios. Nevertheless, technologies like PIR sensors depend on reduced detection precision while exhibiting excellent energy efficiency. Some systems, like smart monitoring cameras, prioritize video storage and processing yet may have significant computational demands that hinder their real-time performance.

Consequently, enhancing the equilibrium between accuracy and efficiency is essential for the advancement of more effective systems. We may offer a hybrid framework that integrates deep learning with multi-sensing approaches. This framework uses deep learning algorithms to evaluate data from many sources, including cameras and sensors such as PIR and RFID. This technology may augment the precision of theft detection by synthesizing the gathered data to refine pattern recognition and human behavior analysis. Moreover, reinforcement learning methodologies can enhance system responsiveness through real-time feedback, hence augmenting the system's efficacy in dynamic contexts.

4.2 Multi-Object Tracking (MOT) Methodologies

Multi-object tracking plays an important function in packages inclusive of site visitor management, crowd monitoring, and autonomous systems, with current advancements specializing in enhancing monitoring accuracy and computational efficiency. Deep learning techniques, such as YOLO (You Only Look Once), Deep SORT, and R-CNN, utilize convolutional neural networks for object detection and re-identification, thereby enabling real-time tracking [66]. The monitoring-through-detection approach finds objects in separate images and connects them over time using methods like correlation filters, Kalman filters, or IoU-based trackers. Optical flow techniques, such as the Lucas-Kanade method, calculate pixel-degree movement to keep item trajectories in dynamic settings [5].

Moreover, hybrid frameworks amalgamate multi-object tracking algorithms with sensor records, such as LIDAR or GPS, to enhance resilience in congested or obstructed environments [40]. These improvements show the ongoing development of methods for tracking multiple items to meet various needs. Numerous research in the domain of object tracking have demonstrated significant variability in performance parameters, including accuracy and computing efficiency. The YOLO model achieves an accuracy of up to 95.6% in object detection using video analysis techniques. Conversely, the Flow-Tracker model, recognized as a motion tracking model, attained superior accuracy with a MOTA metric of 32.1%. Despite the efficacy of DeepSORT and YOLO methods, there is a necessity for enhancement in both accuracy and efficiency.

The enhanced models in this domain demonstrated a rise in MOT accuracy from 50.32% to 76.21%. Combining reinforcement learning with deep learning is another idea. In this method, deep learning figures out where things are in movies, and the reinforcement learning model improves performance based on what the deep learning model learns in its training environment. Thus, the device can simultaneously identify and track, enhancing efficiency.

4.3 Human Pose Estimation (HPE) Methodologies

Human posture estimate (HPE) is important for comprehending human conduct and facilitating programs including activity recognition, healthcare monitoring, and aggression detection. Main strategies include advanced methods for understanding data, such as heatmap-based approaches (like HRNet) and regression models. These use convolutional neural networks to accurately predict key point locations [67]. HPE frameworks can be divided into two methods: top-down and bottom-up. The top-down method first identifies people and then estimates their positions within boxes, which works well for single individuals or small groups. On the other hand, the bottom-up method detects key points in the image and connects them to individuals, making it better for crowded areas [68]. Temporal fashions enhance posture estimation in videos by using time-based connections and methods such as optical flow or recurrent neural networks (RNNs). Additionally, signal-based human pose estimation (HPE), which uses radio frequency signals or radar, helps overcome issues like obstructions and poor visibility [69].

This enhances traditional image-based methods. These advances underscore the adaptability and innovative abilities of HPE throughout diverse packages. Performance indicators, including accuracy and computing efficiency, seem to fluctuate while employing various approaches for estimating human position. The HRNet-W48 model is among the most proficient systems in its domain, with an accuracy of 88.4%. However, it requires 63.6 million parameters and 32.9 GFLOPs for operation, rendering it unsuitable for jobs necessitating rapid execution. Lightweight versions like Lite-HRNet have attained a balance between efficiency and precision, achieving an accuracy of 69.7% with

 ISSN: 2073-9524
Pages:35-51

much fewer parameters than HRNet-W48, rendering it ideal for mobile devices.

Consequently, it is imperative, particularly in practical applications, to achieve an optimal equilibrium between performance and efficiency requirements. A novel potential paradigm for human posture estimation is the integration of deep learning and unsupervised learning methodologies. This method entails producing a preliminary assessment of human posture via a neural network, followed by the use of unsupervised learning models to enhance such assessments. A multi-task technique can be employed by concurrently training the model while estimating and assessing motion. This will improve the capacity to navigate the difficulties of intricate situations, including congestion and fluctuating circumstances.

4.4 Common Techniques Across Domains

Data augmentation, area adaptation, and version compression are essential methodologies for enhancing the overall performance of gadget-studying models. Data augmentation complements the diversity of education datasets by using adjustments such as rotation, scaling, and occlusion simulation, consequently improving model resilience and generalization [70]. Domain edition reconciles the disparity between training and deployment environments, making certain uniform performance throughout varied conditions. Version compression focuses on creating smaller and simpler architectures and using pruning methods to improve performance. This makes it easier to run real-time applications on devices with limited resources [71]. Collectively, those methodologies are crucial in improving machine learning competencies across numerous fields.

Integrating optical data with other sensor data, such as temperature, sound, and RFID, is crucial for enhancing the reliability and accuracy of surveillance systems. This may be accomplished by employing machine learning models to analyze the data and identify any anomalous activities. This analytical process results in improved responses to prospective threats, hence augmenting the efficacy of monitoring and security systems. Investigations in this domain can provide substantial advancements in addressing theft issues and enhancing overall security.

4.5 Challenges and Innovations

Despite primary improvements in surveillance technology strategies, problems remain regarding scalability, adaptability, and ethical concerns. Future studies aim to develop stronger algorithms that can handle tough challenges like blocked views, motion blur, and immediate obstacles. They will also focus on privacy and fairness during implementation [72]. By utilizing these new methodologies, current surveillance systems are gradually evolving, becoming more shrewd, adaptive, and proficient in handling the complexities of the real international environment [73].

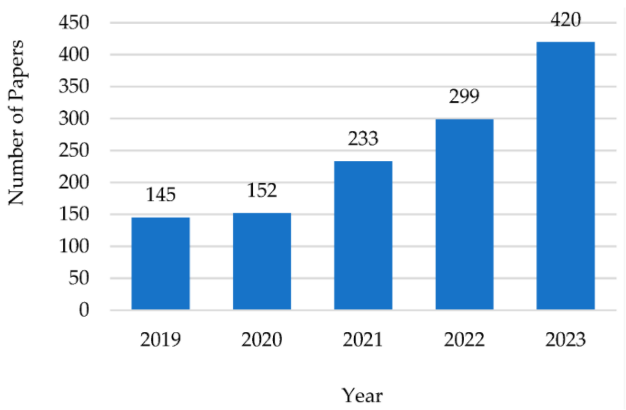

Fig. 1 Trend in the number of publications on deep learning for abnormal human behaviour detection over the past five years (2019-2023) [74].

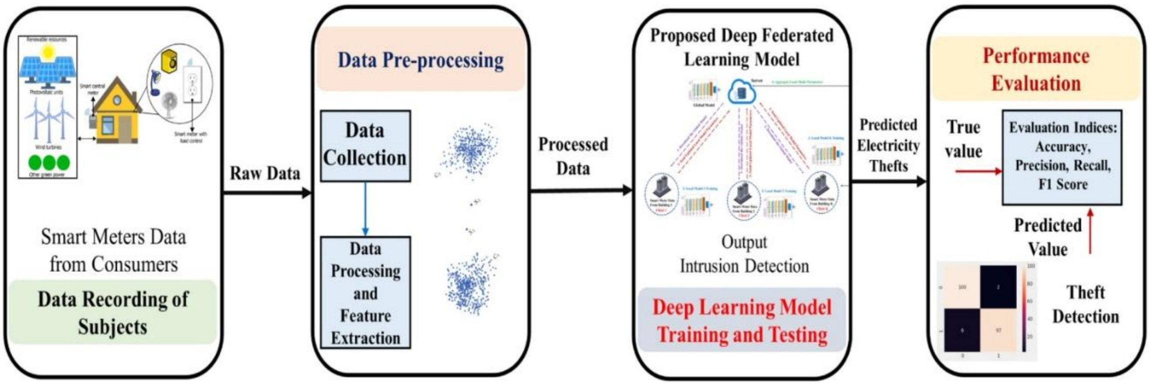

Fig. 2 An illustration of the Theft Detection method for Smart Grids based on Federated Learning [75].

5. Conclusion

This paper reviewed a set of studies dealing with advanced techniques in the field of surveillance and movement recognition. The results showed that many of the developed systems, such as those using deep learning and image processing techniques, achieved high accuracy—up to 96%—in identifying violations and analysing movement. These results indicate the effectiveness of the methods used in improving the efficiency of surveillance systems in complex urban environments. These studies contribute to enhancing knowledge on how to integrate multimedia data and deep learning to develop more accurate and reliable systems. By applying these techniques, performance can be improved in areas such as security and traffic safety, enhancing user experience and reducing accidents. However, further research is needed to explore new ways to overcome current challenges, such as pattern recognition techniques in crowded or complex environments. Proposed solutions could include developing hybrid models that combine deep learning and quantum computing techniques, which could lead to significant performance improvements. It is also recommended to conduct field studies to evaluate the effectiveness of these systems in real-life conditions, which will enable areas for improvement to be accurately identified. Continued research in this area will lead to improvements in existing systems and the development of new technologies that enhance safety and efficiency in multiple fields, opening new horizons for innovation and practical application.

The paper discusses studies about the application of emerging technologies, such as the Internet of Things, artificial intelligence, and data compression in enhancing surveillance and motion detection systems. Studies indicate that these systems may get high accuracy levels between 85% and 95%, demonstrating the efficacy of the ways used to enhance security and safety. This research enhances the understanding of integrating diverse data and sophisticated models to better surveillance systems. Nevertheless, research must persist in investigating innovative methods to enhance performance and decrease costs, including the development of data compression algorithms and the refinement of pattern recognition systems. The deployment of the suggested methods can improve the efficacy of monitoring in sensitive settings, like smart cities and public facilities. Utilising contemporary technology enhances security and mitigates hazards, hence fostering the development of safer societies.

Conflict of Interest

The authors declare that there are no conflicts of interest regarding the publication of this manuscript.

References

[1] Gubbi, J., Buyya, R., Marusic, S., & Palaniswami, M. (2013). Internet of Things

(IoT): A vision, architectural elements, and future directions. Future generation computer systems, 29(7), 1645-1660.

[2] Sarkar, S., Ahire, S., Rahate, S., Barde, R., Agrawal, R., & Sorte, S. (2022, August). Design of Weapon Detection System. In 2022 3rd International Conference on Electronics and Sustainable Communication Systems (ICESC) (pp. 1016-1022). IEEE.

[3] Madore, K. P., & Wagner, A. D. (2019, March). Multicosts of multitasking. In Cerebrum: the Dana forum on brain science (Vol. 2019). Dana Foundation.

[4] Girshick, R. (2015). Fast r-cnn. arXiv preprint arXiv:1504.08083.

[5] Ren, S., He, K., Girshick, R., & Sun, J. (2016). Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE transactions on pattern analysis and machine intelligence, 39(6), 1137-1149.

[6] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 779-788).

[7] Maji, D., Nagori, S., Mathew, M., & Poddar, D. (2022). Yolo-pose: Enhancing yolo for multi person pose estimation using object keypoint similarity loss. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 2637-2646).

[8] Baccouche, M., Mamalet, F., Wolf, C., Garcia, C., & Baskurt, A. (2011). Sequential deep learning for human action recognition. In Human Behavior Understanding: Second International Workshop, HBU 2011, Amsterdam, The Netherlands, November 16, 2011. Proceedings 2 (pp. 29-39). Springer Berlin Heidelberg.

[9] Chen, C. H., & Ramanan, D. (2017). 3d human pose estimation= 2d pose estimation+ matching. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 7035-7043).

[10] Gong, K., Liang, X., Li, Y., Chen, Y., Yang, M., & Lin, L. (2018). Instance-level human parsing via part grouping network. In Proceedings of the European conference on computer vision (ECCV) (pp. 770-785).

[11] Girdhar, R., Gkioxari, G., Torresani, L., Paluri, M., & Tran, D. (2018). Detect-and-track: Efficient pose estimation in videos. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 350-359).

[12] Arun, G., Ajay, B., & Valarmathi, R. (2023, December). IoT Based Anti-Theft Detection System. In 2023 Intelligent Computing and Control for Engineering and Business Systems (ICCEBS) (pp. 1-4). IEEE.

[13] Patil, M. N., Hattaraki, S. M., Ukkali, P. S., Rathod, P. B., Shatgar, D. D., & Metri, P. B. (2024, September). Smart Anti-Theft Control System Using IoT. In 2024 IEEE North Karnataka Subsection Flagship International Conference (NKCon) (pp. 1-5). IEEE.

[14] Al-Hashemi, S. Q. Y., Naghmash, M. S., & Ghandour, A. (2024). Development of Smart Video Surveillance Systems: Technical and Security Challenges in Urban Environments. SHIFRA, 2024, 24-38.

[15] Prajwal, G. K., Sanjaya, D. S., Wandkar, S. M., & Basawaraj. (2024, April). Real-Time Multi-object Detection and Tracking on Zynq UltraScale+ MPSoc. In International Conference on Information and Communication Technology for Intelligent Systems (pp. 71-81). Singapore: Springer Nature Singapore.

[16] Liu, H., Zhao, Y., Dong, P., Guo, X., & Wang, Y. (2025). IOF-Tracker: A Two-Stage Multiple Targets Tracking Method Using Spatial-Temporal Fusion Algorithm. Applied Sciences (2076-3417), 15(1).

[17] Razak, R. N., & Abdullah, H. N. (2024). Improving multi-object detection and tracking with deep learning, DeepSORT, and frame cancellation techniques. Open Engineering, 14(1), 20240056.

[18] Luo, Z., Wang, Z., Huang, Y., Wang, L., Tan, T., & Zhou, E. (2021). Rethinking the heatmap regression for bottom-up human pose estimation. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 13264-13273).

[19] Takagi, S., Xiong, L., Kato, F., Cao, Y., & Yoshikawa, M. (2024). HRNet: Differentially Private Hierarchical and Multi-Resolution Network for Human Mobility Data Synthesization. arXiv preprint arXiv:2405.08043.

[20] Vendrow, E., Le, D. T., Cai, J., & Rezatofighi, H. (2023). Jrdb-pose: A large-scale dataset for multi-person pose estimation and tracking. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 4811-4820).

 ISSN: 2073-9524
Pages:35-51

[21] Kukreja, V., Thapliyal, N., Aeri, M., & Sharma, R. (2024, June). An Advanced YOLOv7 Model for Vehicle Detection to Enhance the Security of Public Places. In 2024 IEEE International Conference on Information Technology, Electronics and Intelligent Communication Systems (ICITEICS) (pp. 1-5). IEEE.

[22] Manasa, A., & Renuka Devi, S. M. (2023, August). An Enhanced Real-Time System for Wrong-Way and Over Speed Violation Detection Using Deep Learning. In International Conference on Image Processing and Capsule Networks (pp. 309-322). Singapore: Springer Nature Singapore.

[23] Tang, Z., Gu, R., & Hwang, J. N. (2018, July). Joint multi-view people tracking and pose estimation for 3D scene reconstruction. In 2018 IEEE international conference on multimedia and expo (ICME) (pp. 1-6). IEEE.

[24] Li, W., Mu, J., & Liu, G. (2019). Multiple object tracking with motion and appearance cues. In Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops (pp. 0-0).

[25] Larasati, H. T., & Kim, H. (2022, August). Trends of quantum computing applications to computer vision. In 2022 International Conference on Platform Technology and Service (PlatCon) (pp. 7-12). IEEE.

[26] Hassan, S., Mujtaba, G., Rajput, A., & Fatima, N. (2024). Multi-object tracking: a systematic literature review. Multimedia Tools and Applications, 83(14), 43439-43492.

[27] Chen, W., Wang, W., Wang, K., Li, Z., Li, H., & Liu, S. (2020). Lane departure warning systems and lane line detection methods based on image processing and semantic segmentation: A review. Journal of traffic and transportation engineering (English edition), 7(6), 748-774.

[28] Bullinger, S., Bodensteiner, C., & Arens, M. (2017, September). Instance flow based online multiple object tracking. In 2017 IEEE International Conference on Image Processing (ICIP) (pp. 785-789). IEEE.

[29] Sakkos, D., Liu, H., Han, J., & Shao, L. (2018). End-to-end video background subtraction with 3d convolutional neural networks. Multimedia Tools and Applications, 77, 23023-23041.

[30] Grant, J. M. (2018). Analysis of Crowd Behavior Based on Optical Flow: Detection, Classification, and Clustering. University of Notre Dame.

[31] Mampilayil, H. R., & Rahamathullah, K. (2019, May). Deep learning based detection of one way traffic rule violation of three wheeler vehicles. In 2019 international conference on intelligent computing and control systems (ICCS) (pp. 1453-1457). IEEE.

[32] Suttiponpisarn, P., Charnsripinyo, C., Usanavasin, S., & Nakahara, H. (2022). An autonomous framework for real-time wrong-way driving vehicle detection from closed-circuit televisions. Sustainability, 14(16), 10232.

[33] Li, X., Wang, K., Wang, W., & Li, Y. (2010, June). A multiple object tracking method using Kalman filter. In The 2010 IEEE international conference on information and automation (pp. 1862-1866). IEEE.

[34] Zhao, D., Fu, H., Xiao, L., Wu, T., & Dai, B. (2018). Multi-object tracking with correlation filter for autonomous vehicle. Sensors, 18(7), 2004.

[35] Liang, H., Song, H., Li, H., & Dai, Z. (2020). Vehicle counting system using deep learning and multi-object tracking methods. Transportation research record, 2674(4), 114-128.

[36] Jodoin, J. P., Bilodeau, G. A., & Saunier, N. (2014, March). Urban tracker: Multiple object tracking in urban mixed traffic. In IEEE Winter Conference on Applications of Computer Vision (pp. 885-892). IEEE.

[37] Zhang, L., Lai, J., Zhang, Z., Deng, Z., He, B., & He, Y. (2020). Multimodal Multiobject Tracking by Fusing Deep Appearance Features and Motion Information. Complexity, 2020(1), 8810340.

[38] Harshitha, R., & Hussain, M. H. S. (2018, April). Surveillance robot using raspberry Pi and IoT. In 2018 International conference on design innovations for 3Cs compute communicate control (ICDI3C) (pp. 46-51). IEEE.

[39] Li, K. Y., Liang, J. M., Fan, C. S., Wu, K. R., Lin, Y. T., Lin, T. Y., & Tseng, Y. C. (2015, April). A web-based, real-time video surveillance system by leveraging PTZ cameras. In 2015 IEEE Tenth International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP) (pp. 1-1). IEEE.

[40] Arora, S., Bhatia, K., & Amit, V. (2016, October). Storage optimization of video surveillance from CCTV camera. In 2016 2nd

International Conference on Next Generation Computing Technologies (NGCT) (pp. 710-713). IEEE.
[41] Lin, H. H., & Tsao, W. Y. (2011, August). Automatic mapping and localization of a tour guide robot by fusing active RFID and ranging laser scanner. In The 2011 International Conference on Advanced Mechatronic Systems (pp. 429-434). IEEE.
[42] Sanoob, A. H., Roselin, J., & Latha, P. (2015). Smartphone enabled intelligent surveillance system. IEEE sensors journal, 16(5), 1361-1367.
[43] Shao, Z., Cai, J., & Wang, Z. (2017). Smart monitoring cameras driven intelligent processing to big surveillance video data. IEEE Transactions on Big Data, 4(1), 105-116.
[44] Booranawong, A., Jindapetch, N., & Saito, H. (2018). A system for detection and tracking of human movements using RSSI signals. IEEE sensors journal, 18(6), 2531-2544.
[45] Ma, R., Hu, F., & Hao, Q. (2016). Active compressive sensing via pyroelectric infrared sensor for human situation recognition. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 47(12), 3340-3350.
[46] Yoon, Y. I., & Chun, J. A. (2014, February). Tracking System for mobile user Based on CCTV. In The International Conference on Information Networking 2014 (ICOIN2014) (pp. 374-378). IEEE.
[47] Singh, D., Zaware, P., & Nandgaonkar, A. (2017, May). Wi-Fi surveillance bot with real time audio & video streaming through Android mobile. In 2017 2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT) (pp. 746-750). IEEE.
[48] Li, M., & Li, Z. (2011, September). Design of embedded wireless network camera. In 2011 3rd International Conference on Awareness Science and Technology (iCAST) (pp. 183-186). IEEE.
[49] Bensky, A. (2019). Short-range wireless communication. Newnes.
[50] Goyal, S., Desai, P., & Swaminathan, V. (2017). Multi-level security embedded with surveillance system. IEEE Sensors Journal, 17(22), 7497-7501.
[51] Pala, Z., & Inanc, N. (2007, September). Smart parking applications using RFID technology. In 2007 1st Annual RFID Eurasia (pp. 1-3). IEEE.
[52] Gami, H. (2017). Movement direction and distance classification using a single PIR sensor. IEEE sensors letters, 2(1), 1-4.
[53] Cao, Z., Simon, T., Wei, S. E., & Sheikh, Y. (2017). Realtime multi-person 2d pose estimation using part affinity fields. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 7291-7299).
[54] Sun, K., Xiao, B., Liu, D., & Wang, J. (2019). Deep high-resolution representation learning for human pose estimation. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 5693-5703).
[55] Xiu, Y., Li, J., Wang, H., Fang, Y., & Lu, C. (2018). Pose Flow: Efficient online pose tracking. arXiv preprint arXiv:1802.00977.
[56] Vakunov, A., Chang, C. L., Zhang, F., Sung, G., Grundmann, M., & Bazarevsky, V. (2020, June). Mediapipe hands: On-device real-time hand tracking. In Workshop on Computer Vision for AR/VR (Vol. 2, No. 4, p. 5).
[57] Poushya, M., Rupasri, K., Supritha, N., Hema, K., & Tejaswini, R. (2018). IoT based vehicle theft detection. IRE Journals, 1(8).
[58] Saranu, P. N., Abirami, G., Sivakumar, S., Ramesh, K. M., Arul, U., & Seetha, J. (2018, February). Theft detection system using PIR sensor. In 2018 4th International conference on electrical energy systems (ICEES) (pp. 656-660). IEEE.
[59] Rajawat, S. S., Som, S., & Rana, A. (2020, June). Iot based theft detection using raspberry pi. In 2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO) (pp. 829-834). IEEE.
[60] Kai, A., Zikriya, M., Mallapur, J. D., & Hiremath, S. C. (2020). Motion Intervene Surveillance Control Using IoT. In ICT Systems and Sustainability: Proceedings of ICT4SD 2019, Volume 1 (pp. 207-214). Springer Singapore.
[61] Dakalbab, F., Talib, M. A., Waraga, O. A., Nassif, A. B., Abbas, S., & Nasir, Q. (2022). Artificial intelligence & crime prediction: A systematic literature review. Social Sciences & Humanities Open, 6(1), 100342.
[62] Jiang, M., Cui, P., & Faloutsos, C. (2016). Suspicious behavior detection: Current trends and future directions. IEEE intelligent systems, 31(1), 31-39.
[63] Bennamoun, M., & Mamic, G. J. (2012). Object recognition: fundamentals and case studies. Springer Science & Business Media.

[64] Arora, J., Bangroo, A., & Garg, S. (2022). Theft Detection and Monitoring System Using Machine Learning. In Emerging Research in Computing, Information, Communication and Applications: ERCICA 2020, Volume 2 (pp. 957-966). Springer Singapore.

[65] Yang, T., & Lv, C. (2021). A secure sensor fusion framework for connected and automated vehicles under sensor attacks. IEEE Internet of Things Journal, 9(22), 22357-22365.

[66] Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. (2020). Yolov4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934.

[67] Sun, K., Geng, Z., Meng, D., Xiao, B., Liu, D., Zhang, Z., & Wang, J. (2020). Bottom-up human pose estimation by ranking heatmap-guided adaptive keypoint estimates. arXiv preprint arXiv:2006.15480.

[68] Zhao, S., Liu, K., Huang, Y., Bao, Q., Zeng, D., & Liu, W. (2022, August). Dpit: Dual-pipeline integrated transformer for human pose estimation. In CAAI International Conference on Artificial Intelligence (pp. 559-576). Cham: Springer Nature Switzerland.

[69] Cheng, Y., Wang, B., & Tan, R. T. (2022). Dual networks based 3d multi-person pose estimation from monocular video. IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(2), 1636-1651.

[70] Pellicer, L. F. A. O., Ferreira, T. M., & Costa, A. H. R. (2023). Data augmentation techniques in natural language processing. Applied Soft Computing, 132, 109803.

[71] Farahani, A., Voghoei, S., Rasheed, K., & Arabnia, H. R. (2021). A brief review of domain adaptation. Advances in data science and information engineering: proceedings from ICDATA 2020 and IKE 2020, 877-894.

[72] Custers, B., & Vergouw, B. (2015). Promising policing technologies: Experiences, obstacles and police needs regarding law enforcement technologies. Computer Law & Security Review, 31(4), 518-526.

[73] Macnish, K. (2017). The ethics of surveillance: An introduction. Routledge.

[74] Wastupranata, L. M., Kong, S. G., & Wang, L. (2024). Deep learning for abnormal human behavior detection in surveillance videos—A survey. Electronics, 13(13), 2579.

[75] Zafar, M. H., Bukhari, S. M. S., Abou Houran, M., Moosavi, S. K. R., Mansoor, M., Al-Tawalbeh, N., & Sanfilippo, F. (2023). Step towards secure and reliable smart grids in Industry 5.0: A federated learning assisted hybrid deep learning model for electricity theft detection using smart meters. Energy Reports, 10, 3001-3019.