International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

A Comprehensive Survey of Multiple Object Tracking Techniques

Hardik Jaiswal[1] , Aditya Gambhir[2] , Laxmi Bewoor*[3] , Nagaraju Bogiri[4]

Submitted: 07/02/2024 **Revised**: 15/03/2024 **Accepted**: 21/03/2024

Abstract: Multiple Object Tracking (MOT) is crucial in computer vision and surveillance, especially for automating traffic control in challenging traffic environments. This review surveys advancements in object detection, tracking algorithms, lane departure warnings, and semantic segmentation, with a specific focus on traffic law enforcement. It covers issues like wrong-way, clearway, and one-way traffic violations, as well as challenges including occlusion and splits. Various methods, such as background subtraction and deep learning, are explored.The review stresses the significance of analyzing recent literature for researchers to bridge gaps, overcome limitations, and create new algorithms. It also touches on hardware, datasets, metrics, and research directions. Future MOT research aims to develop efficient algorithms for dynamic tracking, improve detection accuracy, and reduce real-time processing. The survey's proposed methods offer valuable references for tracking multiple objects in frame sequences.

Keywords: *Multiple Object Tracking (MOT), Traffic environment, Deep Learning*

1. Introduction:

Over the past few years, computer vision has rapidly evolved and become one of the most active research areas in computer science.[1]

Object tracking can be described as the process of tracking an object over successive frames so that its position and direction are known throughout the series. As a result, the tracking task can involve two major sub-tasks: object detection and object identification or re-identification between frames. These two tasks allow us to determine the exact position of the various items in an image as well as the journey they have taken in the field of view represented by the photographs.

Monitoring many moving items in a recording gets more difficult for a human as the number of objects to track increases.[2] On the contrary, computer vision can recognize and track hundreds of items with high accuracy. This is known as multi-object tracking. Item detection and tracking are difficult tasks that necessitate overcoming numerous obstacles. Detection systems must cope with challenges such as scale variance, lighting changes, occlusions, and differences in shape and appearance in addition to recognizing the location and kind of objects. Furthermore, it must deal with the classification of items that have varied degrees of similarity. Meanwhile, tracking algorithms must deal with new issues such as tracking fast-moving objects, dealing with momentary occlusions, and tracking objects with various motion patterns.

They must also consider scale variations, motion blur, and camera movement. Furthermore, tracking systems must account for differences in lighting conditions, which can have a considerable impact on object appearance.

Numerous studies have been developed that use and propose innovative architectures and approaches in computer vision, such as Region-Convolutional Neural Network (R-CNN) [3], Fast R-CNN [4], Faster R-CNN [5], and You Only Look Once (YOLO).[6] These unique approaches are frequently created by large tech corporations, such as Facebook [5, 6], Microsoft [3], Google, and others. As a result, the relevance of computer vision in the real research overview is highlighted. Many of these studies are used in object detection and tracking, which are inextricably linked since object tracking depends on object detection.

These research topics are particularly helpful because they allow for the easy and rapid monitoring of open or populated situations such as shopping centers, public buildings, squares, streets, highways, and so on. In this review, we mainly focus on the research of MOT in traffic environments. The growing demand for automatic traffic control is the underlying reason for this specification. In traffic situations, detection and tracking techniques are critical for monitoring the behavior of road users. This data can be utilized to detect unsafe driving behaviors, improve traffic flow in congested intersections, and increase overall safety. However, traffic scenarios present their issues, such as occlusion, backdrop clutter, and position changes, which must be considered. Furthermore, tracking systems may require integrating data from various sensors or cameras. Despite

[1] [3] *Vishwakarma Institute of Information Technology, Pune-411048,INDIA*
ORCID ID : 0000-0003-3470-421X
[4]*Vishwakarma Institute of Information Technology, Pune-411048,INDIA*
ORCID ID : 0000-0003-3037-4452
** Corresponding Author Email: laxmi.bewoor@viit.ac.in*

the complexity and use of these systems, there is no comprehensive study of the state-of-the-art on this topic at the moment. As a result, it is critical to perform a thorough analysis of recent studies on multi-object tracking in traffic settings. This review covers the latest algorithmic developments, types of datasets available for further investigations, the hardware used, evaluation metrics for analyzing the efficiency of algorithms, and future research directions.

The remainder of the article is organized as follows: Section 2 discusses the major contributions in this domain. Section 3 discusses the more pertinent aspects of Section 2, Section 4 gives the future potential study directions, and Section 5 provides the conclusion.

2. Related Work

Computer vision research spans several sub-areas that have proven to be critical over the years. Image classification was one of the first notable advances in computer vision.[7] The method entails recognizing the objects or entities that appear in an image, and the output of the process is usually the object that is present in the image or the different items that appear in the image if it is a multi-label classification. Image classification has a wide range of applications, including object detection, face recognition, and medical diagnosis. On the other hand, the performance of image classification systems largely depends on the quality of the training dataset, the features retrieved from the images, and the machine learning methods used. As a result, researchers are constantly looking for new ways to increase the accuracy and durability of image categorization systems, such as deep learning models and transfer learning.[8,9] Another critical part of image classification is the evaluation metrics used to quantify the system's performance, such as precision, recall, and F1-score, which are critical in establishing the system's effectiveness.[10]

Although image classification may identify items in an image, it cannot pinpoint their exact location. To overcome this issue, object-detection algorithms have been developed.[6] These approaches can recognize a large range of item kinds and identify the exact position of objects in a picture if they are included in the training process.

This approach, however, has a shortcoming in that it cannot establish the exact location of the items in the image. To overcome this issue, object identification systems that not only classify the objects in an image but also establish their specific location have been created. These techniques have proven effective in a variety of applications, including surveillance, traffic control, and medical imaging. Object identification algorithms may recognize a wide variety of things if they are included in the training data. To locate items in a picture, these techniques use a variety of algorithms, including region proposal-based methods, sliding window-based methods, and deep learning-based methods. Object detection system performance is measured using measures such as accuracy, precision, recall, and F1 score.

This section provides an overview of object detection and tracking methods specifically in the traffic environment context.

Cummaragunta et al.[11] presented a paper proposing the automation of traffic law enforcement using real-time video input and object detection. They used the You Only Look Once (YOLO) model to detect vehicles and create centroids for each object, which they categorized by lane using a divider. They returned the output video and stored violating vehicles' images on the Firebase database. The model achieved a high accuracy of over 95%, making it a fully automated solution without the need for manual intervention.

On a similar ground, a real-time system for detecting wrong-way and clearway violations using video input was presented by Ali Şentaş et al. [12] The system extracted frames and vehicle features and drew a security lane for clearway detection. Vehicles violating rules were marked red. Real-time image processing techniques were used, and background subtraction YOLO was not used, ensuring reliable performance even in severe weather conditions.

Further, Zhang J et al.[13] proposed an improved tracking-by-detection strategy for multi-person tracking. The strategy consisted of a new appearance model, a hierarchy of trackers, and tracker initialization/termination based on the matching rate. Trackers were classified as experts or novices based on the number of templates they had, and different strategies were applied for adapting their models and search areas for matching new detections, resulting in improved tracking performance. False detections were also reduced.

A new multiple object tracking (MOT) framework based on the tracking-by-detection scheme was proposed in a paper by Weiqiang Li et al.[14] The proposed framework used an Intersection Over Union (IoU) tracker and an optical flow network to handle global motion problems, and an auxiliary tracker and a better cascade matching strategy were used for missing detections. The proposed "Flow-Tracker" achieved superior performance compared to baseline methods, with an AP of 30.87 and the highest accuracy in all categories. It also offered a trade-off between accuracy and speed. In line with this context, Frank Ngeni et al.[15] proposed a quantum computing-based approach to improve the accuracy of multiple object detection and tracking using Yolov5 and Deep Simple Online and Real-time Tracking

(DeepSORT). The model integrated the Kalman filter and Hungarian algorithm[16] for optimization and the Alternating Direction Method of Multipliers (ADMM) for constraint handling. The proposed model reduced mismatches, improved vehicle counting and classification, and reduced identity switching. Similar models trained without quantum computing were outperformed by the model, which showed an increase in Multi-Object Tracking Accuracy (MOTA) by 16.03%, Multi-Object Tracking Precision (MOTP) by 5.49%, and F1 by 6.09%. The proposed approach showed promise in improving the accuracy of multiple object detection and tracking.

The work reported by Diego M. Jiménez-Bravo et al.[17] advocated a systematic literature review of recent works in the area of multi-object tracking in traffic environments. The authors formulated 5 important questions, defined advanced search terms, and summarized relevant articles to cover techniques, hardware, datasets, metrics, and open lines of research in this area. The review provided a comprehensive collection of research conducted in MOT until 2022, including the most used techniques and datasets. This paper seems to be a valuable resource for identifying new lines of research in multi-object tracking.

A work proposed by Weiwei Chen et al.[18] that examined lane line departure warning systems, image processing algorithms, and semantic segmentation methods for lane line detection. The study conducted a review of various lane departure warning systems and evaluated their capabilities and limitations using benchmarks and different test cases. The paper identified five major problems in existing systems that affected Lane Departure Warning System (LDWS) reliability and proposed techniques to increase the accuracy and precision of vision-based LDWS. Overall, the study provided insights into the current state of LDWS and outlined future research directions.

A paper proposing an online multiple-object tracking approach was presented by Bullinger et al.[19] The approach utilizes semantic instance segmentations and optical flow cues to track objects' two-dimensional shapes at the pixel level. The authors compared their method with Simple Online Realtime Tracking (SORT) using the MOT 2-Dimensional (2D) 2015 test dataset and evaluated the algorithm on pedestrians. The algorithm's ability to track objects' shapes in subsequent frames provided benefits in tracking objects with high relative motions.

Dimitrios Sakkos et al. [20] presented a paper proposing a completely end-to-end temporal-aware approach for background subtraction with 3-Dimensional (3D) convolutional neural networks. The model was capable of effectively tracking the movement of the foreground.

An automated system for detecting unusual behavior on highways, such as sudden lane changes or wrong-way driving was developed by M. Bazan et al.[21] The Lucas-Kanade method was used to calculate dense optical flow, which detected movements in the direction of vehicles. The proposed technique utilized Lucas-Kanade optical flow to detect vehicles moving in the wrong direction.

Further, an automated system for detecting violations of one-way traffic rules using vehicle trajectory points to determine the direction of movement was proposed by Mampilayil and Rahamathullah.[22] OpenCV and TensorFlow were used to detect three-wheeler objects in video frames, and their centroids were tracked to calculate the direction of movement. The proposed framework did not require sensors and was cost-effective and easily deployable. The system could be used for efficient management of traffic in one-way areas.

Suttiponpisarn et al.[23] proposed a paper presenting the Wrong Way-Lane detection, direction Validation, Detection of vehicles, and Capture(LDVC) framework for detecting wrong-way driving vehicles using Closed-Circuit Television (CCTV) videos. The framework includes three sub-systems for improved road lane boundary detection, distance-based direction detection, and inside boundary image capturing. The proposed framework achieved an accuracy of 95.23% on a personal computer and 94.66% on an embedded system. The article also discusses the challenges of creating a road lane detection algorithm due to varying road conditions. The proposed framework has the potential for future improvement in detecting various road conditions and adding new features. It is possible to implement the framework on edge devices for real-time detection in various areas.

An algorithm was proposed by Xin Li et al.[24] to address challenges in real-time object tracking such as occlusion, background obstacles, splits, and merges. The algorithm establishes information links using feature matching based on the centroid feature, which updates the moving object's model. The updated model was then used as the input for the next frame, for achieving continuous tracking of objects. The proposed algorithm was validated on human and vehicle image sequences and showed efficient tracking of multiple moving objects in confusing situations.

A multi-object tracking approach for autonomous vehicles was proposed by Dawei Zhao et al.[25] to improve both the detection and tracking modules. The proposed approach included a multi-scale object detection module and a compressed Convolutional

Neural Network (CNN) feature-based correlation filtering module that were interleaved to track multiple objects and re-identify lost objects efficiently. The proposed approach was extensively tested on the Karlsruhe Institute of Technology and Toyota Technological Institute (KITTI) and MOT2015 tracking benchmarks, which showed that it outperformed most state-of-the-art tracking approaches with fewer false negatives and a low missing rate. This made it suitable for autonomous vehicles to avoid the miss-detection of important targets.

Recently, Haoxiang Liang et al.[26] developed a method to achieve high accuracy in vehicle detection while maintaining accuracy and precision in tracking multiple objects and obtaining accurate vehicle counting results. The method combined the YOLOv3 object detection algorithm for vehicle detection with an Improved kernel correlation filter (KCF) for vehicle tracking. Different features were integrated with the proposed method to calculate the maximum response for predicting vehicle positions, resulting in high-precision vehicle object detection and suitable multiple object tracking accuracy and precision. The method also maintained high tracking speed.

An Urban Tracker algorithm for detecting and tracking various objects in outdoor urban traffic scenes was presented by Jean-Philippe Jodoin et al.[27] The algorithm involved three main steps, including blob extraction, blob tracking, and object tracking. The paper compared the performance of Urban Tracker (UT) and Track Initiation (TI) on four different videos, and the results showed that UT performed better than TI in terms of Multiple Object Tracking Accuracy (MOTA). The precision of UT was also higher than TI for most objects, except for some cars and bicycles. The differences in performance were attributed to the algorithms' ability to handle occlusions, object size, and other characteristics, rather than appearance alone.

Liwei Zhang et al.[28] proposed a multimodal MOT method to enhance object tracking accuracy. They integrated motion information and deep appearance features, utilizing YOLOv3 for 2D object detection and PointRCNN for 3D object detection. Experimental validation on the KITTI Tracking Benchmark showcased successful tracking in crowded scenes, mitigating false detections. The comparative evaluation confirmed the competitiveness of their approach. In summary, Liwei Zhang et al. presented a robust solution, achieving superior tracking accuracy and reliability by integrating motion information and deep appearance features.

3. Analysis of Computational Experiences

In this review, various popular benchmarks with their respective variations and datasets were considered to assess the performance of multiple object-tracking algorithms. The benchmarks included in this analysis were VisDrone2019-MOT (Visual Drone 2019 Multiple Object Tracking)[29], Open Images version 6 (OIv6)[30], Performance Evaluation of Tracking and Surveillance Scenario - Sequence 2 Level 1 (PETS S2L1), Performance Evaluation of Tracking and Surveillance Scenario - Sequence 2 Level 2 (PETS S2L2), Performance Evaluation of Tracking and Surveillance Scenario - Sequence 2 Level 3 (PETS S2L3)[31], TownCenter, MOT 2015, and MOT 2016.[32]

These benchmarks were chosen due to their widespread use and ability to evaluate different aspects of multiple object-tracking systems. VisDrone2019-MOT focuses on tracking objects in aerial videos, while OIv6 provides a diverse range of object categories in real-world scenes. The PETS benchmarks (S2L1, S2L2, and S2L3) were designed specifically for tracking in surveillance scenarios, and TownCenter is another dataset commonly used for pedestrian tracking in crowded urban environments. MOT 2015 and MOT 2016 are widely employed benchmarks for multi-object tracking.

To compare the performance of various algorithms on these benchmarks, several evaluation metrics were utilized. The evaluation metrics used in this review are Multiple Object Tracking Accuracy (MOTA), Multiple Object Tracking Precision (MOTP), and Recall. MOTA combines multiple factors such as false positives, false negatives, and identity switches to provide an overall measure of tracking accuracy. MOTP, on the other hand, quantifies the localization precision of the tracked objects. Recall assesses the ability of the algorithm to successfully detect and track objects across frames.

By employing these evaluation metrics on the selected benchmarks and datasets, this review aims to provide a comprehensive analysis of the strengths and weaknesses of different multiple object tracking algorithms. The insights gained from this evaluation can contribute to the advancement of tracking systems and aid in the development of more accurate and reliable algorithms for real-world applications.

Cummaragunta et al.[11] used the YOLOv4 model, which was trained on the OIv6 dataset. They tested the model on several YouTube videos and achieved an astonishing accuracy of 95.6%. Despite problems such as camera sensitivity, orientation, false positives, and response time, the project met its objectives by utilizing a dependable and fixed CCTV setup. Blurred imagery and faulty detections could result in false positives, and while Anvil's response time fell short of the ideal, the project's essential functions as a fast and lightweight solution were preserved.

The method proposed by Şentaş et al.[12] extracted frames from video input and extracted features for different vehicles. They incorporated a security lane by drawing a line, specifically for clearway detection. The system detected and tracked vehicles, marking them in red if they violated any rules. However, the system had limitations concerning its sensitivity to lighting, weather, and camera placement, making it challenging to detect small or distant vehicles. Moreover, the system's high computational requirements posed feasibility challenges for certain traffic control centers.

An improved tracking-by-detection strategy for multi-person tracking was proposed by Zhang et al.[13] The strategy involved the introduction of a new appearance model, which utilized an online pruning strategy to discard weaker templates and maintain each ensemble's strength. The implementation also included the use of a Hierarchy of Trackers and employed a matching rate and average matching rate among established trackers to determine the initialization or termination of a tracker. The accuracy and precision results were as follows: for the PETS S2L1 dataset, the accuracy was 90.75% and precision was 68.64%; for the PETS S2L2 dataset, the accuracy was 64.12% and precision was 58.66%; for the PETS S2L3 dataset, the accuracy was 39.67% and precision was 57.75% and for the TownCenter dataset, the accuracy was 72.3% and precision was 66.35%. However, the proposed method had limitations, including the absence of 3D information on pedestrians and camera calibration, challenges in handling variations in crowd density, pedestrian dynamics, and parameter settings, and hence reduced effectiveness of expert trackers in a high-density crowd or occluded scenarios.

Weiqiang Li et al.[14] introduced a novel algorithm, the Flow-Tracker, for multi-object tracking (MOT) based on the IOU tracker. The Flow-Tracker enhanced the IOU tracker by incorporating an optical flow network to compensate for camera motion. It also employed an auxiliary tracker to predict object positions in subsequent frames when detections were missing and utilized a more effective cascade matching strategy to handle mismatching caused by missing detections. The comparative evaluation demonstrated the Flow-Tracker's superiority over both the standard IOU tracker and the DeepSort algorithm. The Flow Tracker achieved an improved MOTA of 32.1 (compared to 10.1 in DeepSort and 12.6 in the IOU tracker) and MOTP of 78.1 (compared to 74.7 in DeepSort and 75.7 in the IOU tracker). However, the experiment was limited by its reliance on the VisDrone dataset, suggesting the need for replication on diverse datasets for generalizability. Additionally, the algorithm's use of time-consuming optical flow estimation constrained its speed to 5 Frames Per Second (FPS) for high-resolution videos.

Frank Ngeni et al.[15] split the dataset into an 80-20 ratio for training and validation, using Nvidia Graphics Processing Unit (GPU) and Compute Unified Device Architecture (CUDA). Yolov5 detected objects, and DeepSORT tracked them using the Kalman filter, Hungarian algorithm, and ADMM for optimization. The study compared YOLOv5-DeepSort with YOLOv3-DeepSort and YOLOv5-DeepSort-Quantum on the MOT17 dataset. The accuracy for MOT improved from 50.32% in YOLOv3-DeepSort to 60.18% in YOLOv5-DeepSort and further to 76.21% in YOLOv5-DeepSort-Quantum. Similarly, the precision increased from 65.46% in YOLOv3-DeepSort to 86.26% in YOLOv5-DeepSort and further to 91.75% in YOLOv5-DeepSort-Quantum.However, limitations included reliance on the MOT17 dataset for pedestrian tracking, lack of vehicle tracking datasets for evaluation, and missing information on algorithm complexity and resource requirements for real-time applications or limited devices.

Bullinger et al. [19] developed a methodology for dense instance segmentation, which involved utilizing an instance-aware segmentation system to predict semantic category labels and instance indices for each pixel. Optical flow or semi-dense matching methods were employed to calculate a pixel offset field, and an instance flow tracker was proposed to associate objects between frames based on affinity scores. The methodology demonstrated improved prediction for dense instance segmentation. Evaluating different algorithms on the KITTI 13 Dataset yielded the following results: MaskRCNN+Simple Online and Realtime Tracking (MNC+SORT) achieved 12.9% accuracy and 65.2 precision, MaskRCNN+Center and Perspective Match (MNC+CPM) achieved 19.2% accuracy and 66.7 precision, MaskRCNN+Deep Matching (MNC+DeepMatch) achieved 16.9% accuracy and 66.8 precision, and MaskRCNN+PolyExp achieved 11.7% accuracy and 66.8 precision. However, limitations of the study included a narrow focus on pedestrians in the evaluation, disregarding other object categories, and the presence of unstable segmentation quality adversely affecting the tracker's performance. Additionally, the study did not account for the influence of different weather conditions on the tracker's performance.

Suttiponpisarn et al. [23] developed a framework that employs CCTV video inputs to detect road lanes. However, the system's ability to detect lane lines on roads with unusual conditions, such as construction, pale road line color, and varying CCTV angles, may be limited. These challenging factors observed in Thai road traffic data influenced the algorithm's development. The Road Lane Boundary-CCTV (RLB-CCTV) and Dynamic Background Difference Detection (DBDD) algorithms were trained and tested on custom datasets with over 1000 images from CCTV cameras, achieving an accuracy of 95.23%. Optimizing the system for Jetson Nano and detecting curved roads can improve performance, although the handling of curved roads remains a limitation. Accurately validating the Modified Background Change Detection and Difference (MBCDD) algorithm requires longer video lengths, potentially affecting embedded system efficiency. Balancing accuracy and efficiency in shorter validation videos poses a challenge that warrants attention.

Zhao et al. [25] proposed an integrated multi-object tracking framework consisting of two modules: multi-scale object detection and compressed CNN feature-based Correlation Filtering (CCF). This framework efficiently tracks multiple objects and incorporates re-identification capabilities for lost objects. The framework was evaluated on the KITTI and MOT15 datasets, achieving an accuracy of 71.27% and precision of 81.83% on KITTI, and an accuracy of 32.7% and precision of 38.9% on MOT15. However, the algorithm's performance evaluation was constrained by limited access to ground-truth data, available only for the training set, as well as the complexity of the utilized scenes.

Liang et al. [26] introduced a vehicle detection and tracking system that integrated YOLOv3 for object detection and an enhanced KCF algorithm for tracking. This system extracted multiple features from vehicle objects, such as grayscale, Histogram of Oriented Gradients (HOG), Convolutional Network (CN), and color histogram features, and employed their integration to predict vehicle positions based on the maximum response. Evaluating this novel model on the KITTI dataset revealed an accuracy of 70.3, precision of 83.9, and recall of 85.6. However, the study faced limitations concerning the dataset's limited scale and the utilization of a single Central Processing Unit (CPU) and GPU for experimentation. The scalability of the algorithm to handle larger datasets and more complex scenarios was not evaluated. Moreover, the algorithm might encounter challenges when dealing with scenarios involving heavy occlusion or a high number of vehicles.

Jodoin et al. [27] developed an algorithm called The Urban Tracker, comprising three steps: blob extraction, blob tracking, and object tracking. The algorithm extracts foreground blobs and calculates blob models, tracks individual blobs across frames, resolves ambiguity, handles occlusion, and updates the object model. The algorithm's performance was evaluated on custom datasets from four streets (Sherbrooke, Rouen, St-Marc, and Rene-Levesque), achieving accuracy and precision values for each street. However, the evaluation had limitations due to a small number of tested video sequences, lack of camera calibration, and reliance on a single evaluation metric (CLEAR MOT). These limitations may have impacted the generalizability and accuracy of the results.

Liwei Zhang et al.[28] proposed a multimodal MOT method for simultaneous object tracking and trajectory recording. Their methodology involved object detection using NMS, motion information extraction, deep appearance feature learning, and object tracking through data association. By combining 2D and 3D object detection techniques, robust detection was achieved. Deep appearance features were learned using a pre-trained CNN, and object tracking relied on associating motion information and appearance features using distance metrics. A threshold controlled motion information confidence, while a binary indicator determined appearance feature relationships. Computational efficiency was optimized by considering objects exiting the scene after a certain number of frames. Evaluation of the KITTI tracking benchmark included a comparison with other algorithms and an ablation study. The method achieved a MOTA of 76.40% and MOTP of 83.50% in the comparison, and a MOTA of 72.36% and MOTP of 83.94% in the ablation study. It should be noted that while 3D object detection was used as auxiliary information, the method primarily focused on 2D tracking and did not fully exploit the potential of 3D object detection for autonomous driving. Further research is required to develop a comprehensive and integrated approach for 3D multitarget tracking, leveraging the capabilities of 3D object detection in complex environments.

4. Future Scope

Analysis of computational experimentation provided multifold opportunities for future research and developments in this domain.

The project of automation could be improved by customizing the simplistic centroid tracking algorithm (which currently uses Euclidean distances) and improving the model's efficiency through better training. Additionally, a custom front-end could be created for the project.[11]

Efforts like using a simple centroid tracking algorithm which currently employs Euclidean distances that might be used to increase the efficiency of the model.

Improvements in the accuracy of multi-object to ensure track algorithms remain a constant goal, and future insights can be gained by exploring new methods and techniques, including deep learning, reinforcement learning, and alternative sensors like Light Detection and Ranging (LIDAR) or radar.[12] To enhance accuracy, future studies can focus on incorporating deep learning algorithms or other sensors such as LiDAR or radar. Addressing the challenge of reducing computational power is crucial for implementing the system in traffic control centers with limited resources. Additionally, efforts can be made to enhance the system's capability in detecting wrong-way driving, a significant road safety concern. This can be achieved by utilizing better-quality cameras, optimizing the object detection model employed by Tensorflow, and extending the system's ability to detect and track various vehicle types. Furthermore, integration with other traffic management systems, such as traffic signal control systems, can be pursued to improve overall traffic management and minimize traffic violations [13, 22].

Future research holds the potential for enhancing the system's robustness by incorporating 3D information and camera calibration data, as well as developing advanced algorithms to handle crowd density variations, pedestrian dynamics, and parameter settings while detecting abnormal behavior in crowded scenes. Optimization of the algorithm and utilization of more powerful hardware can further improve real-time performance. Investigating deep learning techniques like graph neural networks to explicitly model object interactions is also valuable [13, 14]. Addressing the system's limitations requires future research to improve performance on resource-constrained embedded systems, refine the road lane detection algorithm to handle different road conditions and accurately detect curved roads, and reduce validation time for the MBCDD algorithm without compromising accuracy. Extensive studies should also explore system performance under challenging constraints such as varying CCTV camera resolutions, heavy traffic conditions, and high-speed vehicles. Developing effective solutions in these areas ensures robust system performance, and future research can also explore techniques for citizen detection and tracking [23].

The algorithm's performance in practical applications could be improved by addressing challenges such as handling occlusions and other challenging scenarios. In addition, the new algorithm could be integrated with other computer vision techniques, such as object detection, semantic segmentation, and instance segmentation, to enhance the overall performance of the MOT system. Privacy concerns related to the use of MOT systems could also be addressed by exploring techniques to protect privacy while still achieving accurate MOT results. [14, 15]

There are still open lines of research that need further exploration, despite the advances made in the field of MOT. For example, a good-performing algorithm that can handle occlusion and re-identification under different conditions such as rain, non-rain, day, night, and extreme light conditions needs to be found. Multi-view solutions have been proposed, but they need improvement to obtain results that can compete with single-view solutions. Furthermore, as MOT is applied not only in traffic environments but also in other domains, more research could be conducted in this area.[17]

The integration of solutions into commercial vehicles and making them widely available to the public should be the focus of future research. Additionally, the research could explore how to further improve the alarm system to reduce false alarms and improve the user experience.[18]

The performance of the tracker can be increased by combining it with tracker management algorithms to handle occlusions. The potential of using semantic instance segmentations as an alternative to conventional bounding box detections in other domains can be explored in future research. The robustness of the tracker can be improved by investigating the effect of different weather and lighting conditions. Finally, other objects can be considered, and their performance can be evaluated in different domains.[19]

Research can be extended ahead to address the limitations of the proposed model by enhancing the continuity between consecutive frames in the Low Framerate and Pan-Tilt-Zoom (PTZ) categories. Furthermore, the issue of large reflections from car headlights can be tackled to enhance the model's performance in the Night Videos category. To achieve improved results, a more advanced separation between the train and test sets can be developed in future studies. Additionally, real-time applications, particularly in surveillance and security domains, can benefit from extending the proposed model.[20]

Future work could focus on addressing the aperture problem and improving the algorithm's ability to detect small objects. The detection threshold could be adjusted to optimize the detection time, and a validation procedure could be implemented to increase the confidence of malicious behavior detection. Additionally, the information about block directions could be manually adjusted. Future tests could be performed on a larger variety of real-world data, including videos of wrong-way driving, to further evaluate the effectiveness of the algorithm.[21]

Exploring various deep learning techniques is one more way that leads to enhancing tracking accuracy and handling complex scenes more effectively. The algorithm's practicality and performance can be assessed by applying it to real-time tracking applications in surveillance and robotics.[24] Addressing the limitations of the proposed algorithm in future work involves reducing computational resource requirements and enhancing performance evaluation through the collection of more ground-truth data. Furthermore, extending the proposed approach to other areas, such as object detection and recognition, and integrating it with other deep learning models can lead to further performance improvements.[25]

The dataset is one more crucial challenge used in this study and can be expanded to include more complex and diverse scenarios to further evaluate the performance of the proposed algorithm. The scalability of the algorithm can also be tested on larger datasets with more complex scenarios. Other state-of-the-art object detection networks, such as EfficientDet and detection Transformers, can be used to compare their performance with the proposed algorithm. In addition, the proposed algorithm can be extended to handle multiple classes of objects and evaluate its performance on multi-class object tracking. Finally, the algorithm can be optimized for real-time processing, which is essential for practical applications such as autonomous driving.[26]

The performance of Urban Tracker and TI can be evaluated using other metrics such as the F1 score or IoU in future studies. Furthermore, how well they perform in different scenarios can be assessed by evaluating them on a larger dataset. Finally, future research can be conducted to improve the performance of Urban Tracker in scenarios with occlusions.[27]

To advance in this field, future work should focus on developing an adaptable 3D multitarget tracking system. This entails utilizing the accurate position and size estimation capabilities of 3D object detection to enhance the robustness and accuracy of object tracking. Furthermore, incorporating advanced techniques like sensor fusion, which combines data from multiple sensors such as LiDAR and cameras, can further improve tracking performance. The application of such a system would be highly valuable in autonomous driving, where precise and reliable tracking of objects in a 3D space is crucial for safe and efficient navigation. Continued research and advancements in 3D multitarget tracking will contribute to the development of more sophisticated and capable autonomous systems.[28]

5. Conclusion

The literature review on the Multi-Object Tracking (MOT) system reveals that there has been significant progress in developing various techniques for object tracking in recent years. Different approaches have been proposed to address the challenges of object tracking. The paper summarized the related work for MOT in general and traffic environment in particular along with its strengths and limitations. The review also provided the evaluation metrics and benchmark datasets in this context which will be helpful to pursue research ahead in this domain. With computational experimentation, it is observed that deep learning techniques are widely used for object tracking as it effectively makes use of feature extractions and yields better accuracy. Finally, we concluded the paper by providing future directions to work in this domain.

1. References and Footnotes

Author contributions

Hardik Jaiswal, Aditya Gambhir: Data generation, Methodology, Software, Evaluation, Writing-Original draft preparation

Laxmi Bewoor, Nagaraju Bogiri: Conceptualization, Methodology, Revising Original draft, Software, Evaluation and Validation.,

Conflicts of interest

The authors declare no conflicts of interest.

References

[1] Wang, Z., Zheng, L., Liu, Y., Li, Y., Wang, S. (2020). Towards Real-Time Multi-Object Tracking. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, JM. (eds) Computer Vision – ECCV 2020. ECCV 2020. Lecture Notes in Computer Science(), vol 12356. Springer, Cham. https://doi.org/10.1007/978-3-030-58621-8_7

[2] Madore KP, Wagner AD. Multicosts of Multitasking. Cerebrum. 2019 Apr 1;2019:cer-04-19. PMID: 32206165; PMCID: PMC7075496.

[3] Girshick, R., Donahue, J., Darrell, T., Malik, J., 2014. Rich feature hierarchies for accurate object detection and semantic segmentation, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, pp. 580–587. https://doi.org/10.1109/CVPR.2014.81

[4] Girshick, R., 2015. Fast R-CNN, in: Proceedings of the IEEE International Conference on Computer Vision. IEEE, pp. 1440–1448. https://doi.org/10.1109/ICCV.2015.169

[5] Ren, S., He, K., Girshick, R., Sun, J., 2017. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Trans. Pattern Anal. Mach. Intell. 39, 1137–1149. https://doi.org/10.1109/TPAMI.2016.2577031

[6] Redmon, J., Divvala, S., Girshick, R., Farhadi, A., 2016. You only look once: Unified, real-time object detection, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, pp. 779–788. https://doi.org/10.1109/CVPR.2016.91

[7] Ballard, D., Lecun, Y., Boser, B., Denker, J.S., Henderson, D., Howard, R.E., Hubbard, W., Jackel, L.D., 1989. Backpropagation Applied to Handwritten Zip Code Recognition.

[8] Shin, H.C., Roth, H.R., Gao, M., Lu, L., Xu, Z., Nogues, I., Yao, J., Mollura, D., Summers, R.M., 2016. Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning. IEEE Trans. Med. Imaging 35, 1285–1298. https://doi.org/10.1109/TMI.2016.2528162

[9] Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., Torralba, A., n.d. Learning Deep Features for Discriminative Localization.

http://cnnlocalization.csail.mit.edu

[10] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A.C., Fei-Fei, L., 2015. ImageNet Large Scale Visual Recognition Challenge. Int. J. Comput. Vis. 115, 211–252. https://doi.org/10.1007/s11263-015-0816-y

[11] Narayana Cummaragunta, S., S, S.K., Shetty, J., 2022. Wrong Side Driving Detection. Int. J. Technol. Emerg. Sci.

[12] Sentas, A., Kul, S., Sayar, A., 2019. Real-Time Traffic Rules Infringing Determination over the Video Stream: Wrong Way and Clearway Violation Detection, in: 2019 International Conference on Artificial Intelligence and Data Processing Symposium, IDAP 2019. https://doi.org/10.1109/IDAP.2019.8875889

[13] Zhang, J., Lo Presti, L., Sclaroff, S., 2012. Online multi-person tracking by tracker hierarchy, in: Proceedings - 2012 IEEE 9th International Conference on Advanced Video and Signal-Based Surveillance, AVSS 2012. pp. 379–385. https://doi.org/10.1109/AVSS.2012.51

[14] Li, W., Mu, J., Liu, G., 2019. Multiple object tracking with motion and appearance cues, in: Proceedings - 2019 International Conference on Computer Vision Workshop, ICCVW 2019. Institute of Electrical and Electronics Engineers Inc., pp. 161–169. https://doi.org/10.1109/ICCVW.2019.00025

[15] NGENI, F.C., Mwakalonge, J., Siuhi, S., 2022. Multiple Object Tracking (Mot) of Vehicles to Solve Vehicle Occlusion Problems Using Deepsort and Quantum Computing. SSRN Electron. J. https://doi.org/10.2139/ssrn.4183319

[16] Montella, C. (n.d.). The Kalman Filter and Related Algorithms: A Literature Review The Kalman Filter and Related Algorithms A Literature Review. https://www.researchgate.net/publication/236897001

[17] Jiménez-Bravo, D.M., Lozano Murciego, Á., Sales Mendes, A., Sánchez San Blás, H., Bajo, J., 2022. Multi-object tracking in traffic environments: A systematic literature review. Neurocomputing. https://doi.org/10.1016/j.neucom.2022.04.087

[18] 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. J. Traffic Transp. Eng. (English Ed. https://doi.org/10.1016/j.jtte.2020.10.002

[19] Bullinger, S., Bodensteiner, C., Arens, M., 2018. Instance flow-based online multiple object tracking, in Proceedings - International Conference on Image Processing, ICIP. pp. 785–789. https://doi.org/10.1109/ICIP.2017.8296388

[20] Sakkos, D., Liu, H., Han, J., Shao, L., 2018. End-to-end video background subtraction with 3d convolutional neural networks. Multimed. Tools Appl. 77, 23023–23041. https://doi.org/10.1007/s11042-017-5460-9

[21] Bazan, M., Ciskowski, P., Halawa, K., Janiczek, T., Rusiecki, A., Śmigowski, M., 2016. Telematics Telematics Transport System Transport System Archives of Detection of vehicles moving in the wrong direction.

[22] Helen Rose Mampilayil, Rahamathullah K, 2019. 2019 International Conference on Intelligent Computing and Control Systems, ICCS 2019. 2019 Int. Conf. Intell. Comput. Control Syst. ICCS 2019.

[23] Suttiponpisarn, P., Charnsripinyo, C., Usanavasin, S., Nakahara, H., 2022. An Autonomous Framework for Real-Time Wrong-Way Driving Vehicle Detection from Closed-Circuit Televisions. Sustain. 14. https://doi.org/10.3390/su141610232

[24] Li, X., Wang, K., Wang, W., Li, Y., 2010. A multiple object tracking method using Kalman filter, in: 2010 IEEE International Conference on Information and Automation, ICIA 2010. pp. 1862–1866. https://doi.org/10.1109/ICINFA.2010.5512258

[25] Zhao, D., Fu, H., Xiao, L., Wu, T., Dai, B., 2018. Multi-object tracking with correlation filter for autonomous vehicle. Sensors (Switzerland) 18. https://doi.org/10.3390/s18072004

[26] Liang, H., Song, H., Li, H., Dai, Z., 2020. Vehicle Counting System using Deep Learning and Multi-Object Tracking Methods. Transp. Res. Rec. 2674, 114–128. https://doi.org/10.1177/0361198120912742

[27] Jodoin, J.P., Bilodeau, G.A., Saunier, N., 2014. Urban Tracker: Multiple object tracking in urban mixed traffic, in: 2014 IEEE Winter Conference on Applications of Computer Vision, WACV 2014. pp. 885–892. https://doi.org/10.1109/WACV.2014.6836010

[28] 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. https://doi.org/10.1155/2020/8810340

[29] D. Du et al., "VisDrone-DET2019: The Vision

Meets Drone Object Detection in Image Challenge Results," 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), Seoul, Korea (South), 2019, pp. 213-226, doi: 10.1109/ICCVW.2019.00030.

[30] "Open Images V6," Googleapis.com [Online]. https://storage.googleapis.com/openimages/web/index.html.(accessed 31 May 2023)

[31] J. Ferryman and A. Shahrokni, "PETS2009: Dataset and challenge," 2009 Twelfth IEEE International Workshop on Performance Evaluation of Tracking and Surveillance, Snowbird, UT, USA, 2009, pp. 1-6, doi: 10.1109/PETS-WINTER.2009.5399556

[32] Zhu J, Yang H, Liu N, et al., 2018. Computer Vision – ECCV 2018, Online multi-object tracking with dual matching attention networks. Springer International Publishing. https://doi.org/10.1007/978-3-030-01228-1