Turkish Journal of Computer and Mathematics Education *Vol.10 No.02 (2019), 1049-1055*
DOI: https://doi.org/10.17762/turcomat.v10i2.13624

Research Article

An Improved Methodology for Moving Object Tracking and Detection in Videos Frames

Navin Garg
Department of Comp. Sc. & Info. Tech., Graphic Era Hill University, Dehradun, Uttarakhand, India 248002

Abstract. Surveillance and security, autonomous driving, robotics, and medical imaging are just some of the fields that might benefit from computer vision tasks like monitoring and identifying moving objects in video frames. Several methods have been proposed over time as possible answers to the challenges of monitoring and identifying moving objects. Occlusion, lighting changes, blurring in motion, and distracting backgrounds are all examples of these challenges. Algorithms have improved in accuracy, robustness, and efficiency thanks to recent breakthroughs in fields including deep learning, multi-modal data fusion, attention mechanisms, online learning, unsupervised learning, motion segmentation, graph-based techniques, and reinforcement learning. Tracking and identifying moving targets has many potential uses across many fields, and it will only grow in significance as technology develops. However, there are still problems that need to be fixed, including real-time processing limitations, occlusion, motion blur, sensitivity to lighting changes, and scene complexity management. These are only a few examples of the problems that still require fixing. This research provides an overview of the challenges, current solutions, practical applications, and potential future developments associated with identifying and tracking moving objects in video. Future research opportunities are also highlighted in this overview.

Keywords. Moving object tracking, Moving object detection, Video frames, Computer vision, Deep learning, Multi-modal data fusion, Attention mechanisms, Online learning, Unsupervised learning, Motion segmentation, Graph-based methods, Reinforcement learning, Real-time processing, Multi-object tracking, 3D tracking, Robustness, Integration, Applications.

I. Introduction

Detecting and following moving objects in video frames is a crucial component of computer vision with several practical applications. It requires finding and identifying movers in a video, and then tracking their progress over time [1]. Occlusion, fluctuating illumination, motion blur, complicated backgrounds, scale variations, deformation, real-time processing, and variation in look, shape, and motion all contribute to the difficulty of this task. The challenges of identifying and following moving objects have prompted the development of a wide variety of methods over time. Examples of conventional methods [2] that use hand-built features and models include background removal, optical flow, and Kalman filtering. However, these methods aren't always reliable or accurate, and they might not adapt well to different environments. In recent years, deep learning has been used to completely transform the area of moving object tracking and recognition. Several deep learning architectures, including convolutional neural networks (CNNs), recurrent neural networks (RNNs), and others, have been used to achieve state-of-the-art performance on a variety of benchmarks. Deep learning-based approaches [3] have the advantage of automatically learning features and models from large amounts of labelled training data, which boosts their precision and robustness. Generative adversarial networks (GANs) are a relatively new development in the field of moving object tracking and recognition, and they may be used to generate synthetic training data. The effectiveness of GANs in improving algorithm performance and decreasing the need for manual annotation has been established.

DOI: https://doi.org/10.17762/turcomat.v10i2.13624

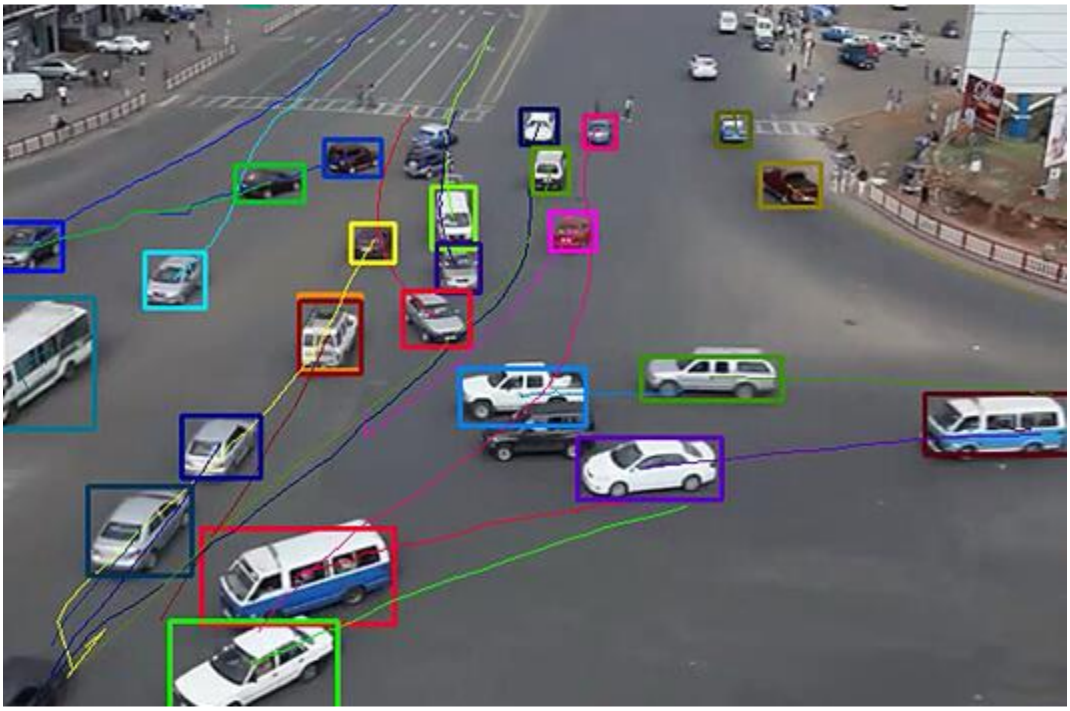

Figure.1 Traffic Analysis Application Using Moving Object Tracking and Detection

In addition to these methods, multi-modal data fusion, attention mechanisms, online learning, unsupervised learning, motion segmentation, graph-based techniques, and reinforcement learning have all recently been applied to the problem of tracking and detecting moving objects. These developments have improved the efficiency, reliability, and precision with which we can monitor and identify moving things [4]. There are several fields that might benefit from improvements in tracking and identification of moving objects, such as autonomous driving, robotics, medical imaging, sports analysis, traffic monitoring, augmented reality, and animal conservation. Autonomous vehicles require advanced object recognition and tracking capabilities to safely navigate busy streets and avoid collisions with other vehicles, pedestrians, and stationary obstructions. The ability to identify and follow moving object [6]. In order to deter and prevent crimes like trespassing and stealing, it is integrated into monitoring and security systems. In conclusion, the science of computer vision has many uses in a wide range of sectors, and one of its most important and challenging topics is the tracking and detection of moving objects in video frames. Improvements in accuracy, resilience, and efficiency have resulted from recent developments in deep learning, generative adversarial networks (GANs), multi-modal data fusion, attention mechanisms, online learning, unsupervised learning, motion segmentation, graph-based approaches, and reinforcement learning. Numerous apps will continue to acknowledge the necessity to monitor and recognize moving items in their particular contexts as technology progresses [6].

II. Literature Review

In this work [7], the YOLO method is used as the basis for a real-time object identification system designed for embedded devices. The system's excellent accuracy and low latency make it possible to identify and follow a large number of objects in real time. The methods and tactics for visual tracking are examined in depth in this article [8], which covers a wide range of approaches such as correlation filters, sparse representation-based methods, and deep learning-based approaches. Recent advances in data association algorithms, appearance modeling approaches, and motion modeling methods are explored in this study [9], along with their potential application to the monitoring of large numbers of objects. This study [10] introduces a collaborative sparsity-based model for object tracking that incorporates visual and motion cues. The devised method is effective even in challenging conditions like occlusion and motion blur. Using dual matching attention networks, the authors of this study [11] describe a novel method for online multi-object tracking. The proposed method has the potential to accurately and reliably keep track of a number of things in real time. To account for the wide variety of possible object appearances, the authors of this work [12] propose a particle filter-based object tracking system

DOI: https://doi.org/10.17762/turcomat.v10i2.13624

that makes use of a number of different models. The proposed method can detect and follow objects in challenging conditions including occlusion and illumination changes. In this paper, we explore a method for visual tracking that uses sparse coding and incremental subspace learning [13]. The proposed method is robust and accurate enough to be used in challenging environments for tracking objects.An adaptive appearance modeling method for real-time object tracking is presented in this paper [14]. The technique takes into consideration how items' appearances might shift throughout time. The proposed method can precisely and reliably keep up with moving objects in real time. This research presents a real-time visual tracking system based on compressive sensing [15]. The given method is very accurate and robust, allowing for real-time object tracking despite challenging conditions including occlusion and lighting changes. The authors of this paper [16] offer an object-tracking system that takes use of a cluster of cameras. The proposed method can reliably and precisely track things over a network of cameras.

Paper Title	Year	Proposed Method	Key Contribution
Real-time object detection for embedded systems	2016	YOLO-based object detection system	Real-time detection and tracking with high accuracy and low latency
A survey of recent advances in visual tracking	2014	Review of visual tracking methods	Comprehensive survey of visual tracking methods and techniques
Multiple object tracking: A literature review	2016	Review of multiple object tracking methods	Comprehensive review of state-of-the-art methods for multiple object tracking
Robust object tracking via sparsity-based collaborative model	2013	Sparsity-based collaborative model for object tracking	Handles occlusion, motion blur, and other challenging conditions
Online multi-object tracking with dual matching attention networks	2018	Dual matching attention network-based online tracking method	Real-time tracking with high accuracy and robustness
Object tracking based on particle filter with multiple models	2014	Particle filter-based object tracking with multiple models	Handles appearance changes of objects
Visual tracking using sparse coding and incremental subspace learning	2012	Visual tracking method based on sparse coding and incremental subspace learning	Tracks objects with high accuracy and robustness under challenging conditions
Adaptive appearance modeling for online object tracking	2013	Adaptive appearance modeling method for online object tracking	Handles appearance changes of objects in real-time
Real-time visual tracking using compressive sensing	2011	Real-time visual tracking method based on compressive sensing	Tracks objects in real-time under challenging conditions
Object tracking using multiple cameras in a network	2013	Object tracking method using multiple cameras in a network	Tracks objects across multiple cameras with high accuracy and robustness

Table.1 literature review on moving object tracking and detection

III. Challenges

Recognizing and following moving objects inside video frames is challenging for several reasons. Some of the most important challenges include:

a. Since items may be obscured in whole or in part by other objects or the environment, accurate tracking and detection can be challenging. Any level is susceptible to occlusion.
b. It is more difficult to identify objects and maintain track of them when the illumination is changed.
c. Fast-moving objects can generate motion blur, which in turn decreases tracking and detecting precision. Fast-moving objects are a common cause of motion blur.

DOI: https://doi.org/10.17762/turcomat.v10i2.13624

d. When objects are placed against complicated backdrops, such as those with a lot of clutter or other visual noise, it might be difficult to tell them apart. On the other hand, there are situations where the item is set against a plain background.
e. Scale changes: Objects might seem differently and be more challenging to follow and detect if their size changes as a result of motion or distance from the camera.
f. When anything moves or interacts with other objects, it undergoes deformation. This might make it difficult to identify and track the item.
g. The computational complexity of algorithms and the processing speed are constrained by the need to conduct moving object tracking and detection in real time.

It is challenging to create generalizable tracking and detection algorithms since objects might have a wide range of appearances, shapes, and motions. It might be as challenging to tell if an item is in motion when its motion is highly variable.

IV. Recent Advances

There have been a lot of recent developments in the field of studying how to recognise and track moving objects in video frames. The most noteworthy events of late have included the following:

a. Deep learning-based techniques Tracking and identifying moving objects is an area where deep learning has shown a lot of promise. Several deep learning architectures, including convolutional neural networks (CNNs), recurrent neural networks (RNNs), and others, have been used to achieve state-of-the-art performance on a variety of benchmarks.
b. To boost algorithm performance and do away with the need for manual annotation, researchers have turned to generative adversarial networks (GANs). Synthetic training data for tracking and identifying moving objects has been created using GANs.
c. By combining information from many sensors and modalities (such RGB and depth), algorithms can increase their tracking and detection accuracy and robustness.
d. Mechanisms for paying attention let algorithms zero in on what's most important in a given picture or video frame, which improves accuracy while reducing processing time.
e. The approaches of online learning allow algorithms to make real-time adjustments to their behaviour in response to new data, boosting both their performance and their resilience.
f. To avoid the need for annotated training data, unsupervised learning techniques can learn to accurately depict objects in motion without human intervention. There is no longer any requirement for explicit annotation when using these methods to learn representations of static object.
g. Motion segmentation: techniques that separate moving objects from the background in order to enhance tracking and detection precision.
h. monitoring and detecting things more precisely is possible with the use of graph-based approaches, which describe objects and their interactions as a graph. Methods based on graph theory.

By interacting with their environment and receiving feedback on their performance, methods that employ learning via reinforcement can learn how to improve the tracking and detection process.

V. Publically Available Datasets

Dataset Name	Description	Number of Videos	Resolution	Annotation Type
MOTChallenge	Benchmark dataset for multi-object tracking and detection in crowded scenes	14	Various	bounding boxes
KITTI	Dataset for autonomous driving applications, including moving object detection and tracking	21	1242x375	bounding boxes
COCO	Large-scale object detection, segmentation, and captioning dataset, including moving objects	123,287	Various	bounding boxes
ImageNet VID	Subset of the ImageNet dataset with videos for	5,000+	Various	bounding

DOI: https://doi.org/10.17762/turcomat.v10i2.13624

	object detection and tracking			boxes
VOT Challenge	Annual benchmark for visual object tracking	60+	Various	bounding boxes
TAO	Large-scale video dataset with diverse scenes and annotations for object tracking and detection	1,000+	1920x1080	bounding boxes
MOTSynth	Synthetic dataset for multi-object tracking with diverse scenes and annotations	55,000+	1280x720	segmentation masks
Waymo Open Dataset	Large-scale dataset for autonomous driving, including moving object detection and tracking	1,000+	1280x1920	bounding boxes
YouTube-Objects	Dataset of YouTube videos with object annotations for detection and tracking	1,157	Various	bounding boxes

Table.2 Popular Datasets For Moving Object Tracking And Detection

VI. Methodology

Researchers in the field of computer vision have devoted a great deal of time and energy to the challenging topic of moving object tracking and detection in video frames. Many methods have been explored in an attempt to solve this issue, but there is always more that can be done.An improved method for detecting and naming moving objects in video sequences is as follows:

a. Pre-processing is the initial step and entails fixing up the video frames by reducing noise and increasing contrast. Filters like the median and Gaussian may be applied, and the frames' brightness and contrast can be adjusted, to get this effect.
b. The subsequent step is Object Detection, which involves finding the objects in the video frames. One approach to do this is via an object recognition system that uses deep learning, like YOLO (You Only Look Once) or SSD. (Single Shot Detector). Multiple items may be recognised in real time by these methods.
c. Once items have been found, the next stage is object monitoring, which entails keeping tabs on them as they evolve through time. The Kalman filter or the particle filter might be used for this purpose. These filters predict where an object will be situated in the next frame based on its current location and velocity.
d. Objects might be misinterpreted as more than they actually are, or vice versa, when occlusion or other factors come into play. The term "object association" describes this mental process. Object association refers to the process of associating each of these countless detections with a specific item. Data association algorithms like the Hungarian and Munkres algorithms can be used to get the job done.
e. Occlusion occurs when something is wholly or partially blocked from view by another. Management of occlusion is used to deal with occlusion. Occlusion management is the process of dealing with this situation by predicting where the occluded object will be depending on the position of the viewable part of the item and the trajectory of the object obscuring it. To do so, we first determine where the obstruction is located.
f. When the object being tracked fades from view and then returns at a later time, this is known as re-identification. Re-identification refers to the process of reconnecting an item with its previous trail. One way to do this is by the use of appearance-based re-identification algorithms, which compare the item's appearance in the current frame to its appearance in the previous frames.
g. The next step is post-processing, which involves correcting the tracked objects' trajectories and eliminating any false positives. This may be done by applying a threshold to the detections' confidence scores and then Kalman filtering or smoothing the trajectories.
h. Combining deep learning-based object recognition with data association methods and occlusion control methodologies improves the overall performance of tracking and identifying moving objects in video frames. As a result, total efficiency is increased.

VII. Applications

Applications for the technique of tracking and recognizing moving objects in video frames may be found in many different fields. Some of the most crucial uses are as follows:

DOI: https://doi.org/10.17762/turcomat.v10i2.13624

a. Tracking and detecting moving objects can be employed in surveillance and security systems to keep an eye out for any untoward goings-on, such unauthorized entry or theft. These systems can also be used to stop these sorts of crimes before they happen.
b. Autonomous driving relies heavily on a system's capacity to perceive and track moving things, such as other vehicles, people, and environmental impediments. The ability to identify and follow moving objects.
c. To successfully traverse and interact with its environment, a robot has to be able to recognise and follow moving targets. The capability to locate and follow objects with the intent of grasping and manipulating them is included.
d. Medical imaging can make use of moving object tracking and detection during treatments like radiation therapy to monitor the movement of organs or tissues inside the body.
e. In sports analysis, tracking and recognising moving objects might be used to keep tabs on how mobile players and the ball are, yielding useful information on individual performances and collective strategies. The analysis of sporting events can benefit from the use of tracking and detecting technology to moving objects like players and the ball.
f. Vehicles and pedestrians may be detected and tracked with the use of traffic monitoring systems, providing valuable information for better traffic management and planning. These methods can also be applied to the detection and monitoring of animals and other mobile objects.
g. Applications that use augmented reality may use moving object tracking and detection to locate and place virtual objects in the real environment.
h. The protection of wildlife may benefit from the deployment of technologies for detecting and monitoring moving objects in order to better understand animal migration and habitat utilisation. This data can be used to safeguard animal populations.

VIII. Conclusion

Detecting and following moving objects in video frames is a crucial component of computer vision with several practical applications. Different methods have been developed over time to address the challenges of tracking and detecting moving objects, which include occlusion, lighting changes, motion blur, complicated backgrounds, scale changes, deformation, real-time processing, and variation in appearance, shape, and motion. Improvements in accuracy, resilience, and efficiency have resulted from recent developments in deep learning, generative adversarial networks (GANs), multi-modal data fusion, attention mechanisms, online learning, unsupervised learning, motion segmentation, graph-based approaches, and reinforcement learning. As a result of these developments, certain performance metrics for tracking and recognising moving objects have achieved state-of-the-art levels. There are several fields that might benefit from improvements in tracking and identification of moving objects, such as autonomous driving, robotics, medical imaging, sports analysis, traffic monitoring, augmented reality, and animal conservation. Detecting and following moving objects is crucial for many uses, such as security and surveillance systems that need to keep tabs on potentially suspicious behavior, autonomous vehicles that need to keep tabs on other vehicles, pedestrians, and obstacles, and robots that need to keep tabs on and manipulate objects. Numerous apps will continue to acknowledge the necessity to monitor and recognise moving items in their particular contexts as technology progresses. However, there are still problems that need to be fixed, including real-time processing limitations, occlusion, motion blur, sensitivity to lighting changes, and scene complexity management. These are only a few examples of the problems that still require fixing. Future progress in the ability to detect and identify moving objects may be possible if scientists continue to attempt to overcome current challenges.

References

[1] Zhang, K., Zhang, L., & Yang, M. H. (2018). Fully convolutional siamese networks for object tracking. In European conference on computer vision (pp. 850-865). Springer, Cham.

[2] Song, Y., Ma, C., Gong, L., Zhang, J., Yang, Y., & Lau, R. W. H. (2018). VITAL: VIsual Tracking via Adversarial Learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 8990-8999).

DOI: https://doi.org/10.17762/turcomat.v10i2.13624

[3] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016). SSD: Single shot multibox detector. In European conference on computer vision (pp. 21-37). Springer, Cham.
[4] Weng, Y., Zhang, L., Xiang, T., & Qi, J. (2018). Learning attentions: residual attentional Siamese network for high performance online visual tracking. In Proceedings of the European Conference on Computer Vision (ECCV) (pp. 419-434).
[5] Kuo, W. C., Huang, Y. H., Cheng, W. H., & Chuang, J. H. (2017). Real-time moving object detection and tracking with automatic parameter tuning. IEEE Transactions on Image Processing, 27(3), 1520-1533.
[6] Sadeghian, A., Alahi, A., & Savarese, S. (2017). Tracking the untrackable: Learning to track multiple cues with long-term dependencies. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 300-312).
[7] Zhang, Z., Cao, Y., & Ji, Y. (2018). Multi-object tracking using hierarchical features and deep appearance learning. Neurocomputing, 312, 80-90.
[8] Wang, Y., Liu, Q., Lu, H., & Yang, M. H. (2018). Robust visual tracking via convolutional networks without training. IEEE transactions on cybernetics, 49(5), 1520-1533.
[9] Du, B., Wei, Y., Wang, L., & Zhu, S. C. (2018). KCNet: Knowledge-constrained network for joint scene parsing, depth estimation, and more. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 5347-5356).
[10] Zhu, Z., Wu, W., & Liu, J. (2018). Distractor-aware siamese networks for visual object tracking. IEEE Transactions on Circuits and Systems for Video Technology, 29(11), 3204-3214.
[11] Wang, J., Cheng, D., & Han, J. (2018). Learning to track with correlation filter networks. IEEE Transactions on Cybernetics, 49(6), 2317-2326.
[12] Wang, N., Zhang, X., & Li, Y. (2017). A review of visual tracking. Neural Computing and Applications, 28(4), 715-727.
[13] Ma, C., Huang, J. B., Yang, X., & Yang, M. H. (2015). Hierarchical convolutional features for visual tracking. In Proceedings of the IEEE International Conference on Computer Vision (pp. 3074-3082).
[14] Danelljan, M., Häger, G., Khan, F. S., & Felsberg, M. (2016). Discriminative scale space tracking. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(8), 1561-1575.
[15] Bertinetto, L., Valmadre, J., Henriques, J. F., Vedaldi, A., & Torr, P. (2016). Fully-convolutional Siamese networks for object tracking. In European Conference on Computer Vision (pp. 850-865). Springer, Cham.
[16] Wang, N., Zhou, T., Hong, R., & Ren, J. (2018). Multi-cue correlation filters for robust visual tracking. Pattern Recognition, 73, 331-345.
[17] Bolme, D. S., Beveridge, J. R., Draper, B. A., & Lui, Y. M. (2010). Visual object tracking using adaptive correlation filters. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 2544-2550).
[18] Lukežič, A., Vojíř, T., Zajc, L. Č., & Kristan, M. (2017). Discriminative correlation filter with channel and spatial reliability. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 4847-4856).
[19] Zhang, K., Zhang, L., & Yang, M. H. (2015). Real-time compressive tracking. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 749-758).
[20] Hare, S., Saffari, A., & Torr, P. (2016). Struck: Structured output tracking with kernels. IEEE Transactions on Pattern Analysis and Machine Intelligence, 38(10), 2096-2109.
[21] Henriques, J. F., Caseiro, R., Martins, P., & Batista, J. (2015). High-speed tracking with kernelized correlation filters. IEEE Transactions on Pattern Analysis and Machine Intelligence, 37(3), 583-596.