IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

National Level Conference – AITCON 2K25

Adarsh Institute of Technology & Research Centre, Vita, Maharashtra

Vol. 12, Special Issue 1, March 2025

Real-time Video Segmentation for Object Tracking

Snehal Dudhane[1], Prof. S.S. Redekar[2]

Student, Computer Science & Engineering, AMGOI, Vathar, India[1]

Assistant Professor, Computer Science & Engineering, Vathar, India[2]

Abstract: In order to maintain the strong security for government offices, organization and individual offices need to work on the security and now a day it is a world-wide concern of maintaining the strong security. In this case the real time video surveillance system is crucial for detecting, tracking and monitoring the activity. This paper describes a comprehensive framework for object detection, tracking and recognizing the objects and it is designed specifically for security surveillance application using YOLO (You Only Look Once) method). This system uses the advanced deep learning algorithms, Component identification, foreground and background subtraction, median filtering to improve the performance. The newly proposed approach of software makes it possible to identify many objects within a dynamic scenario quickly and accurately since it employs YOLO, which detects objects in real time and is also able to track and recognize them. In order to resolve this problem, we propose a multi-stage system that combines three-dimensional and temporal information in order to improve the performance of the segmentation and its resistance to occlusion and variations in illumination. The conducted experiments that are performed on benchmark datasets indicate that the proposed method is faster and more accurate as compared to the existing and the most advanced methods. The system also raises the possible implementations of our findings, including car driving without a driver, interaction with a computer, and espionage.

Keywords: Object detection, Tracking, Segmentation, YOLO

I. INTRODUCTION

The security is a now day's word wide concern. The increasing interconnectedness of the word the corporate offices like government, corporation and private citizen they need to maintain the strong security for offices and homes. The traditional security and monitoring system are less accuracy so that the threats can grow day by day. The need for effective surveillance system that can provide detection, monitoring and tracking the objects from live surveillance system. Modern developments in computer vision and deep learning are changing the surveillance technology environment in response to this demand. The creation of the YOLO (You Only Look Once) algorithm is one of the major innovations in this sector [1]. The YOLO gives remarkable accuracy and speed in detecting and classifying in many item or objects in real time. It is a deep learning model for object identification and has gained the widespread adoption. The typical object identification methods involve multiple phases including region proposal and classification while the YOLO can produce faster results because it processes photos in single pass. The speed and precision of YOLO is good for dynamic real time application like video surveillance where the uses need to detect the objects quickly. It is a great and option for metropolitan cities when monitoring the public areas in real time requires the quick object detection.

The real time object detection and tracking can also have several difficulties like Variations in lighting, quick motions, and occlusions situations in which one object is obscured by other can affect the successful surveillance. To overcome these problem suggested strategy uses YOLO in combination with techniques like median filtering and background subtraction. Background subtraction aims in separating the foreground objects from the background objects. The median filter abrupt the changes and removes the noise present in video frame or photos. The system can increase the accuracy of object recognition and tracking in difficult situations by incorporating these techniques into a multi-stage architecture. The system checks the noise from incoming videos and removes it using median filtering technique. Then system passes the video into the next process. Additionally, the suggested system makes use of both temporal and spatial data, which are critical for monitoring objects over time and managing issues like object displacement and occlusions [2].

While temporal data (such as earlier frames) aids the system in maintaining continuity in tracking objects between frames, even when they briefly vanish due to occlusions, spatial data enables the precise localization of objects within a frame.
The system uses multi-stage architecture, which improves segmentation accuracy and lowers the errors. This research has potential uses that go beyond conventional surveillance systems. For instance, real-time object identification and tracking are essential for autonomous driving to navigate dynamic traffic conditions. Autonomous cars can make decisions and prevent crashes by using real-time prediction, vehicle, and obstacle detection. The real-time monitoring systems can improve user experiences in human-computer interaction by making interactive interfaces and human and object recognition if possible. The study highlights how real-time monitoring technologies can revolutionize security processes globally by offering dependable, quick, and precise detection capabilities.

In summary, YOLO's combination with image processing methods like background subtraction and median filtering provides a viable foundation for tracking and detecting objects in real time. This method can greatly improve the efficiency of surveillance systems in a variety of applications, from autonomous driving and human-computer interaction to urban security, by tackling the issues of occlusion, lighting variations, and speed. This framework has the ability to transform security procedures and help create safer public areas worldwide as the need for sophisticated, real-time monitoring systems increases.

II. LITERATURE REVIEW

In the recent years for real time video segmentation and tracking lot of new techniques and strategies are developed for increasing accuracy and efficiency. The mast track is one of the such method for temporal consistency between frames by combining video object segmentation and tracking using a recurrent neural network. [1] Swift Net is the another technique it achieves the high speed segmentation without comprising the accuracy parameter with the help of neural networks [2]. The object based unsupervised vised video segmentation approach does not require the labelled data for offering the reliable solution for the real time video segmentation. [3] The integration of optical flow and object identification has done well and it improves the real time video segmentation for the moving objects. [4] Other research findings reinforced the success of multi-scale supervision in complex cases of having differing object scale by improving detection and tracking of multiple objects at the same time [5]. In their study, [6] are of the view that, due to the increasing number of autonomous vehicles on the roads, real-time instance segmentation has become necessary for safety-critical applications such as monitoring cars and people in real-world traffic scenarios. Also, hybrid models that integrate deep learning and classical approaches are more effective in handling dynamic scenarios and occlusions [7]. To address the problem of speed and scalability, the use of transformer based architectures models for recognition and segmentation tasks is on the rise, particularly for large scale projects such as the [8] proposed. It has generally been concluded that recent advancements in the application of real-time video object segmentation and tracking algorithms are due to their integration with deep learning and efficient memory and hybrid technologies. Recent developments in this field will impact such areas as video surveillance, self-driving cars, and other systems working with a moving picture.

Table 1 ANALYSIS OF EXISTING SYSTEM

Authors	Paper Title	Advantages	Disadvantages	Future Work
Abba, S., Bizi, A. M., Lee, J. A., Bakouri, S., & Crespo, M. L.[9]	Framework for real-time object tracking, detection, and monitoring in security surveillance systems	- Offers the ability of detecting and tracking objects in real time.	Potential limitations with detection in complex environments. - May require high computational resources for real-time analysis.	- Improving model efficiency for lower-power devices. Expanding framework adaptability to various security contexts.
D. Lohani, C. Crispim-Junior, Q. Barthélemy, S. Bertrand, L. Robinault, L. Tougne Rodet[10]	Video surveillance as a means of perimeter intrusion detection: a review.	Video surveillance techniques and algorithm	Limited focus on specific surveillance contexts	Exploration of AI-based solutions for real-time analysis.
F. Dumitrescu, C.-A. Boiangiu, M.-L. Voncila[11]	Robust and fast people detection from RGB Images	Works well in RGB Images	Reduced performance in low light environment	Integration of multi-modal data (e.g., infrared, depth) for better accuracy.
H. Masood, A. Zafar, M.U. Ali, T. Hussain, M.A. Khan, U. Tariq[12]	Tracking of fixed shape moving object	Fixed Shape object Tracking	Fine tuning for different environment.	Integration with real-time systems for dynamic object tracking.

H. Qiao, X. Wan, Y. Wan, S. Li, W. Zhang[13]	A new change detection technique for segmenting and detecting natural disasters in video sequences	A new method for exploiting video sequences to detect natural disasters.	may encounter difficulties in a complicated or highly dynamic disaster.	Improving accuracy under various environmental conditions.
J. Wang, S. Simeonova, M. Shahbazi[14]	Multi-vehicle tracking and detection from unmanned aerial footage that is orientation- and scale-invariant	Efficient tracking and detection of several vehicles in aerial footage.	Performance may degrade in clustered environments.	Enhancing real-time performance for UAV systems.
J.-H. Park, K. Farkhodov, S.-H. Lee, K.-R. Kwon[15]	Visual object tracking in a virtual environment simulation using a DQN agent method based on deep reinforcement learning	DQN shows improved simulation.	High Cost for training	Exploring more efficient algorithms for faster training and inference.
R. Opromolla, G. Inchingolo, G. Fasano[16]	Cooperative UAV visual recognition and tracking in the air using deep learning	Focuses on cooperative UAVs for improved detection accuracy.	Dependence on high quality data.	Expansion to non-cooperative UAV detection.
S. Zhang, L. Zhuo, H. Zhang, J. Li[17]	Using instance-aware attention networks and multi-feature discrimination, object tracking in unmanned aerial vehicle footage	improves object identification by integrating an instance-aware attention network.	In situations that are cluttered or dynamic, performance may suffer.	Expanding to multi-object tracking in complex UAV scenarios.

III. PROPOSED SYSTEM

The main objective of this proposed project is to construct a system of video segmentation, that can perform the tasks of real time video segmentation and precise tracking of objects in situ. The whole system is organized as a multi-stage pipeline aimed at encompassing spatio-temporal information for improving segmentation executive order in difficult conditions such as occlusion, dynamic illumination and background noise. This framework is designed based on YOLO, which has demonstrated its effectiveness in tackling multiple class types in a single pass object detection regime while performing real time object recognition quite effectively. Added development in speed and precision will augment the rapid detection capability of YOLO and suit it more on zones of quick reaction such as autonomous driving and guarding. On the other hand, adaptive filtering and background subtraction techniques will be used to handle the issues in video feeds such as noise, the shifting background and continuous scene updating. The proposed system, due to these integrated system capabilities, is expected to be able to maintain the level of performance required for real time applications while ensuring high precision and robustness when applied to a range of diverse environments. The following fig 1 shows the working flow of system how data flows from one layer to another layer.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

National Level Conference – AITCON 2K25

Adarsh Institute of Technology & Research Centre, Vita, Maharashtra

Vol. 12, Special Issue 1, March 2025

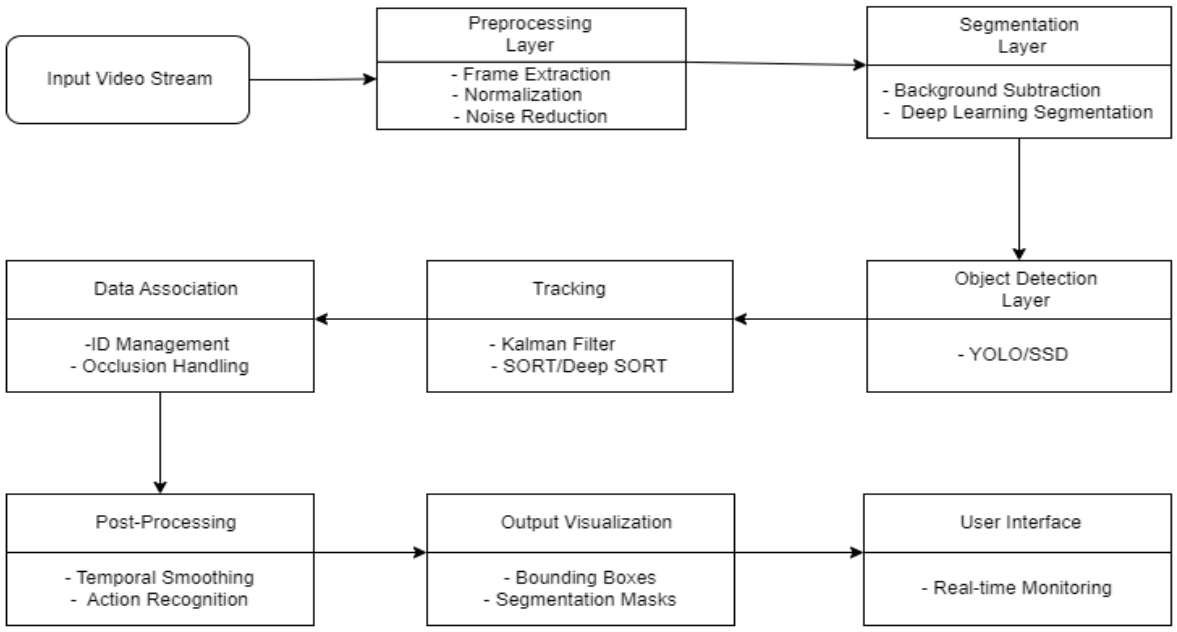

Fig. 1 Flow Chart of working

Working Methodology

The melding of the spatial and temporal features in the suggested multi-stage scheme makes it easier to deal with dynamic factors such as illumination variations and occlusions. To ensure high speed and precise multi-class detection in different environments, the focus is on the detection of objects using the YOLO (You Only Look Once) concept for real time object recognition in the first step of the system. This includes identifying moving objects through motion detection based on background subtraction techniques and dealing with some challenges like noise, shifting backgrounds and slow scene change by using adaptive filtering algorithms. The application of information from previous frames adds temporal coherence which improves the tracking of the objects and reduces false detections over segmentation which in turn improves the accuracy of the computed segmentation. The architecture as a whole is therefore optimized in such a way as to enhance speed, accuracy and efficiency of the system bearing in mind the need to maintain real-time requirements, with testing and tuning done for effectiveness on low resource devices. Finally, the accuracy of the system is determined among other things by the ability to track arbitrary targets of the system and its efficiency assessed on datasets MOT15, MOT16, and MOT17 against the existing techniques with speed and precision being the basis of comparison.

Median Filter :- Median Filter is used for removing noise from video frame.
Operation of Median Filter

1. Window of (Kernel) N * N is slid across each pixel in the image
2. Pixel values are arranged in ascending order.
3. Middle value of sorted list replaces the original pixel.
4. The window moves to the next pixel, and the process is carried out across the image.

Foreground Background Substraction :- Background substraction is a technique to separate the foreground objects from the background in images or videos. This technique is mostly used for object detection and tracking. Background subtraction is a computer vision technique used to separate foreground objects from the background in images or video sequences. It is widely applied in object detection and tracking, particularly in surveillance systems, traffic monitoring, and motion analysis. The core idea is to create a model of the background and then compare each new frame to this model to detect changes, which are considered as moving objects. Various approaches exist for background subtraction, including simple frame differencing, where consecutive frames are compared, and more advanced methods like Gaussian Mixture Models (GMM) that can adapt to dynamic backgrounds. While effective, background subtraction techniques often face challenges such as varying lighting conditions, sudden scene changes, and shadows, requiring adaptive algorithms or deep learning-based enhancements for improved accuracy.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

National Level Conference – AITCON 2K25

Adarsh Institute of Technology & Research Centre, Vita, Maharashtra

Vol. 12, Special Issue 1, March 2025

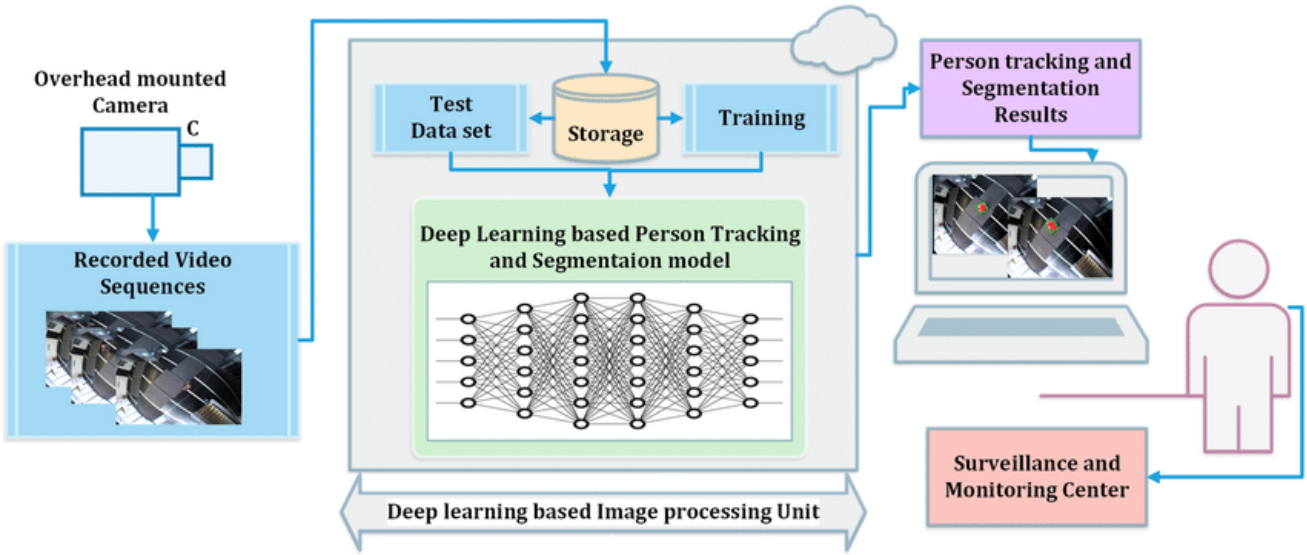

Fig. 2 System architecture

Contrast Limited Adaptive Histogram Equalization is referred to as CLAHE.
CLAHE enhances the local contrast of an image.
Steps:
1. Divide the image into small regions (tiles).
2. Compute the histogram for each tile.
3. Apply the following formula for each pixel:
4. CLAHE ensures no part of the image is over-amplified by clipping the histogram and redistributing clipped values.

Output Pixel = Histogram Equalization (min(pixel intensity, clip limit))……eq(1)

The clip limit ensures no histogram bin exceeds a predefined value.

Dataset

The Multi-Object Tracking Challenge includes the MOT15, MOT16, and MOT17 datasets, which provide benchmarks for assessing multi-target tracking systems. The MOT15 contains eleven film segments of urban pedestrians with comprehensive annotations. Complications are implied when MOT16 expands it to 14 sequences. The MOT17 dataset, which includes 14 sequences with more intricate interactions and occlusions, is the broadest. Every dataset adheres to standard performance metrics, like Multiple Object Tracking Accuracy (MOTA) and Precision (MOTP), which are helpful in advancing tracking accuracy in real-world scenarios.

Advantage & Disadvantage

Advantage
1. Accurate object detection
2. Real time processing
3. Improved motion Analysis

Disadvantages

1. Lighting and environmental changes
2. High computational cost
3. Latency issues

IV. CONCLUSION

The new advancements in technologies such as YOLO Algorithm, median filters and motion background modelling have enhanced object tracking and detection in dynamic situations. This technique improves on the effects of motion blur caused by fast motion, narrow fields of view, and low light conditions that are characteristic of CCTV surveillance systems. Furthermore, the system combines both spatial as well as temporal information to boost the performance of object tracking and segmentation and also aids in increasing the robustness of the system in practical scenarios. Such a system has great consequences in various areas including, but not limited to, man-machine interface, automotive industry, and urban security. In addition to that, this research highlights the usefulness of new more complex scenarios of work of the real-time surveillance system and its significance in enhancing the global practice of national security through the use of new and efficient means in tackling current threats, thus enhancing the safety of public zones and the order in which such zones are managed.

REFERENCES

[1]. Zhang, X., Li, X., & Hu, S. (2020). MaskTrack RNN: Real-Time Video Object Segmentation and Tracking. IEEE Transactions on Pattern Analysis and Machine Intelligence.
[2]. Yuan, X., Zhang, H., & Wu, Y. (2021). SwiftNet: Real-Time Video Object Segmentation. IEEE Transactions on Image Processing.
[3]. Yang, L., Liu, J., & Zhang, Y. (2020). Unsupervised Video Object Segmentation via Object Flow. IEEE Transactions on Image Processing.
[4]. Zhu, S., Han, S., & Li, Y. (2021). Video Object Segmentation via Object Detection and Optical Flow. IEEE Transactions on Image Processing.
[5]. Zhou, S., Sun, S., & Li, X. (2021). Multi-Object Tracking and Segmentation with Multi-Scale Supervision. IEEE Transactions on Pattern Analysis and Machine Intelligence.
[6]. Xie, L., Zhang, C., & Li, T. (2021). Real-Time Instance Segmentation and Tracking for Autonomous Driving. IEEE Transactions on Intelligent Vehicles.
[7]. Li, P., Chen, Z., & Zhang, J. (2022). Fast and Robust Object Tracking and Segmentation Using Hybrid Models. IEEE Transactions on Circuits and Systems for Video Technology.
[8]. Wang, Q., Zhang, W., & Sun, Q. (2023). Real-Time Object Detection and Segmentation with Transformer Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence.
[9]. Abba, S., Bizi, A. M., Lee, J. A., Bakouri, S., & Crespo, M. L. (2024). Real-time object detection, tracking, and monitoring framework for security surveillance systems. Heliyon, e34922. https://doi.org/10.1016/j.heliyon.2024.e34922.
[10]. Lohani, D., Crispim-Junior, C., Barthélemy, Q., Bertrand, S., Robinault, L., & Tougne Rodet, L. (2022). Perimeter intrusion detection by video surveillance: A survey. Sensors, 22(9), 3601. https://doi.org/10.3390/s22093601
[11]. Dumitrescu, F., Boiangiu, C.-A., & Voncila, M.-L. Robust and fast people detection from RGB images.
[12]. Masood, H., Zafar, A., Ali, M. U., Hussain, T., Khan, M. A., & Tariq, U. (Year). Tracking of fixed shape moving object.
[13]. Qiao, H., Wan, X., Wan, Y., Li, S., & Zhang, W. A new change detection technique for segmenting and detecting natural disasters in video sequences.
[14]. Wang, J., Simeonova, S., & Shahbazi, M. Multi-vehicle tracking and detection from unmanned aerial footage that is orientation- and scale-invariant.
[15]. Park, J.-H., Farkhodov, K., Lee, S.-H., & Kwon, K.-R. (2019). Visual object tracking in a virtual environment simulation using a DQN agent method based on deep reinforcement learning.
[16]. Opromolla, R., Inchingolo, G., & Fasano, G. (2020). Cooperative UAV visual recognition and tracking in the air using deep learning.

