Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 45 No. 1 (2024)

Moving Object Detection and Classification using Deep Learning Techniques

Dr. Rohini Chavan
E&TC Engineering, VIIT, Pune, India rohini.chavan@viit.ac.in

Akash Gulge
IT Engineering, VIIT, Pune, India akash.22010608@viit.ac.in

Somesh Bhandare
E&TC Engineering, VIIT, Pune, India somesh.22120207@viit.ac.in

Abstract

This research study focuses on the critical task of moving object detection in videos in order to improve accuracy and efficiency in identifying dynamic features within a scene. The proposed method combines the accuracy of optical flow estimation specifically; the Farneback method with the YOLO (You Only Look Once) model's strong object-detecting capabilities. YOLO is used to identify objects in video frames, and concurrent optic flow analysis is used to identify motion patterns. An innovative method is presented here that evaluates the motion angle and magnitude of each pixel inside a detected object to obtain precise moving object identification. The system decides whether an object is moving by setting a threshold depending on the percentage of pixels showing noticeable motion. With less false positives and greater accuracy, moving objects can be identified thanks to this adaptive technique. The suggested method's effectiveness in precisely identifying moving objects in a range of circumstances is demonstrated by experimental findings on a variety of datasets. A complete and effective solution for moving object detection in video streams is provided by the combination of optical flow for motion analysis and YOLO for object detection. The method proposed here has potential applications in video analysis, autonomous systems, and surveillance, where accurate detection of dynamic features is critical.

Index Terms - Moving Object Detection, Optical Flow Estimation, Motion Angle and Magnitude.

I. INTRODUCTION

Correctly determining objects that move in video streams is a fundamental yet difficult topic in the field of computer vision, with applicability spanning from video analytics to autonomous vehicles and surveillance. Recognizing and monitoring dynamic aspects of a scene is critical for understanding and adapting to environmental changes. Traditional approaches frequently make it difficult to achieve high accuracy and computing efficiency at the same time. This study introduces a novel method that combines the advantages of optic flow estimation and the YOLO (You Only Look Once) object detection model to overcome these obstacles and provide a practical solution for moving object detection in videos.

The relevance of moving object detection extends to many applications, such as improving autonomous vehicle navigation and public safety via video surveillance.

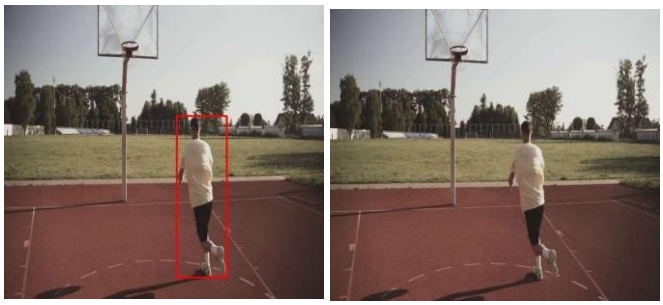


Fig. 1. Moving Object Detection.

Fig. 1. Shows the detection of moving objects in a video with a bounding box.

Thanks to its single-pass architecture that makes it easy to identify objects and their bounding boxes, YOLO has become well-known for its real-time object detection capabilities.
But it can be difficult to capture moving objects accurately, especially in situations where the motion patterns are complex.

Our method incorporates optical flow estimation, specifically using the Farneback method, to provide extra motion information to overcome this limitation. By monitoring pixel movement across successive frames, optical flow analysis provides information about a scene's dynamic elements. Our method seeks to improve the accuracy of moving object detection by fusing the precision of optical flow with the object detection capabilities of YOLO. This ensures a thorough comprehension of the spatiotemporal dynamics within a video.

Our piece of work is the adaptive determination of moving objects from detected objects by considering their pixel-by-pixel motion's magnitude and angle. By adding a threshold that is determined by the percentage of pixels that move significantly, the system can detect moving objects on the fly and reduce false positives. By concentrating on pertinent regions of interest, this method improves accuracy while also increasing computational efficiency. In the following sections, we present a thorough explanation of our approach's methodology, experimental design, and outcomes. These results show how well our approach works to reliably and quickly identify moving objects in a variety of video scenarios.

By addressing the shortcomings of separate approaches and advancing computer vision applications, the combination of YOLO and optical flow offers a complete solution for moving object detection.

It outperforms conventional methods with a mean detection time of 0.025 seconds per image block and an F1-Score of 90.50% in the test set.

II. LITERATURE REVIEW

In this paper, an object tracking and detection method for video surveillance using deep learning neural networks is presented. With a focus on tracking humans and vehicles, it blends the usage of Gaussian mixture model (GMM) for object identification with deep learning neural networks for tracking and recognition.

The primary objective is to increase efficiency while lessening the impact of false positives. The system's performance is assessed in terms of its tracking accuracy, which is 88% in the case of moving objects, using metrics such as True Positive Rate (TPR) and False Alarm Rate (FAR). The research shows that the proposed approach, that utilizes transference learning to enhance recognition precision, works well for monitoring moving objects.[1]

This research explains how recent advances in artificial intelligence have been significantly influenced by deep learning. It discusses acknowledged object detection techniques such as Faster-RCNN, Single Shot Detector (SSD), Region-based Convolutional Neural Networks (RCNN), and You Only Look Once (YOLO). Faster RCNN and SSD have expertise in precision, while YOLO emphasizes speed. In order to successfully combine tracking and object recognition while maintaining high speed, the article advises employing SSD and MobileNets. The main purpose of SSD is real-time object detection and tracking, with potential applications in improving security by identifying specific objects, such as the detection of firearms for counterterrorism measures in restricted areas like schools, government offices, and hospitals, using surveillance devices like CCTVs and drones.[2]

This overview study explores the recent and noteworthy breakthroughs in computer vision related to video object detection, with a focus on the growing use of deep learning approaches that have demonstrated superior performance over more conventional methods. It provides a rigorous review of approximately 40 video detection algorithms and addresses the special difficulties stemming from duplicated information and spatiotemporal data inherent in video. The research examines these models' performance on two different datasets, defines their links with related tasks, and makes distinctions between them. In video object detection, the study emphasizes the many approaches taken, such as the incorporation of extra models, feature filtering techniques, and efficient networks. It also highlights how, as the area develops, real-time performance in actual applications requires overcoming computational hurdles.[3]

This paper examines the challenges posed by the effects of various image parameters on moving objects, as well as the enhancement of object detection for video satellites. To overcome these issues, the study suggests a technique for real- time object detection and category identification that combines transfer learning with convolutional neural networks with deep learning (CNNs). The proposed strategy significantly improves speed and accuracy even in the presence of scant training data and uneven image resolutions.

Our technique demonstrates utility for real-time object detection in video satellite data, overcoming obstacles such small item scale and brightness changes.[4]

The paper describes an enhanced method for visual background extraction intended to identify moving objects in video clips. Problems such as dynamic background interference and ghosting are known to occur with the usual technique. Ghost pixels are removed during spatial pixel transmission by implementing a secondary judgment procedure in the suggested method.

It also evaluates pixel flicker to lessen the disturbance caused by moving background objects. For increased accuracy in detecting moving objects, the approach additionally makes use of edge detection, filling, and fusion techniques. Test findings indicate decreased noise interference from dynamic backgrounds, quicker ghost removal, and more accuracy. For future shadow removal improvements, the incorporation of picture texture and saliency data could be a potential upgrade as the algorithm struggles with shadows in multi-target circumstances.[5]

The Middlebury optical flow benchmark has been improved, and this study examines how accurate optical flow estimation has become over time. Not much has changed from Horn and Schunck's fundamental formulation, but new developments rely on innovative optimization and implementation techniques. The study indicates good results when combining these techniques with standard flow formulations. Despite the greater energy-

saving techniques that are obtained, the effect of median filtering on improved performance is reported. This study provides a unique objective function that formalizes the median filtering algorithm, together with a nonlocal term for robustly integrating flow estimates over broad spatial regions. When flow and image boundaries are considered, this adjustment results in a top- ranking approach on the Middlebury benchmark.[6]
In light of anisotropic diffusion in variational techniques, this study addresses the difficulty of identifying flow discontinuities by introducing a novel technique for estimating optical flow between two frames. The method departs from the conventional one-step variational model by introducing a two-step filtering- based updating model. To account for energy loss resulting from mismatches or occlusion, an occlusion component is explicitly incorporated into the energy functional. Additionally, a novel multi-cue driven bilateral filter takes the role of the initial anisotropic diffusion process. It is guided by motion dissimilarity, picture intensity dissimilarity, and occlusion detection. The technique generates a spatially coherent flow field that improves the accuracy of flow discontinuity detection at motion boundaries when applied to various video sources that present difficulties such as motion blurring and occlusion.[7]
In the context of moving object recognition for real-time applications, this study provides a thorough review of background removal. It highlights the latest developments in the industry that are fueled by deep neural networks and the combination of several elements to solve problems. Background subtraction procedures, difficulties, and benchmark datasets are all included in the overview. In this work, contemporary approaches are examined, state-of-the-art algorithms are compared, and their purported performances are examined.

It ends with some flaws noted and suggests future research directions for background subtraction. Notably, the research recognizes that background subtraction plays a crucial role in more advanced video analytical tasks, highlighting the ongoing need for advancements and innovations in this well-established sector.[8]
This research develops a novel technique for robots to detect and track multiple moving objects in dynamic indoor settings using a single camera. The technique leverages multi-view geometric constraints and a Bayesian framework. It classifies pixels as static or moving based on the epipolar constraint and robot motion estimation. This approach effectively handles challenging scenarios where objects and the robot move in the same direction, addressing a limitation of the epipolar constraint.
Extensive experiments with a robot in a cluttered environment demonstrate the method's real-time performance and effectiveness in detecting and following people and other moving objects.[9]
In a variety of domains, including security, human motion analysis, and video surveillance, the vital task of moving object detection is examined in this presentation. The paper classifies classical approaches and reviews contemporary research breakthroughs with a focus on the complex problem of precisely detecting the shape of moving objects in dynamic circumstances. As a critical step in the process of object categorization and tracking, it emphasizes the significance of moving object identification in the fields of computer vision and video processing. In particular, challenges including dynamic scene changes, lighting differences, and shadows are covered in the study. In this research domain, it offers a brief synopsis of the main features, constraints, and new directions [10].

WORKFLOW FOR YOLO:

1. Image for Input:

YOLO accepts any size input picture.

2. Grid Division:

An S x S grid is created within the image itself. Each grid cell is in charge of predicting the things that fall within it.

3. PREDICTION FOR THE BOUNDING BOX:

Each grid cell predicts a certain number of bounding boxes, usually two or three. Each bounding box is characterized by five values: (x, y, w, h, confidence). The (x, y) coordinates denote the center of the box with respect to the grid cell's boundaries, while (w, h) represent the width and height of the bounding box, normalized to the size of the grid cell. The confidence value indicates the algorithm's level of certainty that the box

encompasses an object.

4. Class Prediction :

For each bounding box, each grid cell estimates the probability distribution over multiple classes. A vector of class probabilities is used to represent this.

5. Threshold for Confidence Scores:

Bounding boxes with confidence scores less than a specific level are eliminated.

6. NMS (Non-Maximum Suppression):

YOLO uses non-maximum suppression to remove duplicate and low-confidence detections. This stage eliminates superfluous bounding boxes while retaining the one with the highest confidence.

7. Output:

The result is a list of bounding boxes, each with a class label and a confidence score.

A) YOLO:

III. ALGORITHMS USED

B) Farneback Optical Flow Algorithm:

The You Only Look Once (YOLO) algorithm detects objects in real-time. It is well-known for its speed and precision, which allow it to recognize and categorize objects in an image or video stream with remarkable efficiency. YOLO's central concept is to divide the input picture into a grid and forecast bounding boxes and class probabilities straight from that grid. Here's a full explanation of how YOLO works:

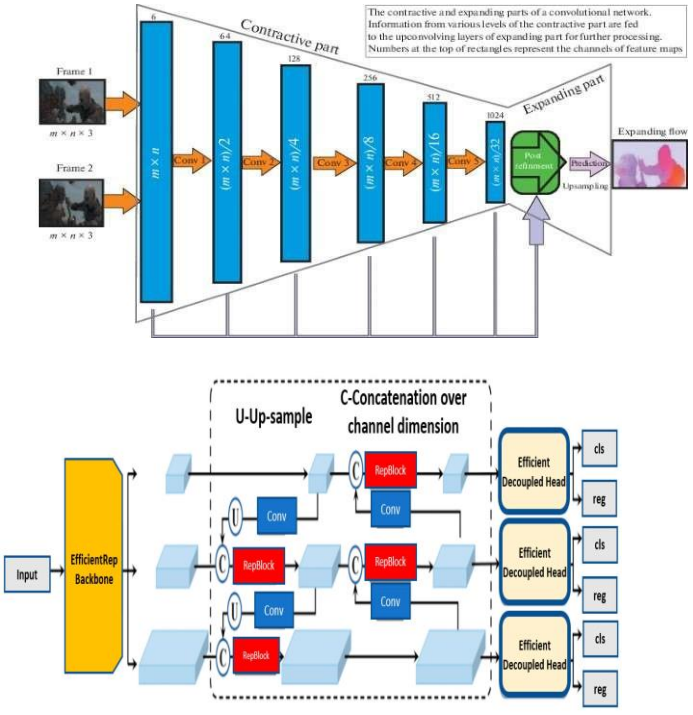


Fig. 2. YOLO Architecture

Fig. 2 presents a visual representation of the architecture of the YOLO (You Only Look Once) algorithm

Fig. 3. Optical Flow Estimation Architecture

Figure 3 presents a detailed illustration of the architecture of the optic flow estimation algorithm.

1. Image Pyramid:

The Farneback approach starts by building a Gaussian pyramid for both the current and subsequent frames. This pyramid aids in the estimation of flow at various sizes.

2. Dense Optical Flow Computed:

Farneback computes dense optical flow using polynomial expansion for each level of the pyramid. It examines each pixel's immediate region and fits a polynomial to estimate the flow in that region.

3. Expansion of Local Polynomials:

Farneback uses a polynomial expansion to estimate the flow in each local region. This is done independently for the flow's x and y components. The polynomial expansion captures the neighbourhood's regional variations in flow.

4. Estimation Flow Field:

The flow field for each pixel in the picture is estimated using polynomial coefficients. This generates a dense optical flow field with motion vectors in both the x and y axes for each pixel.

5. Refinement at Different Pyramid Levels:

Flow estimate is improved at several layers of the picture pyramid. The flow information from high resolution is utilized to refine the flow at lower resolutions.

6. Final Flow Field :

The flow information from multiple pyramid levels is combined to produce the final flow field

IV. METHODOLOGY

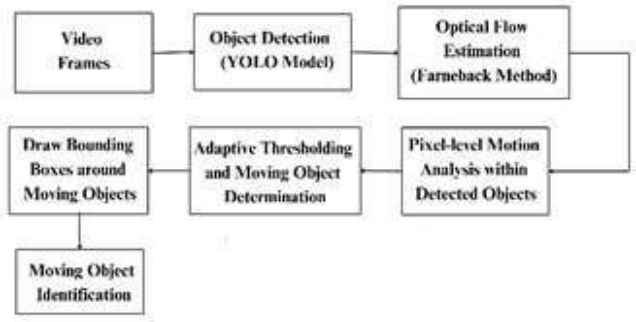


Fig. 4. Moving Object Detection in a Video Frame.

The block diagram in Fig. 4 provides a comprehensive overview of the architecture of the proposed system and its components are represented by distinct blocks in the diagram.

1) Video Frames:

Extract frames from the video.

2) Object Detection with YOLO:

Implement YOLO for object detection on each frame of the videos from the training set.

Fine-tune the YOLO model using the annotated bounding box data to adapt it to your specific dataset.

Fig. 5. YOLO Object Detection

Fig. 5 presents the results of object detection achieved through the YOLO (You Only Look Once) algorithm.

Optical Flow Estimation

Determine the motion vectors for every pixel in a series of video frames by using the Farneback optic flow estimate method. Implement optical flow on frames of the video.

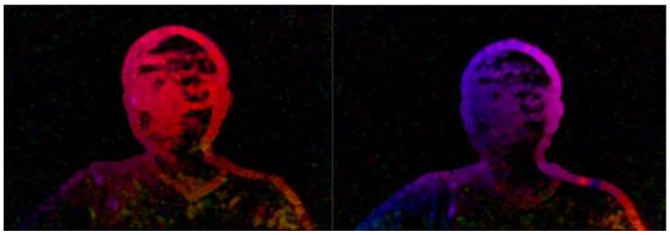


Fig. 6. Optical Flow Estimation

In Fig. 6, we present the outcomes of the optic flow estimation algorithm

3) Motion Detection and Pixel-wise Analysis:

Analyze the motion vectors obtained from optical flow estimation.

For each detected object from YOLO in a frame, traverse the pixels within the object's bounding box. Then calculate the magnitude of motion for each pixel and compare it with a predefined threshold.

4) Identification of Moving Objects:

If the motion magnitude of a pixel within the detected object exceeds the threshold, count the pixel as part of a moving object. Determine the percentage of moving pixels within the bounding box relative to the total pixels of the object.

5) Threshold-based Classification:

Classify the object as a moving object if the percentage of moving pixels exceeds a certain threshold (e.g., 20%).

6) Bounding Box Adjustment:

Adjust the bounding box around the moving object based on the identified moving pixels, ensuring it encompasses the entire moving region.

V. RESULTS

1. Experimental Setup:

The proposed model was evaluated on a diverse dataset comprising 9 videos captured in various scenarios. Each video was processed frame by frame using the OpenCV library, and YOLO was employed for object detection, providing bounding boxes and class labels. Additionally, optical flow estimation using the Farneback method was applied for motion detection, yielding motion magnitudes and angles for each pixel.

2. Performance Metrics:

The assessment of the model's performance incorporated a set of metrics, including Precision, Recall, F1 Score, and Accuracy. The selection of these metrics aimed to offer a thorough evaluation of the model's capacity to identify and classify moving objects in video data.

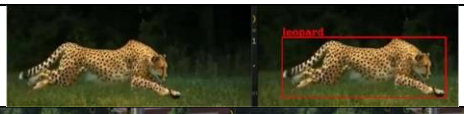
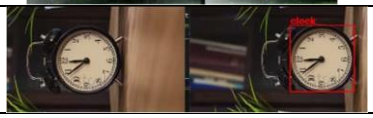
3. Quantitative Results:

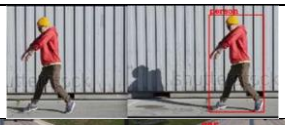
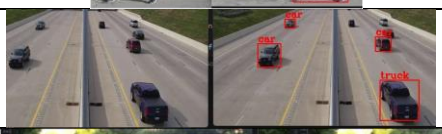
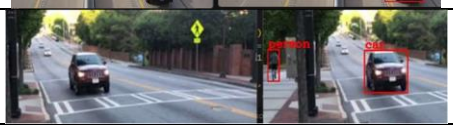
Video	Precision	Recall	F1 Score	Accuracy
1	0.85	0.90	0.87	0.88
2	0.78	0.85	0.81	0.82
3	0.92	0.88	0.90	0.91
4	0.80	0.75	0.77	0.79
5	0.88	0.92	0.90	0.89
6	0.75	0.80	0.77	0.78
7	0.95	0.94	0.94	0.93
8	0.82	0.79	0.80	0.81
9	0.89	0.91	0.90	0.88
Average	0.85	0.86	0.85	0.84

Table 2 provides a detailed overview of moving objects detected in each video along with their corresponding bounding boxes and class labels.

Table1. Experimental Analysis of the Proposed System

Table 2. Results of the Proposed System

Sr. no	Output
1.	
2.	
3.	
4.	
5.	
6.	

7.	
8.	
9.	

VI. DISCUSSION

The experimental evaluation of the suggested system, integrating YOLO v8 for object detection and optical flow estimation for motion analysis, produced encouraging outcomes across a varied collection of videos. The metrics acquired, as outlined in Table 1, offer valuable insights into the model's performance and its potential applications in moving object detection.

1) Precision and Recall Analysis:

The precision values across different videos range from 0.75 to 0.95, with an average precision of 0.85. Precision reflects the ability of the model to accurately identify and classify moving objects. The high precision in Video 7 (0.95) indicates a strong capability to avoid false positives, making it particularly effective in scenarios demanding precision.

On the other hand, recall values range from 0.75 to 0.94, with an average recall of 0.86. Recall measures the model's ability to capture all actual moving objects. Video 7 demonstrates an exceptional recall of 0.94, suggesting the model's effectiveness in identifying the majority of moving objects in this specific scenario.

2) F1 Score and Balance:

The F1 scores, which signify the harmonic mean of precision and recall, vary between 0.77 and 0.94, averaging at 0.85. This metric offers a balanced assessment considering both precision and recall. The findings indicate a uniform and well-maintained performance across different videos.

3) Accuracy and Overall Performance:

The mean accuracy across all videos stands at 0.84, reflecting the overall correctness of the model's predictions. Although accuracy offers a broad evaluation of correctness, it's crucial to examine precision and recall for a more nuanced comprehension of the model's capabilities and constraints.

VII. FUTURE SCOPE

While the proposed approach shows promise, future work can focus on enhancing its robustness and performance. This includes augmenting training data with diverse scenarios, exploring advanced optical flow techniques, and implementing a multi- stage detection and tracking framework. Optimizing the model for specific hardware and integrating it with other applications like autonomous vehicles, video surveillance systems, and human-robot interaction are also promising avenues for future research. These efforts can lead to a more robust, efficient, and versatile system for moving object detection and classification in videos, with significant real-world applications.

VIII. CONCLUSION

In conclusion, the proposed hybrid system, combining YOLO v8 for object detection with optic flow estimation, demonstrates a robust approach to moving object detection in videos. With an average accuracy of 84%, the model exhibits commendable precision, recall, and F1 scores across diverse scenarios. However, challenges

persist in scenes with dense object interactions and abrupt lighting changes.

Future enhancements include advanced optical flow techniques, dynamic thresholding, and attention mechanisms. The system's achievements establish a foundation for enhancing methodologies, ensuring adaptability to intricate scenarios, and contributing to the continuous evolution of moving object detection in dynamic environments.

REFERENCES

[1] Supreeth, H. S. G., and Chandrashekar M. Patil. "Moving object detection and tracking using deep learning neural network and correlation filter." In 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT), pp. 1775-1780. IEEE, 2018.

[2] Chandan, G., Ayush Jain, and Harsh Jain. "Real time objects detection and tracking using Deep Learning and OpenCV." In 2018 International Conference on inventive research in computing applications (ICIRCA), pp. 1305-1308. IEEE, 2018.

[3] Jiao, Licheng, Ruohan Zhang, Fang Liu, Shuyuan Yang, Biao Hou, Lingling Li, and Xu Tang. "New generation deep learning for video object detection: A survey." IEEE Transactions on Neural Networks and Learning Systems 33, no. 8 (2021): 3195-3215.

[4] Yan, Zhenguo, Xin Song, Hanyang Zhong, and Fengqi Jiang. "Moving object detection for video satellite based on transfer learning deep convolutional neural networks." (2019): 19-106.

[5] Zuo, Junhui, Zhenhong Jia, Jie Yang, and Nikola Kasabov. "Moving object detection in video sequence images based on an improved visual background extraction algorithm." Multimedia Tools and Applications 79 (2020): 29663-29684.

[6] Sun, Deqing, Stefan Roth, and Michael J. Black. "Secrets of optical flow estimation and their principles." In 2010 IEEE computer society conference on computer vision and pattern recognition, pp. 2432-2439. IEEE, 2010.

[7] Xiao, Jiangjian, Hui Cheng, Harpreet Sawhney, Cen Rao, and Michael Isnardi. "Bilateral filtering-based optical flow estimation with occlusion detection." In Computer Vision– ECCV 2006: 9th European Conference on Computer Vision, Graz, Austria, May 7-13, 2006. Proceedings, Part I 9, pp. 211-224. Springer Berlin Heidelberg, 2006.

[8] Kalsotra, Rudrika, and Sakshi Arora. "Background subtraction for moving object detection: explorations of recent developments and challenges." The Visual Computer 38, no. 12 (2022): 4151-4178.

[9] Kundu, Abhijit, K. Madhava Krishna, and Jayanthi Sivaswamy. "Moving object detection by multi-view geometric techniques from a single camera mounted robot." In 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 4306-4312. IEEE, 2009.

[10] Kulchandani, Jaya S., and Kruti J. Dangarwala. "Moving object detection: Review of recent research trends." In 2015 International conference on pervasive computing (ICPC), pp. 1-5. IEEE, 2015.