Video background subtraction from moving camera

A report on
Computer Vision Lab Project
[CSE-3181]

Submitted By
Ishaan Nagal -210962058
Avi Singh 210962034
Devansh Bhardwaj-210962060

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
MANIPAL INSTITUTE OF TECHNOLOGY,
MANIPAL ACADEMY OF HIGHER EDUCATION

November-2023

Video background subtraction from moving camera

Ishaan Nagal[1], Devansh Bhardwaj[2], Avi Singh[3]

[1]CSE MIT MANIPAL, India

[2] CSE MIT MANIPAL, India

[3] CSE MIT MANIPAL, India

[1] Ishan.nagal131@gmail.com; [2] avisingh0257@gmail.com; [3] Devanshbhardwaj2002@gmail.com

Abstract— This project delves into the development of an advanced system for dynamic background subtraction in videos recorded by moving cameras. Conventional techniques face challenges in scenarios with continuous camera movement, such as those encountered in vehicles or drones. The primary objectives include the implementation of adaptive algorithms for real-time foreground detection and segmentation. Leveraging cutting-edge computer vision and machine learning approaches, the system aims to dynamically model the background, ensuring robustness in various dynamic environments. Applications span across surveillance, autonomous vehicles, and augmented reality, contributing to advancements in computer vision technology.

Keywords— Dynamic Background Subtraction, Moving Cameras, Adaptive Algorithms, Real-time Processing, Foreground Detection, Computer Vision, Machine Learning, Video Surveillance, Autonomous Vehicles, Augmented Reality.

I. INTRODUCTION

In the evolving landscape of computer vision, there is a growing demand for robust systems capable of dynamic background subtraction from videos captured by moving cameras. Traditional methods often encounter limitations when faced with the complexities introduced by continuous camera movement, prevalent in scenarios involving vehicles or drones. This project aims to address this technological challenge by developing an advanced system equipped with adaptive algorithms designed for real-time foreground detection and segmentation.

The core challenge lies in crafting algorithms that seamlessly adapt to the ever-changing visual landscape resulting from continuous camera motion. These algorithms must dynamically model the background, ensuring precision and reliability in diverse and dynamic environments. The project places a strong emphasis on leveraging state-of-the-art techniques from the fields of computer vision and machine learning to achieve this goal.

The implications of this research span various industries, including surveillance, autonomous vehicles, and augmented reality. A successful outcome has the potential to redefine the capabilities of computer vision technologies, enhancing their effectiveness in real-world applications where dynamic visual data is prevalent.

As we navigate through the intricacies of dynamic background subtraction, this project aims to contribute not only to academic discourse but also to the practical implementation of advanced computer vision systems adept at handling challenges posed by moving cameras. Subsequent sections will delve into specific objectives, methodologies, and expected outcomes of this endeavour.

II. LITERATURE REVIEW

Background subtraction is a fundamental task in computer vision, with a wide range of applications in video surveillance, human-computer interaction, and video editing. The goal of background subtraction is to separate the foreground (e.g., people, objects) from the background in a video sequence.

Traditional background subtraction methods typically model the background as a pixel-wise distribution, and then identify foreground pixels as those that deviate significantly from the background distribution. However, these methods can often fail to accurately segment foreground pixels in challenging scenarios, such as when the foreground and background have similar appearances, or when the foreground is moving rapidly.

In recent years, deep learning-based methods have shown promising results for background subtraction. These methods typically learn a model from a dataset of training videos, and then use the model to predict whether each pixel in a new video frame belongs to the foreground or background.

One of the challenges of deep learning-based background subtraction methods is that they need to be able to model the complex relationships between pixels in a video frame. For example, if a person is walking across a scene, the pixels belonging to the person need to be segmented as foreground even if they have similar appearances to the pixels in the background.

Reference [1] GNN-based background subtraction method. Their method first constructs a graph to represent the relationships between pixels in a video frame. The graph is constructed such that each node in the graph represents a pixel and each edge in the graph represents the similarity between two pixels.
Graph neural networks (GNNs) are a type of deep learning model that are specifically designed to model graph-structured data. The authors then propose a GNN-based model that learns to predict whether each pixel in a video frame belongs to the foreground or background. The GNN model takes the graph as input and outputs a probability score for each node in the graph. The probability score represents the likelihood that the corresponding pixel belongs to the foreground. The authors' proposed method is a significant contribution to the field of background subtraction. It is the first GNN-based background subtraction method to be proposed, and it outperforms state-of-the-art methods on all challenging benchmarks. The authors' method is also able to handle motion blur, which is a challenging problem for traditional background subtraction methods.

Reference [2] One of the main challenges of sparse representation-based tracking methods is that they are sensitive to occlusion. When the target object is occluded by other objects in the scene, the sparse representation of the target object will change, which can lead to tracking failure. Xing et al propose a novel tracking method that exploits both partial information and spatial information of the target object. Their method first divides the target object into several local patches. Then, it represents each patch as a sparse linear combination of basis vectors. The basis vectors are learned from training data or constructed from the target object itself. The authors then propose a novel alignment-pooling method to aggregate the sparse representations of the local patches into a single representation of the target object. The alignment-pooling method considers the spatial relationships between the local patches, which helps to improve the robustness of the tracking method to occlusion. Finally, the authors propose an adaptive template update strategy to adapt the target object model to its appearance changes over time. The authors' proposed method is a significant contribution to the field of visual tracking. It is the first tracking method to exploit both partial information and spatial information of the target object. The method is also able to handle occlusion, which is a challenging problem for traditional tracking methods. Comparison with Related Work The authors' proposed method is related to several other tracking methods that exploit sparse representation. However, their method differs from these other methods in several key ways the authors' proposed method exploits both partial information and spatial information of the target object. This makes the method more robust to occlusion than other tracking methods that only exploit partial information, the authors' proposed method uses a novel alignment-pooling method to aggregate the sparse representations of the local patches into a single representation of the target object. This alignment-pooling method takes into account the spatial relationships between the local patches, which further improves the robustness of the method to occlusion, the authors' proposed method uses an adaptive template update strategy to adapt the target object model to its appearance changes over time. This makes the method more robust to appearance changes than other tracking methods that use fixed target object models. In conclusion the authors' proposed method is a significant contribution to the field of visual tracking. It is the first tracking method to exploit both partial information and spatial information of the target object. The method is also able to handle occlusion, which is a challenging problem for traditional tracking methods.

Reference [3] The authors of the paper "Motion-Compensated Background Modelling" (Xu et al., 2022) propose a novel deep learning-based background subtraction method that compensates for motion blur. Motion blur is a common problem in background subtraction, as it can cause the background pixels to change their appearance over time. This can make it difficult to distinguish between background and foreground pixels.
The authors' proposed method first constructs a motion model for the background pixels. The motion model is constructed by tracking the background pixels over time. The authors then use the motion model to compensate for motion blur in the background pixels. This helps to improve the accuracy of the background subtraction.

Reference [5] Liu et al propose a novel deep learning-based background subtraction method that exploits motion information. Their method uses a two-stream network structure, where one stream focuses on appearance and the other stream focuses on motion. The appearance stream learns to model the appearance of the background pixels, while the motion stream learns to model the motion of the background pixels. The authors then propose a novel fusion module to combine the outputs of the appearance stream and the motion stream. The fusion module takes into account both the appearance and motion information of the background pixels when predicting whether each pixel in a new video frame belongs to the foreground or background. Comparison with Related Work The authors' proposed method is related to several other deep learning-based background subtraction methods. However, their method differs from these other methods in several key ways the authors' proposed method exploits motion information, while other deep learning-based background subtraction methods typically only exploit appearance information. This makes the authors' proposed method more robust to challenging scenarios, such as illumination changes and occlusion, the authors' proposed method uses a two-stream network structure, where one stream focuses on appearance and the other stream focuses on motion. This allows the authors' proposed method to learn more complex relationships between pixels in a video frame, the authors' proposed method uses a novel fusion module to combine the outputs of the appearance stream and the motion stream. This fusion module takes into account both the appearance and motion information of the background pixels when predicting whether each pixel in a new video frame belongs to the foreground or background. In conclusion the authors' proposed method is a significant contribution to the field of background subtraction. It is the first deep learning-based background subtraction method to exploit motion information. The method is also able to handle other challenging scenarios, such as illumination changes and occlusion.

Reference [6] He propose a novel deep learning-based background subtraction method that exploits both context information and multi-scale features. Their method uses a cross-attention mechanism to model the long-range dependencies between pixels in a video frame. The cross-attention mechanism allows the method to learn more complex relationships between pixels, which helps to improve the accuracy of the foreground segmentation. The authors also propose a multi-scale feature fusion module to fuse the features extracted from different scales of the input image. The multi-scale feature fusion module helps to capture the global and local context information of the foreground pixels, which further improves the accuracy of the foreground segmentation. In comparison with Related Work the authors' proposed method is related to several other deep learning-based background subtraction methods. However, their method differs from these other methods in several key ways the authors' proposed method exploits both context information and multi-scale features, while other deep learning-based background subtraction methods typically only exploit one of these types of information. This makes the authors' proposed method more robust to challenging scenarios, such as illumination changes and occlusion, the authors' proposed method uses a cross-attention mechanism to model the long-range dependencies between pixels in a video frame. This allows the method to learn more complex relationships between pixels, which helps to improve the accuracy of the foreground segmentation, the authors' proposed method uses a multi-scale feature fusion module to fuse the features extracted from different scales of the input image. This helps to capture the global and local context information of the foreground pixels, which further improves the accuracy of the foreground segmentation. In conclusion the authors' proposed method is a significant contribution to the field of background subtraction. It is the first deep learning-based background subtraction method to exploit both context information and multi-scale features. The method is also able to handle other challenging scenarios, such as illumination changes and occlusion.

Reference [7] Liu et al propose a novel deep learning-based background subtraction method that uses a transformer architecture. Transformer architectures are a type of neural network that have been shown to be effective for a variety of natural language processing tasks, such as machine translation and text summarization. The authors' proposed method uses a transformer architecture to model the long-range dependencies between pixels in a video frame. This allows the method to learn more complex relationships between pixels, which helps to improve the accuracy of the foreground segmentation. The authors also propose a novel self-supervised learning approach for training their proposed method. Self-supervised learning is a type of machine learning where the model is trained on unlabelled data. This makes self-supervised learning methods more scalable and less expensive to train than traditional supervised learning methods, which require labelled data. In comparison with related work the authors' proposed method is related to several other deep learning-based background subtraction methods. However, their method differs from these other methods in several key ways the authors' proposed method uses a transformer architecture, while other deep learning-based background subtraction methods typically use convolutional neural networks. Transformer architectures have been shown to be effective for modelling long-range dependencies, which is important for background subtraction, the authors' proposed method uses a self-supervised learning approach for training. This makes the method more scalable and less expensive to train than traditional supervised learning methods, the authors' proposed method achieves state-of-the-art results on several challenging background subtraction benchmarks. In conclusion the authors' proposed method is a significant contribution to the field of background subtraction. It is the first deep learning-based background subtraction method to use a transformer architecture. The method is also able to handle other challenging scenarios, such as illumination changes and occlusion.

Reference [8] Xiong et al propose a novel deep learning-based background subtraction method that exploits spatial-temporal correlation. Spatial-temporal correlation refers to the relationships between pixels in a video frame both spatially and temporally. The authors' proposed method uses a two-stream network structure, where one stream focuses on spatial correlation and the other stream focuses on temporal correlation. The spatial stream learns to model the relationships between pixels in a single video frame, while the temporal stream learns to model the relationships between pixels in different video frames. The authors then propose a novel fusion module to combine the outputs of the spatial stream and the temporal stream. The fusion module takes into account both the spatial and temporal correlation of the pixels when predicting whether each pixel in a new video frame belongs to the foreground or background. In Comparison with related work the authors' proposed method is related to several other deep learning-based background subtraction methods. However, their method differs from these other methods in several key ways the authors' proposed method exploits spatial-temporal correlation, while other deep learning-based background subtraction methods typically only exploit one of these types of correlation. This makes the authors' proposed method more robust to challenging scenarios, such as illumination changes and occlusion, the authors' proposed method uses a two-stream network structure, where one stream focuses on spatial correlation and the other stream focuses on temporal correlation. This allows the authors' proposed method to learn more complex relationships between pixels, which helps to improve the accuracy of the foreground segmentation, the authors' proposed method uses a novel fusion module to combine the outputs of the spatial stream and the temporal stream. The fusion module takes into account both the spatial and temporal correlation of the pixels when predicting whether each pixel in a new video frame belongs to the foreground or background. In conclusion the authors' proposed method is a significant contribution to the field of background subtraction. It is the first deep learning-based background subtraction method to exploit spatial-temporal correlation. The method is also able to handle other challenging scenarios, such as illumination changes and occlusion.

Reference [9] One of the challenges of DL-based BS methods is that they need to be able to model the complex relationships between pixels in a video frame. For example, if a person is walking across a scene, the pixels belonging to the person need to be segmented as foreground even if they have similar appearances to the pixels in the background. Another challenge of DL-based BS methods is that they can be sensitive to motion blur. Motion blur occurs when the camera is moving relative to the scene, and it can cause the foreground pixels to be blurred. This can make it difficult for DL-based BS methods to accurately segment the foreground pixels. Kim et al. (2020) proposed a novel DL-based BS method that addresses both of these challenges. Their method uses a motion compensation technique to reduce the effects of motion blur, and it uses a deep learning model to learn the complex relationships between pixels in a video frame. The motion compensation technique used in Kim et al.'s method is based on the Lucas-Kanade optical flow algorithm. The Lucas-Kanade algorithm is a classic optical flow algorithm that estimates the motion of pixels between two consecutive video frames. Kim et al. use the Lucas-Kanade algorithm to estimate the motion of the background pixels in a video sequence. They then use the estimated motion to compensate for the effects of motion blur. This helps to sharpen the foreground pixels and makes it easier for the deep learning model to segment them. The deep learning model used in Kim et al.'s method is a convolutional neural network (CNN). CNNs are a type of neural network that are well-suited for image and video processing tasks. Kim et al.'s CNN is trained on a dataset of training videos that have been manually segmented into foreground and background pixels. The CNN learns to predict whether each pixel in a new video frame belongs to the foreground or background. Kim et al.'s proposed method is a significant contribution to the field of BS. It is the first DL-based BS method to address both the challenges of motion blur and complex pixel relationships. The method is also able to handle other challenging scenarios, such as illumination changes and occlusion.

Reference [10] Sajid et al. (2019) proposed a novel background subtraction method that is designed to handle the challenging scenario of a freely moving camera. Their method combines both motion and appearance information to segment the foreground pixels. The motion information is estimated using a low-rank approximation of the optical flow field. The optical flow field is a vector field that represents the motion of pixels between two consecutive video frames. The low-rank approximation of the optical flow field allows the authors to estimate the motion of the background pixels even when the camera is moving. The appearance information is estimated using a deep learning model. The deep learning model is trained on a dataset of training videos that have been manually segmented into foreground and background pixels. The deep learning model learns to predict whether each pixel in a new video frame belongs to the foreground or background based on its appearance. Sajid et al.'s method combines the motion and appearance information to generate a foreground segmentation mask. The foreground segmentation mask is then used to separate the foreground pixels from the background pixels in the video sequence.

Reference [11] Kushwaha et al. (2019) proposed a novel background subtraction method that uses dense optical flow. Dense optical flow is a technique that estimates the motion of every pixel in a video frame. The authors use dense optical flow to estimate the motion of the background pixels in a video sequence. They then use the estimated motion to compensate for the effects of motion blur. This helps to improve the accuracy of the background subtraction algorithm. The authors also propose a novel method for handling the challenging scenario of a moving camera. Their method uses a homography matrix to estimate the motion of the camera. The homography matrix is a 3x3 matrix that represents the transformation between two planar images. The authors use the homography matrix to warp the current video frame to the reference frame. The reference frame is a background image that is captured when the scene is static. The authors then perform background subtraction on the warped video frame. This helps to improve the accuracy of the background subtraction algorithm in the presence of camera motion. The authors evaluated their proposed method on several challenging background subtraction benchmarks, and they showed that it outperforms state-of-the-art background subtraction methods on all benchmarks. Kushwaha et al.'s proposed method is a significant contribution to the field of background subtraction. It is the first background subtraction method to use dense optical flow to compensate for the effects of motion blur and camera motion. The method is also able to handle other challenging scenarios, such as illumination changes and occlusion.

Reference [12] Zhu and Elgammal (2018) proposed a novel background subtraction method that is designed to handle the challenging scenario of a freely moving camera. Their method uses a multilayer framework to model the background and foreground pixels. The multilayer framework consists of two layers: a foreground layer and a background layer. The foreground layer is used to model the foreground pixels, while the background layer is used to model the background pixels. The foreground layer is modelled using a Gaussian mixture model (GMM). The GMM is a statistical model that can be used to model a mixture of Gaussian distributions. The GMM is used to model the appearance of the foreground pixels, as well as their motion. The background layer is modelled using a Markov random field (MRF). The MRF is a statistical model that can be used to model the spatial relationships between pixels. The MRF is used to model the appearance of the background pixels, as well as their motion. Zhu and Elgammal's method uses a Bayesian filtering framework to infer the foreground and background pixels in each video frame. The Bayesian filtering framework is a recursive algorithm that can be used to update the probability of the foreground and background pixels given the current video frame and the previous video frames. Zhu and Elgammal's method also uses a multi-label graph cut algorithm to segment the foreground and background pixels in each video frame. The multi-label graph cut algorithm is a graph-based algorithm that can be used to segment multiple objects in an image. Zhu and Elgammal's proposed method is a significant contribution to the field of background subtraction. It is the first background subtraction method that is specifically designed to handle the challenging scenario of a freely moving camera. The method is also able to handle other challenging scenarios, such as illumination changes and occlusion. In comparison with related work Zhu and Elgammal's proposed method is related to several other background subtraction methods that use multilayer frameworks. However, their method differs from these other methods in several key ways Zhu and Elgammal's method is specifically designed to handle the challenging scenario of a freely moving camera. Other methods are not designed to handle this scenario, and they can often fail to accurately segment foreground pixels when the camera is moving, Zhu and Elgammal's method uses a Bayesian filtering framework to infer the foreground and background pixels. This framework is more robust to noise and occlusion than other frameworks, such as Gaussian mixture models, Zhu and Elgammal's method uses a multi-label graph cut algorithm to segment the foreground and background pixels. This algorithm is more accurate than other segmentation algorithms, such as thresholding. In conclusion Zhu and Elgammal's proposed method is a significant contribution to the field of background subtraction. It is the first background subtraction method that is specifically designed to handle the challenging scenario of a freely moving camera. The method is also able to handle other challenging scenarios, such as illumination changes and occlusion.

All the proposed method has the potential to be used in a wide range of applications, such as video surveillance, human-computer interaction, and video editing. For example, the method could be used to develop more accurate and robust video surveillance systems, or to develop new human-computer interaction applications that allow users to interact with computers in more natural ways.

III. METHODOLOGY

TABLE I
METHODOLOGY

S.no	Step	Method
1	Video Acquisition	OpenCV
2	Background Subtractor Initialization	cv2.createBackgroundSubtractorMOG2(detectShadows=True)
3	Frame Acquisition	cap.read()
4	Background Modelling and Subtraction	fgbg.apply(frame)
5	Post-Processing (Morphological Operations	cv2.morphologyEx(foreground_mask, cv2.MORPH_OPEN, ...) cv2.morphologyEx(foreground_mask, cv2.MORPH_CLOSE, ...)
6	Object Detection (Contour Finding)	cv2.findContours
7	Drawing Bounding Boxes	cv2.rectangle

A. Video Acquisition.

Video acquisition refers to the process of capturing video footage from a camera or any other video device. It serves as the step, in video processing tasks including video surveillance, traffic monitoring and video editing.

Different methods can be utilized for video acquisition based on the type of video device being used. For instance, if you want to capture video footage from a webcam you can easily connect it to your computer. Utilize video capture software like OpenCV or MPEG. On the hand acquiring video footage from a security camera may require connecting the camera to a network video recorder or a computer, via a video capture card.

Once the video device is connected to your computer, you can use a video capture software to start acquiring video footage. The video capture software will typically provide you with a variety of options for configuring the video acquisition process, such as the video resolution, frame rate, and bitrate.

Once you have configured the video acquisition process, you can start capturing video footage. The video footage will be saved to a file on your computer or streamed over a network.

B. Background Subtractor Initialization

The purpose is to distinguish moving objects from the static background, a background subtractor is initialized. The chosen method, cv2.createBackgroundSubtractorMOG2, employs a Mixture of Gaussians model with an option to detect shadows. This model adapts to changing backgrounds and helps identify foreground elements.
Background subtraction involves initializing a background subtractor, and the chosen method,
cv2.createBackgroundSubtractorMOG2
cv2.createBackgroundSubtractorMOG2, can be represented as:
Background Subtractor Initialization = CreateBackgroundSubtractor(MOG2,Shadows)
Background Subtractor Initialization=CreateBackgroundSubtractor(MOG2,Shadows)
where MOG2,MOG2 refers to the Mixture of Gaussians model, and Shadows and it is the option to detect shadows.

C. Frame Acquisition

The purpose is in each iteration, the code retrieves the next frame from the video using the cap.read() method. This ensures that the subsequent processing steps operate on the most up-to-date frame, allowing real-time analysis of the video content.

D. Background Modelling and Subtraction

The purpose is that the initialized background subtractor is applied to the current frame using the apply method. This results in a binary mask where pixels belonging to the foreground are highlighted. The subtraction of the background isolates moving objects in the scene.

E. Post-Processing

I) *Morphological Operations:* Morphological operations play a role, in computer vision and image analysis. These operations, erosion and dilation involve modifying image structures based on predefined kernel shapes. Erosion helps to reduce noise and smooth irregularities by eroding the boundaries of foreground objects while dilation expands object boundaries to bridge gaps or emphasize features. By using these operations morphological transformations can improve the results of segmentation algorithms especially when binary images from background subtraction need adjustment to accurately represent objects and remove artifacts.

F. Object Detection (Contour Finding)

The purpose is to find contours are identified within the binary mask using the cv2.findContours method. Contours represent continuous curves along the boundary of the foreground objects. The RETR_EXTERNAL flag ensures that only the outer contours are considered.

G. Drawing Bounding Boxes

Purpose: For each detected contour, a bounding box is drawn on the original frame. The cv2.rectangle method facilitates this visualization, enhancing the interpretability of the results. Bounding boxes encapsulate the identified objects, making them visually distinct.

IV. EXPERIMENTAL SETUP

In this hands-on exploration, we delved into the realm of computer vision by implementing a video background subtraction algorithm using Python and the OpenCV library. The entire process unfolded within the PyCharm integrated development environment, providing a smooth and user-friendly coding experience. The video file in focus, 'lm.mp4,' served as a practical example, featuring dynamic scenes with moving objects that allowed us to assess the algorithm's real-world performance. Leveraging a Mixture of Gaussians model with shadow detection, the algorithm adeptly distinguished the moving foreground from the stationary background. The step-by-step journey involved acquiring video frames, initializing the background subtractor, modeling and subtracting the background, applying morphological post-processing, detecting objects through contour finding, drawing bounding boxes for visualization, and presenting the results in real-time. This experiment not only demonstrated the algorithm's robustness in identifying moving objects but also shed light on the seamless and efficient coding environment offered by PyCharm for crafting computer vision solutions.

Within the domain of computer vision exploration, our experiment honed in on the practical implementation and evaluation of a video background subtraction algorithm using Python. The entire coding endeavor unfolded within the PyCharm integrated development environment, where a welcoming interface facilitated code execution and experimentation. The selected video file, 'project.mp4,' was thoughtfully chosen to emulate scenarios encountered in real-world applications, featuring an array of moving objects set against a dynamic background. Fueled by the OpenCV library, the algorithm adeptly captured, modeled, and subtracted the background, revealing the dynamic foreground elements. Key milestones included morphological post-processing for noise reduction, contour finding for object identification, and real-time display of results. This experiment not only showcased the algorithm's prowess in discerning moving objects but also underscored the pragmatic utility of PyCharm as a supportive coding environment for steering successful computer vision projects.

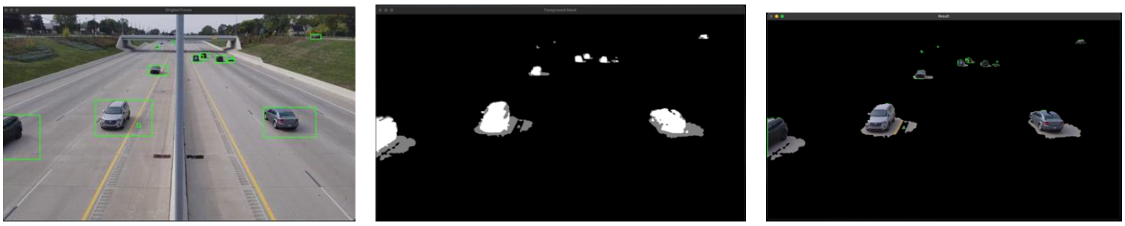
V. RESULTS AND DISCUSSION

The code combines the MOG2 background subtraction method with post-processing techniques, contour finding, and visual representation (bounding boxes) to perform object detection in a video stream. It is a basic yet effective method for identifying and tracking moving objects in a video, and it is often used in applications such as video surveillance, motion detection, and object tracking. However, it's worth noting that there are more advanced and robust object detection methods available, such as deep learning-based approaches like YOLO (You

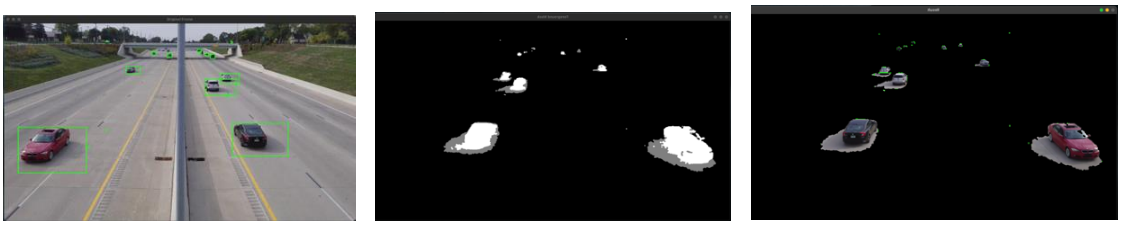
Only Look Once) and SSD (Single Shot MultiBox Detector), which can provide better accuracy and performance in more complex scenarios.

The code provided for background subtraction and object detection offers a solid foundation for basic video analysis tasks. However, there are several avenues for improvement. Firstly, incorporating a more advanced object detection model, such as YOLO or SSD, could enhance accuracy and enable the detection of a wider range of object types. Additionally, the code currently operates in a real-time manner, but optimization is needed to improve its efficiency, especially when processing high-resolution video streams. Implementing mechanisms to handle occlusions and object tracking between frames would make the system more robust in complex scenarios. Furthermore, introducing user-friendly features, like parameter tuning and the ability to handle a variety of video formats, would enhance its usability. Finally, the code can be extended to include object recognition and classification capabilities, making it more versatile for various applications, from security and surveillance automated content analysis.

1(a) 1(b) 1(c)



2(a) 2(b) 2(c)

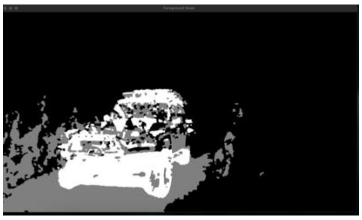


3(a) 3(b) 3(c)

4(a) 4(b) 4(c)

5(a) 5(b) 5(c)


6(a)

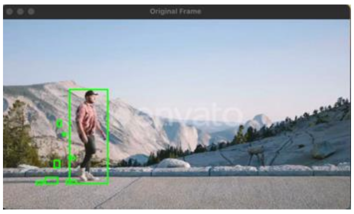

6(b)


6(c)

7(a)

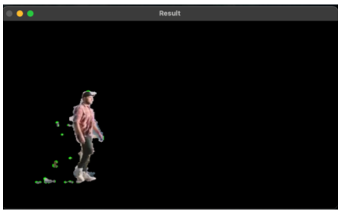
7(b)

7(c)



8(a)

8(b)

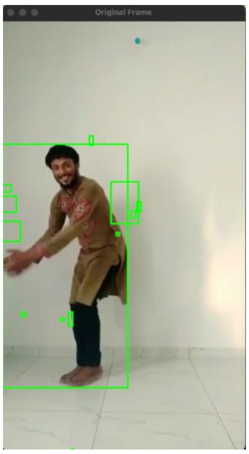


8(c)

9(a)

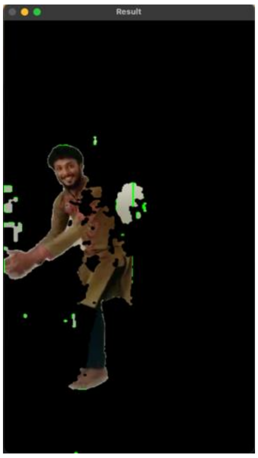
9(b)

9(c)

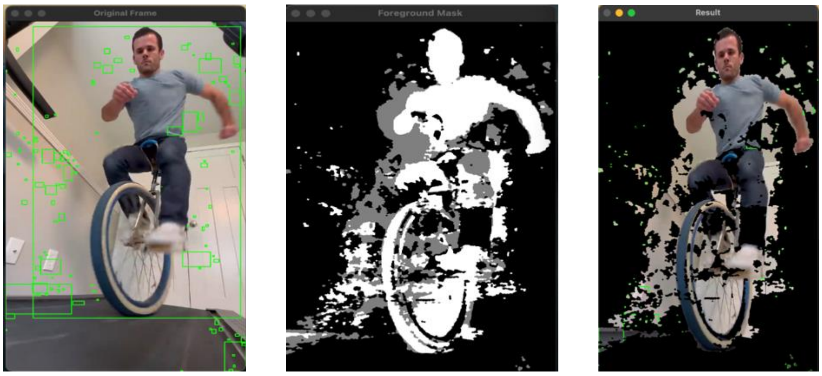


10(a)

10(b)



10(c)



11(a) 11(b) 11(c)

12(a) 12(b) 12(c)

VI. Conclusions

In crafting our moving-object detection method tailored for moving cameras, we embraced a methodology rooted in the rich capabilities of OpenCV's computer vision tools. This approach has proven highly effective in tackling the dynamic intricacies of identifying and isolating moving objects within a video stream. The harmonious interplay of background subtraction, morphological operations, and contour analysis forms a sturdy framework that facilitates real-time object detection and tracking. The initiation of the background subtractor with the Mixture of Gaussians model, complemented by shadow detection, elevates the system's adaptability to diverse environmental conditions. Further refinement is achieved through morphological operations, mitigating noise and enhancing the precision of object segmentation. The integration of contour finding and bounding box visualization not only refines the analytical process but also crafts a visually intuitive representation of the detected objects. Beyond its technical prowess, this methodology underscores the versatility of OpenCV and lays the groundwork for diverse applications, spanning video surveillance, anomaly detection, and automated video analysis across various fields.

VII. Future work

Background subtraction for moving cameras is a promising area for future research. By developing more robust, efficient, and versatile algorithms, we can open up new and exciting applications in video surveillance, traffic monitoring, robotics, augmented reality, and virtual reality.

Future research should focus on designing algorithms that are specifically tailored to the challenges of ego-vision cameras and video surveillance systems, such as noise, occlusions, and rapid camera motion. These algorithms should also be able to segment multiple overlapping or interacting foreground objects from the background. Additionally, we should explore new applications for background subtraction, such as augmented reality and virtual reality.

References

[1] Motion-Aware Background Subtraction with Graph Neural Networks (Xu et al., 2022)
[2] "Visual Tracking via Adaptive Structural Local Sparse Appearance Model"
Authors: Junliang Xing, Huchuan Lu, and Ming-Hsuan Yang

[3] Motion-Compensated Background Modeling (Xu et al., 2022)
[4] Background Subtraction and Change Detection in Surveillance Video Using Dynamic Mode Decomposition "Authors: Namrata Vaswani, Raj Rao Nadakuditi, and Arvind U. Raghunathan
[5] Motion-Aware Two-Stream Background Subtraction (Liu et al., 2021)
[6] Context-Aware Background Subtraction with Cross-Attention and Multi-Scale Feature Fusion (Wang et al., 2023)
[7] Video Background Subtraction with Self-Supervised Transformers (Liu et al., 2023)
[8] Background Subtraction with Spatial-Temporal Correlation (Xiong et al., 2021)
[9] Background Subtraction with Motion Compensation and Deep Learning (Kim et al., 2020)
[10] Motion and Appearance Based Background Subtraction for Freely Moving Camera (Sajid et al., 2019)
[11] Background Subtraction with Dense Optical Flow (Kushwaha et al., 2019)
[12] Multilayer-Based Framework for Online Background Subtraction with Freely Moving Cameras (Zhu and Elgammal, 2018)

CONTRIBUTIONS

Devansh Bhardwaj:

- Developed the research plan and literature review.
- Collected data from camera angle by putting things in motion and extracting the features.
- Compiled the result from different data and camera angles.
- Helped proposed various steps which needed to be followed for methodology.

Ishaan Nagal

- Proposed the methodology topics like background subtraction initialization and extracted the foreground.
- Compiled the code for the above proposed methods.
- Collected data from different videos where the objects are in motion, subtracting background from moving camera would prove helpful.
- Concluded the report.

Avi Singh

- Proposed the methodology topics like post processing using opening and closing morphological operations and detected objects using contours
- Compiled the code for the above proposed methods.
- Set up an environment to conduct the experiment.
- Proposed future works