TrioMobil Internship Project Report - Case 1

1. Introduction

This project focuses on detecting motionless and moving objects in video sequences using classical image processing techniques. Humans can naturally focus on important moving objects in a scene, such as cars or people, while ignoring the background. Teaching a computer to make a similar distinction is challenging, and this study explores how motion-based segmentation can be achieved without deep learning.

The experiments are conducted using the CDNet 2014 dataset, which includes various scenarios such as highways, pedestrians, fountains, boats, and scenes involving shadows. The main objective is to extract regions that contain motion while suppressing static background regions.

All implementations were developed in Python, using OpenCV and NumPy. The project builds on concepts learned in the CE 490 Image Processing course, including digital image fundamentals, spatial filtering, morphological image processing, noise removal, and color image processing.

2. Methodology

As a baseline approach, the MOG2 background subtraction algorithm was used to detect motion at the pixel level. MOG2 models the background over time and labels pixels that deviate from this model as foreground. This method was first applied to baseline sequences to verify correct operation in stable scenes.

After background subtraction, the raw foreground masks often contained noise, holes inside objects, and leakage around object boundaries. To address these issues, morphological operations such as erosion and dilation were applied to remove small artifacts and fill gaps. A median blur was then used to reduce blocky edges and improve the visual quality of the masks and overlays.

Two challenging dataset categories were selected for further analysis:

Shadow category: Scenes where both objects and their shadows move.
Dynamic background category: Scenes where the background itself exhibits motion.

For shadow handling, HSV and YUV color space analysis was explored. These color spaces separate brightness from color information, allowing illumination changes caused by shadows to be analyzed more effectively. Shadow suppression was applied after initial motion detection.

For dynamic background scenes, experiments were performed on sequences such as *fountain01*, where splashing water introduces continuous background motion.

3. Results and Observations

In baseline scenes, the system successfully detected moving objects and produced stable foreground masks.

In shadow sequences such as *copyMachine*, multiple moving people were detected correctly. However, when a person stopped moving, their foreground mask gradually disappeared as the algorithm absorbed the stationary region into the background model. This demonstrates that background subtraction detects motion rather than object presence.

In dynamic background experiments such as *fountain01* and *fountain02*, moving water was partially detected as foreground. Due to irregular and repetitive background motion, the algorithm could not consistently model all water regions, resulting in flickering masks and noisy artifacts. Some background motion remained undetected, while other regions were falsely classified as foreground.

Results were evaluated qualitatively using overlay images, which are included in the report to illustrate typical successes and failure cases.

4. Discussion and Conclusion

The experiments show that classical background subtraction methods perform well in simple, stable environments but struggle in complex scenarios. Shadows introduce illumination failures, while dynamic backgrounds violate the assumption that motion corresponds to meaningful foreground objects. Additionally, motion-based methods cannot preserve object masks once motion stops.

Although post-processing and color-based techniques improve visual quality, these methods cannot fully resolve the limitations of pixel-based motion detection. Overall, this project highlights the strengths and weaknesses of classical image processing approaches and provides insight into why more advanced methods are required for robust real-world scene understanding.