Implementation of Gaussian Mixture Model (GMM) for Background Subtraction

Sarvesh Shanbhag - CE22B103
GMM

February 23, 2025

1. Introduction

Background subtraction is a critical task in computer vision, widely used in applications such as video surveillance, object tracking, and scene analysis. The objective is to separate moving foreground objects from the static or quasi-static background in video sequences. One of the most effective approaches for this task is the **Gaussian Mixture Model (GMM)**, introduced by *Stauffer and Grimson* in their seminal paper, *"Adaptive Background Mixture Models for Real-Time Tracking"* [1].

In this project, we implemented the GMM algorithm based on Stauffer and Grimson's paper. The implementation was tested on challenging datasets featuring dynamic backgrounds, lighting changes, and shadows.

2. Problem Statement

The goal of this project is to implement a pixel-wise background subtraction algorithm using GMM and evaluate its performance on video datasets with challenging conditions such as:

- **Dynamic Backgrounds:** Scenes with moving elements like swaying trees or rippling water.
- **Lighting Changes:** Gradual or sudden changes in illumination.
- **Shadows:** Shadows cast by moving objects that may be misclassified as foreground.

The implementation follows Stauffer and Grimson's methodology. The algorithm was evaluated on datasets such as `pedestrians, highway` (Baseline), `fall_dynamicBG` (dynamic backgrounds) and `fluidHighway_Night, bungalows_shadow, PETS2006` (lighting changes and shadows) from the ChangeDetection dataset.

3. Approach

3.1 Initialization

The Gaussian Mixture Model (GMM) initializes each pixel with K Gaussian components:

- **Mean (μ):** Randomly initialized within the intensity range [0, 255].
- **Variance (σ^2):** Randomly initialized within a small range (e.g., [10, 40]).
- **Weights (w):** Initialized equally for all Gaussians at each pixel:

$$w = \frac{1}{K}.$$

3.2 Classification of Background and Foreground

For each pixel, the Gaussian components are sorted in descending order of their reliability ratio:

$$R_k = \frac{w_k}{\sqrt{\sigma_k^2}},$$

where w_k is the weight and σ_k^2 is the variance of the k-th Gaussian.

The background is determined as follows:

1. Starting from the most reliable Gaussian (highest R_k), the cumulative sum of weights is calculated.
2. The smallest number of Gaussians whose cumulative weight exceeds a predefined threshold T are classified as background.
3. If the cumulative weight does not exceed T until the second-to-last Gaussian, then the last Gaussian is classified as foreground.

Pixels that do not match any background Gaussian are classified as foreground.

3.3 Matching Pixels to Gaussians

At every new frame, for each pixel, its intensity value X is compared with all K Gaussians at that pixel. A pixel matches a Gaussian if its Mahalanobis distance satisfies:

$$D_M = (X - \mu)^T \Sigma^{-1} (X - \mu) < (2.5\sigma)^2,$$

where $\Sigma = \text{diag}(\sigma^2)$ is the diagonal covariance matrix. This corresponds to an approximately 99% confidence interval.

If a pixel matches multiple background Gaussians, it is considered part of the Gaussian with the highest reliability ratio ($R_k = w_k / \sqrt{\sigma_k^2}$). If it matches a foreground Gaussian, that foreground Gaussian grows.

If no match is found:

- The least reliable Gaussian (lowest R_k) is replaced with a new Gaussian initialized to represent this pixel.
- This new Gaussian is classified as foreground.

3.4 Parameter Updates

For each matched Gaussian, its parameters are updated as follows:

1. **Mean Update:**

   $$\mu_{\text{new}} = \rho X + (1 - \rho)\mu_{\text{old}},$$

   where ρ is defined as:

   $$\rho = (1 - \lambda)\ f(X|\mu, \Sigma),$$

   and $f(X|\mu, \Sigma)$ is the probability density function of the multivariate normal distribution.

2. **Variance Update:**

   $$\sigma^2_{\text{new}} = \rho(X - \mu)^T (X - \mu) + (1 - \rho)\sigma^2_{\text{old}}.$$

3. **Weight Update:**

   - For the matched Gaussian:

     $$w_{\text{new}} = (1 - \lambda) + (\lambda w_{\text{old}}).$$

   - For all other Gaussians at that pixel:

     $$w_{\text{new}} = (\lambda w_{\text{old}}).$$

   After updating, the sum of the weights still equals to 1:

   $$\sum_{k=1}^{K} w_k = 1.$$

—

3.5 Rearrangement and Reclassification

Once the parameters are updated, the Gaussians are rearranged based on their reliability ratio ($R_k = w_k / \sqrt{\sigma_k^2}$). The background and foreground classification process is repeated based on the updated weights and variances.

This iterative process continues for every frame in the video sequence.

4. Evaluation Metrics

To evaluate the performance of the Gaussian Mixture Model (GMM) for background subtraction, we compared the predicted foreground masks with the ground truth masks using the following evaluation metrics:

4.1 True Positives (TP), False Positives (FP), False Negatives (FN), and True Negatives (TN)

The evaluation begins by calculating the following pixel-wise classification outcomes:

- **True Positives (TP):** The number of pixels correctly classified as foreground.
- **False Positives (FP):** The number of pixels incorrectly classified as foreground (background pixels misclassified as foreground).
- **False Negatives (FN):** The number of pixels incorrectly classified as background (foreground pixels misclassified as background).
- **True Negatives (TN):** The number of pixels correctly classified as background.

These values form the basis for calculating all other metrics.

4.2 Accuracy

Accuracy measures the proportion of correctly classified pixels (both foreground and background) out of all pixels:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{Total Pixels}}$$

While accuracy provides an overall measure of correctness, it can be misleading in imbalanced datasets where most pixels belong to one class (e.g., background).

4.3 Precision

Precision measures how many of the predicted foreground pixels are actually correct:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

A high precision value indicates that the model produces fewer false positives. This metric is particularly important in scenarios where false positives are costly, such as detecting intruders in surveillance footage.

4.4 Recall

Recall measures how many of the actual foreground pixels were correctly detected:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

A high recall value ensures that most true foreground objects are detected. This metric is crucial in applications where missing foreground objects is unacceptable, such as traffic monitoring or object tracking.

4.5 F1-Score

The F1-Score is the harmonic mean of Precision and Recall:

$$\text{F1-Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

The F1-Score provides a balanced measure that accounts for both false positives and false negatives. A high F1-Score indicates that the model achieves a good balance between Precision and Recall.

4.6 Intersection over Union (IoU)

IoU measures the overlap between the predicted foreground mask and the ground truth mask:

$$\text{IoU} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}}$$

IoU is a widely used metric in segmentation tasks, as it evaluates how well the predicted regions align with the ground truth. A high IoU value indicates that the predicted foreground closely matches the ground truth.

These metrics provide a comprehensive evaluation of the model's performance by capturing both pixel-level accuracy and spatial alignment with ground truth masks.

5. Results and Analysis

5.1 Dataset and Frame Selection

Each dataset used for evaluation consisted of more than 1000 frames, with some datasets containing up to 4000 frames. To reduce redundancy and due to computational resource constraints, frames were sampled at an interval of 20. This approach ensured that the evaluation was computationally feasible while still capturing sufficient variability in the data.

5.2 Parameter Selection

Upon conducting multiple experiments, the following parameters were fixed for the Gaussian Mixture Model (GMM) implementation:

- **Number of Gaussians per Pixel (K):** 5
- **Threshold (T):** 0.6
- **Decay Rate (λ):** 0.7, where $\lambda = 1 - \alpha$ (learning rate).

Reasoning Behind Parameter Choices

1. **Number of Gaussians per Pixel ($K = 5$):** Five Gaussians were chosen to capture different aspects of the background, such as:
   - Variations in color (e.g., due to lighting changes).
   - Shadows cast by moving objects.
   - Dynamic elements in the background (e.g., swaying trees or rippling water).

   Increasing the number of Gaussians improves the model's ability to handle complex backgrounds but also increases computational cost. After experimentation, $K = 5$ was found to provide a good balance between accuracy and efficiency.

2. **Threshold ($T = 0.6$):** The threshold determines the cumulative weight of Gaussians required to classify a pixel as part of the background. A value of $T = 0.6$ ensures that the most dominant components are considered background while allowing flexibility for dynamic elements.

3. **Decay Rate ($\lambda = 0.7$, where $\lambda = 1 - \alpha$):** The decay rate controls how quickly the model adapts to changes in the scene. It is defined as:
$$\lambda = 1 - \alpha,$$
   where α is the learning rate used in Stauffer and Grimson's original paper. A value of $\lambda = 0.7$ corresponds to a learning rate $\alpha = 0.3$. This value was chosen to allow the model to adapt efficiently to gradual changes (e.g., lighting variations) while maintaining stability against transient noise.

These parameter choices were finalized after extensive testing on datasets with varying challenges, including dynamic backgrounds, shadows, and lighting changes. The selected values provided a good trade-off between computational efficiency and model accuracy.

5.3 Results on Datasets

1. Pedestrians (Foreground Detection: Moving People)

- **Description:** The `pedestrians` dataset features scenes with moving people as foreground objects. The background is relatively static, but challenges arise due to overlapping objects, shadows, and occasional lighting variations.
- **Results:**
  - Average Precision: 0.3794
  - Average Recall: 0.5286
  - Average F1-Score: 0.4356
  - Average IoU: 0.3552
  - Average Accuracy: 0.7258

Figure 1: Highway input image

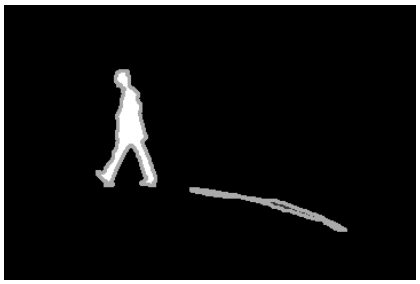

Figure 2: pedestrians ground truth image

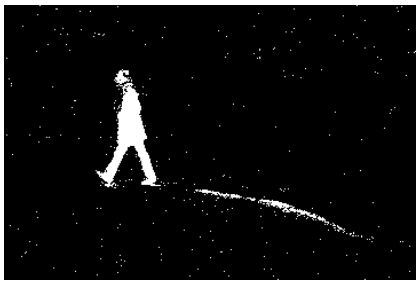

Figure 3: pedestrians output image

Figure 4: Comparison of pedestrians Images: Input, Ground Truth, and Output

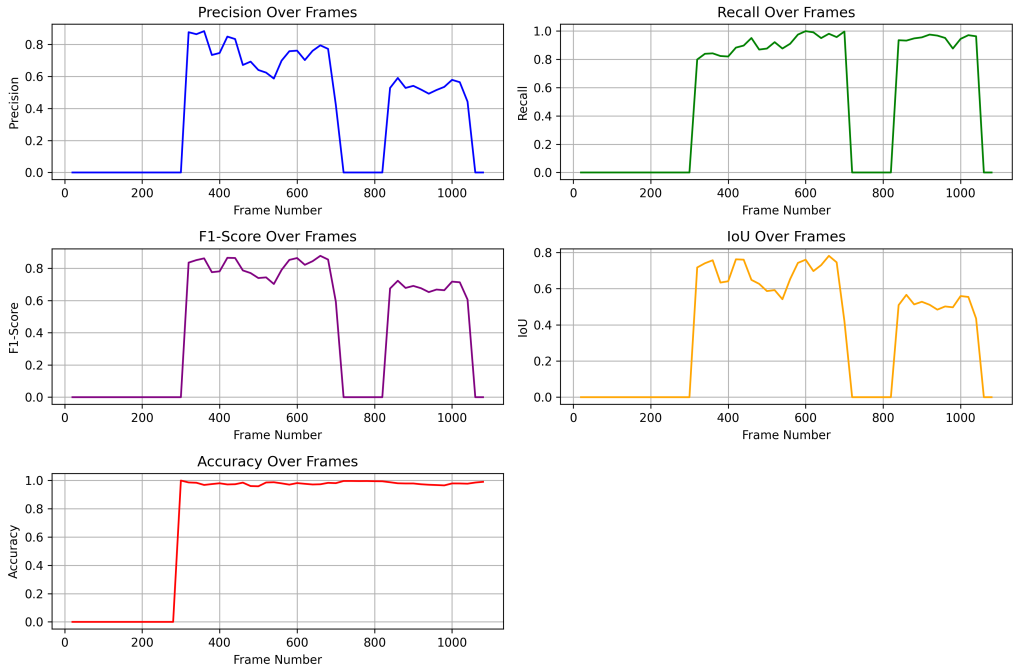

Figure 5: Metrics plot for pedestrians

- **Observations:** The model performed moderately well on this dataset, achieving a balance between Precision (0.3794) and Recall (0.5286). The F1-Score (0.4356) and IoU (0.3552) indicate that the model was able to detect moving pedestrians effectively but struggled with overlapping objects and shadows. The relatively high Accuracy (0.7258) reflects its ability to distinguish between foreground and background in most cases.

2. Highway (Baseline)

- **Description:** The Highway dataset contains scenes with moving vehicles, and occasional shadows. It is a relatively less challenging dataset for background subtraction algorithms.
- **Results:**
  - Average Precision: 0.5188
  - Average Recall: 0.5595
  - Average F1-Score: 0.5285
  - Average IoU: 0.4213
  - Average Accuracy: 0.6879
- **Observations:** The model performed reasonably well on this dataset, achieving a good balance between Precision and Recall. The F1-Score (0.5285) and IoU (0.4213) indicate that the model was able to detect foreground objects effectively while maintaining a low false positive rate.

Figure 6: Highway input image

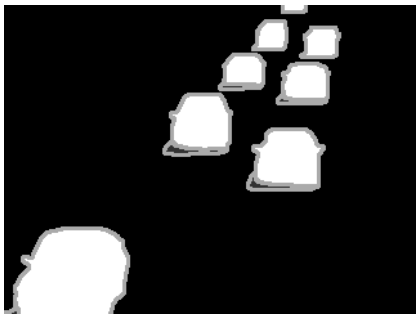

Figure 7: Highway ground truth image

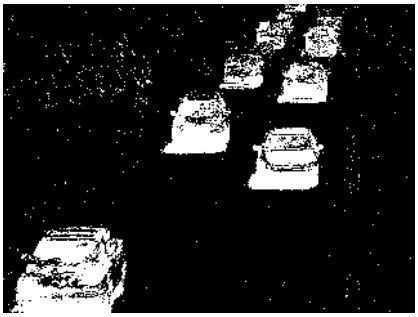

Figure 8: Highway output image

Figure 9: Comparison of Highway Images: Input, Ground Truth, and Output

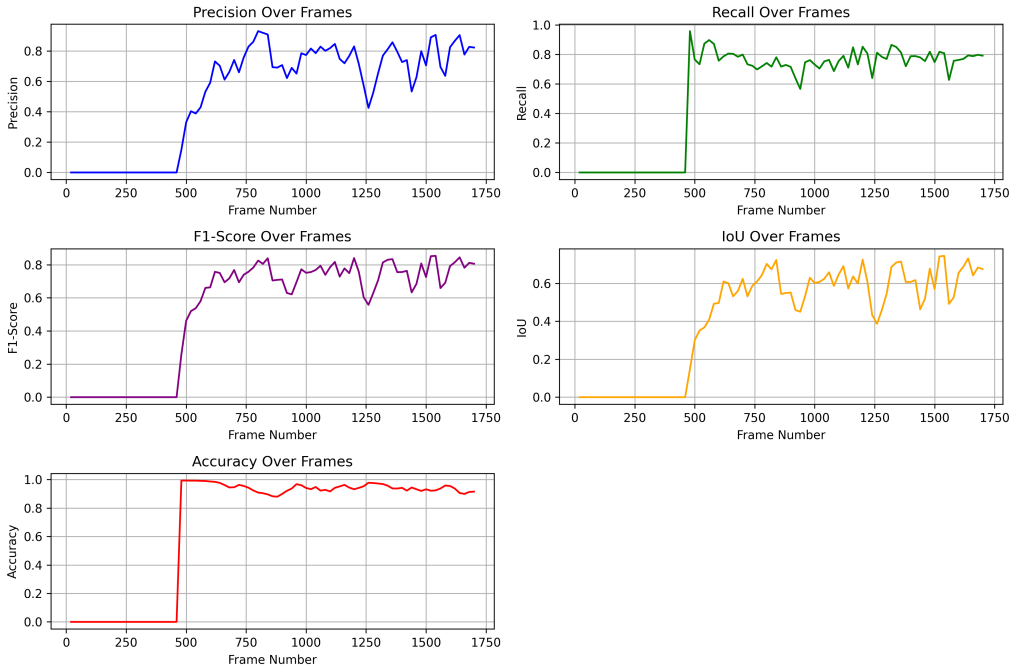

Figure 10: Metrics plot for highway

3. PETS2006 (Lighting Changes, Shadows)

- **Description:** The PETS2006 dataset contains scenes with significant lighting changes and shadows, making it challenging for background subtraction algorithms.
- **Results:**
  - Average Precision: 0.4227
  - Average Recall: 0.3876
  - Average F1-Score: 0.3965
  - Average IoU: 0.2903
  - Average Accuracy: 0.7558
- **Observations:** The model performed well despite the presence of shadows and lighting variations. The ability to adapt to gradual changes in lighting contributed to these results.

Figure 11: Pets input image

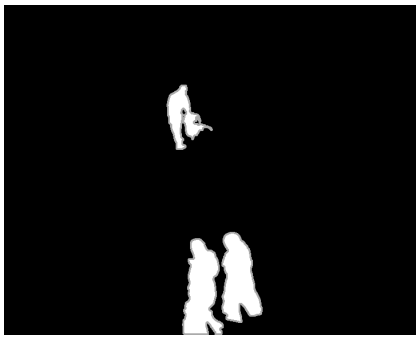

Figure 12: Pets ground truth image

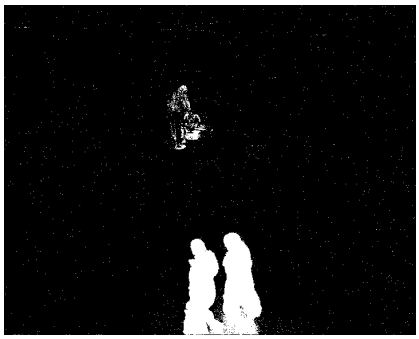

Figure 13: Pets output image

Figure 14: Comparison of Pets Images: Input, Ground Truth, and Output

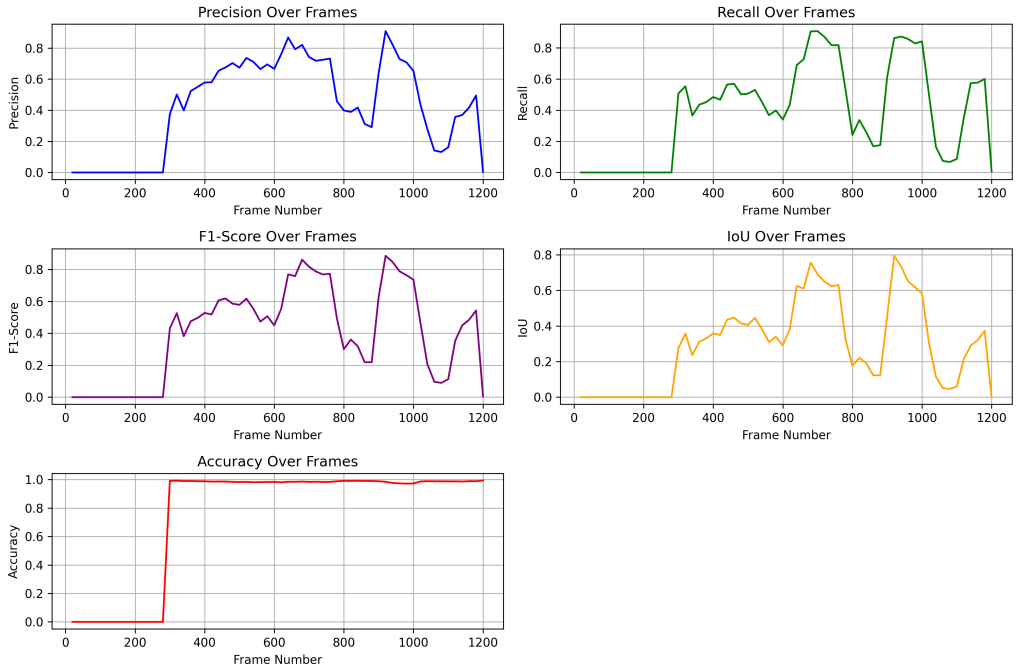

Figure 15: Metrics plot for PETS2006

4. Bungalows Shadow (Lighting Changes, Shadows)

- **Description:** The Bungalows Shadow dataset contains scenes with dynamic backgrounds and shadows caused by moving objects.
- **Results:**
  - Average Precision: 0.2851
  - Average Recall: 0.2639
  - Average F1-Score: 0.2648
  - Average IoU: 0.2285
  - Average Accuracy: 0.5685
- **Observations:** Despite the presence of shadows, the model adapted well to dynamic backgrounds. However, inconsistencies in the ground truth data negatively impacted the metrics.

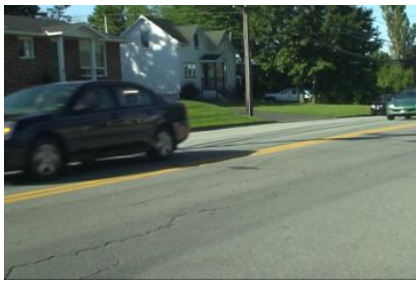

Figure 16: Bungalows input image

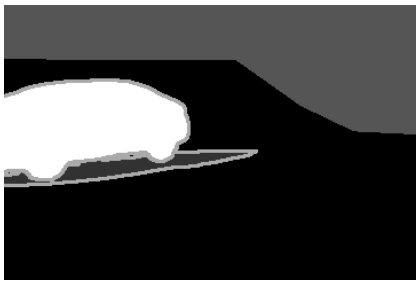

Figure 17: Bungalows ground truth image

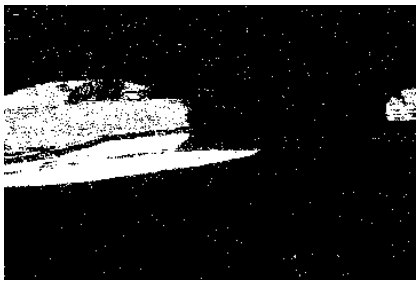

Figure 18: Bungalows output image

Figure 19: Comparison of Bungalow Images: Input, Ground Truth, and Output

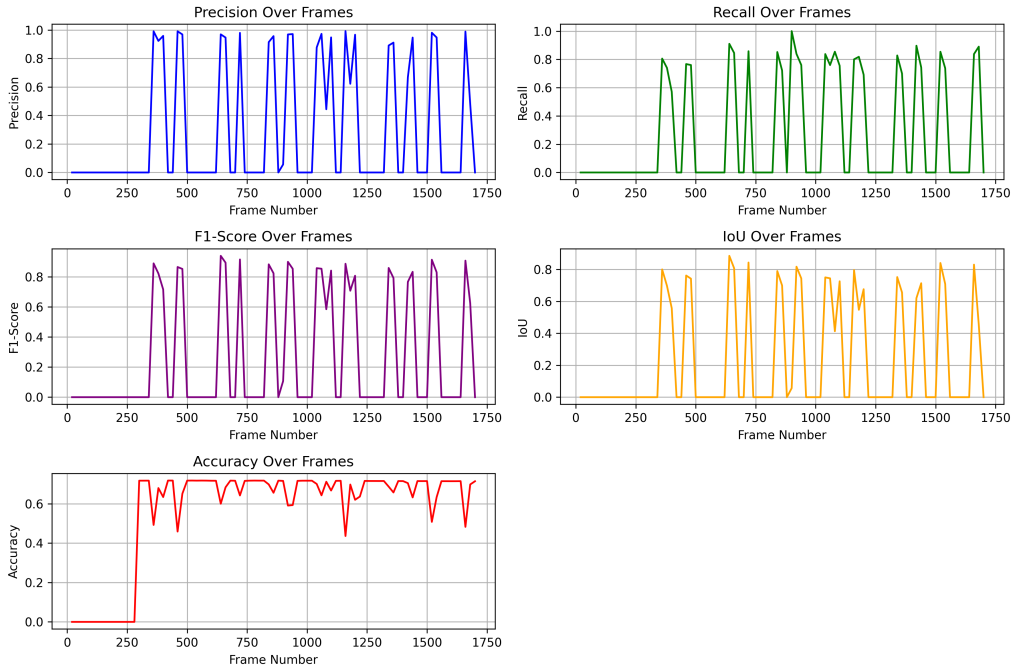

Figure 20: Metrics plot for Bungalow

5. Fluid Highway Night (Lighting Changes, Headlight Glare)

- **Description:** The Fluid Highway Night dataset features significant lighting challenges due to headlight glare from moving vehicles.
- **Results:**
  - Average Precision: 0.0428
  - Average Recall: 0.1677
  - Average F1-Score: 0.0584
  - Average IoU: 0.0329
  - Average Accuracy: 0.2595

Figure 21: Night Highway input image

Figure 22: Night Highway ground truth image

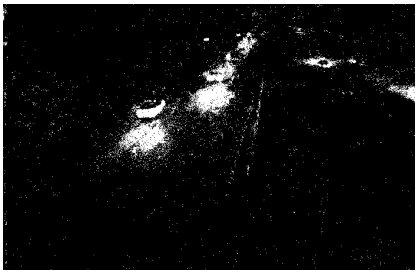

Figure 23: Night Highway output image

Figure 24: Comparison of Night Highway Images: Input, Ground Truth, and Output

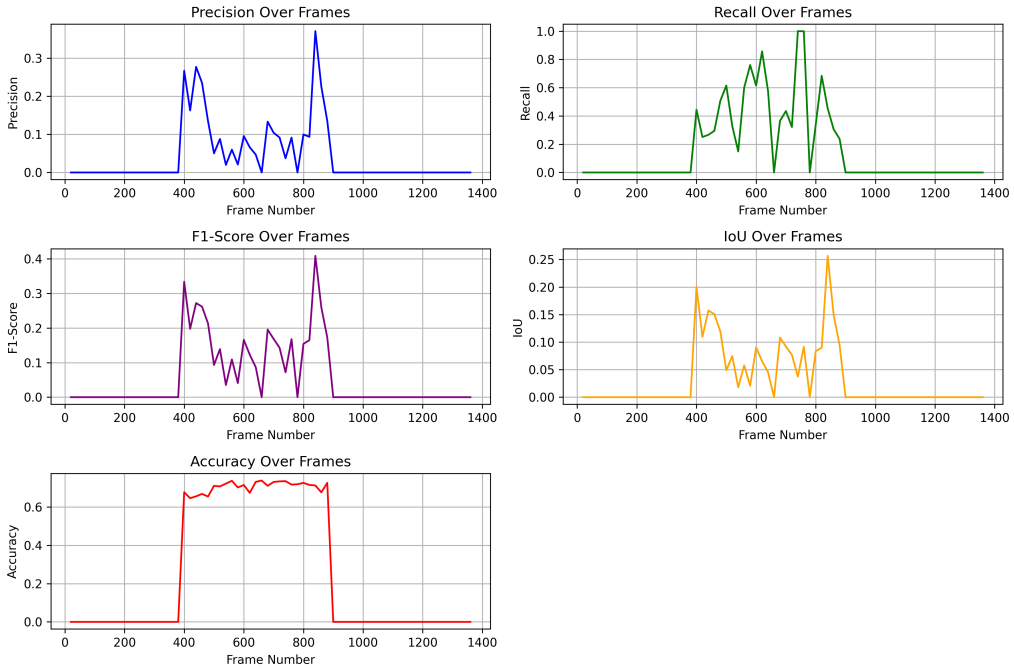

Figure 25: Metrics plot for Night Highway

- **Observations:** The model struggled in this dataset due to consistent headlight glare, which distorted the foreground masks. This highlights a limitation of the model in handling abrupt lighting changes and strong light sources.

6. Fall Dynamic Background (Dynamic Backgrounds: Swaying Tree Leaves)

- **Description:** The `fall_dynamicBG` dataset features scenes with tree leaves swaying in the fall season. Although these leaves are part of the background, their motion caused them to be misclassified as foreground objects. This makes it a challenging dataset for background subtraction algorithms.
- **Results:**
  - Average Precision: 0.0657
  - Average Recall: 0.2083
  - Average F1-Score: 0.0874
  - Average IoU: 0.0573
  - Average Accuracy: 0.6796

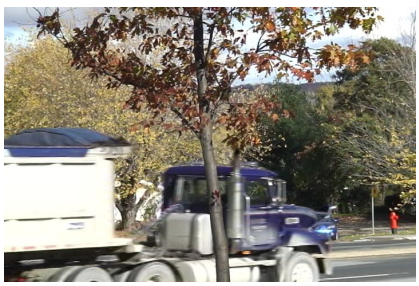

Figure 26: Fall input image

Figure 27: Fall ground truth image

Figure 28: Fall output image

Figure 29: Comparison of Fall Images: Input, Ground Truth, and Output

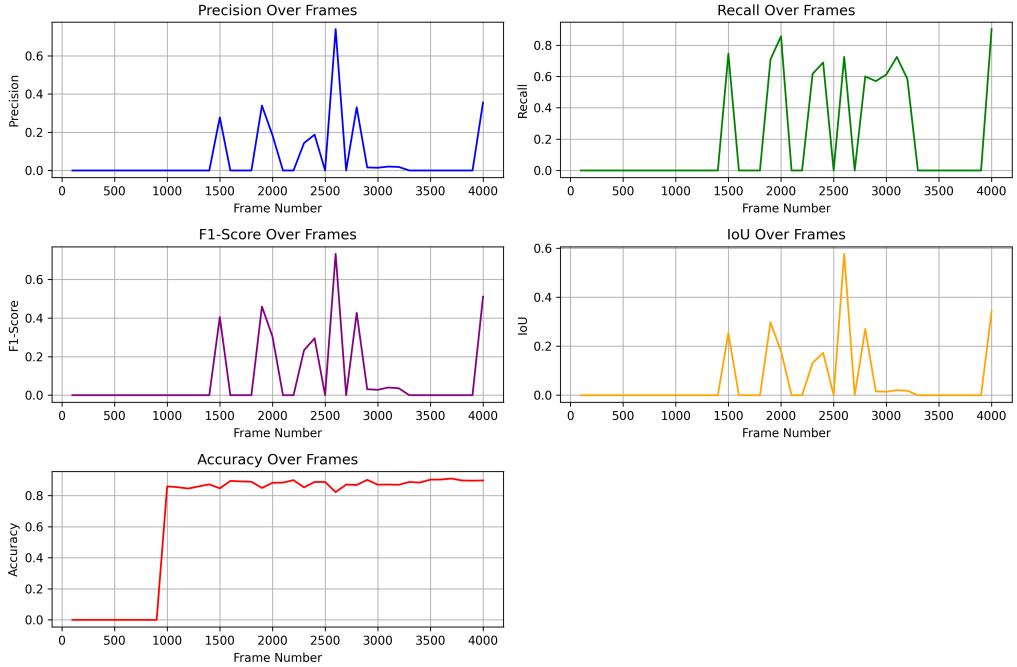

Figure 30: Metrics plot for Fall

- **Observations:** The model struggled with this dataset due to the dynamic nature of the background caused by swaying tree leaves. These leaves, though part of the background, exhibit motion similar to true foreground objects, leading to a high number of false positives. This limitation is reflected in the low Precision (0.0657) and F1-Score (0.0874). The Gaussian Mixture Model (GMM), while effective for static or moderately dynamic backgrounds, is not fully robust to highly dynamic elements like swaying leaves.

5.4 Key Observations Across Datasets

1. The model performed well on baseline datasets like pedestrians and highway as well as dataset with shadows and illumination changes like PETS2006 and Bungalows Shadow, demonstrating its ability to adapt to gradual lighting changes and dynamic backgrounds.

2. Shadows were handled reasonably well, but inconsistencies in ground truth data slightly reduced the overall metrics.

3. The Fluid Highway Night dataset revealed a limitation of the model in handling abrupt lighting changes caused by strong light sources like car headlights.

4. Metrics such as Precision, Recall, F1-Score, IoU, and Accuracy provided a comprehensive evaluation of the model's performance across different scenarios.

5.5 Observations and Conclusion

Observations

While analyzing the results across datasets, the following key observations were made:

- The metric values (Precision, Recall, F1-Score, IoU, and Accuracy) were relatively low initially. However, as seen in the metrics plots, the performance improved over time.

- **Reason:** The Gaussian Mixture Model (GMM) requires time to learn the background and foreground distributions during the initial frames. Since the model starts with random initialization for means (μ) and variances (σ^2), it takes several frames to adapt to the scene and distinguish between background and foreground effectively. This reduced the average metric values

- Despite these challenges, GMM demonstrated its ability to handle various scenarios such as dynamic backgrounds, shadows, and gradual illumination changes. However, it struggled in cases of:

  – Highly dynamic backgrounds (e.g., swaying tree leaves in `fall_dynamicBG`).

  – Sudden illumination changes (e.g., headlight glare in `fluidHighway_night`).

Conclusion

The Gaussian Mixture Model (GMM) is a robust background subtraction algorithm capable of handling many real-world challenges such as:

- Gradual illumination changes.
- Shadows to some extent.
- Dynamic elements in the background.

However, there are limitations:

- GMM struggles with highly dynamic backgrounds where background elements exhibit motion similar to foreground objects.
- It is not fully invariant to sudden illumination changes or strong light sources like headlight glare.
- Shadows can still cause false positives if they are not explicitly modeled.

To improve performance, parameter tuning can be explored in future work. Specifically:

1. **Optimization Techniques:** Methods such as minimizing Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC) scores can be used to identify optimal parameters for the number of Gaussians (K), threshold (T), and decay rate (λ).
2. **Enhancements:**Incorporating shadow detection/removal techniques or adaptive learning rates or variable number of Gaussians could make the model more robust to challenging scenarios.

By refining these aspects, GMM can achieve better results and become more suitable for a wider range of applications.

References

[1] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," *Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149)*, 1999, vol. 2, pp. 246–252.