Sheet 06

Suraj Kuswaha Heldert Villegas

1 Introduction

In this implementation, the pipeline consists of two main stages:

- Background modeling and foreground segmentation using a per-pixel Mixture of Gaussians.
- Post-processing and tracking to count the number of unique people.

2 Mixture of Gaussians Background Model

2.1 Model Description

Each pixel location (i, j) is modeled independently using a Mixture of $K = 3$ Gaussian distributions. For each Gaussian component k, the following parameters are maintained:

- Mean vector $\mu_k \in \mathbb{R}^3$ (RGB color space)
- Scalar variance σ_k^2
- Weight ω_k, where $\sum_k \omega_k = 1$

The background is assumed to consist of stable and frequently occurring colors, while moving objects correspond to foreground regions.

2.2 Initialization

At initialization, all Gaussian means are set to a neutral gray color $[122, 122, 122]$. The variances are initialized to a large value, and the weights are uniformly distributed. This allows the model to adapt gradually to the true background over time.

3 Parameter Update Strategy

3.1 Gaussian Matching

For each pixel value x, the squared Euclidean distance to each Gaussian mean is computed. A Gaussian is considered a match if:

$$\|x - \mu_k\|^2 < 9 \cdot \sigma_k^2$$

If at least one Gaussian matches, the component with the smallest distance is selected for parameter updates.

3.2 Weight Update

When a match occurs, the Gaussian weights are updated using an exponential decay:

$$\omega_k \leftarrow (1-\alpha)\omega_k + \alpha$$

for matched components, followed by normalization to ensure the weights sum to one.

This increases the influence of frequently matching components.

3.3 Dynamic Learning Rate

Instead of using a fixed learning rate for updating the Gaussian parameters, a dynamic learning rate ρ is computed based on the confidence of the match:

$$\rho = \alpha \cdot \exp\left(-\frac{1}{2}\frac{\|x-\mu\|^2}{\sigma^2}\right)$$

This means that pixels close to the Gaussian mean update the model more strongly, while uncertain matches have a smaller influence.

3.4 Mean and Variance Update

For the selected Gaussian, the mean is updated using:

$$\mu \leftarrow (1-\rho)\mu + \rho x$$

The variance is updated using the deviation from the *previous* mean:

$$\sigma^2 \leftarrow (1-\rho)\sigma^2 + \rho\|x-\mu_{\text{old}}\|^2$$

Using the old mean avoids underestimating the variance and improves model stability. A minimum variance value is enforced to prevent Gaussian collapse.

3.5 No Match Case

If no Gaussian matches the pixel, the component with the smallest weight is replaced. Its mean is set to the current pixel value, its variance is initialized to a large value, and its weight is set to a small constant. This allows the model to adapt to new objects entering the scene.

4 Background and Foreground Classification

After updating the Gaussian parameters, the components are sorted by the ratio:

$$\frac{\omega_k}{\sigma_k^2}$$

The top components whose cumulative weight exceeds a predefined threshold are classified as background components. If the pixel matches one of these components, it is labeled as background; otherwise, it is labeled as foreground.

The result is a binary foreground mask where background pixels are set to zero and foreground pixels to 255.

5 People Detection and Tracking

5.1 Extraction

Connected components are extracted from the foreground mask. Bounding boxes are computed and filtered based on size and aspect ratio to remove noise and non-person regions. This step aims to retain only person-like forms.

5.2 Tracking Across Frames

Detected bounding boxes are associated across frames using a simple distance-based matching strategy. Tracks are updated when detections are matched, and new tracks are created for unmatched detections. Tracks that disappear for several frames are removed.

5.3 Counting the Number of People

The tracking information is used to estimate the number of distinct people appearing in the video sequence. Each detected person is represented by a track that stores its bounding box and a unique identifier.

To associate detections across frames, the distance between the centers of bounding boxes is computed. If a detection is sufficiently close to an existing track, it is assigned to that track; otherwise, a new track is created. Tracks that are not updated for several consecutive frames are removed.

To avoid counting noise and false positives, a track is considered a valid person only after it has been detected in multiple frames. The final count is obtained by computing the number of unique confirmed track identifiers.