REAL-TIME TRACKING
REPORT

1.1 PROBLEM STATEMENT:

Implement the real-time object tracking algorithm using background subtraction and a Mixture of Gaussians.

1.2 CONCEPT:

Background modeling: Process of removing the static background from video frames.

Background subtraction of an image or video enables the foreground for subsequent image processing(creating annotations, masks, segmentation). This method may not be applicable for object detection since we are creating masks, but it has a wide range of applications. Typically it's utilized after the backdrop has been modeled.

Background modeling using: **Gaussian Mixture Models**

1.3 DATASET:

Dataset description: A video file is created with its content as a basketball rolling on plain tile.

FPS: 15, Number of frames: 42, Format: **.jpg**

1.4 ALGORITHM:

1. Each pixel is modeled as a mixture of Gaussians (K-Gaussians, K = 4).
2. Recent history of each pixel (X1, X2,...Xt) is modeled by a mixture of K-Gaussian distributions.
3. Probability of observing current pixel value:

$$P(X_t) = \sum_{i=1}^{K} \omega_{i,t} * \eta(X_t, \mu_{i,t}, \Sigma_{i,t})$$

Here,

$$\eta(X_t, \mu, \Sigma) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(X_t-\mu_t)^T\Sigma^{-1}(X_t-\mu_t)}$$

4. Cov matrices are in the form of $\Sigma_{k,t} = \sigma_k^2\mathbf{I}$ for efficient computation.
5. If the pixel process should be considered a stationary process, a standard method for maximizing the likelihood of observed data is expectation maximization(EM), but the pixel process varies over time.
6. Use K-means approximation.
7. Find whether the pixel value matches a gaussian distribution, a match is when the pixel value satisfies standard deviation of 2.5 of a distribution.
8. If none of the K distributions matches the current pixel value, the least probable distribution is replaced with a distribution with the current value as the mean value, initially high variance and low prior weight.
9. The mean and covariance matrix of the unmatched distributions remains the same.

$$\mu_t = (1-\rho)\mu_{t-1} + \rho X_t$$

$$\sigma_t^2 = (1-\rho)\sigma_{t-1}^2 + \rho(X_t - \mu_t)^T(X_t - \mu_t)$$

where the second learning rate[3], ρ, is

$$\rho = \alpha\eta(X_t|\mu_k, \sigma_k)$$

10. Gaussians are ordered in decreasing order by the ratio value (weights/standard deviation).
11. First B Gaussian distributions are chosen for the background model as per the threshold T.

$$B = argmin_b \left(\sum_{k=1}^{b} \omega_k > T \right)$$

12. If the image in a frame matches any of the first B distributions for that pixel, it's modeled as a background. Else, it's a foreground. The background pixels are turned black and then subtracted.

1.5 RESULTS:

Minimum RMSE between our GMM and inbuilt function: 25.06

Maximum RMSE between our GMM and inbuilt function: 142.78

1.6 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, pp. 246-252 Vol. 2, doi: 10.1109/CVPR.1999.784637.