# ADAPTIVE BACKGROUND MIXTURE MODELS FOR REAL-TIME TRACKING

## STAUFFER AND GRIMSON, CVPR 1998

# MOTIVATION

- Video monitoring and surveillance is a challenging task

- Must deal with
  - Cluttered areas, shadows, occlusions, lighting changes, moving elements in scene, slow moving objects, objects (dis)appear

# STANDARD PRACTICE

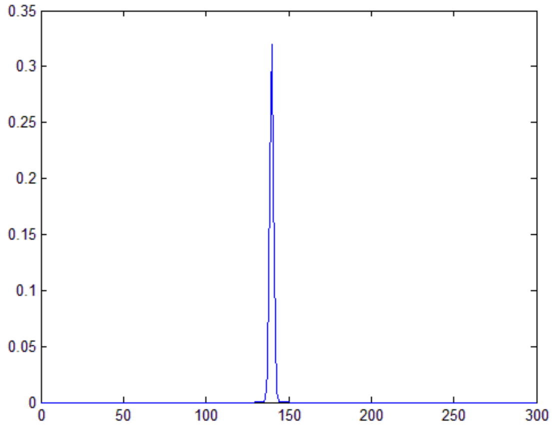
- Use of adaptive background model
  - $B(x, y, t) = (1 - \alpha)B(x, y, t - 1) + \alpha I(x, y, t)$
    - $\alpha$ – is the learning rate
- Strengths: simple and effective in scenes with mostly background and constantly moving objects

- Other techniques try to model the background pixels statistically but cannot deal with bimodal background
  - Kalman filter to track pixel value and has automatic threshold
  - Gaussian distribution for each pixel used to classify as a background or not

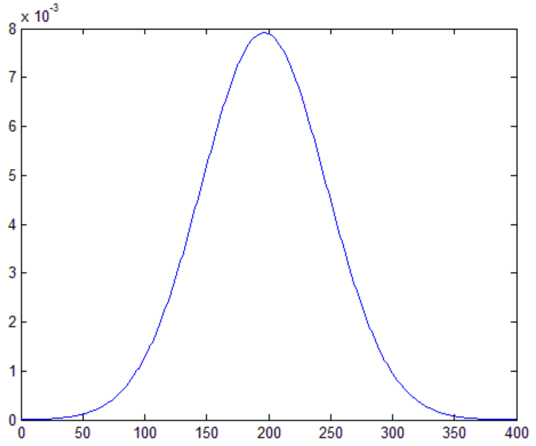
# STANDARD LIMITATIONS

- Weakness: Poor performance for many slow moving objects, recovers slowly, and uses a single threshold for the entire scene

- Example of a rainy day
  - Pixel intensity values over 16 frames (rain occurs halfway through)
    - 139,140,141,141,138,140,140,139,240,241, 243,244,180,141,140,142
  - Model as two different distributions




$\mu_1 = 139.75, \delta_1 = 1.22$



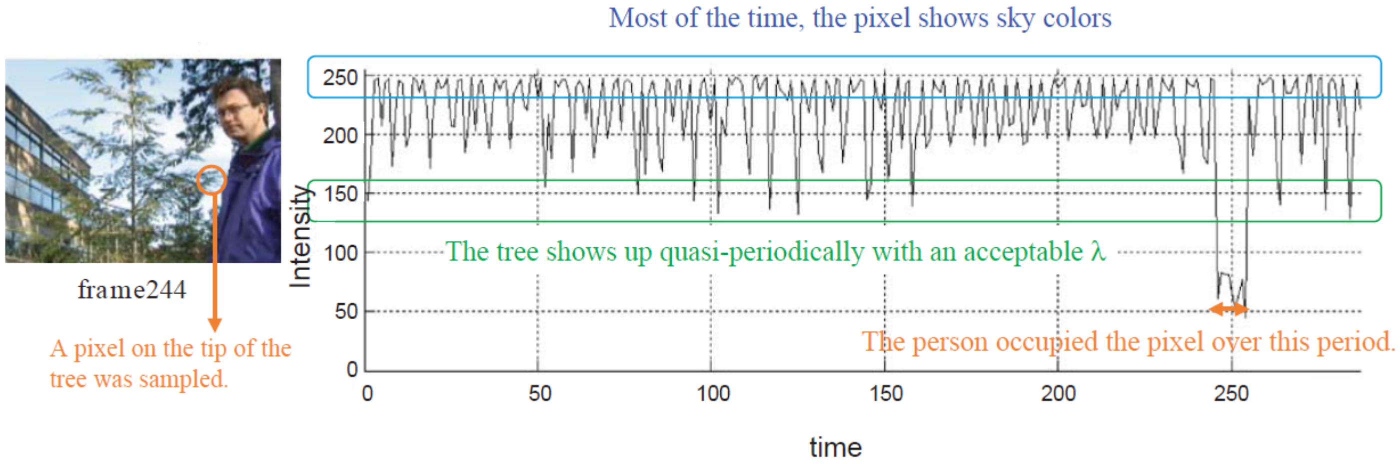

$\mu_2 = 196.37, \delta_2 = 50.43$

# CONTRIBUTIONS

- Develop a computationally efficient background modeling technique
- Pixel intensity distribution modeled using a mixture of Gaussians (MoG) [or Gaussian mixture model (GMM)]
  - Able to model arbitrary distributions (e.g. bimodal)
- Designed an online approximation for computationally efficient update of model

# BACKGROUND DISTRIBUTION

- Single Gaussian distribution is insufficient for real scenes over long periods
  - Mean background assumes a single distribution with the threshold a variance parameter
- Many scenarios with multiple values for a pixel

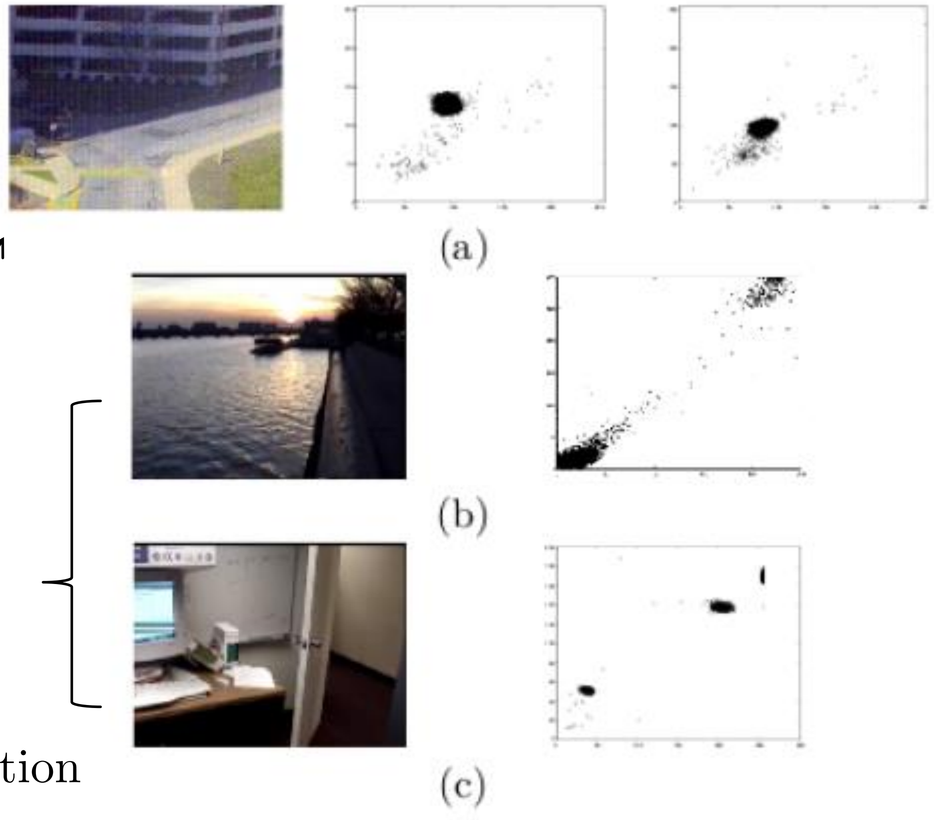



Kyungnam Kim, Thanarat H. Chalidabhongse, David Harwood, Larry Davis, Real-time foreground–background segmentation using codebook model, Real-Time Imaging, Volume 11, Issue 3, June 2005, Pages 172-185

# ROBUST BACKGROUND SUBTRACTION

- Should handle:
  - Lighting changes
    - Adaptive
  - Repetitive motion from clutter
    - Multimodal distribution
  - Long term scene changes
    - Multi-threshold

RG plots of a single pixel




Differing threshold over time

Bimodal distribution over time

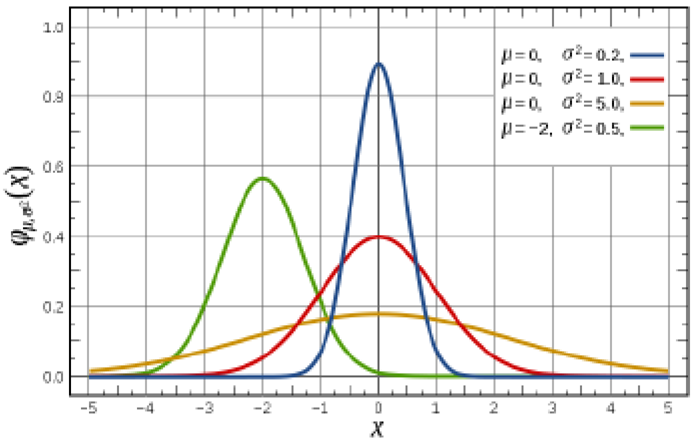
# ALGORITHM OVERVIEW

- Pixel value is modeled as a mixture of adaptive Gaussian distributions
  - Why a mixture?
    - Multiple surfaces appear in a pixel (mean background assumes a single pixel distribution)
  - Why adaptive?
    - Lighting conditions change
- Gaussians are evaluated to determine which ones are most likely to correspond to the background
  - Based on persistence and variance
- Pixels that do not match the background Gaussians are classified as foreground

# GAUSSIAN (NORMAL) DISTRIBUTION

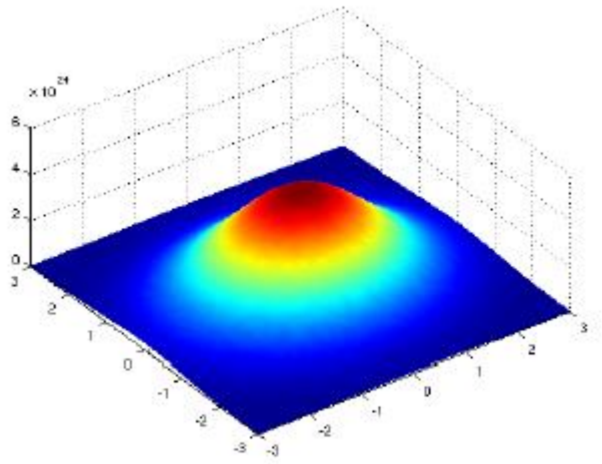
- Univariate

$$\mathcal{N}(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$




- Multivariate

$$\mathcal{N}(\mathbf{x}|\mu,\mathbf{\Sigma}) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\mathbf{\Sigma}|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\mu)^T \mathbf{\Sigma}^{-1}(\mathbf{x}-\mu)}$$



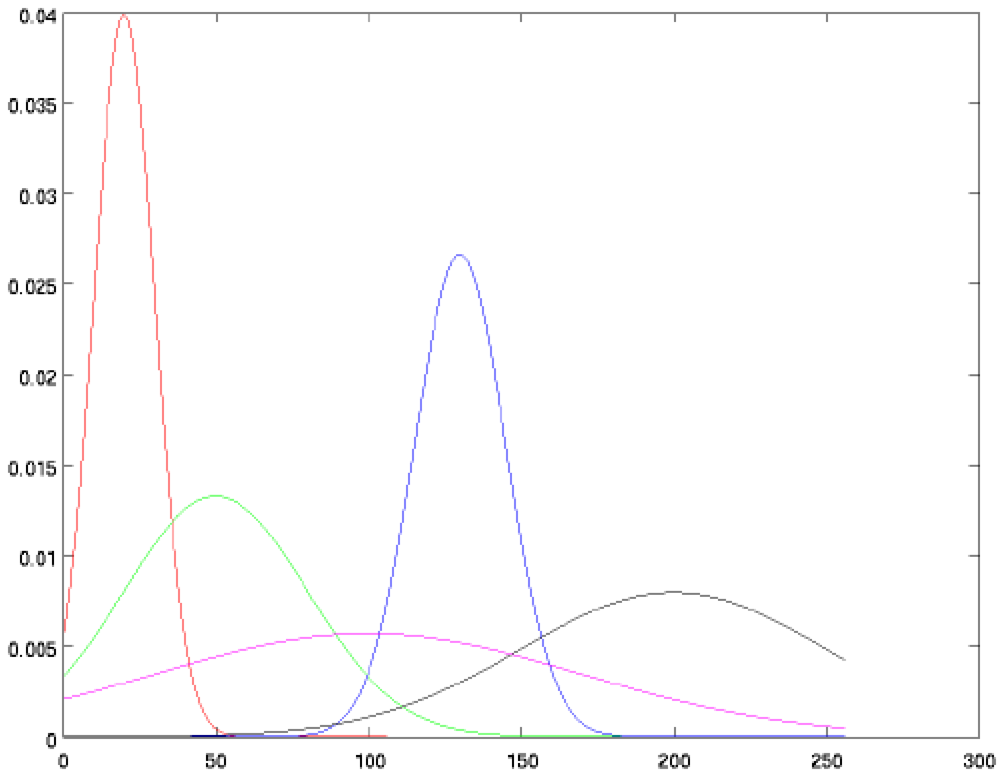
# ONLINE MIXTURE MODEL

- History of a pixel is known up to current time $t$
  - $\{X_1, \ldots, X_t\} = \{I(x_o, y_o, i): 1 \leq i \leq t\}$
- Model the history as a mixture of $K$ Gaussian distributions
  - $P(X_t) = \sum_{i=1}^{K} w_{i,t} \mathcal{N}(X_t | u_{i,t}, \Sigma_{i,t})$
    - $w_{i,t}$ - prior probability (weight) of Gaussians $i$
    - $\Sigma_{i,t} = \sigma_k^2 I$ – simple covariance matrix (independent RGB)

  - What is the dimensionality of the Gaussian?

# MIXTURE MODEL EXAMPLE

- For a grayscale image with $K = 5$
  - Pixel intensity distribution (over time) modeled with five Gaussians



# MODEL ADAPTION I

- Online K-means approximation is used to update the Gaussians
  - Enables fast and efficient model parameter estimation

- Each pixel is compared with its distribution model
  - New pixel $X_{t+1}$ is compared with each of the existing $K$ Gaussians until a match is found
  - Match is defined as a pixel value within $2.5\sigma$ standard deviations of a distribution

# MODEL ADAPTION II

- Match found:

- Update parameters
  - $\mu_{i,t+1} = (1-\rho)\mu_{i,t} + \rho X_{t+1}$
  - $\sigma^2_{i,t+1} = (1-\rho)\sigma^2_{i,t} + \rho\left(X_{t+1} - \mu_{i,t}\right)^2$
    - $\rho = \alpha\mathcal{N}\left(\mathrm{X}_{t+1} \middle| \mu_{i,t}, \sigma^2_{i,t}\right)$
    - $\alpha$ – is a learning rate
- Update Gaussian weights
  - $w_{i,t+1} = (1-\alpha)w_{i,t} + \alpha\left(M_{i,t+1}\right)$
    - $M_{i,t+1} = 1$ for matching Gaussian or $M_{i,t+1} = 0$ for all others
    - Match increases weight

# MODEL ADAPTION III

- No match found:

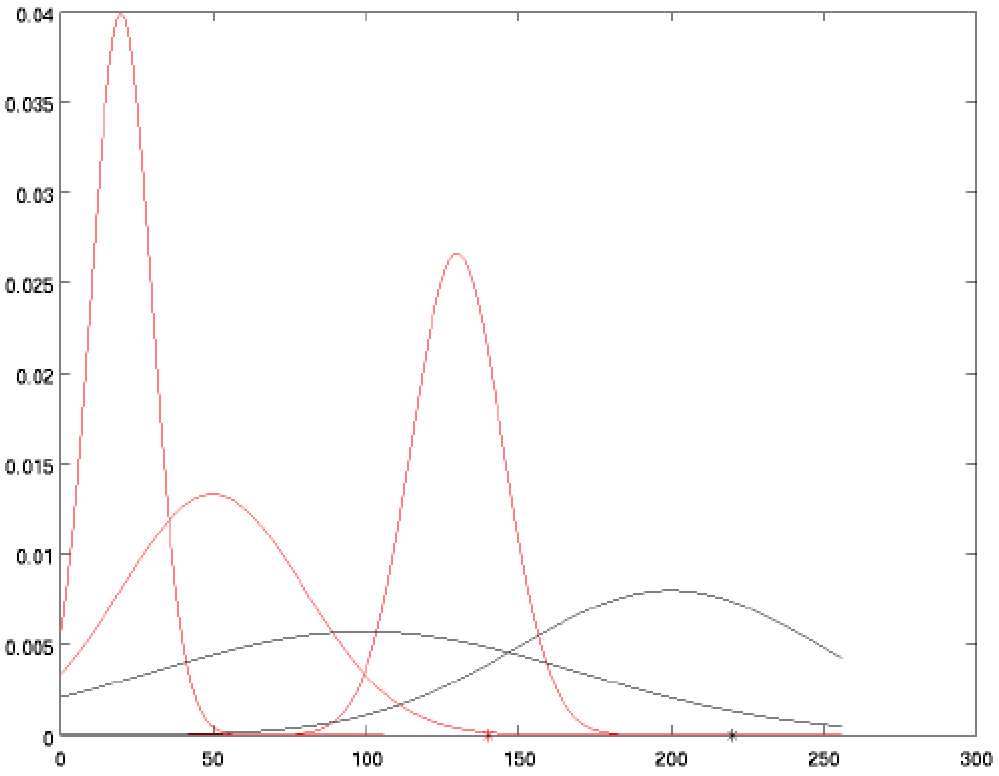
- None of the $K$ Gaussians match pixel value $X_{t+1}$
  - Observed value not well explained by model
- Replace the least probable distribution with a new one
  - Newly created distribution based on current value
    - $\mu_{t+1} = X_{t+1}$
    - Has high variance and low prior weight
  - Least probable in the $\omega/\sigma$ sense (to be explained)

# BACKGROUND MODEL ESTIMATION

- A background pixel value should be consistent
- Heuristic: Gaussians with the most **supporting evidence** and **least variance** should correspond to the background
- Gaussians are ordered by the value of $\omega/\sigma$
  - High support $\omega$ and smaller variance $\sigma$ give larger value
- First $B$ distributions are selected as the background model
  - $B = argmin_b(\sum_{i=1}^{b} w_i > T)$
    - $T$ minimum portion of image expected to be background

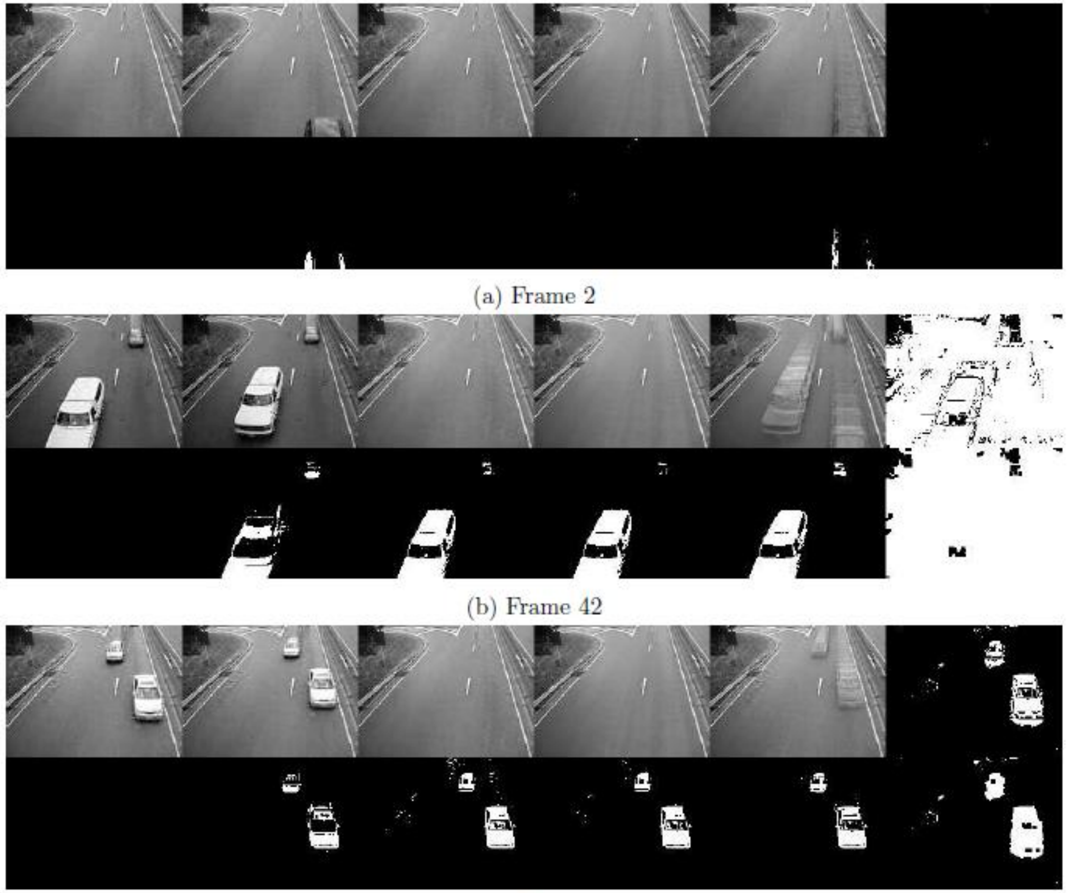
# BACKGROUND ESTIMATION EXAMPLE

- After background estimation, red are the background and black are foreground (not background)



# RESULTS

- Not much in paper, comparison from homework
- Notice frame difference tends to result in "holes" of car
- GMM performs very poorly on frame 42
- Adaptive model (while smeared background) performs quite well
  - This is a very short sequence with limited lighting variation
  - Generally, very difficult to select a single frame as "background"



(a) Frame 2

(b) Frame 42

(c) Frame 92

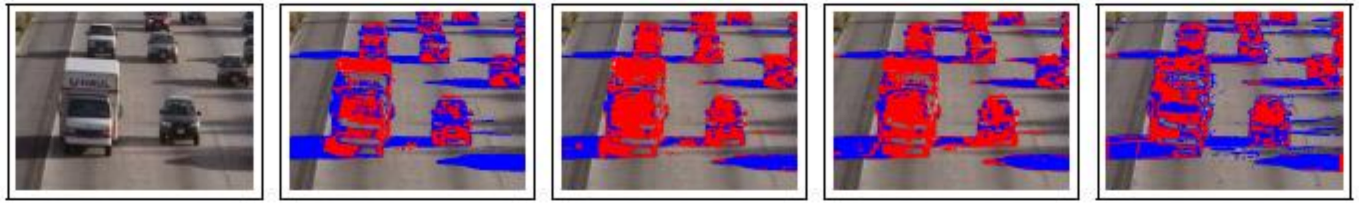
Figure 1: Background subtraction. Left column (raw image), column 2 (frame difference), column 3 (last frame background), column 4 (average background), column 5 (adaptive background), Right column (Gaussian mixture model detections (top), cleaned (bottom)).

# DISCUSSION

- Advantages
  - Different threshold for each pixel
  - Pixel-wise thresholds adapt over time
  - Objects are allowed to become part of the background without destroying the existing background model
  - Provides fast recovery
- Disadvantages
  - Cannot handle sudden, drastic lighting changes
  - Must have good Gaussian initialization (median filtering)
  - There are a number of parameters to tune

# MORE ISSUES?

- Shadows detection
  - [Prati, Mikic, Trivedi, Cucchiara 2003]



(a) Raw image (b) SNP result (c) SP result (d) DNM1 result (e) DNM2 result

- Chen & Aggarwal: The likelihood of a pixel being covered or uncovered is decided by the relative coordinates of optical flow vector vertices in its neighborhood.
- Oliver et al.: "Eigenbackgrounds" and its variations.
- Seki et al.: Image variations at neighboring image blocks have strong correlation.

# SIMPLE IMPROVEMENT

- Incorporate both spatial and temporal information into the background model
- Adaptive background mixture model + 3D connected component analysis [Goo et al.]
  - $3^{rd}$ dimension is time

# SUMMARY

- Simple background subtraction approaches such as fame diff, mean, and median filtering are fast
  - Constant thresholds make them ill-suited for challenging real-world problems
- Adaptive background mixture model approach can handle challenging situations
  - Bimodal backgrounds, long-term scene changes, and repetitive motion
- Improvements include upgrade the approach with temporal information or using region-based techniques

# THANK YOU

- Questions?

Background subtraction implementation using GMM at OpenCV

# REFERENCES

- Reading
  
  - Stauffer, Chris; Grimson, W.E.L., "Adaptive background mixture models for real-time tracking," in Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on. , vol.2, no., pp.252 Vol. 2, 1999
  - Kyungnam Kim, Thanarat H. Chalidabhongse, David Harwood, Larry Davis, "Real-time foreground–background segmentation using codebook model," Real-Time Imaging, Volume 11, Issue 3, June 2005, Pages 172-185
  

- Background Subtraction Datasets
  - https://sites.google.com/site/backgroundsubtraction/test-sequences