Video Segmentation Background modeling

K.K. Delibasis, Department of Computer Science and Biomedical Informatics
Lamia, Greece
kdelibasis@yahoo.com

Contents

Video segmentation with approximated median filtering

- Median filtering along the frames is an efficient estimator of the background
- Due to the complexity of median calculation, an incremental linear approximation is often used
- In the case of RGB video, the algorithm is applied to the Intensity of the pixel
- Symbols: FR: current frame, BG: current background, **p**: current pixel, index I: HSI intensity

Running average

- Μοντελοποιεί το υπόβαθρο βάσει του τρέχοντα μέσου όρου

$$B_i = \frac{1}{K}\sum_{k=0}^{K-1} F_{i-k}$$

- $\mathrm{B_i}$ είναι το υπόβαθρο και $\mathrm{F_i}$ το τρέχον πλαίσιο
- Απαιτεί μεγάλη μνήμη
- Όχι ακριβή αποτελέσματα

Running average with exponential forgetting

- Update of background values

$$\mu_t = aI_t + (1-a)\mu_{t-1}$$
$$\mu_0 = I_0$$

- Condition for background

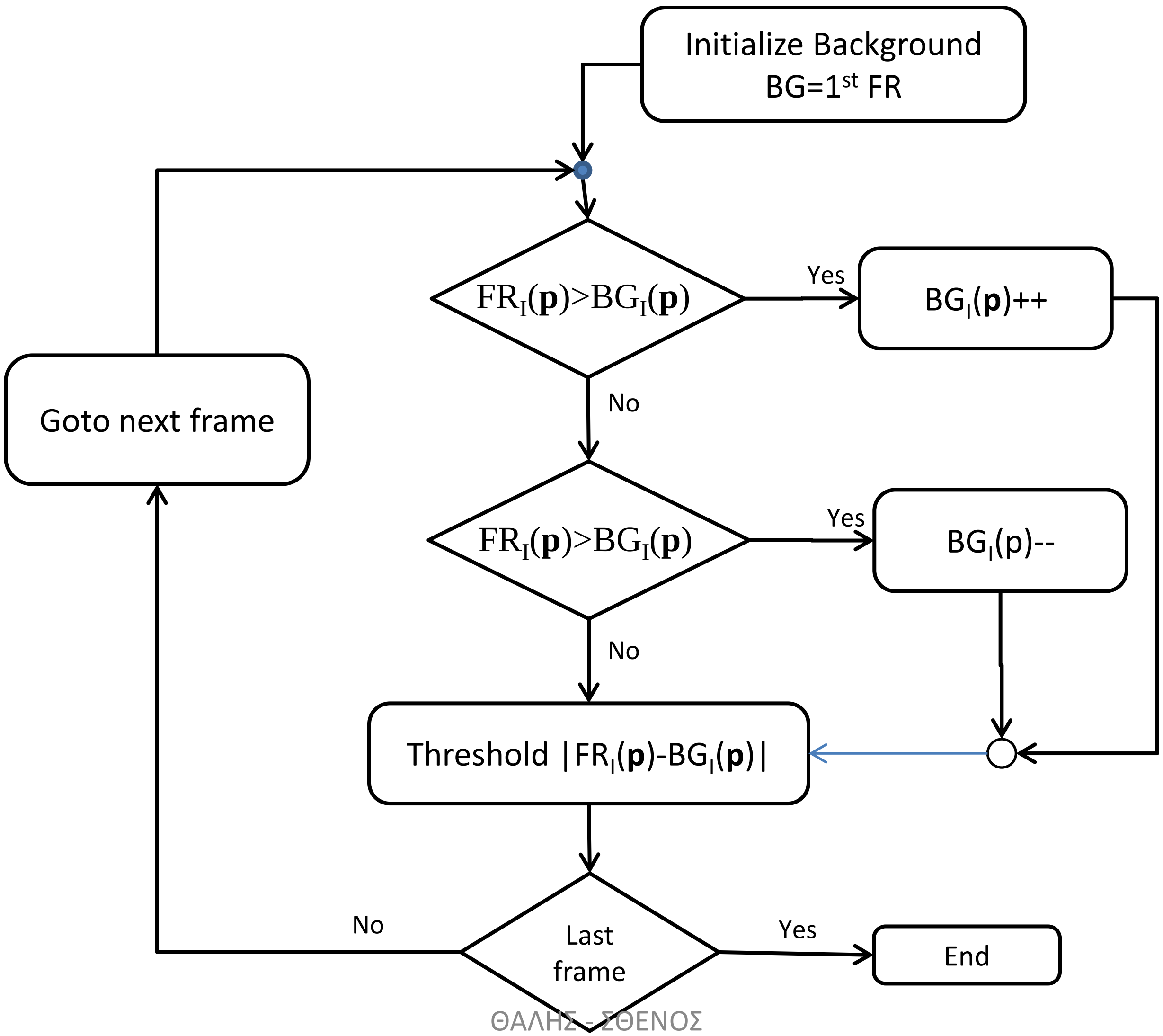
$$\left| I_{ij} - \mu_{t,ij} \right| < k\sigma_{t,ij} \Rightarrow B_{ij} = 1$$

- Selective background update
- a learnng rate- approx. 0.05

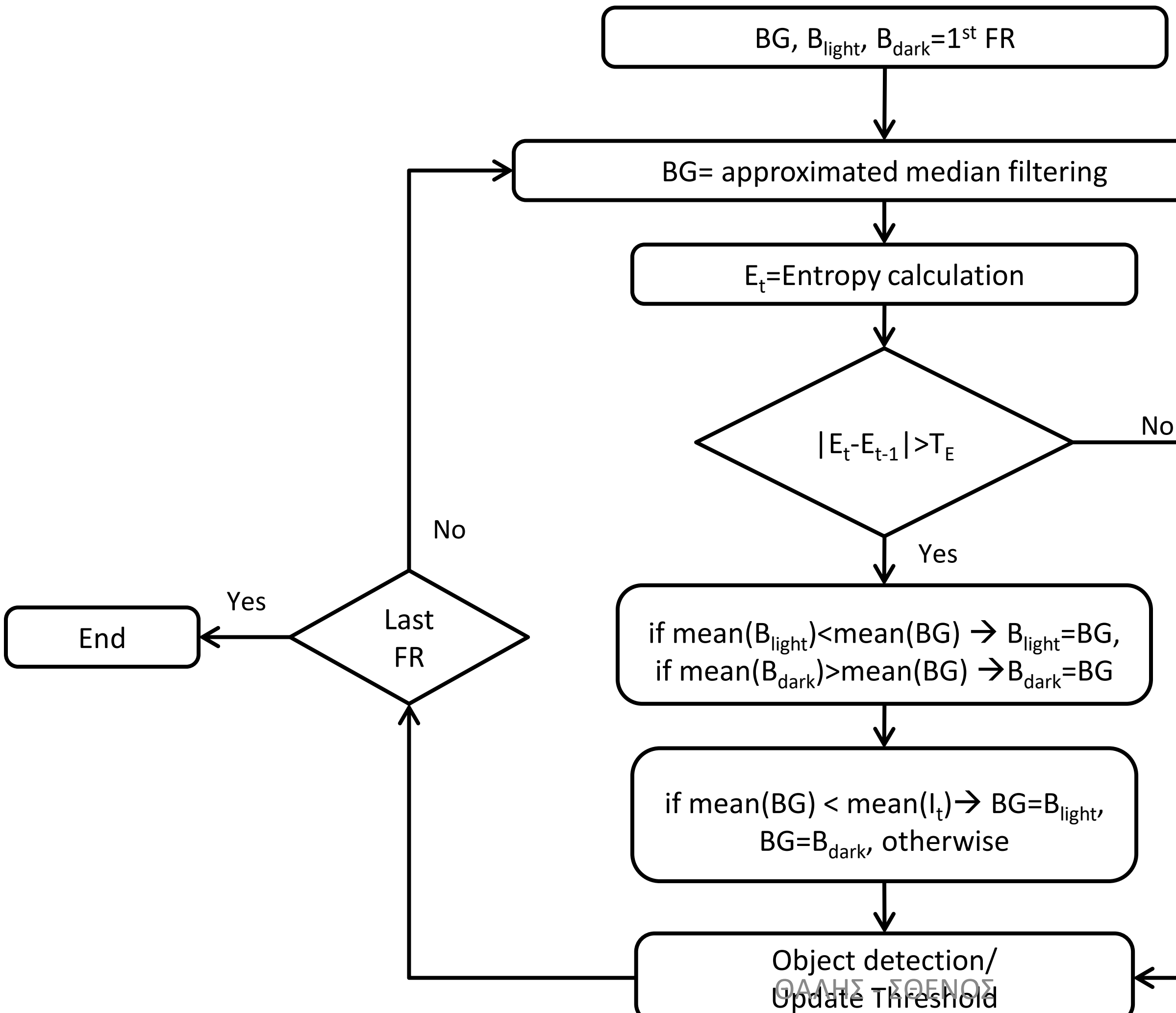
Approximated median filtering

- Κατά προσέγγιση ενδιάμεσο φιλτράρισμα
- Όταν ένα pixel στο τρέχον πλαίσιο έχει τιμή μεγαλύτερη από το pixel του υπόβαθρου το υπόβαθρο αυξάνεται κατά ένα, αν είναι μικρότερη μειώνεται κατά ένα.
- Το υπόβαθρο συγκλίνει σε μια καλή εκτίμηση
- Μεγάλη ακρίβεια, χαμηλές απαιτήσεις μνήμης, μικρή προσαρμοστικότητα
- Εναλλακτική: Mixture of Gaussians (MoG)

Video segmentation with approximated median filtering

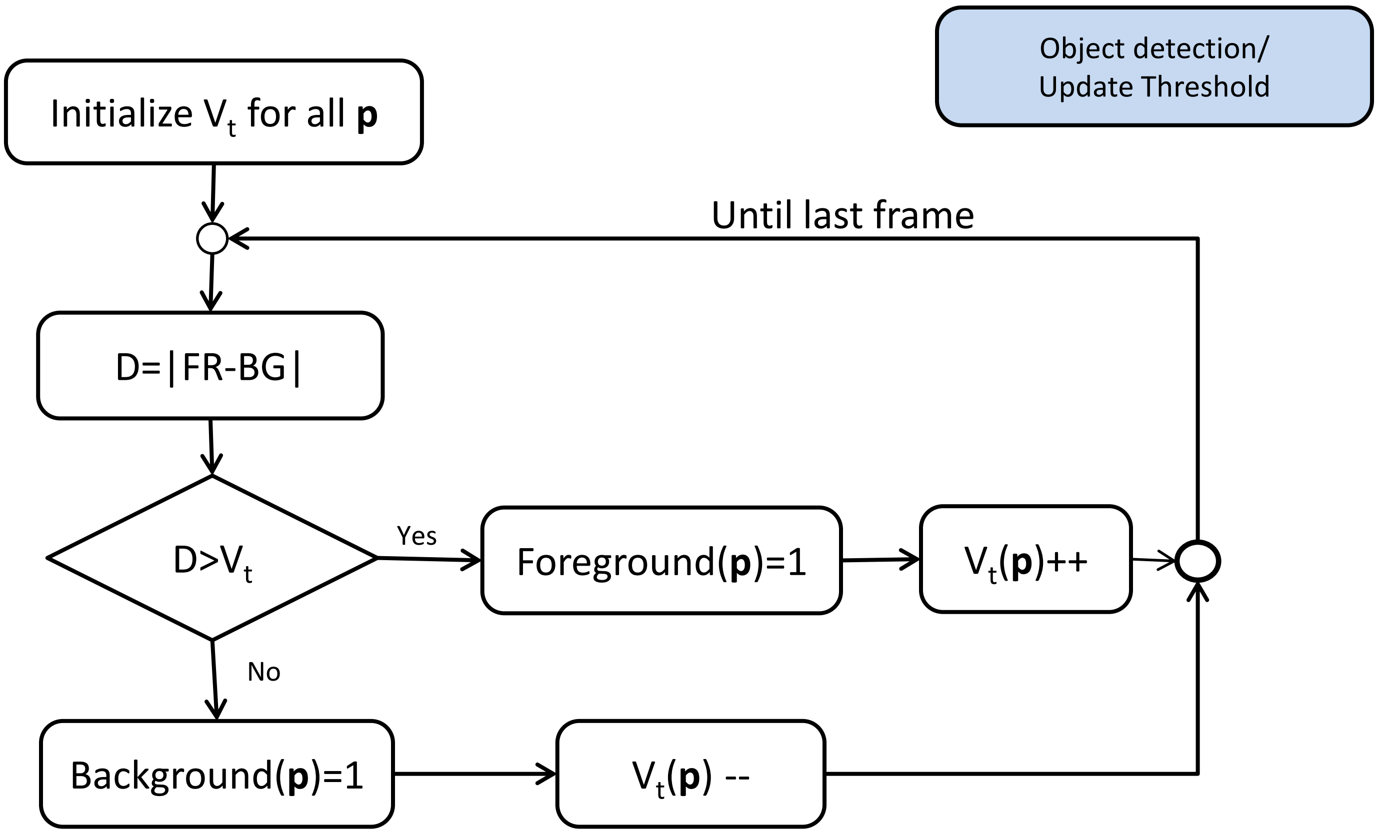


Background modeling using the Illumination sensitive method



F. C. Cheng, S. C. Huang, S. J. Ruan," Illumination-Sensitive Background Modeling Approach for Accurate Moving Object Detection," *IEEE trans. on broadcasting, vol. 57, no. 4, dec. 2011*

Illumination sensitive method: Video segmentation -



Gaussian Mixture models: 1D case

- An unknown distribution may be approximated using a number of Gaussians (=Normal distributions)
- Each Gaussian k is parametrized by μ_k, σ_κ and its weight a_k.
- Problem statement:
  – *Given a set of points, determine the parameters of a mixture of K Gaussians and label each point to a Gaussian component, so that maximize the likelihood of observing these points these points is maximized*

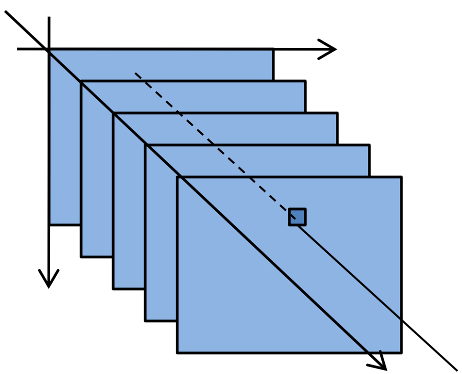
Mixture of Gaussians (MoG) algorithm for video segmentation

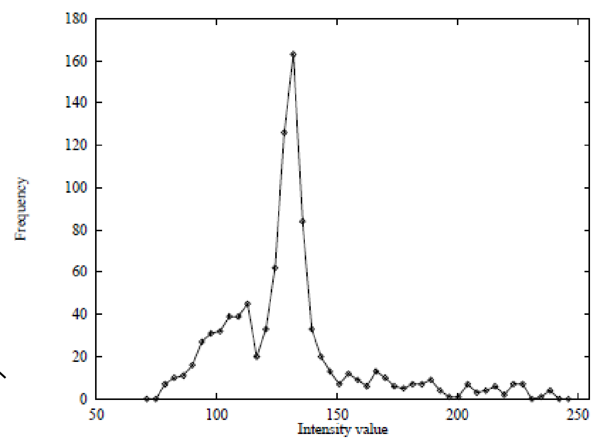
- The histogram of the values of each pixel is assumed to follow a probablility density function (pdf) that is the sum of a number of normalized Gaussian components (GC)
- The MoG algorithm,
  - given the pixel value x at the current frame, it determines iteratively the parameters of each GC: the number of counts, mean and variance (m_k, σ_k^2 , c_k, for the kth GC)
  - Segments the foreground by assuming that the GC with greatest counts and lower varience correspond to background values.

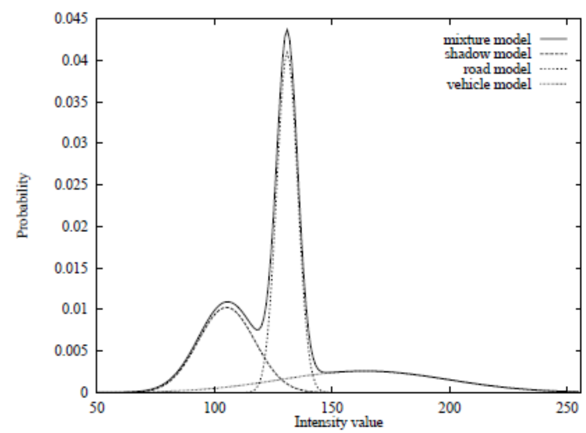
C. Stauffer, E. Grimson, "Adaptive background mixture models for real-time tracking," *IEEE CVPR*, 1999.
C. Stauffer, E. Grimson, "Learning patterns of activity using realtime tracking," *IEEE T PAMI*, 22, 747–757, 2000

Mixtue of Gaussians MoG

- The value of a pixel with time is considered a random variable following a pdf (Prob. Density function) that is modelled by a sum of Gaussian compomentents.
- The labelling of these components as background or foreground is automatic







Estimating the parameters of the Gaussians with EM method

- Let us assume *K* Gaussian components with weight a_c and parameters (μ_c, σ_c), $c=1,2, \ldots K$.
- Responsibility (or Membership) matrix: R: r_{ic} the probability that sample i belongs to cluster c.

$$r_{ic} = \frac{a_c N\left(x_i ; \mu_{\mathrm{c}}, \Sigma_{\mathrm{c}}\right)}{\sum_c a_c N\left(x_i ; \mu_{\mathrm{c}}, \Sigma_{\mathrm{c}}\right)}$$

Expectation Step

- Change the parameters of the distribution to maximize Likelihood

Maximization Step

- Goto Expectation step and iterate until convergence

- **Maximization step** in the case of a 1D Gaussian the (normal) PDF $N(x_i; m_c, \sigma_c)$ with weight a_c

$$N\left(x_i;\mu_c,\sigma_c\right)=\frac{1}{\sigma\sqrt{2\pi}}e^{\frac{\left(x_i-\mu_c\right)^2}{2\sigma^2}}$$

- Update the Gaussian parameters

$$a_c=\frac{N_c}{N}$$

$$\mu_c=\frac{1}{N_c}x_i r_{ic}, N_c=\sum_{i=1}^{N} r_{ic}$$

$$\sigma_c^2=\frac{1}{N_c}\sum_{i=1}^{N} r_{ic}\left(x_i-\mu_c\right)^2$$

- In the case of RGB images, the normal distribution considers 3 color components:

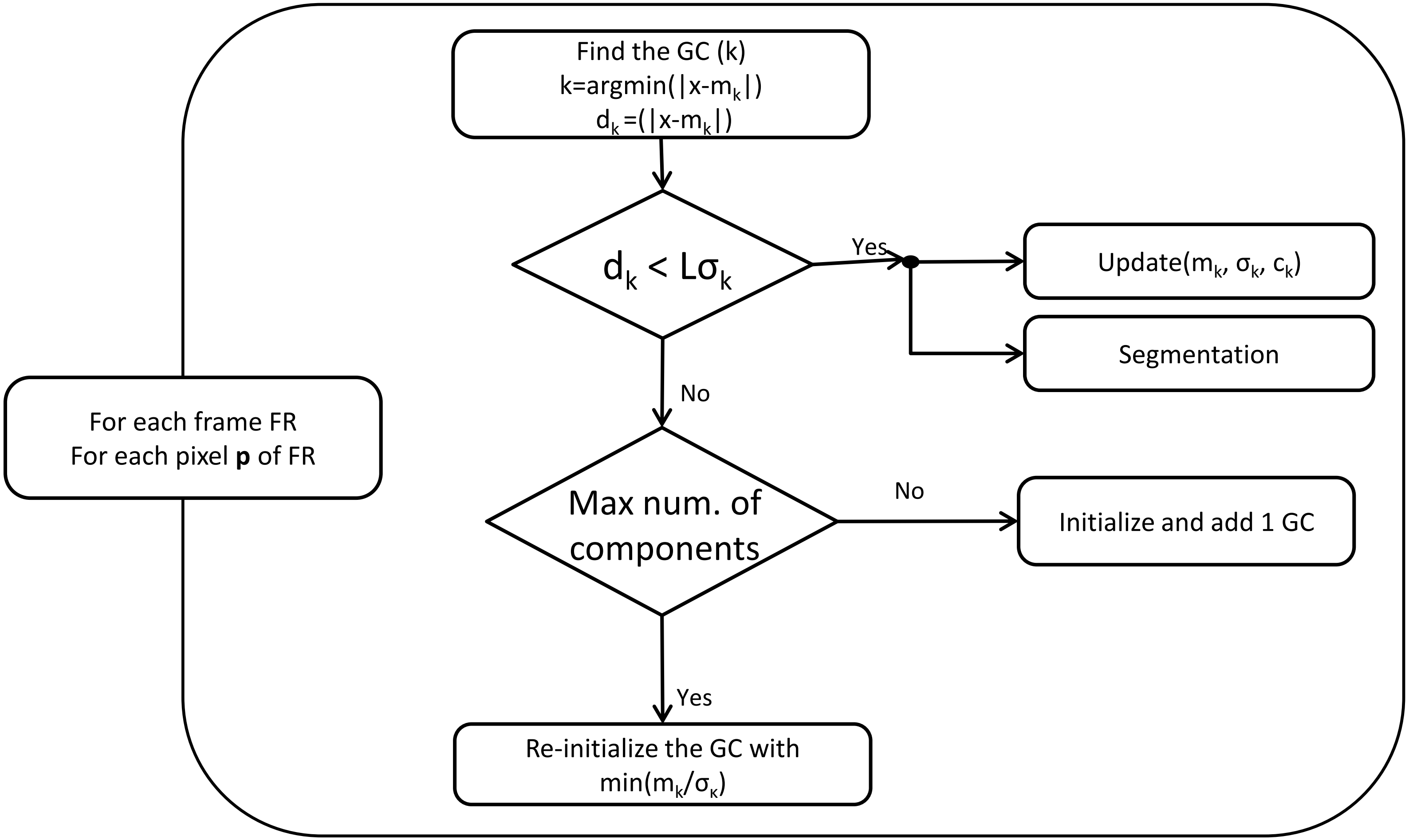
$$N\left(\mathbf{x_i};\boldsymbol{\mu}_\mathbf{c},\Sigma\right)=\frac{1}{(2\pi)^{\frac{3}{2}}\left|\Sigma\right|^{\frac{1}{2}}}e^{-\frac{1}{2}(\mathbf{x_i}-\boldsymbol{\mu}_\mathbf{c})\Sigma^{-1}(\mathbf{x_i}-\boldsymbol{\mu}_\mathbf{c})^T},\ \mathbf{x_i}=\begin{bmatrix}R_i\\G_i\\B_i\end{bmatrix}$$

- In case of independent RGB colors,

$$\Sigma=\begin{pmatrix}\sigma_r & 0 & 0\\0 & \sigma_g & 0\\0 & 0 & \sigma_b\end{pmatrix}$$

- In practice, applying the EM for GMM is not feasible for video segmentation in real time
- Thus, a modified EM for GMM is applied: The determination of the parameters of the gaussians is done incrementally
  - Approximate
  - Very fast
  - Memory efficient

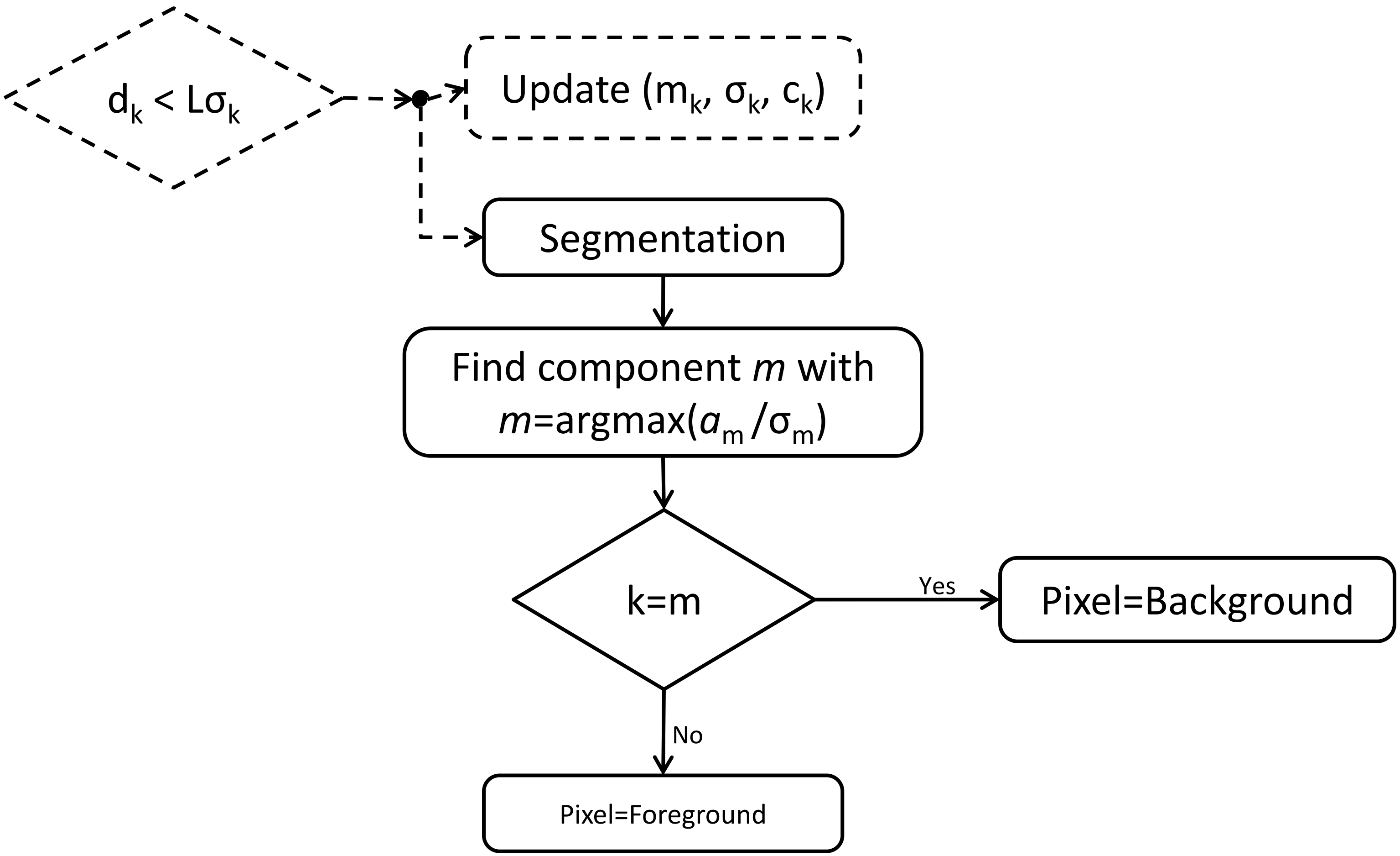
Video segmentation using Modified MoG algorithm



Define foreground

- There are several ways to define the foreground / background:
- for each pixel with value v.
- First the value of all Gaussian components is calculated $y_c = N\left(x_i; \mu_c, \sigma_c\right), c = 1, 2, \ldots, K$
- The matching component is identified:
  - $y_c > T$, or $|v - m_c| < k\,\sigma_c$
- Background is considered:
  - The (1 or more) component with the largest *a*/σ, OR all components (depending on the definition)

Segmentation using MoG (cont.)



Incremental Update of the parameters of the Gaussian components

$$m_k(\mathbf{p}) = (1-a)\,m_k(\mathbf{p}) + a \cdot FR(\mathbf{p})$$

$$c_k = c_k + 1$$

$$\sigma_k = \begin{cases} \sigma_k + \delta\sigma, & \text{if } \left|FR(\mathbf{p}) - m_k\right| > 2\sigma_k \\ \sigma_k - \delta\sigma, & \text{if } \left|FR(\mathbf{p}) - m_k\right| < 0.5\sigma_k \end{cases}$$

α=learning rate

Algorithm modified from
T. Bouwmans, F. El Baf, B. Vachon, "Background Modeling using Mixture of Gaussians for Foreground Detection - A Survey". *Recent Patents on Computer Science* 1, 3 (2008) 219-237

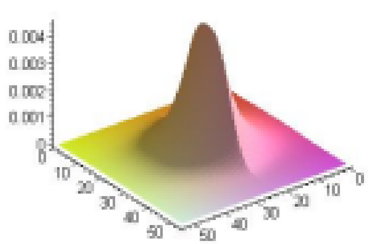
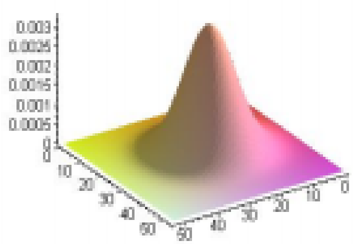
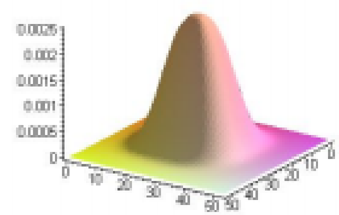
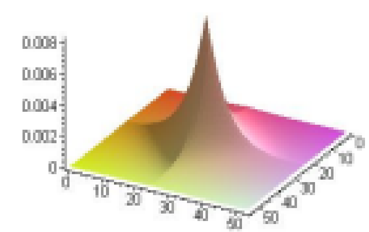
MoG variants

- A large number of variants exist, that
  - Handle spatial pixel relations
  - Change the incremental Gaussian parameter update
  - Handle shadows.
  - Incorporate Markov Random Fields, Hierarchical approaches, Graph cuts etc
- Even with several heuristics, MoG is computationally intensive and cannot be in real time for large video frames without specialized hardware

Thierry Bouwmans, Recent Advanced Statistical Background Modeling for Foreground, Detection - A Systematic Survey

Generalized MoG

- Same algorithm as with the MoG, but the shape of the Gaussian pdf is generalized by extra paramters

- Effective Gaussian mixture learning for video background subtraction, Lee, Dar-Shyang S., PAMI 27(5), 827-832, 2005
- M. Allili, N. Bouguila, D. Ziou,"Finite Generalized Gaussian Mixture Modelling and Application to Image and Video Foreground Segmentation", Journal of Electronic Imaging, 2008

Initial frame

Segm. with MoG

Segm. with generalized MoG

PCA and ICA

- Any frame is considered a mixed result of the background and the foreground object, which are recovered by ICA.
- ICA determines a small matrix (2x2 if one background and 1 foreground is considered), that when multiplied by the input frame produces the background and the foreground image, uder criterions like:
  - Minimizing mutual information between the gistograms of the two component images, or
  - requiring the double histogram to be separable etc.
- Stochastic optimization llike PSO is emploeyd for video segmentation
- Offline training, real time execution for small frames.
- Du-Ming Tsai and Shia-Chih Lai, Independent Component Analysis-Based Background Subtraction for Indoor Surveillance, IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 18(1), JANUARY 2009

- Example video segmentation using ICA

Background frame for training

Backgound component

Frame with forground object

Foreground component

Resulting segmentation

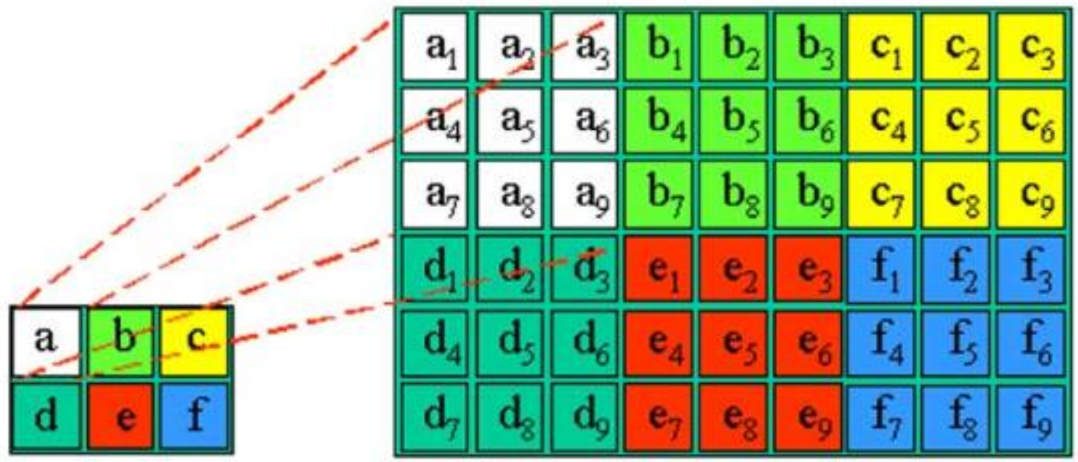
A SOM-based approach

- Self Organizing Maps (SOMs) are a class of **unsupervised** neural networks
- SOM learn by re-enforcing similarity: each neuron contains a number of features.
- Each input to the SOM is compared to all neurons
  - the most similar (winning neuron) is modified to be more similar to the input (controlled by the learning rate *a*)
  - The neighbors of the winner are modified less heavily
  - The size of the neuron neighborhood and a are relaxed with time

A Self-Organizing Approach to Background Subtraction for Visual Surveillance Applications, IEEE T Im Proc, 2008

Video segmentation using SOM

- Self organizing maps are ANN based on competitve learning and require no training set.
- A 2D set of neurons of multiple dimensions of the video frame used to model the background. Each neuron has 3 weights
- Subsequent frames are presented to the SOM so that they model the background.



Frame pixels Neurons corresponding to pixels

- Distance between frame pixels and neurons is defined in HSV

$$d(p_i, p_j) = \|(v_i s_i \cos(h_i), v_i s_i \sin(h_i), v_i) \\ - (v_j s_j \cos(h_j), v_j s_j \sin(h_j), v_j)\|_2^2.$$

- Neuron update

$$A_t(i,j) = (1 - \alpha_{i,j}(t))A_{t-1}(i,j) + \alpha_{i,j}(t)p_t(x,y)$$

- Detecting shadows in HSV (and characterize it as background)

$$\left(\gamma \leq \frac{p_t^V}{c_i^V} \leq \beta\right) \wedge \left(p_t^S - c_i^S \leq \tau_S\right) \wedge \left(|p_t^H - c_i^H| \leq \tau_H\right)$$

- Maddalena L., Petrosino A. A self organizing approach to background subtraction for visual surveillance applications, IEEE Transactions on Image Processing, Volume17, No. 7, pages 1729–1736, 2008

A non-parametric approach

Elgammal, A., Harwood, D., & Davis, L. (2000) Non-parametric model for background subtraction. *ECCV, >3000 citations*.

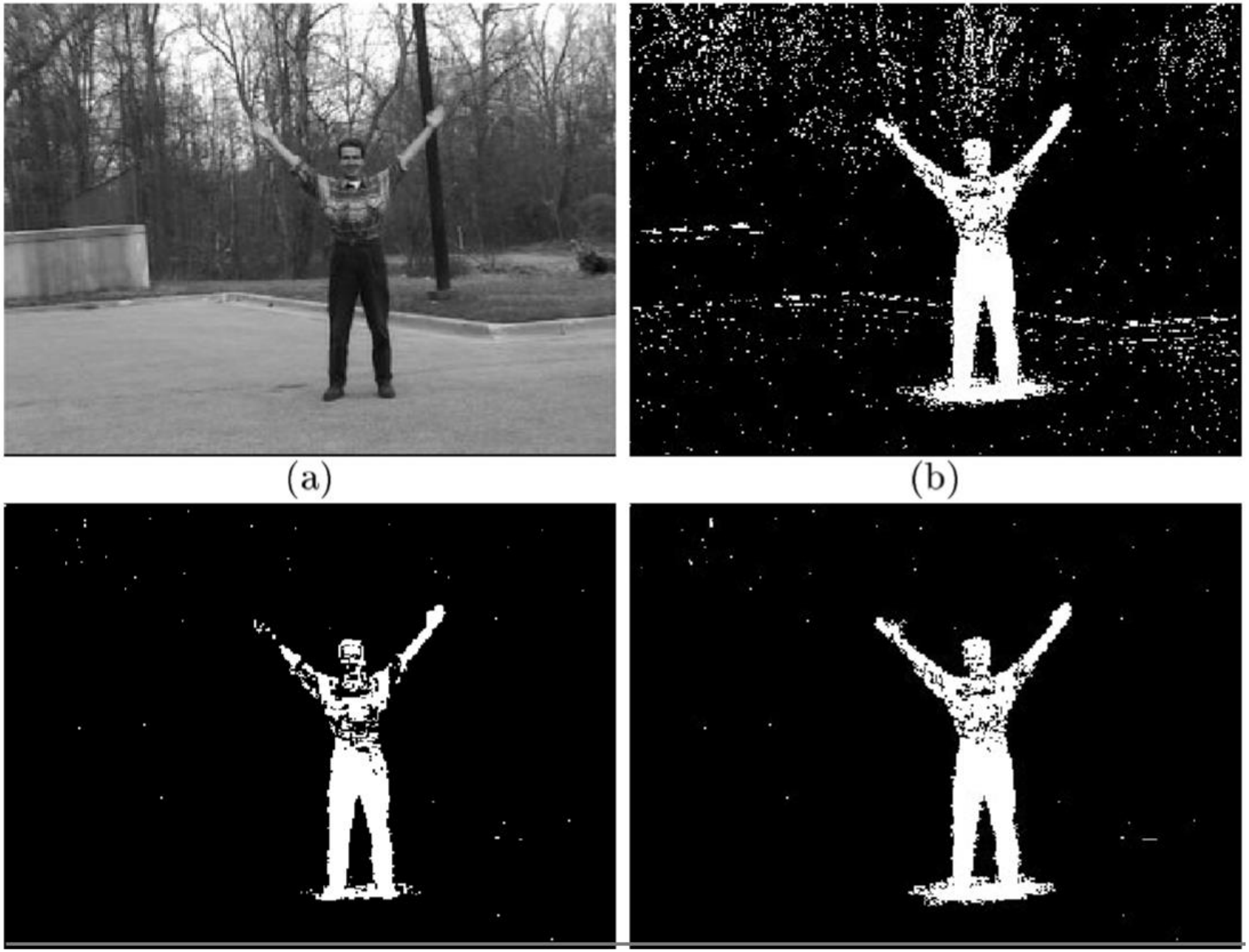
- Given the current P, pixel **p** with value x_t is considered foreground if $P(x_t) < th$, where th is a global threshold
- The width σ of the kernel is estimated for each pixel (and for each color, if applicable) as following:
  - The median value m of $|x_{t+1} - x_t|$ for a number of frames is calculated

$$\sigma = \frac{\mu}{0.68\sqrt{2}}$$

- Reducing false positives:
  - Background objects may move slightly, thus appear in new pixels → be considerate as foreground

- To reduce this effect:
  - For any pixel **p** with value x_t classified as foreground:
    - Calculate the probability of this pixel x_t being background in the neighborhood of p.


(a)
(b)

Shadow suppression

- The shadow suppression algorithm processes the current frame in HSI and produces a binary frame of the human shadow
- The main idea is that shadowed pixels have the same color components with the background pixels, but less intensity.
- Shadowed pixels are detected according to the following criterion

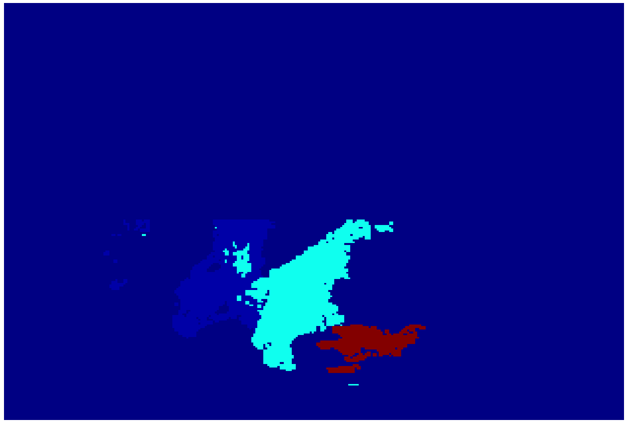
$$a \leq \frac{FR_I(\mathbf{p})}{BG_I(\mathbf{p})} \leq b, D_H(\mathbf{p}) \leq \tau_s, \left|FR_s(\mathbf{p}) - BG_s(\mathbf{p})\right| \leq \tau_{\mathrm{H}}$$

$$D_H(p) = \min\begin{pmatrix} \left|\mathrm{FR}_{\mathrm{H}}(p) - \mathrm{BG}_{\mathrm{H}}(p)\right|, \\ 2\pi - \left|\mathrm{FR}_{\mathrm{H}}(p) - \mathrm{BG}_{\mathrm{H}}(p)\right| \end{pmatrix}$$

Results

Illumination-sensitive

Approximated median



Mixture of Gaussians

The result of the segmentation of frames: 31, 36, 52 from the 1st video after the application of: 1) the illumination-sensitive, 2) the approximated median, 3) the MoG method