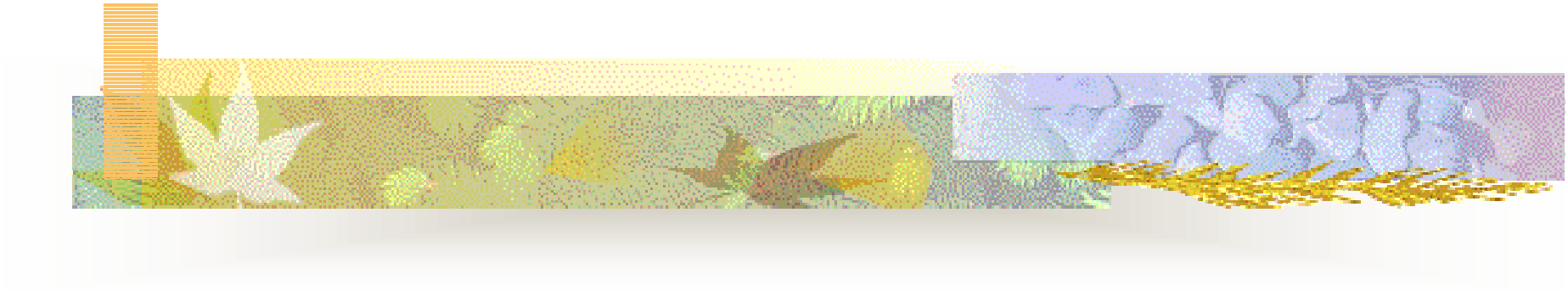
Class 2: Moving Object Detection

Feb. 4, 2008



Instructor: YingLi Tian

Video Surveillance E6998-007 Senior/Feris/Tian

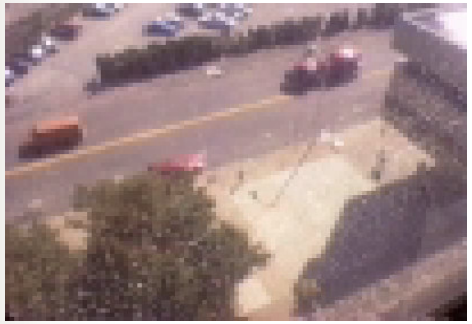
Some contents from Massimo Piccardi, UTS and Christopher Rasmussen, UDEL

Outlines

- Moving Object Detection Problem
  - Static Camera
  - Moving Camera (next class)
- The Basic Methods
- The Advanced Methods (next class)

Moving Object Detection Problem (Static Camera)

- Main goal: detecting moving objects from a video sequence of a fixed camera
- Background – static scene
- Foreground – moving objects
- Approach: detect the moving objects as the difference between the current frame and the image of the scene background.



from C. Stauffer and W. Grimson

Current image

Background image

Foreground pixels

Challenges

- How to automatically obtain the background image?
- Foreground processing
- The background image must adapt to:
  - Illumination changes (gradual and sudden)
  - Distracting motions (camera shake, swaying tree, moving elevators, ocean waves…)
  - Scene changes (parked car)
- Others:
  - Shadows
  - Black/blue screen
  - Bad weathers
  - Foreground fragment

Background Subtraction Methods (BGS)

* Basic BGS
* Running Gaussian average
* Mixture of Gaussians
* Enhanced mixture of Gaussians
* Kernel Density Estimators
* Mean-shift based estimation
* Combined estimation and propagation
* Eigenbackgrounds

Basic BGS – basic idea (1)

- Pixels belongs to foreground if | Current Frame – BG Image| > Threshold
- BG image can be just the previous frame or the average image of a number of frames
- Works only in particular conditions of objects' speed and frame rate
- Very sensitive to the threshold

Basic BGS – example (2)

Current Frame

Difference

High threshold

Low threshold

Basic BGS -- average (3)

- Median/average: use the average or the median of the previous *n* frames
  - Quick but memory consuming
- Running average
  - Alpha is adapt rate (0.05).

$$B_{i+1} = (1-\alpha)B_i + \alpha(F_i)$$

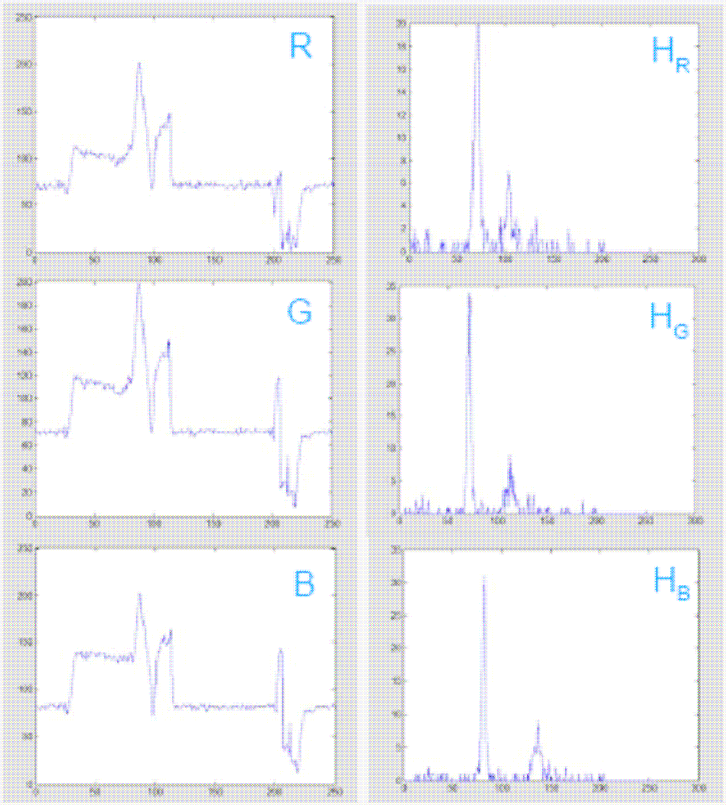
Basic BGS – rationale (4)

- The background model at each pixel location **is based on the pixel's recent history**
- In many works, such history is:
  - just the previous *n* frames
  - a weighted average where recent frames have higher weight
- In essence, the background model is computed as a chronological average from the pixel's history
- No spatial correlation is used between different (neighbouring) pixel locations

Basic BGS – histograms (5)

- Example:

pixel location



sequence of pixel values

histograms

Basic BGS – selectivity (6)

- At each new frame, each pixel is classified as either foreground or background
- What feedback from the classification to the background model?

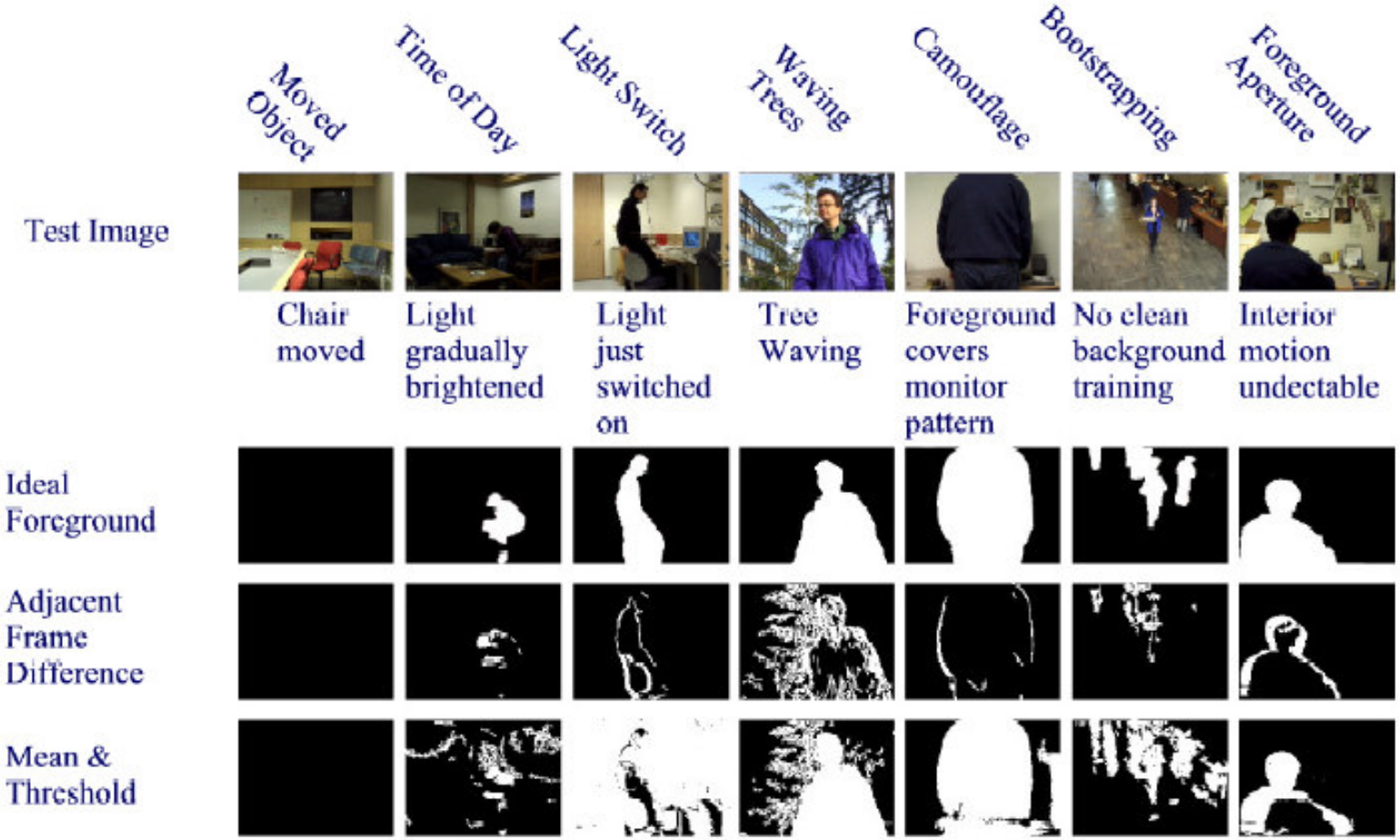
  → if the pixel is classified as foreground, it is ignored in the background model
- In this way, we prevent the background model to be polluted by pixel logically not belonging to the background scene

- Running average with selectivity:

$$B_{i+1}(x,y) = \alpha F_t(x,y) + (1-\alpha) B_t(x,y) \quad \text{if } F_t(x,y) \text{ background}$$

$$B_{i+1}(x,y) = B_t(x,y) \quad \text{if } F_t(x,y) \text{ foreground}$$

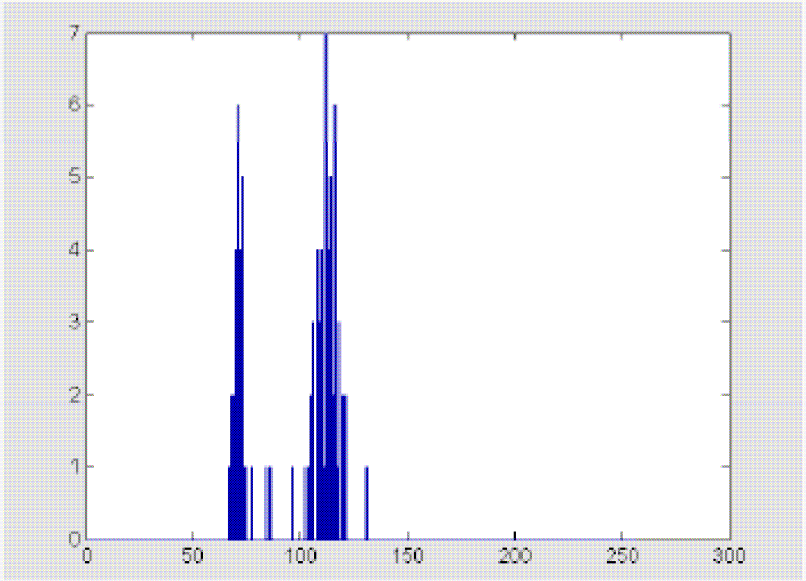
Basic BGS -- results



from K. Toyama et al.

Basic BGS -- limitations

- They do not provide an explicit method to choose the threshold
- Major: Based on a single value, they cannot cope with multiple modal background distributions; example:



Running Gaussian average

- *Pfinder* (Wren, Azarbayejani, Darrell, Pentland, 1997):
  - fitting one Gaussian distribution (μ,σ) over the histogram: this gives the background PDF
  - background PDF update: running average:

$$\mu_{t+1} = \alpha F_t + (1-\alpha)\mu_t$$

$$\sigma_{t+1}^2 = \alpha(F_t - \mu_t)^2 + (1-\alpha)\sigma_t^2$$

  - In test $|F - \mu| > Th$, Th can be chosen as $k\sigma$
  - It does not cope with multimodal backgrounds

Mixture of Gaussians (1)

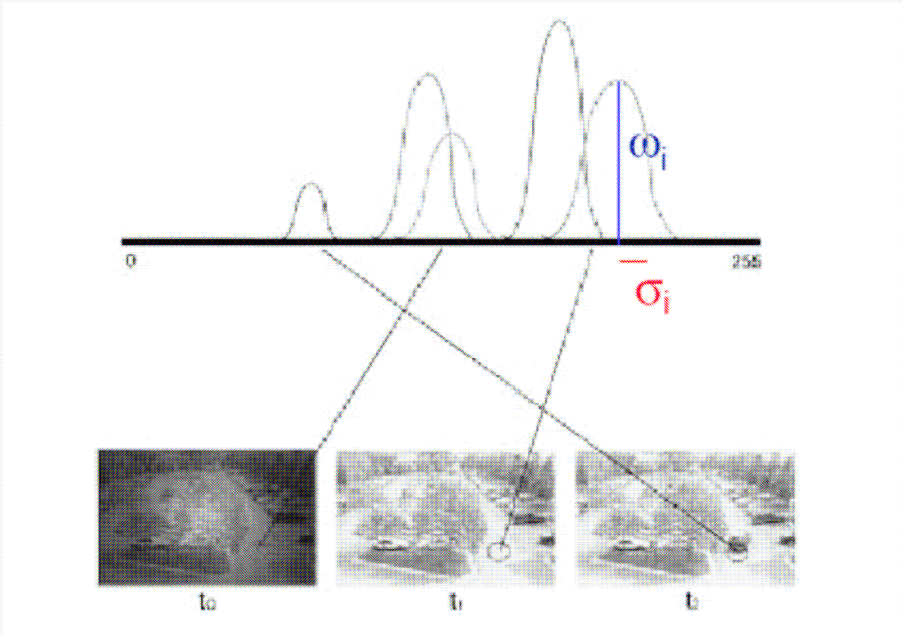
- *Mixture of K Gaussians* $(\mu_i, \sigma_i, \omega_i)$ (Stauffer and Grimson, 1999)
- In this way, the model copes also with multimodal background distributions; however:
  - the number of modes is arbitrarily pre-defined (usually from 3 to 5)
  - how to initialize the Gaussians?
  - how to update them over time?

Mixture of Gaussians (2)

- All weights ω_i are updated (updated and/or normalised) at every new frame
- At every new frame, some of the Gaussians "match" the current value (those at a distance < 2.5 σ_i): for them, μ_i, σ_i are updated by the running average
- The mixture of Gaussians actually models both the foreground and the background: how to pick only the distributions modeling the background?:
  - all distributions are ranked according to their ω_i / σ_i and the first ones chosen as "background"

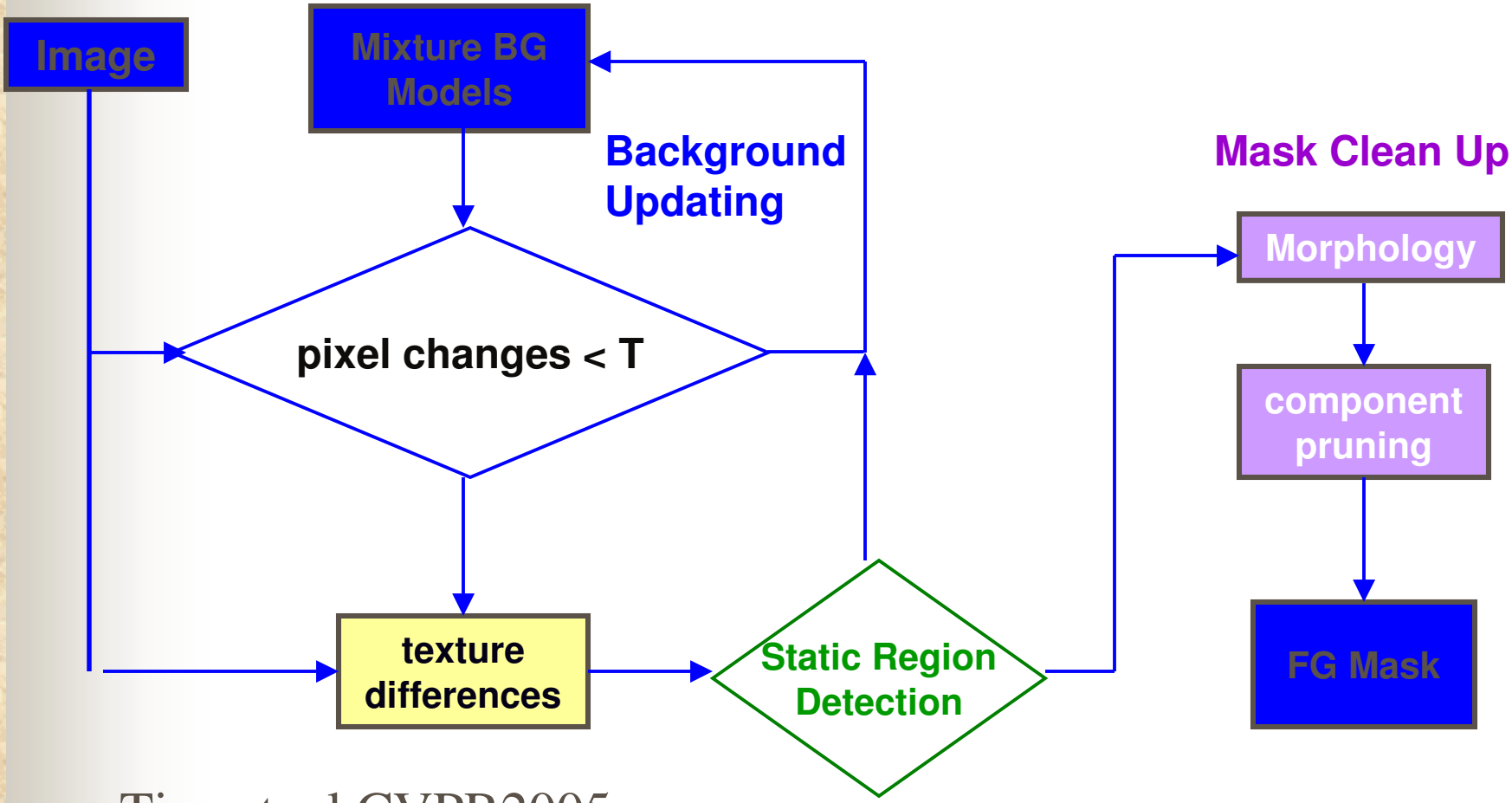
Mixture of Gaussians (3)

- Example:



(from: I. Pavlidis, V. Morellas, P. Tsiamyrtzis, and S. Harp, "Urban surveillance systems: from the laboratory to the commercial world," Proceedings of the IEEE, vol. 89, no. 10, pp. 1478 -1497, 2001)

Enhanced mixture of Gaussians



Tian et. al CVPR2005

Foreground Analysis

For each pixel at time *t*, the probability of the pixel:

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

where

$$\eta(X_t, \mu, \Sigma) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(X_t - \mu_t)^T \Sigma^{-1} (X_t - \mu_t)}, \quad \omega_{i,t} = (1-\alpha)\omega_{i,t-1} + \alpha(M_{k,t}).$$

The first *B* distributions are chosen as the background model:

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

The *B+1* distribution is used for static region detection :

$$pixel \in static \quad region, \quad if\ \omega_{B+1} > T.$$

Texture Integration for Quick Lighting Changes

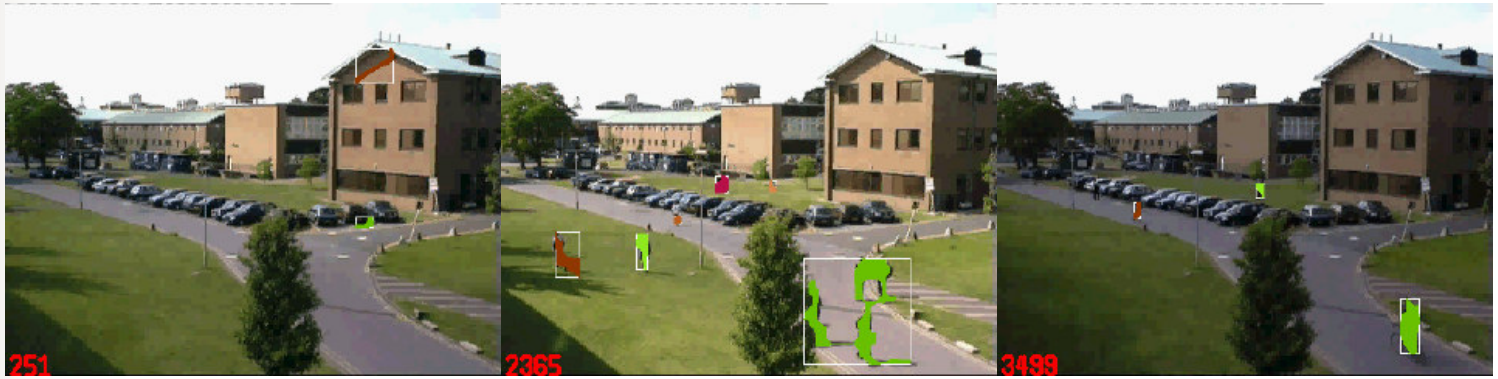
Measure texture similarity at pixel *X* between the current frame and the background image as:

$$S(X) = \frac{\sum_{u \in W_x} 2 \| g(u) \| \cdot \| g_b(u) \| \cos\theta}{\sum_{u \in W_x} (\| g(u) \|^2 + \| g_b(u) \|^2)},$$

In the false positive foreground areas caused by quick lighting changes, there are no texture changes between the current frame and the background.

Handle quick lighting changes

(a) BGS by a mixture of Gaussians only



(b) BGS by a mixture of Gaussians with the intensity and texture information

Intensity Integration for Shadow Removal

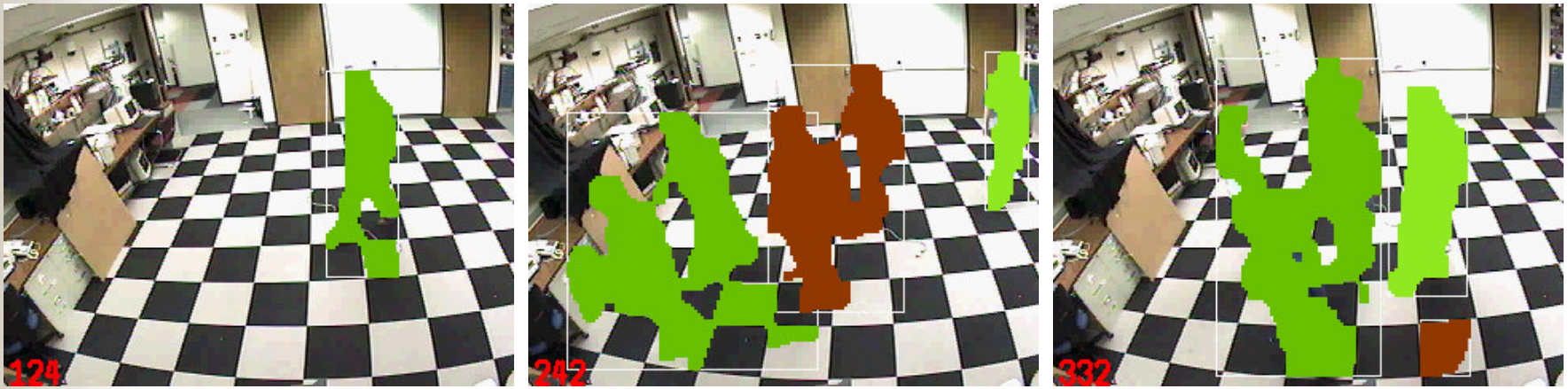
The normalized cross-correlation of the intensities at pixel *X* in the *M by N* neighborhood between the current frame and the background is calculated as :

$$NCC(X) = \frac{\sum_{u \in W_x} I_t(u) \cdot I_b(u) - \frac{1}{MN} \sum_{u \in W_x} I_t(u) \sum_{u \in W_x} I_b(u)}{\sqrt{\left(\sum_{u \in W_x} I_t^{\,2}(u) - \frac{1}{MN}\left[\sum_{u \in W_x} I_t(u)\right]^2\right)\left(\sum_{u \in W_x} I_b^{\,2}(u) - \frac{1}{MN}\left[\sum_{u \in W_x} I_b(u)\right]^2\right)}},$$

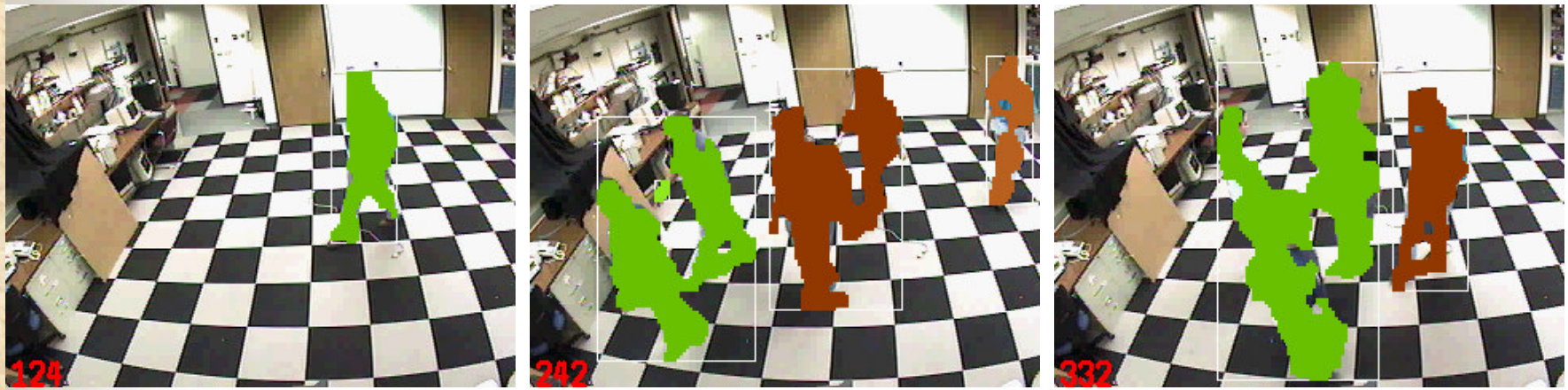
The pixel *X* is shadow if $NCC(X) \geq T_s$

and $I_t(X) \geq T_I$ (No shadow detections in very dark areas.)

Shadow removal



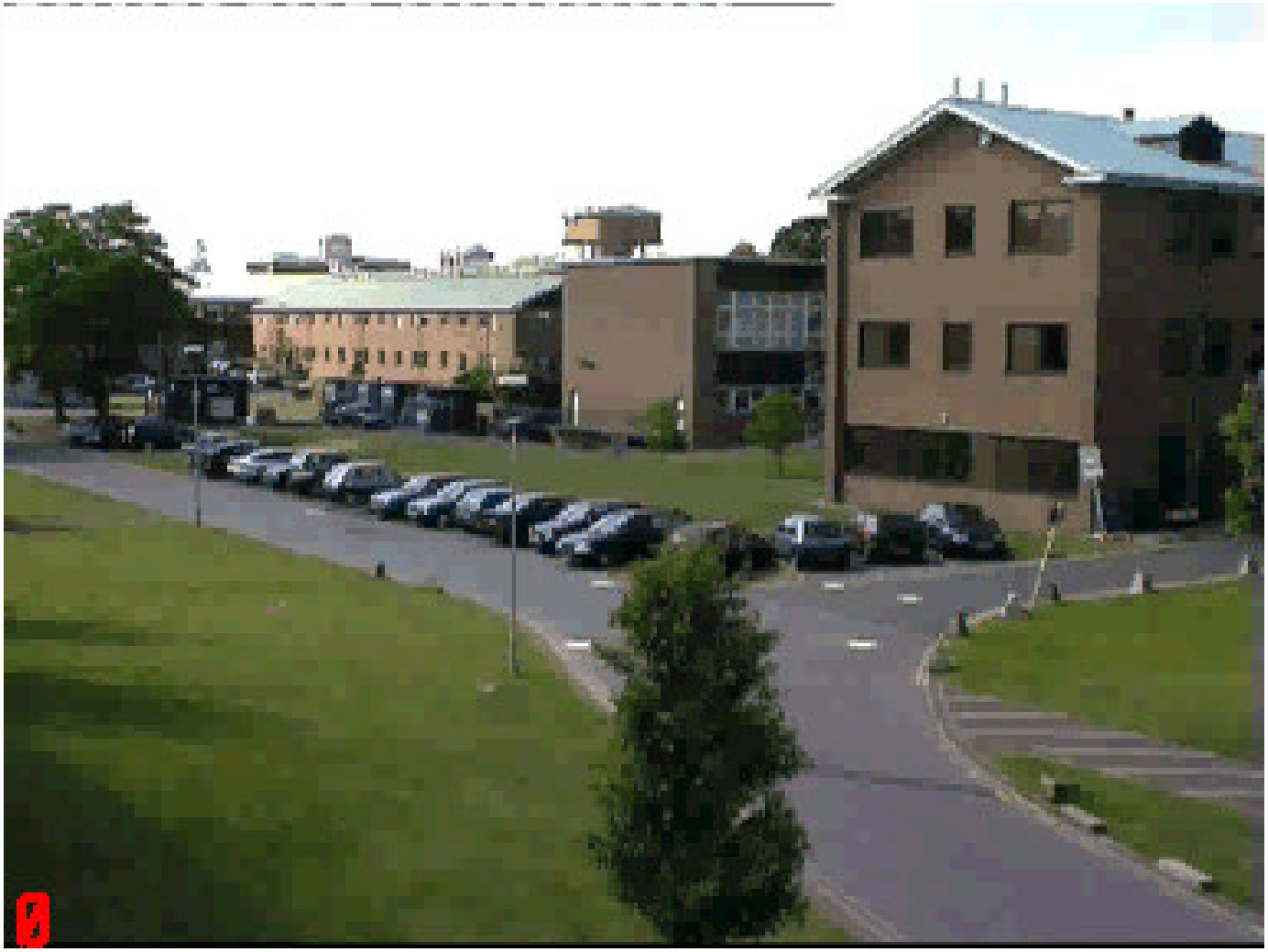
(a) BGS by a mixture of Gaussians only



(b) BGS by a mixture of Gaussians with the intensity and texture information

Results

quick lighting change use intensity



quick lighting change use intensity + texture

Other enhancement of Mixture of Gaussians

- Use region instead of single pixels (Eng, Wang, Kam, and Yau, 2006, more details in next class)
- Use Color and Gradient information (Javed, Shafique, and Shah, 2002)

Kernel Density Estimators – (1)

(Elgammal, Harwood, Davis 2000)

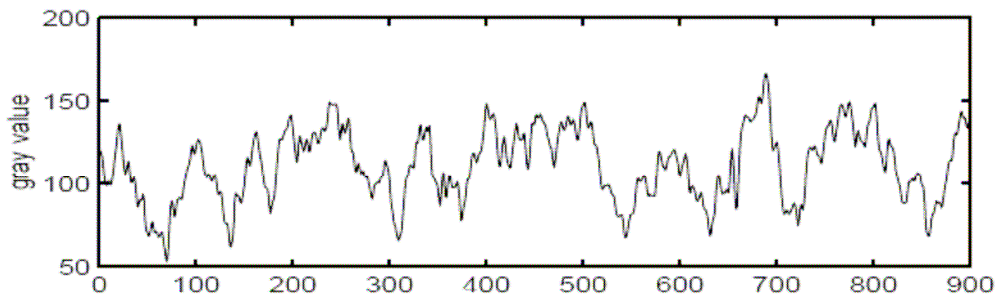


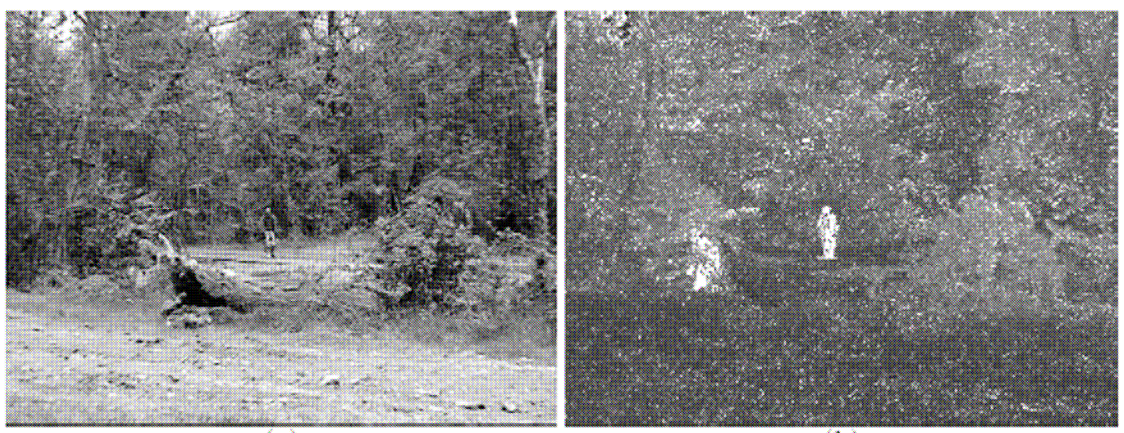
Fig. 1. Intensity value overtime

Fig. 2. Outdoor scene with a circle at the top left corner showing the location of the sample pixel in figure 1

Kernel Density Estimators – (2)

For N samples at a pixel, K is the kernel estimator

$$Pr(x_t) = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{(2\pi)^{\frac{d}{2}} \mid \Sigma \mid^{\frac{1}{2}}} e^{-\frac{1}{2}(x_t - x_i)^T \Sigma^{-1}(x_t - x_i)}$$

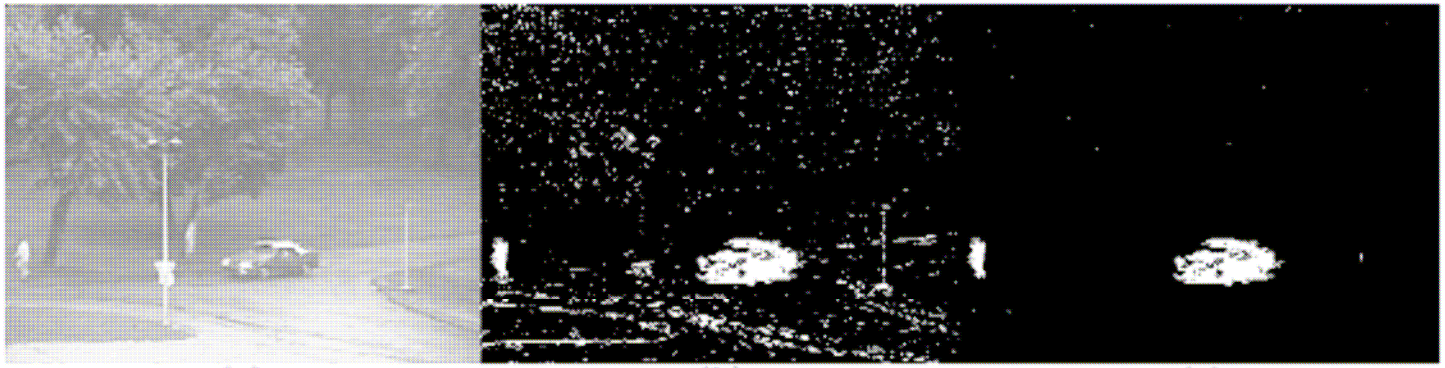


(a) (b)

Original image Estimated probability image

Kernel Density Estimators – (3)

Suppression of False Detection – by using neighborhood pixels



(a) (b) (c)

Kernel Density Estimators – (4)

Remove shadows – by color in chromaticity coordinates

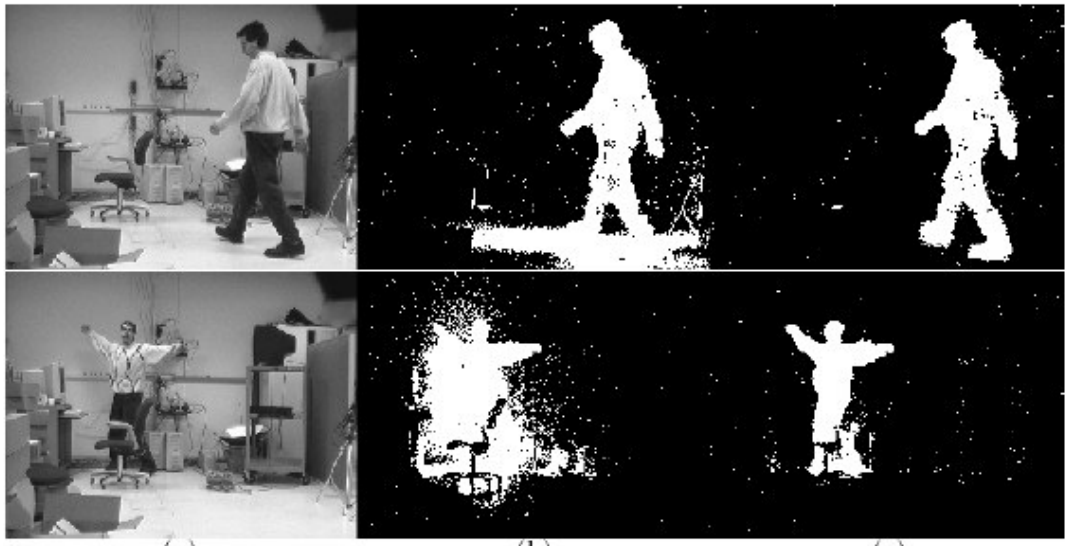


Fig. 8. (b) Detection using (R,G,B) color space (c) detection using chromaticity coordinates (r, g) and the lightness variable s

(b) RBG space, c) chromaticity coordinates

Kernel Density Estimators – (5)

- Histogram or kernel density based BGS – Non-parametric BGS, compare to parametric BGS (Gaussian) or semi-parametric BGS (mixture of Gaussians)
  - Very flexible for complex densities
  - Memory consuming

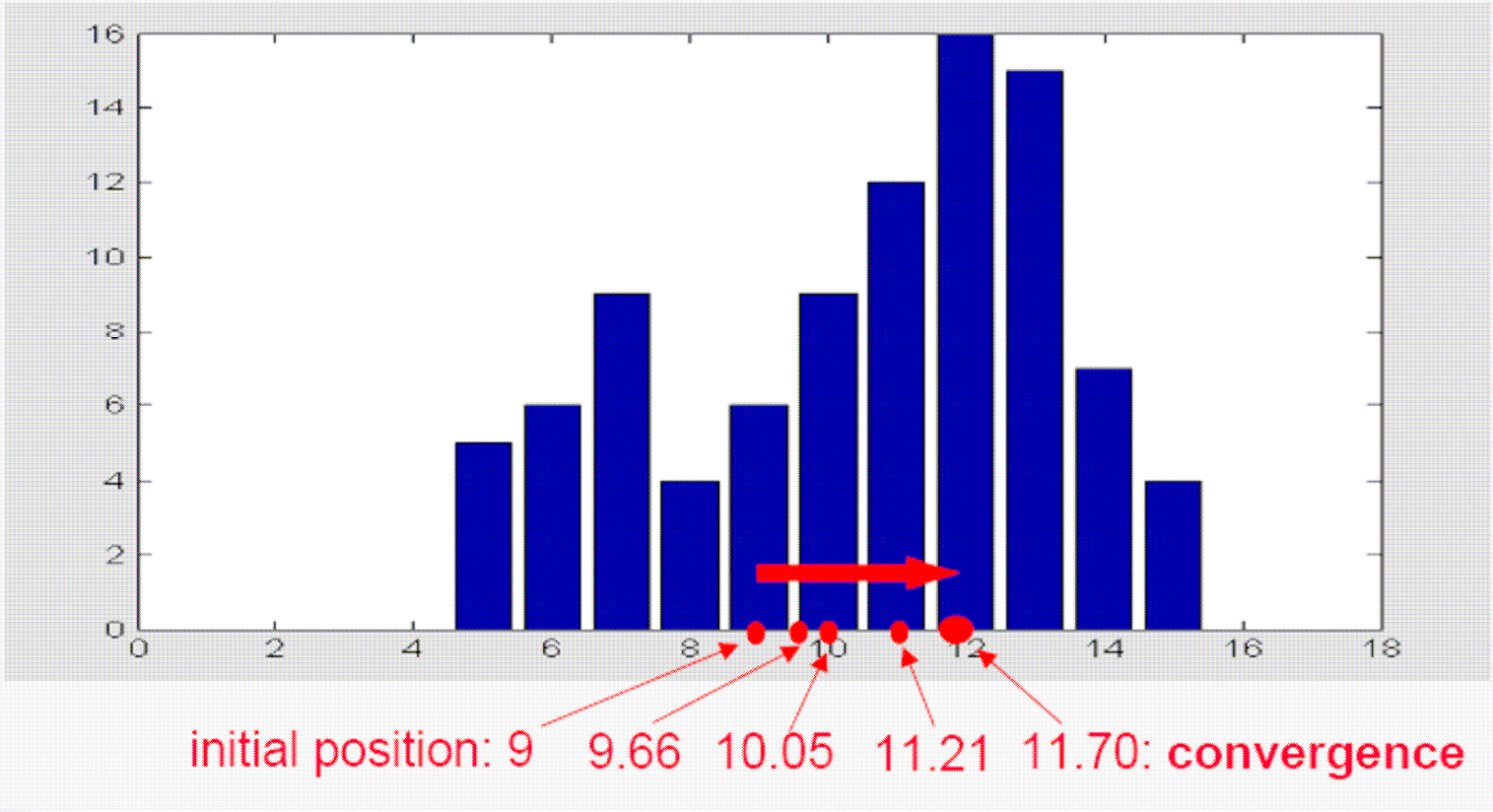
Mean-shift based estimation (1)

- Mean-shift based estimation (Han, Comaniciu, Davis, 2004; Piccardi, Jan, submitted 2004)
  - a gradient-ascent method able to detect the modes of a multimodal distribution together with their covariance matrix
  - iterative, the step decreases towards convergence
  - the mean shift vector:

$$m(x) = \frac{\sum_{i=1}^{n} x_i g((x - x_i / h)^2)}{\sum_{i=1}^{n} g((x - x_i / h)^2)} - x$$

Mean-shift based estimation (2)

- Example of mean-shift trajectory in the data space:



Mean-shift based estimation (3)

- Problems:
  - a standard implementation (iterative) is way too slow
  - memory requirements: *n * size(frame)*
- Solutions:
  - computational optimisations
  - using it only for detecting the background PDF modes at initialisation time; later, use something computationally lighter (mode propagation)

Combined estimation and propagation (1)

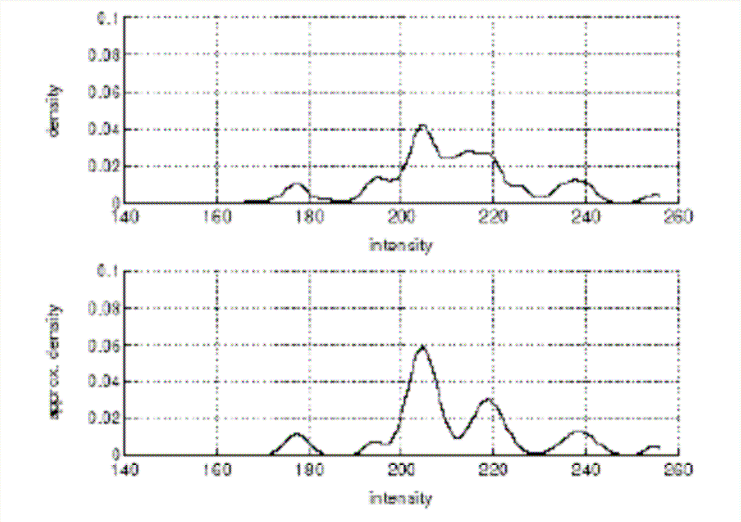
- Sequential Kernel Density Approximation (Han, Comaniciu, Davis, 2004)
  - mean-shift mode detection from samples is used only at initialisation time
  - later, modes are propagated by adapting them with the new samples:

$$PDF(x) = \alpha(new_mode) + (1-\alpha)(\sum existing_modes)$$

  - heuristic procedures are used for merging the existing modes (the number of modes is not fixed a priori)
  - faster than KDE, low memory requirements

Combined estimation and propagation (2)

- Example:
  - above: exact KDE
  - below: Sequential KD Approximation



(from: B. Han, D. Comaniciu, and L.. Davis, "Sequential kernel density approximation through mode propagation: applications to background modeling," Proc. ACCV 2004)

Eigenbackgrounds (1)

- Eigenbackgrounds (N. M. Oliver, B. Rosario, and A. P. Pentland, 2000)
  - Principal Component Analysis (PCA) by way of eigenvector decomposition is a way to reduce the dimensionality of a space
  - PCA can be applied to a sequence of *n* frames to compute the *eigenbackgrounds*
  - The authors state that it works well and is faster than a Mixture of Gaussians approach

Eigenbackgrounds – main steps (2)

1. The n frames are re-arranged as the columns of a matrix, A
2. The covariance matrix, $C = AA^T$, is computed
3. From C, the diagonal matrix of its eigenvalues, L, and the eigenvector matrix, Φ, are computed
4. Only the first M eigenvectors (eigenbackgrounds) are retained
5. Once a new image, I, is available, it is first projected in the M eigenvectors sub-space and then reconstructed as I'
6. The difference $I - I'$ is computed: since the sub-space well represents only the static parts of the scene, the outcome of this difference are the foreground objects

Summary (1)

- Basic BGS
- Running Gaussian average
- Mixture of Gaussians
- Enhanced mixture of Gaussians
- Kernel Density Estimators
- Mean-shift based estimation
- Combined estimation and propagation
- Eigenbackgrounds

Summary (2)

From the data available from the literature

- Speed
  - Fast: average, median, running average
  - Intermediate: Mixture of Gaussians, KDE, eigenbackgrounds, SKDA, optimised mean-shift
  - Slow: standard mean-shift
- Memory requirements
  - High: average, median, KDE, mean-shift
  - Intermediate: Mixture of Gaussians, eigenbackgrounds, SKDA
  - Low: running average

Summary (3)

- Accuracy
  - Impossible to say - an unbiased comparison with a significant benchmark is needed!
  - Some methods seem to be better principled: KDE, SKDA, mean-shift
  - Mixture of Gaussians and eigenbackgrounds certainly can offer good accuracy as well
  - Simple methods such as standard average, running average, median can provide acceptable accuracy in specific applications

Summary (4)

Real Video Surveillance Application:

- No single BGS method can handle all the cases
- Multiple BGS methods are needed
  - Indoor (relative stable lighting)
  - Outdoor
    - Relative stable
    - High dynamic
  - Crowded environment
- Other motion based methods

Homework #1

- Submit 1-page review for the following paper about Moving Object Detection
- Due before: Feb. 11, 2008
- Paper (I'll send out by email and put it in class website):
  - **A. Bugeau and P. Perez, "Detection and segmentation of moving objects in highly dynamic scenes" CVPR07.**

References (1)

* B.P.L. Lo and S.A. Velastin, “Automatic congestion detection system for underground platforms,” Proc. of 2001 Int. Symp. on Intell. Multimedia, Video and Speech Processing, pp. 158-161, 2000
* R. Cucchiara, C. Grana, M. Piccardi, and A. Prati, “Detecting moving objects, ghosts and shadows in video streams”, IEEE Trans. on Patt. Anal. and Machine Intell., vol. 25, no. 10, Oct. 2003, pp. 1337-1342
* D. Koller, J. Weber, T. Huang, J. Malik, G. Ogasawara, B. Rao, and S. Russel, “Towards Robust Automatic Traffic Scene Analysis in Real-Time,” in Proceedings of Int’l Conference on Pattern Recognition, 1994, pp. 126–131.
* C. Wren, A. Azarbayejani, T. Darrell, and A. Pentland, “Pfinder:Real-time Tracking of the Human Body,” IEEE Trans. on Patt. Anal. and Machine Intell., vol. 19, no. 7, pp. 780-785, 1997
* C. Stauffer, W.E.L. Grimson, “Adaptive background mixture modelsfor real-time tracking”, Proc. of CVPR 1999, pp. 246-252.
* C. Stauffer, W.E.L. Grimson, “Learning patterns of activity using real-ime tracking”, IEEE Trans. on Patt. Anal. and Machine Intell., vol. 22, no. 8, pp. 747-757, 2000.
* Elgammal, A., Harwood, D., and Davis, L.S., “Non-parametric Model for Background Subtraction”, Proc. of ICCV '99 FRAME-RATE Workshop, 1999.

References (2)

* B. Han, D. Comaniciu, and L. Davis, "Sequential kernel density approximation through mode propagation: applications to background modeling,“ Proc. ACCV -Asian Conf. on Computer Vision, 2004. Running Gaussian average
* N. M. Oliver, B. Rosario, and A. P. Pentland, “A Bayesian Computer Vision System for Modeling Human Interactions,” IEEE Trans. on Patt. Anal. and Machine Intell., vol. 22, no. 8, pp. 831-843, 2000.
* M. Seki, T. Wada, H. Fujiwara, K. Sumi, “Background detection based on the cooccurrence of image variations”, Proc. of CVPR 2003, vol. 2, pp. 65-72.
* Ying-Li Tian, Max Lu, and Arun Hampapur, “Robust and Efficient Foreground Analysis for Real-time Video Surveillance,” IEEE CVPR, San Diego. June, 2005.
* J. Connell, A.W. Senior, A. Hampapur, Y.-L. Tian, L. Brown, S. Pankanti, “Detection and Tracking in the IBM PeopleVision System,” IEEE ICME Taiwan June, 2004
* O. Javed, K. Shafique, and M. Shah, “A Hierarchical Approach to Robust Background Subtraction using Color and Gradient Information,” IEEE Workshop on Motion and Video Computing, 2002.