Background subtraction techniques: a review

Massimo Piccardi

Computer Vision Research Group (CVRG)
University of Technology, Sydney (UTS)
e-mail: massimo@it.uts.edu

The ARC Centre of Excellence for Autonomous Systems (CAS)

Faculty of Engineering, UTS, April 15, 2004

Agenda

- The problem
- The basic methods
- Running Gaussian average
- Mixture of Gaussians
- Kernel Density Estimators
- Mean-shift based estimation
- Combined estimation and propagation
- Eigenbackgrounds

The problem

- Main goal: given a frame sequence from a fixed camera, detecting all the foreground objects
- Naive description of the approach: detecting the foreground objects as the difference between the current frame *and an image of the scene's static background*:

$$|frame_i - background_i| > Th$$

- First consequent problem: how to automatically obtain the image of the scene's static background?

The problem - requirements

The background image is not fixed but must adapt to:

- Illumination changes
  - gradual
  - sudden (such as clouds)
- Motion changes
  - camera oscillations
  - high-frequencies background objects (such as tree branches, sea waves, and similar)
- Changes in the background geometry
  - parked cars, ...

The basic methods

Frame difference:

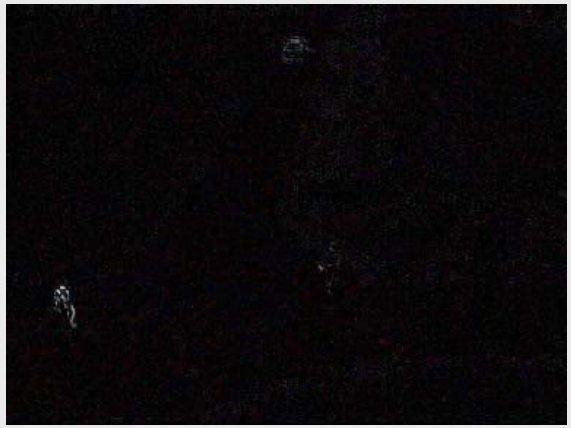
$$| frame_i - frame_{i-1} | > Th$$

- The estimated background is just the previous frame
- It evidently works only in particular conditions of objects' speed and frame rate
- Very sensitive to the threshold Th

The basic methods (2)

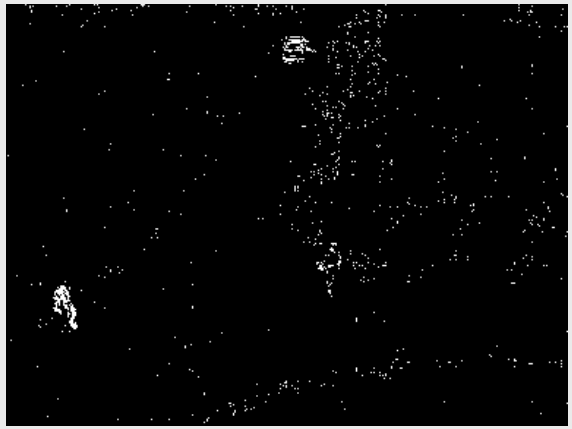
Frame difference: an example

the frame



absolute difference

threshold: too high



threshold: too low

The basic methods (3)

- Background as the ***average*** or the ***median*** (Velastin, 2000; Cucchiara, 2003) of the previous *n* frames:
  - rather fast, but very memory consuming: the memory requirement is *n * size(frame)*
- Background as the ***running average***:

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

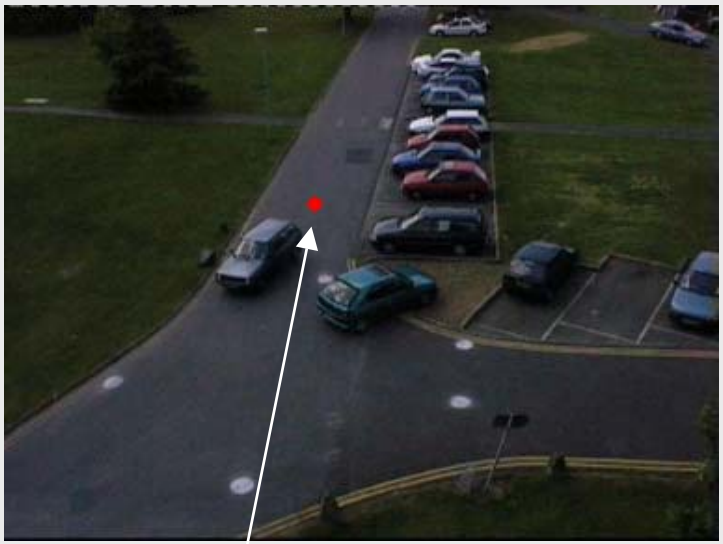
  - α, the *learning rate*, is typically 0.05
  - no more memory requirements

The basic methods – rationale

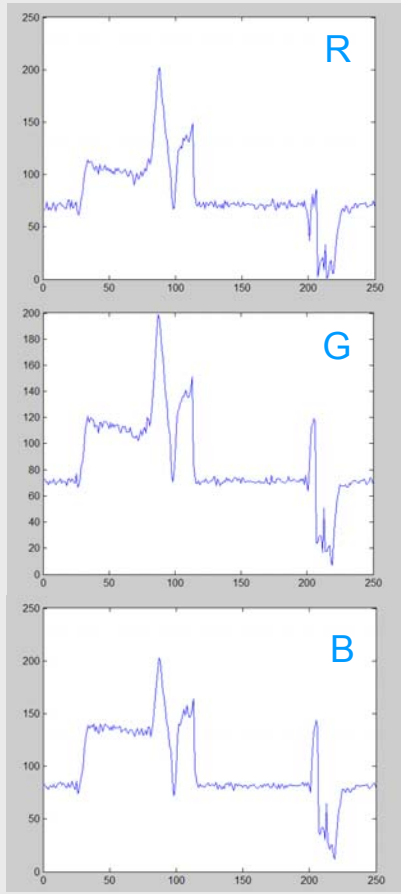
- The background model at each pixel location **is based on the pixel's recent history**
- In many works, such history is:
  - just the previous *n* frames
  - a weighted average where recent frames have higher weight
- In essence, the background model is computed as a chronological average from the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

The basic methods - histograms

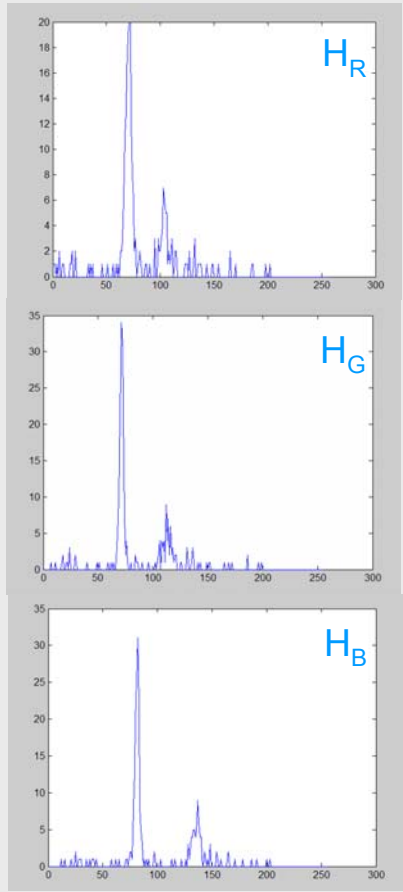
- Example:



pixel location



sequence of pixel values



histograms

The basic methods - selectivity

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

  $\rightarrow$ 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

The basic methods – selectivity (2)

- Running average with selectivity:

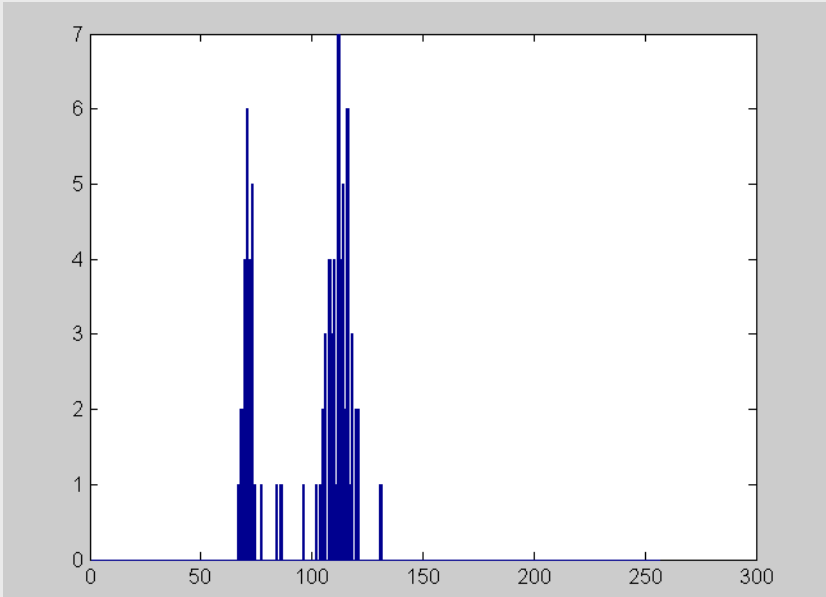
$$B_{i+1}(x,y) = \alpha F_t(x,y) + (1-\alpha) B_t(x,y)$$ if $F_t(x,y)$ background

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

- Similarly for other methods

The basic methods - 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:

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

$$\sigma^2_{t+1} = \alpha(F_t - \mu_t)^2 + (1-\alpha)\sigma^2_t$$

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

Mixture of Gaussians

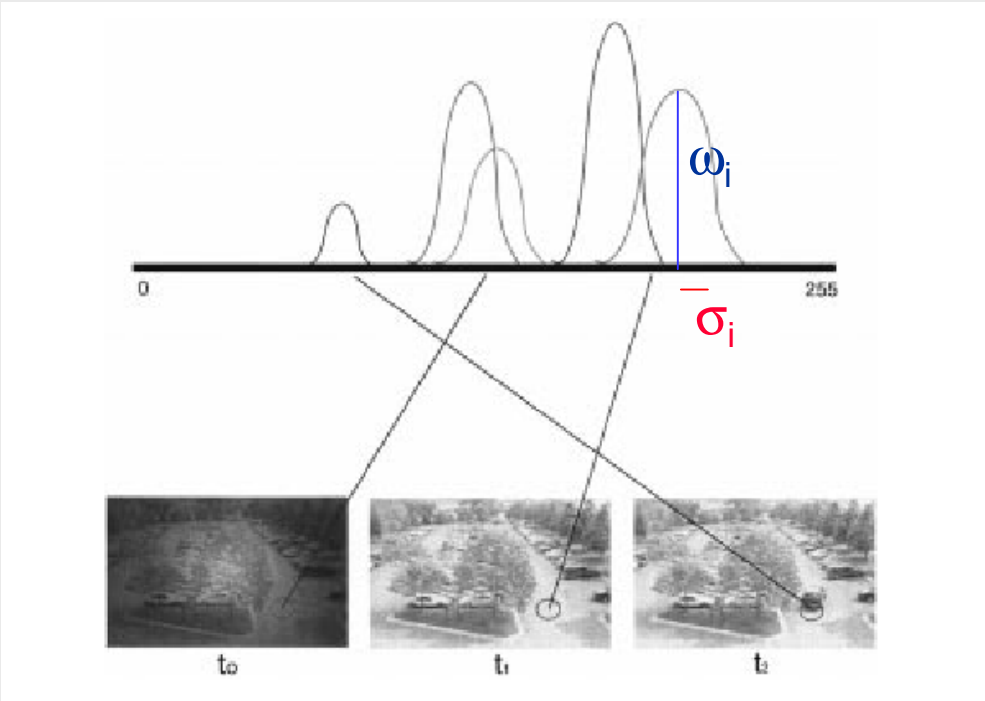
- *Mixture of K Gaussians* ($\mu_i, \sigma_i, \omega_i$) (Stauffer and Grimson, 1999)
- In this way, the model copes also with multimodal background distributions; however:
  - the number of modes is arbitrarily pre-defined (usually from 3 to 5)
  - how to initialize the Gaussians?
  - how to update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

Kernel Density Estimators

- Kernel Density Estimators (Elgammal, Harwood, Davis, 2000):
- The background PDF is given by the histogram of the *n* most recent pixel values, each smoothed with a Gaussian kernel (sample-point density estimator)
- If *PDF(x)* > *Th*, the *x* pixel is classified as background
- Selectivity
- Problems: memory requirement (*n* * *size(frame)*), time to compute the kernel values (mitigated by a LUT approach)

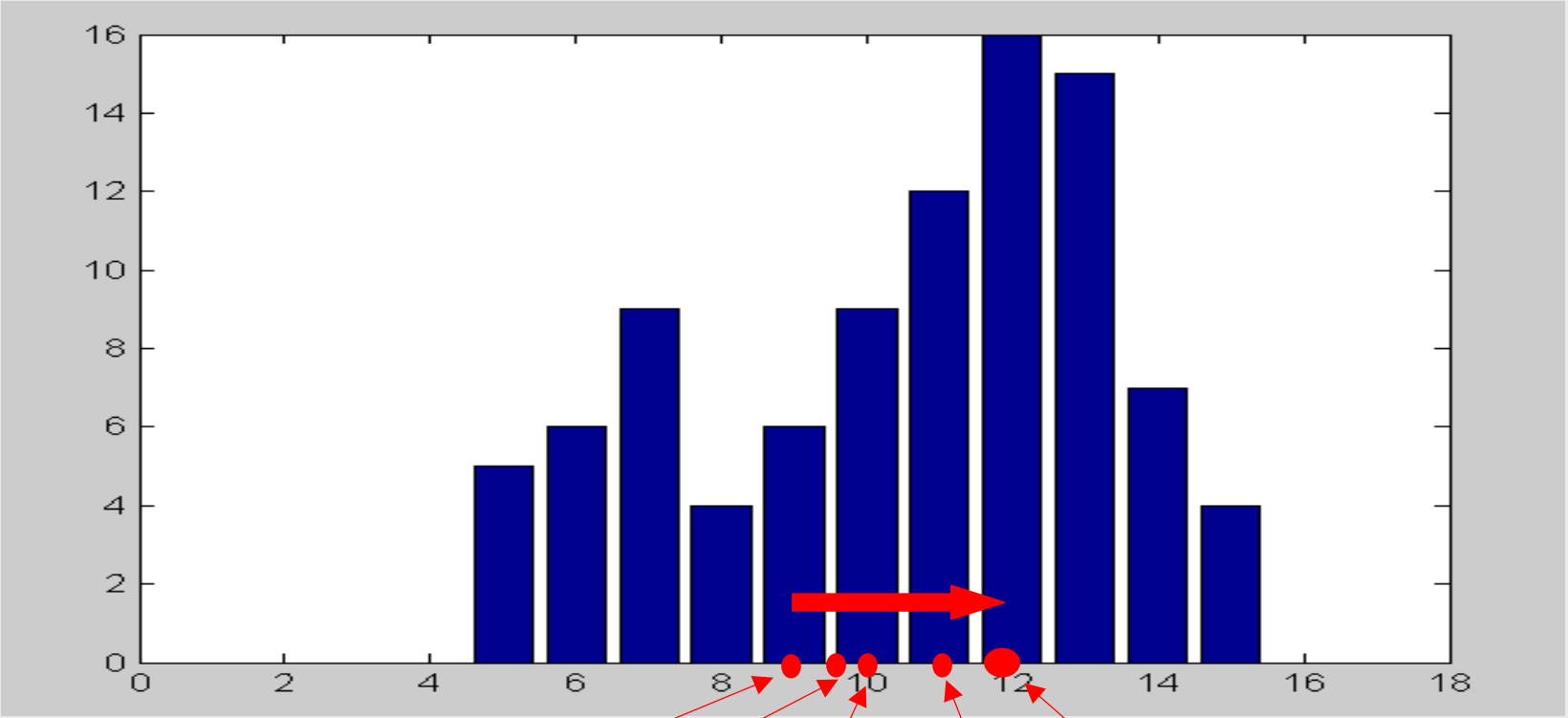
Mean-shift based estimation

- 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

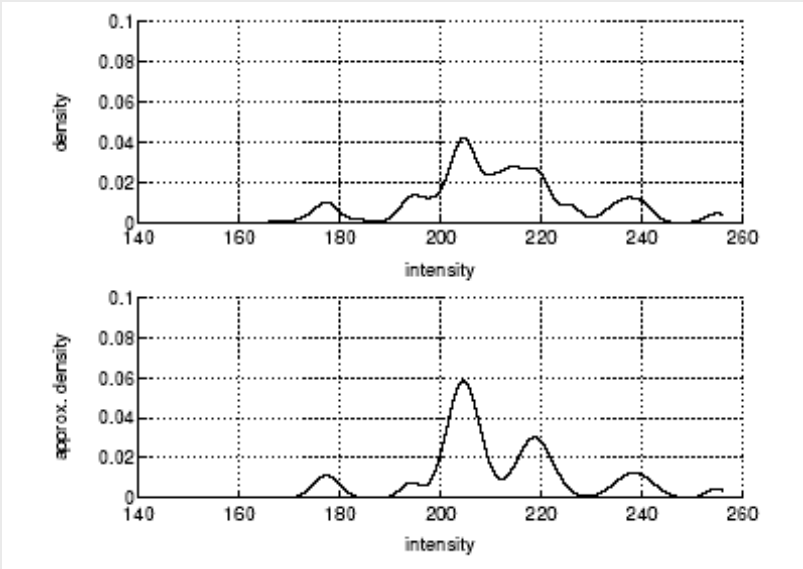
- Sequential Kernel Density Approximation (Han, Comaniciu, Davis, 2004)
  - mean-shift mode detection from samples is used only at initialisation time
  - later, modes are propagated by adapting them with the new samples:

$$PDF(x) = \alpha(new\_mode) + (1-\alpha)(\sum existing\_modes)$$

  - 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

- 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

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

Spatial correlation?

- It can be immediately evident that there exists spatial correlation between neighboring pixels. How can that be exploited?
- Low-end approach: binary morphology applied to the resulting foreground image
- Better principled: at the PDF level (for instance: Elgammal, Harwood, Davis, 2000)
- In the eigenbackground approach, the correlation matrix
- An approach formally exploiting spatial correlation: background detection based on the cooccurrence of image variations (Seki, Wada, Fujiwara, Sumi, CVPR 2003)

Summary

Methods reviewed:

- Average, median, running average
- Mixture of Gaussians
- Kernel Density Estimators
- Mean shift (possibly optimised)
- SKDA (Sequential KD Approximation)
- 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

Main references

- 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 models for 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.

Main 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.
- 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.