Video Segmentation for Surveillance

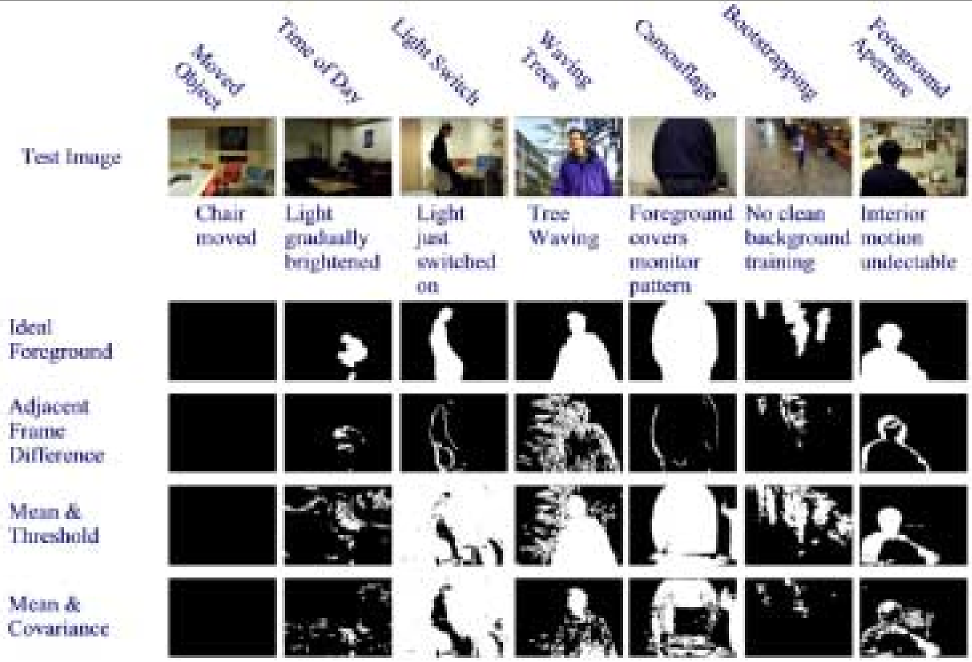
-- A Transform Domain Approach

Juhua Zhu
Dept. of Electrical Engineering
Princeton University

April 23, 2005

Problem Formulation

- Goal: To segment moving objects in video
- Static camera
- Change detection
  - What to compare to?
  - What change is of interest?
- Challenges



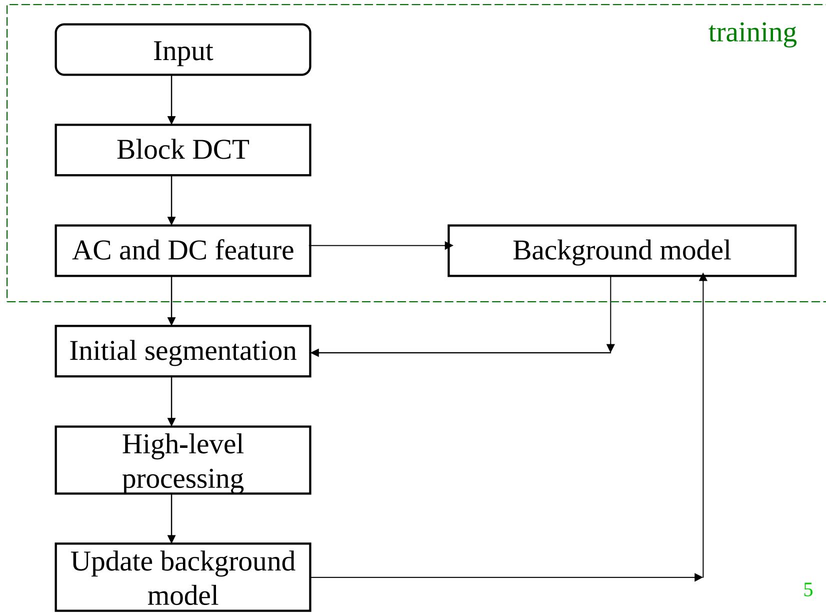
from [K. Toyama 99] -to be cont'd

3

from [K. Toyama 99]

4

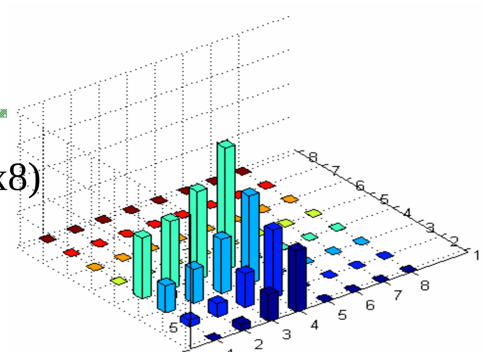
Overview



5

Features

- Block-based DCT(e.g.8x8)
- DC and AC features
  - $f^{DC} = DCT(0,0)$
  - $f^{AC} = \sum_{i=0}^{N/2-1} \sum_{j=0}^{N/2-1} (i^2 + j^2) \cdot |DCT(i,j)|$



- Local information is encoded, intensity and texture.
- Features insensitive to noise, small scene changes, and light shadows
- Foreground objects usually lead to significant changes in AC and DC features

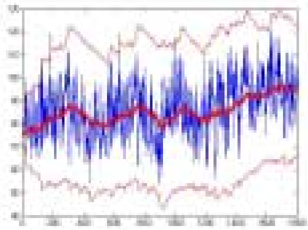
6

Background Blocks

- Features change slowly from frame to frame
- Single Exponential Smoothing
- Predict $\hat{f}_{t+1}$ from $\hat{f}_t$ and f_t

$$\hat{f}_{t+1}^{AC} = \alpha^{AC} \cdot f_t^{AC} + (1-\alpha^{AC})\hat{f}_t^{AC}$$

$$\hat{f}_{t+1}^{DC} = \alpha^{DC} \cdot f_t^{DC} + (1-\alpha^{DC})\hat{f}_t^{DC}$$



- For background blocks $\hat{f}_t$ close to f_t
- Deviation of $\hat{f}_t$ from f_t has mean close to 0 and variance σ^2
- $\sigma_t^2 = \alpha(f_t - \hat{f}_t)^2 + (1-\alpha)\sigma_{t-1}^2$

7

Foreground Blocks

- Features significantly differ from prediction
- Initial Segmentation
  - Declare a block as foreground if

$$| f_t^{DC} - \hat{f}_t^{DC} | \geq \kappa \cdot \sigma_{t-1}^{DC} + \lambda \cdot \hat{f}_t^{DC}$$

or $$| f_t^{AC} - \hat{f}_t^{AC} | \geq \kappa \cdot \sigma_{t-1}^{AC} + \lambda \cdot \hat{f}_t^{AC}$$

2nd term in thresholds due to small # of training frames

- If foreground, do not predict and update σ^2

8

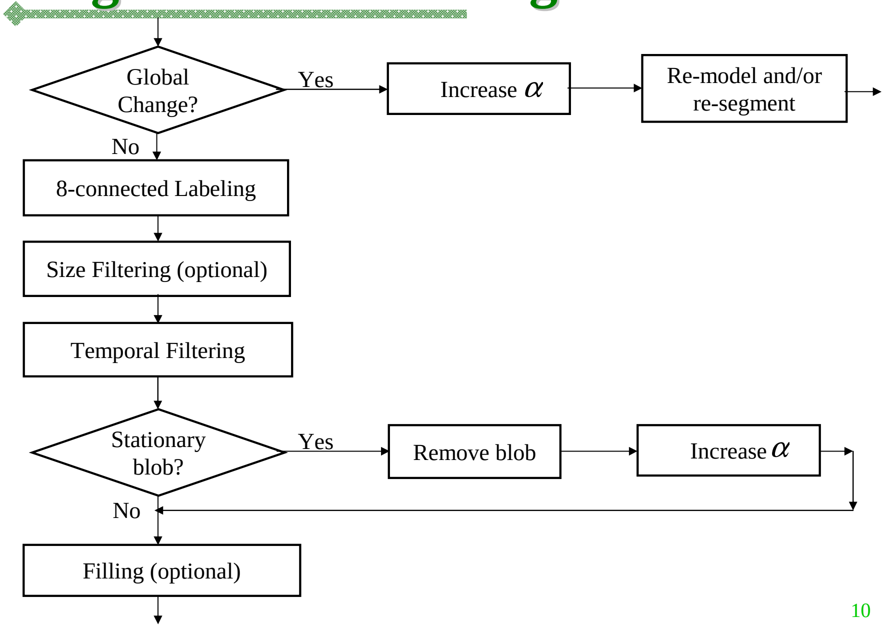
Selection of α

- Initialization
  - Minimize Sum of Squared Error
- Update
  - Based on classification confidence and frame-wise correlation

$$\alpha_t = \begin{cases} 0 & \text{foreground block} \\ \alpha_{\min} & \text{background block} \\ \max(\alpha_{\min}, (0.5)^n \cdot \alpha_{\max}) & \text{fast update needed} \end{cases}$$

9

High-level Processing

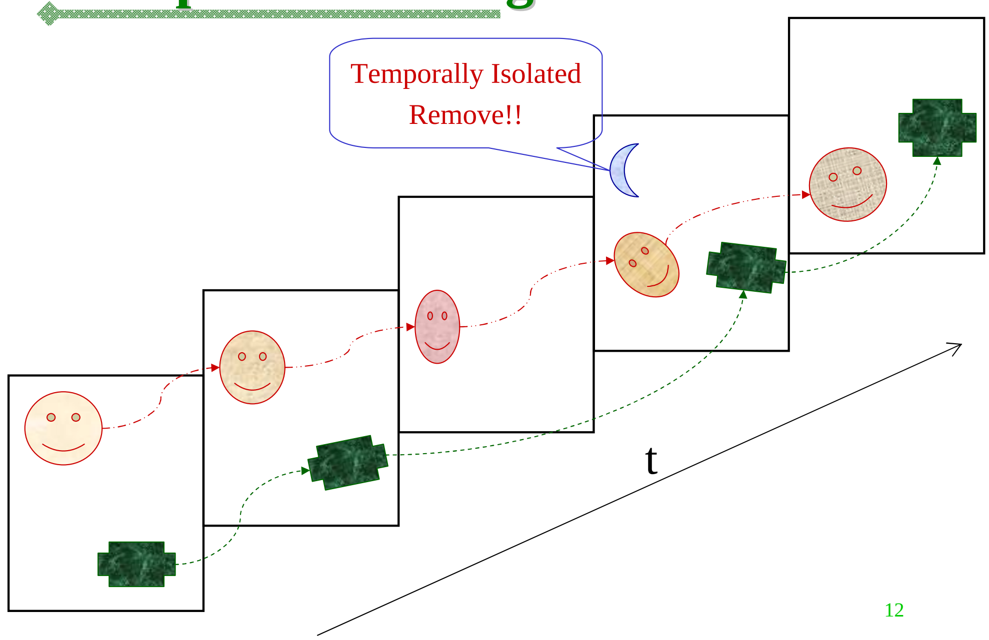


10

Form blobs

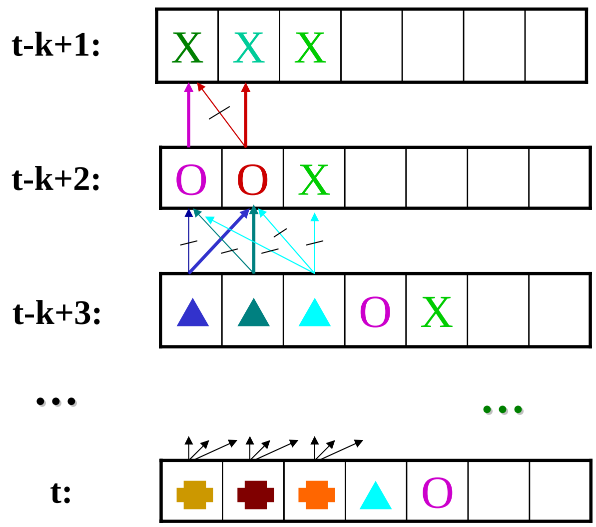
- Group eight-connected foreground blocks as blob
- Tag all blobs in each frame
- Match each blob with blobs in previous *k* frames using proximity
- Link matched blobs from frame to frame
- Apply temporal filtering to remove temporally isolated blobs

Temporal Filtering



Blob Matching

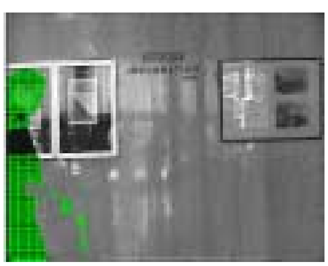
- Keep the blob or not? – continuous motion



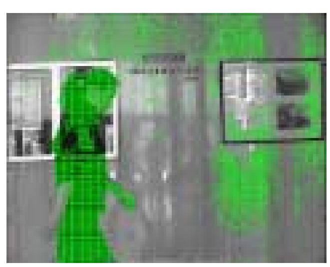
13

13

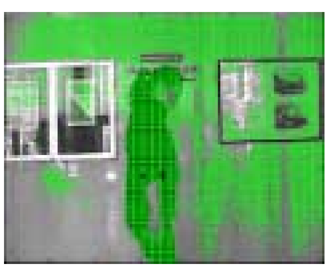
Compare with Pixel-wise MoG



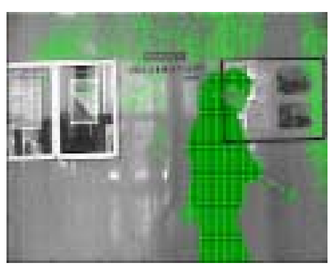
Frame 31



Frame 33



Frame 35



Frame 37

14

Experimental results

A. Strong moving reflections

B. Severe global/local illumination change, strong lighting, mirror-effect

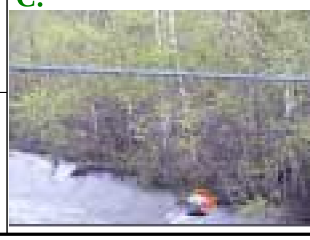
C. Flowing water, swaying branches, occlusions, small objects

D. Waving water, strong reflections, small objects, camouflage

A.

B.

C.



D.

15

Comparison

PROPOSED METHOD	CLASSIC MOG
Texture and intensity	Color information
Block-based DCT	Pixel-wise
Grayscale	RGB or YUV
Modeled by Single Gaussian	Modeled by Mixture of Gaussians
Fast adaptation to changes	Slow adaptation
Robust to noise and small scene changes	Sensitive to noise and small scene changes

16

Summary

- Fast (~40 fps)
- Robust to noise, small scene changes, and illumination changes
- Can handle
  - Moved background objects
  - Foreground aperture
  - Bootstrapping
  - Shadows

Q & A