CSCI 435/MCS9435
Computer Vision

Dynamic Vision (I)

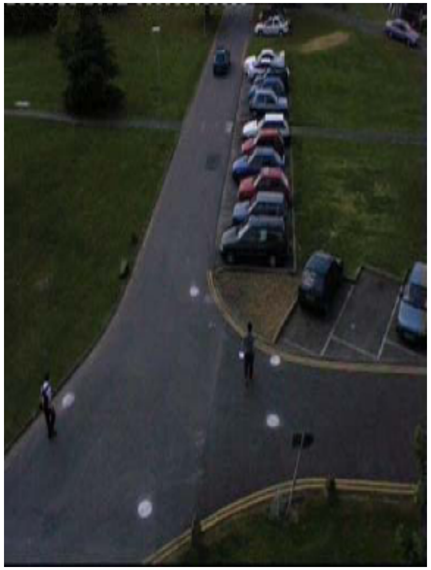
Lecturer: Wanqing Li, PhD
Room: 3.101
email: wanqing@uow.edu.au
Web: http://www.uow.edu.au/~wanqing

Outline

- Static vision vs Dynamic Vision
  - Image vs image sequence
  - Key problems in dynamic vision
- Change Detection
  - Frame difference
- Background Modeling
  - Running average / median
  - Gaussian Mixture Model
- Connected Component Analysis

The need for dynamic vision (DV)

- Dynamic vision provides a way to monitor the scene constantly or at certain time interval
  - Robot vision (autonomous system)
  - Video surveillance
  - Satellite imaging of land cover
  - Motion picture



DV settings

- A dynamic vision captures and analyses
  - A sequence of images
- Four different settings
  - SCSO - Stationary camera, stationary objects
  - SCMO - Stationary camera, moving objects
  - MCSO - Moving camera, stationary objects
  - MCMO - Moving camera, moving objects

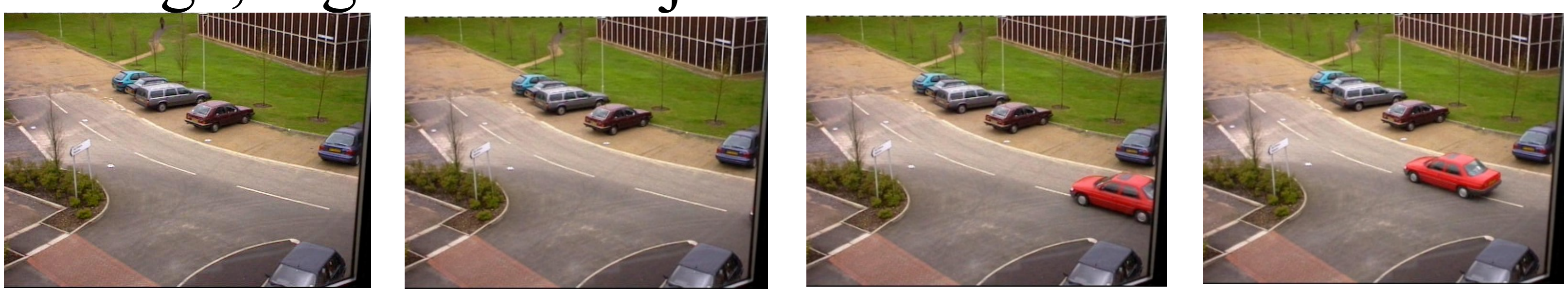
Key problems in DV

- Change detection
  - Most time SCMO
  - Apps- Video surveillance
- Moving object detection and location
  - SCMO or MCSO or MCMO
  - Apps- Cloud tracking, autonomous vehicles, city traffic analysis, target tracking and positioning (military)
- Recovery of 3D object properties
  - MCSO or MCMO

Change Detection

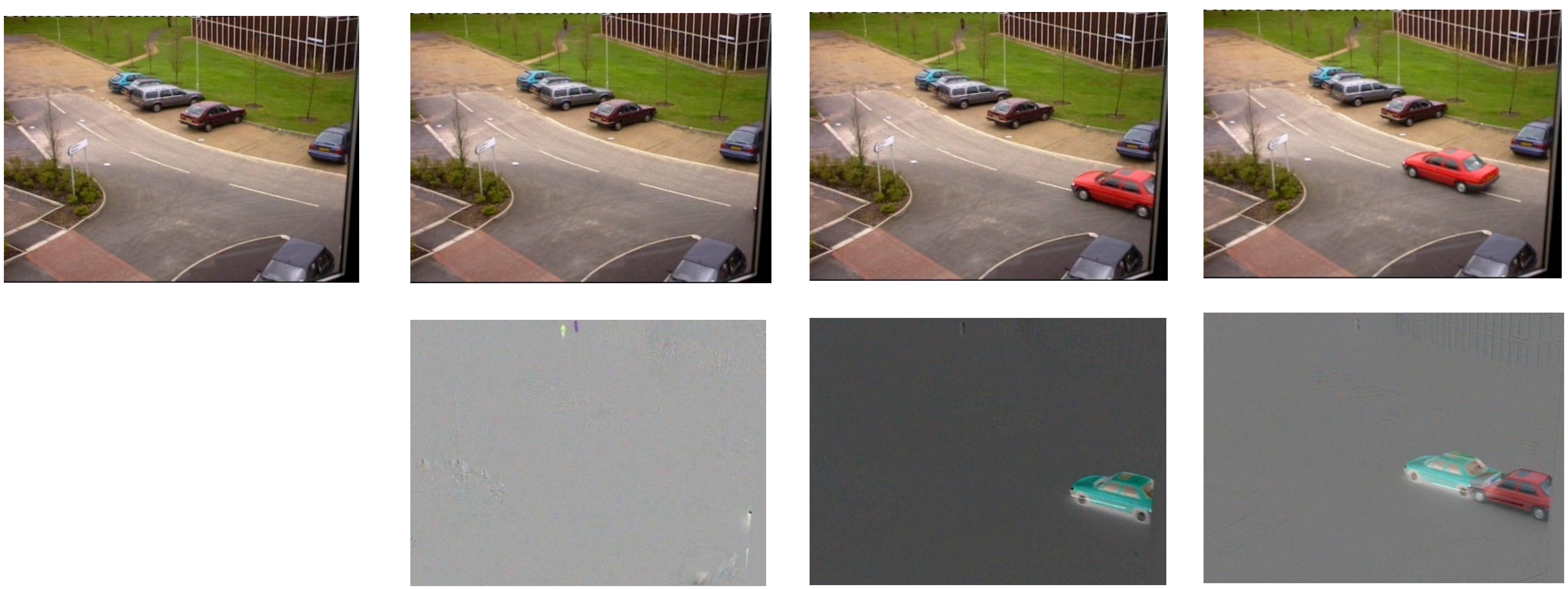
Change Detection (CD)

- Goal
  - Detecting and localizing any "meaningful" changes from a sequence of images taken from a scene
  - Changes can be detected at different levels: pixel, edge, region and object



Frame Difference

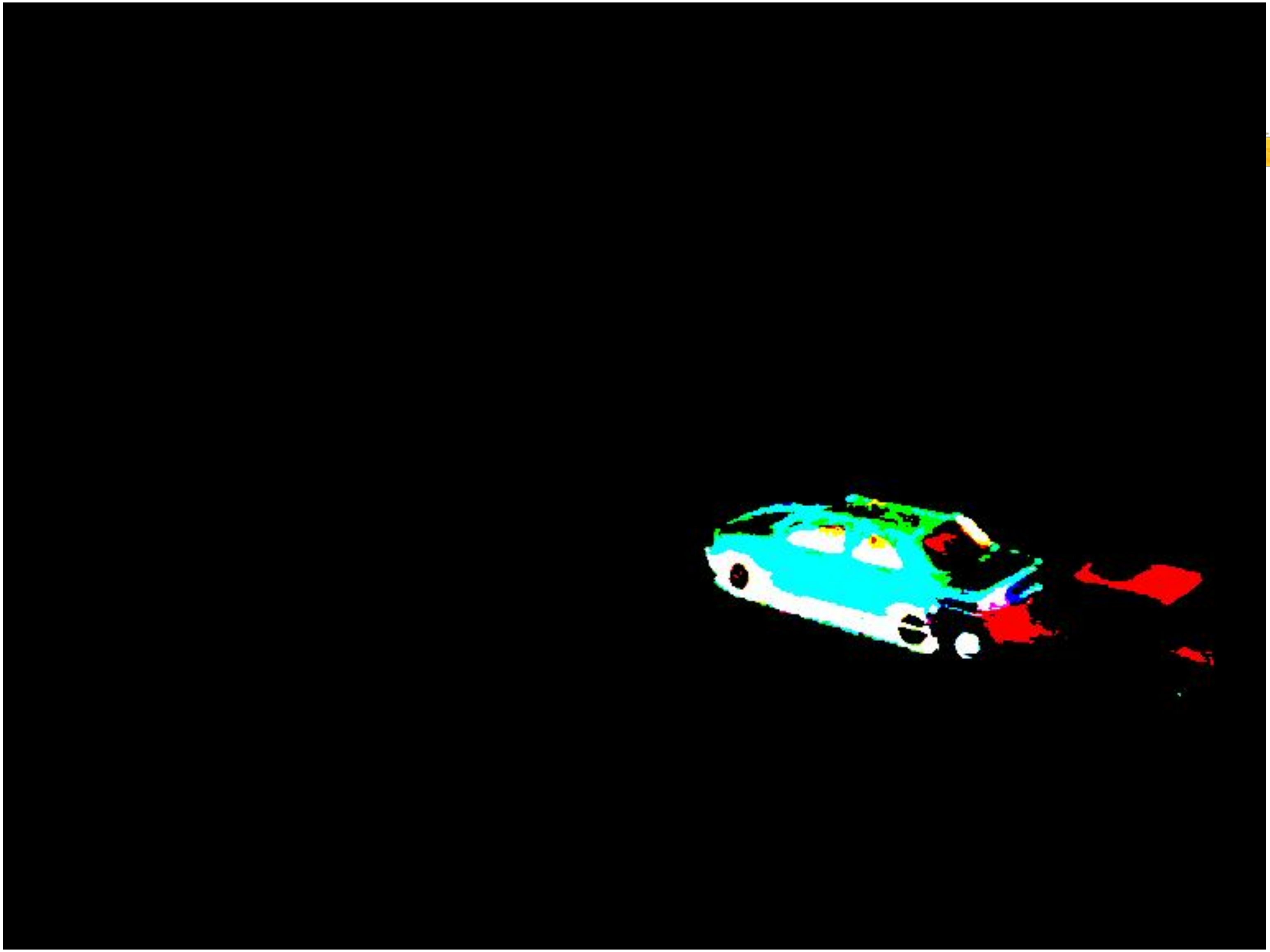
- Simplest approach is to calculate the difference between two consecutive frames



Frame difference…

$$DF_{jk}(x,y) = \begin{cases} 1 & if \left| f(x,y,j) - f(x,y,k) \right| > \tau \\ 0 & otherwise \end{cases}$$

How to choose the threshold ?



Frame difference: an example

Post-processing

- The binarized difference frame can be very noisy
- Post processing that removes the noise is often required
  - Majority voting
  - Morphological operations
  - Size filter
    - ✓ Remove pixels that do not belong to a connected cluster of a minimum size
    - ✓ Connected component analysis

Issues with frame difference

- Sensitive to:
  - noise
  - illumination changes
  - nonsense changes; e.g. movement of tree leaves
  - camera movement
- Trouble to detect slow changes
  - E.g. slow motion

Robust CD

- To make the CD more robust
  - Intensity/color characteristics of regions or group of pixels at the same location in two frames may be compared using either
    - ✓ A statistical approach, or
    - ✓ An approach based on the local approximation of intensity distribution
  - At cost of added computation

Likelihood Ratio based CD

Likelihood Ratio Test (LRT) based CD…

$$\lambda = \frac{\left[\frac{\sigma_1^2 + \sigma_2^2}{2} + \left(\frac{\mu_1 - \mu_2}{2}\right)^2\right]^2}{\sigma_1^2 * \sigma_2^2}$$

$$DF_{jk}(x, y) = \begin{cases} 1 & \lambda > \tau \\ 0 & otherwise \end{cases}$$

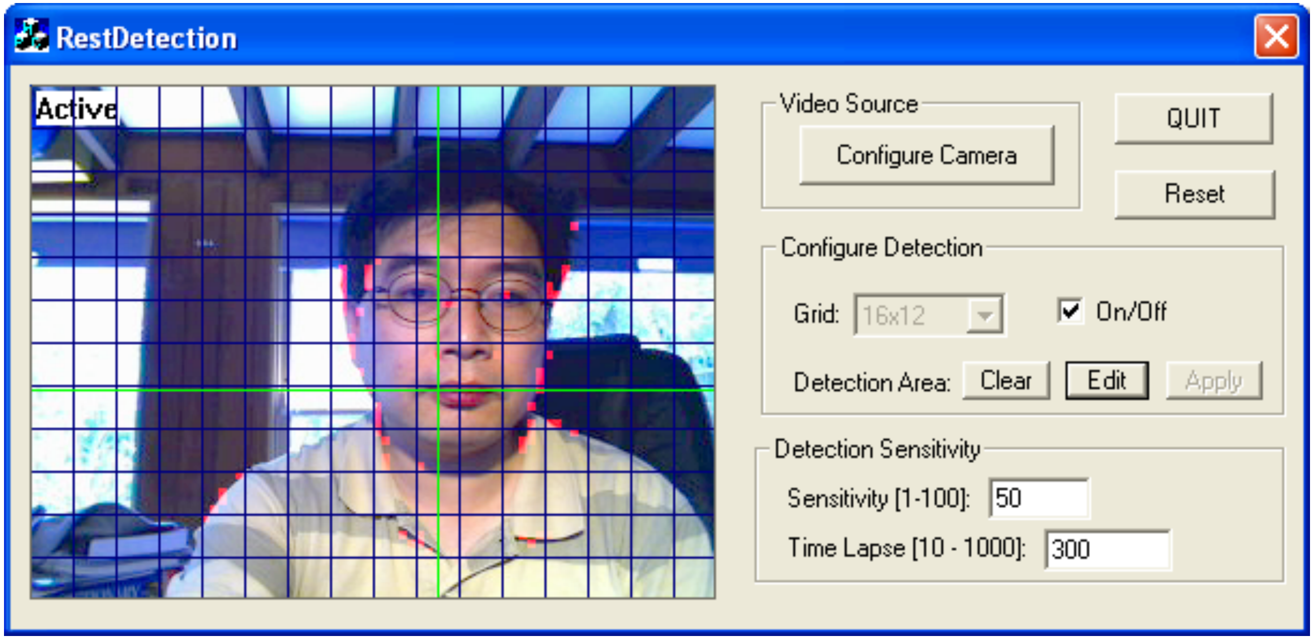
- LRT based CD works well for real world scenes when it is combined with a size filter

Accumulative Frame Difference

- Small or slow motion will fail
  - frame difference & Robust CD
- Analyze the changes over a sequence of frames
  - Accumulative frame difference

$$df_0(x, y) = 0$$
$$df_t(x, y) = \alpha \cdot df_{t-1}(x, y) + (1 - \alpha) \cdot | f_t(x, y) - f_{t-1}(x, y) |$$

An Application



How it works

Background Modeling & Subtraction

The Problem

- Main goal:
  - given a frame sequence from a fixed camera, detect all the foreground objects (SCMO)
- 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:

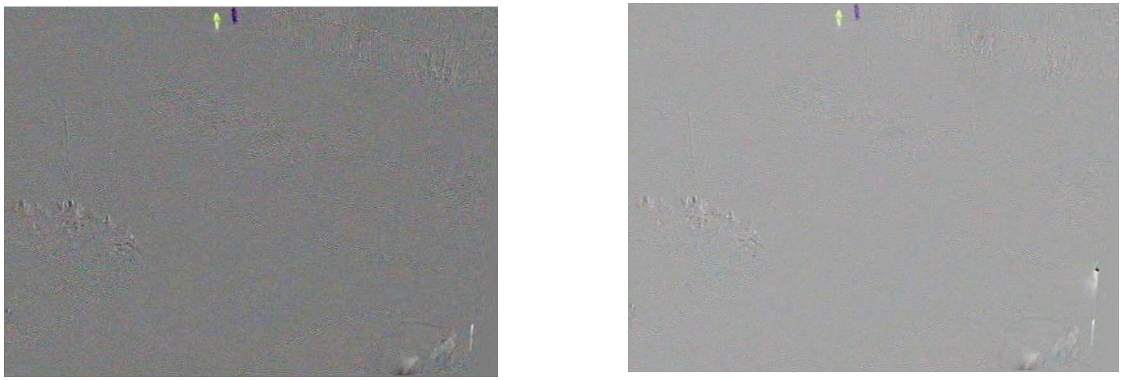
$$| \, frame_i - background_i \, | > Th$$

- 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, ...

Changing Background



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 (neighboring) pixel locations

Basic Methods - Average / Median

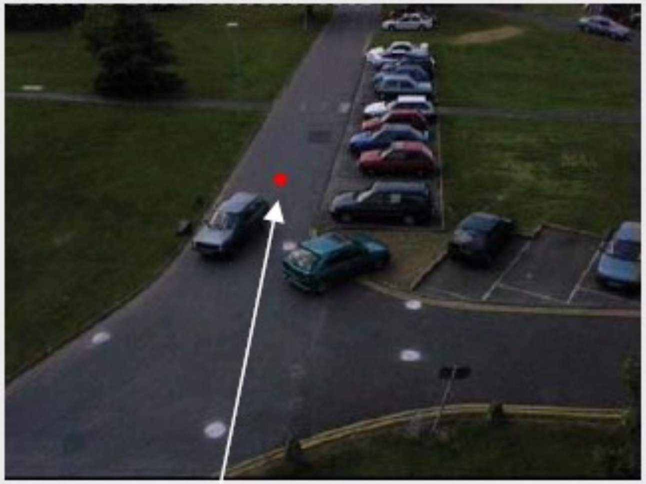
- Background as the **average** or the **median** (Velastin, 2000; Cucchiara, 2003)of the previous *n* frames
  - rather fast, but very memory consuming:
    - ✓the memory requirement is *n*size(frame)*
- Background as the **running average**:

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

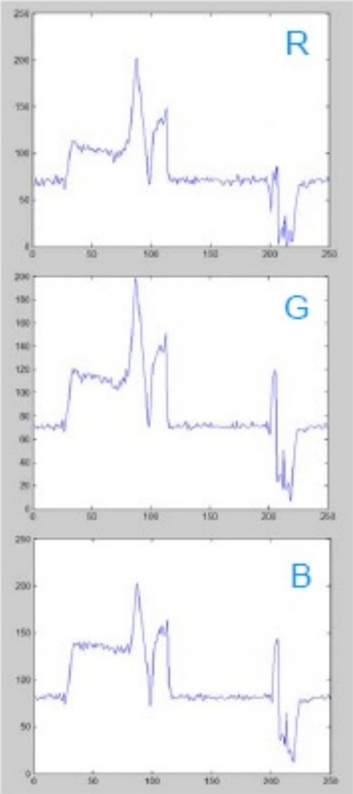
  - α, the learning rate, is typically 0.05
  - no more memory requirements

Basic Methods - Histogram

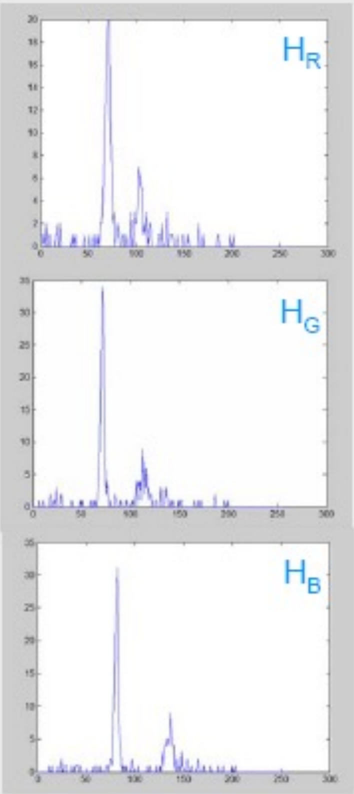
- Example:



pixel location



sequence of pixel values



histograms

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?
  - →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

Basic Methods - Selectivity

- Running average with selectivity:

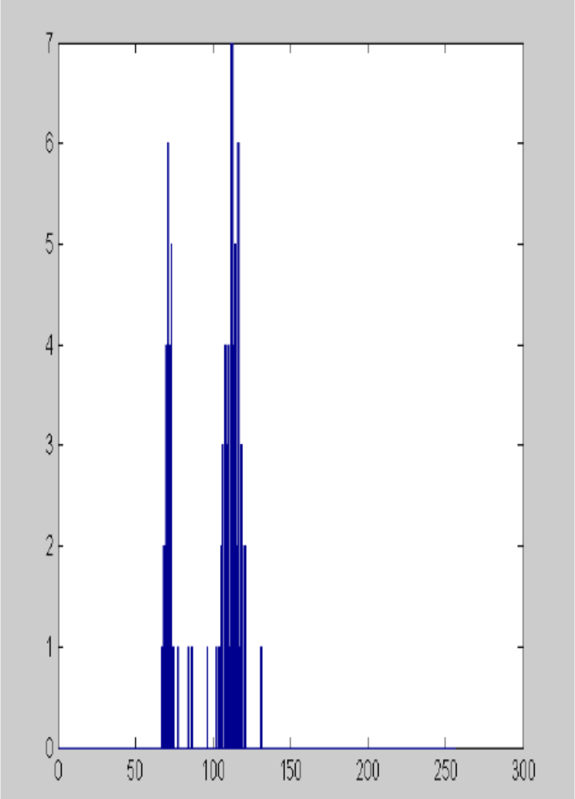
$$B_{i+1}(x,y) = \alpha F_t(x,y) + (1-\alpha) B_t(x,y) \quad \text{if } F_t(x,y) \text{ background}$$

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

- Similarly for other methods.

Basic Methods - Limitations

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



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 -μ| > Th, Th can be chosen as kσ
- It does not cope with multimodal backgrounds

Mixture Gaussians (Stauffer & Grimson'99)

- The probability of observing the current pixel value is

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

  - Gaussian probability density function

$$\eta(X_t, \mu, \Sigma) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(X_t-\mu_t)^T \Sigma^{-1} (X_t-\mu_t)}$$

- Every new pixel value, *Xt*, is checked against the existing K Gaussian distributions
  - A match is defined as a pixel value within 2.5 standard deviations of a Gaussian.

Mixture Gaussians…

- The parameters of the distribution which matches the new observation are updated as follows

$$\mu_t = (1-\rho)\mu_{t-1} + \rho X_t$$

$$\sigma_t^2 = (1-\rho)\sigma_{t-1}^2 + \rho(X_t - \mu_t)^T(X_t - \mu_t)$$

$$\rho = \alpha\eta(X_t|\mu_k, \sigma_k)$$

$$\omega_{k,t} = (1-\alpha)\omega_{k,t-1} + \alpha(M_{k,t})$$

- where $M_{k,t}$ is 1 for matched Gaussian and zero otherwise

Mixture Gaussians…

- **Background Model Estimation**
  - Consider the accumulation of supporting evidence and the relatively low variance for the "background" distributions
  - New object occludes the background object
    - ➔ Increase in the variance of an existing distribution.
  - First, the Gaussians are ordered by the value of ω/σ.

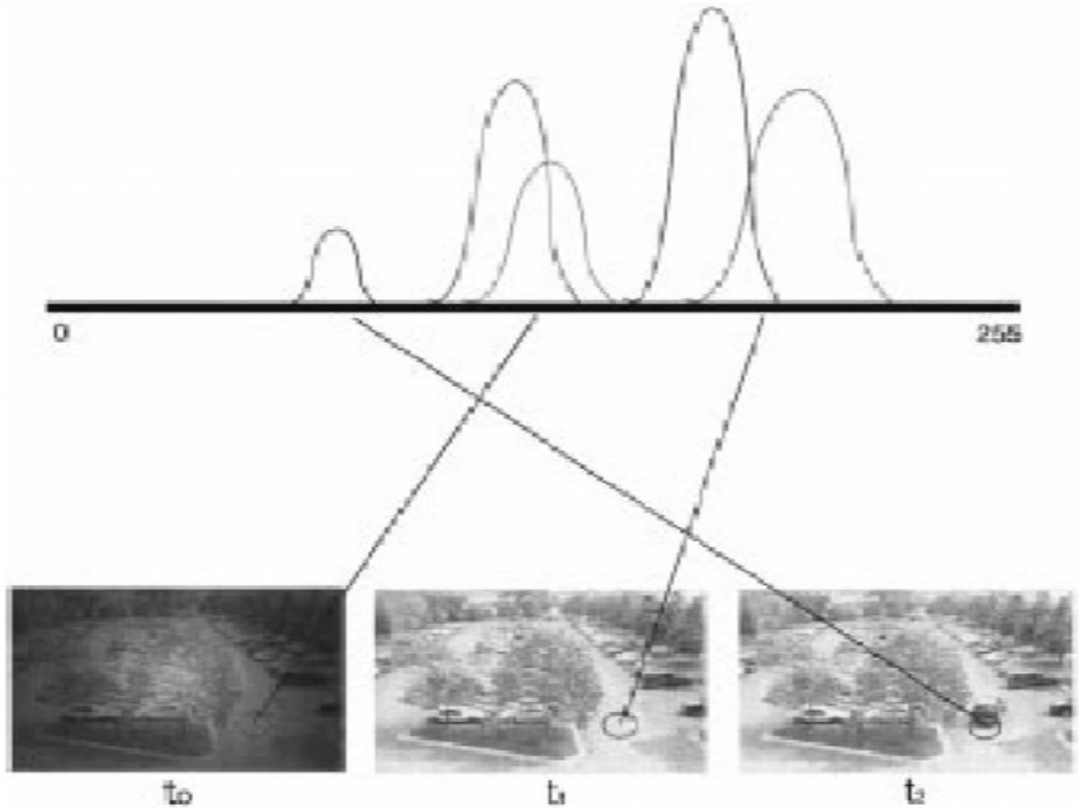
Background Model Estimation

- First, the Gaussians are ordered by the value of ω/σ.
- Then, the first B distributions are chosen as the background model

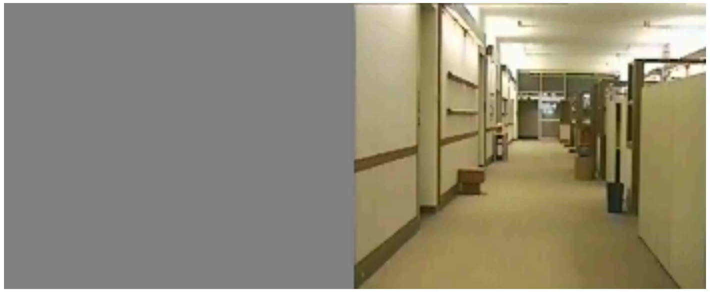
$$B = argmin_b \left(\sum_{k=1}^{b} \omega_k > T \right)$$

- T is a measure of the minimum portion of the data that should be accounted for by the background
  - Small T: unimodal
  - Large T: multi-modal

Mixture of Gaussians…

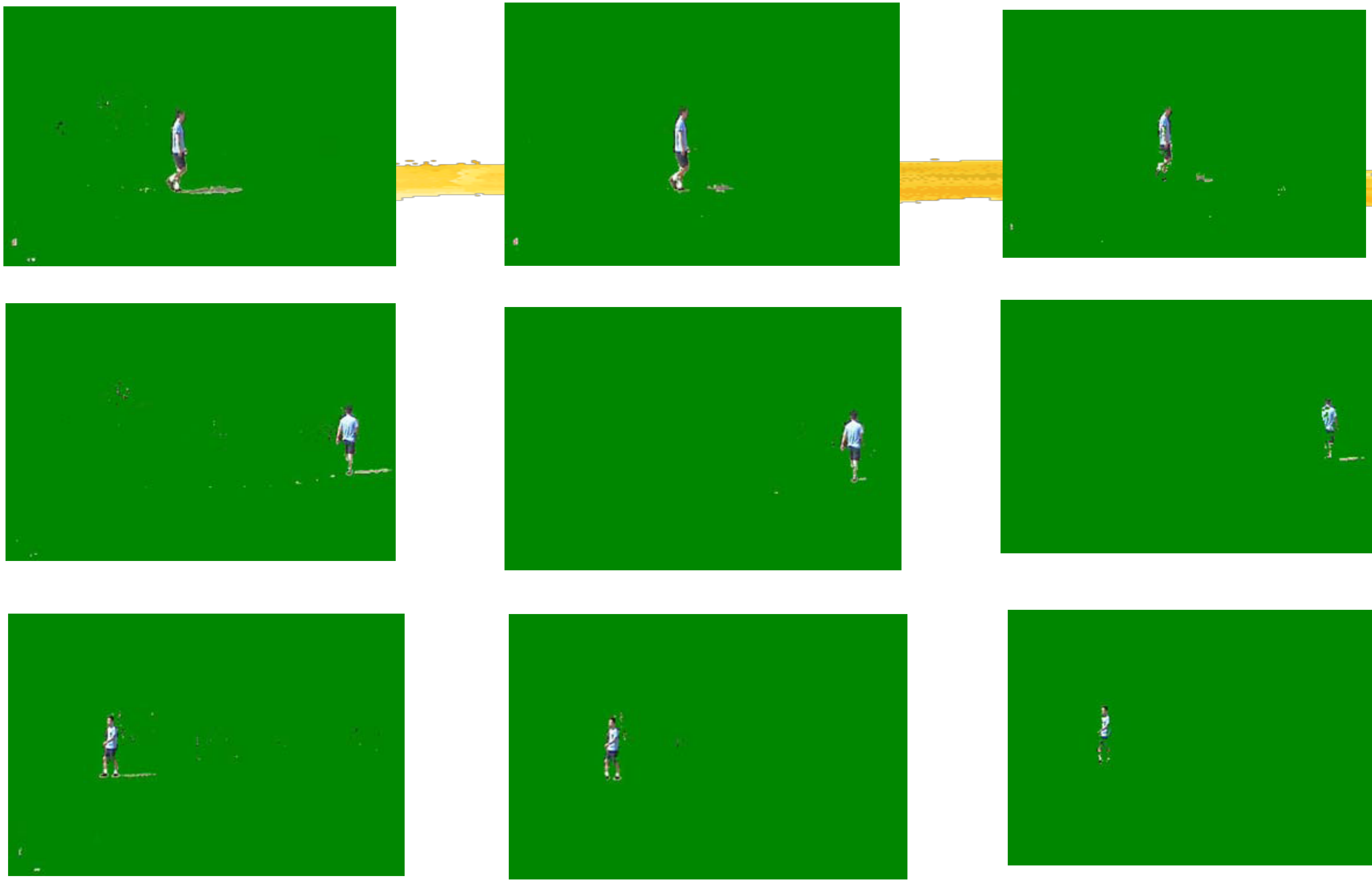


Mixture of Gaussians…

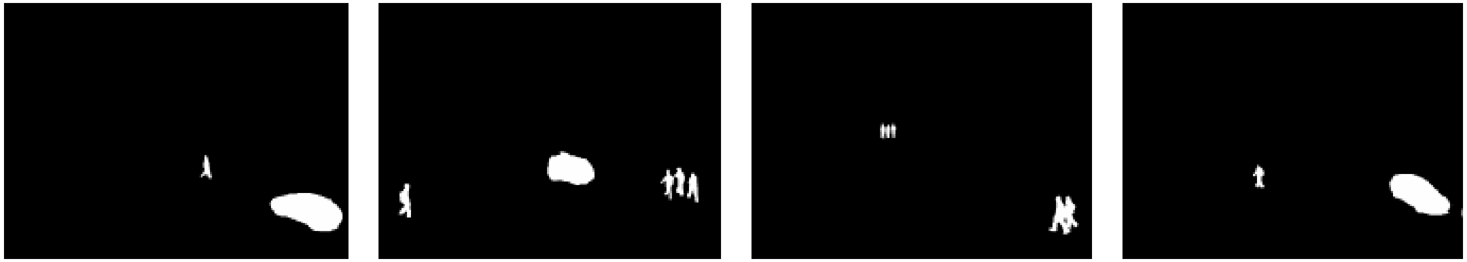


Background (RA, RM, GMM)

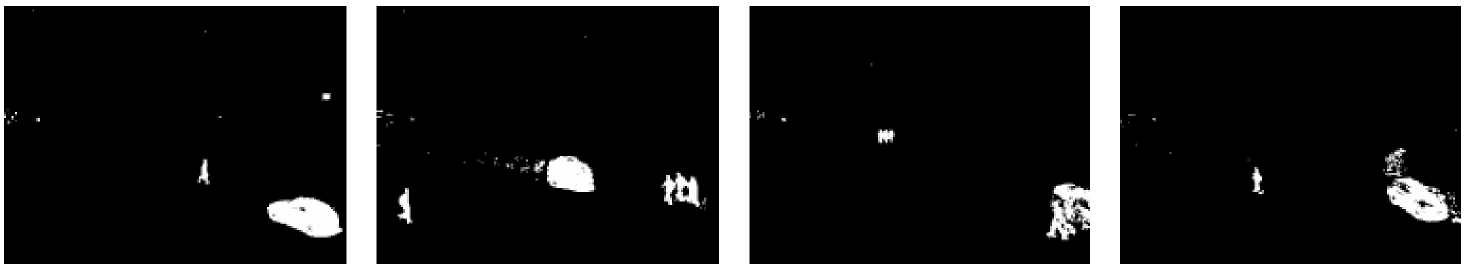
Sample Frames



RA RM GMM



(a) ground-truth mask images

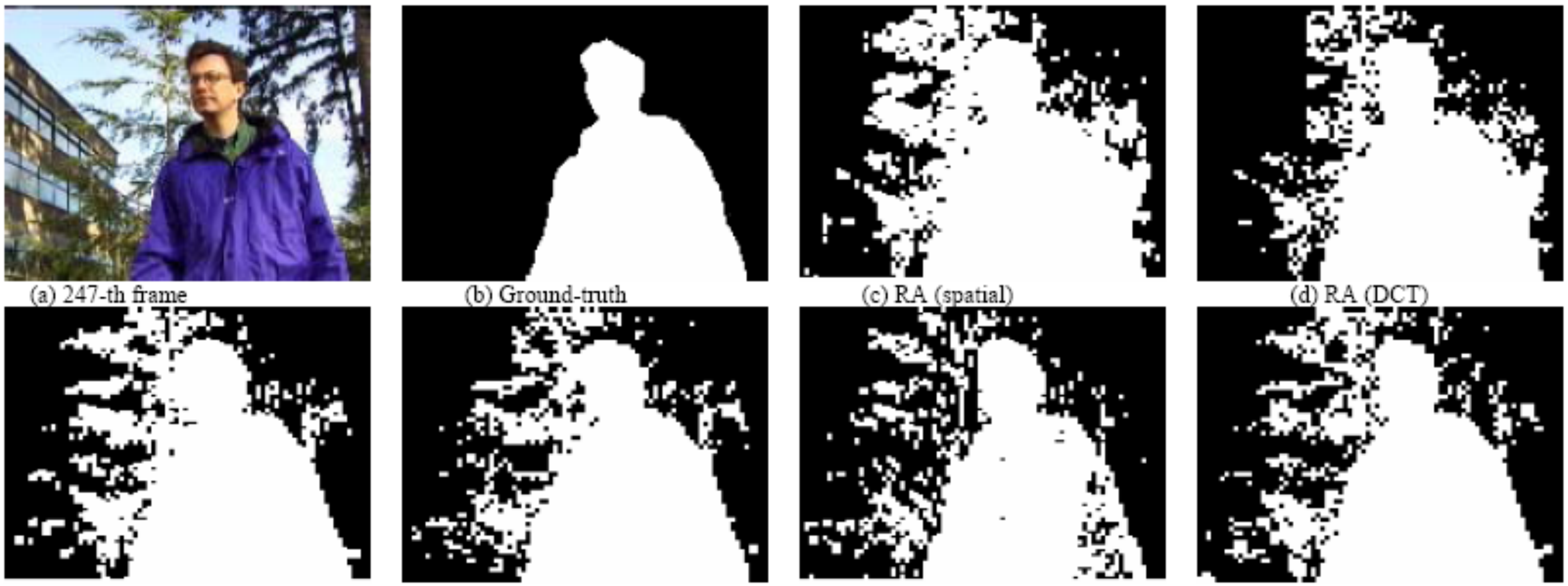


(b) segmentation results generated by the RA algorithm in pixel domain

(d) segmentation results generated by the Median algorithm in pixel domain

(f) segmentation results generated by the MoG algorithm in pxiel domain

Issues in Background Modeling



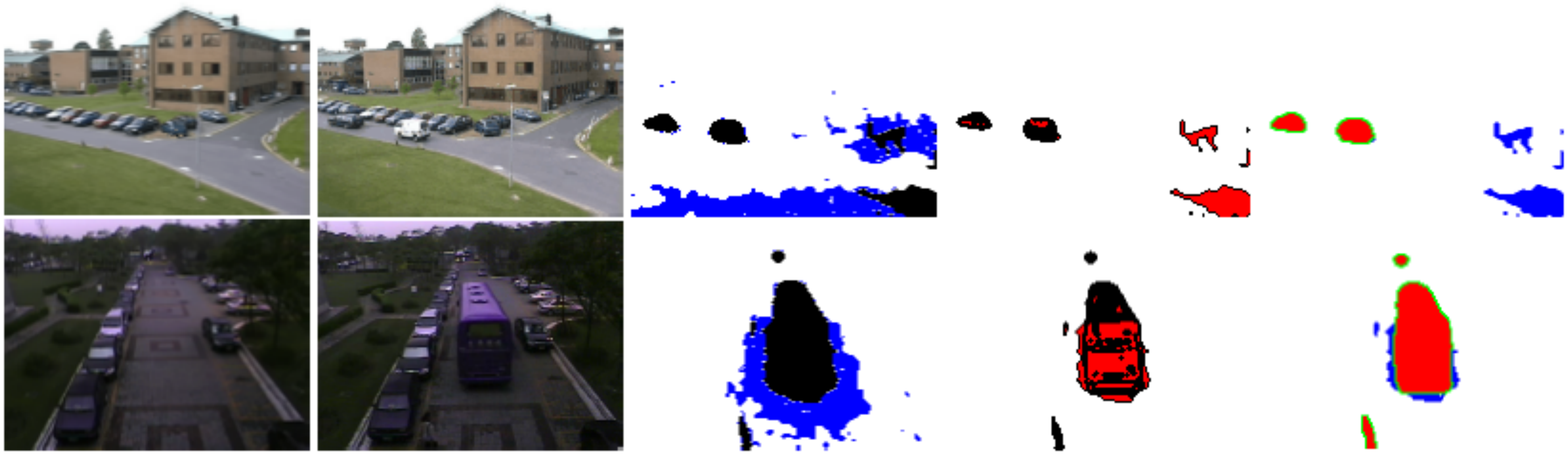
(a) 247-th frame (b) Ground-truth (c) RA (spatial) (d) RA (DCT)

(e) Median(spatial) (f) Median(DCT) (g) MoG (spatial) (h) MoG (DCT)

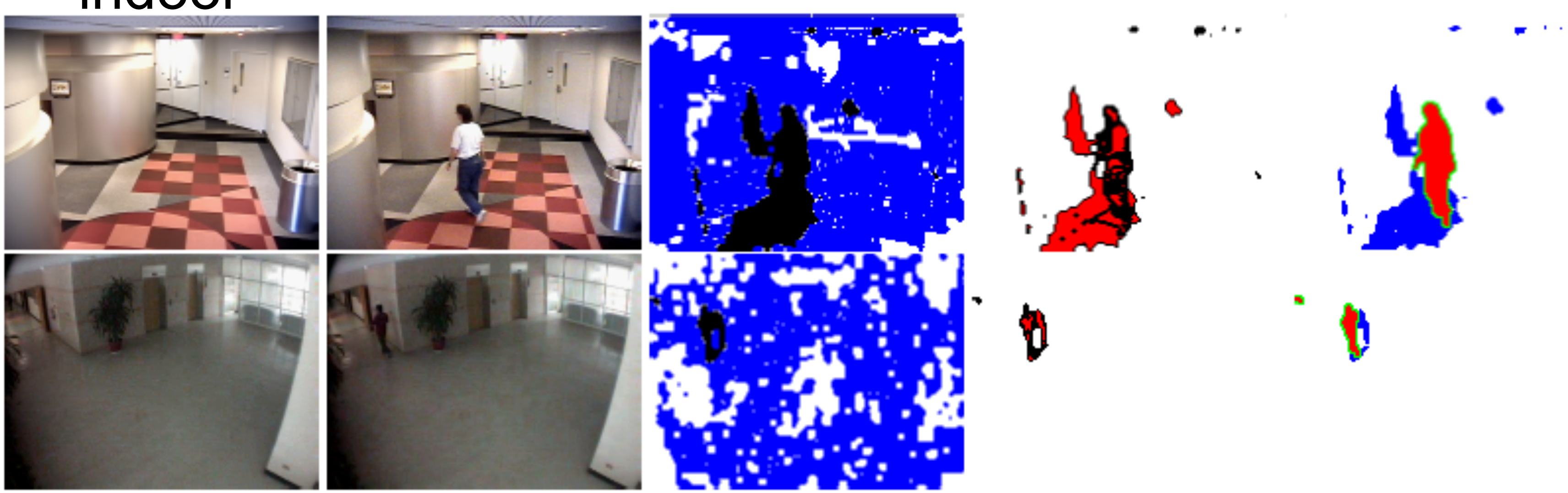
Fig.5. Segmentation results generated by different algorithms in both domains for the "waving tree" sequence

Issues in Background modeling…

- outdoor

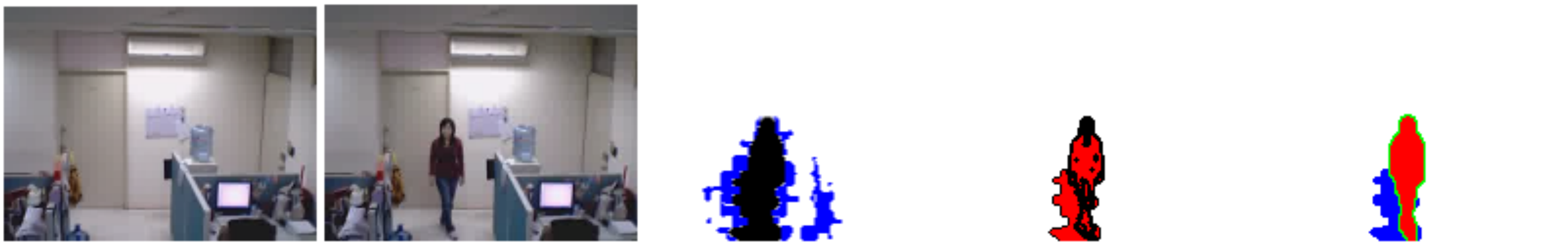


- indoor

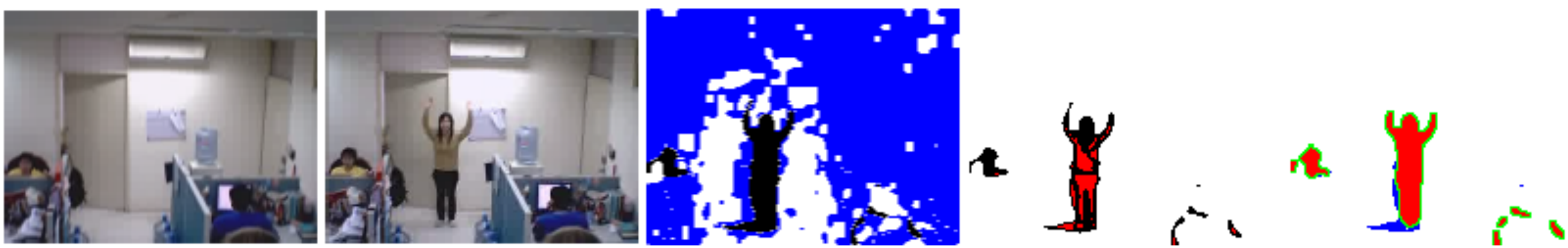


Issues in Background modeling…

- Indoor shadow diffusion

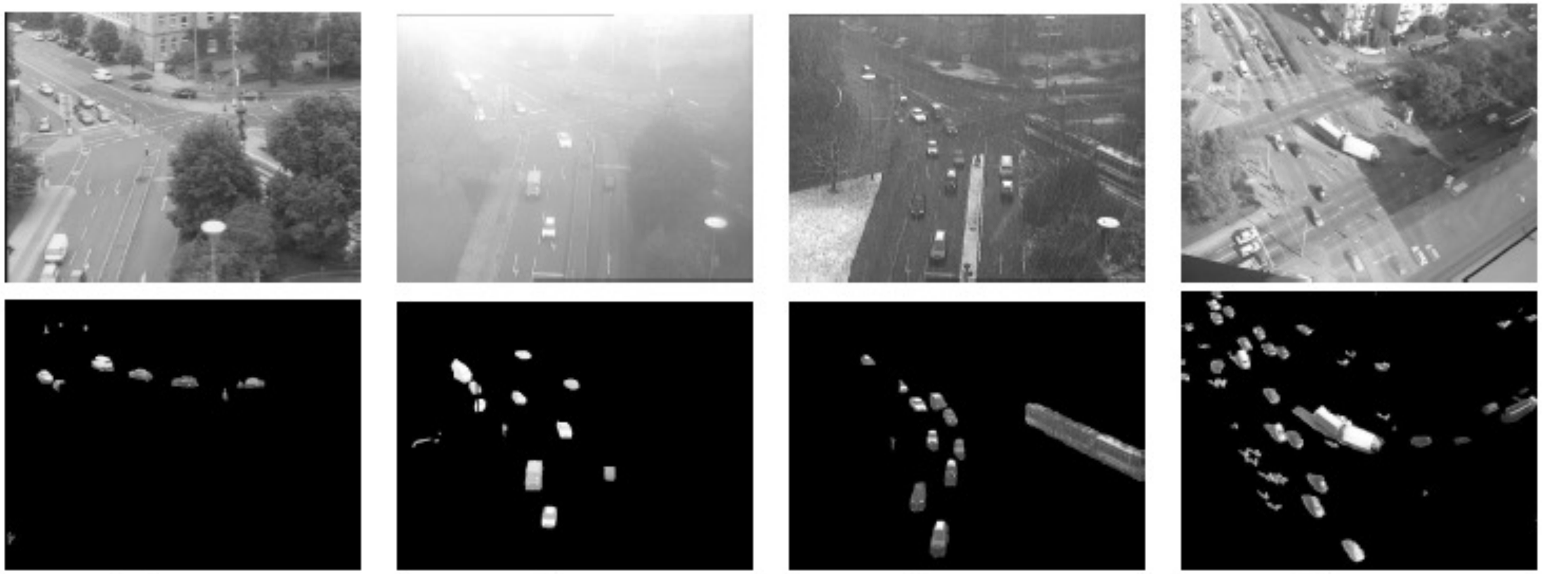


- Auto white balance



- Turning light on/off

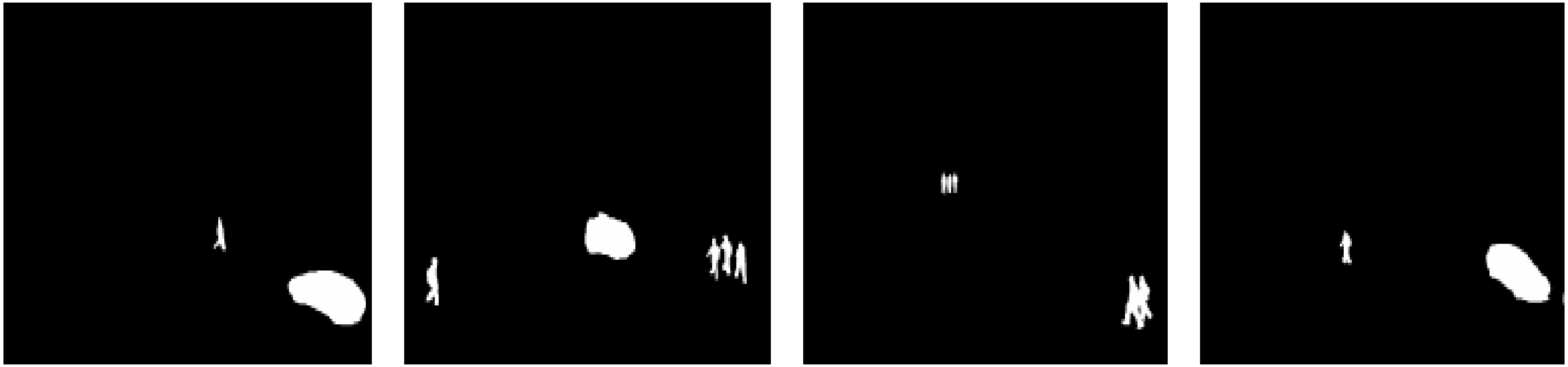
Issues in Background modeling…



(a) Bright (b) Fog (c) Snow (d) Busy

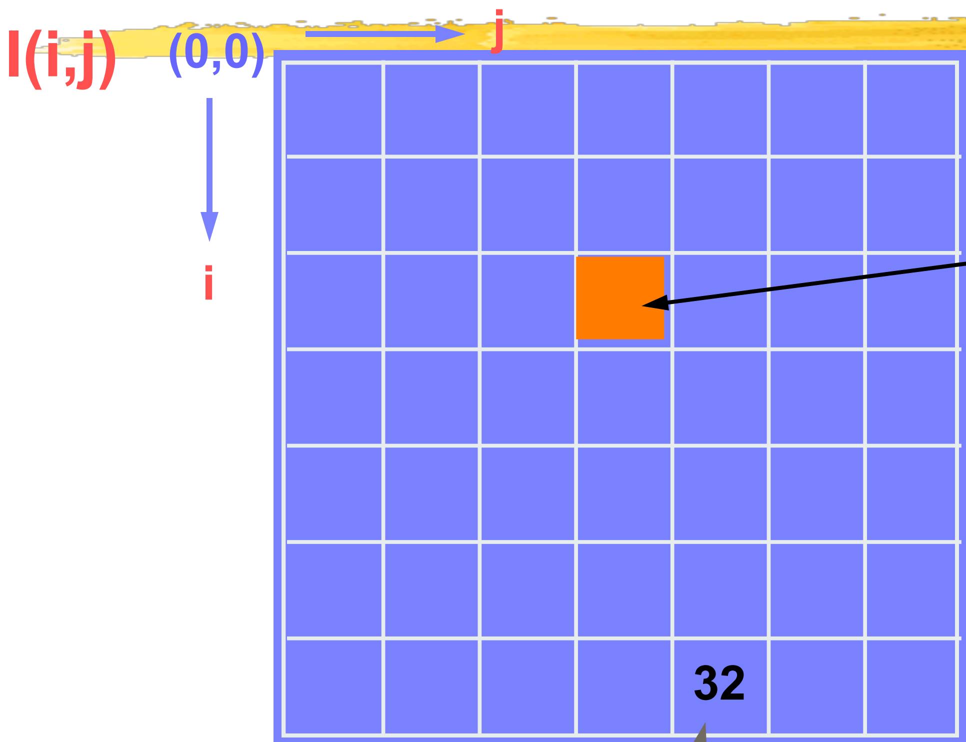
Connected Component Analysis

How many moving objects are there?



(a) ground-truth mask images

Digital Geometry



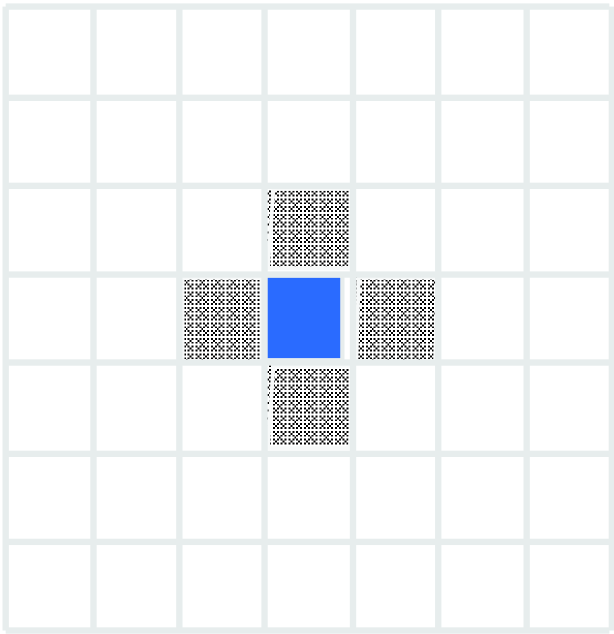
Picture Element or Pixel

- Neighbourhood
- Connectedness

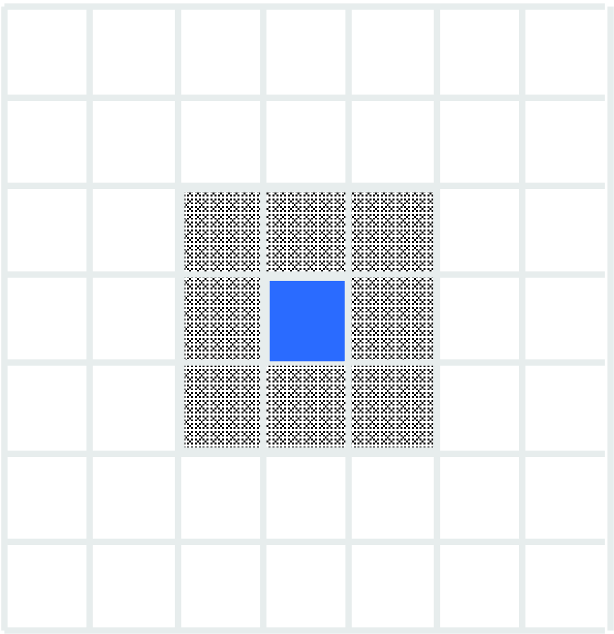
Pixel value I(I,j) =

0,1 Binary Image
0 - K-1 Gray Scale Image
Vector: Multispectral Image

Neighbour



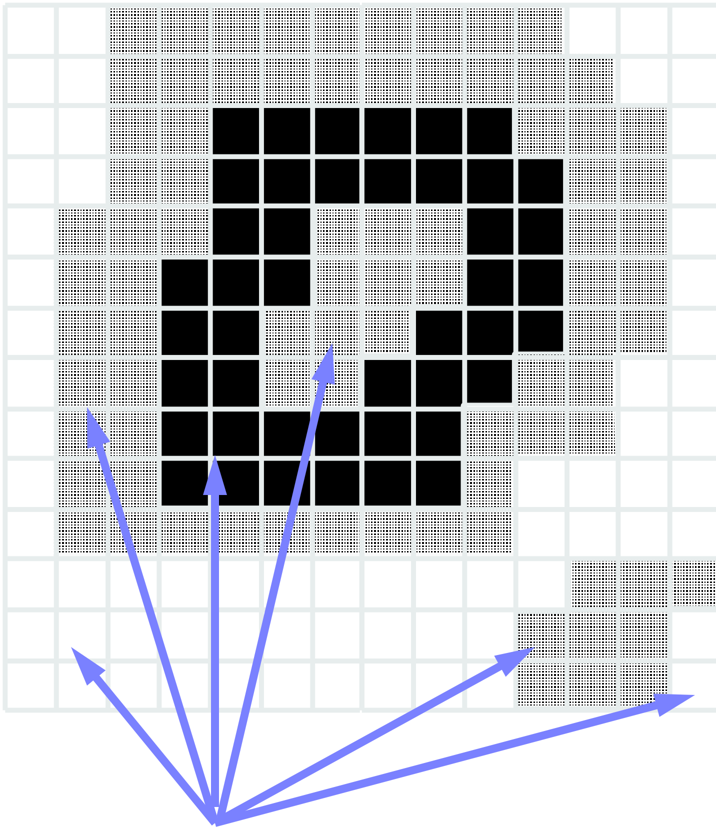
Four-Neighbour



Eight-Neighbour

Connected Component Analysis

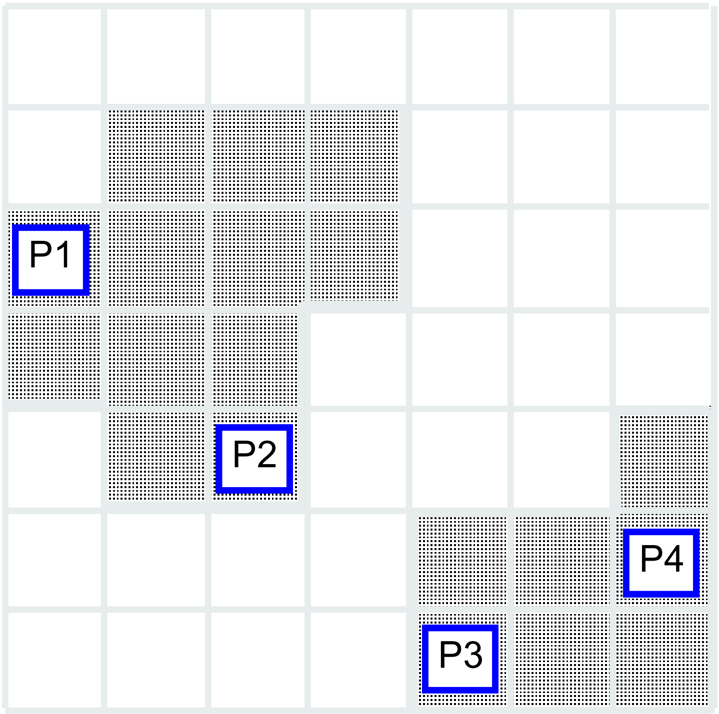
- Binary image with multiple 'objects ' and
- Separate 'objects' must be labeled individually



6 Connected Components

Connected Component Labeling

- Two points in an image are 'connected' if a path can be found for which the value of the image function is the same all along the path.



P_1 connected to P_2

P_3 connected to P_4

P_1 not connected to P_3 or P_4

P_2 not connected to P_3 or P_4

P_3 not connected to P_1 or P_2

P_4 not connected to P_1 or P_2

Recursive CCL Algorithm

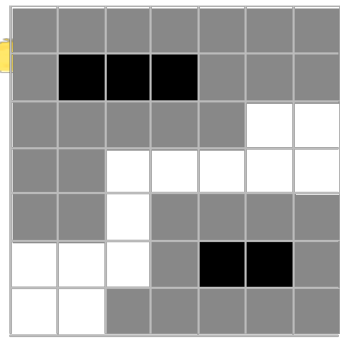
- Algorithm

  1. Scan the image to find an unlabeled ***unity valued*** pixel and assign it a new label L.
  2. Recursively assign a label L to all its unity valued neighbours.
  3. Stop if there are no more unlabeled unity valued pixels.
  4. Go to step 1.

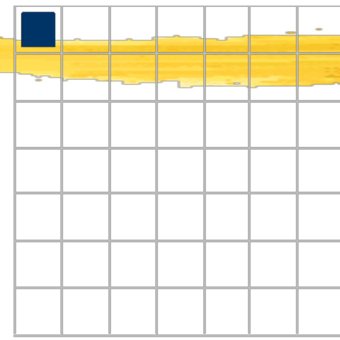
Sequential CCL Algorithm

- Pick any pixel in the image and assign it a label
- Assign same label to any neighbour pixel with the same value of the image function
- Continue labeling neighbours until no neighbours can be assigned this label
- Choose another label and another pixel not already labeled and continue
- If no more unlabeled image points, stop.

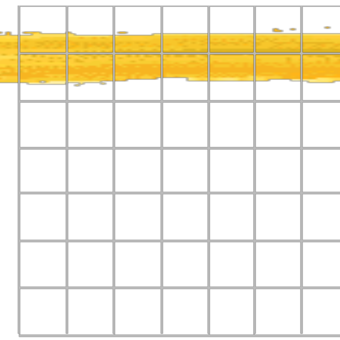
An Example

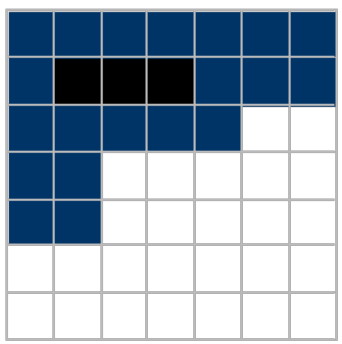


Image

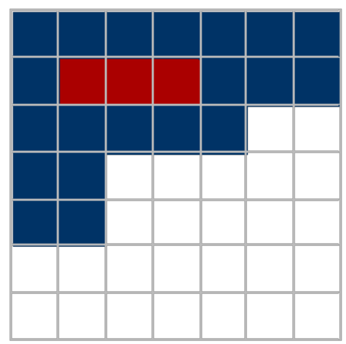


'Label' Image

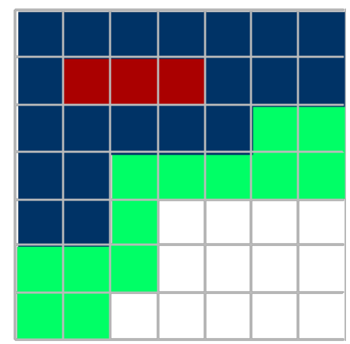




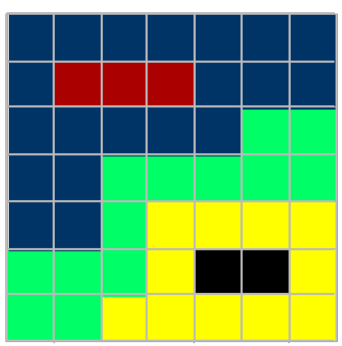
Lab. Im. - 1st Component



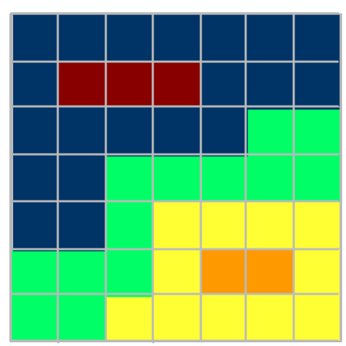
Lab. Im. - 2nd Component



Lab. Im. - 3rd Component

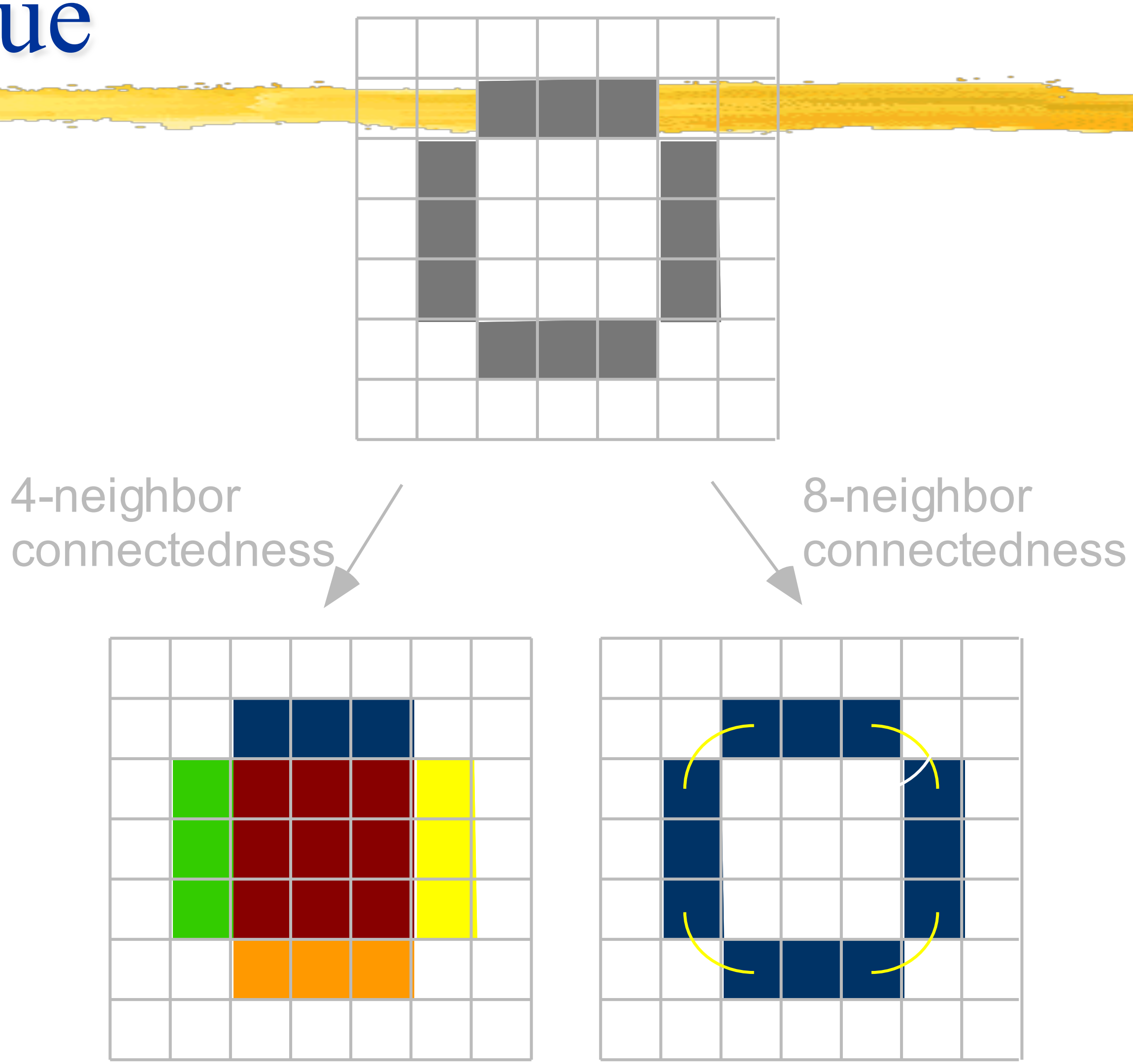


Lab. Im. - 4th Component



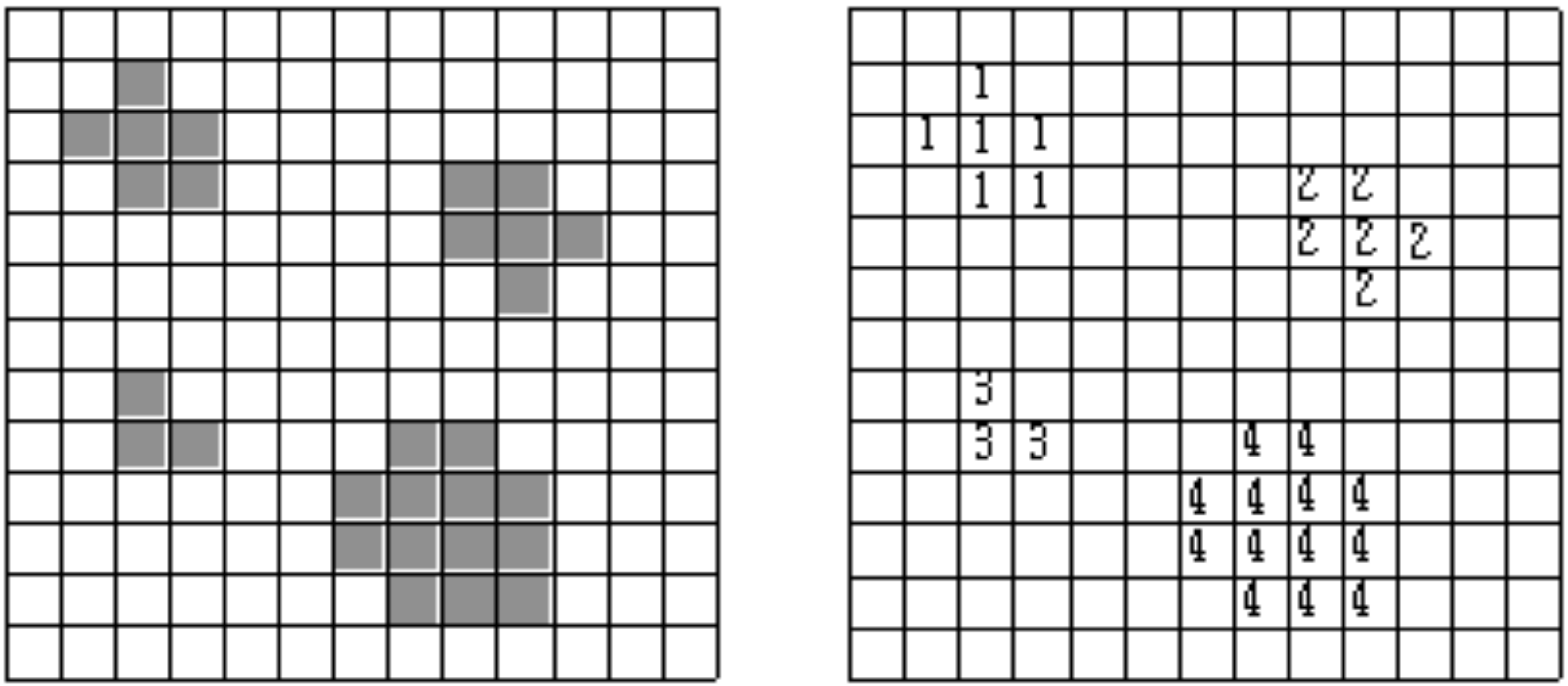
Final Labeling

An Issue



Neither neighborhood definition satisfactory!

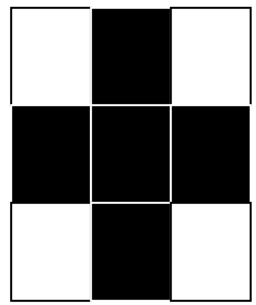
Example One



(a) (b)

Example Two

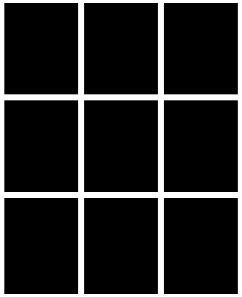
1	1	1	1	1	1
1	0	0	1	1	1
1	1	1	0	1	1
1	2	2	0	0	1
1	2	2	0	0	1



A	A	A	A	A	A
A	B	B	A	A	A
A	A	A	C	A	A
A	D	D	C	C	A
A	D	D	C	C	A

Example Three

1	1	1	1	1	1
1	0	0	1	1	1
1	1	1	0	1	1
1	2	2	0	0	1
1	2	2	0	0	1



A	A	A	A	A	A
A	B	B	A	A	A
A	A	A	B	A	A
A	C	C	B	B	A
A	C	C	B	B	A

References

- 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.
- 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

References

- J. R. Parker, Practical Computer Vision Using C, John Wiley & Sons, Inc., 1993. (Chapter 2)
- L. G. Shapiro and G. C. Stockman, Computer Vision, Prentice Hall, 2001. (Chapter 3)
- H. Samet and M. Tamminen "*Efficient Component Labeling of Images of Arbitrary Dimension Represented by Linear Bintrees*", IEEE Transactions on Pattern Analysis and Machine Intelligence, 10(4), July 1988, pp.579-586.