Background Image Generation Using Boolean Operations

Invited speaker for the 9th Philippine Computing Science Congress (PCSC 2009)
March 2 - 3, 2009 in Silliman University, Dumaguete City

Kardi Teknomo, PhD

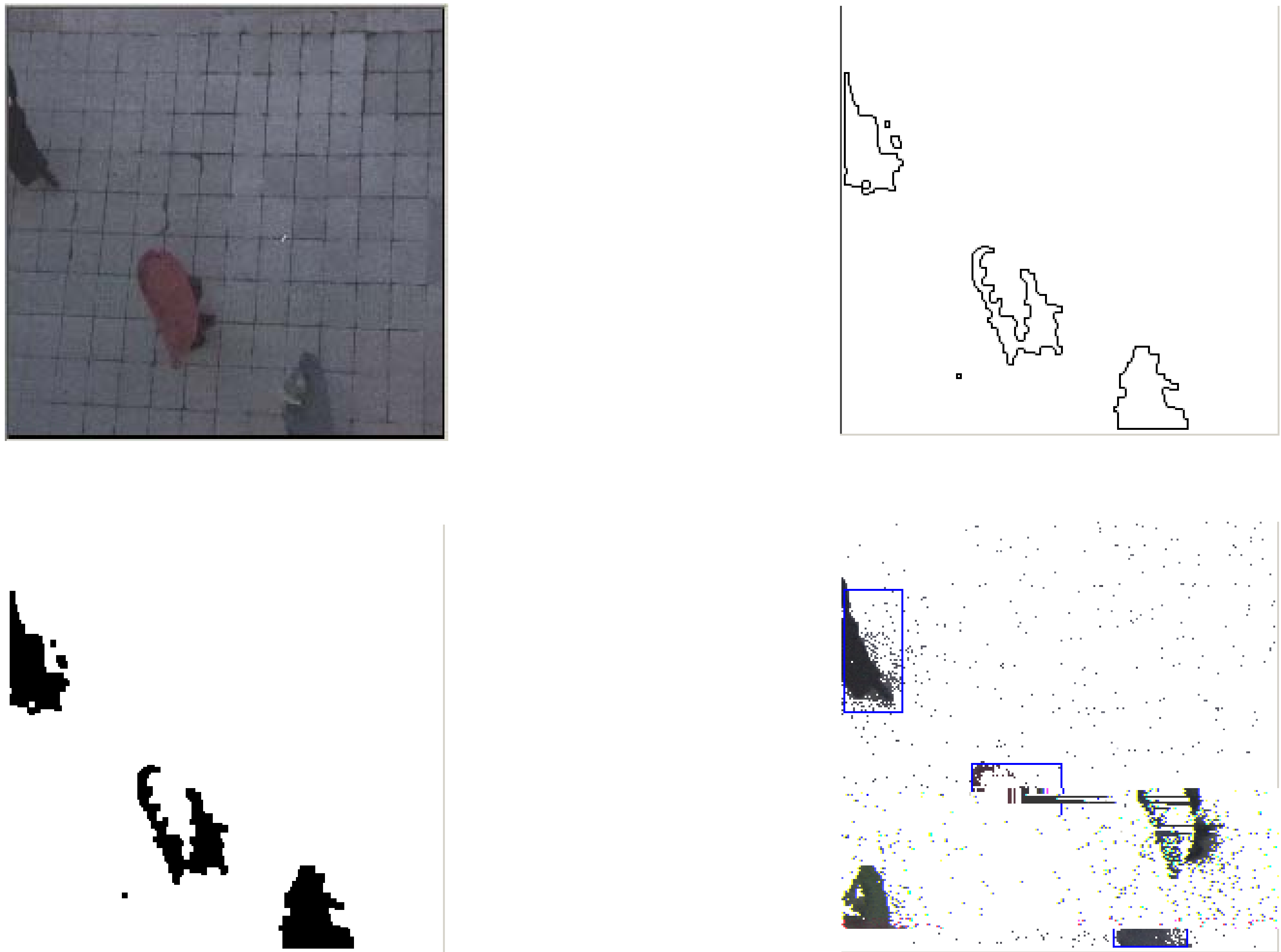
Proceso Fernandez, PhD

Ateneo de Manila University, Quezon City, Philippines

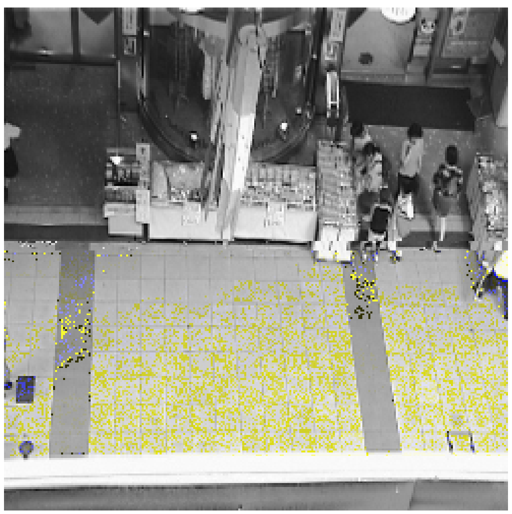
Overview

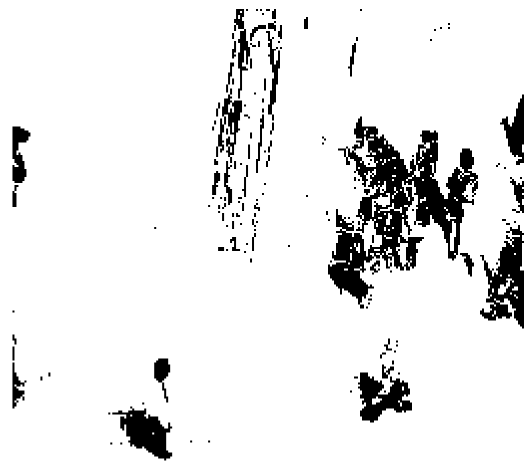
- Motivation
- Background
- State of the Arts
- Problem
- Proposition
- Results & discussion
- Conclusion & further studies

Motivation



Motivation





Background

- Why do we need background image?
  - Background substraction

$$g(x,y)=\begin{cases}0 & \text{if } \left|f(x,y)-b(x,y)\right|<\theta \\ f(x,y) & \text{otherwise}\end{cases}$$

- Better background make further steps of noise filtering and object segmentation and object recognition easier

State of the Arts

- What other people do?
  - Manually, from the video images when there are no objects (Matsuyama, 1999; Tsuchikawa, 1995)
- Background Modeling
  - EM – Gaussian Mixture (Haritaoglu et al, 2000; McKenna et all, 2000)
  - Need Object detection ➔

    Image difference ➔ create unwanted shadow

$$o(x,y,t) = f(x,y,t) - f(x,y,t-1)$$

Image difference

State of the Arts: Mode

- Another way
- Mode of intensity value in each pixel
  - Given an image sequence from a static camera, the mode of intensity level of all pixels in the same location over many image sequence may produce the background image
- Need to find histogram for each pixel over time
  - Very time consuming

Problem

- While mode method is promising, the computational space is 256 . 3 . n . m
- Computational time is O (n . m . t)
- If 25 frames per second ➔ tremendous computation

- How to reduce the computation time and space?

Proposed Approach

- Use Hierarchical Boolean Mode
- Frames are selected randomly
- Basic: only 3 images, Boolean Mode
- Time complexity of
  - Basic mode O(1)
  - Overall: O(R) where R is image size

Assumptions

- At any pixel location (r, c),

$$\#b > \#g + n$$

- where
  - n is average noise at location (x, y)
  - #b = number of background pixels over time T
  - #g = number of foreground pixels over time T
- Since
  - $\#b(r, c) + \#g(r, c) = T$
- Then
  - $\#b(r, c) > \frac{1}{2} T$

Assumptions

- The assumption above can be easily

Simplest Scenario

- T=3 images with non overlapping objects
  - at any pixel location (r, c)
  - at most $\#g = 1$
  - at least $\#b = T - \#g = 3 - 1 = 2$

- What we obtain is actually the largest frequency of the pixel value (i.e. mode) at a pixel location (r, c) over all frames

Proposition

- For any mutually exclusive three frames of image sequence x_1, x_2 and x_3 following our assumption above, we can obtain background image

$$B = x_3(x_1 \oplus x_2) + x_1 x_2$$

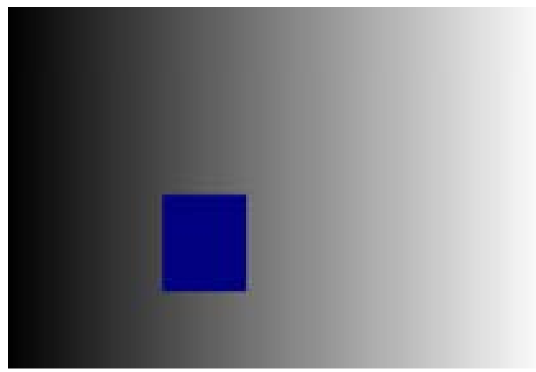
- or

$$b = \left(g_3 \ and \ \left(g_1 \ xor \ g_2\right)\right) or \left(g_1 \ and \ g_2\right)$$

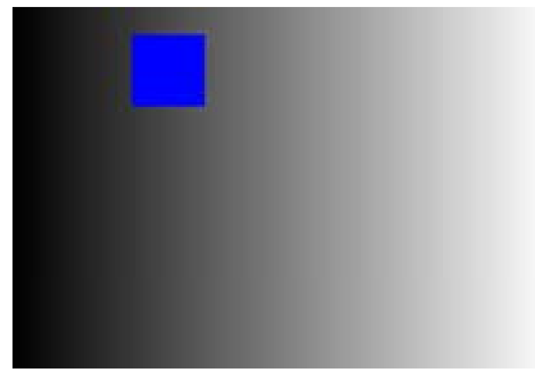
Hypothetical Images

- Simple experimental images with one non-overlapping object on each image

$g_1 = [0, 0, 64]$

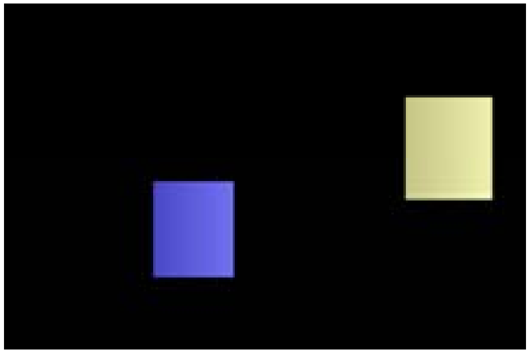


$g_2 = [0, 0, 128]$

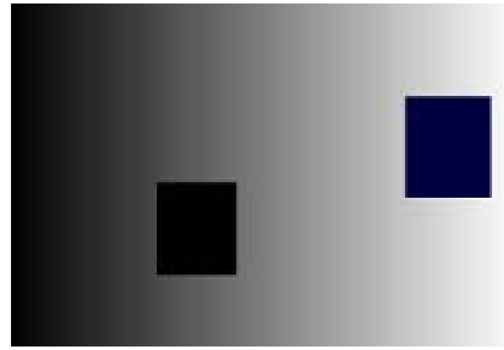


$g_3 = [0, 0, 255]$

Hypothetical Images



g_1 Xor g_2



g_1 And g_2

g_3 And (g_1 Xor g_2)

b = g_3 And (g_1 Xor g_2)) Or (g_1 And g_2)

8 Possibilities of 3 Binary Images

- Simplified version but gives clarity on what actually happens

x_1	0	0	0	0	1	1	1	1
x_2	0	0	1	1	0	0	1	1
x_3	0	1	0	1	0	1	0	1
S	0	0	0	1	0	1	1	1

- Boolean Mode function

$$S = \begin{cases} 1 \quad if \quad \sum_{i=1}^{t} x_i \geq \left\lceil \frac{t}{2} + 1 \right\rceil, \quad t \geq 3 \\ 0 \quad otherwise \end{cases}$$

Derivation

- The background image can be taken as the value of 1 in S

$$B = \bar{x}_1 x_2 x_3 + x_1 \bar{x}_2 x_3 + x_1 x_2 \bar{x}_3 + x_1 x_2 x_3$$

- Or, $B = \left(\bar{x}_1 x_2 + x_1 \bar{x}_2\right) x_3 + x_1 x_2 \left(\bar{x}_3 + x_3\right)$

- But, by definition $x_1 \oplus x_2 = \bar{x}_1 x_2 + x_1 \bar{x}_2$

- Thus, it can be simplified into our proposition

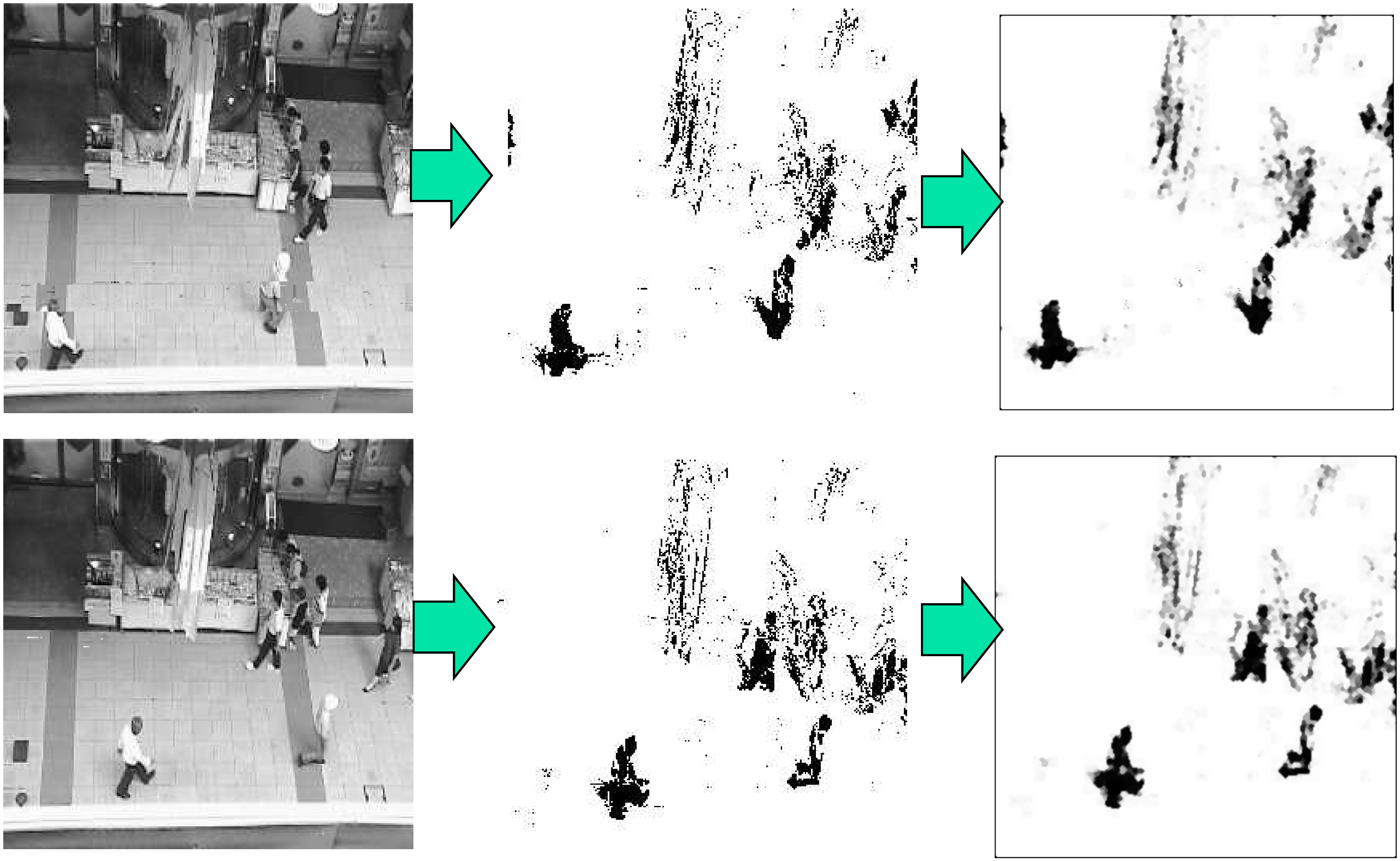
$$B = x_3 (x_1 \oplus x_2) + x_1 x_2$$

Hierarchical Mode

- Multilevel mode:
  - 1st Level: 3 random frames → 1 1st level background frame
  - 2nd Level: 3 of 1st level background → 1 2nd level background frame
  - …
  - Lth Level: 3 of (L-1)th level background → 1 final level background frame

Sample Results (Gray Level)

Subtracted Background



Sample results (Color)

Space Complexity

- Space complexity: $O(RF + R3^L)$
  - R = image size (per frame)
  - F = number of frames

- In practice L ≤ 6 ➔ $O(RF)$ = 27 seconds of video input

Time Complexity

- Basic 3 frames: $O(R)$-time
- Hierarchical: $O(3^L)$

- In practice: L ≤ 3 ➔ $O(1)$
- Overall algorithm: $O(R)$-time

Kardi Teknomo & Proceso Fernandez

Expected Accuracy

- Probability to select a frame $p_0=m/T$
  - m = modal pixel value $\in (0.5T,\ T]$
  - T = total frames in video sequence

- At 1st level, probability to obtain modal pixel of the entire sequence is

$$p_1 = \underbrace{p_0^3}_{\text{Modal value happen in all 3 frames}} + \underbrace{3(p_0)^2(1-p_0)}_{\text{Modal value happens in 2 out of 3 frames}}$$

Expected Accuracy (contd.)

- 2nd Level probability of accurate prediction:
  - $p_2 = p_1^3 + 3(p_1)^2(1 - p_1)$

- Recursive probability of accuracy

$$p_k = (p_{k-1})^3 + 3(p_{k-1})^2(1 - p_{k-1})$$

Expected Accuracy (contd.)

p_0	p_1	p_2	p_3	p_4	p_5	p_6
0.50	0.50	0.50	0.50	0.50	0.50	0.50
0.55	0.575	0.611	0.664	0.737	0.829	0.923
0.60	0.648	0.716	0.803	0.899	0.972	**0.998**
0.65	0.718	0.807	0.902	0.973	0.998	1.000
0.70	0.784	0.880	0.960	0.995	1.000	1.000
0.75	0.844	0.934	0.988	1.000	1.000	1.000
0.80	0.896	0.970	0.997	1.000	1.000	1.000
0.85	0.939	0.989	1.000	1.000	1.000	1.000
0.90	0.972	0.998	1.000	1.000	1.000	1.000
0.95	0.993	1.000	1.000	1.000	1.000	1.000
1	1	1	1	1	1	1

- At low modal bit of 60% of the frame, probability of accurate modal bit determination is already more than 99% at 6 levels (about 3^6 = 729 frames = 27 seconds of video data)

Conclusion

- An alternative fast algorithm to generate background from real world images
- Uses less memory space than simple modal algorithm
- No need to detect the existence of the objects
- At about few seconds of video data, assuming the background image comprises at least 60% of the video, we were able to extract the background with over 99% accuracy