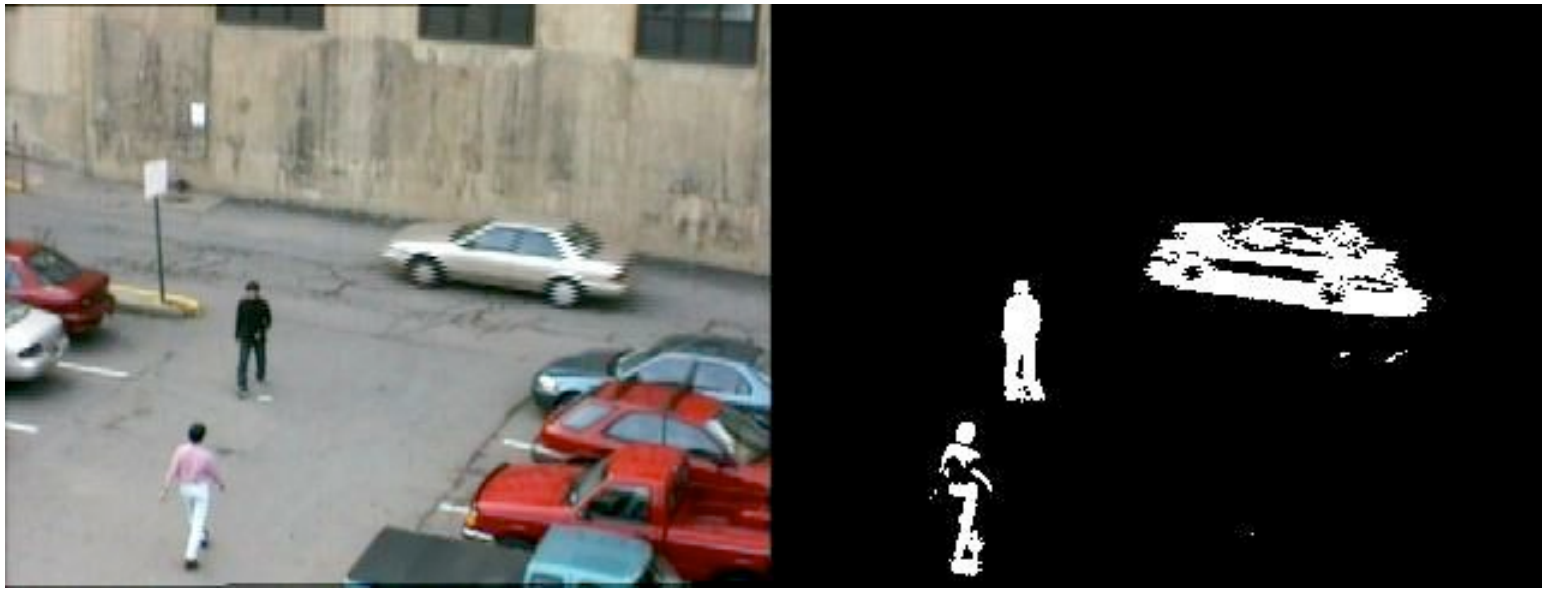
Robert Collins
CSE486, Penn State

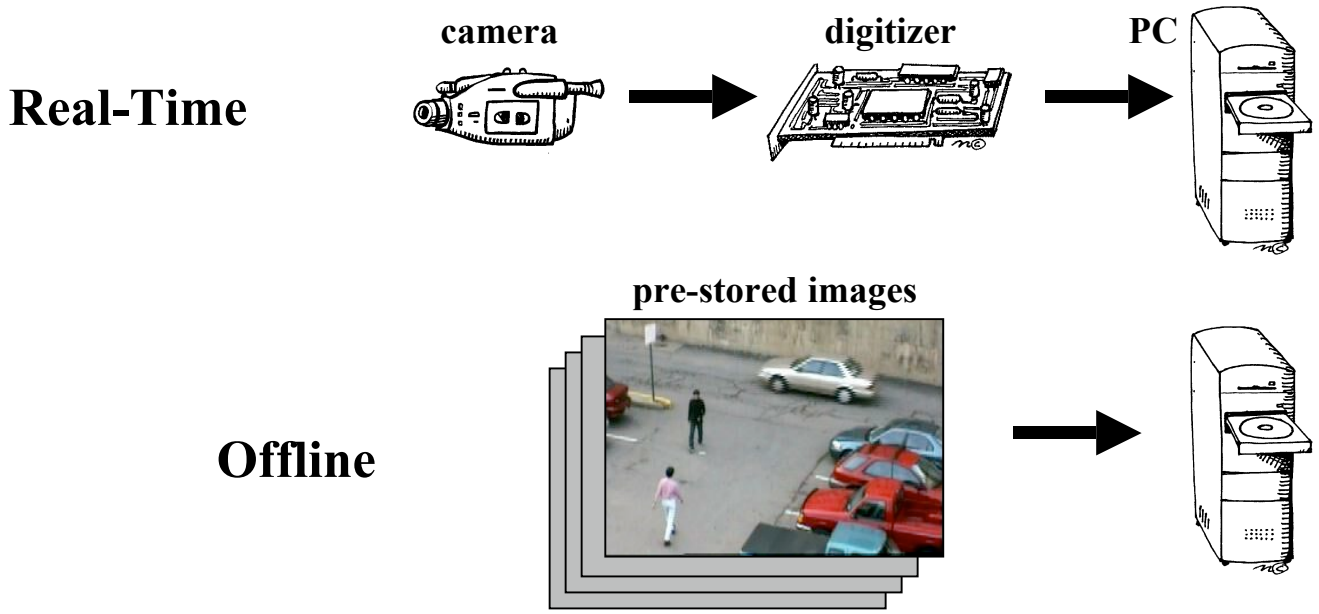
Lecture 24

Video Change Detection



Robert Collins
CSE486, Penn State

Basics of Video



Frames come in 30 times per second. This is not much time to process each image. Real-time algorithms therefore tend to be very simple.

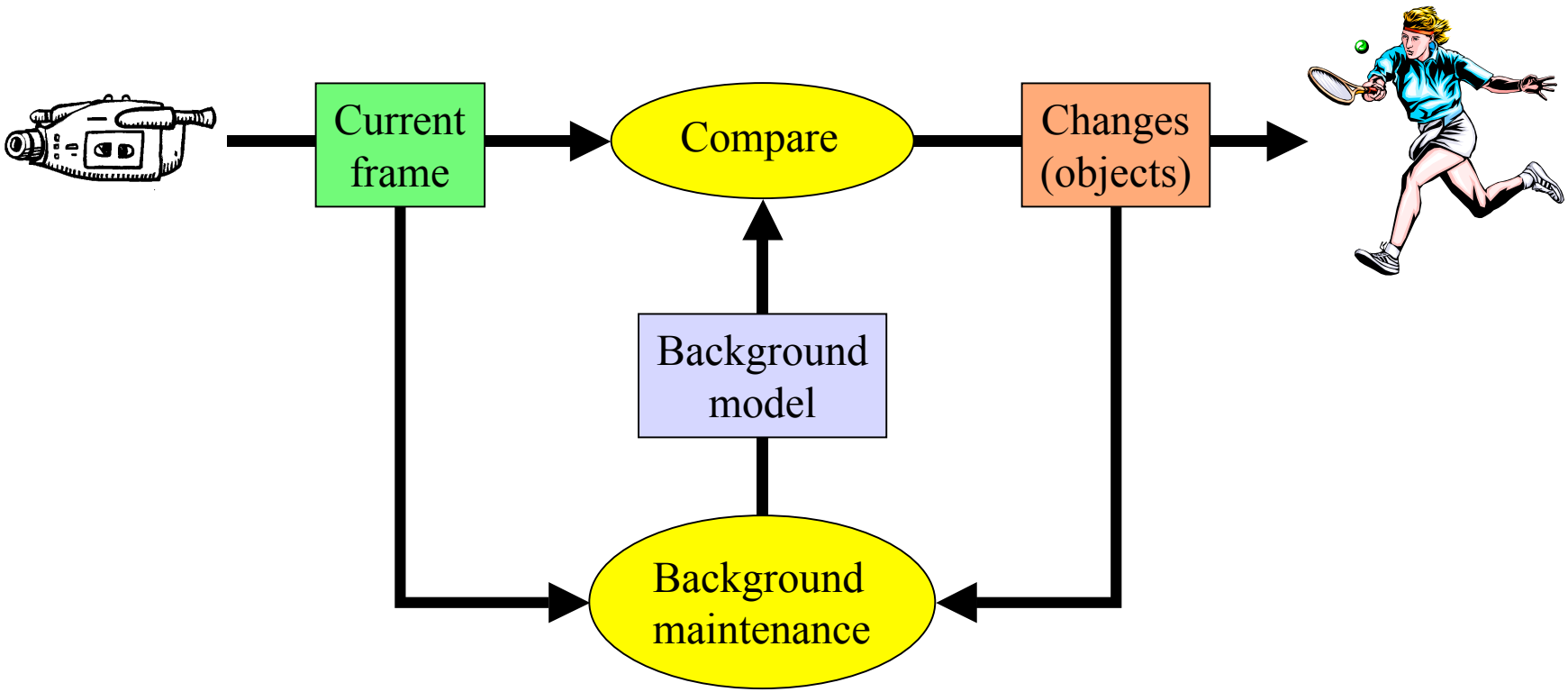
One of the main features of video imagery is the temporal consistency from frame to frame. Not much changes during 1/30 of a second!

Robert Collins
CSE486, Penn State

Detecting Moving Objects

Assumption: objects that move are important (e.g. people and vehicles)

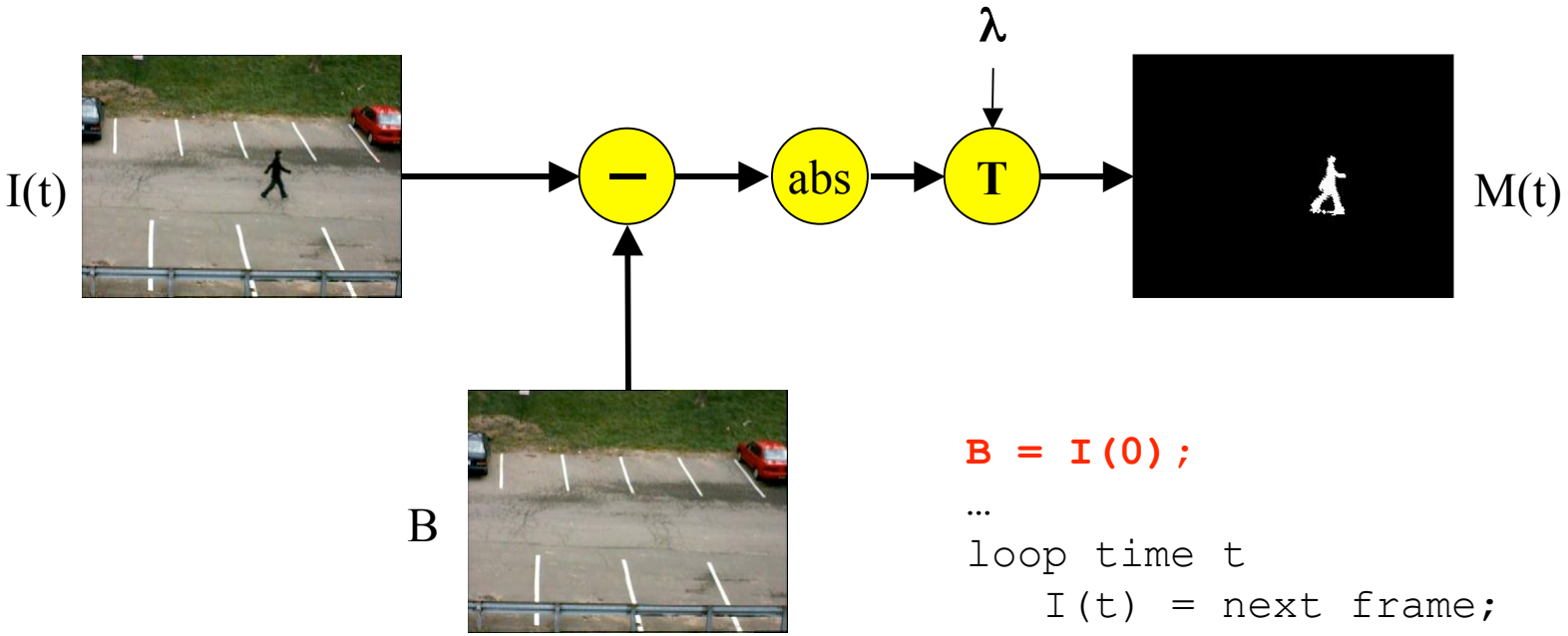
Basic approach: maintain a model of the static background. Compare the current frame with the background to locate moving foreground objects.



Robert Collins
CSE486, Penn State

Simple Background Subtraction

- Background model is a static image (assumed to have no objects present).
- Pixels are labeled as object (1) or not object (0) based on thresholding the absolute intensity difference between current frame and background.




```
B = I(0);
…
loop time t
    I(t) = next frame;
    diff = abs[B - I(t)];
    M(t) = threshold(diff,λ);
    …
end
```

Robert Collins
CSE486, Penn State

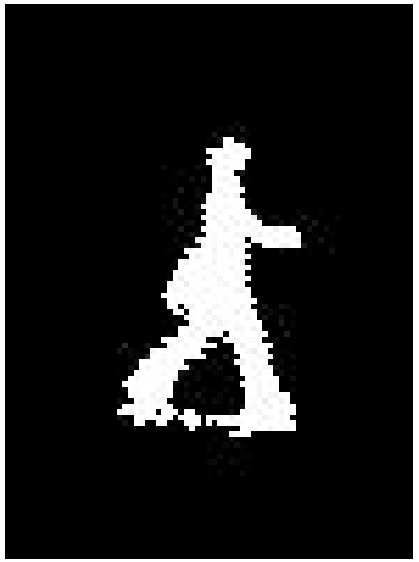
Background Subtraction Results

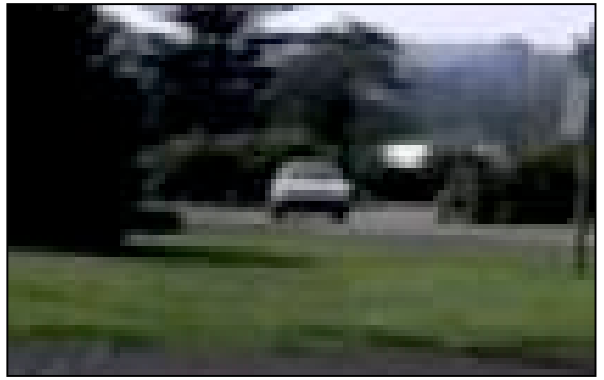
movie

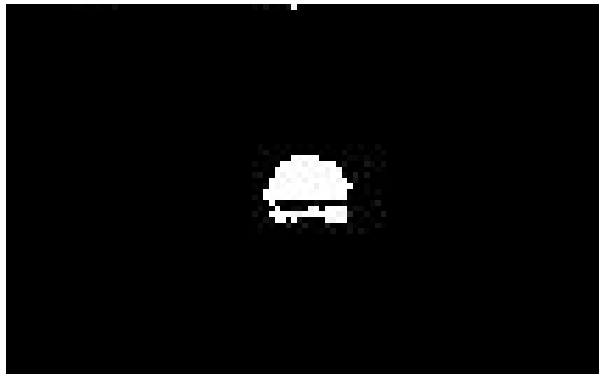
Robert Collins
CSE486, Penn State

BG Observations

Background subtraction does a reasonable job of extracting the shape of an object, provided the object intensity/color is sufficiently different from the background.





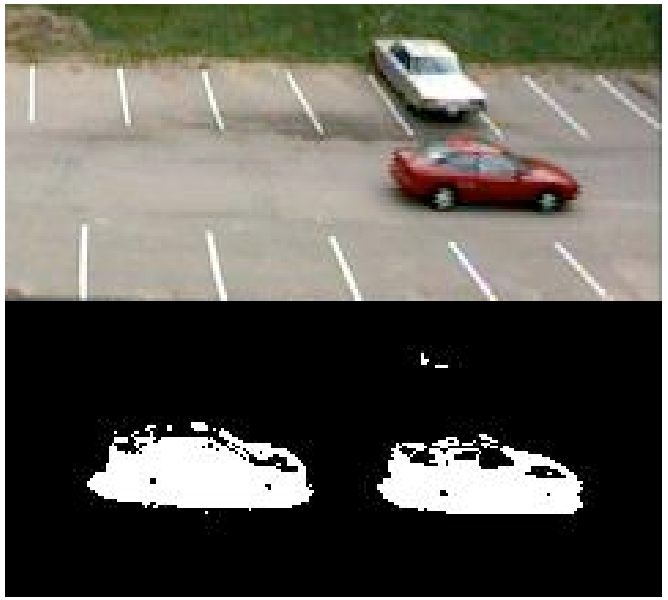


Robert Collins
CSE486, Penn State

BG Observations

Objects that enter the scene and stop continue to be detected, making it difficult to detect new objects that pass in front of them.

If part of the assumed static background starts moving, both the object and its negative ghost (the revealed background) are detected



Robert Collins
CSE486, Penn State

BG Observations

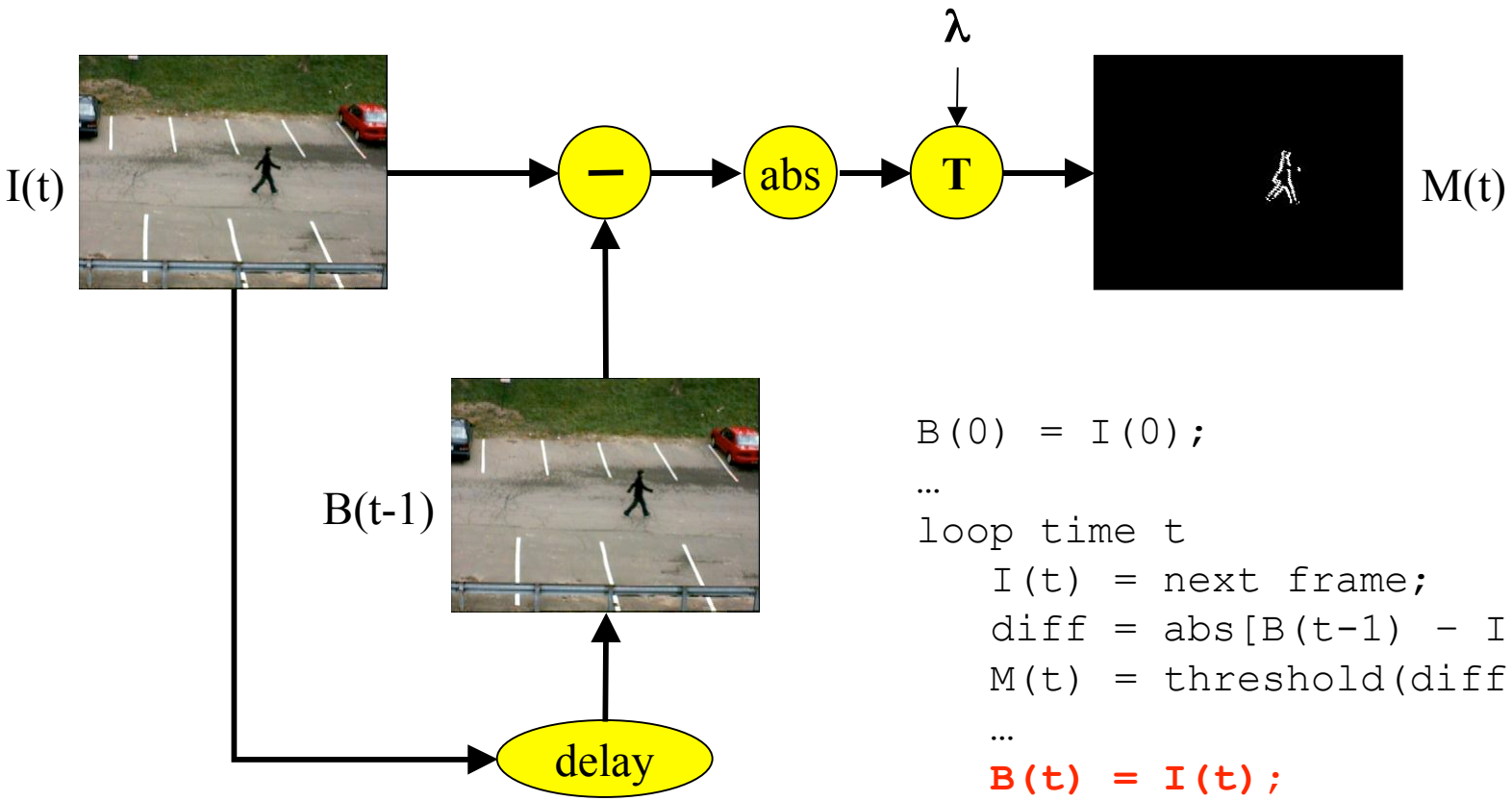
Background subtraction is sensitive to changing illumination and unimportant movement of the background (for example, trees blowing in the wind, reflections of sunlight off of cars or water).

Background subtraction cannot handle movement of the camera.

Robert Collins
CSE486, Penn State

Simple Frame Differencing

• Background model is replaced with the previous image.




```
B(0) = I(0);
…
loop time t
    I(t) = next frame;
    diff = abs[B(t-1) - I(t)];
    M(t) = threshold(diff,λ);
    …
    B(t) = I(t);
end
```

Robert Collins
CSE486, Penn State

Frame Differencing Results

movie

Robert Collins
CSE486, Penn State

FD Observations

Frame differencing is very quick to adapt to changes in lighting or camera motion.

Objects that stop are no longer detected. Objects that start up do not leave behind ghosts.

However, frame differencing only detects the leading and trailing edge of a uniformly colored object. As a result very few pixels on the object are labeled, and it is very hard to detect an object moving towards or away from the camera.

Robert Collins
CSE486, Penn State

Differencing and Temporal Scale

Note what happens when we adjust the temporal scale (frame rate) at which we perform two-frame differencing …

Define D(N) = || I(t) - I(t+N) ||

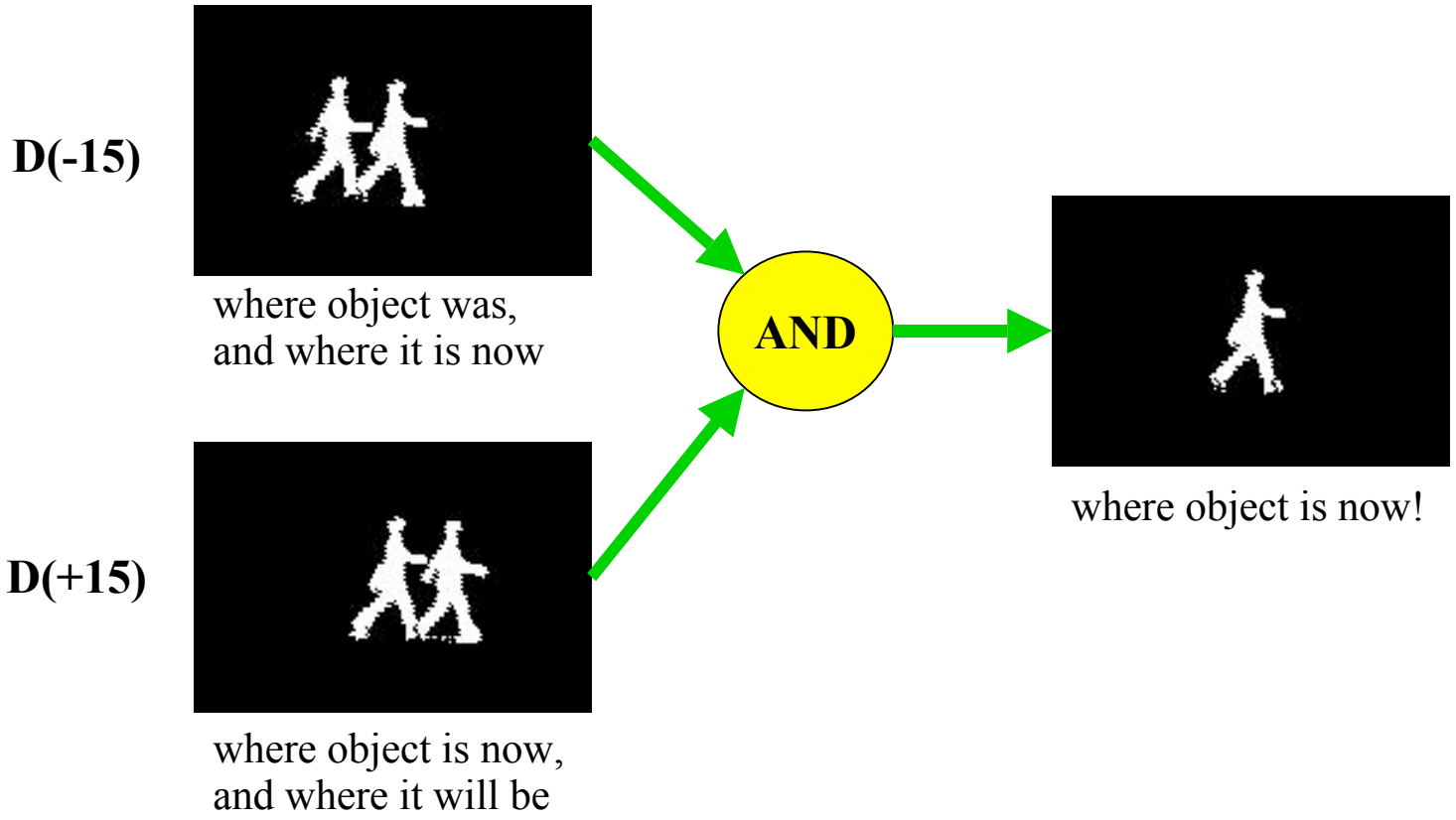
I(t) **D(-1)** **D(-3)** **D(-5)** **D(-9)** **D(-15)**

more complete object silhouette, but two copies
(one where object used to be, one where it is now).

Robert Collins
CSE486, Penn State

Three-Frame Differencing

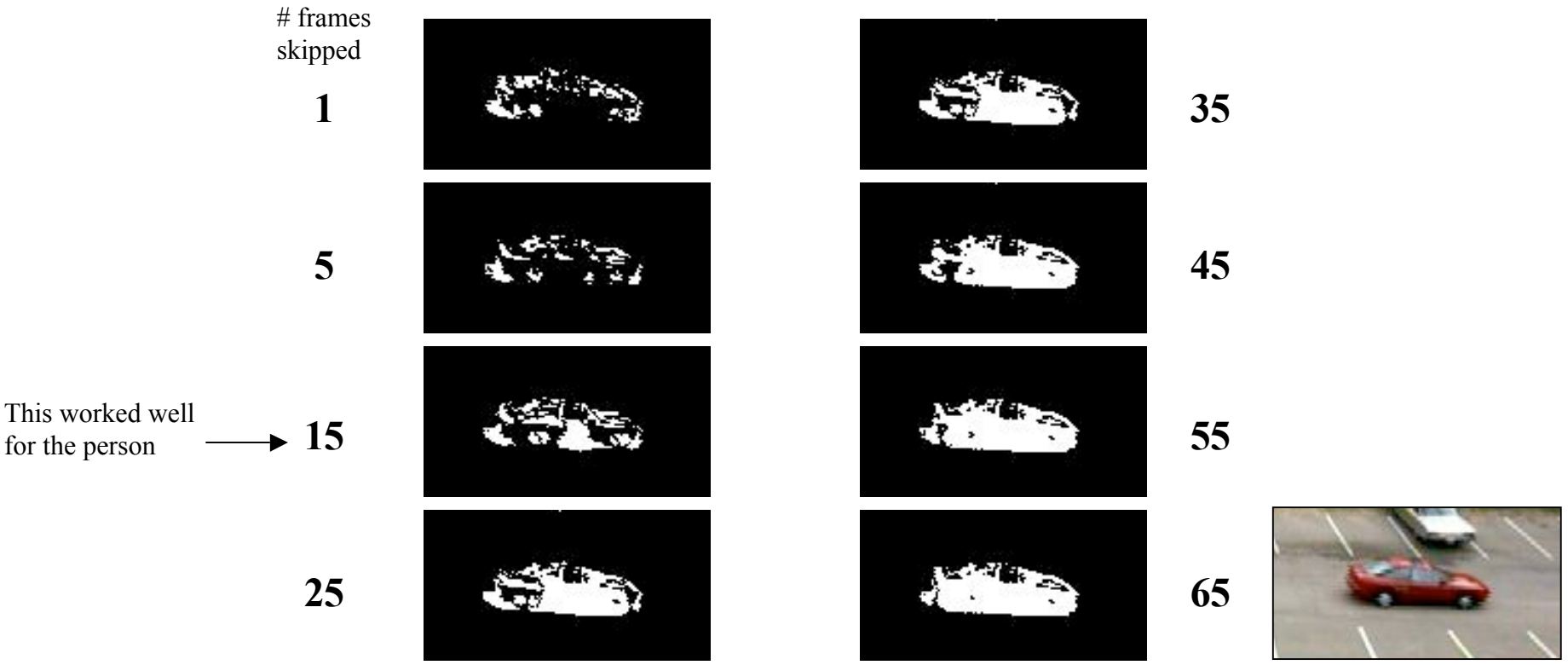
The previous observation is the motivation behind three-frame differencing



Robert Collins
CSE486, Penn State

Three-Frame Differencing

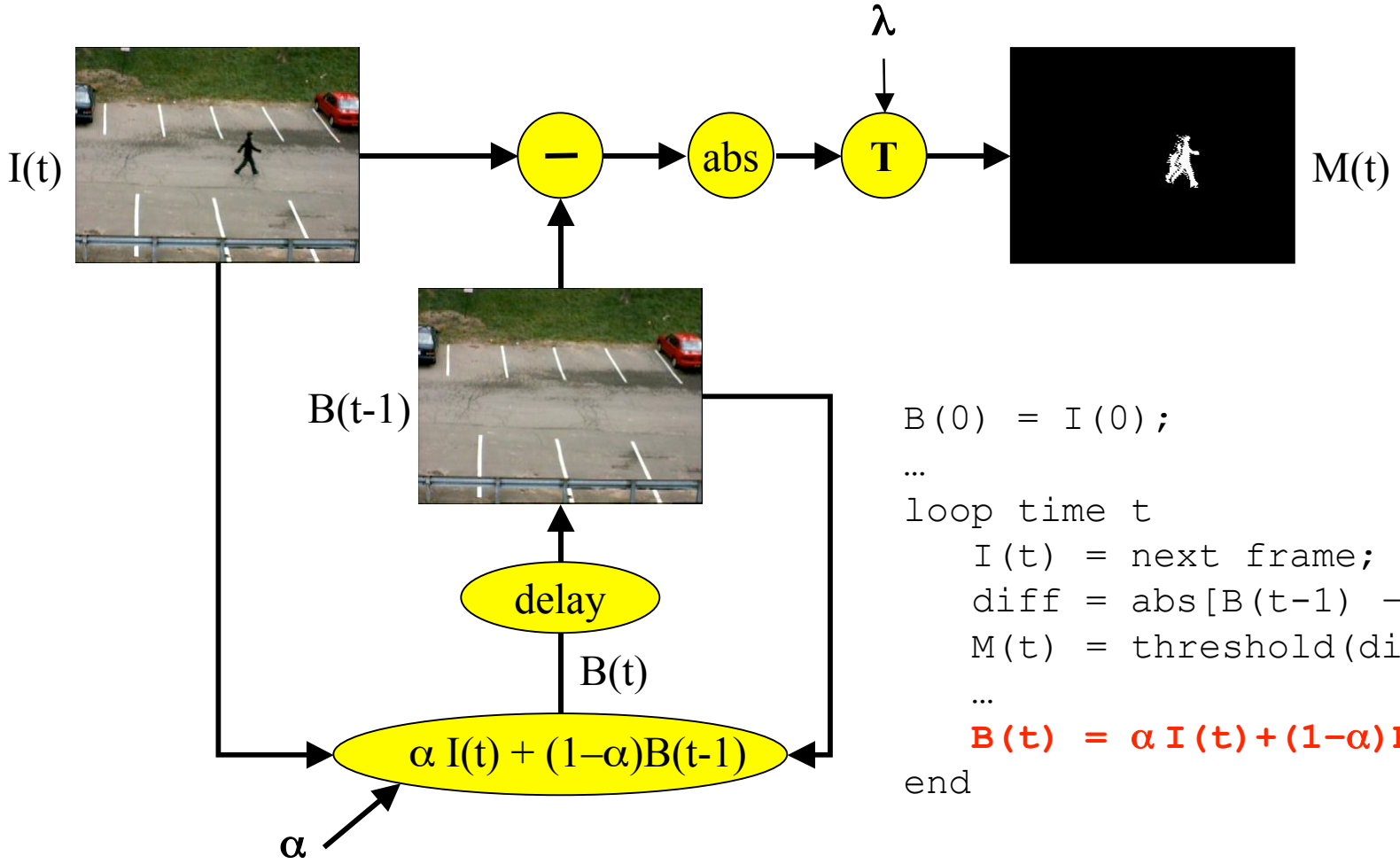
Choice of good frame-rate for three-frame differencing depends on the size and speed of the object



Robert Collins
CSE486, Penn State

Adaptive Background Subtraction

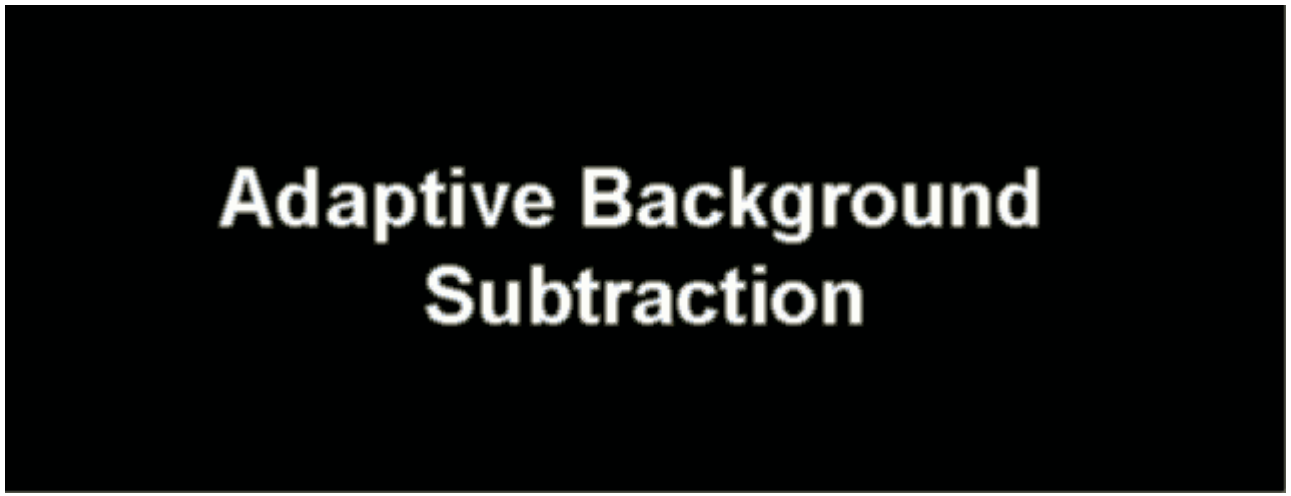
- Current image is "blended" into the background model with parameter α
- $\alpha = 0$ yields simple background subtraction, $\alpha = 1$ yields frame differencing




```
B(0) = I(0);
…
loop time t
    I(t) = next frame;
    diff = abs[B(t-1) - I(t)];
    M(t) = threshold(diff,λ);
    …
    B(t) = α I(t)+(1–α)B(t-1);
end
```

Robert Collins
CSE486, Penn State

Adaptive BG Subtraction Results



movie

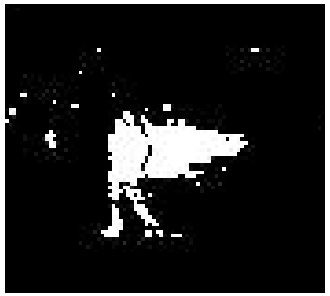
Robert Collins
CSE486, Penn State

Adaptive BG Observations

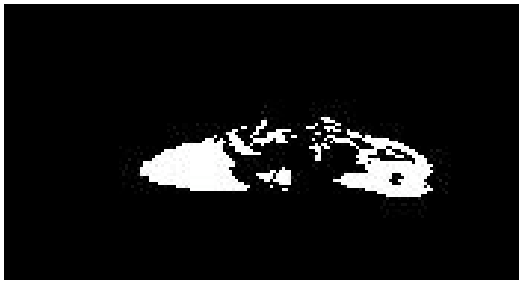
Adaptive background subtraction is more responsive to changes in illumination and camera motion.

Fast small moving objects are well segmented, but they leave behind short "trails" of pixels.

Objects that stop, and ghosts left behind by objects that start, gradually fade into the background.



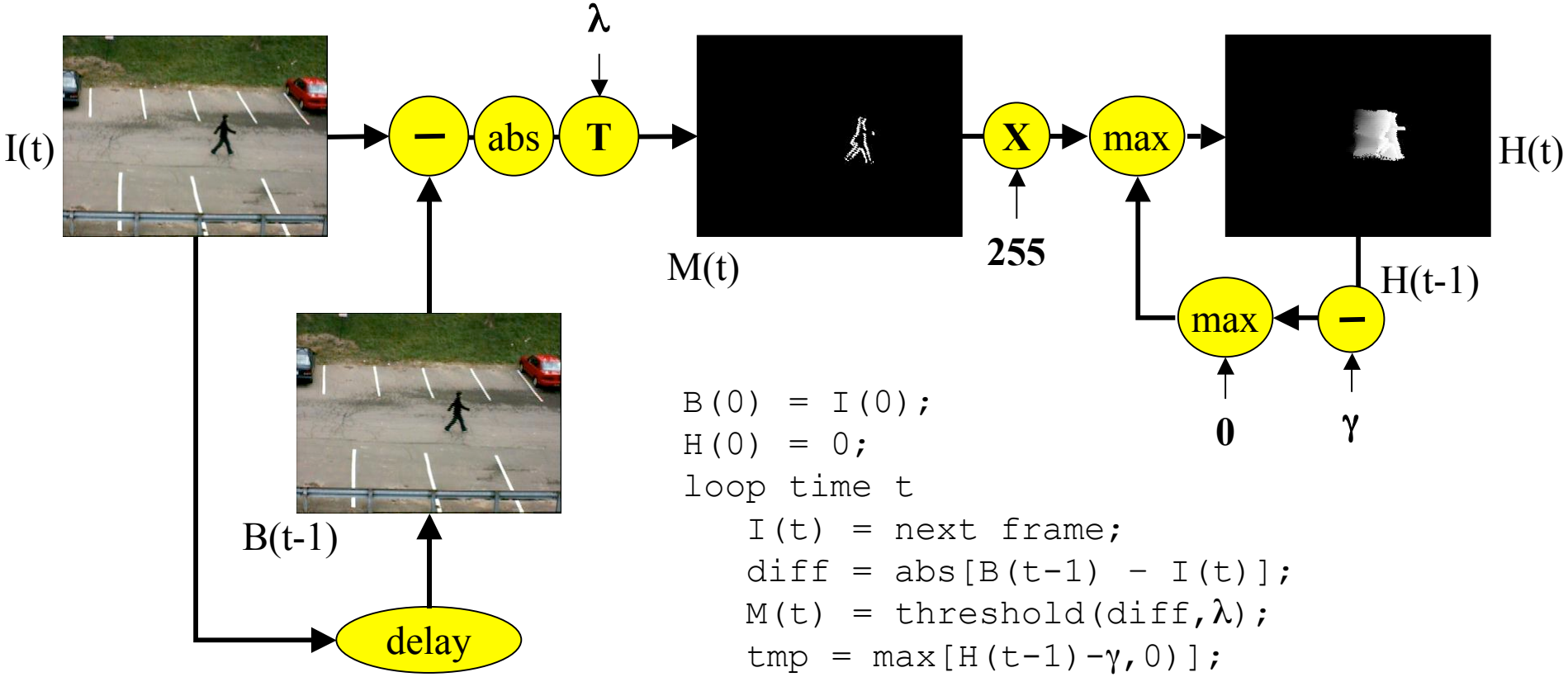
The centers of large slow moving objects start to fade into the background too! This can be "fixed" by decreasing the blend parameter A, but then it takes longer for stopped/ghost objects to disappear.



Robert Collins
CSE486, Penn State

Persistent Frame Differencing

- Motion images are combined with a linear decay term
- also known as motion history images (Davis and Bobick)




```
B(0) = I(0);
H(0) = 0;
loop time t
    I(t) = next frame;
    diff = abs[B(t-1) - I(t)];
    M(t) = threshold(diff,λ);
    tmp = max[H(t-1)-γ,0)];
    H(t) = max[255*M(t),tmp)];
    …
    B(t) = I(t);
end
```

Robert Collins
CSE486, Penn State

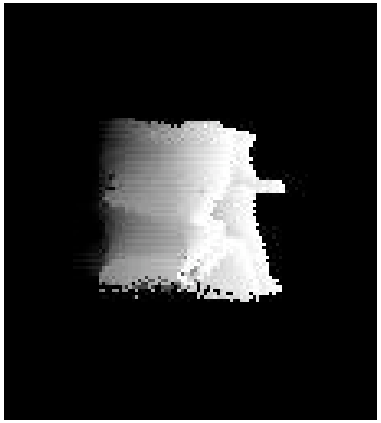
Persistant FD Results

movie

Robert Collins
CSE486, Penn State

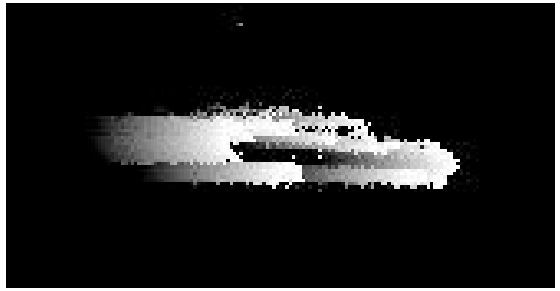
Persistant FD Observations

Persistant frame differencing is also responsive to changes in illumination and camera motion, and stopped objects / ghosts also fade away.



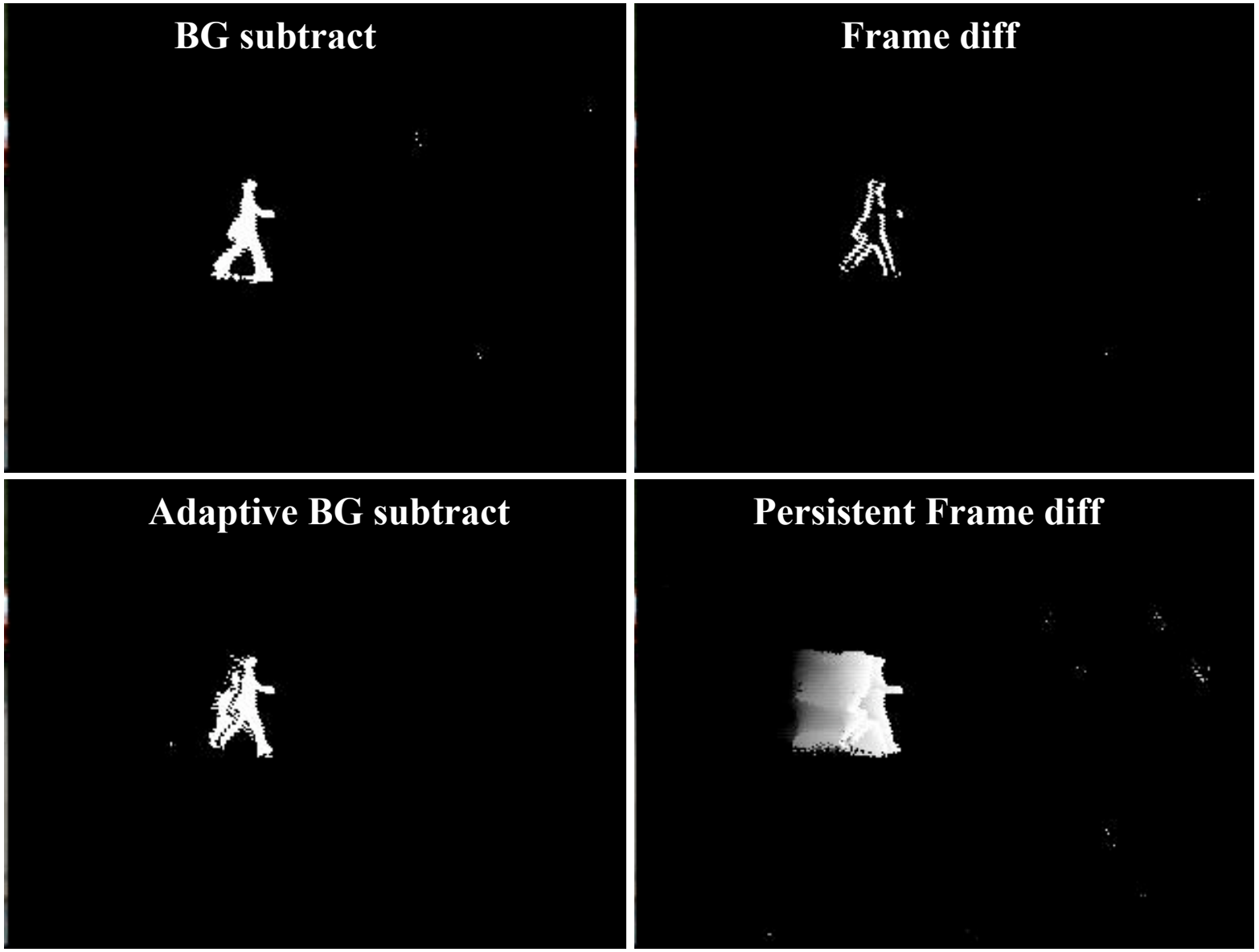
Objects leave behind gradually fading trails of pixels. The gradient of this trail indicates the apparent direction of object motion in the image.

Although the centers of uniformly colored objects are still not detected, the leading and trailing edges are make wider by the linear decay, so that perceptually (to a person) it is easier to see the whole object.



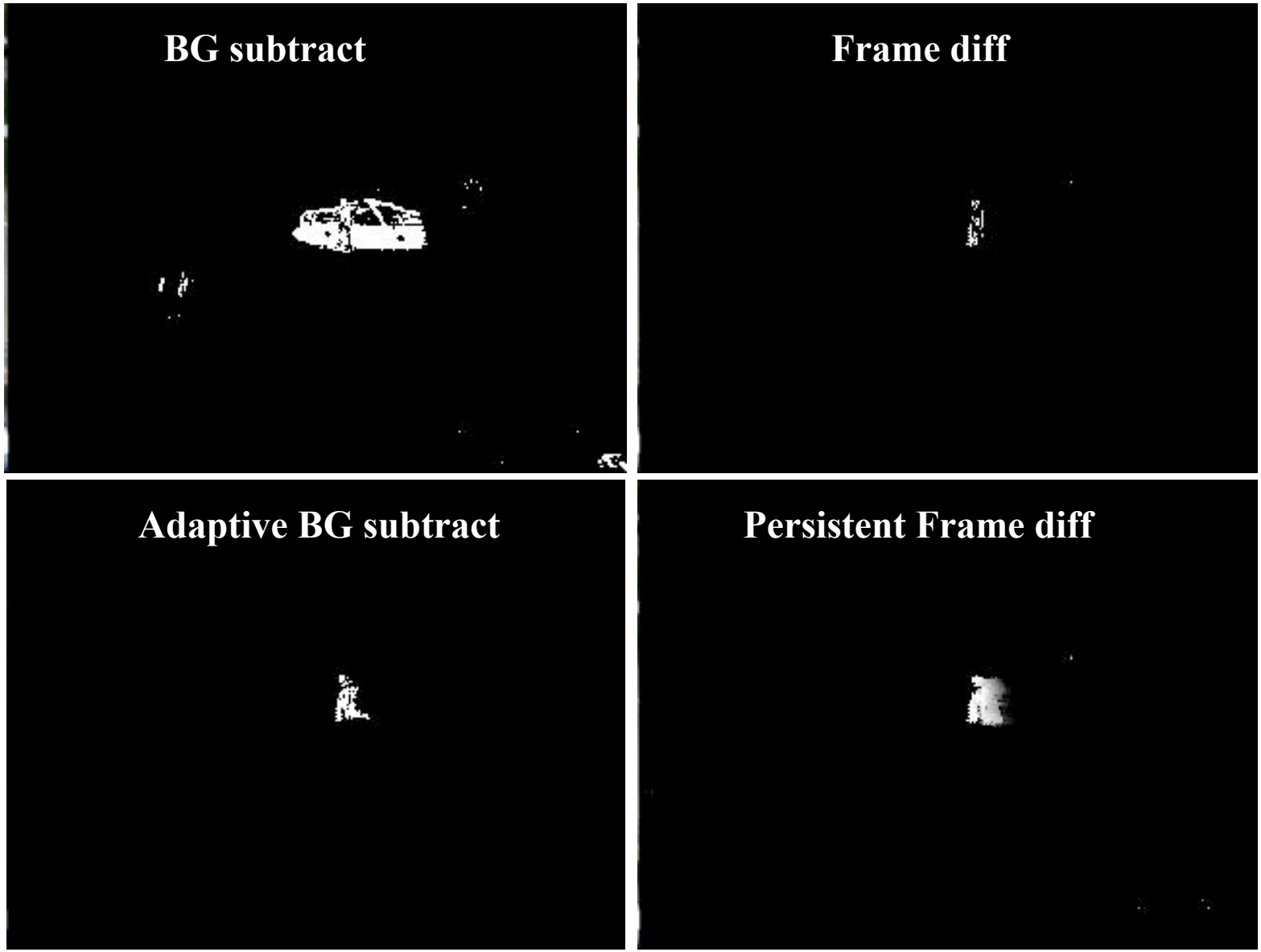
Robert Collins
CSE486, Penn State

Comparisons



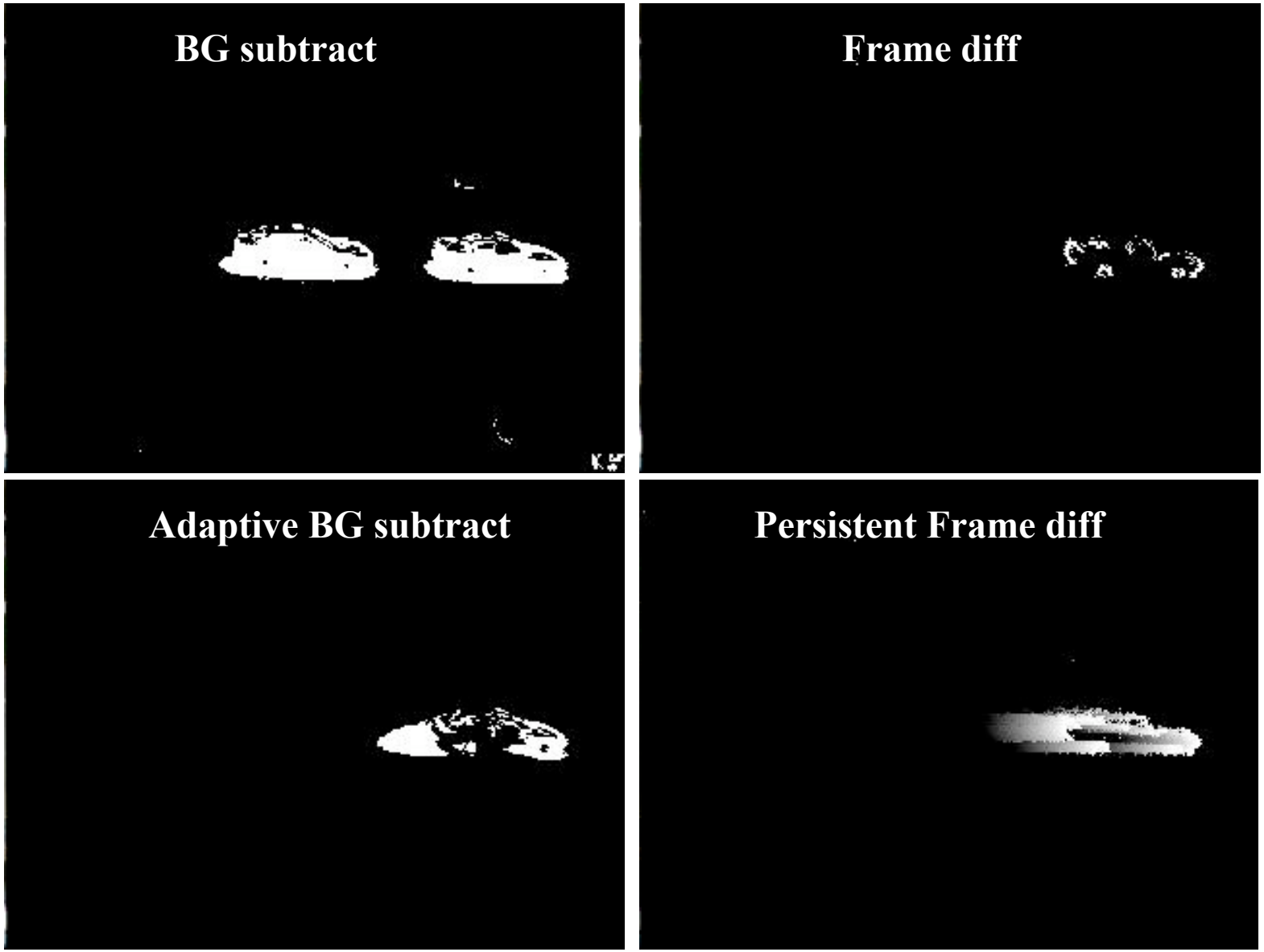
Robert Collins
CSE486, Penn State

Comparisons



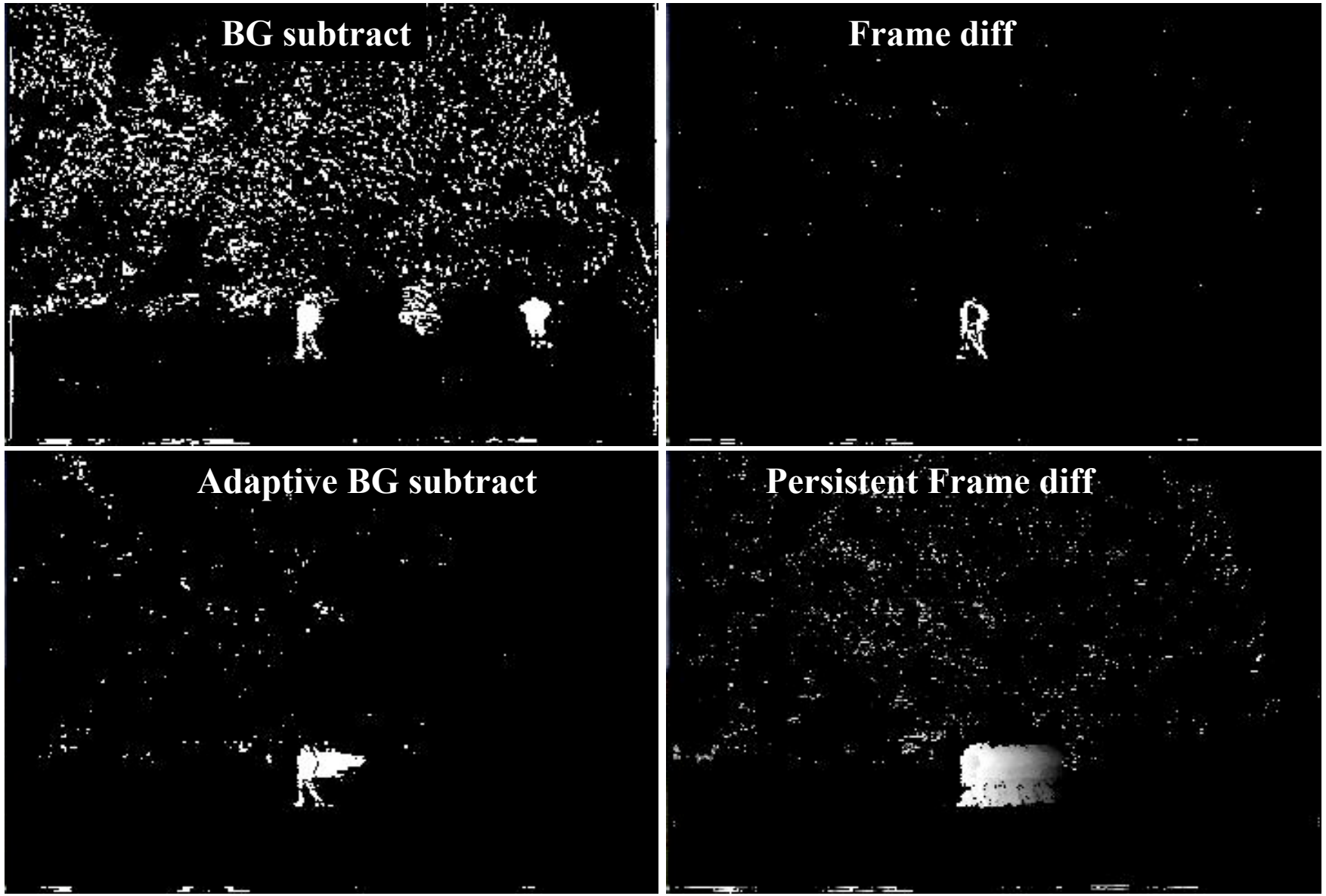
Robert Collins
CSE486, Penn State

Comparisons



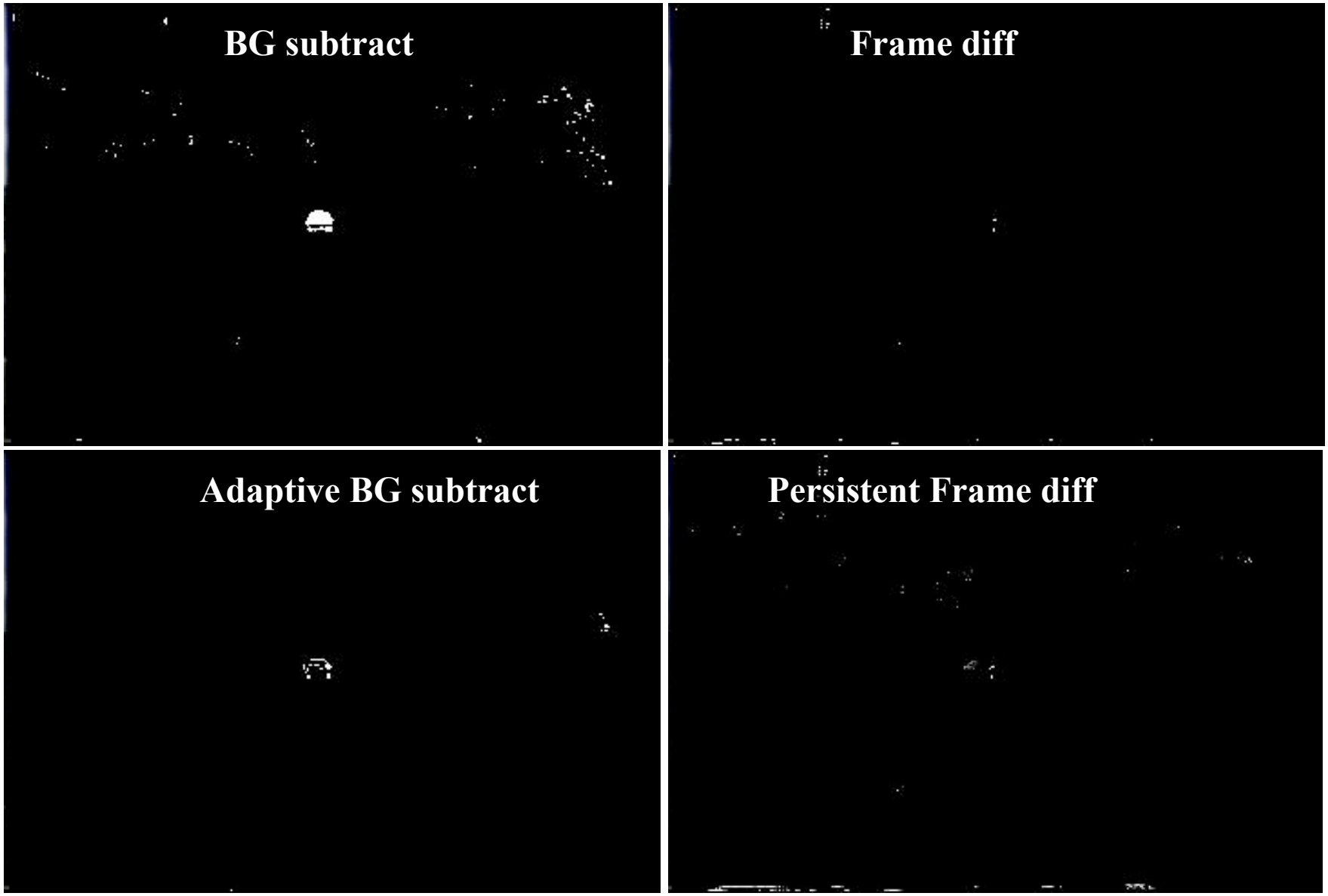
Robert Collins
CSE486, Penn State

Comparisons



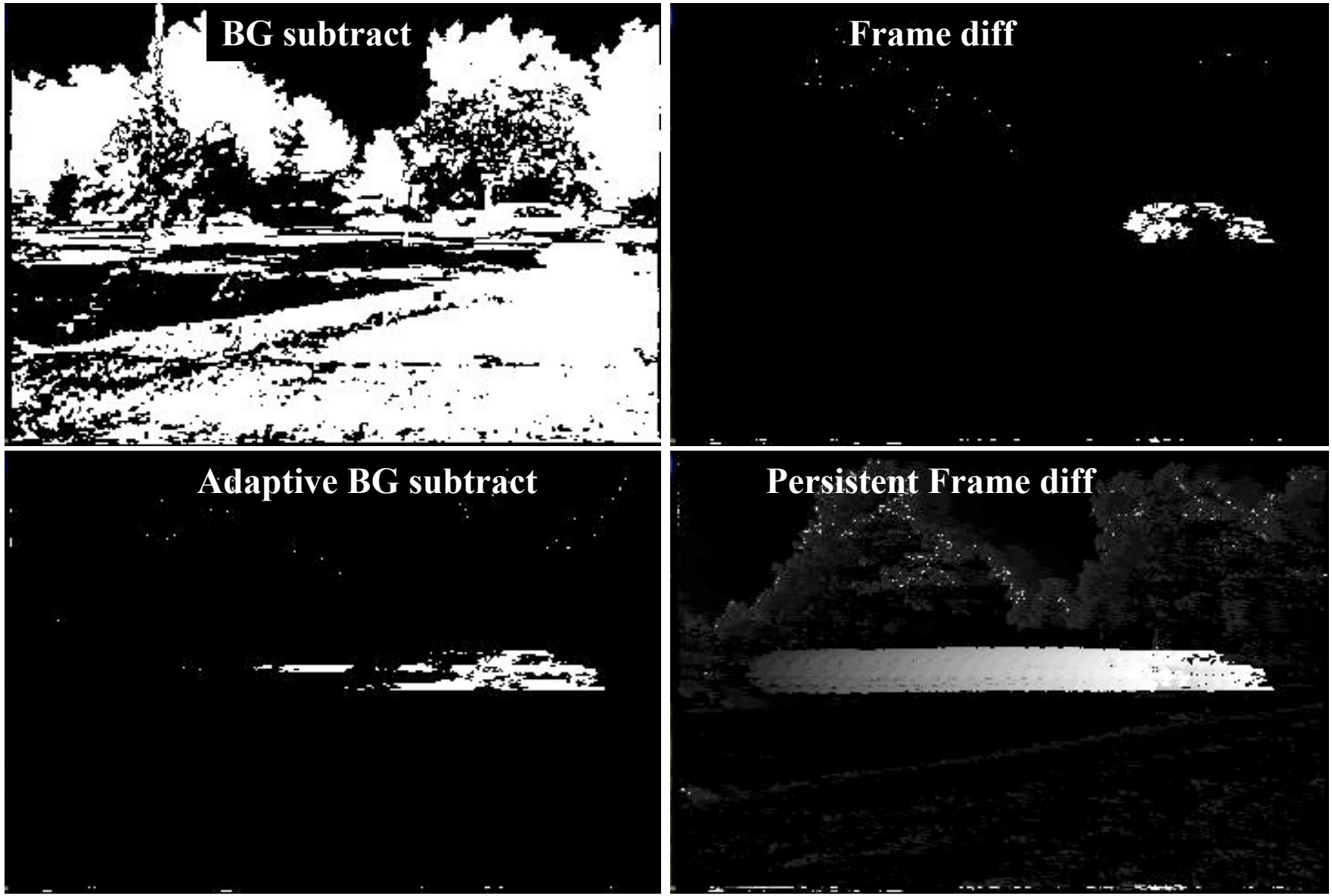
Robert Collins
CSE486, Penn State

Comparisons



Robert Collins
CSE486, Penn State

Comparisons



Robert Collins
CSE486, Penn State

Variants of Basic Background Subtraction

There are lots, and more papers every day.

Robert Collins
CSE486, Penn State

Statistical Background Modeling

Wren, Azarbayejani, Darrell, Pentland, "Pfinder: Real-time Tracking of the Human Body," IEEE Pattern Analysis and Machine Intelligence (PAMI), Vol. 19(7), July 1997, pp.780-785.

Note that $B(t) = \alpha I(t) + (1-\alpha) B(t-1)$
is an Exponential Moving Average IIR filter.
(A recursive approximation to the mean, with more recent samples weighted more highly).

Variant: Compute a mean AND covariance of colors at each pixel. The change detection threshold should now be statistical distance (aka Mahalanobis distance).

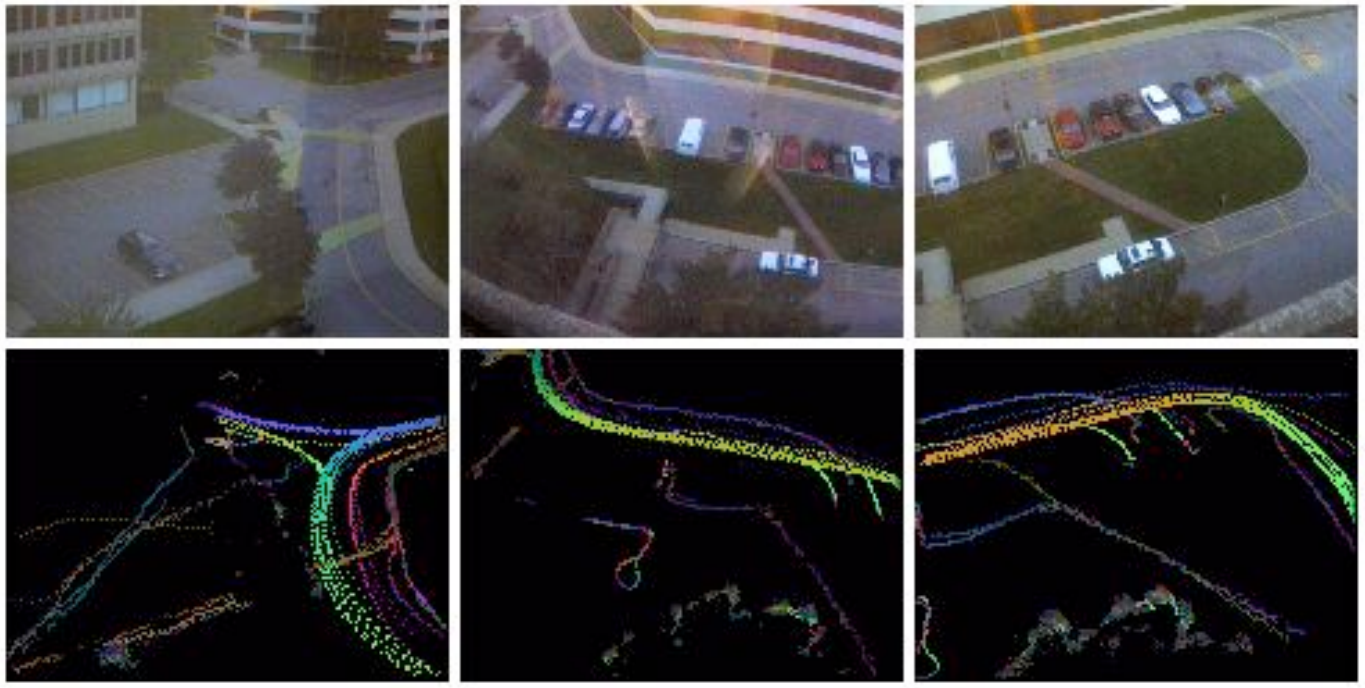
This typically works better, since the threshold adapts.
However, it is based on the assumption that the observed colors at a pixel are unimodal (works best if it is a Gaussian distribution).

Robert Collins
CSE486, Penn State

Statistical Background Modeling

Chris Stauffer and Eric Grimson, "Adaptive Background Mixture Models for Real-time Tracking," IEEE Computer Vision and Pattern Recognition (CVPR), June 1999, pp.246-252.

Observation: the distribution of colors at a pixel is often multi-modal (for example, areas of tree leaves, or ripples on water). They maintain an adaptive color model at each pixel based on a mixture of Gaussians (typically 5 or 6 components).



Statistical Background Modeling

Non-parametric color distribution, estimated via kernel density estimation

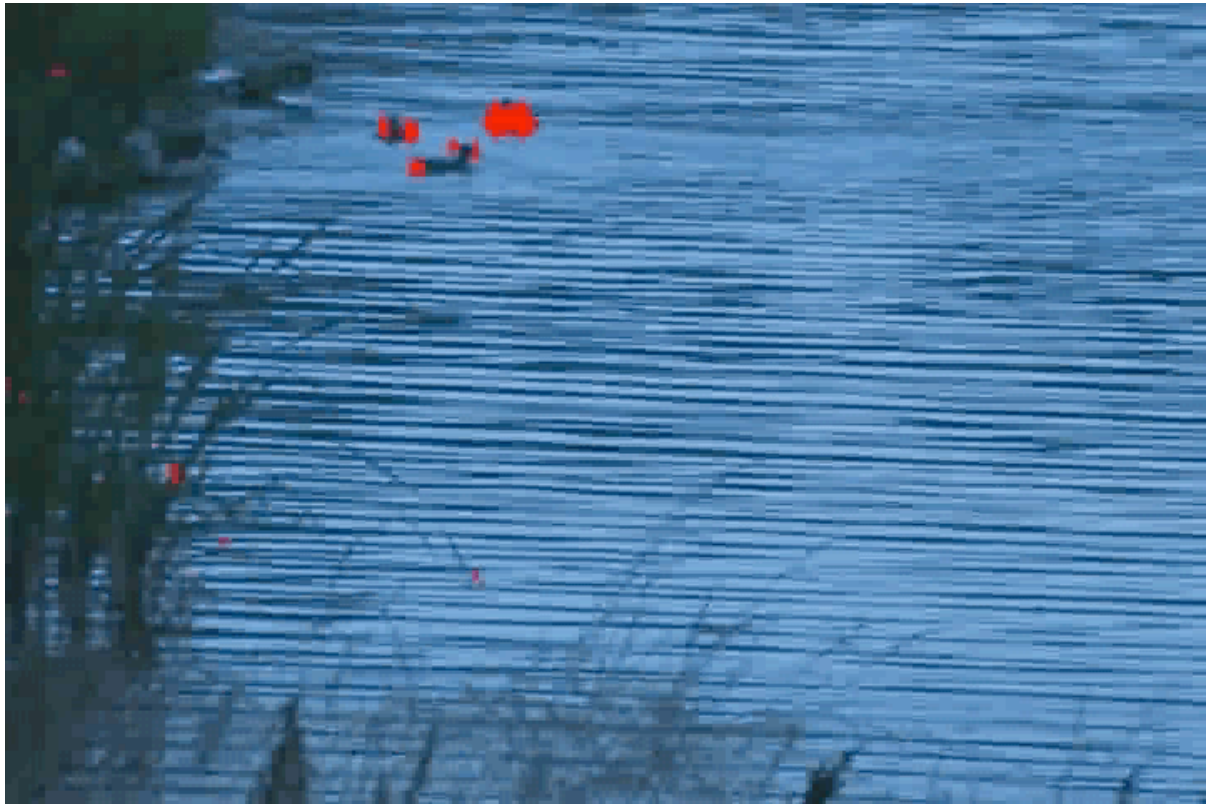
Ahmed Elgammal, David Harwood, Larry Davis "Non-parametric Model for Background Subtraction", 6th European Conference on Computer Vision. Dublin, Ireland, June 2000.

Robert Collins
CSE486, Penn State

Statistical Background Modeling

Use optic flow u,v values at each pixel, rather than intensity/color

R.Pless, J.Larson, S.Siebers, B.Westover, "Evaluation of Local Models of Dynamic Backgrounds," IEEE Computer Vision and Pattern Recognition (CVPR), June 2003



Robert Collins
CSE486, Penn State

Pan/Tilt Work-Arounds (cont)
Master-Slave Servoing

X.Zhou, R.Collins, T.Kanade, P.Metes, "A Master-Slave System to Acquire Biometric Imagery of Humans at a Distance," ACM SIGMM 2003 Workshop on Video Surveillance , Berkeley, CA, Nov 7, 2003, pp.113-120

- **Detect object using stationary master camera**
- **Compute pointing angle for nearby pan/tilt camera**

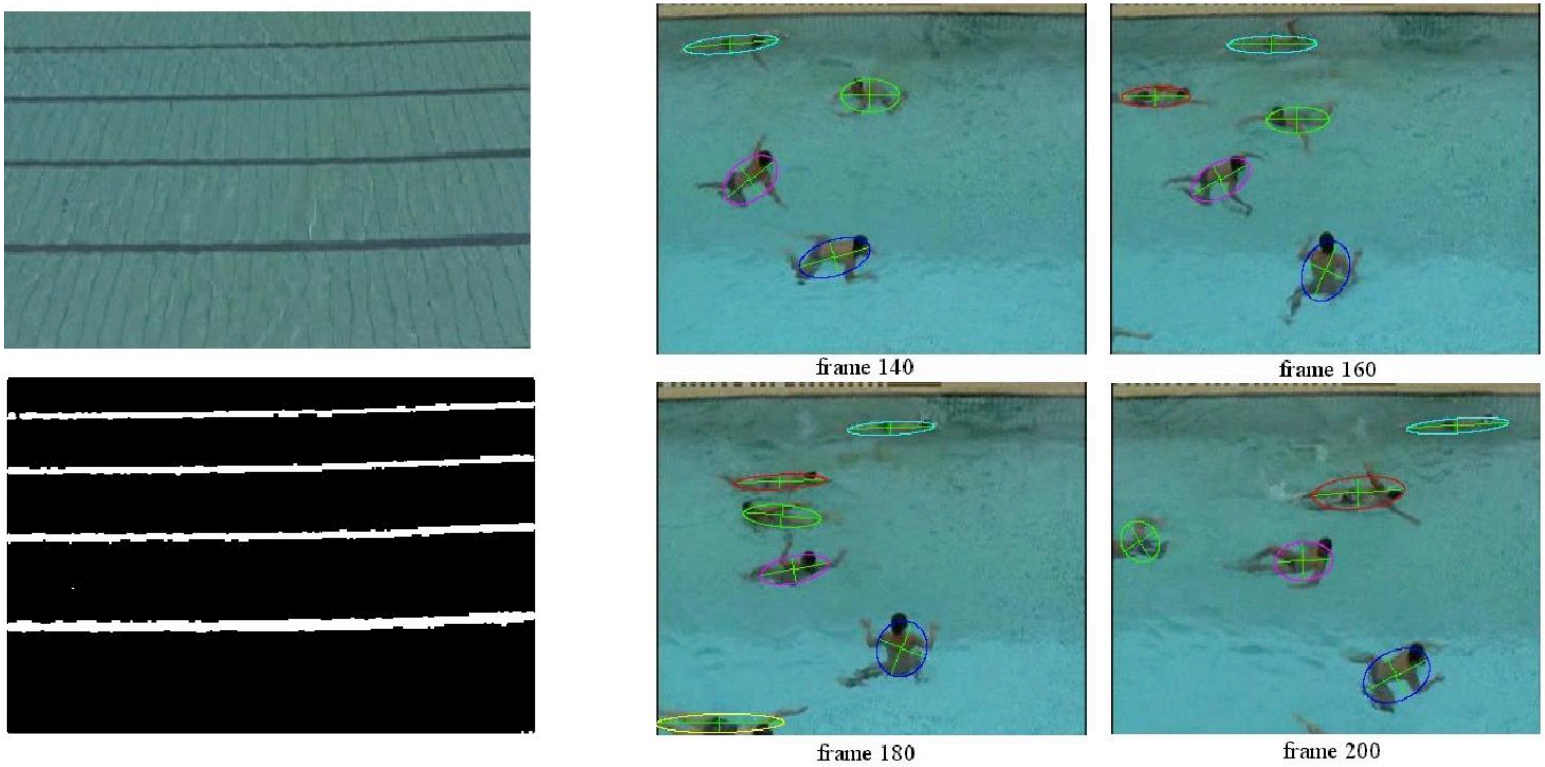
Stationary master

Servoing slave

Robert Collins
CSE486, Penn State

Background Color Models

Wenmiao Lu, Yap-Peng Tan, Weiyun Yau, "Video Surveillance System for Drowning Detection", Technical Sketches, IEEE Conference on Computer Vision and Pattern Recognition, Kauii Hawaii, Dec 2001.

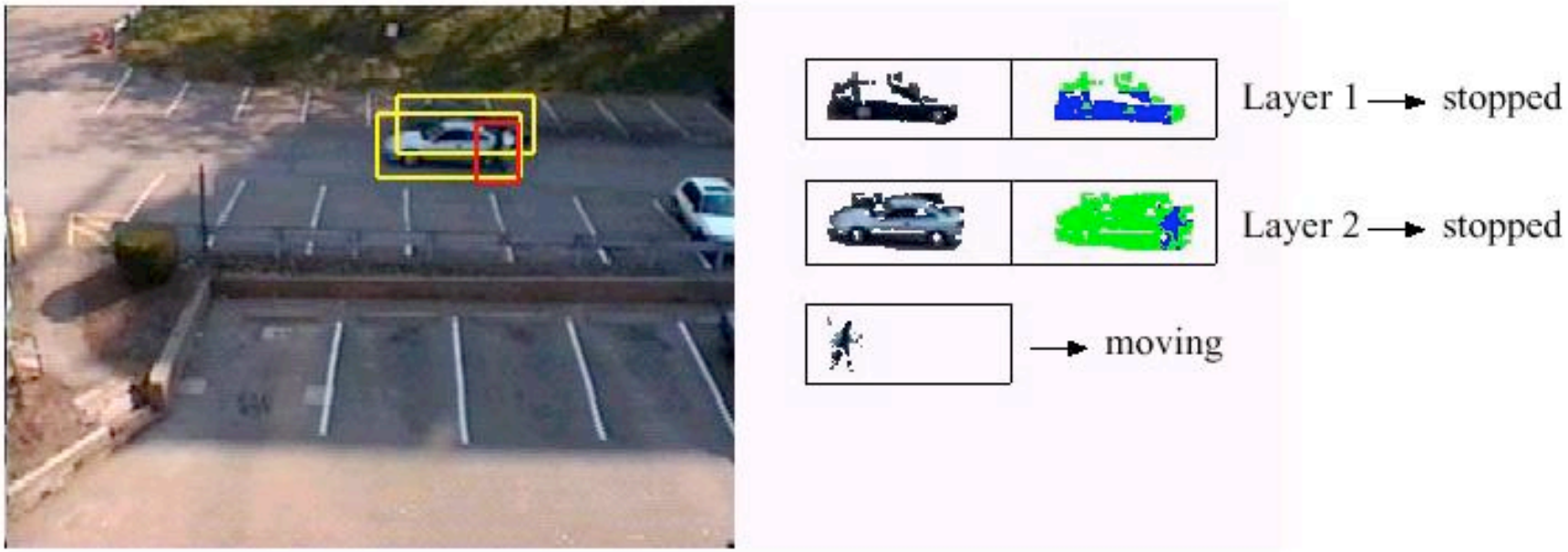


Color segmentation (two-color pool model) to detect swimmers while ignoring intensity changes due to ripples and waves in the water.

Robert Collins
CSE486, Penn State

Layered Detection

R.Collins, A.Lipton, H.Fujiyoshi and T.Kanade, "Algorithms for Cooperative Multi-Sensor Surveillance," Proceedings of the IEEE, Vol 89(10), October 2001, pp.1456-1477.



Allow blobs to be layered, so that stopped blobs can be considered part of the background for new object detection, but they will not leave behind ghosts when they start moving again.

Robert Collins
CSE486, Penn State

Pan/Tilt Work-Arounds

A major limitation of background subtraction is that the camera must be stationary, limiting our ability to do active tracking with a pan/tilt camera.

Work-around #1: Step-and-Stare Tracking

Robert Collins
CSE486, Penn State

Pan/Tilt Work-Arounds (cont)
Stabilization of Camera Motion

Frank Dellaert and Robert Collins, "Fast Image-Based Tracking by Selective Pixel Integration," ICCV Workshop on Frame-Rate Vision, Corfu, Greece, Sept 1999.

Video in | Reference view | Warped video | Subtraction

Apparent motion of a panning/tilting camera can be removed by warping images into alignment with a collection of background reference views.

There is also work on stabilization more general camera motions when the scene structure is predominately planar (Sarnoff papers; also Michal Irani).

Robert Collins
CSE486, Penn State

Pan/Tilt Work-Arounds (cont)
Master-Slave Servoing

X.Zhou, R.Collins, T.Kanade, P.Metes, "A Master-Slave System to Acquire Biometric Imagery of Humans at a Distance," ACM SIGMM 2003 Workshop on Video Surveillance , Berkeley, CA, Nov 7, 2003, pp.113-120

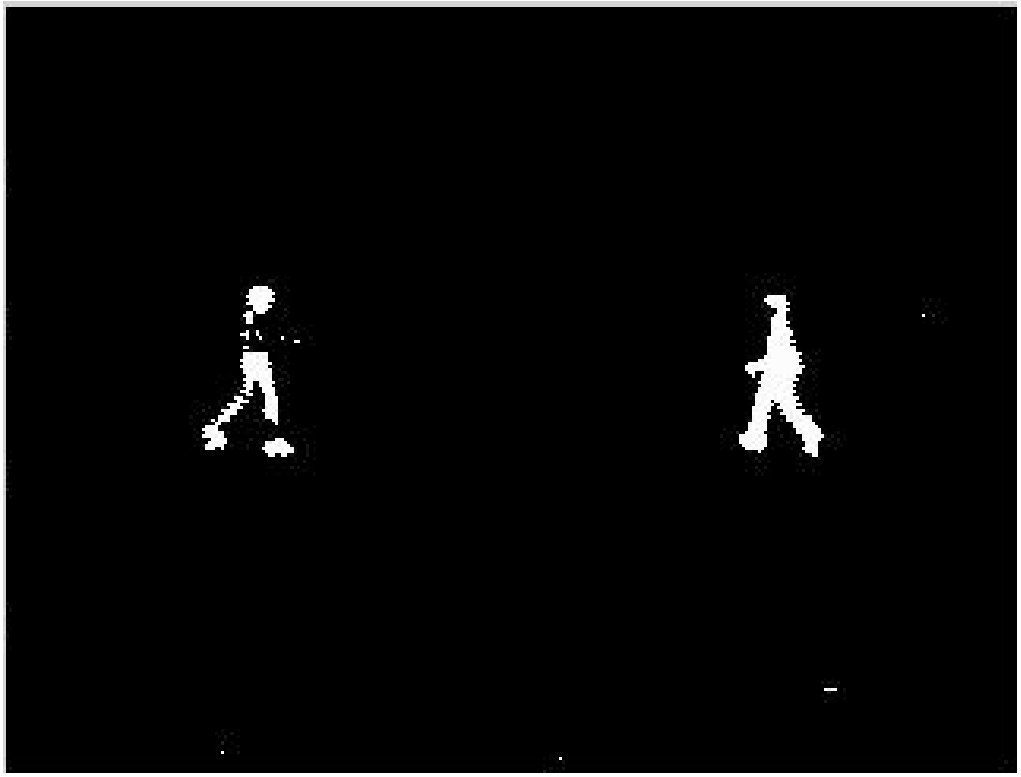
- **Detect object using stationary master camera**
- **Compute pointing angle for nearby pan/tilt camera**

Stationary master

Servoing slave

Robert Collins
CSE486, Penn State

Grouping Pixels into Blobs



Motivation: change detection is a pixel-level process.
We want to raise our description to a higher level of abstraction.

Robert Collins
CSE486, Penn State

Motivation

If we have an object-level blob description, AND we can maintain a model of the scene background, then we can artificially remove objects from the video sequence.

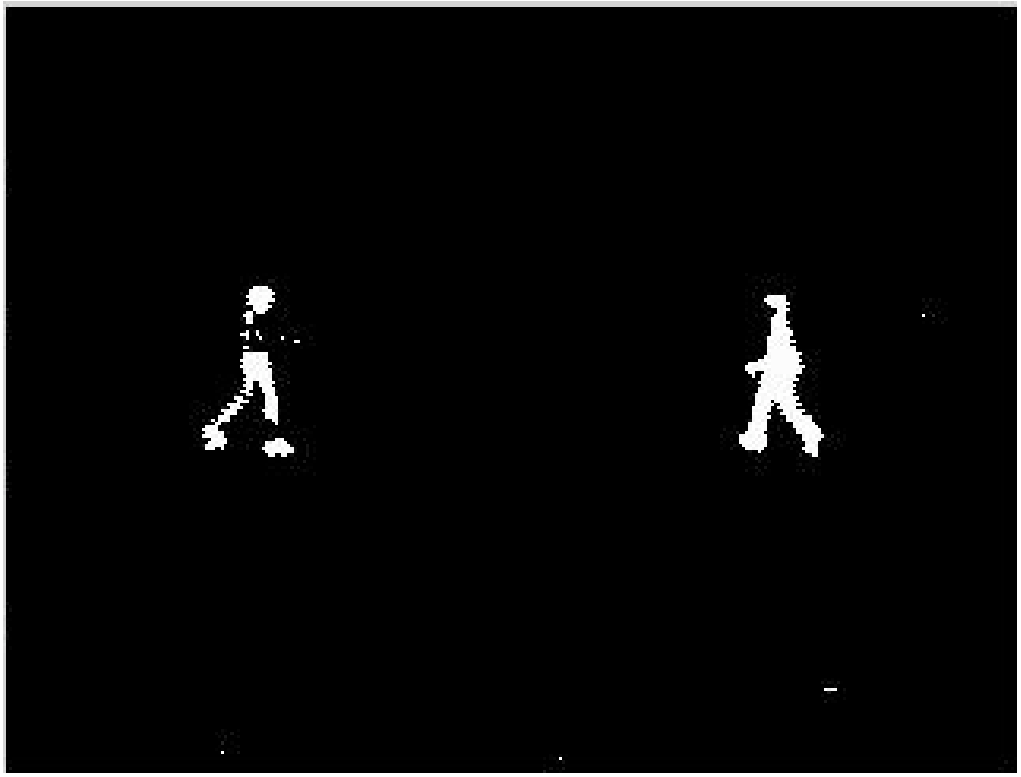
Future video game idea?

movie

Alan Lipton, "Virtual Postman – Real-time, Interactive Virtual Video," IASTED CGIM, Palm Springs, CA, Octover 1999.

Robert Collins
CSE486, Penn State

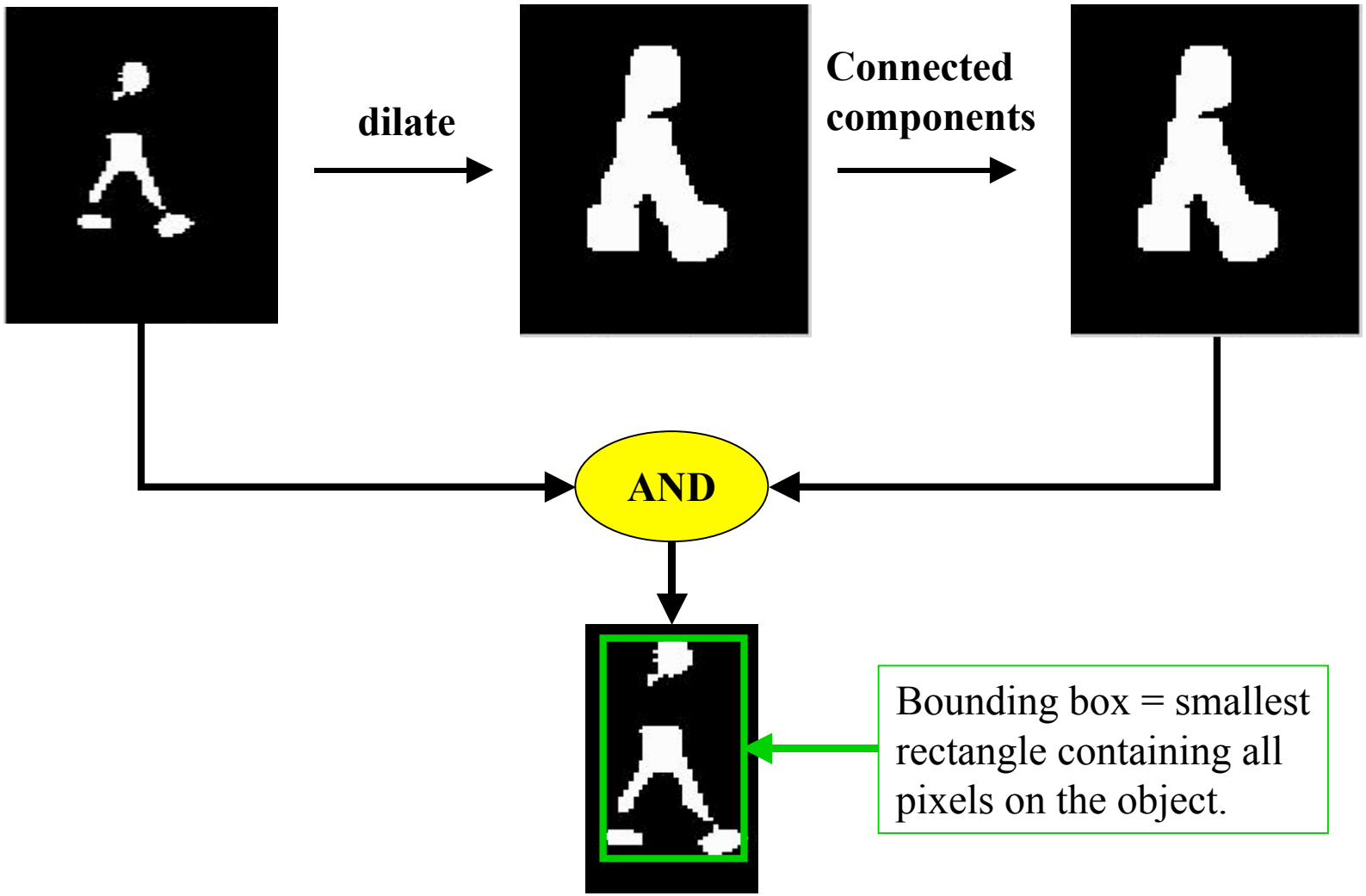
Grouping Pixels into Blobs



- median filter to remove noisy pixels
- connected components (with gap spanning)
- Size filter to remove small regions

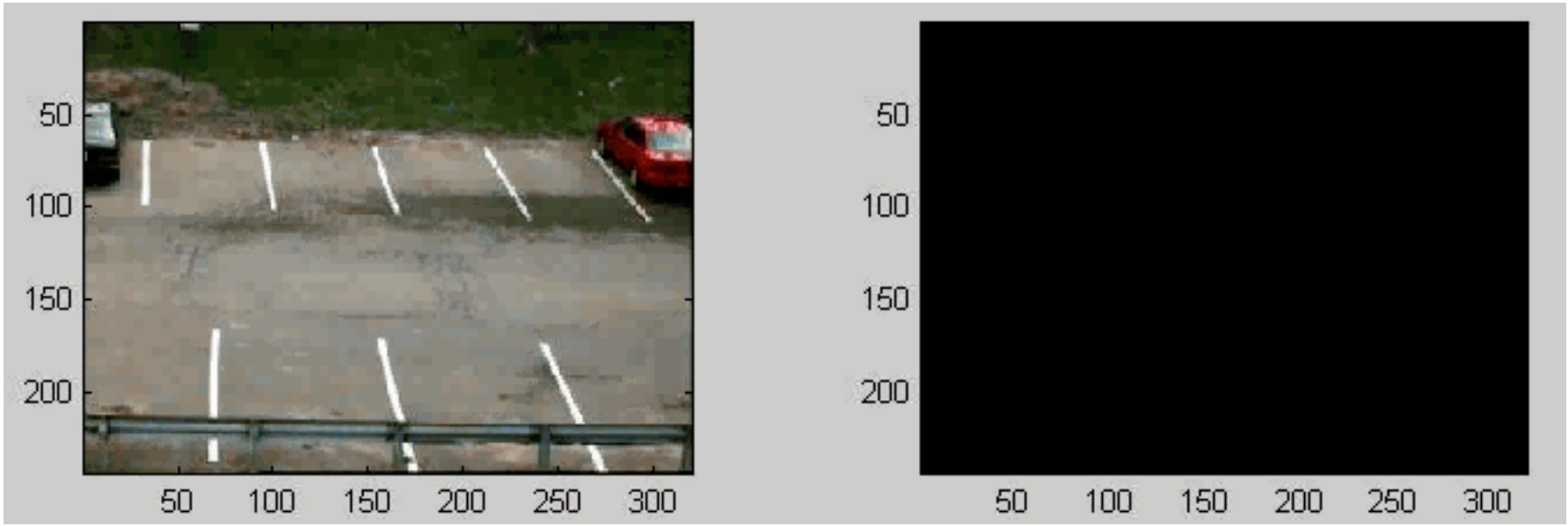
Robert Collins
CSE486, Penn State

Gap Spanning Connected Components



Robert Collins
CSE486, Penn State

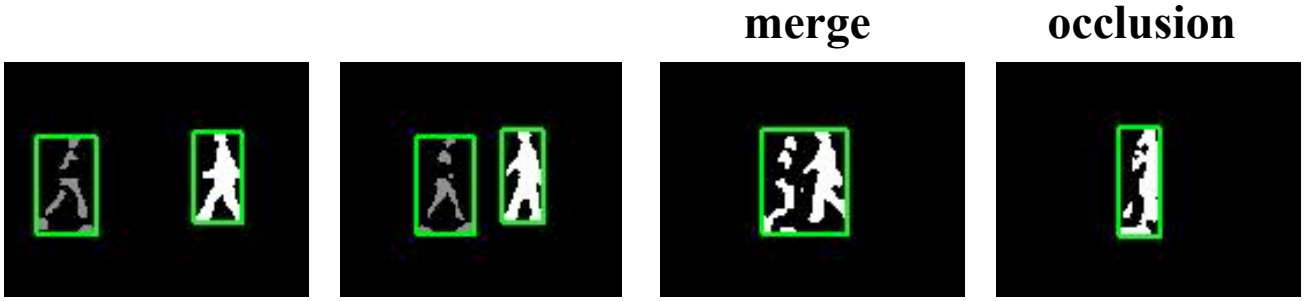
Detected Blobs

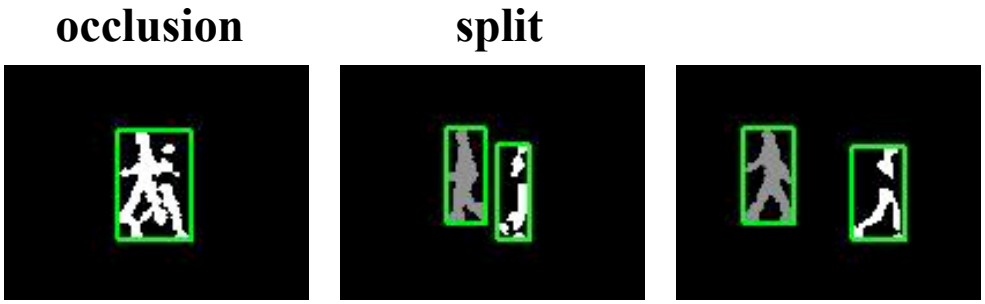


movie

Robert Collins
CSE486, Penn State

Blob Merge/Split





When two objects pass close to each other, they are detected as a single blob. Often, one object will become occluded by the other one. One of the challenging problems is to maintain correct labeling of each object after they split again.

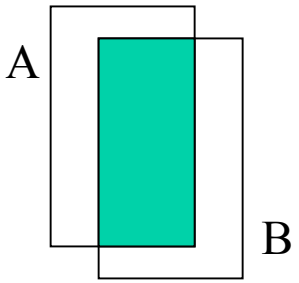
Robert Collins
CSE486, Penn State

Data Association

Determining the correspondence of blobs across frames is based on feature similarity between blobs.

Commonly used features: location , size / shape, velocity, appearance

For example: location, size and shape similarity can be measured based on bounding box overlap:



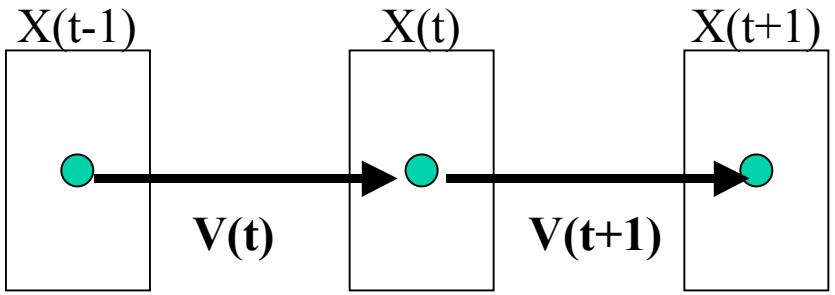
$$\textbf{score} = \frac{\textbf{2 * area(A and B)}}{\textbf{area(A) + area(B)}}$$

A = bounding box at time t
B = bounding box at time t+1

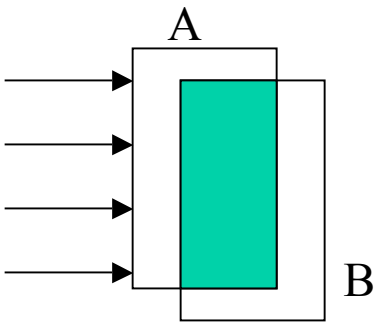
Robert Collins
CSE486, Penn State

Data Association

It is common to assume that objects move with constant velocity



constant velocity
assumes V(t) = V(t+1)



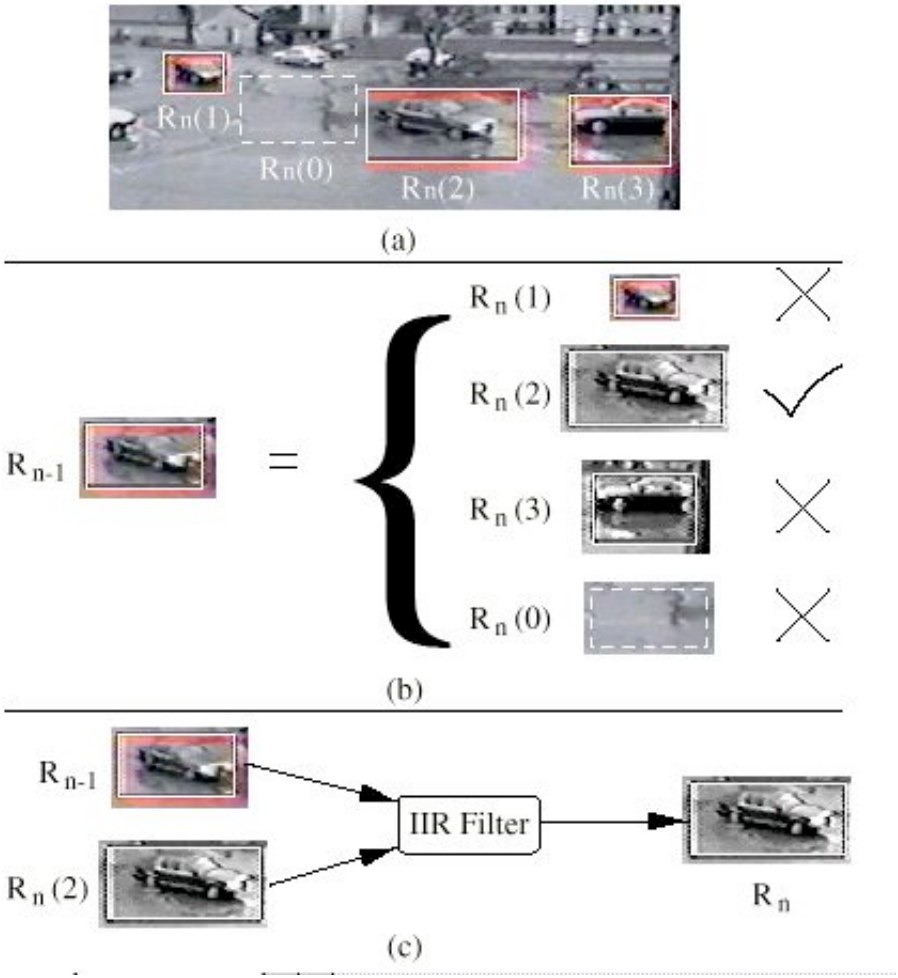
$$\text{score} = \frac{2 * \text{area(A and B)}}{\text{area(A)} + \text{area(B)}}$$

A = bounding box at time t, adjusted by velocity V(t)
B = bounding box at time t+1

Robert Collins
CSE486, Penn State

Appearance Information

Correlation of image templates is an obvious choice (between frames)



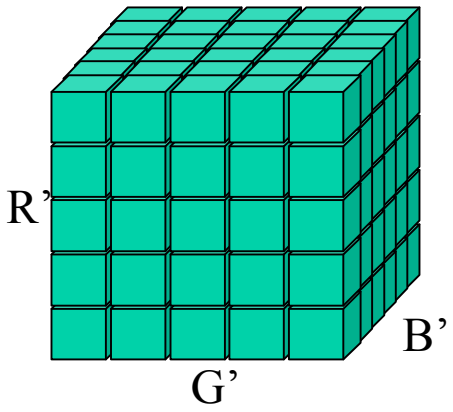
Extract blobs

Data association via normalized correlation.

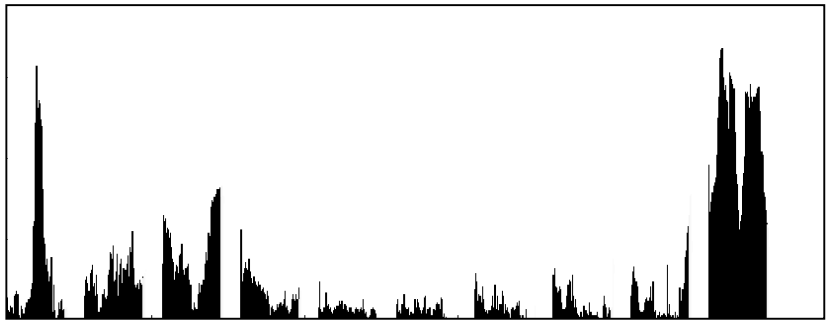
Update appearance template of blobs

Robert Collins
CSE486, Penn State

Appearance via Color Histograms



discretize



Color distribution (1D histogram normalized to have unit weight)

R’ = R << (8 - nbits)
G’ = G << (8 - nbits)
B’ = B << (8 - nbits)

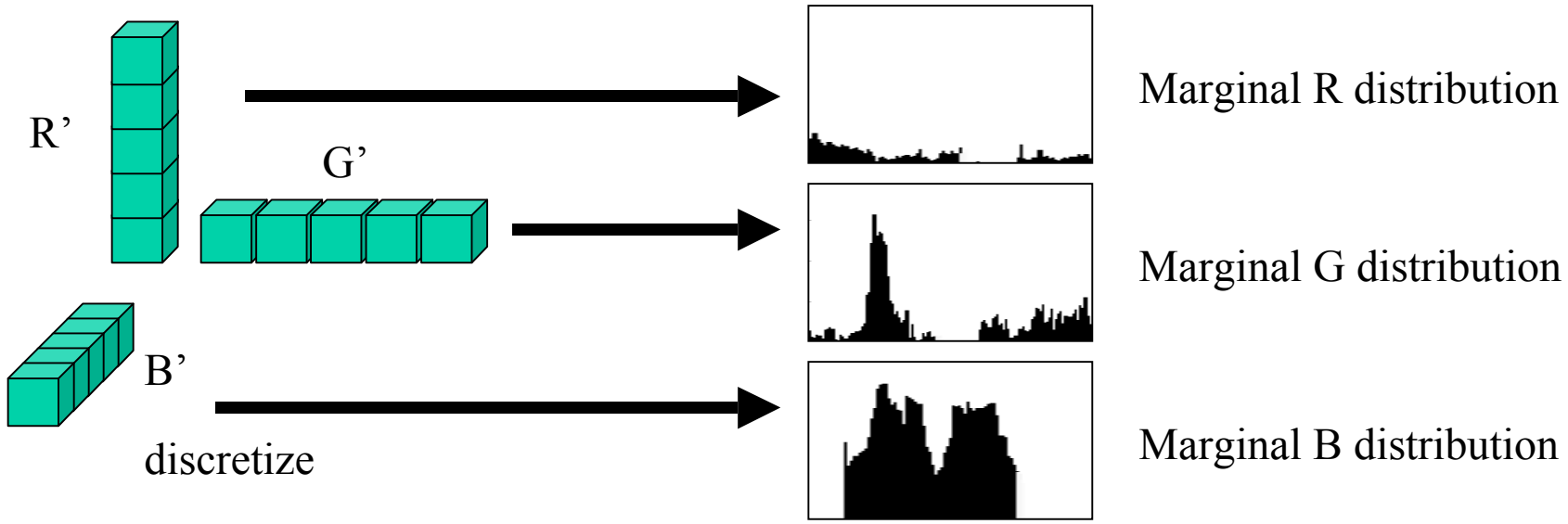
Total histogram size is (2^(8-nbits))^3

example, 4-bit encoding of R,G and B channels yields a histogram of size 16*16*16 = 4096.

Robert Collins
CSE486, Penn State

Smaller Color Histograms

Histogram information can be much much smaller if we are willing to accept a loss in color resolvability.



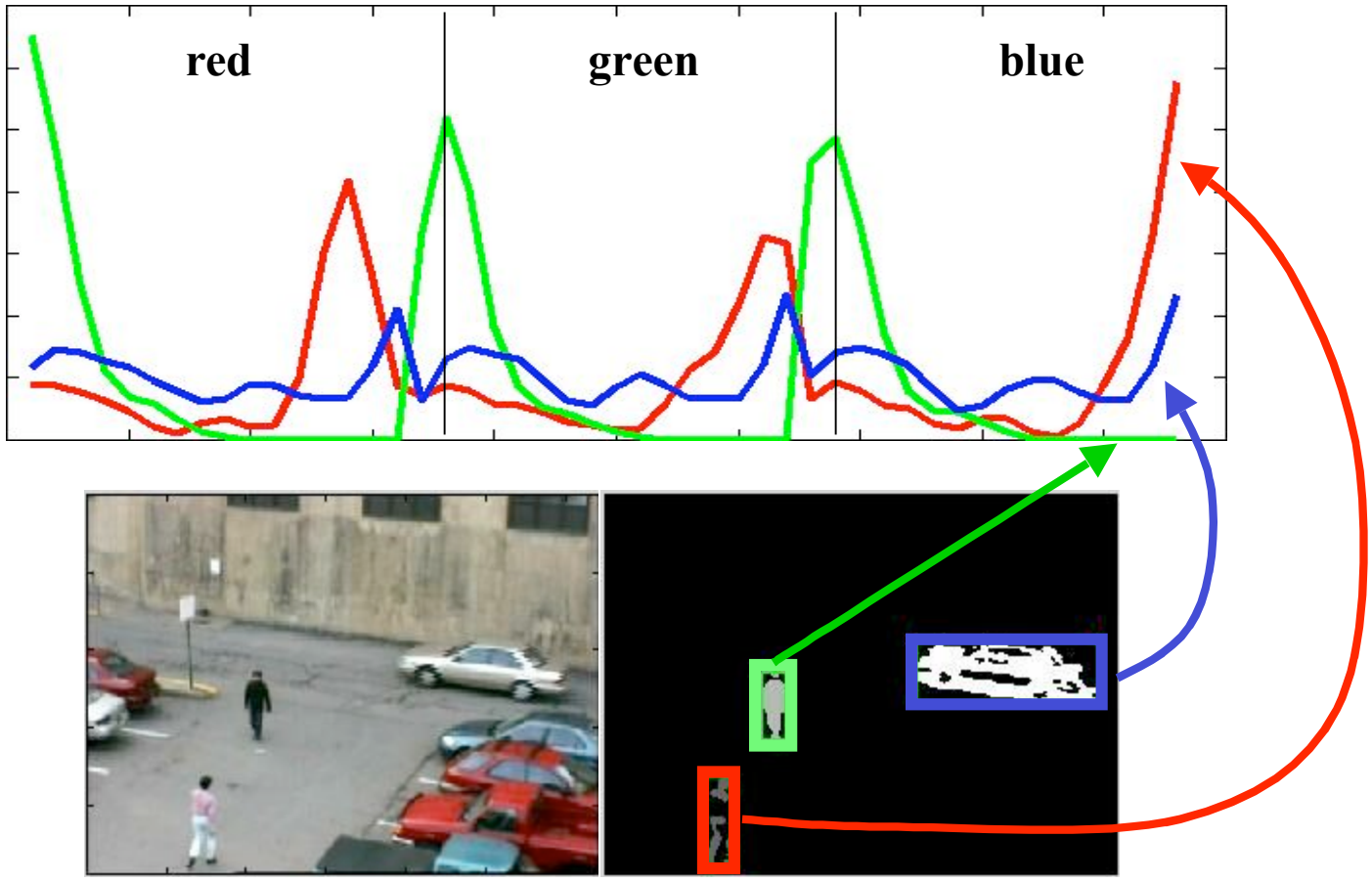
R' = R << (8 - nbits)
G' = G << (8 - nbits)
B' = B << (8 - nbits)

Total histogram size is 3*(2^(8-nbits))

example, 4-bit encoding of R,G and B channels yields a histogram of size 3*16 = 48.

Robert Collins
CSE486, Penn State

Color Histogram Example



Robert Collins
CSE486, Penn State

Comparing Color Distributions

Given an n-bucket model histogram $\{m_i \mid i=1,\ldots,n\}$ and data histogram $\{d_i \mid i=1,\ldots,n\}$, we follow Comanesciu, Ramesh and Meer * to use the distance function:

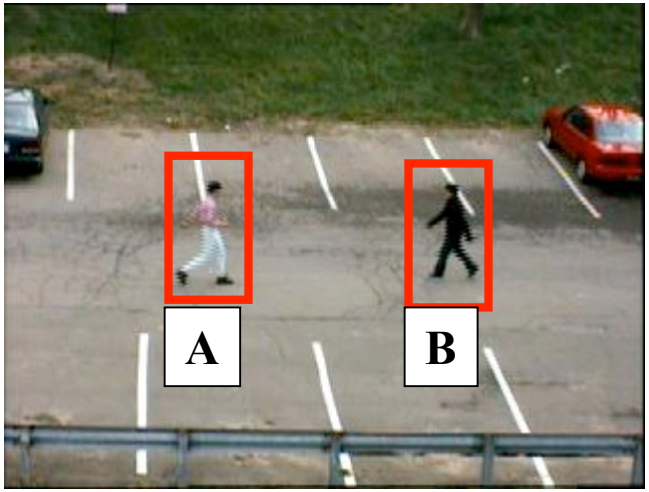
$$\Delta(m,d) = \sqrt{1 - \sum_{i=1}^{n} \sqrt{m_i \times d_i}}$$

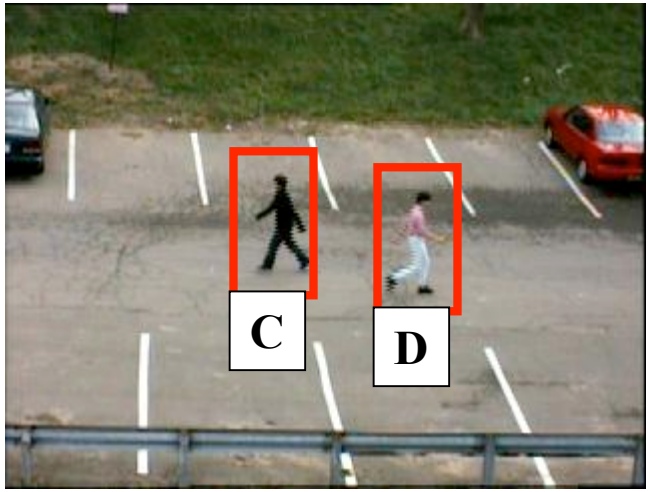
Why?
1) it shares optimality properties with the notion of Bayes error
2) it imposes a metric structure
3) it is relatively invariant to object size (number of pixels)
4) it is valid for arbitrary distributions (not just Gaussian ones)

*Dorin Comanesciu, V. Ramesh and Peter Meer, "Real-time Tracking of Non-Rigid Objects using Mean Shift," IEEE Conference on Computer Vision and Pattern Recognition, Hilton Head, South Carolina, 2000 (best paper award).

Robert Collins
CSE486, Penn State

Example of Data Association After Merge and Split





$\Delta(A,C) = 2.03$
$\Delta(A,D) = 0.39$ ●

A -> D

$\Delta(B,C) = 0.23$ ●
$\Delta(B,D) = 2.0$

B -> C