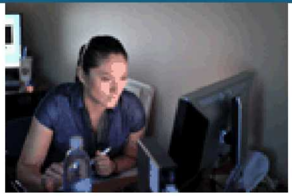
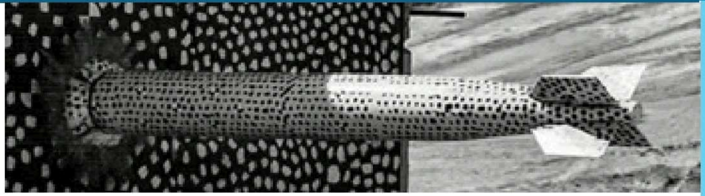
SAND2020-7950PE

Introduction to Remote Sensing Object Detection

Presenter: Tian J. Ma

Distinguished Member of R&D Staff

Data Analysis and Exploitation Department

Sandia National Laboratories, Albuquerque, New Mexico

Email: tma@sandia.gov

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

OUTLINE

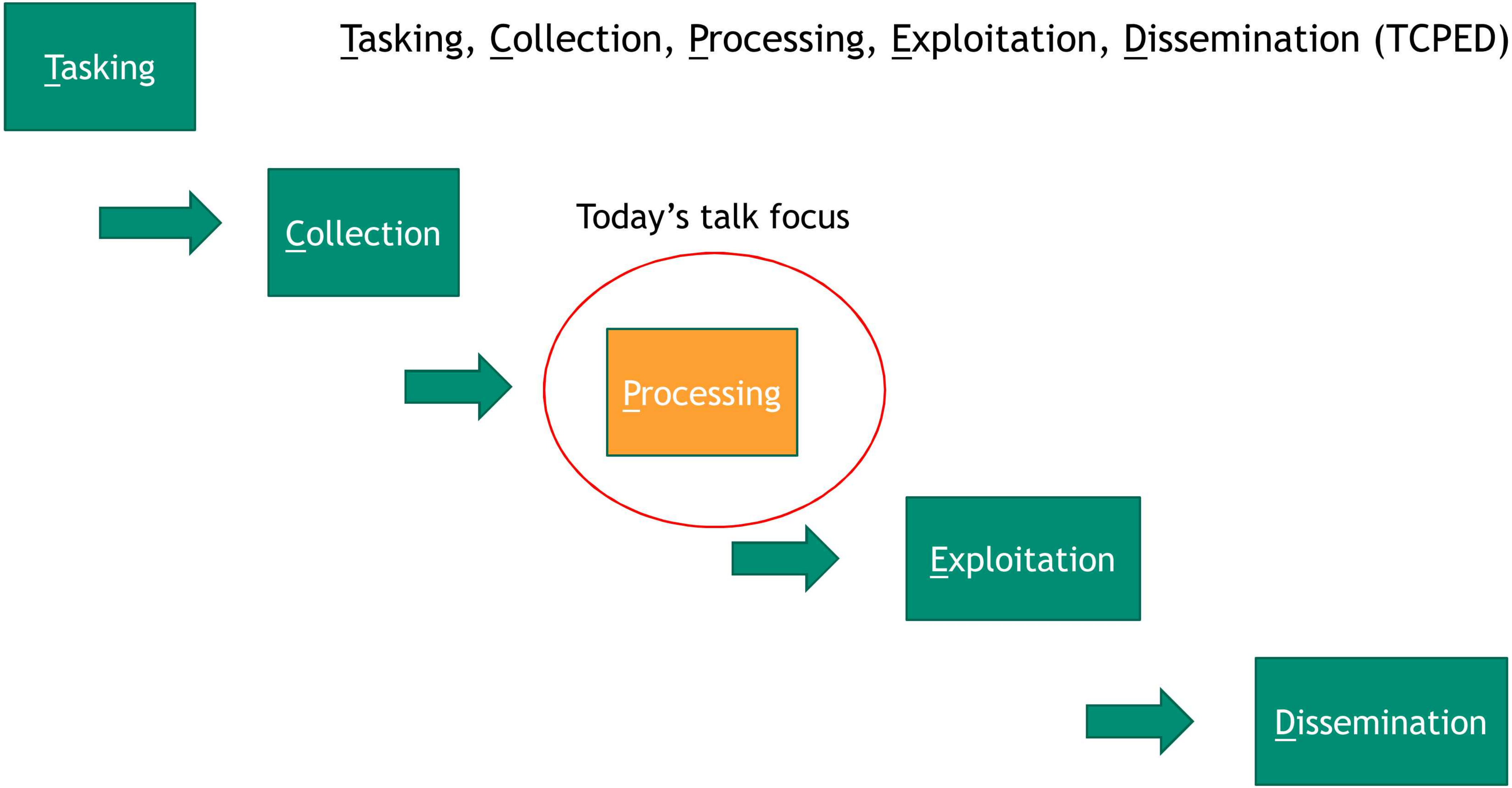
- Overview
- General Framework for Remote Sensing System
- General Framework Detection Processing
- General Algorithm Processing
- Examples From Applications
- Further Reading

Overview

- Real-time remote object detection plays a critical role in Sandia's Global Security Mission.
- In real-time remote sensing applications, frames of data are continuously flowing into the system.
- The capability of detecting objects of interests and tracking them as they move is critical to many critical and challenging national missions.
- Some common applications in this field include: home/business surveillance, environmental monitoring, autonomous sensing, etc..
- This talk will provide audiences a general understanding of the remote object detection problem as well as provide key algorithms and techniques used to solve these types of problems.

General Framework for Remote Sensing System

Tasking, Collection, Processing, Exploitation, Dissemination (TCPED)



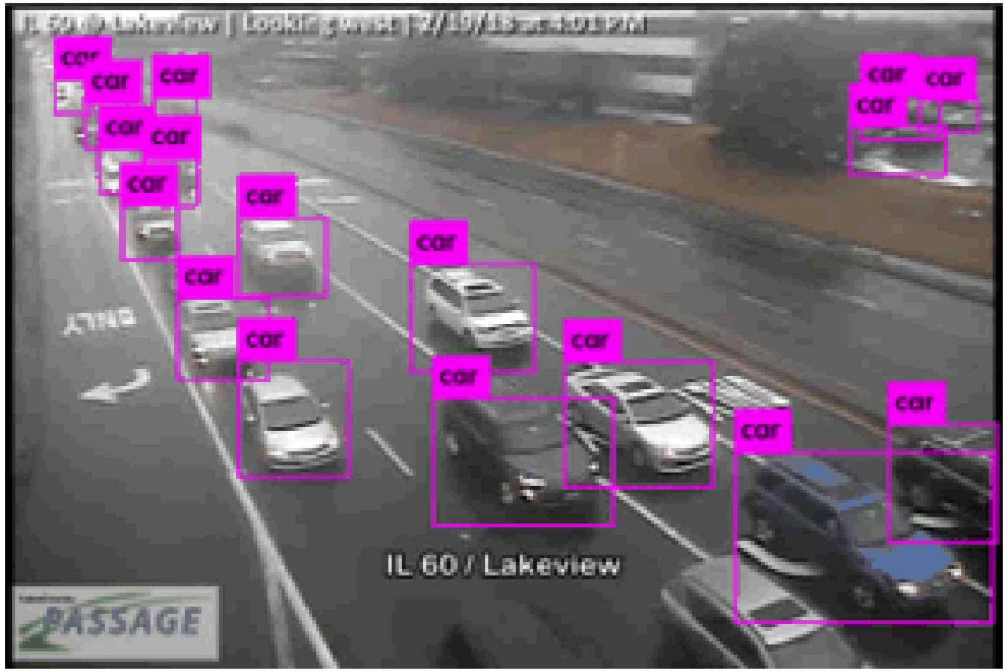
Artificial Intelligence (AI) Processing

- Machine learning and Deep Learning Techniques

Image Frame → AI Object Detector (e.g. YOLO) → Detected Objects

- Advantages
  - Easy to get started (many tools available: TensorFlow, Caffe, PyTorch, etc..)
  - High accuracy
  - Fast decision (operates on one image frame)
    - Populate methods: You Only Look Once (YOLO), Mask R-CNN
- Disadvantages
  - Requires a large number training labels (i.e. usually thousands of examples per target class)
  - Limited explainability
  - Vulnerability (pixel attack)

YOLO Vehicle Detector



Results produced by students from UIUC during the 2018 SNL/UIUC/ARI internship program

Traditional Detection Processing (today's main focus)

Image

Background Estimation and Subtraction

Noise Estimation

Foreground/Background Discrimination

Additional Filtering

Detected Object

Key Advantages:

- Algorithm does not require pre-trained label
- Explainable (strong mathematic and statistical principles)

Disadvantages:

- May requires multiple frames to drive down false alarm rate

Background Estimation and Subtraction

Let $B(t-1)$, represents the background image estimated prior to time t, and $F(t)$, represents the image frame at time t.

To detect new energy at time t, subtract from the frame taken at t, an estimate of the “background” energy in the scene prior to this time

The background subtracted image $D(t)$, is expressed as: $D(t) = F(t) - B(t-1)$

Raw Image

Background Subtracted Image

Background Estimation Algorithm – Pixel Based Techniques

- Pixel-based (tracking temporal change of pixels)
  - Advantages: fast and scalable, relatively straight forward to understand and implement
  - Drawbacks: not very effective in clutter reduction (sensitive to environmental changes, i.e. illumination)
- Common Methods
  - 1st Order Difference – Subtract the current frame from the previous frame

$$1^{st}\text{ Diff }(i,j,t) = Frame\,(i,j,t) - Frame\,(i,j,t-1)$$

  - Mean Difference – Subtract the current frame from the average of the n previous frames

$$\bar{F} = \frac{1}{n}\sum_{k=1}^{n} F_k$$

$$\text{Mean Diff }(i,j,t) = F(i,j,t) - \bar{F}(\mathrm{i},\mathrm{j},\mathrm{t}-1)$$

$k < t$, where t, correspond to the time at the current frame

Background Estimation Algorithm – Pixel Based Techniques (2)

- Infinite Impulse Response (IIR) Low Pass Filter, also known as exponential smoothing, or gaussian average model
- Initialization – initialize background with initial n number of frames

$$B_0 = \bar{F} = \frac{1}{n}\sum_{k=1}^{n} F_k$$

$k < t$, where t, correspond to the time at the current frame

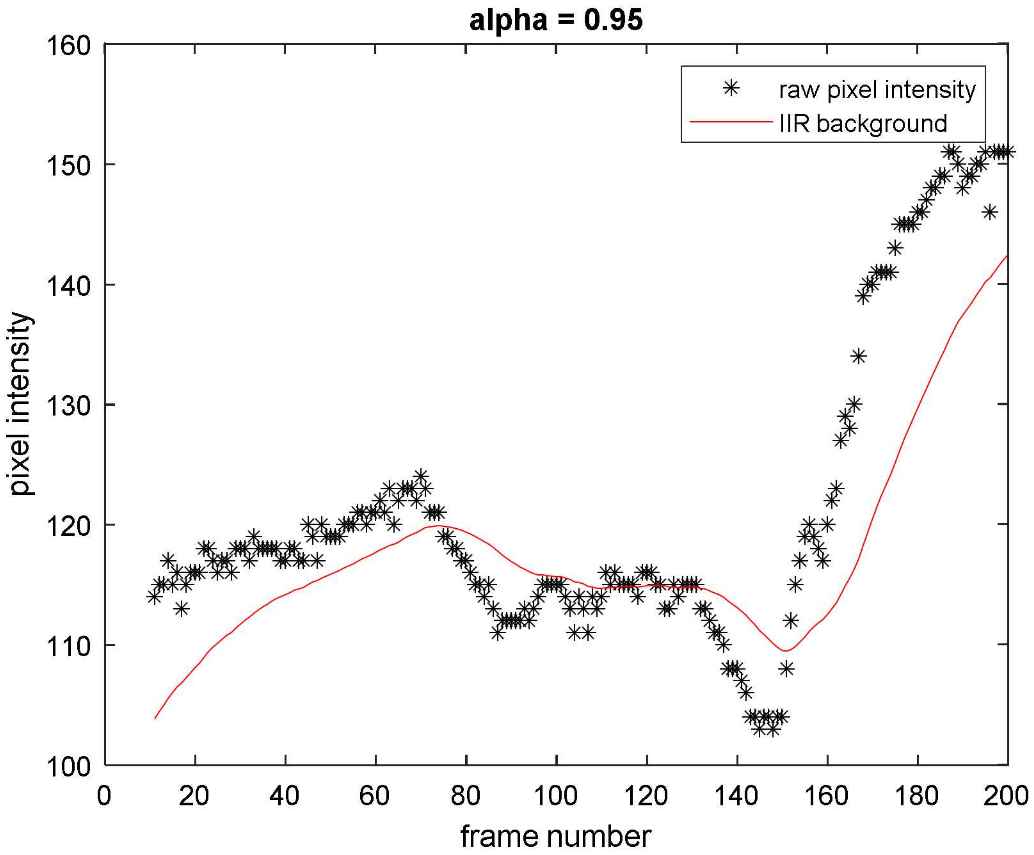
- Update:

$$B(i,j,t)=(1\text{-}\alpha)F(i,j,t)+(\alpha)B(t-1)$$

where $0<\alpha<1$

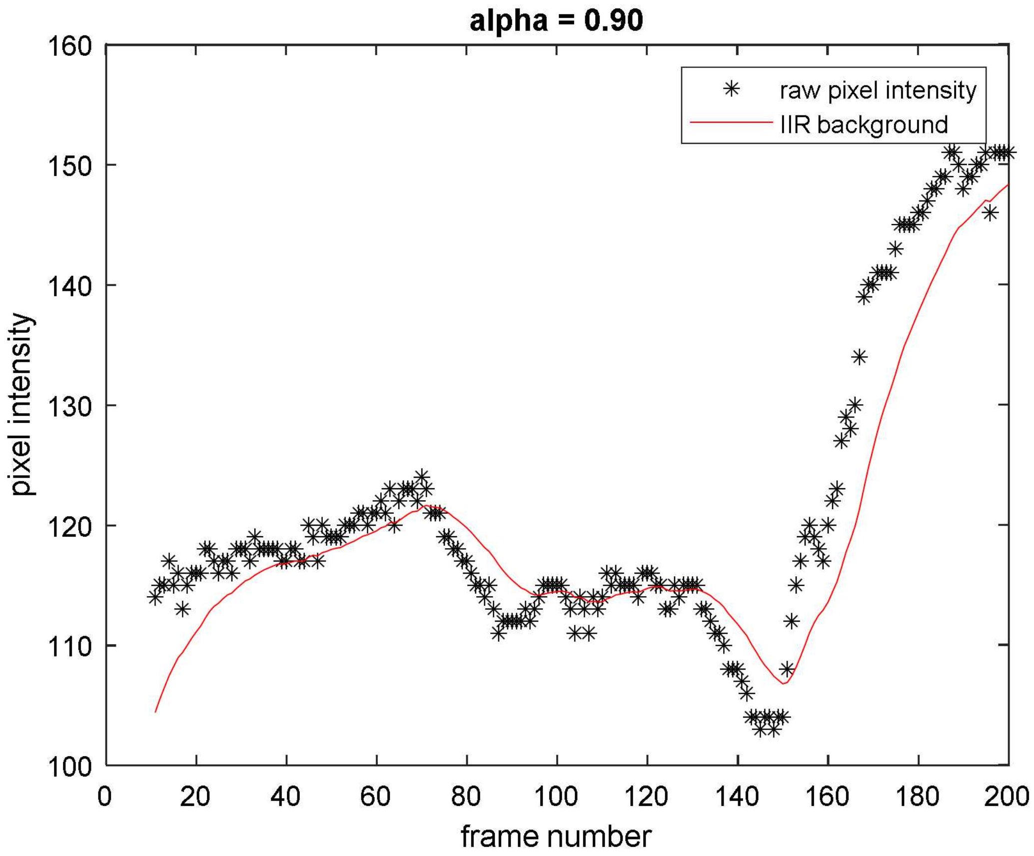
IIR Demonstration

IIR Background Update Rate $\alpha = 0.95$



Less responsive to background changes

IIR Background Update Rate $\alpha = 0.90$



More responsive to background changes

Background Estimation Algorithm – Subspace Tracking Techniques

- Subspace Tracking
  - Advantages: more effectively in tracking environmental changes (i.e. camera drift, illumination changes)
  - Disadvantages: requires more complexity in implementation and removing outliers that could make subspace unstable subspaces

- The goal is to capture the covariance structure of a sequence of frames in a low-dimensional, orthogonal subspace.

- From a sequence of N-dimensional vectors, X(1), X(2), . . . X(t), we could (in theory) compute the $N \times N$ sample covariance matrix, CXX(t).

- Let R be the dimension of the subspace representing background energy in the scene of interest (for jitter, want $R \geq 3$).

  - $W(t)$ is the $N \times R$ matrix whose columns are the basis vector estimates at time t.

- Changes that are consistent with those induced by jitter will lie in (close to) the subspace, while anomalous (target) events will not.

Adaptive Subspace Tracking Technique

- One approach uses the Fast Approximated Power Iteration (FAPI) algorithm for subspace estimation (Badeau et al., 2005).
  - Has low computational cost, O(*NR*), and provides orthogonal basis vectors.
- FAPI tracks the principal subspace of the data covariance, CXX(t), *without ever computing, decomposing, or storing this high-dimensional matrix.*
  - Approximates the principal subspace of a covariance matrix that is recursively updated using exponential weights:

$$C_{XX}(t) = \beta\, C_{XX}(t\text{-}1) + X(t)\, X^T(t)$$

- To track gradual change (pointing drift, cloud motion) the subspace is updated after every frame (can perform less frequently).

- Parameter $\beta \subset [0, 1]$ controls the rate at which new data are incorporated.

Larger values of β give slower update rates.

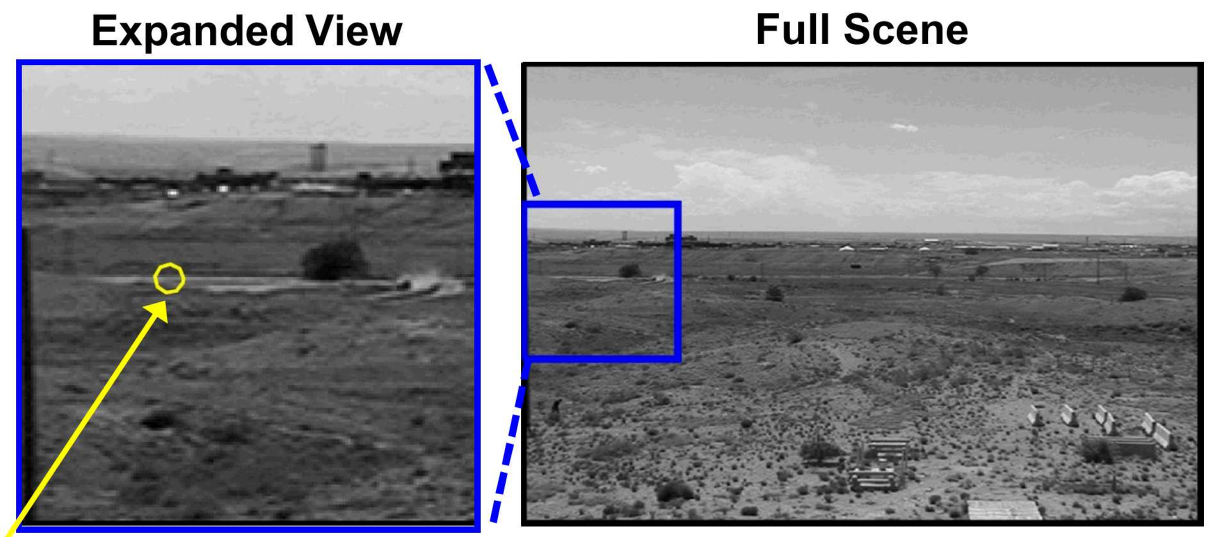
- Can selectively slow the background update rate for pixels with large detections.

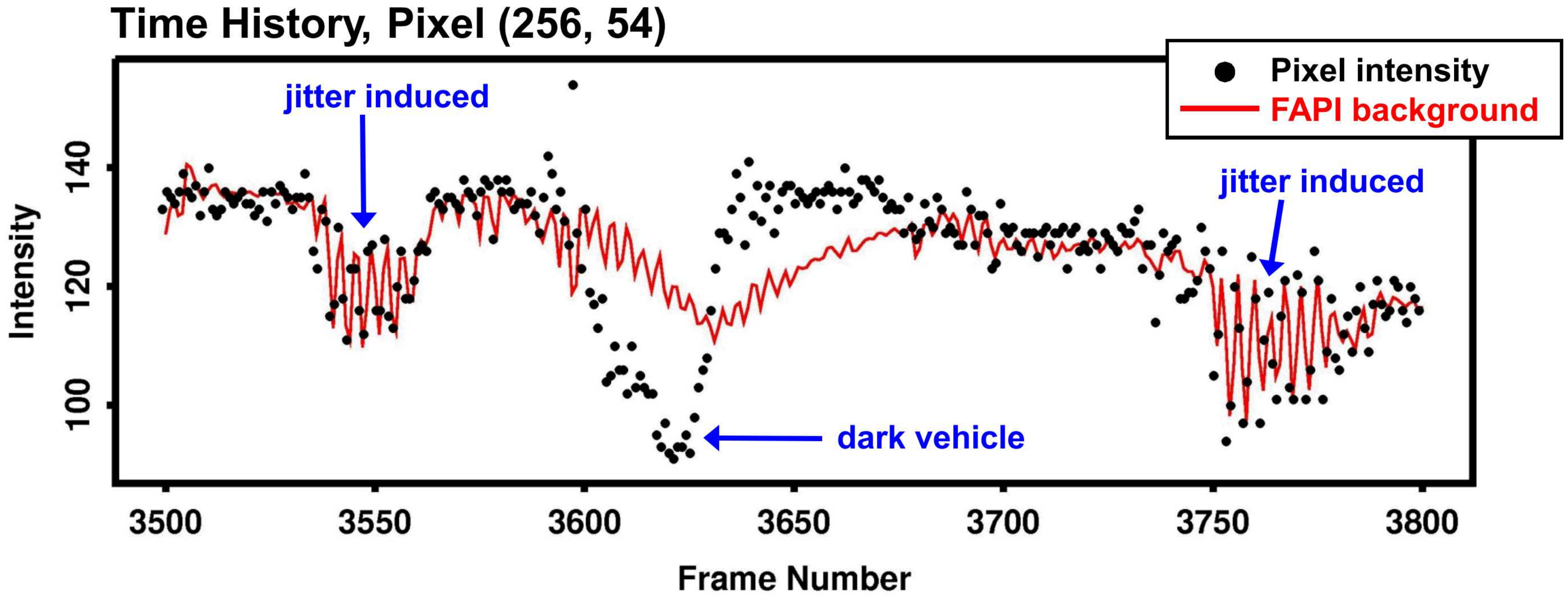
FAPI Background Estimation Example

Slide taken from presentation: Katherine Simonson and Tian Ma, “Real-Time Change Detection for Wide Area Surveillance”, SAND2014-1489C

The highlighted pixel lies along a road, and is subject to change due to both camera jitter and passing traffic.

The FAPI background estimate tracks jitter closely, but gives large residuals when a dark vehicle moves through.





Temporal Noise Estimation

- Motivation: enable "Adaptive" thresholding
  - Avoid hard thresholding, (e.g. apply threshold on difference frame)
  - Normalization of difference frame relatively to various noise levels in the image

Temporal Variance Update

- Provides statistical measure of variability of a given pixel, one way to estimate temporal variance is to use a IIR low pass filter

$$var(i,j,t) = (1-\gamma)\, D(i,j,t)^2 + \gamma\, var(i,j,t-1)$$

where γ is the variance update rate [0,1]

Temporal Standard Deviation σ, is obtain by:

$$\sigma = \sqrt{var(i,j,t)}$$

Foreground and Background Discrimination

- Recall,
  - $B(i,j,t-1)$ correspond to background estimate, pixel location i,j at time $t-1$
  - $F(i,j,t)$ corresponds to frame F, pixel location i,j at time t
  - σ (i, j, t − 1) correspond to the temporal standard deviation of pixel location i,j at time $t-1$
- Foreground/Background Decision Logic
  - Normalize difference frame by its standard deviations

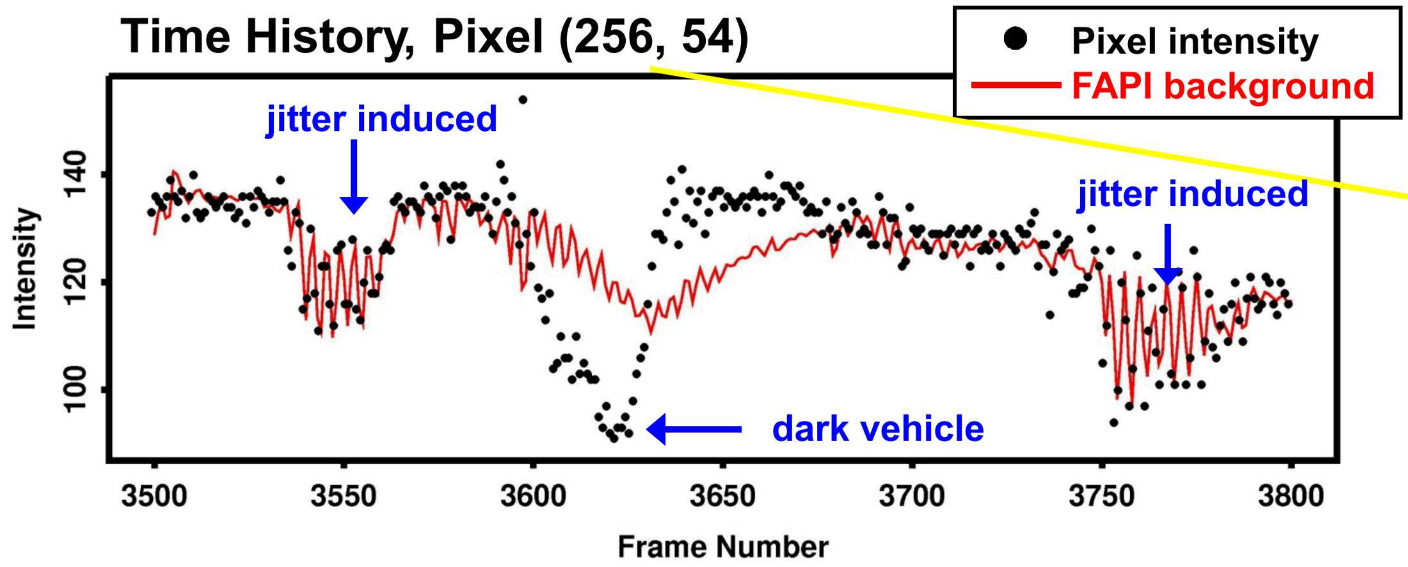
$$\text{if, } \frac{|F(i,j,t)-B(i,j,t-1)|}{\sigma(i,j,t-1)} > Threshold \Rightarrow Detection\ (pixel\ \mathrm{i,j}\ belong\ to\ foreground)$$

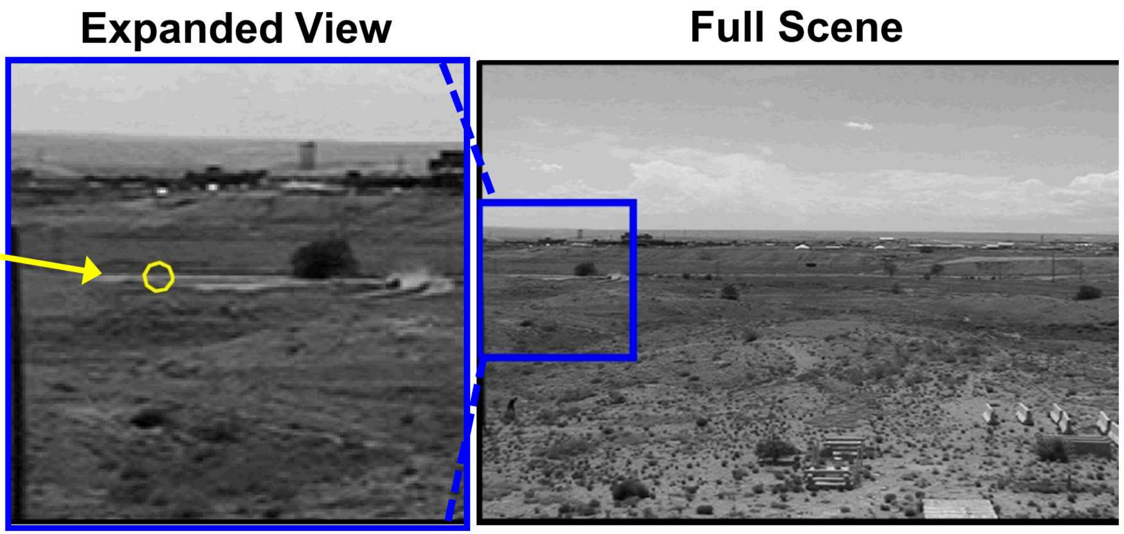
$$\text{else => No Detection } (pixel\ \mathrm{i,j}\ belong\ to\ background)$$

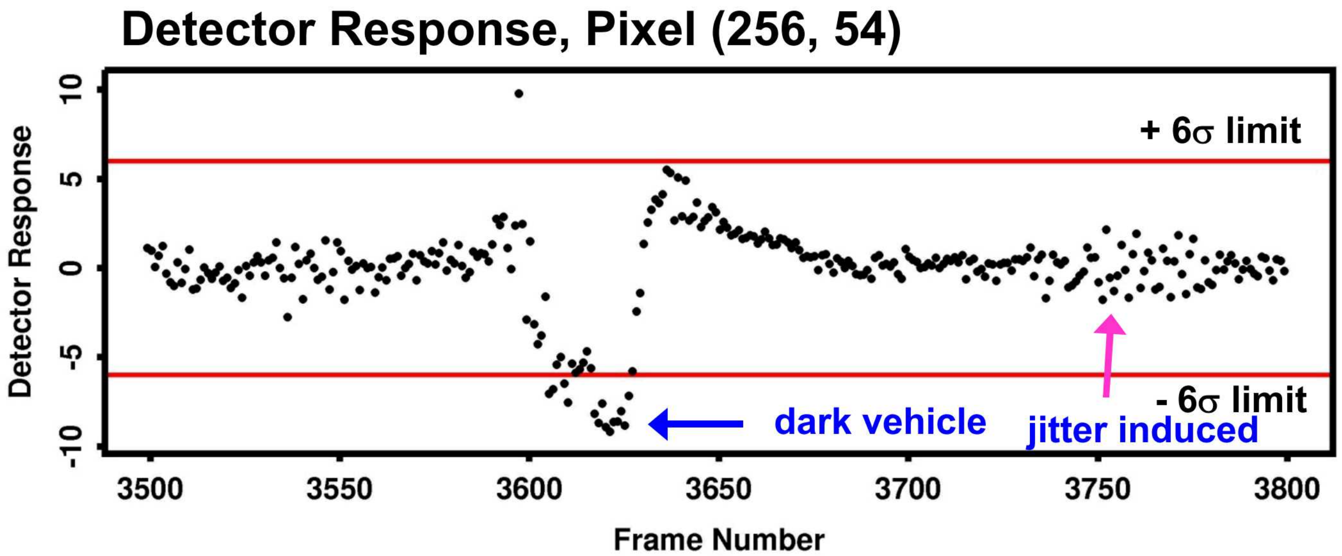
  - .

Target Detection Thresholding

Slide taken from presentation: Katherine Simonson and Tian Ma, “Real-Time Change Detection for Wide Area Surveillance”, SAND2014-1489C







While the decreased intensity due to jitter (pink arrow) is almost as low as the drop due to a dark vehicle passing through the pixel (blue arrow), the detector responds differently to jitter and signal.

Challenging Problems

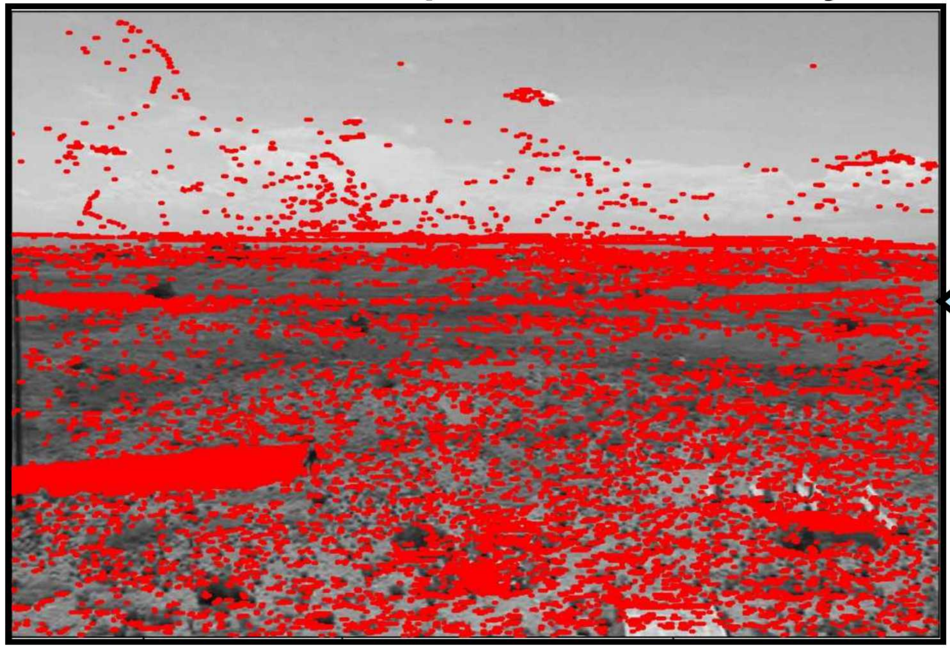
Detection in Presence of Jitter

- Sudden Random Jitter (e.g. wind blowing all of sudden!)
- Temporal model estimation becomes insufficient
- Need for a separate spatial statistical model
- Further reading: Katherine M. Simonson and Tian J. Ma, "Robust Real-Time Change Detection in High Jitter", SAND2009-5546.
  - Sandia was granted a patent for the spatial variance technique. **U.S. Patent No. 8,103,161, K.M. Simonson and T.J. Ma, "Estimating Pixel Variances in the Scenes of Staring Sensors," 24-Jan-2012.**

Kirtland AFB (Presence of Jitter): Single-Model Detections

Slide taken from presentation: Katherine Simonson and Tian Ma, "Real-Time Change Detection for Wide Area Surveillance", SAND2014-1489C

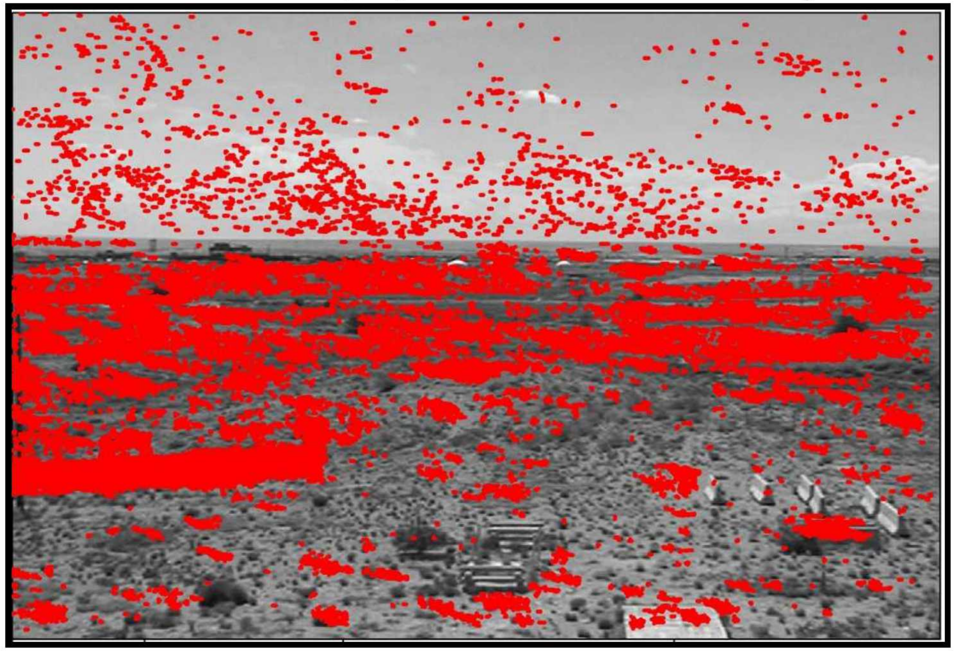
Detections, Temporal Variances Only



Frames 2400 – 3800
Detection Threshold = 6.0

When only temporal estimates of pixel variance are available, false alarms occur at scene edges: bright clouds, roads, vegetation, and the horizon.

Detections, Spatial Variances Only



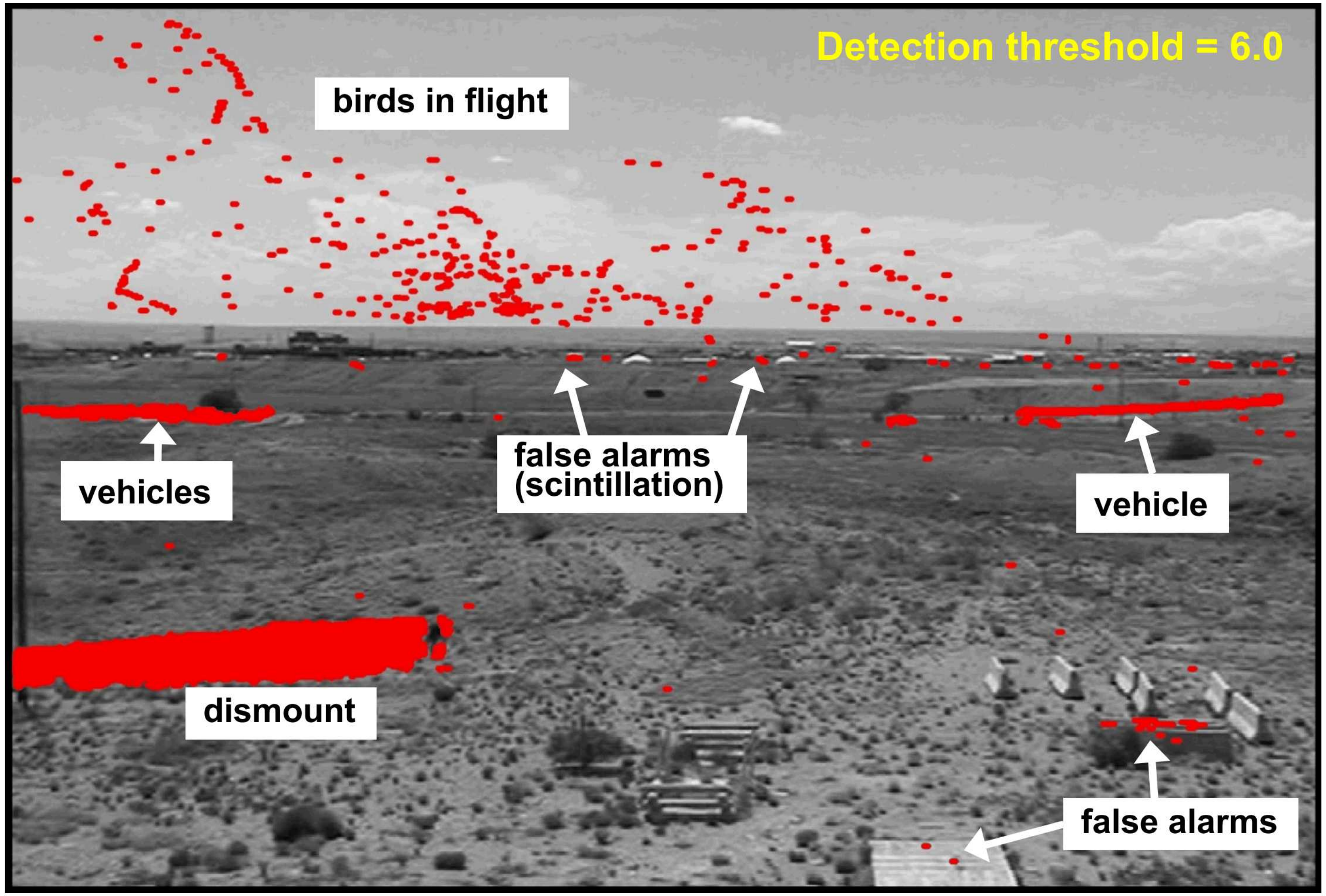
When background differences are normalized with spatial standard deviation estimates only, sensor noise induces false alarms in relatively uniform parts of the scene.

Kirtland AFB (presence of Jitter) - both temporal and spatial model

Slide taken from presentation: Katherine Simonson and Tian Ma, “Real-Time Change Detection for Wide Area Surveillance,” SAND2014-1489C

30 Hz video showing various activities near Sandia’s robotic vehicle range.

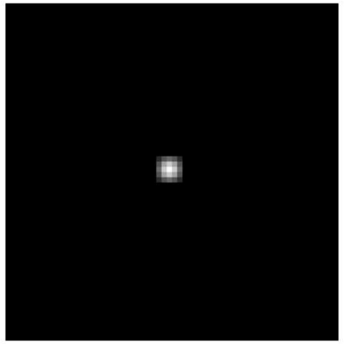
Red dots show pixels with at least one detection in frames 2400 – 3800, using the dual-variance (spatial & temporal) model.



Low Signal and Noise Ratio (SNR) Target Detection

- Low signal-to-noise (SNR) detection
  - Requires signal integration over multiple frames to increase target SNR

Simulated Target
(No added noise)

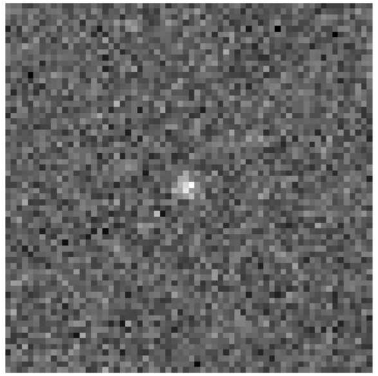


Target – Frame 1
(Added noise)

Target – Frame 2
(Added noise)

Target – Frame 3
(Added noise)

Signal Integration
(over 3 frames)



Integrating signal over multiple frames can make the more easily detectable!

Key Challenge: How to do this efficiently on moving target?

Sandia was granted a patent for the multiple hypothesis signal integration technique. **U.S. Patent No. 10,032,285, T.J. Ma, "Multi-Hypothesis Moving Object Detection System," 24-Jul-2018.**

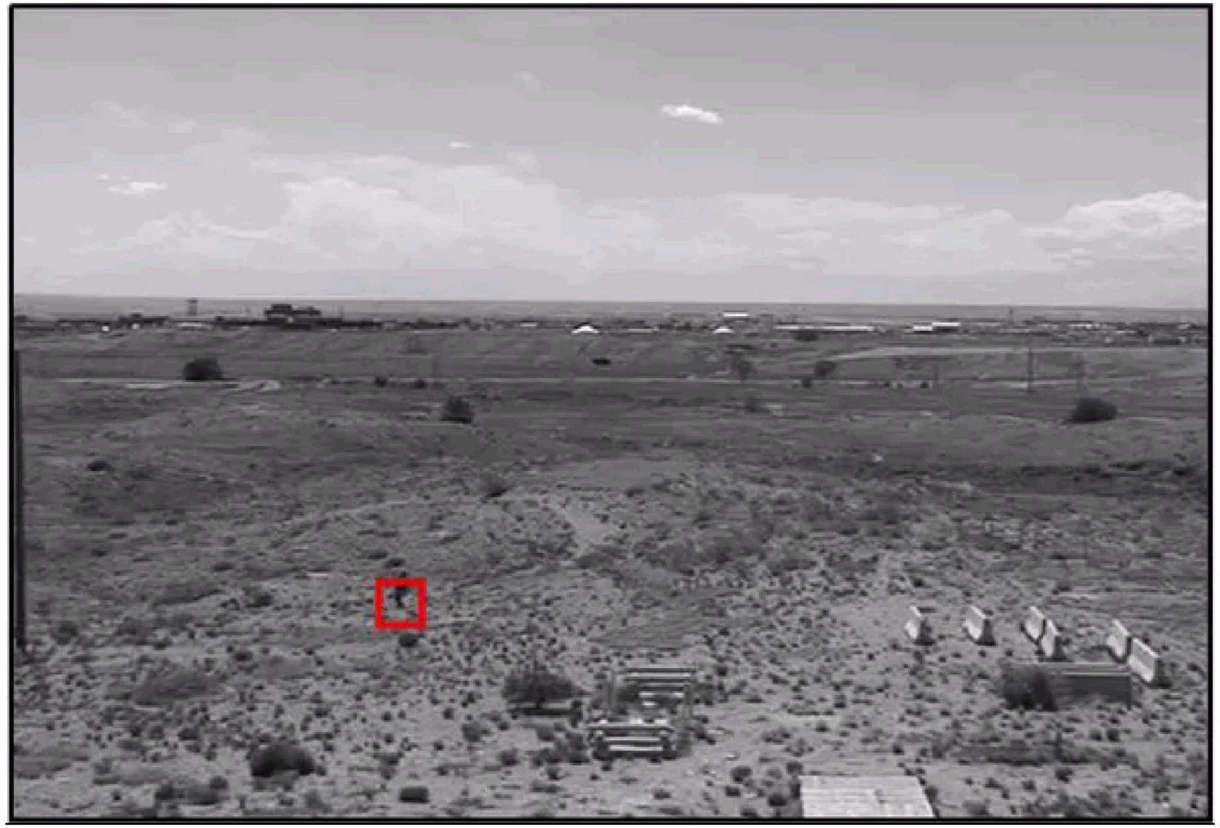
Examples

Examples are taken from presentation:

Katherine Simonson and Tian Ma
“Real-Time Change Detection for Wide Area Surveillance”
SAND2014-1489C

Example 1: KAFB Video with Detections

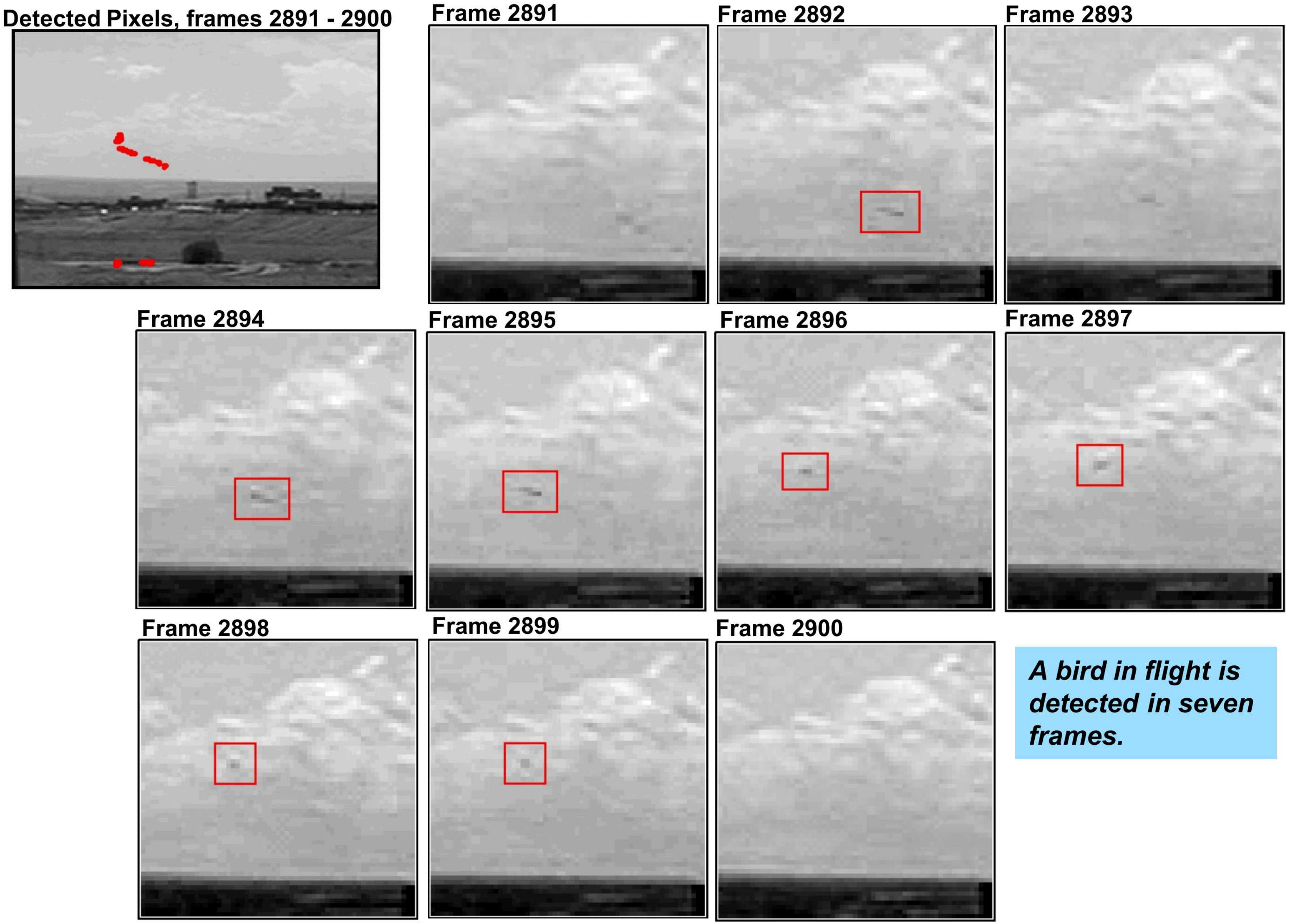
Slide taken from presentation: Katherine Simonson and Tian Ma, “Real-Time Change Detection for Wide Area Surveillance”, SAND2014-1489C



Red boxes indicate pixel detections; no tracker is applied.

Detecting Birds in Flight

Slide taken from presentation: Katherine Simonson and Tian Ma, “Real-Time Change Detection for Wide Area Surveillance”, SAND2014-1489C



A bird in flight is detected in seven frames.

Example 2 - Border Camera Footage

Slide taken from presentation: Katherine Simonson and Tian Ma, “Real-Time Change Detection for Wide Area Surveillance”, SAND2014-1489C

Video from a surveillance camera on the Texas/Mexico border.

- Downloaded from “Virtual Border Watch,” a live video streaming website operated by the Texas Border Sherriff’s Coalition and Bluservo.net.
- Network of pole-mounted surveillance cameras operating in the visible during daytime hours and infrared at night.





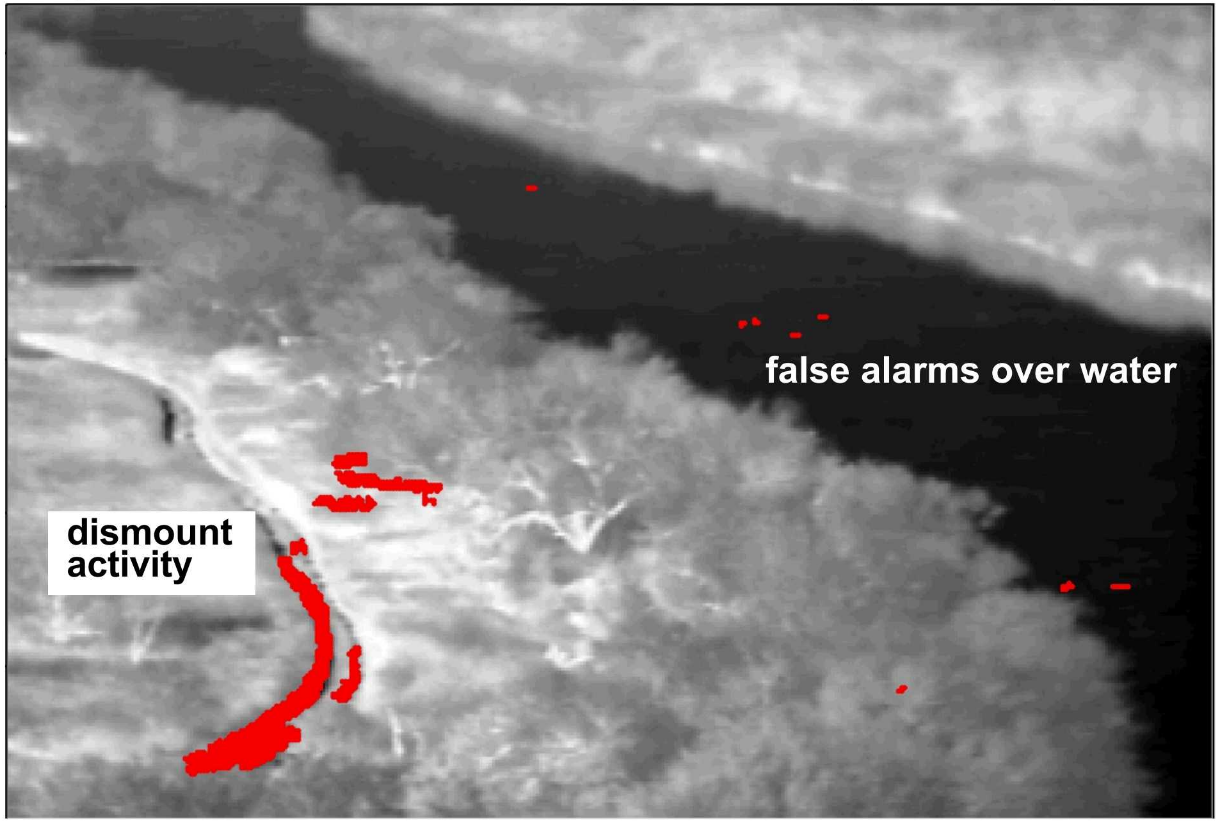

Nighttime Scene Along River

Slide taken from presentation: Katherine Simonson and Tian Ma, “Real-Time Change Detection for Wide Area Surveillance”, SAND2014-1489C

10 Hz infrared video sample; nighttime scene.

- In this example, jitter was artificially induced.
- Detector set to find only positive change: new heat sources.

Red dots show pixels with at least one detection in frames 500 – 1500, using the dual model (temporal & spatial) approach.



Texas Border Sherriff's coalition and www.blueservo.net

Two dismounts emerge from the vegetation along the river, return to the riverside, re-emerge, and proceed down the track and out of the scene.

At times, they are lost in the near-saturated pixels to the right of the track.

Detection threshold = 6.0

Nighttime Scene Along River – Detection

Slide taken from presentation: Katherine Simonson and Tian Ma, “Real-Time Change Detection for Wide Area Surveillance,” SAND2014-1489C

Texas Border Sherriff's coalition and www.blueservo.net

Red boxes indicate pixel detections; no tracker is applied.

Example 3 – ZooCam

Slide taken from presentation: Katherine Simonson and Tian Ma, “Real-Time Change Detection for Wide Area Surveillance,” SAND2014-1489C

- **10 Hz video downloaded from the “Bear Cam” at the Woodland Park Zoo.**
- **Original video was in color – downgraded to greyscale for our analysis.**
- **Stable camera with no jitter – many moving scene elements (running water).**
- **Several birds visit the scene – both the birds and their shadows are detected.**

ZooCam Video With Detection Boxes

Slide taken from presentation: Katherine Simonson and Tian Ma, “Real-Time Change Detection for Wide Area Surveillance,” SAND2014-1489C

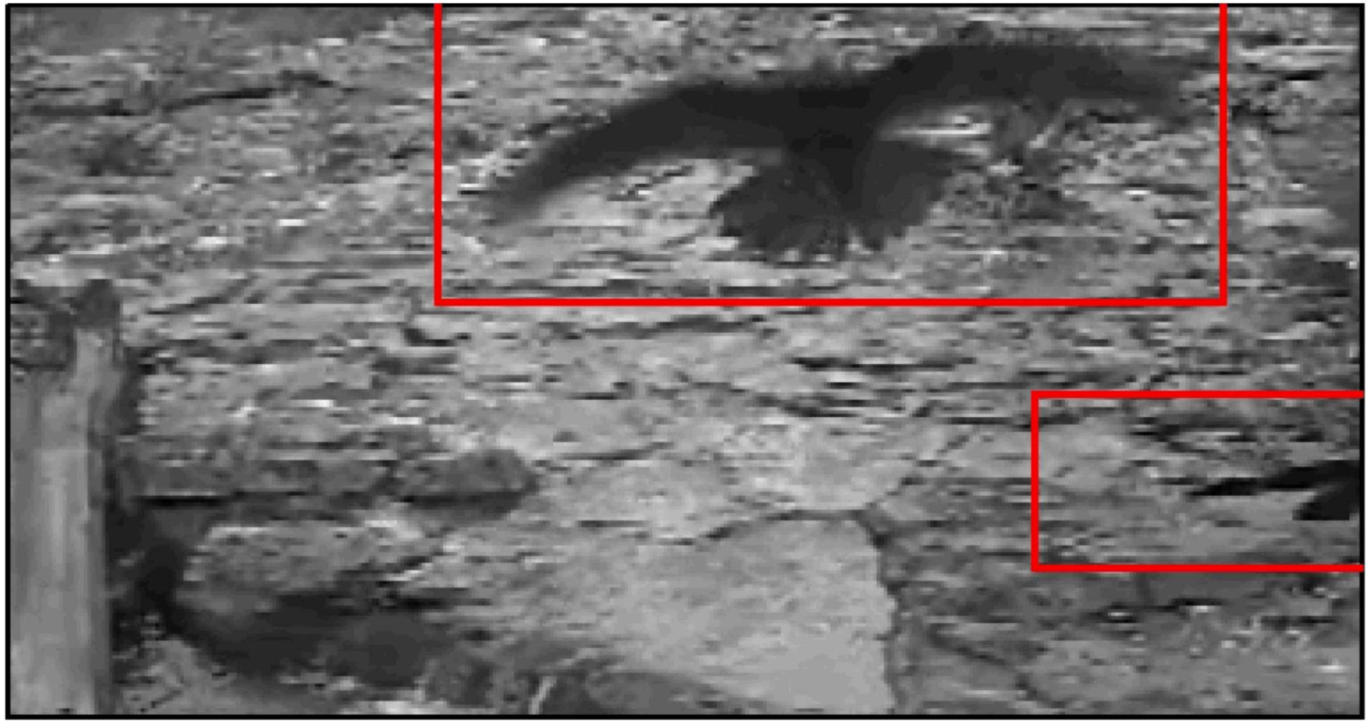
Frames 3380 – 3700, Threshold = 8.0

Red boxes indicate pixel detections; no tracker is applied.

Bird in Foreground

Slide taken from presentation: Katherine Simonson and Tian Ma, “Real-Time Change Detection for Wide Area Surveillance,” SAND2014-1489C

Frame #3691



In the last 25 frames of the video, a bird flies into the foreground of the camera. Both the bird and its shadow are detected.

Detection boxes shown for frame #3691, Dual Model, Threshold = 8.0

Further Readings

- Katherine M. Simonson and Tian J. Ma, "Robust Real-Time Change Detection in High Jitter," SAND2009-5546
- Roland Badeau, Bertrand David, and Gael Richard, "Fast Approximated Power Iteration Subspace Tracking", IEEE Transaction on Signal Processing, VOL. 53, NO. 8, AUGUST 2005.
- Katherine Simonson and Tian Ma, "Real-Time Change Detection for Wide Area Surveillance", SAND2014-1489C

Questions?