SAND2014-15273PE

Real-Time Anomaly Detection for Wide Area Surveillance

Pacific Northwest National Laboratory
01-Jul-2014 Richland, WA

Katherine Simonson and Tian Ma
Sandia National Laboratories

Contact Author: Katherine Simonson
email: kmsimon@sandia.gov
tel: 505-844-6551

Outline

- Background and Mission Constraints
- Mathematics of the Detector
- A Few Examples
- Summary and Conclusions

Research Goal

- Our goal is the ***automatic detection of small changes*** in wide area surveillance.
  - We work primarily with ***low-resolution, staring, radiometric sensors***, which are subject to significant jitter.
  - Frame rates up to 75 Hz; ***algorithms must run causally in real time***.
- We ***work directly with real-world data*** from deployed, operational, surveillance systems.
  - As well as video sequences from a range of unclassified sources.
- We are interested in ***detecting all physical change in the scene***, no matter how small.
  - While rejecting variation due to sensor-related artifact, including pointing drift, jitter, noise, pixel irregularities, and specularities.

Background Subtraction

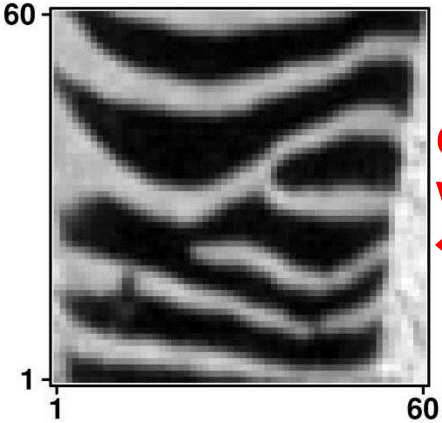
- The standard approach to ***change detection involves some form of subtraction***:
  - To detect new energy at time *t*, subtract from the frame taken at *t* an estimate of the "background" energy in the scene prior to this time.
  - The background estimate may be a single prior frame or a more complex function evaluated over a window of recent frames.

- If the current frame is not properly registered to the background, large values in the difference frame may be caused by intensity gradients in the scene, rather than true (physical) change.

- It follows that ***change detection in a high jitter environment is particularly challenging!***

Mis-Registration

Frame 1

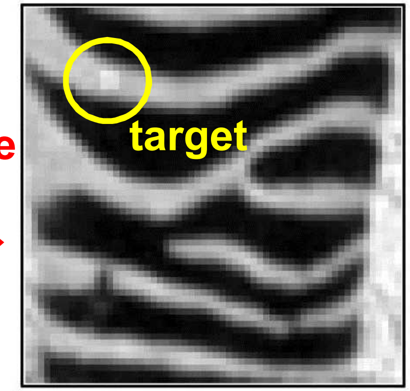


Full Image

Crop and add white noise

Crop, shift (r,c)+0.5, add white noise and 3x3 target

Frame 2

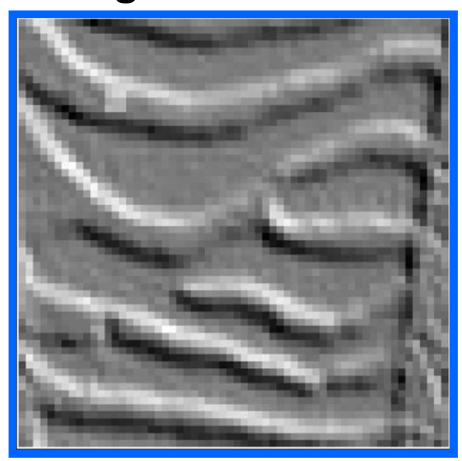


The difference between frames slightly out of alignment is dominated by scene gradients larger than the target change.

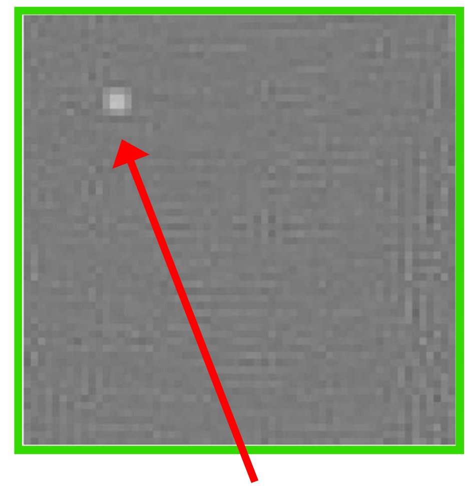
Interpolating the second frame into alignment with the first blurs the target signal.

The two difference frames are plotted in the same greyscale.

Frame 2 – Frame 1, Unregistered



Frame 2 – Frame 1, Registered



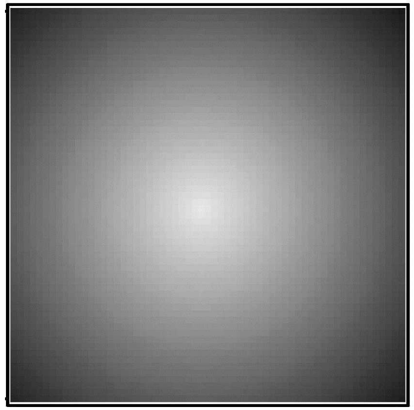
target signal, blurred

Sensor Artifact

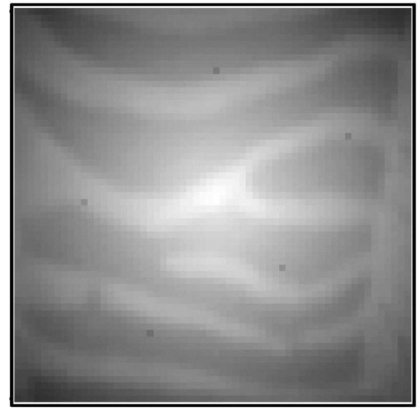
Artifacts in pixel space challenge solutions based on scene registration!

A bias surface was added to the original frames, and reduced responsiveness was simulated in 5 pixels.

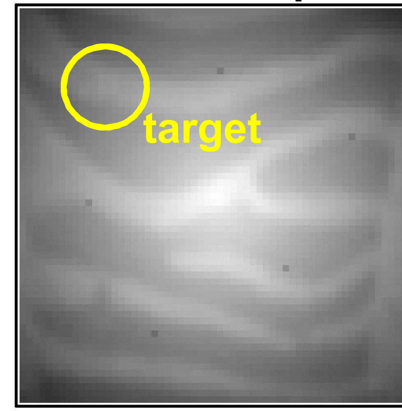
Bias Surface



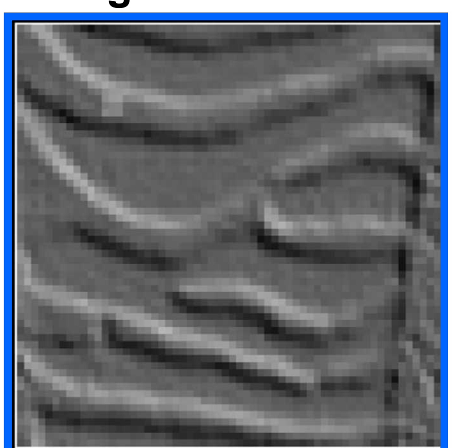
Frame 1 + Bias, Reduced Response



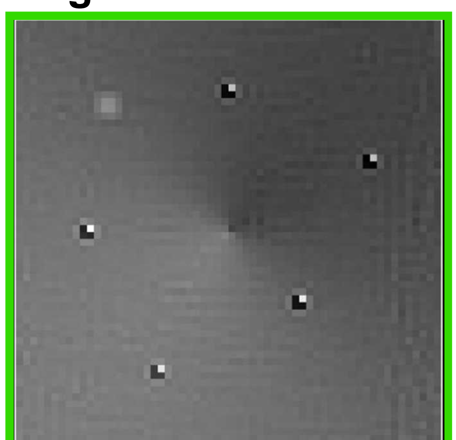
Frame 2 + Bias, Reduced Response



Difference Frame, Unregistered



Difference Frame, Registered



When Frame 2 is translated to register with the scene of Frame 1, the defects move out of alignment, creating large apparent changes in the difference frame.

All such defects must be known and corrected for prior to scene registration.

Algorithm Approach

- ***Frame registration cannot solve the jitter problem*** **in real time:**

    - **Registration to a small fraction of a pixel is required, but this precision is often not achievable at high frame rates for low-quality data.**

    - **Even if jitter-induced offsets are known perfectly, all sensor artifacts (fixed pattern noise, self-emission, over- or under-responsive pixels) have to be corrected prior to frame transformation. This may not be feasible for gradually varying artifacts.**

- **Our approach does not require registration, instead relying on two separate statistical models for variations in pixel intensity.**

    - ***The temporal model*** **handles pixels that are naturally variable due to sensor noise or moving scene elements, along with jitter displacements comparable to those observed in the recent past.**

    - ***The spatial model*** **captures jitter-induced changes that may or may not have been observed previously.**

Outline

- ❖ Background and Mission Constraints
- ❖ Mathematics of the Detector
- ❖ A Few Examples
- ❖ Summary and Conclusions

Normalized Differences

- **For each pixel (*k*, *h*) at time *t*, *we test whether the observed intensity is consistent with the spatial and temporal models*. The decision is based on simple normalized differences.**

$$Z(k,h;t) = \frac{X(k,h;t) - B(k,h;t)}{S(k,h;t-1)}$$

***X* = pixel (*k,h*)'s intensity at time *t*,**

***B* = current background estimate,**

***S* = current standard deviation estimate.**

- **A large (absolute) value of *Z(k, h; t)* means the observed pixel intensity is outside the range anticipated under the current model.**

Decision Logic

$$Z(k,h;t) = \frac{X(k,h;t) - B(k,h;t)}{S(k,h;t-1)}$$

- **Normalized differences, $Z_{SPATIAL}$ and $Z_{TEMPORAL}$ are computed using the same background *B*, but *different standard deviation estimates*.**

- **If min { | $Z_{SPATIAL}$|, | $Z_{TEMPORAL}$| } exceeds a fixed threshold, the observed value of the pixel at time *t* is inconsistent with both models, and a candidate detection occurs.**

  - A one-sided test may be applied if, e.g., only positive deviations are of interest.

- ***Depending on the characteristics of the target changes sought*, downstream logic may be employed to reduce the false alarm rate:**

  - *Area filtering:* Require detection in at least *K* connected pixels.

  - *Duration filtering:* Require detection in at least *M* consecutive frames.

Subspace Methods for Background Estimation

- **Our background estimator models the manner in which pairs of pixels vary together.**
  - ***The goal is to capture the covariance structure of a sequence of frames*** **in a low-dimensional, orthogonal subspace.**
  - **If jitter and/or pointing drift are major contributors to pixel intensity changes, we expect strong patterns of correlation between pixels.**

- **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, $C_{XX}(t)$.**
  - **Then use eigen decomposition (or SVD) to *estimate the principal subspace*.**
  - ***Computational issues (run-time and storage) would be significant!***
  - **For a 2K × 2K image, N = 4,000,000.**

- **If the background varies over time, we require a mechanism to *update the covariance matrix (and basis vectors) throughout the frame sequence*.**
  - **This need arises in many applications and has driven development in the very active field of *adaptive subspace estimation.***

Adaptive Subspace Estimation

- **Many authors have proposed methods of subspace tracking, beginning with Owsley (1978).**

  - Application to jitter suppression dates (at least) to Barry and Klop (1983).

- ***Many papers are published*** **in adaptive subspace estimation:**

  - Frequently cited: Oja and Karhunen (1985); Sanger (1989); Yang (1995); Badeau et al. (2005).

  - Literature reviews: Comon and Golub (1990); Doukopoulos and Moustakides (2008).

- **Approaches differ in terms of computational complexity, desired output (principal or noise subspace), tunable parameters, and orthogonality.**

  - Can add a Gram-Schmidt step, with increased computational cost.

Subspace Background Estimates

- **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 vectors estimates at time t.**

- ***Compute the scene background estimate at time t*** **by projecting data vector $X(t)$ onto the subspace spanned by the columns of $W(t\text{-}1)$:**

$$B(t) = W(t-1)\, W^T(t-1)\, X(t)$$

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

- **Suppose that pixel A correlates highly with pixels B, C, D, E and F in frames 1 to t-1: When A increases or decreases, B – F do the same.**
  - **At time t, if A suddenly increases but B – F do not, the change pattern will be inconsistent with the covariance structure captured in the basis vectors.**
  - **Pixel A will show a large projection residual for frame t.**

FAPI Algorithm for Subspace Tracking

- **Our 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, $C_{XX}(t)$, *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.

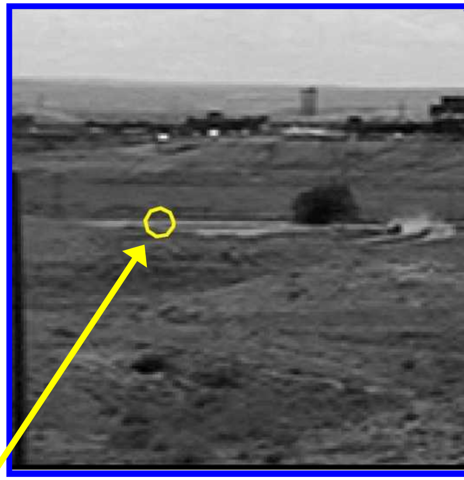
  - Can selectively slow the background update rate for pixels with large detections.

Background Estimation: Example

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

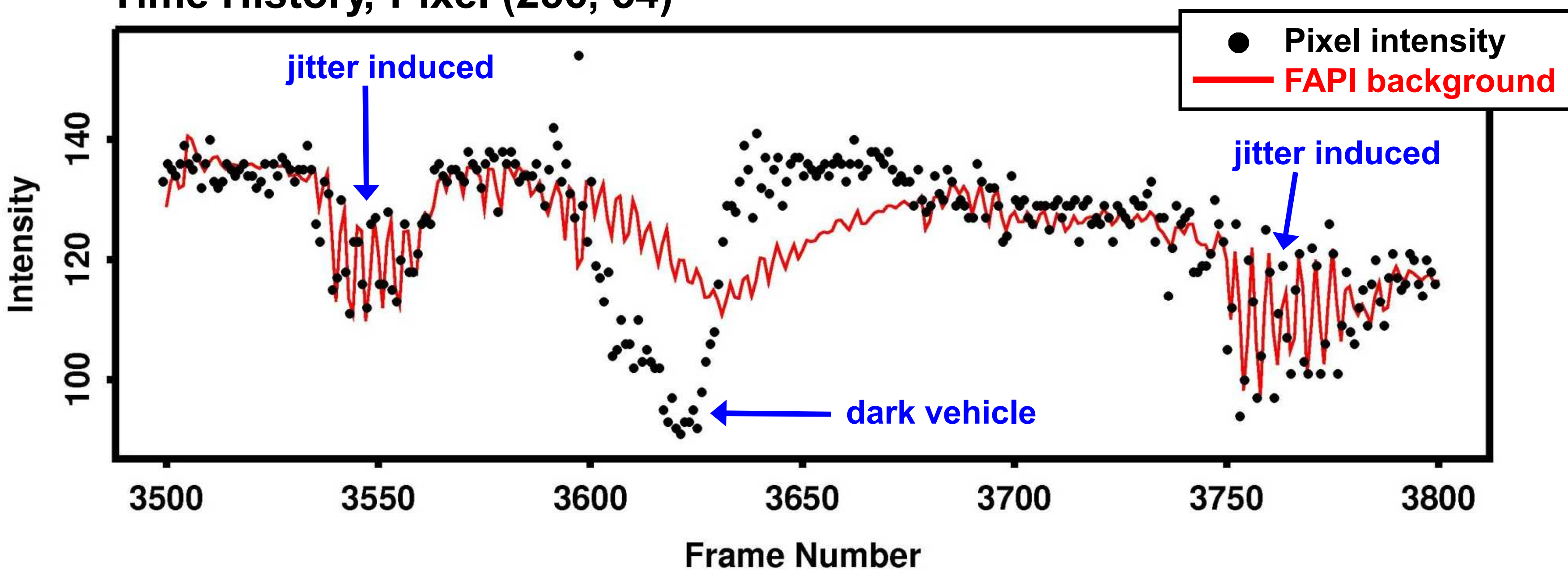
The FAPI background estimate tracks jitter closely, but gives large residuals when a dark vehicle moves through.

Expanded View



Full Scene

Time History, Pixel (256, 54)



Temporal Variances

- "Temporal" estimates of pixel variance are based on a recent time window of projection residuals. They are computed as follows:

  1. Initialize with the sample variance over the first n frames, $V(k,h;n)$.

  2. For subsequent frames, update using:

$$V(k,h;t) = (1 - \gamma)\left[X(k,h;t) - B(k,h;t)\right]^2 + \gamma\, V(k,h;t\text{-}1)$$

- Forgetting factor $\gamma \subset [0,1]$ determines how rapidly the filter responds to new energy.

  - As with the background estimate, the temporal variance estimate for any pixel showing a strong detection can be selectively updated at a slower rate.

Spatial Estimation: Motivation

- **As long as the jitter distribution is relatively stable, the temporal approach to variance estimation provides reasonable scale factors.**

- **For non-stationary jitter, temporal estimates are inadequate: when jitter increases, false alarms occur along scene gradients.**

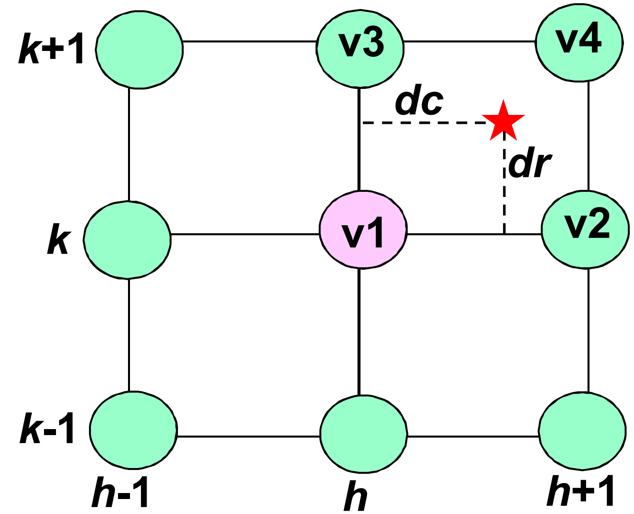
  - Subspace projection alone does not solve this problem !

Key Observation: You do not need to observe line-of-sight jitter to predict which pixels will be influenced !

- **We have developed a new mathematical concept for pixel variance estimation. Our "spatial" approach produces estimates that are robust to non-stationary jitter, based on a single frame.**

Bilinear Interpolation

- **The method operates over a grid of conditional expectations in the vicinity of each pixel.**



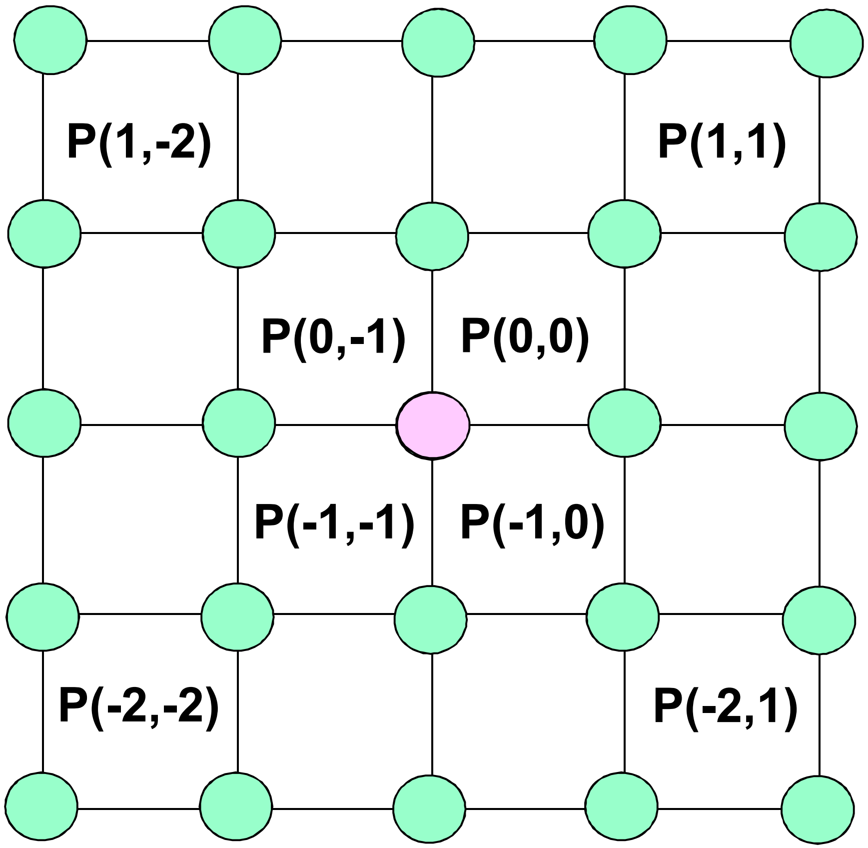
- **At time *t*-1, define:**
  v1 = value at pixel (*k*,*h*)
  v2, v3, v4 = values at nearby pixels

- ***If we knew that jitter between times t-1 and t was exactly dr rows and dc columns*, we could use bilinear interpolation to estimate the background at pixel (*k*,*h*) at time *t*:**

$$E(k,h;\ t) = v1 + dr\,(v3-v1) + dc\,(v2-v1) + dr\cdot dc\,(v1+v4-v2-v3)$$

- **If (*dr*, *dc*) is unknown, we can *use its statistical distribution to estimate the mean and variance* of each pixel at time *t* as a function of pixel values at time *t*-1 (or other previous frame).**

Conditional Expectation



- For each "cell" near (k,h), we use an ***assumed jitter distribution*** to compute:

1) The probability of jittering into this cell at time t, and:
2) The expected pixel value (and its square) at t, given jitter into this cell.

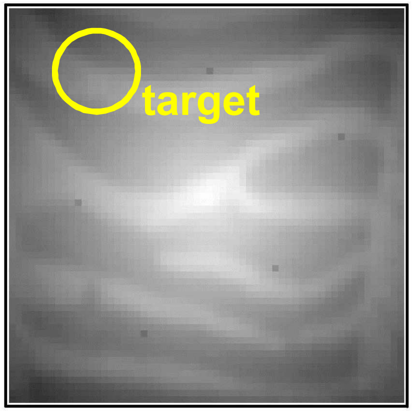
- After much algebra (see SAND report), we ***apply the Law of Total Probability to estimate the variance*** of each pixel at time t.

- Estimates computed in this manner are ***surprisingly robust to mis-specification of the jitter distribution***: They scale roughly linearly with the jitter standard deviation parameter (sigma).

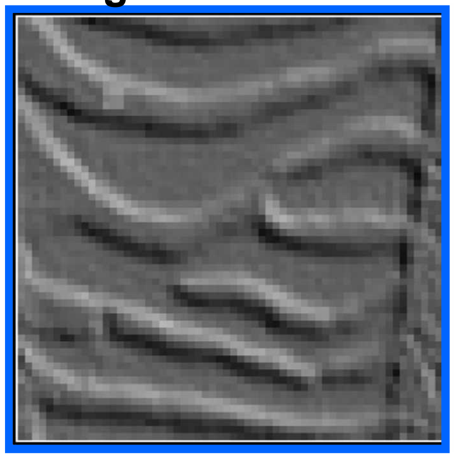
  - A good strategy is to set sigma conservatively (based on the worst jitter expected) and re-scale on a per-frame basis.

Incorporating SSP and Spatial Variances

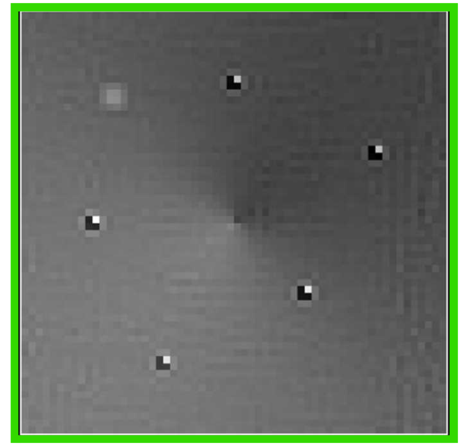
Frame 2 + Bias, Reduced Response



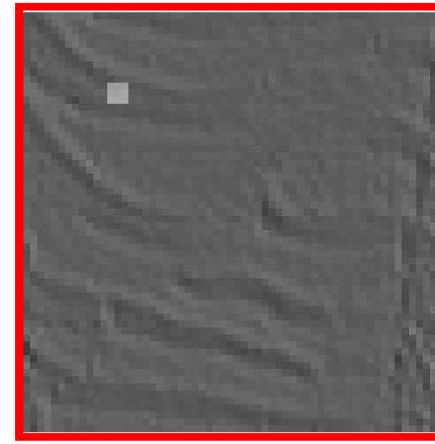
Difference Frame, Unregistered



Difference Frame, Registered



Raw SSP Residuals

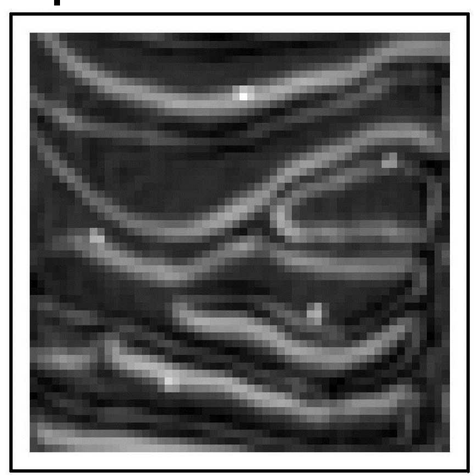


The principal subspace was estimated from 100 simulated jittered, noise-added versions of Frame 1 (with bias surface and reduced responsiveness).

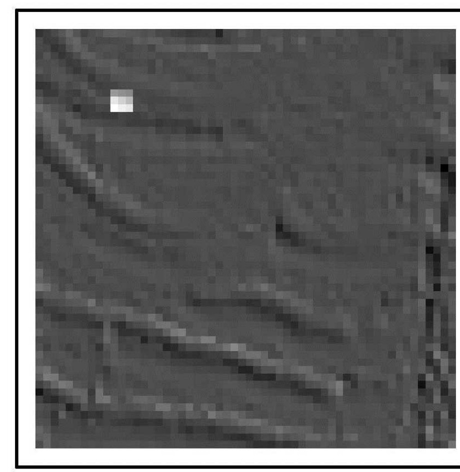
SSP residuals show less scene structure than the unregistered frame differences, and exhibit no sensor artifact.

After division by spatial standard deviations, the nine target pixels have values between 1.51 and 4.55, larger than ALL non-target pixels.

Spatial StDevs



Normalized SSP Residuals



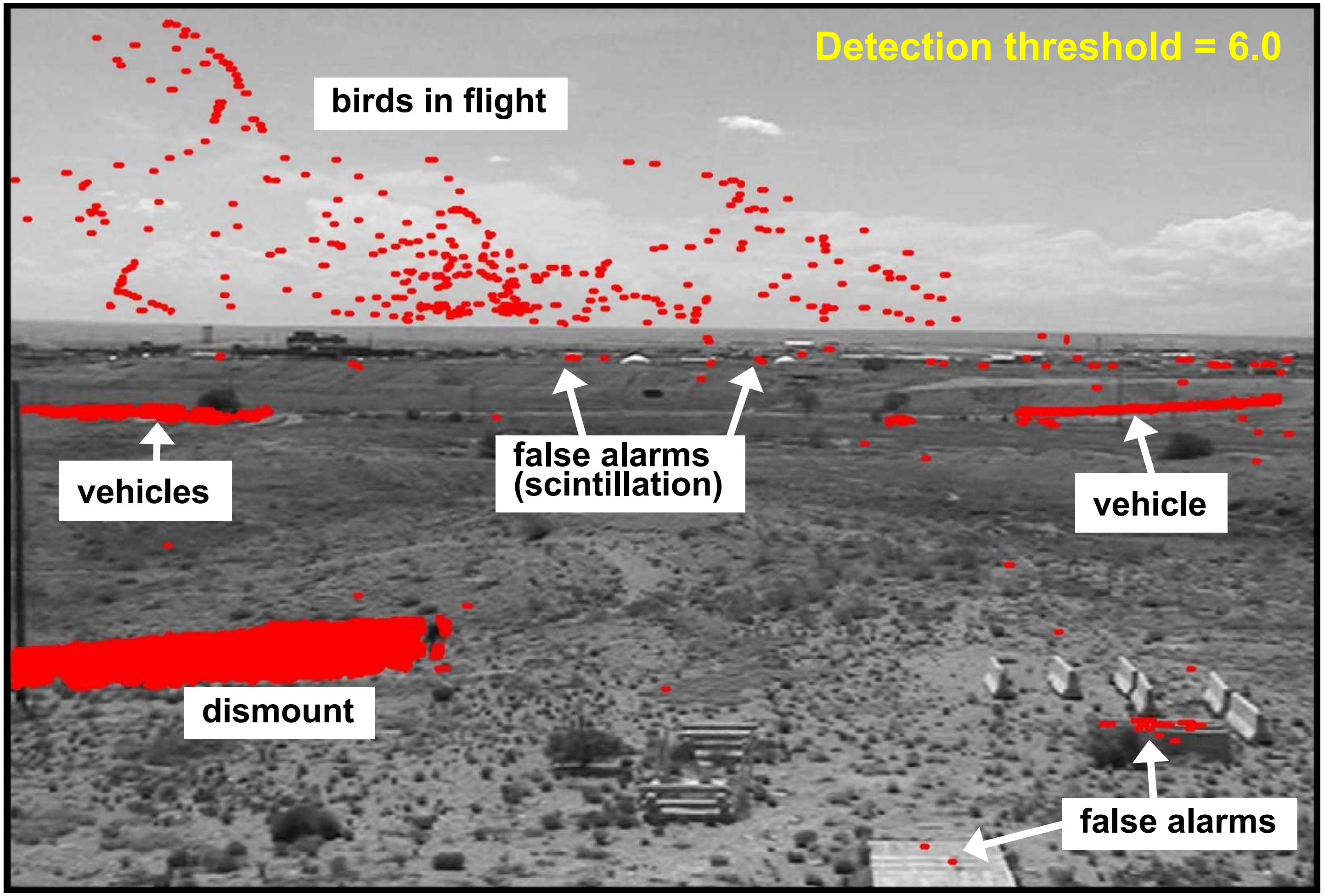
Outline

- ❖ Background and Mission Constraints
- ❖ Mathematics of the Detector
- ❖ A Few Examples
- ❖ Summary and Conclusions

Example 1 – Kirtland AFB

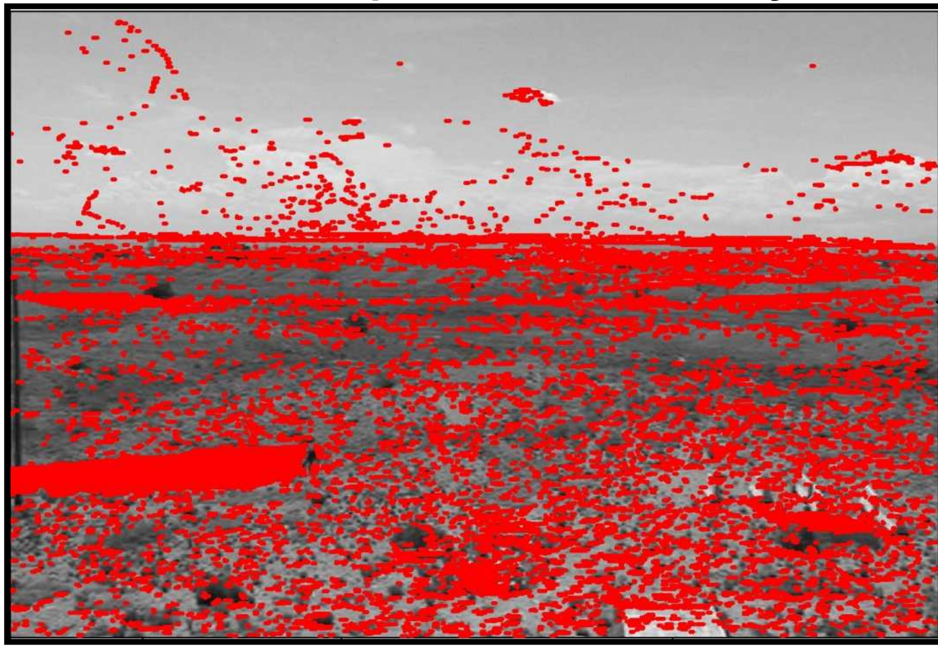
30 Hz video showing various activities near Sandia's robotic vehicle range.

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



KAFB Example: Single-Model Detections

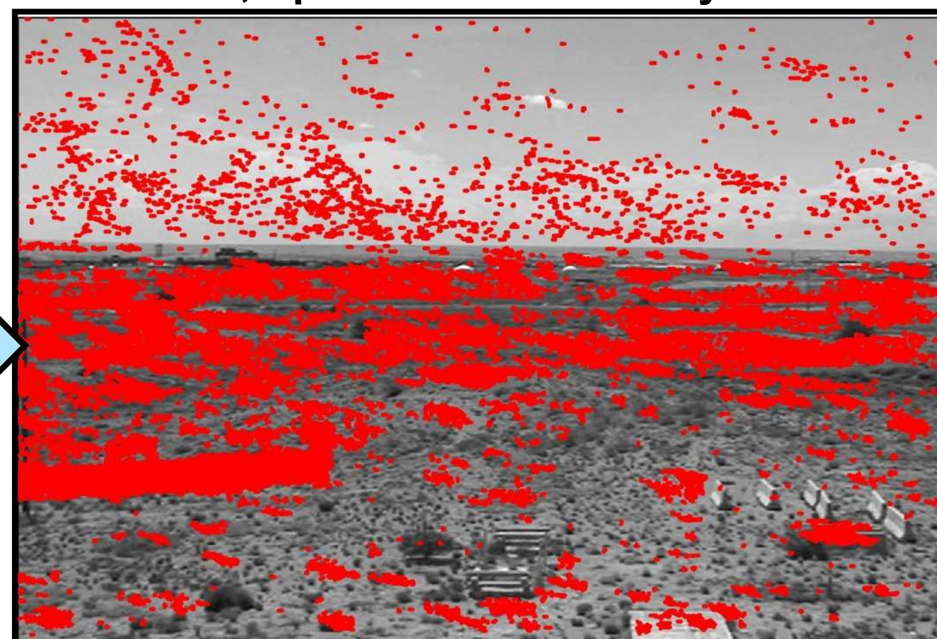
Detections, Temporal Variances Only



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

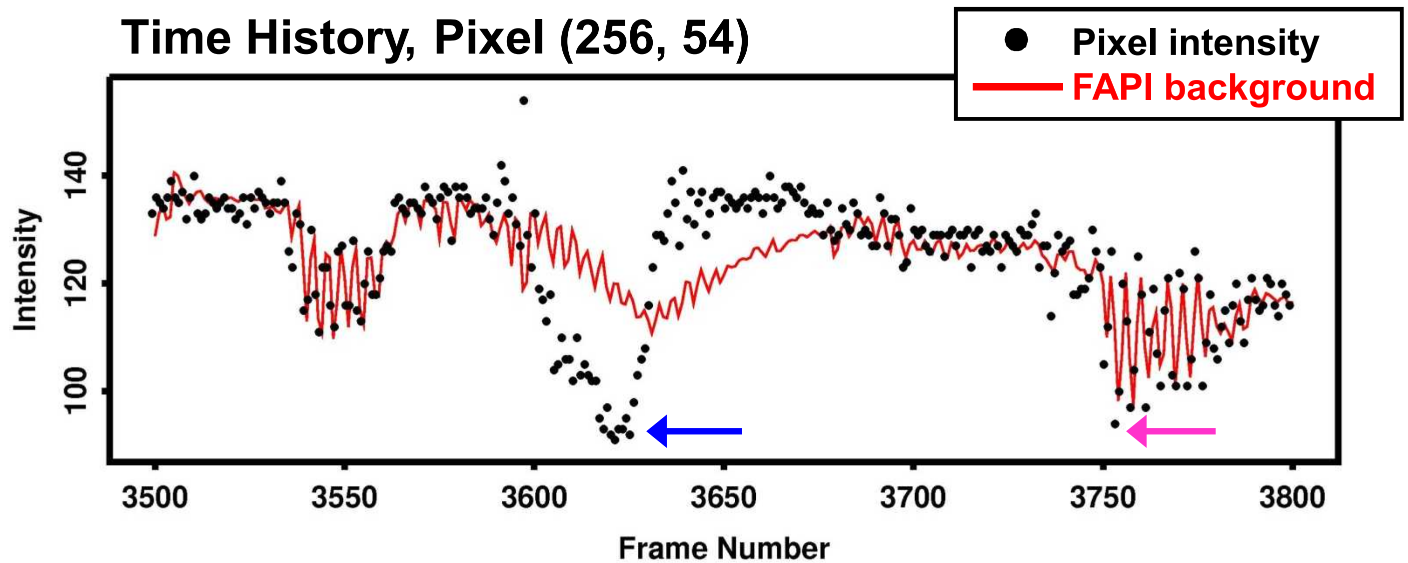


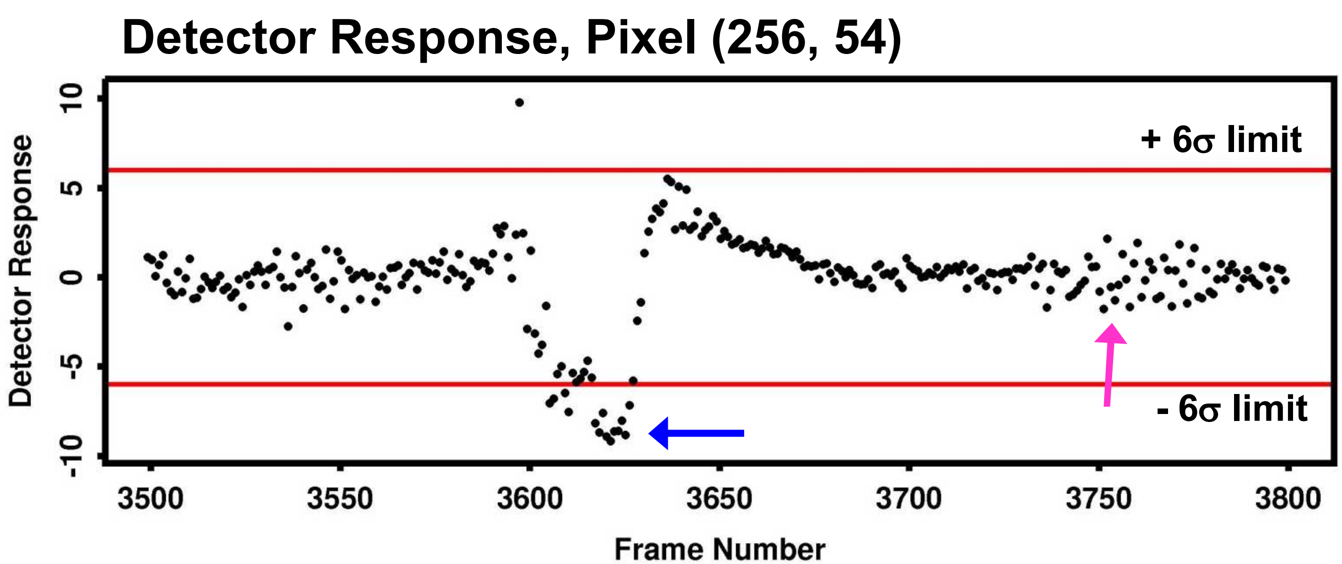
When background differences are normalized with spatial standard deviation estimates only, sensor noise induces false alarms in relatively uniform parts of the scene.

KAFB Video with Detections

Red boxes indicate pixel detections; no tracker is applied.

Detector Response



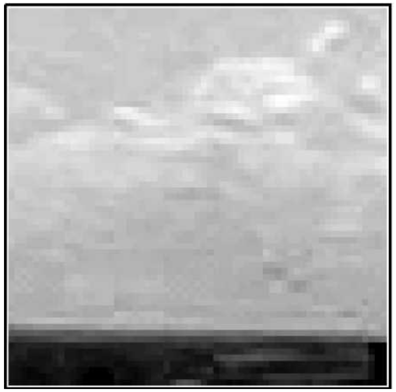


While the decreased intensity due to jitter (pink arrow) is almost as low as the drop due to a dark vehicle passing through the pixel (blue arrow), the detector responds differently to jitter and signal.

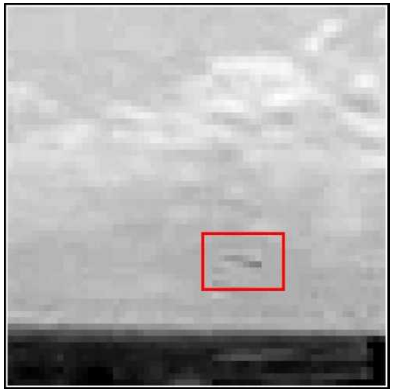
Detecting Birds in Flight

Detected Pixels, frames 2891 - 2900

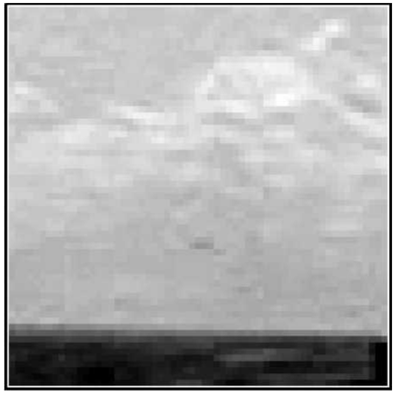
Frame 2891



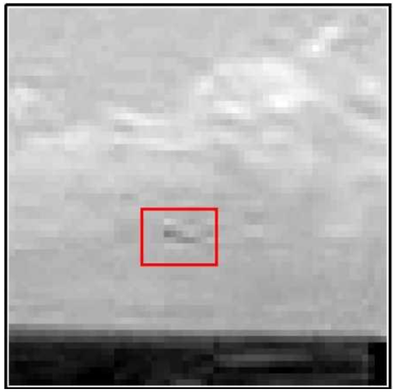
Frame 2892



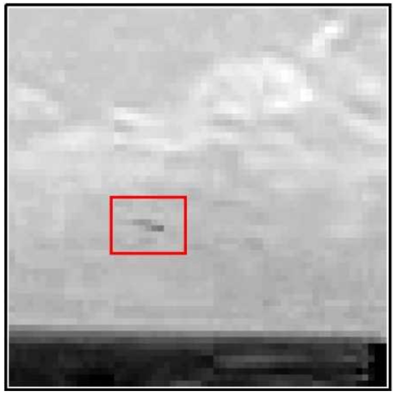
Frame 2893



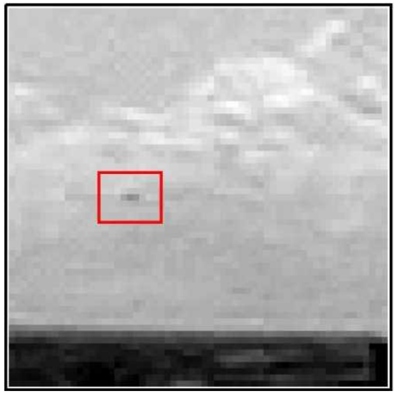
Frame 2894



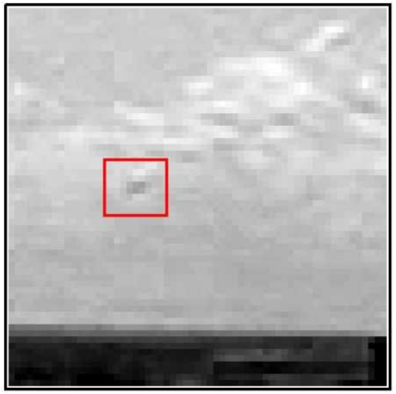
Frame 2895



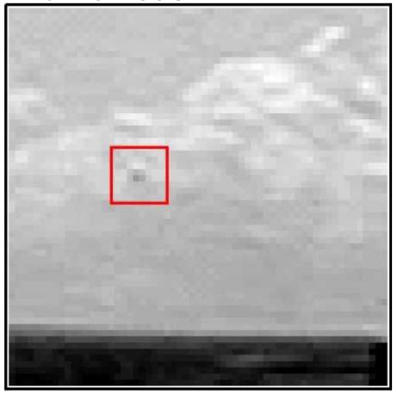
Frame 2896



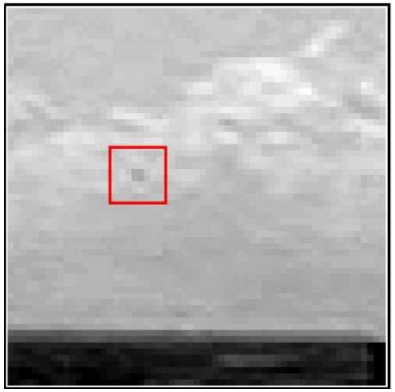
Frame 2897



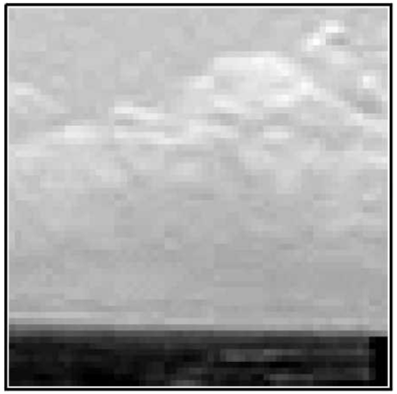
Frame 2898



Frame 2899



Frame 2900



A bird in flight is detected in seven frames.

Example 2 - Border Camera Footage

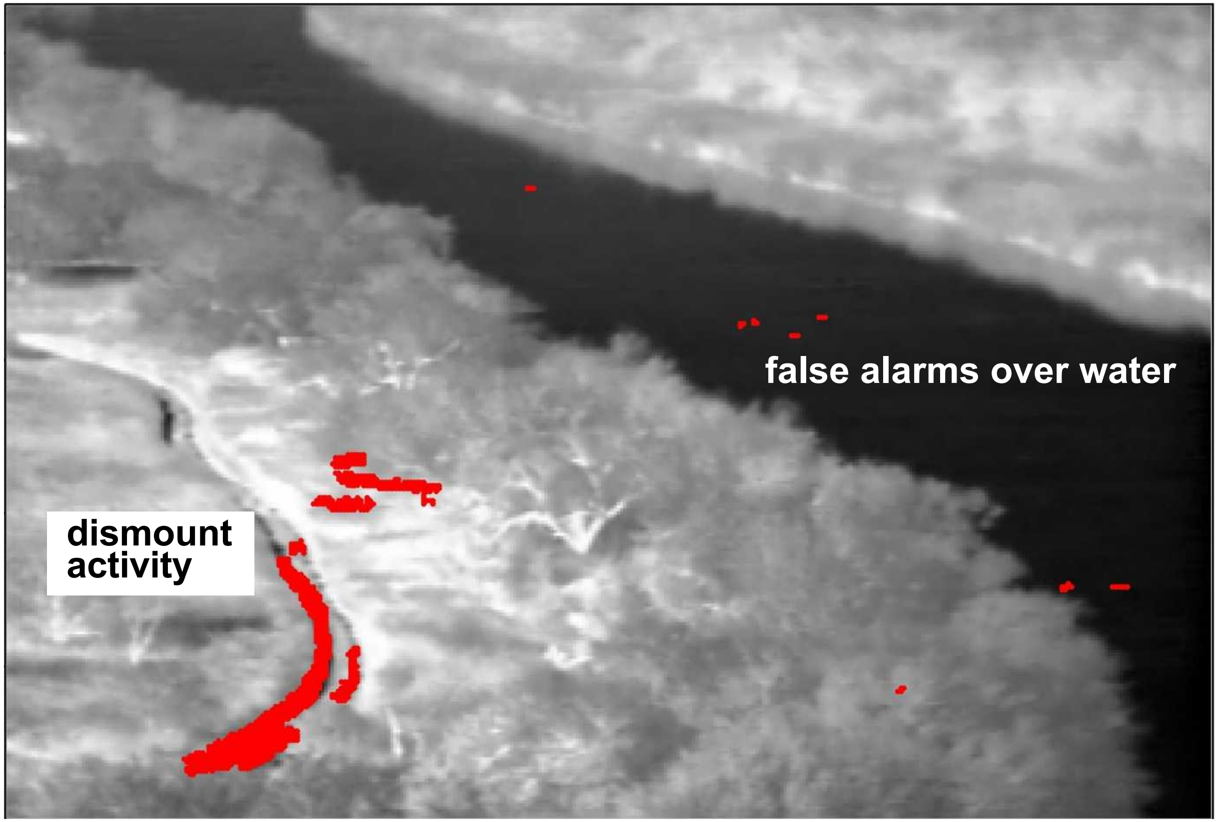
- **Video from a surveillance camera on the Texas/Mexico border.**
  - **- Downloaded from "Virtual Border Watch", a live video streaming website operated by the Texas Border Sherriff's Coalition and Bluservo.net.**
  - **- Network of pole-mounted surveillance cameras operating in the visible during daytime hours and infrared at night.**

Nighttime Scene Along River

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.



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.

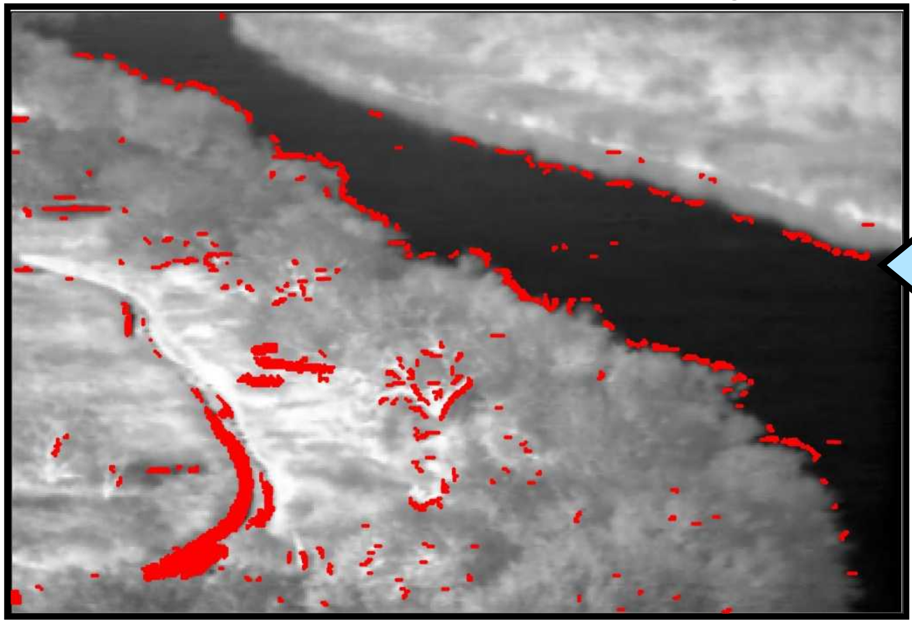
At times, they are lost in the near-saturated pixels to the right of the track.

Detection threshold = 6.0

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

Single-Model Detections

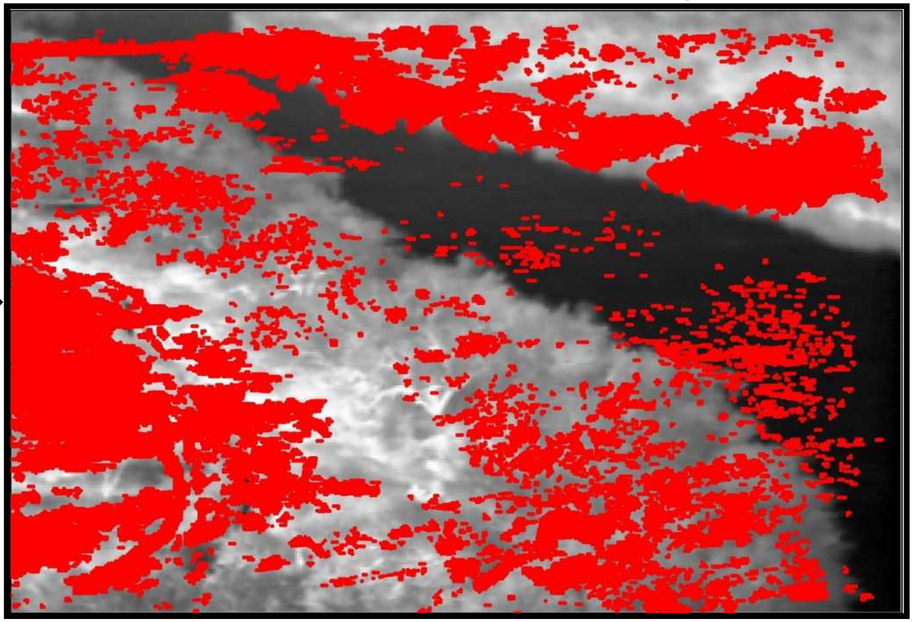
Detections, Temporal Variances Only



Frames 500 - 1500
Detection Threshold = 6.0

When only temporal estimates of pixel variance are available, false alarms occur at scene edges: riverbanks and tree trunks.

Detections, Spatial Variances Only



With only spatial standard deviation estimates, scene and sensor noise induce false alarms in relatively uniform parts of the scene.

Border Video with Detections

Red boxes indicate pixel detections; no tracker is applied.

Example 3 – ZooCam

- 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.

time stamp cropped for analysis

moving water

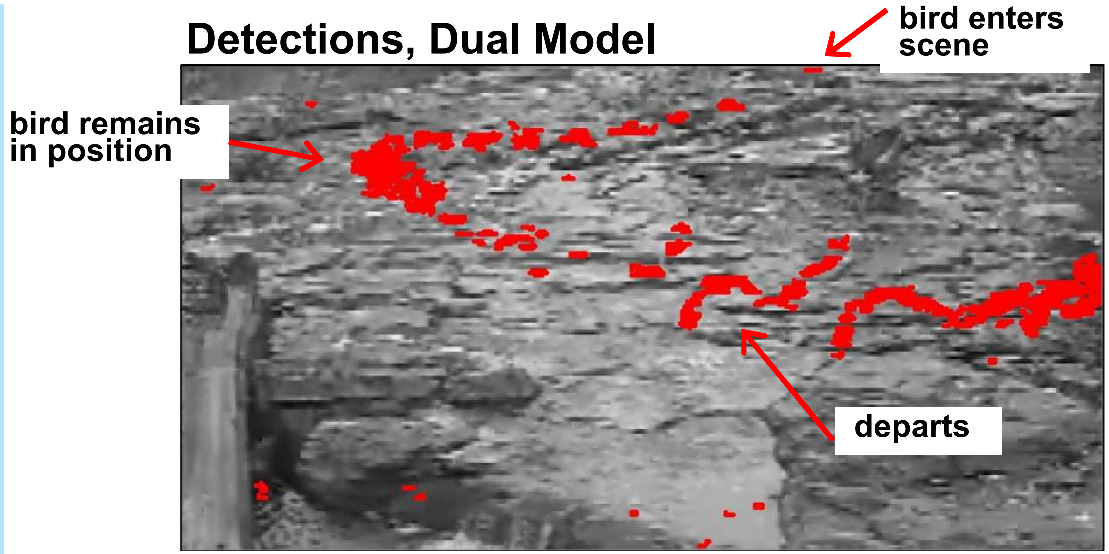
small waterfall

ZooCam Detections

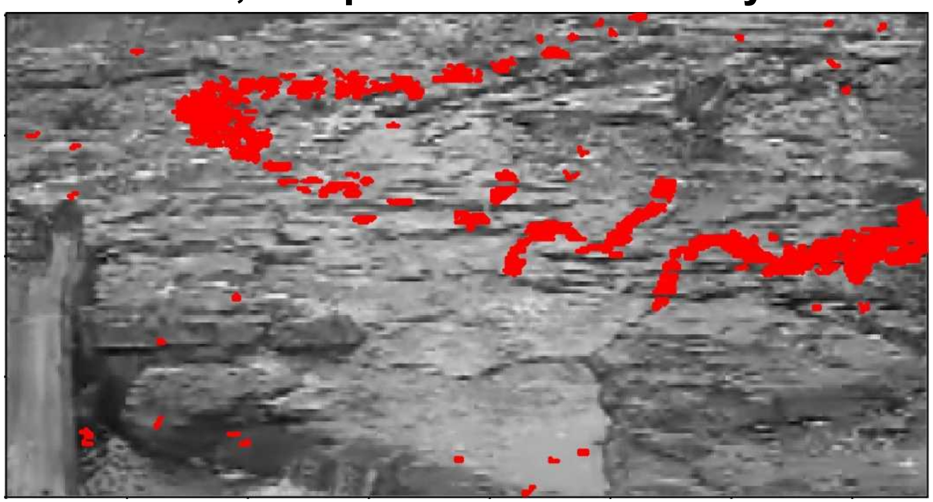
A small dark bird enters the scene in frame #3412 and departs in frame #3669.

With no camera jitter, temporal standard deviation estimates suppress most false alarms. The spatial estimates fail to account for pixel variability on the moving water pixels.

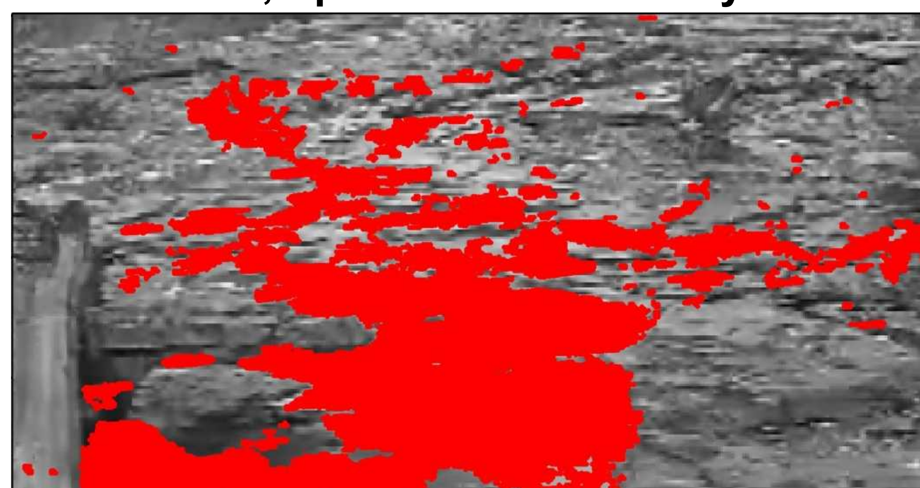
Detections, Dual Model



Detections, Temporal Variances Only



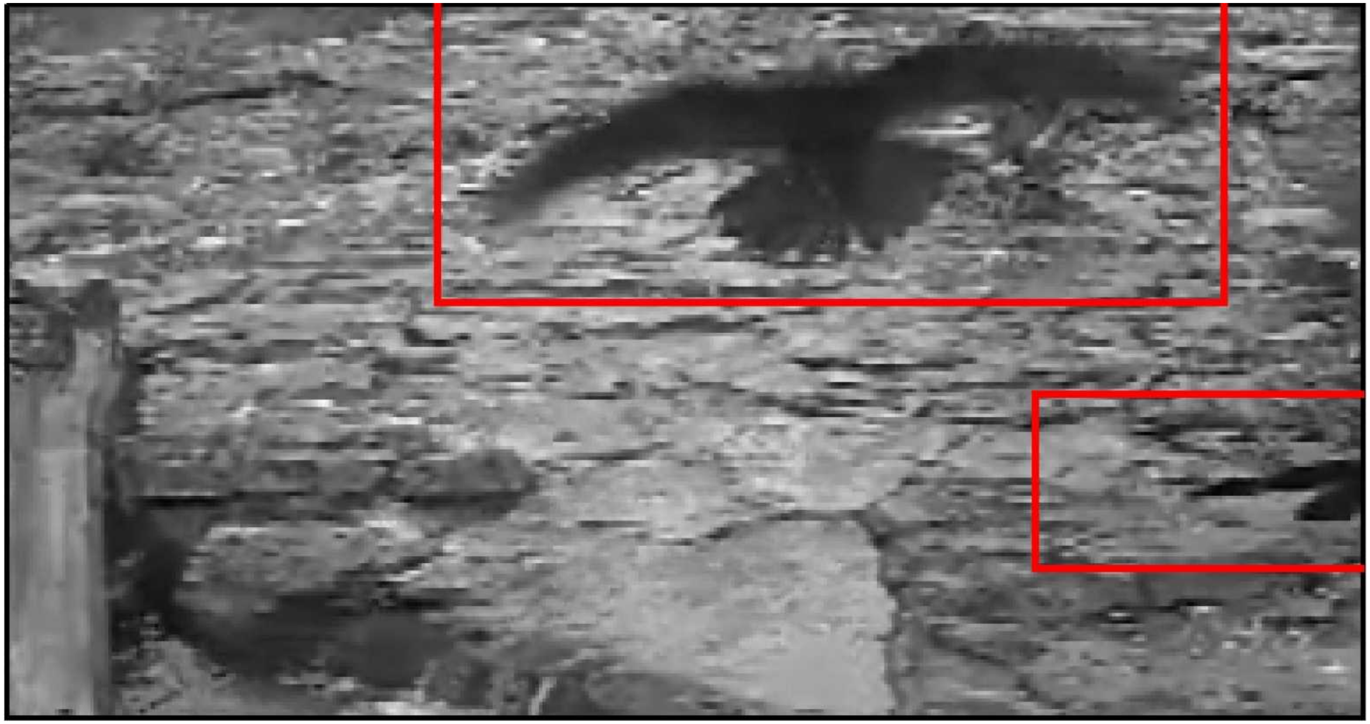
Detections, Spatial Variances Only



Detections shown for frames 3380 – 3675, Threshold = 8.0

Bird in Foreground

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

ZooCam Video With Detection Boxes

Frames 3380 – 3700, Threshold = 8.0

Red boxes indicate pixel detections; no tracker is applied.

Outline

- Background and Mission Constraints
- Mathematics of the Detector
- A Few Examples
- Summary and Conclusions

Run-Time Performance

- ***Real-time implementations*** **of the detection algorithm described here have been utilized for a variety of applications over the past several years.**

    - The software has run on frames as large as 2k × 2k, at frame rates up to 75 Hz.

- **Scene background and temporal variance estimates are *efficiently updated after every frame.***

    - For large frames, adaptive subspace estimation (FAPI) processing runs on specialized GPU hardware.

- **Spatial variance estimates are currently updated once per second.**

    - Sufficient for slowly-changing background or gradual pointing drift.

    - We are planning upgrades to a higher refresh rate, to enable *robust change detection even in the presence of fast pointing drift*.

Summary

- **The algorithm outlined here is designed to provide robust change detection, even in the presence of platform jitter, pointing drift, and significant sensor artifacts.**

- **The three key elements are:**

  1. ***Background modeling*** **via adaptive subspace estimation;**
  2. ***Temporal variance estimates*** **to track historical change;**
  3. ***Spatial variance estimates*** **to model susceptibility to jitter and/or pointing drift.**

- **The approach has *proven performance in real-world operations*.**

- **Sandia was granted a U.S. Patent for the spatial variance estimation 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.

BACKUP SLIDES

Detection Parameters

DETECTOR PARAMETER:	*SETTINGS, KAFB & Border*	*SETTINGS, BearCam*
FAPI Decay Rate	0.975	0.975
FAPI Decay Rate, Suppressed	0.99	0.99
Variance Decay Rate	0.99	0.99
Variance Decay Rate, Suppressed	1.0	1.0
Detection Threshold	6.0	8.0
Background Suppression Threshold	6.0	6.0
Variance Suppression Threshold	3.0	3.0
Jitter Standard Deviation	2.0	0.25
Connected Neighbors	3	5

The same parameter values were used for the KAFB and border videos. However, a one-sided (positive deviations only) threshold was used for the infrared border data, while a two-sided (absolute value) threshold was applied for the visible KAFB data. For the Bear Cam example, the detection threshold was increased and the jitter standard deviation was decreased.

References

Adaptive Subspace Estimation:

1. R. Badeau, B. David, and G. Richard, "Fast Approximated Power Iteration Subspace Tracking", *IEEE Transactions on Signal Processing*, vol. 53, no. 8, pp. 2931-2941, 2005.

2. P.E. Barry and M. Klop, "Jitter Suppression: A Data Processing Approach," *Proceedings of the SPIE*, vol. 366, pp. 2-9, 1983.

3. P. Comon and G.H. Golub, "Tracking a Few Extreme Singular Values and Vectors in Signal Processing," *Proceedings of the IEEE*, vol. 22, no. 8, pp. 1327-1343, 1990.

4. X.G. Doukopoulos and G.V. Moustakides, "Fast and Stable Subspace Tracking," *IEEE Transactions on Signal Processing*, vol. 56, no. 4, pp. 1452-1465, 2008.

5. E. Oja and J. Karhunen, "On Stochastic Approximation of the Eigenvectors and Eigenvalues of the Expectation of a Random Matrix," *J. Math Anal. and Appl.*, vol. 106, iss. 1, pp. 69-84, 1985.

6. N.L. Owsley, "Adaptive Data Orthogonalization," *Proceedings of the IEEE Conference on Acoustics, Speech, and Signal Processing*, pp. 109-112, 1978.

7. T.D. Sanger, "Optimal Unsupervised Learning in a Single-Layer Linear Feedforward Neural Network," *Neural Networks*, vol. 2, pp. 459-473, 1989.

8. B. Yang, "Projection Approximation Subspace Tracking," *IEEE Transactions on Signal Processing*, vol. 43, pp. 95-107, 1995.

Sandia's Anomaly Detector:

1. K.M. Simonson and T.J. Ma, "Robust Real-Time Change Detection in High Jitter," Technical Report SAND2009-5546, Sandia National Laboratories, Albuquerque NM, 2009.