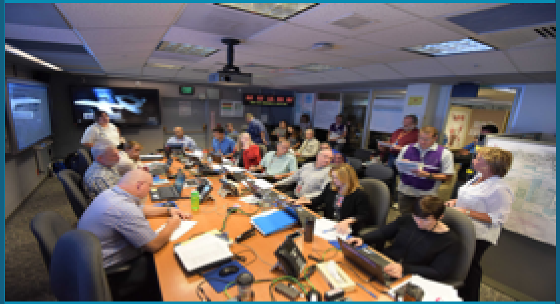
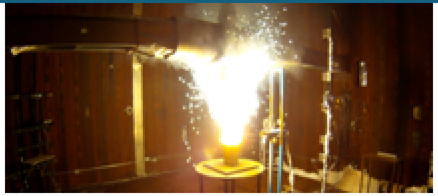
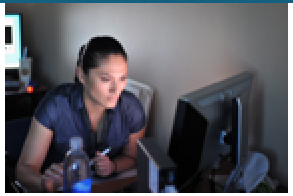
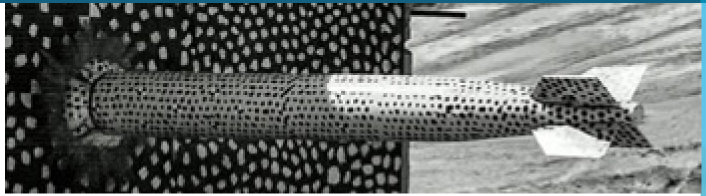
SAND2021-14135PE

Remote Sensing Target Detection Using Constraints









Presented by: Tian J. Ma

Distinguished Member of R&D, S&E Staff

Sandia National Laboratories

505-284-1238

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

- Problem Overview
- Background
- Method
- Result
- Conclusion

Problem Overview

- Remote Sensing Big Data are earth observing data continuously obtained from remote sensors (e.g. satellite, cameras, drones, RADARs, etc.)
  - Big Data Characteristics: Volume and Velocity

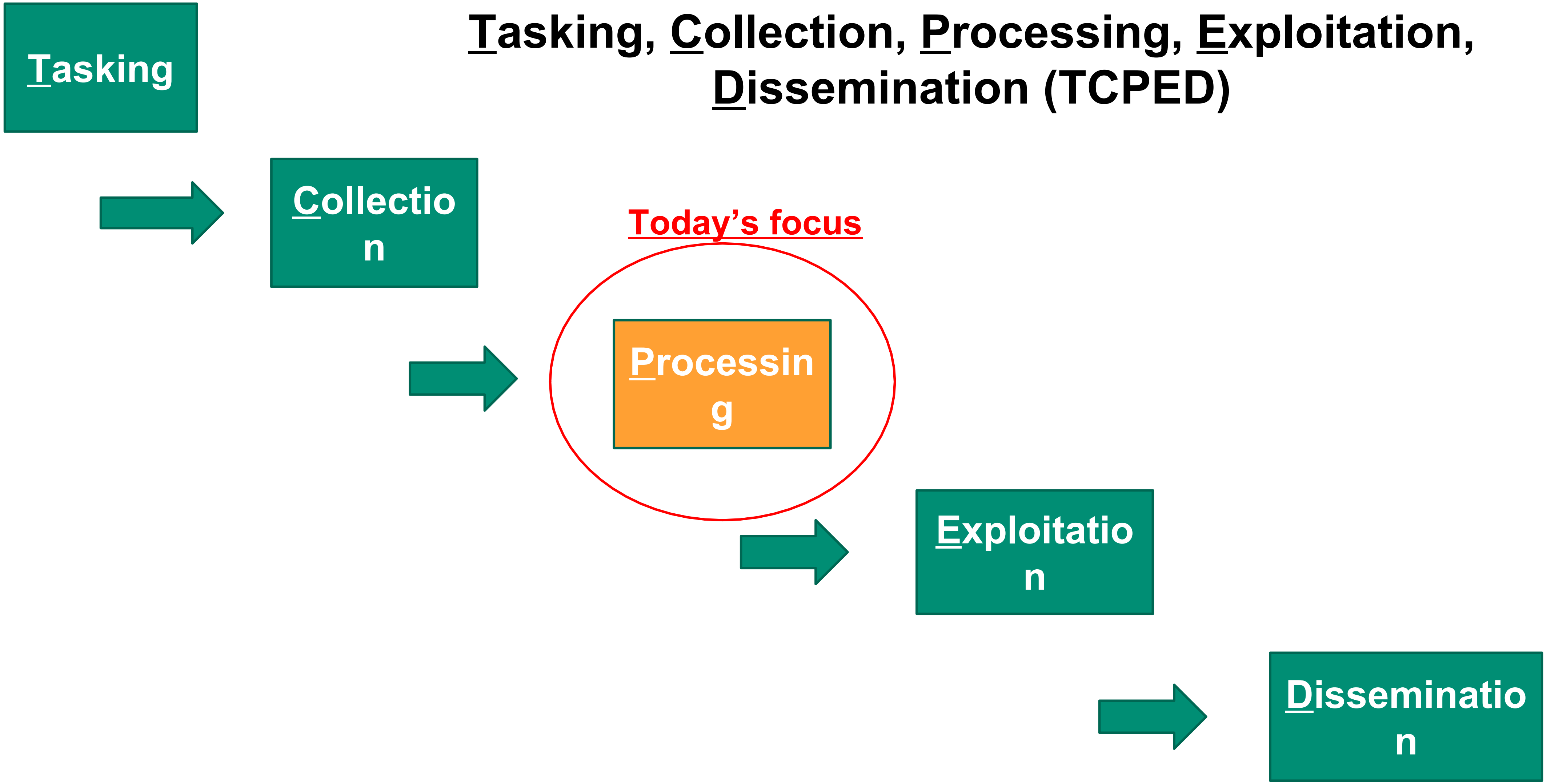
- The capability of detecting objects of interests and tracking them as they move is important to many critical and challenging national security missions

- Common application: home/business surveillance, environmental monitoring, autonomous sensing, etc..

- Key Challenges:
  - Computation Processing
    - High volume of data and detection rate (large field-of-view)
    - Real-time processing requirement
  - Small Object Detection
    - Difficult to detect far away objects (lack of spatial features)

General Framework for Remote Sensing System

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



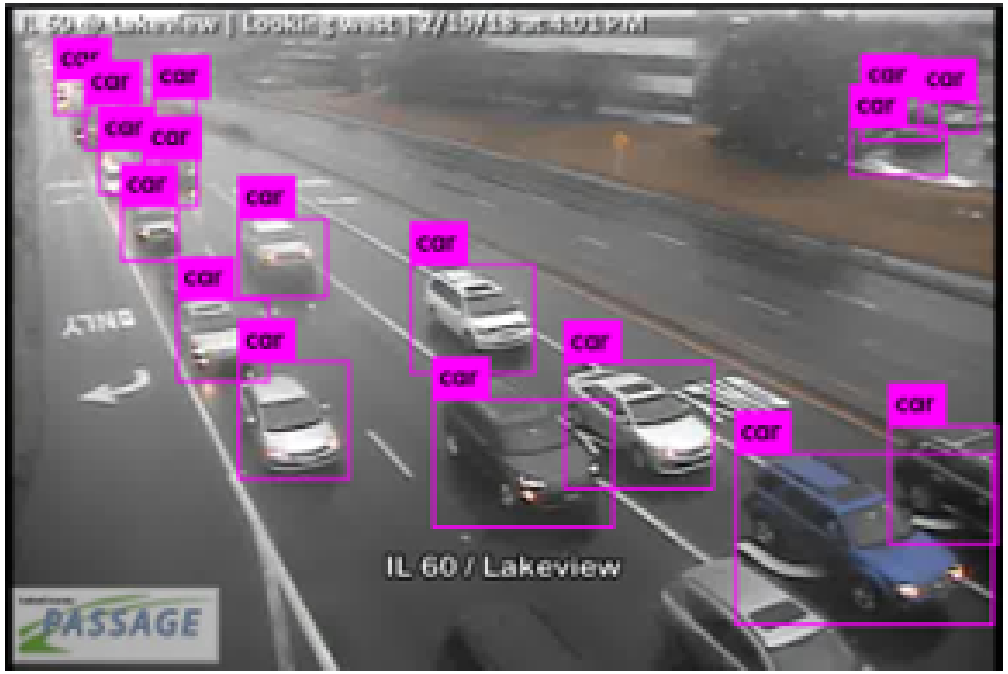
Artificial Intelligence (AI) Processing

- Machine learning and Deep Learning Techniques

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

- Advantages
  - Easy to get started (e.g. TensorFlow, Caffe, PyTorch, etc..)
  - Requires deep quality features in training data
  - Large pre-trained labels
  - High accuracy
  - Fast decision (operates on one image frame)
    - Populate methods: You Only Look Once (YOLO), Mask R-CNN
- Disadvantages
  - Requires a large number of 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

Key Advantages:

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

Disadvantages:

- May require multiple frames to drive down false alarm rate

Image → Background Estimation and Subtraction → Foreground/Backgrou nd Discrimination → Detected Object

Research Problem: Detection of Small Moving Object

- Challenges
  - Large field-of-view sensor placed very far away
  - Small sized target (just a few pixels)
  - Adaptive algorithm (does not require pre-trained labels)
  - Low SNR target

Experiment: Remote Detection of Vehicles from Sandia Peak

Camera Specification

Video Camera	Frame Rates	Image Resolution	Lens focal length
Mysterium X	24 frames per second	3072x1620	72mm

Camera Location

Camera Location	Peak of Sandia Mountain, Albuquerque, New Mexico
Camera Location Elevation	10,379 feet
Ground Target Elevation	6060 feet

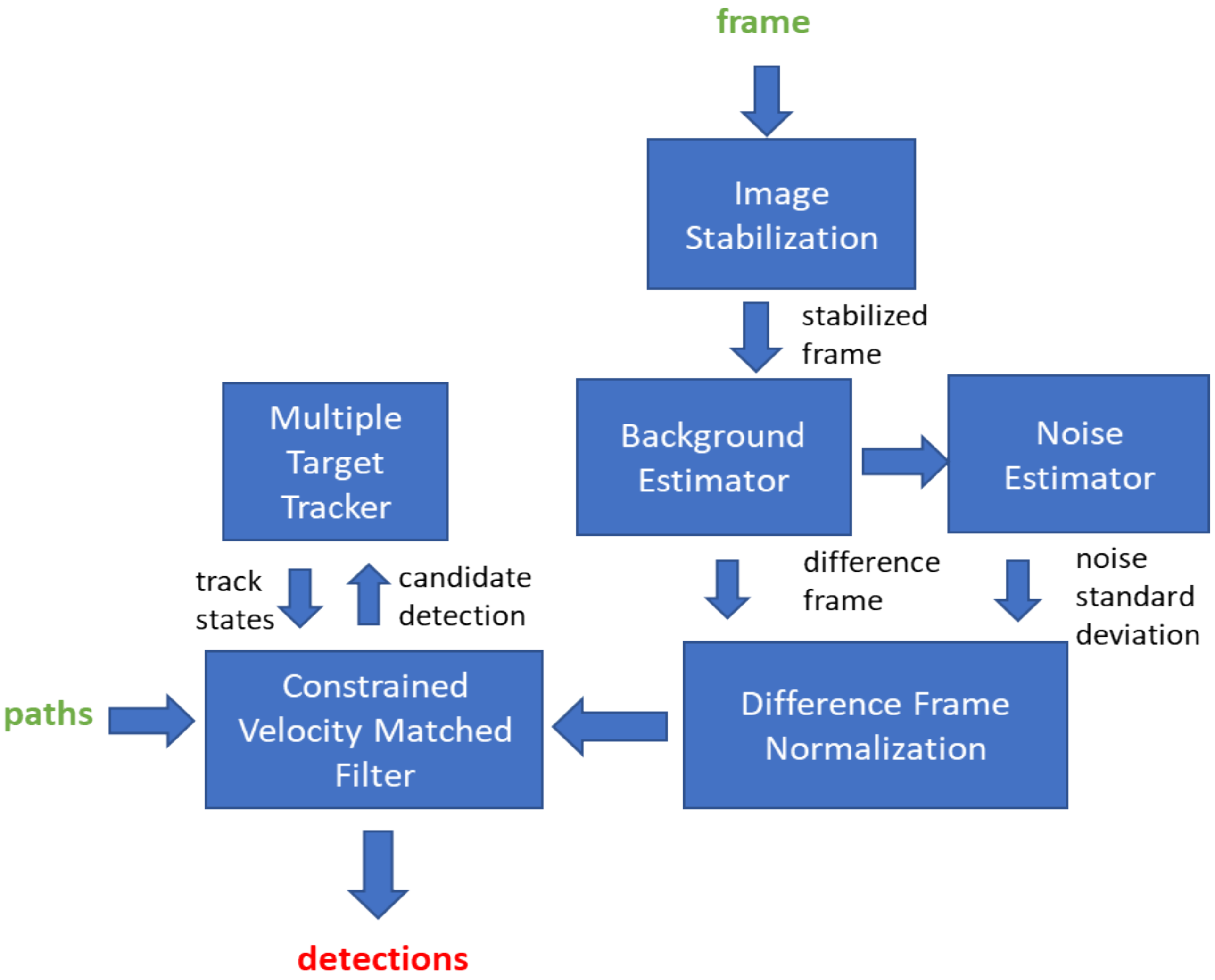
Cropped Raw Image

- **Distance to target is ~4000 ft.**
- **Size of the vehicles range from 4-20 pixels.**
- **Vehicles are barely visible to the human eye.**

Method: Detection Processing using Constraints

- **Key scientific contributions:**
  - An ideal "Normalized Difference Frame" calculation to perform velocity matched filter enhancement
  - Constrained Velocity Matched Filter (CVMF)
    - novel idea
    - combines known physical constraints with the target's dynamic motion constraints to enhance the target SNR

Algorithm Workflow



Normalized Difference Calculation

Let F_s (t) correspond to the stabilized frame at time t, and B(t-1) corresponds to the background computed in the previous time step.

The Difference Frame at time t, can be calculated using the following equation:

$$D(t) = F_s(t) - B(t-1) \quad (1)$$

The Temporal Variance v for each pixel at time t, can be calculated using an Infinite Impulse Response (IIR) filter with the following equation:

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

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

The Temporal Standard Deviation σ for each pixel at time t, can be obtained by using the following equation:

$$\sigma(i,j,t) = \sqrt{v(i,j,t)} \quad (3)$$

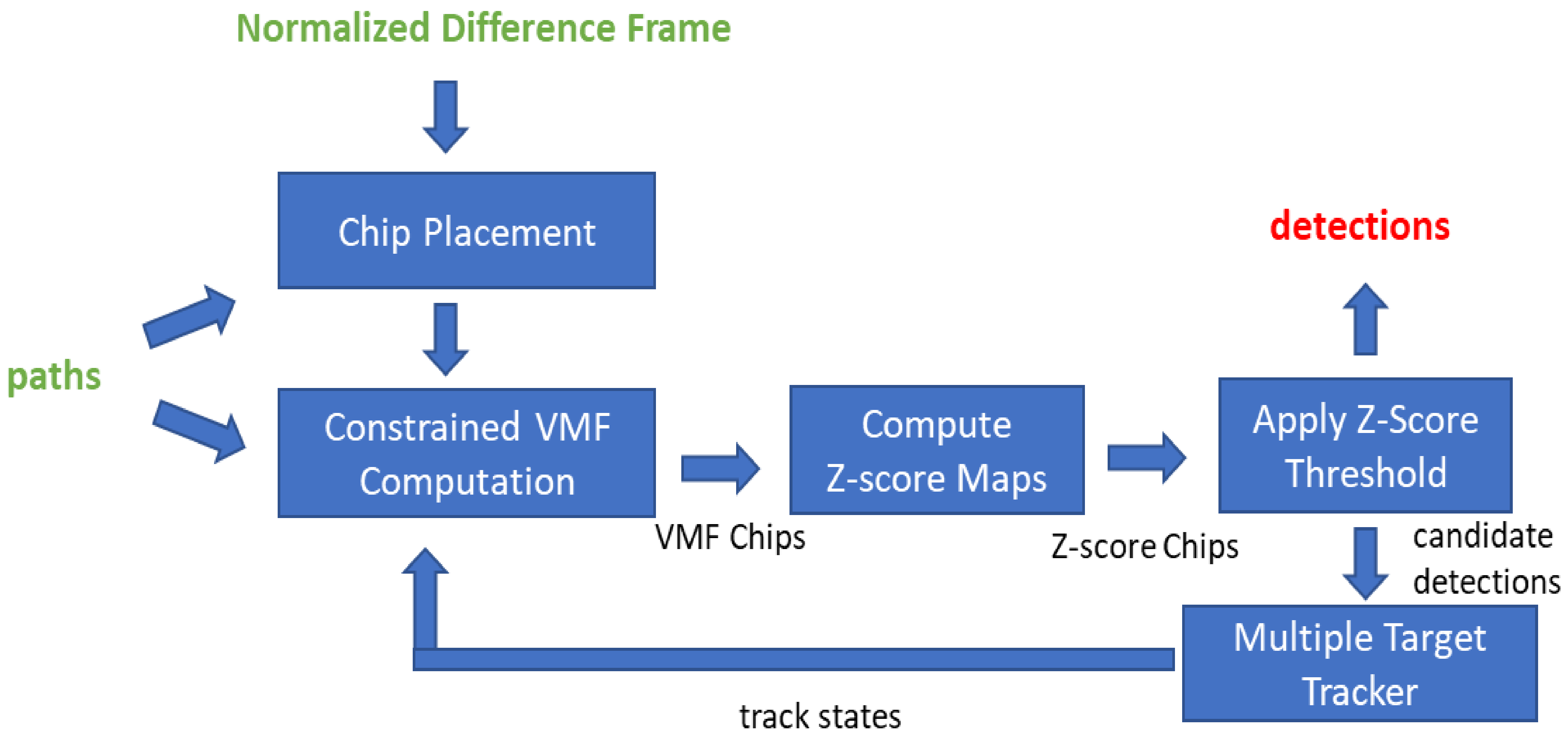
Normalized Difference Calculation (cont.)

The Normalized Difference Frame N_d for frame pixel location (i, j) in time t is expressed as follow:

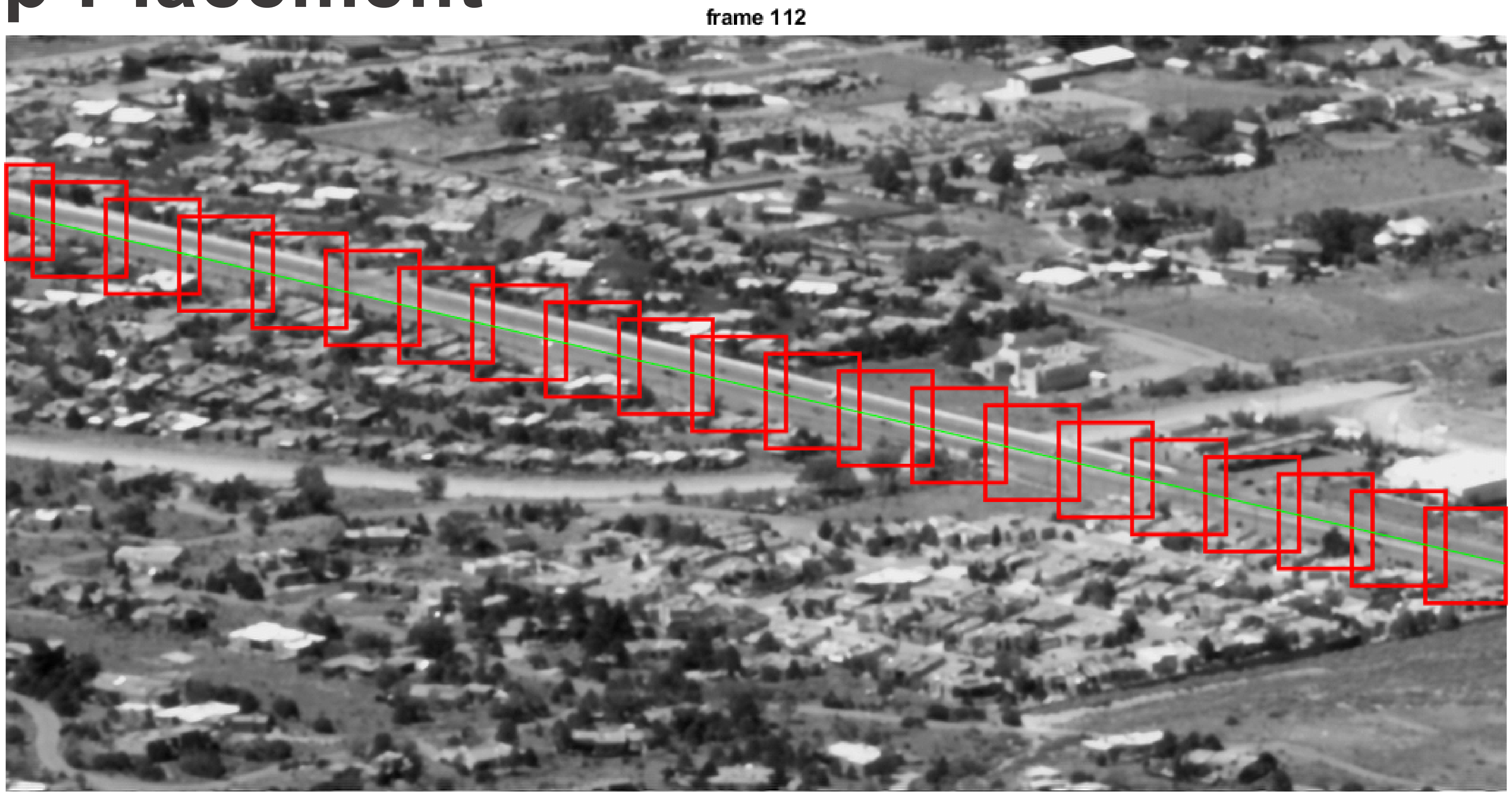
$$N_d = \frac{D(i,j,t)}{\sigma(i,j,t-1)} \quad (4)$$

Key Motivation: Pixels in different parts of an image can have different temporal standard deviation, depending on factors such as the environment and the scene structure. **It is important to normalize frame relative to account for variation of noise levels across an image.**

Constrained Velocity Matched Filter (CVMF) Process



Chip Placement

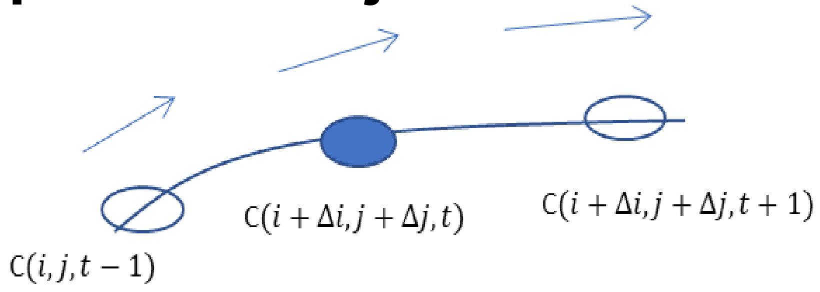


Strategy: Dividing the road into different number of regions (called “chips”)
Key Motivation: Enable parallel processing of individual chips and Z-score calculat

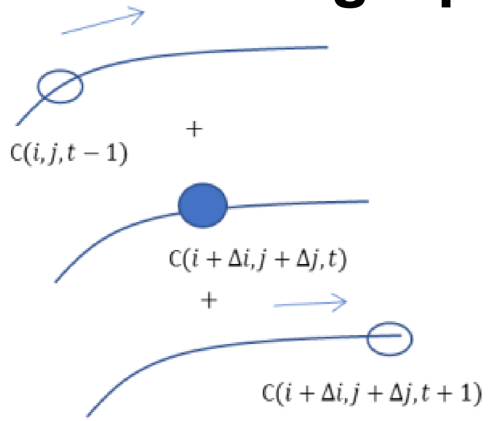
Velocity Matched Filter Processing on "Chips"

The continuous VMF process can be implemented in discreate form, by shift-and-add-operation with different velocity hypotheses along the path.

Example of an object's movement over time



Shifting and Adding Operation



Mathematically, this can be expressed as the following:

$$S_k(i,j,t) = C(i+\Delta i, j+\Delta j, t-w) + \cdots C(i,j,t) + \cdots C(i+\Delta i, j+\Delta j, t+w) \quad (5)$$

where S, is the summation of the pixel (i, j) across multiple frames. $(\Delta i, \Delta j)$ corresponds to the shift positions, and w, represents the frame window for the summation, and k corresponds to the index of the matched hypothesis.

Computation Analysis

The total number of matched hypothesis K can be expressed as:

$$K = M * N \quad (6)$$

where M is the number of directional hypotheses and N is the number of velocity hypotheses. Since the movement of the individual targets are constrained in a pre-determined path, M is 2 in most cases (either forward or backward direction). M can be greater than 2 when the chip is at an intersection. The number of velocities depends on the target's speed.

Processing of each individual chip can be done independently

Z-score Calculation

To find the detection in the sum chip S for a given hypothesis k, we first normalized the sum chip to form a Z-score chip. We can do this by computing the mean μ_s and standard deviation σ_s of the sum chip S. For dense target scenarios, it is recommended that a trim mean is used instead, to avoid high SNR targets inflating the mean estimates.

$$\mu_s = \frac{1}{p}\sum_{p=1}^{P} S(p) \tag{7}$$

$$\sigma_s = \sqrt{\frac{1}{P}\sum_{p=1}^{P} (S(p) - \mu_s)^2} \tag{8}$$

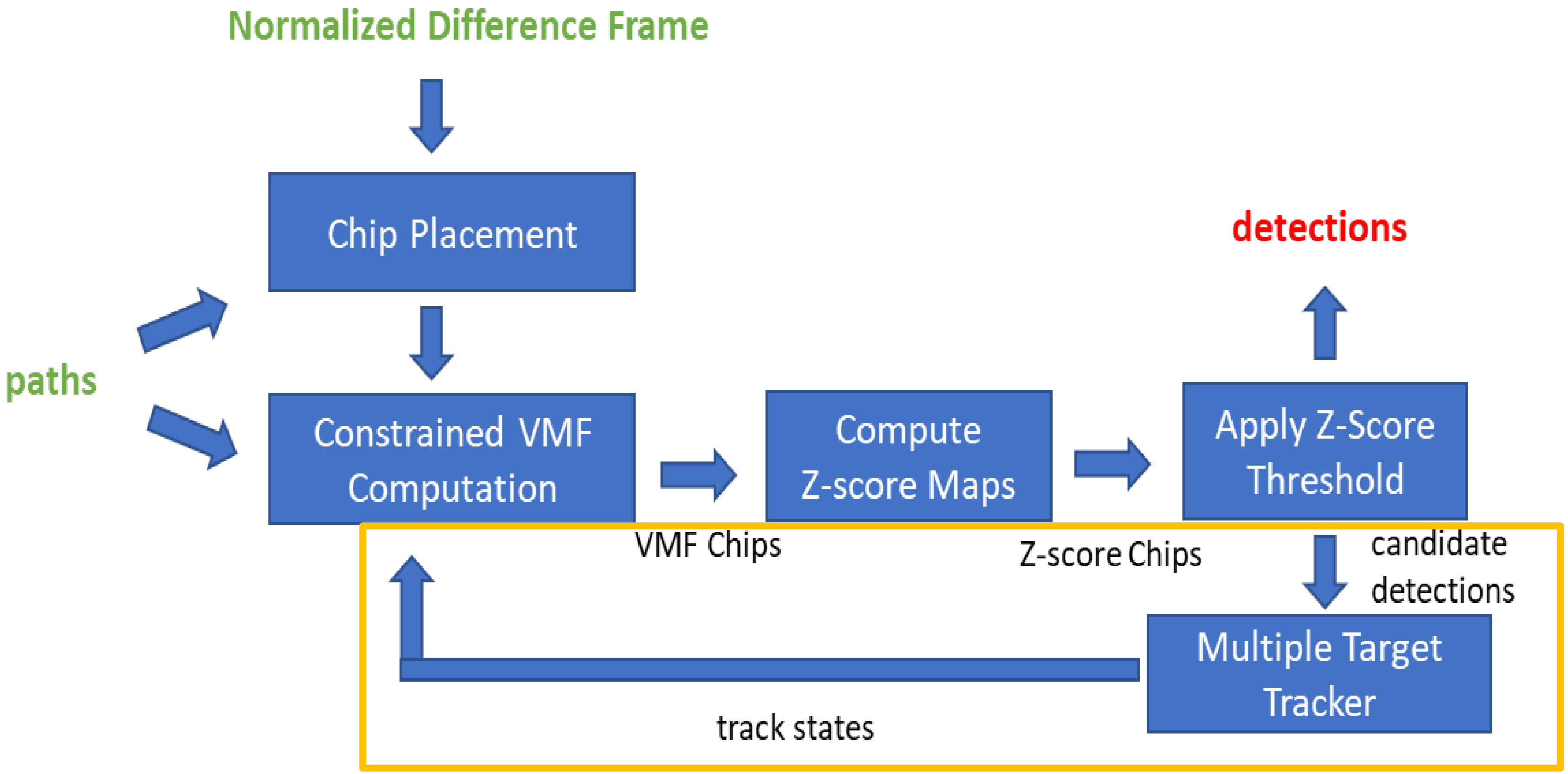
Then, we compute the Z score of the sum chip Z_s for each pixel (i,j) using the following equation:

$$Z_s(i,j) = \frac{S(i,j) - \mu_s}{\sigma_s} \tag{9}$$

Z-Score Thresholding

If $(|Z_s(i,j)| \geq T)$, then pixel (i,j) is a candidate detection.

Tracker Feedback



As the target(s) are being tracked, the state vectors $\hat{\boldsymbol{x}}$ associated with covariance $\boldsymbol{P}$ (motion constraint) are fed back to the CVMF process to fine tune the pre-defined velocity bins and improve the accuracy of matching.

Multiple Target Tracker

An object's dynamic movement can be expressed mathematically using the following equations:

$$x(t) = A\,x(t-1) + q(t-1), \qquad q(t) \sim N(0,Q)$$
$$y(t) = H\,x(t) + r(t), \qquad r(t) \sim N(0,R) \qquad (10)$$

where $\boldsymbol{x}$ corresponds to the state vector, $\boldsymbol{y}$ corresponds to the output vector, $\boldsymbol{A}$ corresponds to the system matrix, and $\boldsymbol{H}$ corresponds to the output matrix. The system includes additive process noise q and measurement noise r, which are modeled as white noise gaussian with zero mean.

The constant velocity model can be expressed in the following form:

$$x_1(t) = x_1(t-1) + \Delta T x_3(t-1) + q_1 \qquad (11)$$
$$x_2(t) = x_2(t-1) + \Delta T x_4(t-1) + q_2$$
$$x_3(t) = x_3(t-1) + q_3$$
$$x_4(t) = x_4(t-1) + q_4$$

State Equation in matrixed form

In matrix form, this can be expressed as:

(12)

$$\boldsymbol{x}(t) = \begin{bmatrix} 1 & 0 & \Delta T & 0 \\ 0 & 1 & 0 & \Delta T \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \boldsymbol{x}(t-1) + \boldsymbol{Q},$$

$$\boldsymbol{y}(t) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \boldsymbol{x}(t) + \boldsymbol{R}$$

where $\boldsymbol{Q}$, is the process noise matrix, and $\boldsymbol{R}$, is the measurement noise matrix. Kalman Filtering can be used to predict and update the state estimates and its covariance estimate $\boldsymbol{P}$ at each time step.

Kalman Filtering

Kalman Filtering is used to update states and covariance for each time step.

Prediction Steps

$$\hat{x}(k|k-1) = A\,\hat{x}(k-1|k-1) \quad (13)$$

$$P\,(k|k\text{-}1) = A\,P(k-1|k-1)A^T + Q$$

Updated Steps:

$$K(k) = P(k|k-1)H^T(HP(k|k-1)H^T + R)^{-1} \quad (14)$$

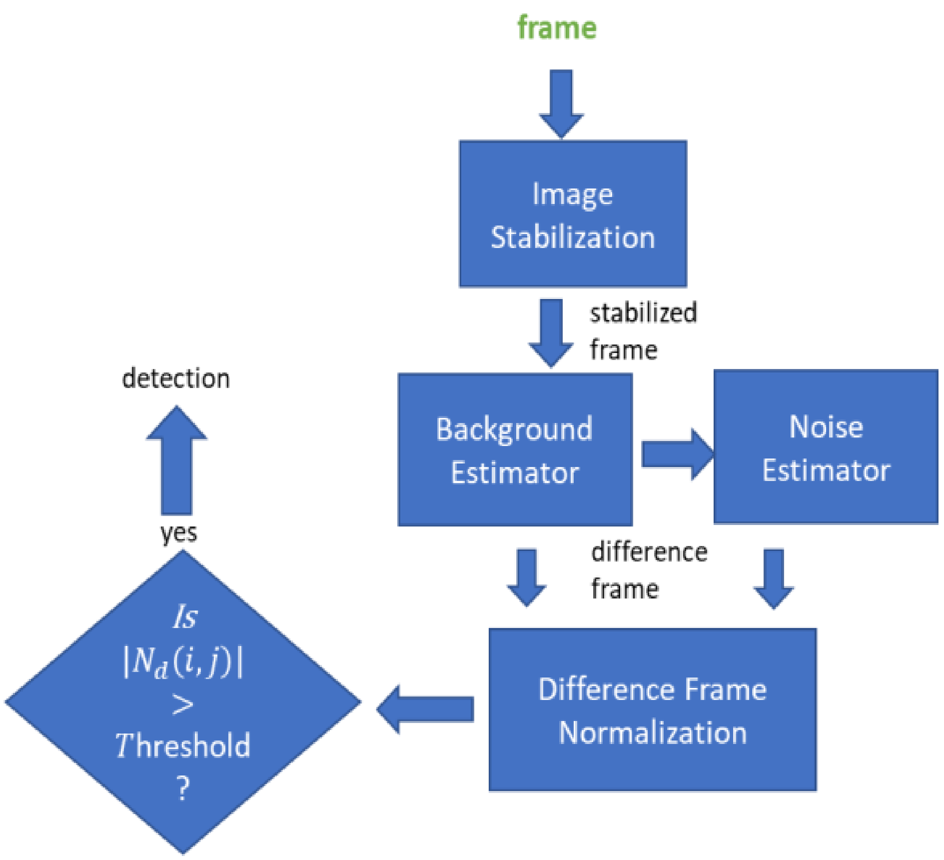
$$\hat{x}(k|k) = \hat{x}(k|k-1) + K\,(y(k) - H)\hat{x}(k|k-1)$$

$$P(k|k) = (I - K(k)H)P(k|k-1)$$

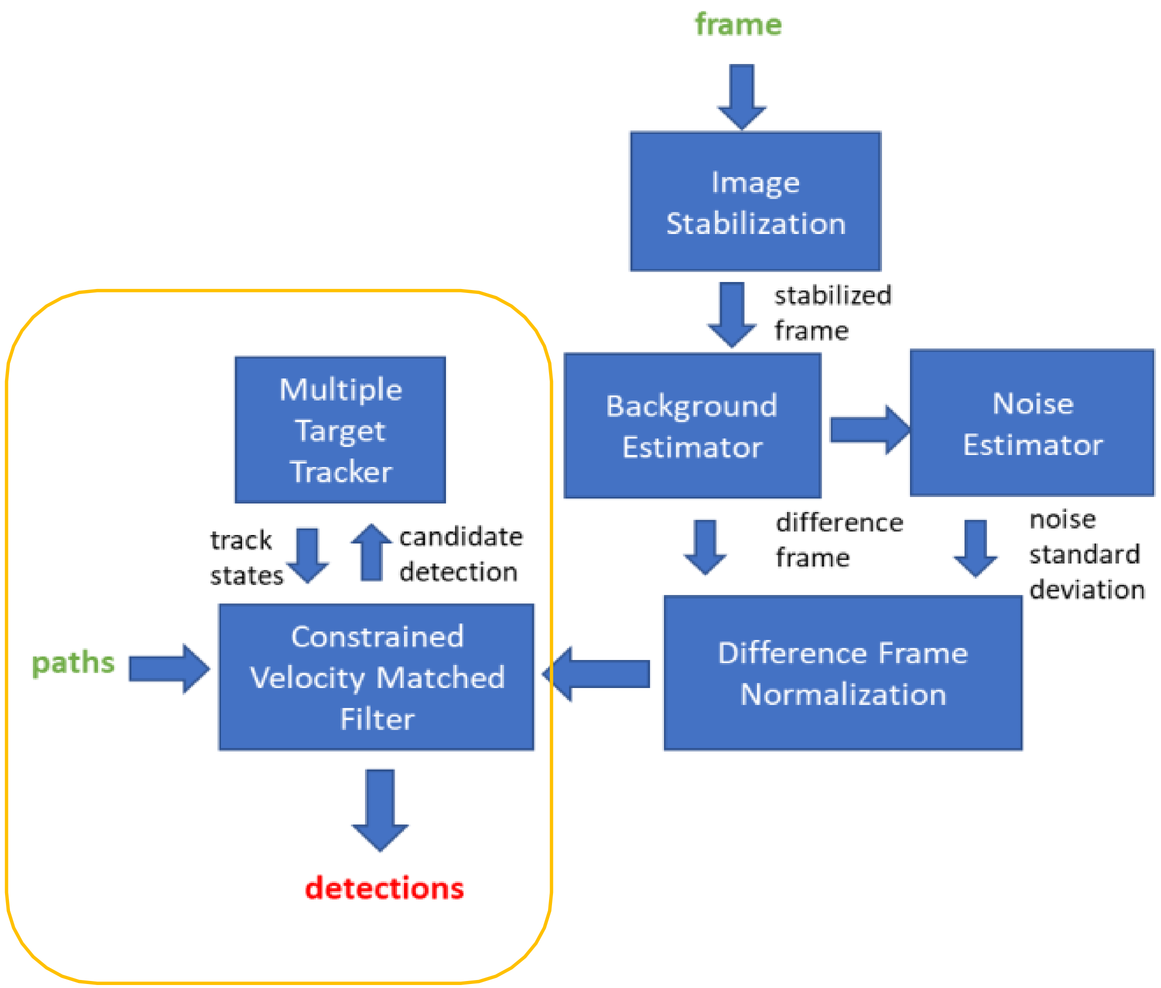
As the target(s) are being tracked, the state vectors $\hat{x}$ associated with covariance P (motion constraint) are fed back to the CVMF process to fine tune the pre-defined velocity bins and improve the accuracy of matching.

Benchmark Comparison

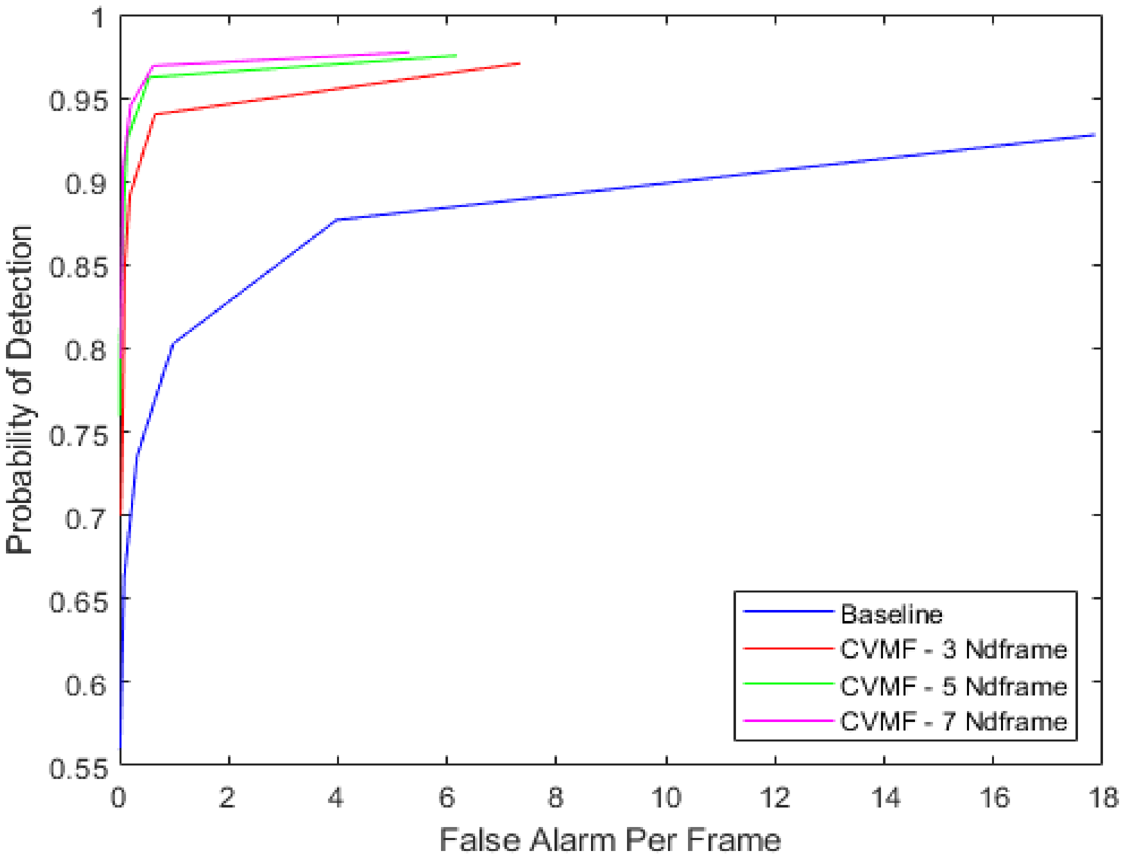
Baseline (No CVMF)

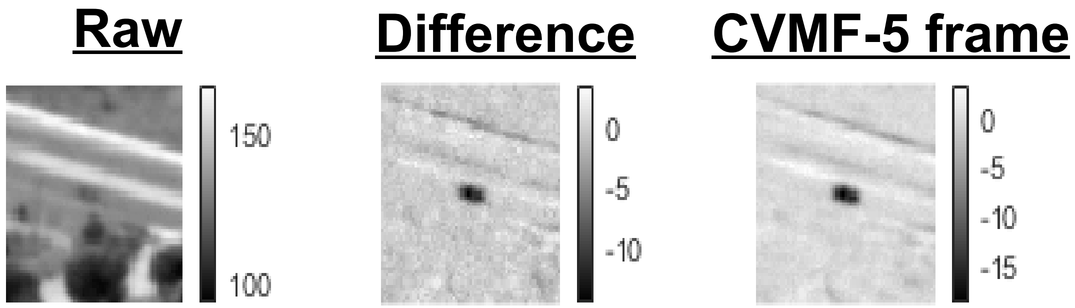


Baseline + CVMF



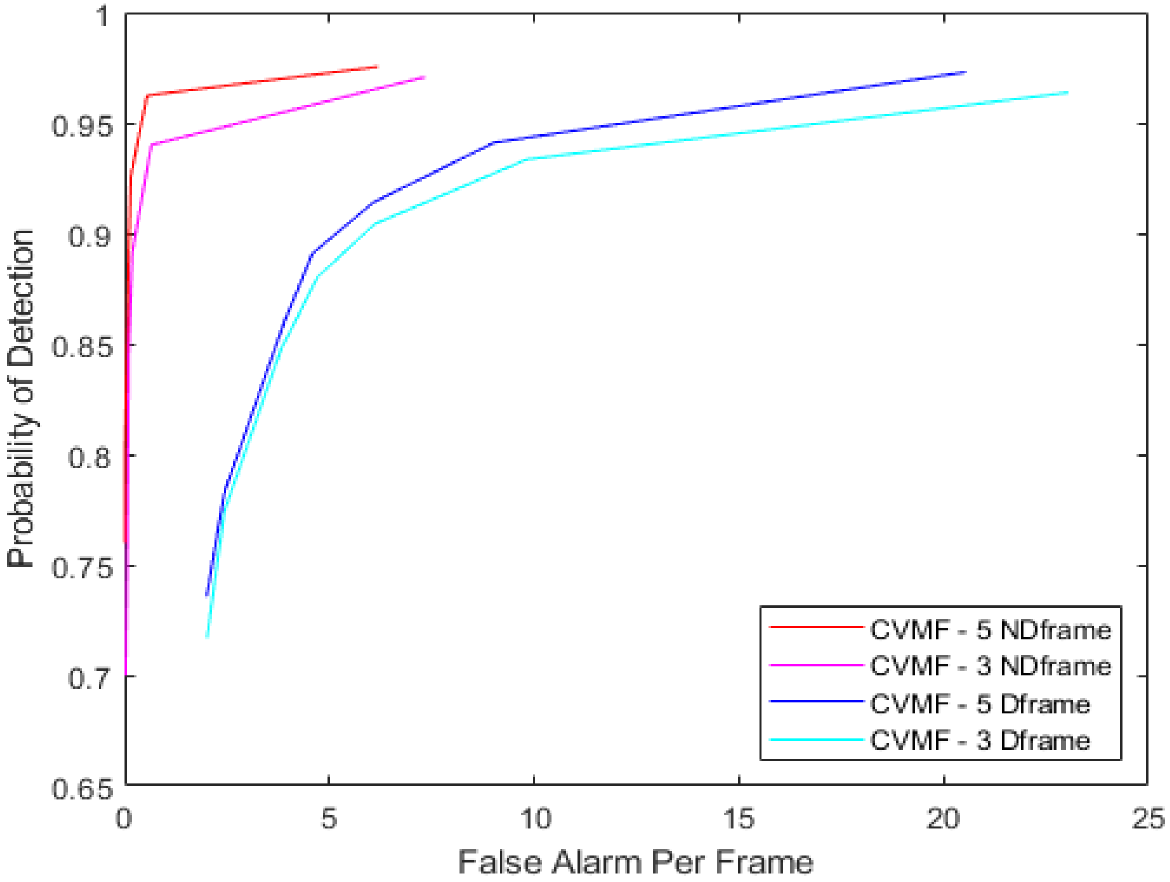
ROC curve comparison





The additional CVMF processing improves baseline ROC curve significantly

CVMF Normalized Difference vs CVMF Difference



Operating CVMF Normalized Difference Frame is much more desirable than on the Difference Frame.

Conclusion

- **Key Scientific Contribution**
  - The addition of CVMF processing significantly improves the ROC curves
  - CVMF should be operating on normalized difference frame

- **Reference Publication:**
  - Tian J. Ma, "Remote Sensing Detection Enhancement", Springer, Journal of Big Data, October 2, 2021, DOI :https://doi.org/10.1186/s40537-021-00517-8