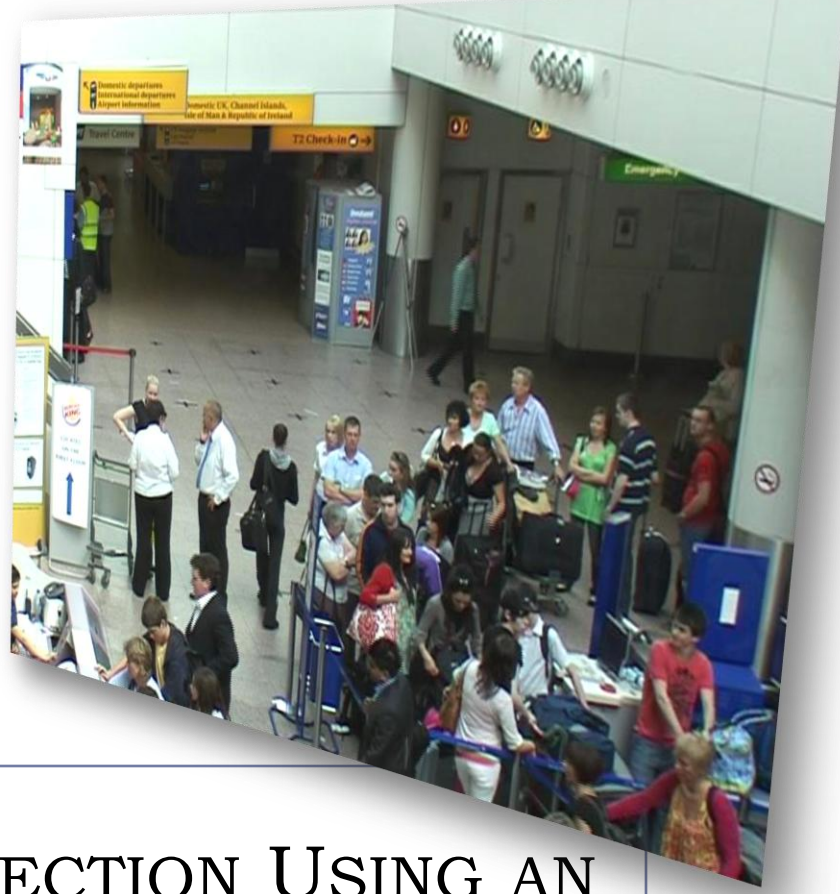
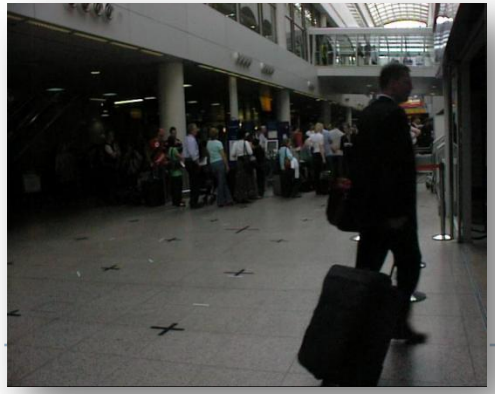


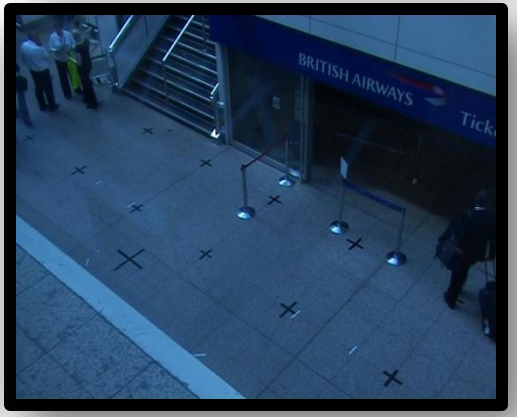
Event Detection Using an Attention-Based Tracker

22 Outubro 2007. Universidade Federal do Paraná.
Gerald Dalley, Xiaogang Wang, and W. Eric L. Grimson

Dataset Description

- Glasgow Airport
- 4 cameras
- 9 video clips
  - 1 for training
  - 8 for testing





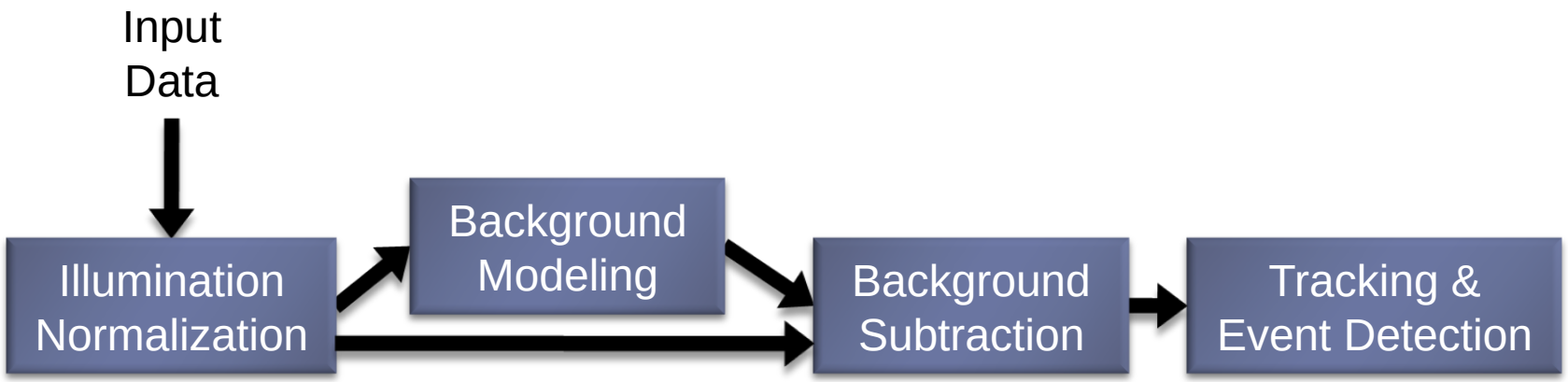
Challenge Problem

- Loitering
  - Staying in the scene for > 1 minute
- Theft
  - Taking someone else's baggage
- Left luggage
  - Leaving the scene without your luggage
- Evaluation
  - Correct event times
  - Correct 3D locations
  - Without false detections

Our Approach

- Basic tracking first
- Detect interesting tracks
- Improve interesting tracks to detect events

Processing Pipeline

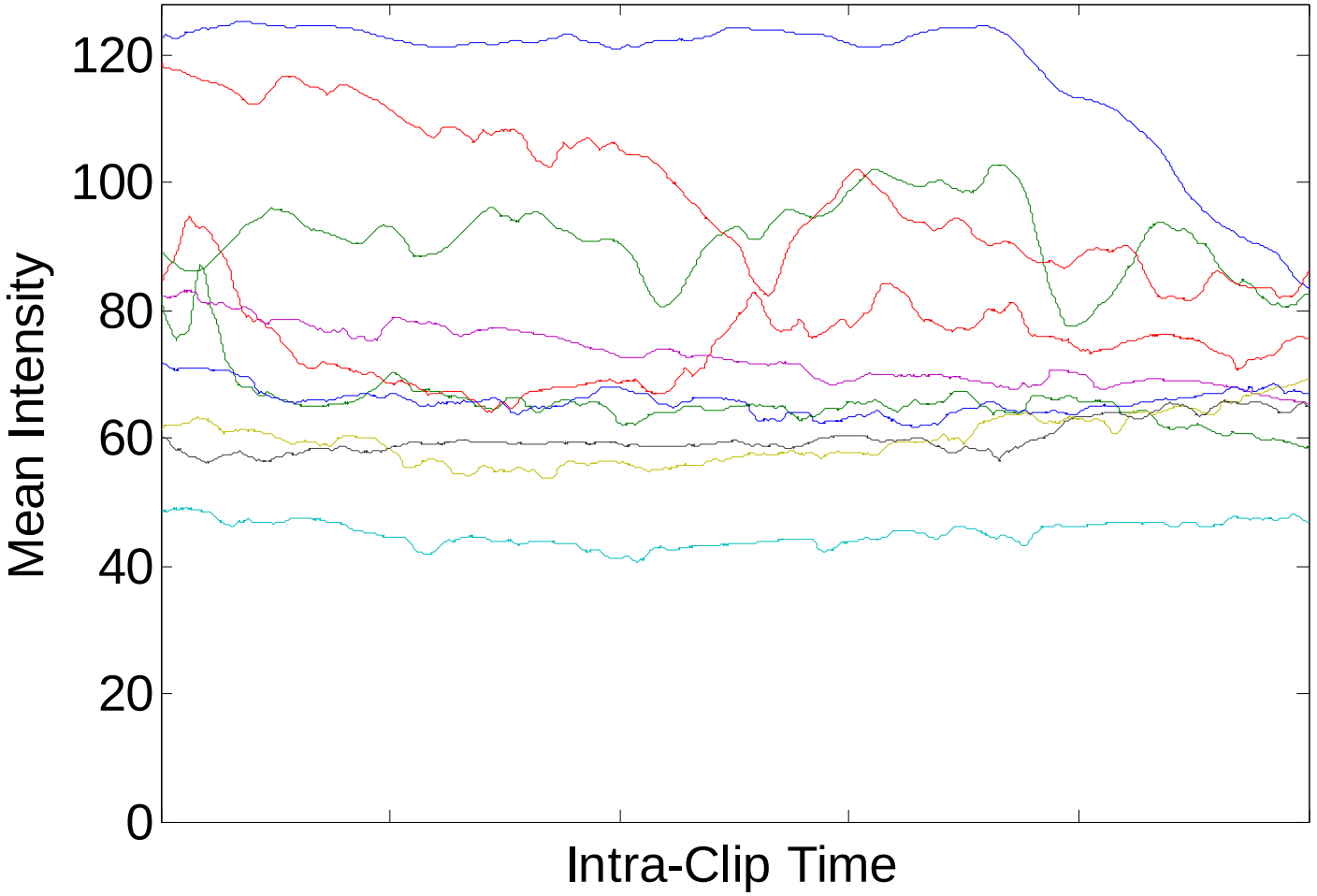


Illumination Changes

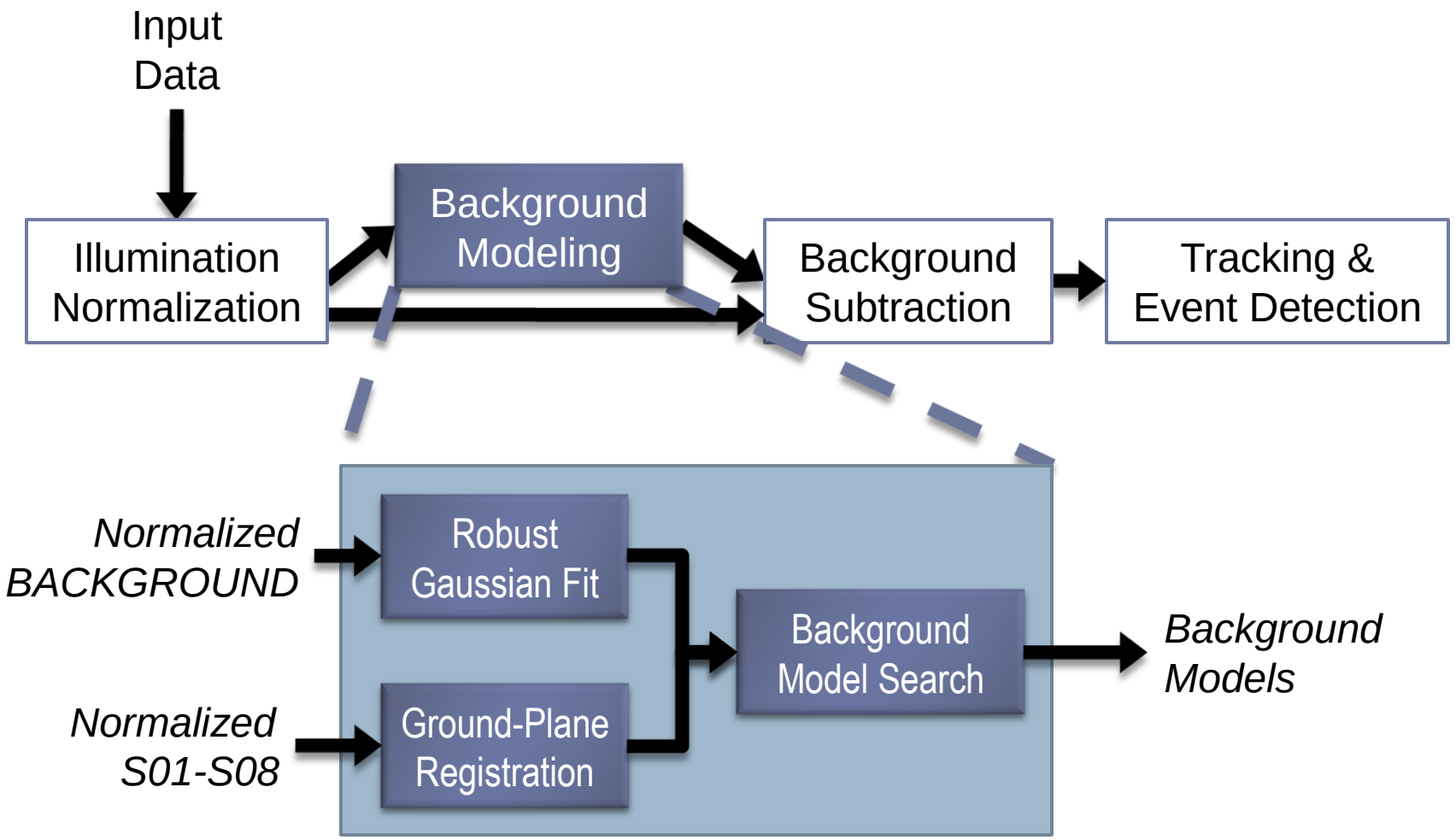
BACKGROUND

S08

Illumination Changes by Clip

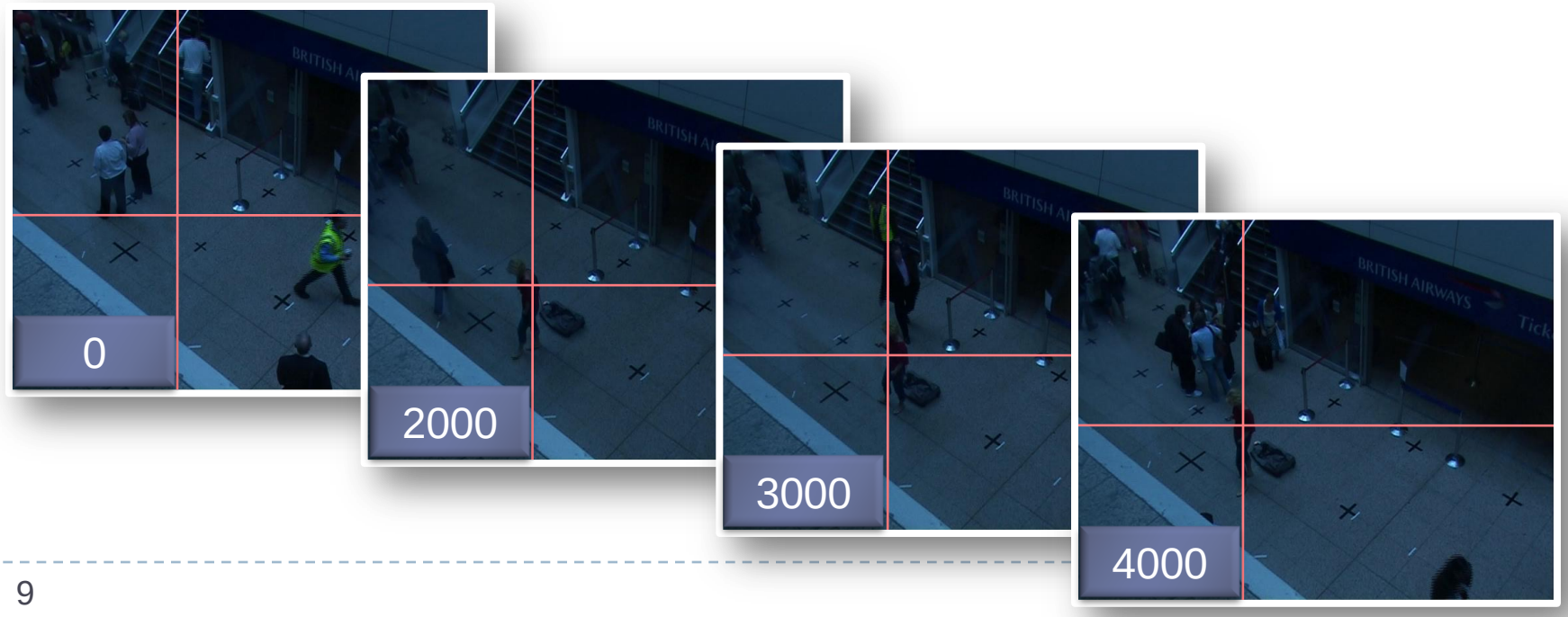


Background Modeling



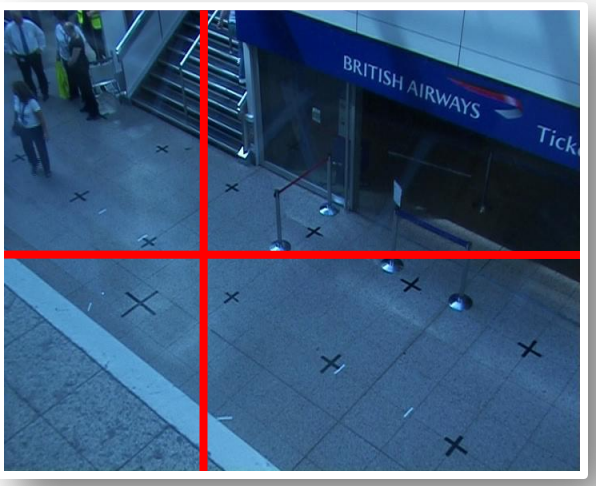
Breaking Adaptive Background Subtraction

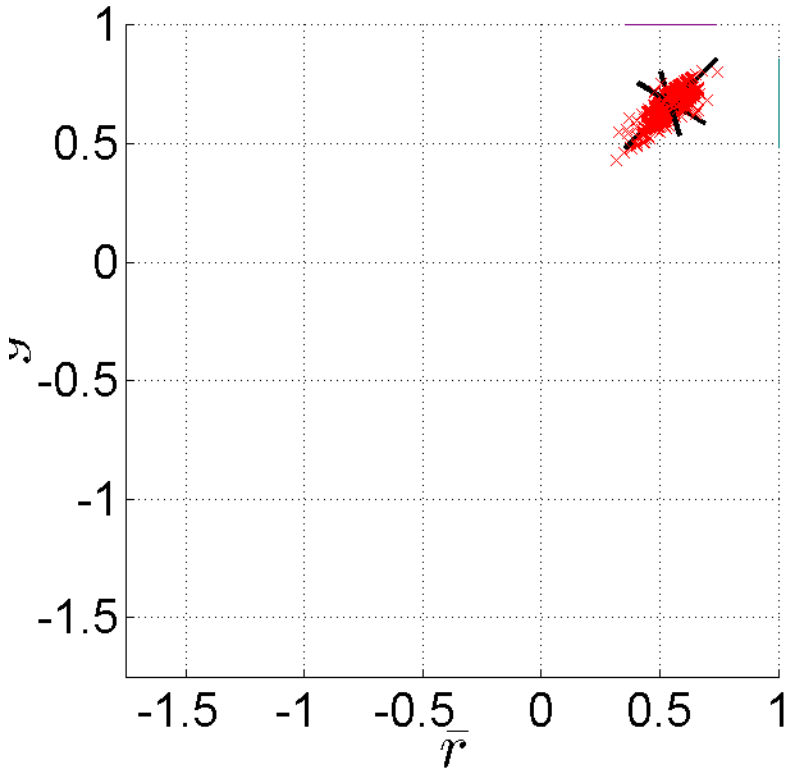
- Fundamental assumption
  - *foreground is rare at every pixel*
- Reality for PETS 2007…
  - *background is rare for the pixels we care about most*
    - Some pixels: **foreground** as much as **90% of the time**



Robust Gaussian Fit (per pixel)

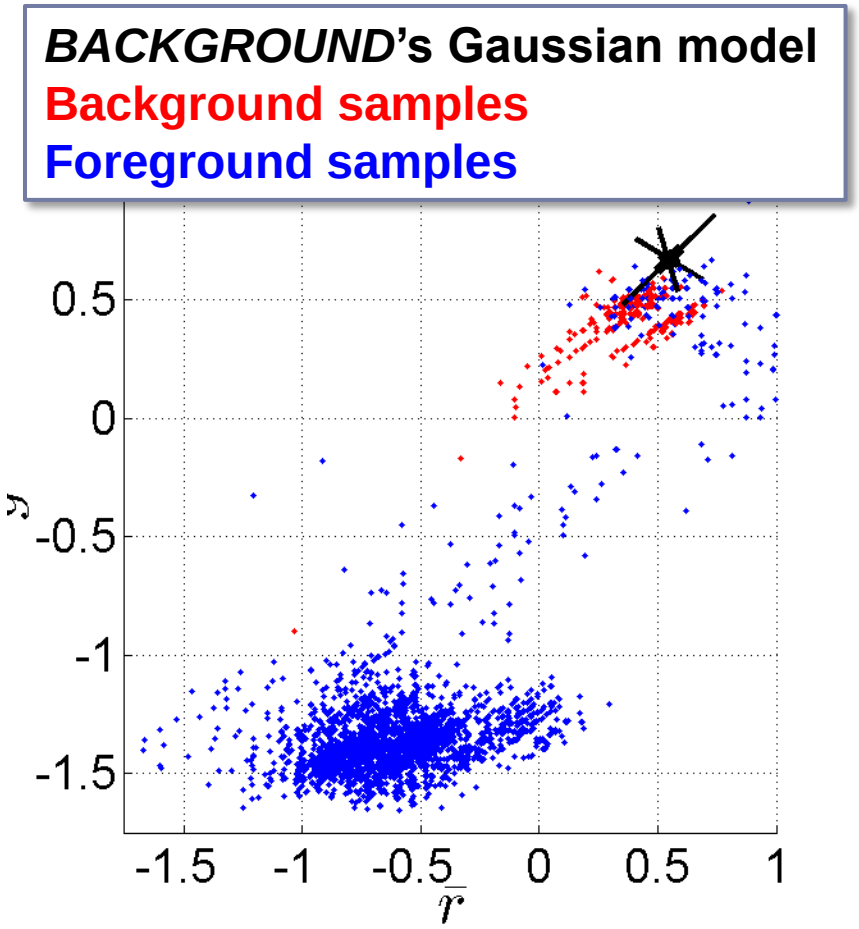
- *BACKGROUND* clip
  - Foreground is rare everywhere
  - Fit a Gaussian
  - Refit to inliers





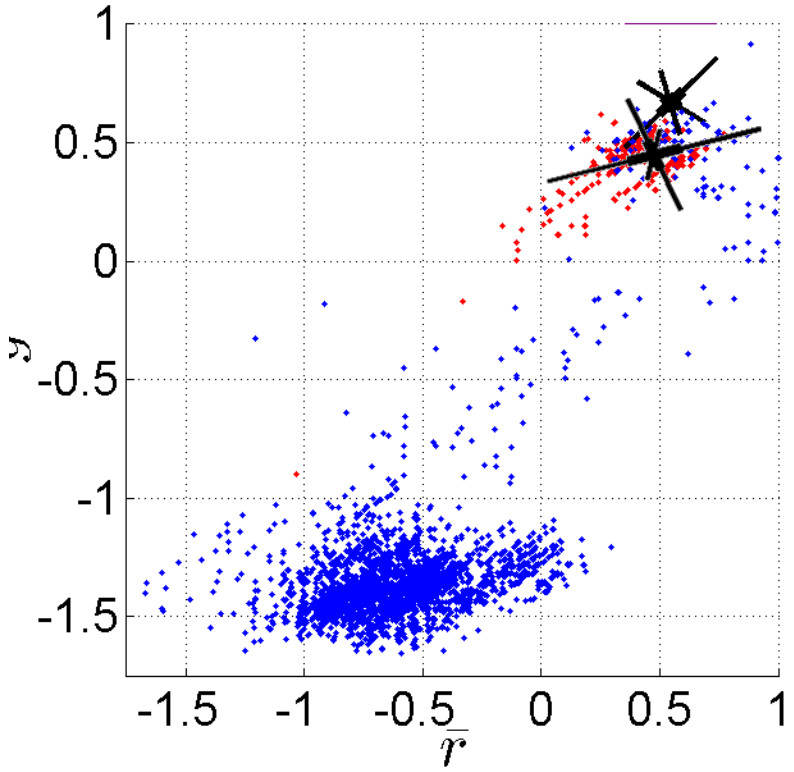
Need for Model Adaptation

- Another clip (*S02*)
  - *BACKGROUND*'s model: Suboptimal fit



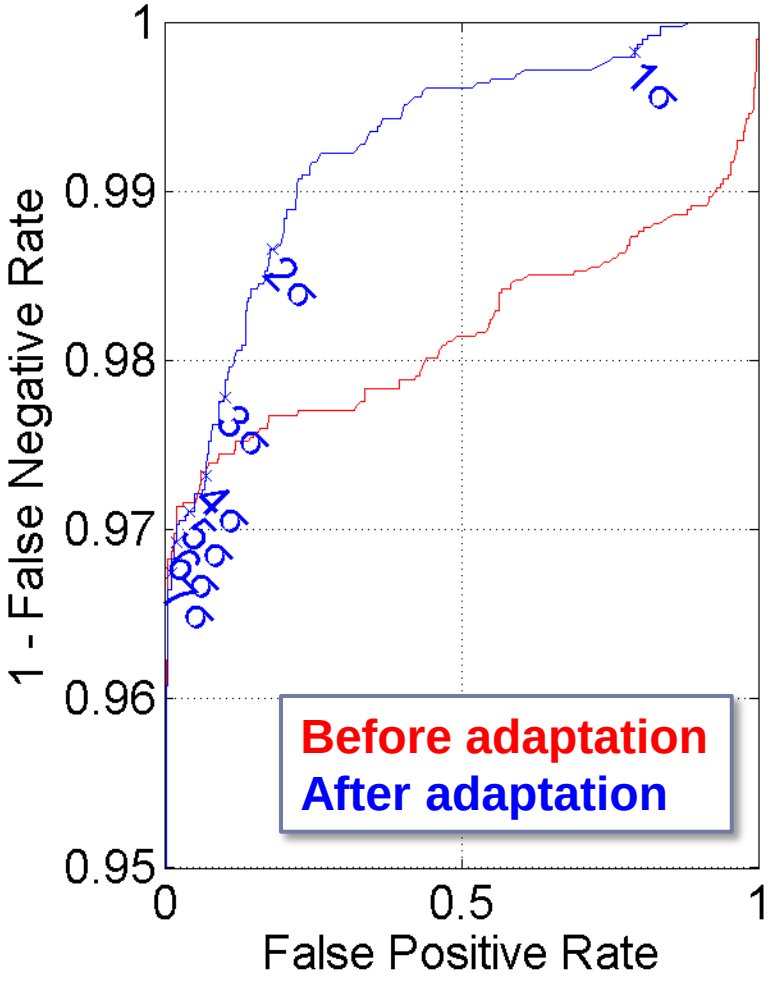
Model Adaptation

- Until convergence
  - Find inliers
  - Shift Gaussian center

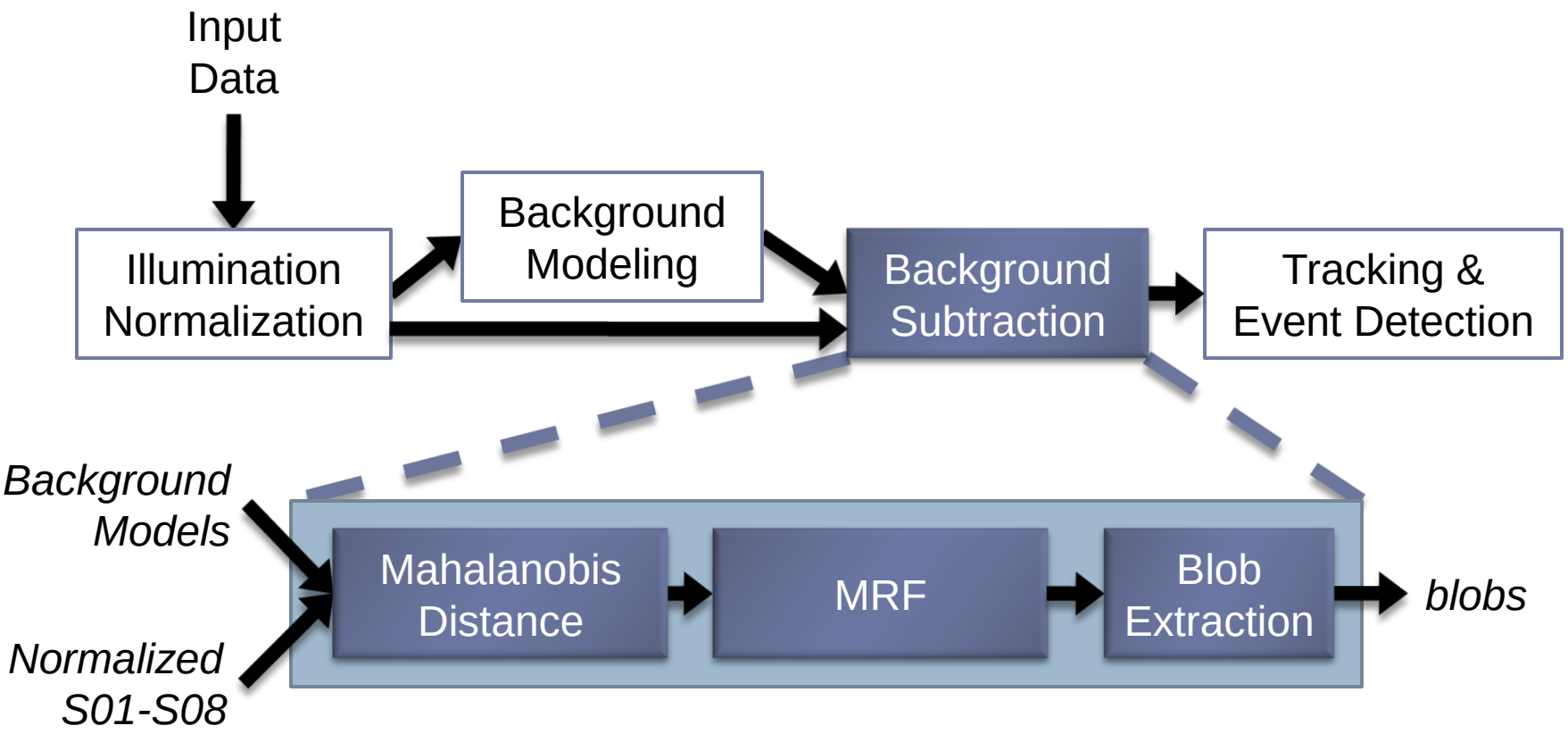


Improvement

- Adaptation
  - Is robust
  - Improves FG/BG classification rates

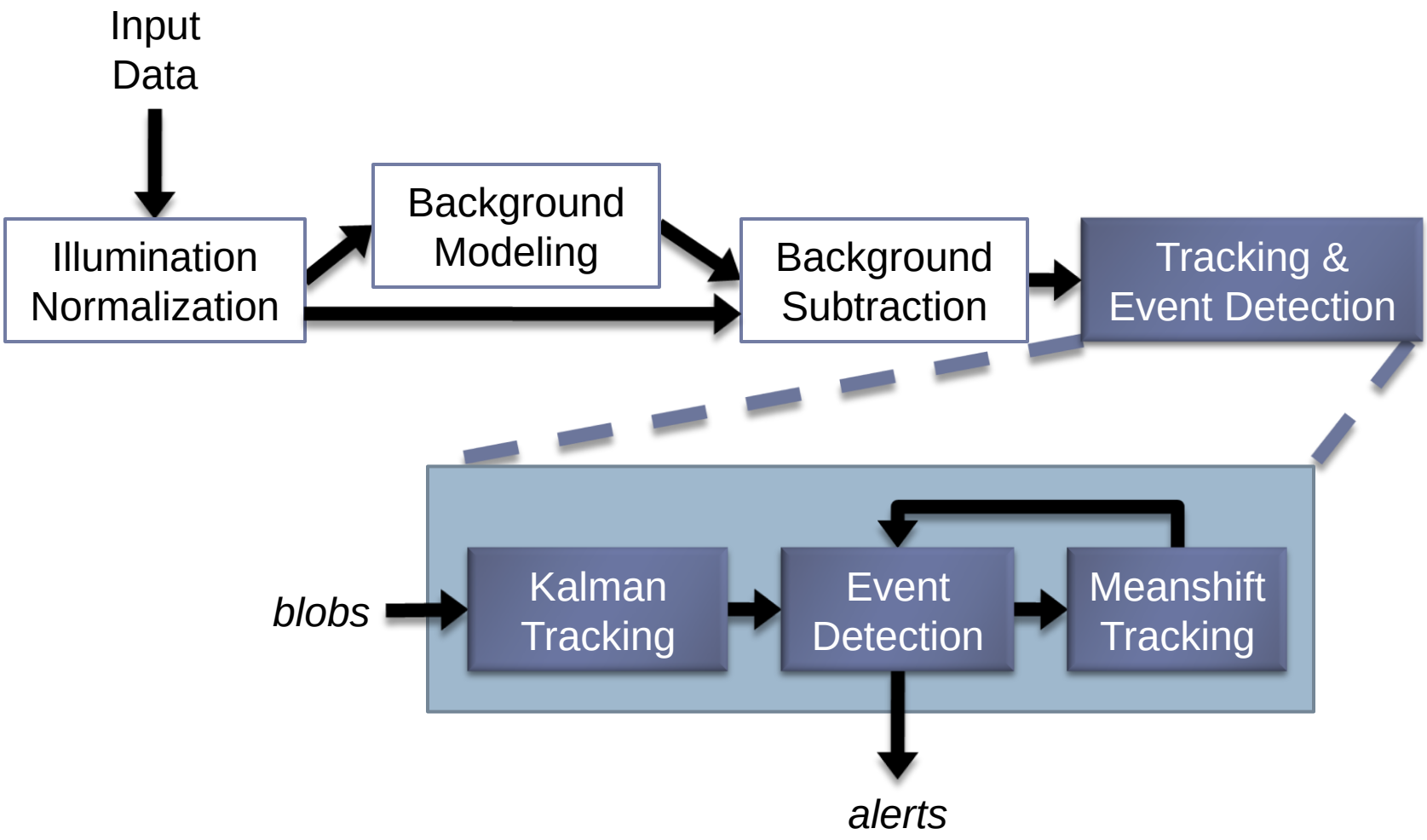


Background Subtraction



Background Subtraction

Tracking & Event Detection



Blob Tracking

- Idea: *Focus on tracking what we care about.*
  - Loitering humans
  - Dropped luggage that becomes dissociated from its owner
- Kalman tracking
  - Constant velocity
  - Low false positive rate

Detecting Humans and Luggage

- Loitering humans
  - Remain in the scene for a long time
  - Likely to create isolated tracks
- Dispossessed luggage
  - Likely to create at least an isolated blob detection

Object Type	Min. Blob Area (% of frame)	Max. Blob Area (% of frame)	Min. Blob Track Length
humans	1.5%	3.0%	16s
luggage	0.2%	1.0%	1 frame

Mean-Shift Tracking

- Blob tracking
  - Yields high-quality tracks (good)
  - Requires isolated blobs (bad)

- Meanshift tracker
  - Learn a model (color histogram) from the good blob tracks
  - Tracks through occlusions

  - Humans
    - Find scene entry/exit times

  - Luggage
    - Find drop/pickup times
    - Associate with human owners

Results

S00 – No Defined Behavior

- No events occur

- None detected

S01 – General Loitering 1 (Easy)

- **Staged loitering**
  - **5.1s late**

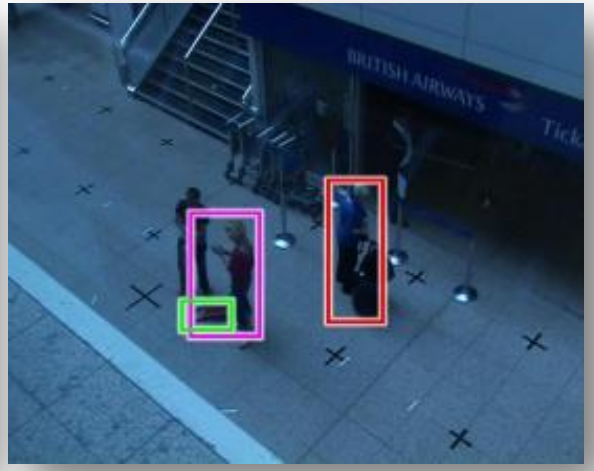
S02 – General Loitering 2 (Hard)





- **Staged loitering**
  - **1.4s late**

S03 – Bag Swap 1 (Easy)


- **Staged loitering**
  - **8.2s late**
- **Dropped luggage**
  - **Should not trigger an alarm**
  - **No alarm triggered**
- **Staged loitering man near purple-outlined woman**
  - **Missed**
- **Unscripted loitering**
  - **Detected**

S04 – Bag Swap 2 (Hard)

Stay close to each other the whole time

S05 – Theft 1 (Easy)

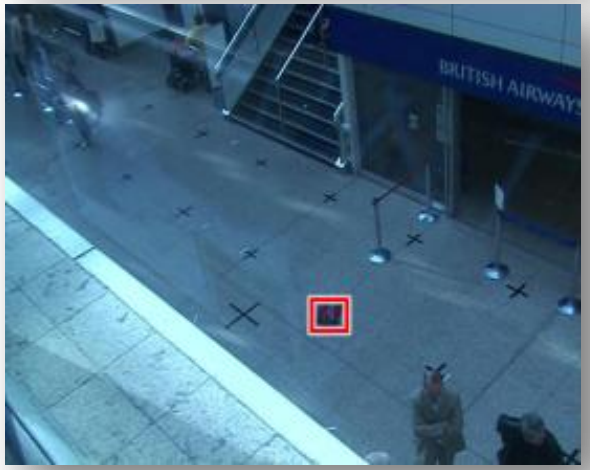




- Victim enters
  - 19.2s late
- Luggage stolen
  - 0.08s late
- Thief exits
  - 0.08s late

S06 – Theft 2 (Hard)

S07 – Left Luggage 1 (Easy)



- **Luggage dropped**
  - **0.12s late**
- **Owner tracked**
- **Luggage taken**
  - **0.08s late**
  - **By owner**

S08 – Left Luggage 2 (Hard)

Websites

- PETS 2007
  - Problem description, datasets, etc.
  - http://www.pets2007.net
- My Website
  - The paper
  - http://people.csail.mit.edu/dalleyg

Perguntas

Illumination Normalization

$$\tilde{c}_{i,t} = \Sigma_t^{-\frac{1}{2}}(c_{i,t} - \bar{c}_t),$$

where

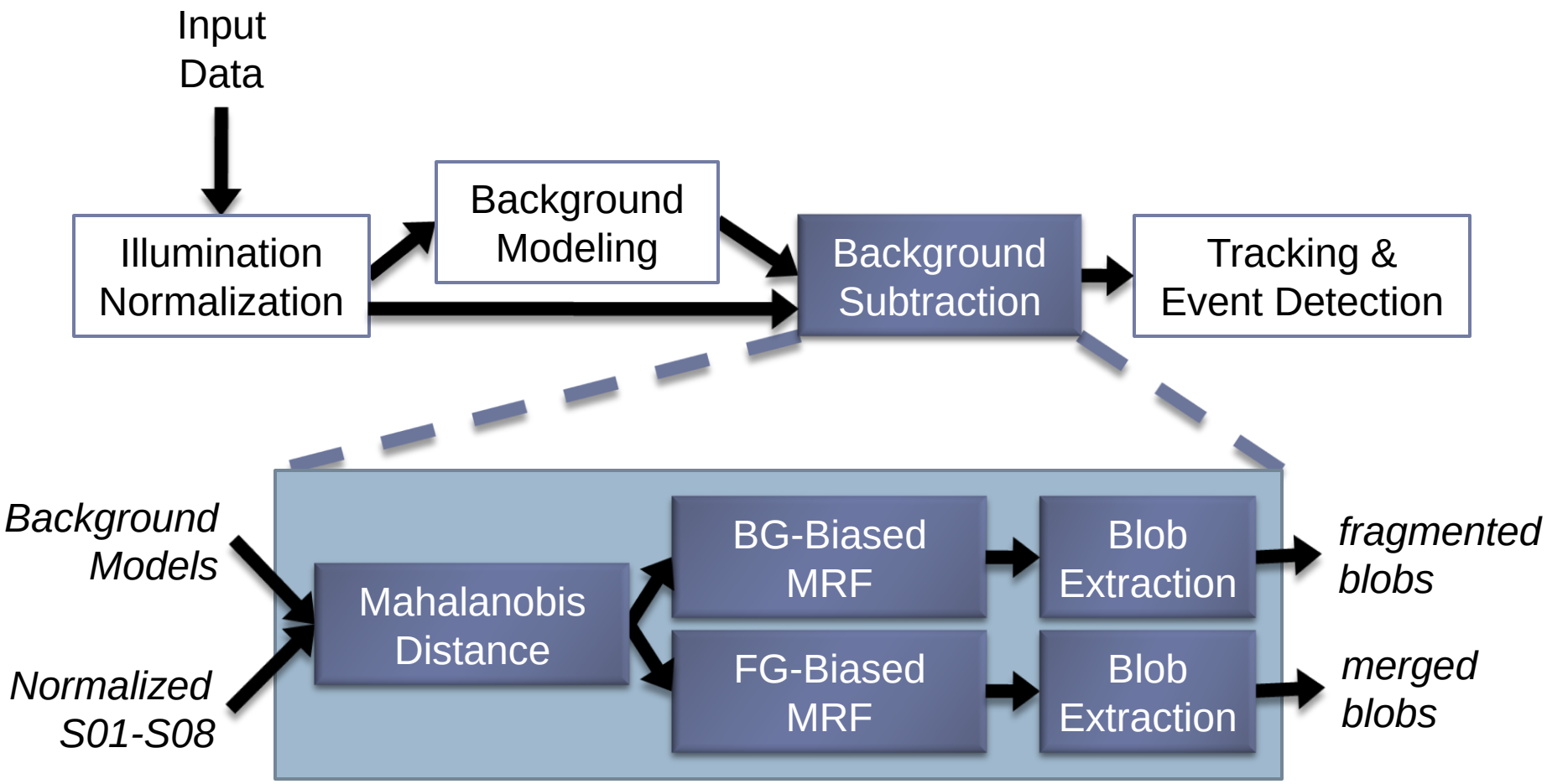
$$c_{i,t} = \text{ the color of pixel } i \text{ in frame } t,$$

$$\bar{c}_t = \frac{1}{N}\sum_{i=1}^{N} c_{i,t}, \text{ and}$$

$$\Sigma_t = \frac{1}{N-1}\sum_{i=1}^{N}(c_{i,t} - \bar{c}_t)^2.$$

Region-of-Interest Masks

Background Subtraction



Dual Background Subtraction

Foreground-Biased Blobs

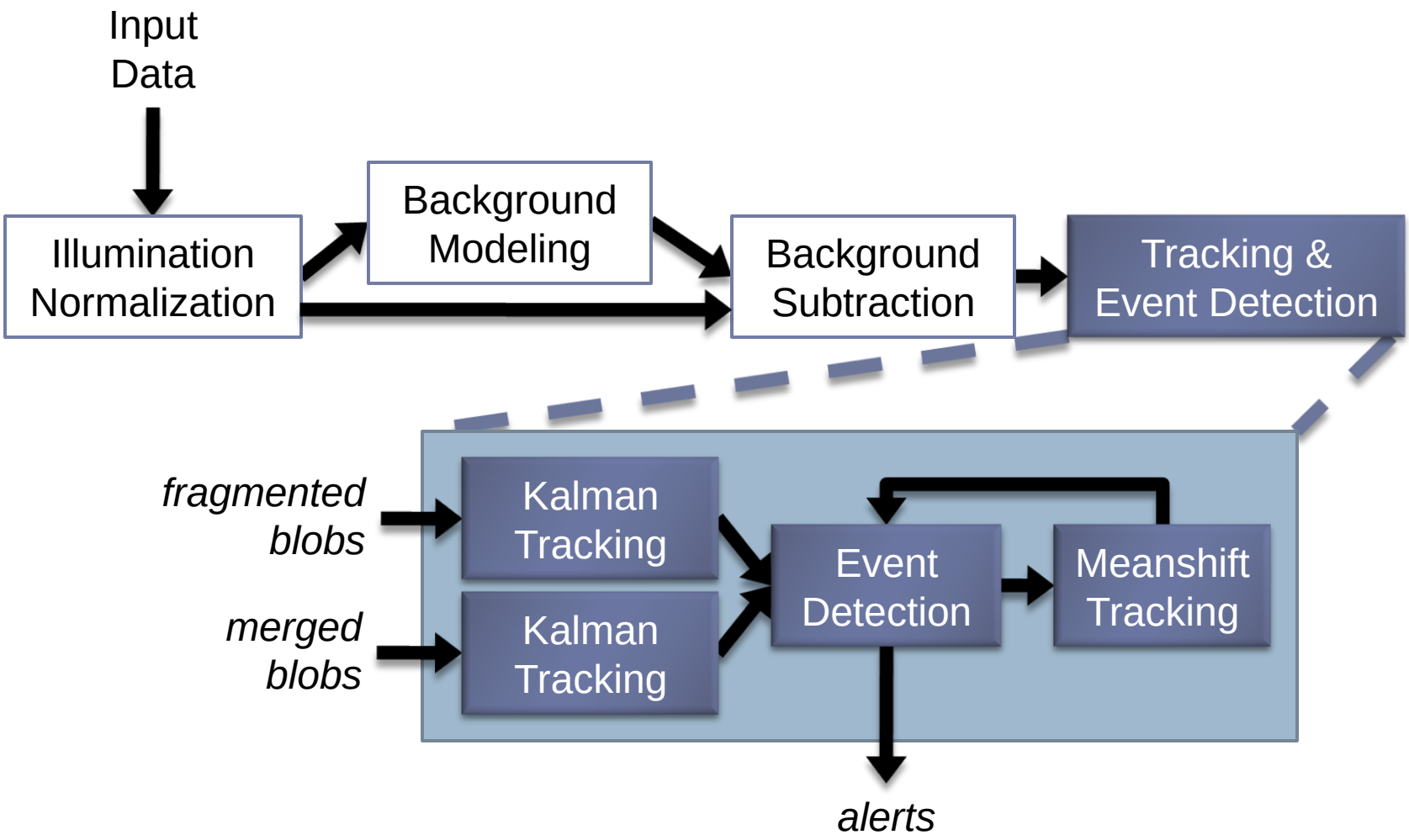
Mahalanobis Distance Map

Background-Biased Blobs

- Low fragmentation
- But blobs merged
- **Good for human tracking**

- Sharp boundaries
- But fragmented blobs
- **Good for dropped luggage detection**

Tracking & Event Detection



Why Dual Trackers/Motion Blobs

- Luggage:
  - it often doesn't travel far from the owner, so we need BG-biased to avoid merging the dropped luggage blob with the owner
  - The floor is more boring (except for specularities), so camouflaging doesn't occur much there, relative to the vertical surfaces
  - Easy to tell people fragments from luggage: small people fragments move
    - They move when they're isolated
    - They move before and after isolation
- Humans
  - With the busyness of the scene, a BG-biased MRF produces a lot of fragments and many-to-many blob matching quickly becomes impractical
  - A FG-biased MRF avoids the fragmentation issue but merges lots of blobs
    - We only care about loiterers
      - Loiterers are in the scene for a long time
      - They're likely to be isolated from other people at least at some point in time

Oriented Ellipsoids