Anomalous event detection from surveillance video

Aggelos K. Katsaggelos

Professor
Joseph Cummings Chair
Northwestern University
Department of EECS
Department of Linguistics
NorthSide University Hospital System
Argonne National Laboratory
Evanston, IL 60208
www.ece.northwestern.edu/~aggk

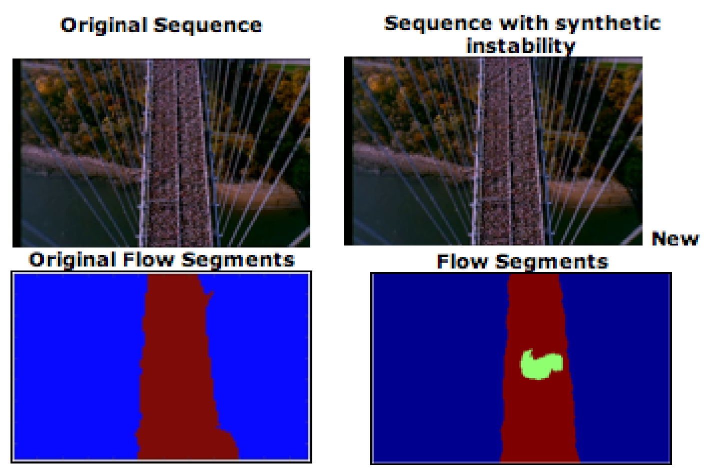
NU Transportation Center, October 26, 2016

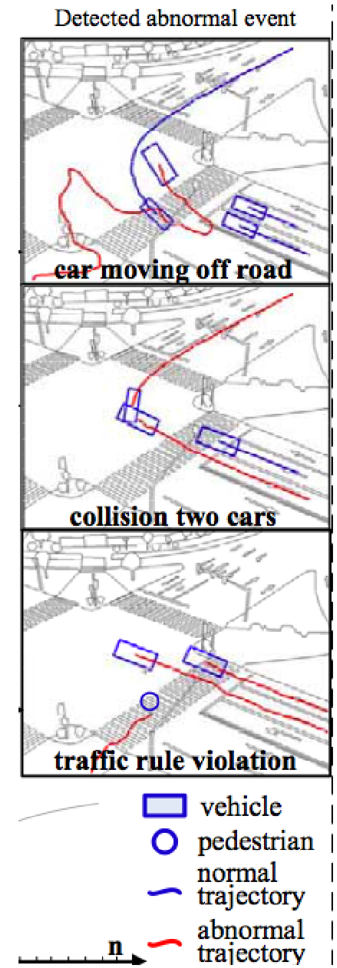
Introduction

- **Wide-scale deployment** of surveillance systems
- **Installation** and **infrastructure** costs are largest barrier to deployment of ubiquitous traffic surveillance
- Major system cost contributors are:
  - **network** requirements (bandwidth)
  - **hardware** requirements (processing power and memory)
  - **system intelligence**

Anomalies in Surveillance Video

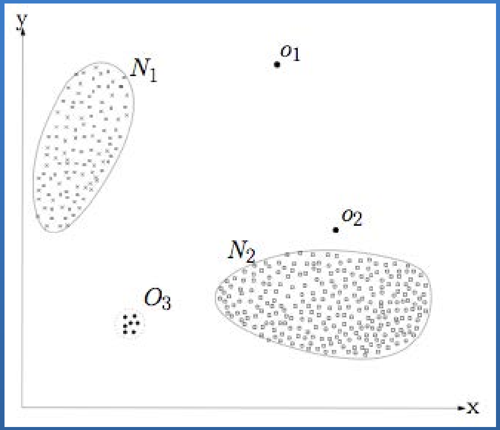
- Intelligent surveillance system
  - Video scene understanding, alarm abnormal behavior
  - Limitation of human observation
- Research problems
  - Object detection classification
  - Motion tracking modeling
  - Behavior analysis



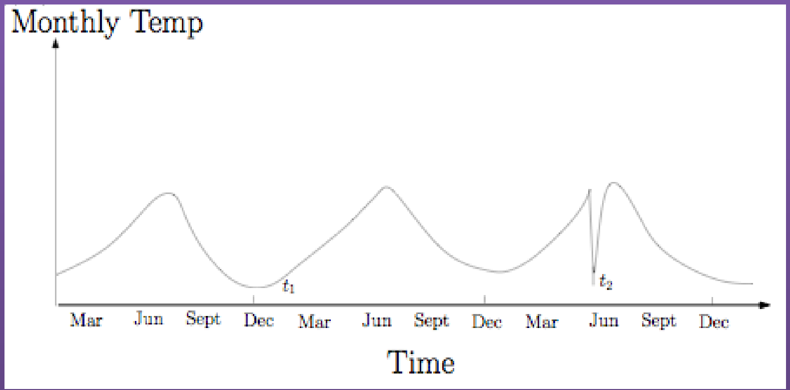


Anomaly Detection

- What are anomalies in data?
- Type of anomaly
  - Point anomaly
  - Contextual anomaly
- No data label
  - Clustering-based approach
  - Data mining approach



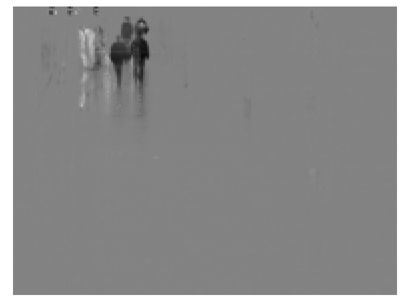
Point anomaly



Contextual anomaly

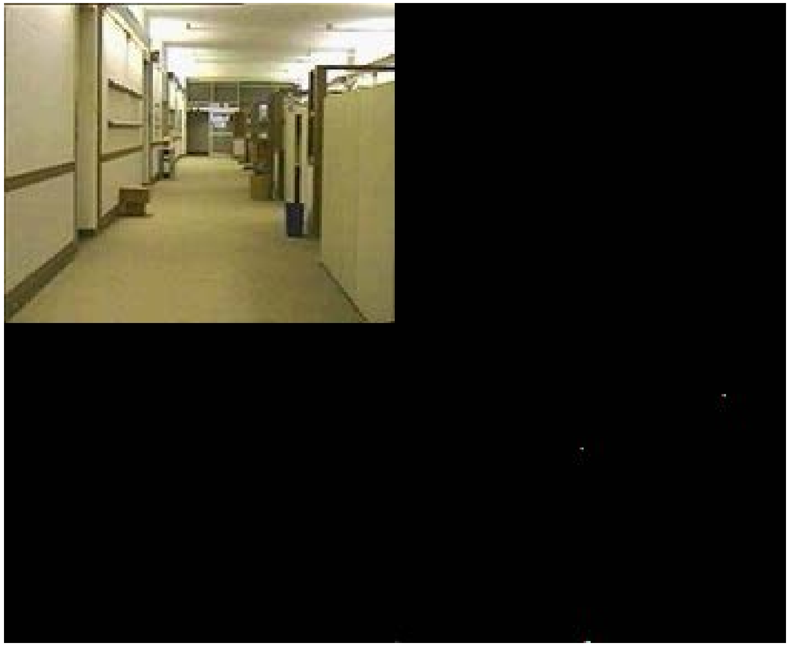
Background Subtraction

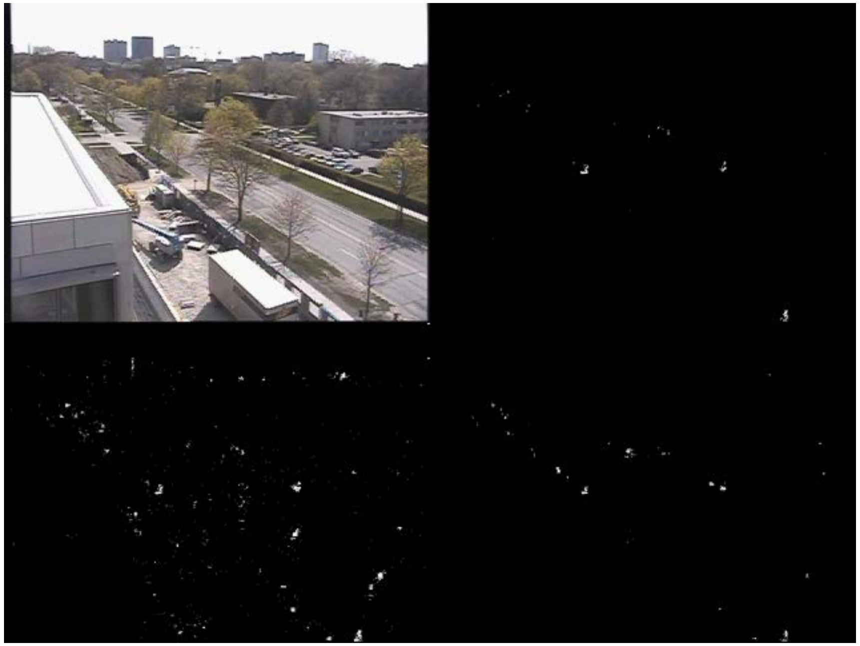
Background
Low-rank matrix



Foreground
Sparse matrix

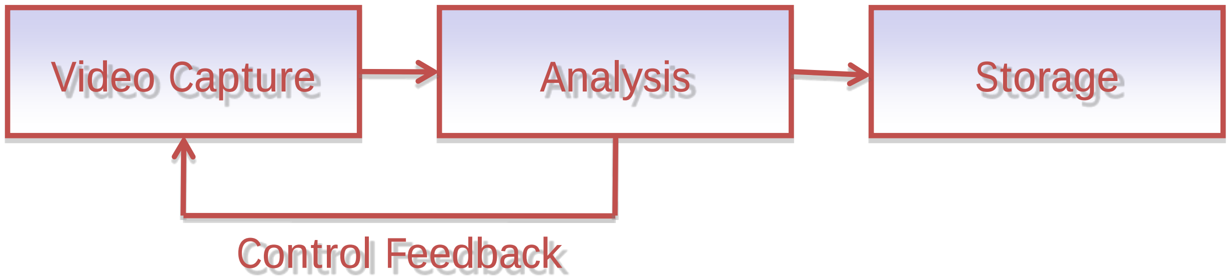
Object Detection and Tracking





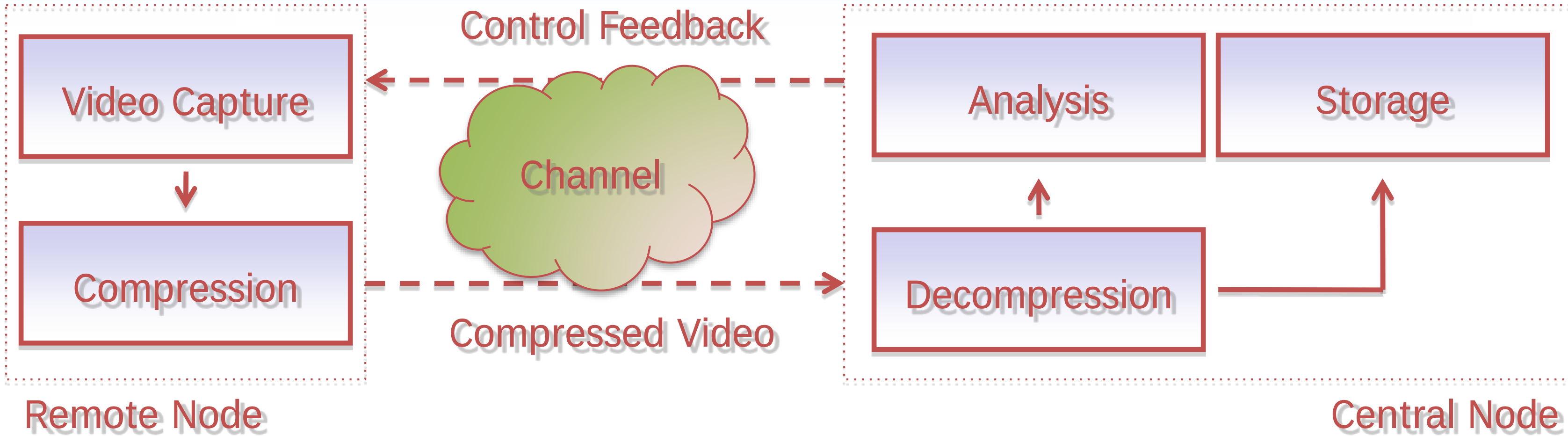
Traffic Video Data

Localized Video Surveillance



- **Localized** systems acquire, process, and store video locally.
- The requirements for these processes make each node costly and difficult to position.

Centrally Controlled Video Surveillance



- **Centrally controlled**
  - simple, low cost **remote nodes**
  - Compress then send
  - more capable **central node**.
- However, they entail
  - high infrastructure costs (**bandwidth**)
  - loss in quality due to bandwidth limitations

Tracking Objects in Compressed Video

- Compression introduces **artifacts**
  - Flicker (motion compensation)
  - Synthetic edges (block based transform)
  - Smoothing (low freq. quantization)
  - Mosquito noise (high freq. quantization)
- Artifacts get **worse** with lower bitrate
- Some artifacts **impact** trackers more severely than others

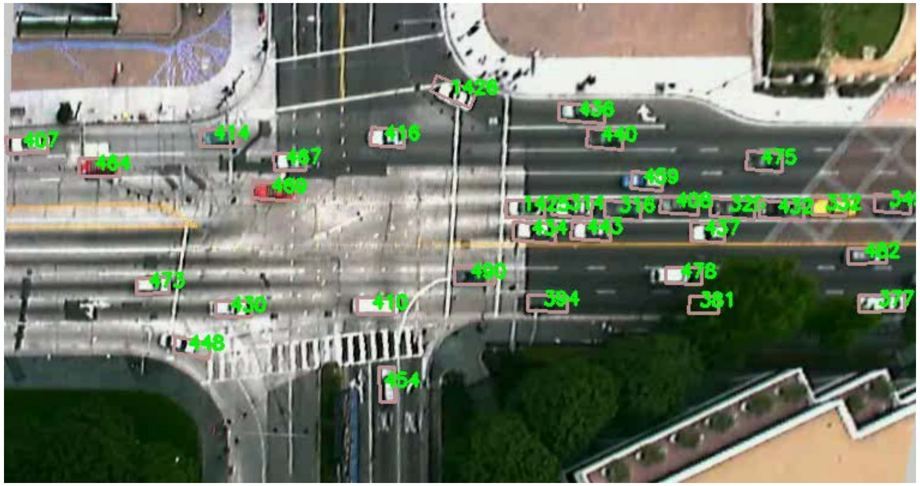
Incorporating Spatiotemporal Context

- 4 categories of anomaly
  - Point Anomaly : anomalous event of single object at specific time instance
  - Sequential Anomaly : anomalous event of single object during a time range
  - Co-occurrence Anomaly : anomalous event of multiple objects at specific time instance
  - Interaction Anomaly : anomalous event of multiple objects during a time range

- F. Jiang, J. Yuan, S. Tsaftaris ,and A. K. Katsaggelos, "Video anomaly detection in spatiotemporal context," *IEEE Int'l Conf. on Image Process.*, Hong Kong, Sept 2010.
- F. Jiang, J. Yuan, S. A. Tsaftaris, and A. K. Katsaggelos, "Anomalous video event detection using spatiotemporal context," *Computer Vision and Image Understanding*, 2011.

Study Case

- Surveillance video : traffic at road intersection
  - Traffic controlled by traffic lights
  - Traffic lights information unknown
- Task :
  - Discover motion patterns followed by most vehicles
  - Detect anomalous traffic motion

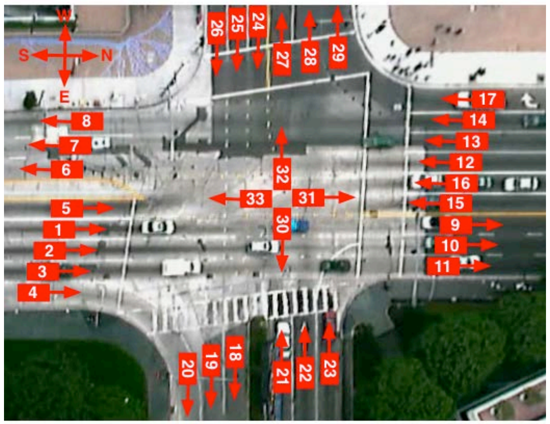


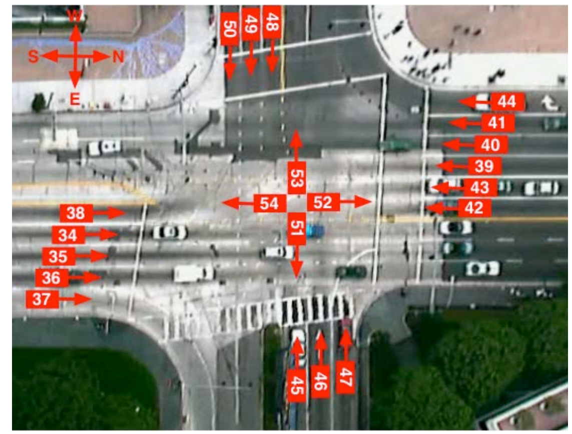
Point Anomaly Detection

- Atomic event $e_a(i,t)$
  - Single object i, time t
  - Location (lane #)
  - Direction (N/S/W/E)
  - Velocity (move/stop)
- Computing 3-D histogram of all $e_a(i,t)$
  - Normal patterns (frequent events) : high bins
  - Point anomalies (rare events) : low bins

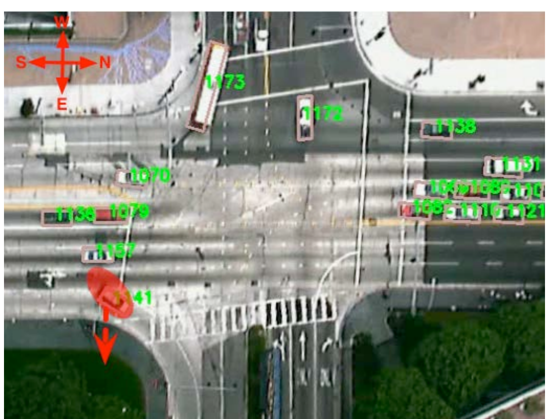
Results

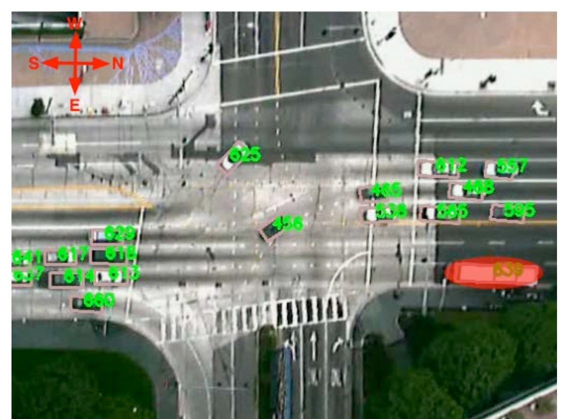
- Normal pattern





- Point anomaly



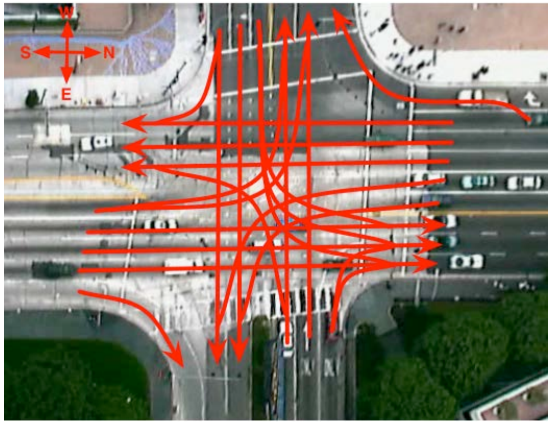


Sequential Anomaly Detection

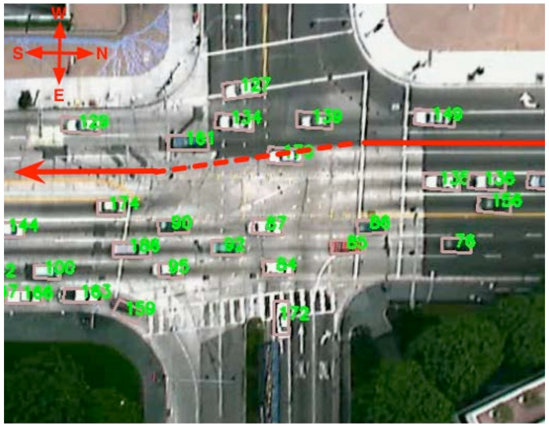
- Sequential event $e_s(i)$
  - Single object i, complete duration time
  - A sequence of atomic events :
  - ($e_a(i,1)$, $e_a(i,2)$, $e_a(i,4)$, …)
- Frequent subsequence mining
  - Detect 44 normal patterns
- Classify every $e_s(i)$ to closest normal pattern
  - Edit distance
- Detect parts different to normal pattern as sequential anomaly

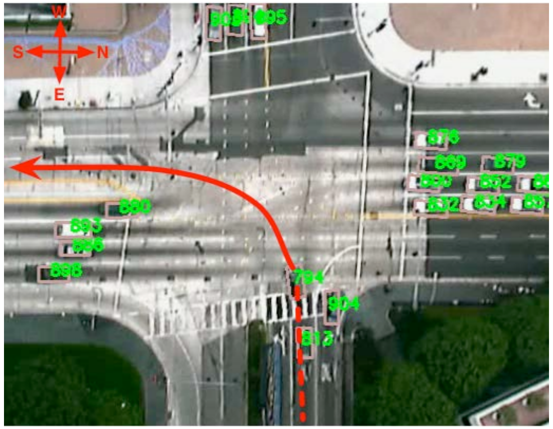
Results

- Normal pattern



- Sequential anomaly





Co-occurrence Anomaly Detection

- Co-occurrence event $e_c(t)$
  - Multiple objects, time t
  - An itemset of sequential events

    $\{ e_s(i) \mid \text{all } i \text{ appearing at } t \}$
- Frequent Itemset Mining
  - Detect 5 normal patterns
  - Regard as 5 traffic states
- Model state transition by HMM
- Classify every $e_c(t)$ by HMM decoding
- Detect parts different to normal pattern as sequential anomaly

Results

- Normal pattern

- Co-occurence anomaly

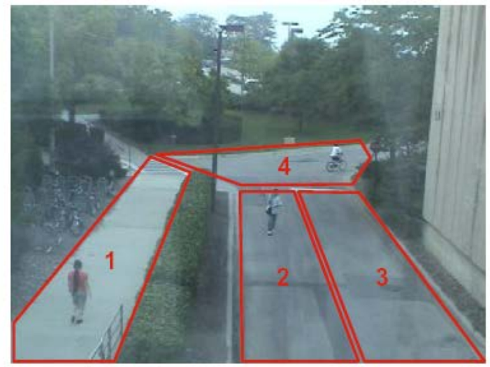
System Performance

Table 1
Statistical results of video event detection (three types).

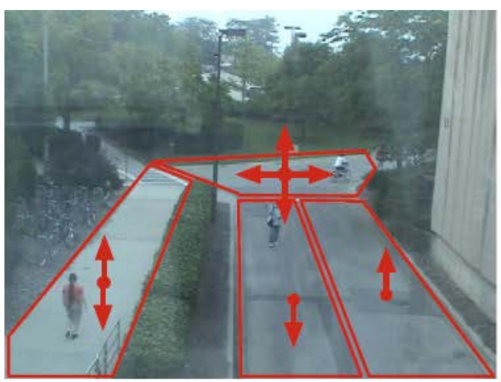
	Event type		
	Atomic	Sequential	Co-occurrence
Total #	7643	2230	21689
Anomaly (ground truth)	103	67	643
Anomaly (true positive)	95	58	504
Anomaly (false positive)	11	12	188
Detection rate (%)	92.2	86.6	78.5
False alarm rate (%)	10.7	17.9	29.2

Pedestrian Examples

- Walking Scenario



- Point anomaly



Pedestrian Examples

- Sequential Anomaly

A Different Approach

- The goal is to understand activities and interactions in a complicated scene, e.g., a crowded traffic scene.
  - Find typical single-agent activities (e.g., car makes a U-turn) and multi-agent interactions (e.g., vehicles stop waiting for pedestrians to cross the street) in this scene;
  - Label short video clips in a long sequence by interaction, and localize different activities involved in an interaction;
  - Show abnormal activities, e.g., pedestrians crossing the road outside the crosswalk; and abnormal interactions, e.g., jay-walking (people cross the road while vehicles pass by)
  - Support queries about an interaction that has not yet been discovered by the system.

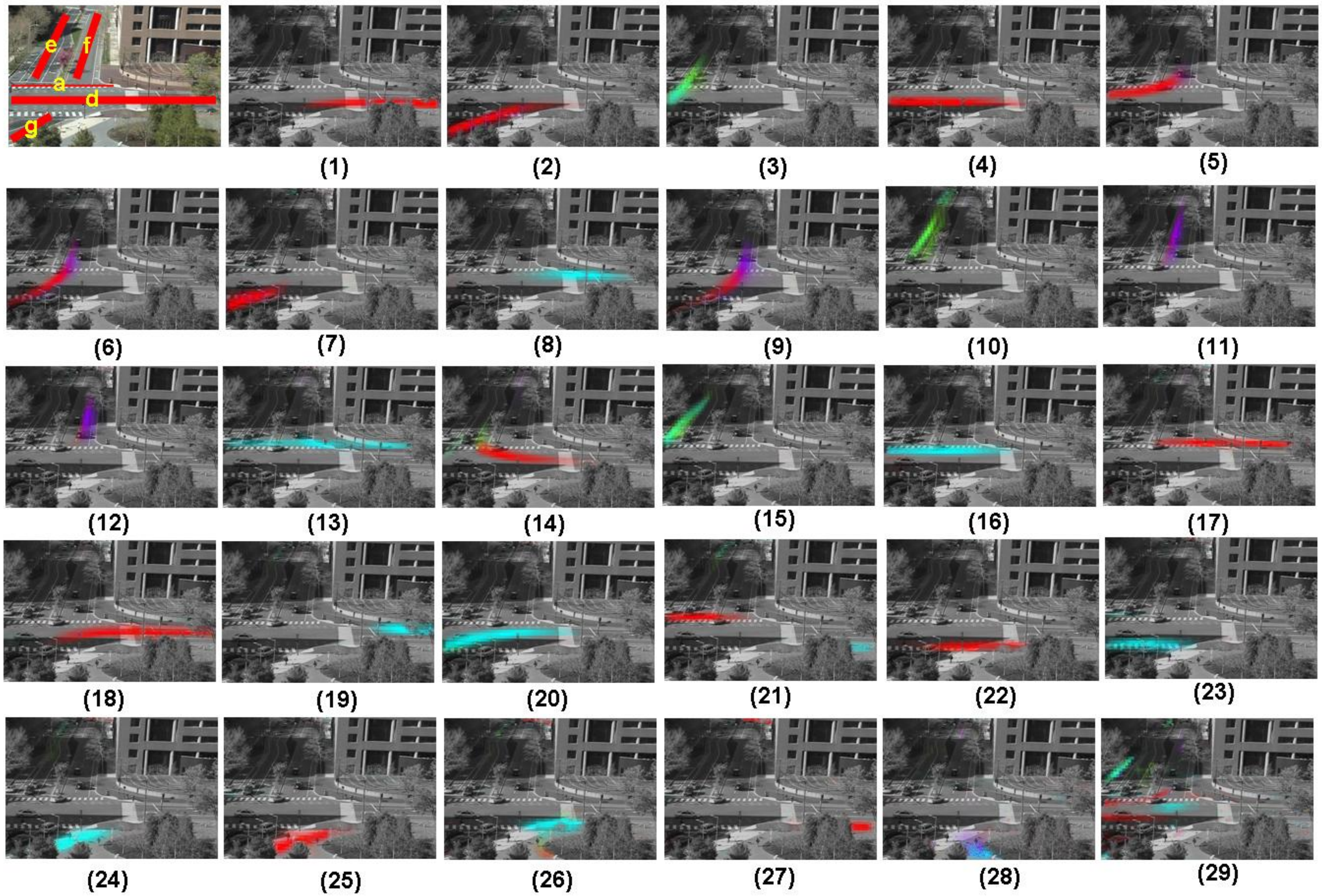
L. Song, F. Ziang, Z. Shi, R. Molina, and A. K. Katsaggelos, "Dynamic scene understanding by hierarchical motion pattern mining", *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, issue 3, June 2014.

Bayesian Hierarchical Models

- Compute low-level visual features
  - Local motion (moving pixels indexed by location and direction)
- Word-document analysis
  - Quantizing local motion into visual words and dividing the long video sequence into short clips as documents
- Hierarchical Bayesian model
  - Atomic activities are modeled as distributions over low-level visual features
  - Interactions are modeled as distributions over atomic activities

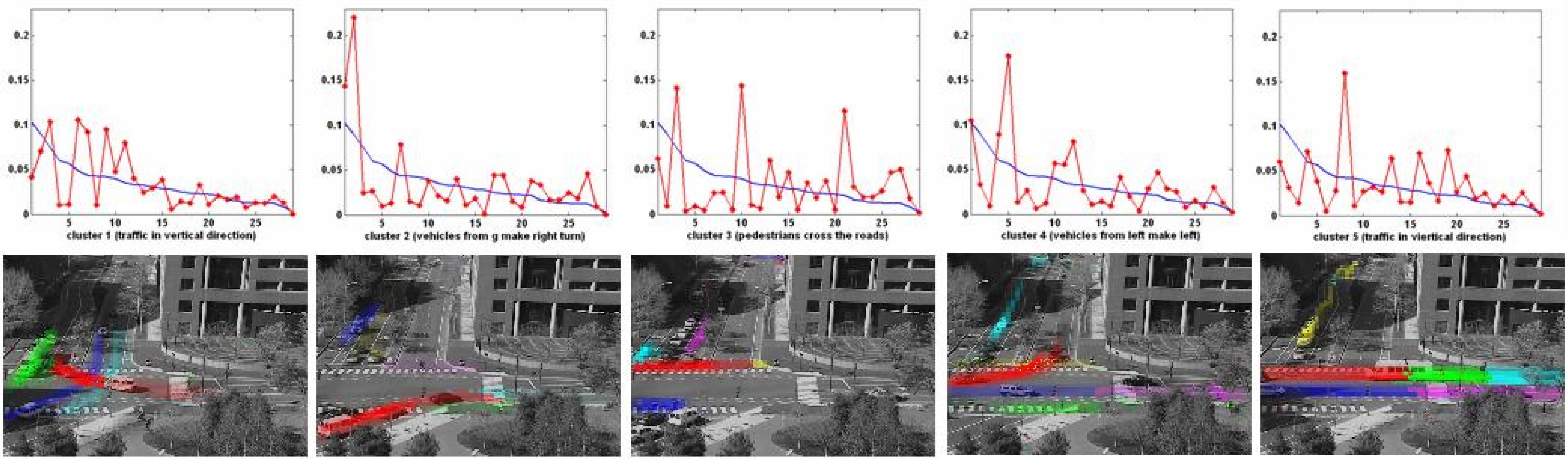
Discover Atomic Activities

- 29 atomic activities (4 colors: 4 motion directions)



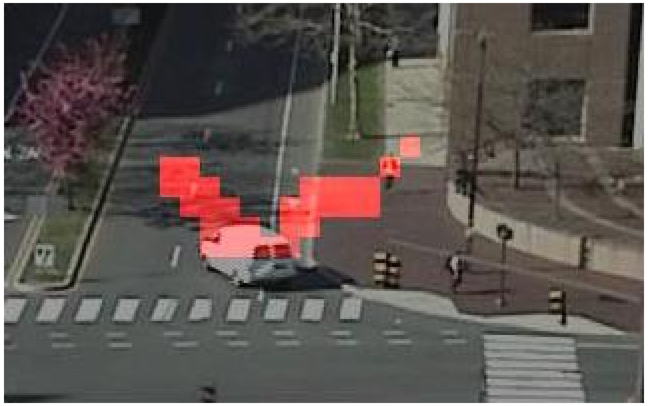
Discover Interactions

- 5 different interactions
  - First row: the interaction distributions over 29 atomic activities
  - Second row: a video clip as an example for each interaction (the motions of the 5 largest atomic activities marked)



Abnormality Detection

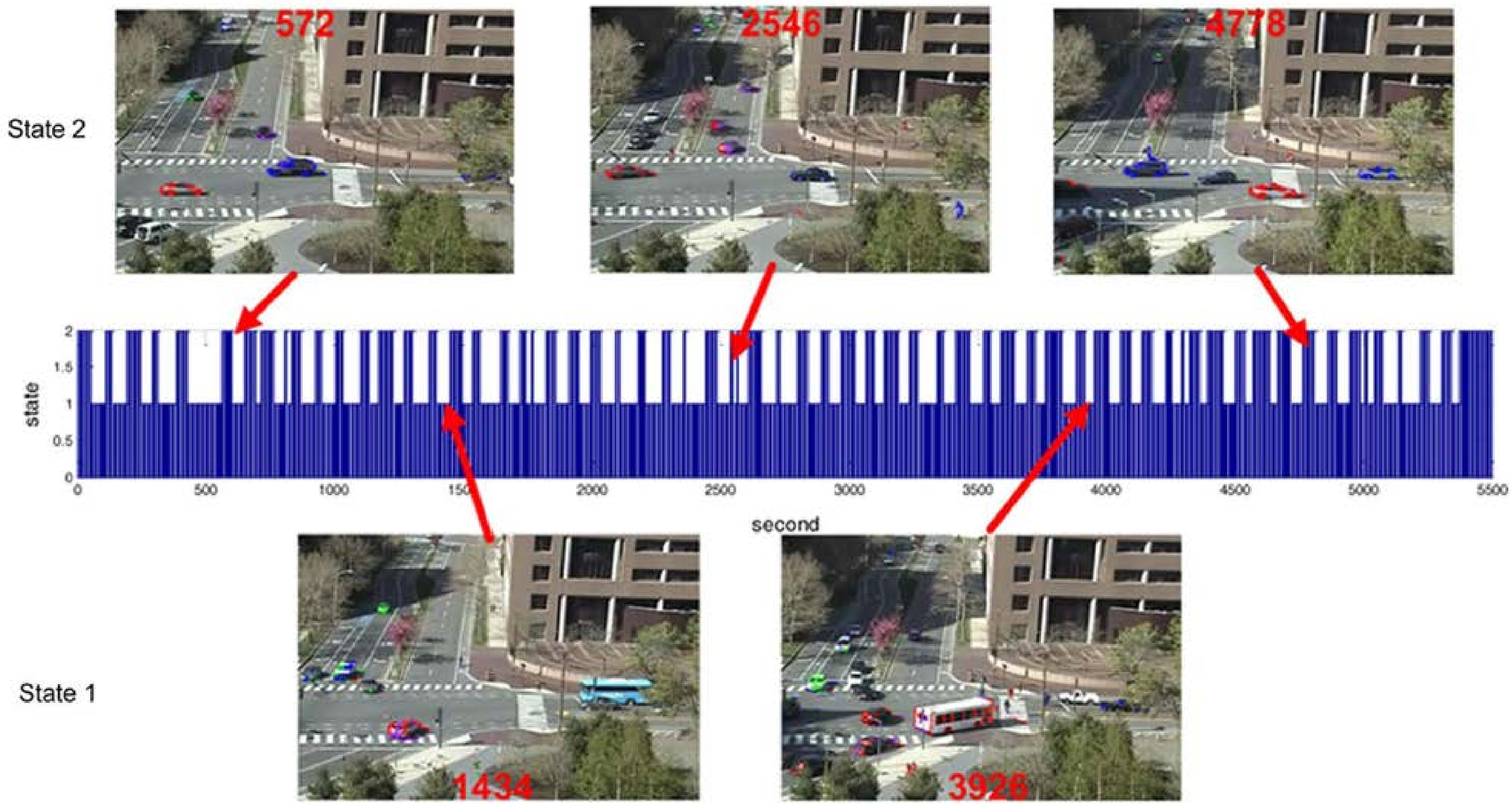
- Under the Bayesian models, abnormality detection is based on the marginal likelihood of every video clip or motion



Example1:
Pedestrian crossing the street while vehicle is passing

Example2:
Pedestrian crossing the street while the red light is on

Segmentation



Closing Thoughts

- Transportation problems rich in applying ML
- Developed techniques applicable to other areas
- It is only the beginning