Toward Building a Robust and Intelligent Video Surveillance System: A Case Study

Edward Chang and Yuan-Fang Wang

Douglas R. Lanman

CS 295-1: Sensor Data Management

28 Sept. 2005

BROWN

Outline

- *Introduction to Video Surveillance*
- UCSB Hardware Configuration
- Event Detection and Data Fusion
- Event Classification
- Conclusion

Introduction to Video Surveillance

Driving Factors

- Inexpensive cameras
- Large-capacity disk storage
- Ubiquitous broad-band communication networks

Motivation: Fully Automated Drudgery

Target Application Areas

- Infrastructure surveillance
  (e.g., airports, bridges, trains, etc.)
- Crime prevention and forensic evidence
- Environmental monitoring

Current Limitations

- Human-in-the-loop
- Semi-autonomous operation

Desired Capabilities

- Robust event detection and data fusion
- Fully automatic *semantic* labeling
- Low latency and limited false negatives

Outline

- Introduction to Video Surveillance
- *UCSB Hardware Configuration*
- Event Detection and Data Fusion
- Event Classification
- Conclusion

UCSB Surveillance System

System Configuration

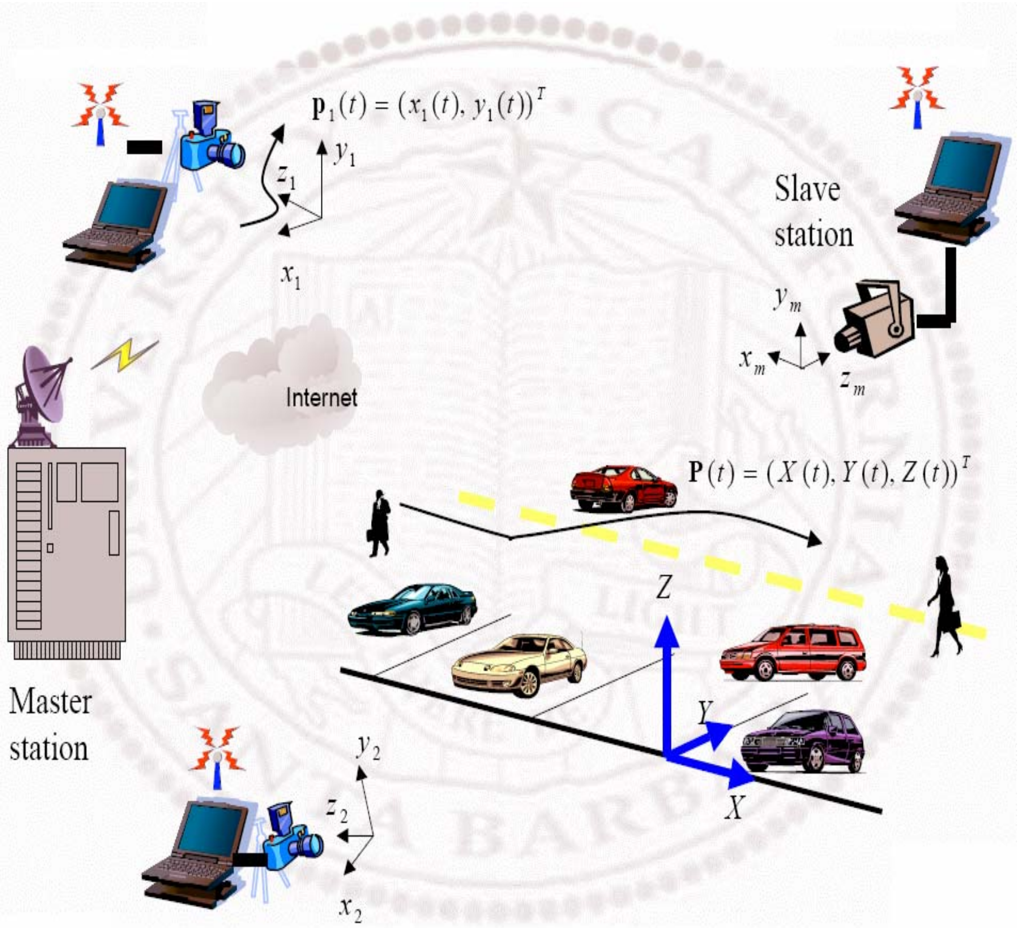
- Master server (central archive)
- Multiple surveillance terminals
- PTZ camera platforms

Operator Interface

- Supports real-time stream retrieval and video playback (rewind, forward, slow-motion)
- On-line meta-data queries
- Alerts issued at master server

Modular Architecture

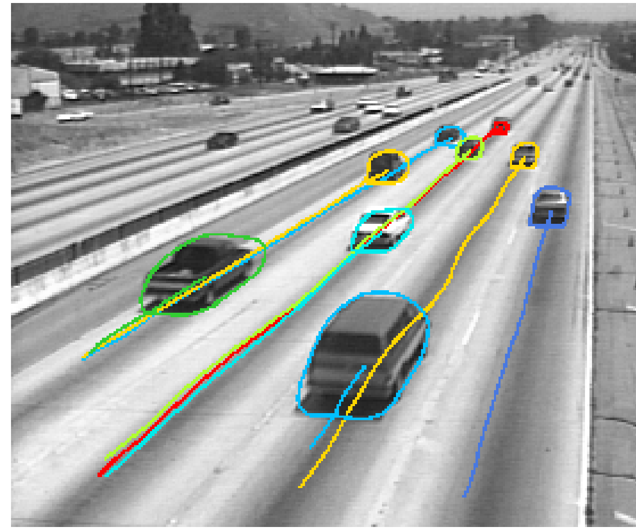
- *Unlimited arbitrary cameras**
- Heterogeneous networks



Outline

- Introduction to Video Surveillance
- UCSB Hardware Configuration
- *Event Detection and Data Fusion*
  - Background Subtraction
  - Camera Calibration and Temporal Registration
  - Sensor Data Fusion
- Event Classification
- Conclusion

Introduction to Event Detection

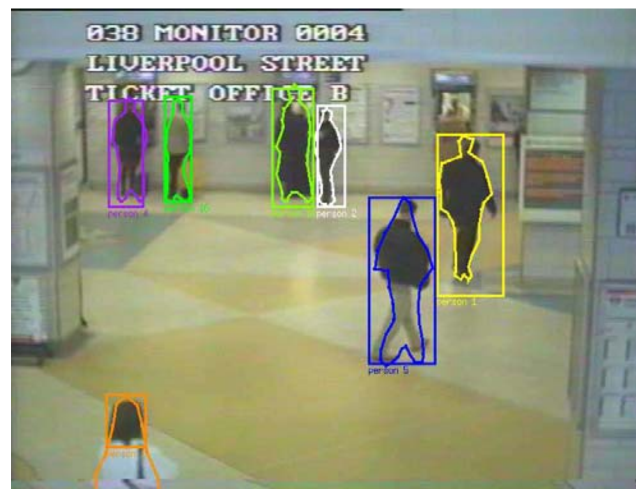


Central Challenge

- From multiple video streams, form a *hierarchical* and *invariant* description of scene activities

Required Processing Stages

- Background subtraction
- Camera calibration
- Temporal synchronization
- Data fusion and dissemination



System Limitations

- Limited spatial coverage and overlap
- Misalignment of temporal time stamps
- Object occlusions and missing data
- *Latency and bandwidth utilization**

Moving Object Segmentation

Background Subtraction

- Compare pixel *intensity* and *color* in adjacent frames

Key Challenge: *Saliency*

- Lighting changes, shadows, and "environmental" motion

Object Tracking

What is a Kalman Filter?

- Used to estimate an object's state (3D track) from a set of observations
- Gaussian state prior and noise model
- Allows *real-time* state updates

Limitations of Kalman Filtering

- Difficult to track through "crossing" events (i.e., intersecting paths)

"Hypothesis-Verification" Tracking

- Arbitrary noise model and non-linear state transition
- Allows multiple hypotheses to be used to track through merging, crossing, or other difficult events
- More computations than Kalman filtering

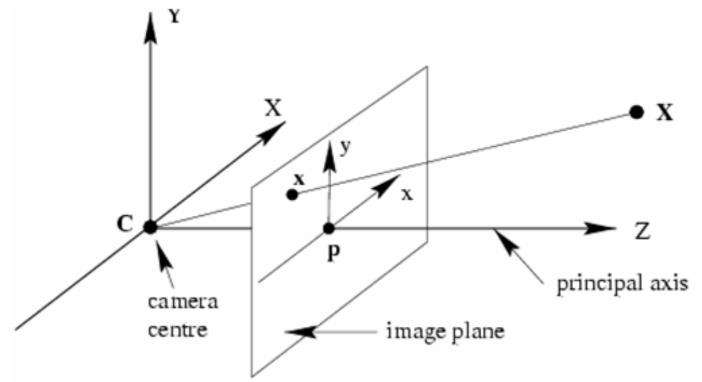
Overview of Camera Calibration

Intrinsic Calibration

- Maps points to a normalized image plane
  (*focal length, skew, and distortion effects*)
- Typically done off-line

$$(X, Y, Z)^T \mapsto (fX/Z, fY/Z)^T$$

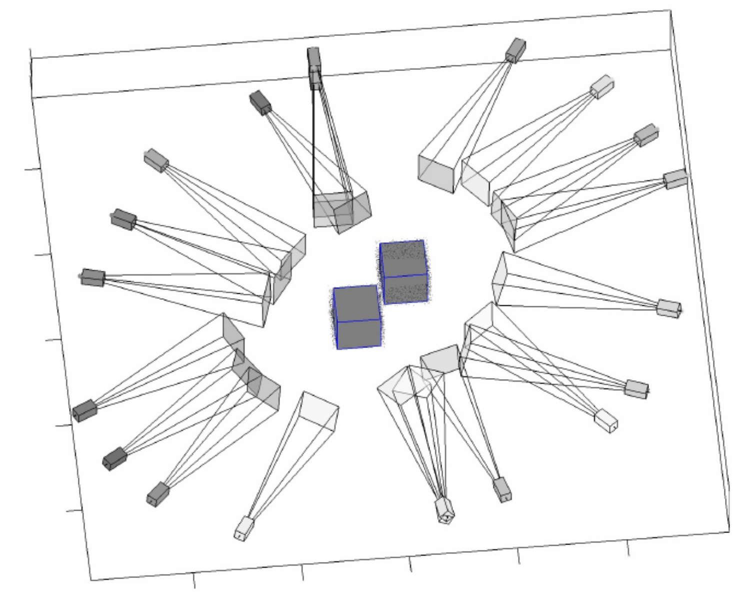
$$\begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \mapsto \begin{pmatrix} fX \\ fY \\ Z \end{pmatrix} = \begin{bmatrix} f & & & 0 \\ & f & & 0 \\ & & 1 & 0 \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$



Extrinsic Calibration

- Pose of camera relative to a fixed world coordinate system (*translation and rotation*)
- Updated continuously

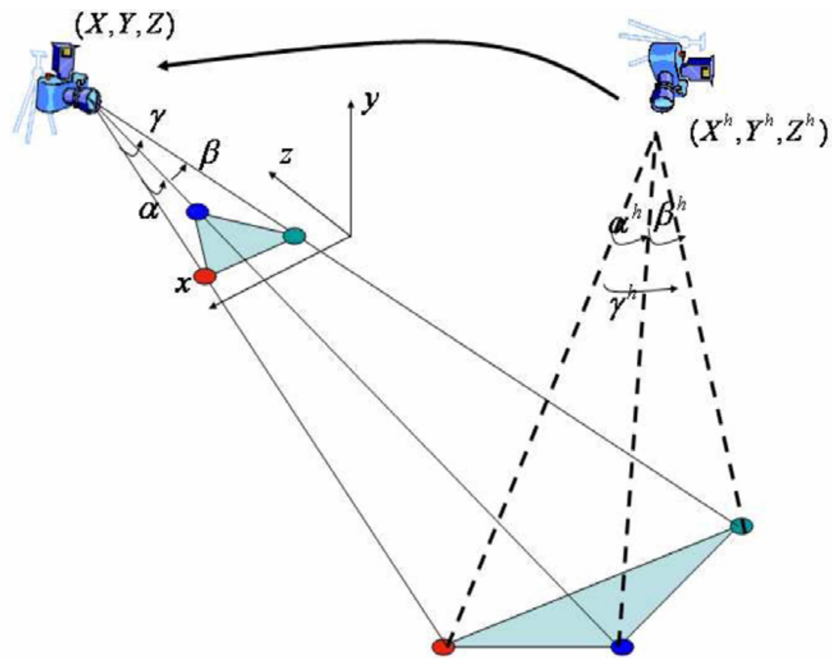
$$\mathrm{X}_{\mathrm{cam}} = \begin{bmatrix} \mathbf{R} & \mathrm{t} \\ 0 & 1 \end{bmatrix} \mathrm{X}_{\mathrm{obj}}$$



Church's Algorithm

General Extrinsic Calibration Requirements

- Each camera must observe *six* known landmarks
  (i.e., six degrees-of-freedom: {x, y, z} and {roll, pitch, yaw})
- Occlusions or limited knowledge of the environment requires calibration with fewer landmarks



Church's Algorithm

- Pose estimation with *three* landmarks
- Face angles in spatial coordinates equal face angles in the image plane
- Thousands of pose updates per second
- Invented by Earl Church for aerial photogrammetry (1945)

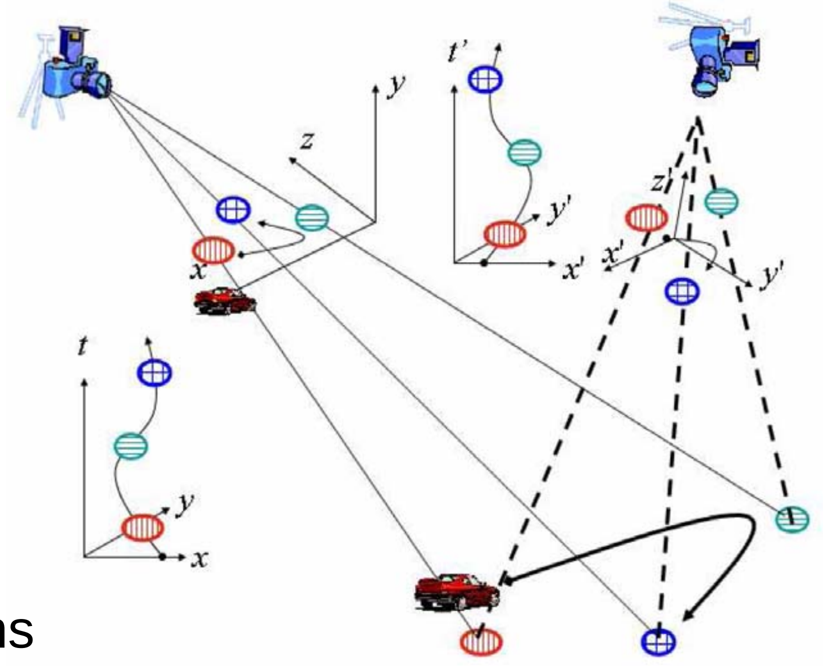
Temporal Alignment from Image Invariants

Key Problem

- Same trajectory appears differently due to projection
- Correlation of observations requires a unique time stamp
- Clocks on surveillance stations may not be synchronized
- Need an observable that is invariant to projection

Observations

- Differential geometry: curve is described (up to rigid motion) by its *curvature* and *torsion* vectors w.r.t. arc length
- Projective geometry: affine projection preserves area ratios



UCSB Solution

- Normalized curvature and torsion ratios used to synchronize multiple observations

Introduction to Sensor Data Fusion

Combining Observations

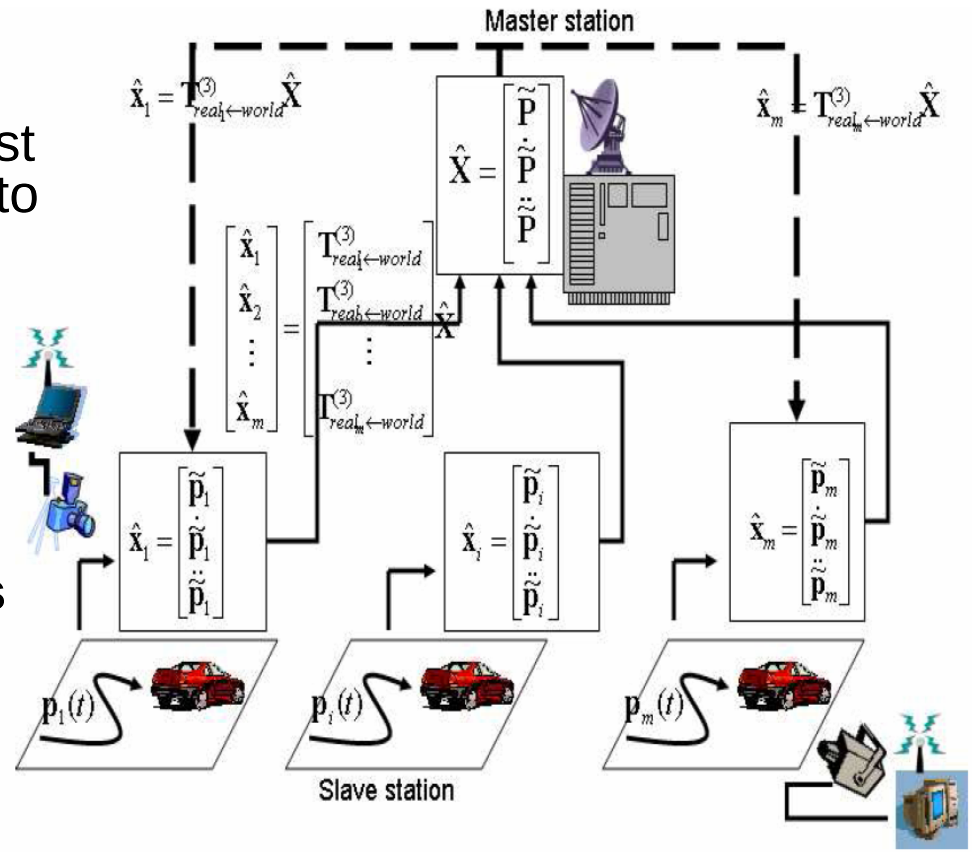
- Local trajectories must be fused into a global representation
- Pose and temporal synchronization required for sensor data fusion

Key Challenges

- Projection of object trajectory must be observed from multiple views to synthesize 3D information
- Occlusion, missing data, and synchronization errors will complicate synthesis (e.g., must track through gaps in coverage)

UCSB Solution: Two Components

- "Bottom-up" analysis
- "Top-down" cueing



Outline

- Introduction to Video Surveillance
- UCSB Hardware Configuration
- Event Detection and Data Fusion
- *Event Classification*
- Conclusion

Semantic Event Classification

Recognizing Events

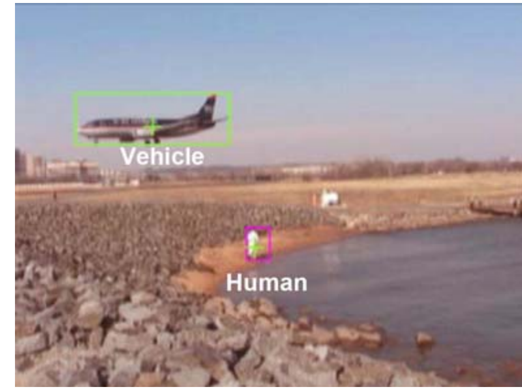
- Given a global representation (3D track), provide *semantic* descriptions of events (e.g., running, walking, crawling, etc.)
- From sequences of semantic event labels and tracks, recognize specific event classes (e.g., waiting for train, missed train, loitering)

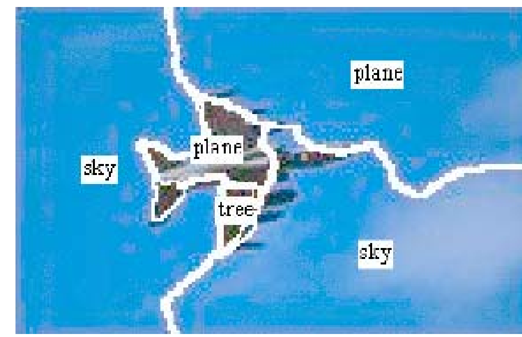
Humans Back in the Loop

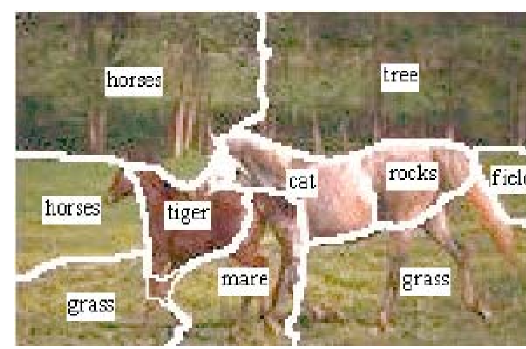
- Issue warning to base station when a prohibited event occurs (e.g., car idling or circling, unattended item, etc.)

Issues

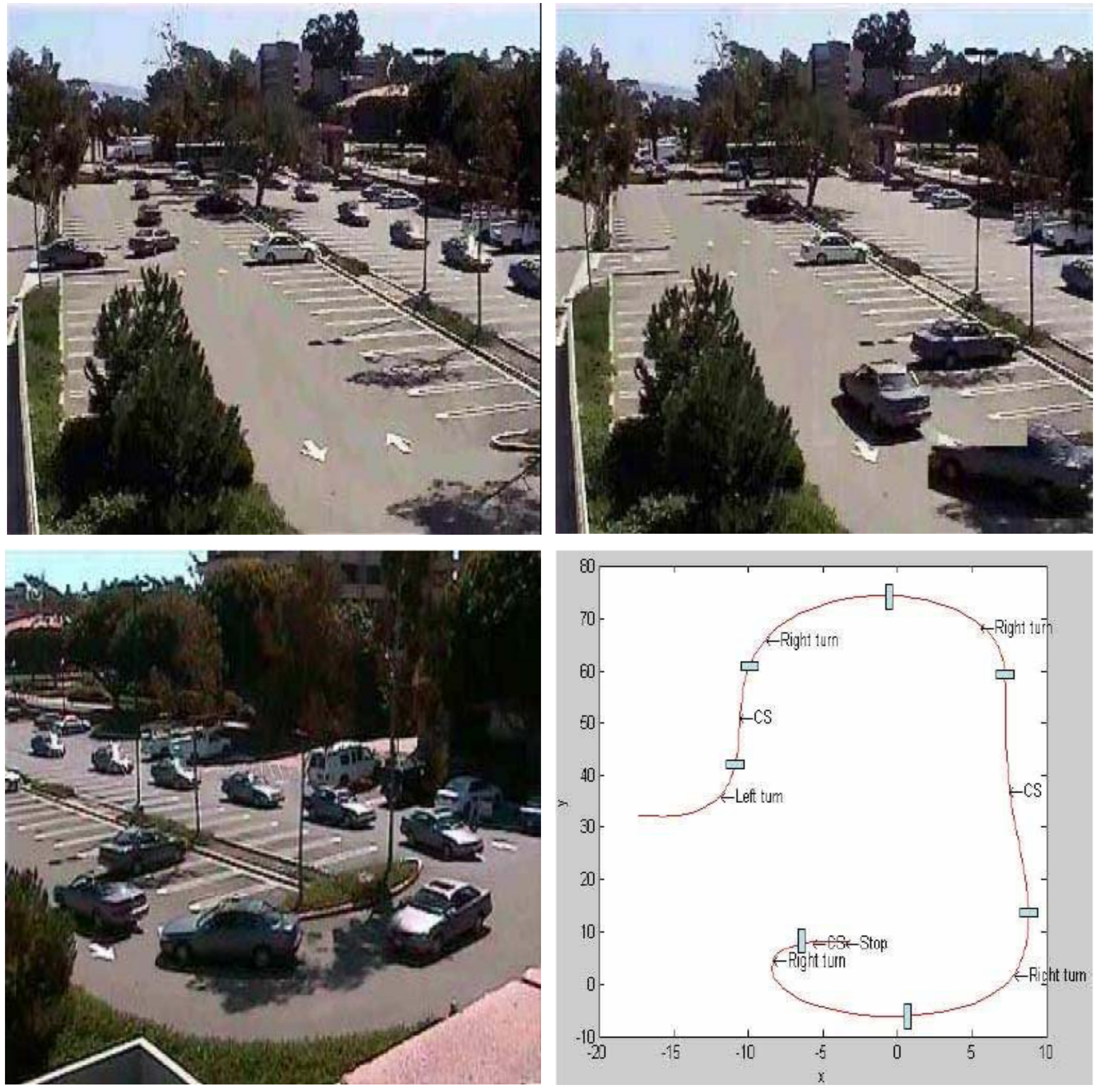
- Latency and false negatives/positives
- Limited training data for threat classes







Example: Vehicle Motion Recognition



Sequence Alignment Learning

Recognizing Event Classes

- Global information: velocity and acceleration statistics
- Semantic information: “turning”, “driving straight”, “stopped”, etc.

Sequence Assignment Learning

- First compare the semantic labels
- Further refine using secondary variables (velocity, etc.)
- Combine into a sequence-alignment kernel through the tensor product of the two similarity metrics

Sequence Representation

Numeric-valued Representation

Symbolic-valued Representation

DFT

Wavelet

Piece-wise Linear

SVD

Natural Language

Strings

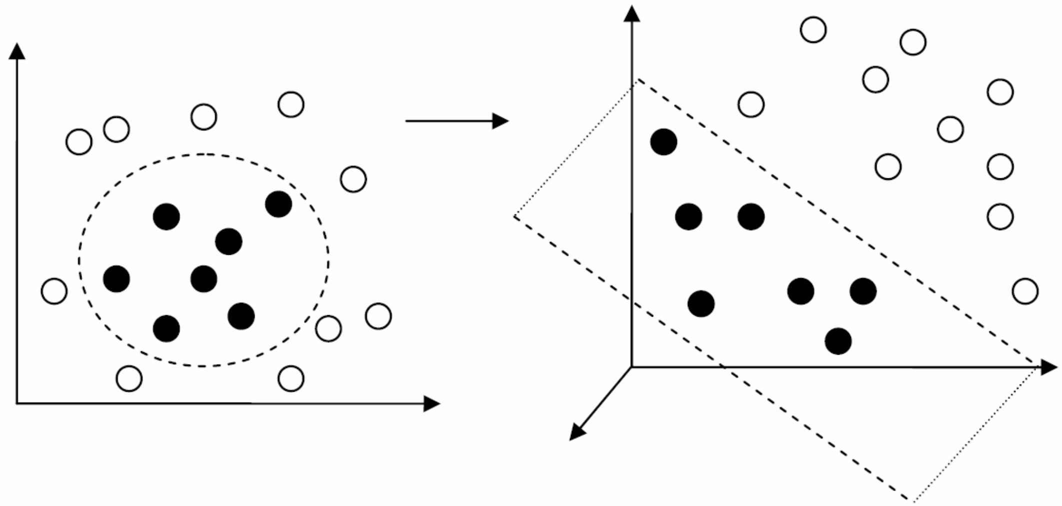
Critical Challenge: “Imbalanced” Learning

Key Issues

- Suspicious (positive) events more frequent than benign (negative)
- Claim: *risk of a false negative outweighs that of a false positive**

Implications from Machine Learning

- Imbalanced training data → skewed class boundary
- Conformal transformation used to reduce skew
- Bias classifier towards negative result to prevent overly frequent alerts



Outline

- Introduction to Video Surveillance
- UCSB Hardware Configuration
- Event Detection and Data Fusion
- Event Classification
- *Conclusion*

Conclusion

The Emergence of Video Surveillance Systems

- Broad application set (e.g., infrastructure, environment, forensics)
- Hardware both *economically* and *technologically* feasible

Key Limitations

- State-of-the-art image and video processing lags far behind hardware technology
- Scalability: *UCSB system applies the "leader-worker" model**

Future Research Areas

- Truly distributed algorithms:

  (1) calibration, (2) event detection, and (3) semantic labeling
- Distributed storage and retrieval
- Reducing latency and false positives

References

1. E. Chang and Y-F. Wang, "Toward Building a Robust and Intelligent Video Surveillance System: A Case Study," Proc. of the IEEE Multimedia and Expo Conference, Taipei, Taiwan, 2004.
2. E. Chang, "Event Sensing on Distributed Video-Sensor Networks," Basenets 2004, in cooperation with ACM/IEEE Conf. on Broadband Networks, San Jose, October 2004.
3. L. Jiao, G. Wu, E. Chang, and Y-F. Wang, "The Anatomy of a Multi-Camera Video Surveillance System," ACM Multimedia System Journal, 2004.
4. E. Mahoney and J. Helperin, "Caught! Big Brother May Be Watching You With Traffic Cameras," Edmunds.com, http://www.edmunds.com/ownership/driving/articles/42961/article.html, 2004.
5. On-line, Midlands CCTV Birmingham Ltd., http://www.midlands-cctv.co.uk/img/website%20picture%20camera%203.jpg, 2005.
6. On-line, http://www.ilexikon.com/images/5/58/London_tube_Charing_Cross.jpg, 2005.
7. On-line, Appian Technology PLC., http://www.appian-tech.com/applications/cctv.html, 2005.
8. On-line, http://www.halfdone.com/SOTW/MBTA_HQ.jpg, 2004.
9. On-line, http://http.cs.berkeley.edu/~pm/RoadWatch/, 2005.
10. N. Siebel, *Design and Implementation of People Tracking Algorithms for Visual Surveillance Applications*, doctoral thesis, Dept. of Computer Science, The University of Reading, 2003.
11. A. Elgammal, D. Harwood, and L. Davis, "Non-Parametric Model for Background Subtraction," http://www.cs.rutgers.edu/~elgammal/Research/BGS/research_bgs.htm, 2005.
12. On-line, IBM Research: *PeopleVision* Project, http://www.research.ibm.com/peoplevision/, 2005.

References

13. M. Pollefeys, “3D Photography: Camera Model and Calibration,” On-line, http://www.unc.edu/courses/2004fall/comp/290/089/, 2004.
14. D. Devarajan and R. Radke., “Distributed Metric Calibration for Large-Scale Camera Networks,” First Workshop on Broadband Advanced Sensor Networks 2004, San Jose, CA, 2004.
15. I. Cohen, “Detection and Tracking of Moving Objects,” On-line, http://iris.usc.edu/~icohen/projects/vace/detection.htm, 2005.
16. A. Lipton, C. Heartwell, N. Haering, and D. Madden, “Critical Asset Protection, Perimeter Monitoring, and Threat Detection Using Automated Video Surveillance,” IEEE 36th Annual International Carnahan Conference on Security Technology, 2002.
17. K. Barnard, P. Duygulu, N. de Freitas, D. Forsyth, D. Blei, and M. I. Jordan, "Matching Words and Pictures," Journal of Machine Learning Research, Vol. 3, pp. 1107-1135, 2003.
18. B. Lovell and C. Walder, “Support Vector Machines for Business Applications,” in *Business Applications and Computational Intelligence*, 2005.
19. “Unlocking The Potential of Wireless Video Networks”, Virginia Tech Department of Electrical and Computer Engineering, Annual Report, 2003.