Tracking: Where has it been and where is it going?

Bob Collins
Penn State University

BMTT-PETS Workshop
Honolulu HI, July 2017

True Story...

1997-2000 Darpa funds the VSAM project in US.

The BAA prohibits proposing tracking research, because “tracking is a solved problem.”

Every funded effort did some tracking research.

Explanation

- Why would Darpa in the 1990's think tracking was a solved problem?
- *"Military intelligence"* ☺
- Radar-based tracking (point-like "objects") was pretty much a solved problem.
- Kalman/EKF/particle filter; JPDAF; MHT were all well-understood.

Vision-based Tracking

- "Tracking" means different things to different people.
- Passive, vision-based "extended object tracking" involves the study of
  - Appearance as well as movement
  - Detection as well as association
- What kind of tracking works depends on data-specific factors.

To Consider: Discriminability

How easy is it to discriminate one object from another?

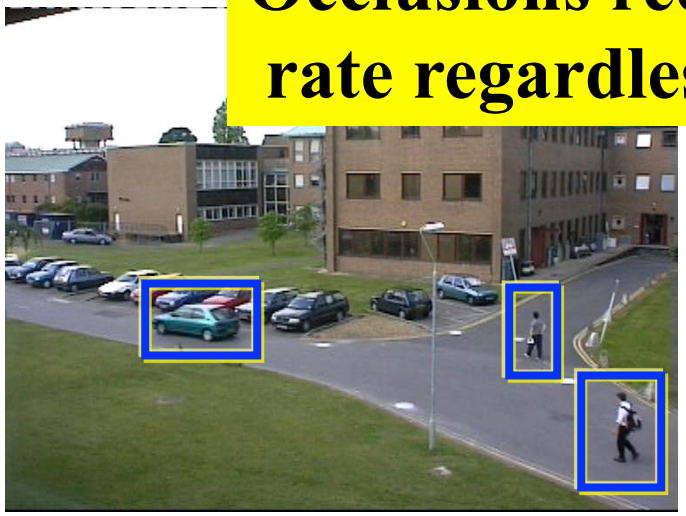
appearance models can do all the work

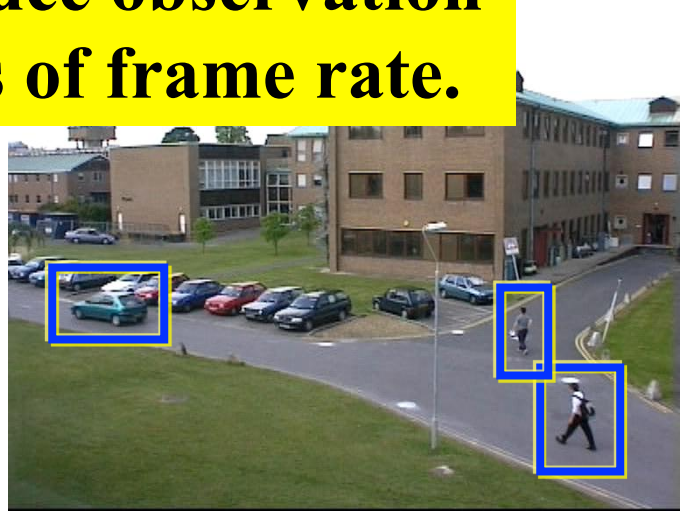
constraints on geometry and motion become crucial

To Consider: Observation Rate

Occlusions reduce observation rate regardless of frame rate.

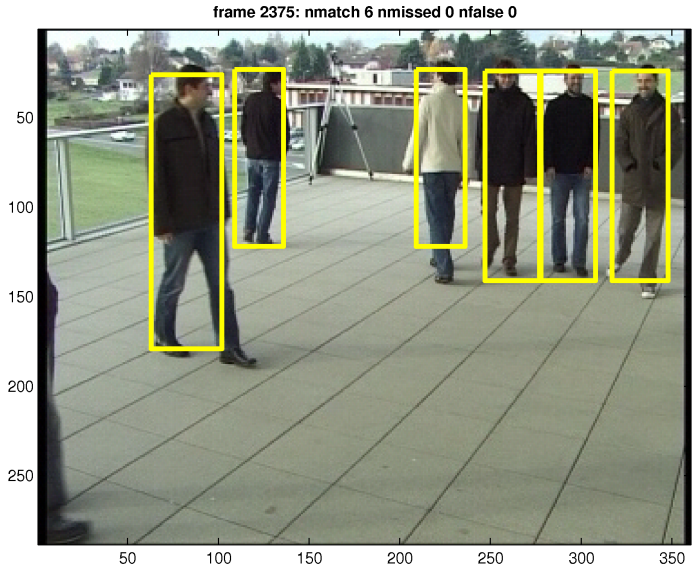
HIGH





gradient ascent (e.g. mean-shift) works OK

LOW



much harder search problem. data association

Other Factors to Consider

single target vs multiple targets (VOT vs MOT)

single camera vs multiple cameras

on-line vs batch-mode (more about this later)

do we have a good generic detector?
(e.g. faces; pedestrians)

does object have multiple parts?

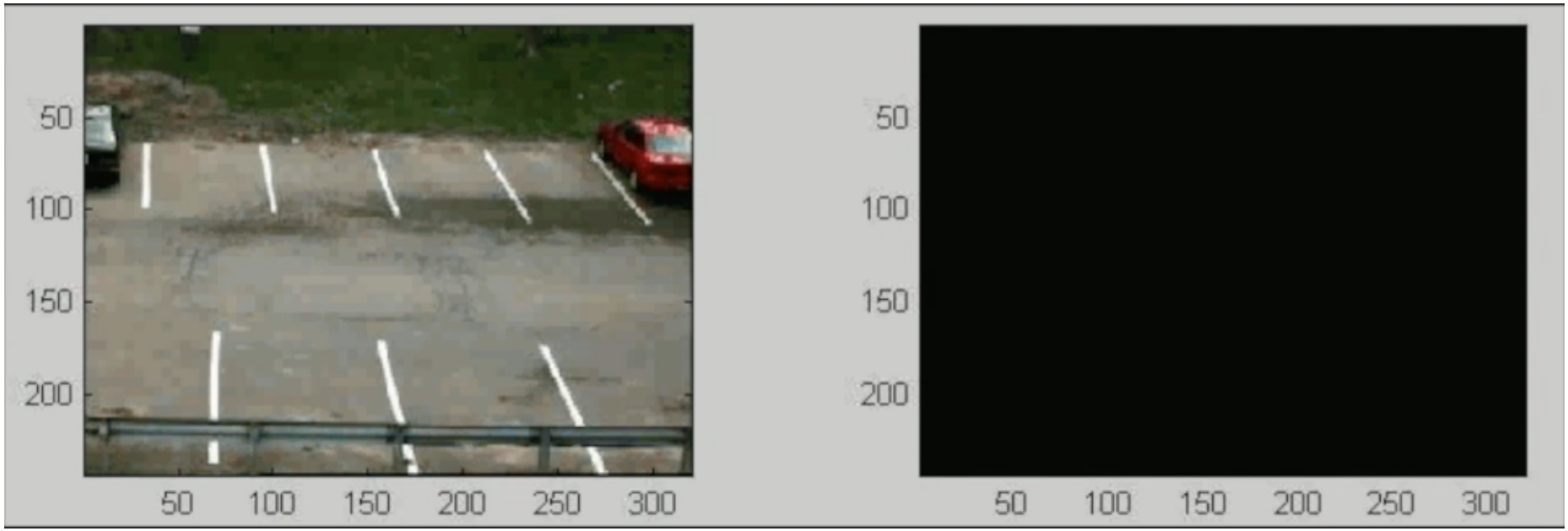
Cavaet

- This is not a survey or literature review.
- Trying to identify rough trends in detection, appearance modeling and data association algorithms for tracking.
- It won't necessarily be a source of good future research problems for you to work on.

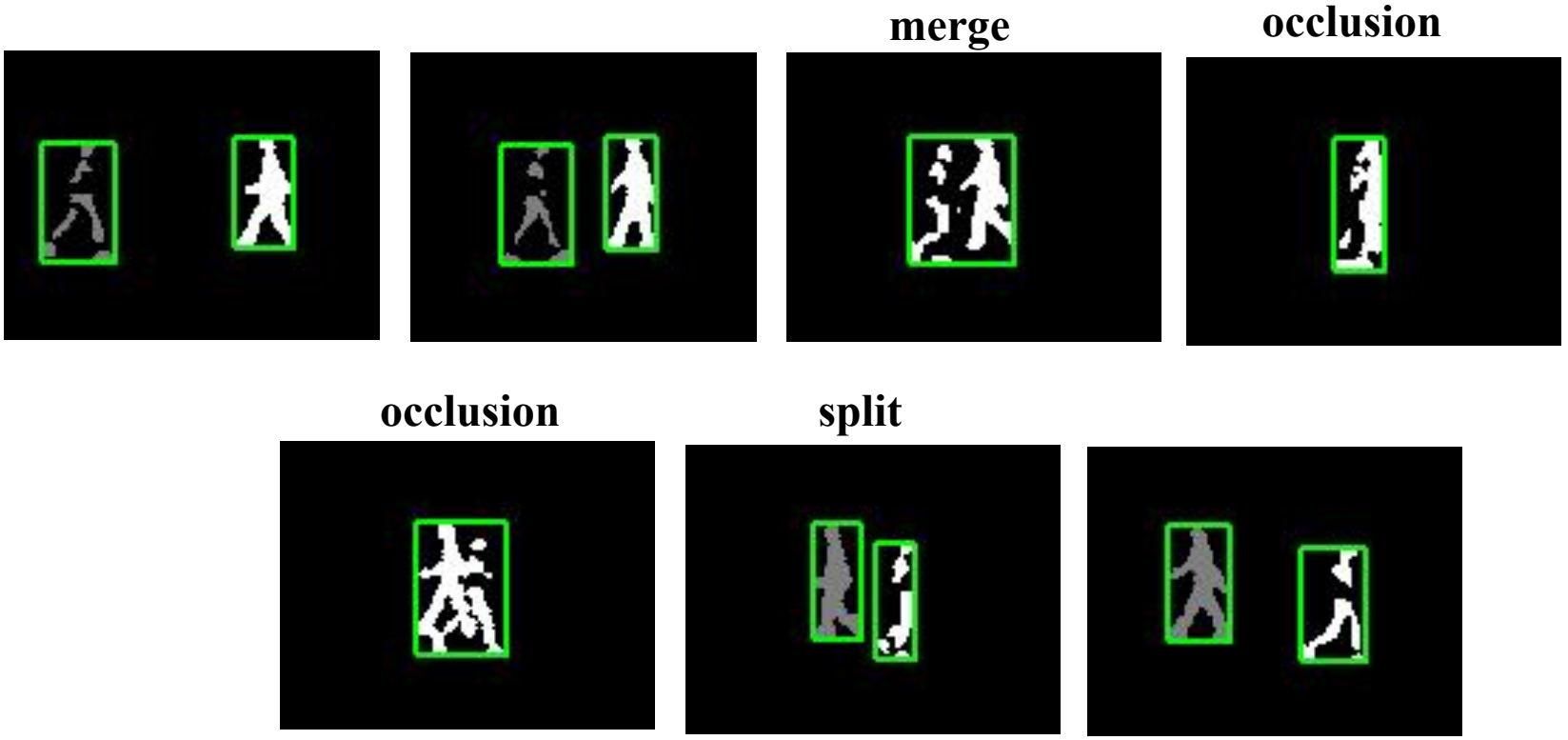
Detector Evolution

Motion Blobs

background subtraction or frame difference



Blob Merge/Split

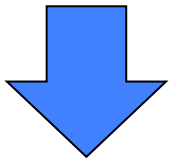


Something I’m glad to never think about again.

Detector Evolution

Motion Blobs

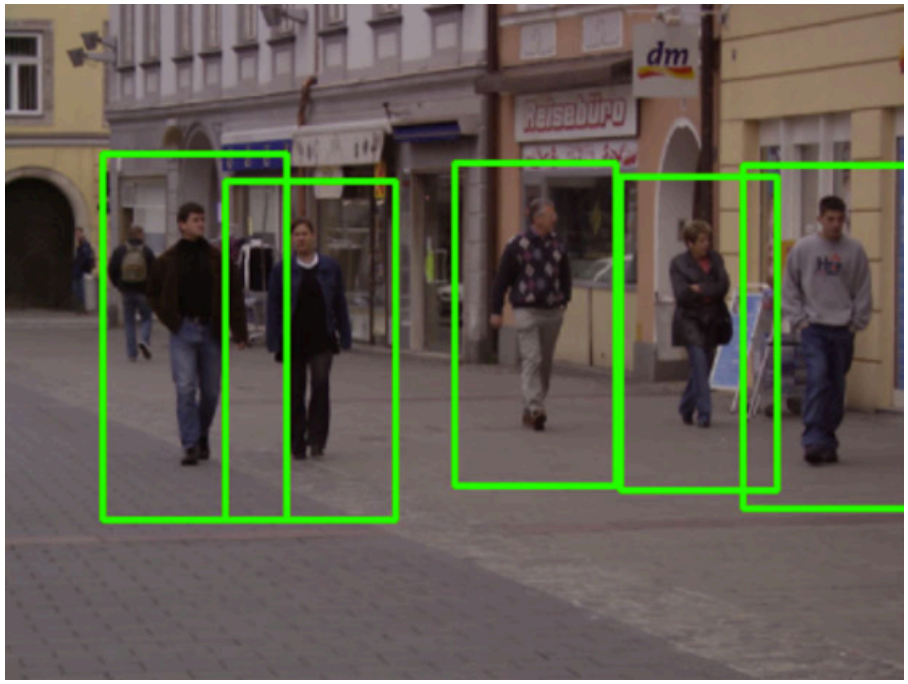
background subtraction or frame difference



Category Location

e.g. pedestrian; car

bounding box representation

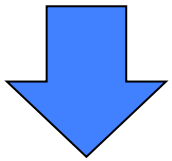


OpenCV detector - based on Dalal and Triggs 2005

Detector Evolution

Motion Blobs

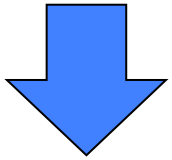
background subtraction or frame difference



Category Location

e.g. pedestrian; car

bounding box representation

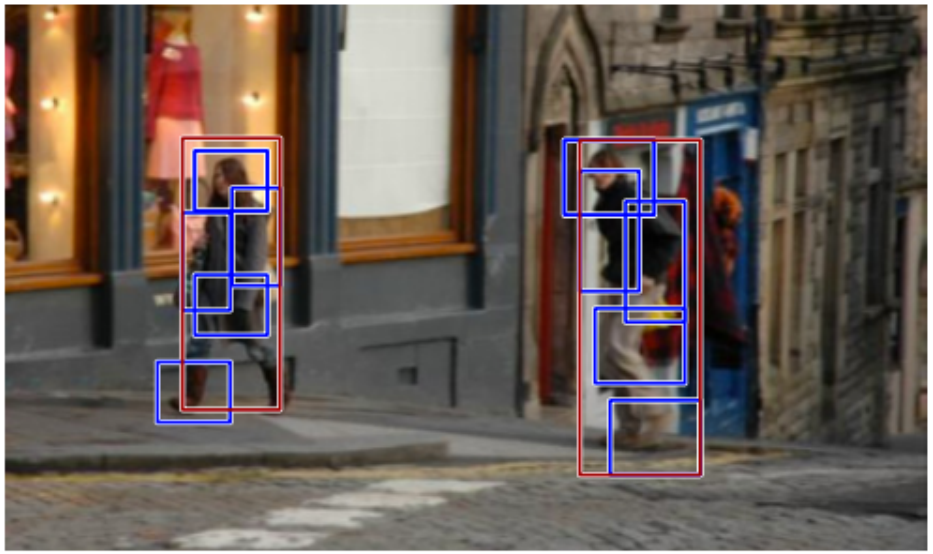


Category Pose

Deformable parts model (Felzenswalb et.al.)

Convolutional pose machines (Wei et.al.; Cao et.al.)

DPM, Felzenswalb et.al. CVPR'08



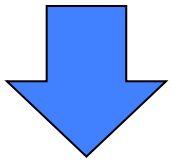
Realtime MultiPerson 2D Pose Estimation using Part Affinity Fields

Cao, Simon, Wei and Sheikh, CMU [CVPR 2017]

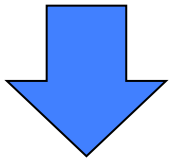
https://github.com/ZheC/Realtime_Multi-Person_Pose_Estimation

Detector Evolution

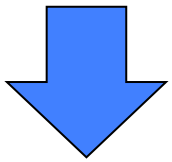
Motion Blobs



Category Location



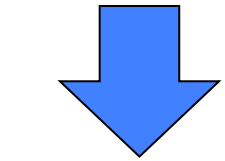
Category Pose



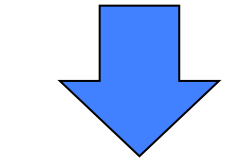
?

Detector Evolution

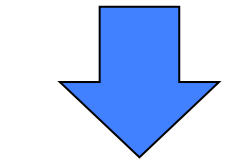
Motion Blobs



Category Location



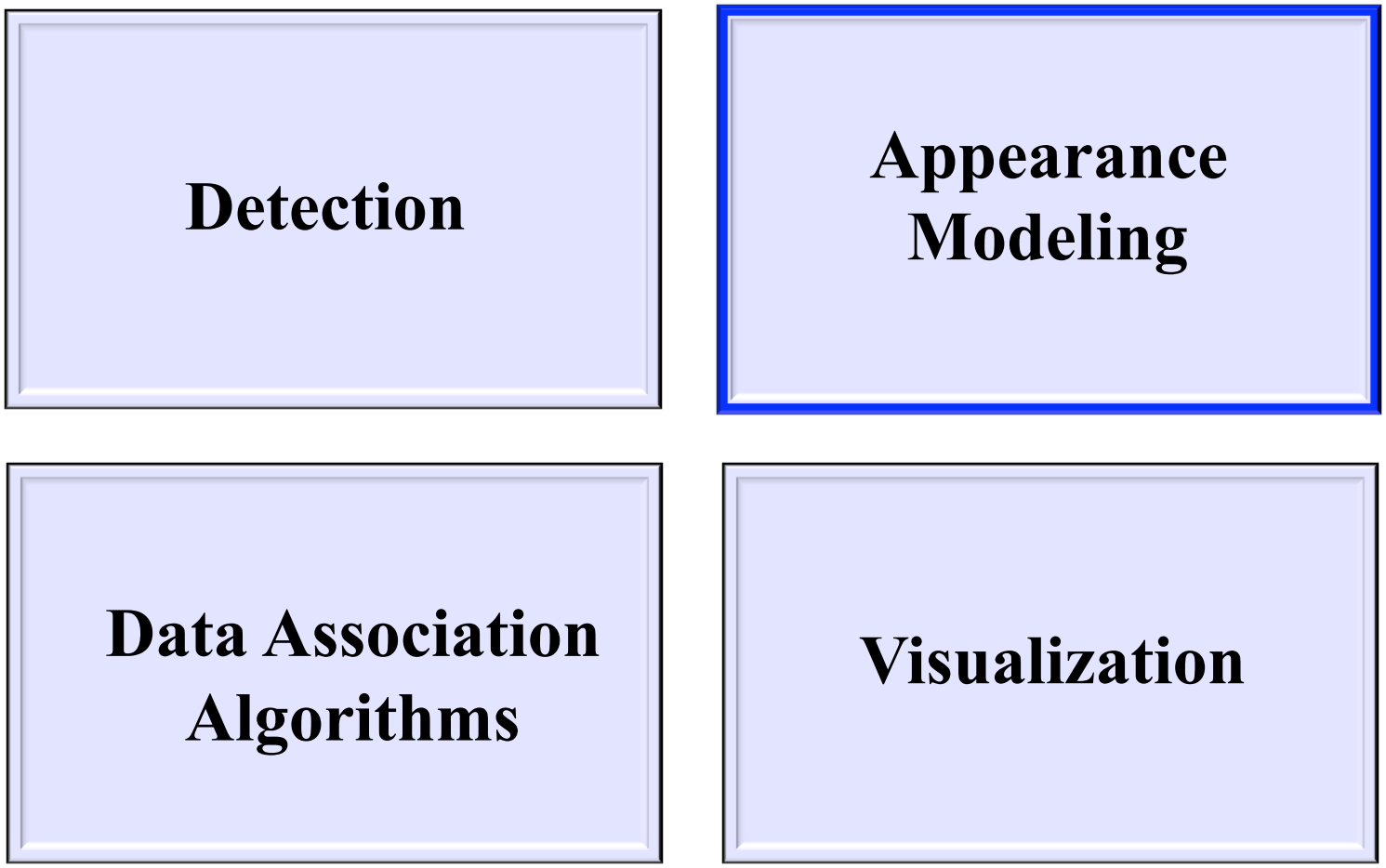
Category Pose



Specific Individual

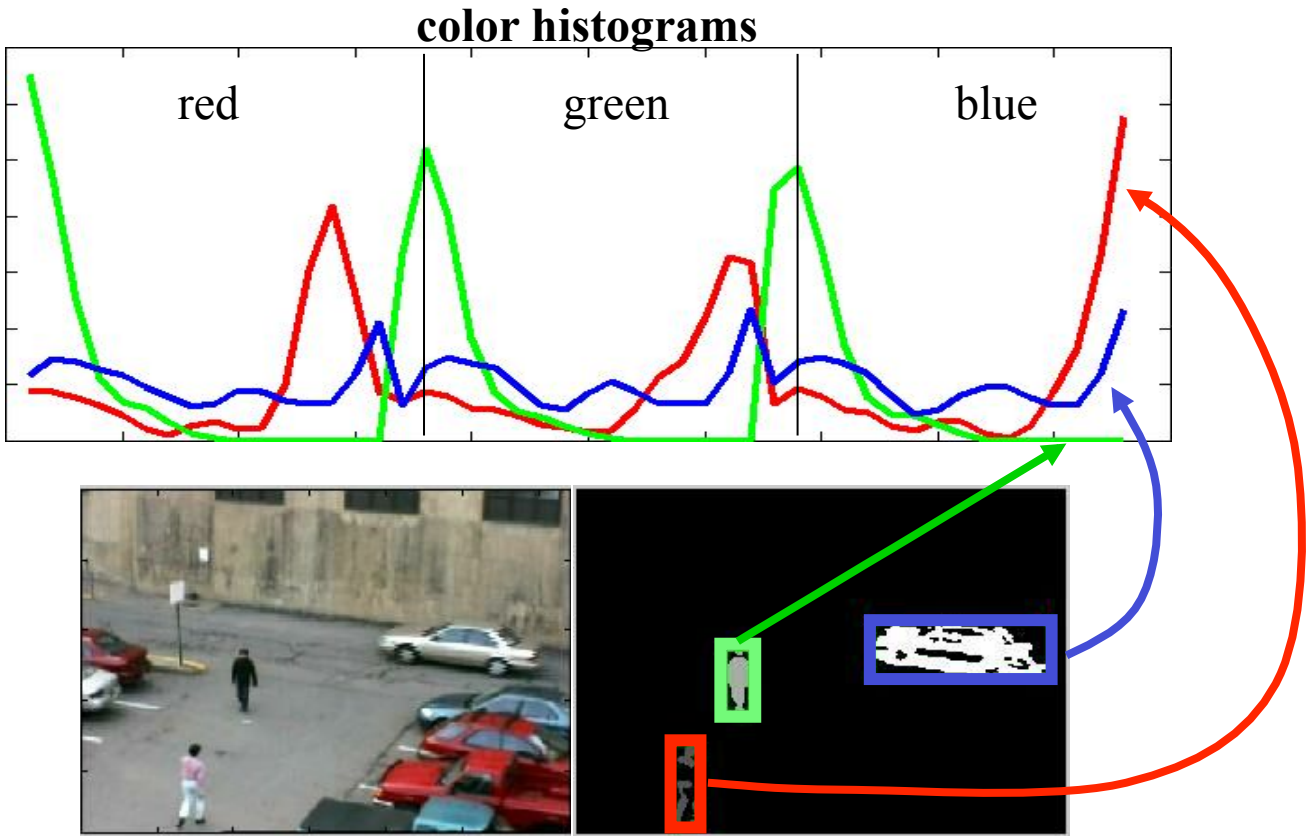
(e.g. Anton Milan detector)

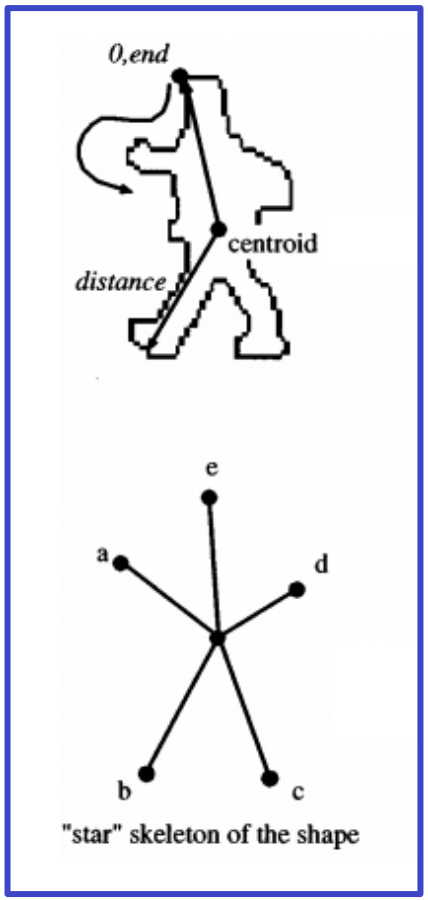
Roadmap


Detection
Appearance Modeling
Data Association Algorithms
Visualization

Appearance Modeling

- Early methods described color, shape of blobs





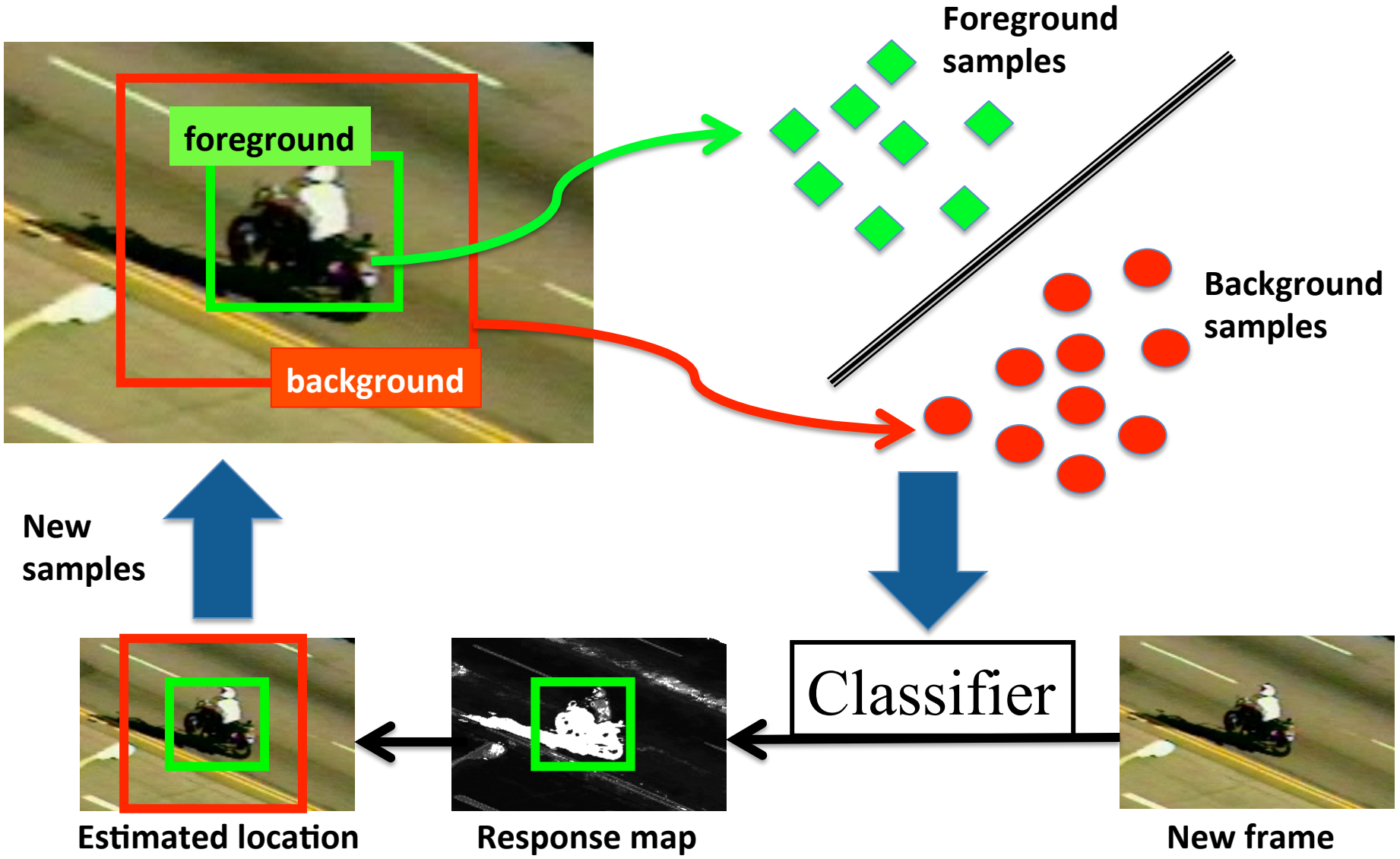
Tracking as Classification

- Target tracking treated as a binary classification problem that discriminates foreground object from scene background.

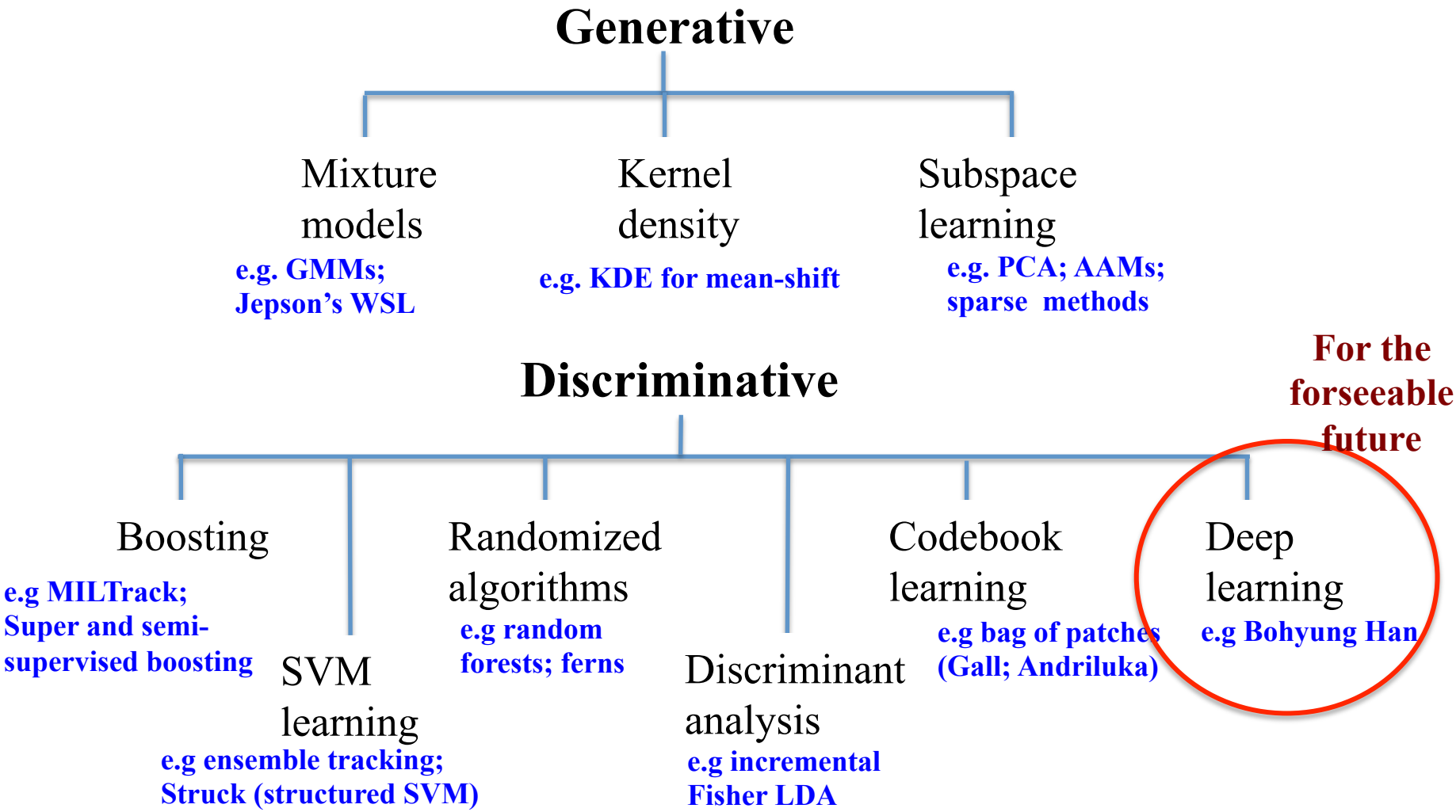
- This point of view opens up a wide range of classification and feature selection techniques that can be adapted for use in tracking.

- Some early works:
  - Collins and Liu, "Online Selection of Discriminative Tracking Features," ICCV'03; PAMI'05
  - Avidan, "Ensemble Tracking," CVPR'05; PAMI'07
  - Grabner, Grabner, and Bischof, "Real-time tracking via on-line boosting," BMVC'06.

Tracking as Classification:



Statistical Appearance Modeling for Tracking by Detection



Adapted from Li et.al., A Survey of Appearance Models in Visual Object Tracking, 2013

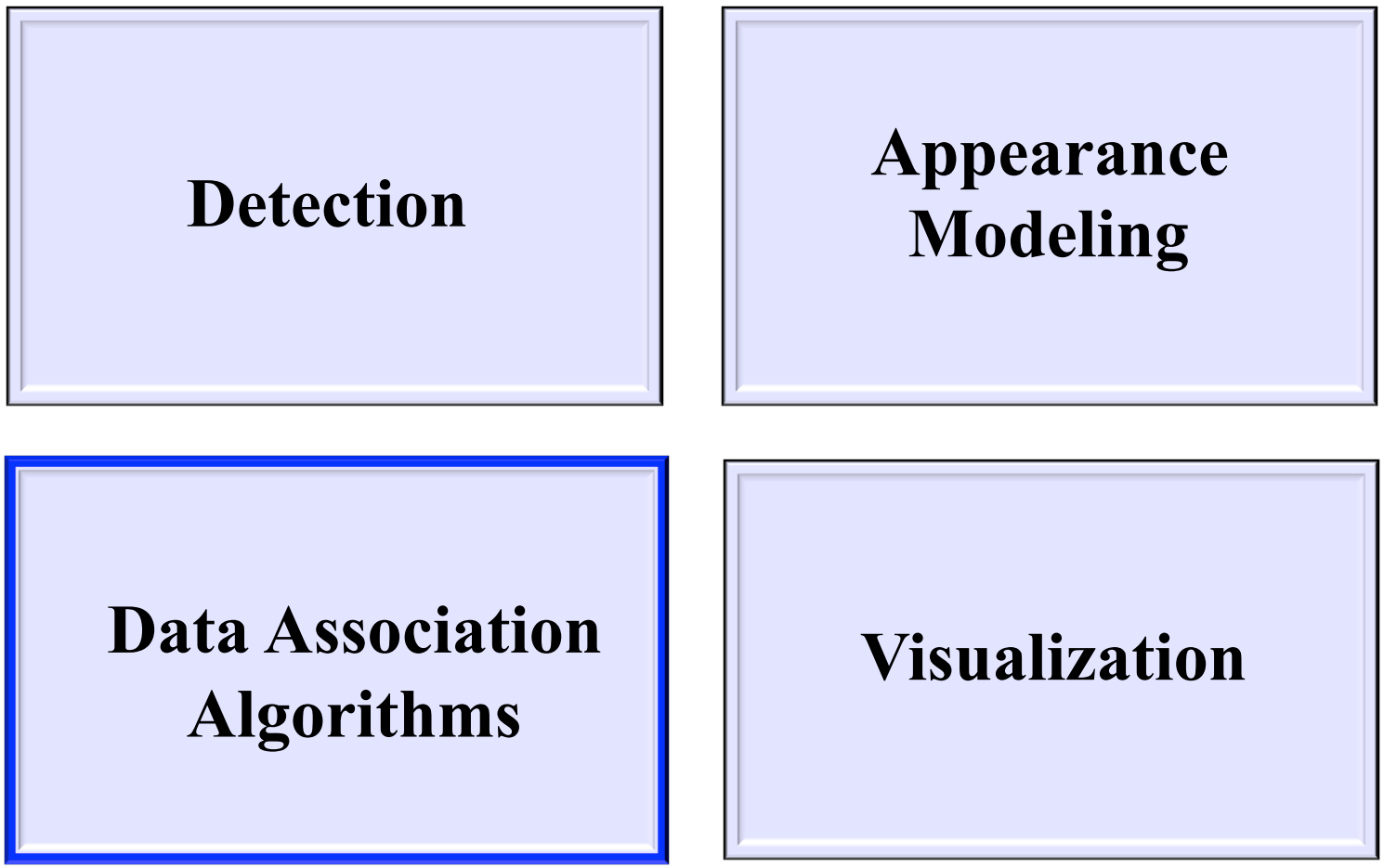
Mean-Shift Nostalgia

Real-time blob tracking based on color distributions

Gary Bradski's Camshift, 1998

Real-time camera control, circa 2001

Roadmap


Detection
Appearance Modeling
Data Association Algorithms
Visualization

Tracking Algorithms
Filtering vs Data Association

- Filtering
  - Bayesian; recursive
  - (continuous) Probability Theory
  - Kalman filter; particle filter; mean-shift; …

usually single object

- Data Association
  - Assignment problems
  - (discrete) Combinatorics
  - Kuhn-Munkres; network flow; ...

usually multiple objects

Discrete-Continuous

- Early precursor (and still a good baseline)

  Kalman filter predictions
  Data association between predictions and observations in next frame
  Update KF trajectories

Blackman and Popoli, *Design and Analysis of Modern Tracking Systems*, 1999.

On-line vs Batch-mode

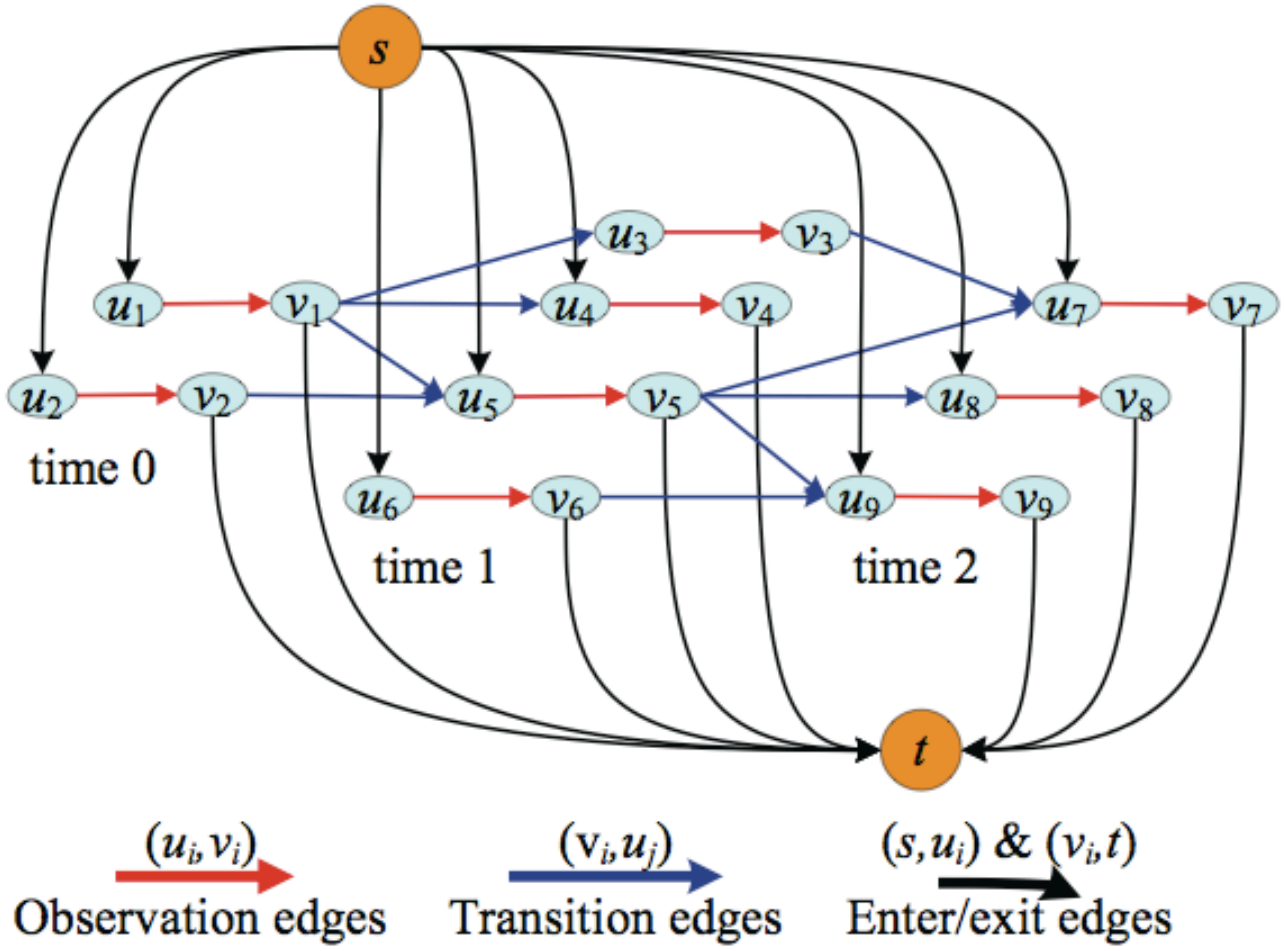
You can afford to do more computation in batch.

However, it becomes tempting to look for the

.Globally.Optimal.Solution.

After which time, nearly everything you want to do becomes NP-hard.

Important Example: Network Flow



picture from Zhang, Li and Nevatia, "**Global Data Association for Multi-Object Tracking Using Network Flows,"** CVPR 2008.

See also Berclaz et.al. 2011 and Pirsiavash et.al. 2011 (successive shortest path algs)

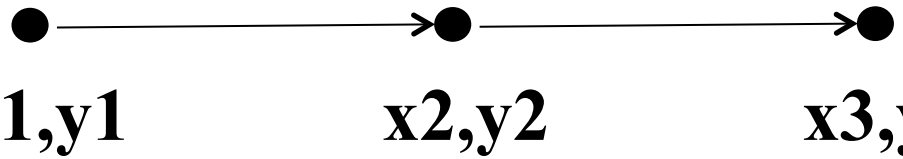
Limitations of Network Flow

Pros:

- Efficient (polynomial time)
- Uses all frames to achieve a global batch solution

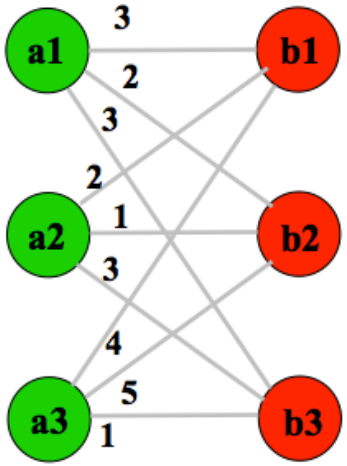
Cons:

- Data association cost functions limited to pairwise terms
- Cannot represent constant velocity or other higher-order motion models
- Will therefore have trouble when appearance information is not discriminative and/or frame rate is low

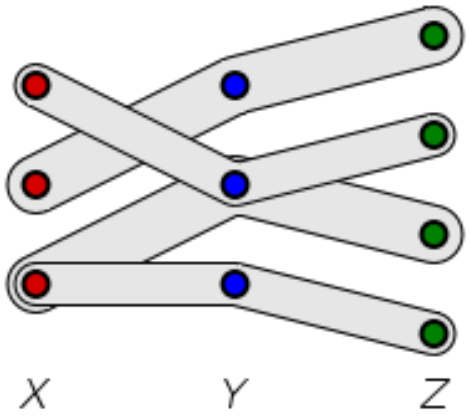


Why is nearly everything else NP-hard?

- Multi-dimensional assignment is NP-hard, including tri-partite (3 frame) matching
- Integer linear or quadratic programming is in general NP-hard

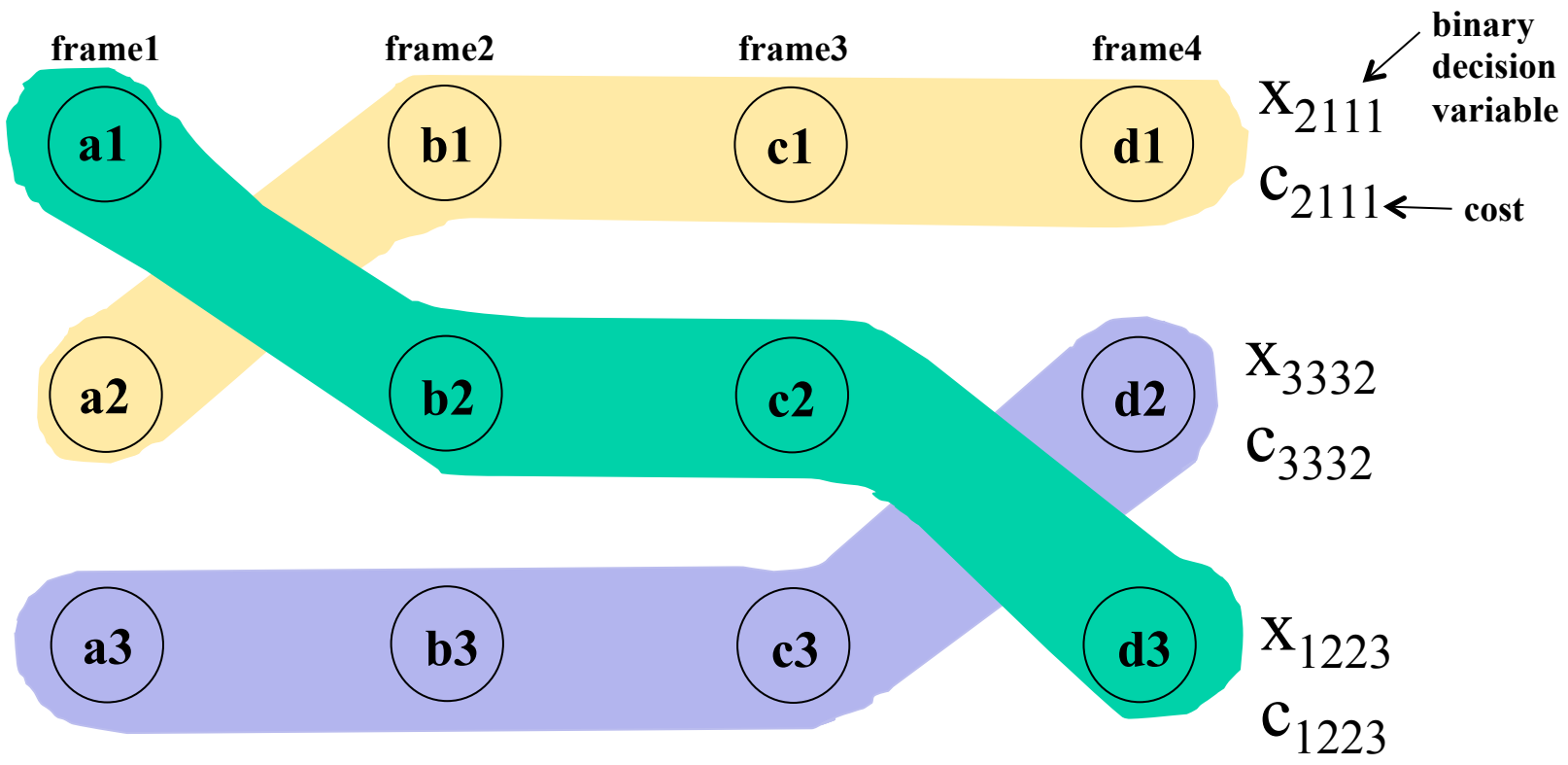


Easy



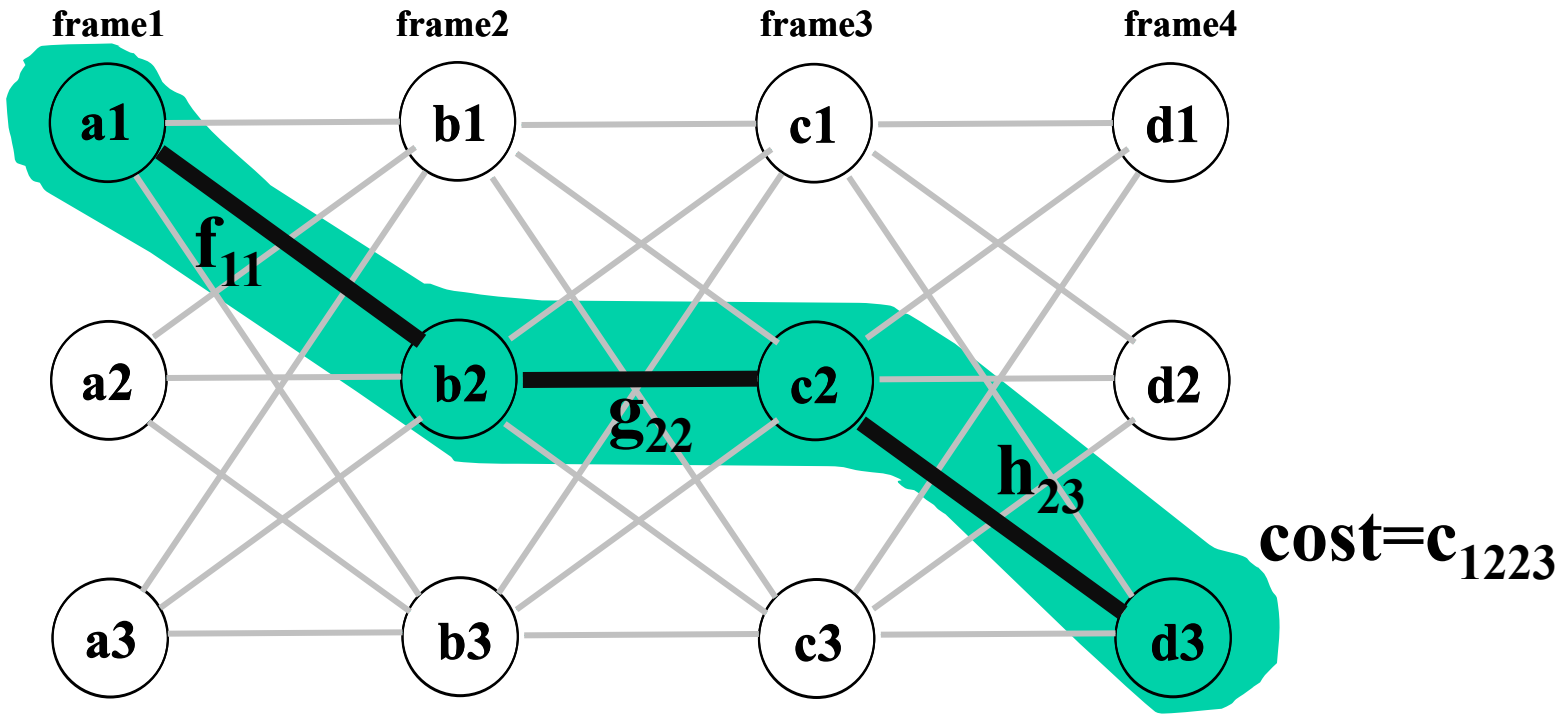
Hard

Multi-Dimensional Assignment



Alternative to network flow allowing higher-order cost functions. Costs and binary decision variables defined over hyperedges rather than edges. NP-hard.

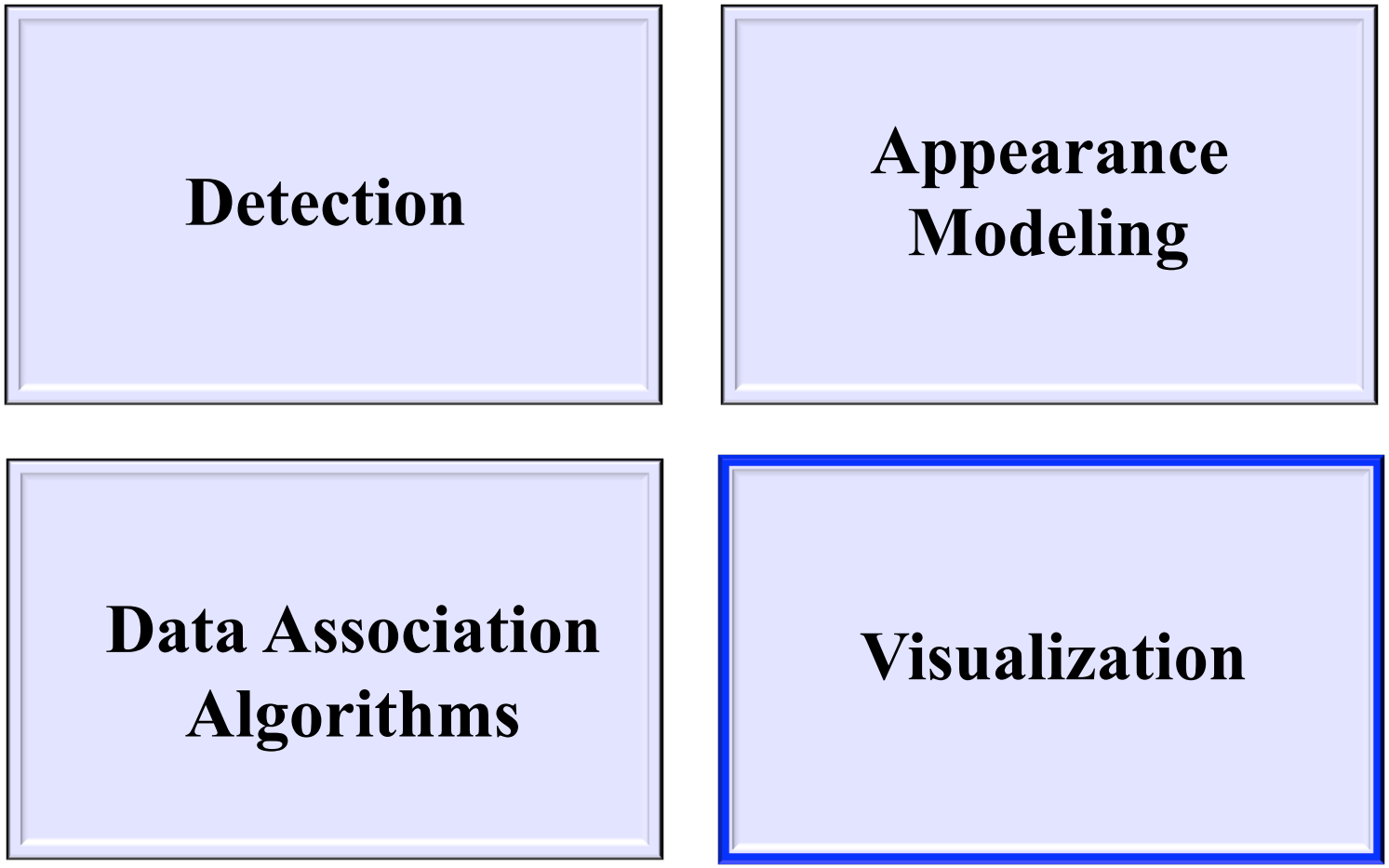
An Interesting Hybrid Model



Decision variables factor pairwise. Allows local updates.
Costs costs remain unfactored. Allows higher-order costs.

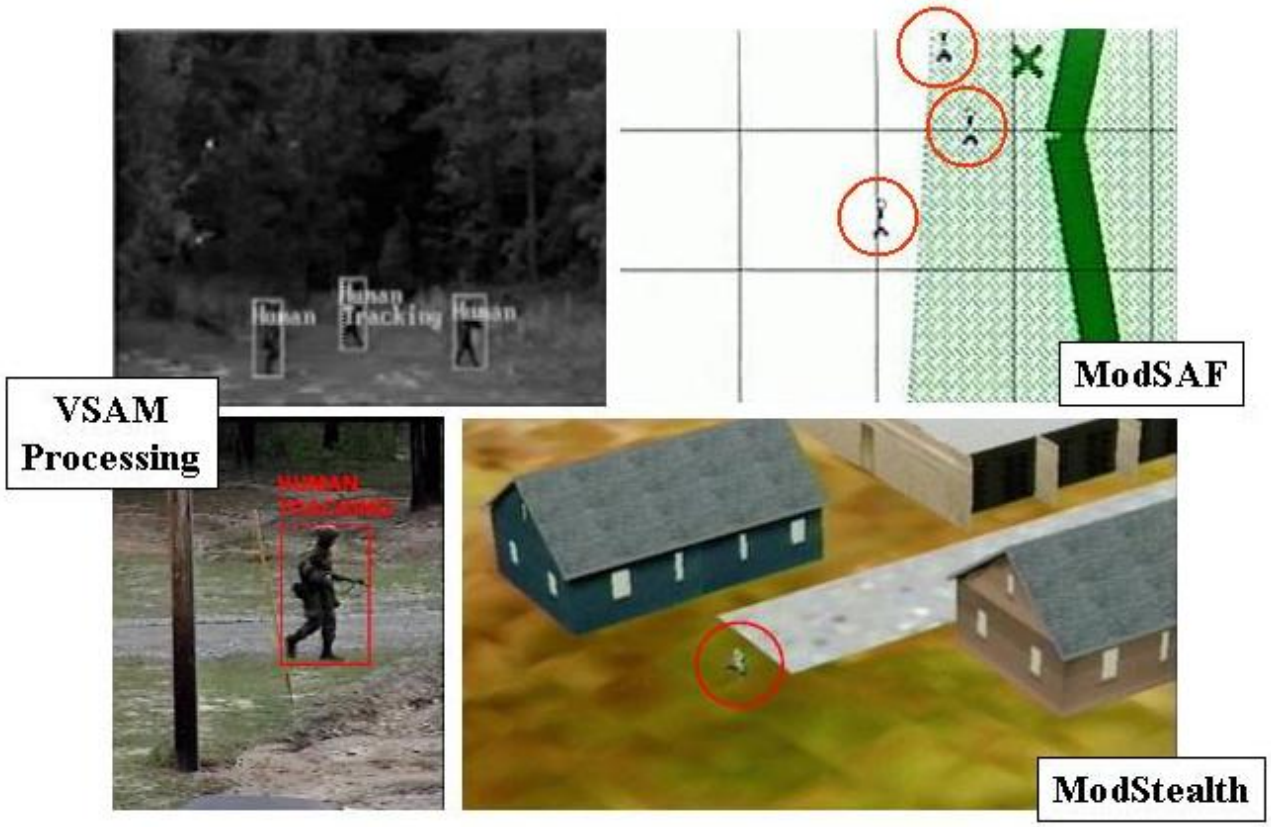
Collins CVPR'12; Butt and Collins CVPR'13

Roadmap


Detection
Appearance Modeling
Data Association Algorithms
Visualization

Visualization

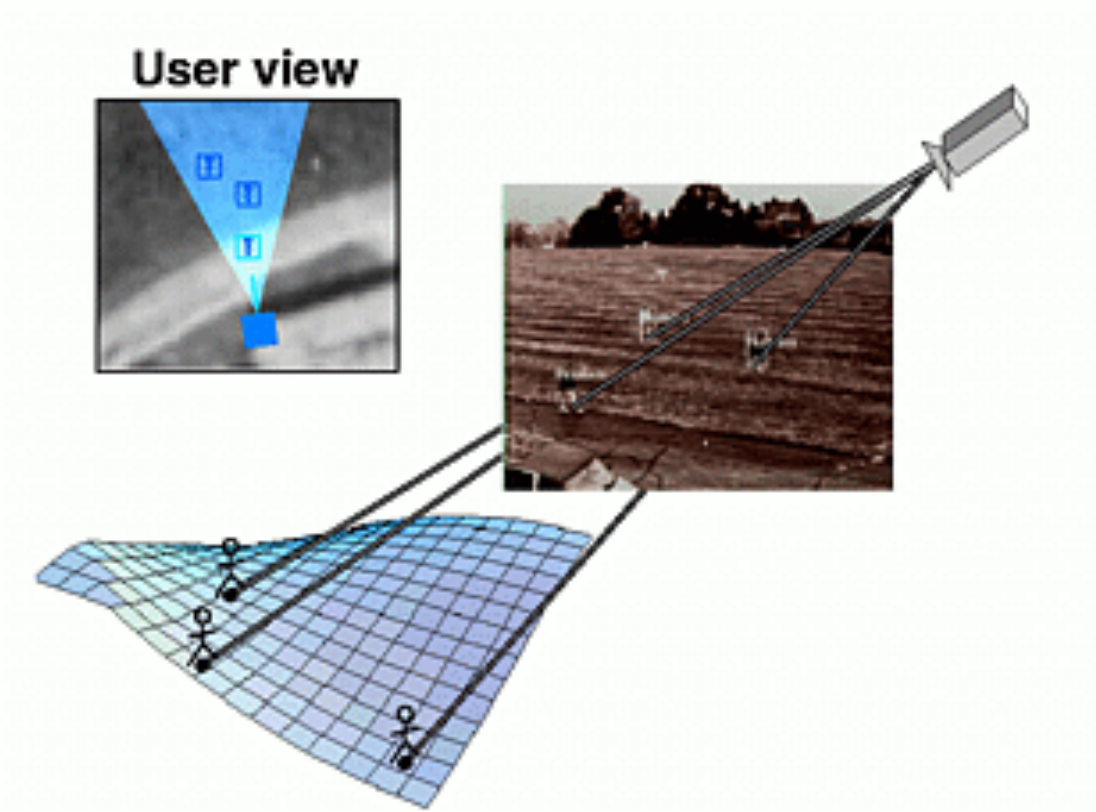
Methods for intuitively exploring output from a tracking/surveillance system.



VSAM project, 1997-2000

Visualization

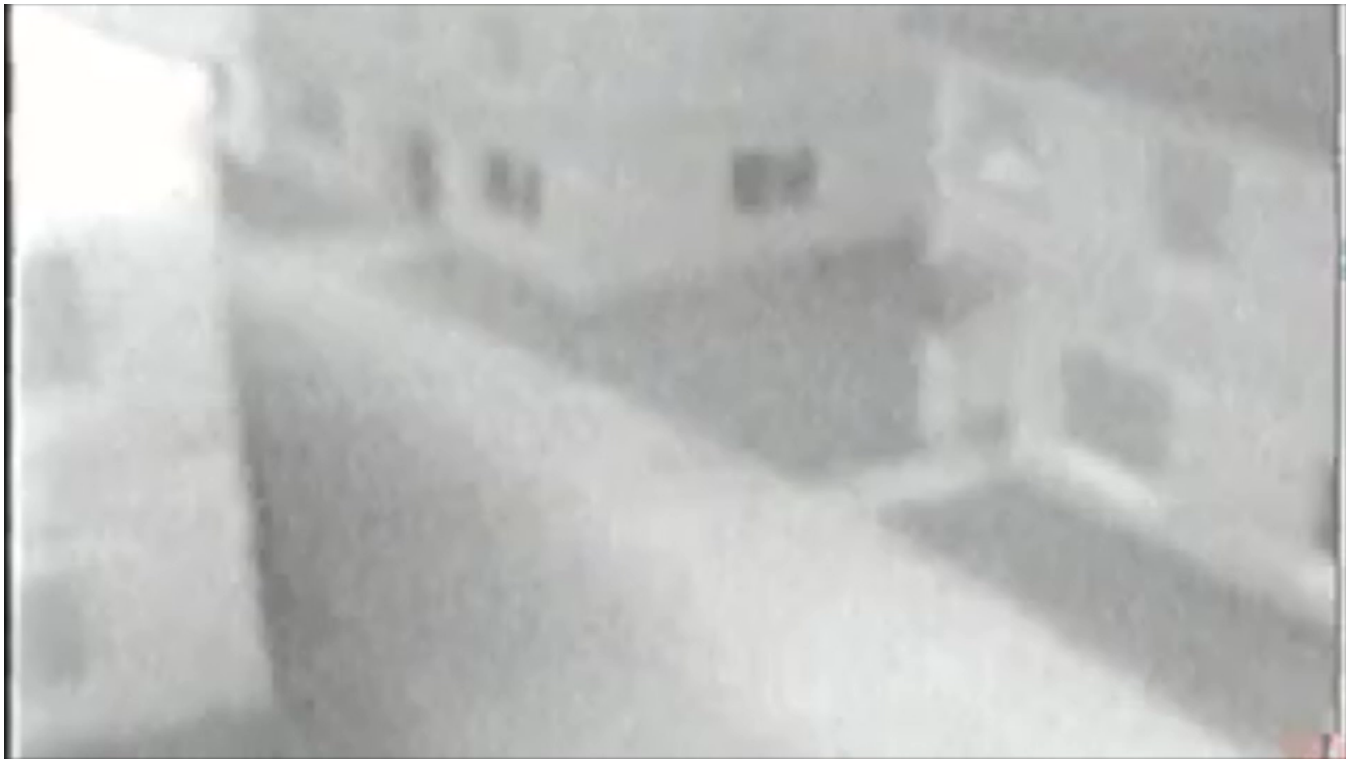
Methods for intuitively exploring output from a tracking/surveillance system.



VSAM project, 1997-2000

Visualization

Methods for intuitively exploring output from a tracking/surveillance system.



VSAM project, 1997-2000

Visualization

We could do a much better job today, and mostly automatically, by combining GPS, camera pose estimation; Google Earth and Street View models.

See for example Park, Luo, Collins and Liu 2014

Where are we going

- Specific individual detectors for absolute ID.
- Specializing generic into specific object detectors for re-ID.
- Incorporate body pose evolution into tracking.
- Embrace deep learning...
- Seek provable guarantees for approximate solutions to NP-hard batch-mode problems.
- Get on board the AR/VR wave wrt visualization.