Practical Tour of Visual tracking

David Fleet and Allan Jepson

January, 2006

Designing a Visual Tracker:

What is the state?

- pose and motion (position, velocity, acceleration, …)
- shape (size, deformation, articulation, …)
- appearance (subspace, colours, …)

Dynamics?

- hand crafted or learned from training examples

Which image properties should we use?

- intensity, colour, texture, motion, edges, motion, …
- template, adaptive appearance, region statistics

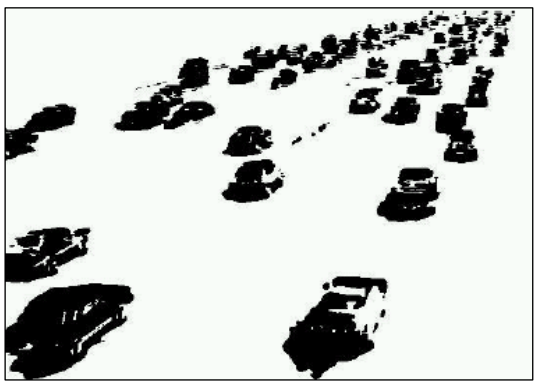
Designing a Visual Tracker:

What might simplify the problem?

- known/stationary background (e.g., track blobs)
- use of color (e.g., skin)
- multiple cameras (often 2 or 3)
- manual initialization
- strong dynamics models
- prior knowledge of the number of objects and object types
- limited (or no) occlusion
- structured noise (e.g., shadows, background clutter)

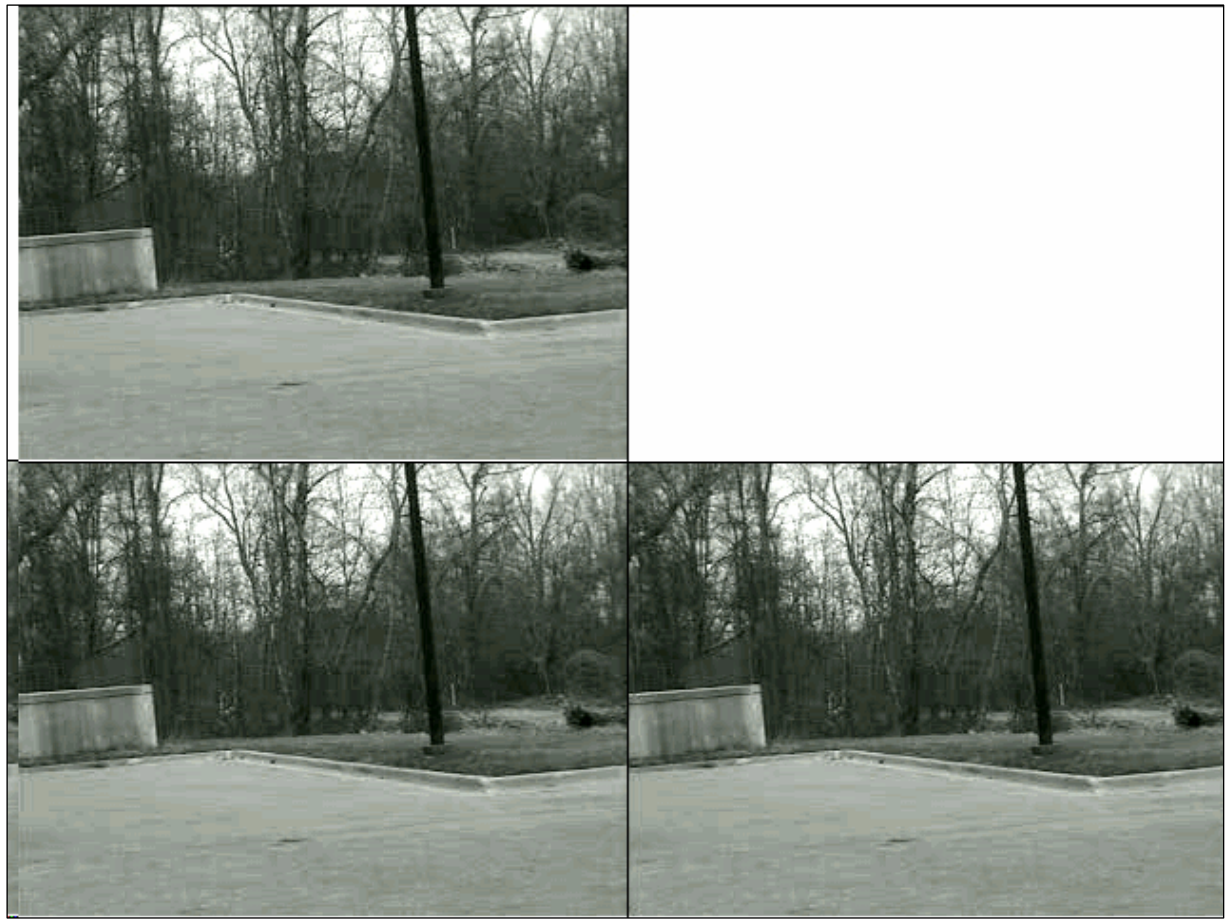
Keep these in mind for the assignment.

Vehicle Tracking: Background Subtraction



[Koller, Weber & Malik, "Robust multiple car tracking with occlusion reasoning." Proc ECCV , 1994]

People Tracking: Background Subtraction



[Haritaoglu, Harwood & Davis, "W4: Who, when, where, what: A real-time system for detecting and tracking people." Proc Face & Gesture Conf, 1998]

Background Modeling

Median filter (or robust fitting over time)

- good for static background and transient foregrounds

Fit background with parametric model (over space-time)

- e.g., with polynomial, spline, or RBF basis set

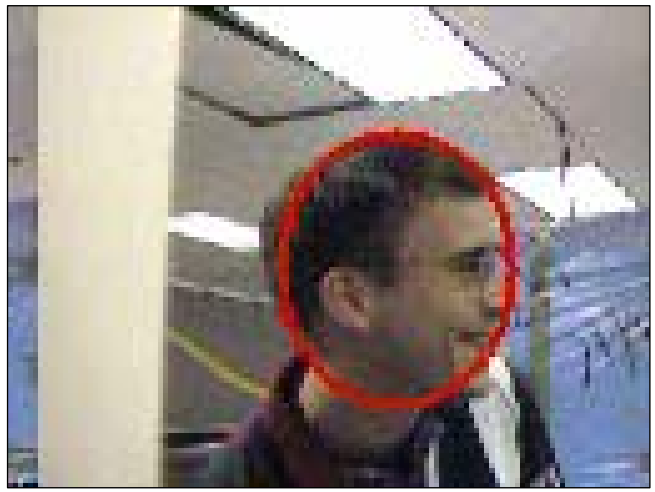
Gaussian mixture models

- probabilistic models help account for natural variation
- adaptation to handle (smoothly) time-varying backgrounds

Brightness and contrast normalization help provide some degree of illumination invariance.

Resources: see the *mixture model tutorial* in utvis Matlab toolbox

Tracking with Histograms



State: 2D image location, size of ellipse, velocity

Appearance model: colour histogram and elliptical contour

Estimation: search over discrete locations and sizes

[Birchfield, "Elliptical head tracking using intensity gradients and color histograms." Proc CVPR, 1998]

Mean Shift with Color/Space Histograms

Appearance model: 5D histogram of 3D colour and 2D location

Estimation: Parzen window density estimation and continuous hill-climbing to find modes

[Comaniciu, Ramesh & Meer, "Kernel-based tracking", IEEE Trans PAMI, 2003]

2.1D Blob Tracking

State: number of people, their positions/velocities on ground plane, and simple shape models (10 dimensions / person)

Appearance: filter response histograms for background, and foreground people

Dynamics: damped 2nd-order model for position/velocity, 1st-order for shape model

Inference: particle filter (1 person ~500 particles, 2-3 people >10,000)

[Isard and MacCormick, "Bramble: A Multiple Blob Bayesian Tracker." Proc ICCV, 2001]

Histogram Tracking

Salient region properties:

- intensity, textons, edge strength/orientation, color direction, …

Histogram matching

- Bhattacharya coefficient, squared distance (L_2), EMD, KL divergence, …

Pros / Cons

- fast
- good for large variations in shape/appearance, 3D rotations
- limited fidelity in pose estimation

Resources: see the *Canny edge tutorial* in the utvis Matlab toolbox.

Motion-Based Tracking

[Shi and Tomasi, “Good features to track.” Proc IEEE CVPR, 1994]

2D Tracking with Adaptive Appearance Model

[Jepson, Fleet, & El-Maraghi, "Robust, on-line appearance models for visual tracking." IEEE Trans. PAMI, 2003]

Tracking with Subspace Appearance Model

Bee appearance is constrained to a linear subspace of basis images that is learned prior to tracking.

[Khan, Balch, & Dellaert, "A Rao-Blackwellized particle filter for eigentracking" Proc CVPR, 2004]

Template Appearance Models

Appearance models (templates)

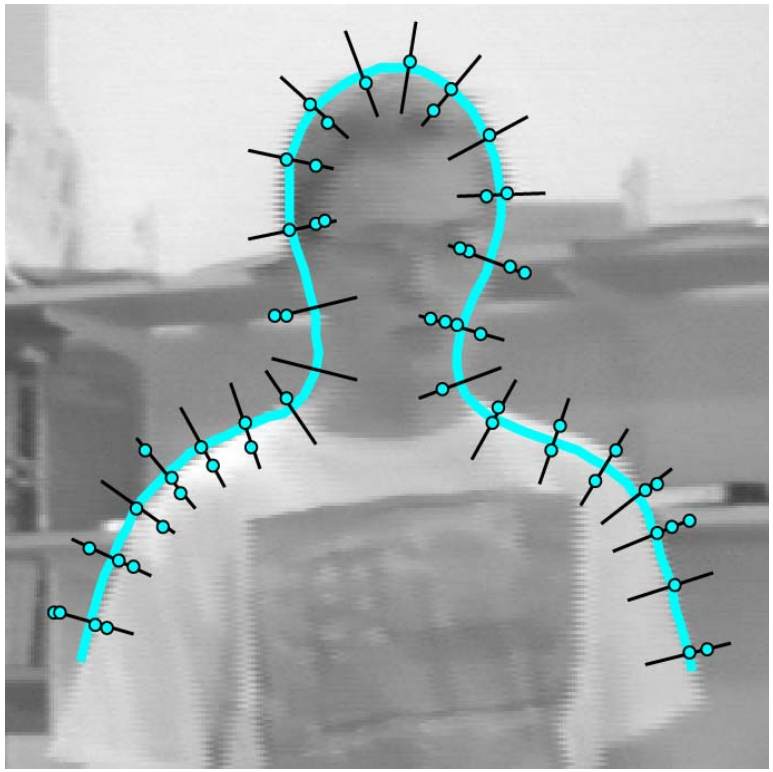
- Fixed: extract an image of the object in frame 1
  - *Problem:* variations in appearance
- Flow-based – template is given by previous image / state
  - *Problem:* drift
- Adaptive – slowly adapt the template over time (e.g., IIR filter)
- Subspace models
  - *Problem:* prior learning required

Tracking:

- Optical flow estimation with parametric flow models (e.g., affine)

Resources: see the *motion tutorial* in the utvis Matlab toolbox.

Contour Tracking



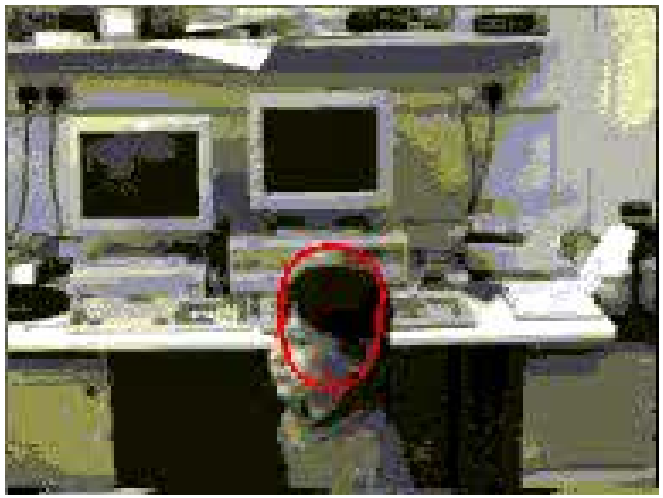
State: control points of spline-based contour representation

Measurements: strength of nearest edgel on line segment perpendicular to contour

Dynamics: 2nd–order Markov (often learned)

[Isard & Blake, "Condensation - conditional density propagation for visual tracking." IJCV, 1998]

2D Contour Tracking

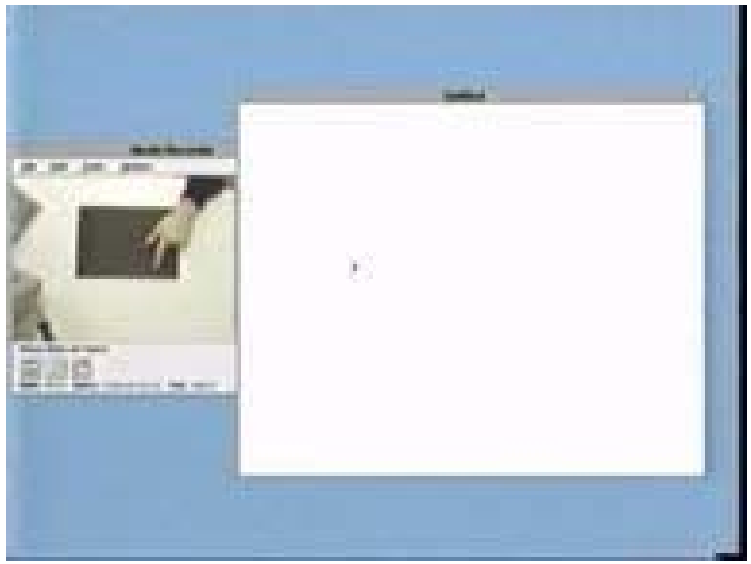


(6D affine state, 100 particles)

(6D affine state, 1200 particles)

[Isard & Blake, "Condensation - conditional density propagation for visual tracking." IJCV, 1998]

Contour Tracking



[Isard & Blake, "Condensation - conditional density propagation for visual tracking." IJCV, 1998]

Contour Trackers

Shape Models

- hand-crafted or learned from examples
- low-dimensional representations are useful
  - e.g., spline control points, subspace basis
- invariance to motion model (deformation class)

Resources: see the *Canny edge tutorial* in the utvis Matlab toolbox.

Multiple Cues

Different properties of image appearance (e.g. bounding contour, motion, and texture (histograms)) are often complementary.

Particle Filter (500 samples, 12D state space)

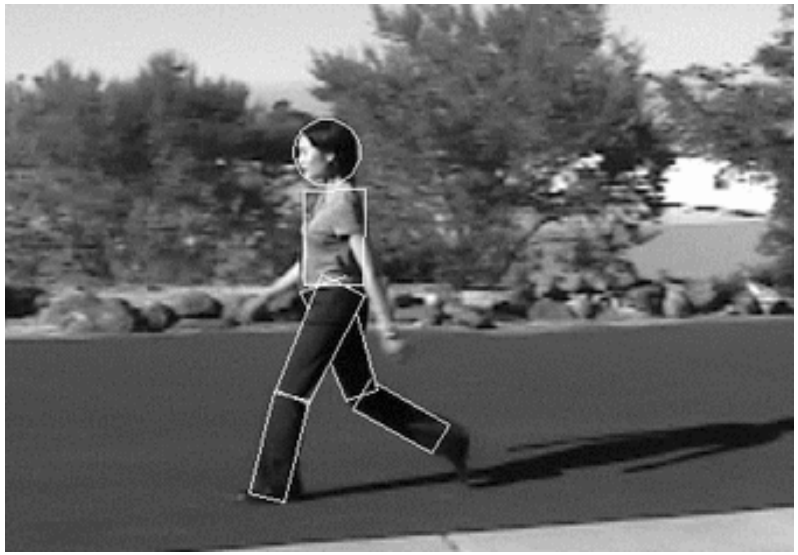
brightness constancy (optical flow)

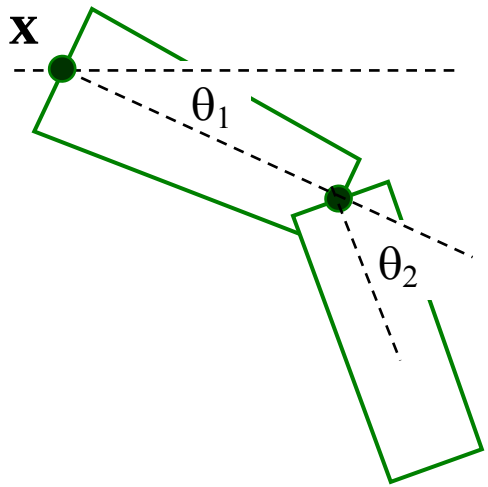
boundary edge strength

motion + edges

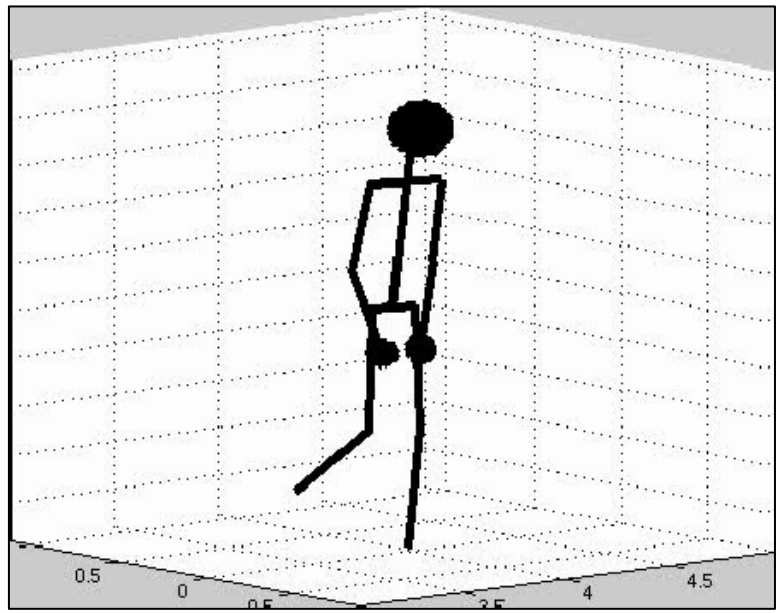
Parametric Models & Articulation

For many complex objects it is often necessary to specify (a priori) or infer a richer parameterization of the object shape.





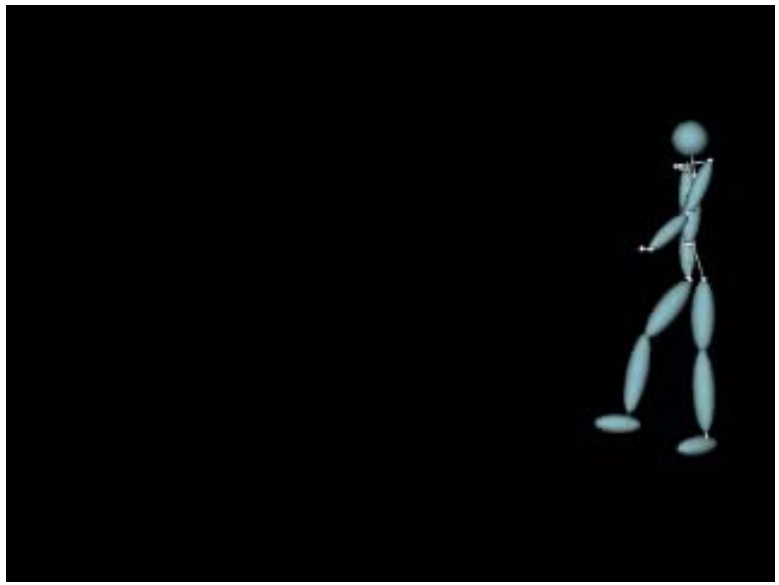
Parametric Models & 3D Tracking



Mean posterior state shown from two viewpoints.
(15000 particles, manual initialization)

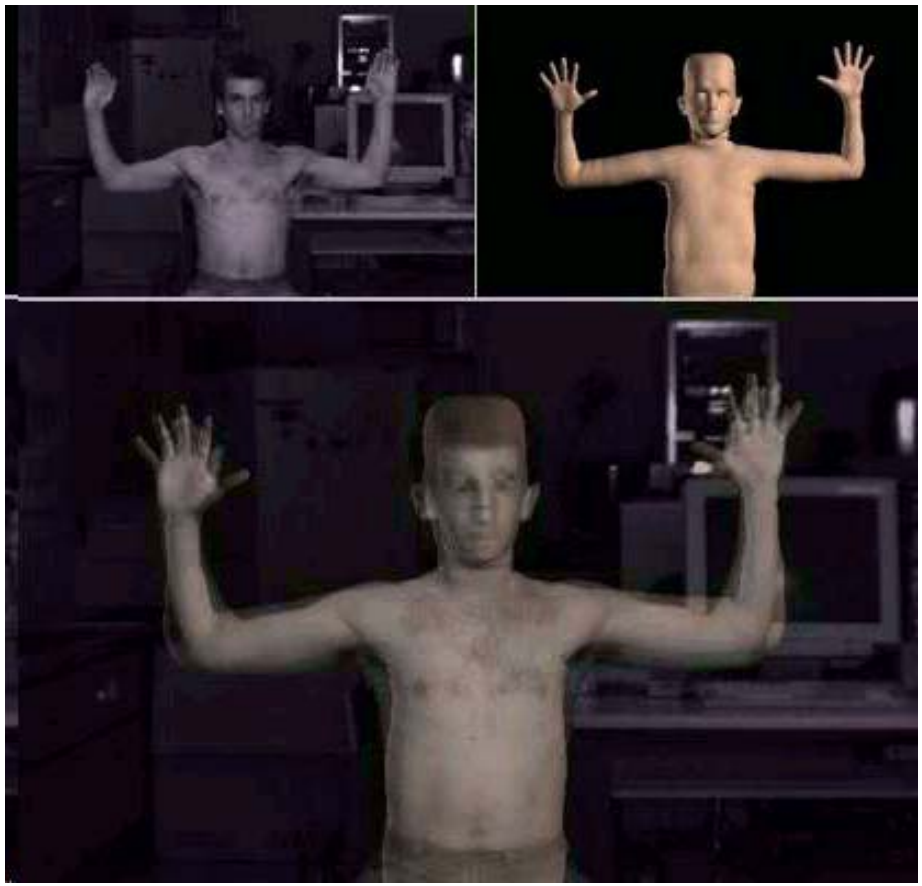
[Sidenbladh, Black & Fleet, "Stochastic tracking of 3D human figures using 2D image motion." Proc ECCV, 2000]

Looking at People



[Urtasun, Fleet & Fua, "Gaussian Process dynamical models for 3D people tracking." (submitted, 2006)]

Multiple Cameras Help



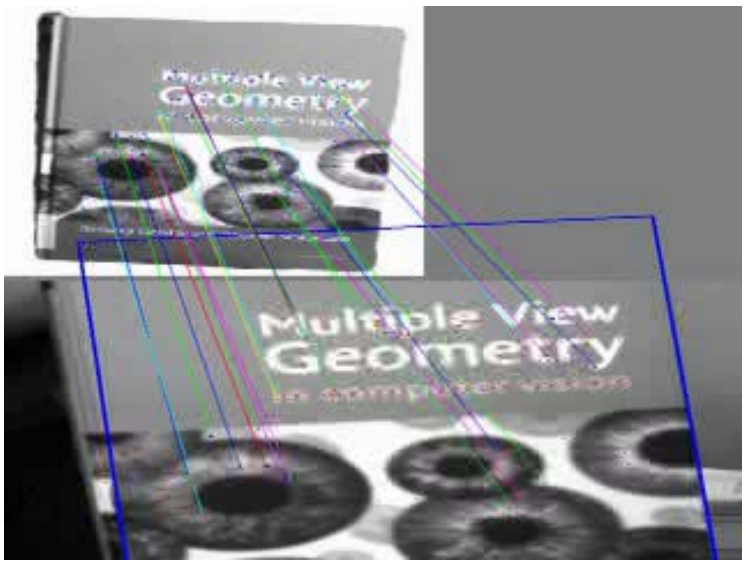
(Hill-climbing on hand-crafted objective function)

[Plankers & Fua, "Articulated soft objects for multiview shape and motion capture." IEEE Trans PAMI, 2003]

Feature-Based Detection (Tracking)

Multiple training images are given, with known object shape and pose, from which feature points are learned.

Real-time feature detection and pose estimation:



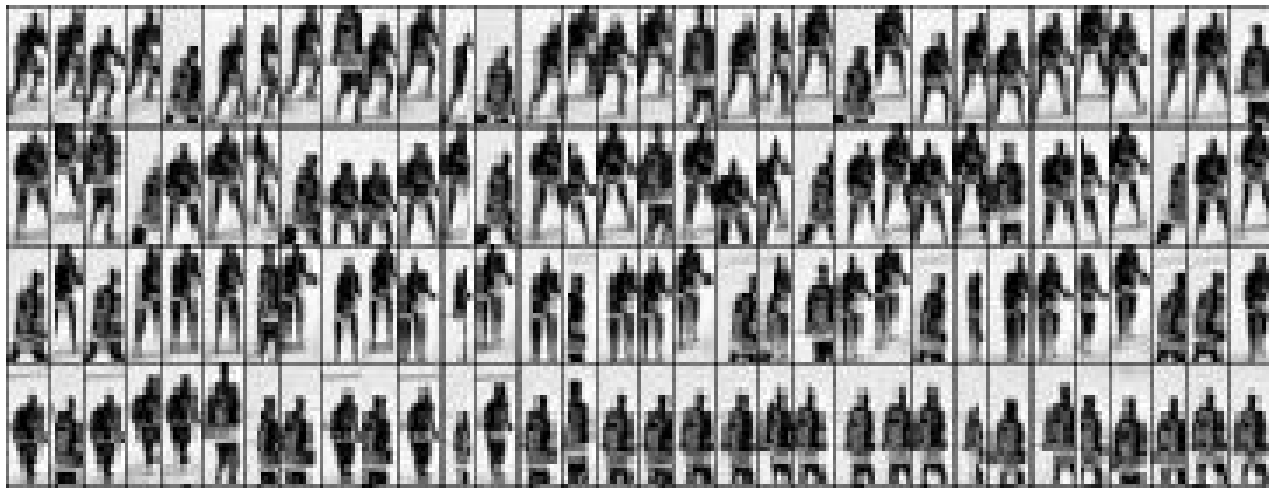
[Lepetit, Pilet & Fua. Point matching as a classification problem and robust object pose estimation. Proc IEEE CVPR 2004]

Resources: see the *SIFT tutorial* in the utvis Matlab toolbox.

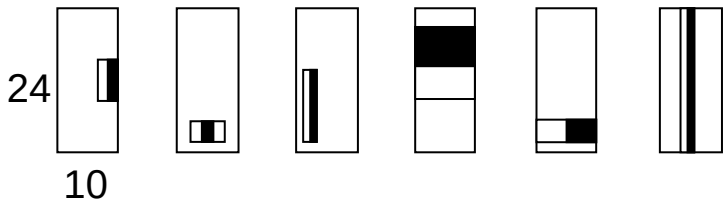
Weak Classifiers

Adaboost used to train a 23 layer classifier to detect hockey players:

- 2000 negative examples from locations on rink without players
- 200 positive examples

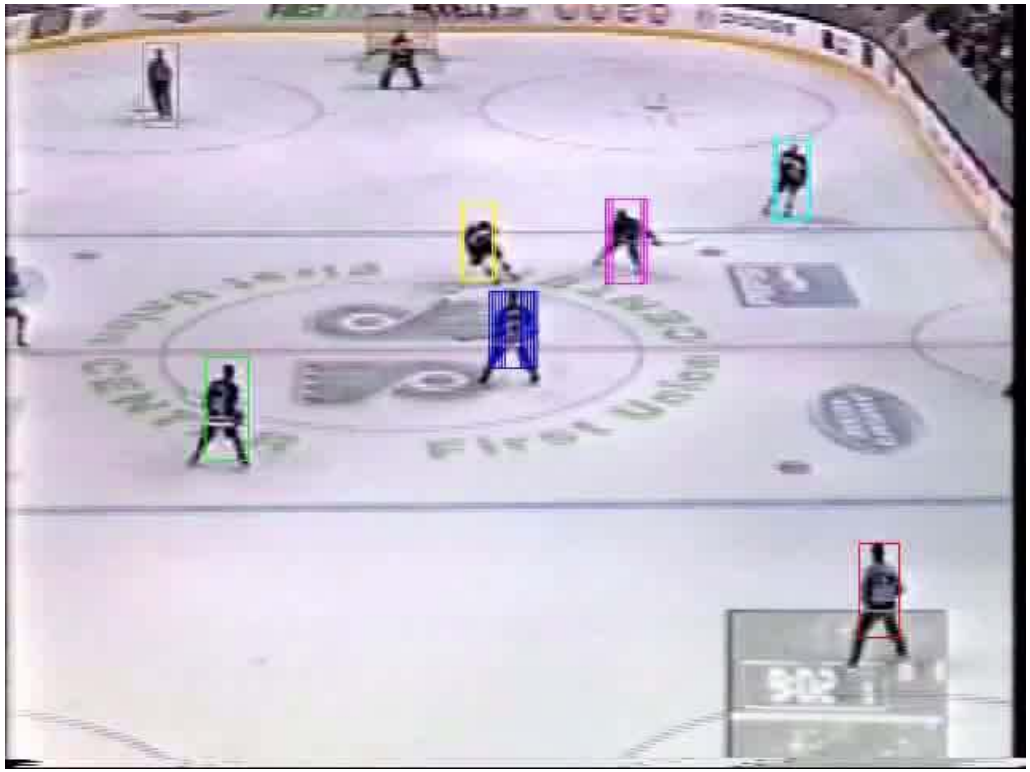


- Key (Haar) features:



[Okuma et al., "Boosted Particle Filter." Proc. ECCV 2004]

Tracking Hockey Players



State: number of players, positions & velocities (in rink coords)

Appearance: color histograms for top & bottom of body

Factored Posterior: independent filters applied to players (unless players in close proximity)

[Okuma et al., "Boosted Particle Filter." Proc. ECCV 2004]

Classifiers

Useful for

- initialization
- tracking loss and recovery
- proposals for efficient search

Remarks:

- E.g., boosted weak classifiers, density models, SVMs, …
- training data often difficult to obtain
- tracking loss and recovery

Resources: see the *eigen tutorial* in the utvis Matlab toolbox.

Back to the Assignment

Don’t panic!