Abnormal "Shape Activity" Detection and Tracking

Namrata Vaswani

Dept. of Electrical and Computer Engineering

Iowa State University

http://www.ece.iastate.edu/~namrata

Collaborators

- Rama Chellappa
- Bi Song, Amit Roy-Chowdhury
- Yogesh Rathi, Allen Tannenbaum
- Anthony Yezzi

Problem Formulation

Problem Formulation

- Modeling activity performed by a group of moving and interacting point "objects" ("landmarks").
- "Objects": People, Vehicles, Robots, Human body parts.
- **Changing configuration of the group: moving & deforming shape**
- **"Shape Activity": model activity performed by a group of moving & interacting "objects" by its shape dynamics**
- **"Abnormal Activity": change in learned shape dynamical model, which could be slow or sudden and whose parameters are unknown**

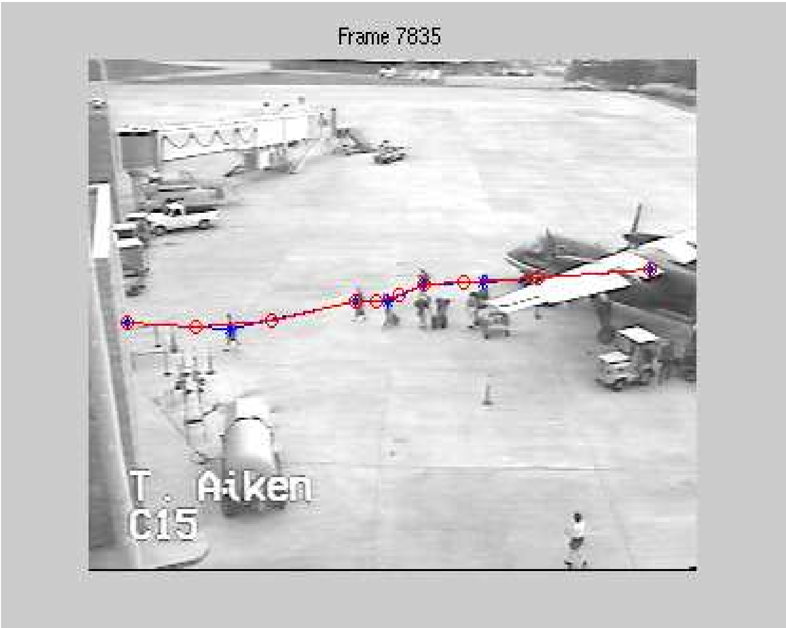
Landmark Shape

- **Shape: geometric information that remains when location, scale & rotation effects are filtered out [Kendall]**

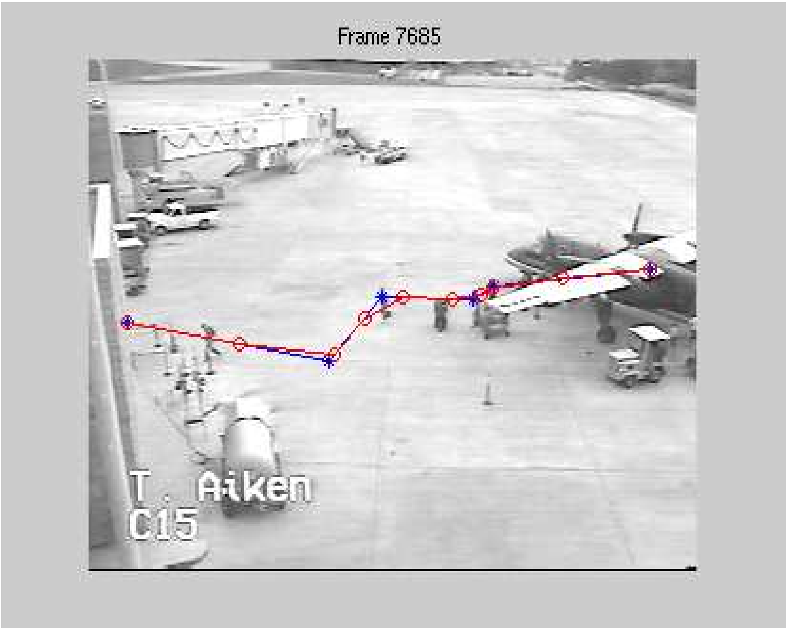
- **Shape of k landmarks in 2D**
  - Represent the X and Y coordinates of the k points as a k-dimensional complex vector: **Configuration**
  - Translation Normalization: **Centered Configuration**
  - Scale Normalization: **Pre-shape**
  - Rotation Normalization: **Shape**

- **Landmarks in 3D: represent by a $k \times 3$ matrix**

Example: Group of Passengers Deplaning



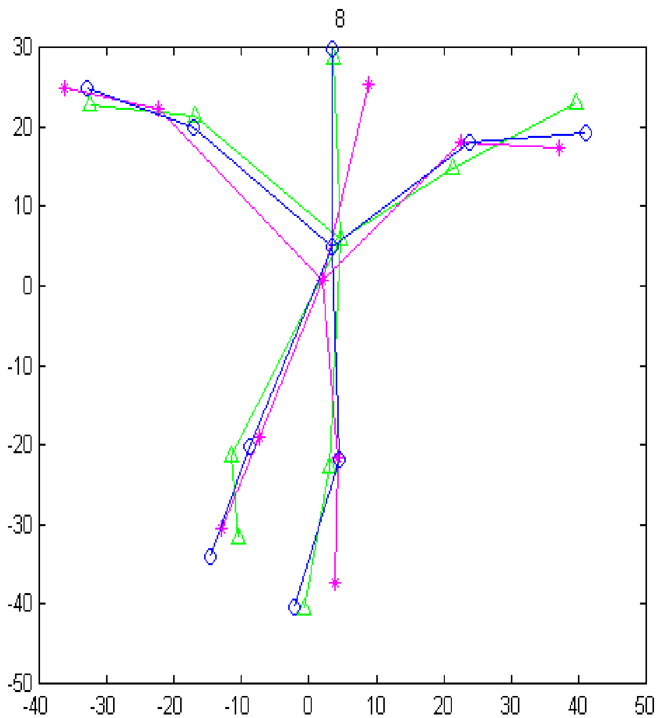
A 'normal activity' frame



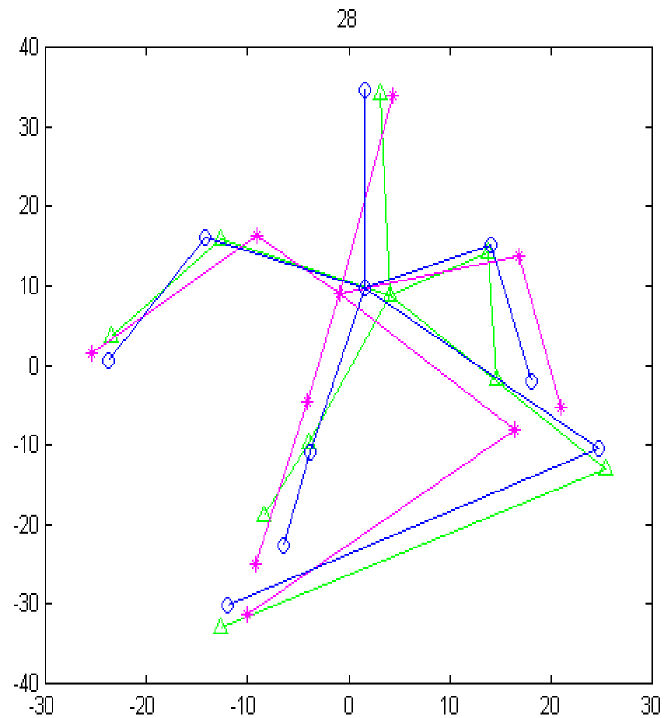
Abnormality

Example: Group of Robots

Human Actions

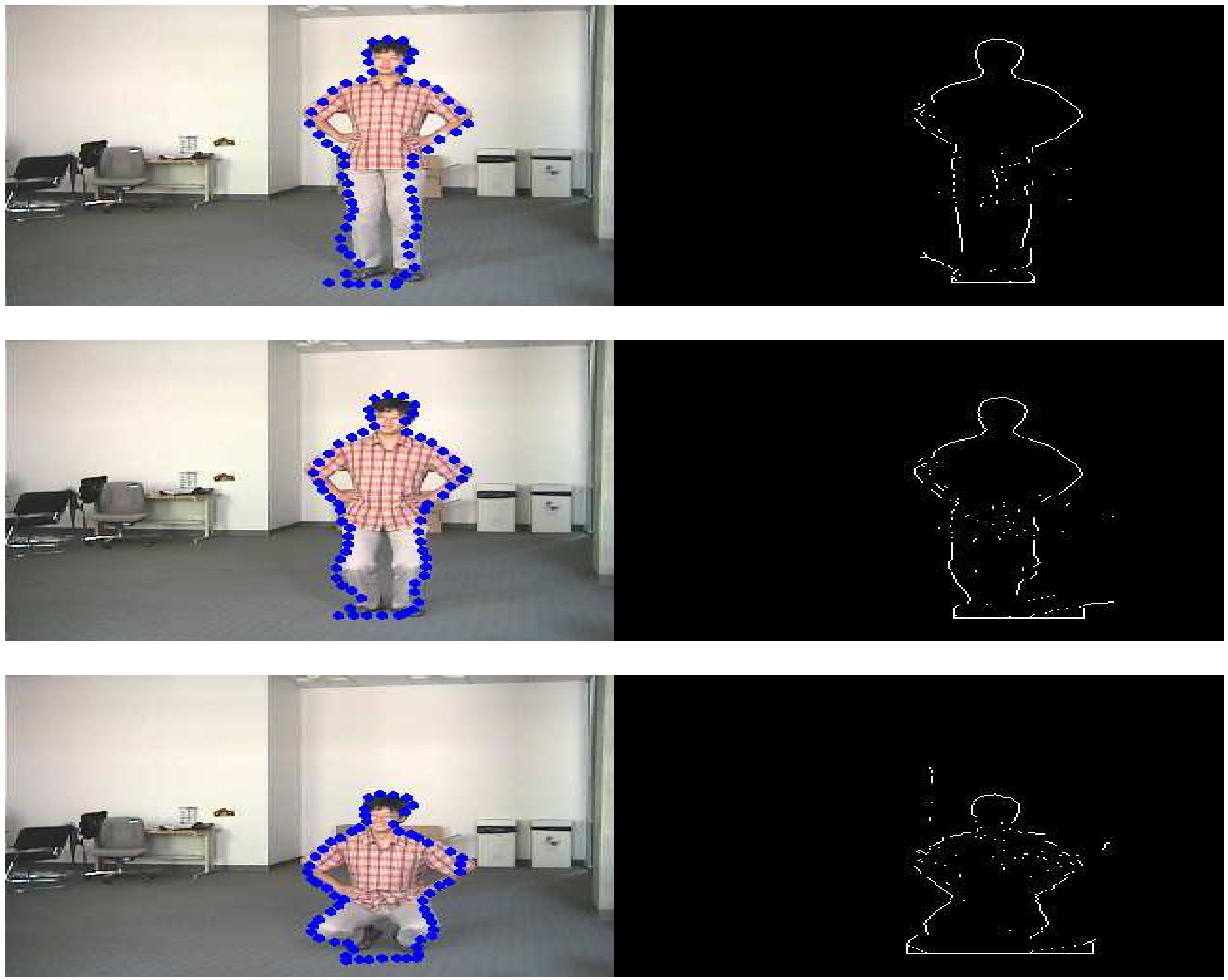


Normal action



Abnormality

Human Actions



Motivation

Make the tracking and recognition invariant to global scale, rotation, translation in image, may occur due to

- **Global scale change of activity**, e.g. person taller/shorter
- **Scaled orthographic camera motion**
  - Small field of view PTZ camera, far from scene, rotated to align with line of sight.
  - UAV looking straight down at activity
  - Activity center on camera's principal axis, no out of plane rotation
- **Track 2 shape activities occurring one behind the other**

A Common Framework for...

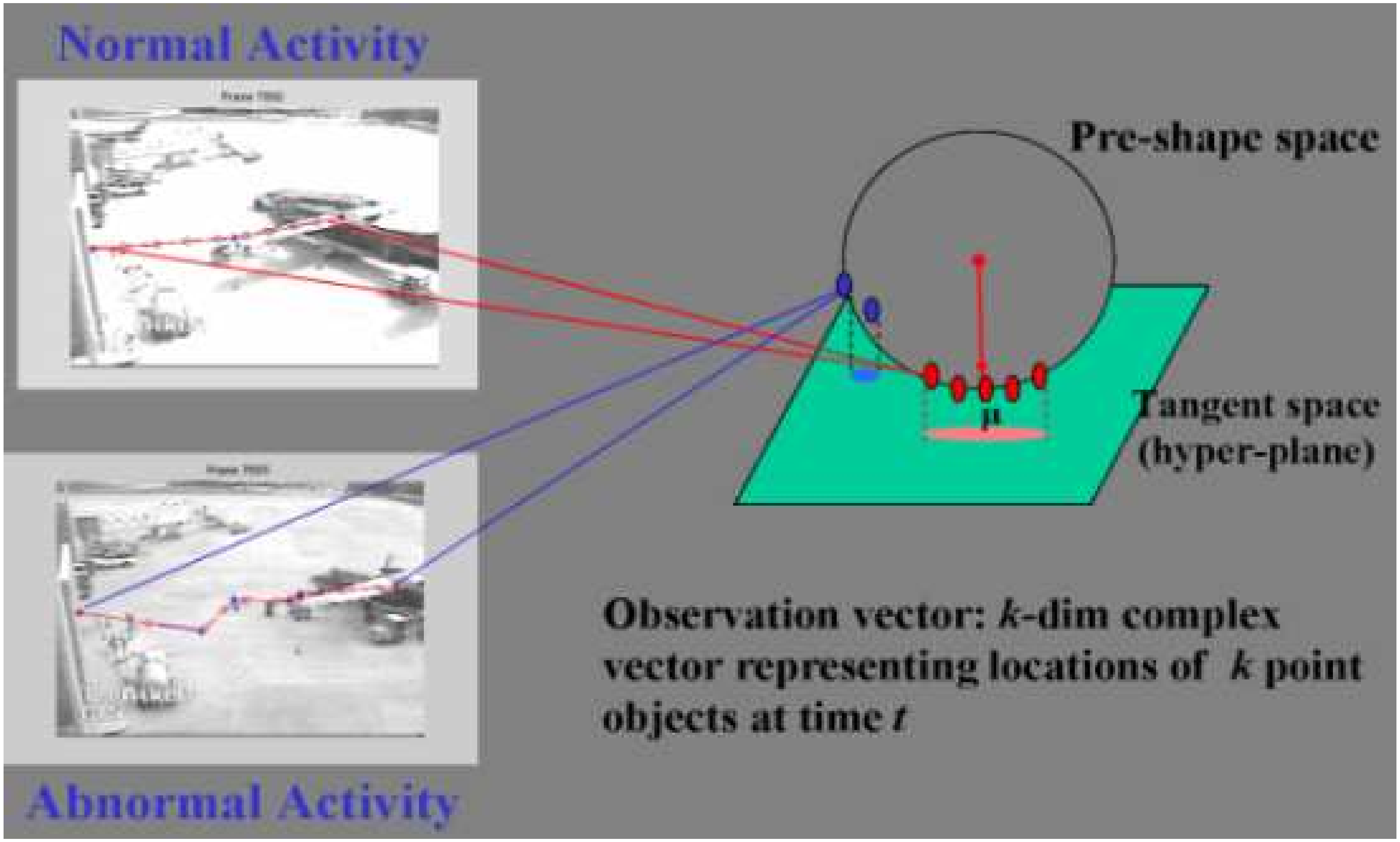
- **Tracking Groups of Moving/Interacting "Objects"**
  - Human action tracking: head,hand,torso - landmarks
  - Activities by groups of people or vehicles: low resol video
  - Biomedical applications: track "landmarks" of interest
- **Abnormal Activity Detection & Tracking**
  - Suspicious behavior detection, Lane change detection in traffic
  - Abnormal Human Action detection, e.g. motion disorders
- **Sequence Id & Tracking**
  - Sequence of human actions, track & summarize video

- **Activity Segmentation & Tracking**
  - Video coding + summarization:
    - Track 2D landmark shape, transmit only shape vector.
    - Detect scene change/abnormality, send more information when scene change
  - Unsupervised learning of activity models
- **Sensor independent approach**
  - Audio, infra-red or radar sensors, fuse different sensors
- **Robotics: robot formation tracking/control**

Existing Work

- **Joint tracking and event recognition**
  - DBN (or FS-HMM) tracked using a Rao-Blackwellized PF, e.g. Condensation for gesture tracking/recognition, figure tracking/recognition, traffic monitoring
  - Assume p.w. constant mode, sample from prior on mode, compute posterior, e.g. [Zhou et al]
- **Tracking groups of moving/interacting objects**, e.g. data association (JPDAF), Schulz et al, robot formation control, Condensation
- **Activity/Action Recognition**, e.g. space-time shapes, shape based factorization, view invariant approaches, multiple levels of zoom, DBN, co-occurrence statistics

Tracking Landmark Shapes


Normal Activity
Pre-shape space
Tangent space
(hyper-plane)
μ
Observation vector: k-dim complex vector representing locations of k point objects at time t
Abnormal Activity

Dynamical Model [Trans IP,Oct'05, CDC'05]

- **Observation:** Observed object locations, centroid subtracted (Centered Configuration) or the Edge Image
- **State:** [Shape, Scale, Rotation, Shape Velocity]

- **Observation Model:**
  Observation = h_t(Shape,Scale,Rotation) + observation noise

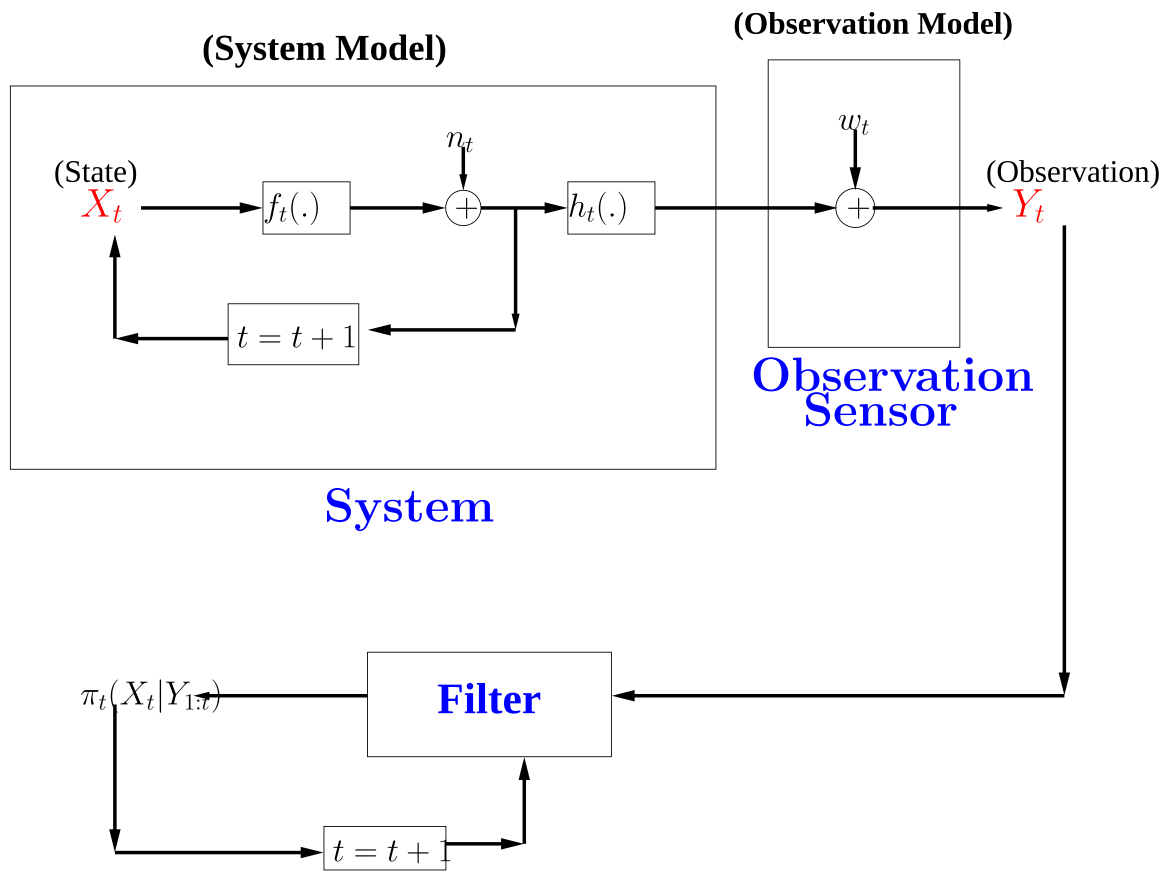
- **System model: Dynamics of shape, scale, rotation**
  - Shape "velocity": defined in tangent plane at current shape.
  - Move on current tangent plane by "velocity".
  - Project back to shape space: shape at next time.

- **Gauss-Markov model on shape "velocity", parallel transported to tangent space at current shape.**

- Appropriate model for global scale, rotation variation.

- **Track the observed landmark locations (or the edge image), to estimate posterior shape and shape velocity distribution, $\pi_t(x_t|Y_{1:t})$.**
  - **Use a particle filter (PF): computationally efficient & provably stable solution for nonlinear, multimodal, large dim state tracking.**
  - Other options: EKF, GSF, UKF, MHT, Grid-based, Quadrature, MCMC, Sequential IS.

Hidden state X_t=[Shape,velocity,scale,rotation], Y_t=Observed object locations or Image, Estimate posterior, $\pi_t(X_t|Y_{1:t})$



Particle Filter (PF) [Gordon et al'93]: Basic Idea

- **Sequential Monte Carlo method, approx. true filter as number of Monte Carlo samples ("particles"),** $N \to \infty$

- Given π_{t-1}^N, **perform importance sampling & weighting,** followed by **resampling** to approx. the Bayes' recursion to get π_t^N

$$\pi_{t-1}^N \xrightarrow[\text{Imp. Samp.}]{\tilde{x}_t^i \sim q_t} \pi_{t|t-1}^N \xrightarrow[\text{Weight}]{Y_t \;\; w_t^i \propto g_t(Y_t|\tilde{x}_t^i)} \bar{\pi}_t \xrightarrow[\text{Resample}]{} \pi_t^N$$

- Using $\gamma_t(x_t|x_{1:t-1}^{(i)}, Y_{1:t}) = q_t(x_t|x_{t-1}^{(i)})$ as importance density

Abnormal Activity Detection [TIP,Oct'05]

- **"Normal Activity": Modeled as a landmark shape dynamical model**
  - **Partially observed system(observations are noisy nonlinear functions of state), satisfying HMM property**

- **"Abnormality": Change w.r.t. learned shape dynamics**
  - **Parameters of changed system unknown**
  - **Change can be slow or sudden**

- **Detect changes in shape using the PF estimate of posterior of shape and/or shape velocity.**

Change Detection

Change Detection Problem

Abnormal activity detection provides the problem definition: **Given the observations $Y_1, Y_2, ...Y_t$, detect, as quickly as possible, if a change occurred in the dynamics of the state X_t**

- **Change parameters unknown**
  - **Cannot use CUSUM** (or its modifications [Azimi et al]).
  - **Generalized CUSUM** intractable [Andrieu et al'04].
  - **Residue statistics [Basseville] for fault detection**, e.g.
    - Tracking Error (TE) [Bar-Shalom]
    - negative log of Observation Likelihood (OL)
    - Score function [Basseville]
- **"Slow" or sudden change**
  - TE, OL, score fn detect sudden changes but miss slow changes

Slow and Sudden Changes

- **Slow change: small change magnitude per unit time, "tracked" by the tracker, i.e. error b/w estimate of posterior using the tracker with unchanged system model and the true posterior is small**

- **Sudden change: mostly "filtered out" ("loses track")**
  - Duration much smaller than "response time" of filter.

- **Quantify "rate of change", r, w.r.t. a filter: For an additive change with magnitude b per unit time, $r^2 = b^T \Sigma_{sys}^{-1} b$.**

Notation

$$
\begin{tikzcd}
X_{t-1} \arrow[r, "q_t"] \arrow[d, "g_{t-1}"] & X_t \arrow[d, "g_t"] \\
Y_{t-1} & Y_t
\end{tikzcd}
$$

- **Prior: Given no observations,** $X_t \sim p_t(.)$
- Posterior: $X_t|Y_{1:t} \sim \pi_t(.)$
- **Superscripts:** ^{0} **(unchanged system),** ^{c} **(changed system)**
- $X_t^0 \sim p_t^0(.), \qquad X_t^c \sim p_t^c(.)$

Slow change detection, Unknown parameters

- **Fully observed state (state= h_t^{-1}(observation):**
  - Use **negative Log Likelihood of state of unchanged system** to detect change, e.g. [Kulhavi,00]

$$-\log p_t^0(X_t) = -\log p_t^0(h_t^{-1}(Y_t))$$

    * Most commonly used when have a set of i.i.d. observations, compute average LL

- **Partially observed state (significant observation noise):**
  - Why not use Min. Mean Square Error estimate of this ?
  - **Our statistic is exactly this MMSE estimate:**

$$\mathbf{ELL}(\mathbf{Y_{1:t}}) \triangleq \mathbb{E}[-\log \mathbf{p_t^0}(\mathbf{X})|\mathbf{Y_{1:t}}]$$

Defining the Statistics [Vaswani, ACC'04]

- **Expected (negative) Log Likelihood of state (ELL)**

$$\mathbf{ELL}(\mathbf{Y_{1:t}}) = \mathbb{E}[-\log \mathbf{p_t^0}(\mathbf{X_t})|\mathbf{Y_{1:t}}] = \mathbb{E}_{\pi_\mathbf{t}}[-\log \mathbf{p_t^0}(\mathbf{X})]$$

- For sudden changes, can use
  - **Tracking Error (TE)** [Bar-Shalom] or its sum over τ past times

$$\mathbf{TE} = ||\mathbf{Y_t} - \hat{\mathbf{Y}}_\mathbf{t}||^2, \ \hat{\mathbf{Y}}_\mathbf{t} = \mathbb{E}[\mathbf{Y_t}|\mathbf{Y_{1:t-1}}]$$

  - **(negative) log of Observation Likelihood (OL)** or its sum

$$\mathbf{OL}(\mathbf{Y_{1:t}}) = -\log \mathbf{p_Y}(\mathbf{Y_t}|\mathbf{Y_{1:t-1}}) = -\log \mathbb{E}_{\pi_{\mathbf{t|t-1}}}[\mathbf{g_t}(\mathbf{Y_t}|\mathbf{X})]$$

  - OL $\approx$ TE (to first order) for white Gaussian observation noise

Computing ELL

- Consider **a linear and Gaussian system model**:

$$X_0 \sim \mathcal{N}(x; 0, \sigma_0^2), \quad X_t = AX_{t-1} + n_t, \quad n_t \sim \mathcal{N}(0, \sigma_n^2)$$

  - $A < 1$ **(stationary)**: $p_t^0(x) = \mathcal{N}(x; 0, \sigma_0^2)$

$$-\log p_t^0(X) = \frac{X^2}{2\sigma_0^2} + const$$

$$ELL(Y_{1:t}) = \frac{1}{N}\sum_{i=1}^{N}[-\log p_t^0(x_t^i)], \quad x_t^i \sim \pi_{t|t}(x)$$

  - $A = 1$ **(nonstationary)**: $p_t^0(x) = \mathcal{N}(x; 0, \sigma_0^2 + t\sigma_n^2)$
    - **Problem: variance of p_t^0 increases with t.**

- **If nonlinear, Gaussian system: linearize f_t to approx p_t^0.**
- **If training sequence available, learn a p.w. constant $p_t^0(x)$.**

- **Replace p_t^0 by Δ-step ahead prediction, $\pi^0_{t|t-\Delta}$.** Approx as:
  - Approx. PF estimate of $\pi^0_{t-\Delta|t-\Delta}$ by a Gaussian (or mixture).
  - Approx. $\pi^0_{t|t-\Delta}$ by applying linearized system model Δ times on each Gaussian mixture mean.
  - **Variance remains bounded & able to detect multiple changes.**

- **Other extensions:**
  - **Sum ELL over finite past: Modified CUSUM**
  - Large dim state: choose "classification" directions intelligently

Detection Thresholds [Vaswani, ACC'04]

- **ELL Threshold:** $\mathbf{Th_{ELL}} = \mathbb{E}_{\mathbf{Y^0_{1:t}}}[\mathbf{ELL^0}] + \mathbf{k}\sqrt{\mathbf{Var(ELL^0)}}$

$$\mathbb{E}_{\mathbf{Y^0_{1:t}}}[\mathbf{ELL^0}] = \mathbf{h(p^0_t)} = \mathbf{h(X^0_t)}$$

  $\mathbf{h}(.)$: entropy

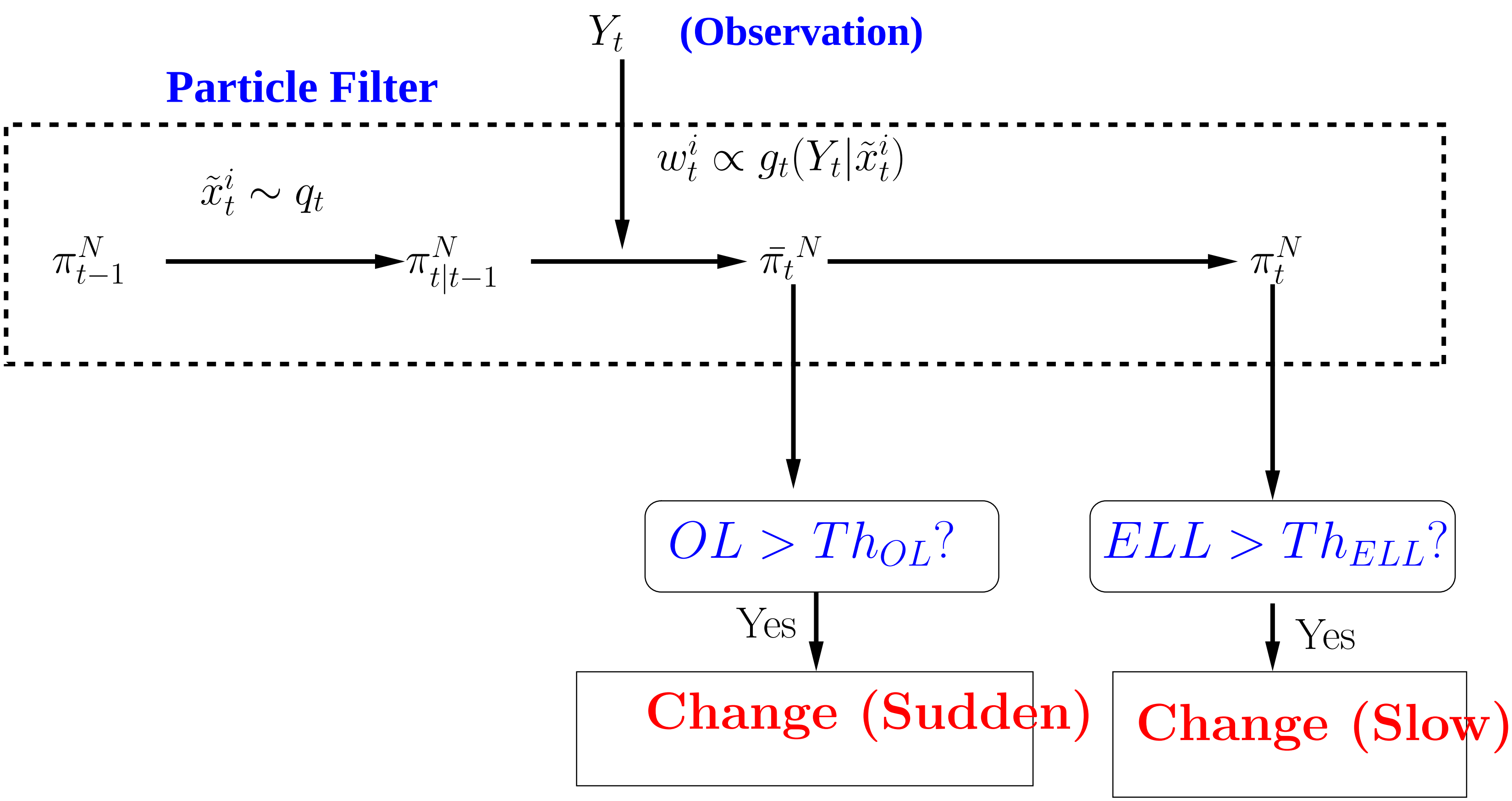
- **OL Threshold:** $\mathbf{Th_{OL}} = \mathbb{E}_{\mathbf{Y^0_{1:t}}}[\mathbf{OL^0}] + \mathbf{k}\sqrt{\mathbf{Var(OL^0)}}$

$$\mathbb{E}_{\mathbf{Y^0_{1:t}}}[\mathbf{OL^0}] = \mathbf{h(Y^0_t | Y^0_{1:t-1})}, \quad \text{compute empirically}$$

- Choose k based on allowed false alarm probability

- **Declare a change if either ELL or OL exceeds its threshold**

Change Detection Algorithm



ELL v/s OL (or TE)

- **OL & TE rely on loss of track to detect a change**
- **ELL detects based on "tracked part of the change"**
- **ELL detects change before loss of track: very useful**
- **Slow Change:**
  - PF: stable under mild assumptions, tracks slow change well
  - **Loss of track small: OL, TE fail or take longer**
  - Estimated posterior close to true posterior of changed system
  - **ELL detects as soon as change becomes "detectable"**
- **Sudden Change: PF loses track**
  - **OL & TE detect immediately, ELL fails/takes longer**

Summarizing [Vaswani, ACC'04, ICASSP'04,'05]

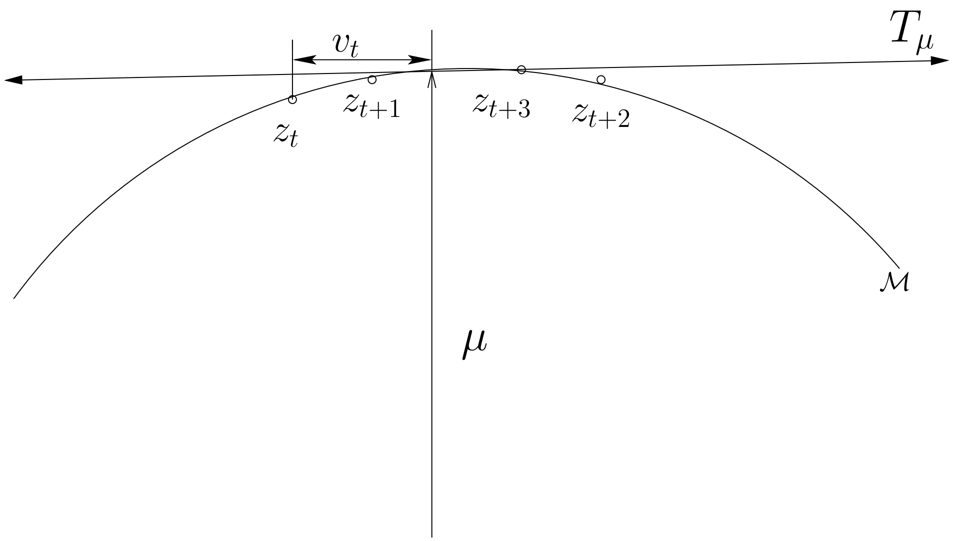
- **ELL detects a change before loss of track (very useful). OL or Tracking Error detect after partial loss of track.**
- **Proposed practical modifications of ELL and OL**
- Have shown:
  - **Error in ELL estimate upper bounded by increasing function of OL estimate: Complementariness**
  - **Stability of total ELL approx error for large N**
  - **Relation to Kerridge Inaccuracy and a sufficient condition for the class of detectable changes using ELL**

Applications

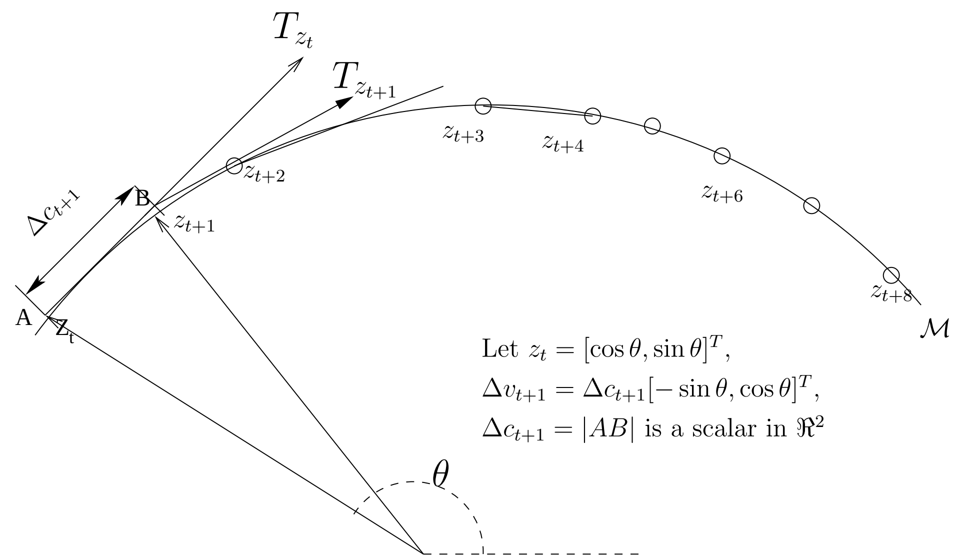
- Abnormal Activity Detection, Sequence Id, Segmentation
- **Any system model change detection w/o tracker losing track**
- **Change detection in bearings-only tracking**
- **Neural signal processing** (changes in STRFs of auditory neurons)
- **Acoustic tracking** (changes in target motion model)
- **Background model change detection**
- **Video shot segmentation**

Shape Activities

Landmark Shape Dynamics



Stationary Sequence



Non-Stationary Sequence

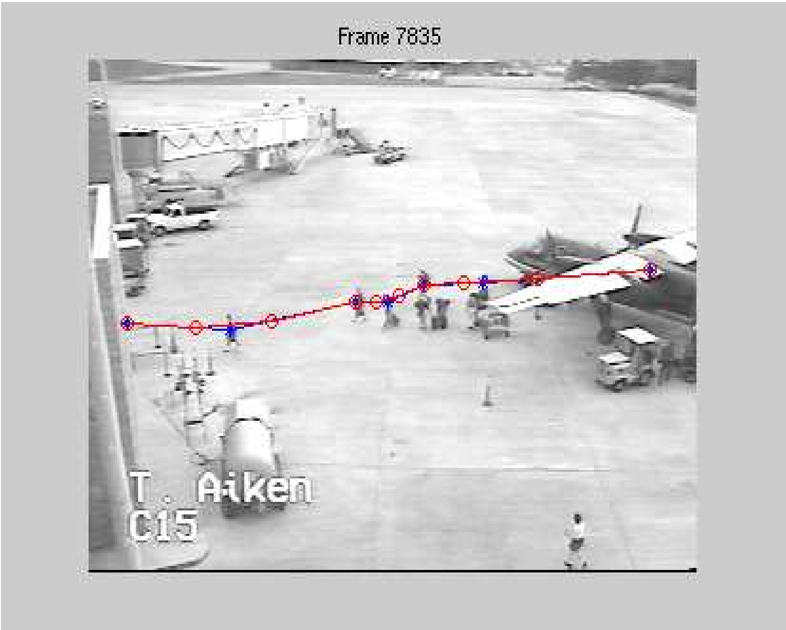
Different Shape Activity Models

- **Stationary SA (SSA) [TIP,Oct'05, CVPR'03]**
  - **Strict Sense SSA**: AR model on deviations about a "mean"
  - **Constant Mean SA (CMSA)**: random walk model on deviations about a "mean shape"
  - **Abnormal activity detection**

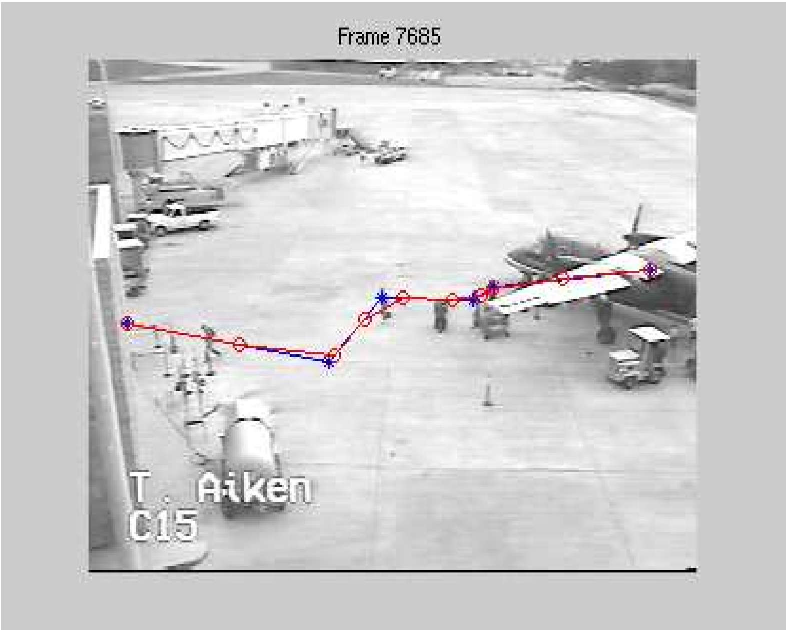
- **NonStationary SA (NSSA) [CDC'05]**
  - Mean shape different at every time
  - Markov model on shape velocity: "moves" current shape
  - **Track as well as detect abnormal activity**
  - **Activity segmentation** (use ELL w.r.t. $\pi^0_{t|t-\Delta}$)

- **Piecewise CMSA [CDC'05]**
  - Slow mean shape change: approx as piecewise constant
  - Sequence of CMSAs with nonstationary transition period
  - **Activity sequence identification** (use ELL to detect change, recognize new activity)

Group of People: Use SSA



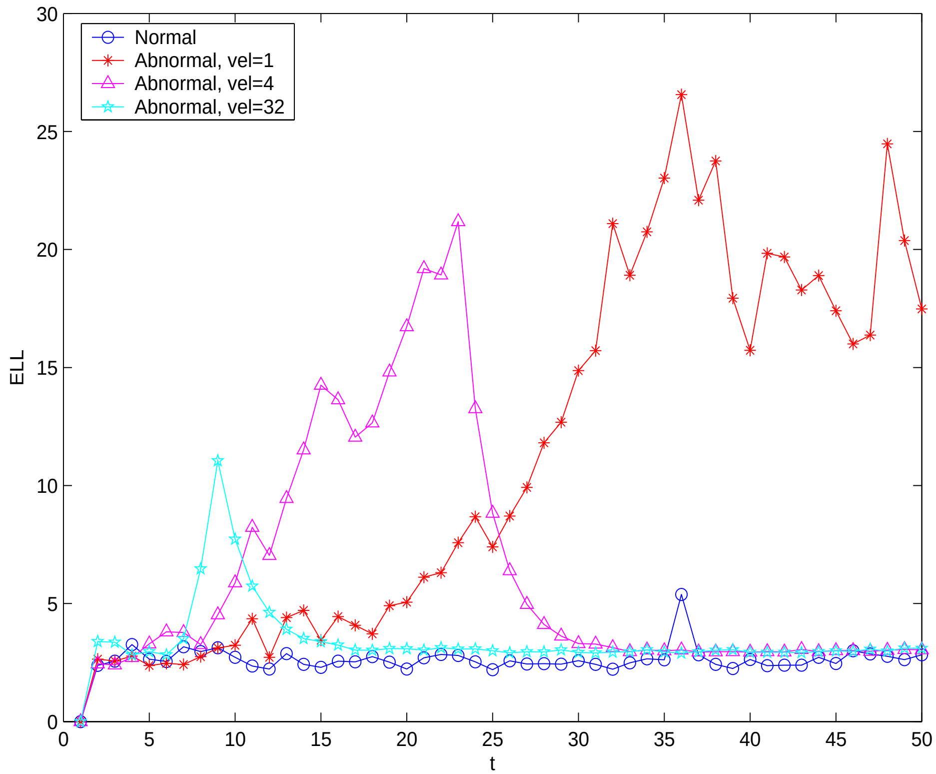
A ‘normal activity’ frame



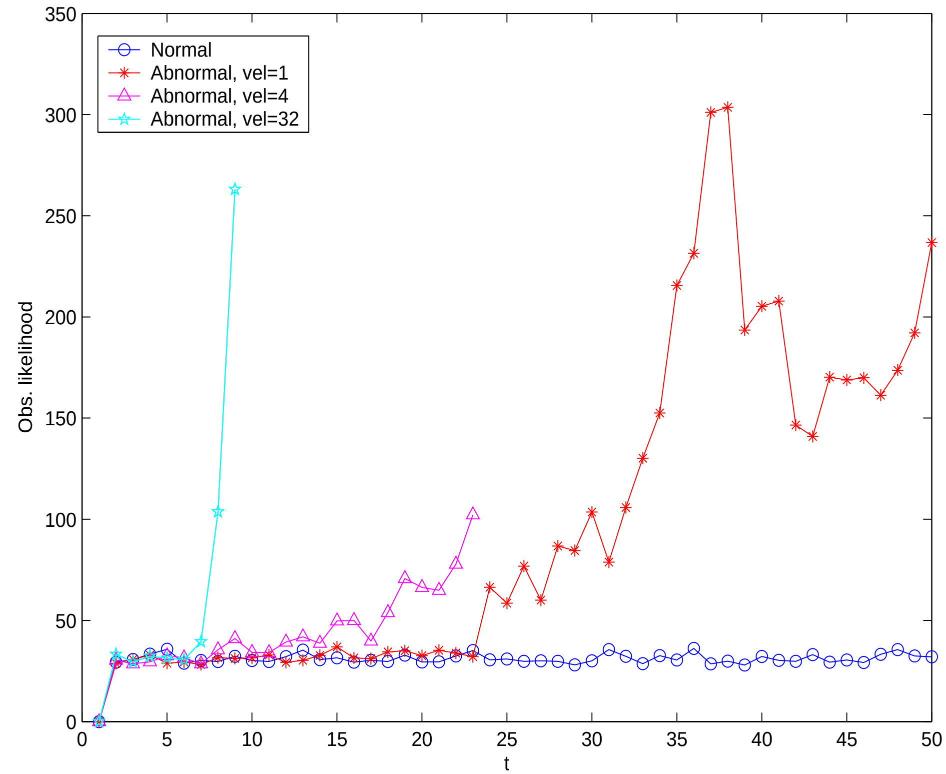
Abnormality

Group of People: Abnormality Detection Using SSA

Abnormality (one person walking away) begins at $t = 5$

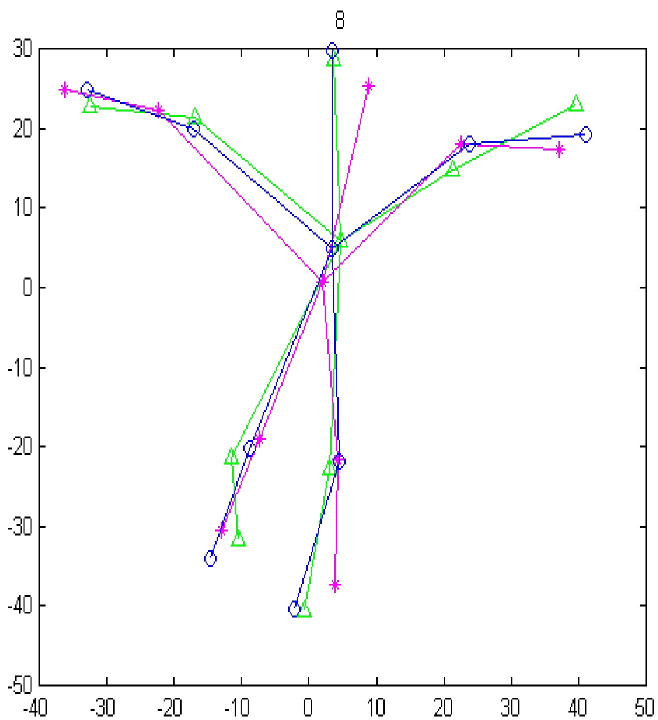


ELL

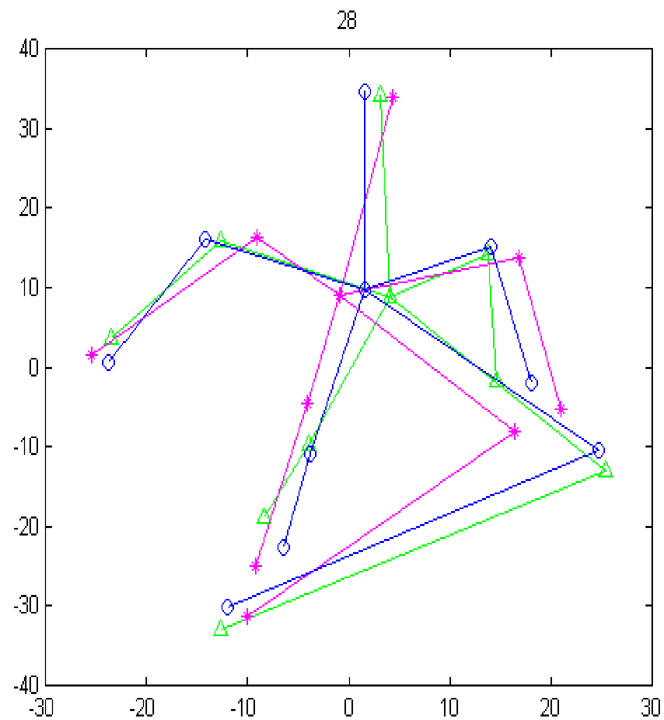


OL

Human Actions: Tracking Using NSSA



Normal action

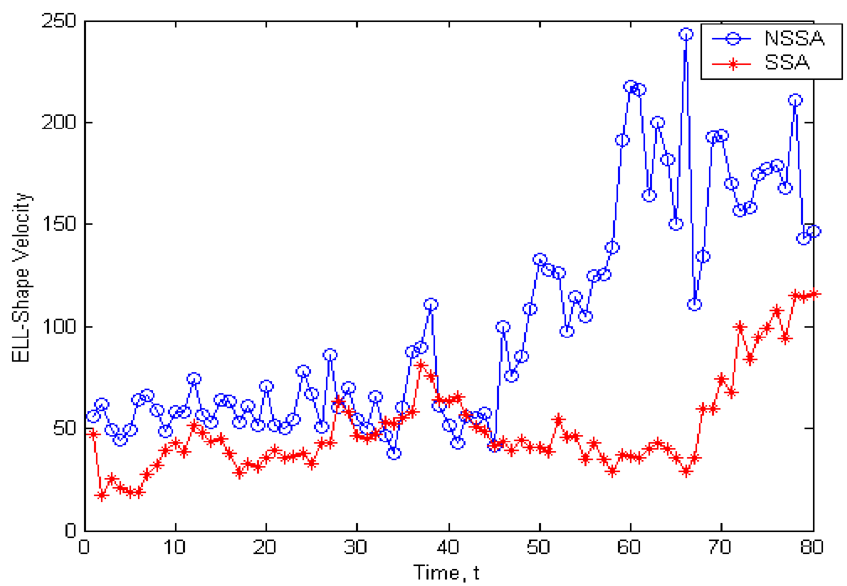


Abnormality Tracked

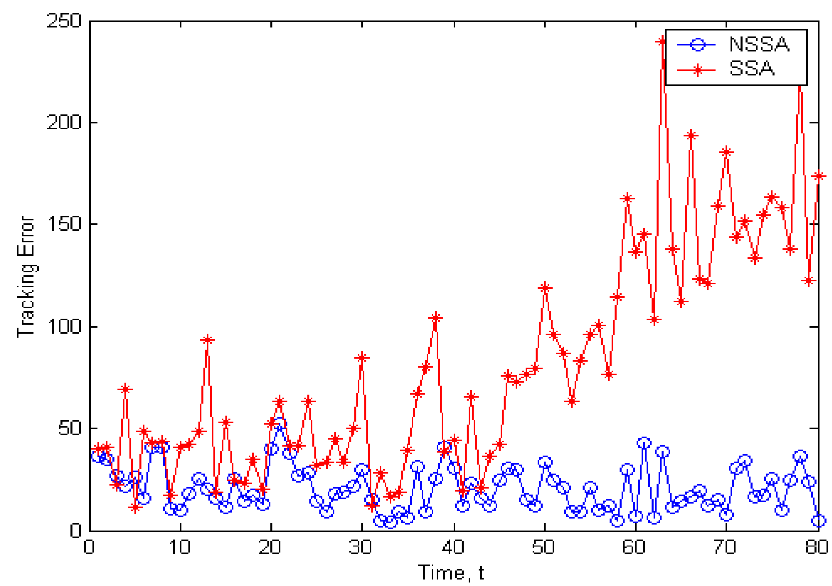
Green: Observation, Blue: Ground Truth, Magenta: Tracked

Human Actions: Abnormality Detection Using NSSA,SSA

- **Abnormality begins at $t = 20$, SSA only detects using TE**
- **NSSA detects using ELL and does not lose track**

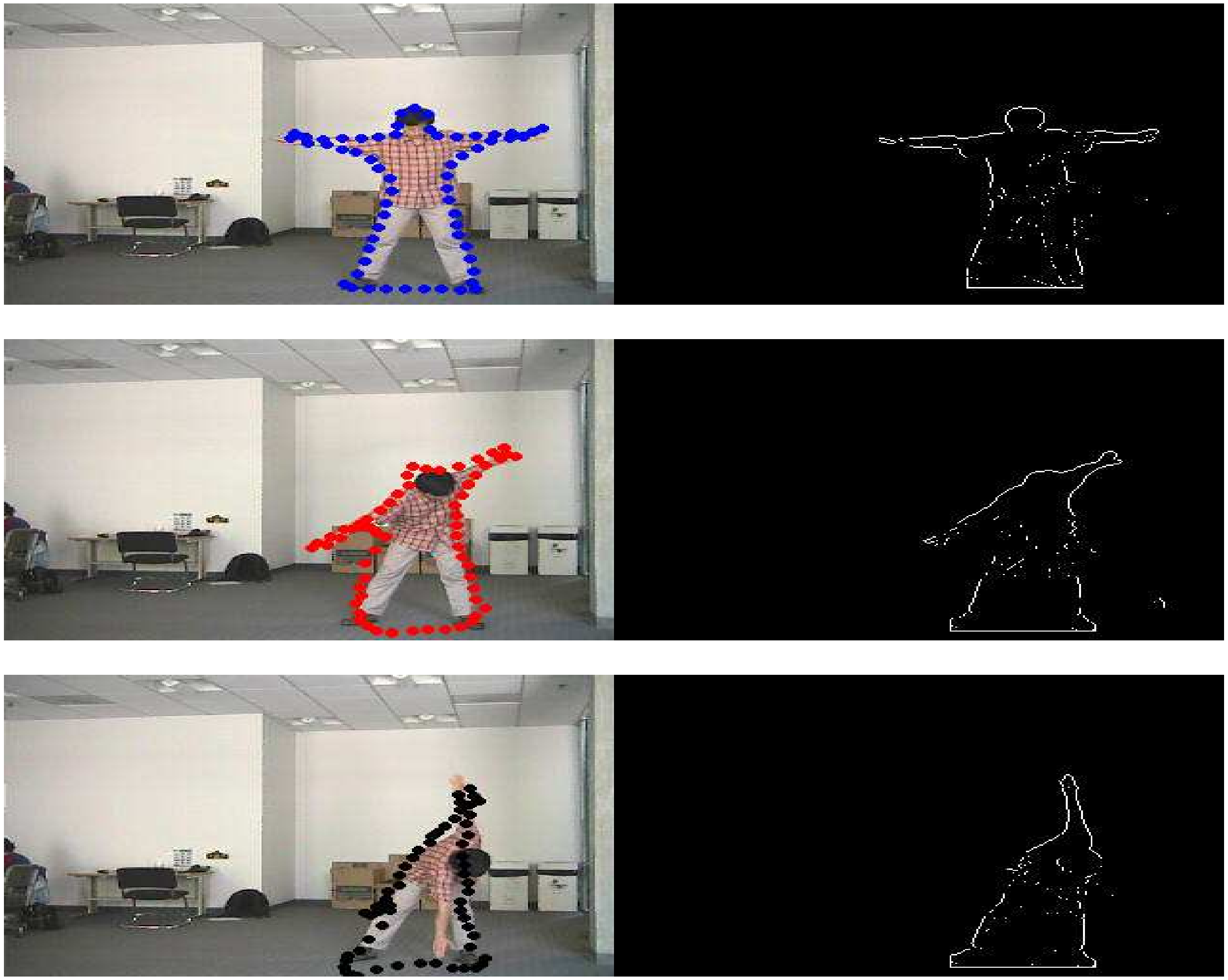


ELL

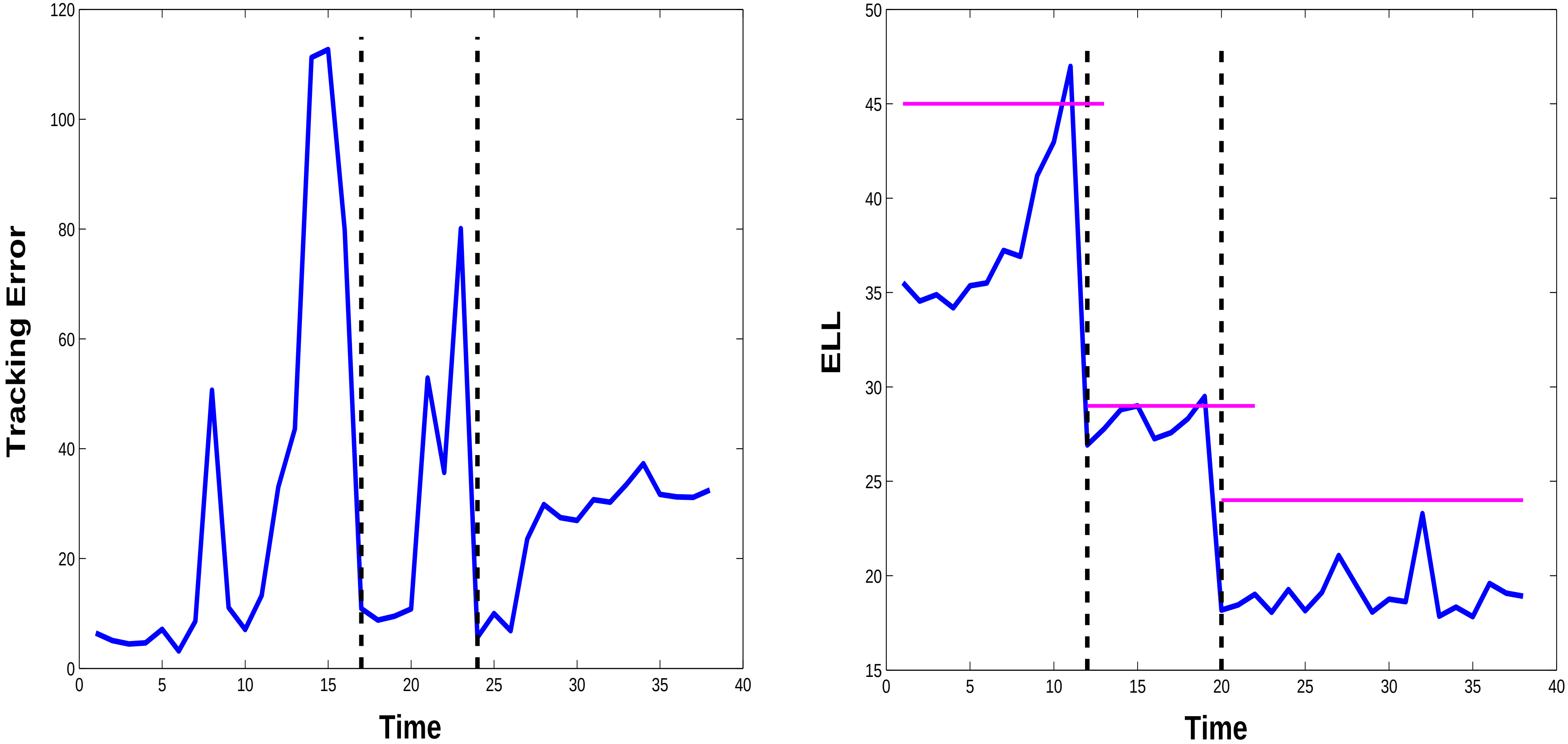


Tracking Error

Human Actions: Tracking a Sequence Using PCMSA



Human Actions: Sequence Identification Using PCMSA



A Common Framework for...

- **Tracking Groups of Moving/Interacting "Objects"**
  - Human action tracking: head,hand,torso - landmarks
  - Activities by groups of people or vehicles: low resol video
  - Biomedical applications: track "landmarks" of interest

- **Abnormal Activity Detection & Tracking**
  - Suspicious behavior detection, Lane change detection in traffic
  - Abnormal Human Action detection, e.g. motion disorders

- **Sequence Id & Tracking**
  - Sequence of human actions, track & summarize video

- **Activity Segmentation & Tracking**
  - Video coding + summarization:
    - Track 2D landmark shape, transmit only shape vector.
    - Detect scene change/abnormality, send more information when scene change
  - Unsupervised learning of activity models
- **Sensor independent approach**
  - Audio, infra-red or radar sensors, fuse different sensors
- **Robotics: robot formation tracking/control**

Summary

- SSA + Abnormality detection (ELL + OL): airport sequence
- NSSA + Abnormality detection (ELL): skater sequence
- PCMSA + Sequence Id (ELL + TE): sequence of human actions
- **Ongoing, Future Directions**
  - **Measurement method: Obtaining landmarks**
  - **NSSA + Activity Segmentation**
  - **PTZ camera control to "follow" activity**
  - **Multiple simultaneous activities**
  - **Learning activity sequence dynamics (DBN)**
  - **3D SA, 2D affine SA, Time varying no of landmarks**
  - **Disease progression models, detect abnormality**

Contour Tracking

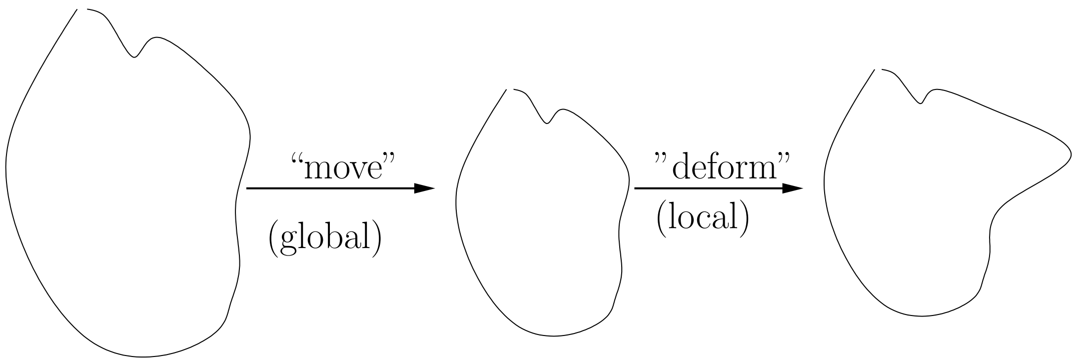
The Problem

- **Track a moving & deforming object from an image sequence**
- **State, X_t:**
  - **Object Contour** (C_t),
    - Represented using the level set method (infinite dim).
    - In practice: large but finite dim: M_t dim at t.
  - **Affine velocity** $(v_{t,s})$, **Deformation velocity** $(v_{t,r})$
- **Observation, Y_t:**
  - **Image (noisy nonlinear function of contour)**

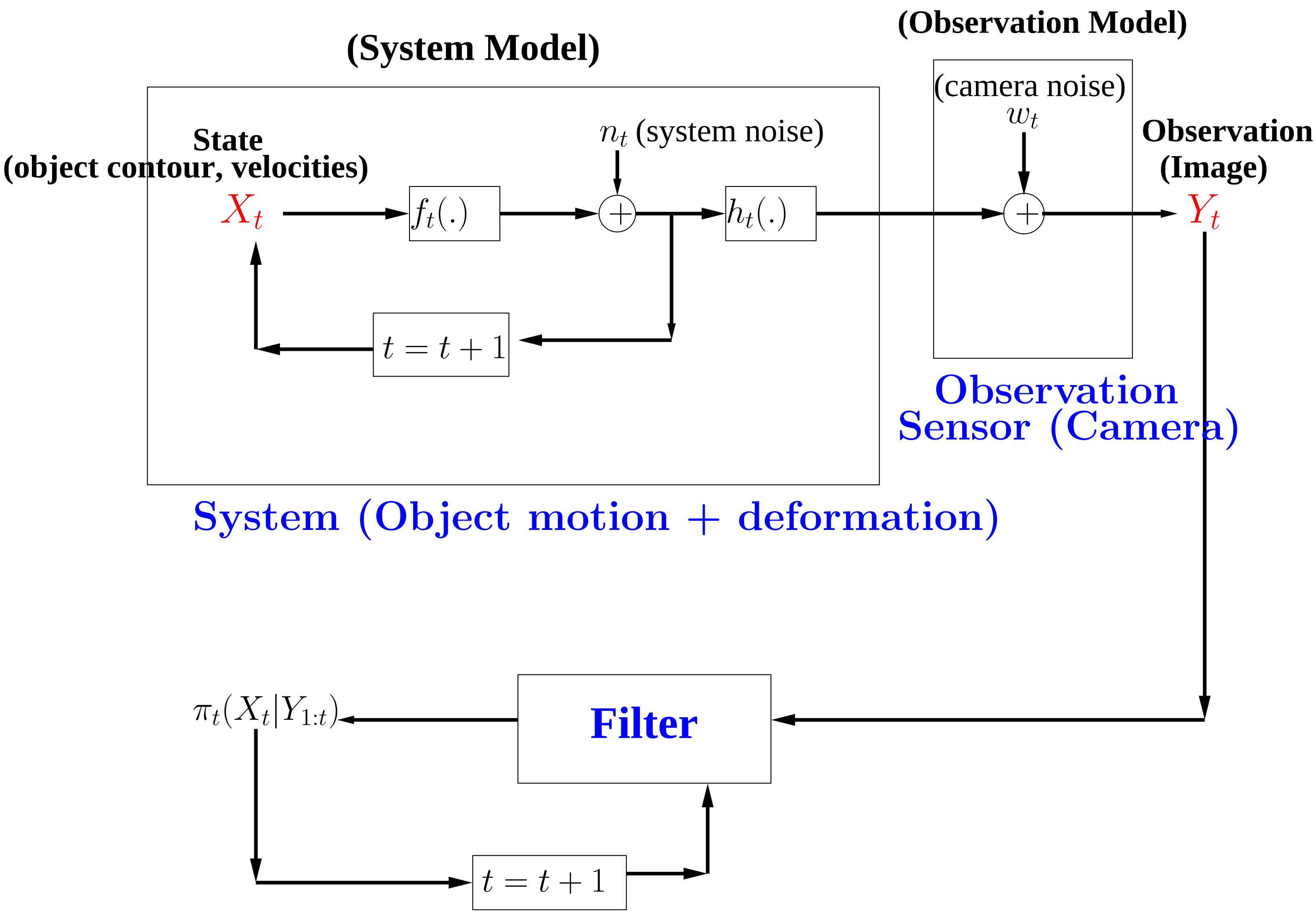
- **Goal: Estimate the posterior of contour & of velocities, given all past images**

Motion and Deformation [Yezzi,Soatto'02]

- **"Motion"**: global motion, a finite dimensional group e.g. Affine
- **"Deformation"**: local shape deformations, infinite dimensional



- **Examples:**
  - **Fish** can move in space and also deform its shape
  - **Human heart** deforms, **Human hand** moves & deforms
  - **Frequent viewpoint changes, partial occlusions**


(System Model)
(Observation Model)
State
(object contour, velocities)
X_t
$f_t(.)$
n_t (system noise)
$h_t(.)$
(camera noise)
w_t
Observation
(Image)
Y_t
$t = t + 1$
Observation
Sensor (Camera)
System (Object motion + deformation)
$\pi_t(X_t|Y_{1:t})$
Filter
$t = t + 1$

Existing Work

- **Finite dim contour - Condensation, deformable snakes,...**
  - Cannot handle large changes in contour length, topology
- **Infinite dim repr: [Brockett,..] [Neithammer,..] [Jackson,..]**
  - Defined approximate linear observers: require
    * Observed contour as the observation
    * $p(X_t|X_{t-1}, Y_t)$ unimodal: may not hold if E_{obs} non-convex
    * Uncoupled observers for contour and velocity
  - We address these by using a particle filter (nonlinear observer)
- **Particle filtering for large dim state spaces: expensive**
  - Sample N times from a large dim noise distrib. at each t
  - N required for accurate PF increases with noise dim

Figure 1: Frequent viewpoint changes + Unreliable observations (images)

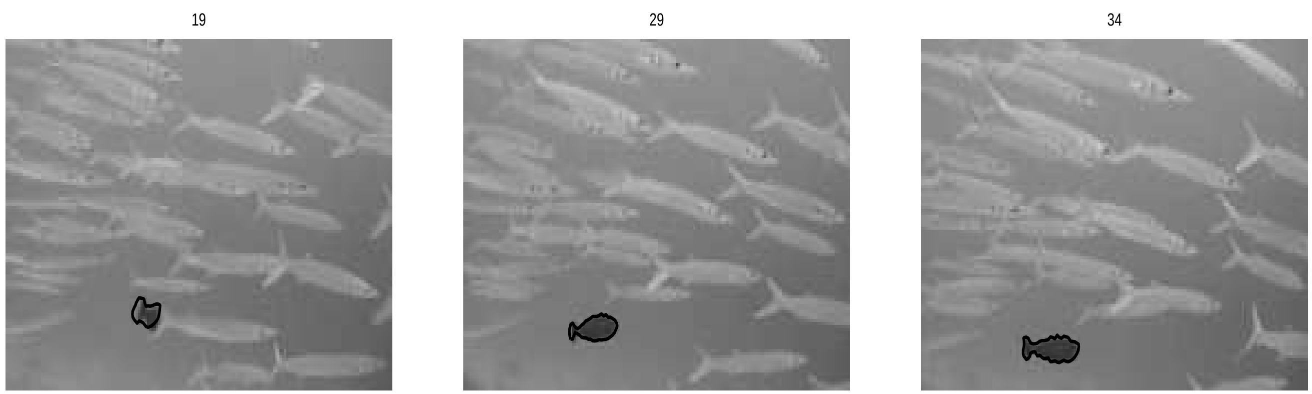


Figure 2: Multiple objects + Deforming objects, partial occlusions.

Figure 3: Partial occlusion due to street light. 3 possible contours.

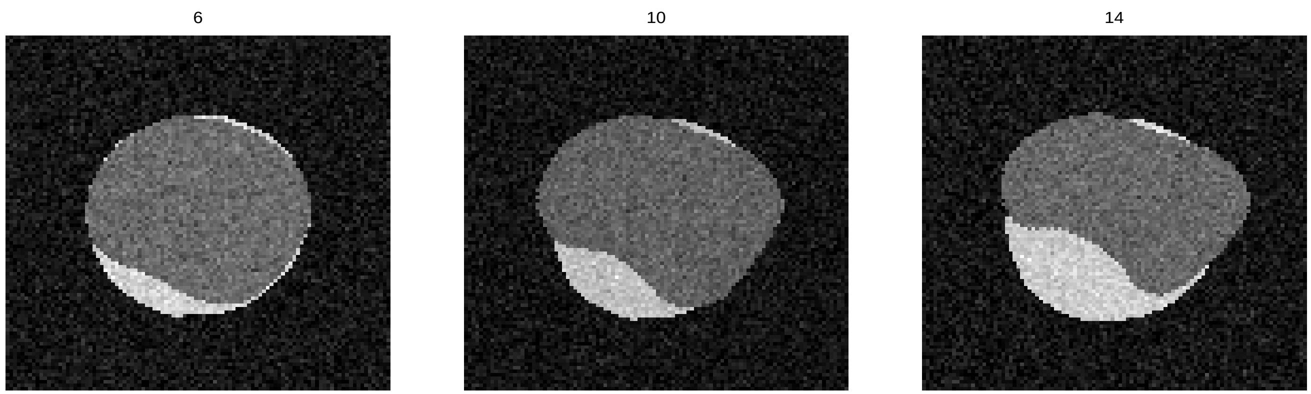


Figure 4: Partial occlusion. Track top (grey) object

Our Solution [CVPR'05, Rathi,Vaswani,Yezzi,Tannenbaum]

- For large dim state spaces, $p^* \stackrel{\triangle}{=} p(X_t|X_{t-1}, Y_t)$ is usually not unimodal. But it is fair to assume that conditioned on a small part of the state space (affine velocity, $v_{t,s}$), the rest of the local deformation is small or that $p(X_t|X_{t-1}, Y_t, v_{t,s})$ is unimodal.

- Based on this assumption, we propose an approx Rao-Blackwellized PF (RBPF): run a regular PF for affine velocity, replace the Kalman filter of RBPF by an approx linear observer for the local deformation. Approx linear observer implemented by running "some" iterations of gradient descent to minimize E_{obs}.