Deformable Contour Tracking & System Identification

Namrata Vaswani
Dept. of Electrical & Computer Engineering
Iowa State University
http://www.ece.iastate.edu/~namrata

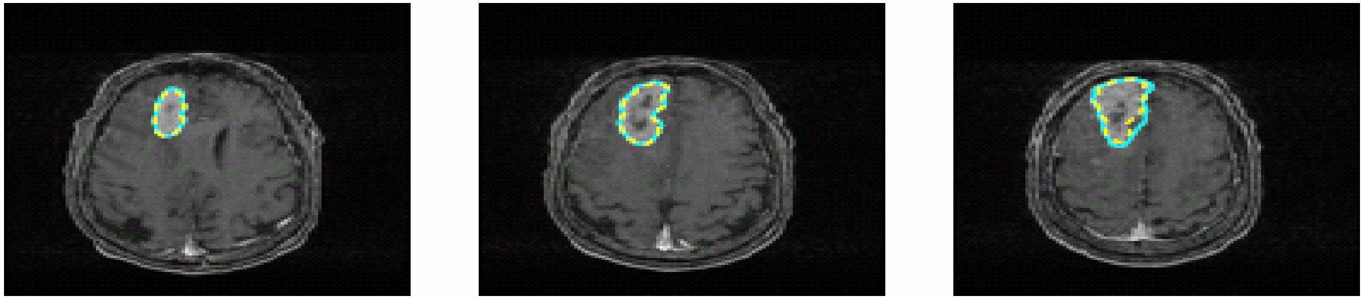
Iowa State University
Becoming the best.

Goals

- Sequentially segment deforming objects or Regions of Interest (ROIs) from video or from spatial image sequences
  - Do this offline ("smoothing") or online ("tracking")
  - O/P of a tracker is input to a smoothing algorithm

- Given a sequence of contours, estimate effective dimension of deformation & sequence dynamics
  - "System Identification"

Brain MRI slices: Tumor sequence (actual deformations)



Partial occlusion of car by street light

Perspective effect: Plane tracked by UAV (frequent camera viewpoint changes)

Outline

- The Tracking Problem
  - Main Issues, Particle Filtering
  - Existing Work & Our Key Ideas
  - Proposed Solutions: Affine & Deform PF-MT
  - Extension to Smoothing (offline segmentation)
- System Identification: Spatial PSD
- Summary & Open Issues

The Tracking Problem

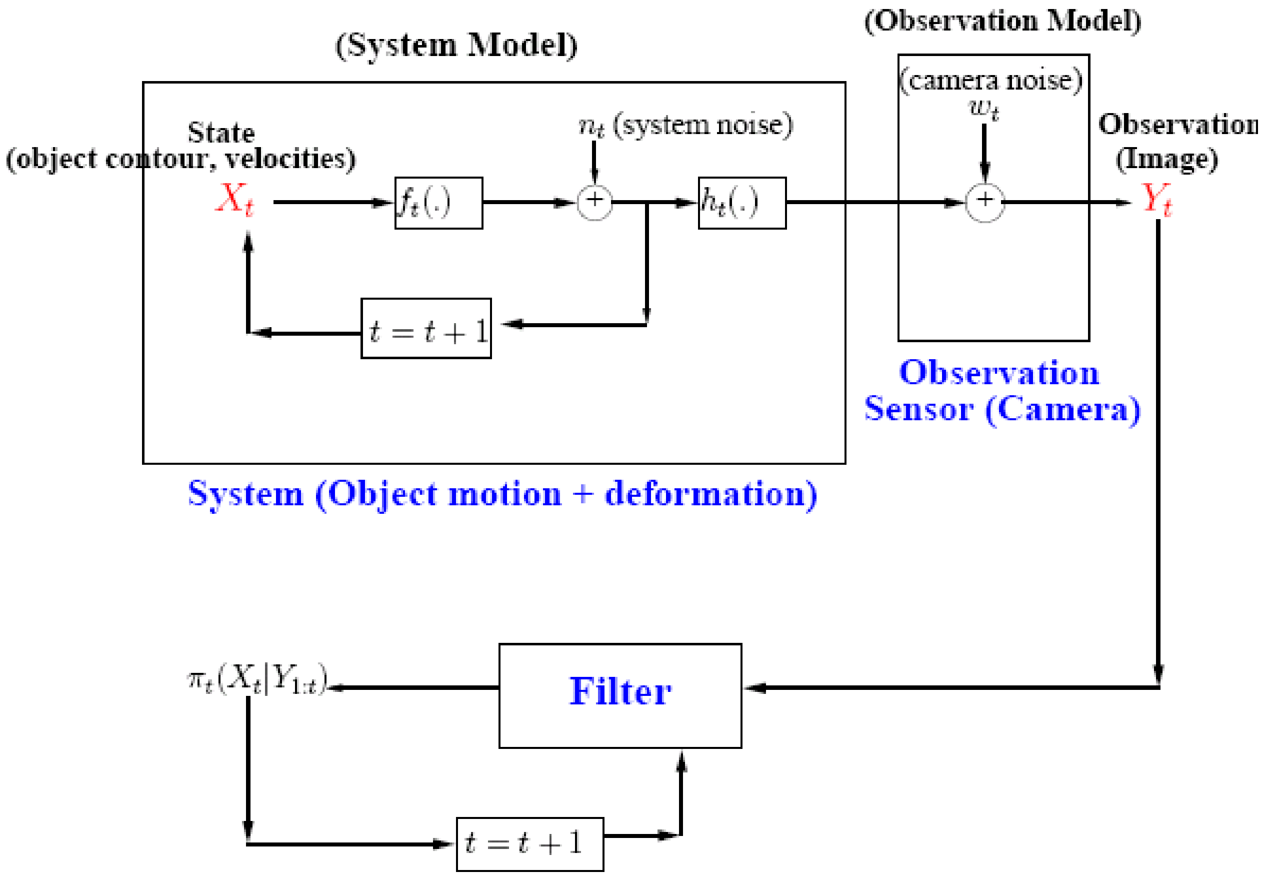
- Causally segment a moving & deforming object from a sequence of images
- Formulate as a "tracking problem": estimate the state at time t from observations until t when
  - States (contour, contour velocity): Markov model
    - State transition prior (STP) known: $p(X_t|X_{t-1})$
  - Observation (image or edge map) at t depends only on state at t
    - Observation Likelihood (OL) known: $p(Y_t|X_t)$

State Space Model

- System Model (State Dynamics). State, $X_t=[C_t,v_t]$
  - Contour = Previous contour + velocity
  - Gauss-Markov model on contour velocity
  - Contour velocity: global (affine) + local deformation

- Observation Model. Observation, Y_t
  - Image, Y_t = noisy & nonlinear function of contour, C_t
  - OL: $p(Y_t|X_t) = p(Y_t|C_t) \propto \exp[-E(Y_t,C_t)/\sigma^2]$
    - E = any segmentation energy functional
    - e.g. E = Chan-Vese, edge energy, or sum of both

Tracking Framework



Definitions

- State transition prior (STP): $p(X_t \mid X_{t-1})$
  - PDF of X_t conditioned on a value of X_{t-1}
- Observation Likelihood (OL): $p(Y_t \mid X_t)$
  - Probability of Y_t taking a certain value conditioned on a value of X_t
  - "OL multimodal": OL has multiple local maxima (modes) as a function of X_t

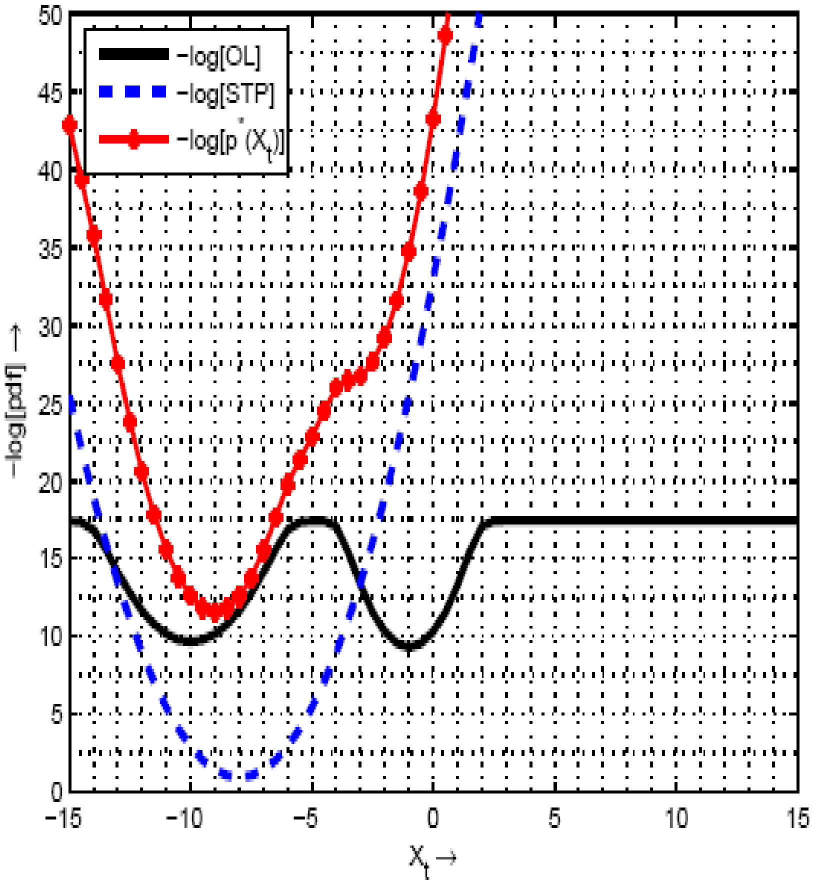
Main Issues, Particle Filtering

Main Issues

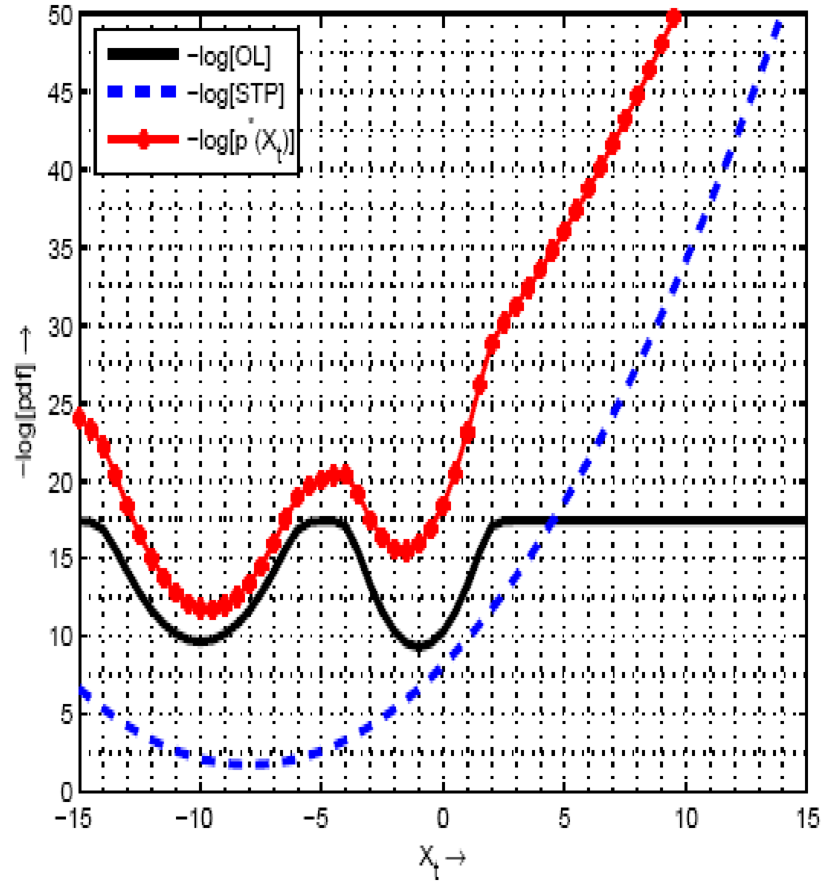
- Observation Likelihood (OL) is often multimodal
  - e.g. clutter, occlusions, low contrast images
  - If STP narrow enough, posterior is unimodal: adapt KF
  - If STP broad (fast deforming sequence): require a Particle Filter (PF)

- Deforming contours: Large dim state space (LDSS)
  - If constrained motion, e.g. rigid/affine: easy to use PF
  - LDSS: PF expensive (requires impractically large N)

Narrow STP: Unimodal posterior



Broad STP: Multimodal posterior



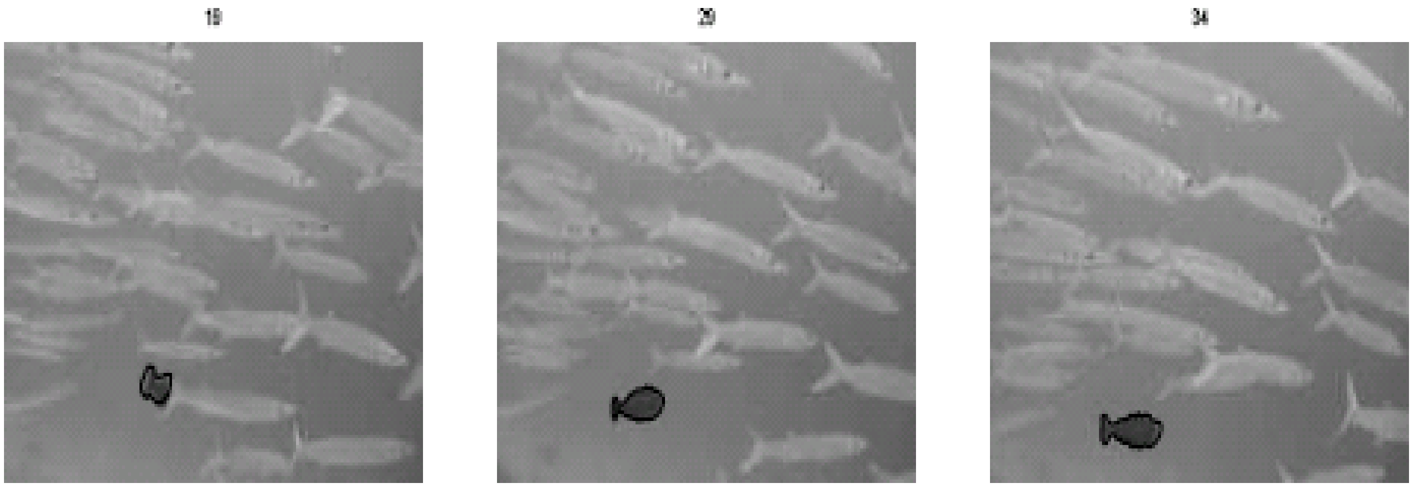
Examples: Multimodal OL

- As a function of affine deformation
  - Background clutter due to separate objects
  - Background clutter due to concentric contours
- As a function of non-affine deformation
  - Overlapping background clutter
  - Partial occlusions
  - Low contrast (weak edges: multiple edge responses)
  - Outliers

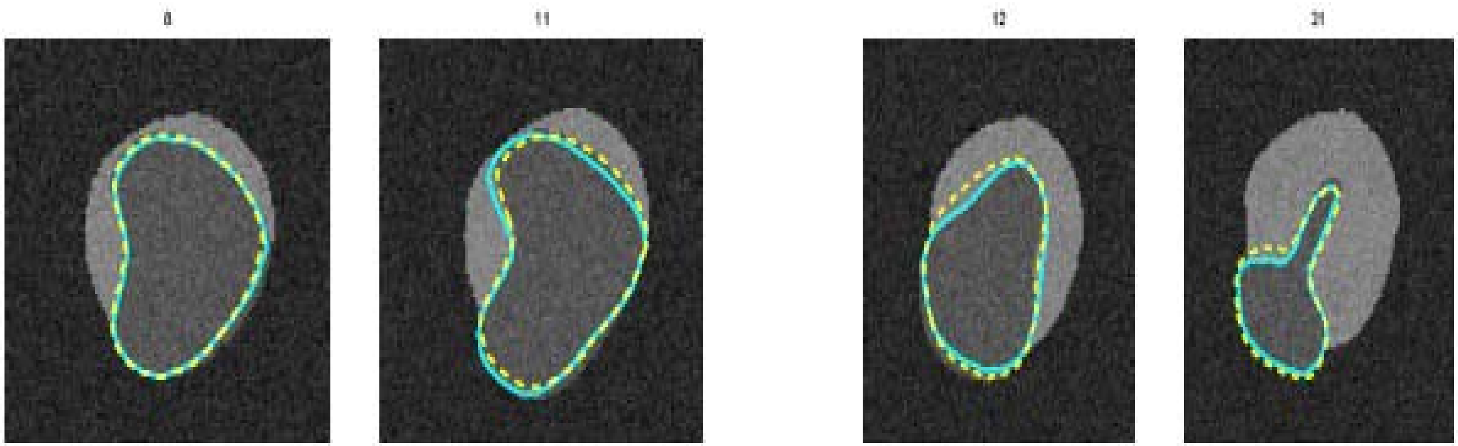
Examples: Deforming contours

- Actual deformations: biological images
  - Human tracking: surveillance, sports videos,...
  - Animals such as a fish
  - Medical sequences: ROIs in brain or heart
- Changing region of partial occlusions
  - Automatic vehicle navigation
  - Robot navigation
- Frequently changing camera viewpoint
  - Tracking using a UAV

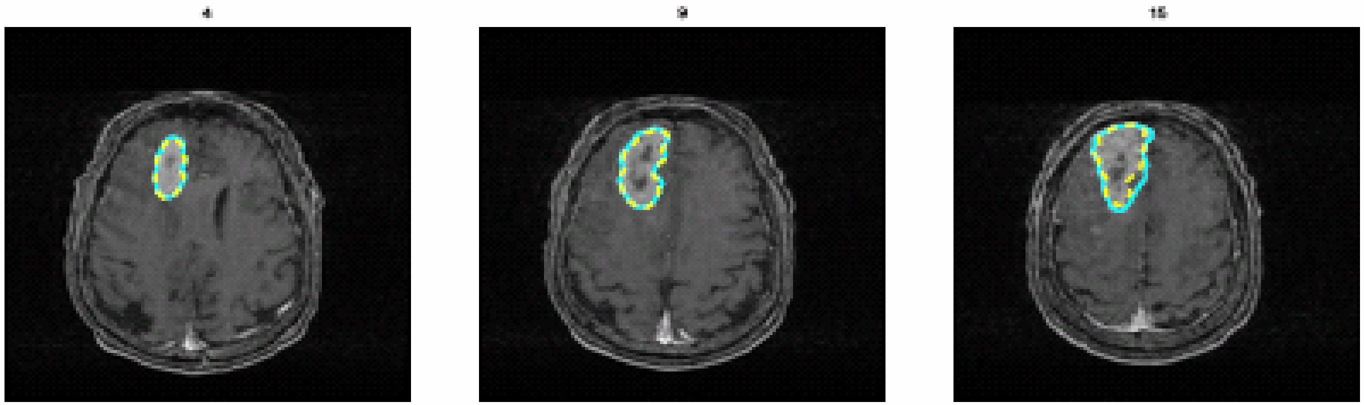
Separate clutter (multiple fishes) + deformation



Overlapping clutter (light grey object) + deformation

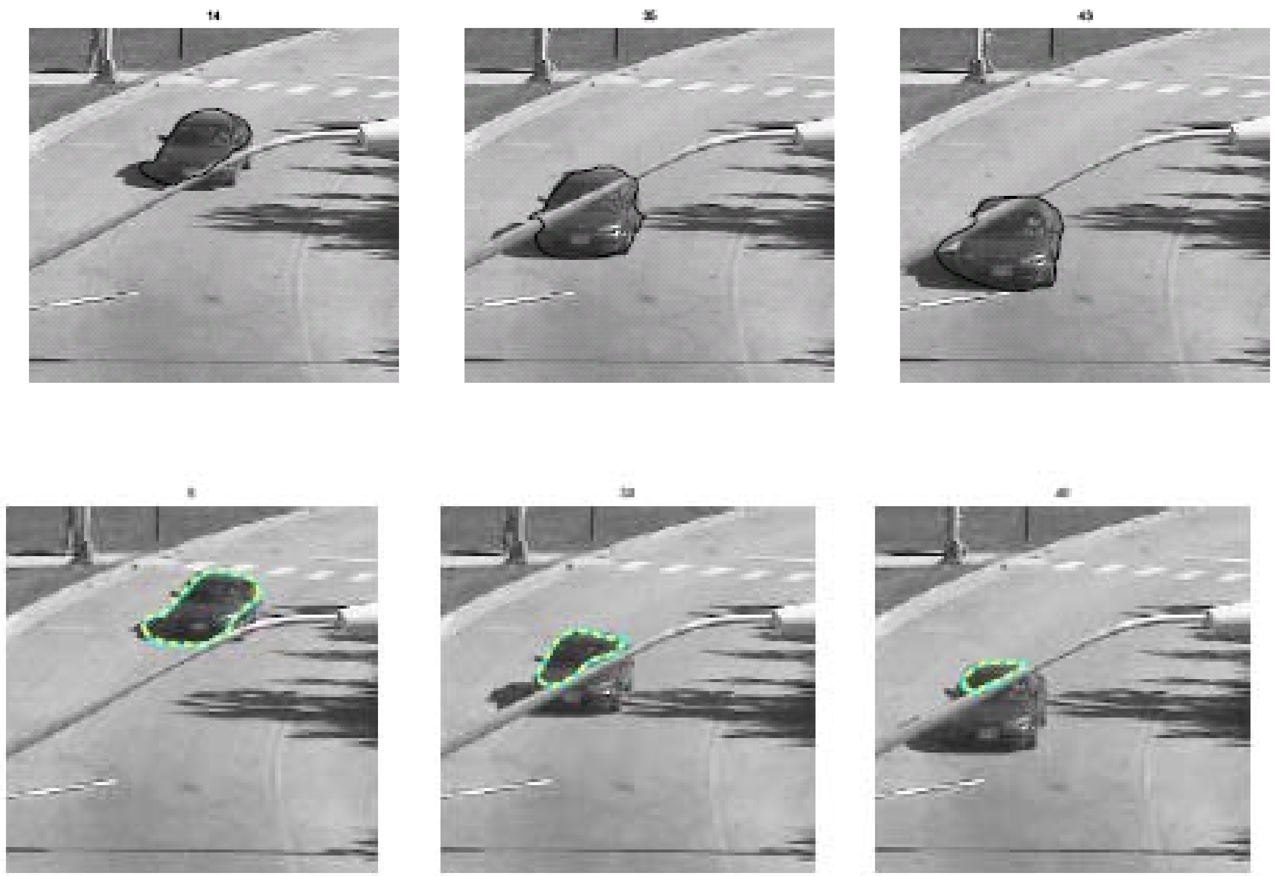


Low contrast + deforming contours (large deformation per frame)



Low contrast + Frequent viewpoint changes (small deformation per frame)

Partial occlusion of car by street light:
3 contour modes, 2 are deforming contours



Other LDSS problems

- Image Sequences
  - Spatially varying illumination change of moving objects
  - Optical flow (motion of each pixel)
- Sensor Networks
  - Spatially varying physical quantities, e.g. temperature
  - Boundary of a chemical spill or target emissions
- Time-varying system transfer functions
  - Time-varying STRF: repr. for neuronal transfer functions
  - Time varying AR model for speech (e.g STV-PARCOR)

Particle Filter (PF) [GSS'93]

- A sequential Monte Carlo technique to approx Bayes' recursion for computing the posterior $\pi_t(X_{1:t}) = p(X_{1:t}|Y_{1:t})$
- Does this sequentially at each t using Sequential Importance Sampling along with a Resampling step (to throw away particles with very small importance weights)

Monte Carlo, Importance Sampling

- Goal: compute $E_p[\gamma(X)] = \int_x \gamma(x)\, p(x)\, dx$

(compute expected value of any function of X, when $X \sim p$)

- Monte Carlo:

$$E_p[\gamma(X)] = \int_x \gamma(x)\, p(x)\, dx$$
$$\approx (1/N) \sum_i \gamma(X^i), \quad X^i \sim p$$

- Imp Sampling: If cannot compute p (e.g. p is a posterior), or cannot sample efficiently from it:

$$E_p[\gamma(X)] = E_q[\gamma(x)\, p(x)/q(x)]$$
$$\approx (1/N) \sum_i \gamma(X^i)\, w^i, \quad X^i \sim q$$
$$w^i \propto p(X^i) / q(X^i)$$

PF: Seq Importance Sampling

Seq Imp Sampling to approx $p(X_{1:t} \mid Y_{1:t})$

- Choose Imp Sampling density s.t. it factorizes as
  $q_{t,Y_{1:t}}(X_{1:t}) = q_{t-1,Y_{1:t-1}}(X_{1:t-1}) \; q_{X_{t-1},Y_t}(X_t)$
  - Allows for recursive computation of weights

- Seq Imp Sample: At each t, for each particle i,
  - Importance Sample: $X_t^i \sim q_{X_{t-1}^i,Y_t}(X_t)$
  - Weight:
    $w_t^i \propto w_{t-1}^i \, p(Y_t \mid X_t^i) \, p(X_t^i \mid X_{t-1}^i) / q_{X_{t-1}^i,Y_t}(X_t^i)$

Existing Work & Our Key Ideas

Existing Work

- Kalman Filter [BIR,CDC'94][TZ'92][P,PAMI'99][BB,CDC'94]
  - Finite dim contour rep, Assumed posterior unimodal
- Particle Filter [Condensation,ECCV'96]
  - Handled multimodal posterior (modeled clutter & occlusion probability in OL). But tracked only on a 6-dim space of affine deformations.
- Approx linear observer + level set [JYS,CDC'04][NT,CDC'4]
  - Infinite dim contour rep. (level set method). But assumed posterior unimodal.
- Exemplars + PF [TB, ICCV'01][ZF,ICCV'03]
  - Choose from a set of exemplars of possible deformations

Existing Work: Problems

- Finite dim. rep. + linear observer: not handle changes in contour length/ topology, or multimodal posterior

- Condensation: handled multimodal posteriors, but PF tracked only on 6-dim space of affine deformation

- Approx. linear observers + level set method
  - Level set method: handled infinite dim deformation
  - Did not handle multimodal posteriors
  - Uncoupled observers for global & local deformation

- Exemplars: very restrictive

Condensation fails

Rathi_PAMI2005_Plane_30_particles_condensation.avi

Rathi_PAMI2005_Plane_1200_particles_condensation.avi

Non-affine deformation due to frequent viewpoint changes. Multimodal OL (due to low contrast)

A Possible Solution

- Use level set representation
- Replace the approx. linear observer by a particle filter
  - Can track nonlinear systems: coupled observer
  - Can handle multimodal posteriors
- But brute force PF on space of deforming contours (very large dim space) expensive

Key Idea 1: “LDSS” [Vaswani etal,ICASSP’06]

- Even though space of contour deformation is very large dim, in most cases,
  - At any given time, most of the contour deformation occurs in a small # of dims (effective basis) while the deformation in the rest of the dims (residual space) is small
    - Different from dimension reduction (PCA) assumption
  - Effective basis dim can change with time

Key Idea 2:"Unimodality" [Vaswani,ICASSP'07]

- If residual deformation small enough (its STP narrow enough) compared to distance b/w OL modes (contour modes in image), can show that the "residual posterior" is unimodal
  - "residual posterior": posterior of residual deformation conditioned on effective basis states

- This is ensured by choosing enough dims as part of effective basis

Key Idea 3: "IS-MT" [Vaswani, ICASSP'07]

- If residual deformation still smaller (its STP still narrower), the residual posterior is unimodal & also narrow

- If an importance sampling (IS) density is unimodal & narrow, any sample from it is close to its mode with high probability
  - A valid approx is to just use its mode as the sample: Mode Tracking (MT) approx of IS or IS-MT
  - Resulting algorithm is called PF-MT

Affine PF-MT

Affine PF-MT [Rathi et al, CVPR'05, PAMI'07]

- Contour represented using level sets
- Use Importance Sampling to track on the 6-dim space of affine deformations (effective basis)
- For each affine deformed contour particle, track the unique mode of the posterior of non-affine deformation (residual space): Mode Tracking (MT)
- Very efficient. Importance sampling dimension was only K=6: small N sufficed for given accuracy

Affine PF-MT algorithm

At each t, for each particle i, do

- Importance Sample
  - Sample on 6D space of affine deformations
  - Apply affine deformation, A_t^i to each contour particle, C_{t-1}^i, to get $\hat{C}_t^i$
- Mode Track on residual (non-affine) deformation
  - Compute the single mode of $p(C_t | \hat{C}_t^i, Y_t)$

    $$m_t^i = \arg\min_C \; [\, E(C) + d^2(C, \hat{C}_t^i) / \Delta \,]$$

  - Set $C_t^i = m_t^i$ (replacing IS by MT)
- Weight & Resample

  $$w_t^i \propto w_{t-1}^i \exp[-E(C_t^i)] \exp[- d^2(C_t^i, \hat{C}_t^i)/\Delta_r]$$

Affine PF-MT assumes

- Assumes posterior of non-affine deformation (conditioned on affine def) is unimodal
  - **Much weaker than assuming posterior unimodal**

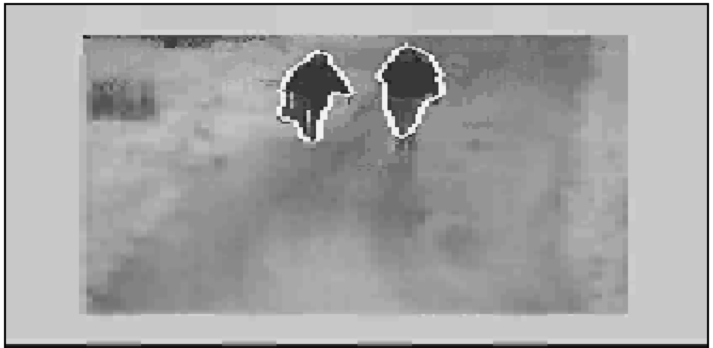
- This is satisfied whenever either
  - Small non-affine deformation per frame

  OR
  - OL modes (contour modes in image) separated only by translation or scale or other affine deformation

Examples where assumption holds

- Images w/ non-overlapping clutter
- Images w/ cluttering object separated by scale
- Overlapping clutter/partial occlusions/low contrast (multiple non-affine OL modes), but small non-affine deformation per frame
  - e.g. rigid body viewed under camera viewpoint changes in low contrast imagery
  - e.g. human body contour from a distance

Tracking multiple slow deforming objects from low contrast images



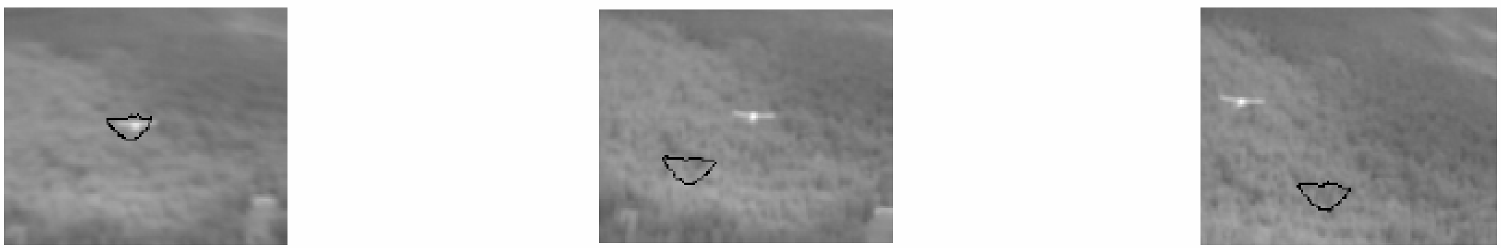
Rathi_PAMI2005_couple_sequence.avi

Posterior multimodal

Low contrast, viewpoint changes

- Deformation due to perspective camera effects (changing viewpoints), e.g. UAV tracking a plane

Condensation fails



Affine PF-MT works

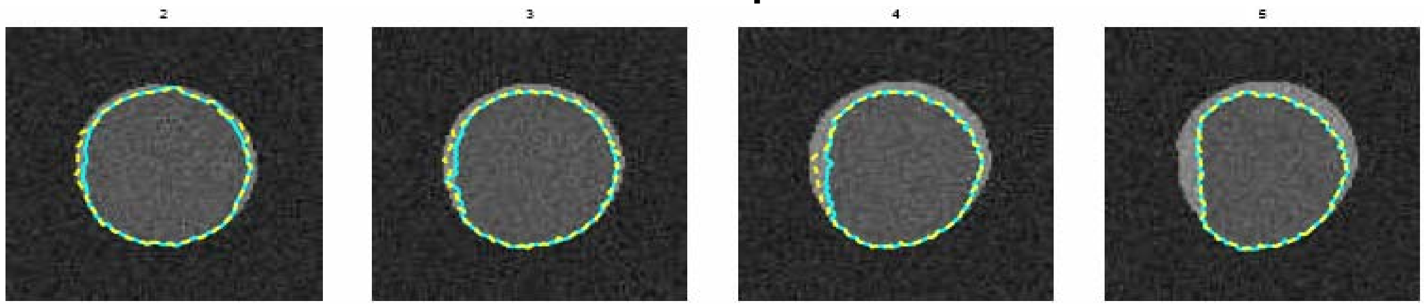
Assumption does not hold…

- Assumption fails when large non-affine deformation per frame (fast deforming sequence) and OL multimodal as a function of non-affine deformation (low contrast images, overlapping clutter or partial occlusions)

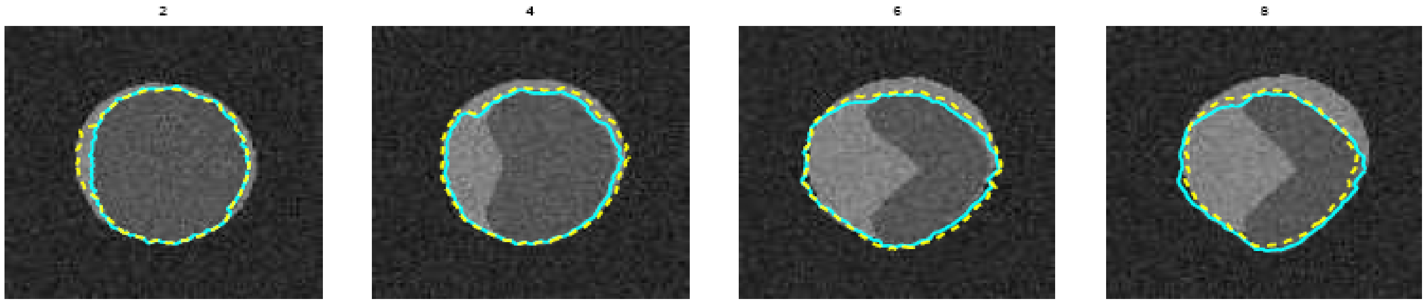
- Results in multimodal residual posterior of non-affine deformation

Overlapping Background clutter

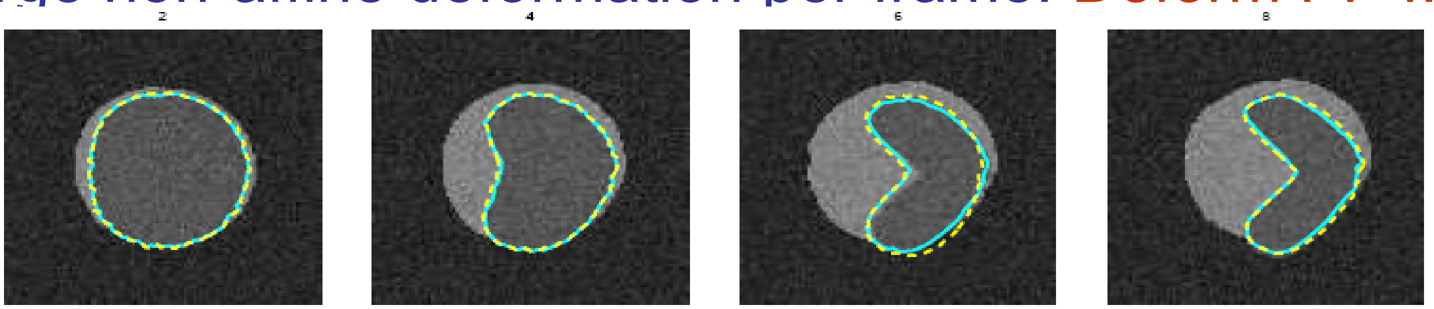
Small non-affine deformation per frame: Affine PF-MT works



Large non-affine deformation per frame: Affine PF-MT fails



Large non-affine deformation per frame: Deform PF-MT works



Deform PF-MT

Deform PF-MT [Vaswani et al,CDC'06]

- For multimodal posterior of non-affine deformation, need an importance sampling step in PF that also samples on space of local deformations
  - This is a very large dim space: regular PF inefficient

- Again use PF-MT but with deformation at a subsampled set of K contour points (basis points) & translation as the effective basis

- K = # of basis points: fixed or time varying

Deform PF-MT Algorithm

At each t, for each particle i, do

- Importance Sampling (IS)
  - IS on translation & move contour (its level set fn)
  - IS deformation at K subsampled basis points
  - Interpolate to get deformation at all contour points
  - Obtain extension velocity & use it to move level set fn
- Mode Tracking (MT)
  - Compute unique mode of residual deformation posterior (assumed unimodal)
  - Set contour particle equal to this mode
- Weight & Resample

Many Implementation Issues

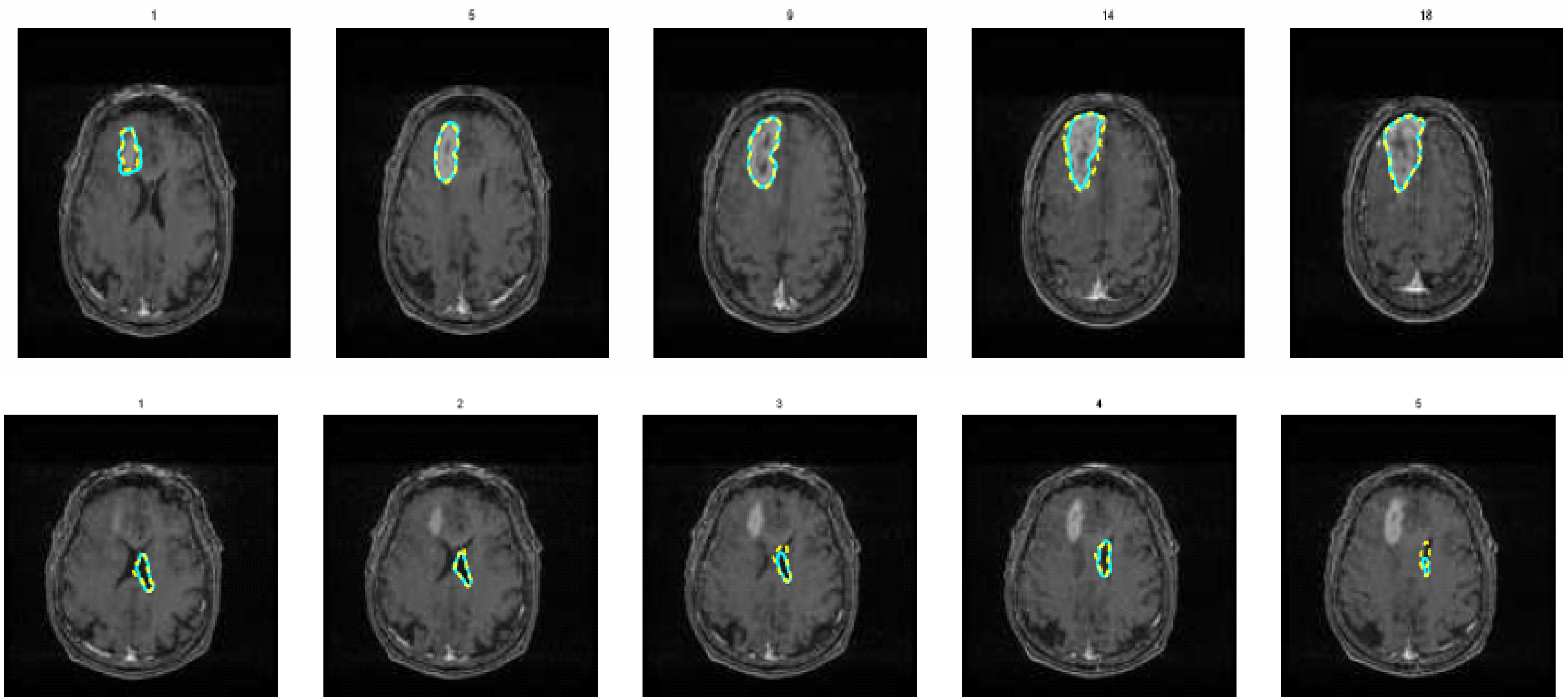
- Imp. Sample + Level Set Rep of contour: CFL?
  - Interpolate & compute extension velocity efficiently
  - Using multiple iterations to deform a contour: slow
- How to parameterize contour deformation?
  - As as a function of arclength: Expensive to implement using level sets, Cannot handle topology change
  - As function of radial/tangent angle: Fails if 2 contour points far along arclength are close along angle
- Estimate/Change effective basis dimension?
  - Use spatial freq response of deformation "signal"

Deform PF-MT assumes

- Residual deformation variance small enough compared to distance between OL modes, s.t.
  - Its posterior is unimodal **and**
  - Its posterior is narrow enough to justify IS-MT
- May need to change K over time to satisfy assumptions

Low contrast images, large def per frame: Brain MRI (Tumor, Ventricle)

- Multiple nearby modes due to low contrast

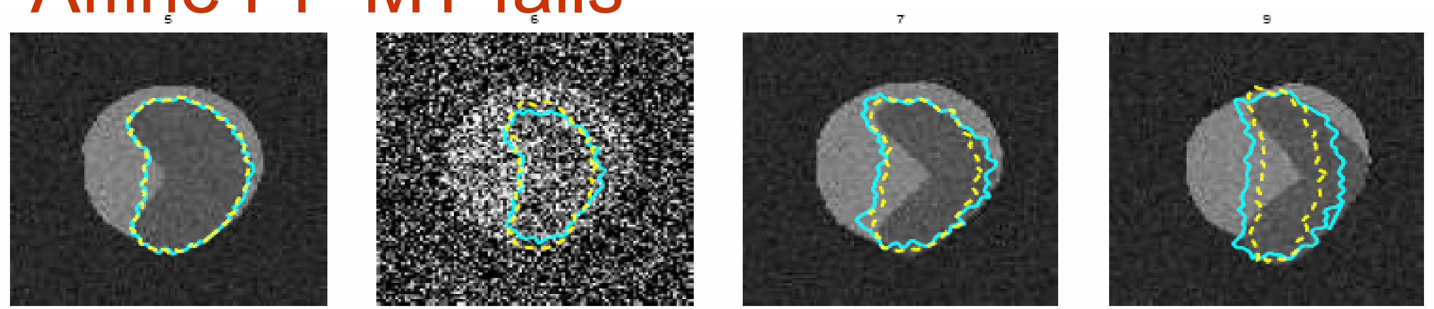


(b) Attempt to track the right ventricle (black region in the center) using Algorithm 2. Notice the low contrast imagery.

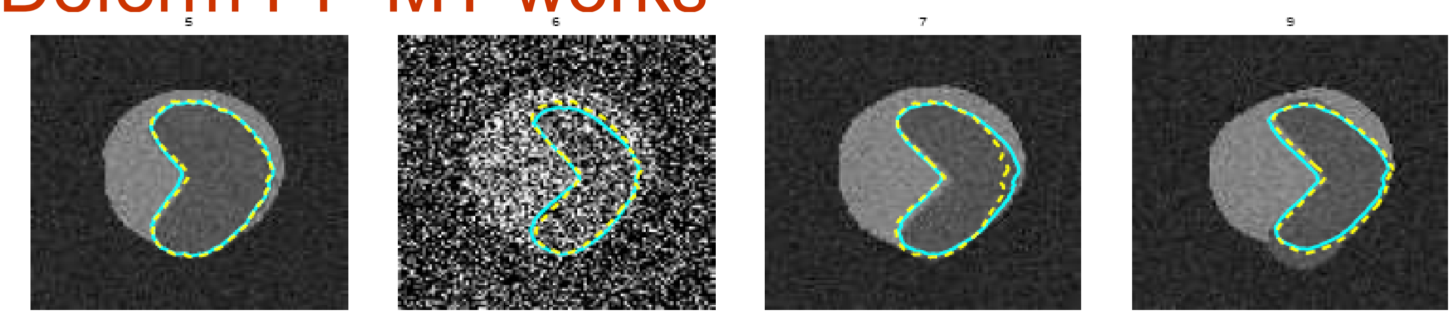
Outlier: multiple nearby modes

- Every even frame: outlier frame
- Multiple nearby modes separated by non-affine deformation

Affine PF-MT fails



Deform PF-MT works



Relation to Other Work

- PF-MT
  - Extension of PF-Doucet [Doucet'98]
  - Approx to Rao-Blackwellized PF [Chen-Liu'00]
- PF for tracking Heart LV [Sun et al, MICCAI'04]
  - PF-MT with PCA effective basis + retaining MAP particle
- Approx linear observer + level sets [JYS'04]
  - PF-MT with zero dimensional effective basis
- Condensation [IB'96]
  - PF-MT with zero dimensional residual space
- Stochastic Active Contours
  - Annealing for segmentation

PF Smoother

- Implemented tracking, even though sequential segmentation is often an offline problem
- Smoother: gives a non-causal estimate, better if algorithm is accurate enough
  - Use [Godsill et al,JASA'04] to approx $p(X_{1:T}|Y_{1:T})$. X_t = contour, deformation velocity
  - Uses tracker output as starting point.

System Identification

System Id Problem

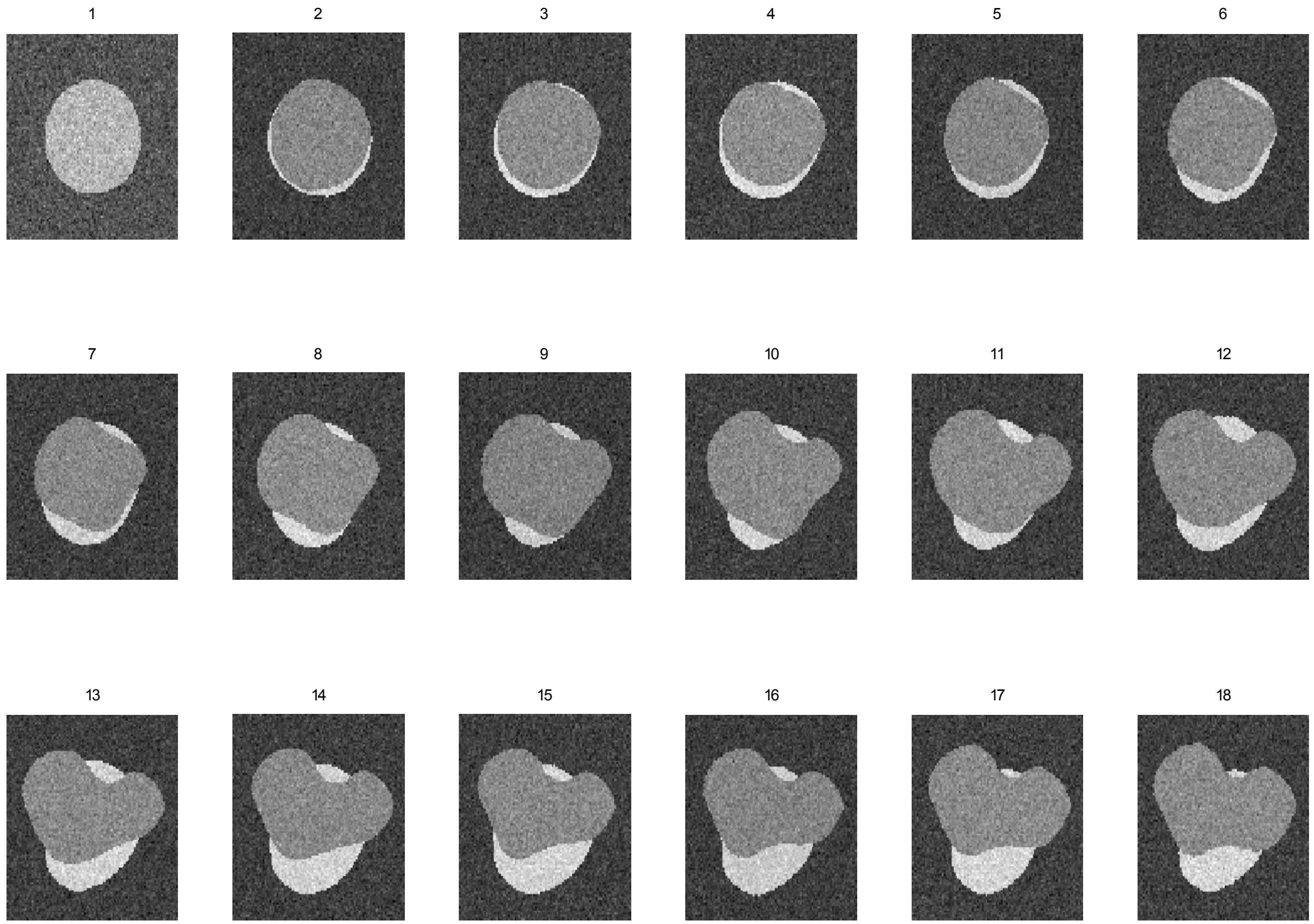
- Contour deformation sequence: time sequence of periodic spatial "signals"
  - "signal": spatially stationary or p.w. stationary

- Given a deformation sequence
  - Estimate effective dimension (K)
  - Learn temporal dynamics

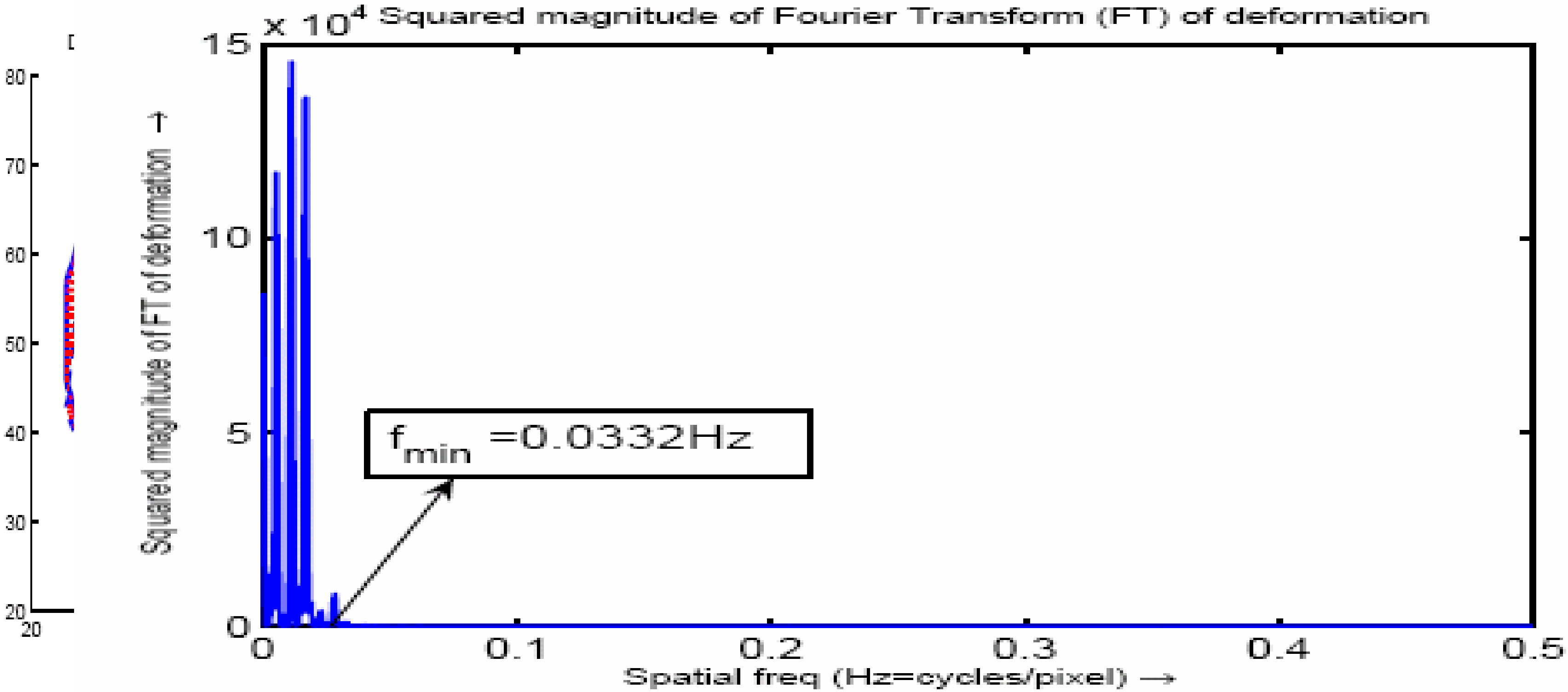
Estimating Effective Dimension

- Assume $v_t = C_t - C_{t-1}$ temporally stationary
- Eigen effective basis
  - Problems: data dependent basis, need fixed dim of v_t
- Fourier effective basis
  - Periodic & spatially stationary: PSD = eigenvalue
  - Compute PSD of v_t. Choose f_{min} s.t. sum of residual PSD small enough
  - Nyquist: $\alpha_s = 1/(2f_{min})$, $K = \lceil L / \alpha_s \rceil = \lceil L \,.\, 2f_{min} \rceil$

Computing PSD & K: Simulated Seq

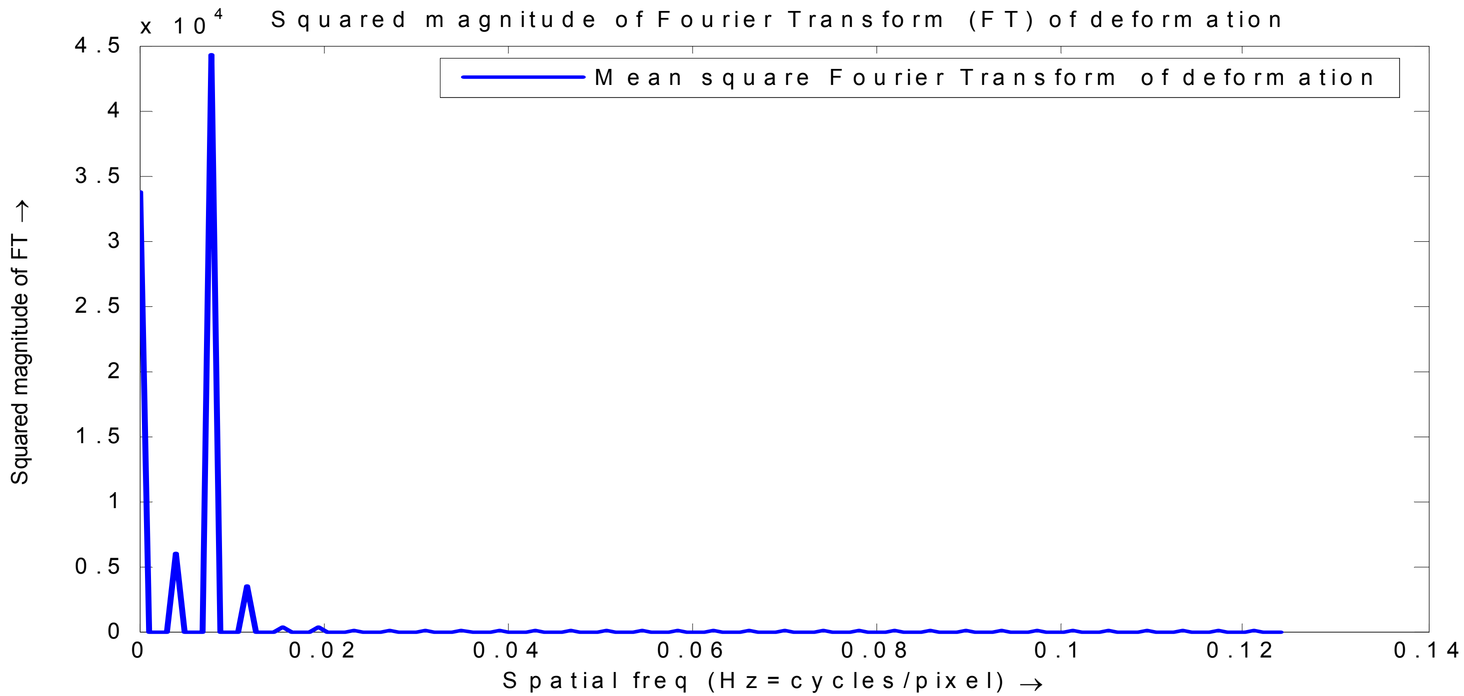


PSD computation steps



For 0.05% residual deformation, f_{min} = 0.0332 Hz.
M = L = 178, K = $\lceil M.\ 2f_{min} \rceil$ = 12

Background object: spatially stationary 10 frame PSD estimate



PSD using v_3 to v_{12} stretched & resampled to M = 256 points. For 5% residual deformation, f_{min} = 0.012Hz, $K = \lceil M.2f_{min} \rceil = 6$

Temporal Dynamics

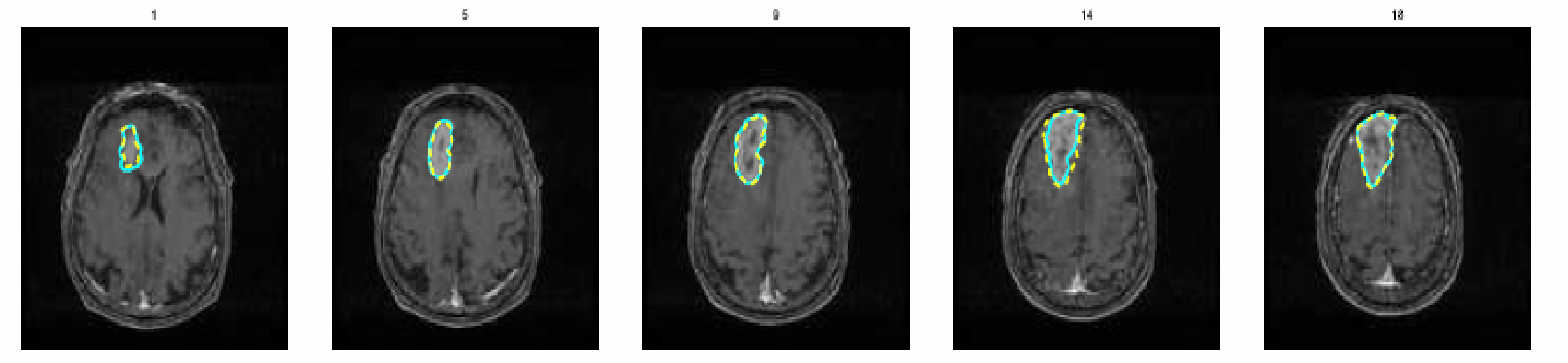
- System model in Fourier domain
  - AR model for time series of FT coefficients
  - Periodic & spatially stationary: FT = KLT
    - FT coefficients uncorrelated: separate AR model for each coefficient

- Get equivalent model in the space domain

Applications of Spatial PSD

- Estimate K offline or while tracking/smoothing
- Recognition & Change detection
  - Recognition by model comparison, e.g. disease progression models (schizophrenia, tumor shape change)
  - Changes in a sequence, by detecting change in K, in entire PSD or in temporal dynamics, e.g. detect abnormal changes in heart beat patterns or in brain shape deformation during surgery

Tumor contour sequence



Summary & Ongoing Work

- Affine PF-MT
  - Use when small non-affine deformation per frame
- Deform PF-MT & Smoother: Fixed or changing K
  - Human body contour tracking
  - Heart LV, Brain ROIs, Lung ROIs
- System Id + Recognition or Change Detection
  - Models for disease progression, e.g. schizophrenia
  - Heartbeat patterns: identifying abnormality
  - Abnormal brain shape deformations during surgery?

Open Issues

- PF smoother for offline sequence segmentation
- Extensions to surface tracking
- Observation models, tracking intensity variations
- Level Set Rep. + Imp Sampling: speedup, CFL

- System id
  - Parametrizing deformation: arclength or angle?
  - Warping of spatial axis (change in arclength) over time
  - Spatially or temporally nonstationary deformation

Collaborators

- Affine & Deform PF-MT
  - Yogesh Rathi, Anthony Yezzi & Allen Tannenbaum at Georgia Tech
- System Identification
  - Ongoing work with my student, Wei Lu

Other applications of PF-MT

- Spatially varying illumination change of moving objects
  - Moving into lighted room, face tracking [Kale et al, ICASSP'07]
  - Vehicle tracking through changing illuminations
- Change in spatially varying physical quantities using sensor networks
  - Tracking temperature change [Vaswani, ICASSP'07]
- Deformations of shapes of landmark points using the nonstationary shape activity model

Illumination Tracking: PF-MT [Kale et al'07]

- State = Motion (3 dim) + Illumination (7dim)
- PF on motion (3 dim) & MT on illumination
  - Illumination change very slow
  - OL usually unimodal as a function of illumination
  - If OL multimodal (e.g. occlusions), modes usually far apart compared to illumination change variance

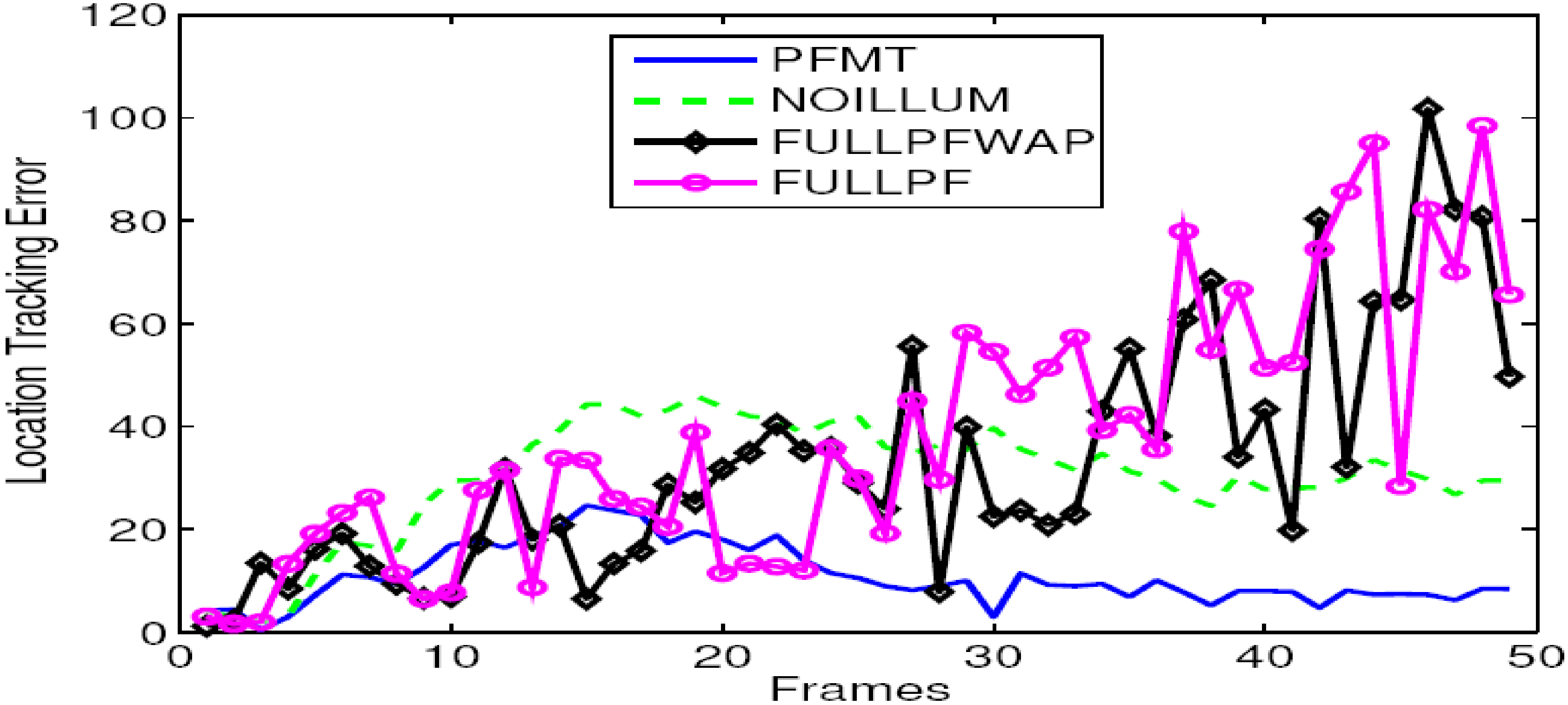
Face tracking results

PF-MT

3 dim PF (no illum)

10-dim Auxiliary PF

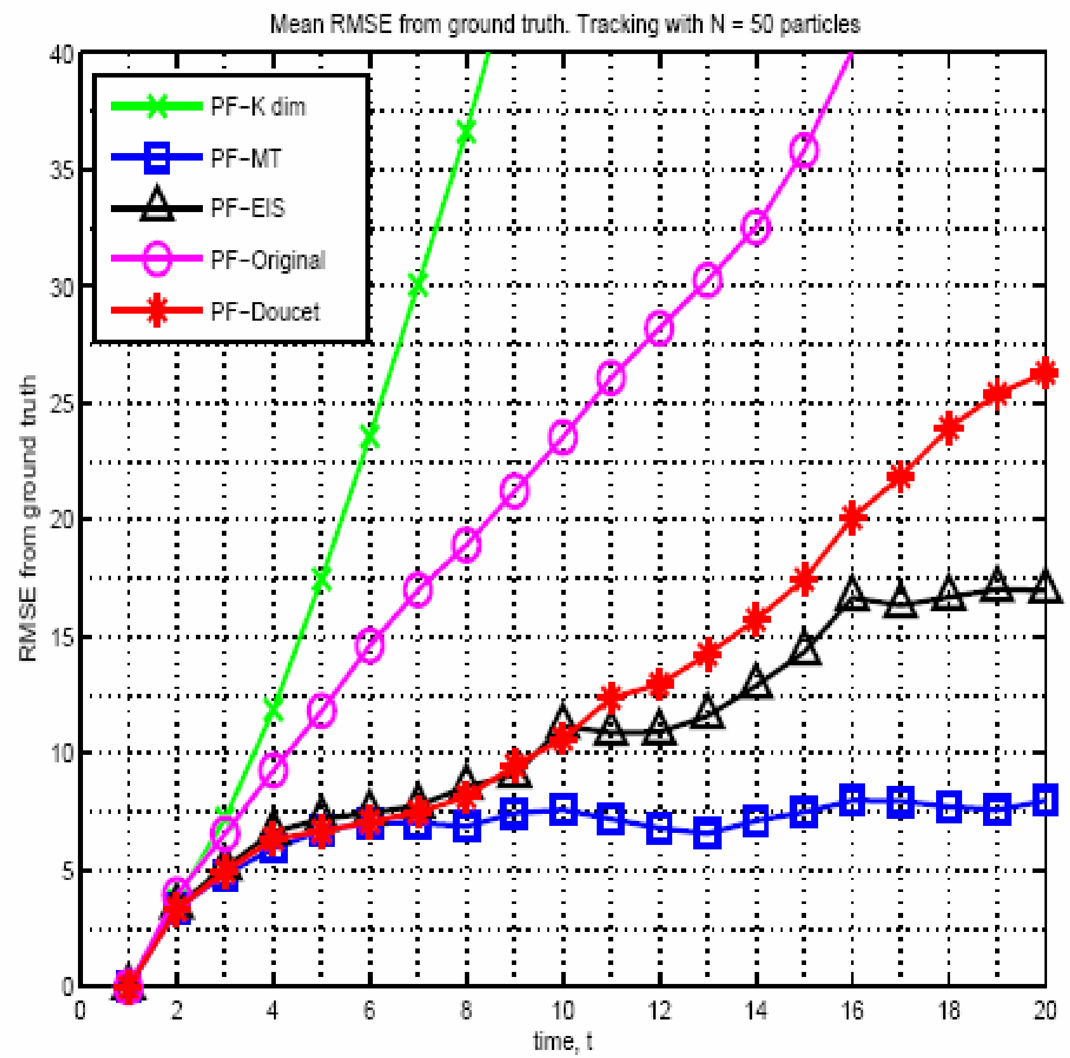
Error from ground truth



Comparing with 10 dim regular PFs (original, Auxiliary) & with PF- K dim (not track illumination at all)

Sensor nets: Temperature tracking

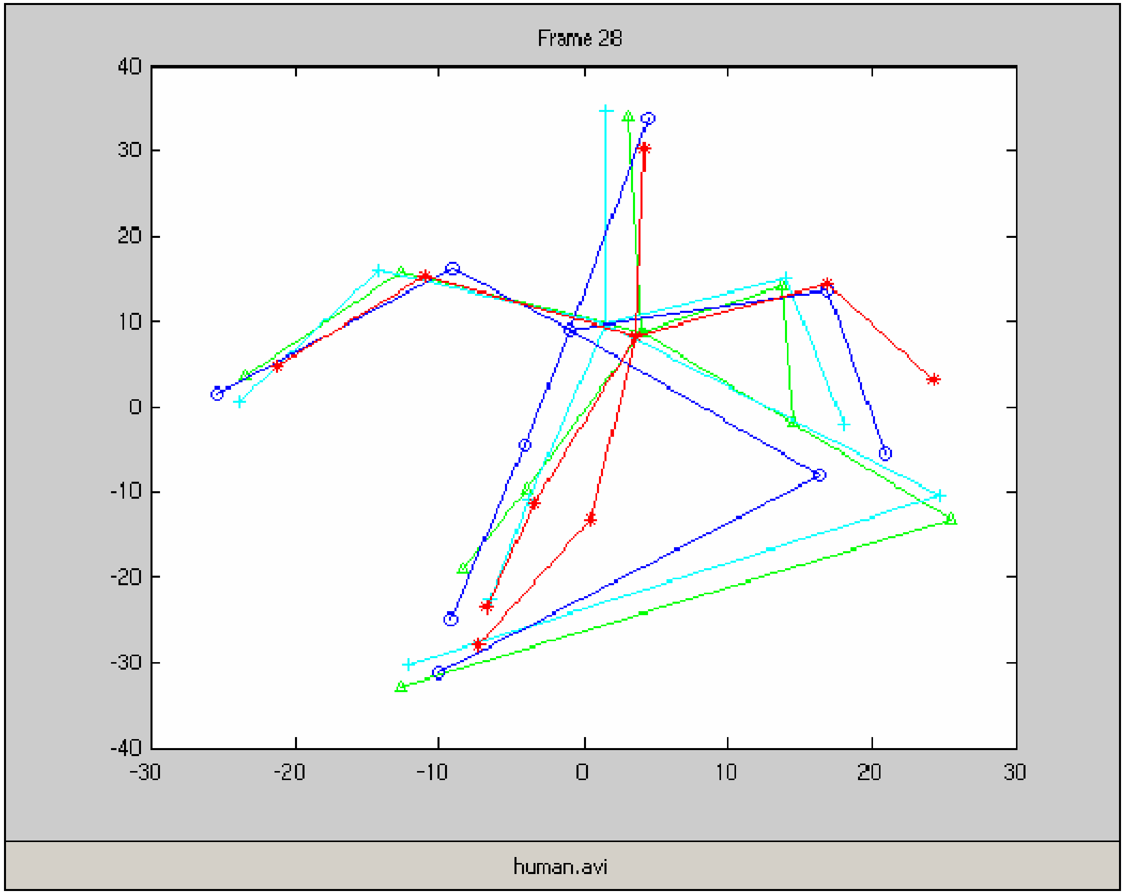
- $\text{Dim}(X_t) = 10$
- K = 1, i.e. Δ_s=10, Δ_r =1, & OL multimodal
- N = 50 particles
- Plotting RMSE from ground truth
- PF-MT better than all full PFs (PF-EIS, PF-D) & PF-K dim (dim reduced PF)



Landmark Shape Tracking

- Tracking deformations of shapes of landmark points using nonstationary SA (NSSA) model
- NSSA better models larger & nonstationary shape changes than existing methods (ASMs)
  - Existing ASM work uses piecewise ASMs to track long sequences, e.g separate ASM for systolic & diastolic heart motion, or hierarchical ASMs
  - Cannot model transitions b/w pieces very well
  - Cannot detect change while tracking

Landmark Shape Tracking



Landmark Shape Tracking

- Compared modeling error of our method (NSSA) with Active Shape Models for CMU MOCAP dataset (human action sequences)
- For all sequences, modeling error of our method much smaller than ASM

Modeling Error Comparison

- Defined 10 dimensional PCA space for ASM and for shape velocity (our method)
- Defined AR model for ASM & for shape velocity. Total modeling error

  Crawl: ASM: 0.00870, Shape Velocity: 0.00030

  Sit: ASM: 0.00760, Shape Velocity: 0.00005

  Interview: ASM: 0.00450, Shape Velocity: 0.00020