PF with Efficient Importance Sampling (EIS) and Conditional Posterior Mode Tracking (MT)

Namrata Vaswani

Dept of Electrical & Computer Engineering

Iowa State University

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

IOWA STATE UNIVERSITY
Becoming the best.

Hidden Markov Model & Goal

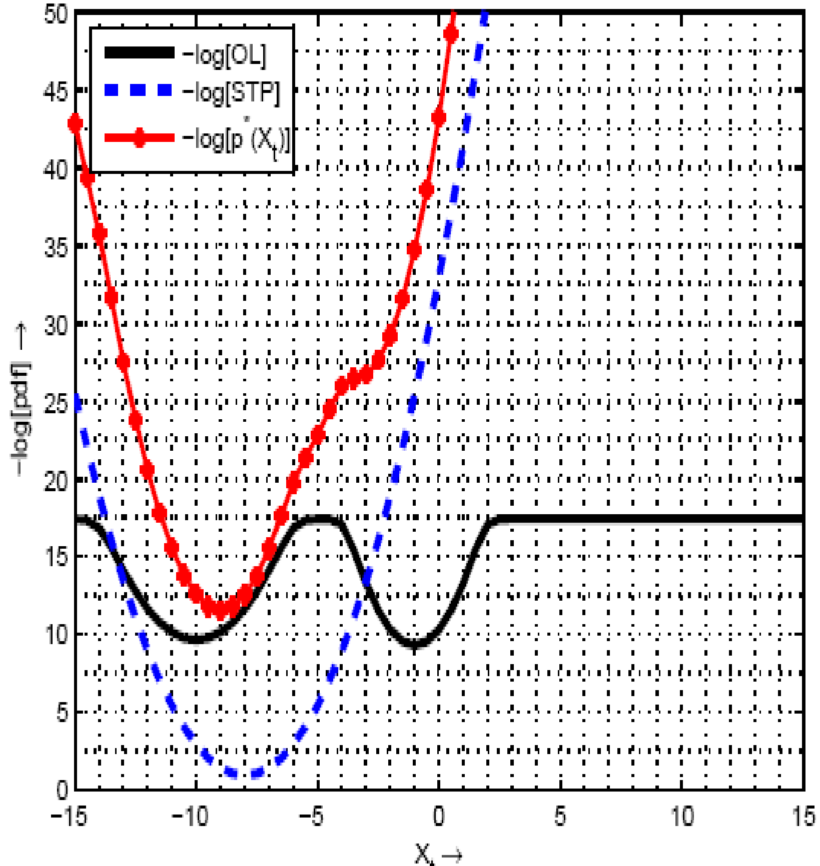
- hidden state sequence: $\{X_t\}$, observations: $\{Y_t\}$
  - state sequence, $\{X_t\}$, is a Markov chain
  - Y_t conditioned on X_t independent of past & future
  - $p(x_t|x_{t-1})$: state transition prior (known)
  - $p(y_t|x_t)$: observation likelihood (known)

- **Goal:** recursively get the optimal estimate of X_t at each time, t, using observations, $Y_{1:t}$
  - compute/approximate the posterior, $\pi_t(X_t) := p(X_t|Y_{1:t})$
  - use π_t to compute any "optimal" state estimate, e.g. MMSE, MAP,…

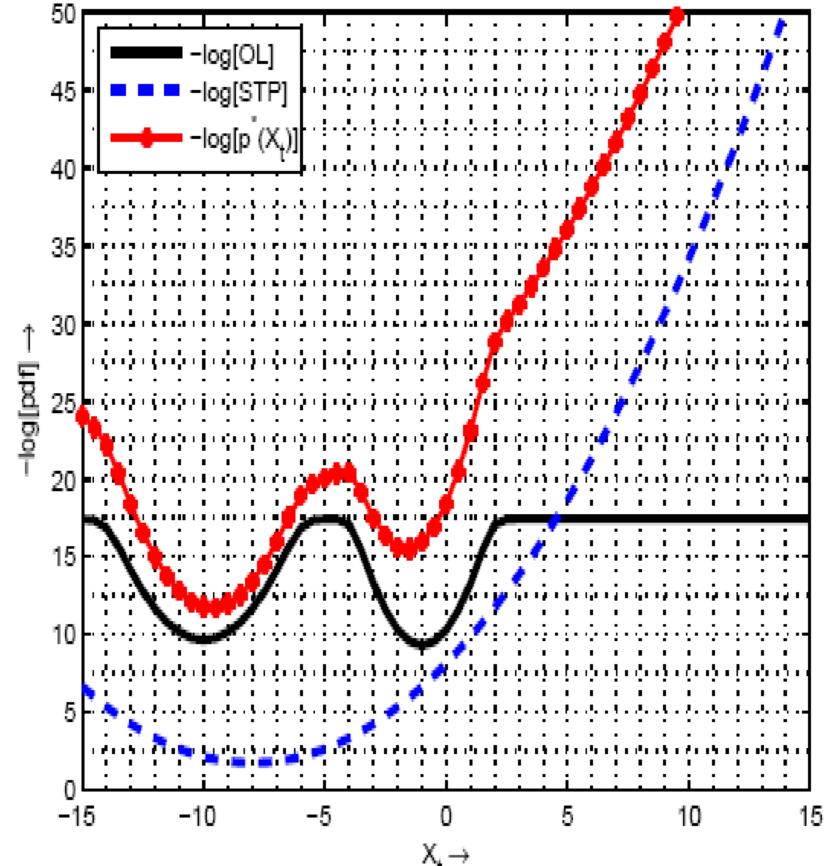
Problem Setup

- Observation Likelihood is often multimodal or heavy-tailed
  - e.g. some sensors fail or are nonlinear
  - e.g. clutter, occlusions, low contrast images
  - If the state transition prior is narrow enough, posterior will be unimodal: can adapt KF, EKF
    - If not (fast changing sequence): req. a Particle Filter
- Large dimensional state space
  - e.g. tracking the temperature field in a large area
  - e.g. deformable contour tracking
  - PF expensive: requires impractically large N

Narrow prior: Unimodal posterior



Broad prior: Multimodal posterior



Temperature measured with 2 types of sensors, each with nonzero failure probability

Multimodal likelihood examples – 1

- Nonlinear sensor [Gordon et al'93]
  - sensor measuring the square of temperature corrupted by Gaussian noise

    $Y_t = X_t^2 + w_t, \ w_t \sim N(0,\sigma^2)$

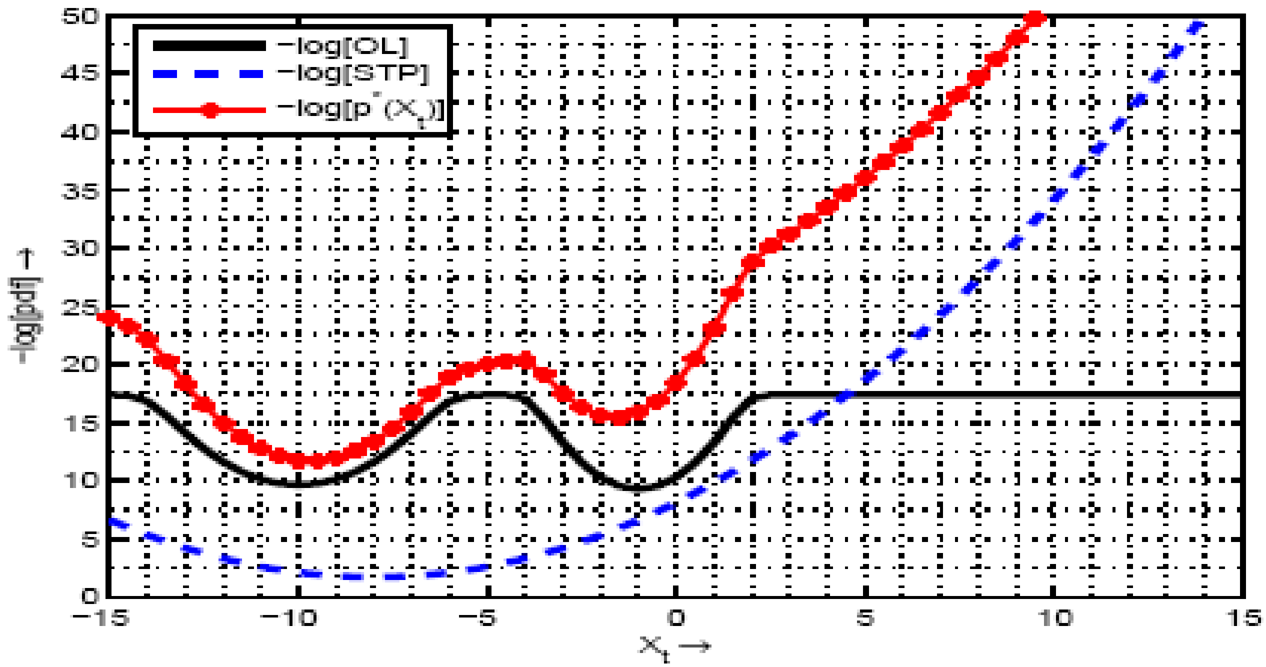
    - whenever $Y_t > 0$, $p(Y_t|X_t)$ is bimodal as a function of X_t with modes at $X_t = Y_t^{1/2}$, $-Y_t^{1/2}$

- More generally, if observation = many-to-one function of state + noise [Kale-Vaswani, ICASSP'07]
  - $Y_t = h_1(X_{t,1})\, h_2(X_{t,2}) + w_t$: h_1, h_2 monotonic

Multimodal likelihood examples – 2

- Sensors with nonzero failure probability
  - temperature measured with 2 sensors, each with some probability of failure, α, conditionally indep.

    $Y_{t,i} \sim (1-\alpha)N(X_t,\sigma^2) + \alpha\, N(0, 100\,\sigma^2)$, i=1,2
  - bimodal likelihood if any of them fails

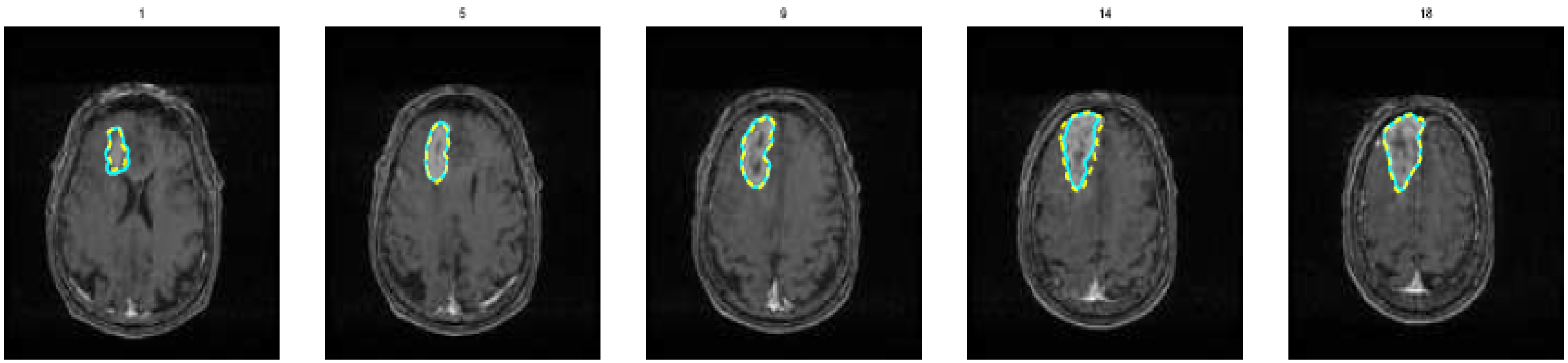


IOWA STATE
Becoming

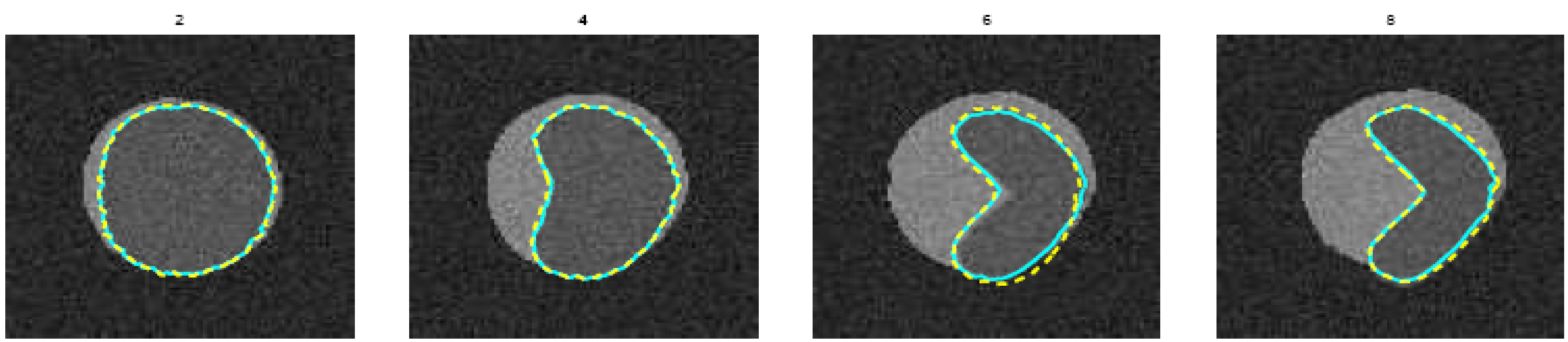
Multimodal likelihood examples – 3

- Deformable contour tracking [Isard-Blake'96][Vaswani et al'06]

through low contrast images (tumor region in brain MRI)



through overlapping background clutter



Particle Filter [Gordon et al'93]

- Sequential Monte Carlo technique to approx the Bayes' recursion for computing the posterior $\pi_t(X_{1:t}) = p(X_{1:t}|Y_{1:t})$
  - Approx approaches true posterior as the # of M.C. samples ("particles")→ ∞, for a large class of nonlinear/non-Gaussian problems

- Does this sequentially at each t using Sequential Importance Sampling along with a Resampling step (to eliminate particles with very small importance weights)

Outline

- In this talk, I will focus on
  – efficient importance sampling (EIS)
  – conditional posterior mode tracking (MT)

  – PF with EIS & PF with MT: easy extension
  – PF-MT for deformable contour tracking

Existing Work – 1

- PF-Original: Importance Sample from prior [Gordon et al'93]
  - always applicable but is inefficient
- Optimal IS density: $p^*(x_t) := p(x_t \mid x_{t-1}, y_t)$ [D'98][older works]
  - cannot be computed in closed form most cases
- When the optimal IS density, p*, is unimodal
  - Adapt KF, EKF, PMT [Brockett et al'94][TZ'92][Jackson et al'04]
    - Possible if the posterior is unimodal too
  - PF-D: IS from Gaussian approx to p* [Doucet'98]
  - Unscented PF [VDDW,NIPS'01]: UKF to approx to p*
- MHT, IMM, Gaussian Sum PF [Kotecha-Djuric'03], …
  - practical only if # of modes is small & known

Existing Work – 2

- If a large part of state space conditionally linear Gaussian or can be vector quantized
  - use Rao Blackwellized PF [Chen-Liu'00][SGN,TSP'05]
- If a large part of state space is asymp. stationary
  - marginalize over it using MC [Chorin et al'04][Givon et al'08]
- If cannot do either: need **PF-EIS w/ Mode Tracker**
- Resampling modifications
  - Look ahead resampling: Auxiliary PF [Pitt-Shepherd'99]
  - Repeated resampling within a single t [Oudjane et al'03]

Corresponding static problem

- Compute the MMSE estimate of a **large dimensional** state/signal, X, from its observation, Y
- Study problems where Y is a nonlinear and non-Gaussian noise corrupted function of X: resulting in **frequently multimodal or heavy-tailed likelihoods**
- The MMSE estimate, $E[X|Y=y] = \int xp(x|y)dx$, requires computing the posterior,

$$p^*(x) := \underbrace{p(x|y)}_{\text{posterior}} \propto \underbrace{p(y|x)}_{\text{likelihood}} \underbrace{p(x)}_{\text{prior}}$$

When p^* cannot be computed analytically: use importance sampling

Example Applications:

- Temperature, pressure or other random field estimation from a set of unreliable and noisy sensor measurements
- Segment deforming objects from clutterred, low contrast or partly occluded images

Issues

- Sample from prior, weight by likelihood: inefficient
- Sample from Gaussian approx to posterior, p^*: valid only if p^* is (effectively) unimodal
- Sample from Gaussian or other mixture density approx's to p^*: practical only if number of possible modes of p^* is small
- Marginalize over part of the state space: can be done only in certain special cases
- **Large dimensional problems with multimodal likelihoods**
  - if the prior is broad, the posterior, p^*, will be multimodal
  - number of possible modes of p^* often increase exponentially with dimension
  - effective sample size reduces as dimension increases

Key proposed ideas

- Use the fact that in most large dim. problems, **the prior is broad in only a few dimensions (multimodal states)**
- If in the rest of the dimensions, the prior is unimodal and "narrow enough", **the posterior conditioned on the multimodal states (conditional posterior) will be unimodal**
- If the conditional posterior is also very narrow, there is little error in **replacing imp. sampling by posterior Mode Tracking (MT)**
  - MT: use conditional posterior mode as the sample
  - MT is an approx of imp sampling: introduces some extra error
  - But reduces sampling dimension by a large amount: improves effective sample size
  - **Net effect: smaller error when number of samples, N, is small**

Efficient importance sampling (EIS)

- Split the state, X, into a small dimensional multimodal part, X_s, and the rest of the states, X_r (s.t. conditional posterior of X_r is unimodal "mostly")
- **IS-prior:** For $i = 1, \dots N$, sample x_s^i from its prior, $p(x_s)$
- **EIS:** For $i = 1, \dots N$, sample x_r^i from a Gaussian approx to the conditional posterior, $p^{**,i}(x_r)$

$$p^{**,i}(x_r) \;:=\; p(x_r|y, x_s^i) \;\propto\; p(y|x_r, x_s^i)\; p(x_r|x_s^i)$$

denote the Gaussian approx. by $\mathcal{N}(x_r; m_r^i, \Sigma^i)$

- For $i = 1, \dots N$, weight appropriately: $w^i \;\propto\; \frac{p(y|x_r^i, x_s^i)\; p(x_r^i|x_s^i)}{\mathcal{N}(x_r^i; m_r^i, \Sigma^i)}$

Computing the Gaussian approx of the conditional posterior:

- Compute the mode of $p^{**,i}(x_r)$ as

$$m_r^i = \arg\min_{x_r} \underbrace{[\underbrace{-\log p(y|x_r, x_s^i)}_{E_y(x_r)} + \underbrace{-\log p(x_r|x_s^i)}_{D(x_r)}]}_{L(x_r)}$$

- Set Σ^i equal to the Hessian of $L(x_r)$ computed at m_r^i
- The Gaussian approx of $p^{**,i}$ is $\mathcal{N}(x_r; m_r^i, \Sigma^i)$

Conditional posterior mode tracking (MT)

- **IS-prior:** For $i = 1, \ldots N$, sample x_s^i from its prior, $p(x_s)$
- **IS-MT:** For $i = 1, \ldots N$, set $x_r^i = m_r^i$ (conditional posterior mode) where

$$m_r^i = \arg\min_{x_r} \left[-\log p(y|x_r, x_s^i) + -\log p(x_r|x_s^i) \right]$$

- For $i = 1, \ldots N$, weight appropriately: $w^i \propto p(y|x_r^i, x_s^i)\, p(x_r^i|x_s^i)$

EIS-MT

- Split X into a small dimensional multimodal part, X_s, and the rest of the states, X_r (s.t. conditional posterior of X_r is most likely to be unimodal)
- Split X_r into $X_{r,s}$ (larger prior variance) and $X_{r,r}$ (smaller prior variance)
- **IS-prior on X_s**
- **EIS on $X_{r,s}$**
- **IS-MT on $X_{r,r}$**
- Weight appropriately

Simulation results

- Simulated a temperature field sensing application when sensors are unreliable and prone to occasional outlier noise
- Modeled failure or outlier noise as a second Gaussian mixture component of sensor noise with very large variance. Cond. indep. measurements.

$$Y_j \sim (1-\alpha_j)\mathcal{N}(X_j, \sigma^2) + \alpha_j \mathcal{N}(0, 100\sigma^2)$$

$$Y_j \sim (1-\alpha_j)\mathcal{N}(X_j, \sigma^2) + \alpha_j \mathcal{N}(X_j, 100\sigma^2)$$

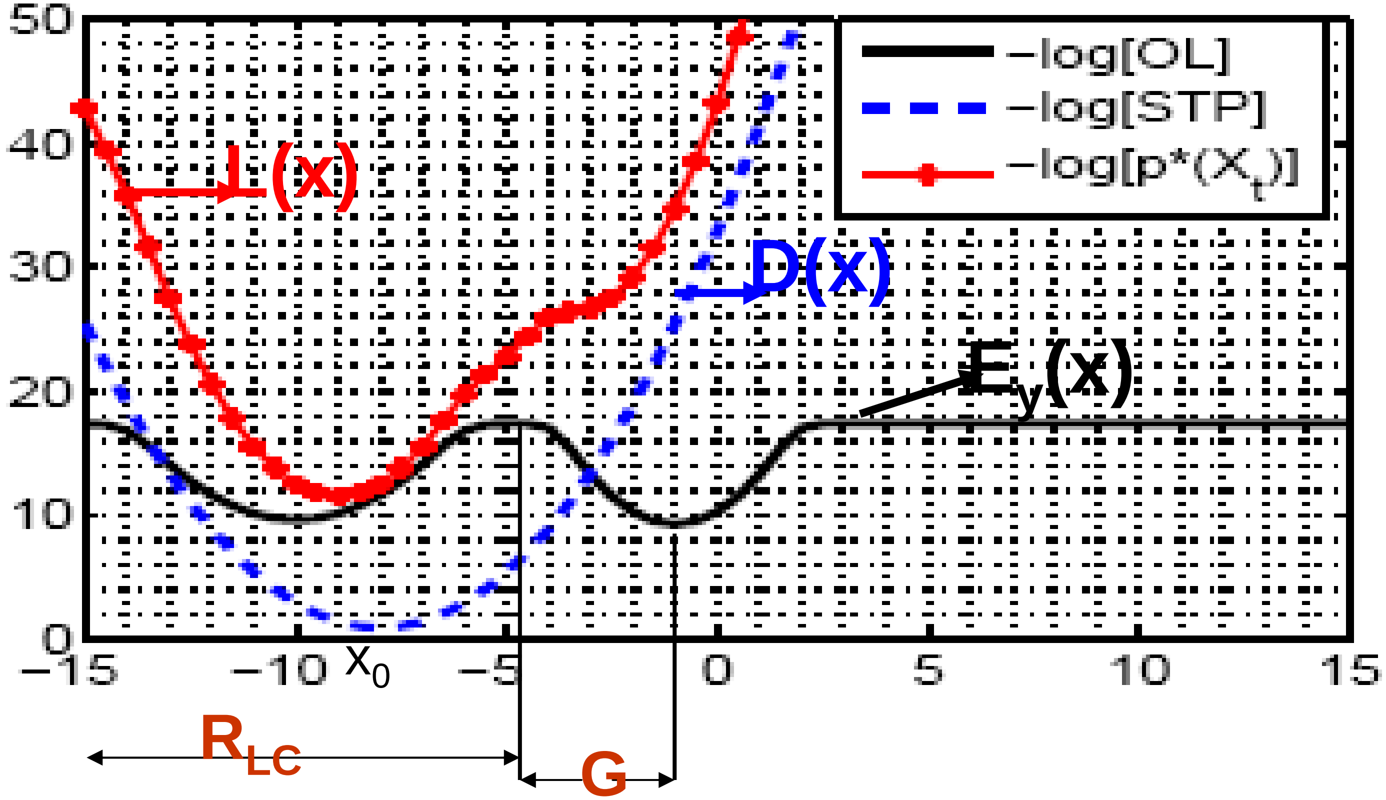
- Simulated a 7 dimensional system. Compared EIS, EIS-MT against IS-prior and IS-Gaussian
- When N=30 samples are used EIS-MT has best performance
- When N=100 samples are used EIS has the best performance

Importance Sampling method	**Err(N=100)**
EIS-MT ($X_s = [V_1]$, $X_{r,s} = [V_2, V_3]$, $X_{r,r} = [V_4, V_5, V_6, V_7]$)	0.0375
EIS ($X_s = [V_1]$, $X_{r,s} = [V_2, V_3, V_4, V_5, V_6, V_7]$, $X_{r,r} = empty$)	**0.0368**
IS-Gaussian ($X_s = empty$, $X_{r,s} = [V]$, $X_{r,r} = empty$)	0.0587
IS-prior ($X_s = [V]$, $X_{r,s} = empty$, $X_{r,r} = empty$)	0.0599

Importance Sampling method	**Err(N=30)**
EIS-MT ($X_s = [V_1]$, $X_{r,s} = [V_2, V_3]$, $X_{r,r} = [V_4, V_5, V_6, V_7]$)	**0.0416**
EIS ($X_s = [V_1]$, $X_{r,s} = [V_2, V_3, V_4, V_5, V_6, V_7]$, $X_{r,r} = empty$)	0.0449
IS-Gaussian ($X_s = empty$, $X_{r,s} = [V]$, $X_{r,r} = empty$)	0.0610
IS-prior ($X_s = [V]$, $X_{r,s} = empty$, $X_{r,r} = empty$)	0.0733

Conditional posterior unimodality

- The likelihood is multimodal ($E_y(x_r)$ has multiple minima)
- **How narrow should the prior be (spread of $D(x_r)$ be) so that the conditional posterior is unimodal ($L(x_r)$ has one minimizer)?**



Main idea of result

- Assume that
  - Prior is strongly log-concave, e.g. Gaussian ($D(x_r)$ strongly convex)
  - The unique minimizer of $D(x_r)$, x_r^*, is close enough to a minimizer of $E_y(x_r)$ to ensure that $E_y(x_r)$ is convex at x_r^*
  - R_{LC}: largest continuous region around x_r^* where E_y locally convex
- Inside R_{LC}, $L(x_r) = D(x_r) + E_y(x_r)$ is strongly convex, i.e. it has at most one minimizer
- **We need to bound the variance of the prior (spread of $D(x_r)$) so that outside R_{LC}, L has no stationary points (no minimizers)**

- **We need to bound the variance of the prior (spread of $D(x_r)$) so that outside R_{LC}, L has no stationary points (no minimizers)**
- Outside R_{LC}, ∇L can be zero only at points where, in all dimensions, ∇D and ∇E_y have different signs (or are both zero): call this region $\mathcal{G}$
- Notice that $D(x_r)$ has no stationary points outside R_{LC}
- If we can ensure that the sign of ∇L follows the sign of ∇D, in at least one dimension, at all points in $\mathcal{G}$, we will be done
- A sufficient condition for the above is that the **prior be Gaussian and the eigenvalues of its covariance be smaller than Δ^* where**

$$\boldsymbol{\Delta}^* := \inf_{\mathbf{x_r} \in \mathcal{G}} \max_{\mathbf{p}=\mathbf{1},\ldots\mathbf{M}} \frac{|[\mathbf{x_r} - \mathbf{x_r^*}]_\mathbf{p}|}{\epsilon_\mathbf{0} \pm |[\nabla \mathbf{E_y}(\mathbf{x_r})]_\mathbf{p}|}$$

(use $+$ where $[\nabla D]_p = [x_r - x_r^*]_p$, $[\nabla E_y]_p$ have different signs, use $-$ elsewhere)

The final result [Vaswani, TSP, Oct'08]

The posterior is unimodal if

- The prior is strongly log-concave, e.g. Gaussian, with unique mode x_r^*
- x_r^* is close enough to a mode of the likelihood to ensure that the likelihood is locally log-concave in its neighborhood: *call the largest such region R_{LC}*
- **The eigenvalues of the covariance of the Gaussian prior are less than Δ^* where**

$$\mathbf{\Delta}^* := \inf_{\mathbf{x_r} \in \mathcal{G}} \max_{\mathbf{p=1,\ldots M}} \frac{|[\mathbf{x_r} - \mathbf{x_r^*}]_\mathbf{p}|}{\epsilon_\mathbf{0} \pm |[\nabla \mathbf{E_y}(\mathbf{x_r})]_\mathbf{p}|}$$

(use $+$ where $[x_r - x_r^*]_p$, $[\nabla E_y]_p$ have different signs, use $-$ elsewhere)

- $E_y(x_r) := -\log p(y|x_s^i, x_r)$
- $\mathcal{G}$: *region outside R_{LC}, in which, in all dimensions, ∇E_y, ∇D either have different signs, or are both zero*

The exact result

- The posterior is unimodal if
  - the prior strongly log-concave, e.g. Gaussian
  - its unique mode, x_0, is close enough to a likelihood mode s.t. likelihood is locally log-concave at x_0
  - spread of the prior narrow enough s.t. $\exists$ an ϵ_o > 0 s.t.

$$[\inf_{x \in \cap_p (A_p \cup Z_p)} \max_p \gamma_p(x)] > 1$$

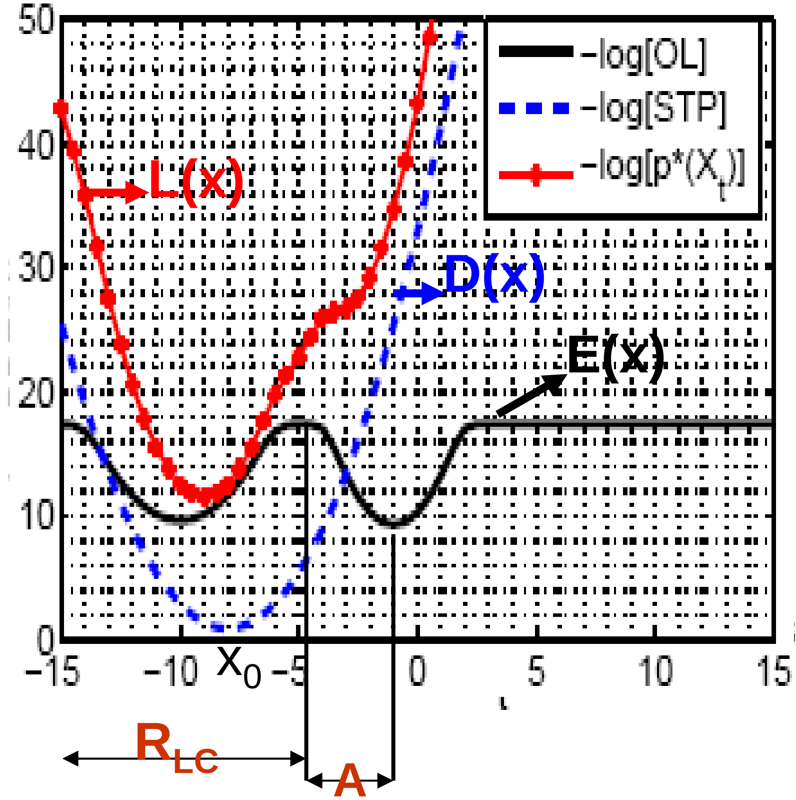
$$\gamma_p(x) := \begin{cases} \frac{|[\nabla D(x)]_p|}{\epsilon_0 + |[\nabla E(x)]_p|} & x \in A_p \\ \frac{|[\nabla E(x)]_p|}{\epsilon_0 - |[\nabla E(x)]_p|} & x \in Z_p \end{cases}$$

$$Z_p := {R_{LC}}' \cap \{x : [\nabla E]_p \cdot [\nabla D]_p \geq 0, |[\nabla E]_p| < \epsilon_0\}$$

$$A_p := {R_{LC}}' \cap \{x : [\nabla E]_p \cdot [\nabla D]_p < 0\}$$

Implications [Vaswani, TSP, Oct'08]

- Need a Gaussian prior with
  - the mode, x_0, close enough to a likelihood mode
  - max. variance small enough compared to distance b/w nearest & second-nearest likelihood mode to x_0
  - allowed max variance bound increases with decreasing strength of the second-nearest mode



PF-EIS algorithm [Vaswani, TSP, Oct'08]

- Split $X_t = [X_{t,s}, X_{t,r}]$
- At each t, for each particle i
  - **IS-prior:** Importance Sample $x_{t,s}^i \sim p(x_{t,s}^i | x_{t-1}^i)$
  - Compute mode of posterior conditioned on $x_{t,s}^i$, x_{t-1}^i
    $m_t^i = \arg\min_x -[\log p(y_t \mid x) + \log p(x \mid x_{t,s}^i, x_{t-1}^i)]$
  - **EIS:** Importance Sample $x_{t,r}^i \sim N(m_t^i, \Sigma_t^i)$
  - Weight
    $w_t^i \propto w_{t-1}^i \, p(y_t \mid x_t^i) \;\; p(x_{t,r}^i \mid x_{t,s}^i, x_{t-1}^i) \,/\, N(x_{t,r}^i \,;\, m_t^i, \Sigma_t^i)$
- Resample

An example problem

- State transition model: state, $X_t = [C_t, v_t]$
  - temperature vector at time t, $C_t = C_{t-1} + Bv_t$
  - temperature change coefficients along eigen-directions, (v_t): spatially i.i.d. Gauss-Markov model
  - Notice that temp. change, Bv_t, is spatially correlated

- Likelihood: observation, Y_t = sensor measurements

$$Y_{t,j} \sim (1-\alpha_j)\, \mathcal{N}(C_{t,j}, \sigma^2) + \alpha_j\, \mathcal{N}(0, 100\sigma^2)$$

  - diff. sensor measurements conditionally independent
  - with probability α_j, sensor j can fail
  - Likelihood heavy-tailed (raised Gaussian) w.r.t. $[C_t]_j$, if sensor at node j fails

Choosing multimodal state, $X_{t,s}$

Practical heuristics motivated by the unimodality result

- Get the eigen-directions of the covariance of temperature change
- If one node has older sensors (higher failure probability) than other nodes:
  – choose temperature change along eigen-directions most strongly correlated to temperature at this node and having the largest variance (eigenvalues) as $X_{t,s}$
- If all sensors have equal failure probability:
  – choose the K eigen-directions with largest variance (evals)

PF-EIS with Mode Tracking

- If for a part of the unimodal state ("residual state"), the conditional posterior is narrow enough,
  - it can be approx. by a Dirac delta function at its mode
- Mode Tracking (MT) approx of Imp Sampling (IS)
  - MT approx of IS: introduces some error
  - But it reduces IS dimension by a large amount (improves effective particle size): much lower error for a given N, when N is small
  - Net effect: lower error when N is small

PF-EIS-MT algorithm design

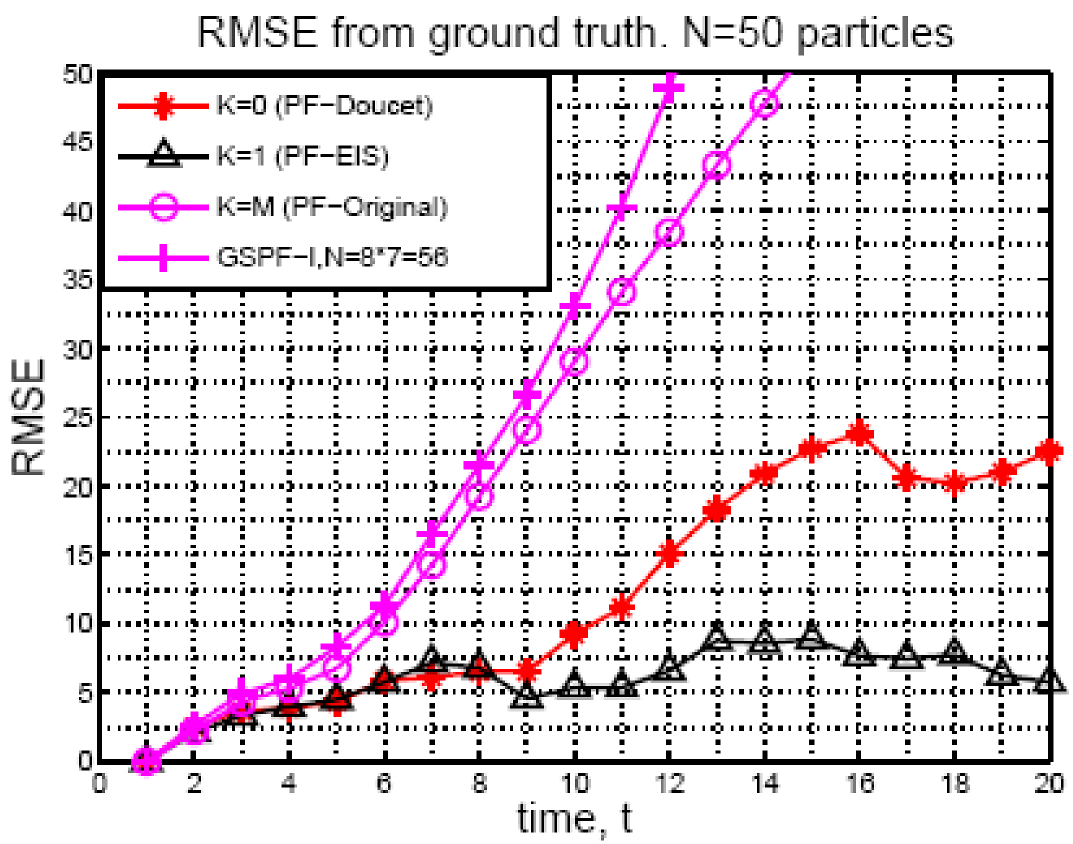
- Select the multimodal state, $X_{t,s}$, using heuristics motivated by the unimodality result
- Split $X_{t,r}$ further into $X_{t,r,s}$, $X_{t,r,r}$ s.t. the conditional posterior of $X_{t,r,r}$ (residual state) is narrow enough to justify IS-MT

PF-EIS-MT algorithm [Vaswani, TSP, Oct'08]

At each t, split $X_t = [\, X_{t,s}\, ,\, X_{t,r,s},\, X_{t,r,r}\,]$ &

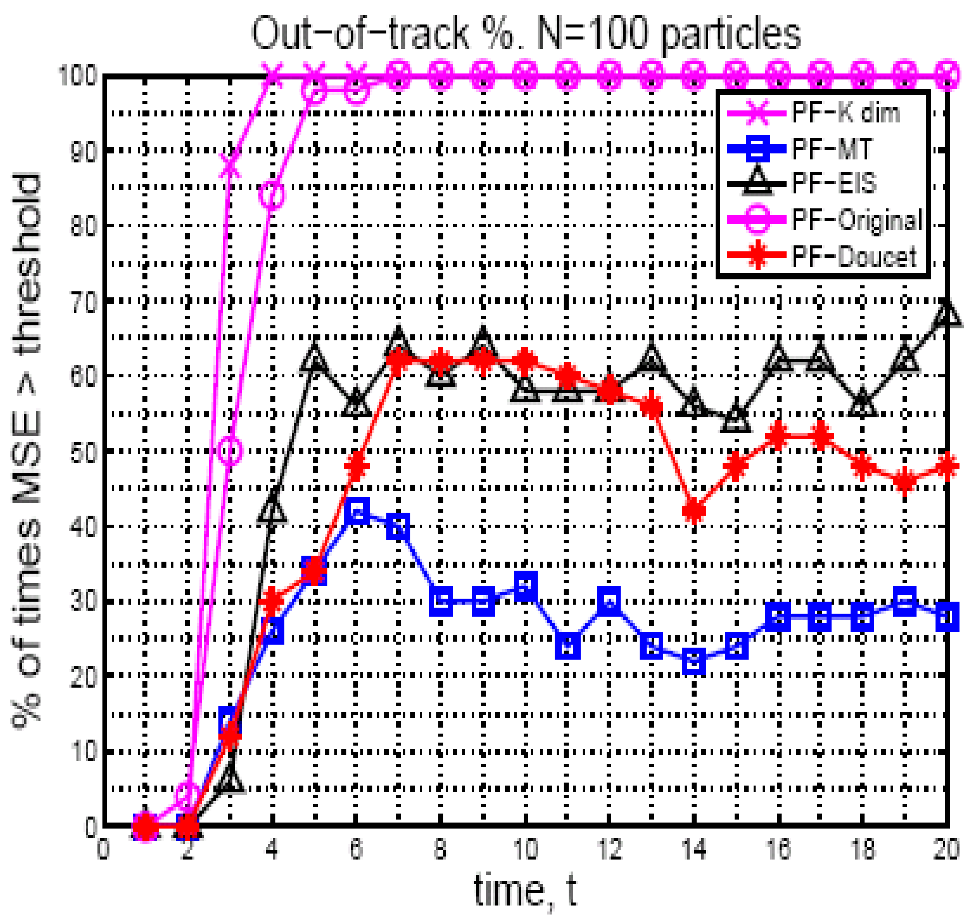
- for each particle, i,
  - sample $x_{t,s}^i$ from its state transition prior
  - compute the conditional posterior mode of $X_{t,r}$
  - sample $x_{t,r,s}^i$ from Gaussian approx about mode
  - compute mode of conditional posterior of $X_{t,r,r}$ and set $x_{t,r,r}^i$ equal to it
  - weight appropriately
- resample

Simulation Results: Sensor failure



- Tracking temperature at M=3 sensor nodes, each with 2 sensors
- Node 1 had much higher failure probability than rest
- PF-EIS: $X_{t,s} = v_{t,1}$
- PF-EIS (black) outperforms PF-D, PF-Original & GSPF

Simulation Results: Sensor failure



- Tracking on M=10 sensor nodes, each with two sensors per node. Node 1 has much higher failure prob than rest
- PF-MT (blue) has least RMSE
  - using K=1 dim multimodal state

- N. Vaswani, Particle Filtering for Large Dimensional State Spaces with Multimodal Observation Likelihoods, IEEE Trans. Signal Processing, Oct 2008

- N. Vaswani, Y. Rathi, A. Yezzi, A. Tannenbaum, Deform PF-MT: Particle Filter with Mode Tracker for Tracking Non-Affine Contour Deformation, IEEE Trans. Image Processing, to appear

- Y. Rathi, N. Vaswani A. Tannenbaum, A. Yezzi, Tracking Deforming Objects using Particle Filtering for Geometric Active Contours, IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI), pp. 1470-1475, August 2007

- S. Das and N. Vaswani, Nonstationary Shape Activities: Dynamic Models for Landmark Shape Change and Applications, IEEE Trans. PAMI, to appear

- A. Kale and N. Vaswani, Generalized ELL for Detecting and Tracking Through Illumination Model Changes, IEEE Intl. Conf. Image Proc. (ICIP), 2008

Open Issues

- Parallel implementations, speed-up posterior mode comp.
- Current conditions for posterior unimodality expensive to verify, depend on previous particles & current observation
  - develop heuristics based on the result to efficiently select multimodal states on-the-fly, or
  - modify the result s.t. unimodality can be checked offline (select multimodal states offline), find states to ensure unimodality w.h.p.
- **Residual space directions usually change over time**
  - How do we select the MT directions on-the-fly?
    - **can we use Compressed Sensing or Kalman filtered CS [Vaswani, ICIP'08] on the state change vector to do this?**
- Analyze the IS-MT approx, prove stability of PF-MT

Deformable Contour Tracking

- State: contour, contour point velocities
- Observation: image intensity and/or edge map
- Likelihood: - exponential of segmentation energies
  - Region based: observation = image intensity
    - Likelihood = probability of image being generated by the contour
    - **Multimodal in case of low contrast images**
  - Edge based: observation = edge locations (edge map)
    - Likelihood = probability of a subset of these edges being generated by the contour; of others being generated by clutter or being missed due to low contrast
    - **Multimodal due to clutter or occlusions or low contrast**

Two proposed PF-MT algorithms

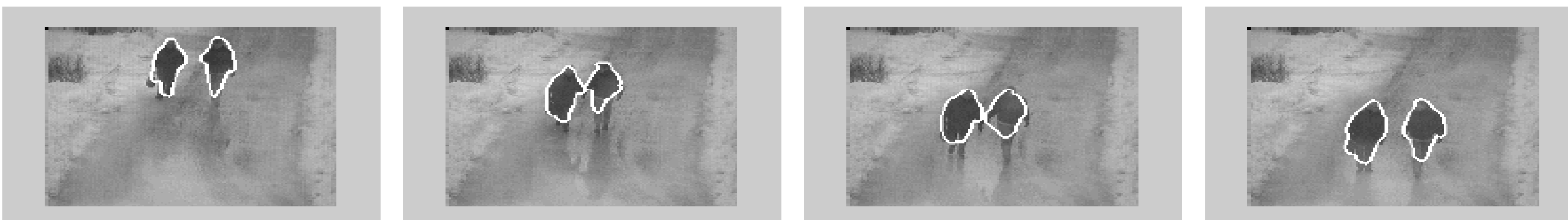
- Affine PF-MT [Rathi et al, CVPR'05, PAMI, Aug'07]
  - Effective basis sp: 6-dim space of affine deformations
  - Assumes OL modes separated only by affine deformation **or** small non-affine deformation per frame
- Deform PF-MT [Vaswani et al, CDC'06, Trans IP (to appear)]
  - Effective basis sp: translation & deformation at K sub-sampled locations around the contour. K can change
  - Useful when OL modes separated by non-affine def (e.g. due to overlapping clutter or low contrast) & large non-affine deformation per frame (fast deforming seq)

Background clutter & occlusions

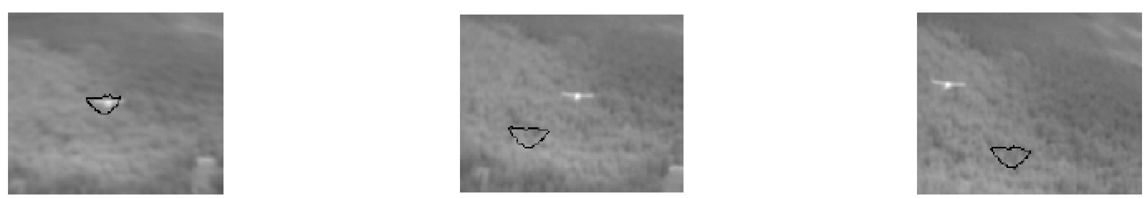
- Need edge based OL: if do not know occluding or background object intensities or if intensities change over the sequence
- 3 dominant modes (many weak modes) of edge based OL due to background clutter
- Overlapping clutter or partial occlusions: OL modes separated by non-affine deformation

Low contrast images, small deformation per frame: use Affine PF-MT

- Tracking humans from a distance (small def per frame)
- Deformation due to perspective camera effects (changing viewpoints), e.g. UAV tracking a plane

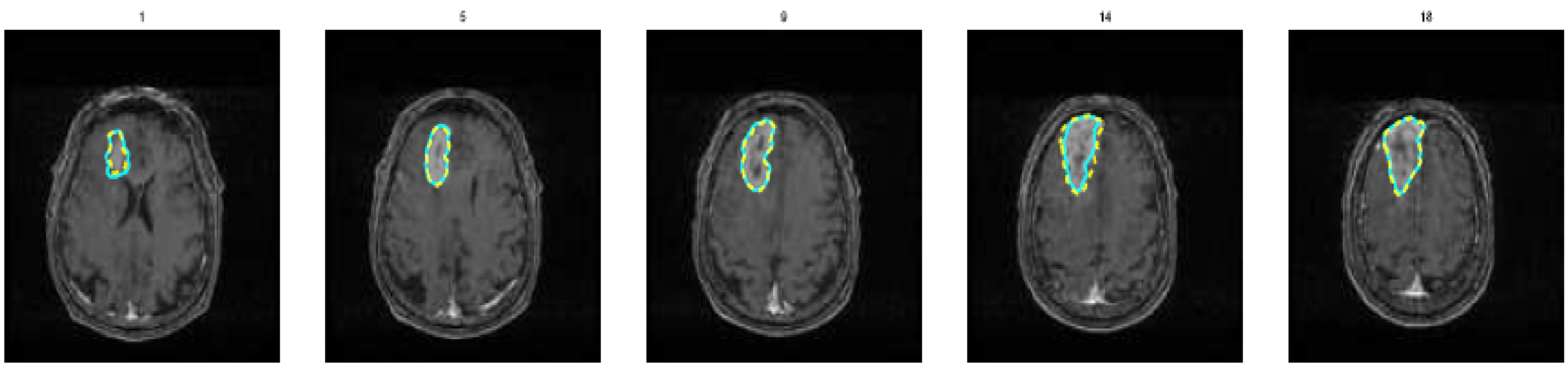


Condensation (PF 6-dim) fails



Low contrast images, large deformation per frame: use Deform PF-MT

- Brain slices, track the tumor sequence
- Multiple nearby likelihood modes of non-affine deformation: due to low contrast



Collaborators

- Deformable contour tracking
  – Anthony Yezzi, Georgia Tech
  – Yogesh Rathi, Georgia Tech
  – Allen Tannenbaum, Georgia Tech
- Illumination tracking
  – Amit Kale, Siemens Corporate Tech, Bangalore
- Landmark shape tracking
  – Ongoing work with my student, Samarjit Das

Summary

- Efficient Importance Sampling techniques that do not require unimodality of optimal IS density
- Derived sufficient conditions to test for posterior unimodality
  - developed for the conditional posterior, $p^{**}(X_{t,r}) := p(X_{t,r} \mid X_{t,s}^i, X_{t-1}^i, Y_t)$
  - used these to guide the choice of multimodal state, $X_{t,s}$, for PF-EIS
- If the state transition prior of a part of $X_{t,r}$ is narrow enough, its conditional posterior will be unimodal & also very narrow
  - approx by a Dirac delta function at its mode: IS-MT
  - improves effective particle size: net reduction in error
- Demonstrated applications in
  - tracking spatially varying physical quantities using unreliable sensors
  - deformable contour tracking, landmark shape tracking, illumination