Introduction to Computer Vision

Michael J. Black

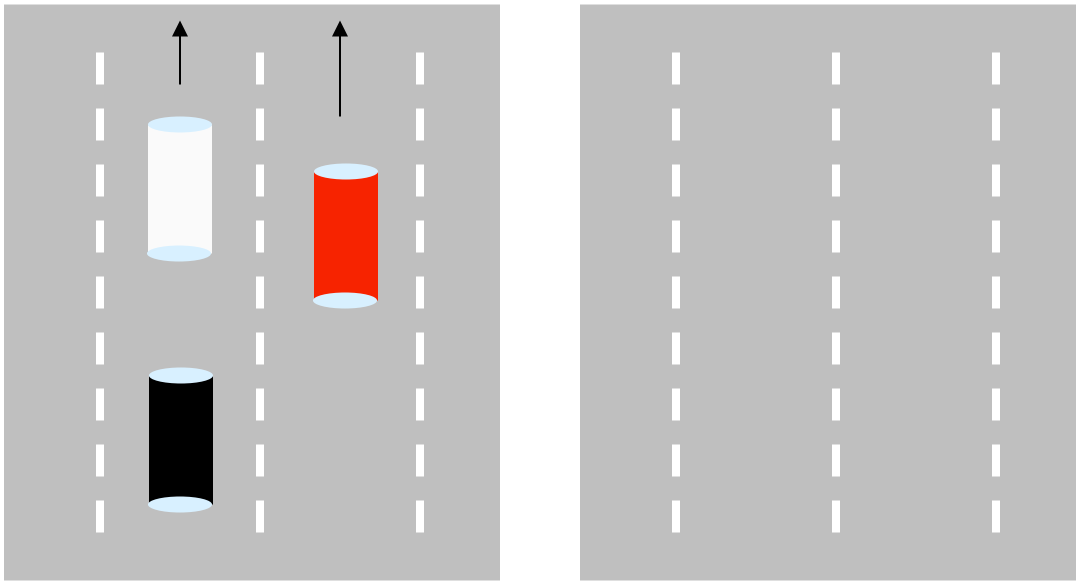
Nov 2009

Tracking and Particle Filtering

Goals

- Today
  - Particle filtering
- Wednesday
  - Binocular stereo
- Friday and beyond
  - Advanced topics – state of the art
  - Short intro to object recognition

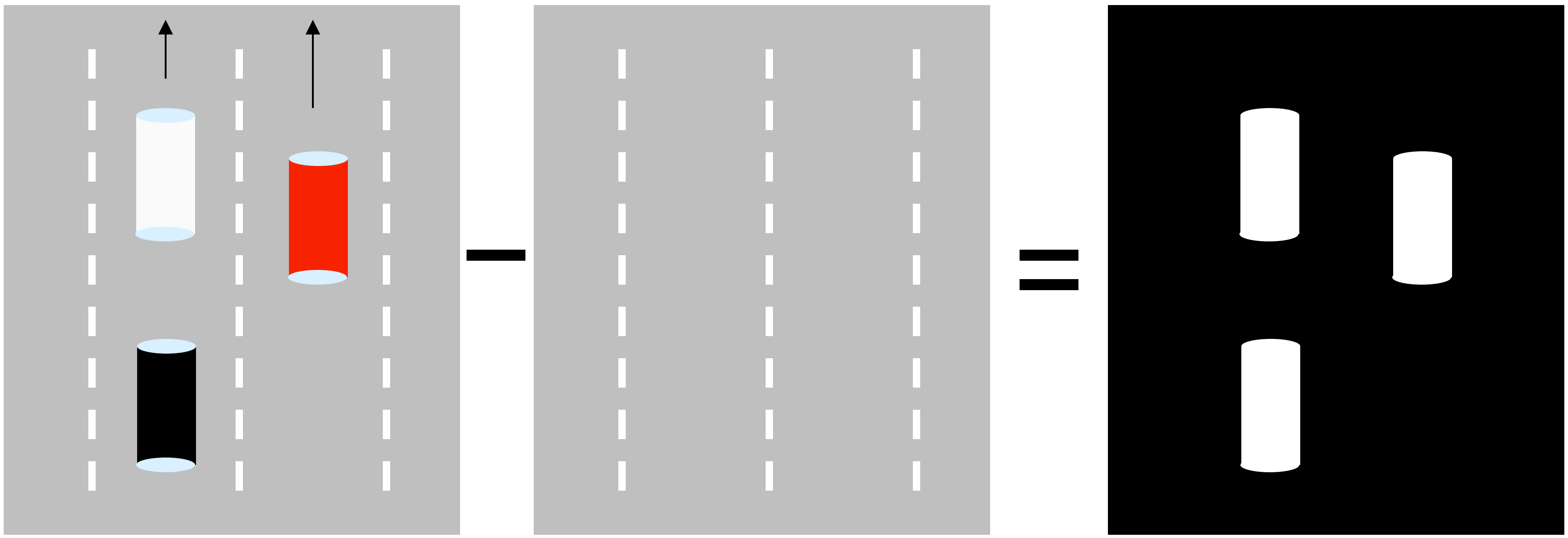
Mathematical Formulation



Goal: estimate car positions at each time instant

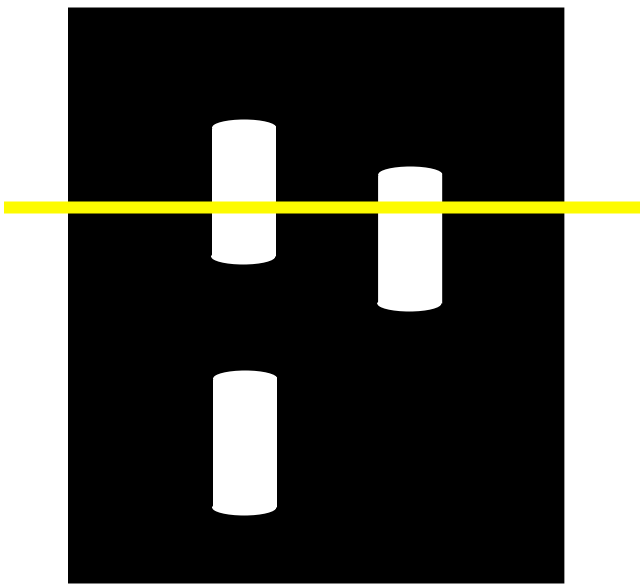
Observations: image sequences and known background

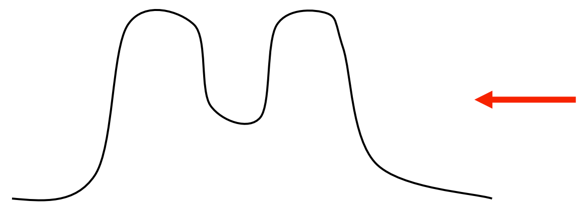
Mathematical Formulation



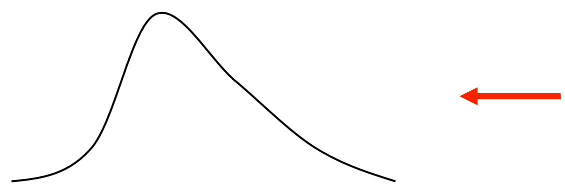
Define image likelihood: p(fg | car=(x,y))

Mathematical Formulation

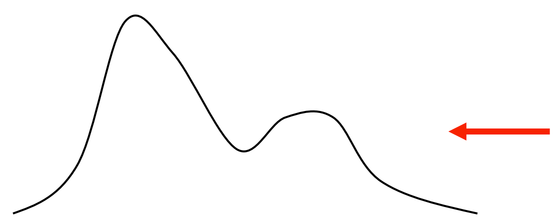




Likelihood:
noisy observation
p(fg | car=(x,y))



Prior:
p(car=(x,y))



Posterior:
Bayesian update
p(car=(x,y)|fg)

system states: car positions

observations: images

Notation

- $\mathbf{x}_k \in \mathbf{R}^d$: internal state at k^{th} frame (hidden random variable, e.g. position of the object in the image).

  $\mathbf{X}_k = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k]^T$: history up to time step k

- $\mathbf{z}_k \in \mathbf{R}^c$: measurement at k^{th} frame (observable random variable, e.g. the given image).

  $\mathbf{Z}_k = [\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_k]^T$:
  history up to time step k

Goal

Estimating the posterior probability $p(\mathbf{x}_k \mid \mathbf{Z}_k)$

How ???

One idea: recursion $\quad p(\mathbf{x}_{k-1} \mid \mathbf{Z}_{k-1}) \;\Rightarrow\; p(\mathbf{x}_k \mid \mathbf{Z}_k)$

- How to realize the recursion ?
- What assumptions are necessary ?

Recursive Formula: Approximation

$$p(\mathbf{x}_k|\mathbf{Z}_k)$$

$$= p(\mathbf{x}_k|\mathbf{Z}_{k-1},\mathbf{z}_k)$$

$$\propto p(\mathbf{z}_k|\mathbf{x}_k,\mathbf{Z}_{k-1})p(\mathbf{x}_k|\mathbf{Z}_{k-1})$$

$$\propto p(\mathbf{z}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{Z}_{k-1})$$

$$\propto p(\mathbf{z}_k|\mathbf{x}_k)\int p(\mathbf{x}_k|\mathbf{x}_{k-1},\mathbf{Z}_{k-1})p(\mathbf{x}_{k-1}|\mathbf{Z}_{k-1})d\mathbf{x}_{k-1}$$

$$\propto p(\mathbf{z}_k|\mathbf{x}_k)\int p(\mathbf{x}_k|\mathbf{x}_{k-1})p(\mathbf{x}_{k-1}|\mathbf{Z}_{k-1})d\mathbf{x}_{k-1}$$

Bayes rule:
$$p(a \mid b) = p(b \mid a)p(a)/p(b)$$

Assumption:
$$p(\mathbf{z}_k \mid \mathbf{x}_k,\mathbf{Z}_{k-1}) = p(\mathbf{z}_k \mid \mathbf{x}_k)$$

Integration:
$$p(a) = \int p(a \mid b)p(b)db$$

Assumption:
$$p(\mathbf{x}_k \mid \mathbf{x}_{k-1},\mathbf{Z}_{k-1}) = p(\mathbf{x}_k \mid \mathbf{x}_{k-1})$$

Bayesian Formulation

$$p(\mathbf{x}_k | \mathbf{Z}_k) = \kappa \, p(\mathbf{z}_k | \mathbf{x}_k) \int p(\mathbf{x}_k | \mathbf{x}_{k-1}) p(\mathbf{x}_{k-1} | \mathbf{Z}_{k-1}) d\mathbf{x}_{k-1}$$

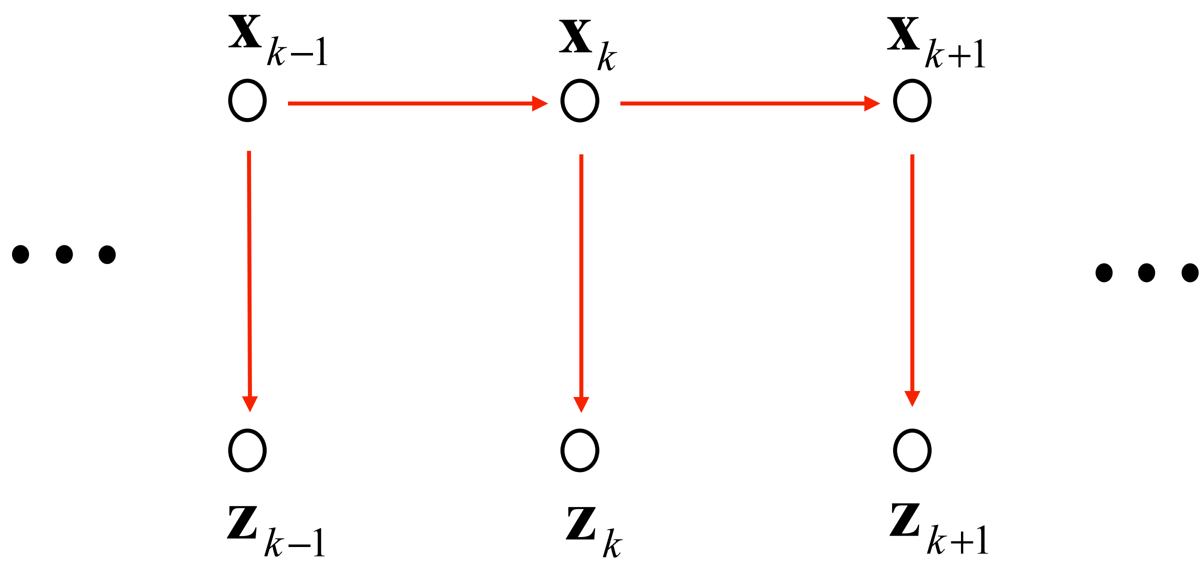
$p(\mathbf{z}_k \mid \mathbf{x}_k):$ likelihood

$p(\mathbf{x}_k \mid \mathbf{x}_{k-1}):$ temporal prior

$p(\mathbf{x}_{k-1} \mid \mathbf{Z}_{k-1}):$ posterior probability at previous time step

$\kappa:$ normalizing term

Bayesian Graphical Model



Assumptions:

$$p(\mathbf{z}_k | \mathbf{x}_k, \mathbf{Z}_{k-1}) = p(\mathbf{z}_k | \mathbf{x}_k), \quad p(\mathbf{x}_k | \mathbf{x}_{k-1}, \mathbf{Z}_{k\text{-}1}) = p(\mathbf{x}_k | \mathbf{x}_{k-1})$$

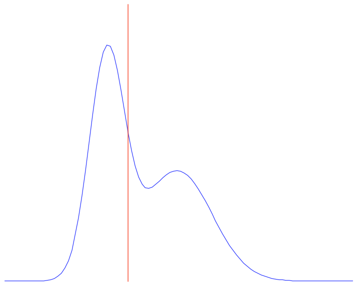
$$p(\mathbf{x}_k \mid \mathbf{X}_{k-1}) = p(\mathbf{x}_k \mid \mathbf{x}_{k-1})$$

Estimators

Assume the posterior probability $p(\mathbf{x}_k \mid \mathbf{Z}_k)$ is known:

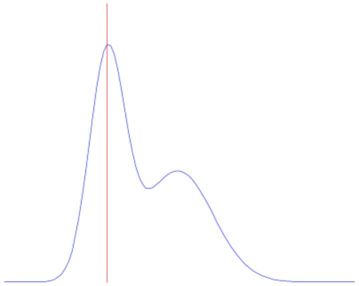
- posterior mean

$$\hat{\mathbf{x}}_k = E(\mathbf{x}_k \mid \mathbf{Z}_k)$$



- maximum *a posteriori* (MAP)

$$\hat{\mathbf{x}}_k = \arg\max p(\mathbf{x}_k \mid \mathbf{Z}_k)$$



$p(\mathbf{x}_k \mid \mathbf{Z}_k)$

General Model

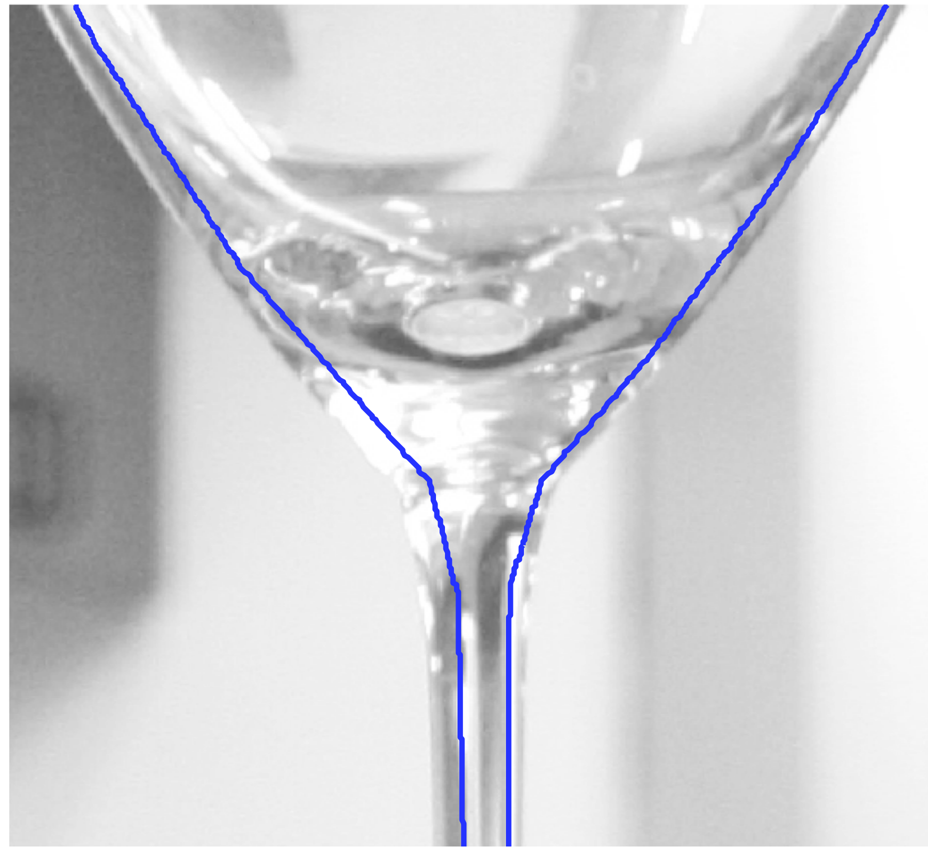
- $p(\mathbf{x}_k \mid \mathbf{Z}_k)$ can be an arbitrary, non-Gaussian, multi-modal distribution.

- The recursive equation has no explicit solution, but can be numerically approximated using Monte Carlo techniques.

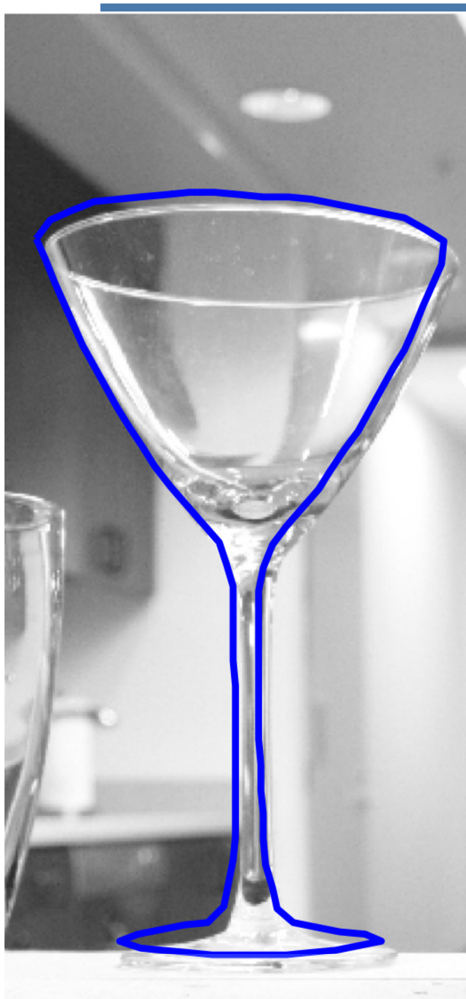
- If both *likelihood* and *prior* are Gaussian, the solution has closed form and the two estimators (posterior mean & MAP) are the same. Such model is known as the Kalman filter. **(Kalman, 1960)**

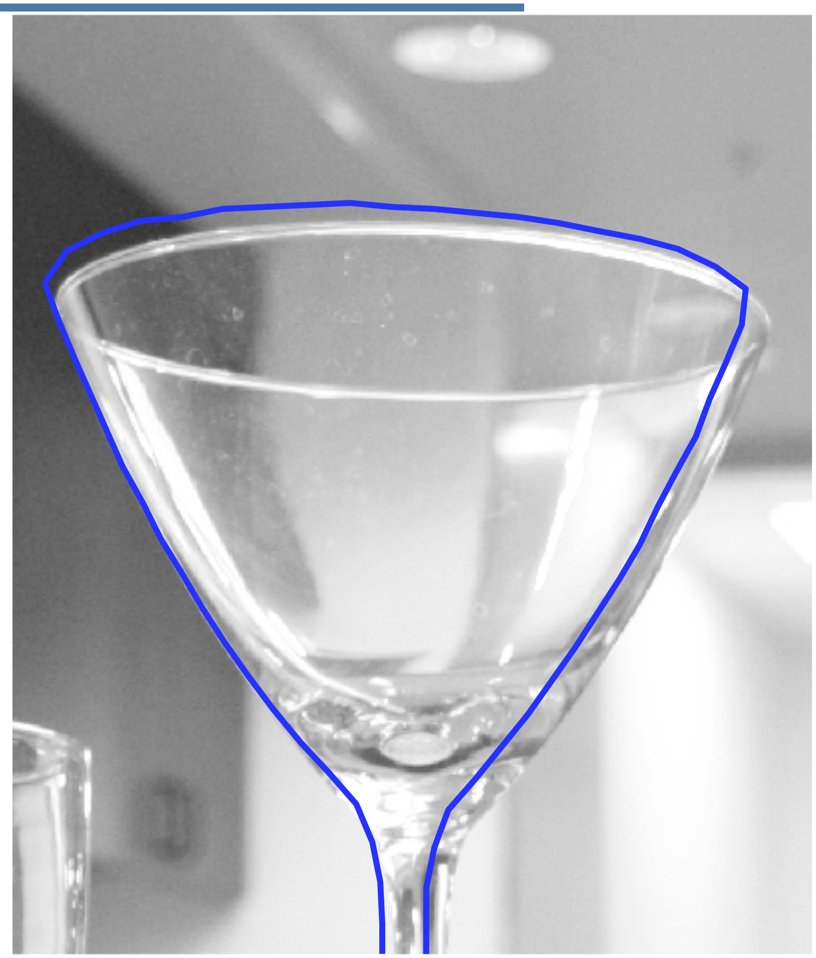
Where's the edge of the glass?

Multi-modal likelihood



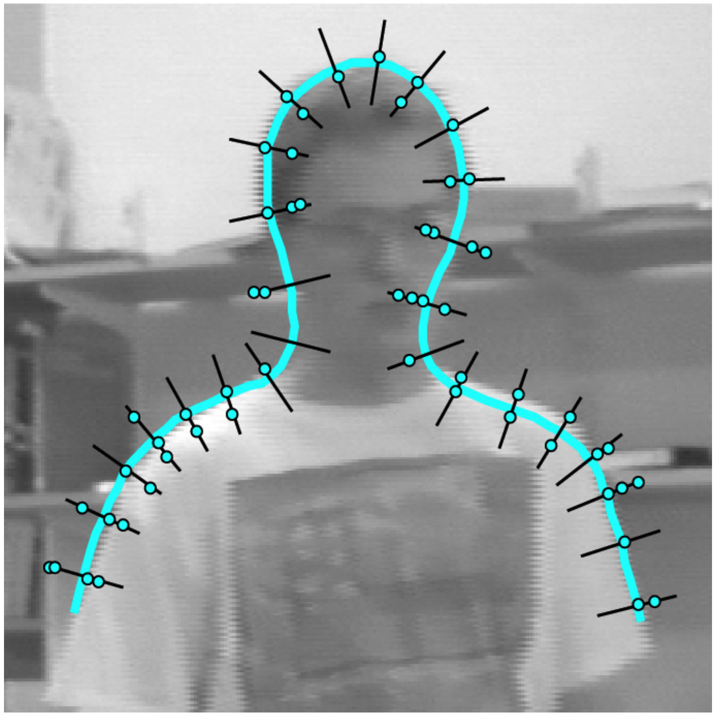
Multi-modal likelihood



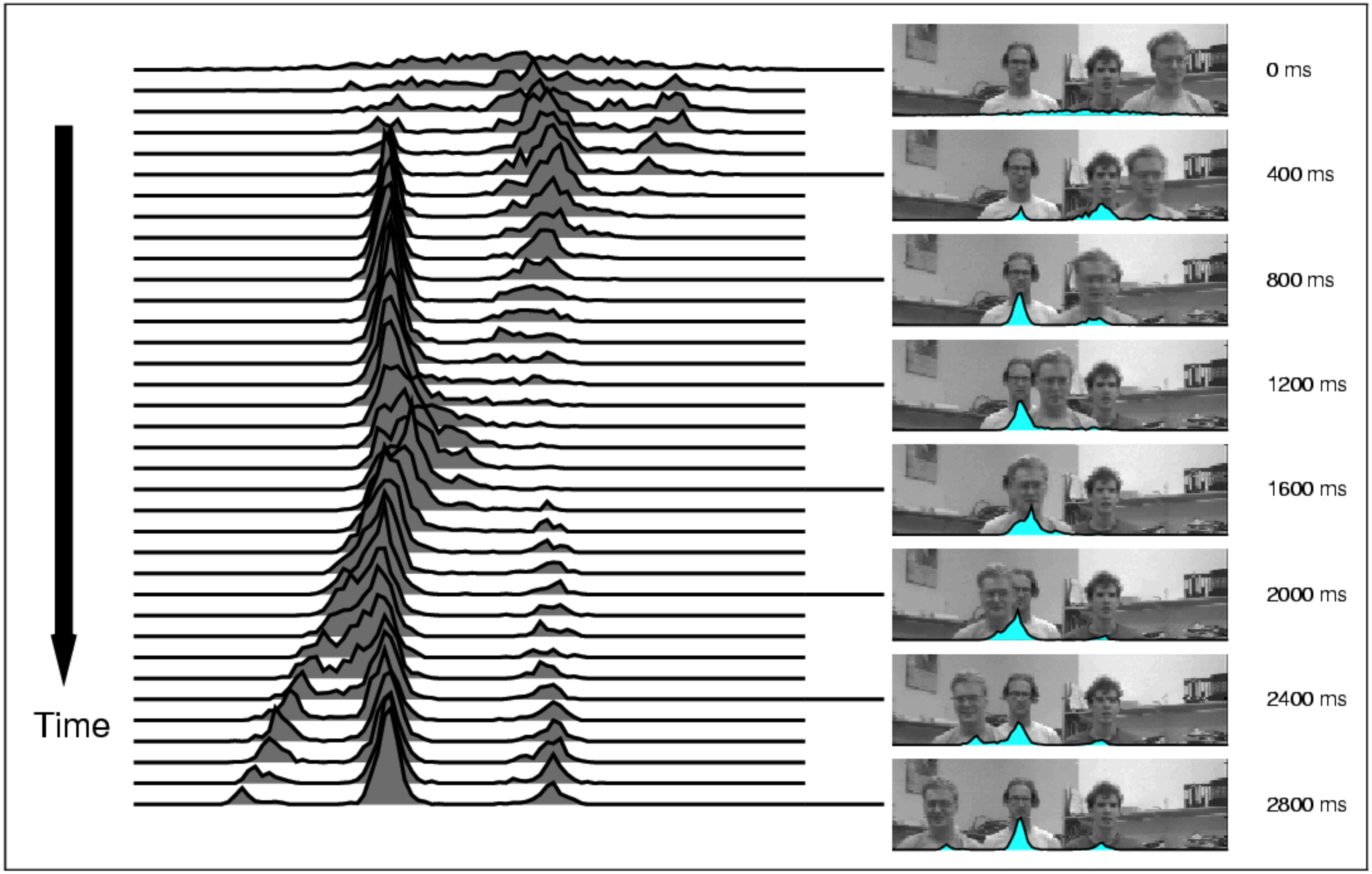


Multi-Modal Likelihoods

Measurement clutter in natural images causes likelihood functions to have multiple, local maxima.

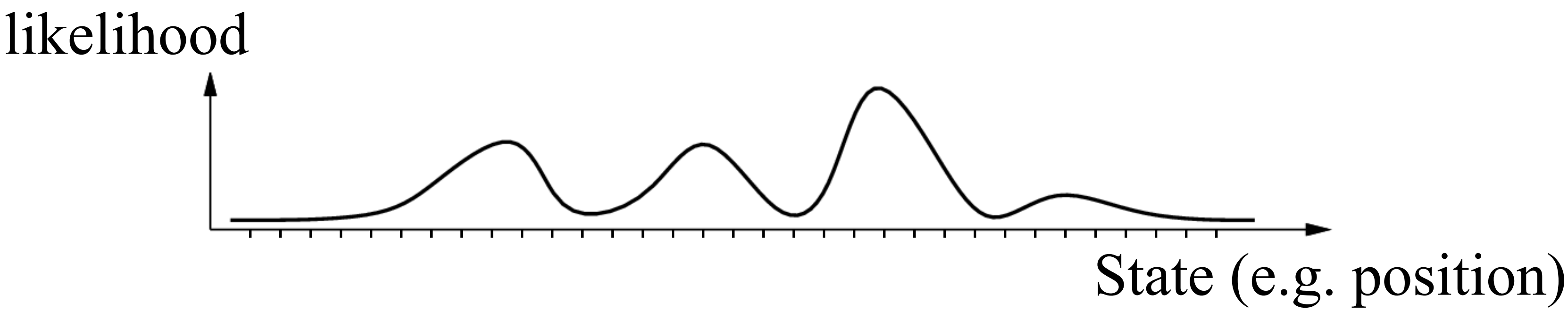


[Isard & Blake, “Condensation - conditional density propagation for visual tracking.” IJCV, 1998]



Michael Isard

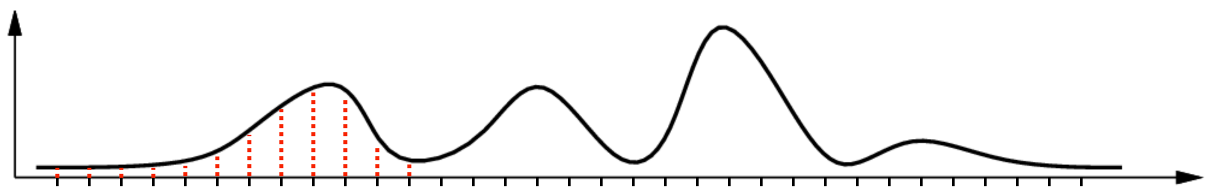
Multi-Modal Likelihood



How can we represent this?

Non-Parametric Approximation

- We could sample at regular intervals



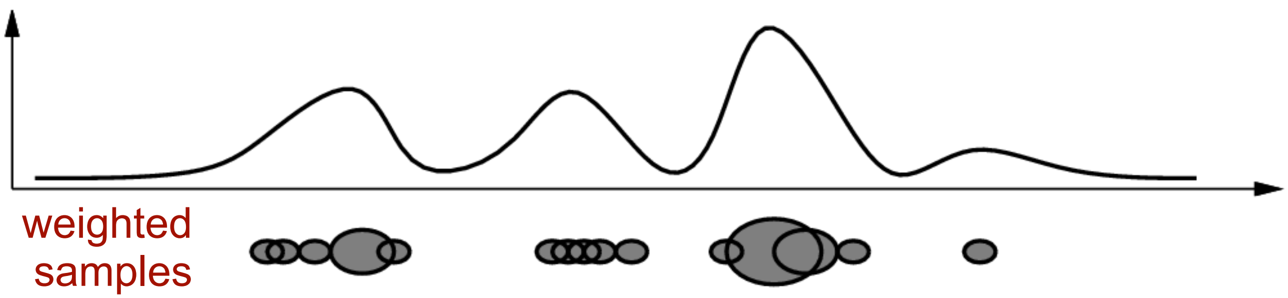
Problems?

most samples have low probability – wasted computation

How finely to discretize

High dimensional space – discretization impractical

Factored Sampling



Weighted samples $S = \{(\mathbf{x}^{(i)}, w^{(i)}); i = 1 \ldots N\}$

Normalized likelihood: $$w_t^{(n)} = \frac{p(\mathbf{z}_t | \mathbf{x}_t^{(n)})}{\sum_{i=1}^{N} p(\mathbf{z}_t | \mathbf{x}_t^{(i)})}$$

Bayesian Tracking

Posterior over model parameters given an image sequence.

$$p(\mathbf{x}_t \mid \mathbf{Z}_t) =$$

Temporal model (prior)

$$\kappa\, p(\mathbf{z}_t \mid \mathbf{x}_t) \int \left(p(\mathbf{x}_t \mid \mathbf{x}_{t-1})\, p(\mathbf{x}_{t-1} \mid \mathbf{Z}_{t-1}) \right) d\mathbf{x}_{t-1}$$

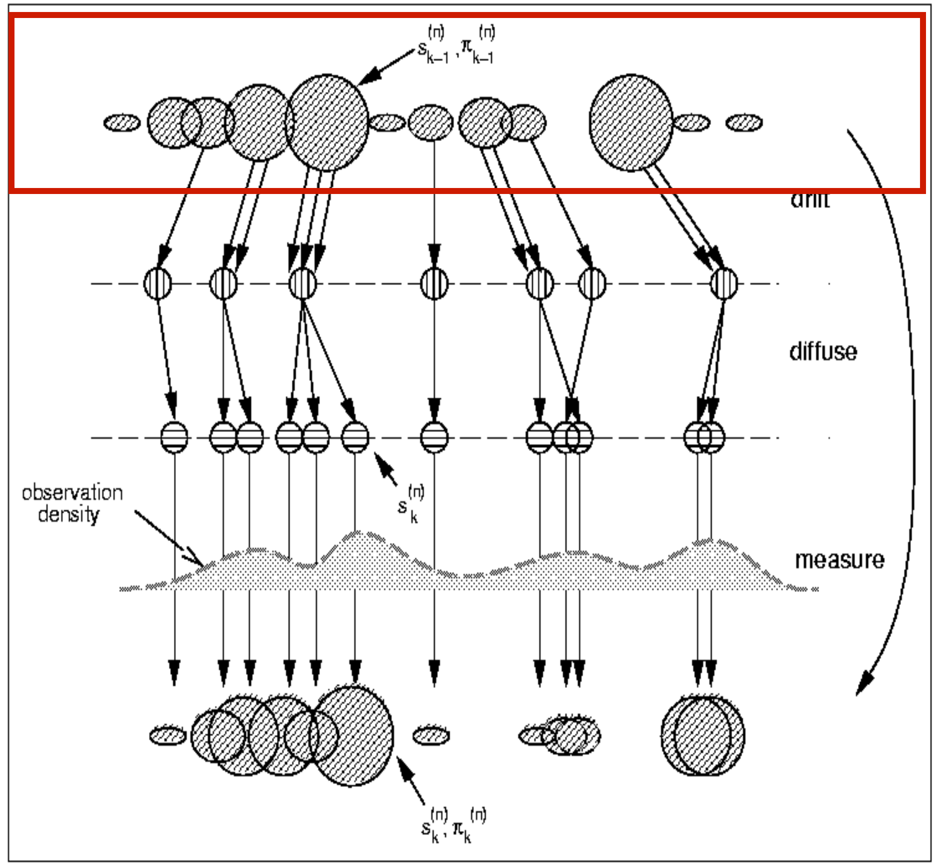
Posterior from previous time instant

Likelihood of observing the image features given the model parameters

Approximate by Monte Carlo sampling

Particle Filter

Posterior $p(\mathbf{x}_{t-1} | \mathbf{Z}_{t-1})$

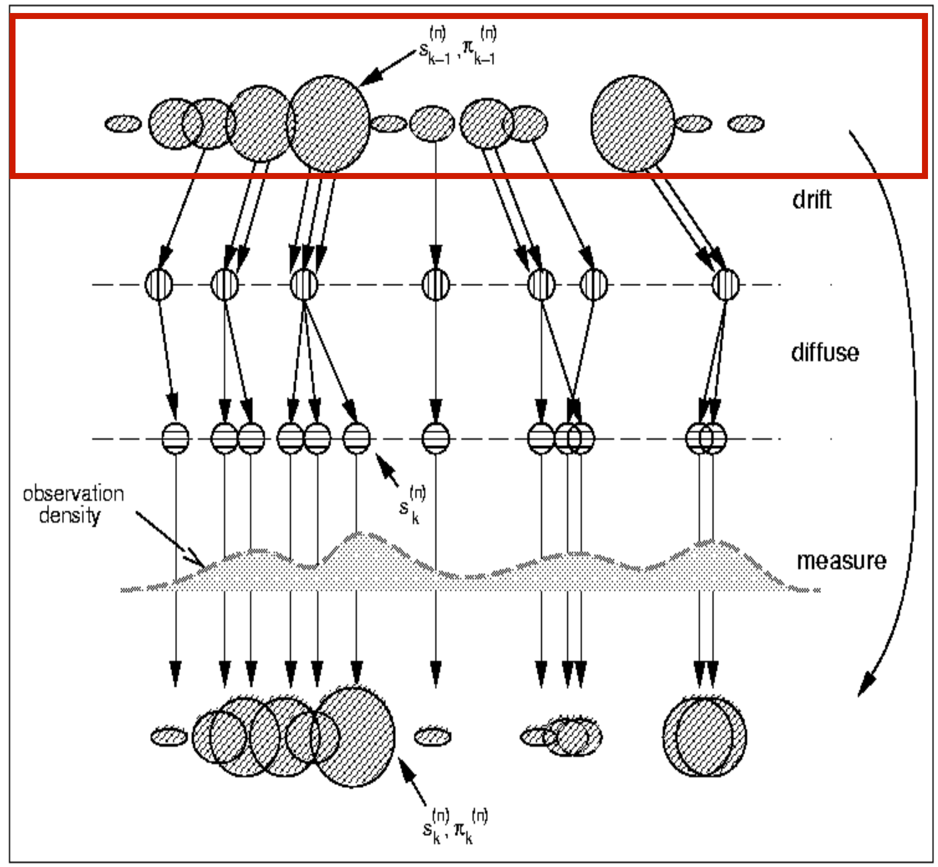


Isard & Blake '96

Particle Filter

Posterior $p(\mathbf{x}_{t-1} | \mathbf{Z}_{t-1})$

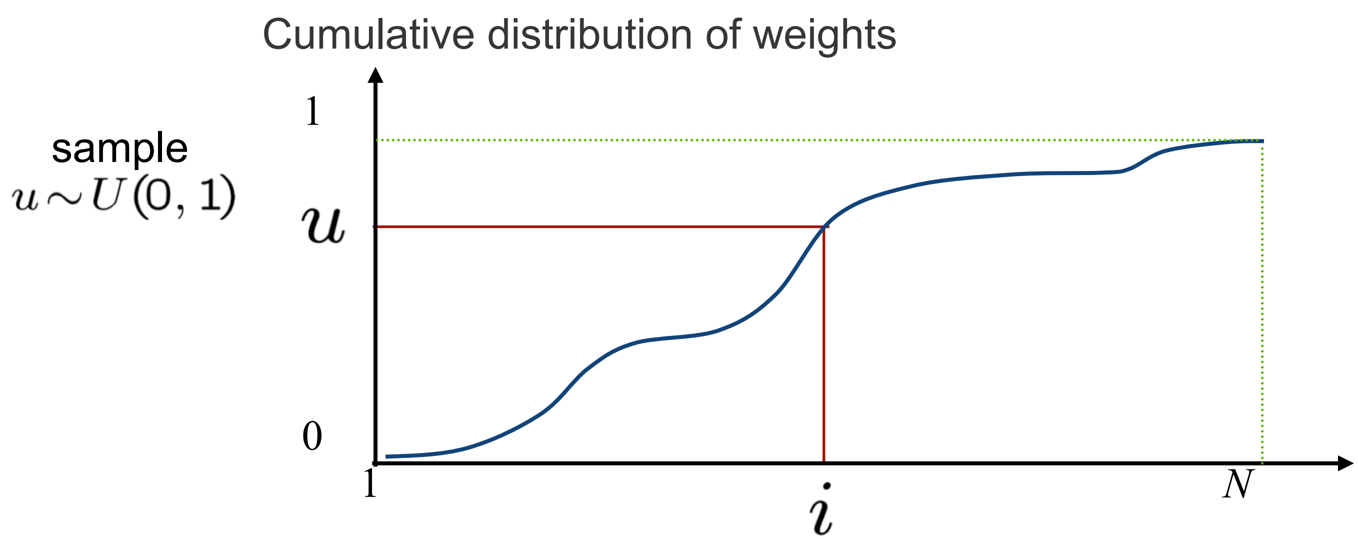
sample



Isard & Blake '96

Monte-Carlo Sampling

Given a weighted sample set $S = \{(\mathbf{x}^{(i)}, w^{(i)});\ i = 1 \ldots N\}$



Particle Filter

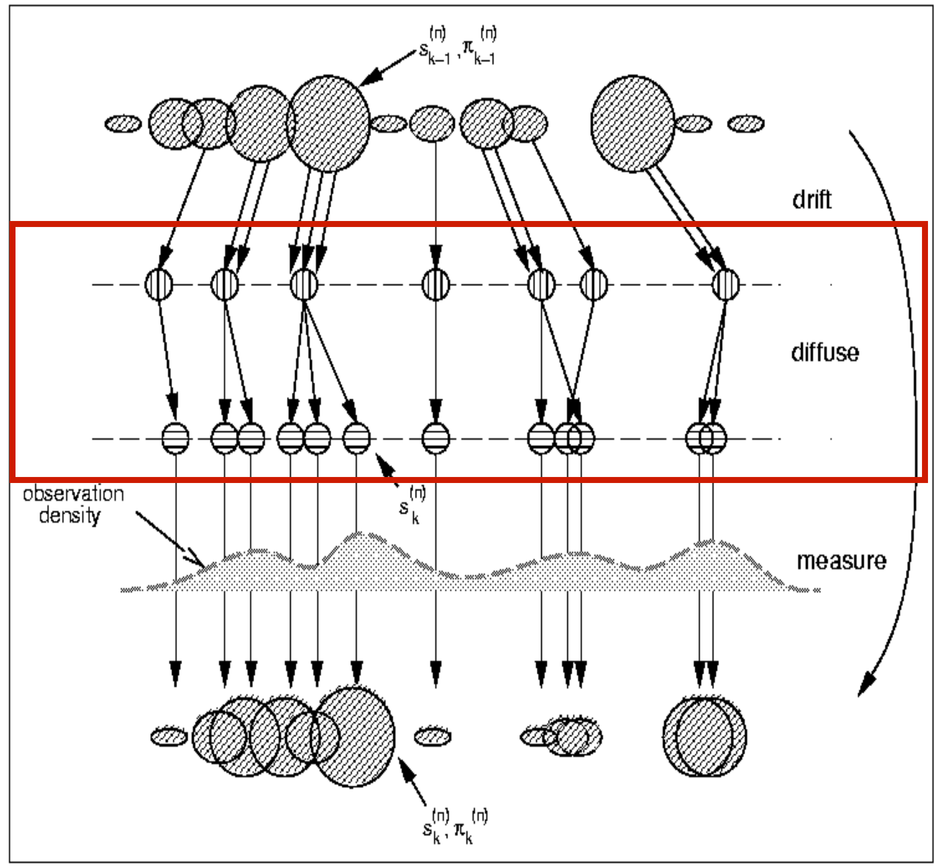
Posterior $p(\mathbf{x}_{t-1} | \mathbf{Z}_{t-1})$

sample

Temporal dynamics

$p(\mathbf{x}_t | \mathbf{x}_{t-1})$

sample (add noise)



Isard & Blake '96

Particle Filter

Posterior $p(\mathbf{x}_{t-1} | \mathbf{Z}_{t-1})$

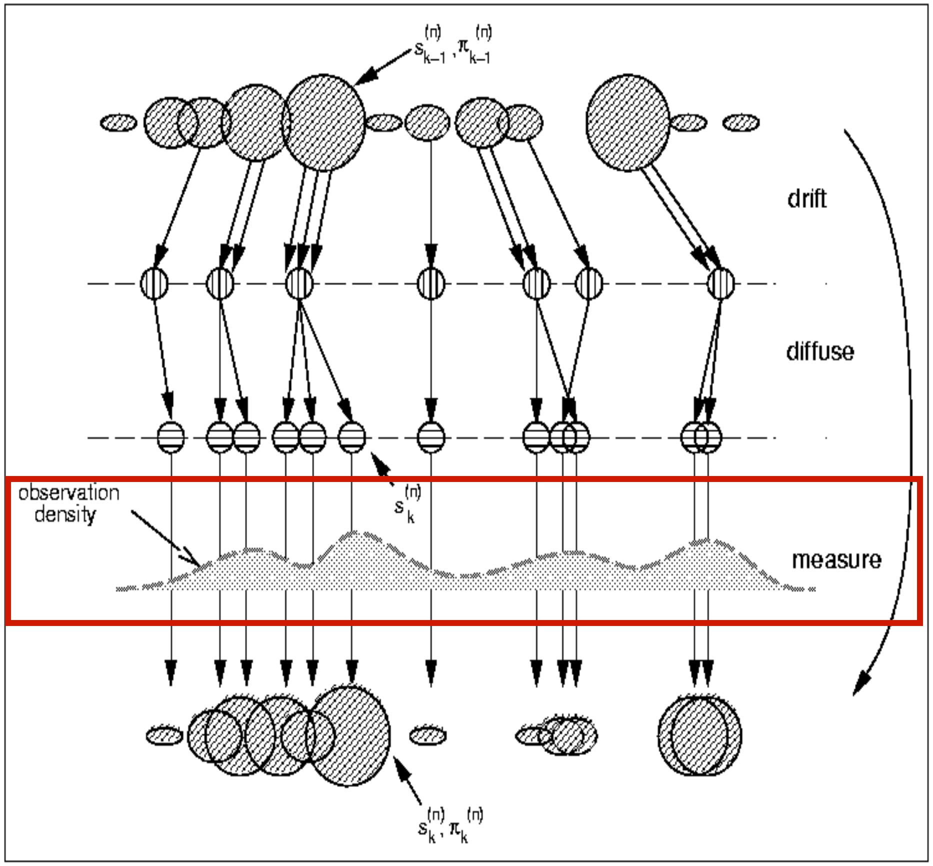
sample

Temporal dynamics

$p(\mathbf{x}_t \mid \mathbf{x}_{t-1})$

sample

Likelihood $p(\mathbf{z}_t | \mathbf{x}_t)$



Isard & Blake '96

Particle Filter

Posterior $p(\mathbf{x}_{t-1} | \mathbf{Z}_{t-1})$

sample

Temporal dynamics

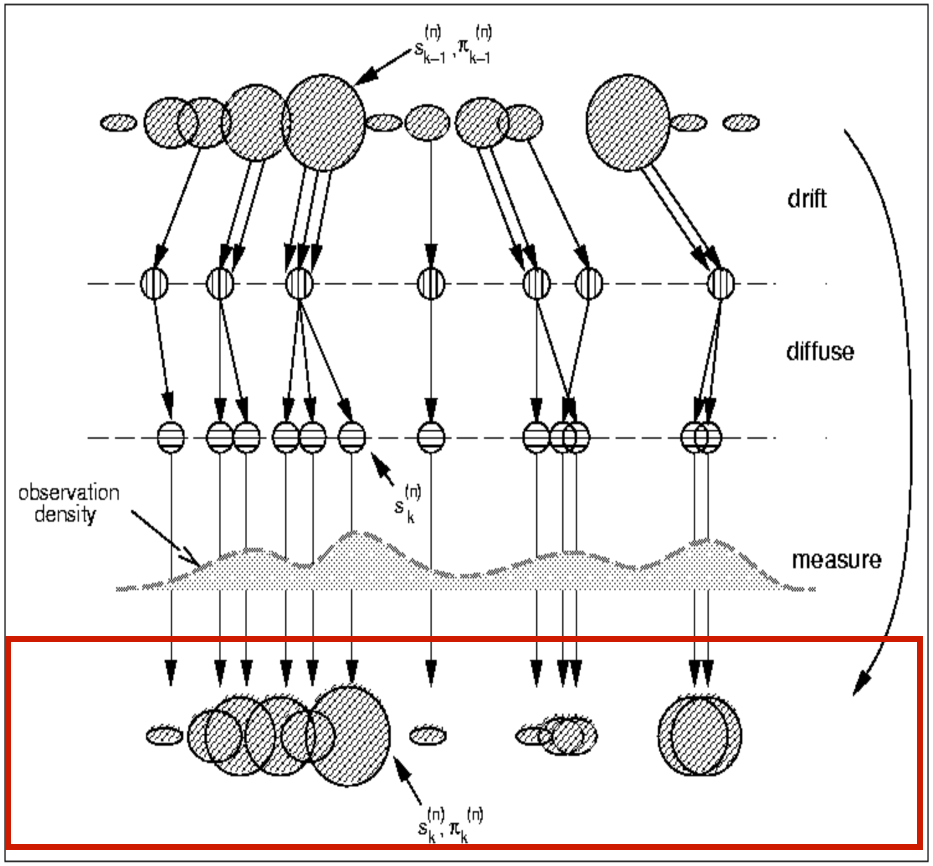
$p(\mathbf{x}_t | \mathbf{x}_{t-1})$

sample

Likelihood $p(\mathbf{z}_t | \mathbf{x}_t)$

normalize

Posterior $p(\mathbf{x}_t | \mathbf{Z}_t)$



Isard & Blake '96

Pseudocode

condense1step

% generate cumulative distribution for posterior at t-1

….

% generate a vector of uniform random numbers.

% if a the number is greater than refreshRate then

- % generate a vector of uniform random numbers
- % use these to search the cumulative probability
- % find the indices of the corresponding particles
- % for each of these particles, predict the new state
- % for each of these new states compute the ***log*** likelihood

% else generate a particle at random and compute its log likelihood.

% find the maximum log likelihood and subtract it from all the other log likelihoods

% construct the posterior at time t by exponentiating all the log likelihoods and normalizing so they sum to 1.

Particle Filter

Michael Isard

Tracking in Clutter

Isard, M., Blake, and A., Condensation — Conditional density propagation for visual tracking, in Int. J. Computer Vision, vol. 28, no. 1, pp. 5–28, 1998