Chapter 4

Bayes Filters

Kalman and Particle Filters

Statistical Methods and Learning in Computer Vision

WS 2012

Claudia Nieuwenhuis
Lehrstuhl für Computer Vision and Pattern Recognition
Fakultät für Informatik
Technische Universität München

Overview

1. **Bayesian Filters**
2. **Kalman Filters**
3. **Particle Filter**

Overview

1. **Bayesian Filters**
2. Kalman Filters
3. Particle Filter

Objective

Task: Prediction of successive states of one or several objects given some kind of control input and measurements for each state.

For example: We want to automatically park a car. We only approximately know its initial location. We have sensor information (measurements) indicating the distance to the nearest obstacles and control commands which move the car. The sensor information contain errors. The location of the car after moving is not exactly known. So we have uncertainties in the measurements and in the actions. We want to estimate a probability distribution over all possible locations (states) of the car.

Objective

We have:

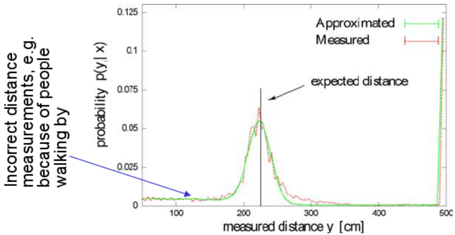
- A system whose state cannot be measured directly (the car's location)
- Measurements of the system (obstacle distances)
- Control commands which influence the system (steering commands)
- Uncertainty in mesurements and actions

We search for:
The current state of the system (the location of the car) if all measurements and actions in previous time steps are known.

Defitions

- State: x_k
  State of the system at time step k
- Measurement z_k
  Measurement of specific aspects of the system at time step k
- Control commands u_k
  Actions or control commands given to the system at time step k

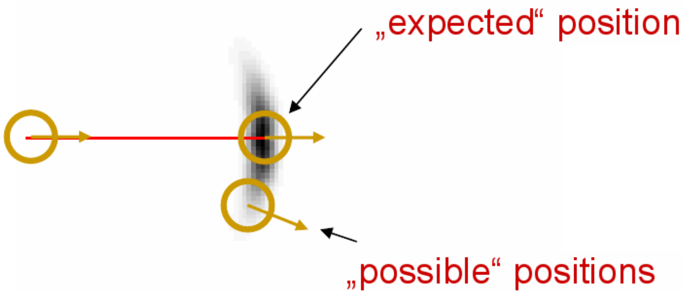
Measurement Model



Example: *sensor model of a distance sensor (laser)*

- **vertical line:** actual distance to the obstacle
- **read line:** measure distances (probability density - PD)
- **green line**: approximated PD for measurements to be expected ➔ distribution is described by the **sensor model**

Motion Model



Dynamic systems can be described by two equations.

State Transition Equation

The system is a Markov Process, which means that the probability of the current state of the system is defined **only** by its previous state x_{k-1}, the last motion control command u_{k-1} and state noise v_{k-1}. f is called the evolution equation, which describes the transition from state x_{k-1} to x_k under uncertainty.

$$x_k = f_k(x_{k-1}, u_{k-1}, v_{k-1})$$

Measurement Equation

The measurement z_k which we obtain at a state x_k depends on this state, the control command u_k and measurement noise n_k under uncertainty.

$$z_k = h_k(x_k, u_k, n_k)$$

Two Assumptions

We assume a first order Markov Process, which means that the current state only depends on the previous state instead of all previous states. For a given $P(x_0)$ that means:

$$P(x_k|x_1, ..., x_{k-1}) = P(x_k|x_{k-1})$$

We assume that the observations z_k are conditionally independent given the state

$$P(z_k|z_{k-1}, x_k) = P(z_k|x_k)$$

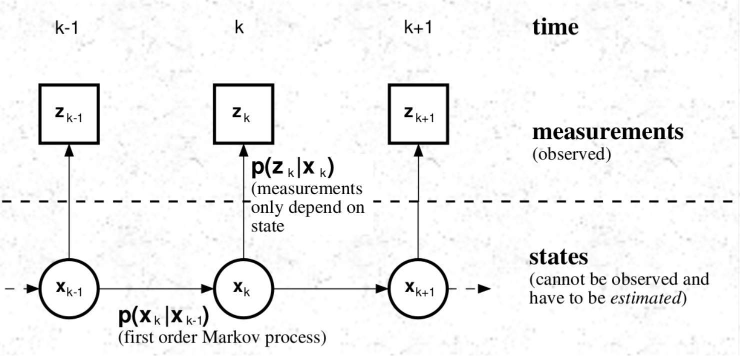
Bayesian Approaches

A Bayesian Approach constructs the posterior probability density function of a state given all measurements.

$$P(x_k|z_1, ..., z_k)$$

Based on this posterior, we can compute expectation values, MAP-estimates, mean values, the modes of the posterior,... to estimate the state of the system.

State Transition Equations



The Posterior Distribution

$$P(x_k|z_1,...,z_k) = \frac{P(z_k|x_k)\;P(x_k|z_1,...,z_{k-1})}{P(z_k|z_1,...,z_{k-1})}$$

The Posterior Distribution

$$P(x_k|z_1,...,z_k) = \frac{P(z_k|x_k)\; P(x_k|z_1,...,z_{k-1})}{P(z_k|z_1,...,z_{k-1})}$$

By the Chapman-Kolmogorov equation we obtain for the prior

$$P(x_k|z_1,...,z_{k-1}) = \int P(x_k|x_{k-1})\; P(x_{k-1}|z_1,...,z_{k-1})\; dx_{k-1}$$

The Posterior Distribution

$$P(x_k|z_1,...,z_k) = \frac{P(z_k|x_k)\ P(x_k|z_1,...,z_{k-1})}{P(z_k|z_1,...,z_{k-1})}$$

By the Chapman-Kolmogorov equation we obtain for the prior

$$P(x_k|z_1,...,z_{k-1}) = \int P(x_k|x_{k-1})\ P(x_{k-1}|z_1,...,z_{k-1})\ dx_{k-1}$$

We finally obtain the posterior probability

$$P(x_k|z_1,...,z_k) = \frac{P(z_k|x_k) \int P(x_k|x_{k-1})\ P(x_{k-1}|z_1,...,z_{k-1})\ dx_{k-1}}{P(z_k|z_1,...,z_{k-1})}$$

The Posterior Distribution

$$P(x_k|z_1,...,z_k) = \frac{P(z_k|x_k)\ P(x_k|z_1,...,z_{k-1})}{P(z_k|z_1,...,z_{k-1})}$$

By the Chapman-Kolmogorov equation we obtain for the prior

$$P(x_k|z_1,...,z_{k-1}) = \int P(x_k|x_{k-1})\ P(x_{k-1}|z_1,...,z_{k-1})\ dx_{k-1}$$

We finally obtain the posterior probability

$$P(x_k|z_1,...,z_k) = \frac{P(z_k|x_k) \int P(x_k|x_{k-1})\ P(x_{k-1}|z_1,...,z_{k-1})\ dx_{k-1}}{P(z_k|z_1,...,z_{k-1})}$$

The posterior probability consists of

- $P(z_k|x_k)$: given by measurement model
- $P(x_k|x_{k-1})$: given by motion model
- $P(x_{k-1}|z_1,...,z_{k-1})$: posterior distribution from previous time step

Representing the Probability Distribution

$$P(x_k|z_1,...,z_k) = \frac{P(z_k|x_k) \int P(x_k|x_{k-1}) \, P(x_{k-1}|z_1,...,z_{k-1}) \, dx_{k-1}}{P(z_k|z_1,...,z_{k-1})}$$

Representing the Probability Distribution

$$P(x_k|z_1,...,z_k) = \frac{P(z_k|x_k) \int P(x_k|x_{k-1}) \, P(x_{k-1}|z_1,...,z_{k-1}) \, dx_{k-1}}{P(z_k|z_1,...,z_{k-1})}$$

This is only a theoretical solution, the integrals are not tractable!

Representing the Probability Distribution

$$P(x_k|z_1,...,z_k) = \frac{P(z_k|x_k) \int P(x_k|x_{k-1}) \, P(x_{k-1}|z_1,...,z_{k-1}) \, dx_{k-1}}{P(z_k|z_1,...,z_{k-1})}$$

This is only a theoretical solution, the integrals are not tractable!

Two different concepts:

- Make restricted assumptions on the posterior distribution: Kalman Filter
- Represent the posterior distribution by a set of randomly chosen weighted samples: Particle Filter

Bayes Filters

The different implementations of Bayes Filters strongly differ in their representation of the posterior probability $P(x_k|z_1, ..., z_k)$.

Continuous Representation

- Kalman Filter (approximates posterior by a single Gaussian, assumes linear system dynamics)
- Extended Kalman Filter (approximates posterior by a single Gaussian, assumes non-linear system dynamics)

Discrete Representation

- Particle Filter (approximates posterior by a set of particles whose density represents the posterior)
- Grid Representation (approximates posterior by a grid with piecewise constant patches)

Kalman Filters

- Most widely used variant of Bayes Filters
- Approximate posterior by Gaussian (which is fully described by its mean and covariance matrix)
- Optimal under the following assumptions
  - Initial pdf $P(x_0)$ is unimodal Gaussian
  - Measurement equation ($z_k = h_k(x_k, u_k, n_k)$) and state transition equation ($x_k = f_k(x_{k-1}, u_{k-1}, v_{k-1})$) are linear with independent Gaussian noise

Advantages
Efficiency: complexity is polynomial in the dimensionality of state space and observations

Disadvantages
Allows to represent unimodal posteriors only $\rightarrow$ restriction to single-object tracking

Particle Filter

- Represent posterior by a set of particles - the more densely distributed the higher the posterior at this location
- Converge to the true posterior even in non-Gaussian, non-linear systems

Advantages

- Ability to represent arbitrary posteriors $\rightarrow$ multi-object tracking
- Very efficient since they focus their particles in regions of space with high likelihood
- Much faster than Kalman Filters (near realtime)
- Much simpler than Kalman Filters

Disadvantages

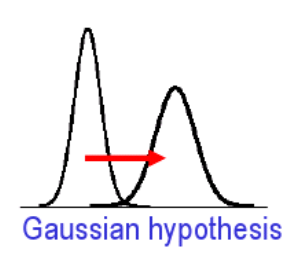
Complexity grows exponentially in the dimension of the state space

Kalman Filter Idea

Kalman filters yield optimal solutions to the state estimation problem, but only if strong assumptions hold.

- The prior distribution $P(x_{k-1}|z_1, ..., z_{k-1})$ is Gaussian.
- State noise v_k and measurement noise n_k are Gaussian.
- The transition equations are linear and thus can be expressed by matrices.

$$\begin{aligned} x_k &= Fx_{k-1} + v_{k-1} \\ z_k &= Hx_k + n_k \end{aligned}$$



$$P(x_k|z_1,...,z_k) = \eta\, P(z_k|x_k)\, P(x_k|z_1,...,z_{k-1})$$

Kalman Filter Assumptions

$$\begin{aligned} x_k &= F x_{k-1} + v_{k-1} \\ z_k &= H x_k + n_k \\ v_k &\sim \mathcal{N}(0, Q) \\ n_k &\sim \mathcal{N}(0, R) \end{aligned}$$

We assume a Gaussian distribution for the prior probability

$$P(x_{k-1}|z_1,...,z_{k-1}) = \mathcal{N}(\mu_{k-1}, \Sigma_{k-1})$$

Then the transition equation effects the mean and covariance of the distribution and we obtain

$$\begin{aligned} P(x_k|z_1,...,z_{k-1}) &= \mathcal{N}(\bar{\mu}_k, \bar{\Sigma}_k) \\ \bar{\mu}_k &= F\mu_{k-1} + 0 \\ \bar{\Sigma}_k &= F\Sigma_{k-1}F^T + Q \end{aligned}$$

$$P(x_k|z_1,\dots,z_k) = \eta\, P(z_k|x_k)\; P(x_k|z_1,\dots,z_{k-1})$$

Kalman Filter Assumptions

$$\begin{aligned} x_k &= F x_{k-1} + v_{k-1} \\ z_k &= H x_k + n_k \\ v_k &\sim \mathcal{N}(0, Q) \\ n_k &\sim \mathcal{N}(0, R) \end{aligned}$$

$$\begin{aligned} P(z_k|x_k) &= \mathcal{N}(Hx_k, R) \\ P(x_k|z_1,\dots,z_k) &= \mathcal{N}(\mu_k, \Sigma_k) \\ \mu_k &= \bar{\mu}_k + K \underbrace{(z_k - H\bar{\mu}_k)}_{\text{estimation error of } z_k} \\ \Sigma_k &= \bar{\Sigma}_k - KH\bar{\Sigma}_k \\ K &= \bar{\Sigma}_k H^T (\underbrace{H\bar{\Sigma}_k H^T + R}_{Cov(z_k)})^{-1} \end{aligned}$$

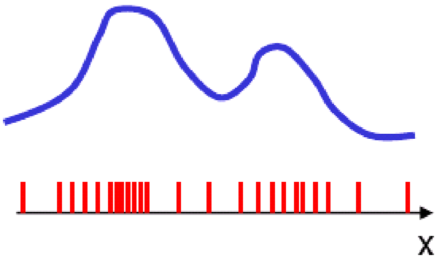
Particle Filter

$$P(x_k|z_1,...,z_k) = \eta P(z_k|x_k) \int P(x_k|x_{k-1}) P(x_{k-1}|z_1,...,z_{k-1}) dx_{k-1}$$

We assume again that the measurement model $P(z_k|x_k)$ and the motion model $P(x_k|x_{k-1})$ is given.

Particle Filter Idea

The prior distribution $P(x_{k-1}|z_1,...,z_{k-1})$ cannot be computed. For Particle Filtering the idea is to represent it by a large set of samples, which approximate the true distribution.



$$\underbrace{P(x_k|z_1,...,z_k)}_{(4)} = \eta \underbrace{P(z_k|x_k)}_{(3)} \int \underbrace{P(x_k|x_{k-1})}_{(2)} \underbrace{P(x_{k-1}|z_1,...,z_{k-1})}_{(1)} dx_{k-1}$$

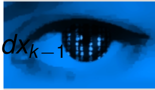
Algorithm

(1) Draw particle $x_{k-1}^{(i)}$ from prior distribution $P(x_{k-1}|z_1,...,z_{k-1})$ of particles

(2) Predict new particle location $x_k^{(i)}$ from motion model $P(x_k|x_{k-1})$

(3) Compute importance factor for each sample $x_k^{(i)}$ from measurement model $P(z_k|x_k)$

(4) Sample from the new particle distribution $\left(x_k^{(i)}\right)_{i=1,...,n}$ based on the particle's importance factors to obtain the posterior distribution $P(x_k|z_1,...,z_k)$



Importance Factor Computation:

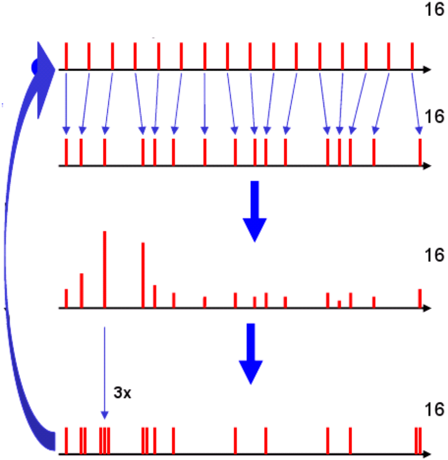
The weight of a particle (its importance factor) is computed according to the measurement model. The higher the probability $P(z_k|x_k)$ of a particle at location x_k for producing the current measurement z_k the higher is its importance weight.

Resampling:

Sample (with replacement!) n times (= number of particles, which stays the same) from the importance factor weighted particle distribution. A particle is chosen with probability of its weight (its importance factor). Particles with heigh weights may be chosen several times as identical copies of the same sample (giving larger weight to the posterior at this location). Particles with low weights may not be chosen and disappear from the distribution. The generated new sample set represents the posterior distribution, which is the prior for the time step $k+1$.

Particle Filter Algorithm

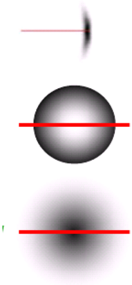
(1) Prior distribution represented by samples, e.g. uniform distribution
(2) The particles move according to the motion model, e.g. Gaussian here
(3) Importance factor computation based on current measurement $P(z_k|x_k)$
(4) Resampling



Multi-Person-Tracking

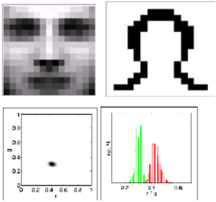
Typical Motion Models

- if both motion distance and direction are known
- if motion distance is known, but direction not
- if motion distance and direction is unknown

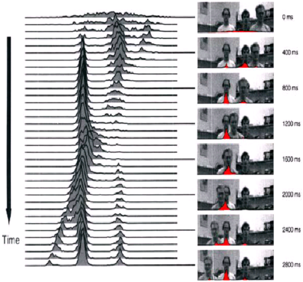


Typical Measurement Models

- Face structure (e.g.Principal Component Analysis)
- Head-shoulder silhouette
- Skin color detector

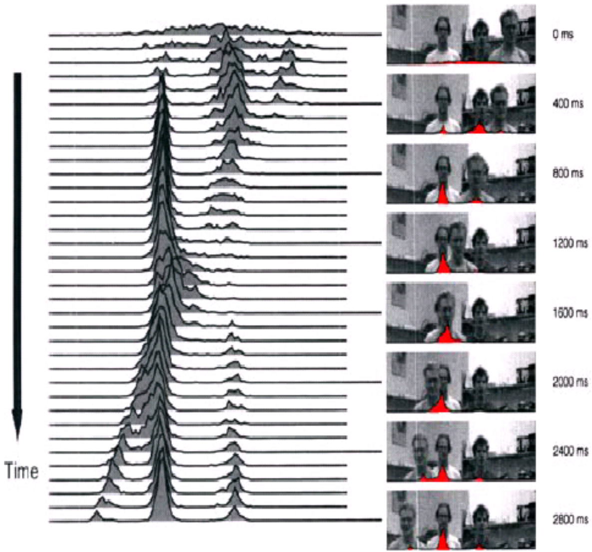


Multi-Person-Tracking

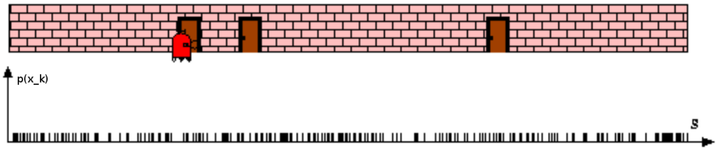


- Shows the projection of the samples onto the horizontal translation axis
- Distribution of $n = 1000$ samples per time-step is used
- Depicts how the state denstiy evolves as tracking progresses
- Initialization with Gaussian at time step 0
- Distribution rapidly collapses down to three peaks - maintained even during temporal occlusion
- Algorithm takes account of motion of all three people, represents multimodal distribution effectively

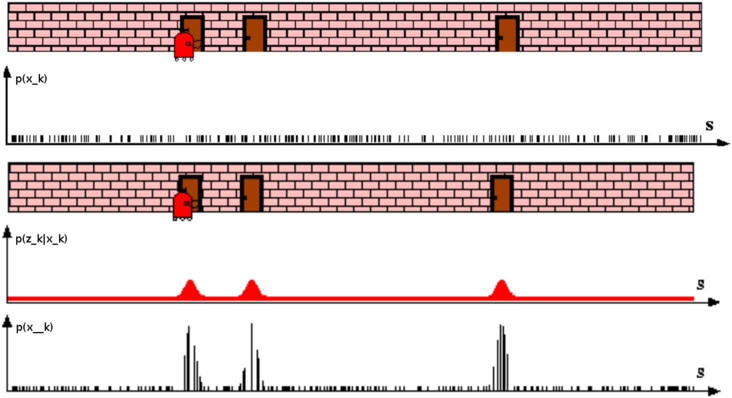
Multi-Person-Tracking



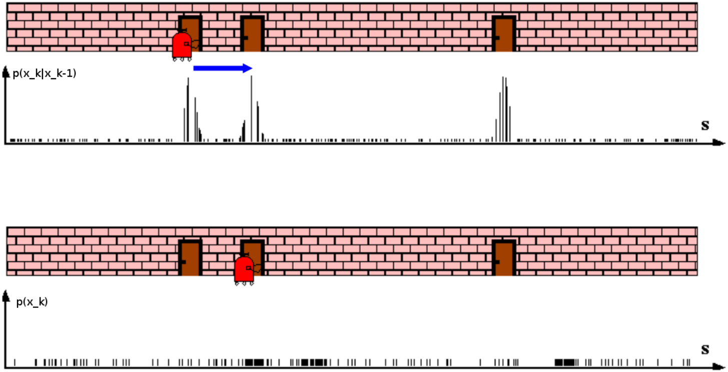
A robot is located in a hall with three doors. It can recognize doors, but it does not know its initial location. At the beginning the robot's position is uncertain over the hall.



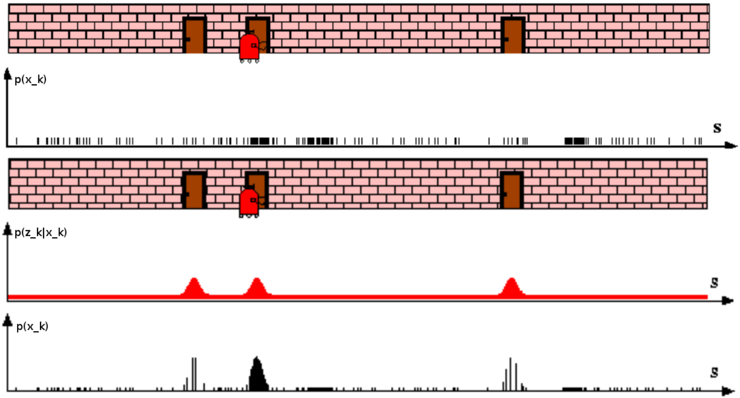
The robot can recognize doors, which is reflected by the sensor model in the middle. Since it knows that it is standing in front of a door, the particles receive stronger weights at the door locations due to the measurement model (importance factors).



The robot moves to the right according to the motion model, so the particles move to the right as well. By means of importance sampling locations with particles with high importance receive more particles.



The sensor model indicates the door locations. Only at the location of the second do the measurement model and a large amount of particles fall together. So the probability is very high that the robot is standing in front of the second door.



The robot moves again to the right and the particles with it.