retour sur innovation

www.onera.fr
Robustness and accuracy in particle filtering

Application to navigation and tracking

Christian Musso (Onera/DTIM)

PhD thesis:

Achille Murangira. Supervision Onera & UTT. Grant DGA/Onera
Paul Bui Quang. Supervision Onera & INRIA. Grant Onera
Outline

• Behavior of the PF in case of multimodality
• Mixture filters
• Application to terrain navigation

• Analysis of the Monte Carlo error with the Laplace method
• Importance density
• Application to tracking
Multimodality: behavior of the PF

In theory, PF deals with multimodality when $N$ is « large » enough.

In practice, we observe mode losses:
- divergence (navigation), loss of a target (multi-target in radar,
  one target with low SNR)

**Cause:** successive resampling

Theoretical analysis in a simple case: bimodal density
Multimodality: behavior of the PF

\[ p = \frac{1}{2} N(-1,0.1) + \frac{1}{2} N(1,0.1) \]

Resampling (with replacement)

\((X_1^1, \ldots, X_N^1)\)
\((X_1^2, \ldots, X_N^2)\)
\((X_1^k, \ldots, X_N^k)\)

How many iterations before losing a mode? Wright Fisher model (genetic)
Multimodality: behavior of the PF

$A_k$: sample size from mode #1

Markov chain:

$$p_{mj} = P(A_k = j / A_{k-1} = m) = C_N^j \left( \frac{m}{N} \right)^j \left( 1 - \frac{m}{N} \right)^{N-j}$$

Absorbtant states: $\{N\}$ and $\{0\}$

Absorption time:

$$\tau = \inf\{k / A_k \in \{0,N\}\}$$
Multimodality: behavior of the PF

Mean time of absorption: \( E_i[\tau] = E[\tau / A_0 = i] \)

Denote \( Z_n = A_n / N \). Computing \( t(x) = E[\tau / Z_0 = x] \) (N large)

\[
\begin{align*}
Z_0 &= x \\
Z_1 &= x + X \\
Z_k &= \{0, 1\}
\end{align*}
\]

Markov: \( t(x) = 1 + E[t(x + X)] \)

With: \( E[X] = 0 \), \( E[X^2] = \frac{x(1-x)}{N} \)

\( N(x + X) \) : binomial distribution

\[
t(x + X) \approx t(X) + t'(x)X + \frac{t''(x)}{2} X^2 \quad \Rightarrow \quad t''(x) = \frac{-2N}{x(1-x)}
\]
Multimodality: behavior of the PF

Expectation time of absorption: \( t(x) \approx -2N[x \log(x) + (1-x) \log(1-x)] \)

Variance time of absorption (x=1/2): \( \left( \frac{2}{3} \pi^2 - 8 \log(2) \right) N^2 + 2 \log(2) N \)

N=100
Multimodality: mixture filters

PF used in the sequel: **regularized PF (RPF)**

\[
\hat{p}_h(x) = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{h^d} K \left( \frac{A^{-1}(x - X_i)}{h} \right)
\]

- Sample \(X_k\) according the weights among \((X_1,\ldots,X_N)\)
- Sample \(\xi\) according to the kernel \(K\)
- Add noise after whitening \(X_k^* = X_k + hA^{-1}\xi\)

Distortion of the multimodal density
Multimodality: mixture filters

Maintaining the modes: mixture of particle filters

\[ p_k = p(X_k | Y_{0:k}) = \sum_{l=1}^{L} \alpha_{l,k} p_l(X_k | Y_{0:k}) \]

where \( \sum_{l=1}^{L} \alpha_{l,k} = 1 \) and \( \alpha_{l,k} \) likelihood of component \( l \)

- **Propagation**
  \[ p_{k+1|k} = p(X_{k+1} | Y_{0:k}) = \sum_{l=1}^{L} \alpha_{l,k+1} p_l(X_{k+1} | Y_{0:k}) \]

- **Correction/regularization**

  \[ p_{k+1} = \sum_{l=1}^{L} \alpha_{l,k+1} p_l(X_{k+1} | Y_{0:k+1}) \quad \alpha_{i,k+1} = \frac{\alpha_{l,k} p_l(Y_{k+1} / Y_k)}{\sum_{j=1}^{L} \alpha_{j,i} p_j(Y_{k+1} / Y_k)} \]

Correction and prediction are done locally
Multimodality: mixture filters

- Mixture component mix \( p_l \) \( \Rightarrow \) estimated by \( \hat{p}_l \) with \( N_l \) particules

\[
\hat{p}_l = \sum_{i=1}^{N_l} \omega_{l,k}^i \delta_{x_{i,k}^l} \quad \forall l = 1...L
\]

\[
\Rightarrow \hat{p}_k^N = \sum_l \alpha_{l,k} \left( \sum_{i=1}^{N_l} \omega_{l,k}^i \delta_{x_{i,k}^l} \right)
\]

- Total number of particules remains unchanged \( \sum_{i=1}^{L} N_l = N \)

- main pros: local resampling for each component of \( \hat{p}_l \)
  \( \Rightarrow \) Maintaining the mode associated insured if \( \alpha_{l,k} > \text{threshold} \)

- In terrain navigation the number of modes evolves with time
  \( \Rightarrow \) Grouping particules to the same mode with clustering algorithm
Multimodality: mixture filters

*Mean shift clustering* used after correction step: grouping particles associated to the same mode.

Meanshift vector

Grouping particles around the mode
Multimodality: mixture filters

- Nb of modes changes: clusterisation with meanshift at each correction
- Nb of particles constant: \( \sum_{l=1}^{L} N_l = N \)
- Threshold on the likelihood of the mode

\[
\hat{p}^N_k = \sum_{l=1}^{L} \alpha_{l,k} \sum_{i=1}^{N_l} \omega_{i,k} \delta_{x_{i,k}} \\
\hat{p}^{N+1}_k = \sum_{l=1}^{L} \alpha_{l,k} \sum_{i=1}^{N_l} \omega_{i,k+1} \delta_{x_{i,k+1}} \\
\hat{p}^N_{k+1} = \sum_{l=1}^{L} \beta_{l,k+1} \sum_{i=1}^{N_l} \nu_{i,k+1} \delta_{x_{i,k+1}}
\]

Prediction / Correction  
Clustering

Actual cost actuel x 2 but alternative in progress (morphological): x 1.3
Application to terrain navigation

• State model
  \[ X_k = f(X_{k-1}) + w_k \]

\[ X_k = (x_k, y_k, z_k, \dot{x}_k, \dot{y}_k, \dot{z}_k)^T \]

• Measurement model
  \[ Y_k = z_k - MNT(x_k, y_k) + v_k = h(X_k) + v_k \]

h : strongly nonlinear

Aim of PF: compute recursively the conditional density \( p(x_k | y_{0:k}) \)

Ambiguous map: multimodal density
Performances in 3 scenarios: Roughness (weak, average, high)

Standard algorithm: RPF (Regularized Particle Filter)

Proposed algorithm: MRPF (Mixture Regularized Particle Filter)
Application to terrain navigation

<table>
<thead>
<tr>
<th>Roughness</th>
<th>MRPF</th>
<th>RPF</th>
</tr>
</thead>
<tbody>
<tr>
<td>weak</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>average</td>
<td>3%</td>
<td>7%</td>
</tr>
<tr>
<td>high</td>
<td>1%</td>
<td>2%</td>
</tr>
</tbody>
</table>

Divergence rate of the filters

- MRPF (Mixture Regularized Particle Filter) converges faster on terrain with high roughness
- MRPF is more robust on ambiguous terrains
  - local redistributions
  - local regularization
- Additional tuning parameters for clustering algorithm
- Reasonable additional computing time when dealing with the same N
Main error: correction step. Analysis of the local error. Large error if weak overlapping likelihood/prediction.

\[ p(x) \propto p(y/x)q(x) = g(x)q(x) \]
Analysis of the Monte Carlo error with the Laplace method

Samping according to \( q \): \((X_1, \ldots, X_N) \sim q\)

Weighting: \( w_i \propto g(X_i) \)

\[ \{w_i, X_i\} : \text{approximate sample of } p \text{ (asymptotically unbiased)} \]

Example: posterior expectation \( E[X / Y] \) estimated by \( \sum_{i=1}^{N} w_i X_i \)

Asymptotic Variance (weight): reliable criterion for the MC error

\[
\frac{1}{N} \frac{\int g^2(x)q(x)dx}{\left(\int g(x)q(x)dx\right)^2} \approx \sum_{i=1}^{N} w_i^2 - \frac{1}{N} \\
\geq 1
\]
Analysis of the Monte Carlo error with the Laplace method

**Laplace approximation** \((\lambda \to +\infty)\)

\[
\int_{\mathbb{R}^d} e^{-\lambda h(x)} \, dx = (2\pi)^{d/2} e^{-\lambda h(\hat{x})} \det [\lambda \nabla^2 h(\hat{x})]^{-1/2} (1 + O(\lambda^{-1}))
\]

\(\hat{x} = \arg \min(h(x))\)

- Global unique minimum, regularity of \(h\)
- Formula related to saddle-point, tilted density
- In general very accurate even if \(\lambda\) small

\[
V = \frac{\int g^2(x)q(x) \, dx}{(\int g(x)q(x) \, dx)^2}
\]
Analysis of the Monte Carlo error with the Laplace method

Thanks to **Laplace approximation**:

$$V = \frac{\int g^2(x)q(x)\,dx}{\left(\int g(x)q(x)\,dx\right)^2} \propto \det \left[ \frac{1}{2} J(\hat{x}) \right]^{1/2}$$

where $\quad J(x) = -(\log g)'(x) - (\log q)'(x)$ is the information matrix a posteriori

and $\hat{x} = \arg \max \{g(x)q(x)\}$ the MAP

**Highly informative model** (small measurement std or high measurement variation)

- Large Monte carlo error
- Accurate Laplace approximation (information plays role of $\lambda$)
Importance density: shifting and scaling

Find an importance function $\tilde{q}$ which reduces the variance

$$V(\tilde{q}) = \int g^2(x) \frac{q^2(x)}{\tilde{q}(x)} dx$$

Approximate the optimal ($V=1$): $\tilde{q}(x) = \frac{g(x)q(x)}{\int g(x)q(x) dx} = p(x)$

By shifting and scaling:

$$q \left[ \frac{x - E(X/Y)}{\sigma(X/Y)} \right]$$

Easy to sample from
Importance density: a method to compute posterior moments

Computing method for posterior expectation

\[ E[X / Y] = \frac{\int x \ g(x) \ q(x) \ dx}{\int g(x) \ q(x) \ dx} = \frac{dm_x}{da}(0) \]

\[ m_x(a) = E[e^{aT X}] = \frac{\int e^{aT x} \ g(x) \ q(x) \ dx}{\int g(x) \ q(x) \ dx} \approx m_x^{Lap}(a) = \frac{e^{aT \tilde{x}} g(\tilde{x}) q(\tilde{x})}{g(\tilde{x}) q(\tilde{x})} \left( \frac{\text{det}[J(\tilde{x})]}{\text{det}[J(\tilde{x})]} \right)^{1/2} \]

\[ \tilde{x} = \tilde{x}(a) = \text{arg max} \ \{e^{aT x} g(x) q(x)\} \quad \hat{x} = \text{arg max} \ \{g(x) q(x)\} \quad \tilde{x}(0) = \hat{x}(0) \]

\[ \frac{dm_x^{Lap}}{da} = \frac{\partial m_x^{Lap}}{\partial \tilde{x}} \frac{d\tilde{x}}{da} + \frac{\partial m_x^{Lap}}{\partial a} \]

1. \[ a + \frac{\partial \log g}{\partial x}(\tilde{x}(a)) + \frac{\partial \log q}{\partial x}(\tilde{x}(a)) = 0 \quad \rightarrow \quad \frac{d\tilde{x}}{da} = -J(\tilde{x}(a))^{-1} \]

2. \[ \frac{\partial \text{det}(J(x))}{\partial x_i} = \text{det}(J(x)) \text{tr}[J^{-1} \frac{\partial J(x)}{\partial x_i}] \]

3. \[ \frac{\partial m_x^{Lap}}{\partial a}(0) = \hat{x} \]
Importance density: a method to compute posterior moments

Computing method for posterior expectation

$$
E[X/Y] = \frac{\int x \, g(x) \, q(x) \, dx}{\int g(x) \, q(x) \, dx}
$$

$$
E[X/Y] \approx \hat{x} - \frac{1}{2} \frac{d}{dx}(\hat{x})
$$

$\left(\begin{array}{c}
tr[J^{-1}(\hat{x}) \frac{\partial J}{\partial x_i}(\hat{x})] \\
tr[J^{-1}(\hat{x}) \frac{\partial J}{\partial x_d}(\hat{x})]
\end{array}\right)
$

Related to skewness

$$
V[X/Y] \approx J^{-1} + \frac{1}{2} J^{-1} \left[ \frac{dJ}{dx} \right]^T (J^{-1} \otimes J^{-1}) \frac{dJ}{dx} J^{-1} + \frac{1}{2} \left( I_d \otimes vec(J^{-1})^T \frac{dJ}{dx} \right) (J^{-1} \otimes J^{-1}) \frac{dJ}{dx} J^{-1} - \frac{1}{2} J^{-1} \left( I_d \otimes vec(J^{-1})^T \right) \frac{d^2J}{dx^2} J^{-1}
$$

Accurate approximation as function of the MAP
Importance density: a method to compute posterior moments

Computing method for posterior expectation, an example: gamma distribution

\[ p(x) = \frac{x^{k-1} e^{\frac{x}{\theta}}}{\Gamma(k) \theta^k} \]

\[ J = -\frac{d^2 \log(p)}{dx^2} = \frac{k-1}{x^2} \]

\[ \frac{dJ}{dx} = -\frac{2(k-1)}{x^3} \]

\[ \frac{d^2J}{dx^2} = \frac{6(k-1)}{x^4} \]

\[ E[X/Y] \approx \hat{x} - \frac{1}{2} J^2(\hat{x}) \frac{dJ}{dx}(\hat{x}) = k\theta \]

\[ V[X/Y] \approx \frac{J^4(\hat{x})}{2} \left( 2J^2 + 2 \left( \frac{dJ}{dx} \right)^2 - J \frac{d^2 J}{dx^2} \right) \]

True values
An other example: triangulation

2 sensors

Prior confidence domain

Measurements

Target

$z_1$

$z_2$
Importance density: a method to compute posterior moments

Gain in terms of accuracy and computing time
Importance density: shifting and scaling the prior

Illustration on a simple example

\[ p(x) = x^2 e^{-\frac{x^2}{2\theta^2}} \]

\[ \tilde{q}(x) = e^{-\frac{(x-\mu)^2}{2\sigma^2}} \]

\[ \mu_{opt} = 0 \]
\[ \sigma_{opt} = \sqrt{3}\theta \]

(coincide with the posterior mean and std).

Asympt. variance = 1.5 (optimal = 1)

Warning: strong degradation of the asympt. var. if \( \sigma < \sigma_{opt} \)
Importance density: Laplace-aided particle filter

Application to PF

\[ p(x) \propto p(y/x)q(x) = g(x)q(x) \]

- Depends on a small number of components of the state vector
- Computed with the predicted particles
- MAP computed on a low-dimensional state space

q a priori: predicted density considered for the MAP as a Gaussian
Application to tracking

Bearing only application with accurate measurements
Application to tracking

\[ N = 3000 \quad \sigma = 0.05^\circ \]

Root Mean Square Error : x-position

Root Mean Square Error : y-position

With Laplace (LRPF): 0% divergence
Without Laplace (RPF) : 12% divergence
Summary

Improved particle filters

• Robustness in case of multimodality

• With a shifted/rescaled importance function