Tracking using CONDENSATION: Conditional Density Propagation


Goal

• Model-based visual tracking in dense clutter at near video frame rates
Example of CONDENSATION Algorithm

Approach

• Probabilistic framework for tracking objects such as curves in clutter using an iterative sampling algorithm
• Model motion and shape of target
• Top-down approach
• Simulation instead of analytic solution
Probabilistic Framework

- Object dynamics form a temporal Markov chain
  \[ p(x_t | X_{t-1}) = p(x_t | x_{t-1}) \]

- Observations, \( z_t \), are independent (mutually and w.r.t process)
  \[ p(Z_{t-1}, x_t | X_{t-1}) = p(x_t | X_{t-1}) \prod_{i=1}^{t-1} p(z_i | x_i) \]

- Use Bayes’ rule

Notation

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>( X )</td>
<td>State vector, e.g., curve's position and orientation</td>
</tr>
<tr>
<td>( Z )</td>
<td>Measurement vector, e.g., image edge locations</td>
</tr>
<tr>
<td>( p(X) )</td>
<td>Prior probability of state vector; summarizes prior domain knowledge, e.g., by independent measurements</td>
</tr>
<tr>
<td>( p(Z) )</td>
<td>Probability of measuring ( Z ); fixed for any given image</td>
</tr>
<tr>
<td>( p(Z</td>
<td>X) )</td>
</tr>
<tr>
<td>( p(X</td>
<td>Z) )</td>
</tr>
</tbody>
</table>
Tracking as Estimation

- Compute state posterior, $p(X|Z)$, and select next state to be the one that maximizes this (Maximum a Posteriori (MAP) estimate)
- Measurements are complex and noisy, so posterior cannot be evaluated in closed form
- Particle filter (iterative sampling) idea: Stochastically approximate the state posterior with a set of $N$ weighted particles, $(s, \pi)$, where $s$ is a sample state and $\pi$ is its weight
- Use Bayes’ rule to compute $p(X|Z)$

Factored Sampling

- Generate a set of samples that approximates the posterior $p(X|Z)$
- Sample set $s=\{s^{(1)},...,s^{(N)}\}$ generated from $p(X)$; each sample has a weight (“probability”)

$$\pi_i = \frac{p_z(s^{(i)})}{\sum_{j=1}^{N} p_z(s^{(j)})}$$

$$p_z(x) = p(z|x)$$
Factored Sampling

- CONDENSATION for one image

Estimating Target State

State samples

Mean of weighted state samples
Bayes’ Rule

\[ p(X | Z) = \frac{p(Z | X) p(X)}{p(Z)} \]

This is what you can evaluate

This is what you want. Knowing \( p(X | Z) \) will tell us what is the most likely state \( X \).

This is what you may know a priori, or what you can predict

This is a constant for a given image

CONDENSATION Algorithm

1. **Select**: Randomly select \( N \) particles from \( \{s_{t-1}^{(n)}\} \) based on weights \( \pi_{t-1}^{(n)} \); same particle may be picked multiple times (*factored sampling*)

2. **Predict**: Move particles according to deterministic dynamics (*drift*), then perturb individually (*diffuse*)

3. **Measure**: Get a likelihood for each new sample by comparing it with the image’s local appearance, i.e., based on \( p(z_t | x_t) \); then update weight accordingly to obtain \( \{(s_t^{(n)}, \pi_t^{(n)})\} \)
Notes on Updating

- Enforcing plausibility: Particles that represent impossible configurations are discarded.
- Diffusion modeled with a Gaussian.
- Likelihood function: Convert “goodness of prediction” score to pseudo-probability.
  - More markings closer to predicted markings → higher likelihood.
Object Motion Model

- For video tracking we need a way to propagate probability densities, so we need a “motion model” such as
  \[ X_{t+1} = A X_t + B W_t \]
  where \( W \) is a noise term and \( A \) and \( B \) are state transition matrices that can be learned from training sequences.
- The state, \( X \), of an object, e.g., a B-spline curve, can be represented as a point in a 6D state space of possible 2D affine transformations of the object.

Evaluating \( p(Z \mid X) \)

\[
p(z \mid x) = q p(z \mid \text{clutter}) + \sum_{m=1}^{M} p(z \mid x, \phi_m) p(\phi_m)
\]

where \( \phi_m = \{ \text{true measurement is } z_m \} \)
  for \( m = 1, \ldots, M \), and \( q = 1 - \sum_m p(\phi_m) \) is the probability that the target is not visible.

\[
\phi_m = \begin{cases} 
  \left| x_m - z_m \right|^2 & \text{if } \left| x_m - z_m \right| < \delta \\
  \rho & \text{otherwise}
\end{cases}
\]
Dancing Example

Hand Example
Pointing Hand Example

Glasses Example

- 6D state space of affine transformations of a spline curve
- Edge detector applied along normals to the spline
- Autoregressive motion model
3D Model-based Example

• 3D state space: image position + angle
• Polyhedral model of object

Minerva

• Museum tour guide robot that used CONDENSATION to track its position in the museum
Advantages of Particle Filtering

• Nonlinear dynamics, measurement model easily incorporated
• Copes with lots of false positives
• Multi-modal posterior okay (unlike Kalman filter)
• Multiple samples provides multiple hypotheses
• Fast and simple to implement