Visual Tracking 101

Ying Wu

Electrical Engineering and Computer Science
Northwestern University, Evanston, IL 60208

yingwu@northwestern.edu

http://www.eecs.northwestern.edu/~yingwu
Outline

Getting Start

What is Tracking
  A Probabilistic Framework
  An Example of Observation Model
  Tracking as Probability Propagation

Recursive Estimation and Kalman Filtering
  Kalman Filtering for Linear Dynamic Systems

Sequential Monte Carlo and Particle Filtering
  Factored Sampling
  Importance Sampling
Template Matching

- Assume the target is an image template $T(x, y)$
- Need to match this template over consecutive frames.
- Define good **match**
  - sum-of-squared-difference (SSD)
  - cross-correlation
  - normalized cross-correlation (NCC)
- Perform good **search**
  - exhaustive search
  - gradient-based search
  - random search
- Is this tracking?
Outline

Getting Start

What is Tracking
  A Probabilistic Framework
  An Example of Observation Model
  Tracking as Probability Propagation

Recursive Estimation and Kalman Filtering
  Kalman Filtering for Linear Dynamic Systems

Sequential Monte Carlo and Particle Filtering
  Factored Sampling
  Importance Sampling
A Dynamic System

- The states of the target \( X_t \)
- Its visual observation (or measurement) \( Z_t \)
- The history of the states \( \underline{X}_t = (X_1, \ldots, X_t) \)
- The history of the measurements \( \underline{Z}_t = (Z_1, \ldots, Z_t) \).
- Tracking is a Bayesian Inference problem:

\[
 p(X_{t+1}|\underline{Z}_{t+1}) \propto p(Z_{t+1}|X_{t+1})p(X_{t+1}|\underline{Z}_t) \tag{1}
\]

\[
 p(X_{t+1}|\underline{Z}_t) = \int p(X_{t+1}|X_t)p(X_t|\underline{Z}_t)dX_t \tag{2}
\]

- \( p(X_{t+1}|\underline{Z}_t) \) the prediction prior
- \( p(Z_{t+1}|X_{t+1}) \) the observation likelihood
- \( p(X_{t+1}|X_t) \) is the dynamic model
A Probabilistic Graphical Model

- This probabilistic system can be represented by a graphical model

- It is a dynamic Bayesian network
- $Z_t$ is independent of previous states $X_{t-1}$ and $Z_{t-1}$, given current state $X_t$,

\[ p(Z_t|X_t, Z_{t-1}) = p(Z_t|X_t) \]

- the states have Markov property:

\[ p(X_t|X_{t-1}) = p(X_t|X_{t-1}) \]
Four Elements in Tracking

- **$X_t$** target “state” to be estimated or tracked
  - shape, motion, appearance, etc

- **$Z_t$** visual evidence or observations
  - e.g., image edges are the evidence of the contour of a shape
  - e.g., the differences from the template are the evidence of a motion

- $p(Z_t|X_t)$ hypotheses measurement, or matching
  - a likelihood measure
  - can be *ad hoc*, or principled

- $p(X_{t+1}|X_t)$ hypotheses generating, or dynamics
  - it stipulates the prediction
  - the evolution of the dynamic process.
An Example of Observation Model: Shape Matching

- We can represent a shape either free-form or parametric.

- Its visual observations are image edge points.
- A measurement line is a normal line located on the contour.
- Take a number of such measurement lines.
- For each measurement line, perform 1-D edge detection.
- Thus, observation reduces to a set of scalar positions $\mathbf{z} = (z_1, \ldots, z_M)$, due to the presence of clutter.
An Example of Observation Model: Shape Matching

- The true observation $\tilde{z}$ can be any one of them. So,

$$p(z|x) = qp(z|\text{clutter}) + \sum_{m=1}^{M} p(z|x, \tilde{z} = z_m)P(\tilde{z} = z_m)$$

where $x$ is the point on the shape contour and $q = 1 - \sum_{m} P(\tilde{z} = z_m)$.

- When we assume
  - any true observation is normally distributed with standard deviation $\sigma$, $P(\tilde{z} = z_m) = p$ for all $z_m$,
  - the clutter is a Poisson process with density $\lambda$

- then we can have an explicit likelihood model

$$p(z|x) \propto 1 + \frac{1}{\sqrt{2\pi}\sigma q\lambda} \sum_{m} \exp - \frac{(z_m - x)^2}{2\sigma^2}$$
Tracking as Probability Propagation

- When everything is Gaussian
  - use mean and covariance to represent those densities
  - it becomes Kalman filtering
- What if they are not Gaussian?
Outline

Getting Start

What is Tracking
  A Probabilistic Framework
  An Example of Observation Model
  Tracking as Probability Propagation

Recursive Estimation and Kalman Filtering
  Kalman Filtering for Linear Dynamic Systems

Sequential Monte Carlo and Particle Filtering
  Factored Sampling
  Importance Sampling
Traditional Tracking

- Radar tracking
- when the likelihood model is explicit or straightforward
- Tracking is traditionally posed as an recursive estimation problem
- i.e., estimate the current motion given
  - the estimate at the previous time instant
  - the current observation (or measurement)
- This is feasible when the dynamics is known.
- The estimation is a tradeoff between the fidelity to the observation (i.e., likelihood) and the preference to the dynamics (i.e., the prior).
A Linear Dynamic System

- A dynamic system can be represented by a set of differential equations, e.g.,
  \[ \dot{X}(t) = f(X(t)) \]

- E.g., the dynamics of constant velocity motion is:
  \[ x(t + \tau) = x(t) + \frac{dx}{dt} \tau \]

- Let \( S = \begin{bmatrix} x \\ \dot{x} \end{bmatrix} \), we have,
  \[
  S(t + \tau) = \begin{bmatrix} x(t + \tau) \\ \dot{x}(t + \tau) \end{bmatrix} = \begin{bmatrix} 1 & \tau \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x(t) \\ \dot{x}(t) \end{bmatrix} + \begin{bmatrix} n_1 \\ n_2 \end{bmatrix}
  \]

  \[
  y(t) = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} x(t) \\ \dot{x}(t) \end{bmatrix} + \eta
  \]
A Linear Dynamic System: State Equations

- Generally, by introducing the state vector $S$,
- A dynamic systems can be represented by:

$$S_{i+1} = H_i S_i + n_i$$
$$y_i = F_i S_i + \eta_i$$

where

$$E[n_i] = 0, \ E[n_i n_i] = Q_i, \ E[n_i n_j] = 0,$$
$$E[\eta_i] = 0, \ E[\eta_i \eta_i] = \Lambda_i, \ E[\eta_i \eta_j] = 0,$$
Kalman Filtering

- Kalman filtering is a prediction-correction procedure:
  - Prediction of State Vector:
    \[ \hat{S}_{i|i-1} = H_{i-1}\hat{S}_{i-1} \]
  - Prediction of State Covariance:
    \[ P_{i|i=1} = H_{i-1}P_{i-1}H_{i-1}^T + Q_{i-1} \]
  - Kalman Gain:
    \[ K_i = P_{i|i-1}F_i^T(F_iP_{i|i-1}F_i^T + \Lambda_i)^{-1} \]
  - Correction of State Vector:
    \[ \hat{S}_i = \hat{S}_{i|i-1} + K_i(y_i - F_i\hat{S}_{i|i-1}) \]
  - Correction of State Covariance:
    \[ P_i = (I - K_iF_i)P_{i|i-1} \]
Understanding Kalman Filtering

- The Kalman Gain can also be written by:

\[ K_i = P_i F_i^T \Lambda_i^{-1} \]

- when the estimation is confident (i.e., \( P_i \) is small), or the observations are not confident (i.e., \( \Lambda_i \) is big), the Kalman gain will be small, it is better not to adjust.

- if the estimation is not confident, or the observations are confident, the Kalman innovation, i.e., \( y_i - F_i \hat{S}_{i|i-1} \), will contain critical information, and we need to use large Kalman gain to update.

- We have:

\[ P_i^{-1} = P_{i|i-1}^{-1} + F_i^T \Lambda_i^{-1} F_i \]

- when the observation applies, the confidence of the estimation of state vector increases.

- because \( P_i \) will always be smaller than \( P_{i|i-1} \)

- the observation will always reduce the error.
Discussion

- How can we use Kalman filtering for visual tracking?

- What prevents Kalman filters from being used?
Outline

Getting Start

What is Tracking
   A Probabilistic Framework
   An Example of Observation Model
   Tracking as Probability Propagation

Recursive Estimation and Kalman Filtering
   Kalman Filtering for Linear Dynamic Systems

Sequential Monte Carlo and Particle Filtering
   Factored Sampling
   Importance Sampling
Sequential Monte Carlo

- It is clear that tracking can be formulated as:

\[ p(X_{t+1}|Z_{t+1}) \propto p(Z_{t+1}|X_{t+1})p(X_{t+1}|Z_t) \]

- An analytical solution can only be found in special case.
- Monte Carlo methods can be used to approximate the inference.
- Sampling techniques are widely used to approximate a complex probability density.
- A set of weighted random samples \( \{(s^{(n)}, \pi^{(n)})\}, n = 1, \ldots, N \) is properly weighted with respect to the distribution \( f(X) \) if for any integrable function \( h(\cdot) \),

\[
\lim_{N \to \infty} \frac{\sum_{n=1}^{N} h(s^{(n)})\pi^{(n)}}{\sum_{n=1}^{N} \pi^{(n)}} = E_f(h(X))
\]

- \( f(X) \) is approximated by a set of samples \( s^{(n)} \), each having a probability proportional to its weight \( \pi^{(n)} \).
Factored Sampling

- $p(\mathbf{X}_t|\mathbb{Z}_t)$ by $\{(s_t^{(n)}, \pi_t^{(n)})\}$

- It evolves into a new particle set $\{(s_{t+1}^{(n)}, \pi_{t+1}^{(n)})\}$ representing $p(\mathbf{X}_{t+1}|\mathbb{Z}_{t+1})$ at time $t+1$.

- To represent $p(\mathbf{X}_t|\mathbb{Z}_t)$,
  - samples $\{\mathbf{X}_t^{(n)}, n = 1, \ldots, N\}$ are drawn from a prior $p(\mathbf{X}_t|\mathbb{Z}_{t-1})$
  - and weighted by their measurements, i.e.,
    $\pi_t^{(n)} = p(\mathbf{Z}_t|\mathbf{X}_t = \mathbf{X}_t^{(n)})$
  - such that $p(\mathbf{X}_t|\mathbb{Z}_t)$ is represented by a particle set $\{s_t^{(n)}, \pi_t^{(n)}\}$.

- It can be shown that such a sample set is properly weighted.
Discussion

- Computational cost?

- What issue do you expect?

- Solution?
Importance Sampling

- It might be difficult to sample $f(X)$, but easy to sample $g(X)$
- Draw samples from $g(X)$ and adjust their weight, so as to represent $f(X)$
- importance sampling:

  - Samples $s^{(n)}$ are drawn from $g(X)$
  - but weights are compensated as
  
  $$
  \pi^{(n)} = \frac{f(s^{(n)}) \tilde{\pi}^{(n)}}{g(s^{(n)})}
  $$

  - the sample set $\{s^{(n)}, \pi^{(n)}\}$ is still properly weighted with respect to $f(X)$. 
Importance Sampling for Tracking

- Let \( f_t(\mathbf{X}_t^{(n)}) = p(\mathbf{X}_t = \mathbf{X}_t^{(n)}|\mathbf{Z}_{t-1}) \), the prediction prior.
- Sampling another distribution \( g_t(\mathbf{X}_t) \), and adjust weights

\[
\pi_t^{(n)} = \frac{f_t(\mathbf{X}_t^{(n)})}{g_t(\mathbf{X}_t^{(n)})} p(\mathbf{Z}_t|\mathbf{X}_t = \mathbf{X}_t^{(n)})
\]

- To evaluate \( f_t(\mathbf{X}_t) \), we have

\[
f_t(\mathbf{X}_t^{(n)}) = p(\mathbf{X}_t = \mathbf{X}_t^{(n)}|\mathbf{Z}_{t-1}) = \sum_{k=1}^{N} \pi_{t-1}^{(k)} p(\mathbf{X}_t = \mathbf{X}_t^{(n)}|\mathbf{X}_{t-1} = \mathbf{X}_{t-1}^{(k)})
\]

- To approximate \( p(\mathbf{X}_t|\mathbf{Z}_t) \)
  - Samples \( s^{(n)} \) are drawn from another source \( g_t(\mathbf{X}_t) \)
  - The weight of each sample is

\[
\pi_t^{(n)} = \frac{f_t(s_t^{(n)})}{g_t(s_t^{(n)})} p(\mathbf{Z}_t|\mathbf{X}_t = s_t^{(n)})
\]

where \( f_t(s_t^{(n)}) = p(\mathbf{X}_t = s_t^{(n)}|\mathbf{Z}_{t-1}) \).
- The selection of the importance function can be quite flexible.