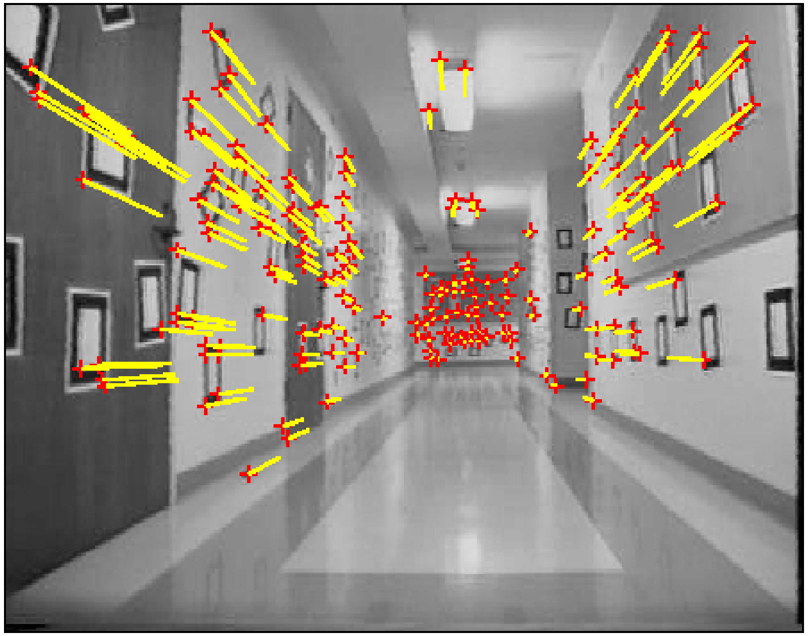
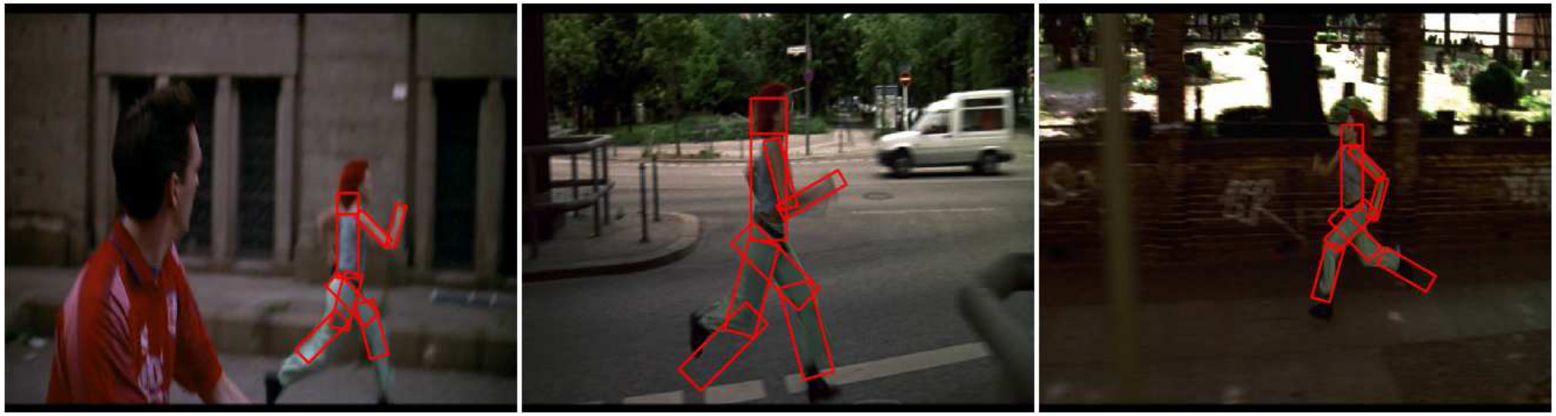
Tracking





Many slides adapted from Kristen Grauman, Deva Ramanan

Tracking with dynamics

- Key idea: Given a model of expected motion, predict where objects will occur in next frame, even before seeing the image
  - Restrict search for the object
  - Improved estimates since measurement noise is reduced by trajectory smoothness

General model for tracking

- The moving object of interest is characterized by an underlying *state* X
- State X gives rise to *measurements* or *observations* Y
- At each time t, the state changes to X_t and we get a new observation Y_t

Steps of tracking

- **Prediction:** What is the next state of the object given past measurements?

$$P\left(X_t \middle| Y_0 = y_0, \ldots, Y_{t-1} = y_{t-1}\right)$$

Steps of tracking

- **Prediction:** What is the next state of the object given past measurements?

$$P\left(X_t \middle| Y_0 = y_0, \ldots, Y_{t-1} = y_{t-1}\right)$$

- **Correction:** Compute an updated estimate of the state from prediction and measurements

$$P\left(X_t \middle| Y_0 = y_0, \ldots, Y_{t-1} = y_{t-1}, Y_t = y_t\right)$$

Steps of tracking

- **Prediction:** What is the next state of the object given past measurements?

$$P\left(X_t \middle| Y_0 = y_0, \ldots, Y_{t-1} = y_{t-1}\right)$$

- **Correction:** Compute an updated estimate of the state from prediction and measurements

$$P\left(X_t \middle| Y_0 = y_0, \ldots, Y_{t-1} = y_{t-1}, Y_t = y_t\right)$$

- Tracking can be seen as the process of propagating the posterior distribution of state given measurements across time

Simplifying assumptions

- Only the immediate past matters

$$P\left(X_t \mid X_0, \ldots, X_{t-1}\right) = \boxed{P\left(X_t \mid X_{t-1}\right)}$$

dynamics model

Simplifying assumptions

- Only the immediate past matters

$$P(X_t | X_0, \ldots, X_{t-1}) = \boxed{P(X_t | X_{t-1})}$$

dynamics model

- Measurements depend only on the current state

$$P(Y_t | X_0, Y_0 \ldots, X_{t-1}, Y_{t-1}, X_t) = \boxed{P(Y_t | X_t)}$$

observation model

Simplifying assumptions

- Only the immediate past matters

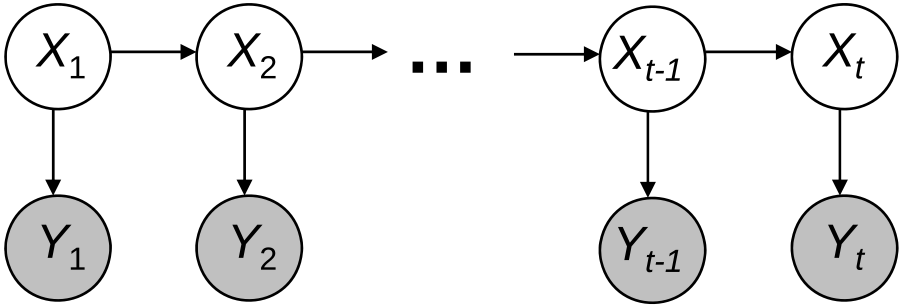
$$P(X_t | X_0, \ldots, X_{t-1}) = \boxed{P(X_t | X_{t-1})}$$

dynamics model

- Measurements depend only on the current state

$$P(Y_t | X_0, Y_0 \ldots, X_{t-1}, Y_{t-1}, X_t) = \boxed{P(Y_t | X_t)}$$

observation model



Tracking as induction

- Base case:
  - Assume we have initial prior that predicts state in absence of any evidence: $P(X_0)$
  - At the first frame, *correct* this given the value of $Y_0 = y_0$

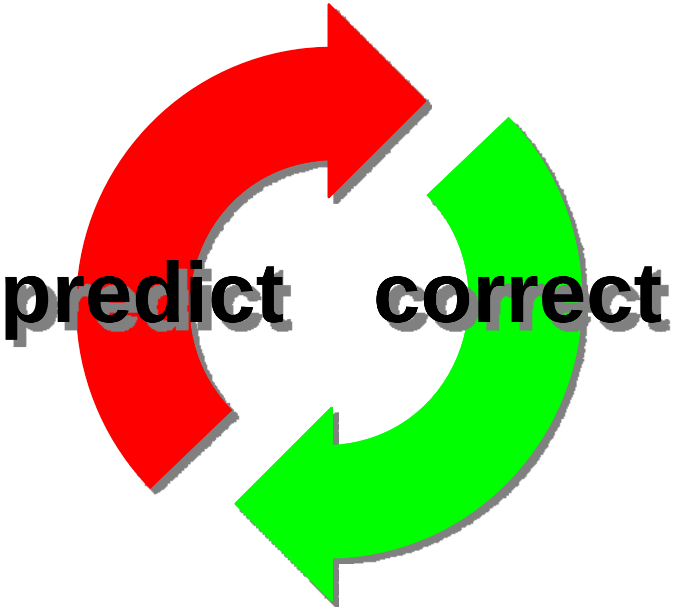
Tracking as induction

- Base case:
  - Assume we have initial prior that predicts state in absence of any evidence: $P(X_0)$
  - At the first frame, *correct* this given the value of $Y_0=y_0$

$$P(X_0 \mid Y_0 = y_0) = \frac{P(y_0 \mid X_0)P(X_0)}{P(y_0)} \propto P(y_0 \mid X_0)P(X_0)$$

Tracking as induction

- Base case:
  - Assume we have initial prior that predicts state in absence of any evidence: $P(X_0)$
  - At the first frame, *correct* this given the value of $Y_0 = y_0$
- Given corrected estimate for frame t:
  - Predict for frame $t+1$
  - Correct for frame $t+1$



Prediction

- Prediction involves representing $P\left(X_t \middle| y_0, \ldots, y_{t-1}\right)$ given $P\left(X_{t-1} \middle| y_0, \ldots, y_{t-1}\right)$

Prediction

- Prediction involves representing $P(X_t | y_0, \ldots, y_{t-1})$ given $P(X_{t-1} | y_0, \ldots, y_{t-1})$

$$P(X_t | y_0, \ldots, y_{t-1})$$

$$= \int P(X_t, X_{t-1} | y_0, \ldots, y_{t-1}) dX_{t-1}$$

Law of total probability

Prediction

- Prediction involves representing $P(X_t | y_0, \ldots, y_{t-1})$ given $P(X_{t-1} | y_0, \ldots, y_{t-1})$

$$
\begin{aligned}
&P(X_t | y_0, \ldots, y_{t-1}) \\
&= \int P(X_t, X_{t-1} | y_0, \ldots, y_{t-1}) dX_{t-1} \\
&= \int P(X_t \mid X_{t-1}, y_0, \ldots, y_{t-1}) P(X_{t-1} \mid y_0, \ldots, y_{t-1}) dX_{t-1}
\end{aligned}
$$

Conditioning on X_{t-1}

Prediction

- Prediction involves representing $P\left(X_t \middle| y_0, \dots, y_{t-1}\right)$ given $P\left(X_{t-1} \middle| y_0, \dots, y_{t-1}\right)$

$$
\begin{aligned}
&P\left(X_t \middle| y_0, \dots, y_{t-1}\right) \\
&= \int P\left(X_t, X_{t-1} \middle| y_0, \dots, y_{t-1}\right) dX_{t-1} \\
&= \int P\left(X_t \mid X_{t-1}, y_0, \dots, y_{t-1}\right) P\left(X_{t-1} \mid y_0, \dots, y_{t-1}\right) dX_{t-1} \\
&= \int P\left(X_t \mid X_{t-1}\right) P\left(X_{t-1} \mid y_0, \dots, y_{t-1}\right) dX_{t-1}
\end{aligned}
$$

Independence assumption

Prediction

- Prediction involves representing $P\left(X_t \middle| y_0, \ldots, y_{t-1}\right)$ given $P\left(X_{t-1} \middle| y_0, \ldots, y_{t-1}\right)$

$$
\begin{aligned}
&P\left(X_t \middle| y_0, \ldots, y_{t-1}\right) \\
&= \int P\left(X_t, X_{t-1} \middle| y_0, \ldots, y_{t-1}\right) dX_{t-1} \\
&= \int P\left(X_t \mid X_{t-1}, y_0, \ldots, y_{t-1}\right) P\left(X_{t-1} \mid y_0, \ldots, y_{t-1}\right) dX_{t-1} \\
&= \int \underbrace{P\left(X_t \mid X_{t-1}\right)}_{\text{dynamics model}} \underbrace{P\left(X_{t-1} \mid y_0, \ldots, y_{t-1}\right)}_{\text{corrected estimate from previous step}} dX_{t-1}
\end{aligned}
$$

Correction

- Correction involves computing $P\left(X_t \middle| y_0, \ldots, y_t\right)$ given predicted value $P\left(X_t \middle| y_0, \ldots, y_{t-1}\right)$

Correction

- Correction involves computing $P\left(X_t \middle| y_0, \ldots, y_t\right)$ given predicted value $P\left(X_t \middle| y_0, \ldots, y_{t-1}\right)$

$$P\left(X_t \middle| y_0, \ldots, y_t\right)$$

$$= \frac{P\left(y_t \mid X_t, y_0, \ldots, y_{t-1}\right) P\left(X_t \mid y_0, \ldots, y_{t-1}\right)}{P\left(y_t \mid y_0, \ldots, y_{t-1}\right)}$$

Bayes rule

Correction

- Correction involves computing $P\left(X_t \middle| y_0, \ldots, y_t\right)$ given predicted value $P\left(X_t \middle| y_0, \ldots, y_{t-1}\right)$

$$P\left(X_t \middle| y_0, \ldots, y_t\right)$$

$$= \frac{P\left(y_t \mid X_t, y_0, \ldots, y_{t-1}\right) P\left(X_t \mid y_0, \ldots, y_{t-1}\right)}{P\left(y_t \mid y_0, \ldots, y_{t-1}\right)}$$

$$= \frac{P\left(y_t \mid X_t\right) P\left(X_t \mid y_0, \ldots, y_{t-1}\right)}{P\left(y_t \mid y_0, \ldots, y_{t-1}\right)}$$

Independence assumption
(observation y_t depends only on state X_t)

Correction

- Correction involves computing $P\left(X_t \middle| y_0, \ldots, y_t\right)$ given predicted value $P\left(X_t \middle| y_0, \ldots, y_{t-1}\right)$

$$P\left(X_t \middle| y_0, \ldots, y_t\right)$$

$$= \frac{P\left(y_t \mid X_t, y_0, \ldots, y_{t-1}\right) P\left(X_t \mid y_0, \ldots, y_{t-1}\right)}{P\left(y_t \mid y_0, \ldots, y_{t-1}\right)}$$

$$= \frac{P\left(y_t \mid X_t\right) P\left(X_t \mid y_0, \ldots, y_{t-1}\right)}{P\left(y_t \mid y_0, \ldots, y_{t-1}\right)}$$

$$= \frac{P\left(y_t \mid X_t\right) P\left(X_t \mid y_0, \ldots, y_{t-1}\right)}{\int P\left(y_t \mid X_t\right) P\left(X_t \mid y_0, \ldots, y_{t-1}\right) dX_t}$$

Conditioning on X_t

Correction

- Correction involves computing $P\left(X_t \middle| y_0,\ldots,y_t\right)$ given predicted value $P\left(X_t \middle| y_0,\ldots,y_{t-1}\right)$

$$P\left(X_t \middle| y_0,\ldots,y_t\right)$$

$$= \frac{P\left(y_t \mid X_t, y_0,\ldots,y_{t-1}\right)P\left(X_t \mid y_0,\ldots,y_{t-1}\right)}{P\left(y_t \mid y_0,\ldots,y_{t-1}\right)}$$

$$= \frac{P\left(y_t \mid X_t\right)P\left(X_t \mid y_0,\ldots,y_{t-1}\right)}{P\left(y_t \mid y_0,\ldots,y_{t-1}\right)}$$

observation model

predicted estimate

$$= \frac{P\left(y_t \mid X_t\right)P\left(X_t \mid y_0,\ldots,y_{t-1}\right)}{\int P\left(y_t \mid X_t\right)P\left(X_t \mid y_0,\ldots,y_{t-1}\right)dX_t}$$

normalization factor

Summary: Prediction and correction

- Prediction:

$$P(X_t \mid y_0, \ldots, y_{t-1}) = \int \underbrace{P(X_t \mid X_{t-1})}_{\text{dynamics model}} \underbrace{P(X_{t-1} \mid y_0, \ldots, y_{t-1})}_{\text{corrected estimate from previous step}} dX_{t-1}$$

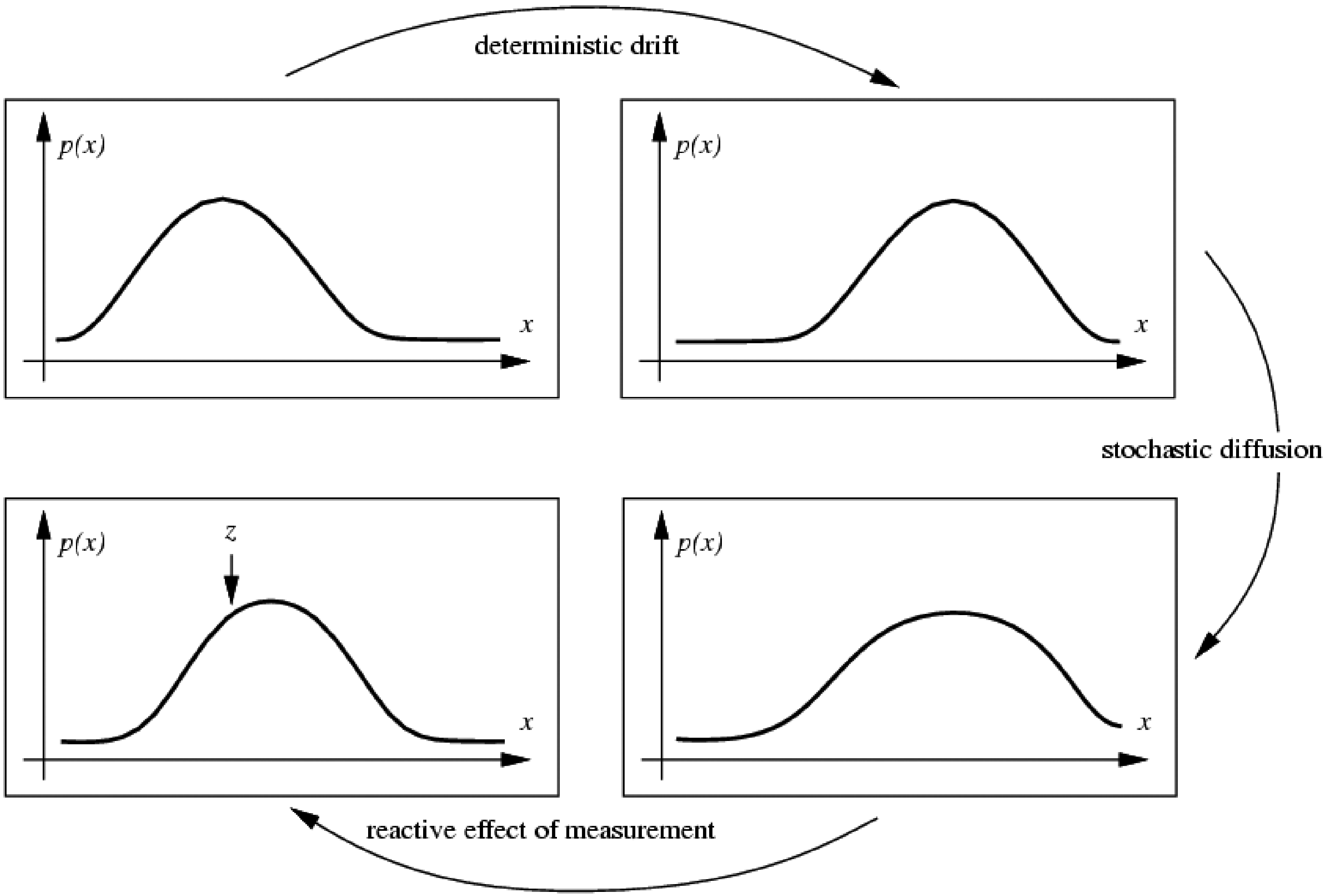
- Correction:

$$P(X_t \mid y_0, \ldots, y_t) = \frac{\overbrace{P(y_t \mid X_t)}^{\text{observation model}} \overbrace{P(X_t \mid y_0, \ldots, y_{t-1})}^{\text{predicted estimate}}}{\int P(y_t \mid X_t) P(X_t \mid y_0, \ldots, y_{t-1}) dX_t}$$

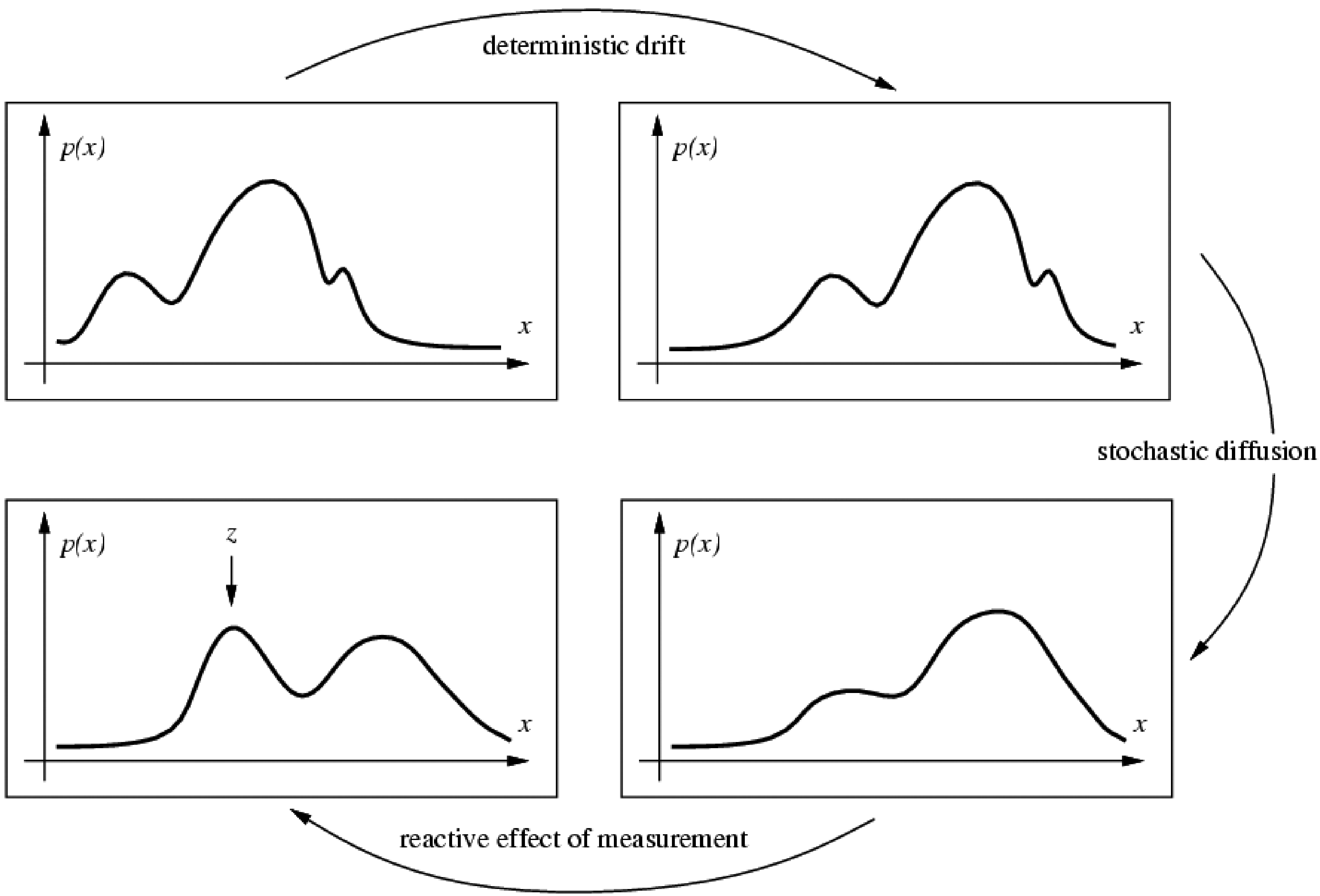
The Kalman filter

- Linear dynamics model: state undergoes linear transformation plus Gaussian noise
- Observation model: measurement is linearly transformed state plus Gaussian noise
- The predicted/corrected state distributions are Gaussian
  - You only need to maintain the mean and covariance
  - The calculations are easy (all the integrals can be done in closed form)

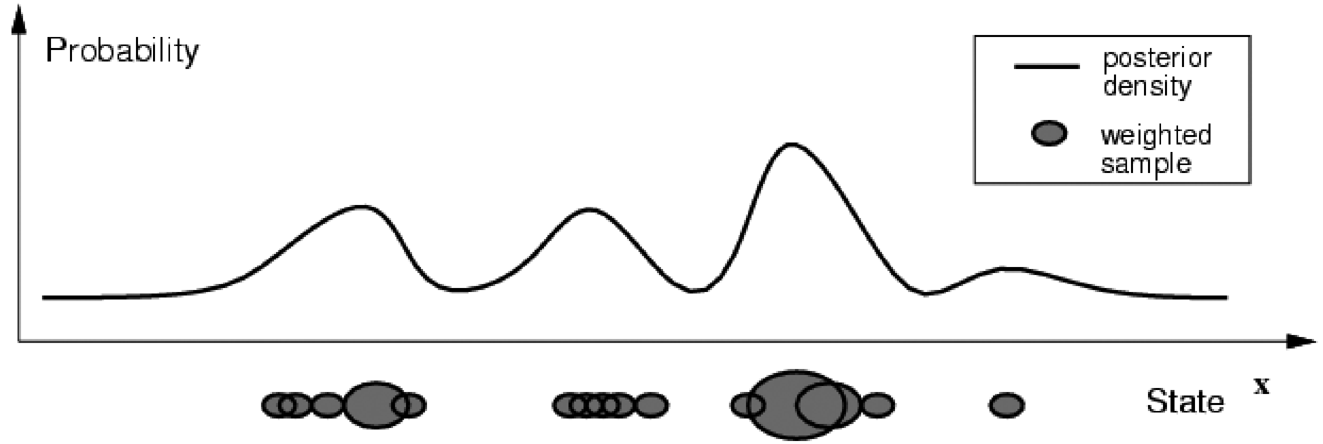
Propagation of Gaussian densities



Propagation of general densities



Factored sampling

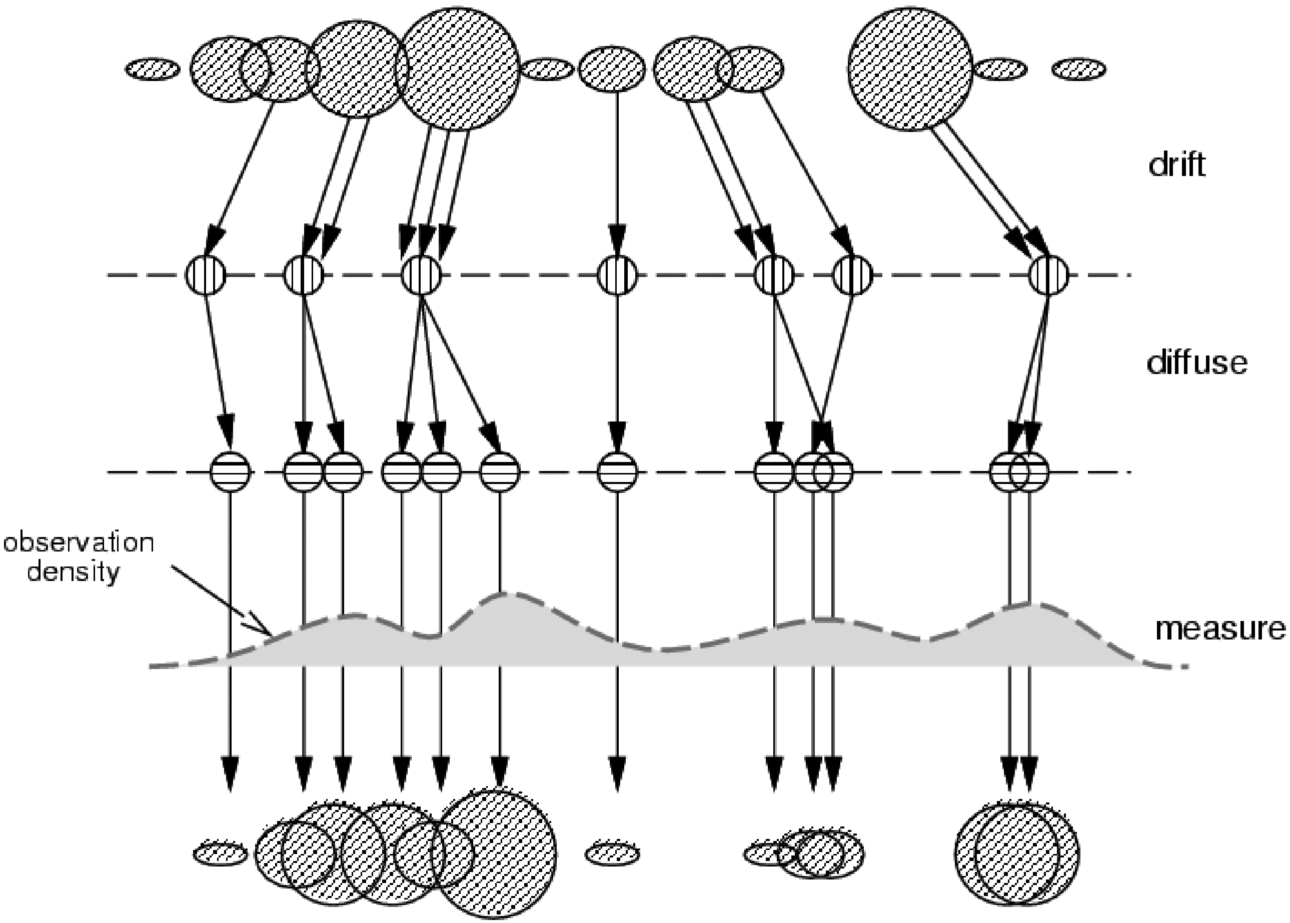


- Represent the state distribution non-parametrically
  - Prediction: Sample points from prior density for the state, *P*(*X*)
  - Correction: Weight the samples according to *P*(*Y*|*X*)

$$P\left(X_t \mid y_0,\ldots,y_t\right)=\frac{P\left(y_t \mid X_t\right)P\left(X_t \mid y_0,\ldots,y_{t-1}\right)}{\int P\left(y_t \mid X_t\right)P\left(X_t \mid y_0,\ldots,y_{t-1}\right)dX_t}$$

M. Isard and A. Blake, [CONDENSATION -- conditional density propagation for visual tracking](), IJCV 29(1):5-28, 1998

Particle filtering



Start with weighted samples from previous time step

Sample and shift according to dynamics model

Spread due to randomness; this is predicted density $P(X_t|Y_{t-1})$

Weight the samples according to observation density

Arrive at corrected density estimate $P(X_t|Y_t)$

M. Isard and A. Blake, [CONDENSATION -- conditional density propagation for visual tracking](), IJCV 29(1):5-28, 1998

Particle filtering results

http://www.robots.ox.ac.uk/~misard/condensation.html

Tracking issues

- Initialization
  - Manual
  - Background subtraction
  - Detection

Tracking issues

- Initialization
- Obtaining observation and dynamics model
  - Generative observation model: “render” the state on top of the image and compare
  - Discriminative observation model: classifier or detector score
  - Dynamics model: learn (very difficult) or specify using domain knowledge

Tracking issues

- Initialization
- Obtaining observation and dynamics model
- Prediction vs. correction
  - If the dynamics model is too strong, will end up ignoring the data
  - If the observation model is too strong, tracking is reduced to repeated detection

Tracking issues

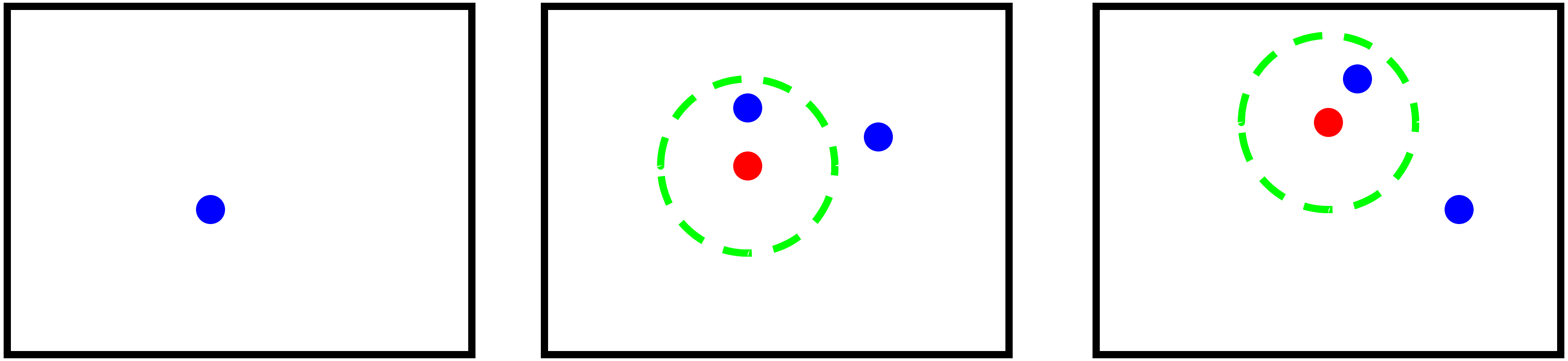
- Initialization
- Obtaining observation and dynamics model
- Prediction vs. correction
- Data association
  - What if we don’t know which measurements to associate with which tracks?

Data association

- So far, we've assumed the entire measurement to be relevant to determining the state
- In reality, there may be uninformative measurements (clutter) or measurements may belong to different tracked objects
- **Data association**: task of determining which measurements go with which tracks

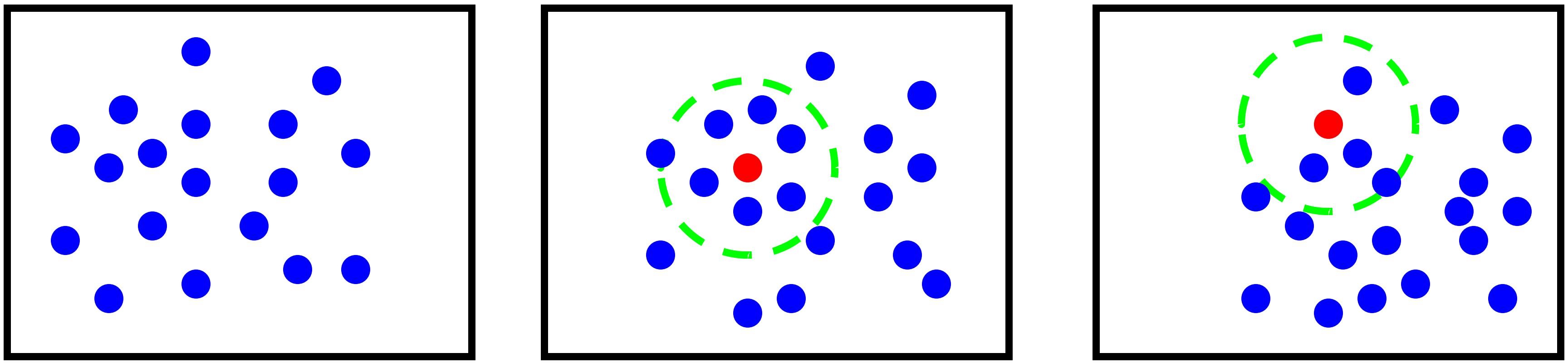
Data association

- Simple strategy: only pay attention to the measurement that is “closest” to the prediction



Data association

- Simple strategy: only pay attention to the measurement that is “closest” to the prediction



Doesn’t always work…

Data association

- Simple strategy: only pay attention to the measurement that is “closest” to the prediction
- More sophisticated strategy: keep track of multiple state/observation hypotheses
  - Can be done with particle filtering
- This is a general problem in computer vision, there is no easy solution

Recall: Generative part-based models

$$P(image \mid object) = P(appearance, shape \mid object)$$

Model

Candidate parts

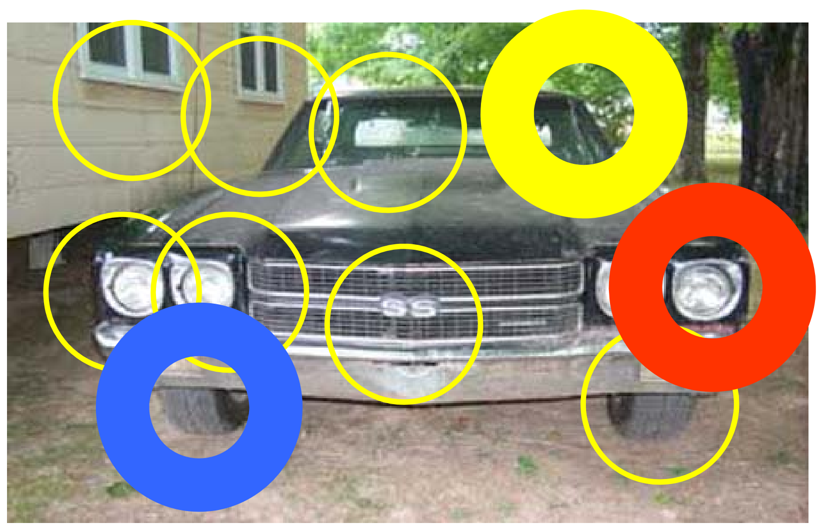
Recall: Generative part-based models

$$P(image \mid object) = P(appearance, shape \mid object)$$

$$= \max_h P(appearance \mid h, object)\, p(shape \mid h, object)\, p(h \mid object)$$

h: assignment of features to parts

Model

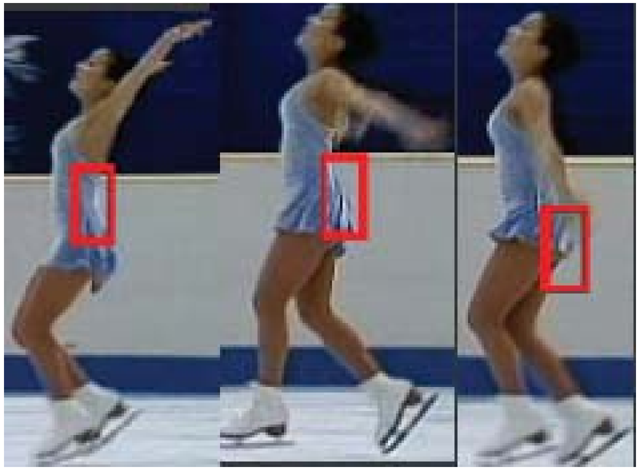


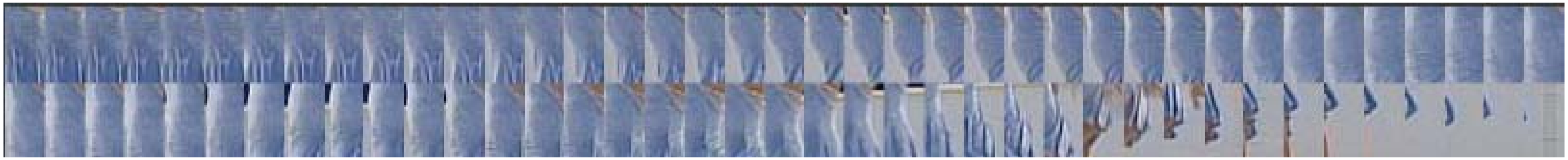
Candidate parts

Tracking issues

- Initialization
- Obtaining observation and dynamics model
- Prediction vs. correction
- Data association
- Drift
  - Errors caused by dynamical model, observation model, and data association tend to accumulate over time

Drift





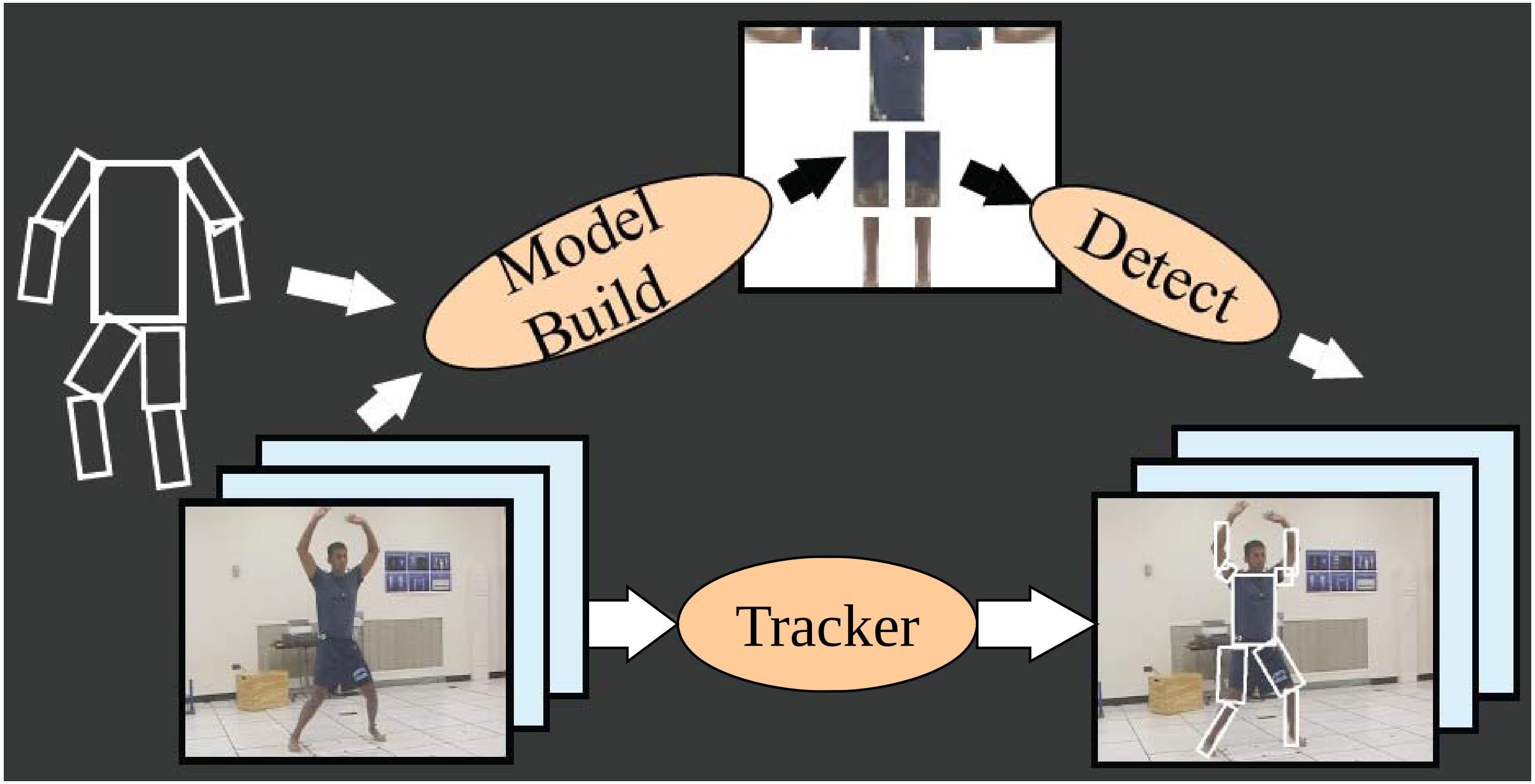
D. Ramanan, D. Forsyth, and A. Zisserman. [Tracking People by Learning their Appearance](). PAMI 2007.

Tracking people by learning their appearance

- Person model = appearance + structure (+ dynamics)
- Structure and dynamics are generic, appearance is person-specific
- Trying to acquire an appearance model "on the fly" can lead to drift
- Instead, can use the whole sequence to initialize the appearance model and then keep it fixed while tracking
- Given strong structure and appearance models, tracking can essentially be done by repeated detection (with some smoothing)

D. Ramanan, D. Forsyth, and A. Zisserman. Tracking People by Learning their Appearance. PAMI 2007.

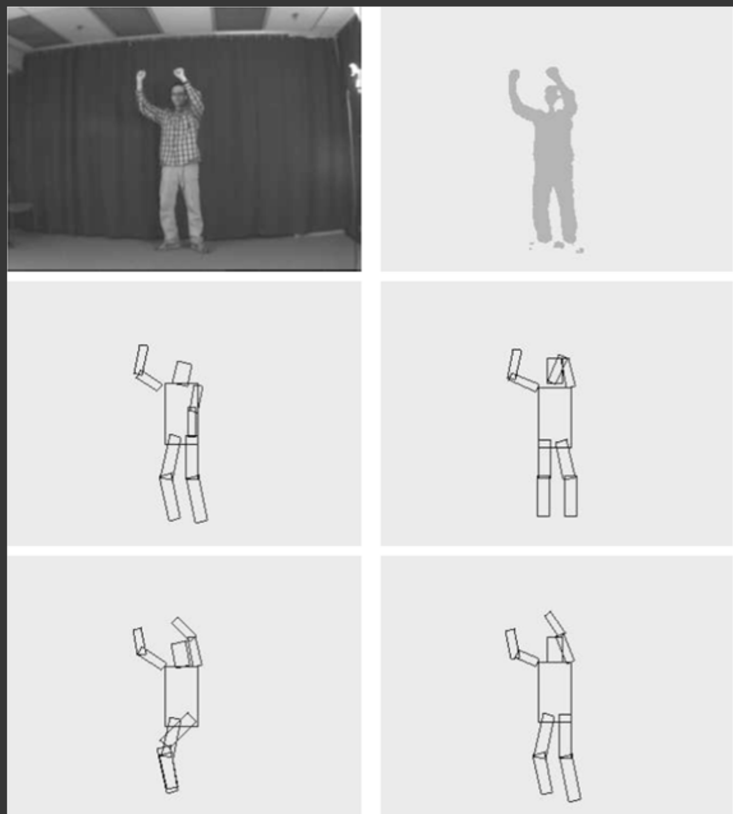
Tracking people by learning their appearance

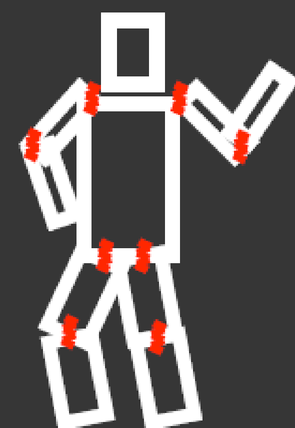


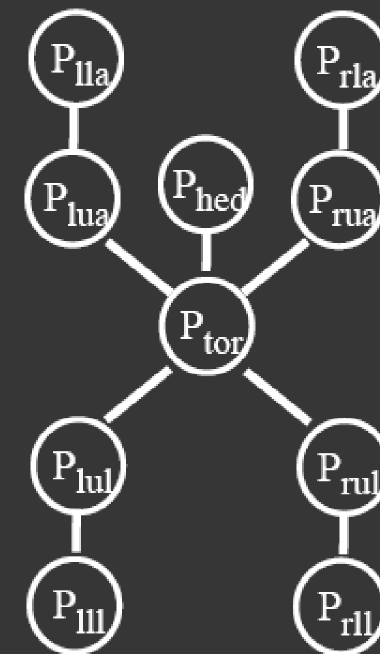
D. Ramanan, D. Forsyth, and A. Zisserman. [Tracking People by Learning their Appearance](). PAMI 2007.

Pictorial structure model

Fischler and Elschlager(73), Felzenszwalb and Huttenlocher(00)





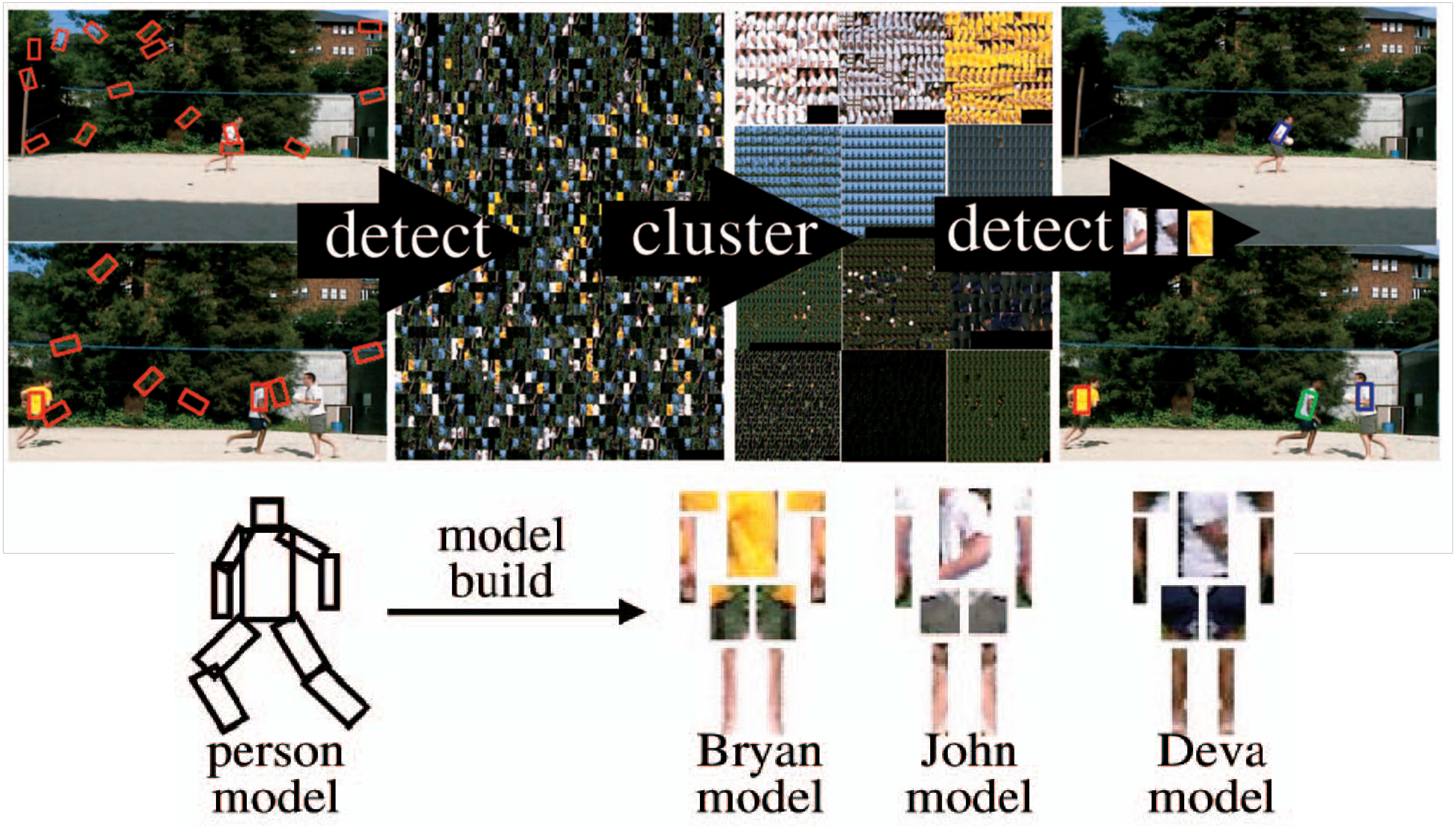


$$\Pr(P_{tor}, P_{arm}, \ldots | \mathrm{Im}) \propto \prod_{i,j} \Pr(P_i \mid P_j) \prod_i \Pr(\mathrm{Im}(P_i))$$

part geometry

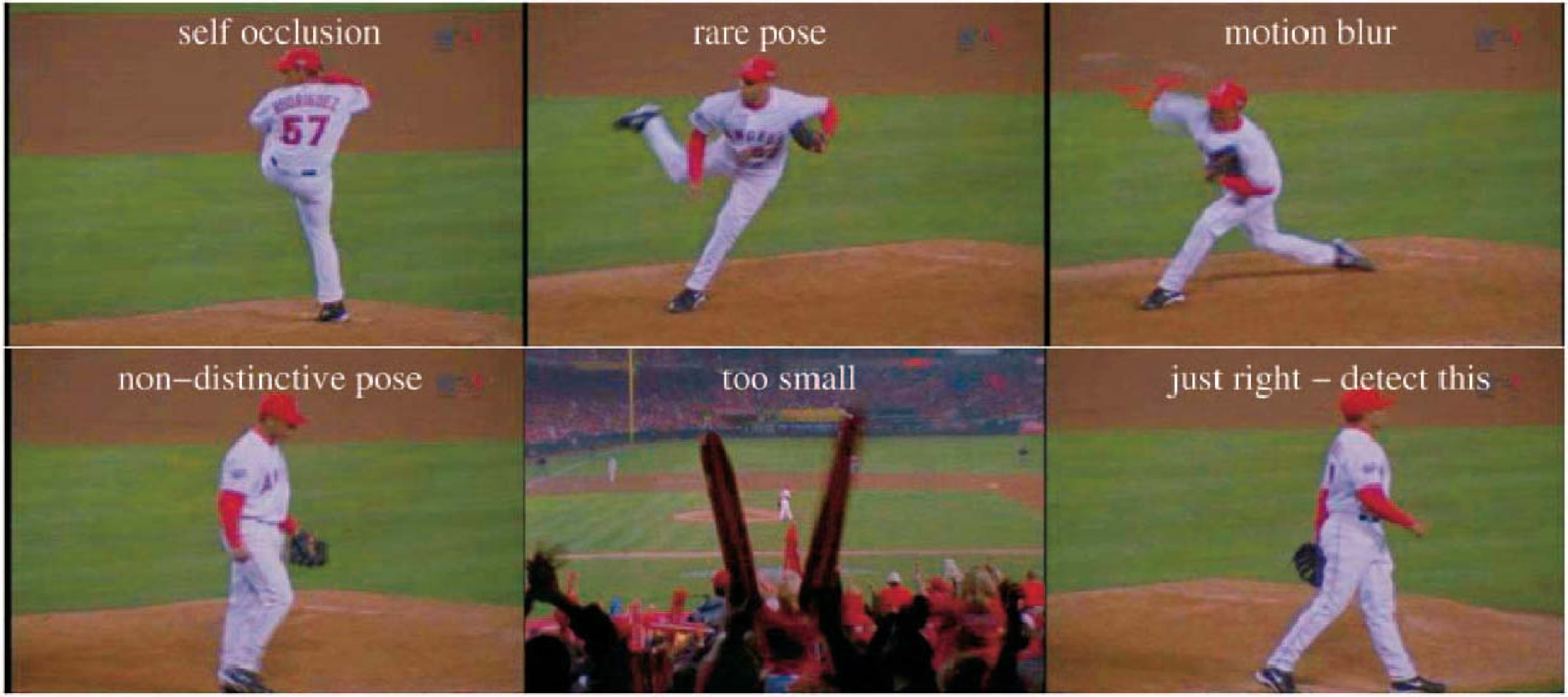
part appearance

Bottom-up initialization: Clustering



D. Ramanan, D. Forsyth, and A. Zisserman. [Tracking People by Learning their Appearance](). PAMI 2007.

Top-down initialization: Exploit “easy” poses



D. Ramanan, D. Forsyth, and A. Zisserman. [Tracking People by Learning their Appearance](). PAMI 2007.

Example results

http://www.ics.uci.edu/~dramanan/papers/pose/index.html