Tracking Objects with Dynamics

Computer Vision
CS 543 / ECE 549
University of Illinois

Derek Hoiem

Slides from Lana Lazebnik, Kristen Grauman, Deva Ramanan

Misc

- Categorization problem in HW 3

- Time for the poster session
  - 12-2pm not possible
  - 10:30-1pm, with pizza at 12pm?
  - 2:30-5pm, with snacks of some sort?

Recent classes

- Point correspondences for
  - Image stitching
  - Stereo and depth estimation
  - Tracking points
  - Structure from motion

Today: Tracking Objects

Goal: Detect and link positions/pose of objects across video frames

Traffic camera

http://www.youtube.com/watch?v=j2C99h6ndS8

Football

http://www.youtube.com/watch?v=odbp6Cg5mC4

Tennis

http://www.youtube.com/watch?v=T7PLgsQHibg

Why do we want to track objects?

- Motion capture / animation:
  http://www.youtube.com/watch?v=eYSXaU6eKm4
- Scene understanding
  - Action recognition
  - Security, traffic monitoring
  - Video summarization

Things that make visual tracking difficult

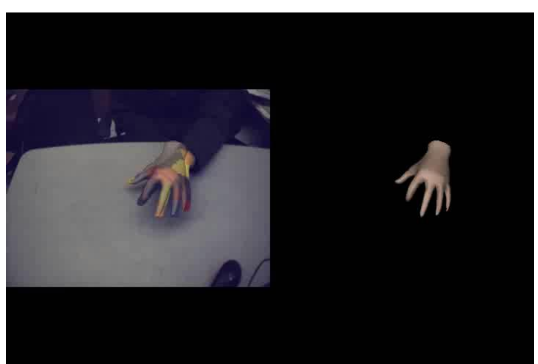
- Objects that are hard to detect
  - Small, few visual features
- Erratic movements
- Moving very quickly
- Occlusions
- Objects may leave and come back
- Surrounding similar-looking objects

This class

- Overview of probabilistic tracking
- Kalman filter
- Particle filter
- Example of tracking people

Strategies for tracking

- Tracking by repeated detection
  - Works well if object is easily detectable (e.g., face or colored glove) and there is only one
    http://people.csail.mit.edu/rywang/handtracking/
  - Might want to update appearance model to fit a particular instance
  - Need some way to link up detections
  - Best you can do, if you can't predict motion



Tracking with dynamics

- Key idea: Based on a model of expected motion, predict where objects will occur in next frame, before even seeing the image
  - Restrict search for the object
  - Improve estimates; measurement noise is reduced by trajectory smoothness
  - Robustness to missing or weak observations

Following slides adapted from Lana Lazebnik, who adapted them from Kristen Grauman and Deva Ramanan

Strategies for tracking

- Tracking with motion prediction
  - Predict the object’s state in the next frame
  - Kalman filtering: next state can be linearly predicted from current state (Gaussian)
  - Particle filtering: sample multiple possible states of the object (non-parametric, good for clutter)

“[Tracking Bees in a Hive]”, Veeraraghavan and Chellappa

General model for tracking

- The moving object of interest is characterized by an underlying ***state*** $\boldsymbol{X}$
  - State could be any combination of position, pose, viewpoint, velocity, acceleration, etc.
- State X gives rise to *measurements* or ***observations*** $\boldsymbol{Y}$
- 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 probability 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

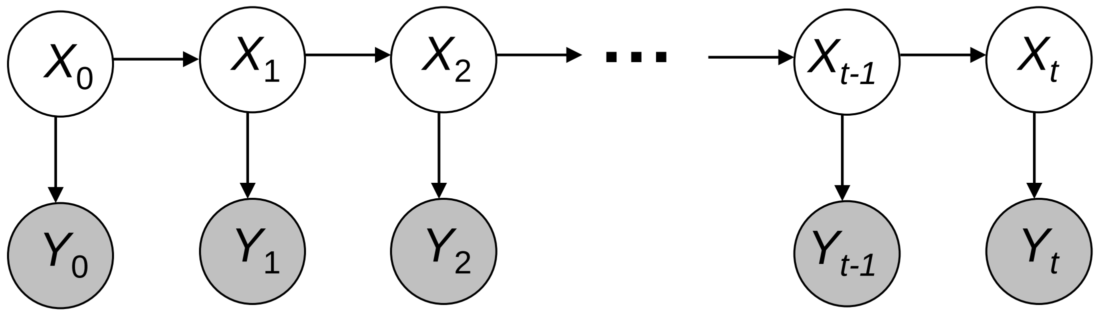
- Only the immediate past matters

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

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



Problem statement

- We have models for

  Likelihood of next state given current state: $P\left(X_t \middle| X_{t-1}\right)$

  Likelihood of observation given the state: $P\left(Y_t \middle| X_t\right)$

- We want to recover, for each t: $P\left(X_t \middle| y_0, \ldots, y_t\right)$

Tracking as induction

- Base case:
  - Start with initial prior that predicts state in absence of any evidence: $P(X_0)$
  - For the first frame, *correct* this given the first measurement: $Y_0 = y_0$

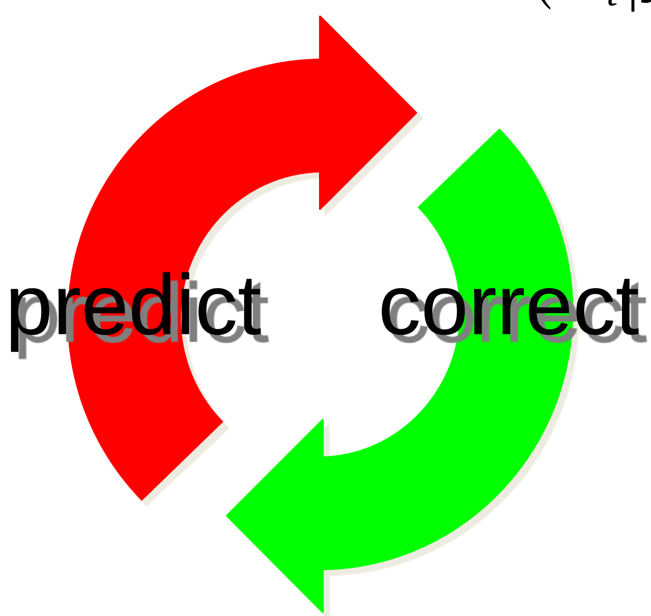
Tracking as induction

- Base case:
  - Start with initial prior that predicts state in absence of any evidence: $P(X_0)$
  - For the first frame, *correct* this given the first measurement: $Y_0 = y_0$

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

Tracking as induction

- Base case:
  - Start with initial prior that predicts state in absence of any evidence: $P(X_0)$
  - For the first frame, *correct* this given the first measurement: $Y_0 = y_0$
- Given corrected estimate for frame *t-1*:
  - Predict for frame *t* ➔ $P\left(X_t \middle| y_0, \ldots, y_{t-1}\right)$
  - Observe y_t; Correct for frame *t* ➔ $P\left(X_t \middle| y_0, \ldots, y_{t-1}, y_t\right)$



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

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

Conditioning on X_{t-1}

Prediction

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

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

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

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

Bayes' Rule

Correction

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

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

Independence assumption
(observation y_t directly depends only on state X_t)

Correction

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

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

Conditioning on X_t

Correction

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

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

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

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

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)

Kalman filter

Dynamic Model:

$$\begin{aligned} \boldsymbol{x}_i &\sim N(\mathcal{D}_i \boldsymbol{x}_{i-1}, \Sigma_{d_i}) \\ \boldsymbol{y}_i &\sim N(\mathcal{M}_i \boldsymbol{x}_i, \Sigma_{m_i}) \end{aligned}$$

Start Assumptions: $\overline{\boldsymbol{x}}_0^-$ and Σ_0^- are known
Update Equations: Prediction

$$\begin{aligned} \overline{\boldsymbol{x}}_i^- &= \mathcal{D}_i \overline{\boldsymbol{x}}_{i-1}^+ \\ \Sigma_i^- &= \Sigma_{d_i} + \mathcal{D}_i \sigma_{i-1}^+ \mathcal{D}_i \end{aligned}$$

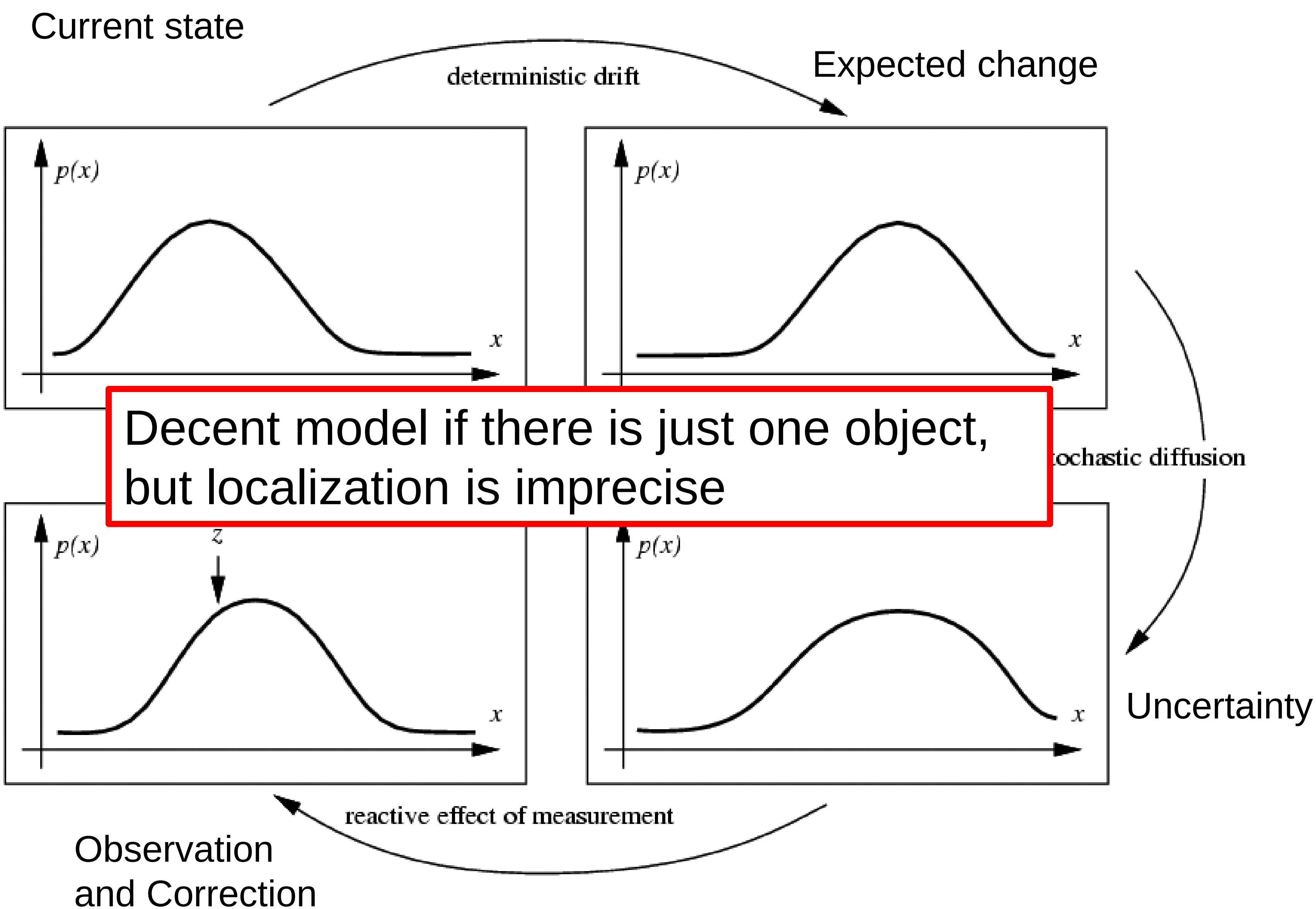
Update Equations: Correction

$$\begin{aligned} \mathcal{K}_i &= \Sigma_i^- \mathcal{M}_i^T \left[\mathcal{M}_i \Sigma_i^- \mathcal{M}_i^T + \Sigma_{m_i} \right]^{-1} \\ \overline{\boldsymbol{x}}_i^+ &= \overline{\boldsymbol{x}}_i^- + \mathcal{K}_i \left[\boldsymbol{y}_i - \mathcal{M}_i \overline{\boldsymbol{x}}_i^- \right] \\ \Sigma_i^+ &= \left[Id - \mathcal{K}_i \mathcal{M}_i \right] \Sigma_i^- \end{aligned}$$

Example: Suppose that

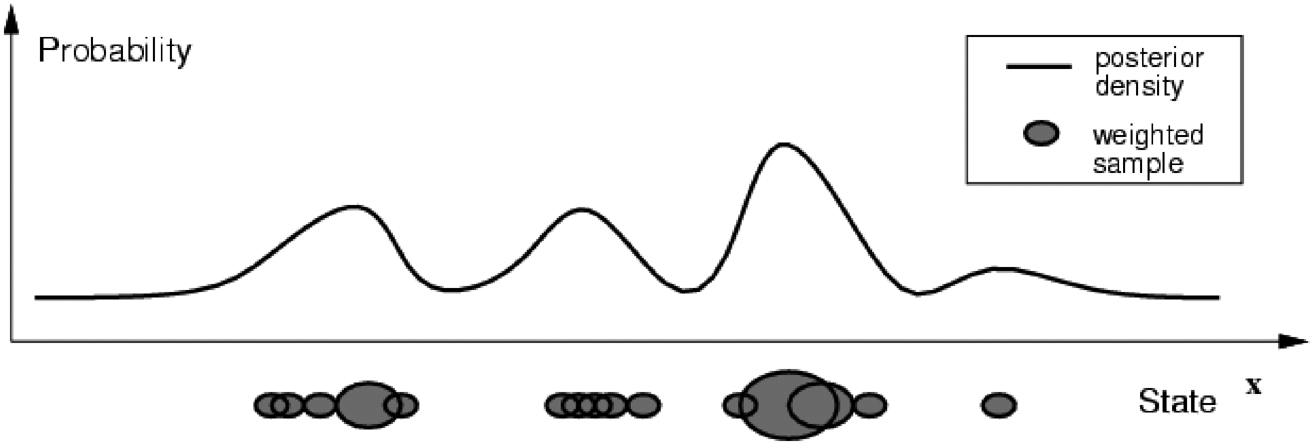
- State X is a vector of {position, velocity, acceleration}
- Observation Y is a position (with some noise)

Propagation of Gaussian densities



Decent model if there is just one object, but localization is imprecise

Particle filtering

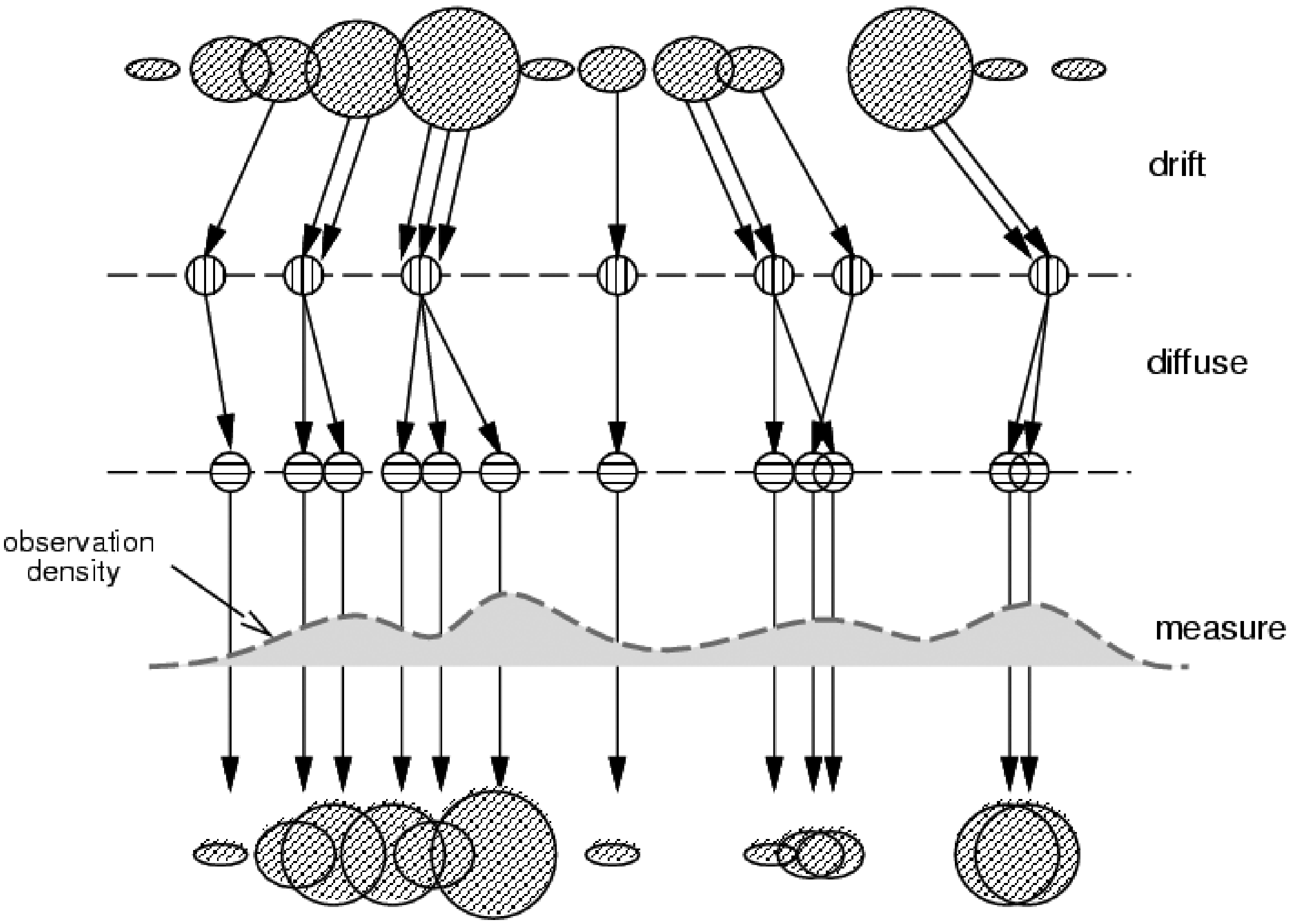


Represent the state distribution non-parametrically

- Prediction: Sample possible values X_{t-1} for the previous state
- Correction: Compute likelihood of X_t based on weighted samples and $P(y_t|X_t)$

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

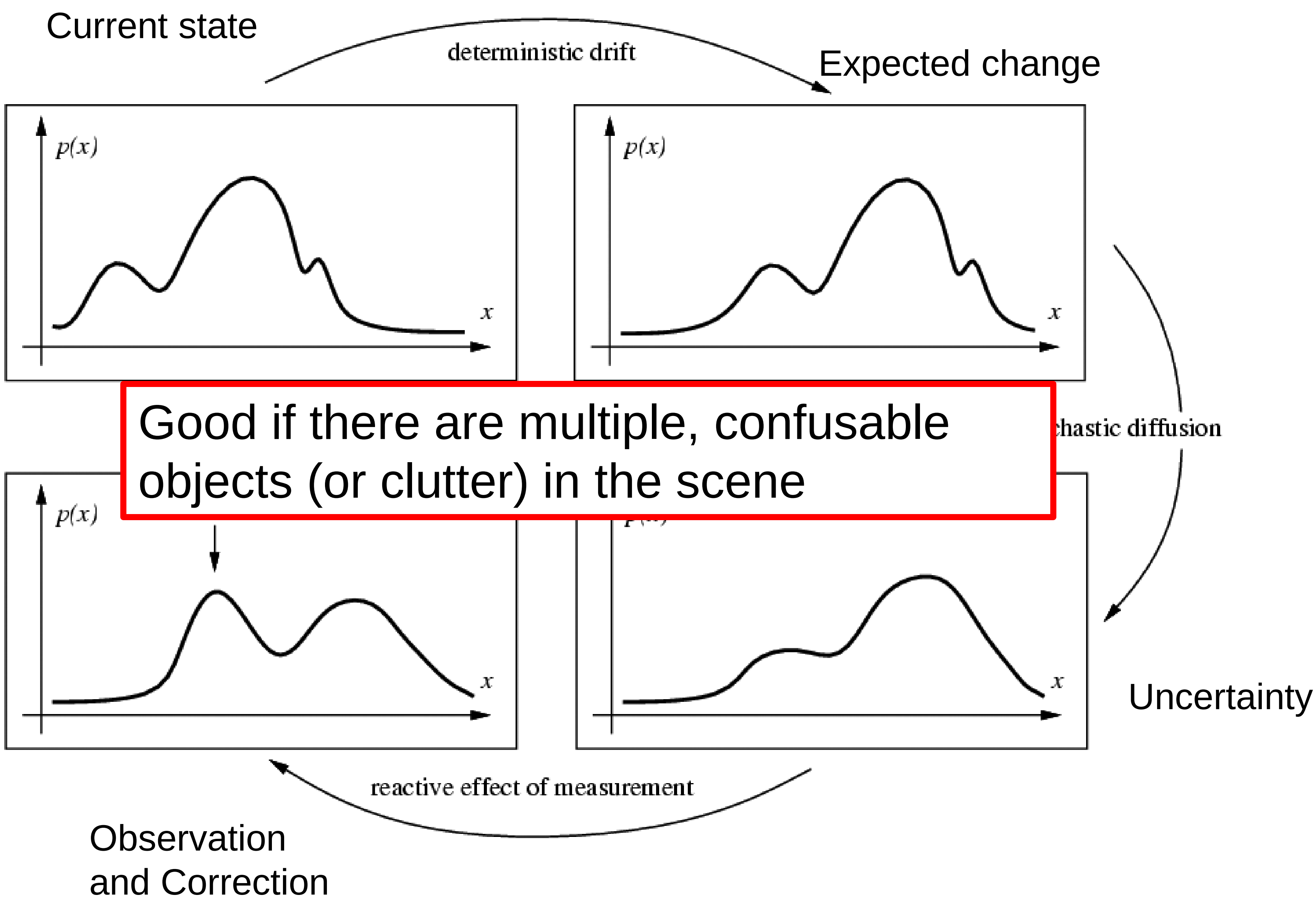
Spread due to randomness; this is predicted density $P(X_t|Y_{t-1})$

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

Propagation of non-parametric densities



Good if there are multiple, confusable objects (or clutter) in the scene

Particle filtering results

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

Tracking issues

- Initialization
  - Manual
  - Background subtraction
  - Detection

Tracking issues

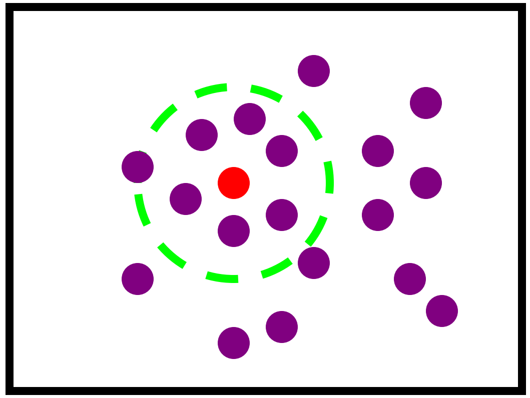
- Initialization
- Getting observation and dynamics models
  - Observation model: match a template or use a trained detector
  - Dynamics model: usually specify using domain knowledge

Tracking issues

- Initialization
- Obtaining observation and dynamics model
- Uncertainty of 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

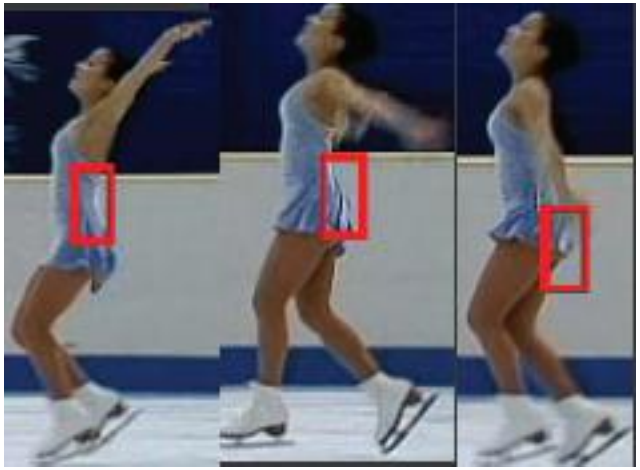
- Initialization
- Getting observation and dynamics models
- Prediction vs. correction
- Data association
  - When tracking multiple objects, need to assign right objects to right tracks (particle filters good for this)



Tracking issues

- Initialization
- Getting observation and dynamics models
- Prediction vs. correction
- Data association
- Drift
  - Errors can accumulate over time

Drift



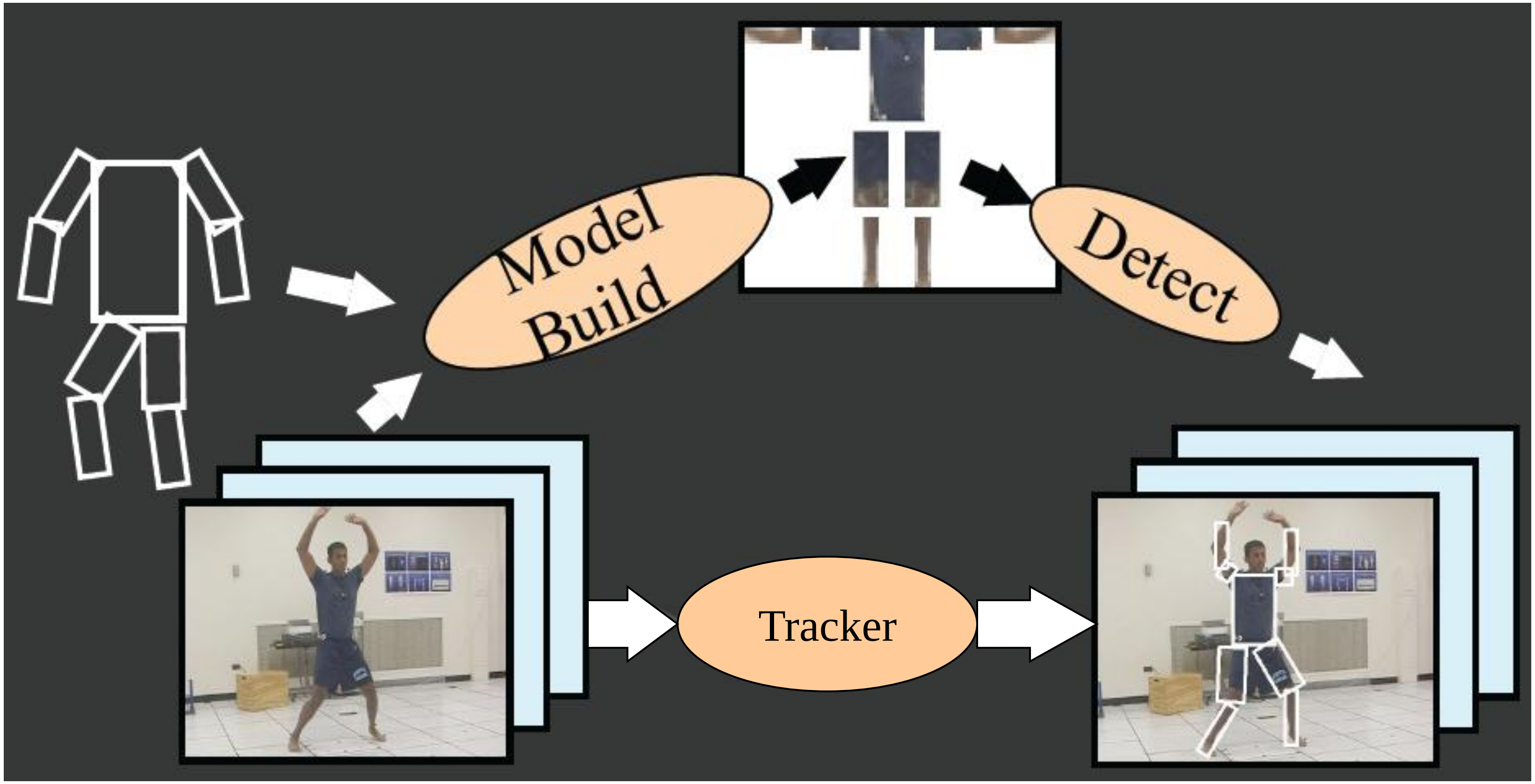
D. Ramanan, D. Forsyth, and A. Zisserman. [Tracking People by Learning their Appearance](). PAMI 2007.

Example: Tracking people

- Person model = appearance + structure (+ dynamics)
- Structure and dynamics are general, 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)

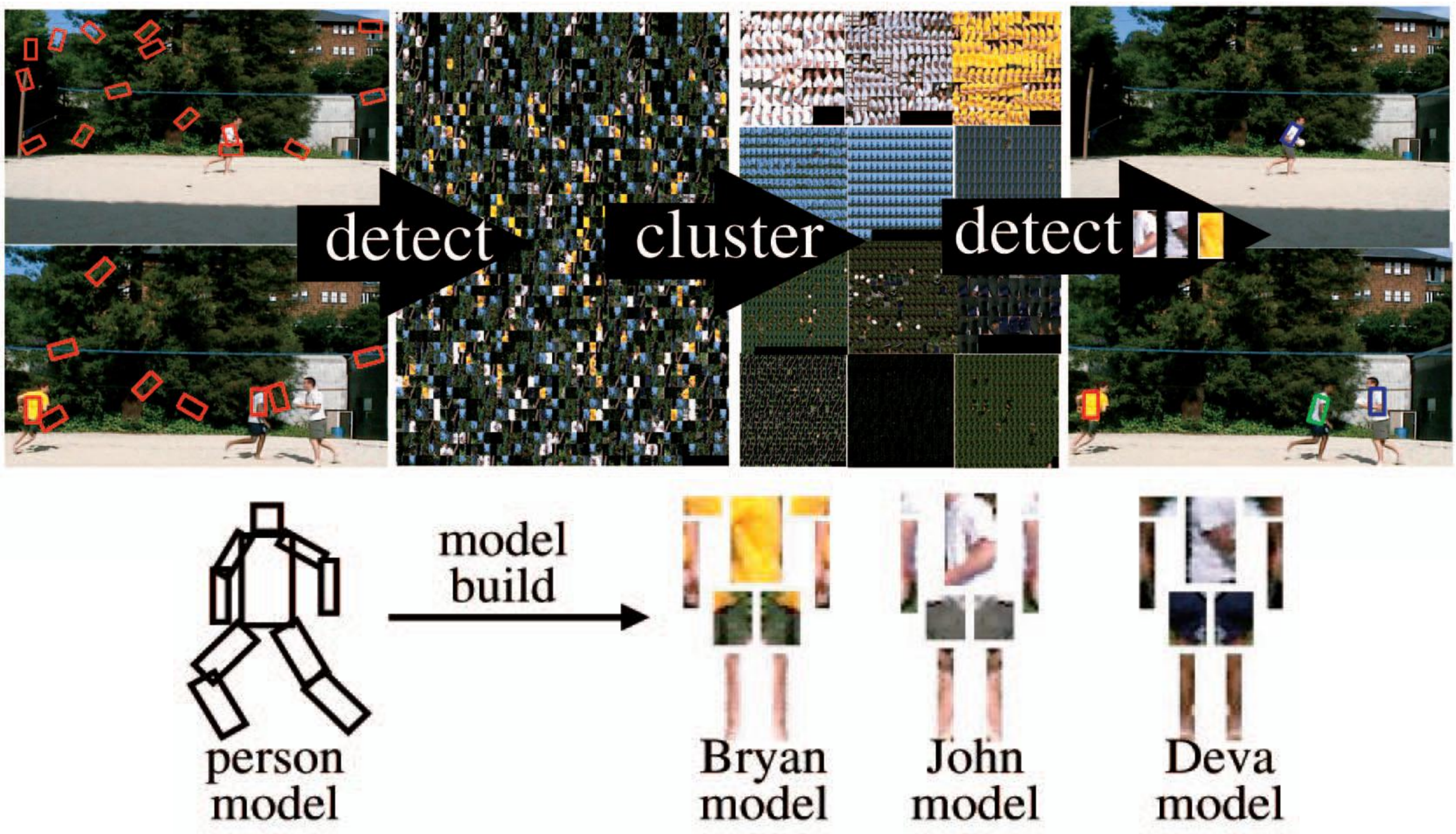
D. Ramanan, D. Forsyth, and A. Zisserman. Tracking People by Learning their Appearance. PAMI 2007.

Tracking people by learning their appearance



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

Bottom-up method to build model: Cluster



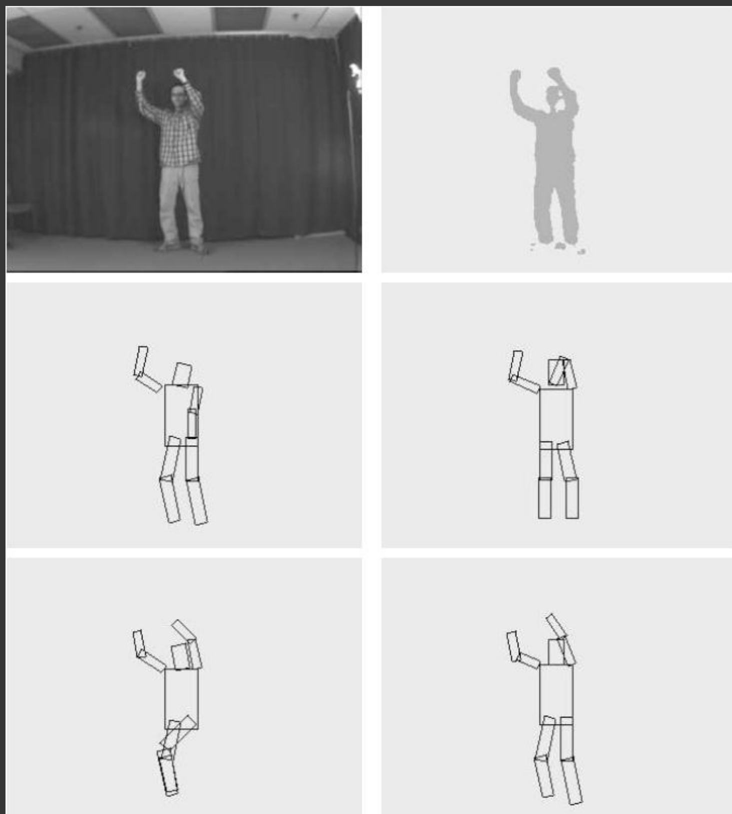
D. Ramanan, D. Forsyth, and A. Zisserman. [Tracking People by Learning their Appearance](). PAMI 2007.

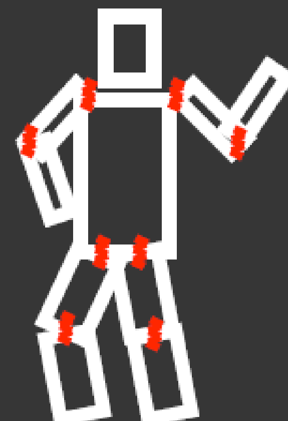
Top-down method to build model:
Exploit "easy" poses

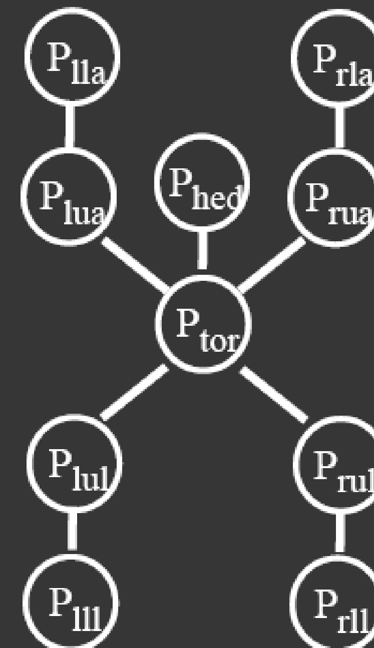
D. Ramanan, D. Forsyth, and A. Zisserman. [Tracking People by Learning their Appearance](). PAMI 2007.

Pictorial structure model

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







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

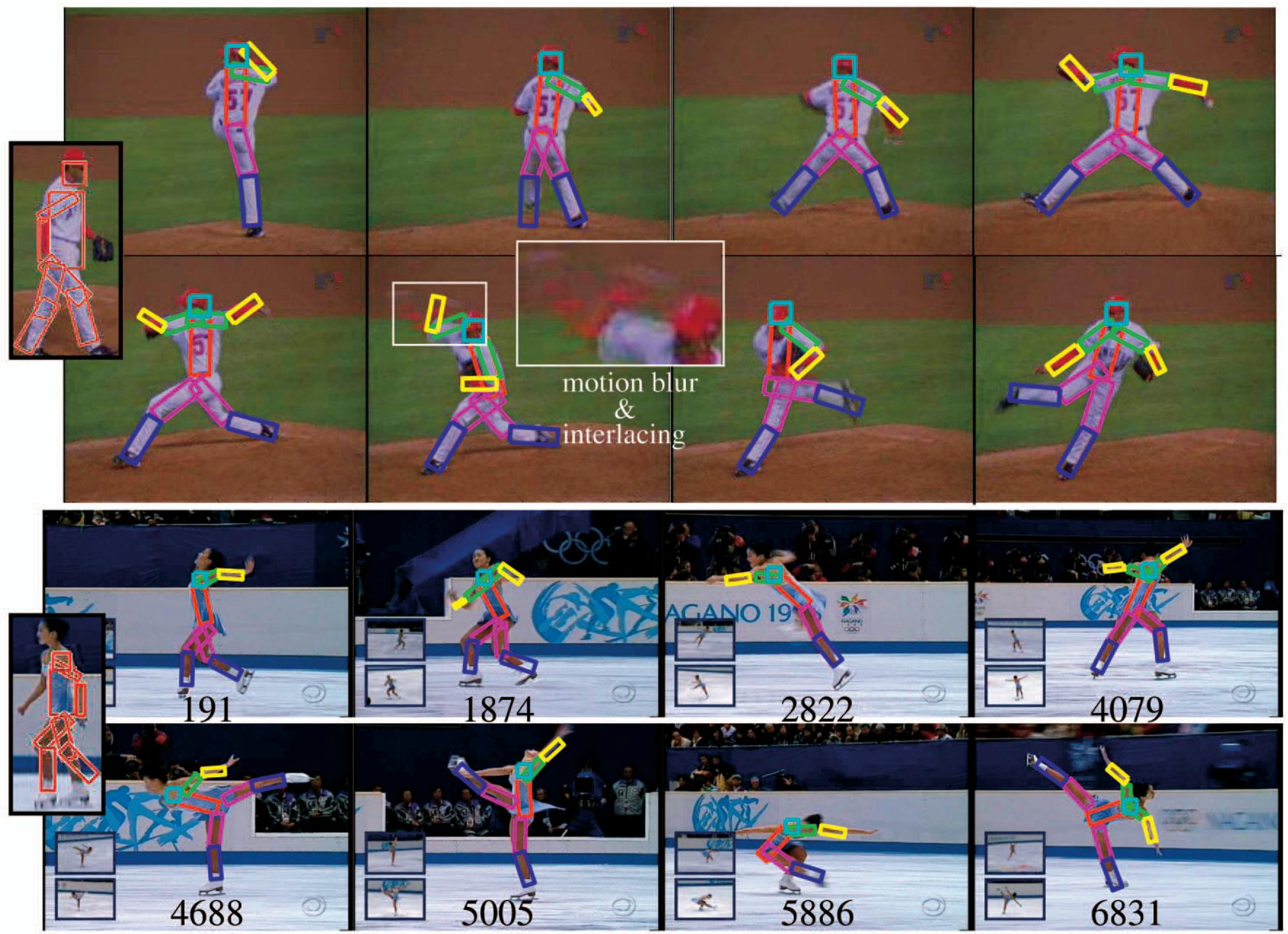
part geometry

part appearance

Temporal model

- Parts cannot move too far

Example results



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

Video

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

Things to remember

- Tracking objects = detection + prediction

- Probabilistic framework
  - Predict next state
  - Update current state based on observation

- Two simple but effective methods
  - Kalman filters: Gaussian distribution
  - Particle filters: multimodal distribution

Next class: action recognition

- Action recognition
  - What is an “action”?
  - How can we represent movement?
  - How do we incorporate motion, pose, and nearby objects?