Robotic 3D Vision

Lecture 17: 3D Object Tracking

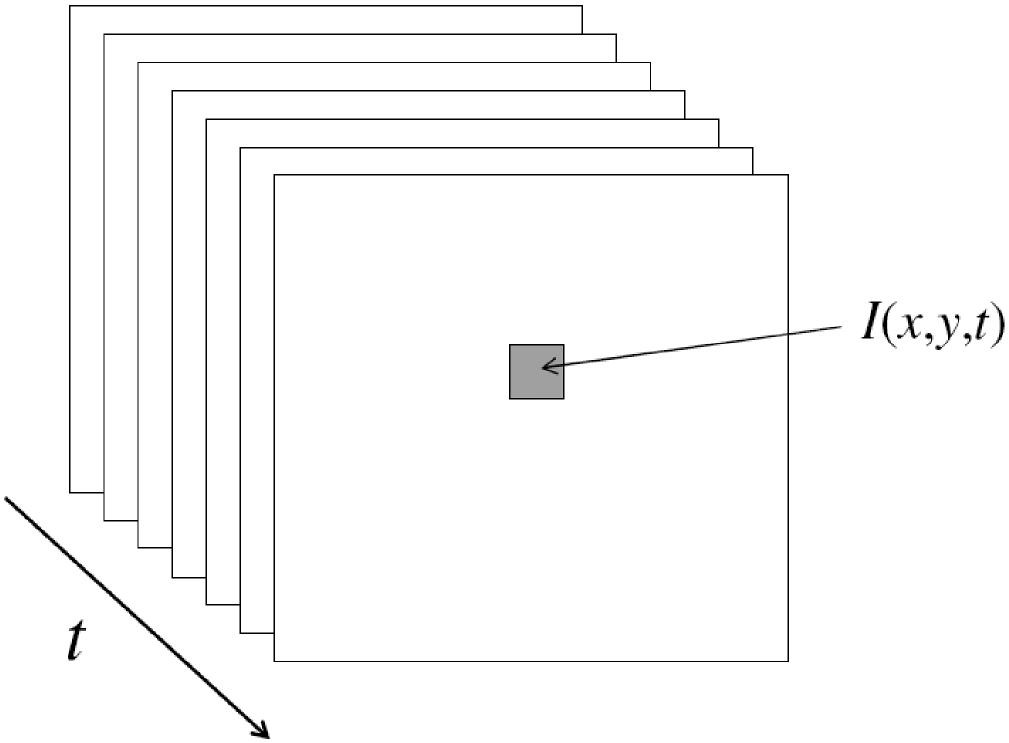
Prof. Dr. Jörg Stückler
Computer Vision Group, TU Munich
http://vision.in.tum.de

What We Will Cover Today

- Taxonomy of object tracking methods
- 3D object tracking using signed distance functions
- Multi-object tracking based on filtering and gated nearest neighbor data association

Motion Requires Video

- A video is a sequence of frames captured over time
- Our image data is a function of space (x, y) and time (t)



Slide credit: Svetlana Lazebnik

What is Object Tracking?

- Goal
  - *Estimate the number and state of objects in a region of interest*

- Number
  - 1: Single-target tracking
  - 0 or 1: Detection and tracking
  - N: Multi-target detection and tracking

Slide adapted from Robert Collins

What is Object Tracking?

- Goal
  - *Estimate the number and state of objects in a region of interest*
- State
  - We are using the term state to describe a vector of quantities that characterize the object being tracked.

E.g.	[x, y]	(location)
	[x, y, dx, dy]	(location + velocity)
	[x, y, appearance-params]	(location + appearance)

  - Because observations are typically noisy, estimating the state vector is a statistical estimation problem.

Slide adapted from Robert Collins

What is Object Tracking?

- Goal
  - *Estimate the number and state of objects in a region of interest*

- Objects
  - Variety of objects to track (including persons)
  - 3D tracking: Tracking the camera pose wrt. the object
  - Articulated tracking: e.g. tracking body pose

Image sources: Kristen Grauman, Michael Breitenstein, Ahmed Elgammal

What is Object Tracking?

- Goal
  - *Estimate the number and state of objects in a region of interest*

- What distinguishes tracking from "typical" statistical estimation (or machine learning) problems?
  - Typically a strong temporal component is involved.
  - Estimating quantities that are expected to change over time (thus, expectations of the dynamics play a role).
  - Interested in current state X_t for a given time step t.
  - Usually assume that we can only compute information seen at previous time steps 1, 2, ..., t-1. (*Can't look into the future!*)
  - Usually we want to be as efficient as possible, even "real-time".

$\Rightarrow$ These concerns lead naturally to recursive estimators.

Slide credit: Robert Collins

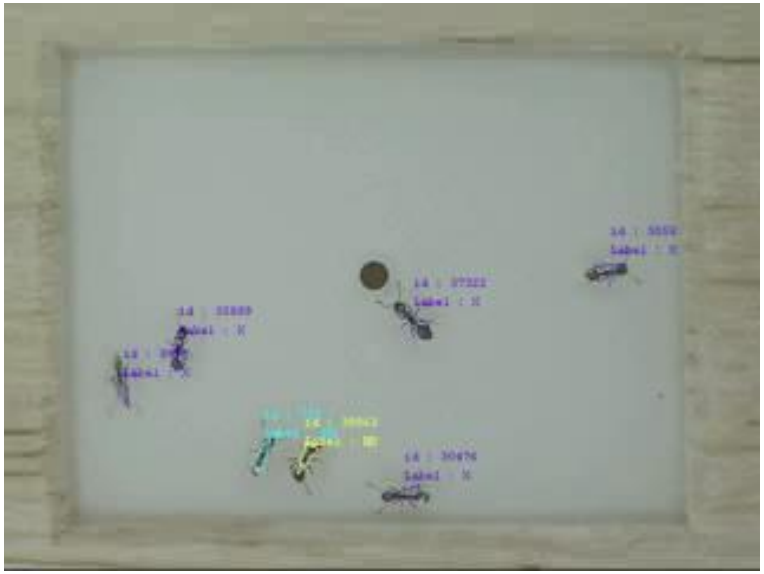
Types of Tracking

- Single-object tracking
  - Focuses on tracking a single target in isolation.

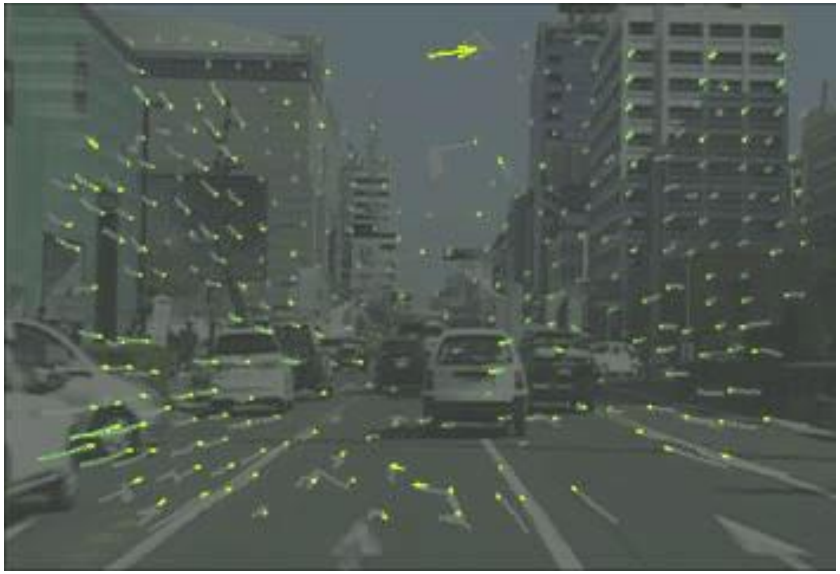
[Z. Kalal, K. Mikolajczyk, J. Matas, PAMI'10]

Types of Tracking

- Multi-object tracking
  - Tries to follow the motion of multiple objects simultaneously.



Ant behavior, courtesy of Georgia Tech biotracking

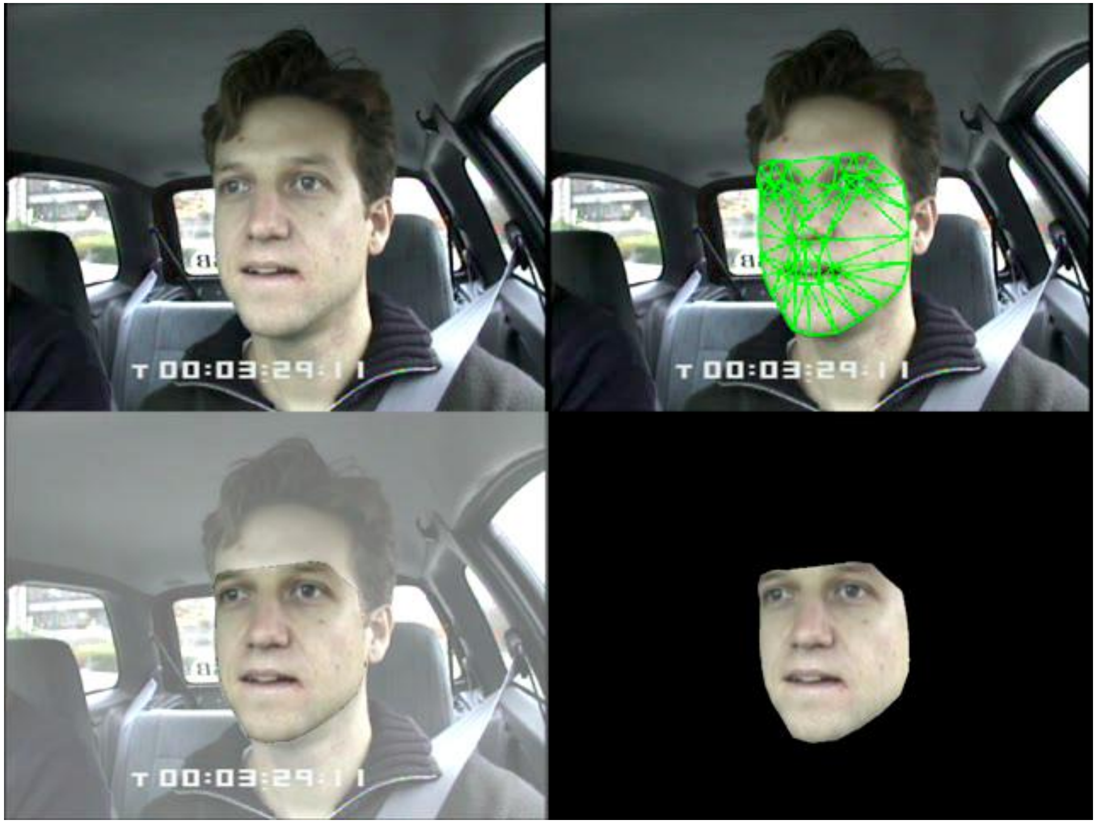


“Objects” can be corners, and tracking gives us optical flow.

Slide credit: Robert Collins

Types of Tracking

- Articulated tracking
- Tries to estimate the motion of objects with multiple, coordinated parts



[I. Matthews, S. Baker, IJCV'04]

Slide credit: Robert Collins

Types of Tracking

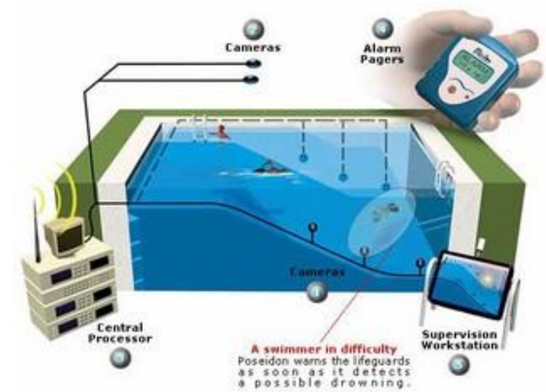
- Active tracking
  - Involves moving the sensor in response to motion of the target. Needs to be real-time!

Slide credit: Robert Collins

Applications: Safety & Security

Autonomous robots

Driver assistance



Monitoring pools
(Poseidon)

Pedestrian detection
[MERL, Viola et al.]

Surveillance

Slide credit: Kristen Grauman

Applications: Human-Computer Interaction

Games
(Microsoft Kinect)

Assistive technology systems
Camera Mouse
(Boston College)

Slide adapted from Kristen Grauman

Applications: Visual Effects

MoCap for *Pirates of the Carribean*, Industrial Light and Magic

Slide adapted from Steve Seitz, Svetlana Lazebnik, Kristen Grauman

Why Are There So Many Papers on Tracking?

- Because what kind of tracking “works” depends on problem-specific factors...

image source: Microsoft

Factors: Discriminability

- How easy is it to discriminate one object from another?

Appearance models can do all the work

Constraints on geometry and motion become crucial

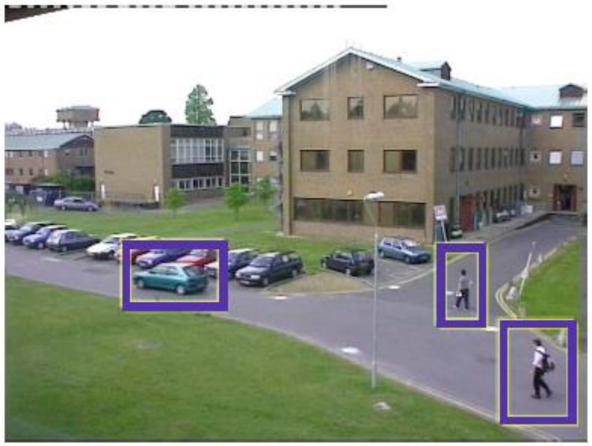
Slide credit: Robert Collins

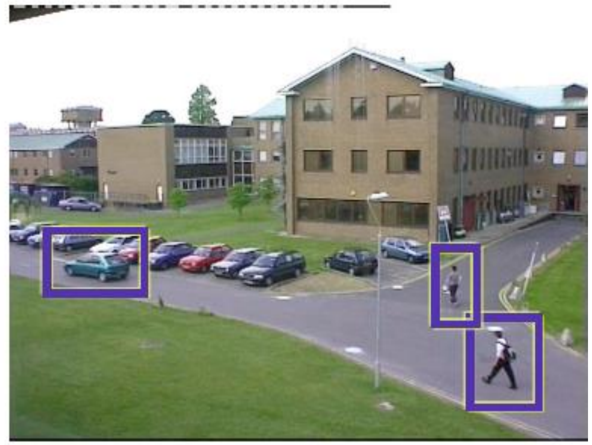
Factors: Frame Rate

frame n

frame n+1

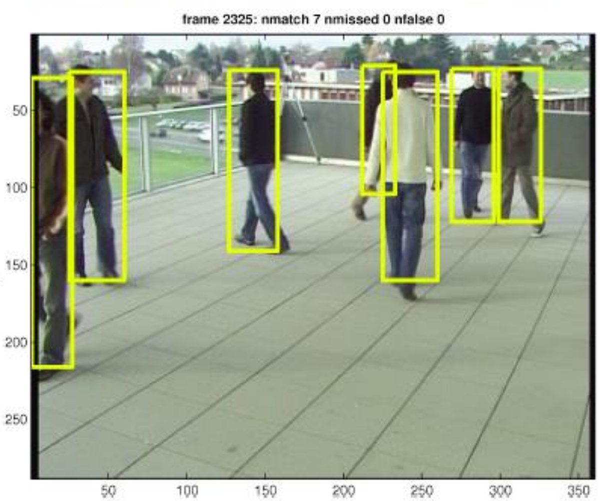
HIGH

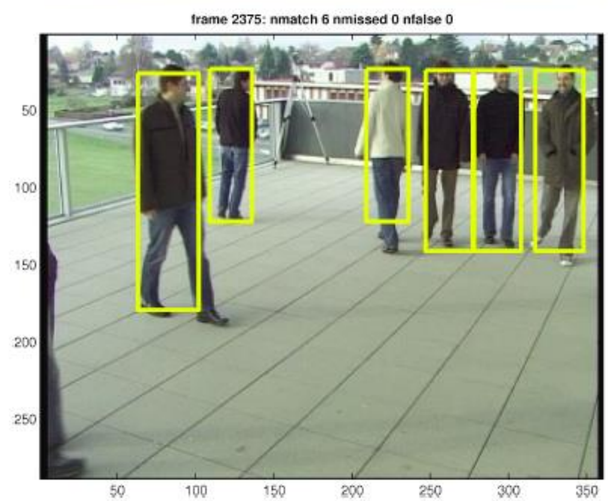




Gradient ascent (e.g. mean-shift) works OK

LOW





Much harder search problem. Good data association becomes crucial.

Slide credit: Robert Collins

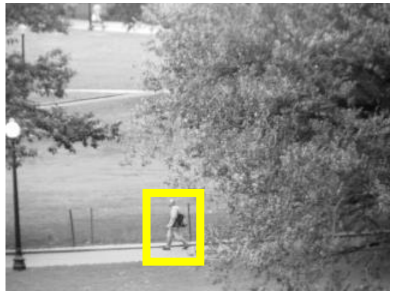
Other Factors

- Single target *vs.* multiple targets
- Single camera *vs.* multiple cameras
- On-line *vs.* batch mode
- Do we have a good generic detector?
  (e.g., faces, pedestrians)
- Does the object have multiple parts?
- ...

Slide credit: Robert Collins

Elements of Tracking

t=1



t=2

…

t=20

t=21

- Detection
  - Find the object(s) of interest in the image.

Image credit: Kristen Grauman

Elements of Tracking

t=1 t=2 … t=20 t=21

- Detection
  - Find the object(s) of interest in the image.
- Association
  - Determine which observations come from the same object.

Image credit: Kristen Grauman

Elements of Tracking

t=1 t=2 … t=20 t=21

- Detection
  - Find the object(s) of interest in the image.
- Association
  - Determine which observations come from the same object.
- Prediction
  - Predict future motion based on the observed motion pattern.
  - Use this prediction to improve detection and data association in later frames.

Image credit: Kristen Grauman

Elements of Tracking

t=1 t=2 … t=20 t=21

- Detection
  - Find the object(s) of interest in the image.
- Association
  - Determine which observations come from the same object.
- Prediction
  - Predict future motion based on the observed motion pattern.
  - Use this prediction to improve detection and data association in later frames.

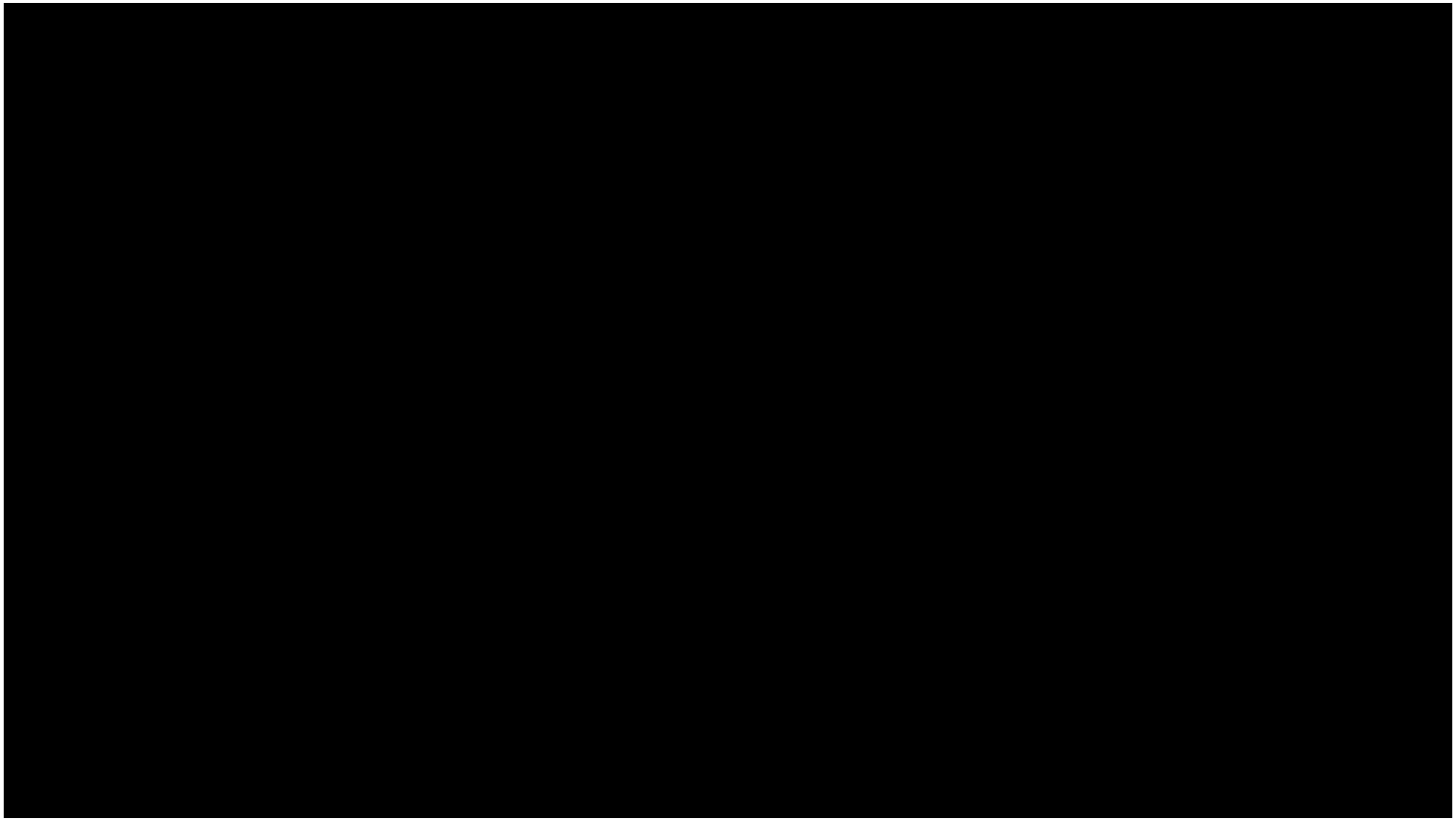
Image credit: Kristen Grauman

3D Object Tracking Approaches

- This lecture:
  - Focus on single-object tracking
  - 6-DoF pose tracking of objects
  - Tracking a known object model (model-based 3D tracking)

- Strategy 1: Tracking-by-detection
  - Detect object in each frame individually
- Strategy 2: Tracking-by-filtering
  - Detect object as measurement within probabilistic filter
- Strategy 3: Tracking-by-registration
  - From an initial guess (detection) perform incremental registration

Model-based Tracking-by-Registration



Tracking-by-Registration

- Consider the following approach:
  - Describe object as a set of points $\mathcal{X} = \{\mathbf{x}_i\}_{i=1}^N$ in its reference frame
  - Optimize for the pose $\boldsymbol{\xi} \in se(3)$ that aligns object points with measurements $\mathcal{Y} = \{\mathbf{y}_j\}_{j=1}^M$ at each time step

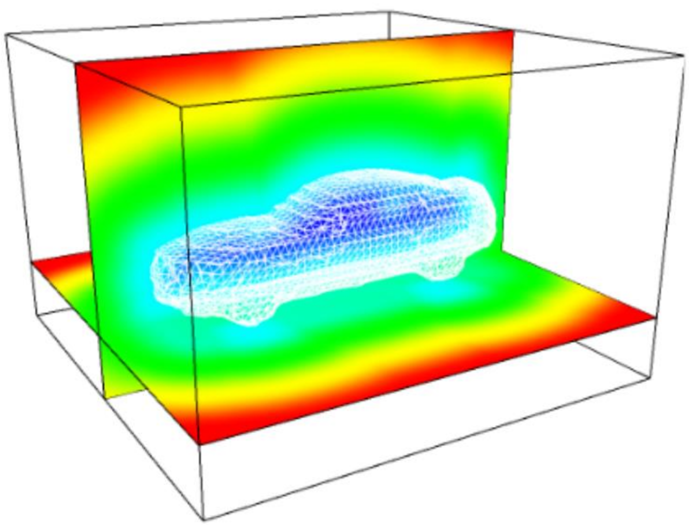
$$E(\boldsymbol{\xi}) = \frac{1}{2} \sum_{(i,j) \in \mathcal{C}} \left\| \mathbf{x}_i - \mathbf{y}_j \right\|_2^2$$

  - Non-linear least squares…
- However this requires to decide
  - which scene points belong to the object (segmentation)
  - which object and scene points correspond to each other
- Could be solved using an ICP-like approach

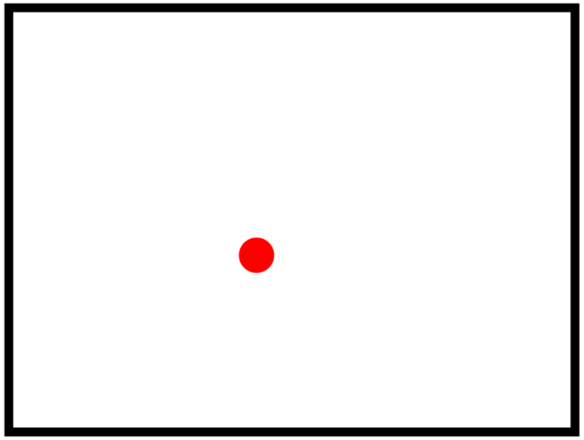
Tracking-by-Registration using Signed Distance Functions

- Represent object model with 3D signed distance function (SDF)
- SDF $\Phi(\mathbf{x}) \mapsto \mathbb{R}$ maps 3D points to their closest distance to object surface
- Sign of the distance specifies "inside" or "outside" of object
- Can be represented and precomputed in a 3D voxel grid
- The surface of the object is given by the zero level-set $\Phi(\mathbf{x}) = 0$
- Ideally, the measured points are on the surface
- We can define the error function as

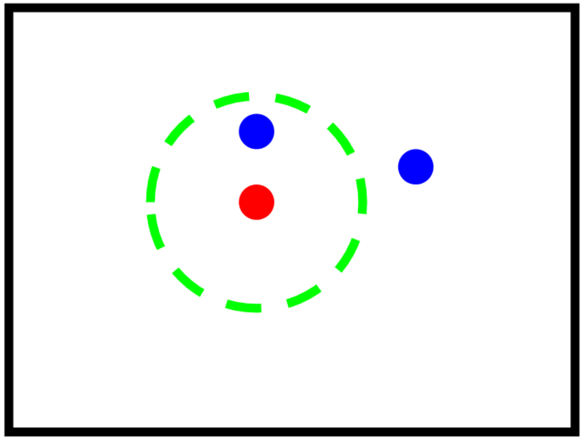
$$E(\boldsymbol{\xi}) = \frac{1}{2} \sum_{i=1}^{N} \Phi(\mathbf{T}(\boldsymbol{\xi})\overline{\mathbf{x}}_i)^2$$



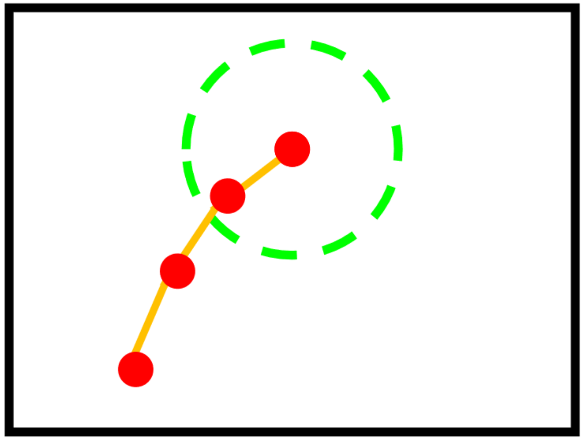
Elements of Tracking



Detection



Data association

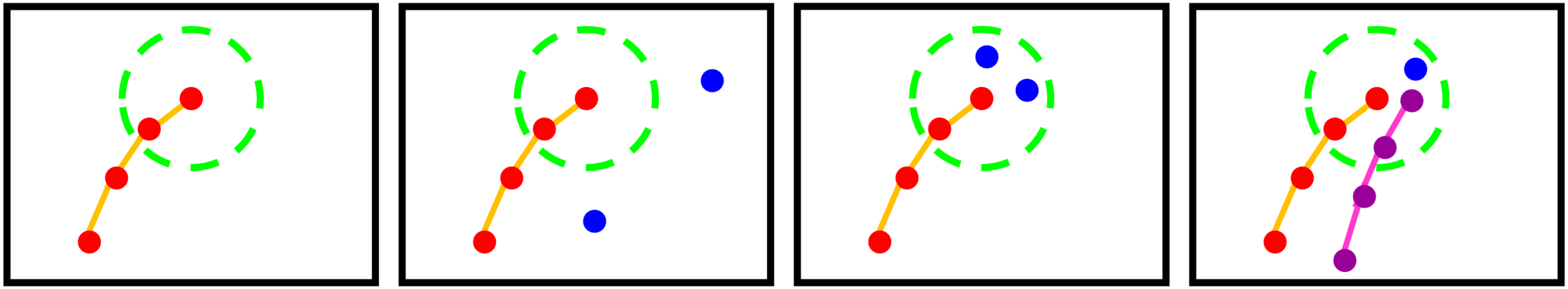


Prediction

- Detection: Where are candidate objects?
- Data association: Which detections belong to the same object?
- Prediction: Where will a tracked object be in the next time step?

Slide credit: Bastian Leibe

What Makes Multi-Object Tracking Difficult?

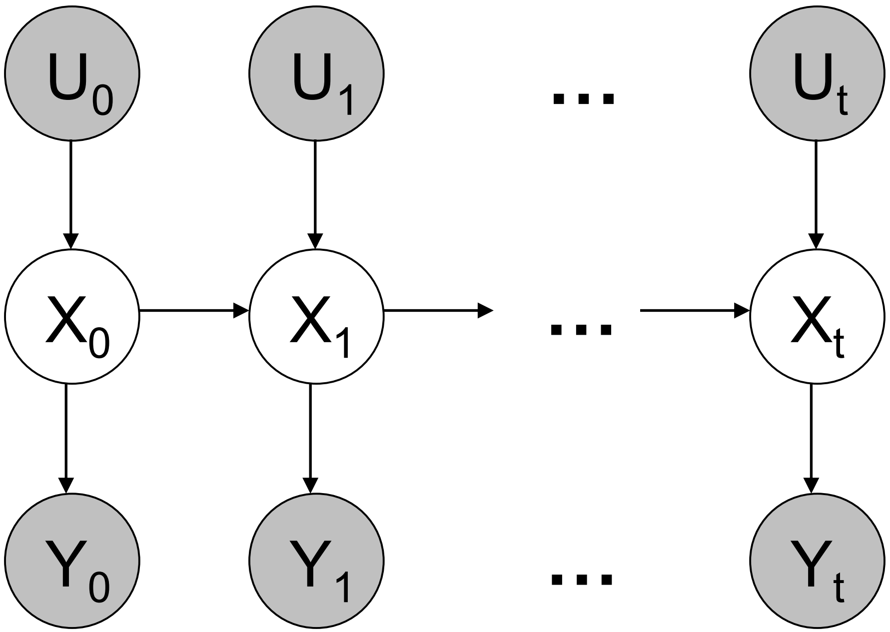


- Predictions may not be supported by detections
  - Occlusion or end of track?
- Unexpected measurements
  - New objects or outliers?
- Correspondence ambiguity for a prediction
  - Which measurement is the correct one?
- Correspondence ambiguity for a measurement
  - Which object track shalle the measurement belong to?

Slide credit: Bastian Leibe

Recap: Probabilistic Model of Time-Sequential Processes

- Hidden state X gives rise to noisy observations Y
- At each time t,
  - the state changes stochastically from X_{t-1} to X_t
  - state change depends on action U_t
  - we get a new observation Y_t



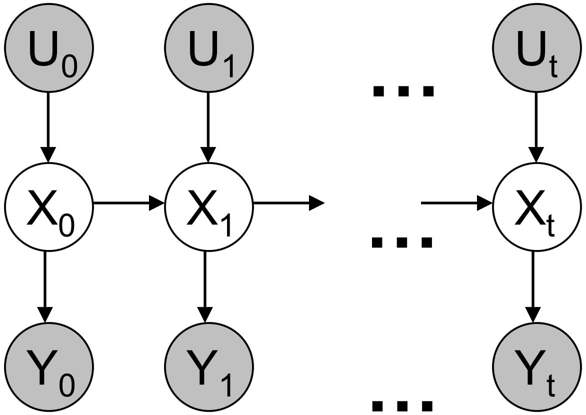
Recap: Markov Assumptions

- Only the immediate past matters for a state transition

$$p\left(X_t \mid X_{0:t-1}, U_{0:t}\right) = \boxed{p\left(X_t \mid X_{t-1}, U_t\right)}$$ **state transition model**

- Observations depend only on the current state

$$p\left(Y_t \mid X_{0:t}, U_{0:t}, Y_{0:t-1}\right) = \boxed{p\left(Y_t \mid X_t\right)}$$ **observation model**



Recap: Predict-Correct Cycle

- Prediction:

$$p\left(X_t \mid y_{0:t-1}, u_{0:t}\right) = \int p\left(X_t \mid X_{t-1}, u_t\right) p\left(X_{t-1} \mid y_{0:t-1}, u_{0:t-1}\right) dX_{t-1}$$

observation

y_t

action

u_t

- Correction:

$$p\left(X_t \mid y_0, \ldots, y_t\right) = \frac{p\left(y_t \mid X_t\right) p\left(X_t \mid y_{0:t-1}, u_{0:t}\right)}{\int p\left(y_t \mid X_t\right) p\left(X_t \mid y_{0:t-1}, u_{0:t}\right) dX_t}$$

Multi-Object Tracking by Filtering

- Approach: probabilistic filtering of position, velocity, etc. of each object track (state) $\mathbf{x}$ based on measurements

$$\mathcal{Y} = \{\mathbf{y}_1, \ldots, \mathbf{y}_M\}$$

- Data association before correction step
  - How?

- Unassociated measurements create new tracks

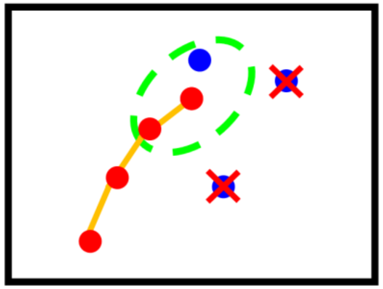
- Discard tracks that cannot be associated to measurements

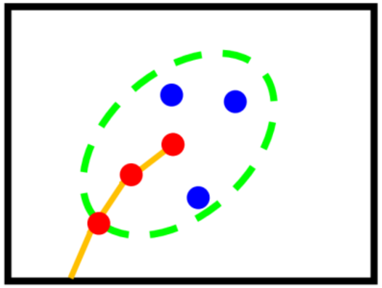
Gating Nearest Neighbor Data Association

- Gating
  - Only consider measurements within a certain area around the predicted location

  ⇒ Large gain in efficiency, since only a small region needs to be searched

- Nearest Neighbor Association
  - Among the candidates in the gating region, only take the one closest to the prediction





Slide adapted from Bastian Leibe

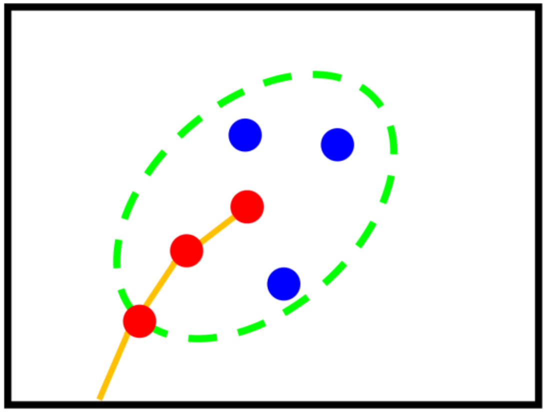
Gating with Mahalanobis Distance

- Recall: Kalman Filter
  - Maintains a Gaussian state estimate $\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t$

- Perform gating based on the distribution of the "innovation"

$$\mathcal{N}(\mathbf{y}_t - h(\boldsymbol{\mu}^-{}_t), \boldsymbol{\Sigma}_{m_t} + \mathbf{H}_t \boldsymbol{\Sigma}_t^- \mathbf{H}_t^\top$$

- Gating volume is ellipsoidal
- E.g. choose volume that corresponds to 95% of probability mass
- Side note: Mahalanobis distance is χ^2-distributed, look up threshold in χ^2-distribution table



Slide adapted from Bastian Leibe

Problems with NN Assignment

- Limitations
  - For NN assignments, there is always a finite chance that the association is incorrect, which can lead to serious effects

  $\Rightarrow$ If a Kalman filter is used, a falsely assigned measurement may lead the filter to lose track of its target

  - The NN filter makes assignment decisions only based on the current frame
  - More information is available by examining subsequent images

  $\Rightarrow$ Data association decisions could be postponed until a future frame will resolve the ambiguity

Slide credit: Bastian Leibe

Other Multi-Object Tracking Approaches

- More powerful approaches
  - Multi-Hypothesis Tracking (MHT)
    - Well-suited for KF, EKF approaches
  - Joint Probabilistic Data Association Filters (JPDAF)
    - Well-suited for PF approaches

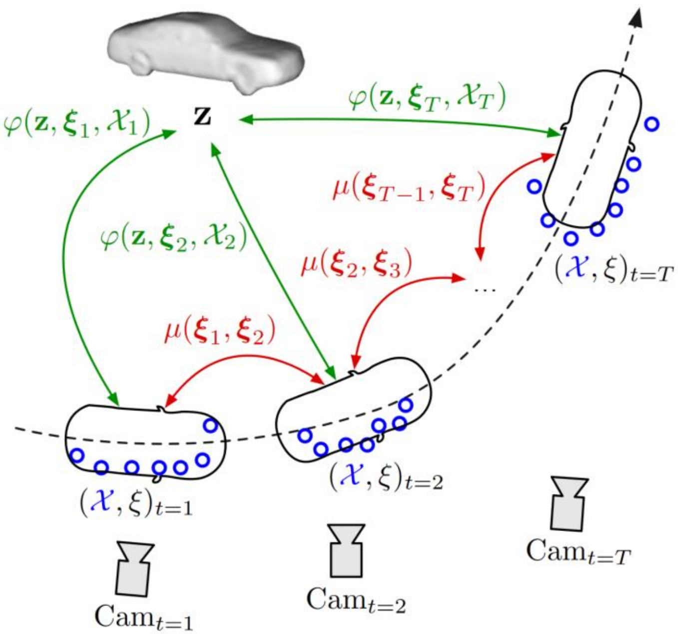
- Data association as convex optimization problem
  - Bipartite Graph Matching (Hungarian algorithm)
  - Network Flow Optimization

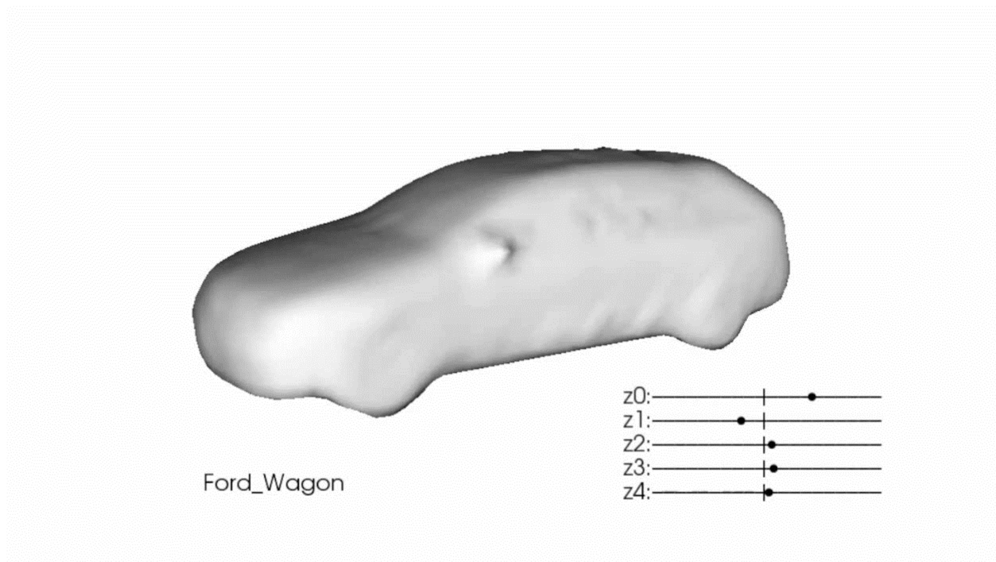
  => Efficient, globally optimal solutions for subclass of problems

Slide credit: Bastian Leibe

Shape Priors for 4D Stereo Reconstruction

Approach: impose shape and motion priors for spatio-temporal reconstruction of vehicles

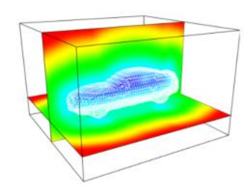




$$E(\mathbf{z}, \boldsymbol{\xi}) = \frac{1}{T} \sum_t [\underbrace{\mu(\boldsymbol{\xi}_t, \boldsymbol{\xi}_{t-1})}_{\text{motion term}} + \underbrace{\varphi(\mathbf{z}, \boldsymbol{\xi}_t, \mathcal{X}_t)}_{\text{data term}}] + \underbrace{\kappa(\mathbf{z})}_{\text{shape prior}}$$

Learned 3D SDF shape embedding

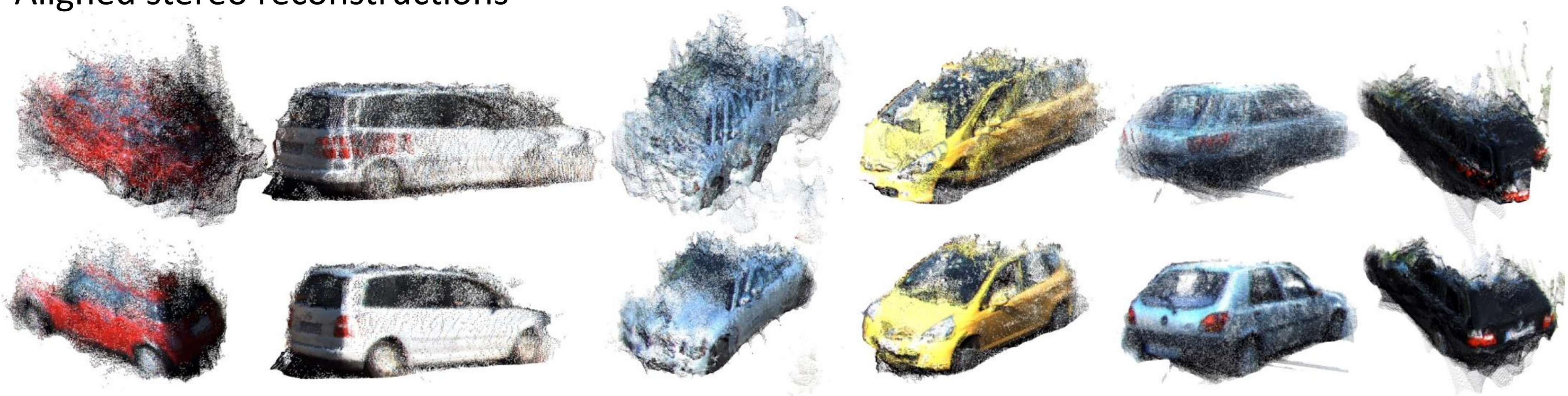
$$\varphi(\mathbf{z}, \boldsymbol{\xi}, \mathcal{X}) = \frac{1}{N} \sum_{\mathbf{x} \in \mathcal{X}} \rho\left(\frac{\phi_{\mathbf{z}}(\mathbf{T}_{\boldsymbol{\xi}}\, \mathbf{x})}{\sigma_{d_{\mathbf{x}}}}\right)$$



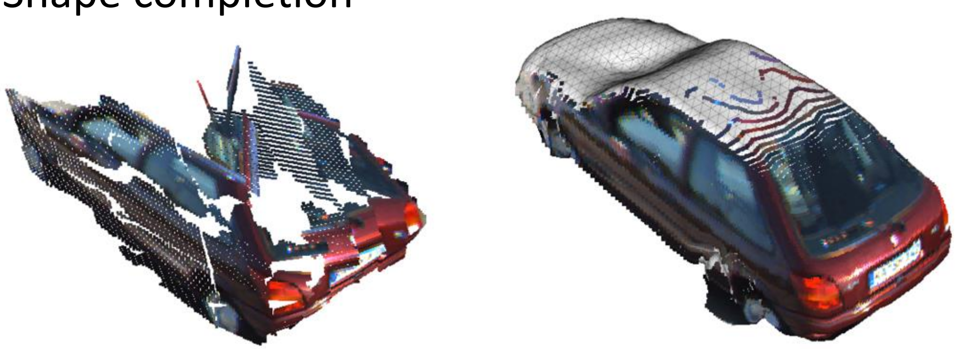
(Engelmann, S, Leibe, WACV 2017)

Shape Priors for 4D Stereo Reconstruction

Aligned stereo reconstructions



Shape completion



(Engelmann, S, Leibe, WACV 2017)

Lessons Learned Today

- Object tracking involves detection, motion estimation (prediction) and data association over time
- 3D object tracking of an object model through registration
  - ICP-based tracking-by-registration
  - SDF-based tracking-by-registration
- Multi-object tracking involves a harder data association problem
  - Gated Nearest Neighbor filter
  - More sophisticated methods f.e. based on convex optimization

Thanks for your attention!