Multi-object tracking.
Formulating the problem

- Goal: track objects in a video
- Input: observations per frame (O_i)
  - Image
- Output: State per object (S_i)
  - Location, speed, color
- Formally, maximize $P(S_{1:t}|O_{1:t})$.
  - Find optimal sequence of states
- Evaluate on false positives, false negatives, switches
- Upstream task for navigation

Video of tracking. Idea is that you want to maintain some sort of identity for each object.
Types of trackers

- Online vs. offline: can we see the future? Note: this is unrelated to the notion of
- Detection based vs. detection free: what is part of tracking?
- 2D vs. 3D: how do we model the world?

Q: (Hugh) - What is the difference between online vs. offline tracking?
Key operations:

- Predict: $P(S_{t+1}|S_t)$
- Update: $P(S_t|O_t)$
- Initiation: When have we seen a new object?
- Termination: When has an object left?

Predict:

- Want to model state of object
- Track position, velocity, acceleration, appearance and predict future states
- Filters (ie Kalman, particle) allow tracking with uncertainty

Example: Uber GPS tracking.
Update:

- Update motion models given object observation
- Which object is which?
- Worker assignment problem with Hungarian algorithm (well understood with cubic runtime).
- And dummy node for unmatched detections and tracks

Cost matrices

- How good is a track-detection pair?
- IOU of states
- Visual appearance similarity
  - Color histogram: well studied but loses spatial info
  - Optical flow: powerful but suceptible to occlusion
  - Learned approach: embed into some vector space
- How do we combine?
  - Concatenate, sum, product, cascade, learned

Initiation:

- High recall: look for n matches in a row before confirming object
- High precision detector: always initiate on detection
- Estimate start state (optical flow with velocity)

Termination:

- High recall: miss for 1 frame
- High precision: miss for n frames
- State based: uncertainty limit based approach

Open ended questions:

- Hard constraints: cars can't exist in a wall
- Social cues: objects respond to each other
- Occlusion handling: how are states and cost matrices affected?

Tracking on JR

- Cute robot
- 2D stereo camera, 3D LIDAR data

- Drives around people and tries not to hit them

References:

- Probabilistic robotics
- The Gradient's 3D perception / vision article

Slides
Potential work