Tracking

CSE 6367 – Computer Vision
Vassilis Athitsos
University of Texas at Arlington

What Is Tracking?

What Is Tracking?

- We are given:
  - the *state* of one or more objects in the previous frame.
- We want to estimate:
  - the *state* of those objects in the current frame.

What Is Tracking?

- We are given:
  - the *state* of one or more objects in the previous frame.
- We want to estimate:
  - the *state* of those objects in the current frame.
- State can be:
  - Location.
  - Velocity.
  - Shape.
  - Orientation, scale, 3D orientation, 3D position, …

Why Do We Care About Tracking?

Why Do We Care About Tracking?

- Improves speed.
  - We do not have to run detection at all locations, all scales, all orientations.

Why Do We Care About Tracking?

- Improves speed.
  - We do not have to run detection at all locations, all scales, all orientations.
- Allows us to establish correspondences across frames.
  - Provides representations such as “the person moved left”, as opposed to “there is a person at (i1, j1) at frame 1, and there is a person at (i2, j2) at frame 2”.
  - Needed in order to recognize gestures, actions, activity.

Example Applications

- Activity recognition/surveillance.
  - Figure out if people are coming out of a car, or loading a truck.
- Gesture recognition.
  - Respond to commands given via gestures.
  - Recognize sign language.
- Traffic monitoring.
  - Figure out if any car is approaching a traffic light.
  - Figure out if a street/highway is congested.
- In all these cases, we must track objects across multiple frames.

Related Problem: Motion Estimation

- Different versions:
  - For every pixel in frame t, what is the corresponding pixel in frame t+1?
  - For every object in frame t, what is the corresponding region in frame t+1?
  - How did a specific pixel, region, or object, move?
- If we know the answers to the above questions, tracking is easy.
- Tracking is inextricably connected with motion estimation.

Estimating Motion of a Block

- What is a block?
  - A rectangular region in the image.
  - In other words, an image window.
- Given a block at frame t, how can we figure out where the block moved to at frame t+1?

Estimating Motion of a Block

- What is a block?
  - A rectangular region in the image.
  - In other words, an image window.
- Given a block at frame t, how can we figure out where the block moved to at frame t+1?
- Simplest method: normalized correlation.

Tracking Main Loop

1. read current frame.
2. find best match of object in current frame.
3. (optional) update object description.
4. advance frame counter.

5 goto 1.

- What is missing to make this framework fully automatic?

Initialization

1. read current frame.
2. find best match of object in current frame.
3. (optional) update object description.
4. advance frame counter.

5 goto 1.

- What is missing to make this framework fully automatic?
- Detection/initialization:
  - find the object, obtain an initial object description.

Initialization

1. read current frame.
2. find best match of object in current frame.
3. (optional) update object description.
4. advance frame counter.

5 goto 1.

- Tracking methods ignore the initialization problem.
- Any detection method can be used to address that problem.

Source of Efficiency

1. read current frame.
2. find best match of object in current frame.
3. (optional) update object description.
4. advance frame counter.

5 goto 1.

- Why exactly is tracking more efficient than detection? In what lines is that used?

Source of Efficiency

1. read current frame.
2. find best match of object in current frame.
3. (optional) update object description.
4. advance frame counter.

5 goto 1.

- Why exactly is tracking more efficient than detection? In what lines is that used?
  - Line 2. Finding best match is faster because:
    - We can use simpler detection methods.
      - We know very precisely what the object looks like.
    - We search few locations, few scales, few orientations.

Updating Object Description

1. read current frame.
2. find best match of object in current frame.
3. (optional) update object description.
4. advance frame counter.

5 goto 1.

- How can we change our implementation to update the object description?

Updating Object Description

1. read current frame.
2. find best match of object in current frame.
3. (optional) update object description.
4. advance frame counter.

5 goto 1.

- How can we change our implementation to update the object description?
  - Update the *block* variable, based on the match found at the current frame.

Drifting

1. read current frame.
2. find best match of object in current frame.
3. (optional) update object description.
4. advance frame counter.

5 goto 1.

- The estimate can be off by a pixel or so at each frame.
  - Sometimes larger errors occur.
- If we update the appearance, errors can accumulate.

Changing Appearance

- Sometimes the appearance of an object changes from frame to frame.
  - Example: left foot and right foot in *walkstraight* sequence.
- If we do not update the object description, at some point the description is not good enough.
- Avoiding drift while updating the appearance are conflicting goals.

Occlusion

- The object we track can temporarily be occluded (fully or partially) by other objects.
- If appearance is updated at each frame, when the object is occluded it is unlikely to be found again.

Improving Tracking Stability

- Check every match using a detector.
  - If we track a face, then the best match, in addition to matching the correlation score, should also have a good detection score using a general face detector.
  - If the face is occluded, the tracker can figure that out, because no face is detected.
  - When the face reappears, the detector will find it again.

Improving Tracking Stability

- Remembering appearance history.
  - An object may have a small number of possible appearances.
    - The appearance of the head depends on the viewing angle.
  - If we remember each appearance, we minimize drifting.
    - When the current appearance is similar to a stored appearance, we do not need to make any updates.

Improving Tracking Stability

- Multiple hypothesis tracking.
  - Real-world systems almost always maintain multiple hypotheses.
  - This way, when the right answer is not clear (e.g., because of occlusions), the system does not have to commit to a single answer.