ISTA 352

Lecture 14

Multiple views in space and time (fancy cameras)

Administrivia

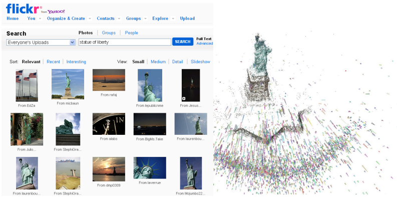
- Don’t forget to follow instructions handing in you homework
- Quiz next Friday
  - You can use a single sheet of notes (one side)
  - I aim to get a practice quiz out this weekend
  - Material through next lecture (Monday, September 24)
- Monday is our first guest lecture (Mary Peterson)
  - It is important that you attend
  - Guest lecture material will show up on quizzes

Fancy digital cameras

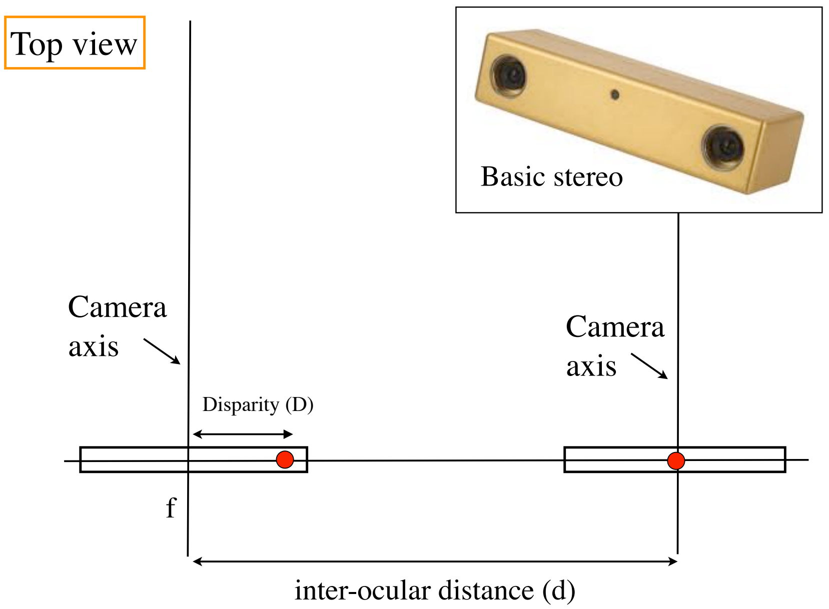
Basic stereo

The kinect

Movie camera



The internet camera



Top view

Basic stereo

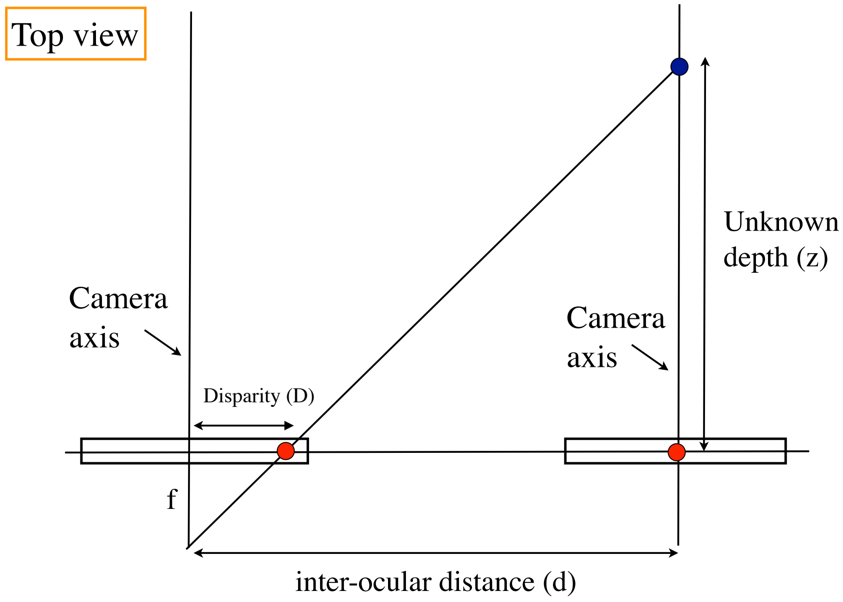
Camera axis

Camera axis

Disparity (D)

f

inter-ocular distance (d)



Top view

Camera axis

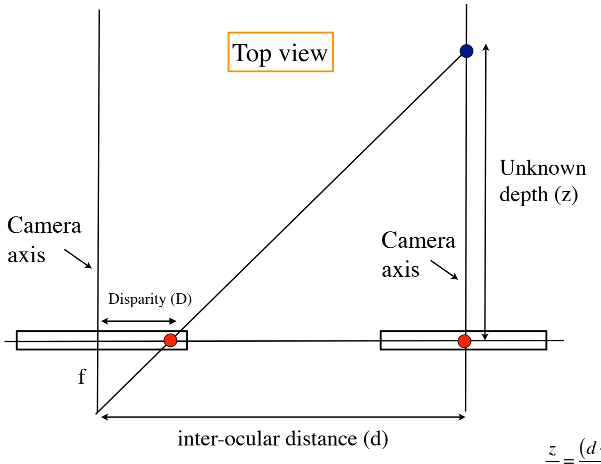
Camera axis

Unknown depth (z)

Disparity (D)

f

inter-ocular distance (d)



Top view

Camera axis

Camera axis

Unknown depth (z)

Disparity (D)

f

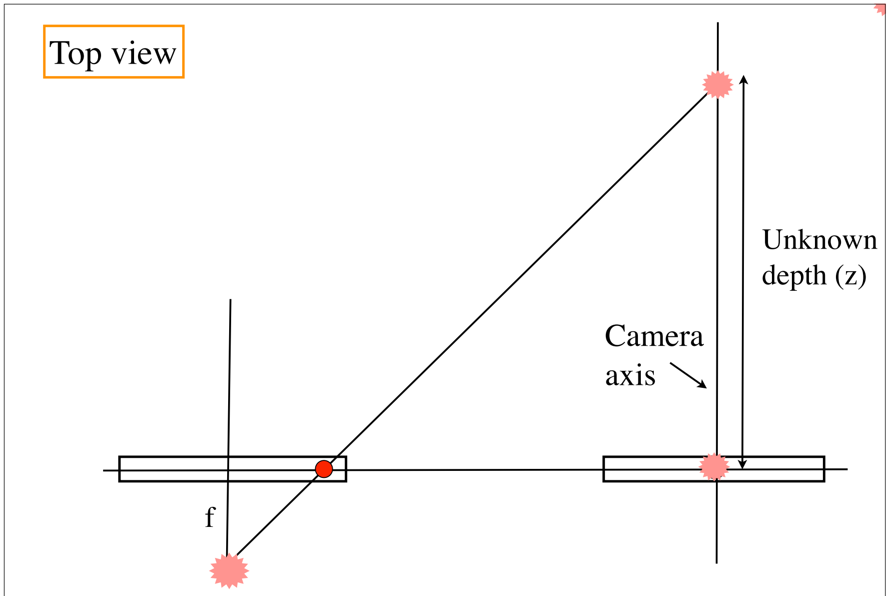
inter-ocular distance (d)

$$\frac{z}{f} = \frac{(d-D)}{D}$$ (by similar triangles)

$$z = f \cdot \frac{(d-D)}{D}$$

Active stereo

- Replace one of those cameras with a light source
- One of the technologies used in the Kinect
  - The Kinect uses an infrared source and sensor
  - It also has a color camera (not used for stereo)



Top view

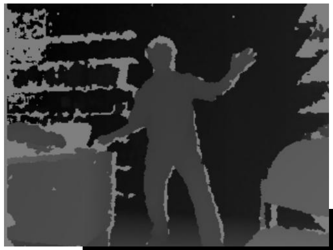
Camera axis

Unknown depth (z)

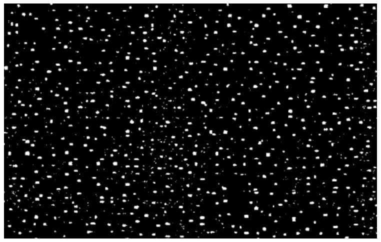
f

Kinect

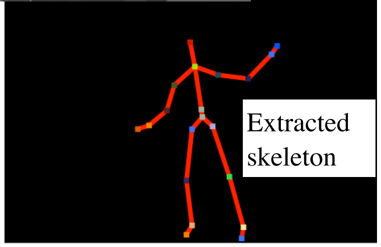
(Images are from a lecture by John MacCormick)



Depth map



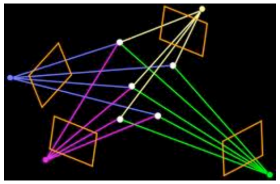
Infrared light speckle pattern



Extracted skeleton

Structure from Motion

- If the world is not changing quickly, then you do not need two cameras, just **move** the camera
- A key issue is that we do not know the camera matrix
  - In graphics we know the camera (we invent it)
  - In the basic stereo case, the two cameras are bolted together
  - Now we need to do stereo and work out the cameras at the same time
  - Many views (not just two) help a lot with this (and occlusion)



The internet camera

- In the structure from motion example it is convenient, but not necessary, that the images come from the same camera
- If the "object" is static why not use photos from the web?
  - Additional camera parameters need to be inferred
  - We need to figure out that they are (mostly) of the same thing
  - A popular application is historic/architectural landmarks.
- Instead of the first step in the previous movie, we could search for images of "rome" or the "room"&"colosseum"

Moving the world instead of the camera

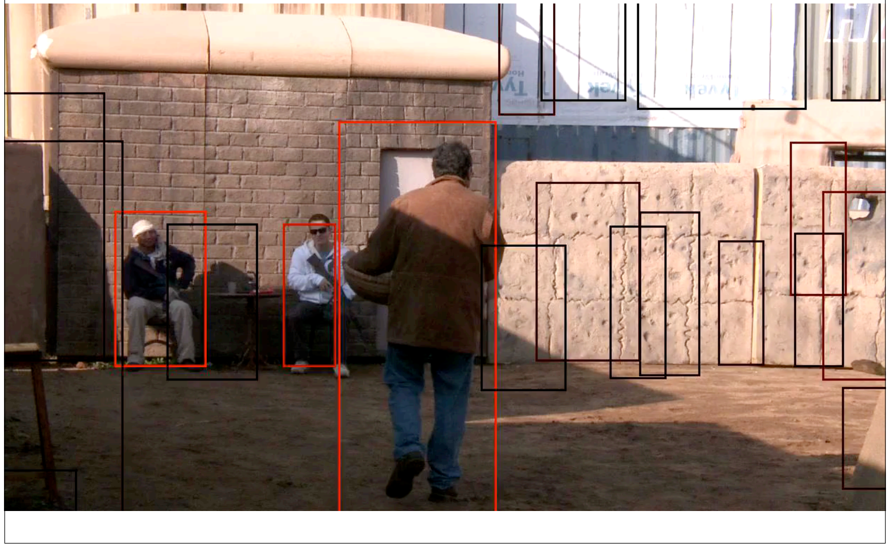
Interpreting movies

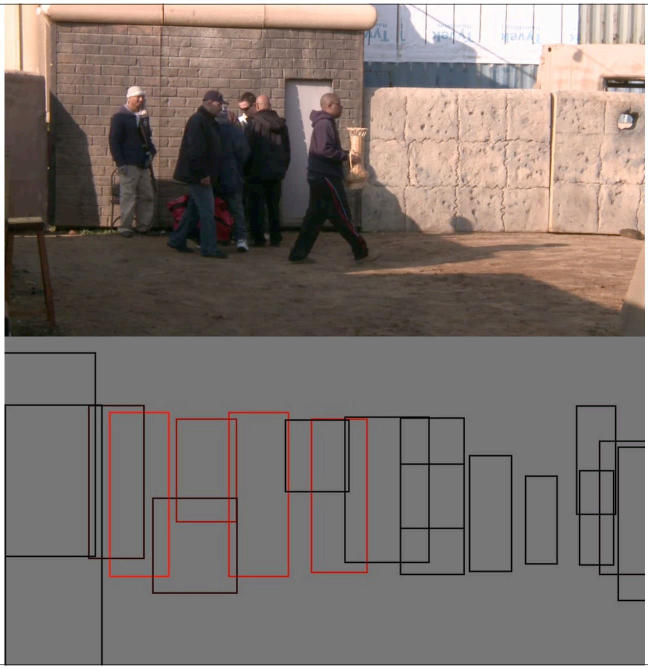
- The eye captures data in discrete chunks
  - It must be this way because noticing anything requires capturing a bunch of photons
- But we "see" motion
  - Understanding motion in a continuous sense is an evolutionary advantage
    - You can effectively "predict" the future
  - Because the brain creates motion from discrete chunks, movies work
  - Discrete image sequences are interpreted as smooth motion
    - 24 to 30 frames per second is more or less adequate
    - 60 frames per second has very few artifacts

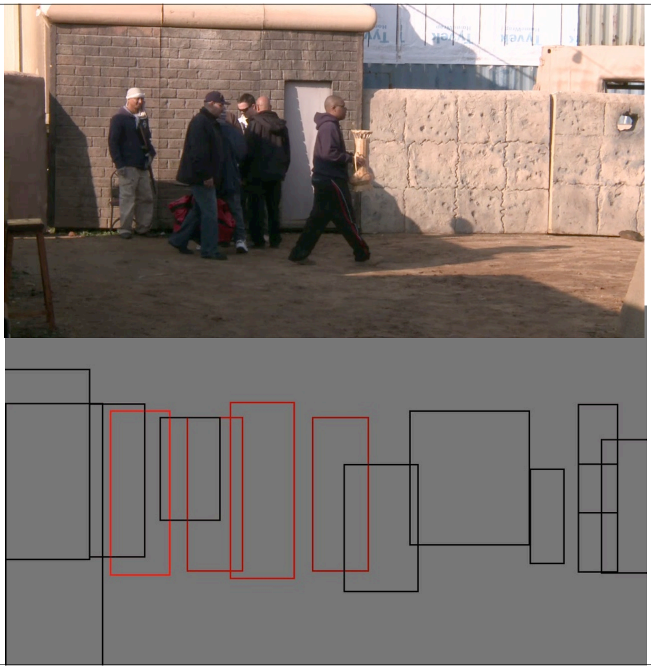
Interpreting movies

- Movies (as we have seen) have more information than single images
- But we need to know which bits of first image correspond to which bits of the second image (and the third, forth, ...)
  - Again a key issue in image understanding is **correspondence**
  - Linking moving objects across frames is **tracking**

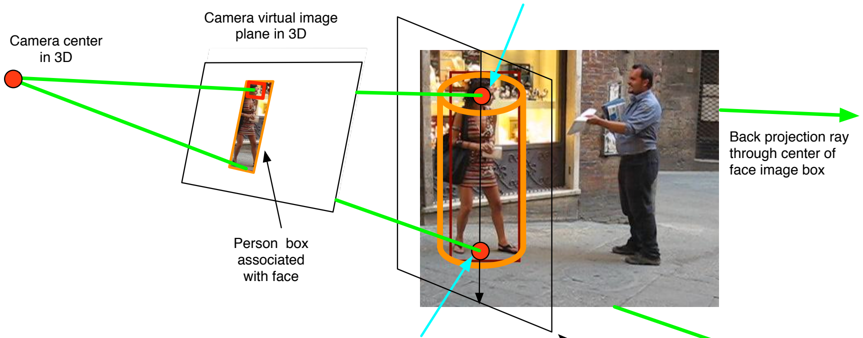
3D tracking (person detection evidence)







Model for 3D tracking (representation)



3D head center is found by intersecting face ray with a plane through cylinder and perpendicular to ground and parallel to camera plane.

Camera virtual image plane in 3D

Camera center in 3D

Back projection ray through center of face image box

Person box associated with face

Intersection with ground plane (known) sets bottom of cylinder

Body plane associated with cylinder

Back projection ray through middle of bottom of person image box

Example of tracking