What is computer vision about?

- Building computers that can see
- Understanding the **content** of images
  - What are the objects in the scene?
  - What are their shapes?
  - How far from the camera are they?
  - Are they moving? If so, how?
- Very different from image processing

Image formation

Cameras gather light scattered from objects to create 2D images

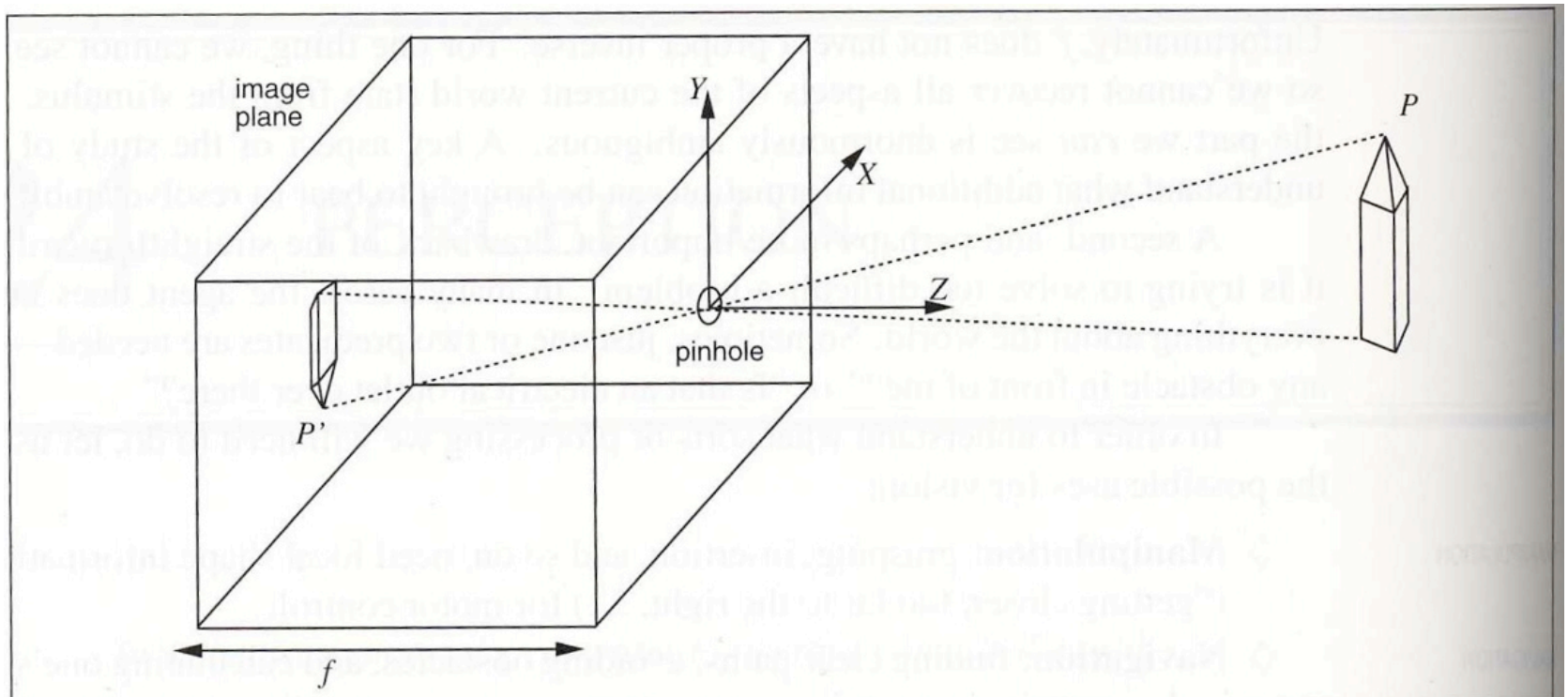


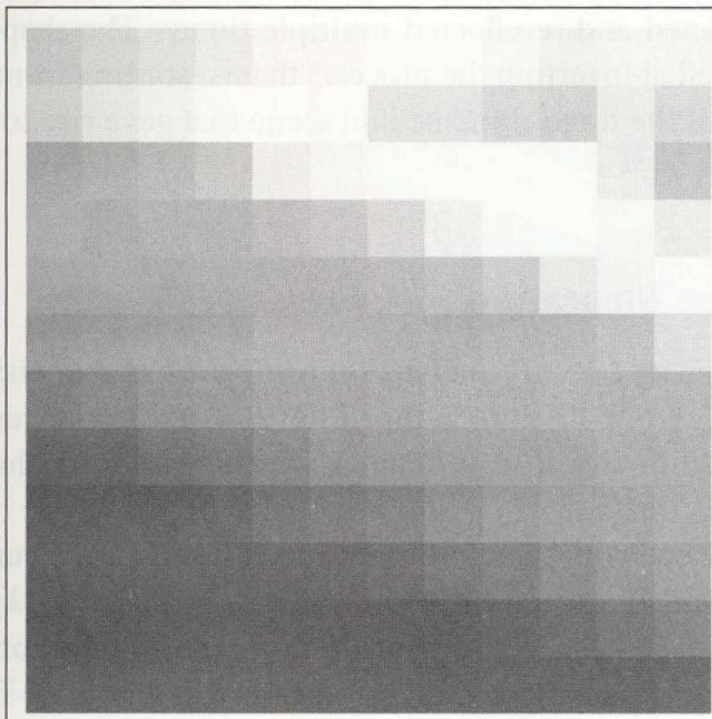
Figure 24.1 Geometry of image formation in the pinhole camera.

Much information is lost

Objects occlude each other

Digital images

The input to a vision algorithm is an array of intensities or color values



195	209	221	235	249	251	254	255	250	241	247	248
210	236	249	254	255	254	225	226	212	204	236	211
164	172	180	192	241	251	255	255	255	255	235	190
167	164	171	170	179	189	208	244	254	255	251	234
162	167	166	169	169	170	176	185	196	232	249	254
153	157	160	162	169	170	168	169	171	176	185	218
126	135	143	147	156	157	160	166	167	171	168	170
103	107	118	125	133	145	151	156	158	159	163	164
095	095	097	101	115	124	132	142	117	122	124	161
093	093	093	093	095	099	105	118	125	135	143	119
093	093	093	093	093	093	095	097	101	109	119	132
095	093	093	093	093	093	093	093	093	093	093	119

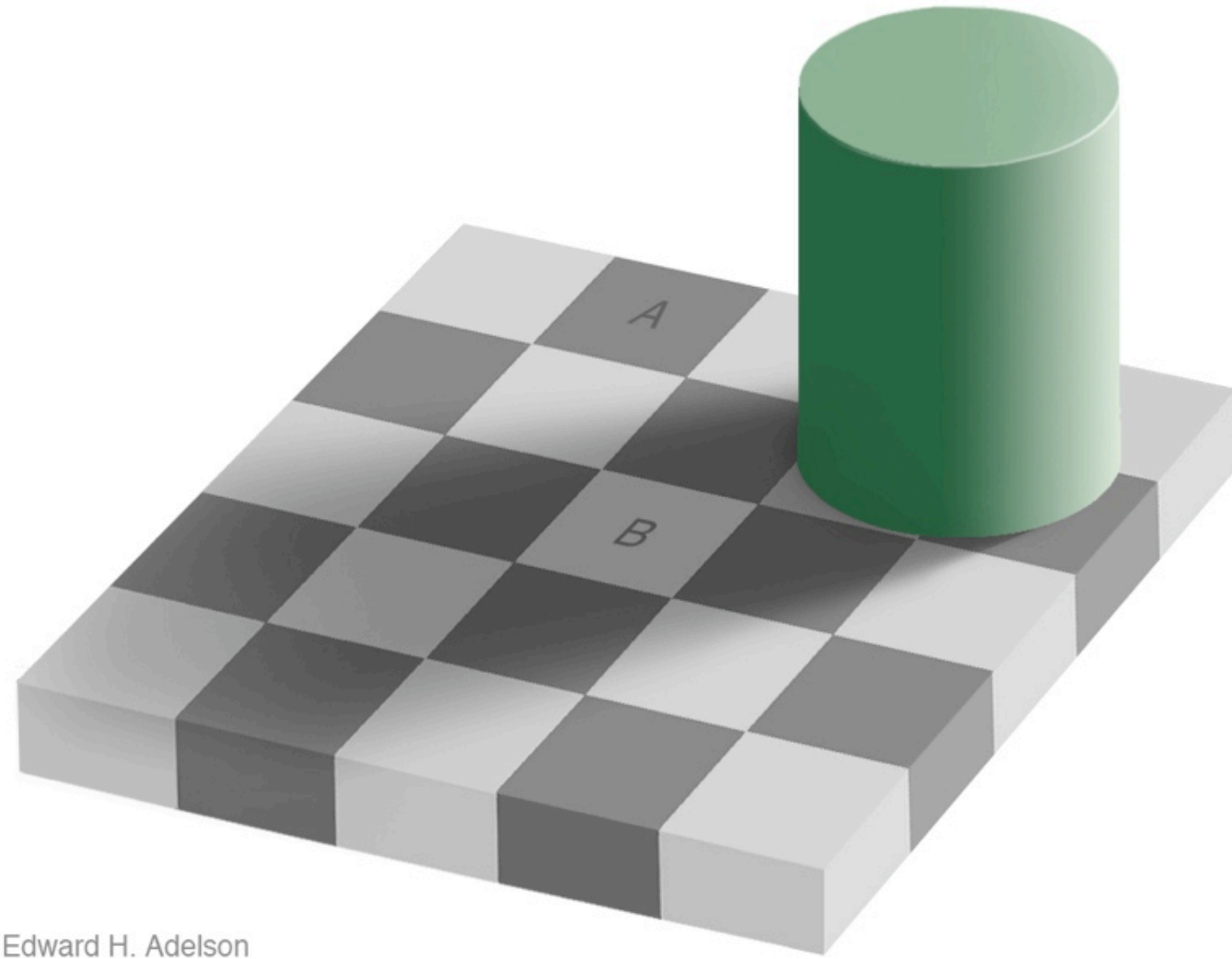
(a) (b)

Figure 24.4 (a) Magnified view of a 12 × 12 block of pixels from Figure 24.3. (b) The associated image brightness values.

Image formation

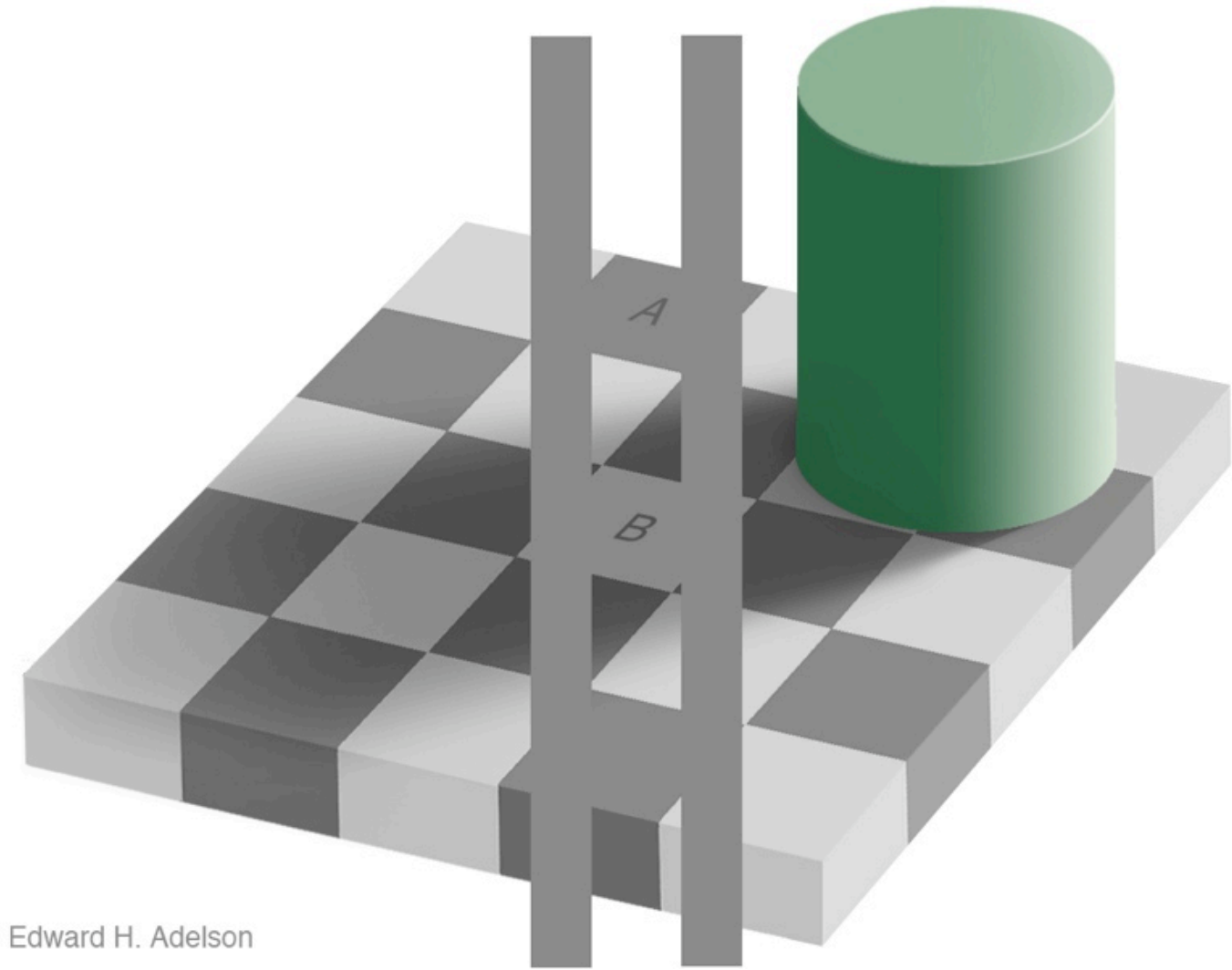
- Intensity of a pixel depends on:
  - Geometry of the scene
  - Reflectance of surfaces (material properties)
  - Illumination of the scene
  - Camera viewpoint
  - Noise
- Our visual system is able to factor out these different aspects very well

Lightness perception



Edward H. Adelson

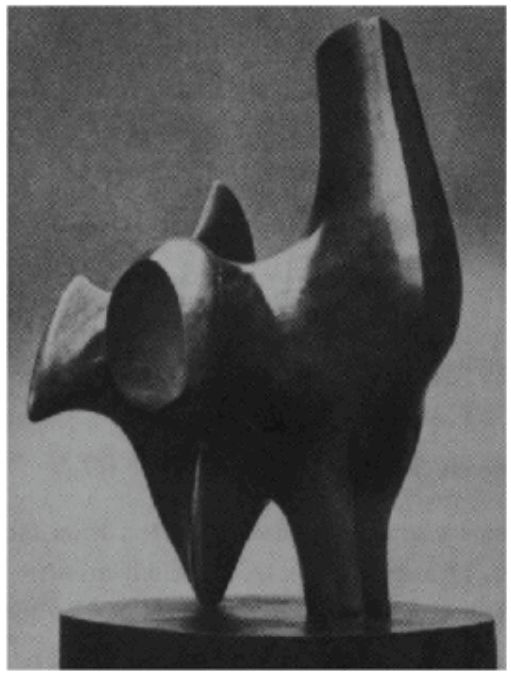
Lightness perception

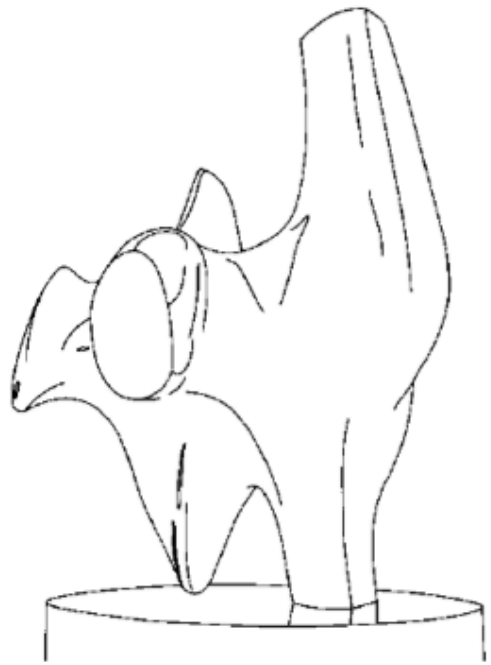


First problem: Edge detection

Goal: compute something like a line drawing of a scene

Input: grayscale image
Output: binary image where 1 = “edge”





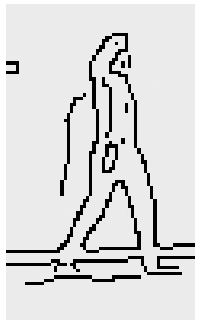
Why?

- Edges reflect intrinsic properties of a scene
  - Capture shape information
  - Independent of illumination
- The human visual system does something like this!
- Good initial step for solving other problems
  - recognition, tracking, etc.

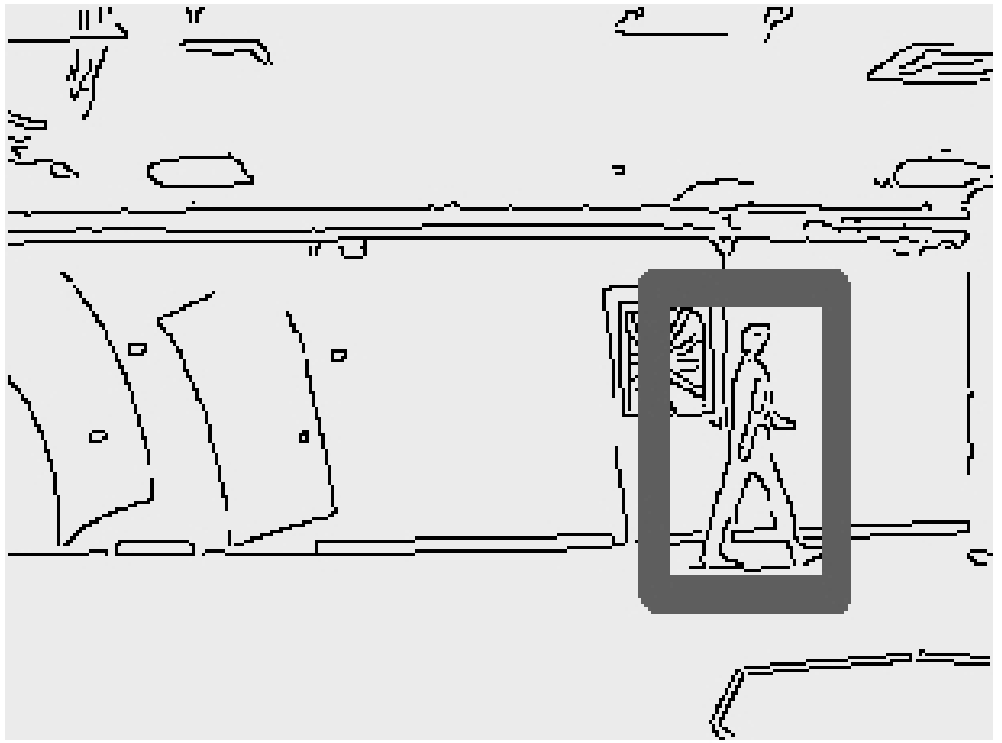
Issues

- No precise problem formulation
  - This is true of almost every vision problem
  - Different algorithms compute different things
- Much harder than it seems to be
- Edge detectors usually work by detecting “big changes” in image intensity
  - Works well for synthetic images!
  - Misses a lot of “boundaries” that we see

Matching edge templates



Model



Best match

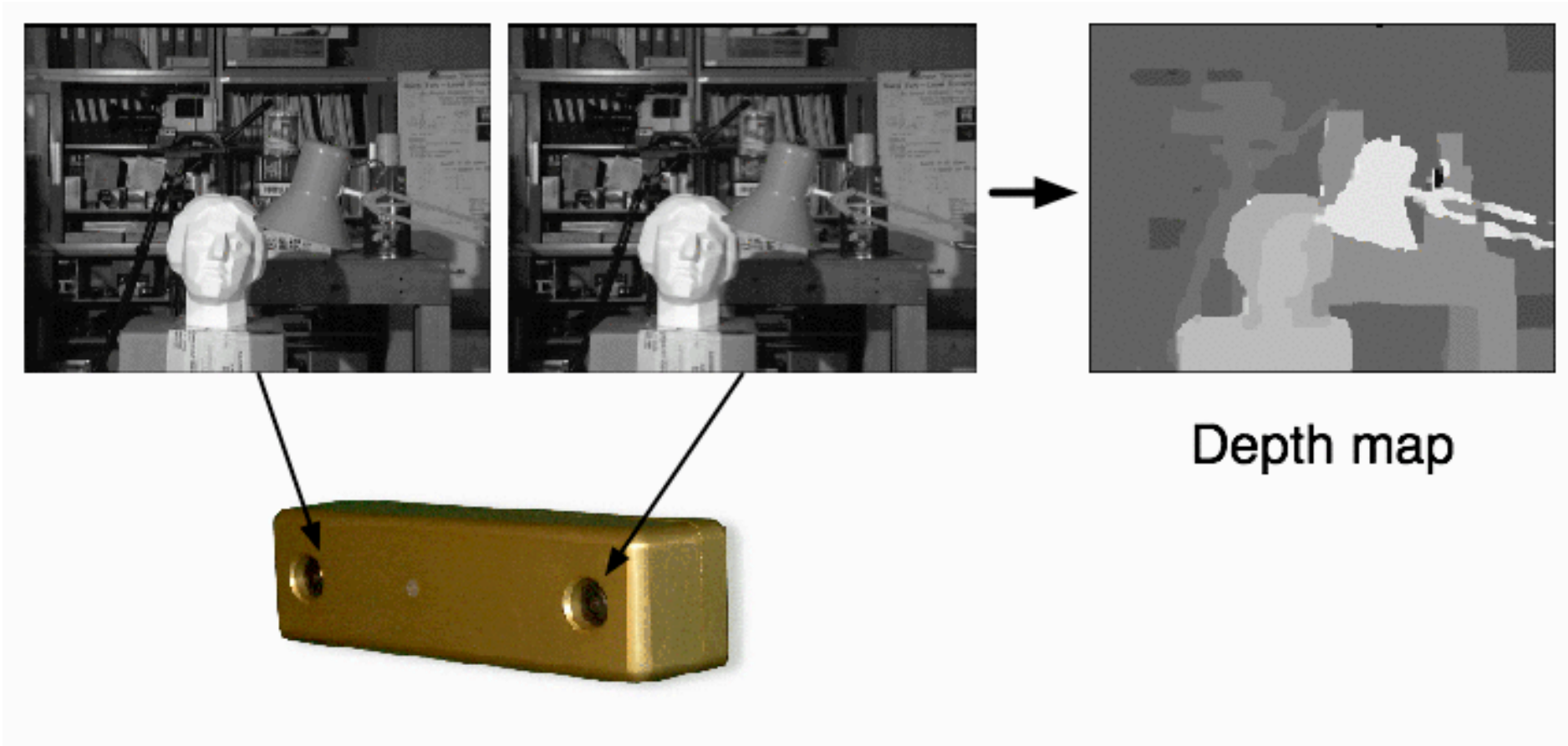
Search the image for the
best match of the model

- how to measure match quality?
- how to search for best match quickly?

Stereo vision / 3D reconstruction

Goal: recover 3D information from multiple 2D images

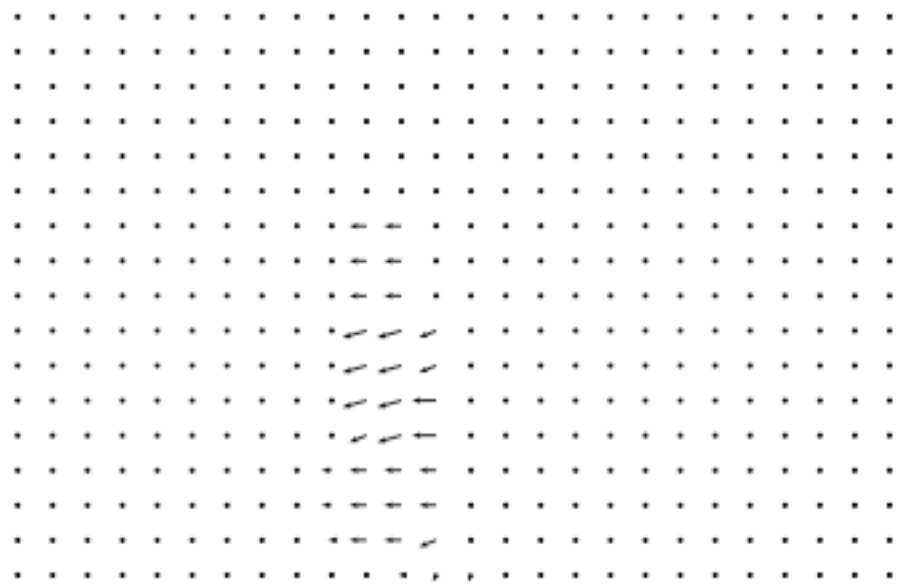
Input: multiple images
Output: depth map or 3D model



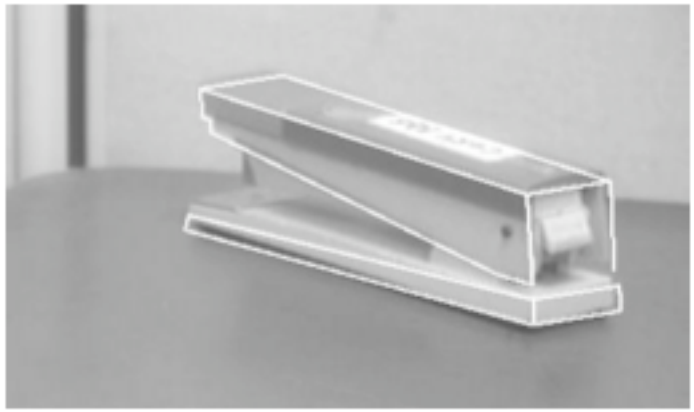
Motion estimation

Goal: estimate the motion of objects

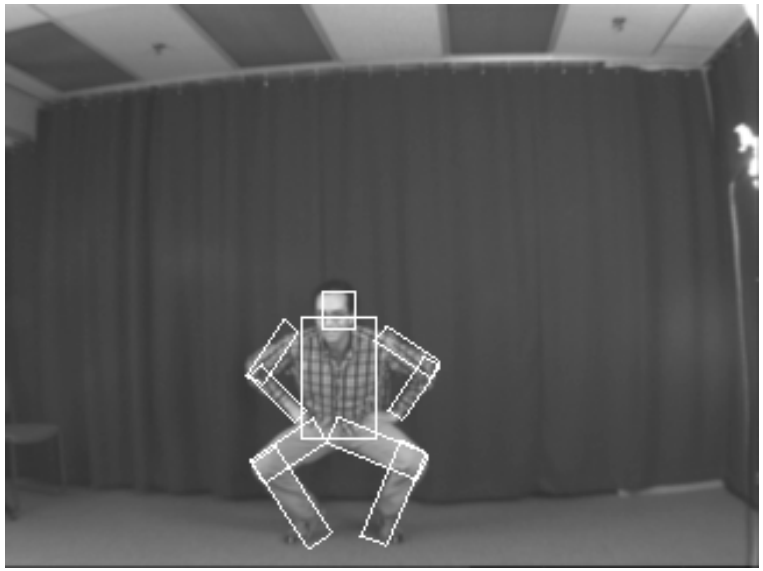
Input: sequence of images
Output: velocity field



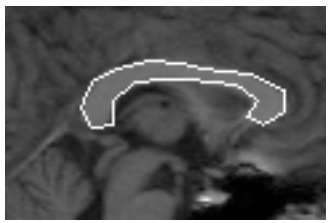
Object detection/recognition

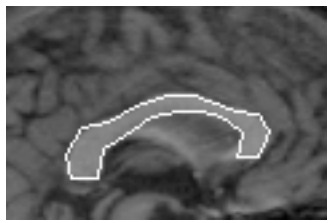


3D models



Articulated objects

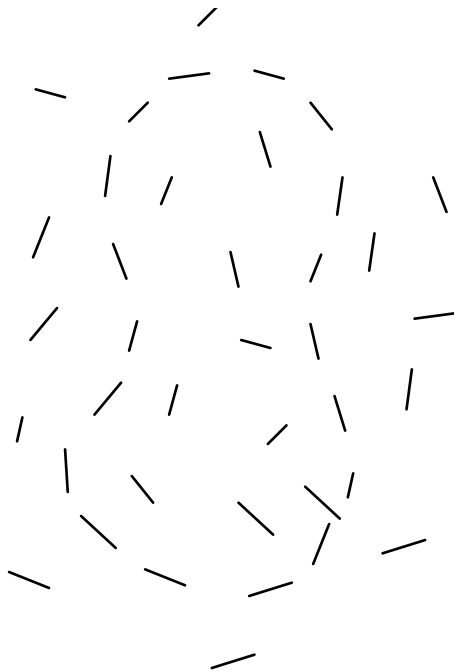




Deformable shapes

Perceptual grouping

Goal: group pixels or tokens into meaningful objects



Face detection

Scene understanding