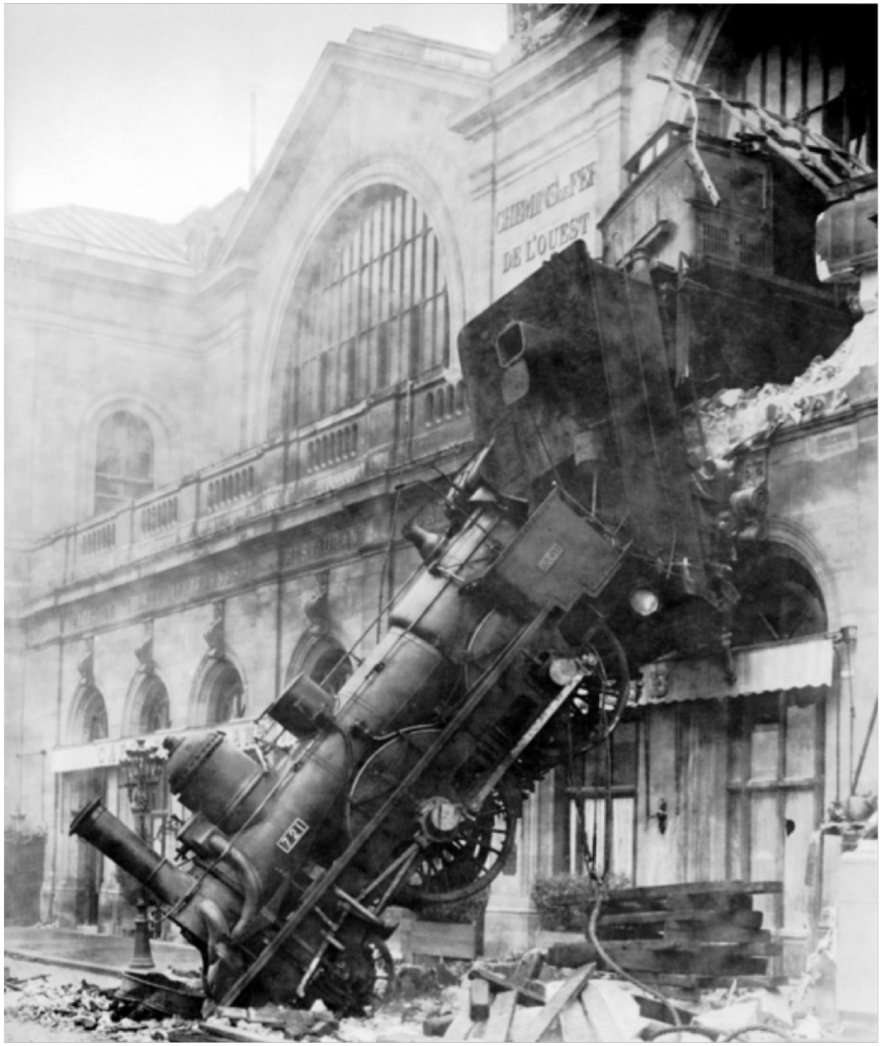
Review - Computer Vision

Saurabh Gupta

Many slides adapted from B. Hariharan, L. Lazebnik, N. Snavely, Y. Furukawa.

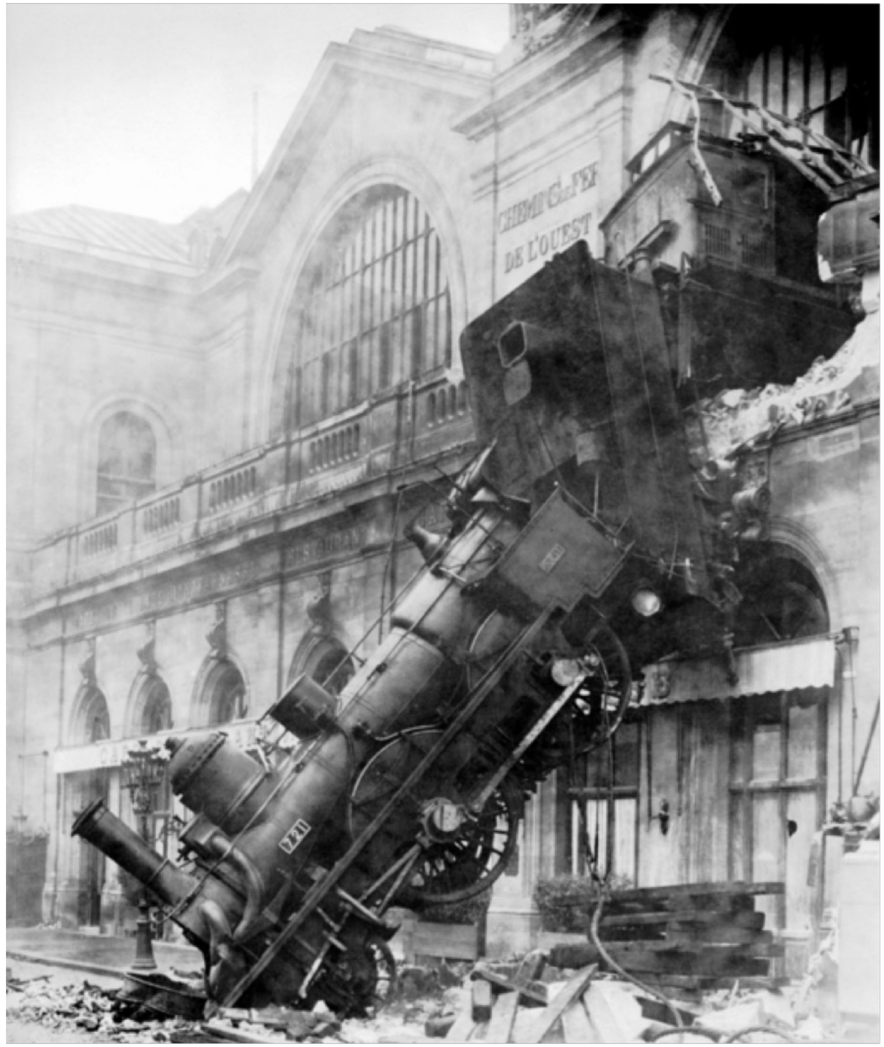
The goal(s) or computer vision



- What is the image about?
- What objects are in the image?
- Where are they?
- How are they oriented?
- What is the layout of the scene in 3D?
- What is the shape of each object?

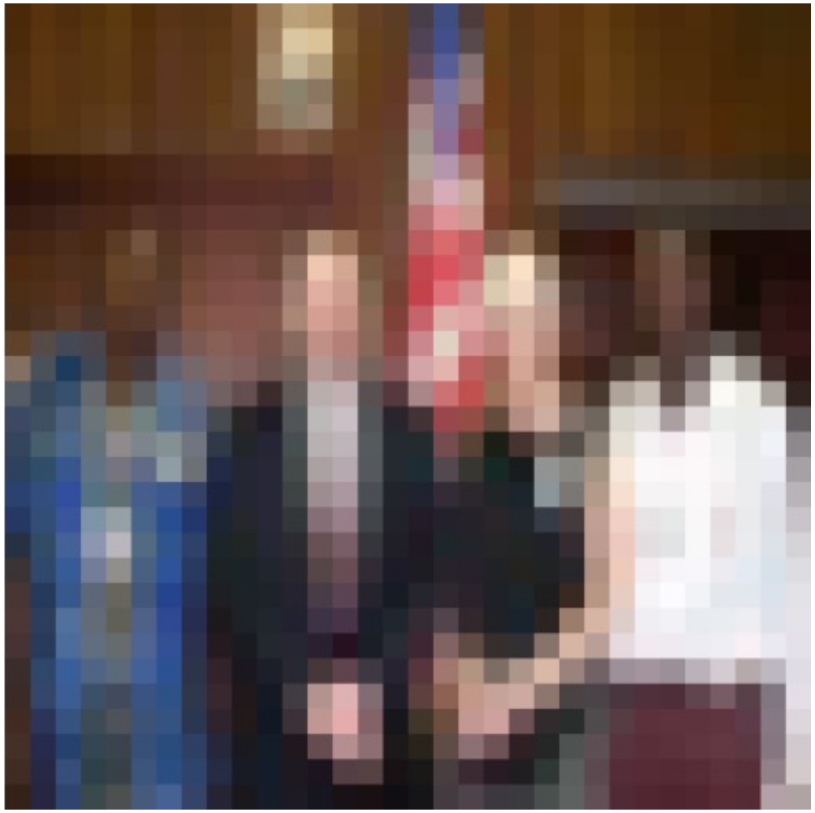
Source: B. Hariharan

Vision is easy for humans



Source: B. Hariharan

Vision is easy for humans



Source: L. Lazebnik

Source: “80 million tiny images” by Torralba et al.

Vision is easy for humans

Attneave's Cat

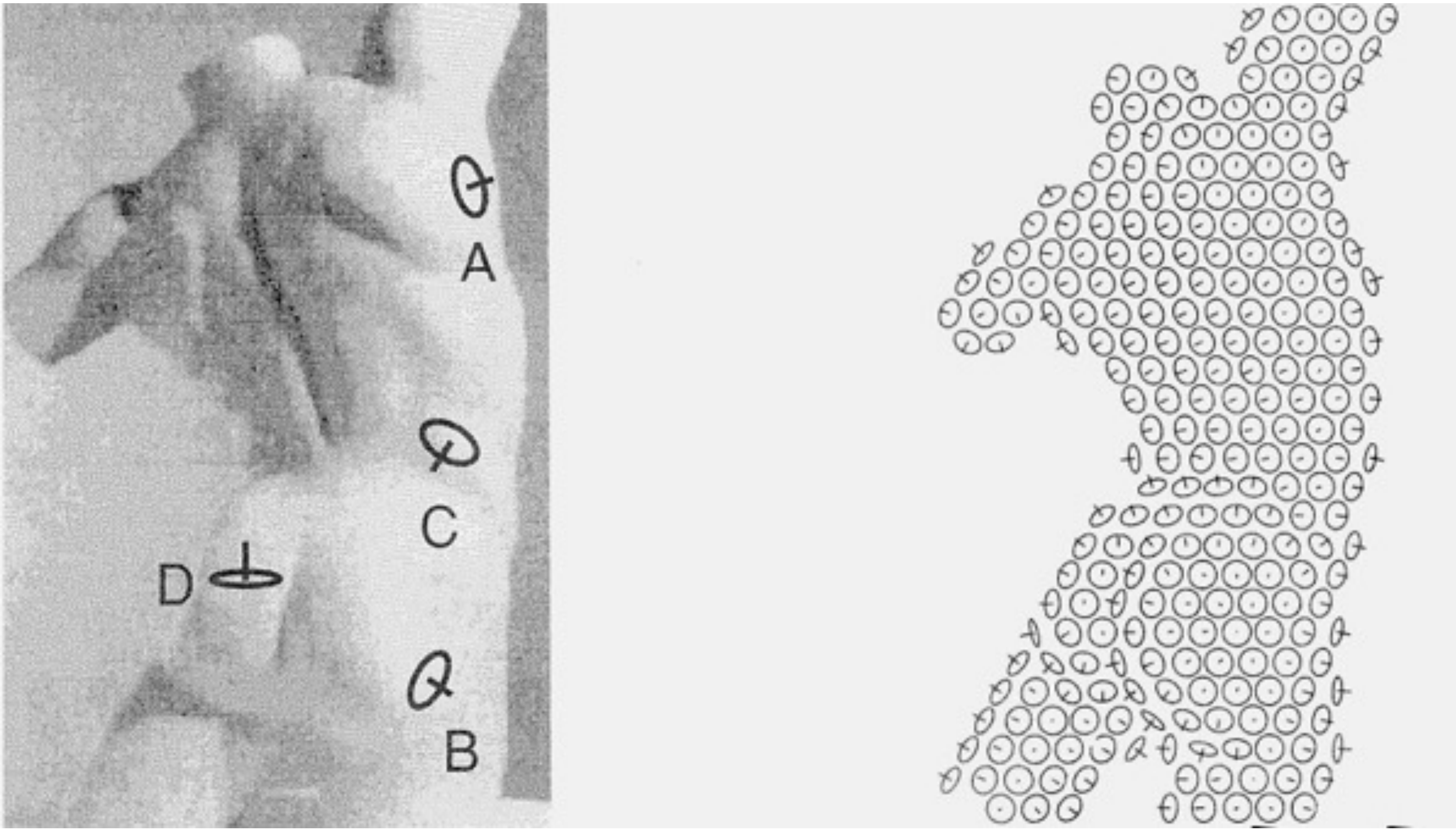
Source: B. Hariharan

Vision is easy for humans

Mooney Faces

Source: B. Hariharan

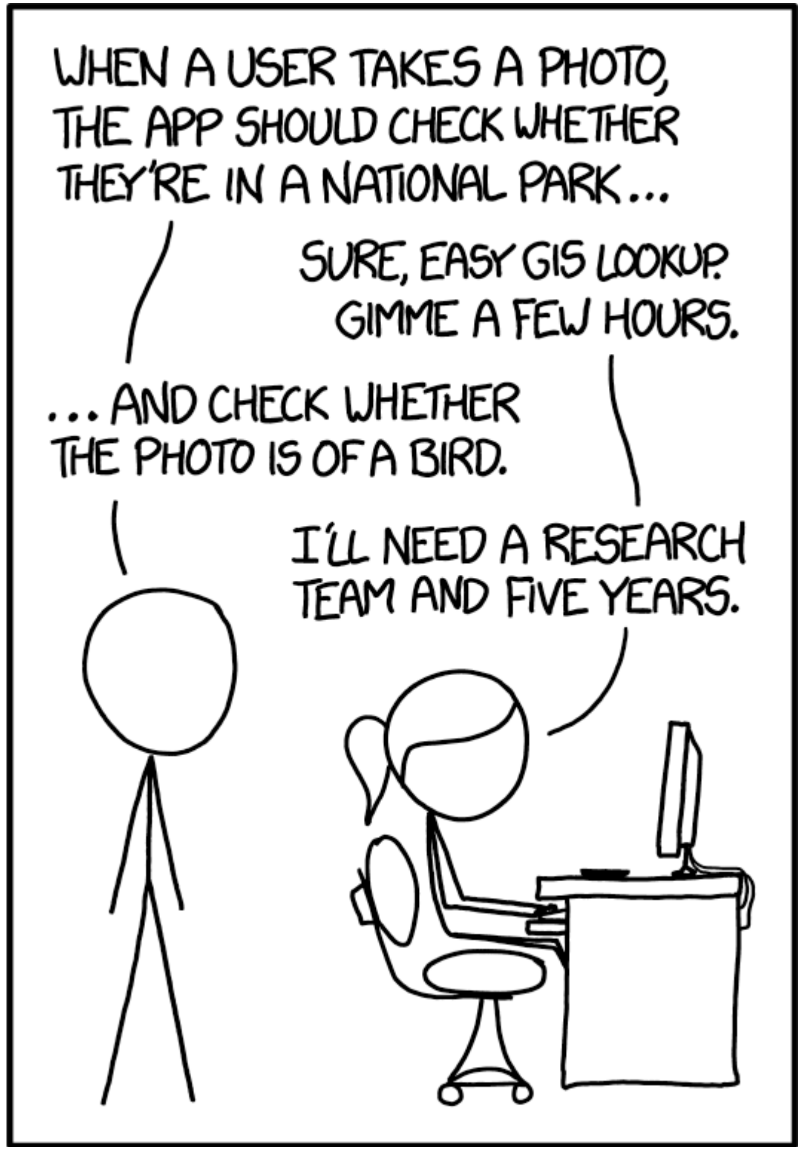
Vision is easy for humans



Surface perception in pictures. Koenderink, van Doorn and Kappers, 1992

Source: J. Malik

Remarkably Hard for Computers



Source: XKCD

Vision is hard: Objects Blend Together

Source: B. Hariharan

Vision is hard: Objects Blend Together

Source: B. Hariharan

Vision is hard: Intra-class Variation

Viewpoint variation

Illumination

Scale

Source: B. Hariharan

Vision is hard: Intra-class Variation

Shape variation

Occlusion

Background clutter

Source: B. Hariharan

Vision is hard: Intra-class Variation

Source: B. Hariharan

Vision is hard: Concepts are subtle

Tennessee Warbler

Orange Crowned Warbler

Source: B. Hariharan

https://www.allaboutbirds.org

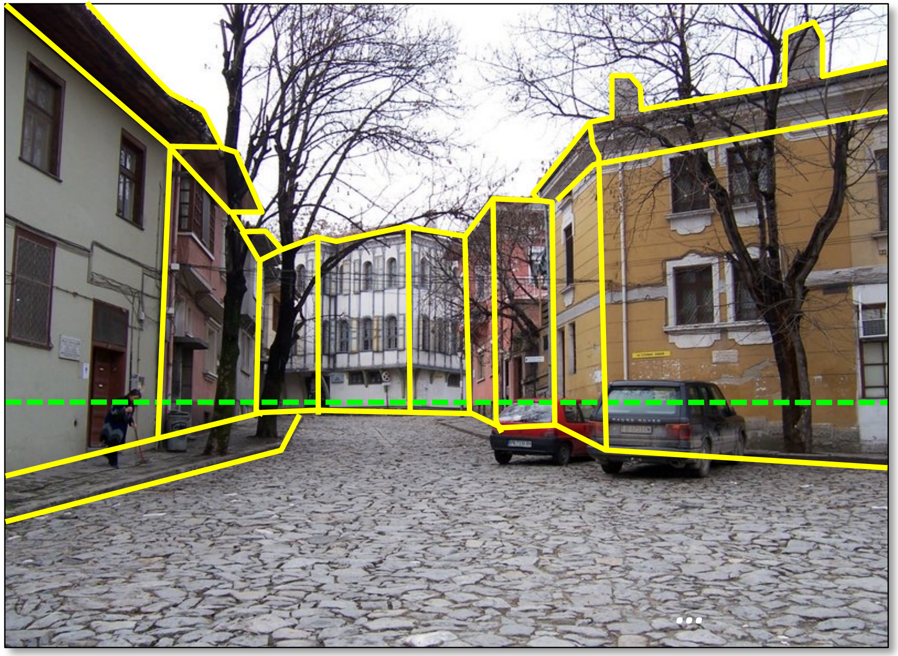
Vision is hard: Images are ambiguous

Source: B. Hariharan

What kind of information can be extracted from an image?

Source: L. Lazebnik

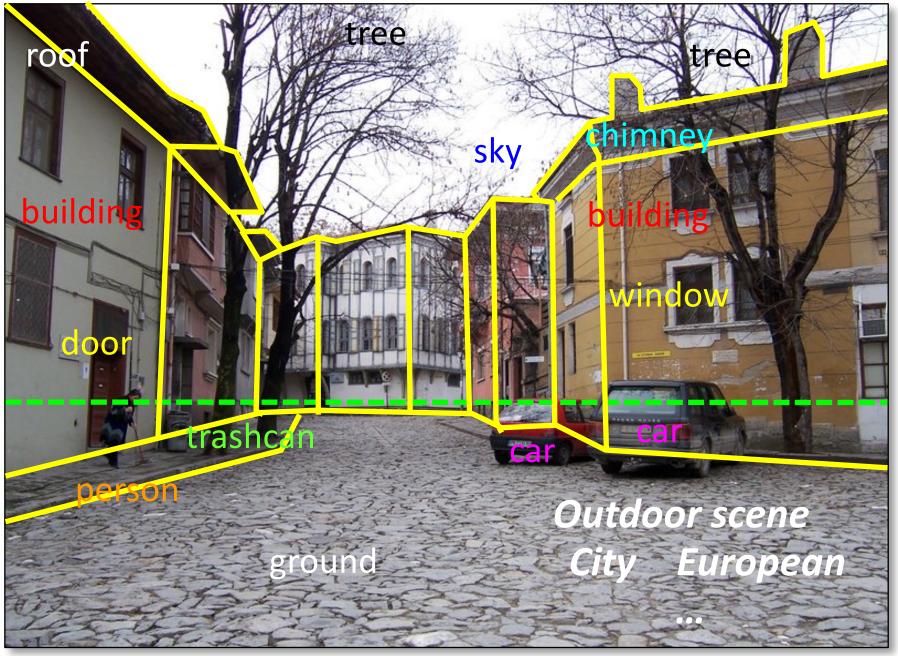
What kind of information can be extracted from an image?



Geometric information

Source: L. Lazebnik

What kind of information can be extracted from an image?



Geometric information

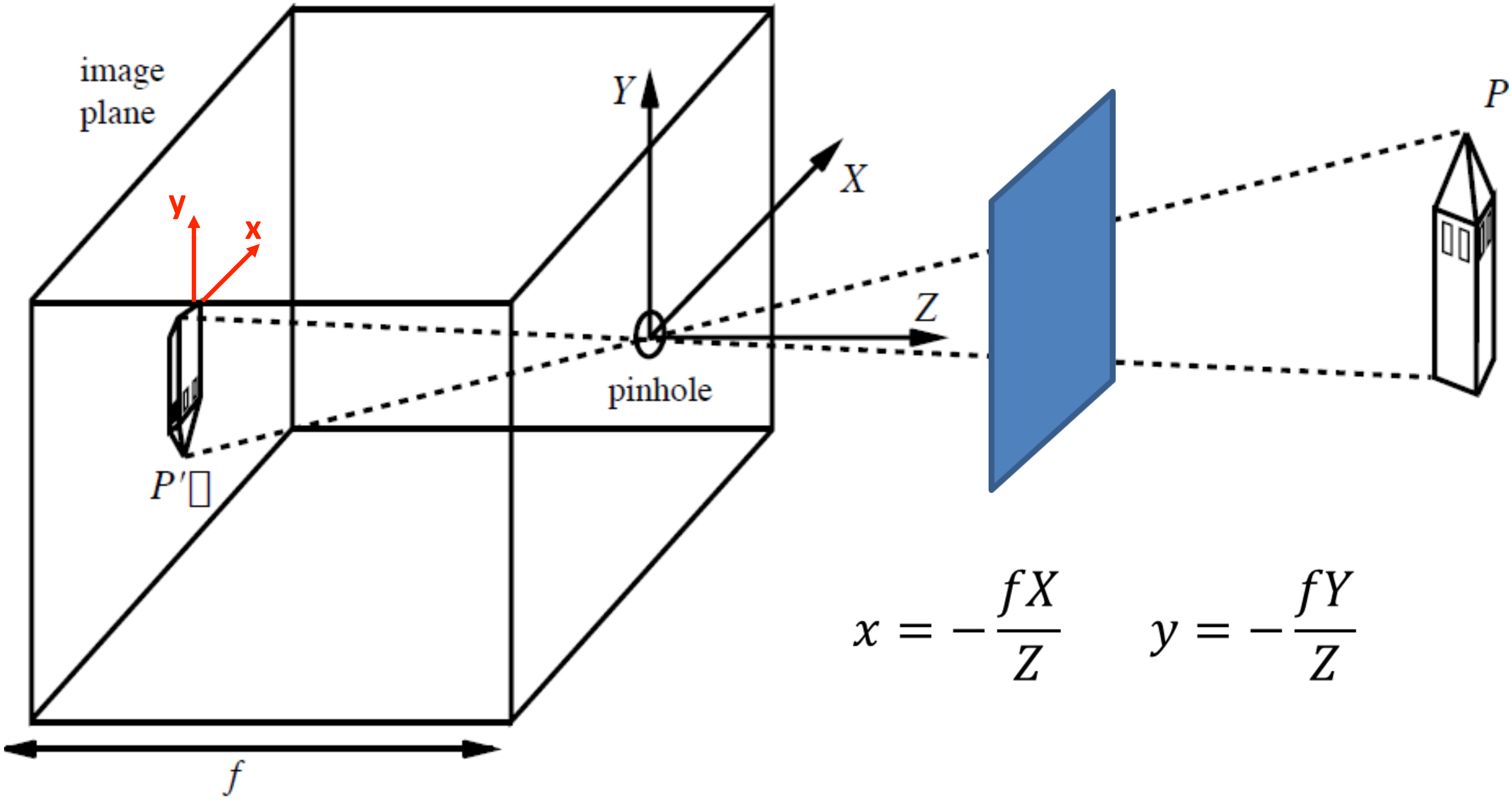
Semantic information

Source: L. Lazebnik

Vision is hard: Images are ambiguous

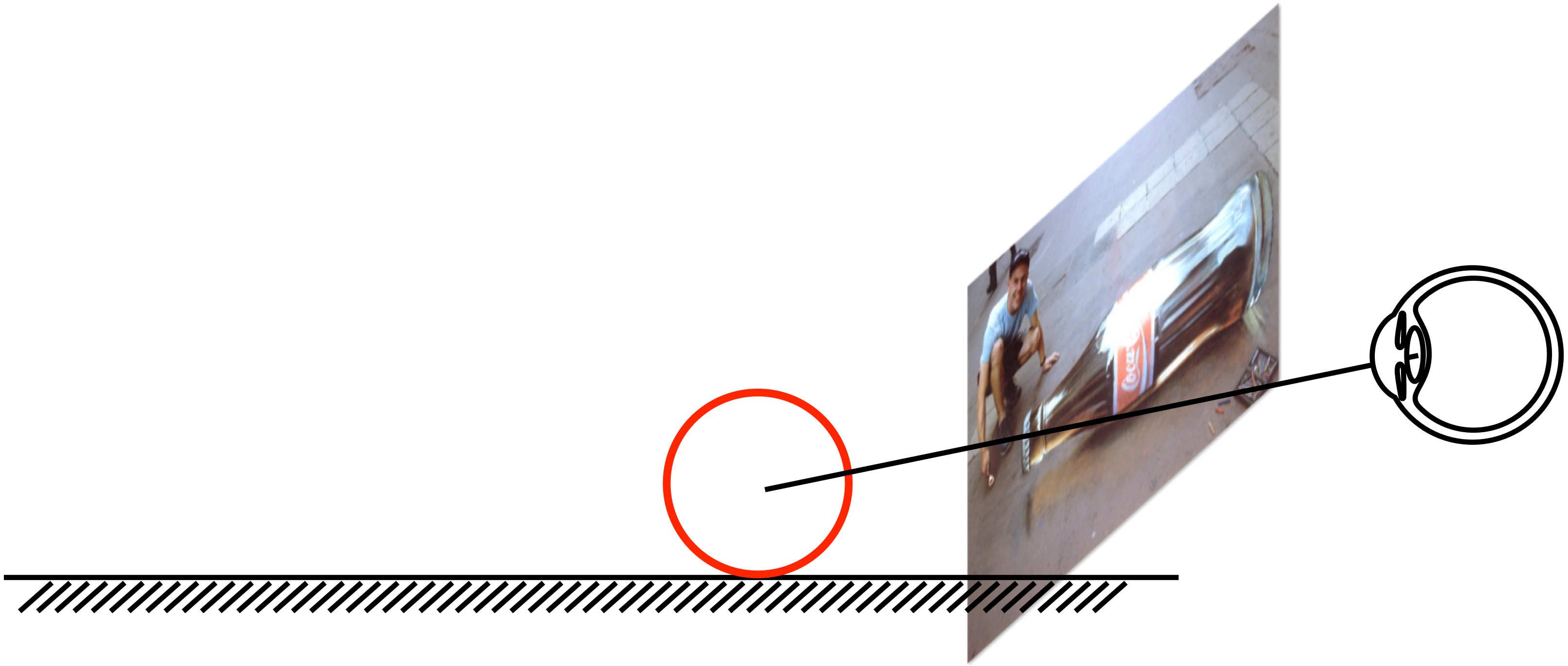
Source: B. Hariharan

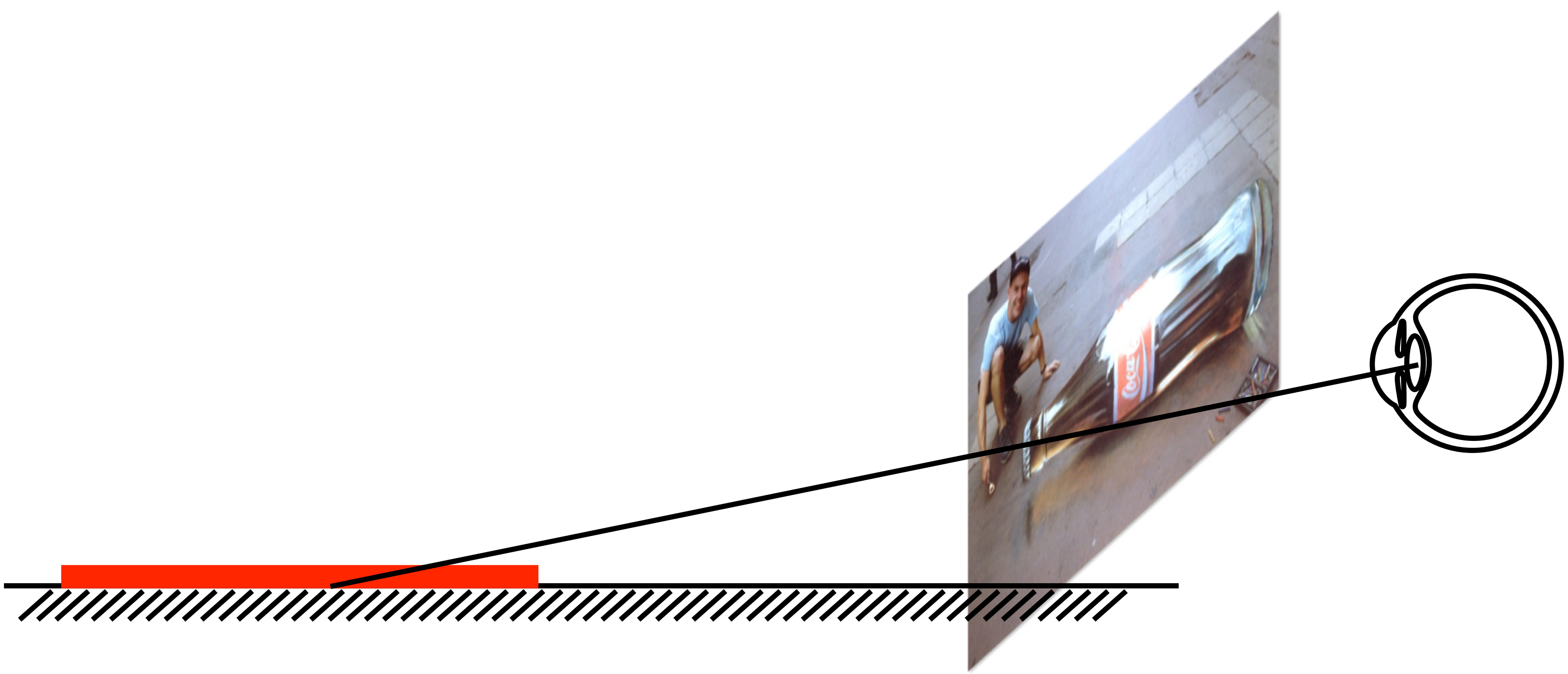
The Pinhole Camera

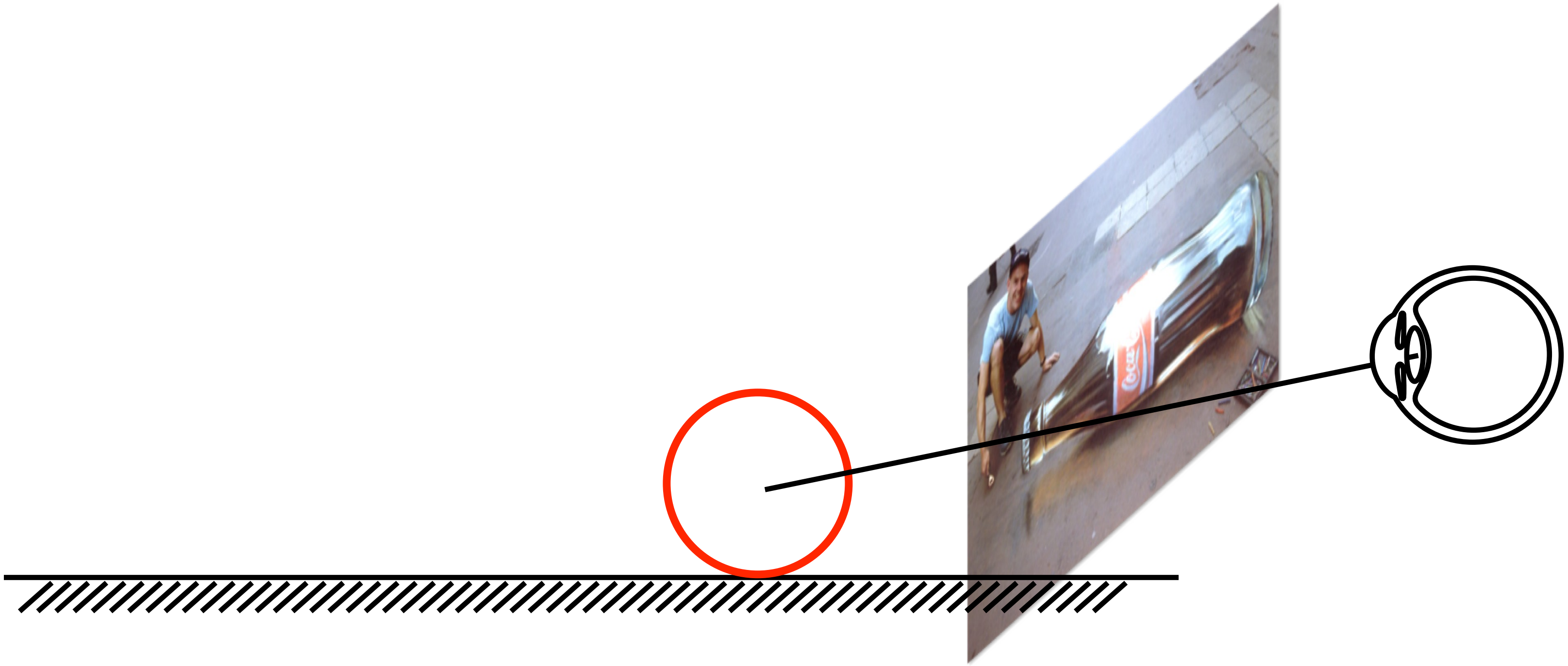


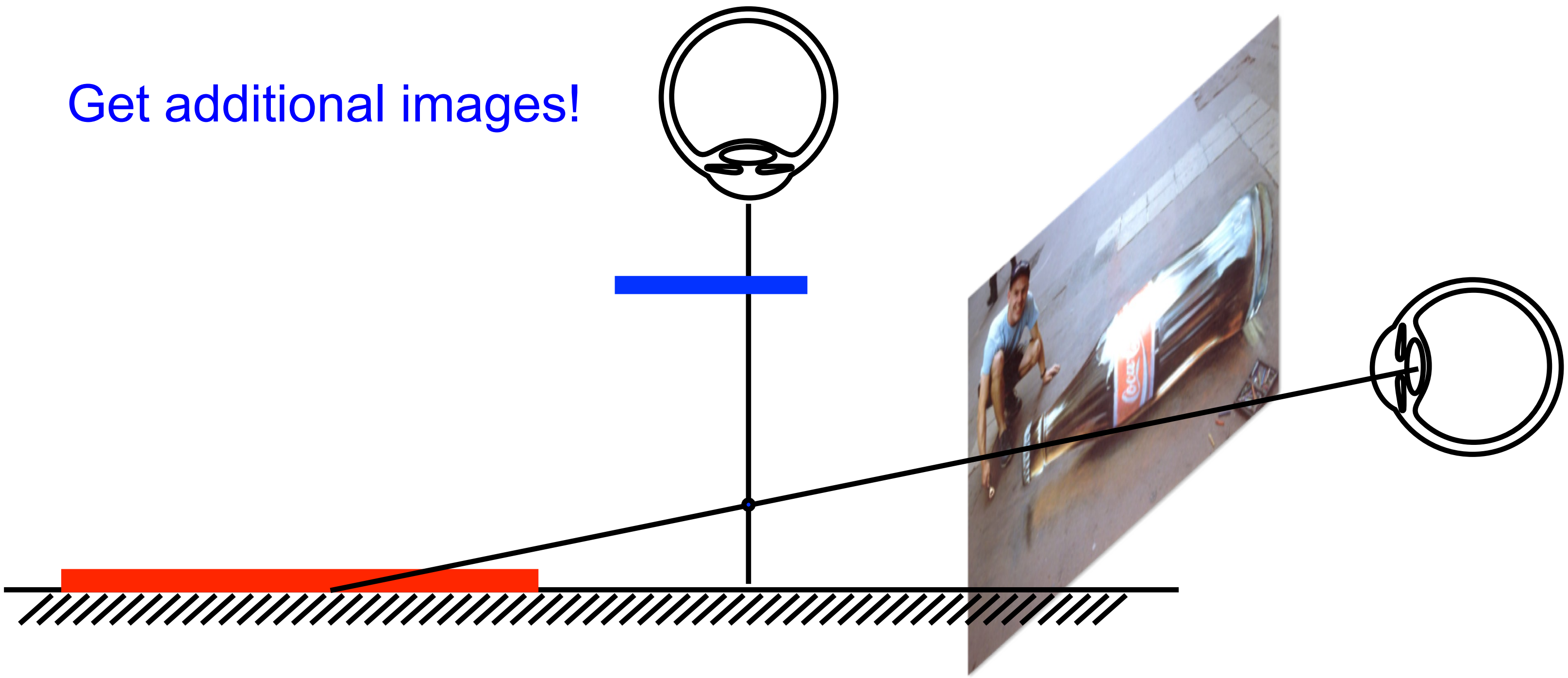
$$x = -\frac{fX}{Z} \qquad y = -\frac{fY}{Z}$$

Source: J. Malik

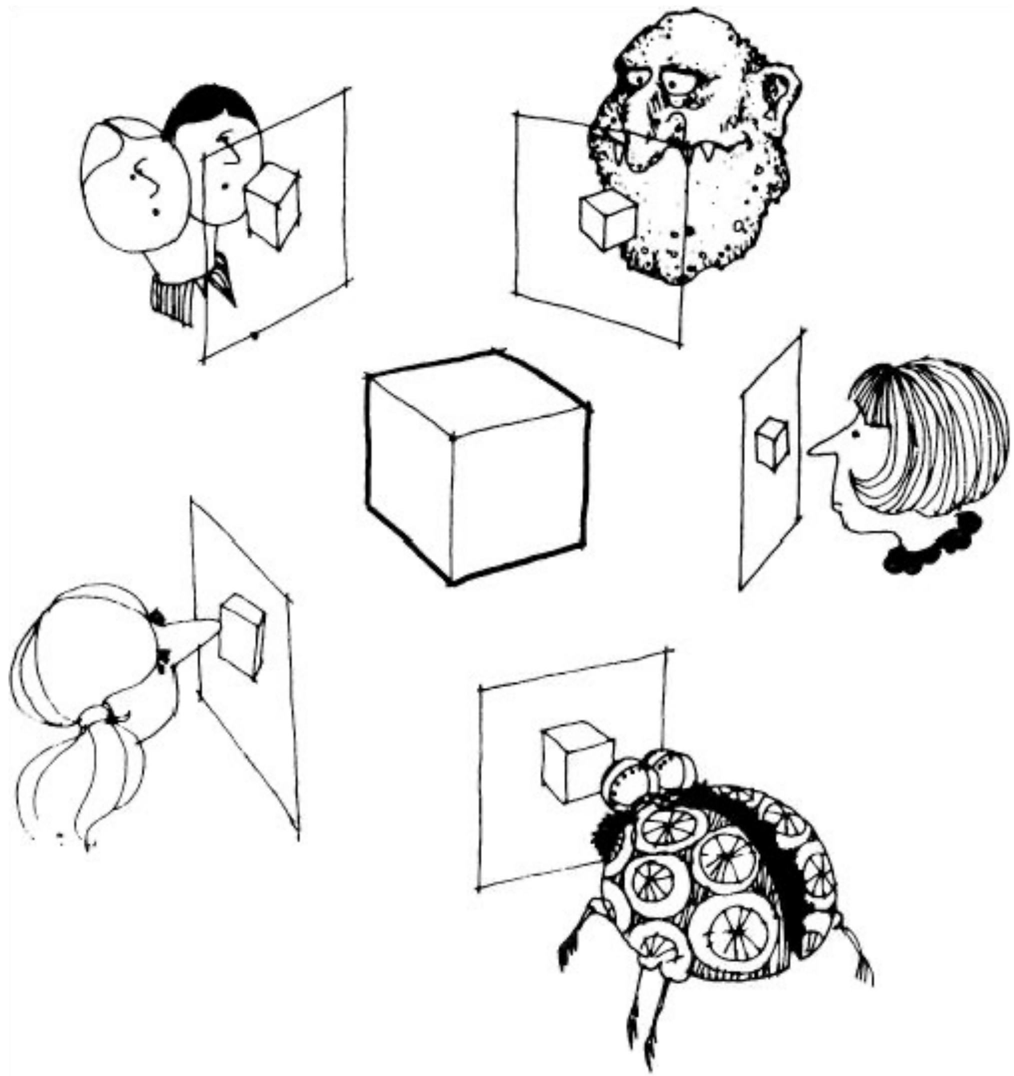







Get additional images!

Structure from Motion



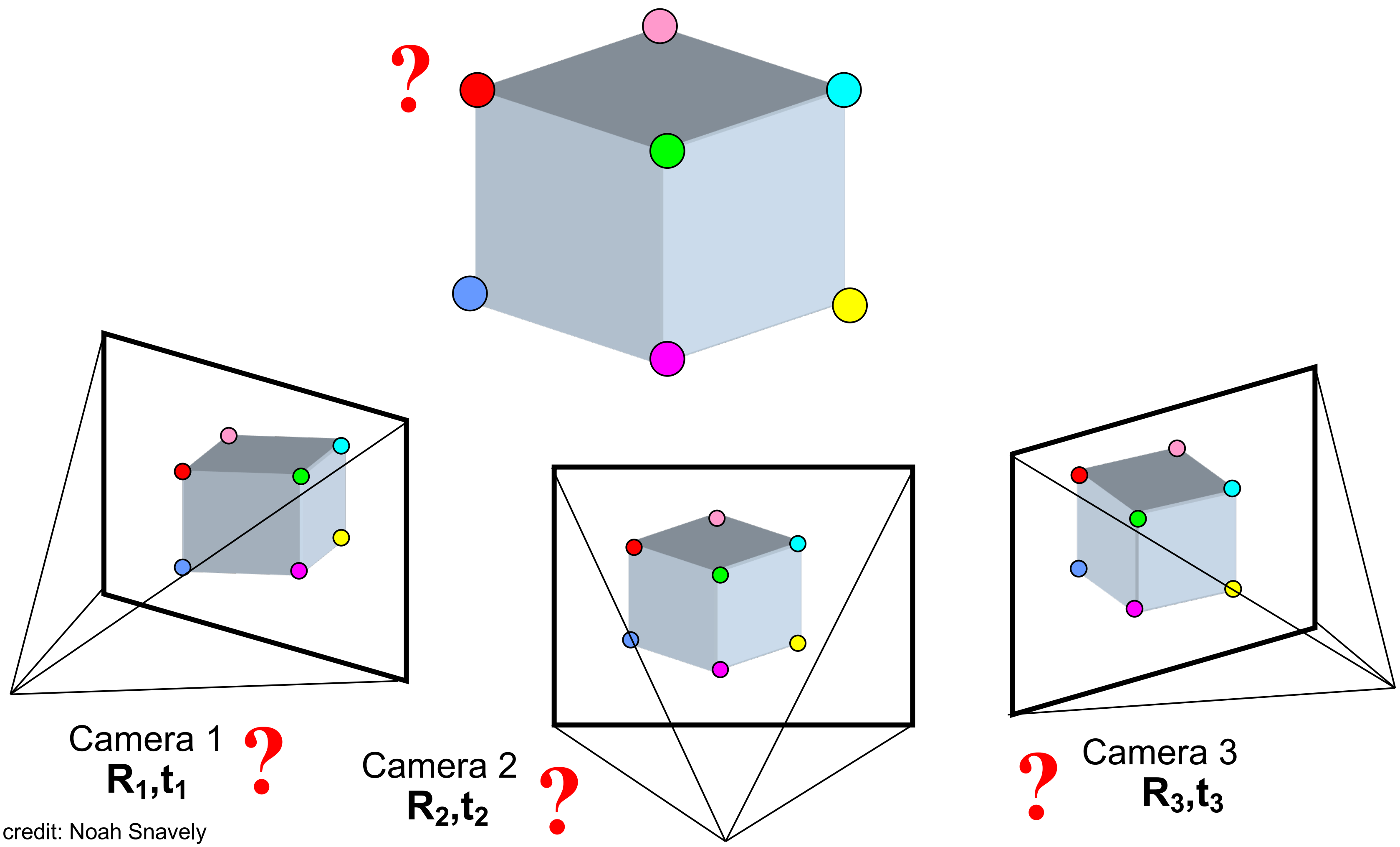
Many slides adapted from S. Seitz, Y. Furukawa, N. Snavely

Structure from motion

- Generic problem formulation: given several images of the same object or scene, compute a representation of its 3D shape
- Images of the same object or scene
  - Arbitrary number of images (from two to thousands)
  - Arbitrary camera positions (special rig, camera network or video sequence)
  - Camera parameters may be known or unknown

Structure from motion

- Given a set of corresponding points in two or more images, compute the camera parameters and the 3D point coordinates



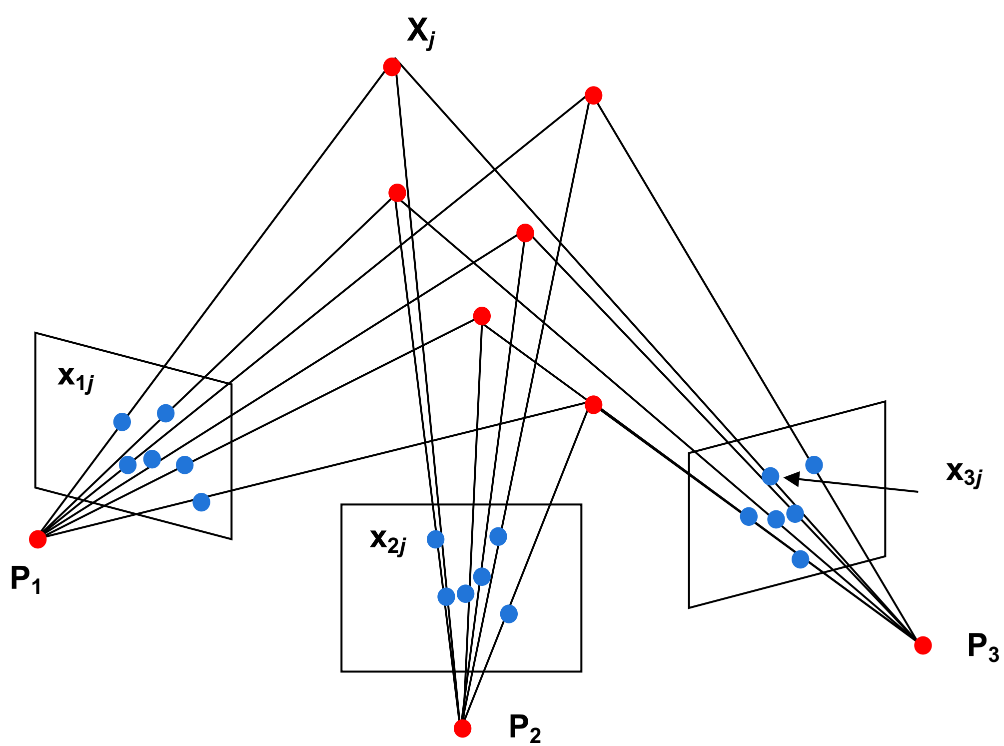
Slide credit: Noah Snavely

Structure from motion

- Given: *m* images of *n* fixed 3D points

$$\lambda_{ij}\mathbf{x}_{ij} = \mathbf{P}_i\mathbf{X}_j, \qquad i = 1, \dots, m, \quad j = 1, \dots, n$$

- Problem: estimate *m* projection matrices $\mathbf{P}_i$ and *n* 3D points $\mathbf{X}_j$ from the *mn* correspondences $\mathbf{x}_{ij}$



Structure from motion

- Triangulation

- Camera calibration

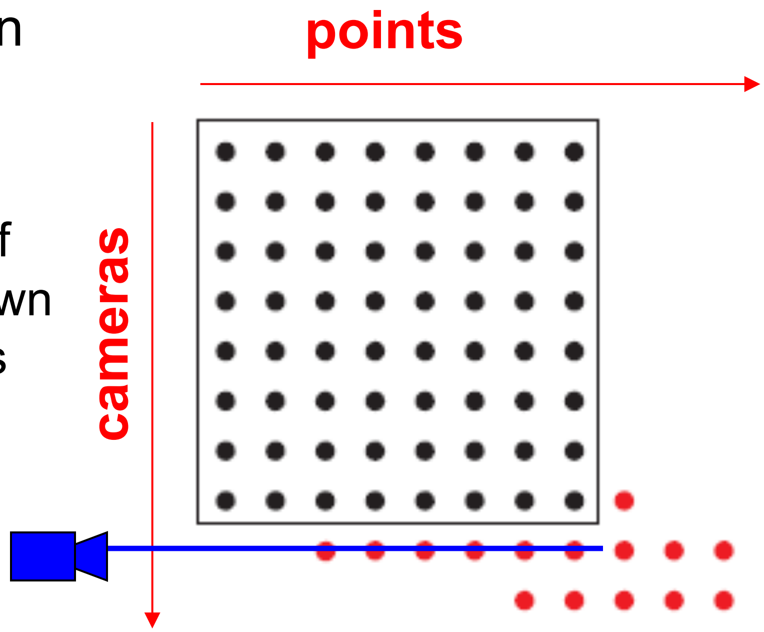
Incremental structure from motion

•Initialize motion from two images using fundamental matrix

•Initialize structure by triangulation

•For each additional view:

- Determine projection matrix of new camera using all the known 3D points that are visible in its image – *calibration*

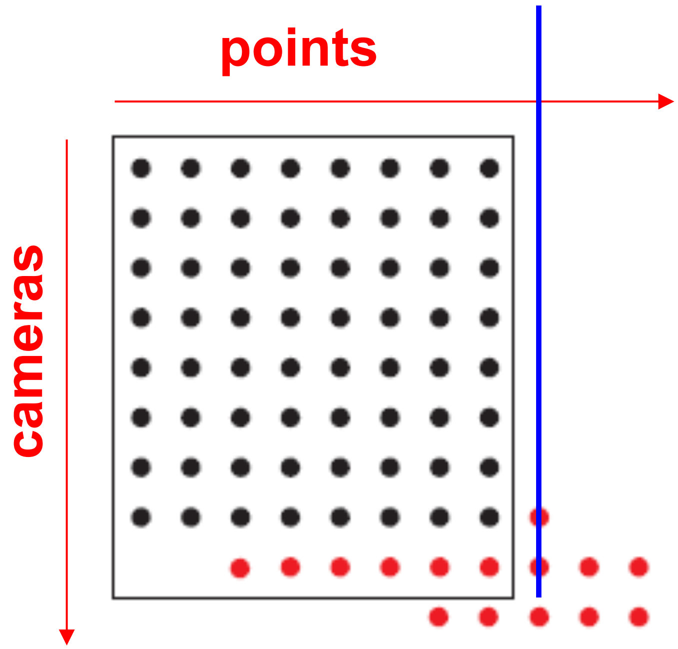


Incremental structure from motion

- Initialize motion from two images using fundamental matrix

- Initialize structure by triangulation

- For each additional view:
    - Determine projection matrix of new camera using all the known 3D points that are visible in its image – *calibration*
    - Refine and extend structure: compute new 3D points, re-optimize existing points that are also seen by this camera – *triangulation*



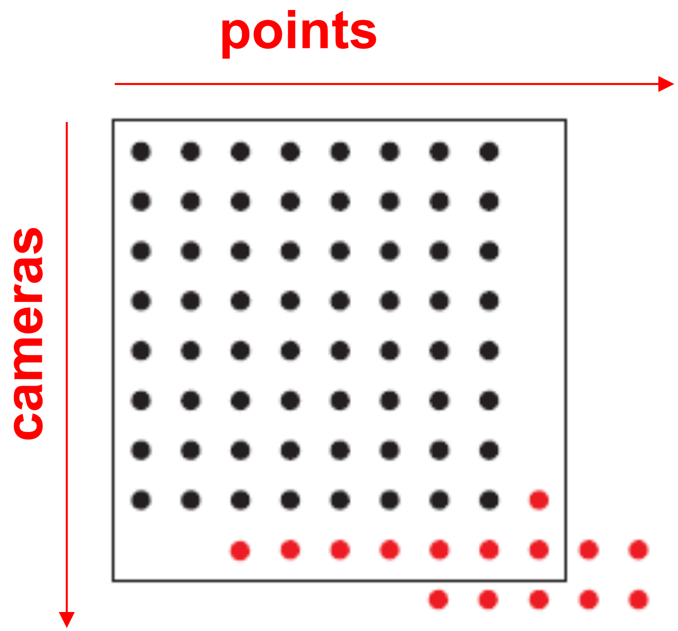
Incremental structure from motion

- Initialize motion from two images using fundamental matrix

- Initialize structure by triangulation

- For each additional view:
  - Determine projection matrix of new camera using all the known 3D points that are visible in its image – *calibration*
  - Refine and extend structure: compute new 3D points, re-optimize existing points that are also seen by this camera – *triangulation*

- Refine structure and motion: bundle adjustment

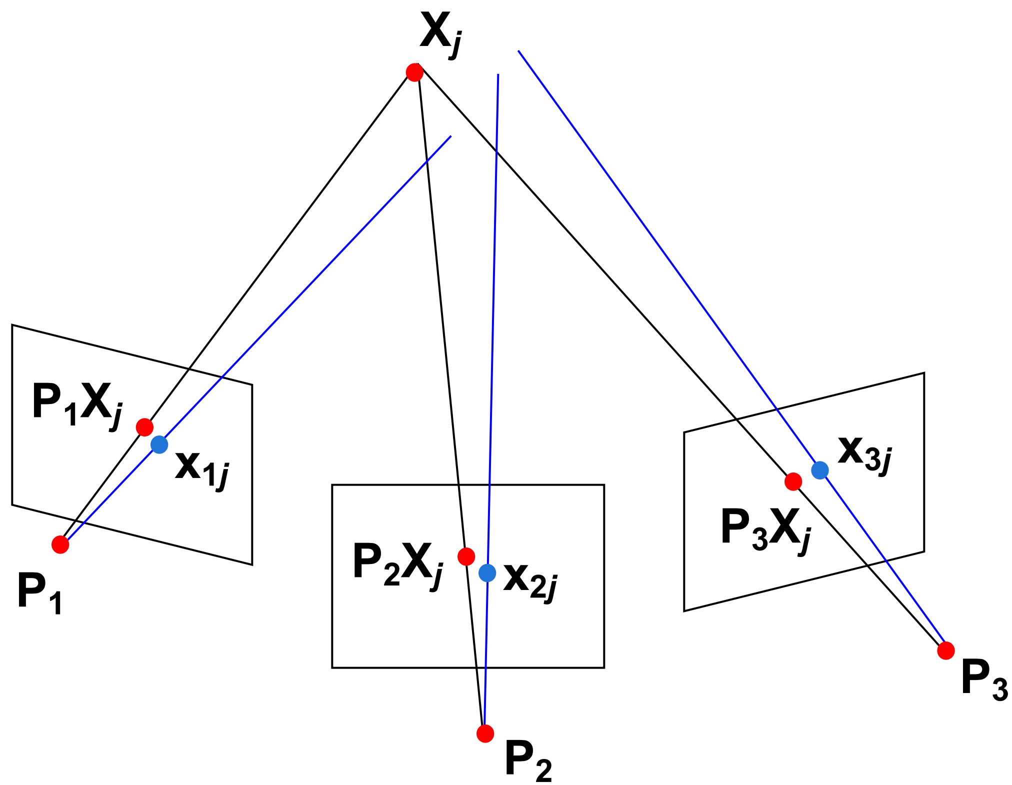


Bundle adjustment

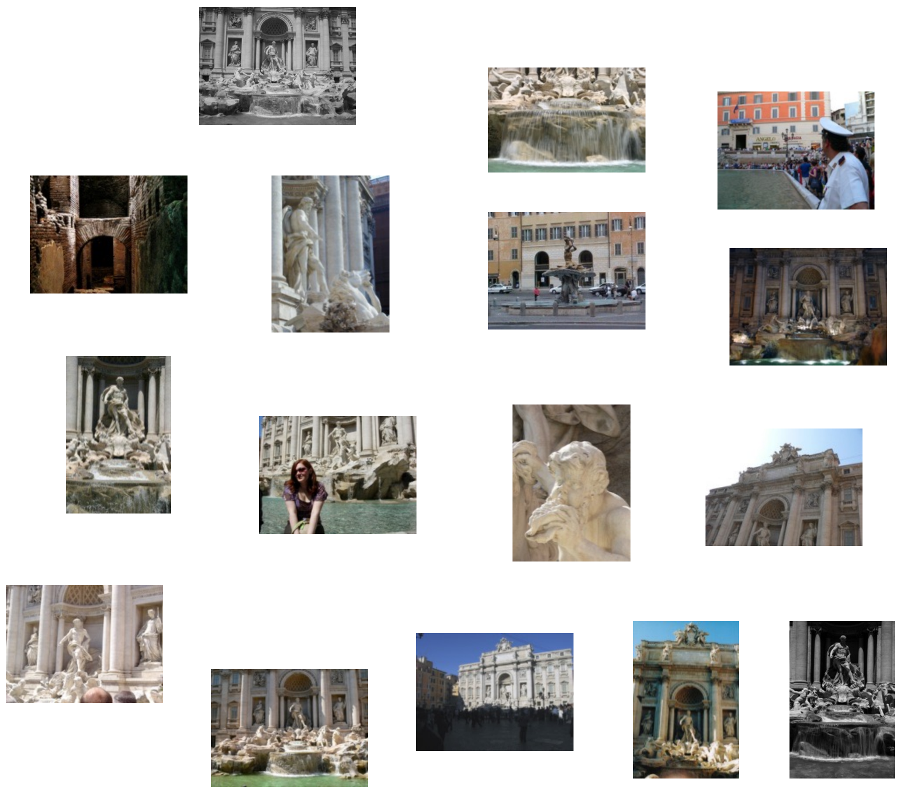
- Non-linear method for refining structure and motion
- Minimize reprojection error

$$\sum_{i=1}^{m}\sum_{j=1}^{n} w_{ij}\left\|\mathbf{x}_{ij}-\frac{1}{\lambda_{ij}}\mathbf{P}_i\mathbf{X}_j\right\|^2$$

visibility flag: is point j visible in view i?



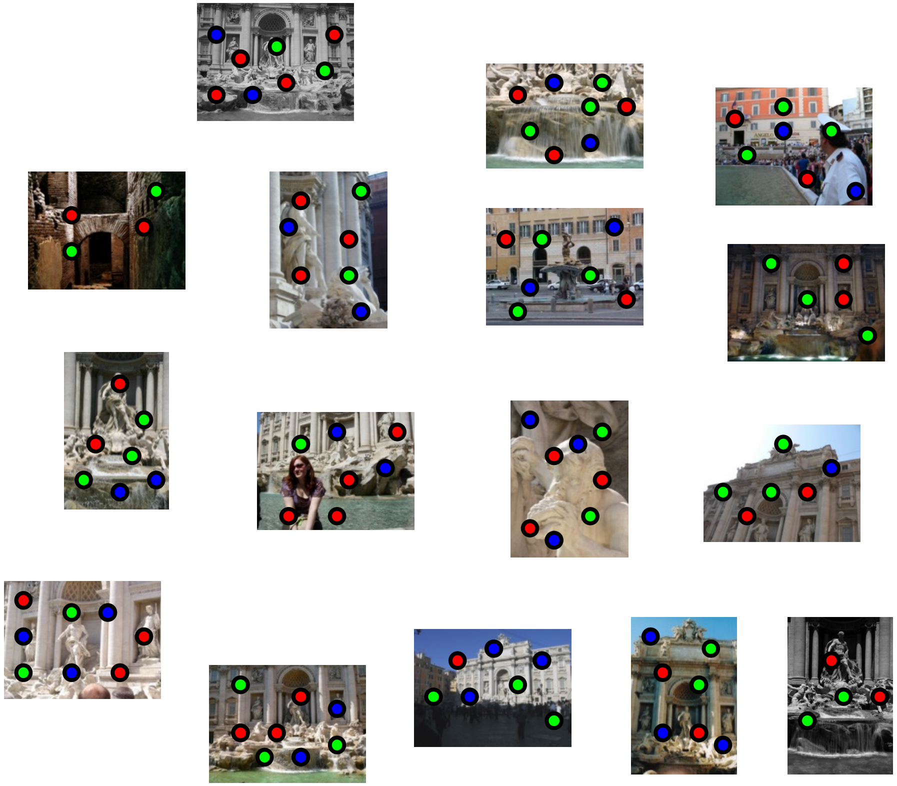
Feature detection



Source: N. Snavely

Feature detection

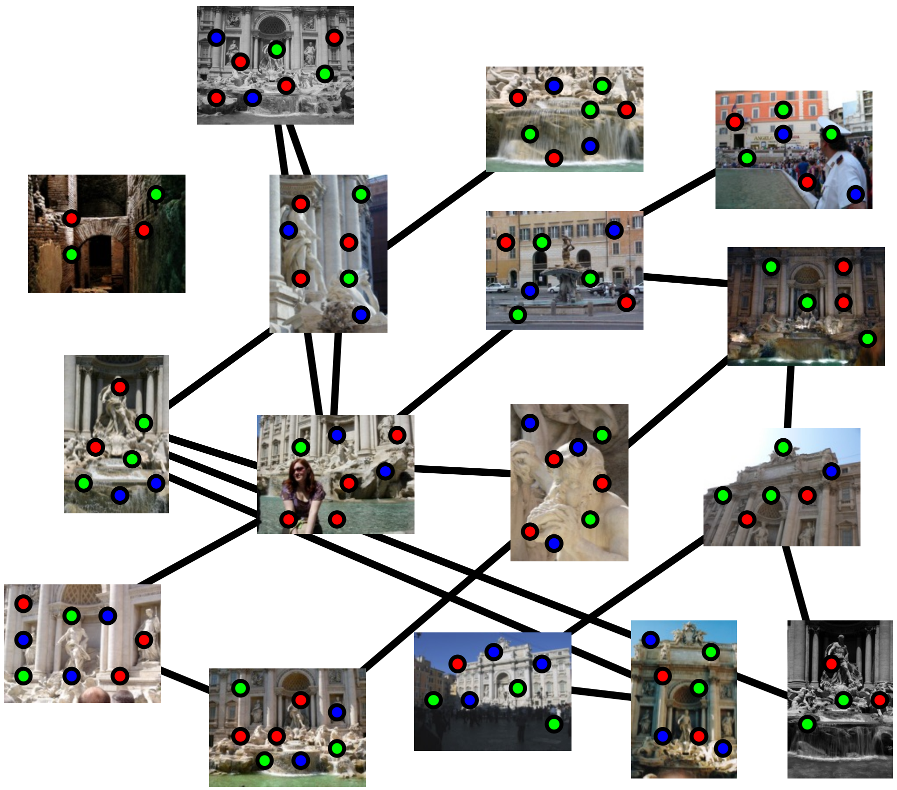
Detect SIFT features



Source: N. Snavely

Feature matching

Match features between each pair of images



Source: N. Snavely

The devil is in the details

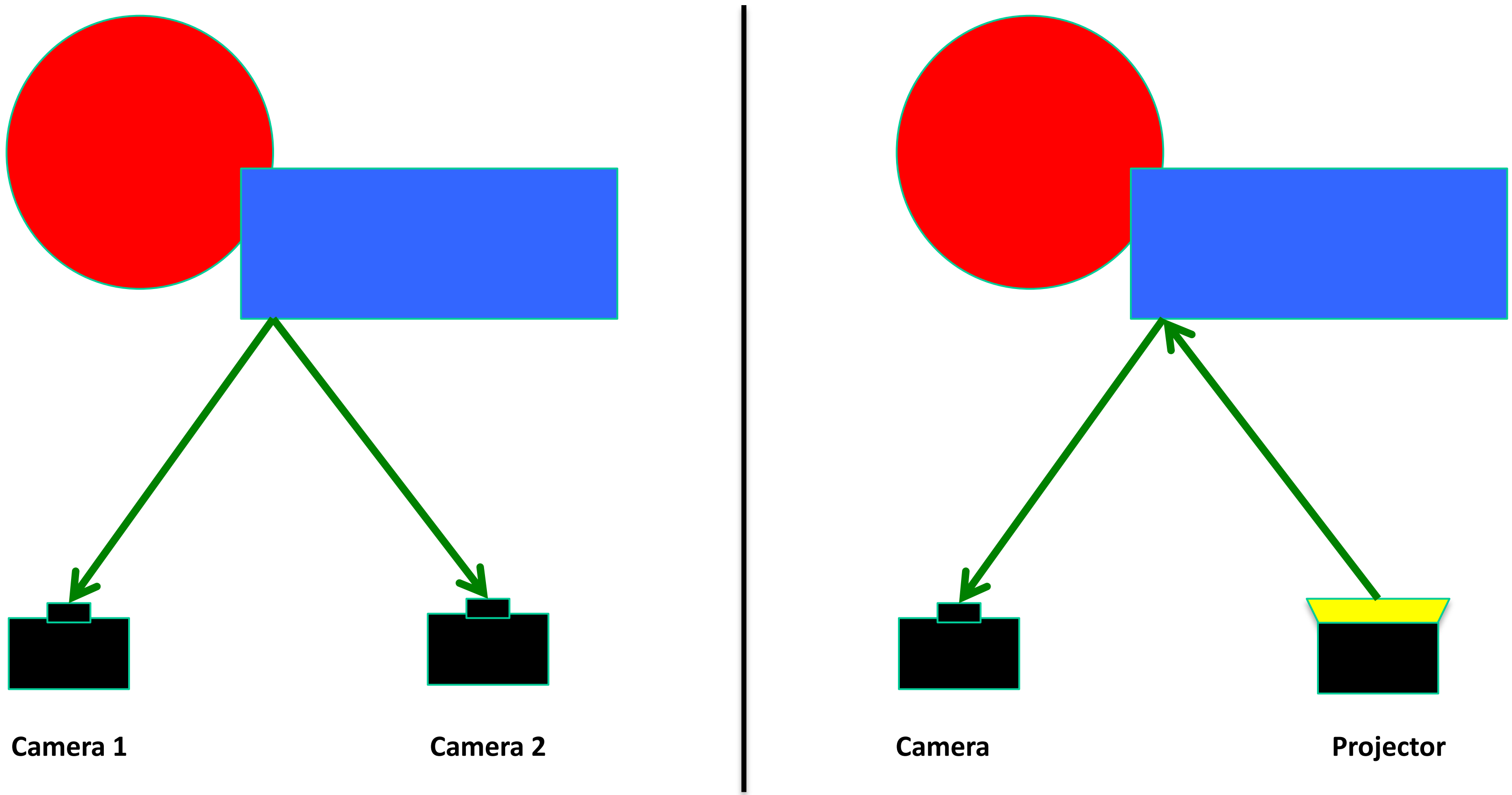
- Handling ambiguities
- Handling degenerate configurations (e.g., homographies)
- Eliminating outliers
- Dealing with repetitions and symmetries

Photo Tourism

N. Snavely, S. Seitz, and R. Szeliski, Photo tourism: Exploring photo collections in 3D, SIGGRAPH 2006.

http://phototour.cs.washington.edu/, http://grail.cs.washington.edu/projects/rome/

Depth from Triangulation



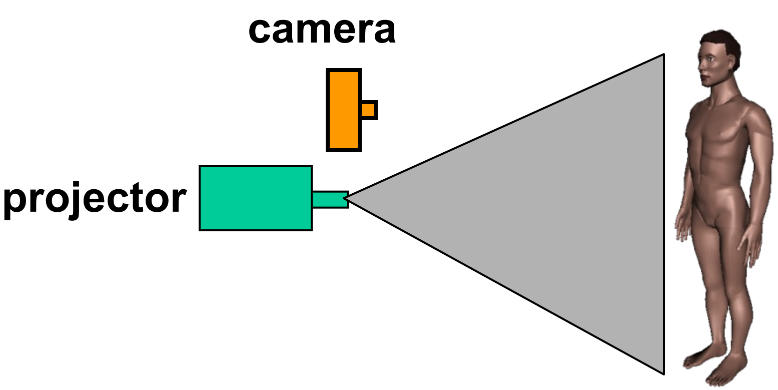
Passive Stereopsis

Active Stereopsis

Active sensing simplifies the problem of estimating point correspondences

Active stereo with structured light

- Project “structured” light patterns onto the object
  - Simplifies the correspondence problem
  - Allows us to use only one camera



L. Zhang, B. Curless, and S. M. Seitz. Rapid Shape Acquisition Using Color Structured Light and Multi-pass Dynamic Programming. *3DPVT* 2002

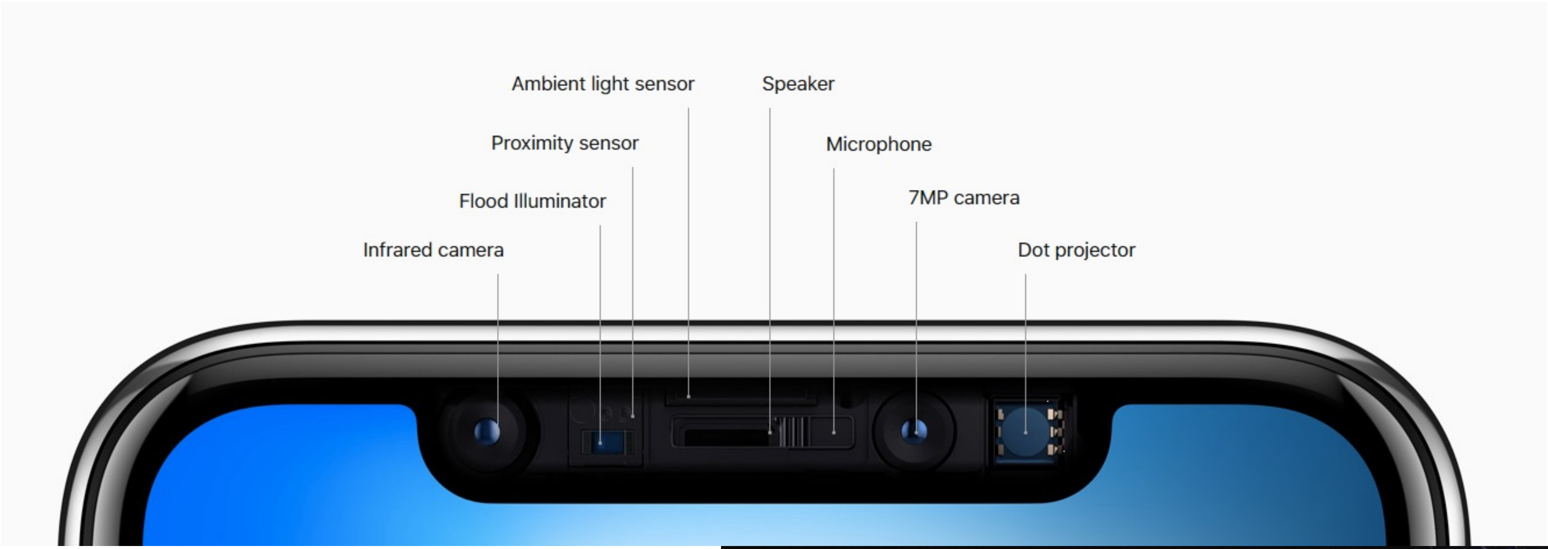
Slide from L. Lazebnik.

Kinect: Structured infrared light

http://bbzippo.wordpress.com/2010/11/28/kinect-in-infrared/

Slide from L. Lazebnik.

Apple TrueDepth



https://www.cnet.com/news/apple-face-id-truedepth-how-it-works/

Slide from L. Lazebnik.

SFM software

- Bundler
- OpenSfM
- OpenMVG
- VisualSFM
- Colmap
- See also Wikipedia’s list of toolboxes

Basis for SLAM

- Specialized sensors
- Approximately know camera location
- Need dense reconstructions for path-planning
- Needs to be fast

KinectFusion: Real-time 3D Reconstruction and Interaction Using a Moving Depth Camera*

Shahram Izadi[1], *David Kim*[1,3], *Otmar Hilliges*[1], *David Molyneaux*[1,4], *Richard Newcombe*[2], *Pushmeet Kohli*[1], *Jamie Shotton*[1], *Steve Hodges*[1], *Dustin Freeman*[1,5], *Andrew Davison*[2], *Andrew Fitzgibbon*[1]

[1]Microsoft Research Cambridge, UK [2]Imperial College London, UK
[3]Newcastle University, UK [4]Lancaster University, UK [5]University of Toronto, Canada

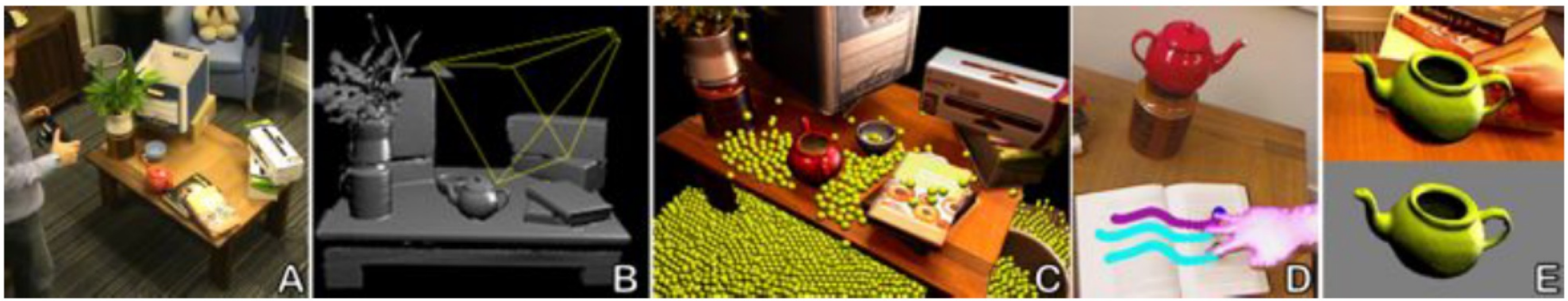


Figure 1: KinectFusion enables real-time detailed 3D reconstructions of indoor scenes using only the depth data from a standard Kinect camera. A) user points Kinect at coffee table scene. B) Phong shaded reconstructed 3D model (the wireframe frustum shows current tracked 3D pose of Kinect). C) 3D model texture mapped using Kinect RGB data with real-time particles simulated on the 3D model as reconstruction occurs. D) Multi-touch interactions performed on any reconstructed surface. E) Real-time segmentation and 3D tracking of a physical object.

Paper link (ACM Symposium on User Interface Software and Technology, October 2011)

YouTube Video

Reconstruction in construction industry

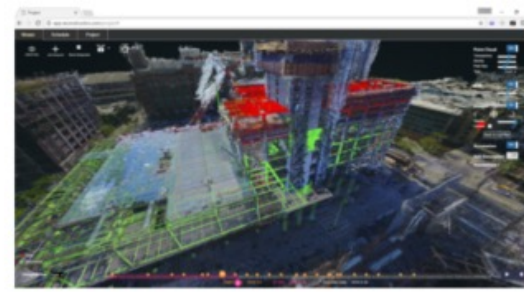
RECONSTRUCT INTEGRATES REALITY AND PLAN

Visual Asset Management

Reconstruct 4D point clouds and organize images and videos from smartphones, time-lapse cameras, and drones around the project schedule. View, annotate, and share anywhere with a web interface.

4D Visual Production Models

Integrate 4D point clouds with 4D BIM, review "who does what work at what location" on a daily basis and improve coordination and communication among project teams.



Predictive Visual Data Analytics

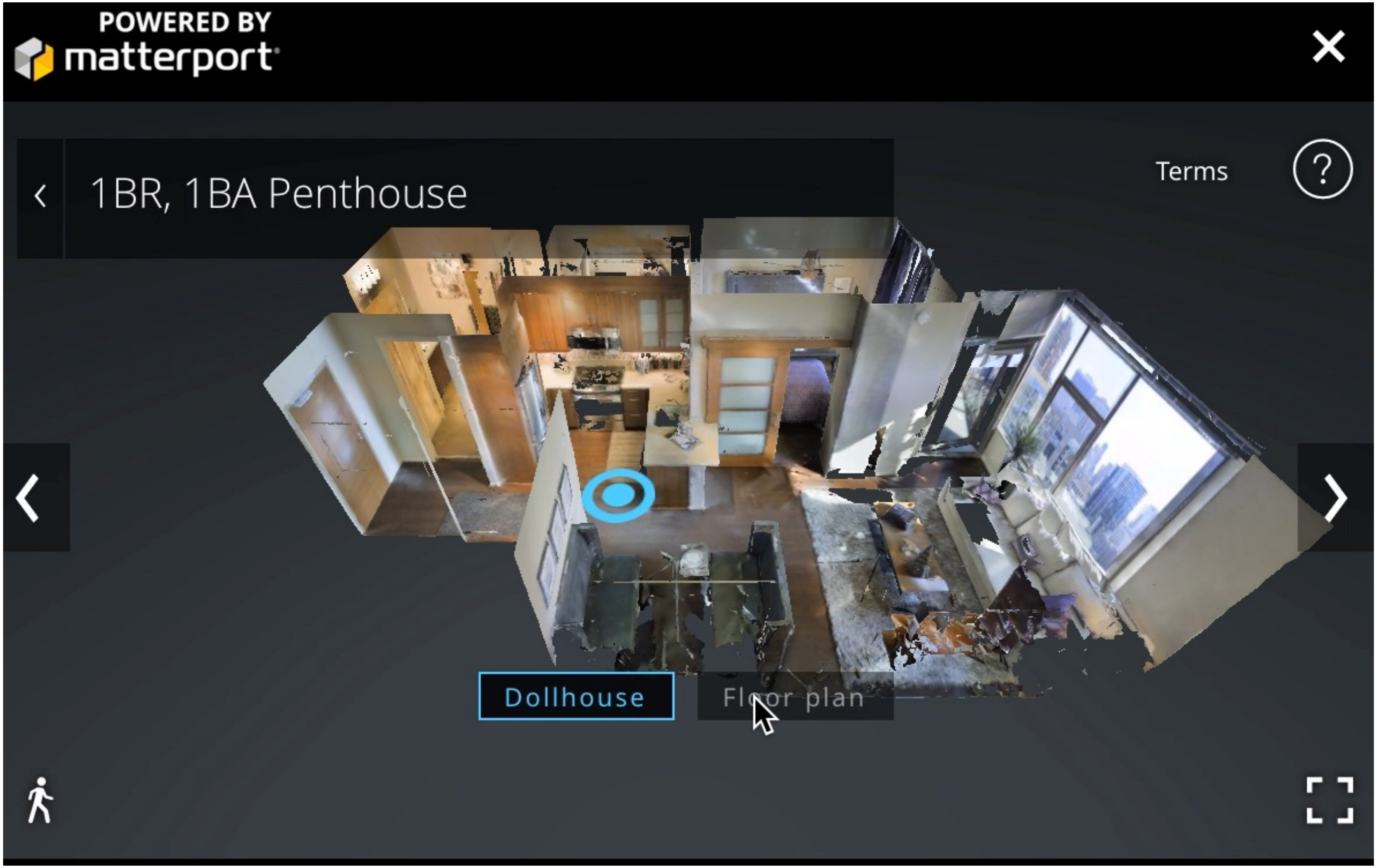
Analyze actual progress deviations by comparing Reality and Plan and predict risk with respect to the execution of the look-ahead schedule for each project location, to offer your project team with an opportunity to tap off potential delays before they surface on your jobsite.

reconstructinc.com

Source: L. Lazebnik

Source: D. Hoiem

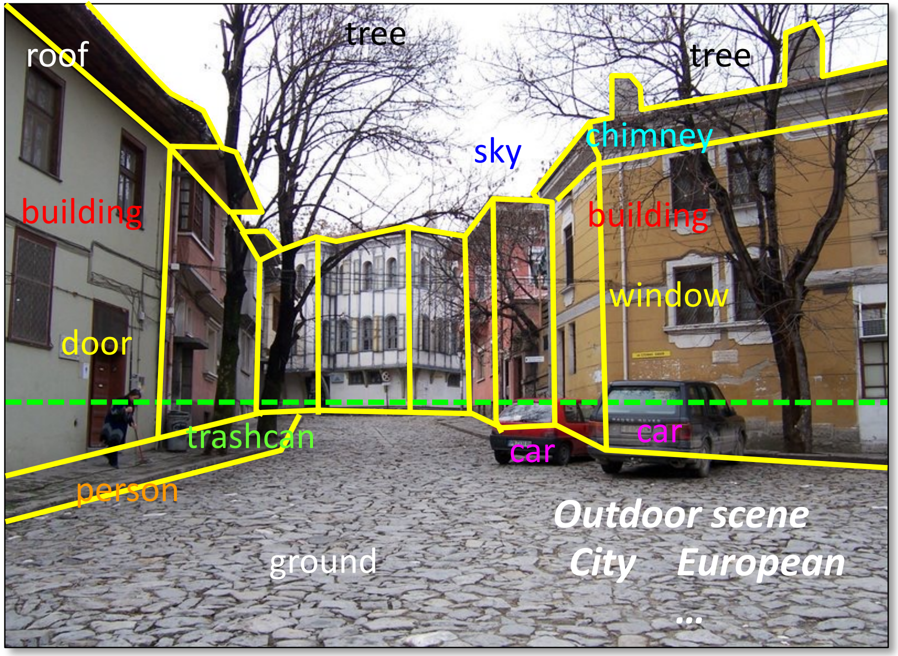
Applications



Source: N. Snavely

Interactive Example : https://matterport.com/en-gb/media/2486

What kind of information can be extracted from an image?



Geometric information

Semantic information

Source: L. Lazebnik