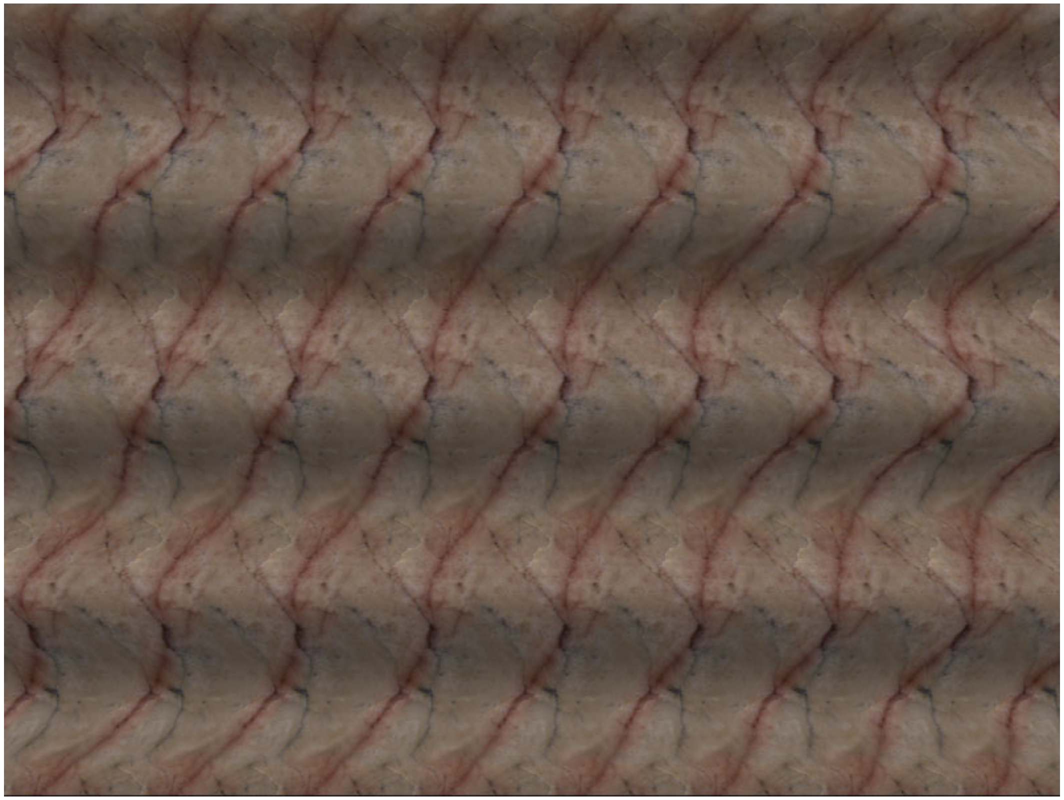
CS4670/5670: Computer Vision

Kavita Bala

Lec 25: Stereo

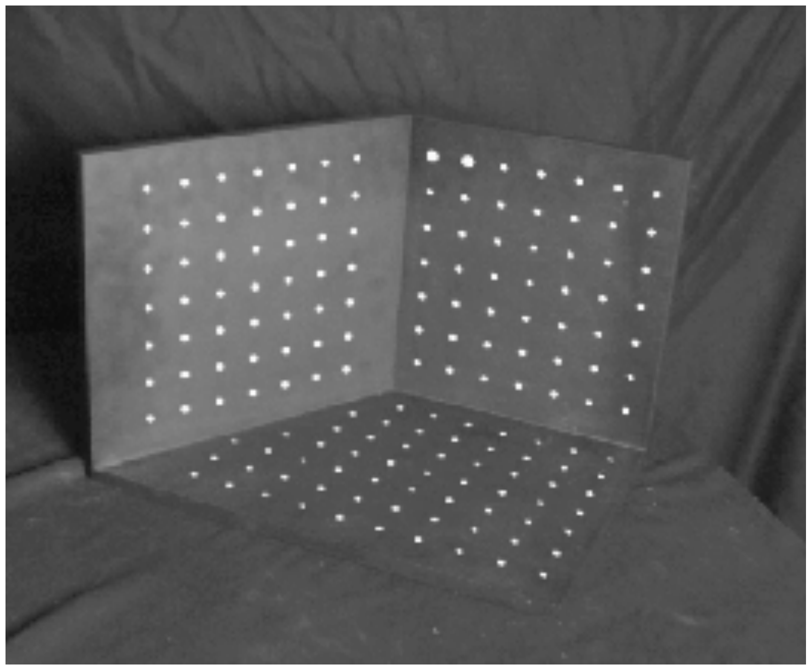


Announcements

- Prelim grades out
  - Solutions on CMS
  - Regrades on Problem 4 and 6
    - Please drop off prelim in hand back room or in office hours with TAs

Calibration using a reference object

- Place a known object in the scene
  - identify correspondence between image 2D and scene 3D
  - compute mapping from scene to image



Issues

- must know geometry very accurately
- must know 3D->2D correspondence

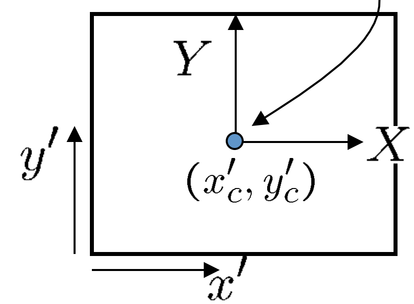
Camera parameters

A camera is described by several parameters

- Translation T of the optical center from the origin of world coords
- Rotation R of the image plane
- focal length f, principal point (x'_c, y'_c), pixel size (s_x, s_y)
- blue parameters are called "extrinsics," red are "intrinsics"

Projection equation

$$\mathbf{X} = \begin{bmatrix} sx \\ sy \\ s \end{bmatrix} = \begin{bmatrix} * & * & * & * \\ * & * & * & * \\ * & * & * & * \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = \Pi \mathbf{X}$$



- The projection matrix models the cumulative effect of all parameters
- Useful to decompose into a series of operations

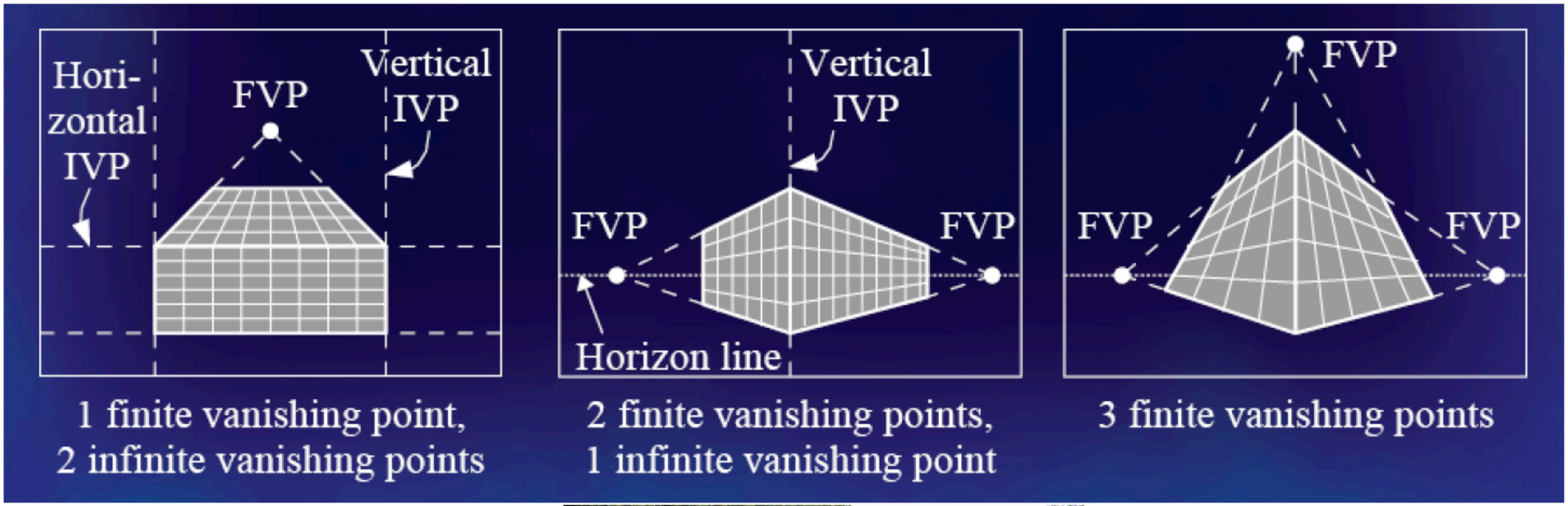
identity matrix

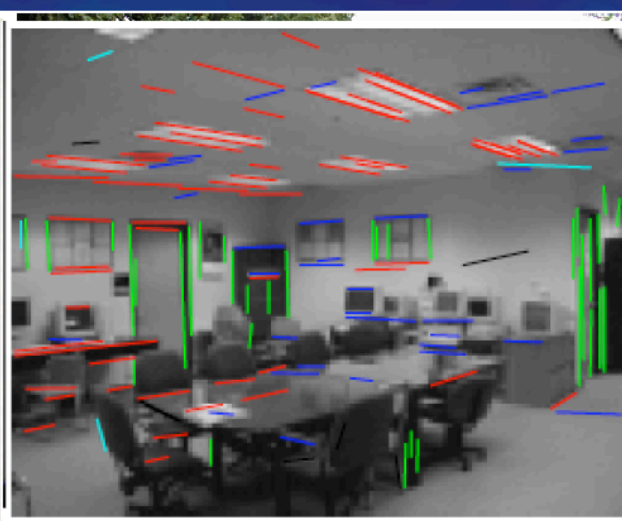
$$\mathbf{\Pi} = \begin{bmatrix} -fs_x & 0 & x'_c \\ 0 & -fs_y & y'_c \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{R}_{3x3} & \mathbf{0}_{3x1} \\ \mathbf{0}_{1x3} & 1 \end{bmatrix} \begin{bmatrix} \mathbf{I}_{3x3} & \mathbf{T}_{3x1} \\ \mathbf{0}_{1x3} & 1 \end{bmatrix}$$

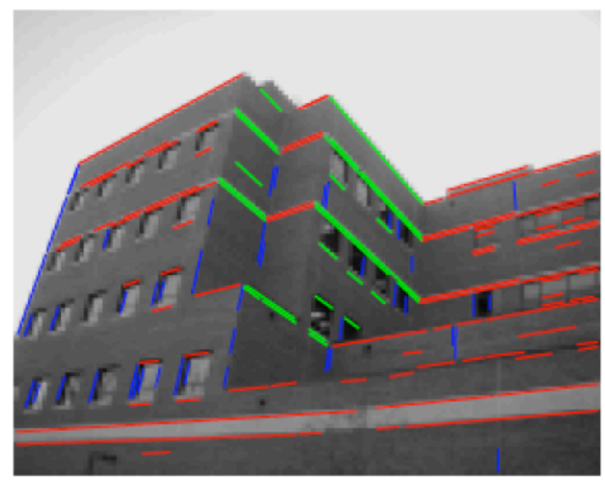
intrinsics projection rotation translation

- The definitions of these parameters are **not** completely standardized
  - especially intrinsics—varies from one book to another

Calibration from vanishing points







Calibration from 3 vanishing points

3 orthogonal world space axes

$$e_i = [1, 0, 0, 0]^T e_j = [0, 1, 0, 0]^T e_k = [0, 0, 1, 0]^T$$

Vanishing points and relation to axes

$$v_i = KRe_i, v_j = KRe_j, v_k = KRe_k$$

$$e_i^T e_j = 0$$

$$\mathbf{v}_i^T K^{-T} R R^T K^{-1} \mathbf{v}_j = \mathbf{v}_i^T K^{-T} K^{-1} \mathbf{v}_j \quad = 0$$

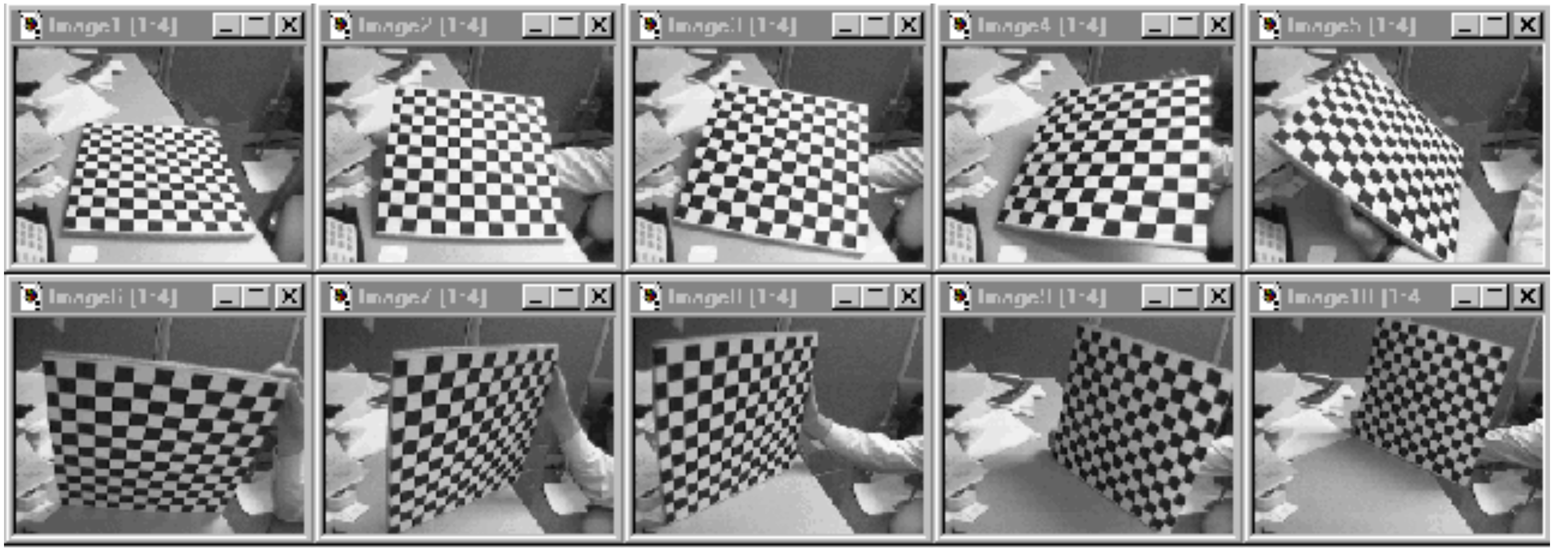
$$K = \begin{bmatrix} f & 0 & u_0 \\ 0 & f & v_0 \\ 0 & 0 & 1 \end{bmatrix} \quad K^{-1} = \begin{bmatrix} 1/f & 0 & -u_0/f \\ 0 & 1/f & -v_0/f \\ 0 & 0 & 1 \end{bmatrix}$$

$$v_i^T K^{-T} K^{-1} v_j = 0$$
$$v_j^T K^{-T} K^{-1} v_k = 0$$

$$v_i^T K^{-T} K^{-1} v_k = 0$$

- 3 finite vanishing points: get f, u0, v0

Alternative: multi-plane calibration

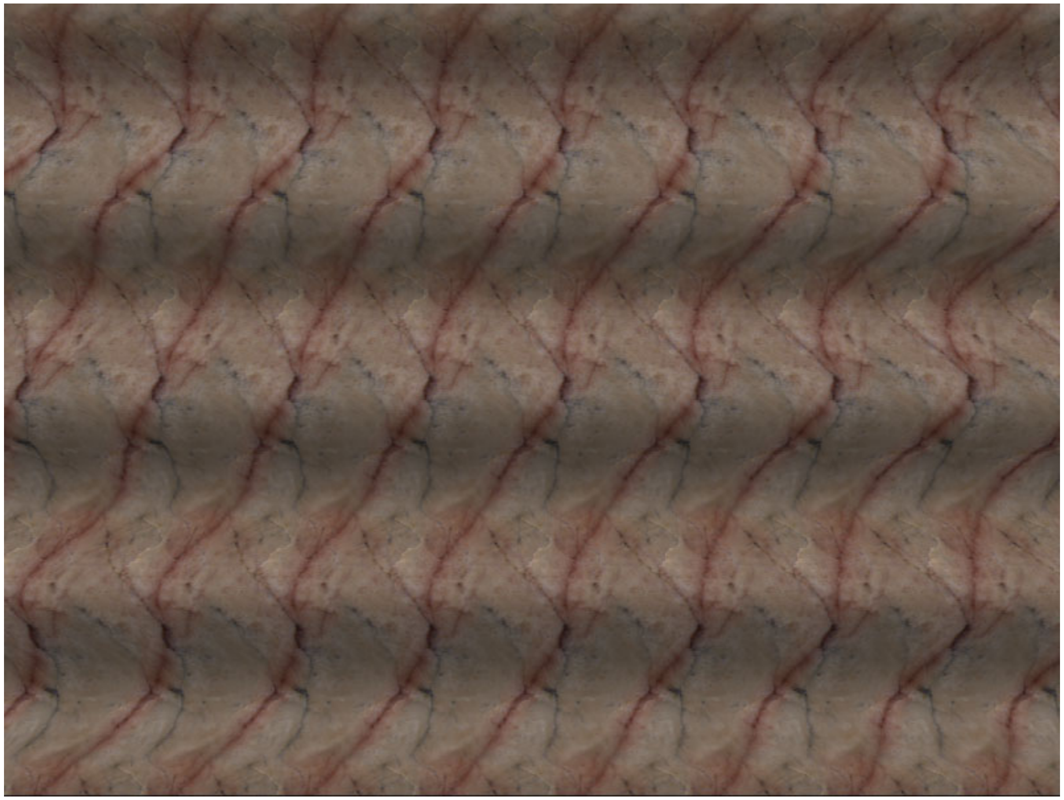


Images courtesy Jean-Yves Bouguet, Intel Corp.

Advantage

- Only requires a plane
- Don't have to know positions/orientations
- Good code available online! (including in OpenCV)
  - Matlab version by Jean-Yves Bouget: http://www.vision.caltech.edu/bouguetj/calib_doc/index.html
  - Zhengyou Zhang's web site: http://research.microsoft.com/~zhang/Calib/

Stereo



Single image stereogram, by Niklas Een

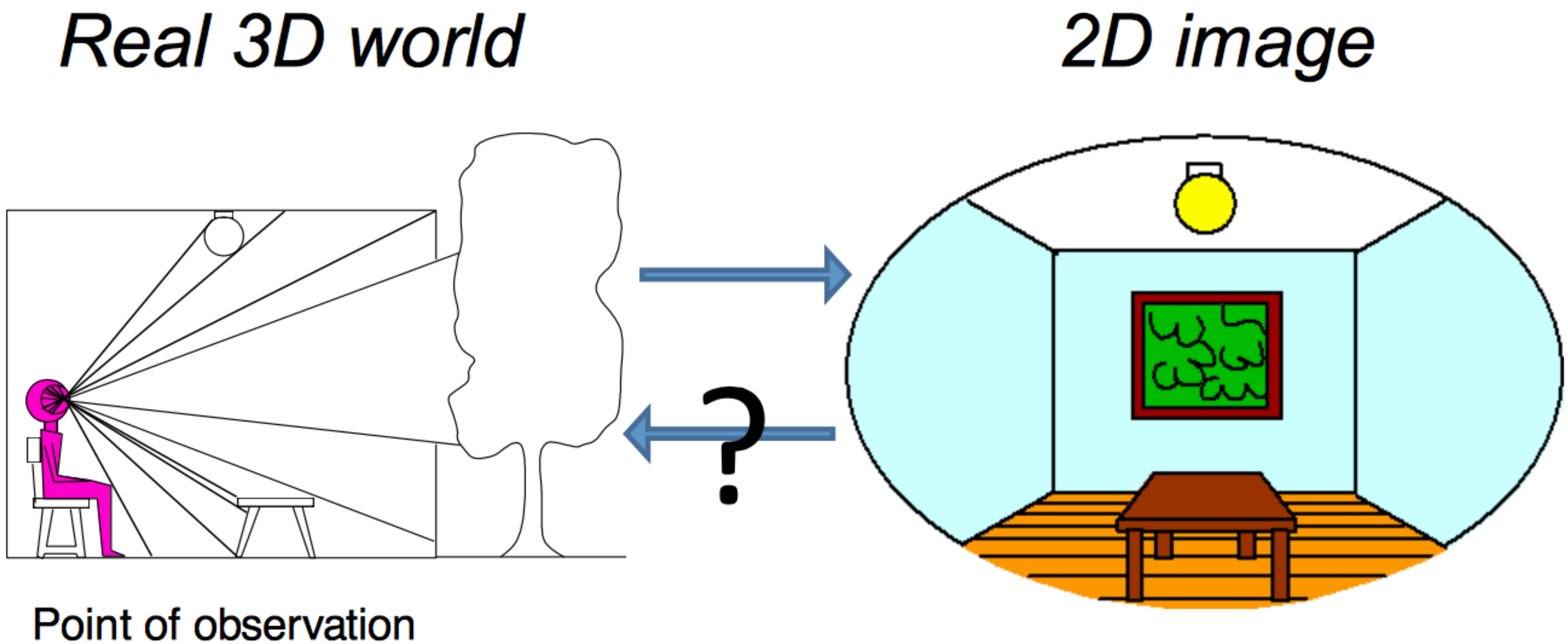
Announcements

- Readings
  - Szeliski, Chapter 7.1 and 7.2
  - Fundamental matrix song

Next two lectures

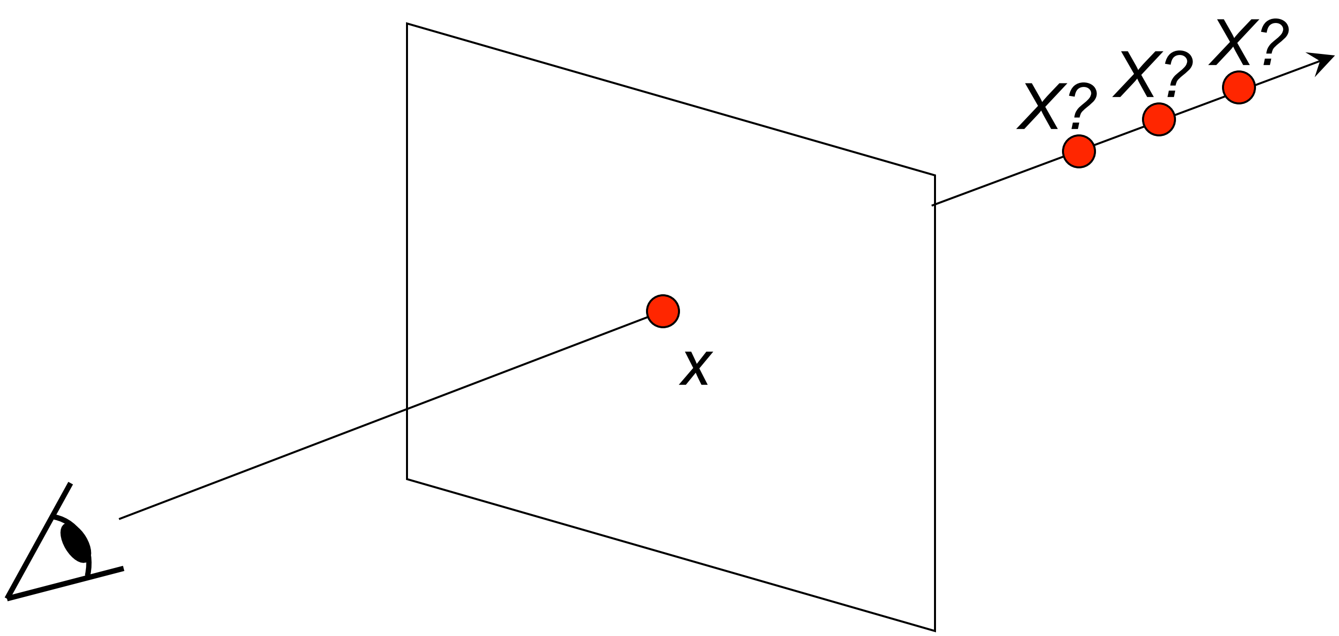
- Intro to stereo vision
- Epipolar geometry
- Fundamental matrix

Recovering 3D from images



Projective Geometry

- Recovery of structure from one image is inherently ambiguous



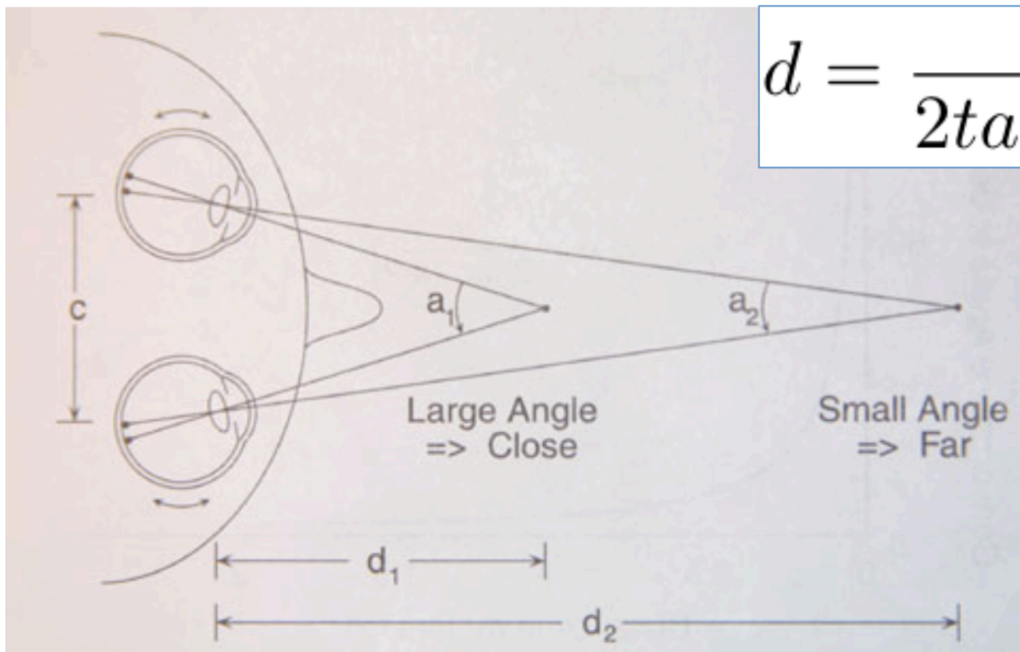
Visual Cues for 3D

- Shading
- Texture
- Shadows
- Focus
- Motion

- Shape from X: X is shading, texture, motion, ….

Stereo Reconstruction

- Shape from 2 (or more) images
- Motivated by human visual system

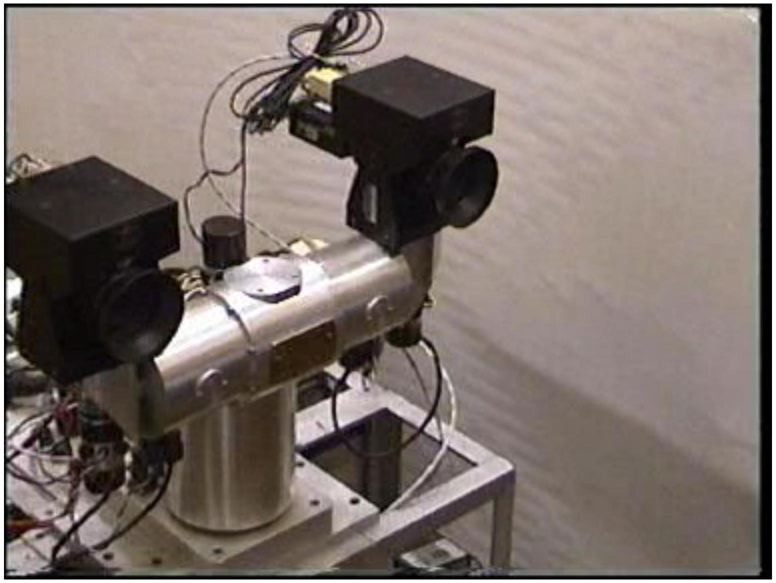


$$d = \frac{c}{2tan(a/2)}$$

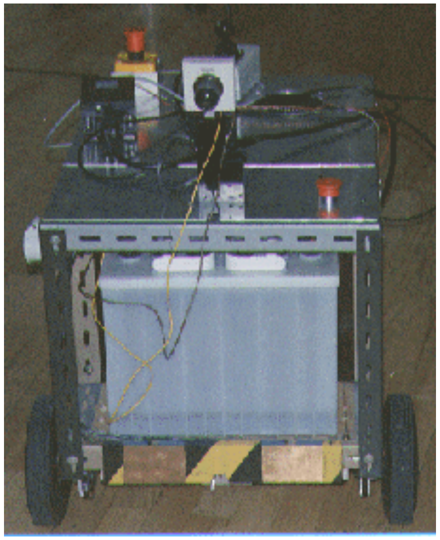
Human performance: up to 6-8 feet

Slide credit: J. Hayes

Stereo head



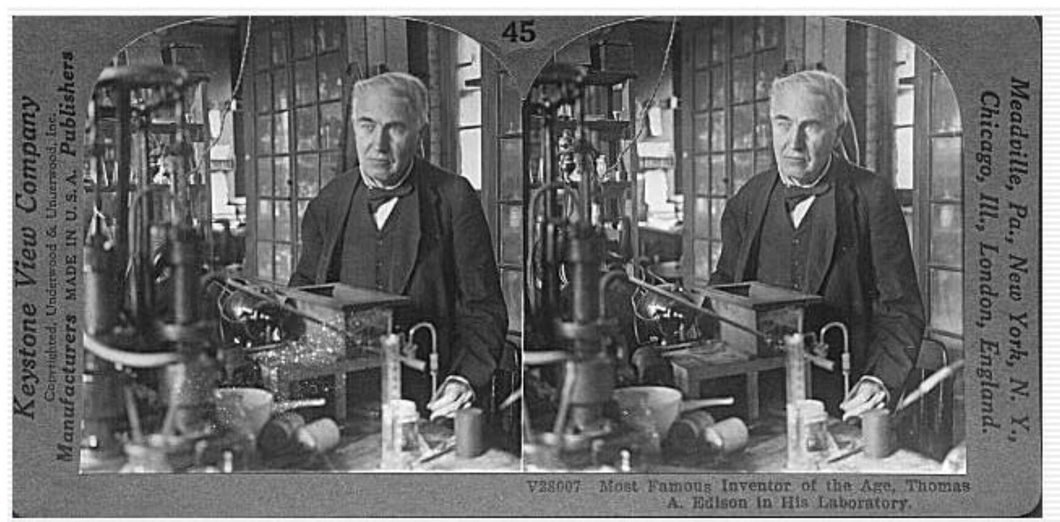
Camera on a mobile vehicle

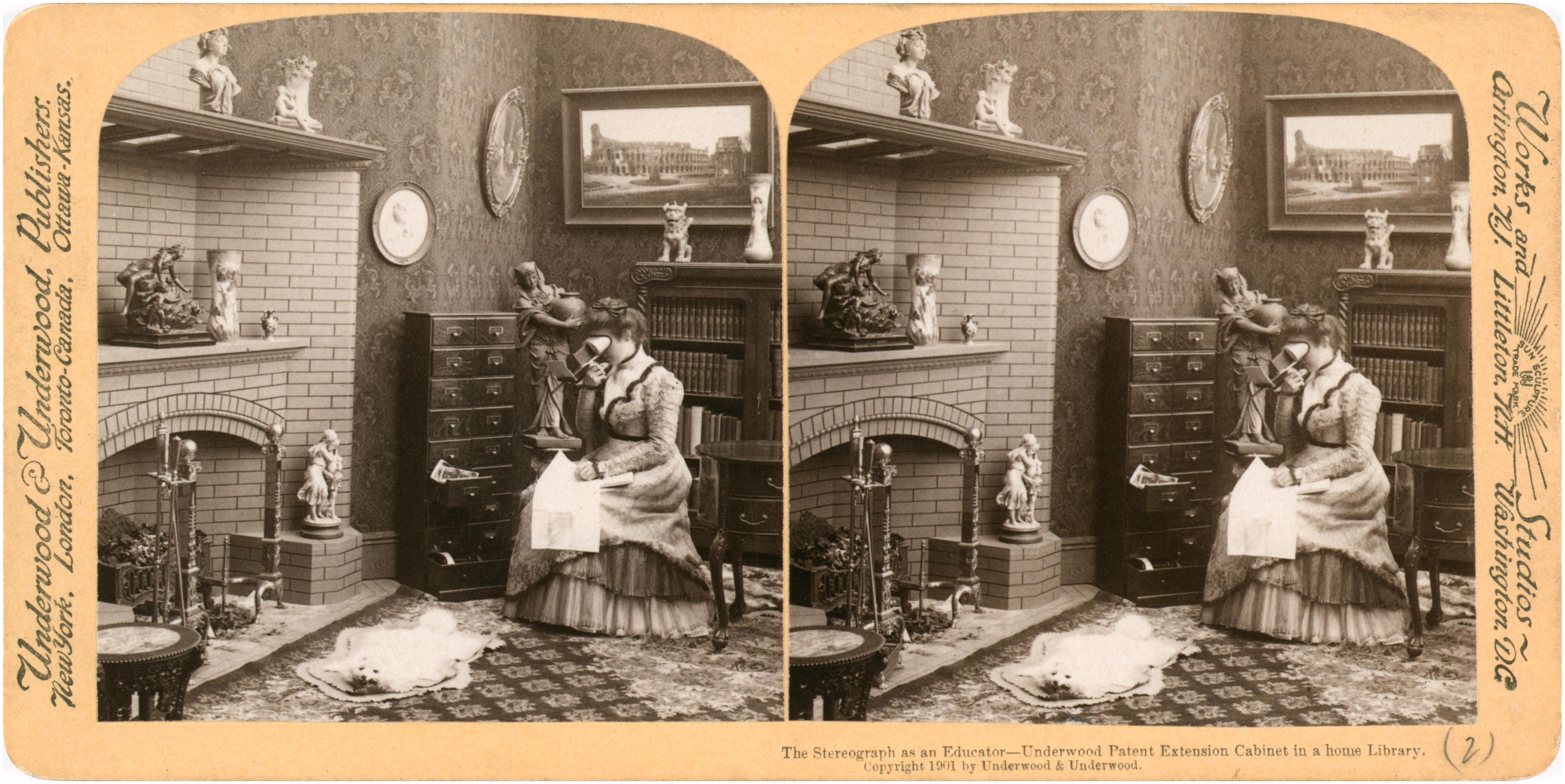


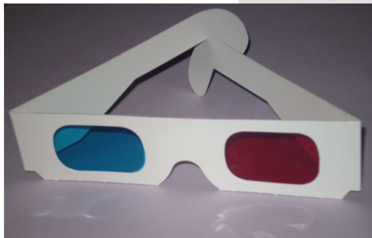
(COURTESY SONY)

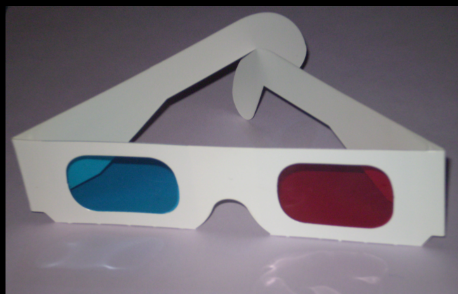
Stereograms

- Invented by Sir Charles Wheatstone, 1838



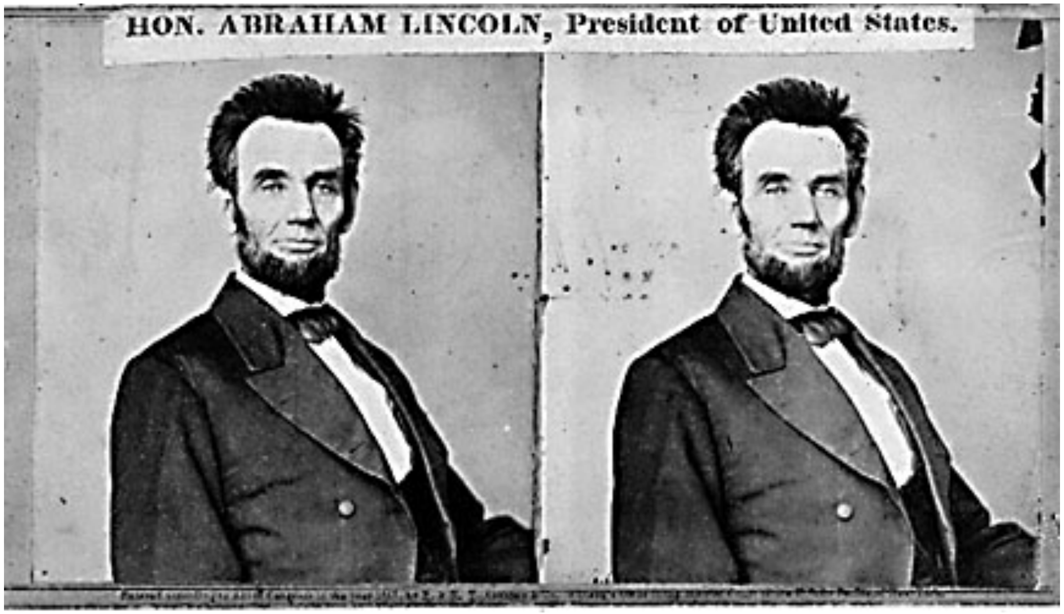

Underwood & Underwood, Publishers.
New York, London, Toronto-Canada, Ottawa-Kansas.
Works and Studios
Arlington, N.J. Littleton, N.H. Washington, D.C.
The Stereograph as an Educator—Underwood Patent Extension Cabinet in a home Library.
Copyright 1901 by Underwood & Underwood.
(2)



Mark Twain at Pool Table", no date, UCR Museum of Photography

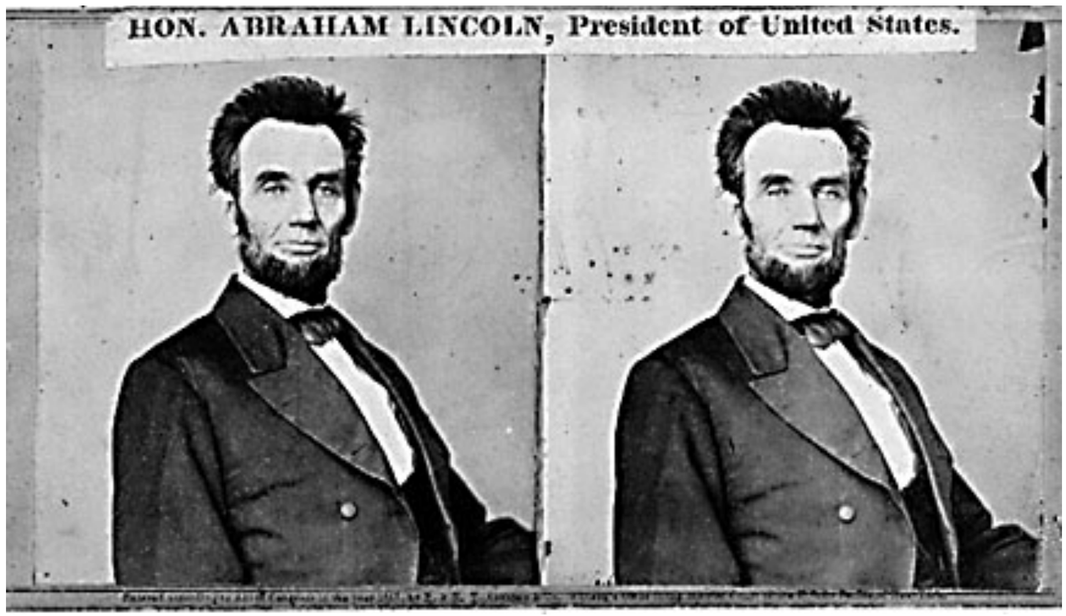
Stereo



- Given two images from different viewpoints
  - How can we compute the depth of each point in the image?
  - Based on *how much each pixel moves* between the two images

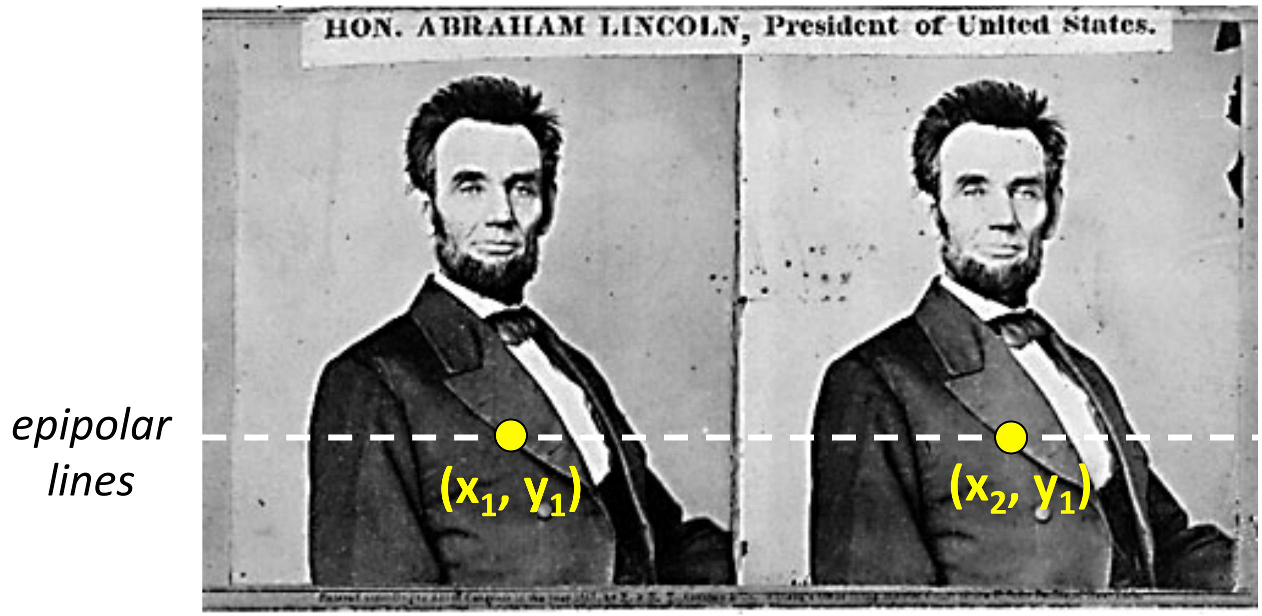
- https://github.com/404

Stereo



Two images captured by a purely horizontal translating camera
(*rectified* stereo pair)

Epipolar geometry

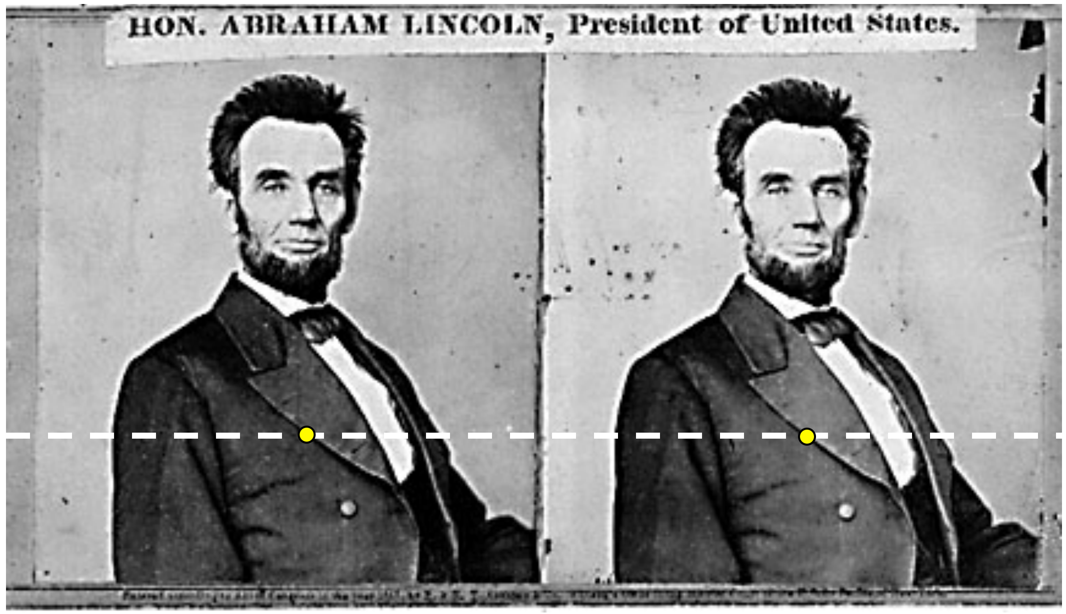


Two images captured by a purely horizontal translating camera
(*rectified* stereo pair)

x_2 -x_1 = the *disparity* of pixel (x_1, y_1)
Disparity is inversely proportional to depth

Disparity and Depth

Your basic stereo algorithm



For each epipolar line

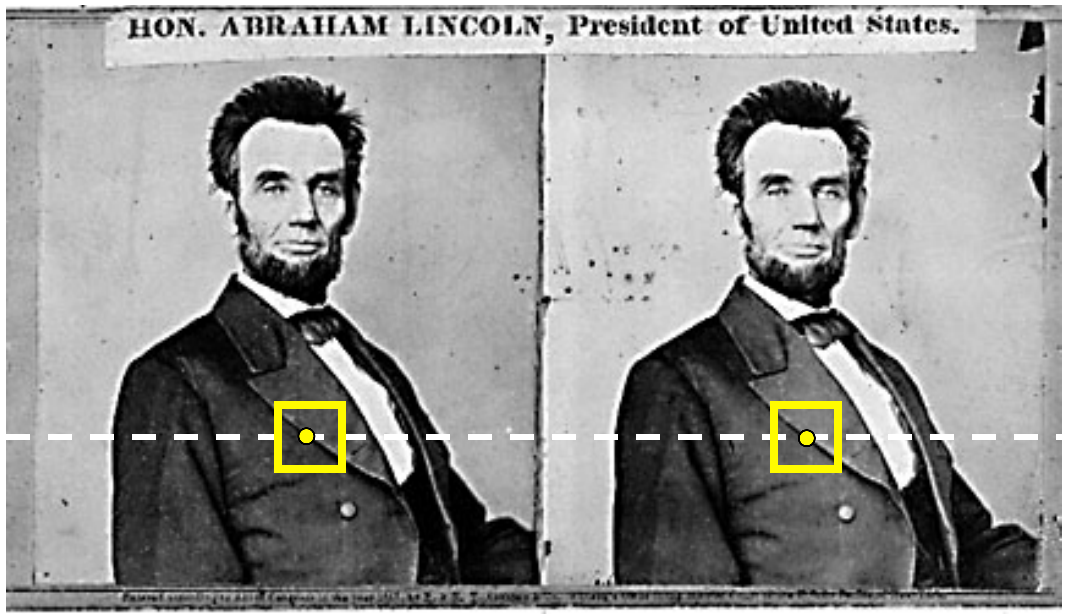
For each pixel in the left image

- compare with every pixel on same epipolar line in right image
- pick pixel with minimum match cost
- Compute disparity

Stereo matching algorithms

- Match Pixels in conjugate epipolar Lines
  - Assume brightness constancy
  - This is a tough problem
  - Numerous approaches
    - A good survey and evaluation: http://www.middlebury.edu/stereo/

Your basic stereo algorithm



For each epipolar line

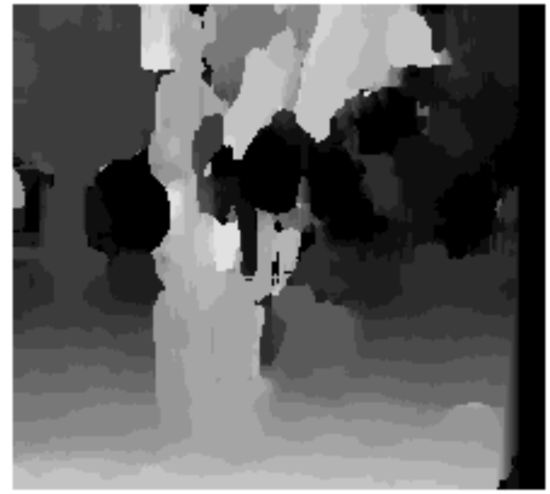
For each pixel in the left image

- compare with every pixel on same epipolar line in right image
- pick pixel with minimum match cost

Improvement: match ***windows***

Window size

W = 3



W = 20

Effect of window size

- Smaller window
  - + more detail
  - • more noise
- Larger window
  - + less noise
  - • less detail

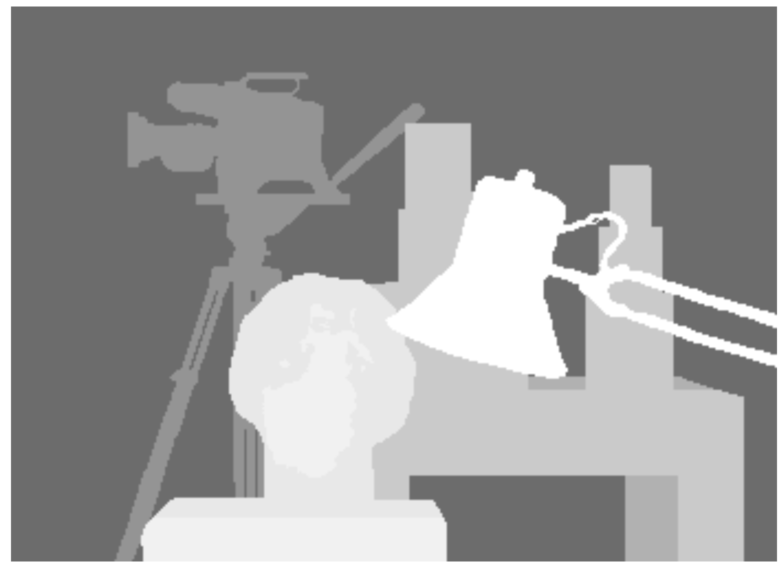
Better results with *adaptive window*

- T. Kanade and M. Okutomi, *A Stereo Matching Algorithm with an Adaptive Window: Theory and Experiment*,, Proc. International Conference on Robotics and Automation, 1991.
- D. Scharstein and R. Szeliski. Stereo matching with nonlinear diffusion. International Journal of Computer Vision, 28(2):155-174, July 1998

Stereo results

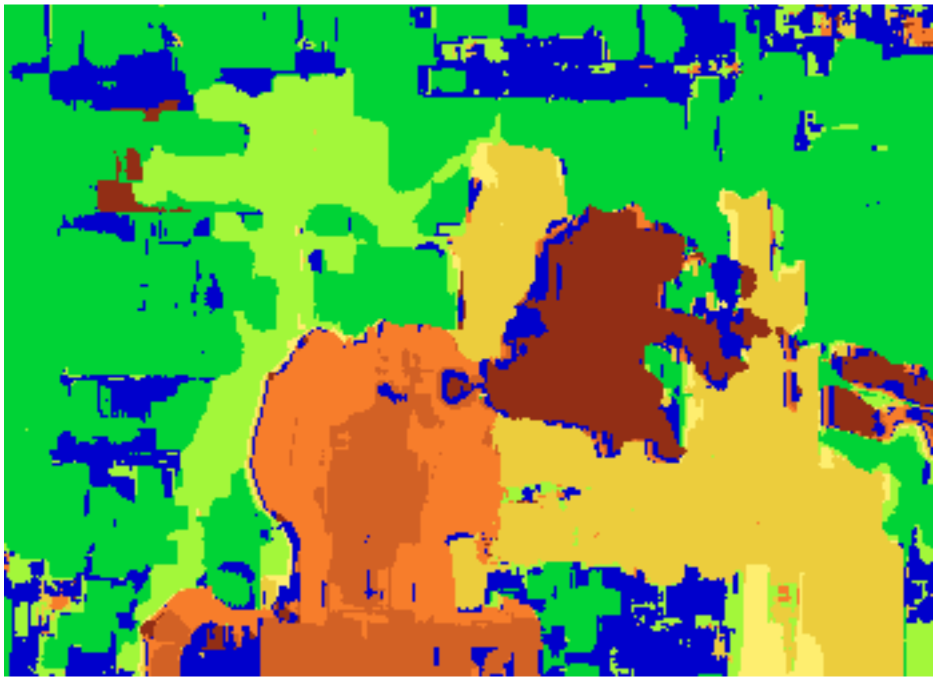
– Data from University of Tsukuba
– Similar results on other images without ground truth

Scene



Ground truth

Results with window search



Window-based matching
(best window size)

Ground truth

Better methods exist...

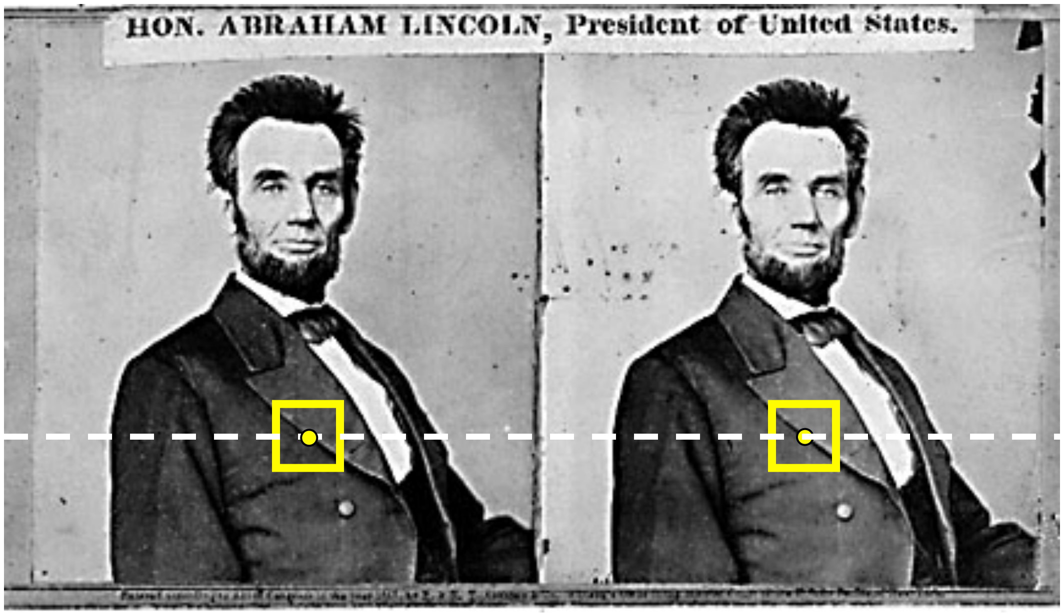
State of the art method

Boykov et al., [Fast Approximate Energy Minimization via Graph Cuts](), International Conference on Computer Vision, September 1999.

Ground truth

For the latest and greatest: http://www.middlebury.edu/stereo/

Back to stereo



- Where do epipolar lines come from?

- Have a good break!