

Stereo

EECS 442 – Prof. David Fouhey
Winter 2019, University of Michigan

http://web.eecs.umich.edu/~fouhey/teaching/EECS442_W19/

Congrats to P2 Kaggle Winners

- (Gold) William Dong (0.673)
- (Silver) Chockalingam Ravi Sundaram (0.669)
- (Bronze) VentusYue (0.662)
- Top-10: Hongzhi Yang, junzhexu, Shadow Demon, Izzy P, Mega Flygon, Naive, and goblue

Two-View Stereo



Stereo



How Two Photographers Unknowingly Shot the Same Millisecond in Time

MAR 07, 2018

RON RISMAN

PetaPixel



<https://petapixel.com/2018/03/07/two-photographers-unknowingly-shot-millisecond-time/>

How Two Photographers Unknowingly Shot the Same Millisecond in Time

PetaPixel

MAR 07, 2018

RON RISMAN

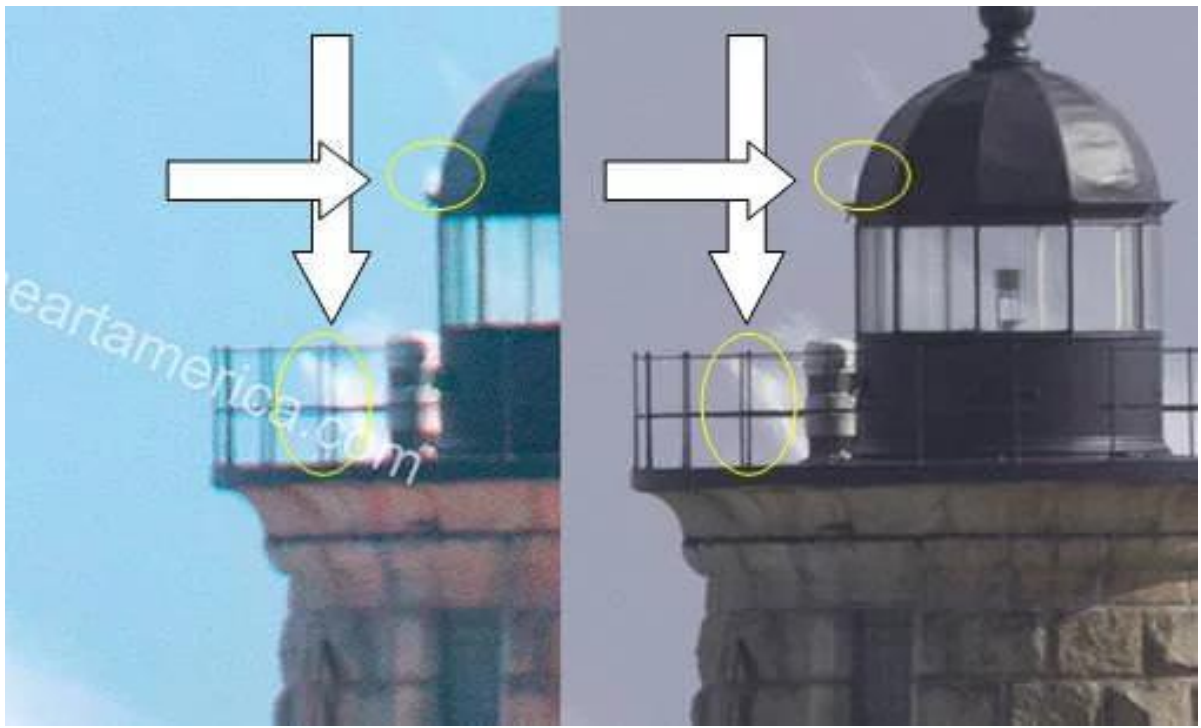


<https://petapixel.com/2018/03/07/two-photographers-unknowingly-shot-millisecond-time/>

How Two Photographers Unknowingly Shot the Same Millisecond in Time

MAR 07, 2018 RON RISMAN

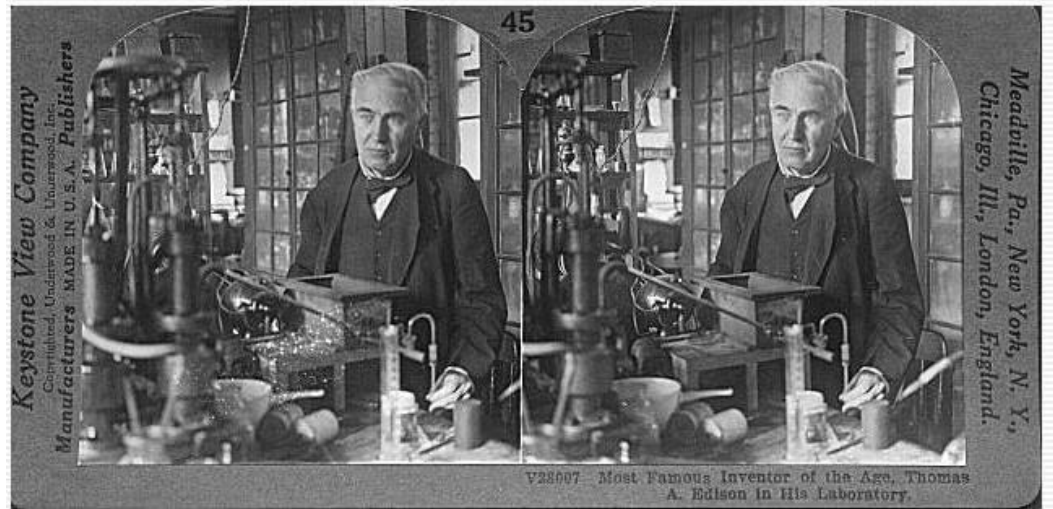
PetaPixel



<https://petapixel.com/2018/03/07/two-photographers-unknowingly-shot-millisecond-time/>

Stereograms

Humans can fuse pairs of images to get a sensation of depth



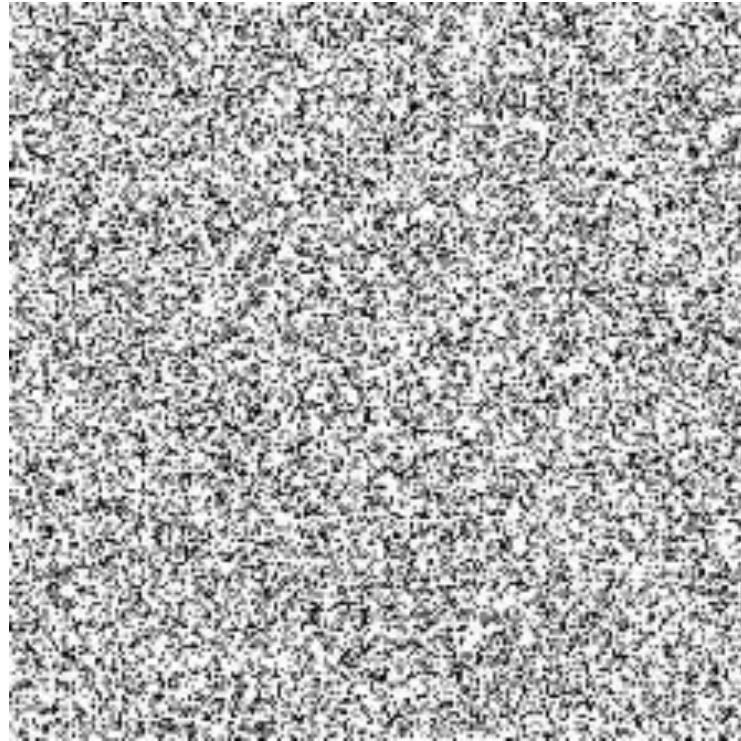
Stereograms: Invented by Sir Charles Wheatstone, 1838

Stereograms



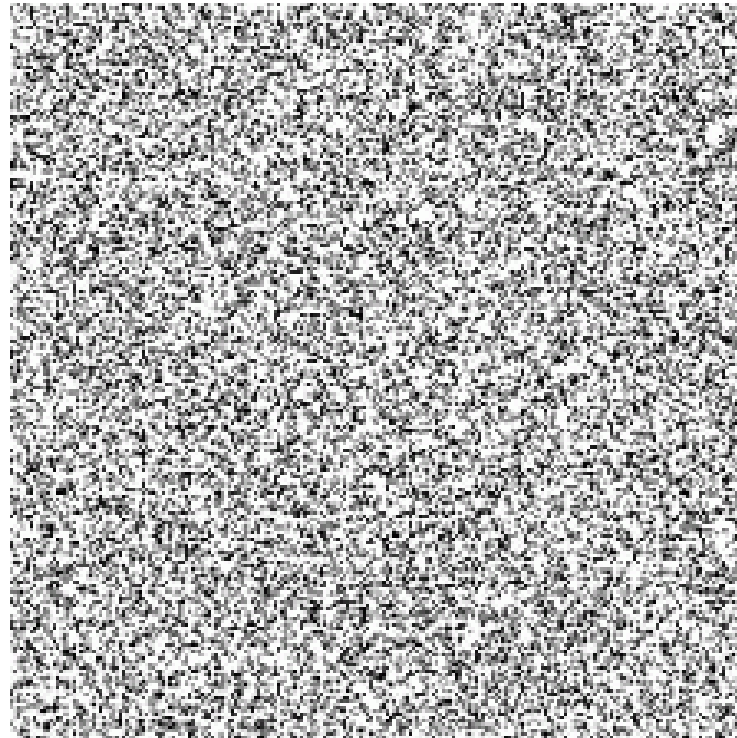
Stereograms

What about this?



Stereograms

Bela Julesz: Random Dot Stereogram
Shows that stereo can operate *without* recognition



Stereograms

Humans can fuse pairs of images to get a sensation of depth



Autostereograms: www.magiceye.com

Stereograms

Humans can fuse pairs of images to get a sensation of depth



Autostereograms: www.magiceye.com

Problem formulation

Given a calibrated binocular stereo pair, fuse it to produce a depth image

image 1



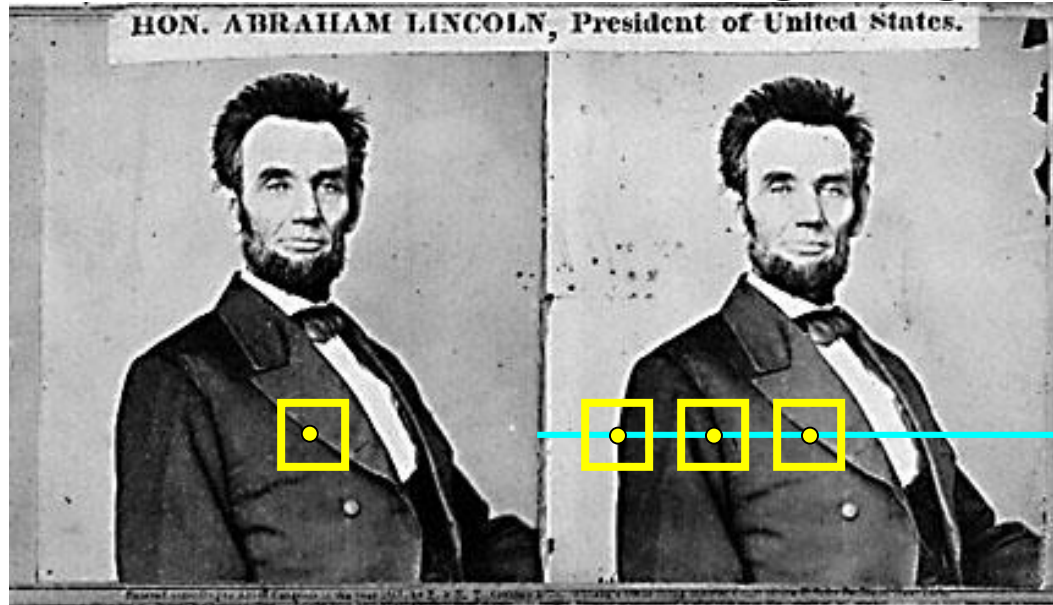
image 2



Dense depth map

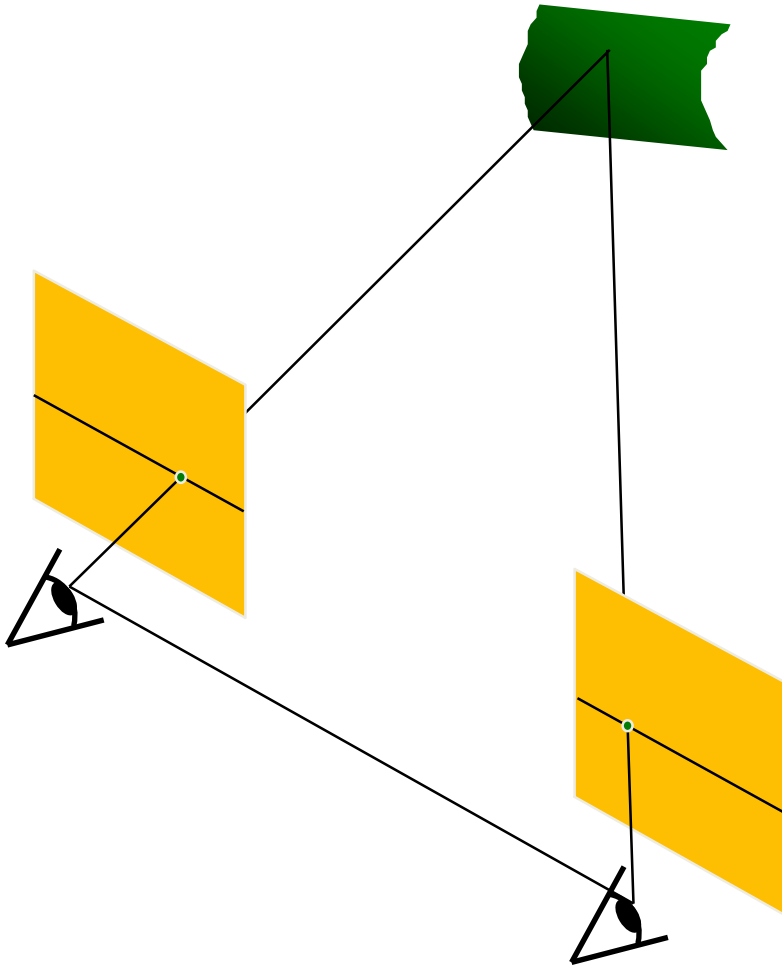


Basic stereo matching algorithm



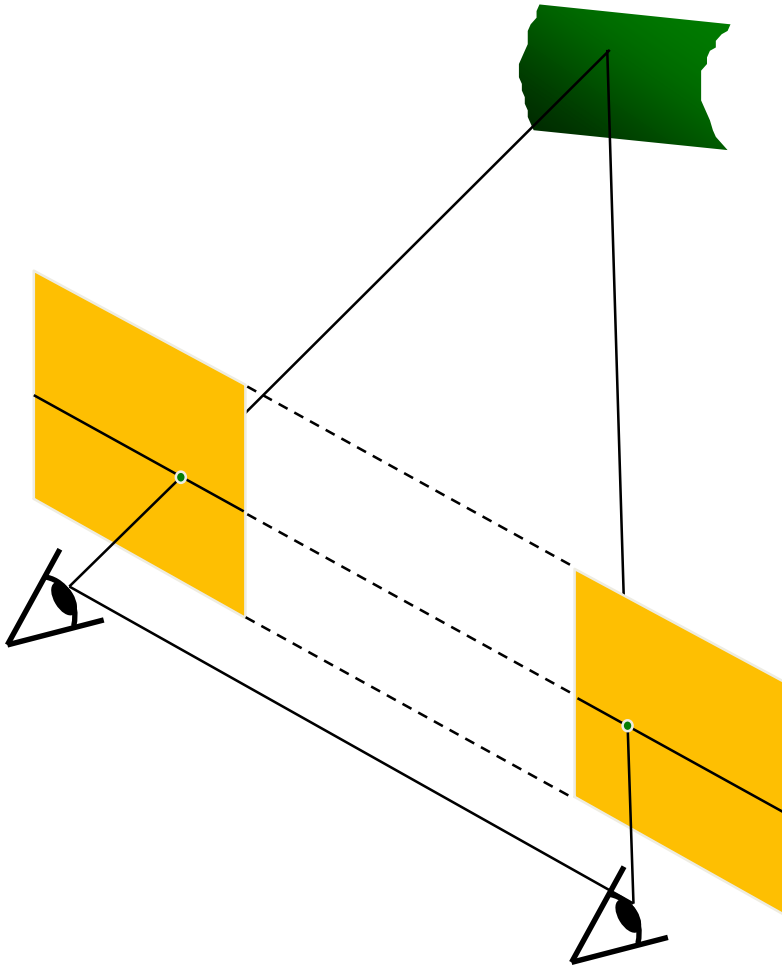
- For each pixel in the first image
 - Find corresponding epipolar line in the right image
 - Examine all pixels on the epipolar line and pick the best match
 - Triangulate the matches to get depth information
- Simplest case: epipolar lines = corresponding scanlines
 - When does this happen?

Simplest Case: Parallel images



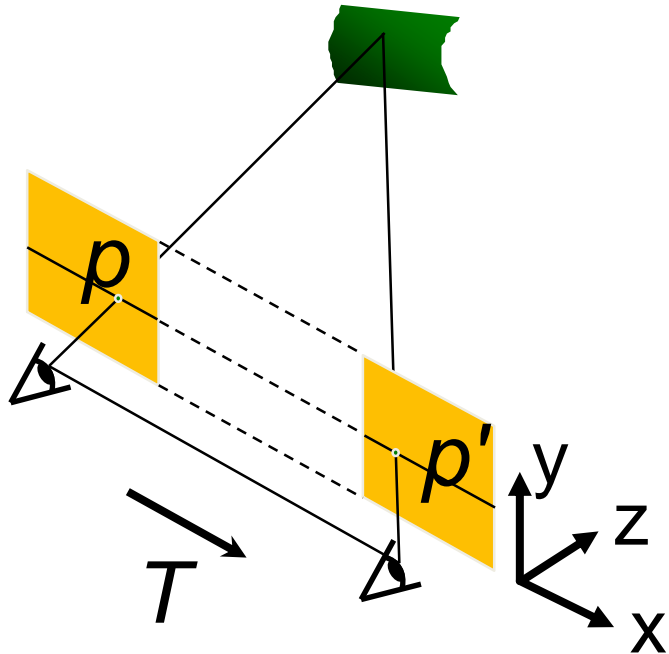
- Image planes of cameras are parallel to each other and to the baseline
- Camera centers are at same height
- Focal lengths the same

Simplest Case: Parallel images



- Image planes of cameras are parallel to each other and to the baseline
- Camera centers are at same height
- Focal lengths the same
- Then epipolar lines fall along the horizontal scan lines of the images

Essential matrix for parallel images



$$pEp' = 0 \quad E = [t_x]R$$

What's R? What's t?

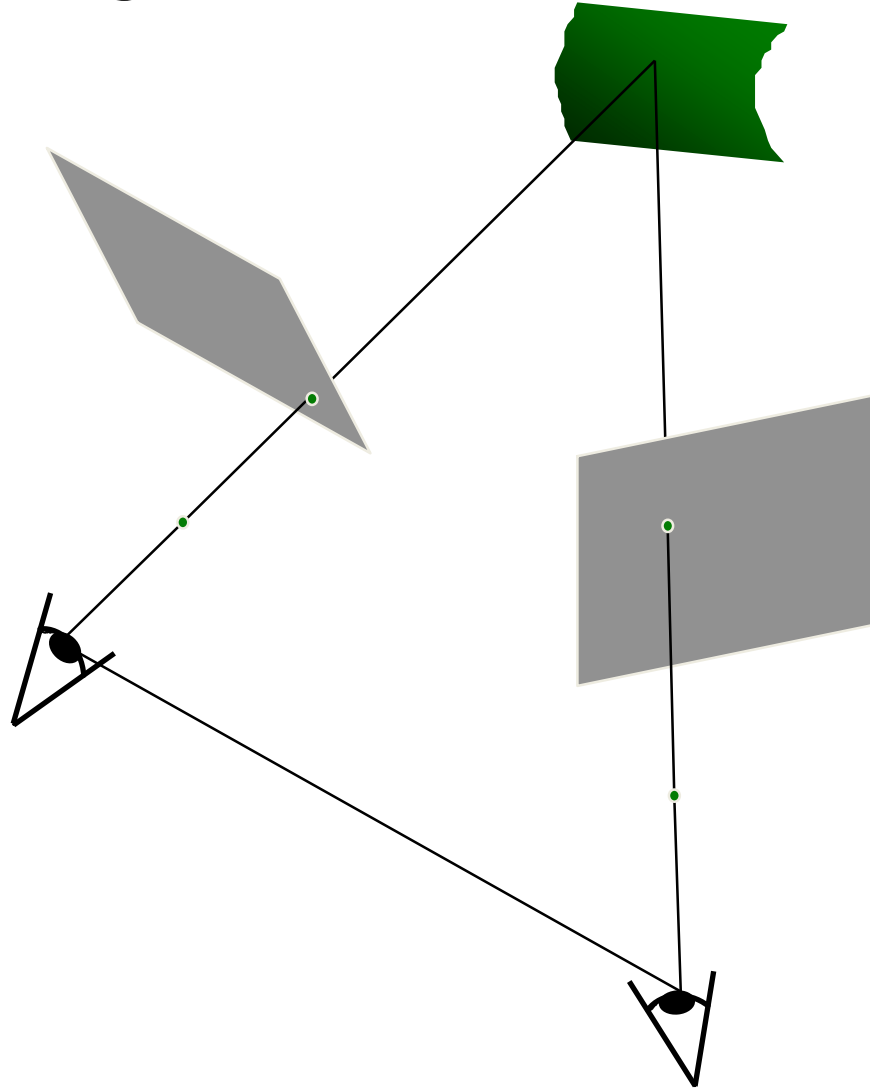
$$R = I \quad t = [T, 0, 0]$$

$$E = [t_x]R = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -T \\ 0 & T & 0 \end{bmatrix}$$

$$[u \ v \ 1] \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -T \\ 0 & T & 0 \end{bmatrix} \begin{bmatrix} u' \\ v' \\ 1 \end{bmatrix} = 0 \quad \rightarrow \quad [u \ v \ 1] \begin{bmatrix} 0 \\ -T \\ T'v \end{bmatrix} = 0 \quad \rightarrow \quad \begin{matrix} -Tv + T'v = 0 \\ T'v = Tv \end{matrix}$$

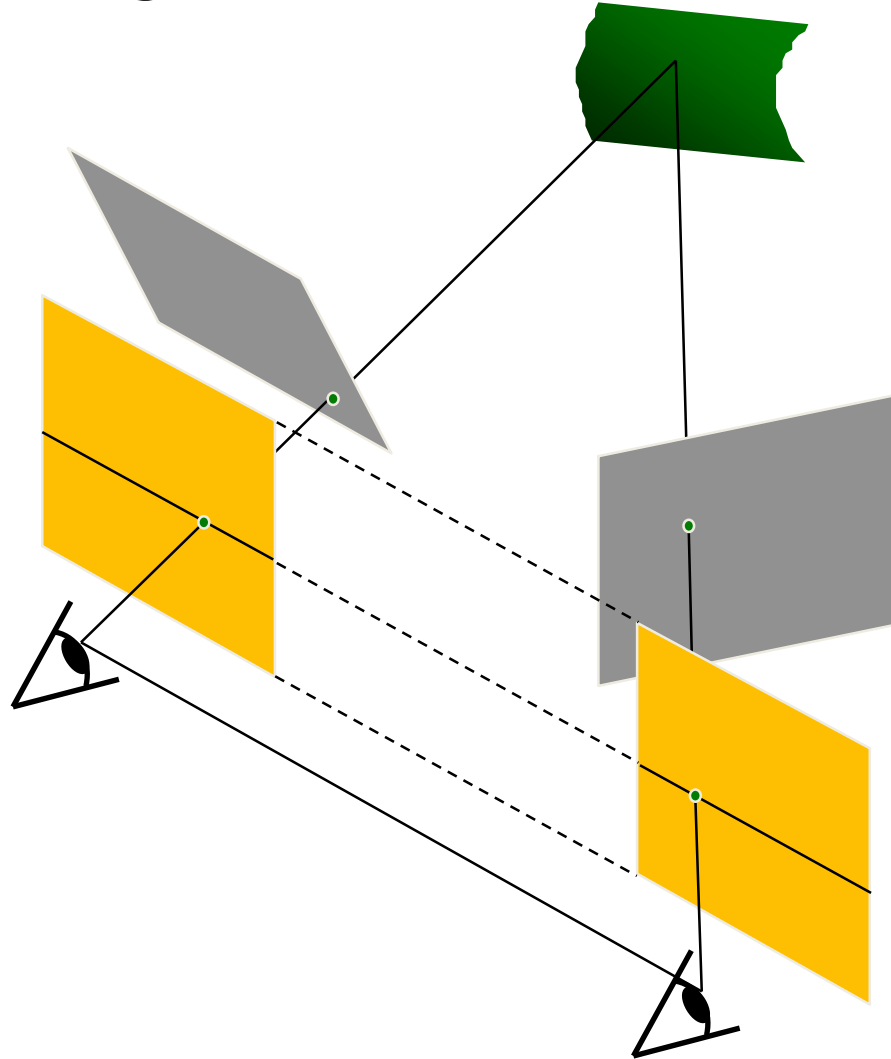
The y-coordinates of corresponding points are the same!

Stereo image rectification



Stereo image rectification

Reproject image planes onto a common plane parallel to the line between optical centers



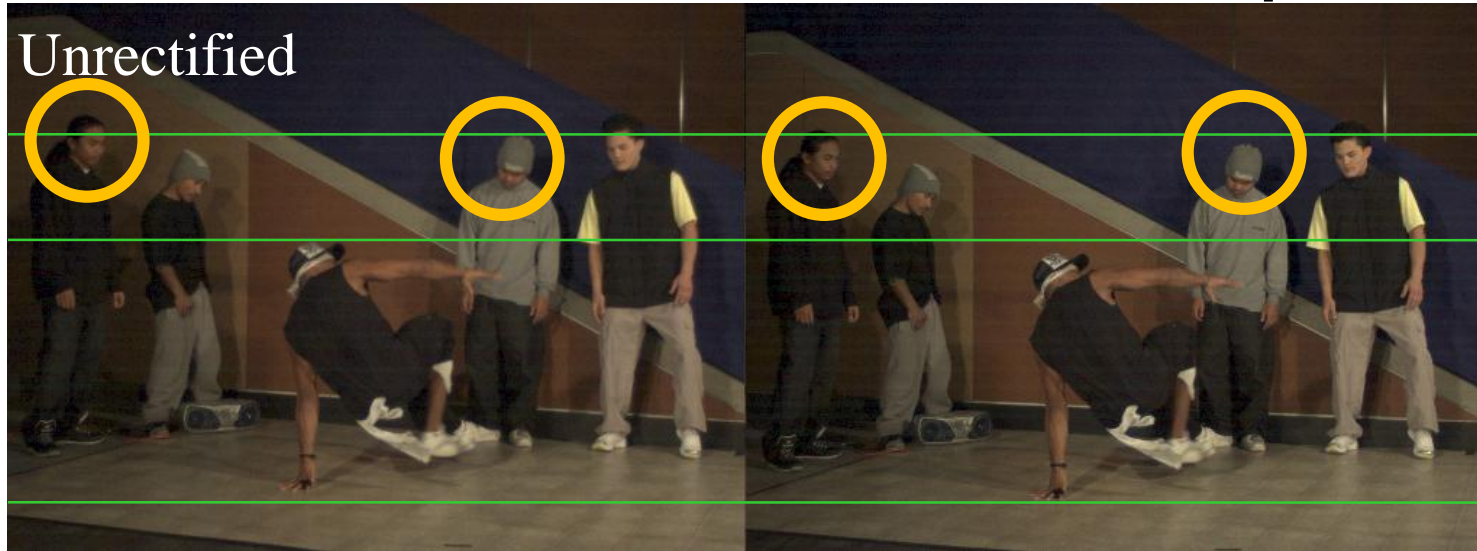
C. Loop and Z. Zhang. [Computing Rectifying Homographies for Stereo Vision](#). CVPR 1999

Slide credit: S. Lazebnik

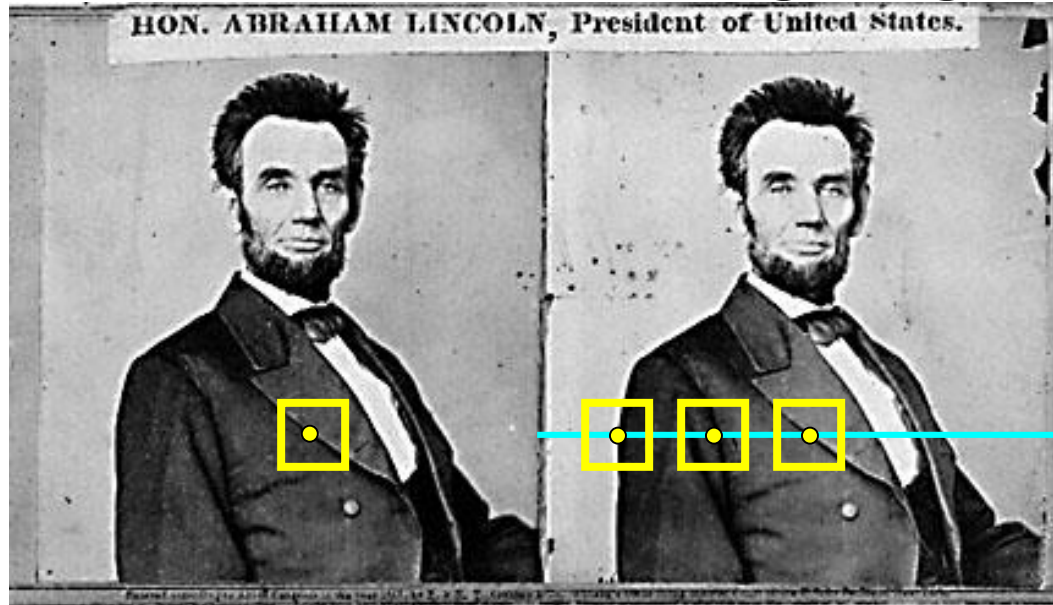
Rectification example



Another rectification example

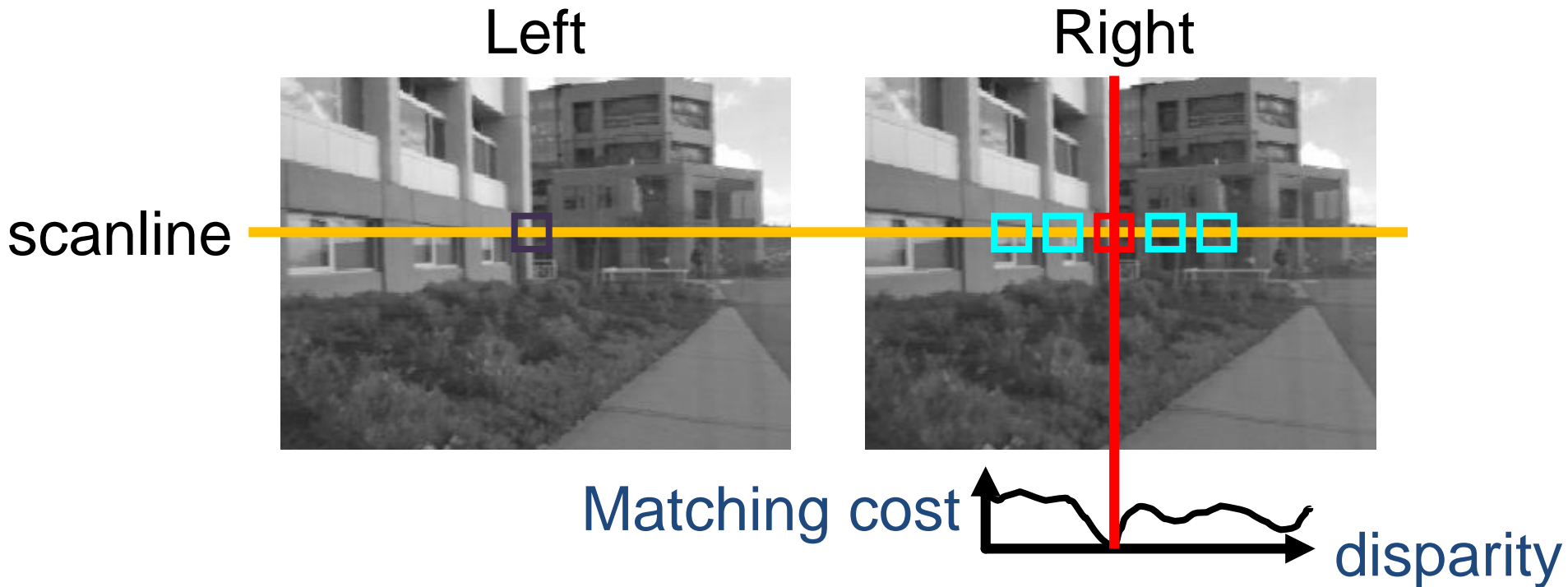


Basic stereo matching algorithm



- If necessary, rectify the two stereo images to transform epipolar lines into scanlines
- For each pixel in the first image
 - Find corresponding epipolar line in the right image
 - Examine all pixels on the epipolar line and pick the best match

Correspondence Search



Slide window along the right scanline, compare contents of that window with reference window on left

Matching cost: SSD or normalized correlation

Correspondence Search

Left

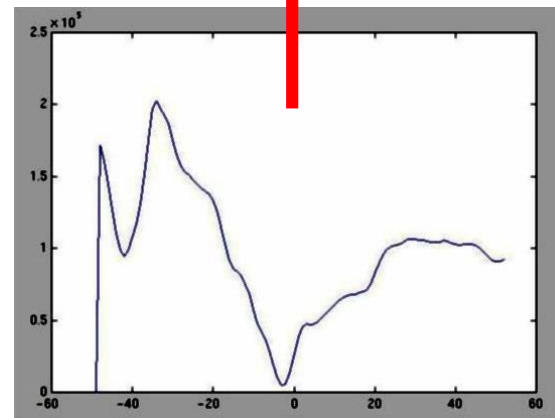
Right

scanline



Matching cost
Sum of squared differences

$$\sum_i (l_i - r_i)^2$$



Disparity

Correspondence Search

Left

Right

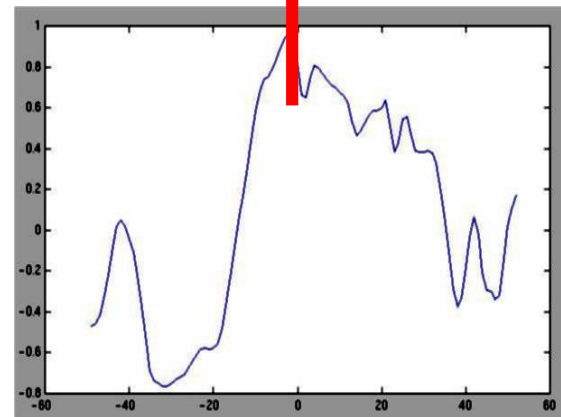
scanline



Matching cost
Normalized correlation

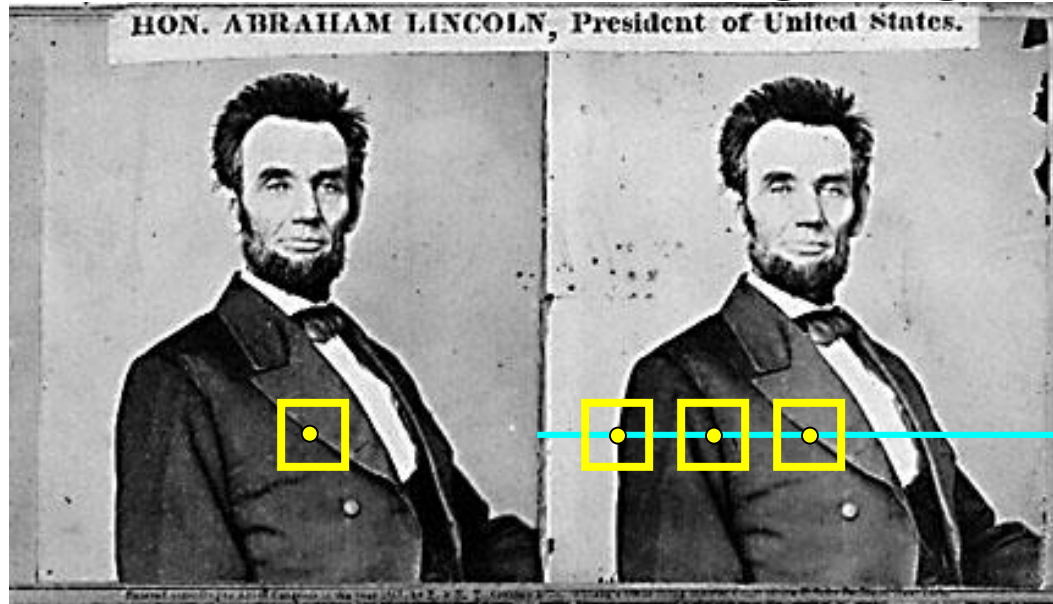
$$\hat{x}_i = \frac{x_i - \text{mean}(x)}{\text{std}(x)}$$

$$\hat{l} \cdot \hat{r}$$



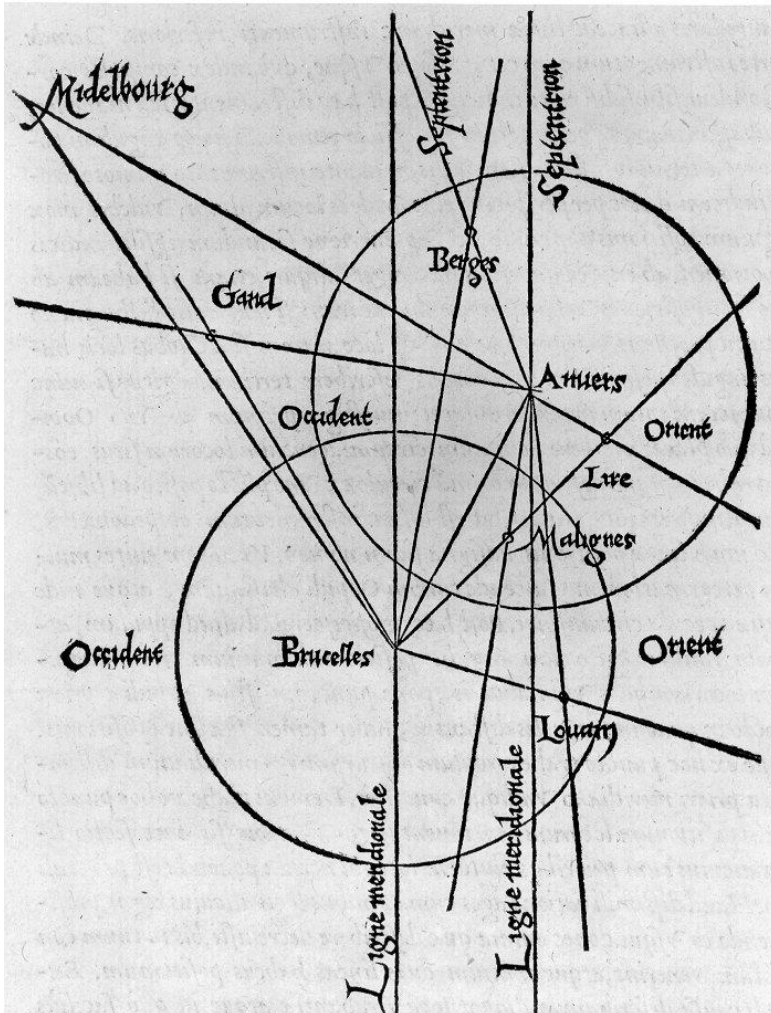
Disparity

Basic stereo matching algorithm



- If necessary, rectify the two stereo images to transform epipolar lines into scanlines
- For each pixel x in the first image
 - Find corresponding epipolar scanline in the right image
 - Examine all pixels on the scanline and pick the best match x'
 - Triangulate the matches to get depth information

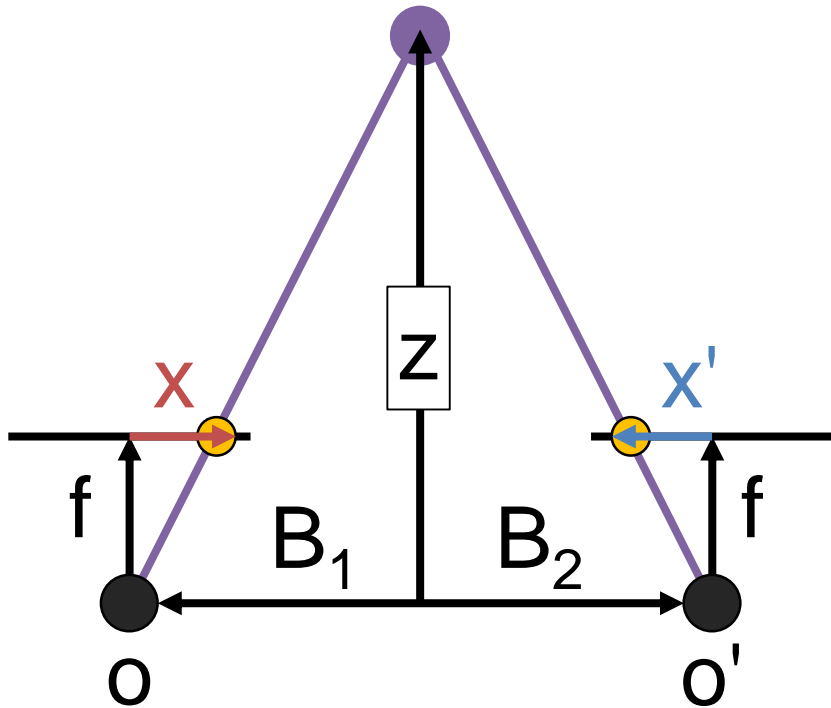
Triangulation: Older History



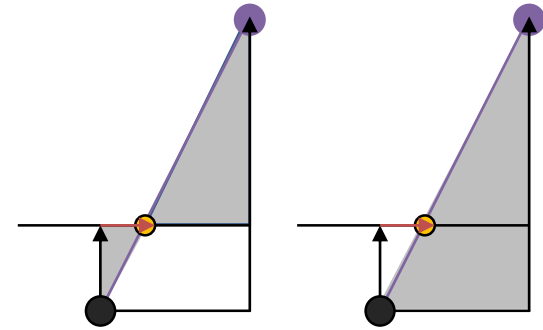
- From [Wikipedia](#): Gemma Frisius's 1533 diagram introducing the idea of triangulation into the science of surveying. Having established a baseline, e.g. the cities of Brussels and Antwerp, the location of other cities, e.g. Middelburg, Ghent etc., can be found by taking a compass direction from each end of the baseline, and plotting where the two directions cross. This was only a theoretical presentation of the concept — due to topographical restrictions, it is impossible to see Middelburg from either Brussels or Antwerp. Nevertheless, the figure soon became well known all across Europe.

Triangulation: Modern History

Depth from disparity



$$\frac{x}{f} = \frac{B_1}{z}$$

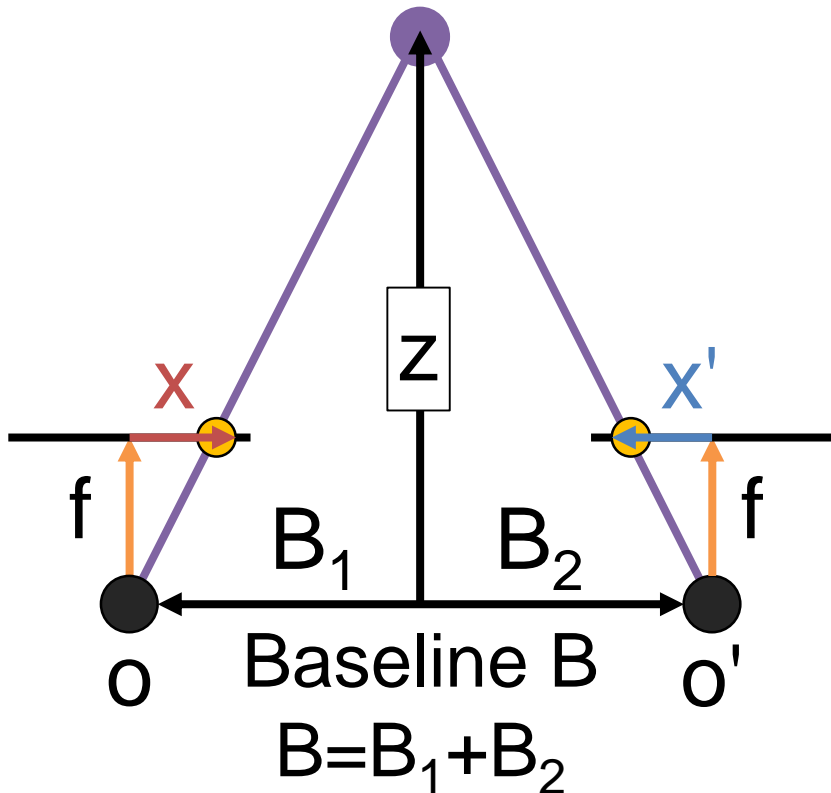


By similar triangles

$$\frac{-x'}{f} = \frac{B_2}{z}$$

Similarly by
similar triangles

Depth from disparity



$$\frac{x}{f} = \frac{B_1}{z} \quad \frac{-x'}{f} = \frac{B_2}{z}$$

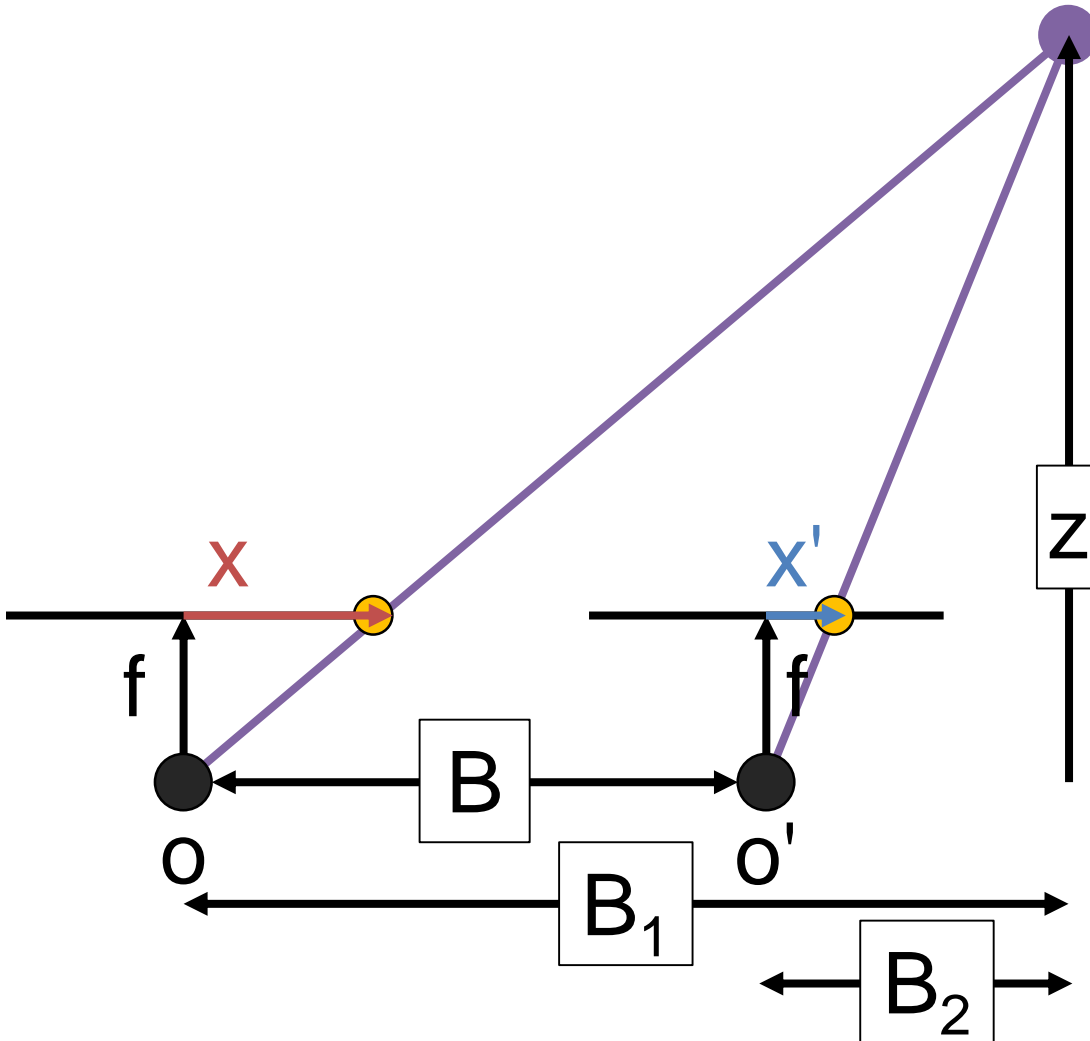
Add them

$$\frac{x - x'}{f} = \frac{B_1 + B_2}{z}$$

$$\underbrace{x - x'}_{\text{Disparity}} = \frac{fB}{z}$$

Disparity

Depth from disparity



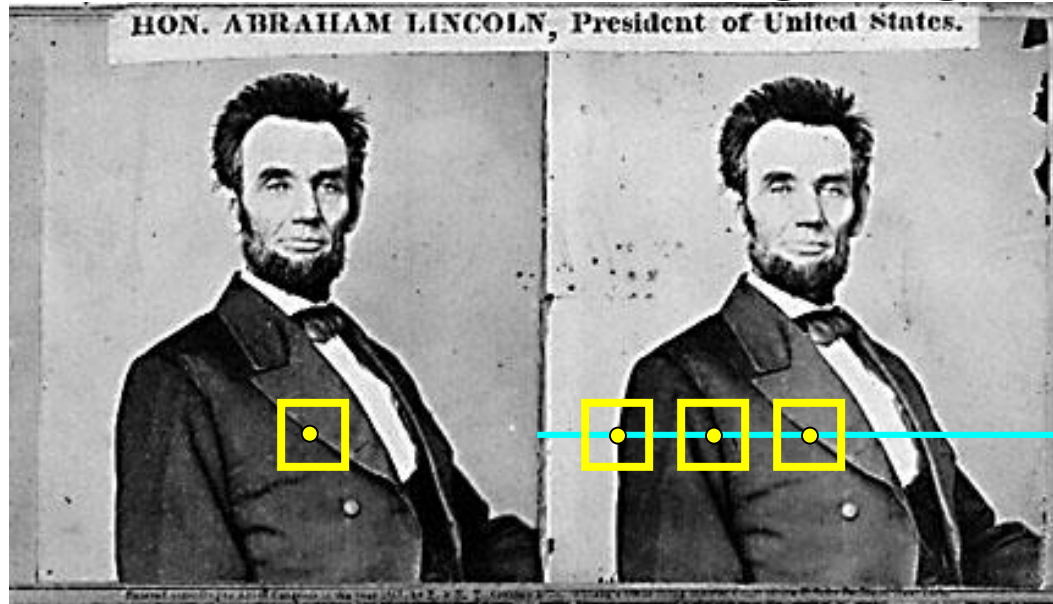
$$\frac{x}{f} = \frac{B_1}{z} \quad \frac{x'}{f} = \frac{B_2}{z}$$

Subtract them

$$\frac{x - x'}{f} = \frac{B_1 - B_2}{z}$$

$$x - x' = \frac{fB}{z}$$

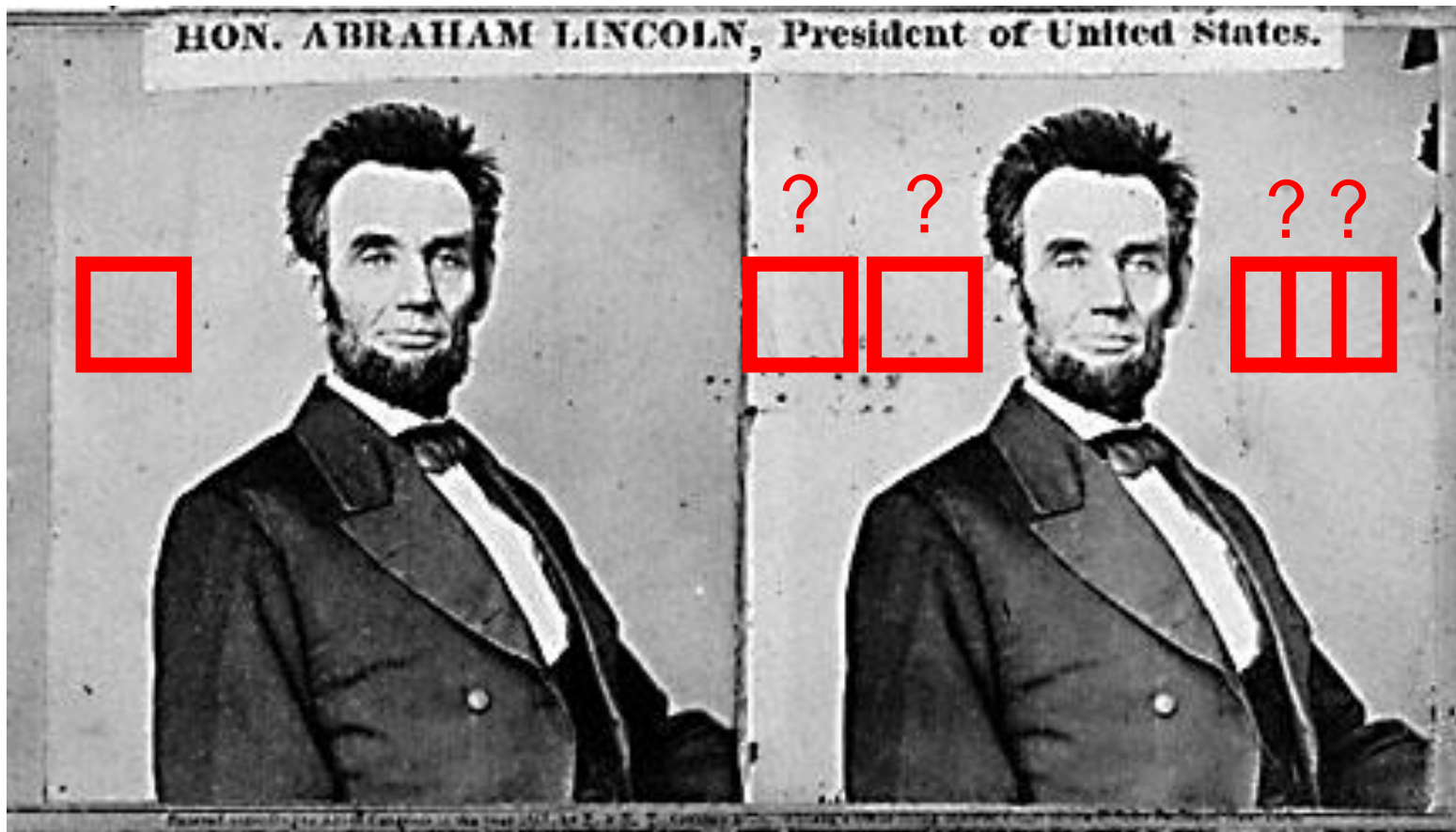
Basic stereo matching algorithm



- If necessary, rectify the two stereo images to transform epipolar lines into scanlines
- For each pixel x in the first image
 - Find corresponding epipolar scanline in the right image
 - Examine all pixels on the scanline and pick the best match x'
 - Compute disparity $x-x'$ and set $\text{depth}(x) = B * f / (x-x')$

Failures of Correspondence Search

Textureless regions. **Why?**



Failures of Correspondence Search

Repeated Patterns. **Why?**

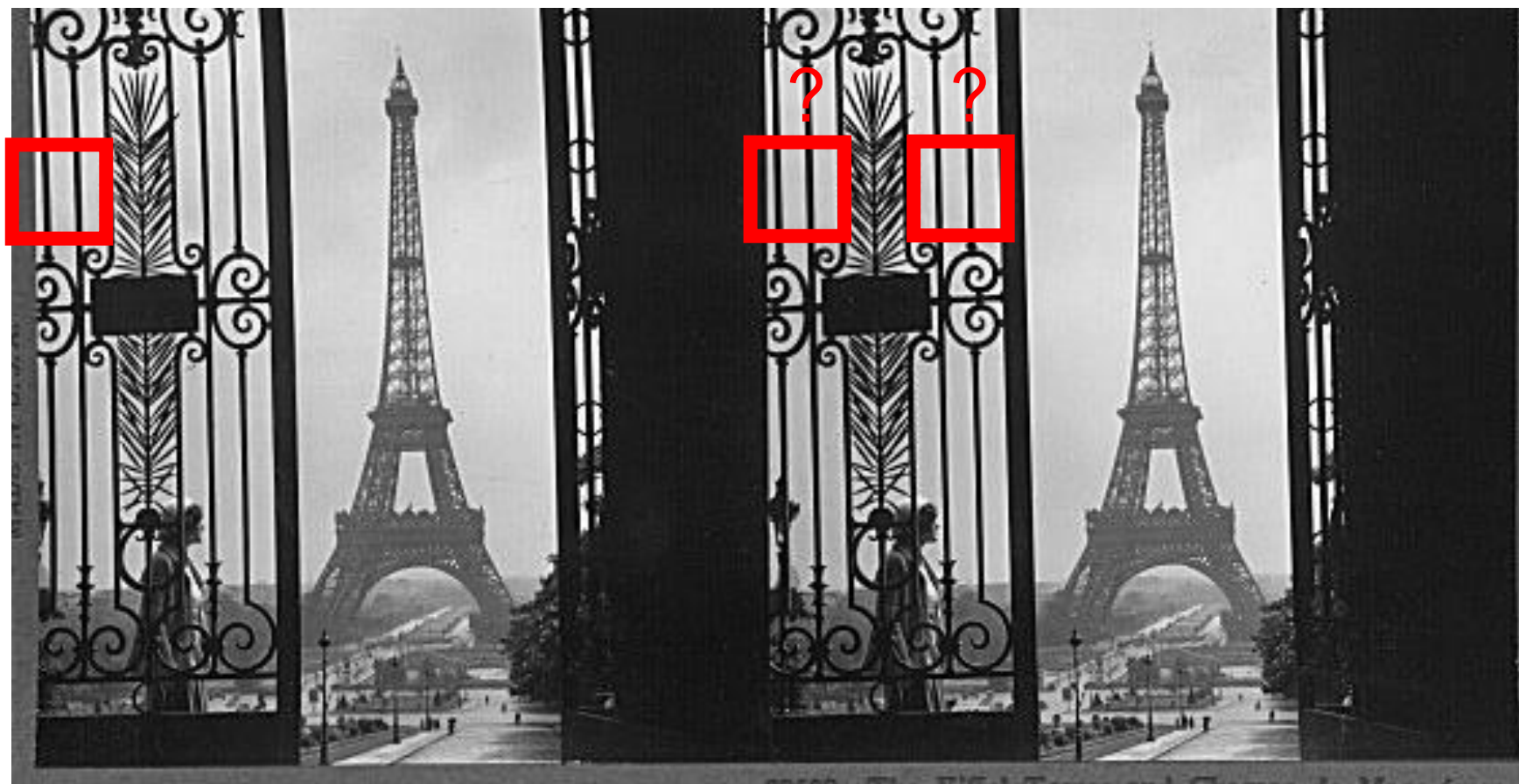


Image credit: S. Lazebnik

Failures of Correspondence Search

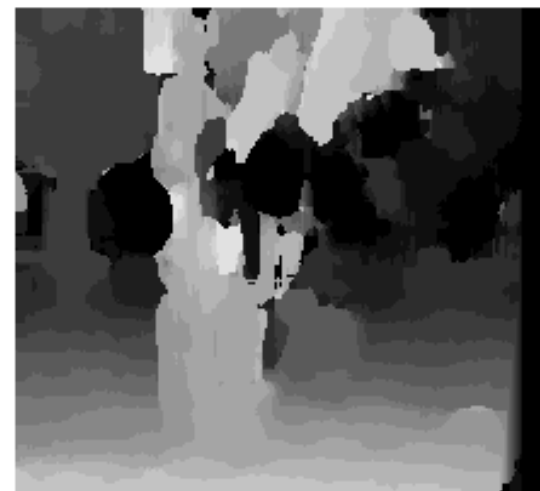
Specular Surfaces. **Why?**



Effect of window size



$W = 3$



$W = 20$

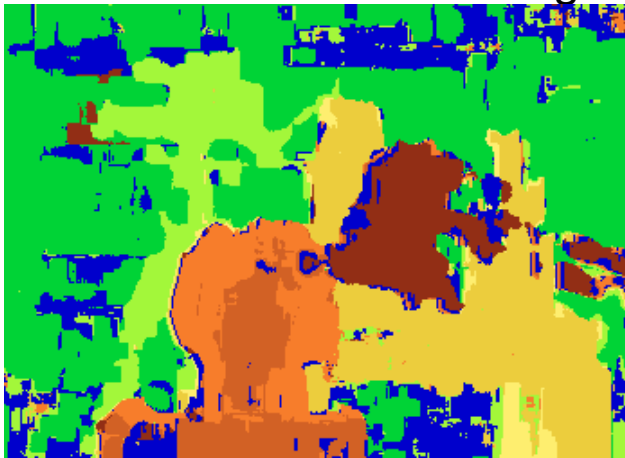
- Smaller window
 - + More detail
 - More noise
- Larger window
 - + Smoother disparity maps
 - Less detail

Results with window search

Data



Window-based matching



Ground truth



Better methods exist...



Graph cuts



Ground truth

Y. Boykov, O. Veksler, and R. Zabih, [Fast Approximate Energy Minimization via Graph Cuts](#), PAMI 2001

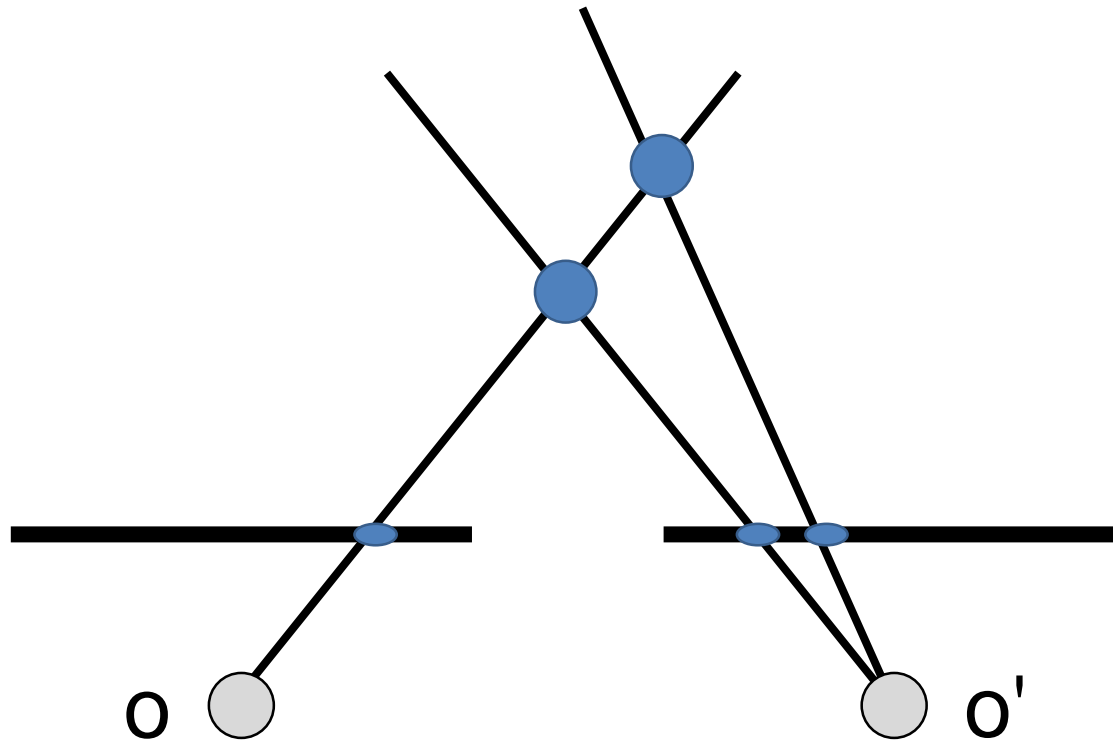
For the latest and greatest: <http://www.middlebury.edu/stereo/>

Improving Window-based Matching

- Similarity is **local** (each window independent)
- Need non-local correspondence constraints / cues.

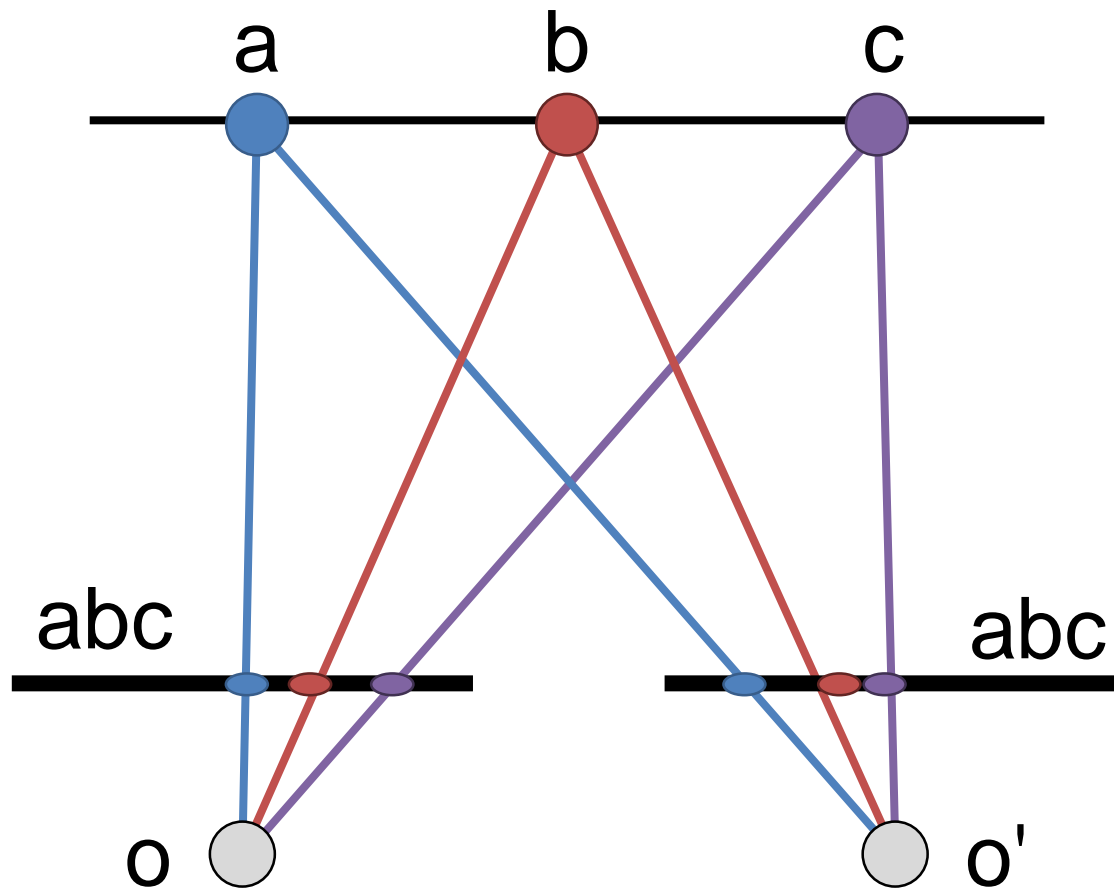
Uniqueness

- Each point in one image should match at most one point in other image.
- **When might this not be true?**



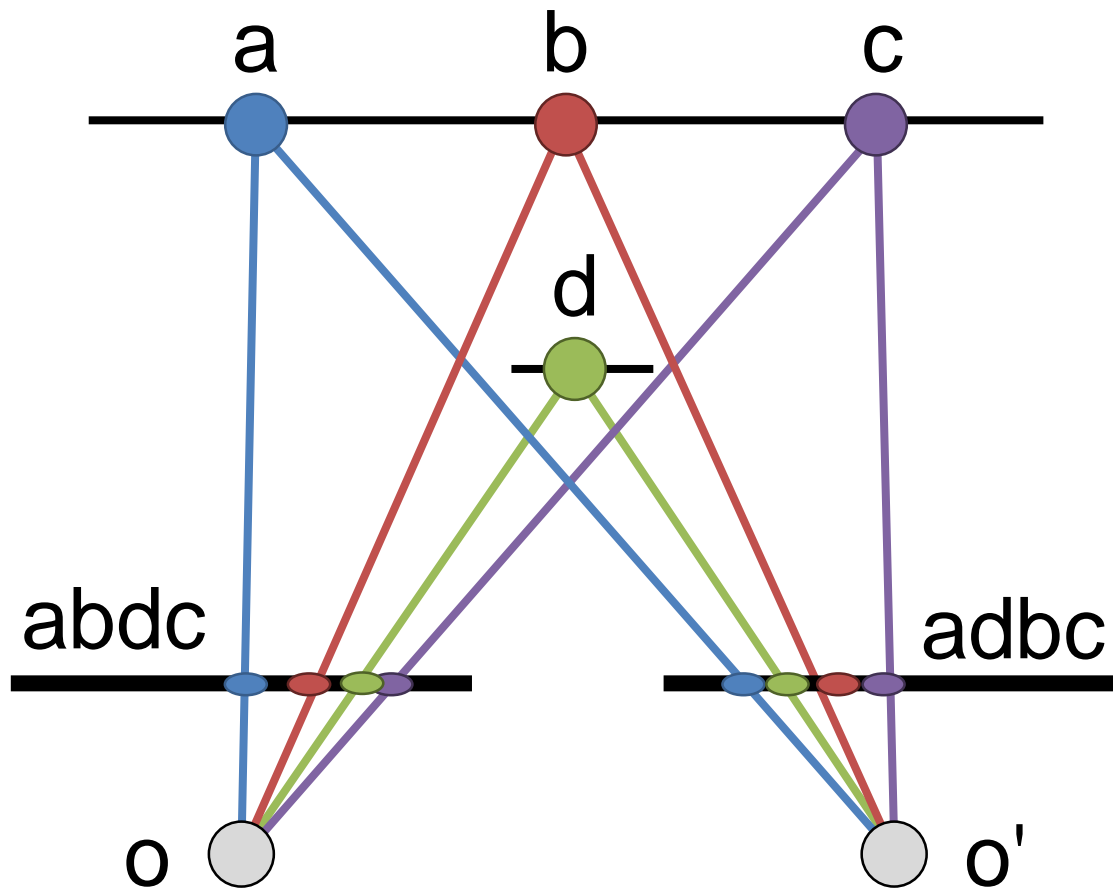
Ordering

- Corresponding points should be in same order



Ordering

- Not always true!

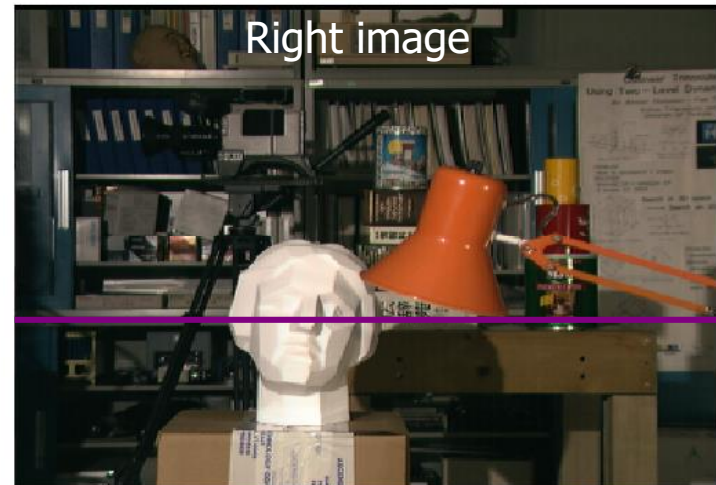
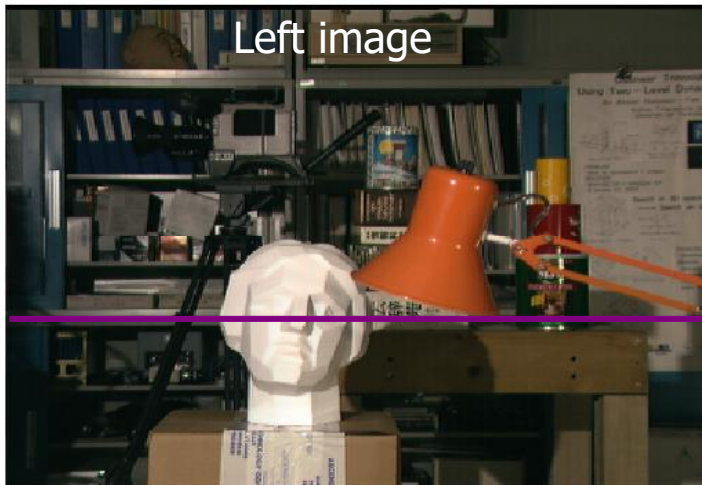


Smoothness

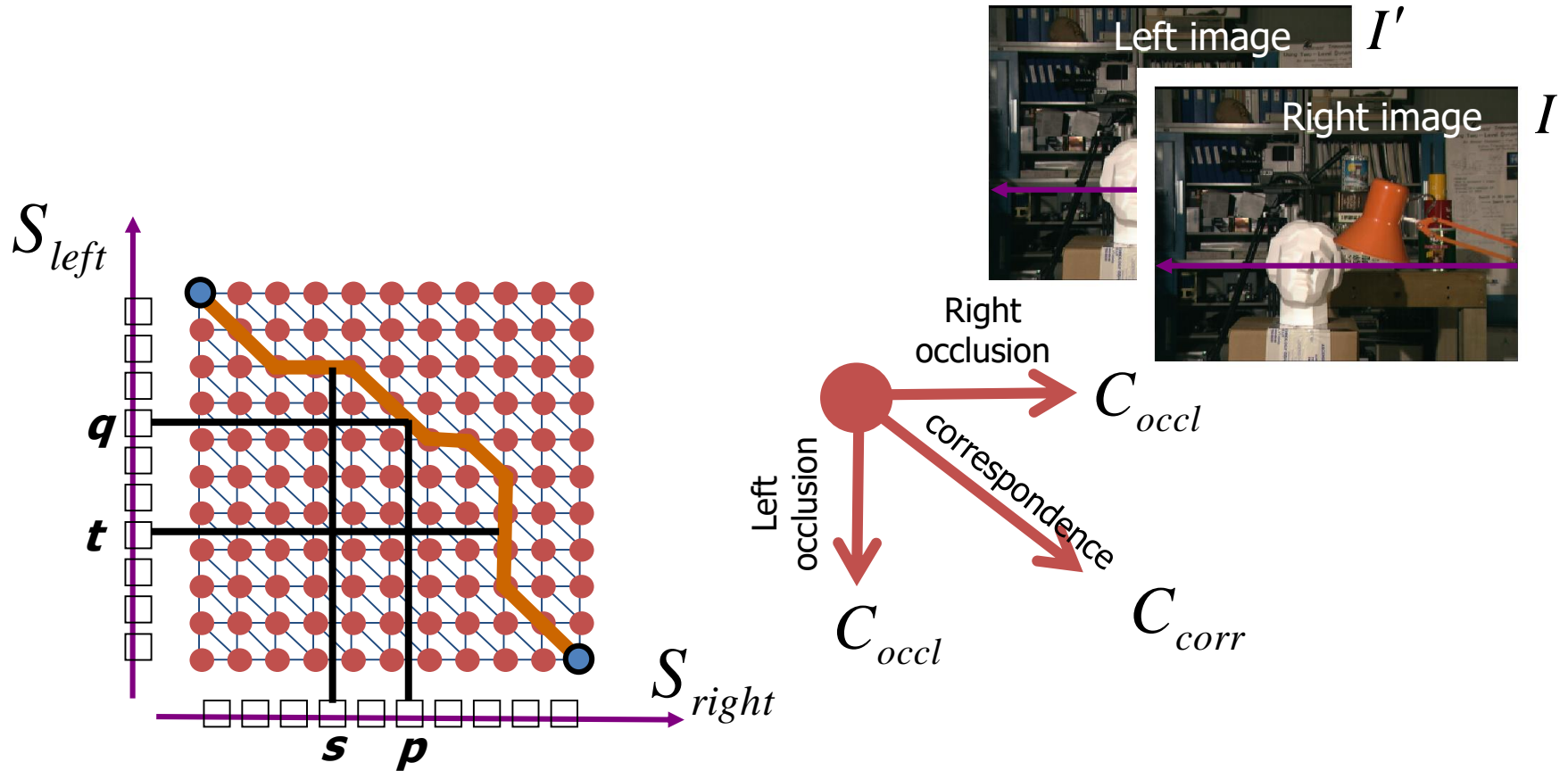
- We expect disparity values to change slowly (for the most part)
- **When is this not true?**

Scanline Stereo

- Try to coherently match pixels on the entire scanline
- Different scanlines are optimized (by dynamic programming) independently



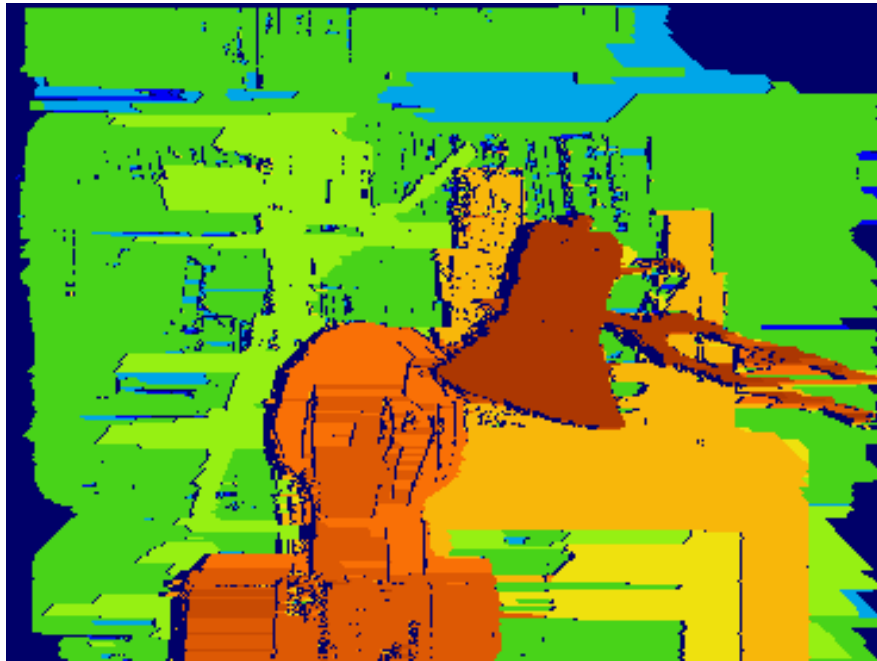
“Shortest paths” for scan-line stereo



Can be implemented with dynamic programming
Ohta & Kanade '85, Cox et al. '96

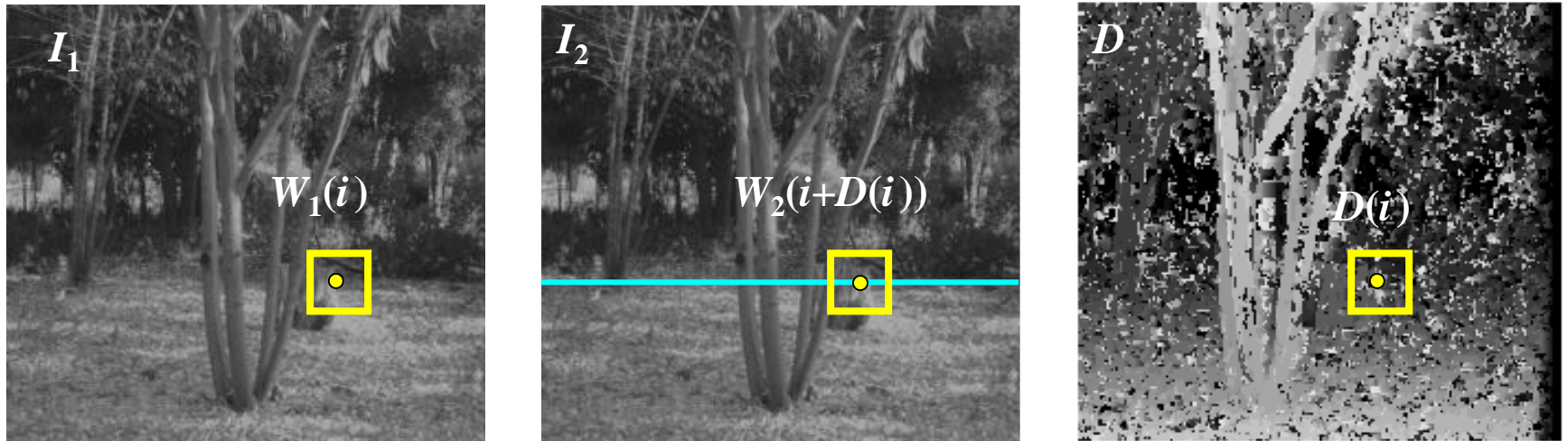
Coherent Stereo on 2D Grid

- Scanline stereo generates streaking artifacts



- Can't use dynamic programming to find spatially coherent disparities on a 2D grid

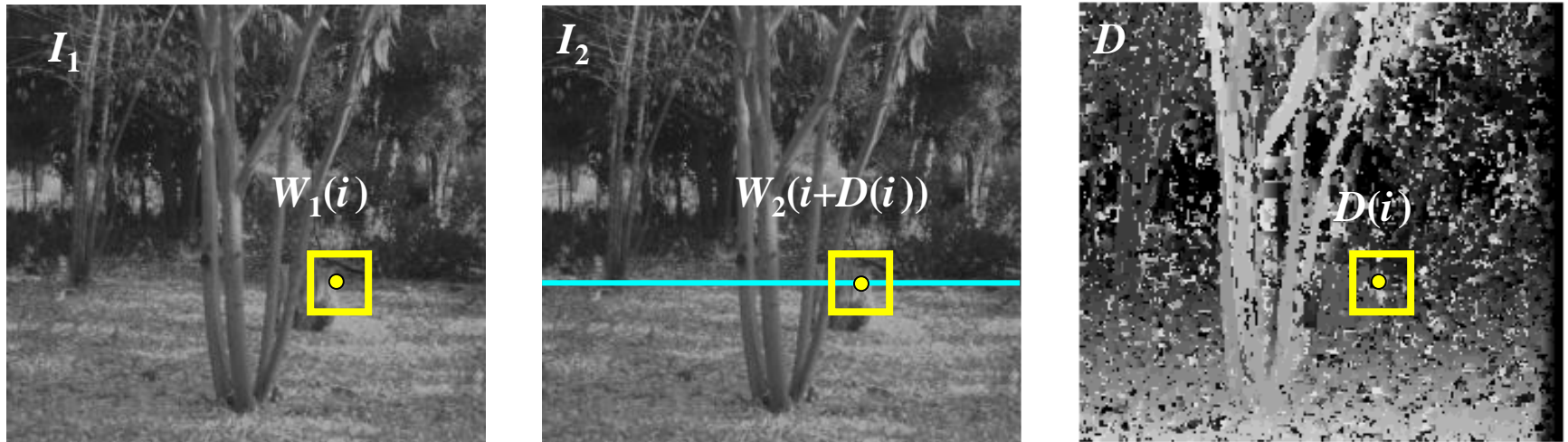
Stereo Matching as Optimization



$$E(D) = \underbrace{\sum_i \left(W_1(i) - W_2(i + D(i)) \right)^2}_{\text{Data term}} + \lambda \underbrace{\sum_{\text{neighbors } i,j} \rho(D(i) - D(j))}_{\text{Smoothness term}}$$

Solvable by graph cuts for certain smoothnesses ρ

Is This Doable by Deep Network?

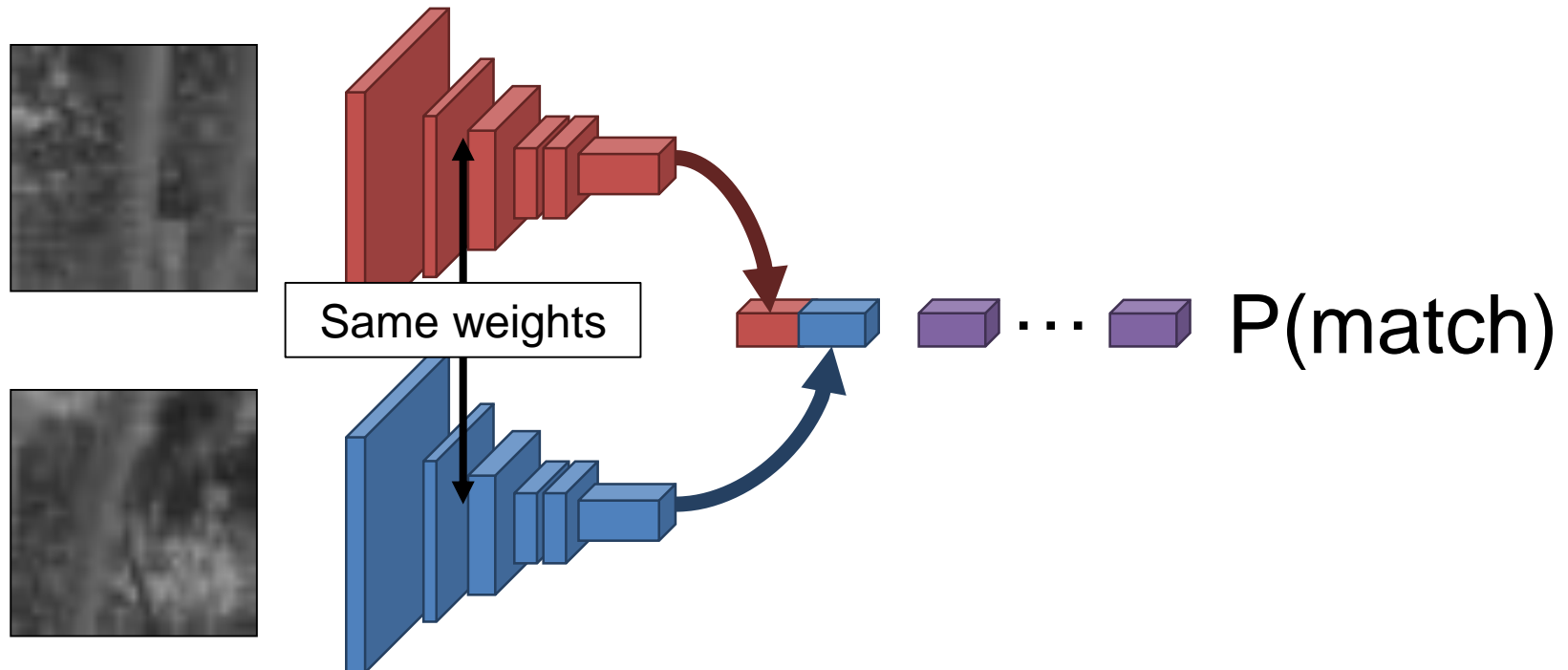


$$E(D) = \sum_i \left(W_1(i) - W_2(i + D(i)) \right)^2 + \lambda \underbrace{\sum_{\text{neighbors } i,j} \rho(D(i) - D(j))}_{\text{Smoothness term}}$$

Easy solution: replace the data term with a network

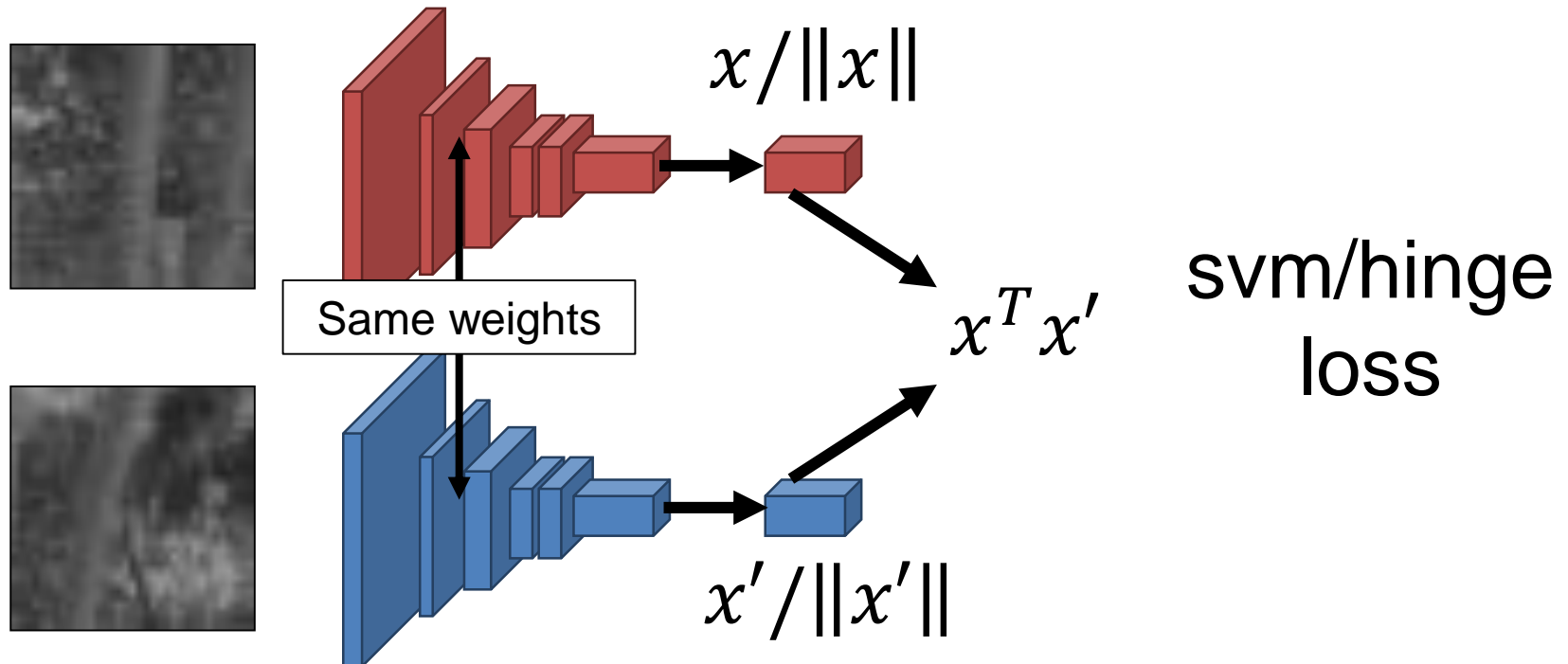
Deep Learning For Stereo

- Feed in two images to identical networks, concatenate outputs, learn multilayer perception
- Slow: **why?**



Deep Learning For Stereo

- Normalize outputs; treat dot product as prediction of match/no match
- Fast: **why?**

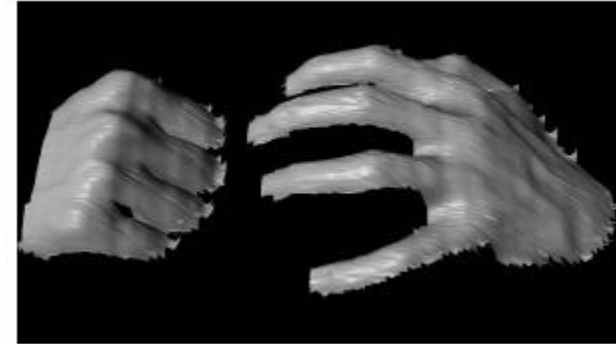
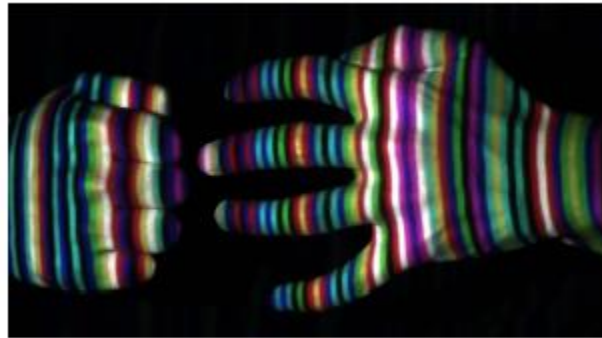


Stereo datasets

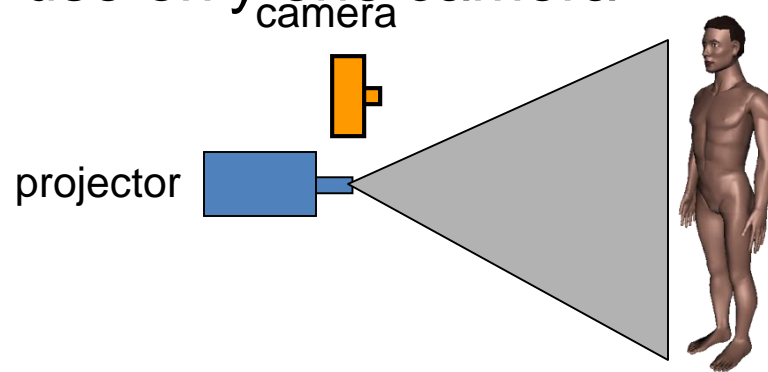
- [Middlebury stereo datasets](#)
- [KITTI](#)
- [Synthetic data?](#)



Active stereo with structured light



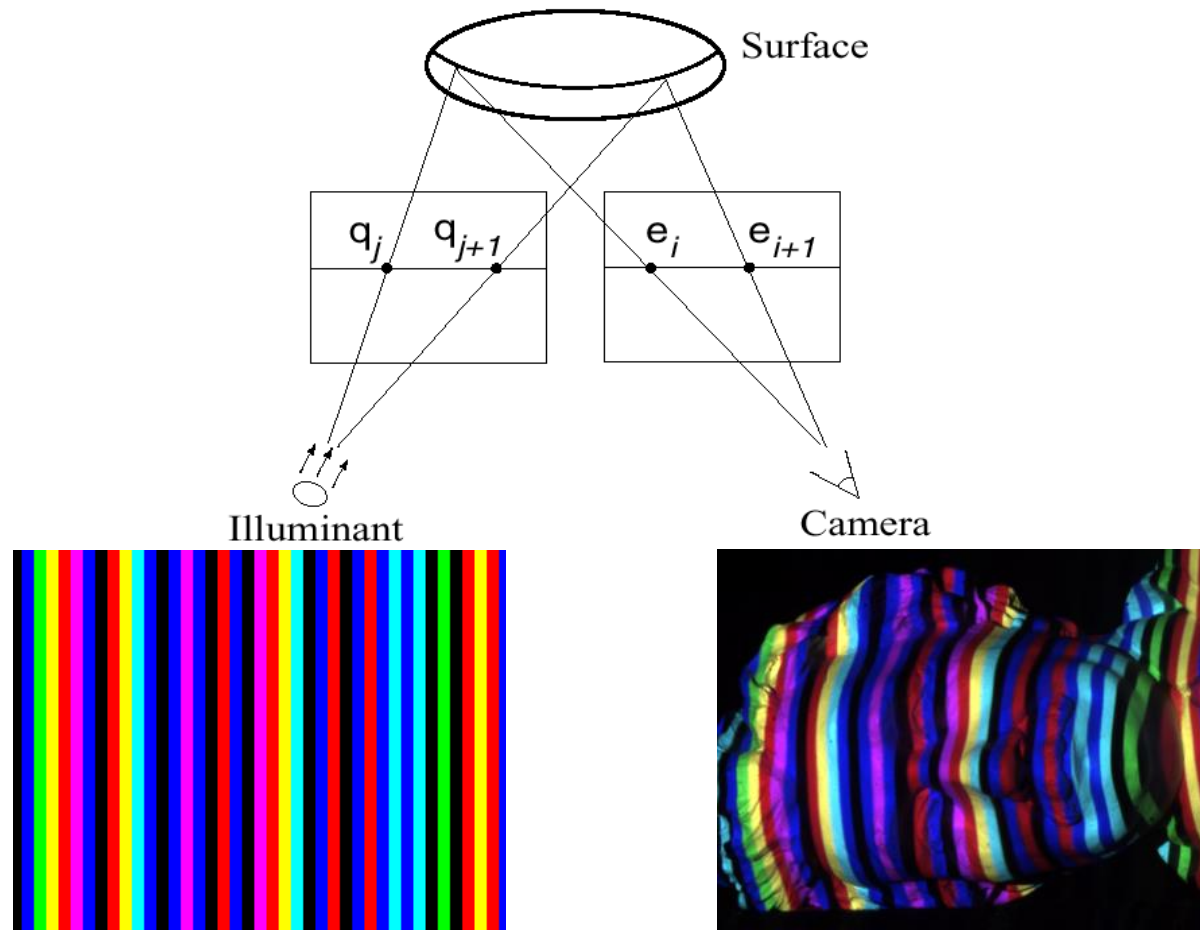
- Project “structured” light patterns onto the object
 - Simplifies the correspondence problem
 - Allows us to use only one camera



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

Slide credit:
S. Lazebnik

Active stereo with structured light



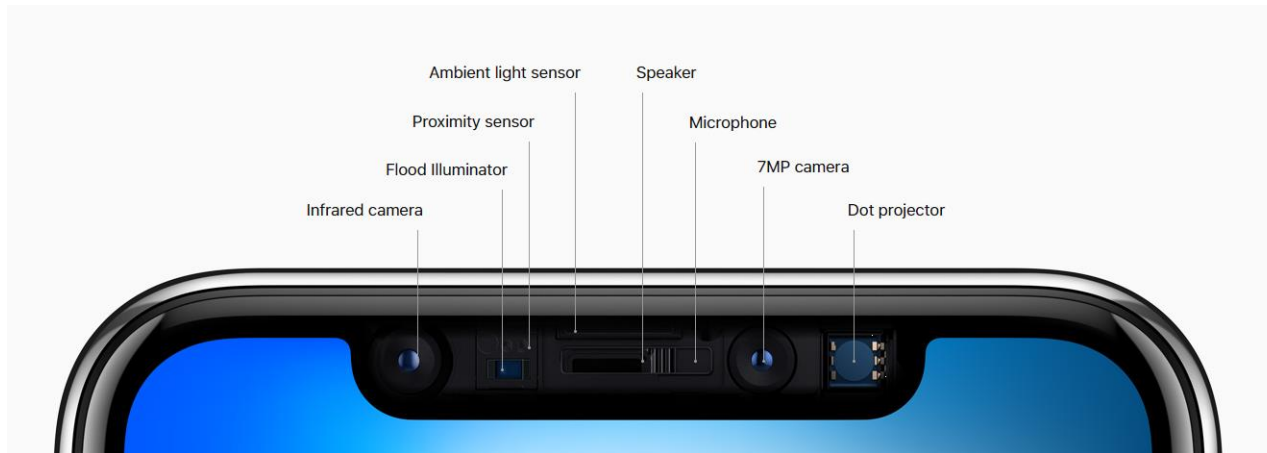
L. Zhang, B. Curless, and S. M. Seitz. [Rapid Shape Acquisition Using Color Structured Light and Multi-pass Dynamic Programming.](#) 3DPVT 2002

Slide credit:
S. Lazebnik

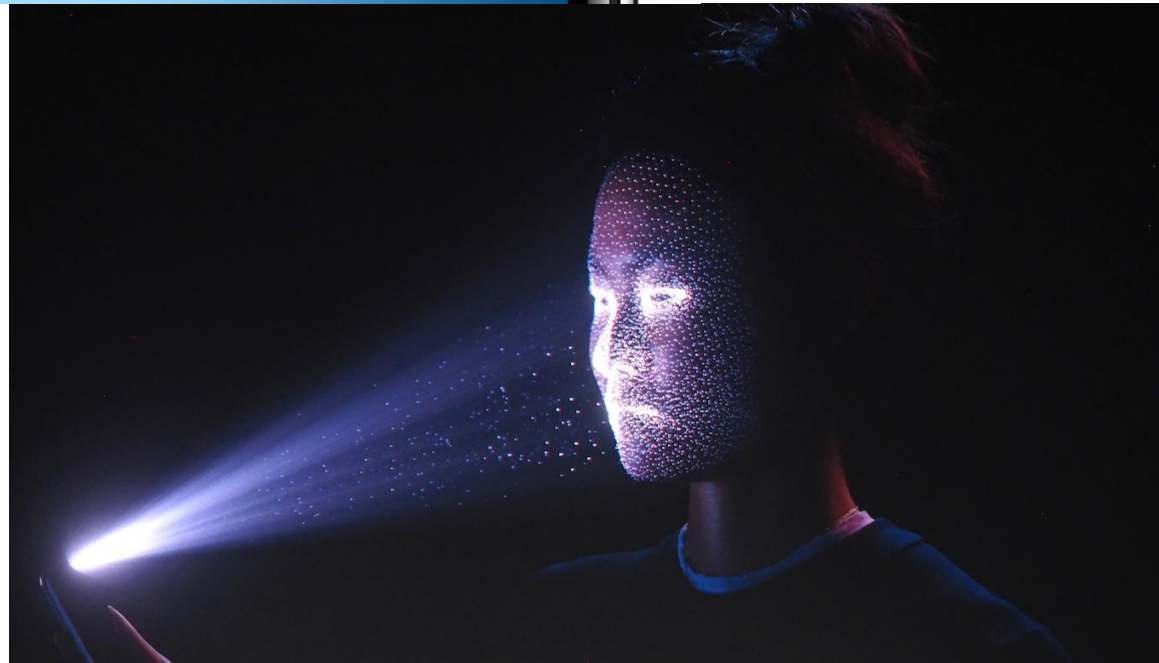
Kinect: Structured infrared light



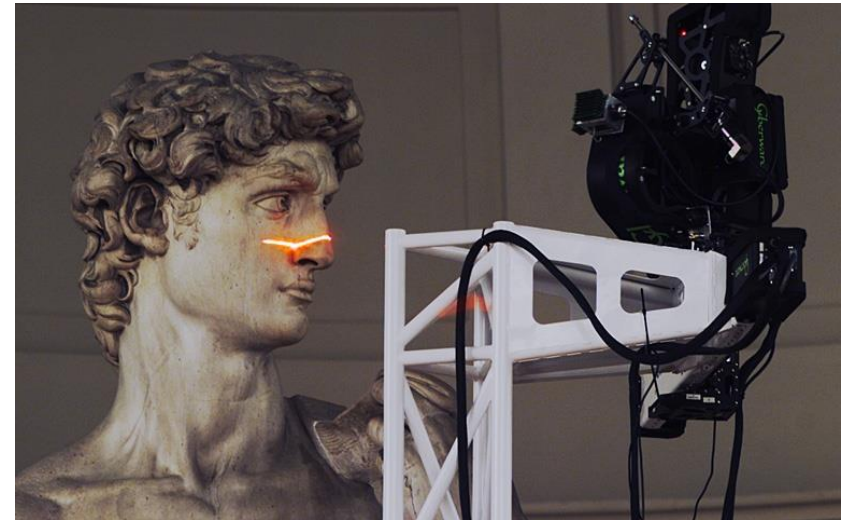
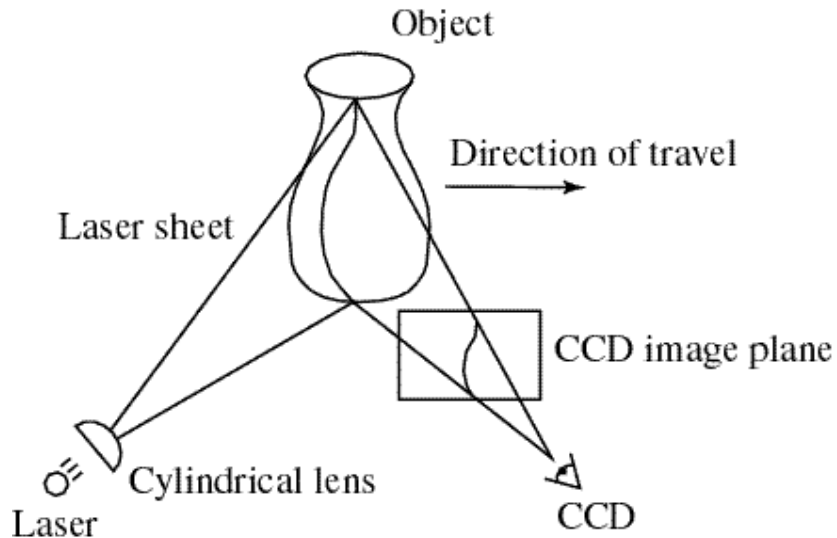
Apple TrueDepth



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



Laser scanning



Digital Michelangelo Project
Levoy et al.

<http://graphics.stanford.edu/projects/mich/>

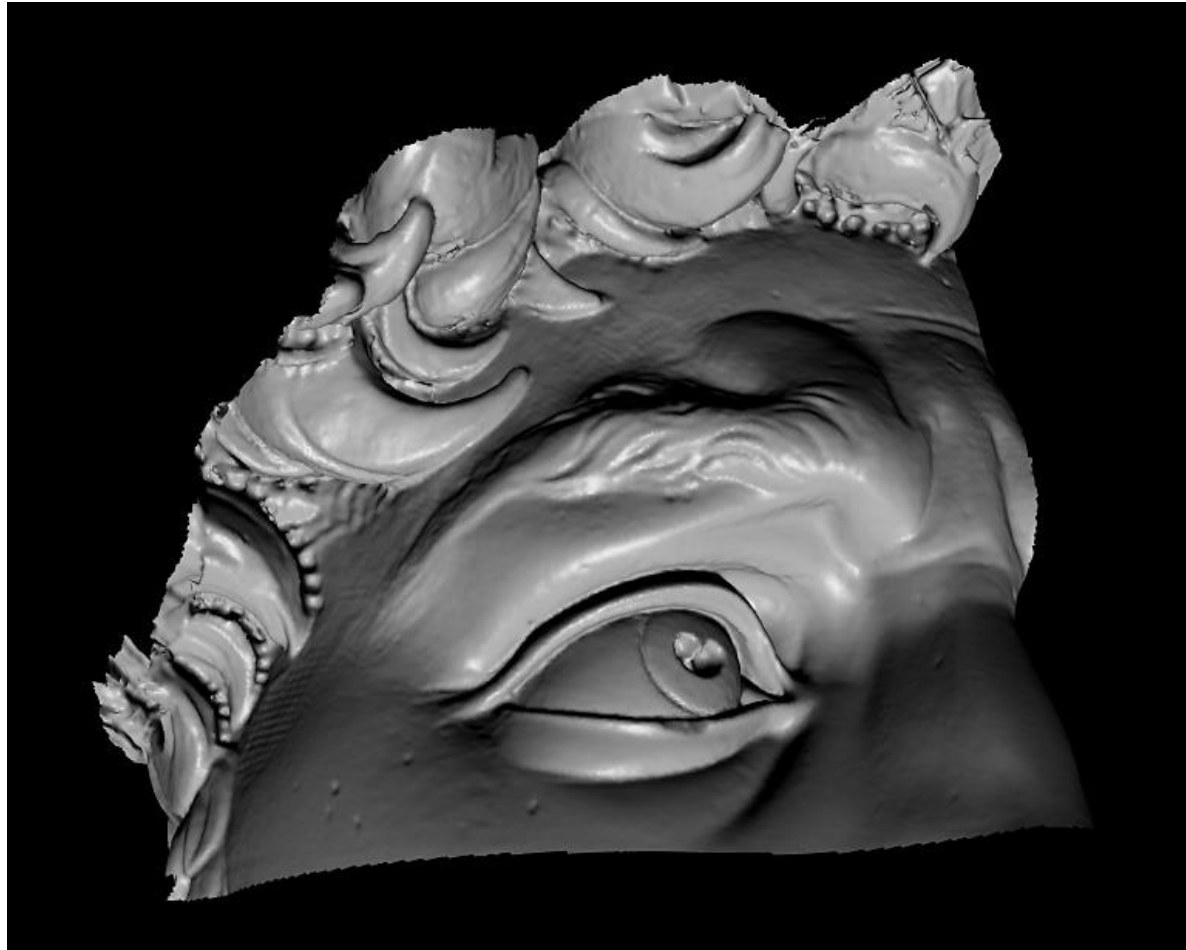
- Optical triangulation
 - Project a single stripe of laser light
 - Scan it across the surface of the object
 - This is a very precise version of structured light scanning

Laser scanned models



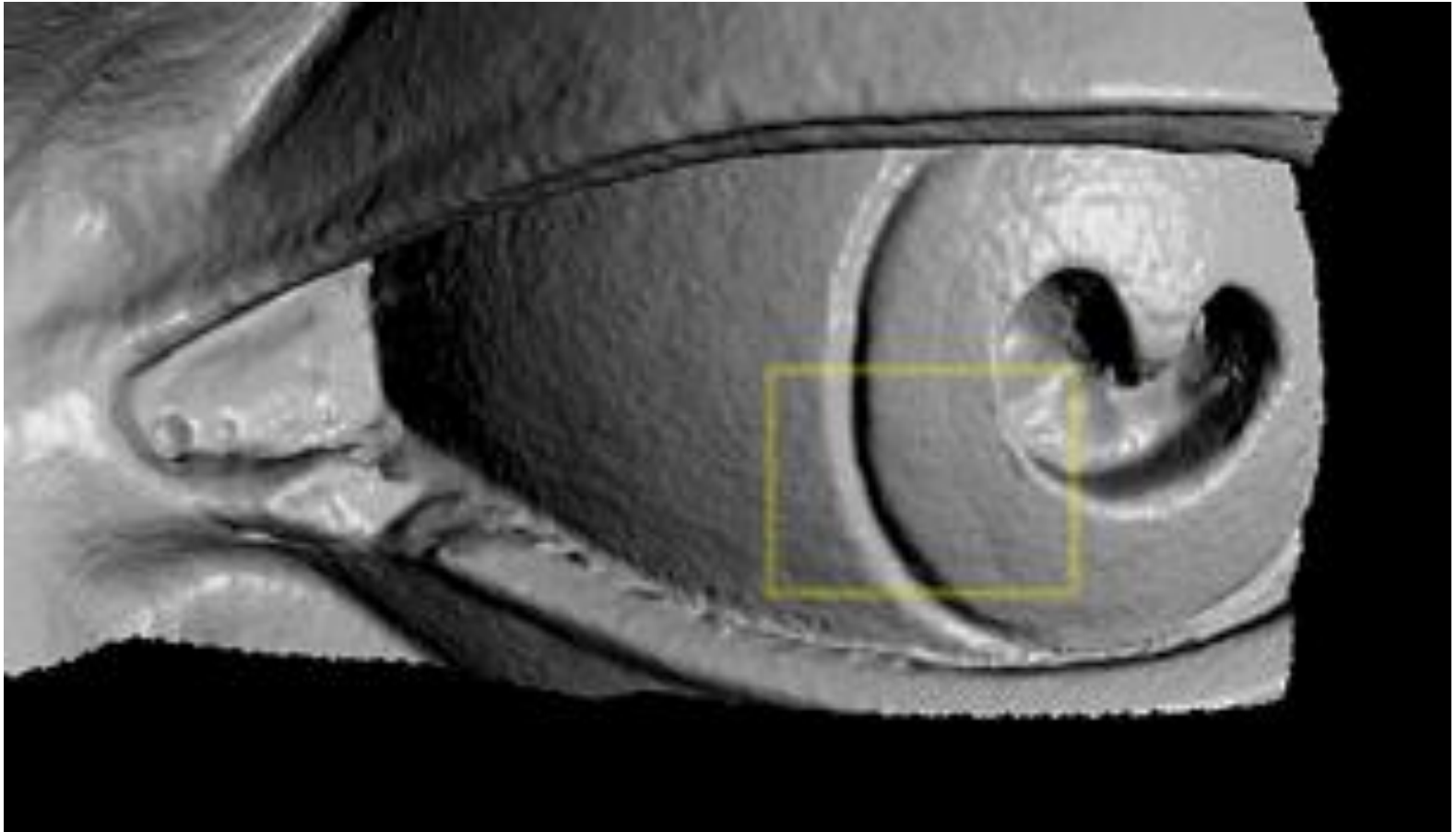
The Digital Michelangelo Project, Levoy et al.

Laser scanned models



The Digital Michelangelo Project, Levoy et al.

Laser scanned models

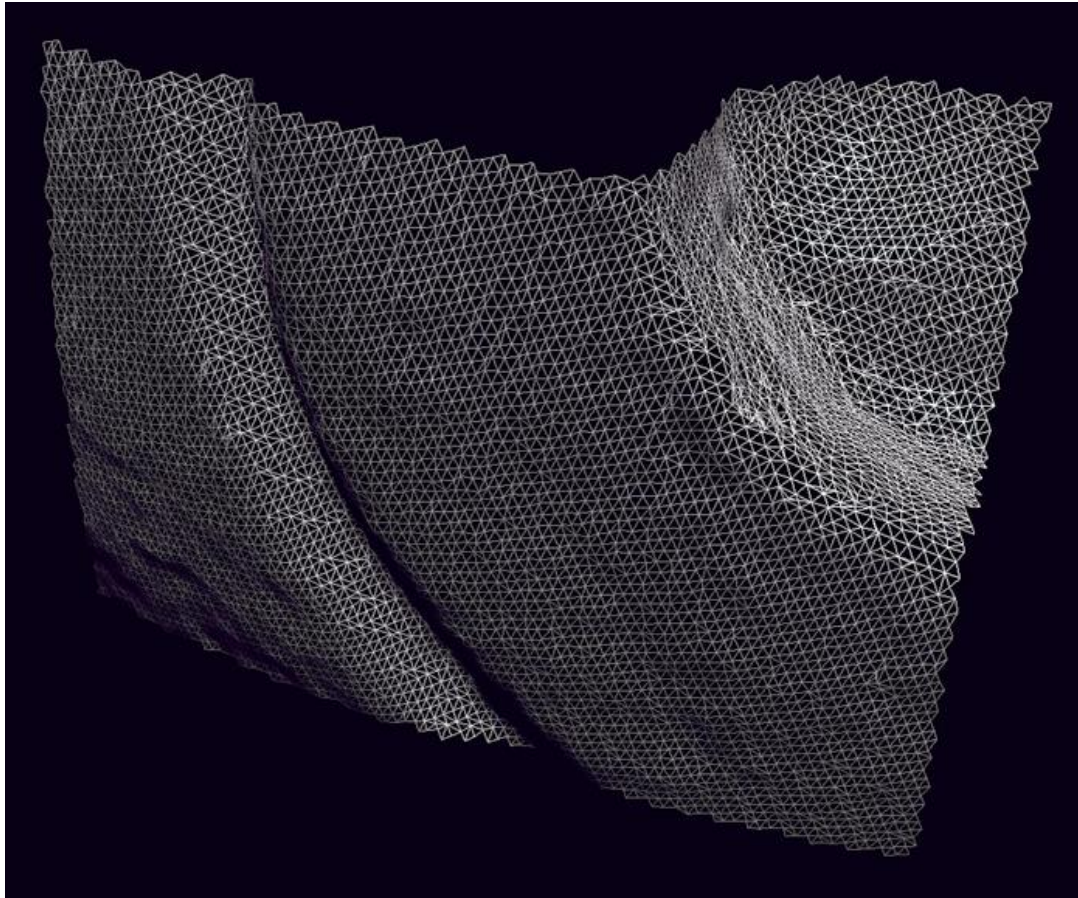


The Digital Michelangelo Project, Levoy et al.

Source: S. Seitz

Laser scanned models

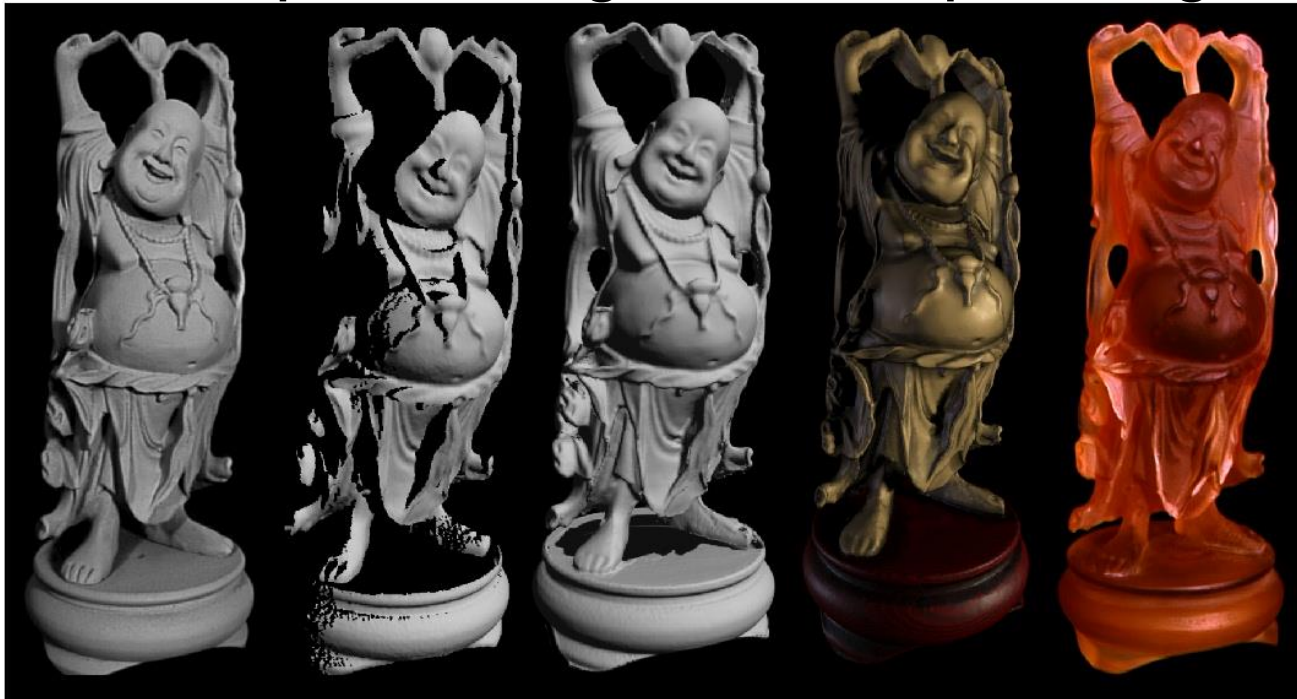
1.0 mm resolution (56 million triangles)



The Digital Michelangelo Project, Levoy et al.

Aligning range images

- One range scan not enough for complex surfaces
- Need techniques to register multiple range images



B. Curless and M. Levoy, [A Volumetric Method for Building Complex Models from Range Images](#), SIGGRAPH 1996