

Stereo



Many slides adapted from Steve Seitz

Binocular stereo

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



Where does the depth information come from?

Binocular stereo

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

image 1



image 2

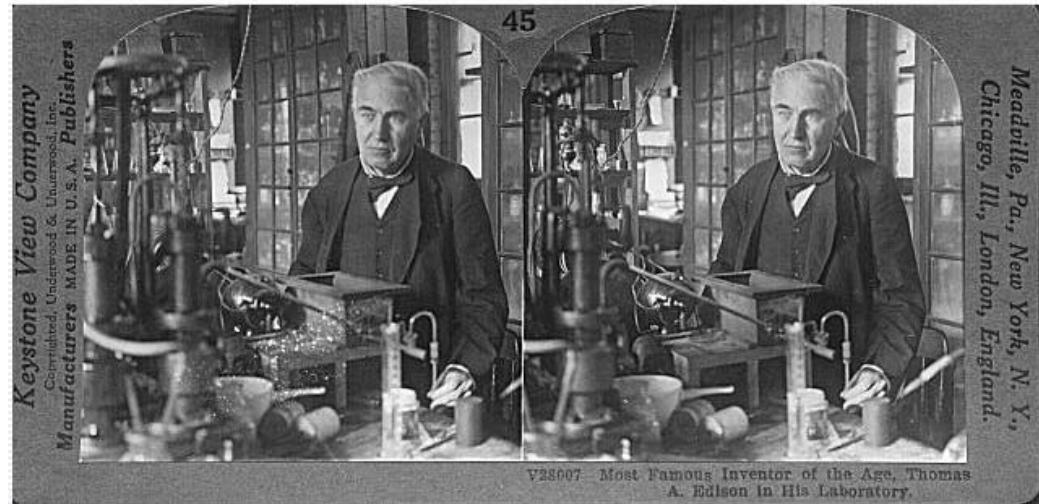


Dense depth map



Binocular stereo

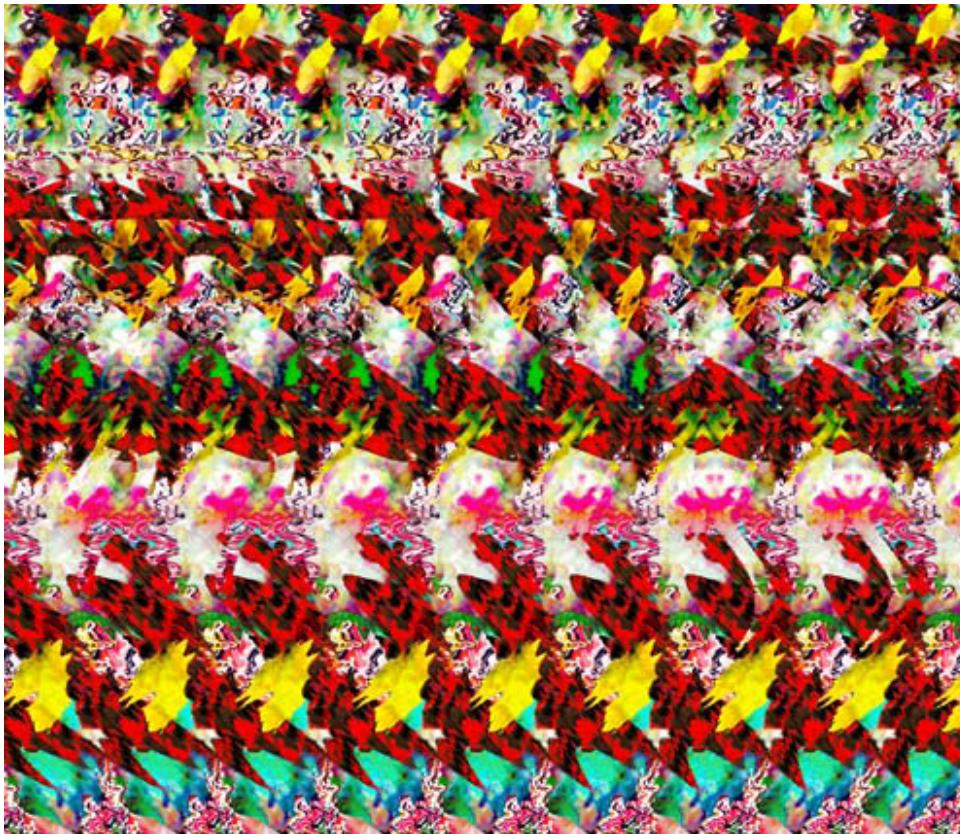
- Given a calibrated binocular stereo pair, fuse it to produce a depth image
 - Humans can do it



Stereograms: Invented by Sir Charles Wheatstone, 1838

Binocular stereo

- Given a calibrated binocular stereo pair, fuse it to produce a depth image
 - Humans can do it



Autostereograms: www.magiceye.com

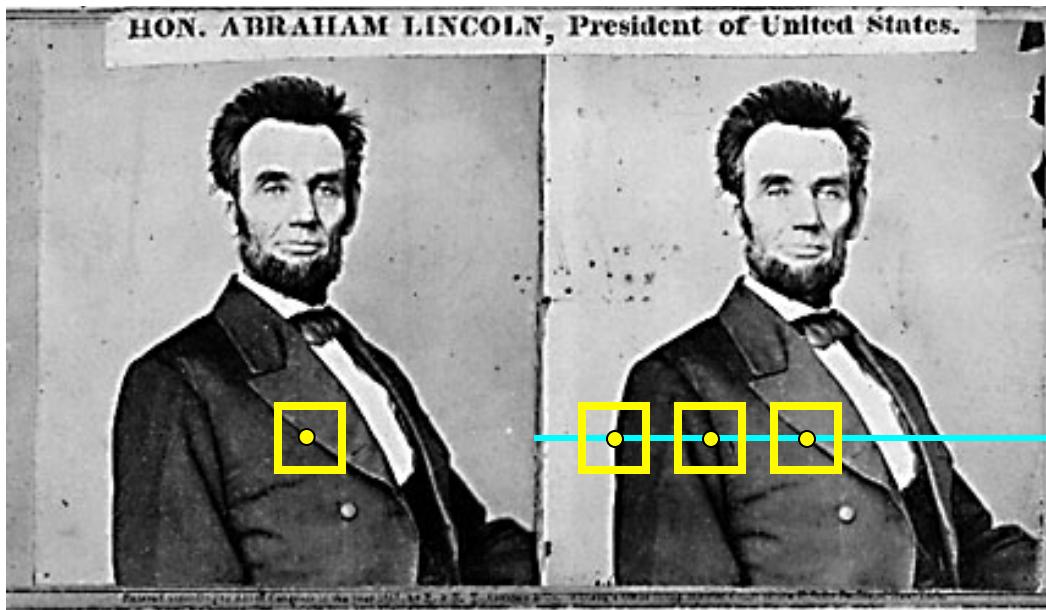
Binocular stereo

- Given a calibrated binocular stereo pair, fuse it to produce a depth image
 - Humans can do it



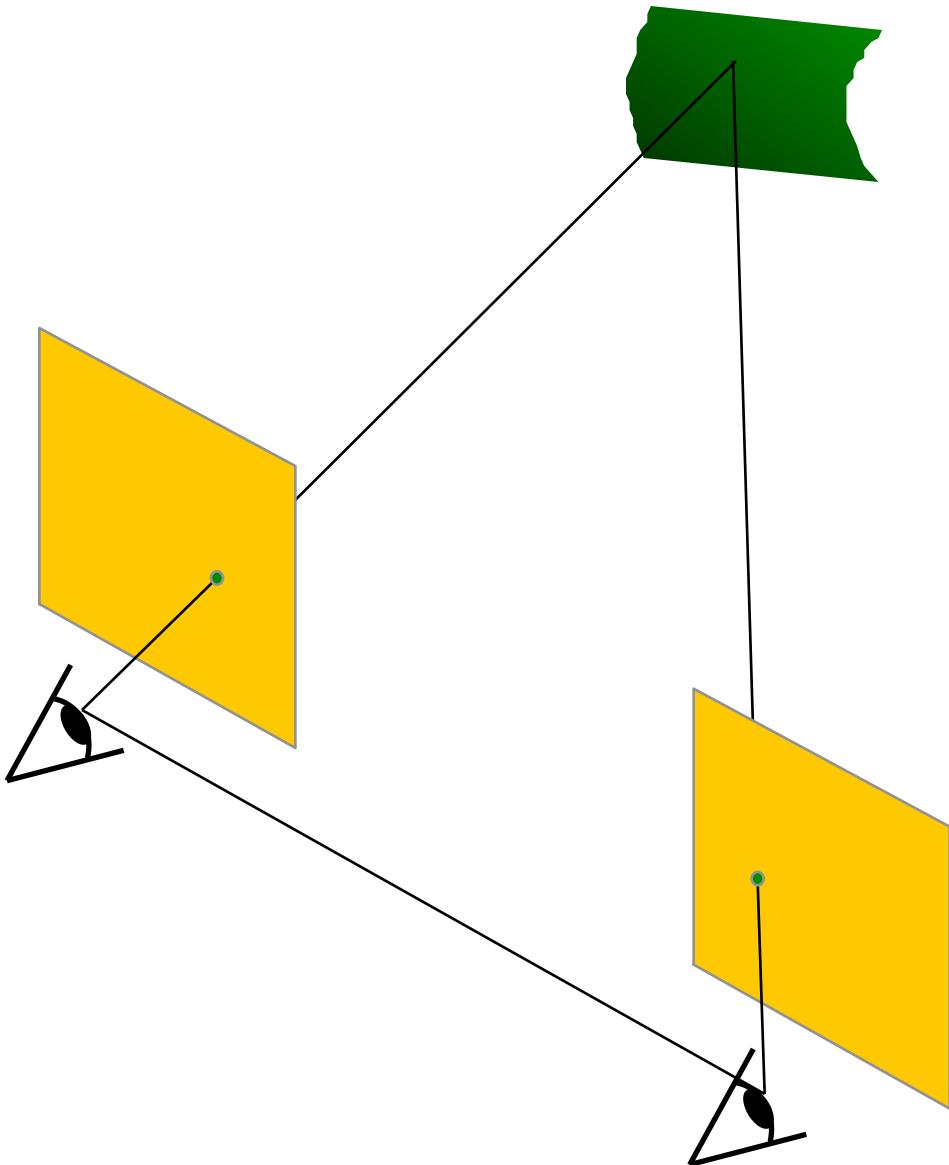
Autostereograms: www.magiceye.com

Basic stereo matching algorithm



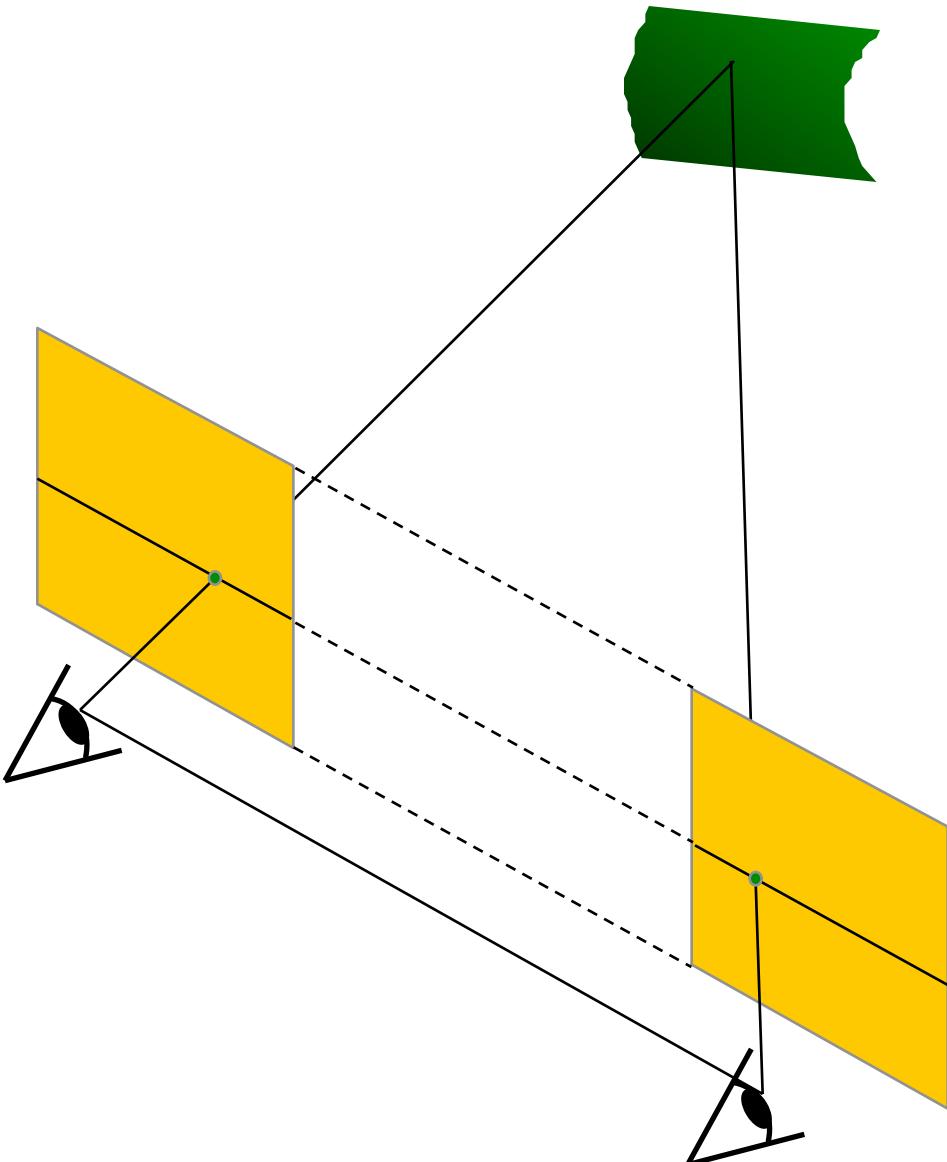
- For each pixel in the first image
 - Find corresponding epipolar line in the right image
 - Examine all pixels on the epipolar line and pick the best match
 - Triangulate the matches to get depth information
- Simplest case: epipolar lines are corresponding scanlines
 - When does this happen?

Simplest Case: Parallel images



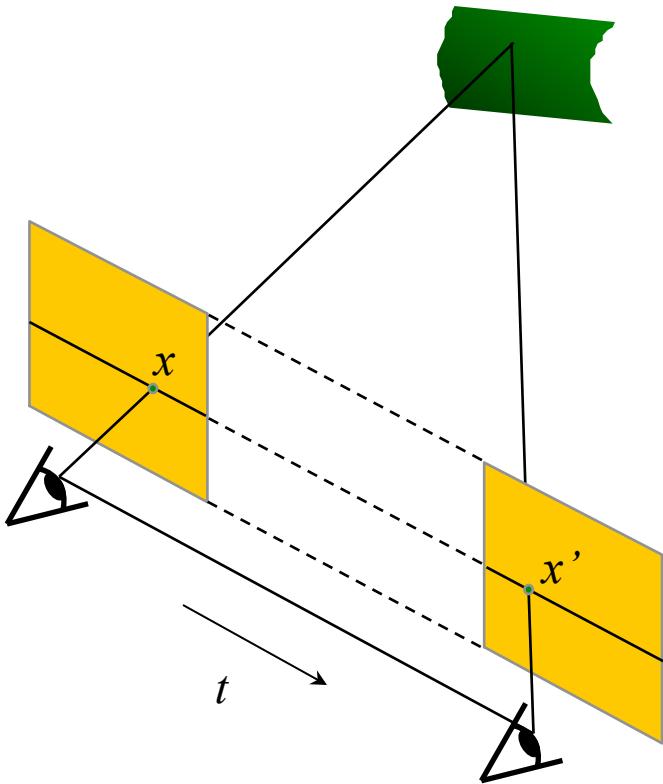
- Image planes of cameras are parallel to each other and to the baseline
- Camera centers are at same height
- Focal lengths are the same

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 are the same
- Then, epipolar lines fall along the horizontal scan lines of the images

Essential matrix for parallel images



Epipolar constraint:

$$\mathbf{x}'^T \mathbf{E} \mathbf{x} = 0, \quad \mathbf{E} = [\mathbf{t}_x] \mathbf{R}$$

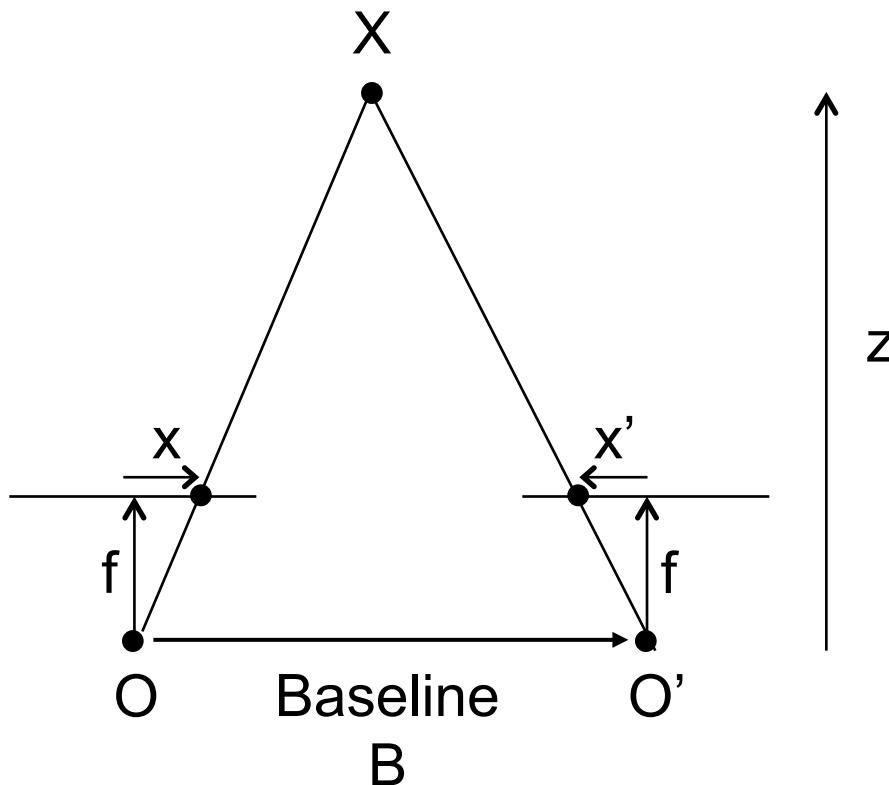
$$\mathbf{R} = \mathbf{I} \quad \mathbf{t} = (T, 0, 0)$$

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

$$(u' \quad v' \quad 1) \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -T \\ 0 & T & 0 \end{bmatrix} \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = 0 \quad (u' \quad v' \quad 1) \begin{pmatrix} 0 \\ -T \\ Tv \end{pmatrix} = 0 \quad Tv' = Tv$$

The y-coordinates of corresponding points are the same!

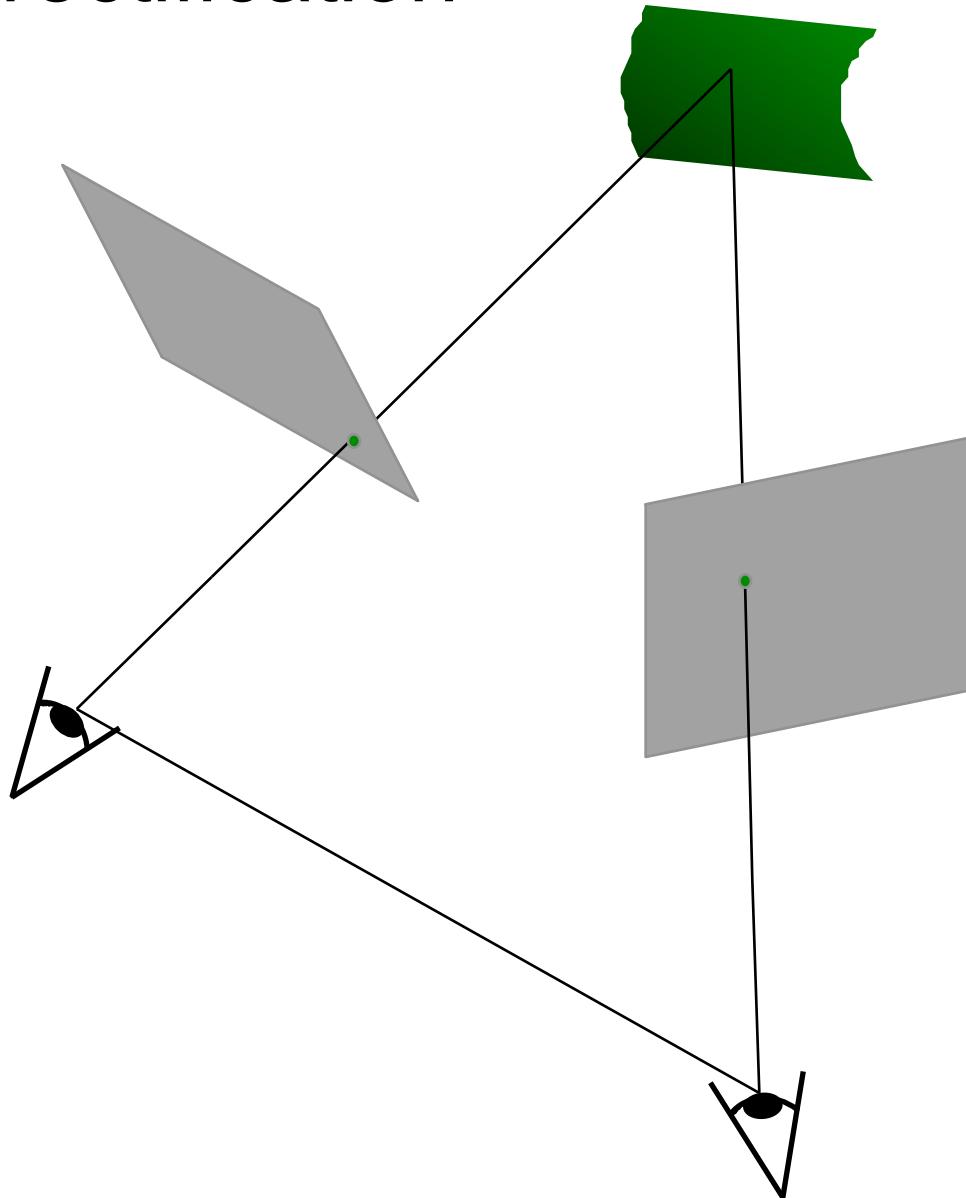
Depth from disparity



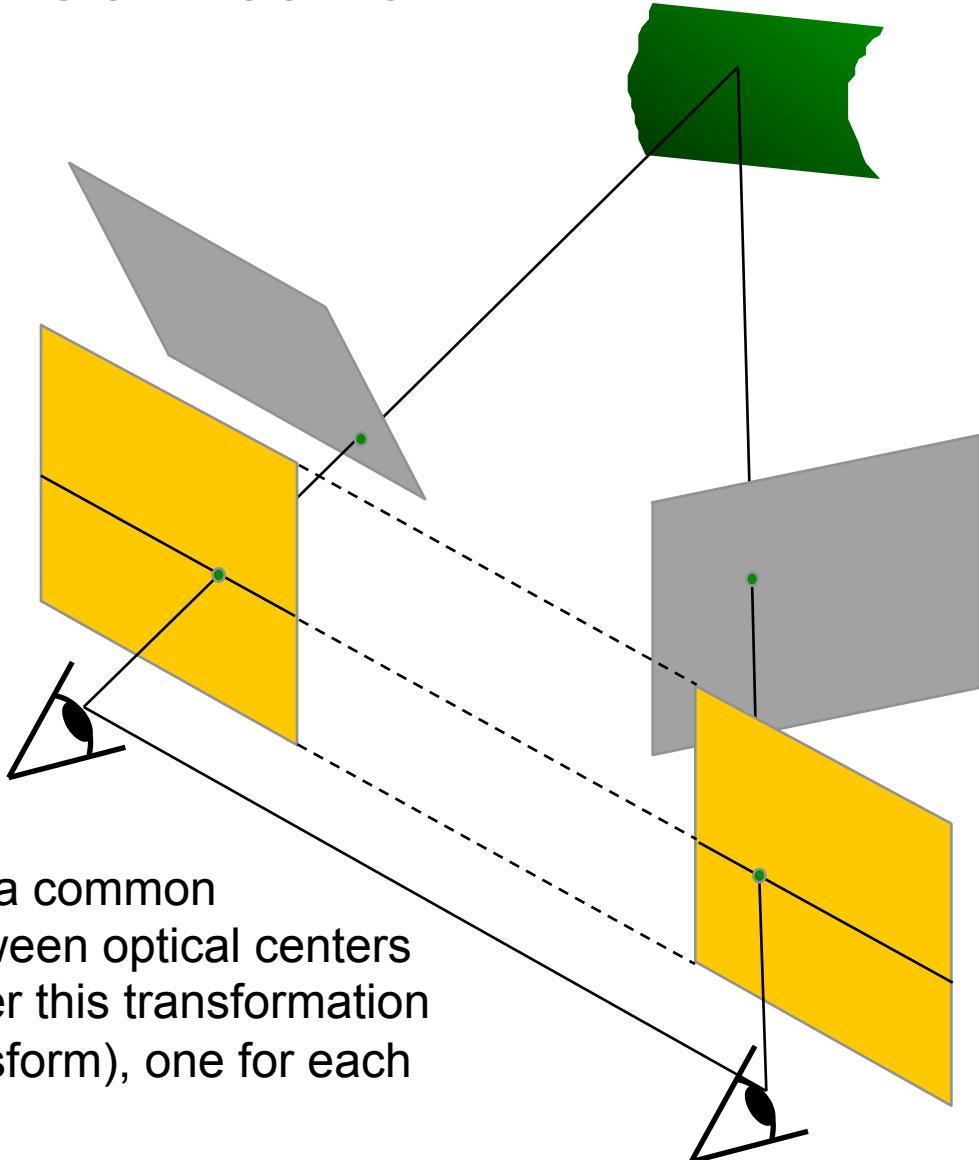
$$\text{disparity} = x - x' = \frac{B \cdot f}{z}$$

Disparity is inversely proportional to depth!

Stereo image rectification

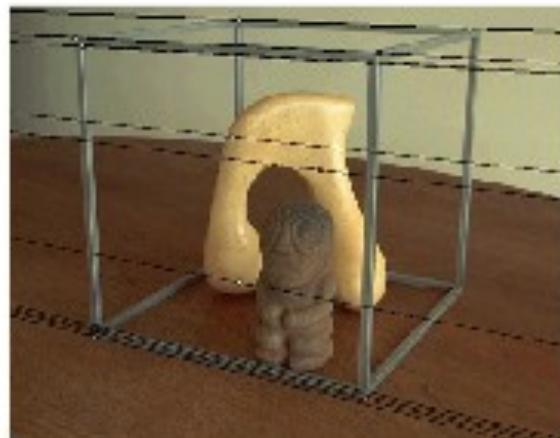


Stereo image rectification

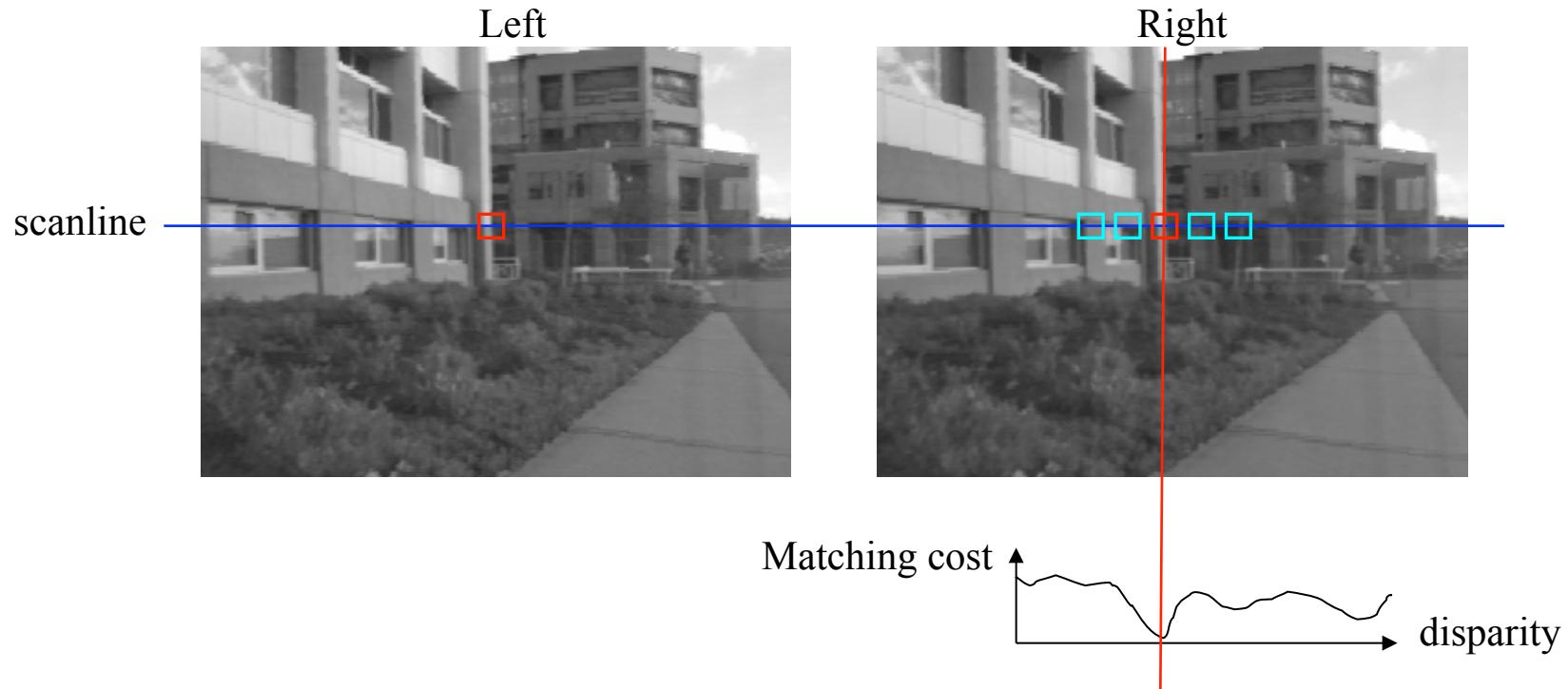


- reproject image planes onto a common plane parallel to the line between optical centers
 - pixel motion is horizontal after this transformation
 - two homographies (3x3 transform), one for each input image reprojection
- C. Loop and Z. Zhang.
[Computing Rectifying Homographies for Stereo Vision](#). IEEE Conf. Computer Vision and Pattern Recognition, 1999.

Rectification example

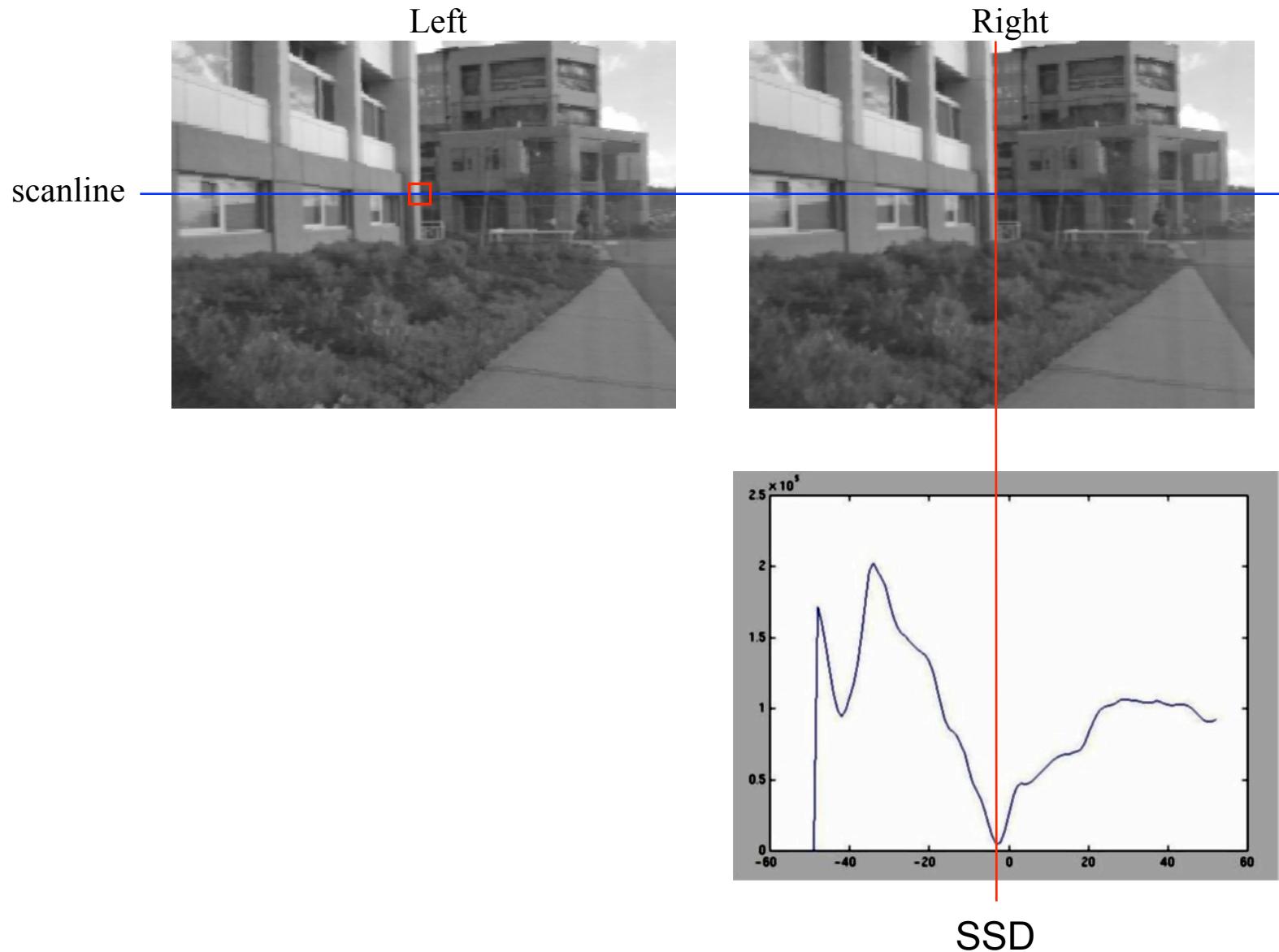


Correspondence search

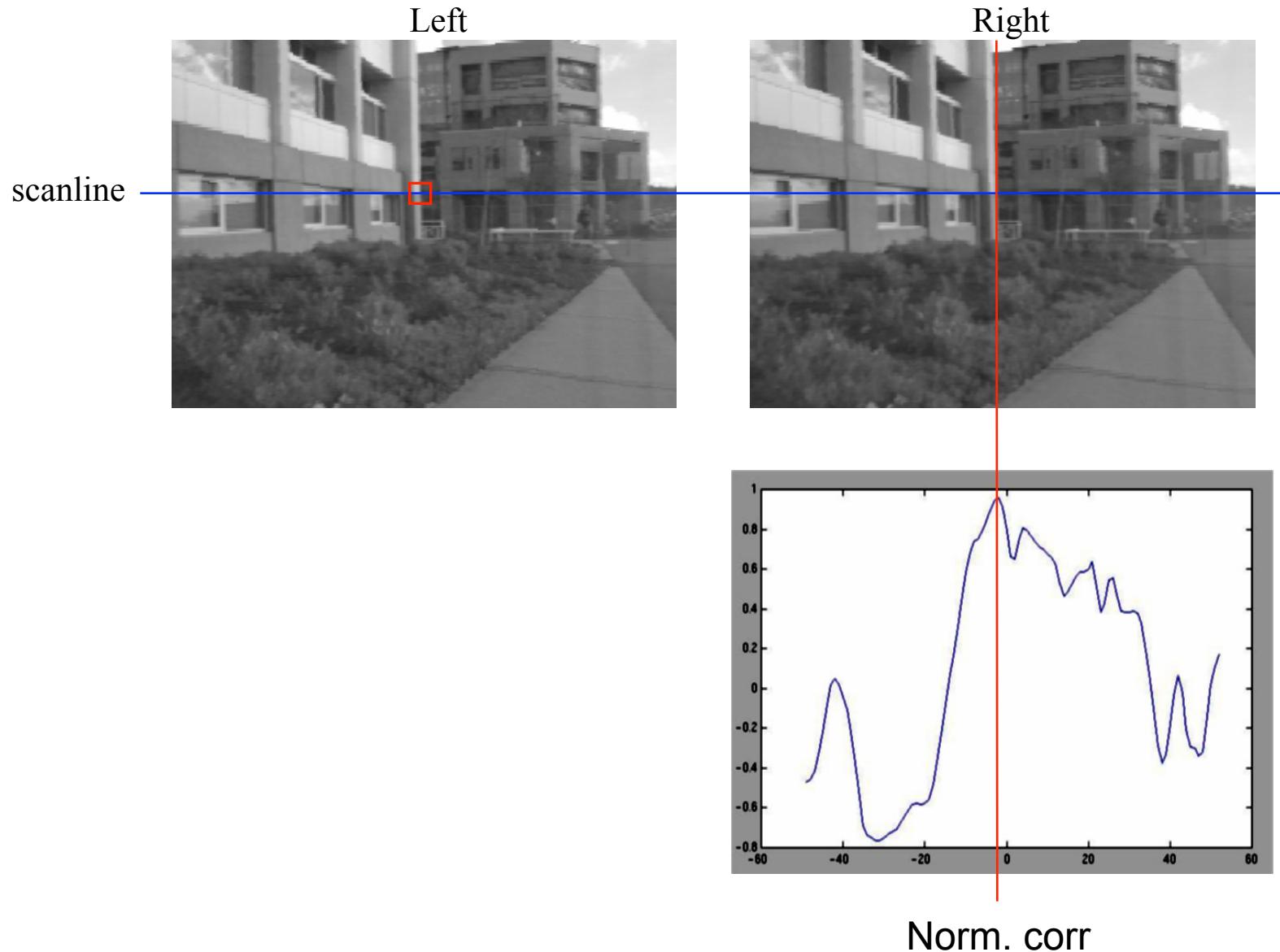


- Slide a window along the right scanline and compare contents of that window with the reference window in the left image
- Matching cost: SSD or normalized correlation

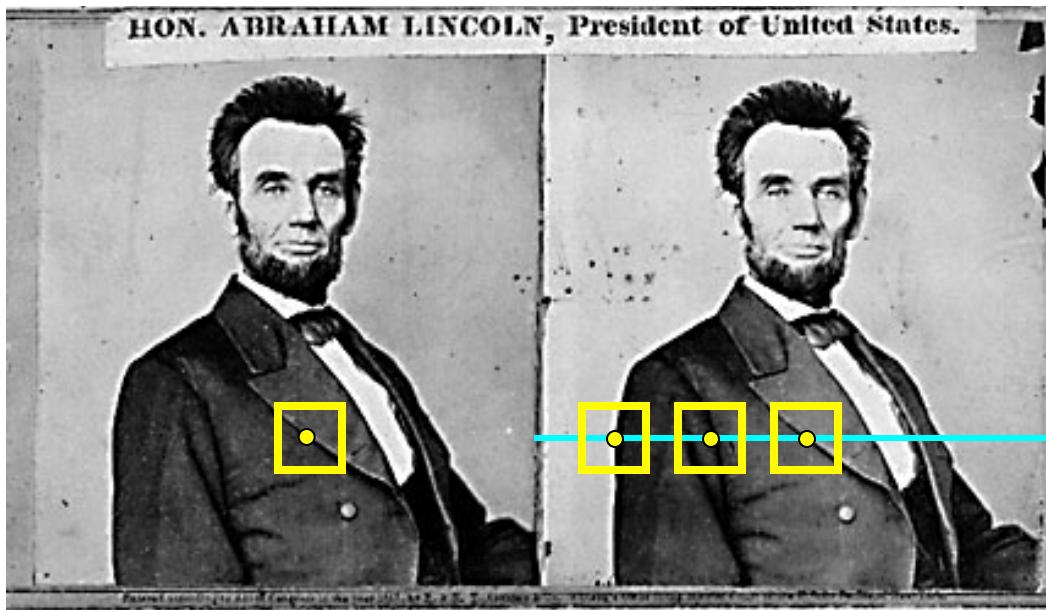
Correspondence search



Correspondence search

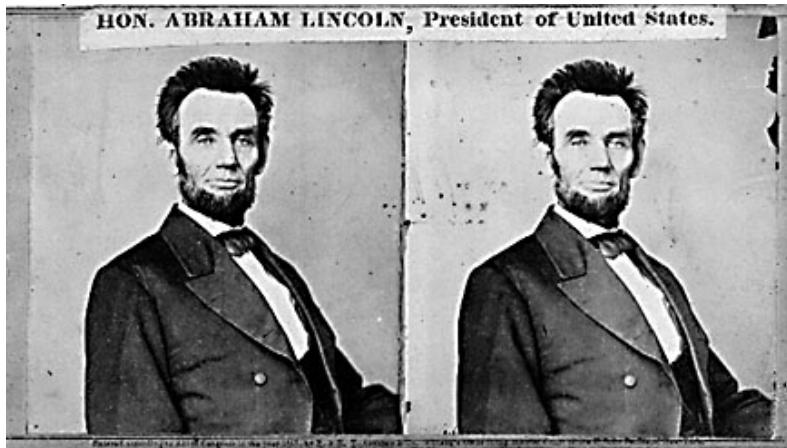


Basic stereo matching algorithm



- 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 surfaces



Occlusions, repetition

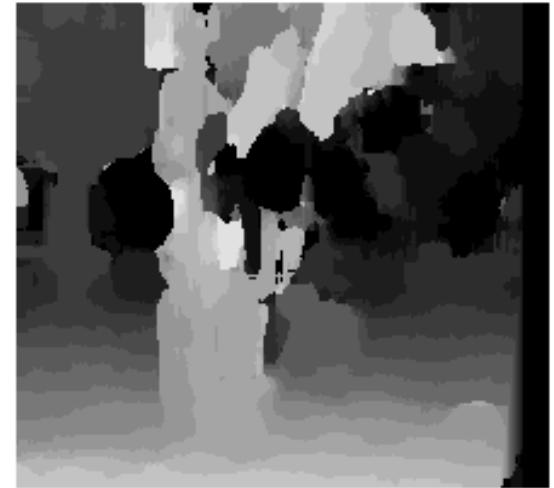


Non-Lambertian surfaces, specularities

Effect of window size



$$W = 3$$



$$W = 20$$

- Smaller window
 - + More detail
 - More noise

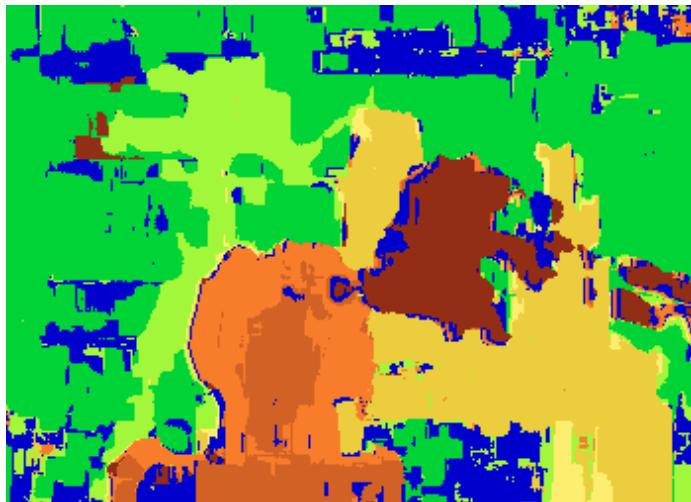
- Larger window
 - + Smoother disparity maps
 - Less detail

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

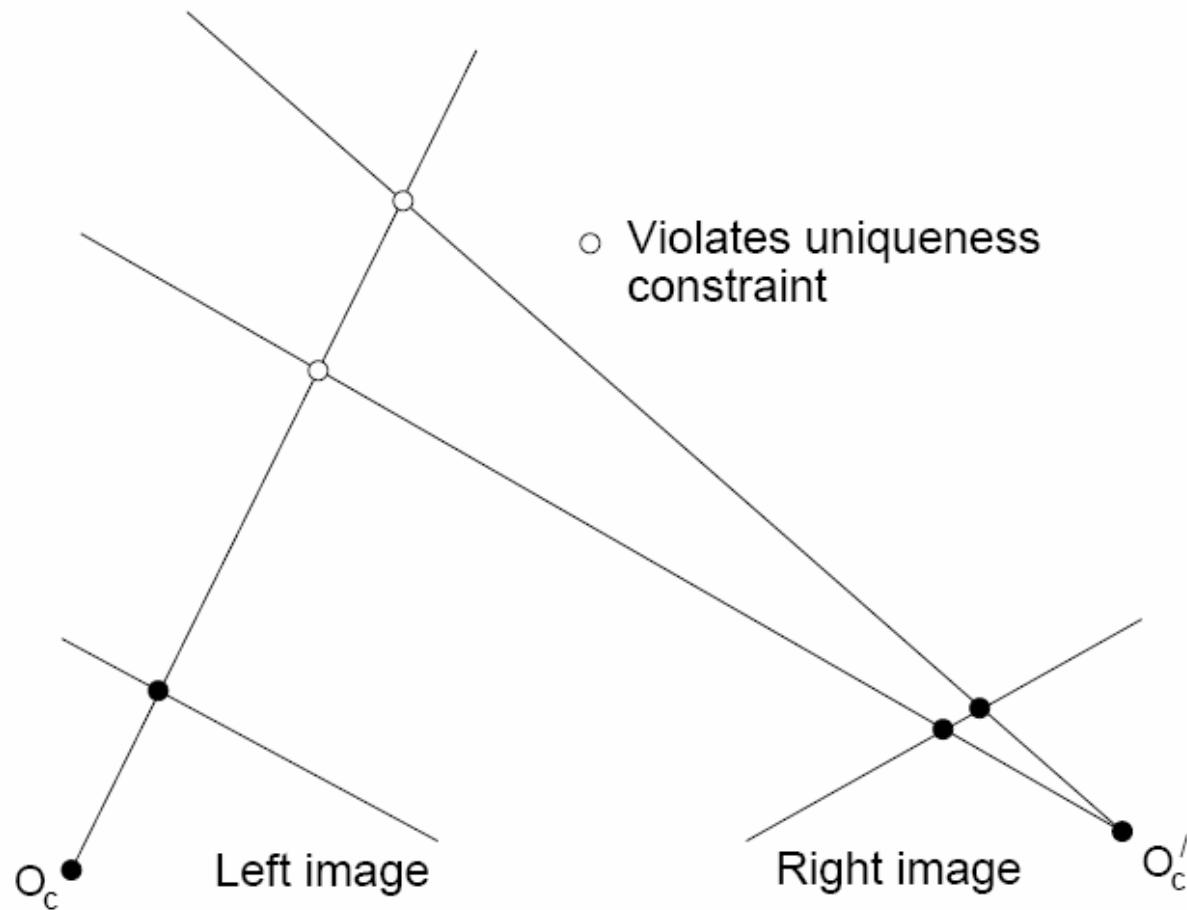
For the latest and greatest: <http://www.middlebury.edu/stereo/>

How can we improve window-based matching?

- The similarity constraint is **local** (each reference window is matched independently)
- Need to enforce **non-local** correspondence constraints

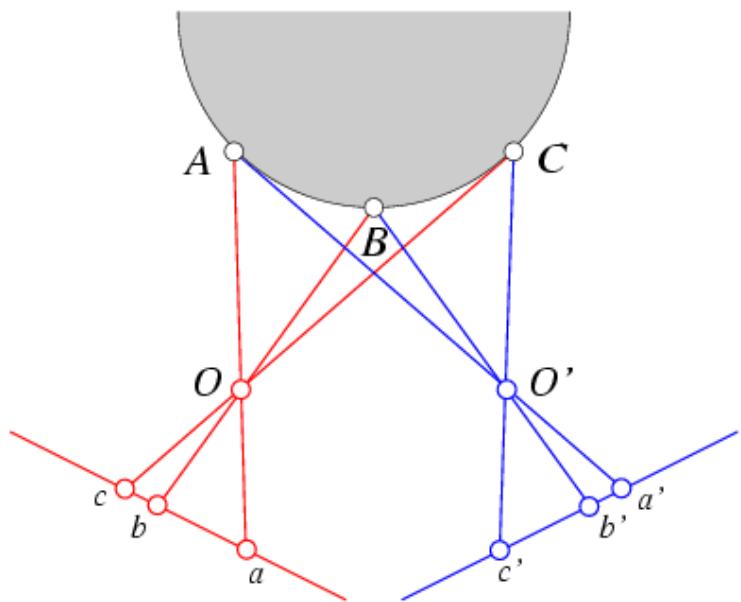
Non-local constraints

- Uniqueness
 - For any point in one image, there should be at most one matching point in the other image



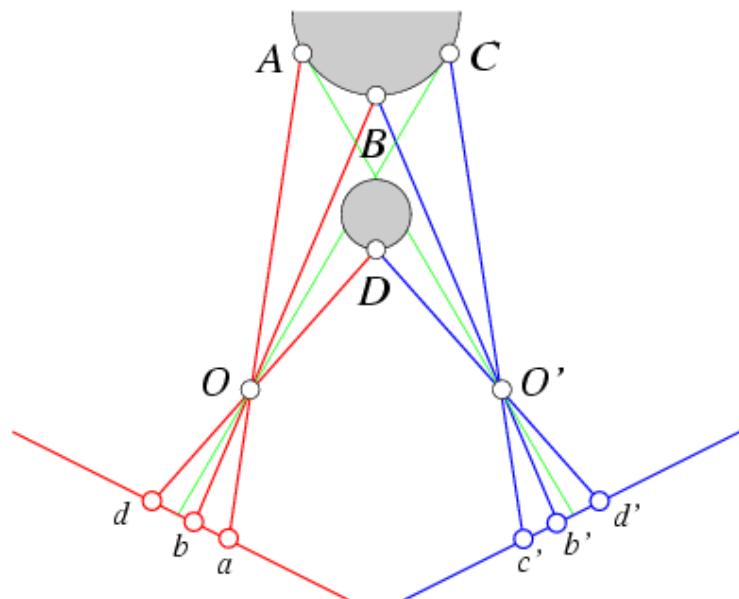
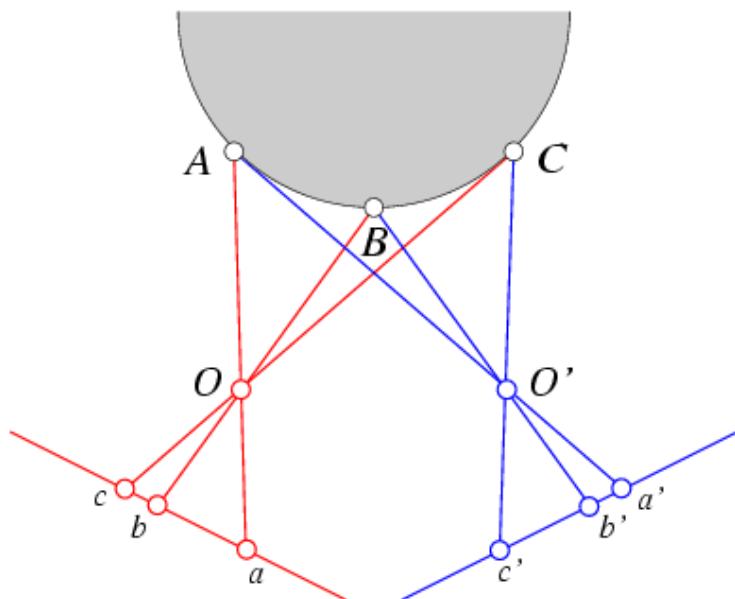
Non-local constraints

- Uniqueness
 - For any point in one image, there should be at most one matching point in the other image
- Ordering
 - Corresponding points should be in the same order in both views



Non-local constraints

- Uniqueness
 - For any point in one image, there should be at most one matching point in the other image
- Ordering
 - Corresponding points should be in the same order in both views



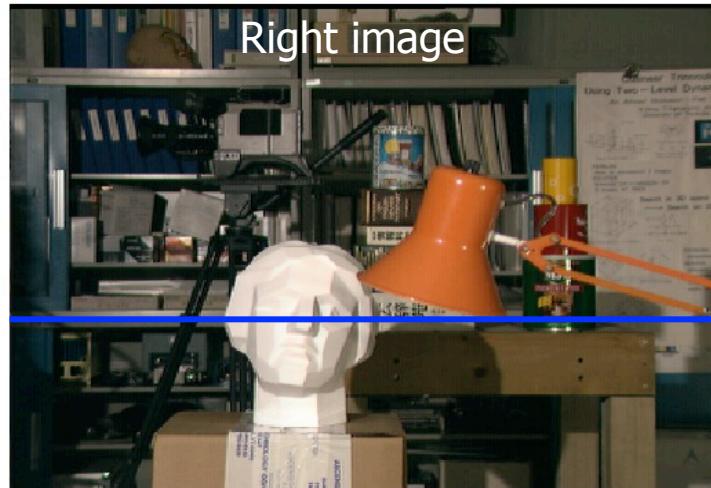
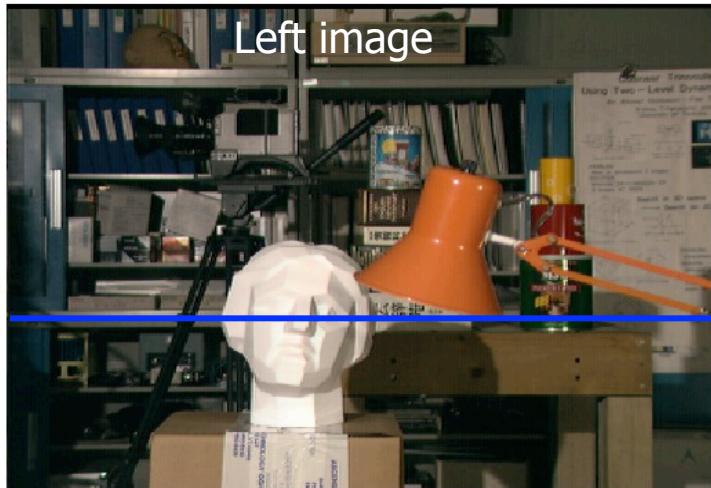
Ordering constraint doesn't hold

Non-local constraints

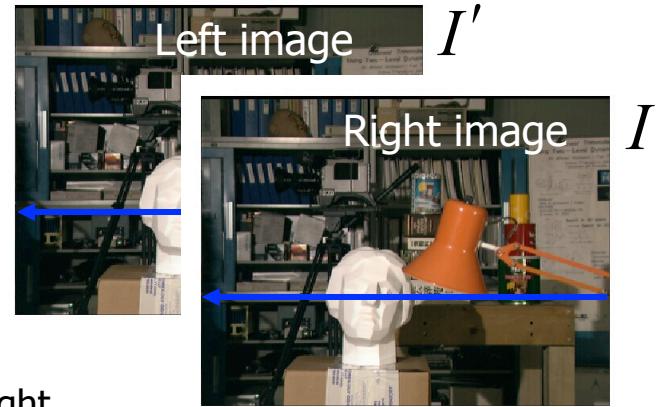
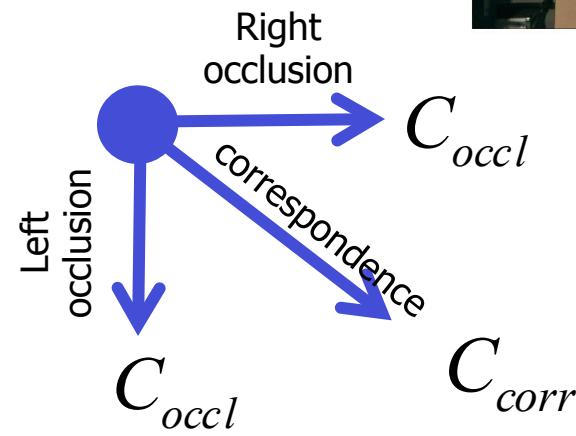
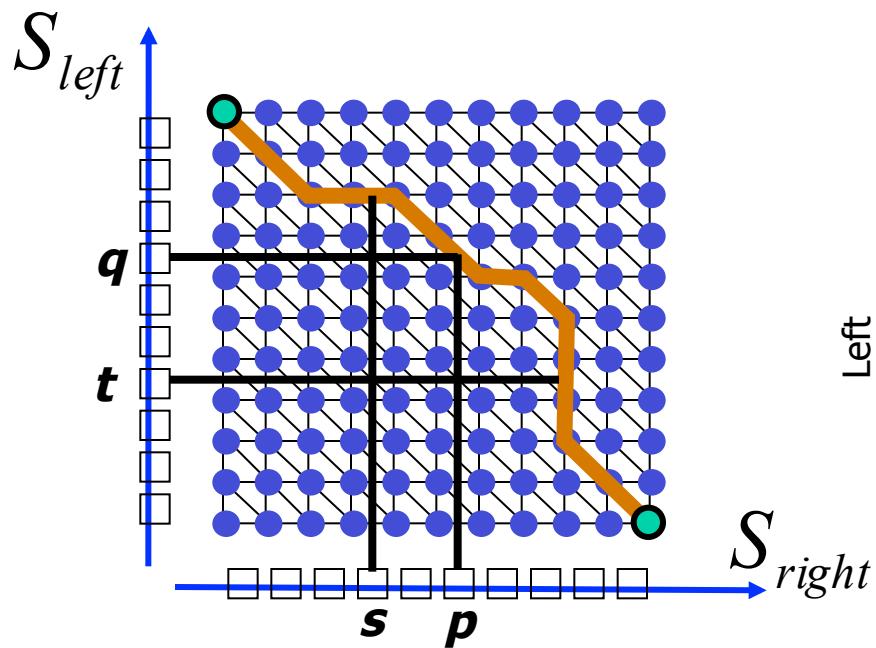
- Uniqueness
 - For any point in one image, there should be at most one matching point in the other image
- Ordering
 - Corresponding points should be in the same order in both views
- Smoothness
 - We expect disparity values to change slowly (for the most part)

Scanline stereo

- Try to coherently match pixels on the entire scanline
- Different scanlines are still optimized independently



“Shortest paths” for scan-line stereo



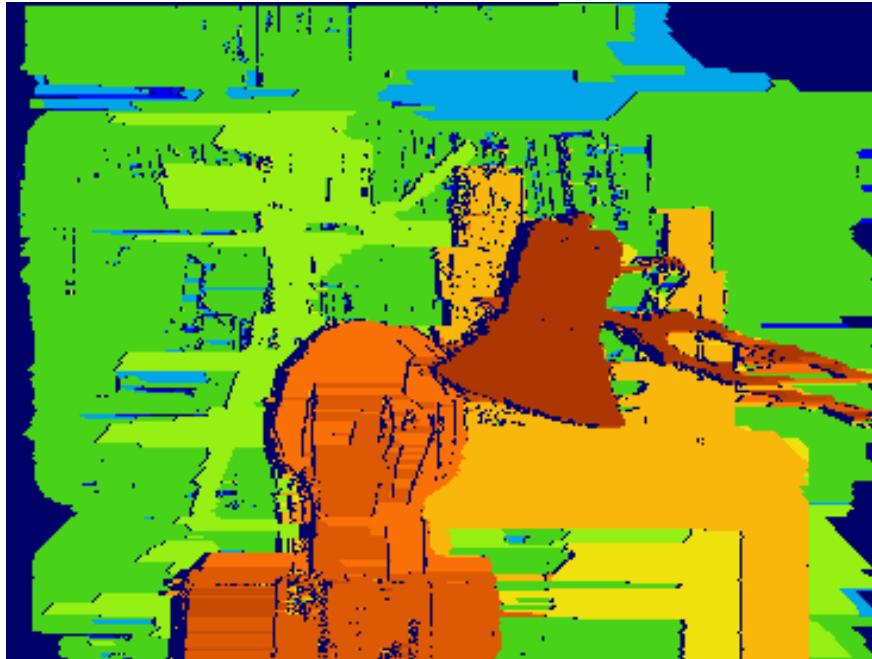
Can be implemented with dynamic programming

Ohta & Kanade '85, Cox et al. '96

Slide credit: Y. Boykov

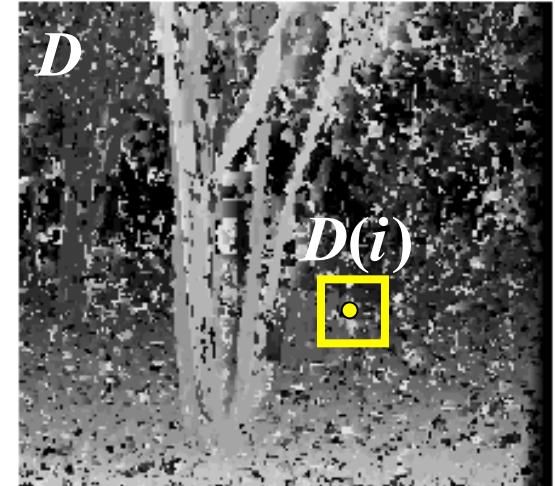
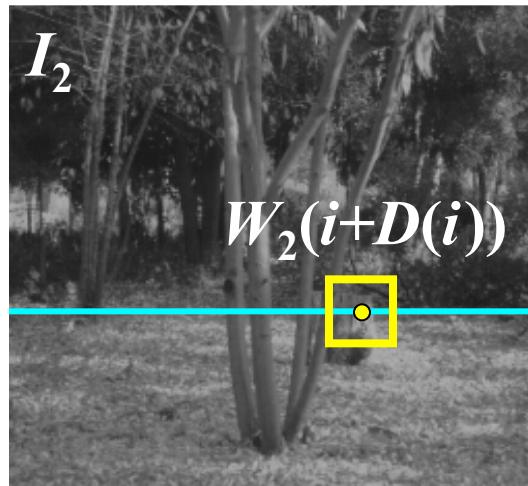
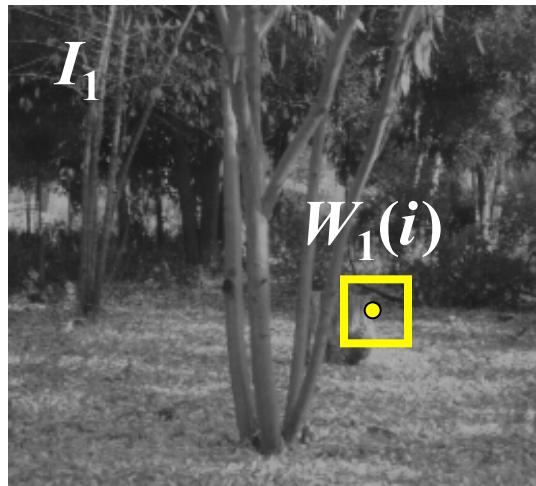
Coherent stereo on 2D grid

- Scanline stereo generates streaking artifacts



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

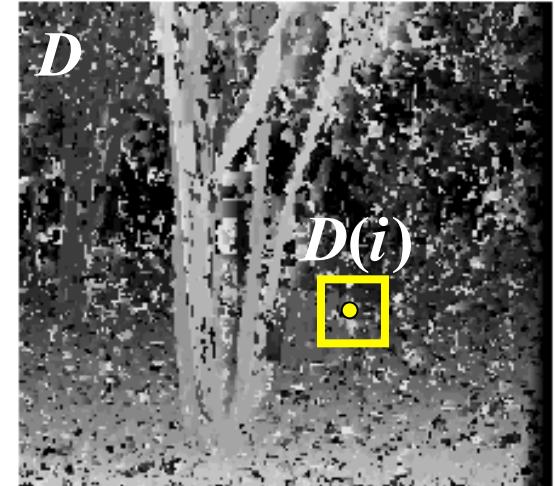
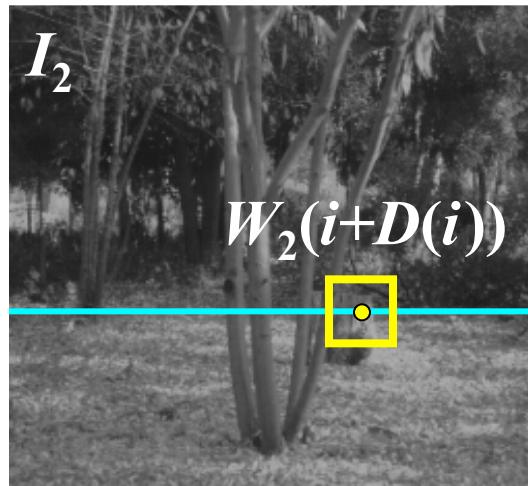
Stereo matching as energy minimization



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

- Energy functions of this form can be minimized using *graph cuts*

Stereo matching as energy minimization



- Probabilistic interpretation: we want to find a Maximum A Posteriori (MAP) estimate of disparity image D :

$$P(D | I_1, I_2) \propto P(I_1, I_2 | D)P(D)$$

$$-\log P(D | I_1, I_2) \propto -\log P(I_1, I_2 | D) - \log P(D)$$

$$E = E_{\text{data}}(I_1, I_2, D) + \lambda E_{\text{smooth}}(D)$$

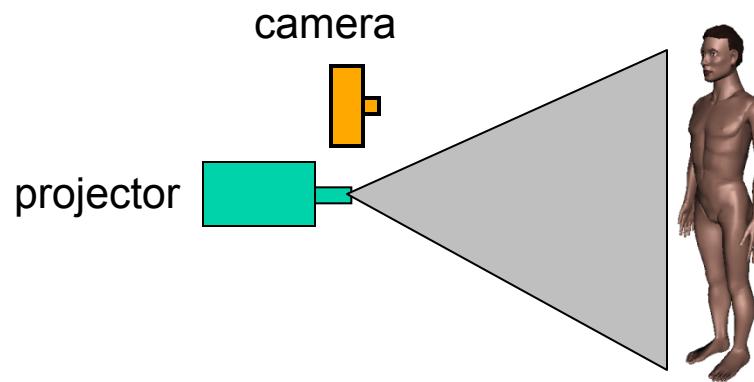
Stereo matching as energy minimization

- Note: the above formulation does not treat the two images symmetrically, does not enforce uniqueness, and does not take occlusions into account
- It is possible to come up with an energy that does all these things, but it's a bit more complex
 - Defined over all possible sets of matches, not over all disparity maps with respect to the first image
 - Includes an *occlusion term*
 - The smoothness term looks different and more complicated

Active stereo with structured light



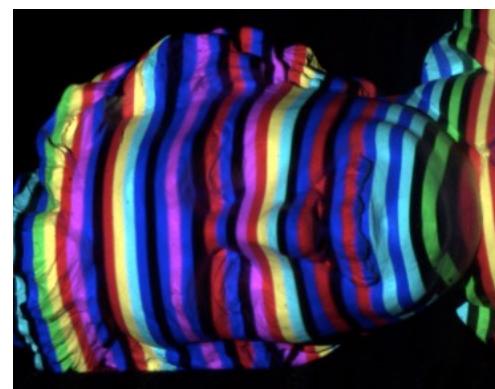
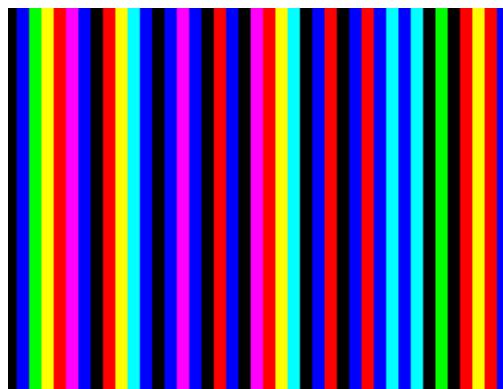
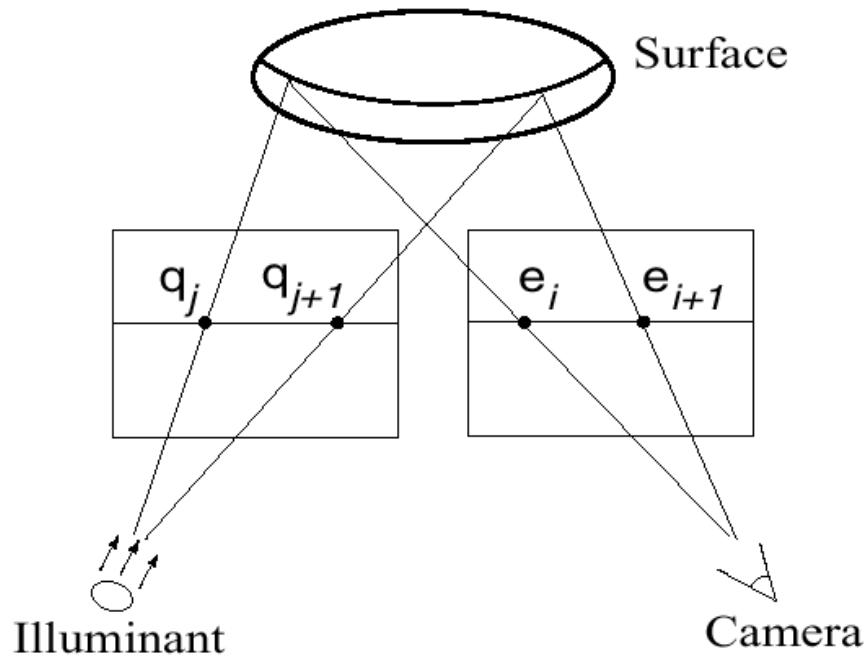
- Project “structured” light patterns onto the object
 - Simplifies the correspondence problem
 - Allows us to use only one camera



L. Zhang, B. Curless, and S. M. Seitz.

[Rapid Shape Acquisition Using Color Structured Light and Multi-pass Dynamic Programming. 3DPVT 2002](#)

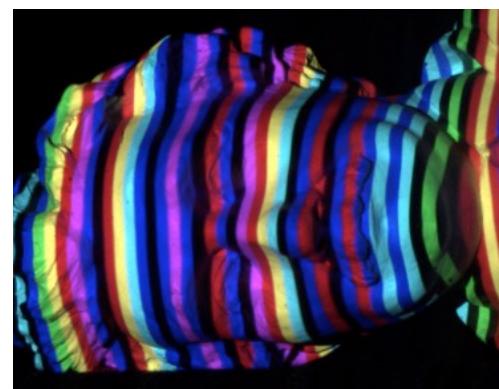
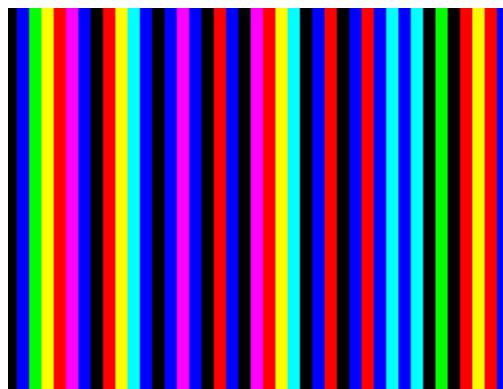
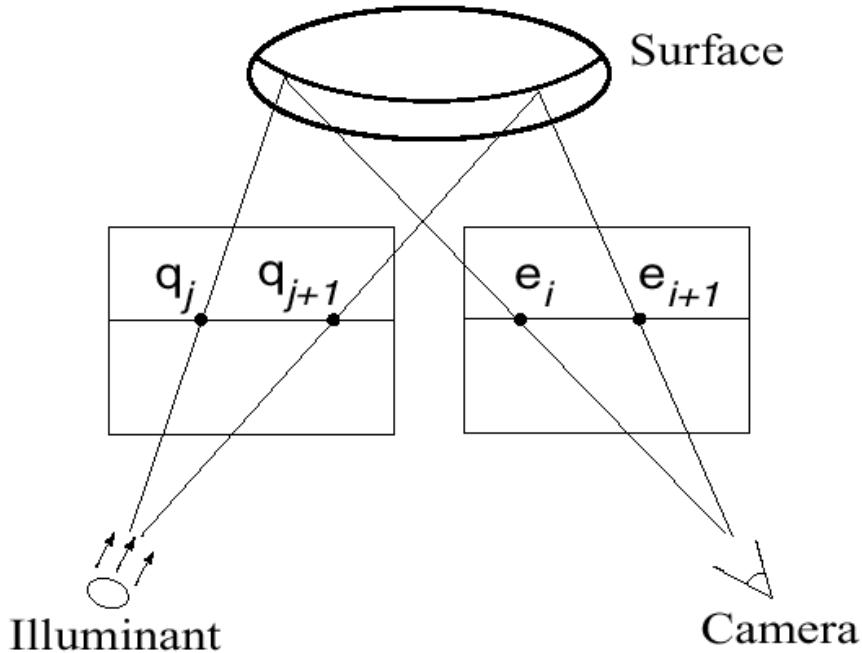
Active stereo with structured light



L. Zhang, B. Curless, and S. M. Seitz.

[Rapid Shape Acquisition Using Color Structured Light and Multi-pass Dynamic Programming.](#) 3DPVT 2002

Active stereo with structured light



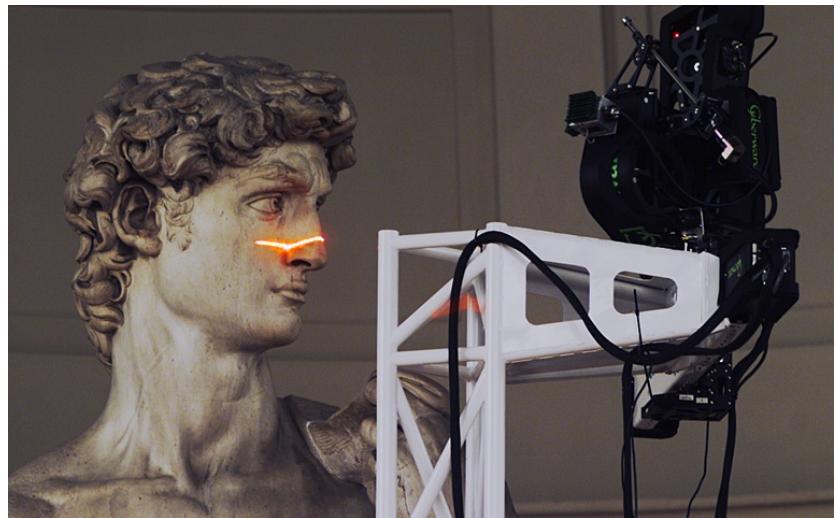
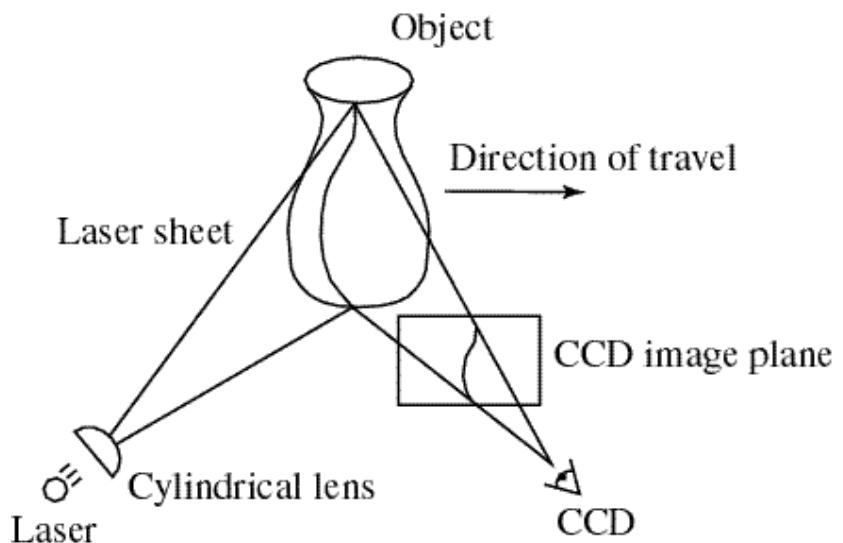
http://en.wikipedia.org/wiki/Structured-light_3D_scanner

Kinect: Structured infrared light



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

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.

Source: S. Seitz

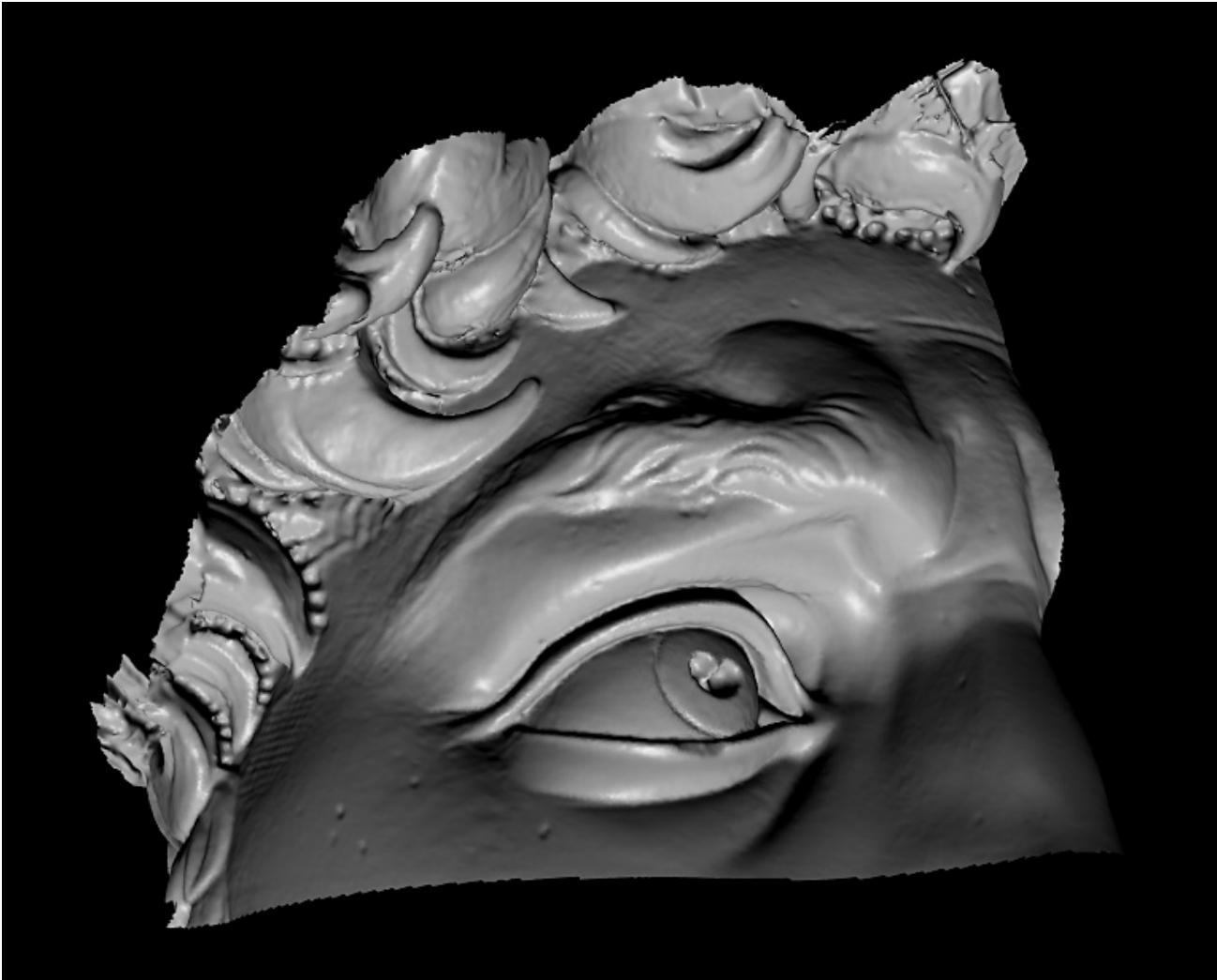
Laser scanned models



The Digital Michelangelo Project, Levoy et al.

Source: S. Seitz

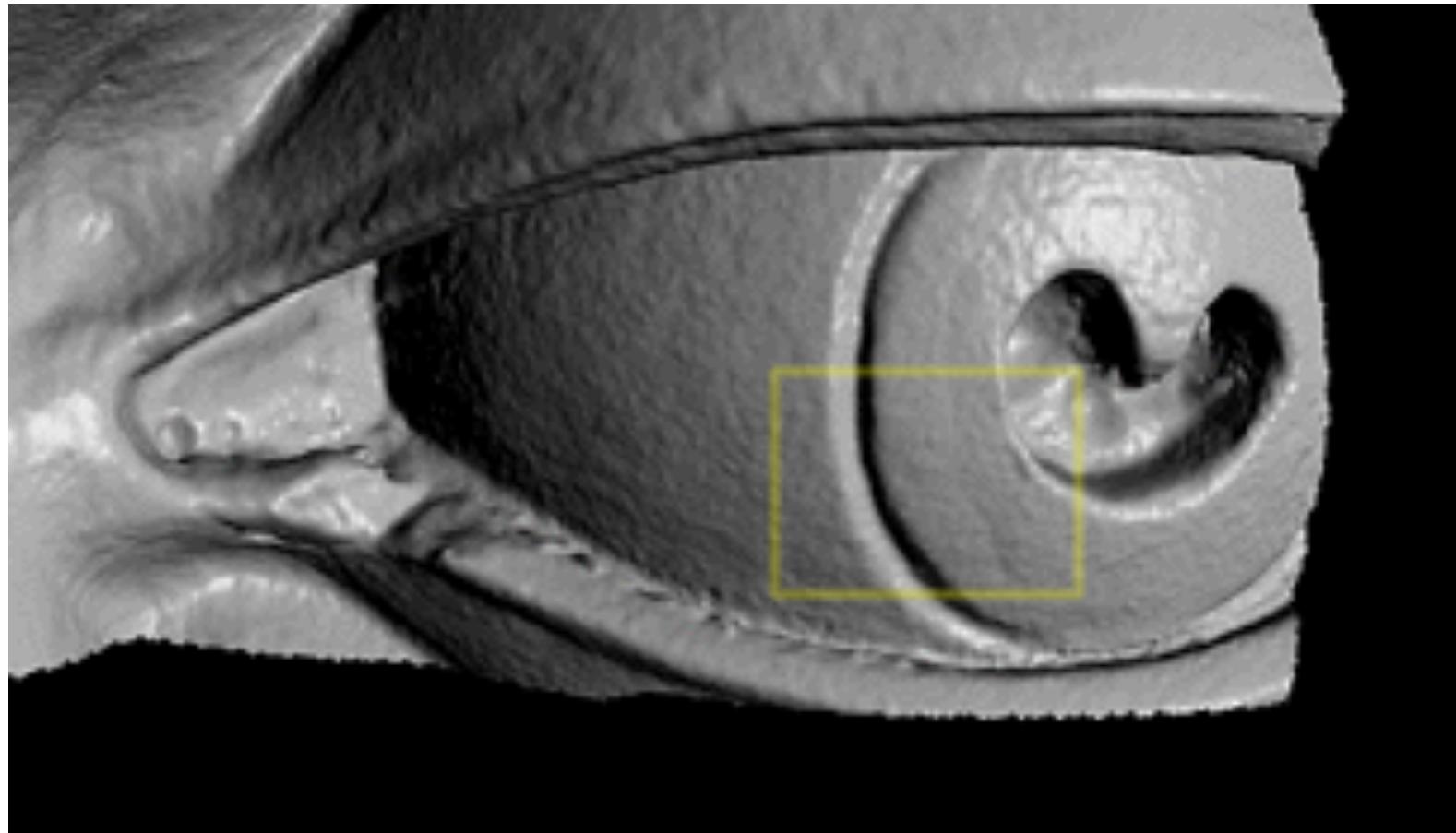
Laser scanned models



The Digital Michelangelo Project, Levoy et al.

Source: S. Seitz

Laser scanned models

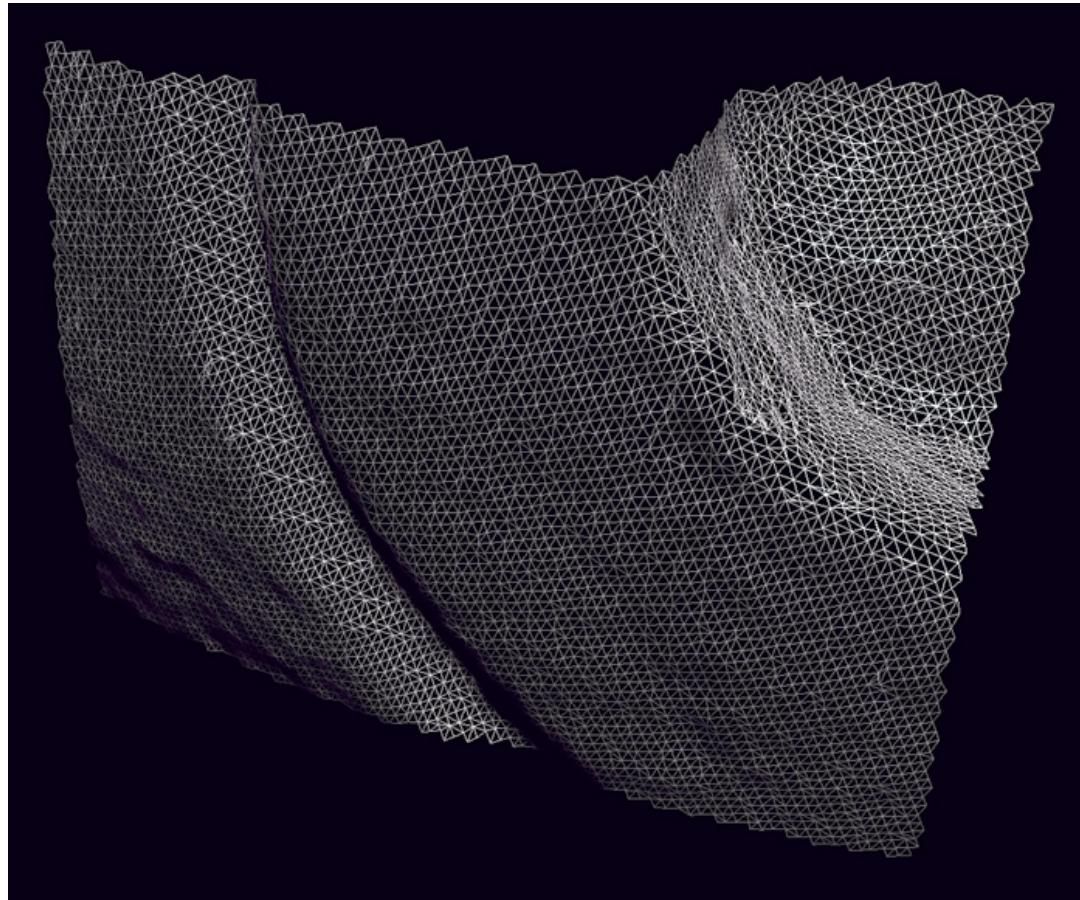


The Digital Michelangelo Project, Levoy et al.

Source: S. Seitz

Laser scanned models

1.0 mm resolution (56 million triangles)



The Digital Michelangelo Project, Levoy et al.

Source: S. Seitz

Aligning range images

- A single range scan is not sufficient to describe a complex surface
- Need techniques to register multiple range images



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

Aligning range images

- A single range scan is not sufficient to describe a complex surface
- Need techniques to register multiple range images
 - ... which brings us to *multi-view stereo*