Stereo vision

Many slides adapted from Steve Seitz

What is stereo vision?

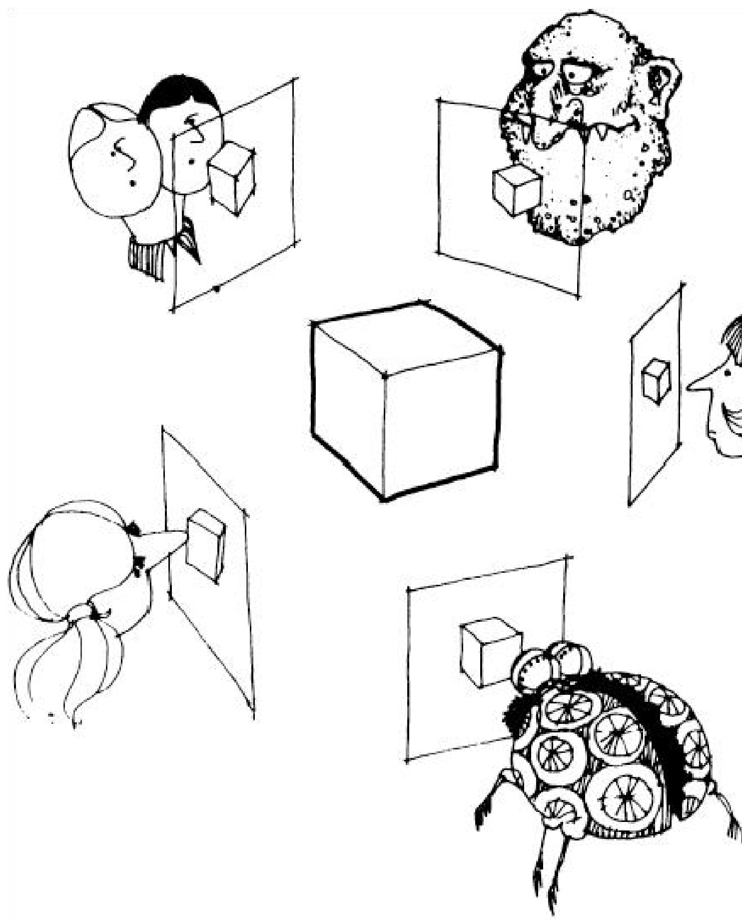
- Generic problem formulation: given several images of the same object or scene, compute a representation of its 3D shape

What is stereo vision?

- 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 (isolated cameras or video sequence)
  - Cameras can be calibrated or uncalibrated
- “Representation of 3D shape”
  - Depth maps
  - Meshes
  - Point clouds
  - Patch clouds
  - Volumetric models
  - Layered models

What is stereo vision?

- Generic problem formulation: given several images of the same object or scene, compute a representation of its 3D shape



What is stereo vision?

- Narrower formulation: given a calibrated binocular stereo pair, fuse it to produce a depth image

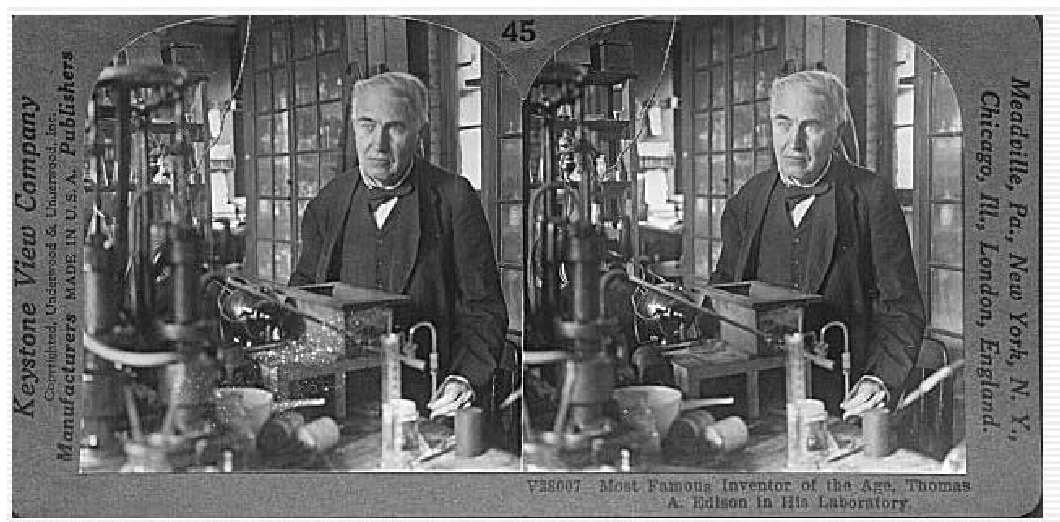
image 1

image 2

Dense depth map

What is stereo vision?

- Narrower formulation: given a calibrated binocular stereo pair, fuse it to produce a depth image
  - Humans can do it



Stereograms: Invented by Sir Charles Wheatstone, 1838

What is stereo vision?

- Narrower formulation: given a calibrated binocular stereo pair, fuse it to produce a depth image
  - Humans can do it

Autostereograms: www.magiceye.com

What is stereo vision?

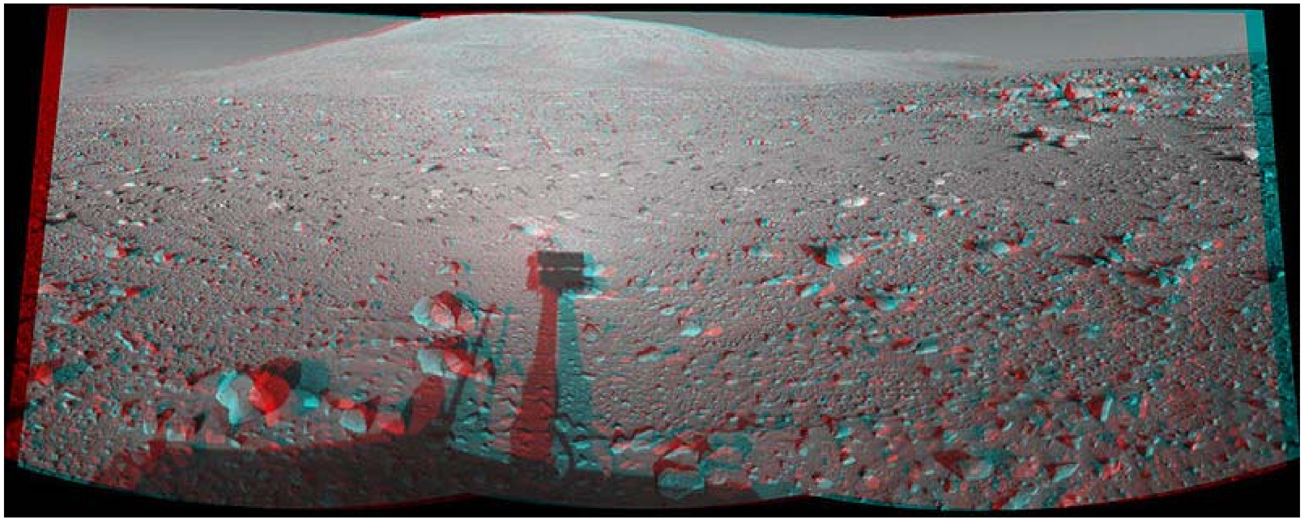
- Narrower formulation: given a calibrated binocular stereo pair, fuse it to produce a depth image
  - Humans can do it

Autostereograms: www.magiceye.com

Application of stereo: Robotic exploration

Nomad robot searches for meteorites in Antartica

Real-time stereo on Mars



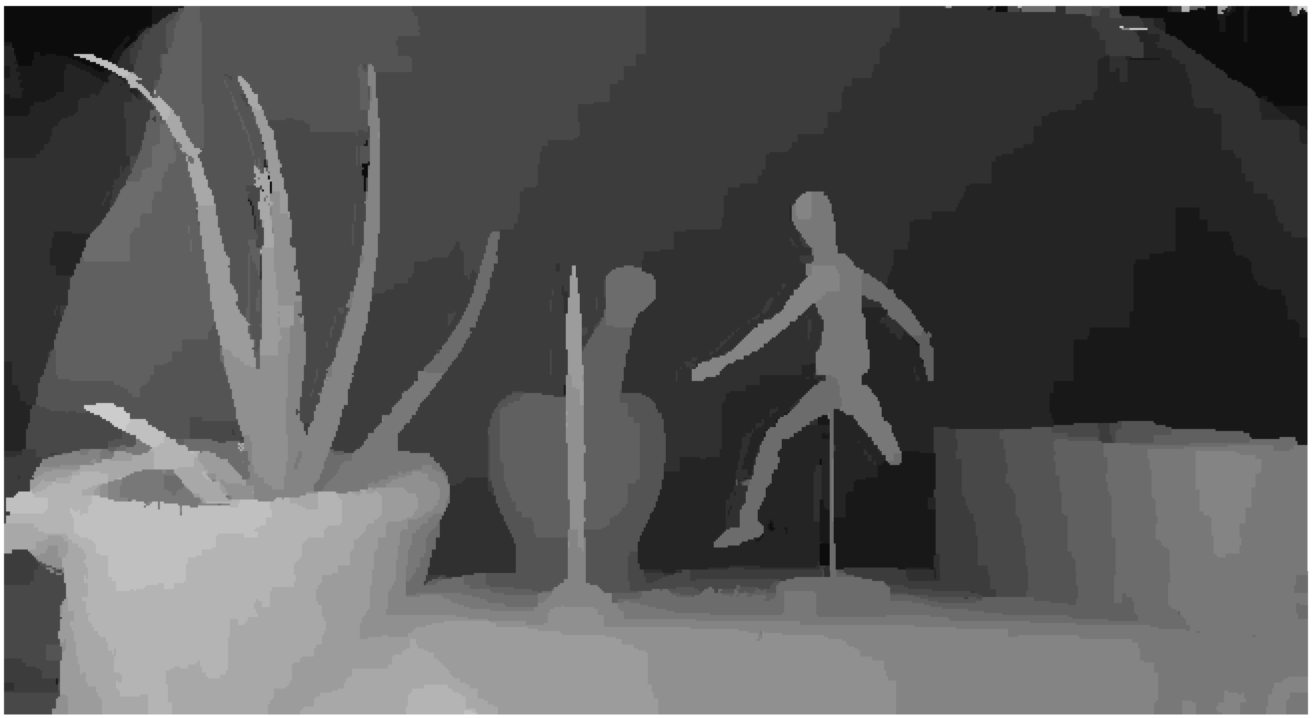
Application: View Interpolation

Right Image

Application: View Interpolation

Left Image

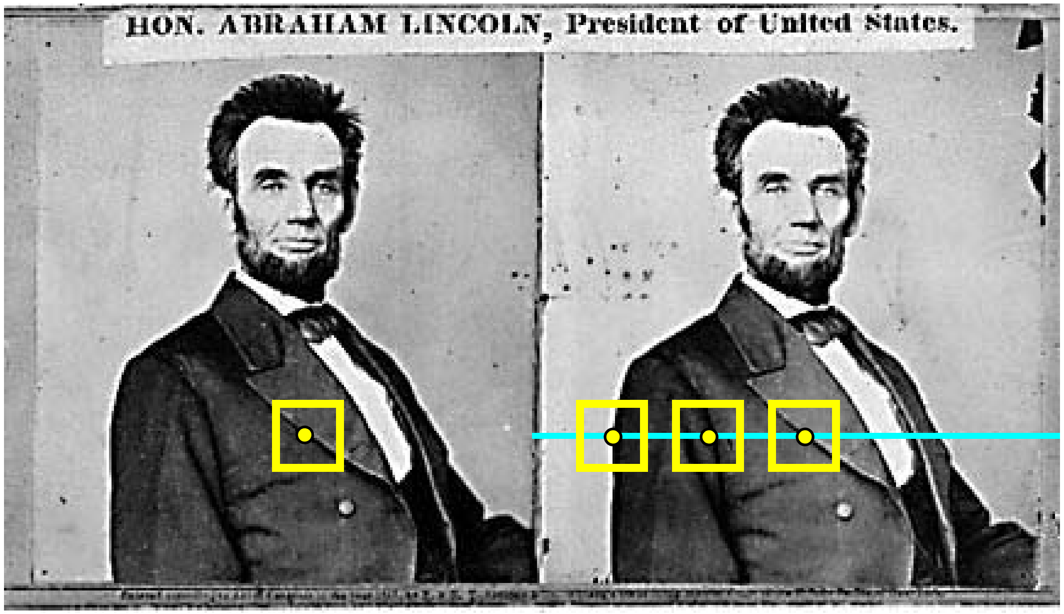
Application: View Interpolation



Disparity

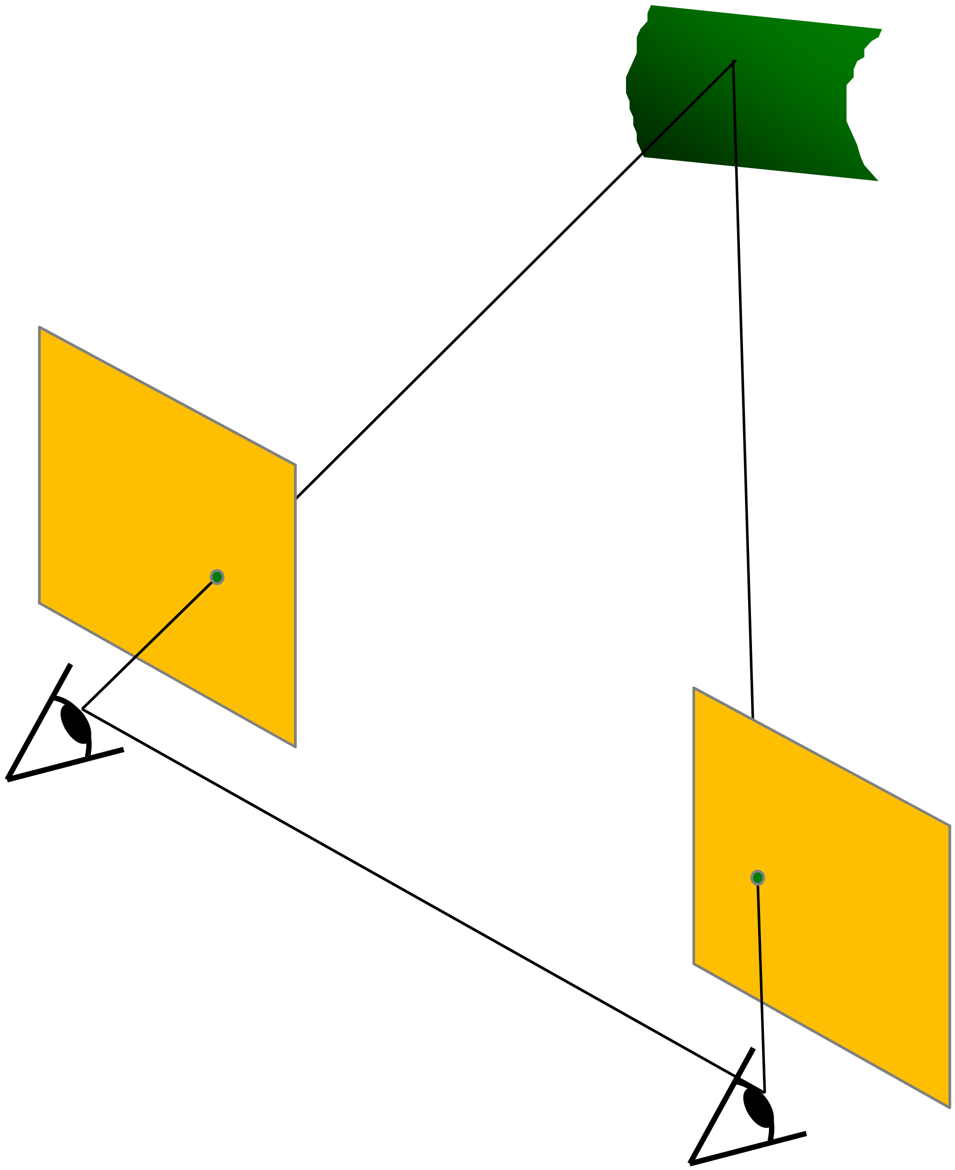
Application: View Interpolation

Basic stereo matching algorithm



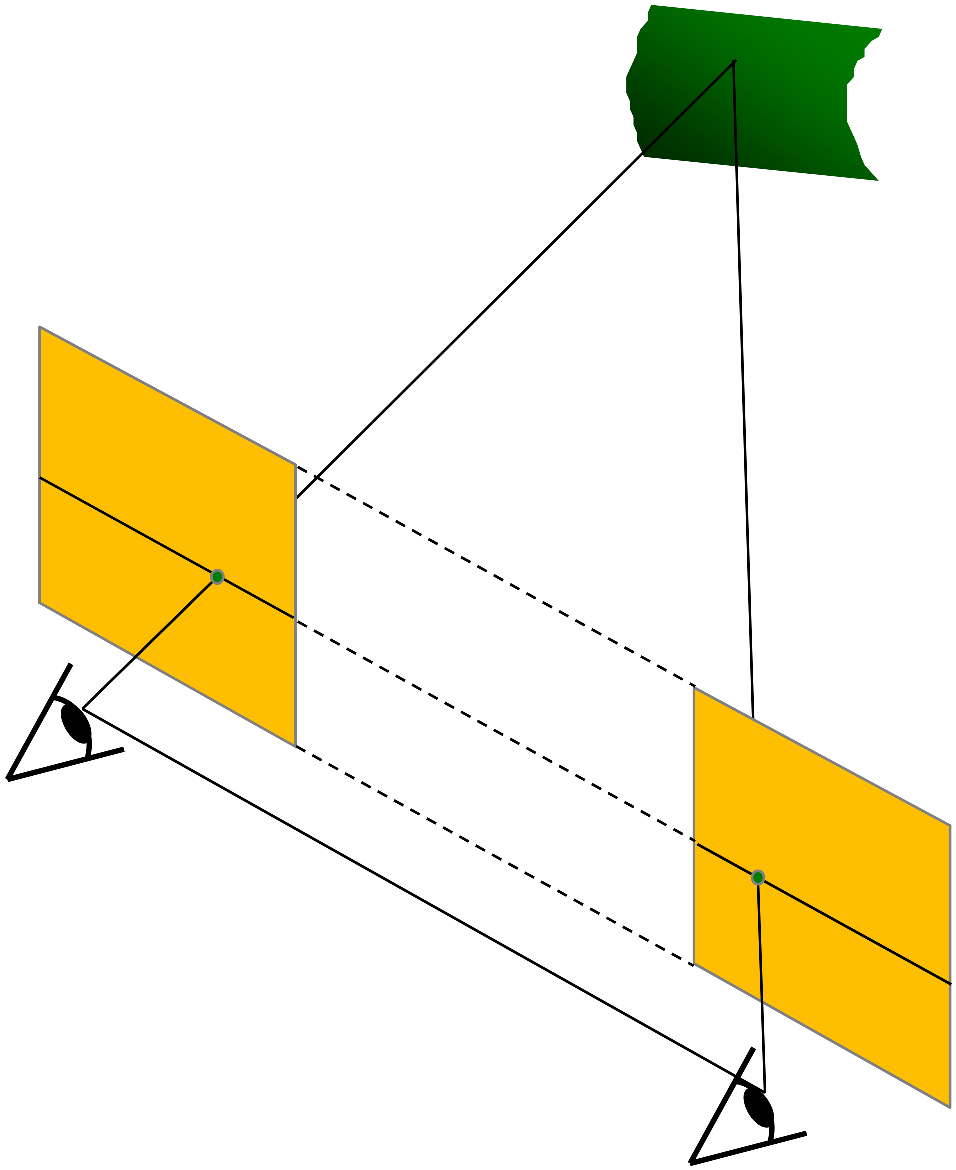
- For each pixel in the first image
  - Find corresponding epipolar line in the right image
  - Examine all pixels on the epipolar line and pick the best match
  - Triangulate the matches to get depth information
- Simplest case: epipolar lines are scanlines
  - When does this happen?

Simplest Case: Parallel images



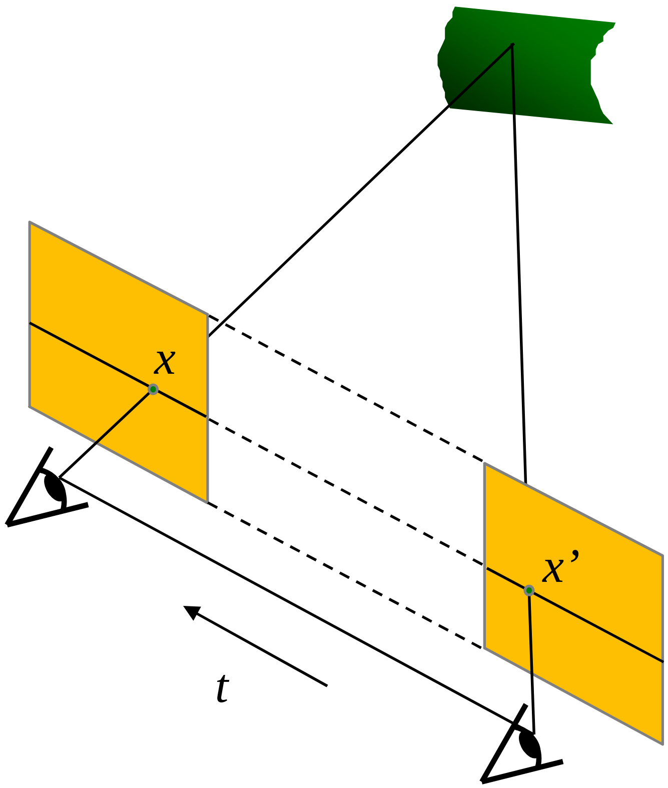
- Image planes of cameras are parallel to each other and to the baseline
- Camera centers are at same height
- Focal lengths are the same

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:

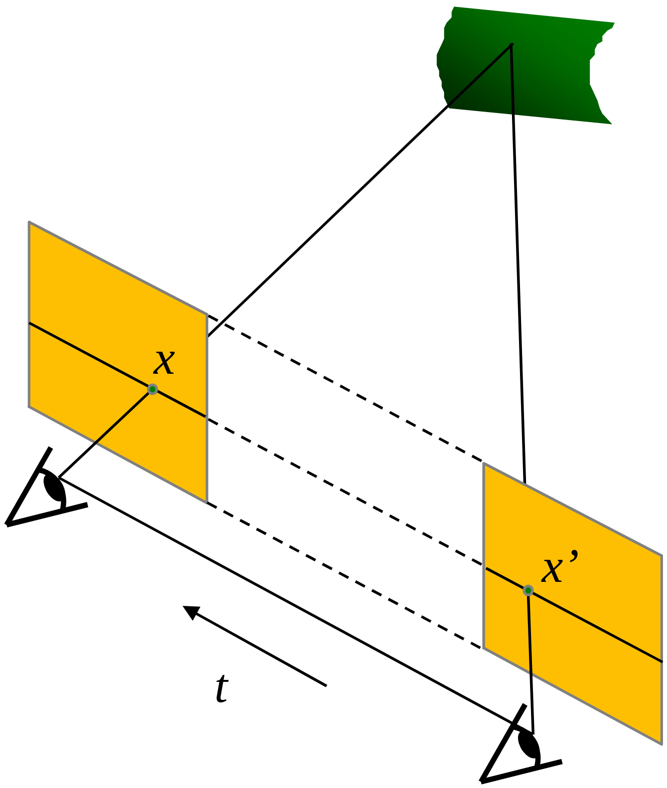
$$x^T E x' = 0, \quad E = [t_\times]R$$

$$R = I \qquad t = (T, 0, 0)$$

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

$$[a_\times] = \begin{bmatrix} 0 & -a_z & a_y \\ a_z & 0 & -a_x \\ -a_y & a_x & 0 \end{bmatrix}$$

Essential matrix for parallel images



Epipolar constraint:

$$x^T E x' = 0, \quad E = [t_\times]R$$

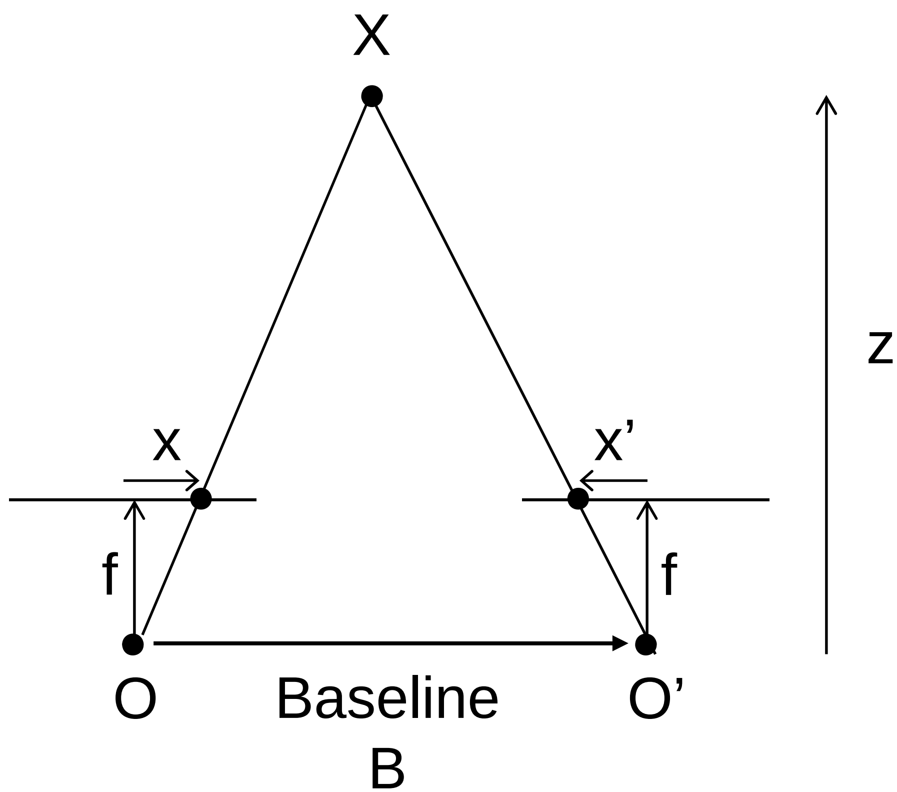
$$R = I \qquad t = (T, 0, 0)$$

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

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

The y-coordinates of corresponding points are the same!

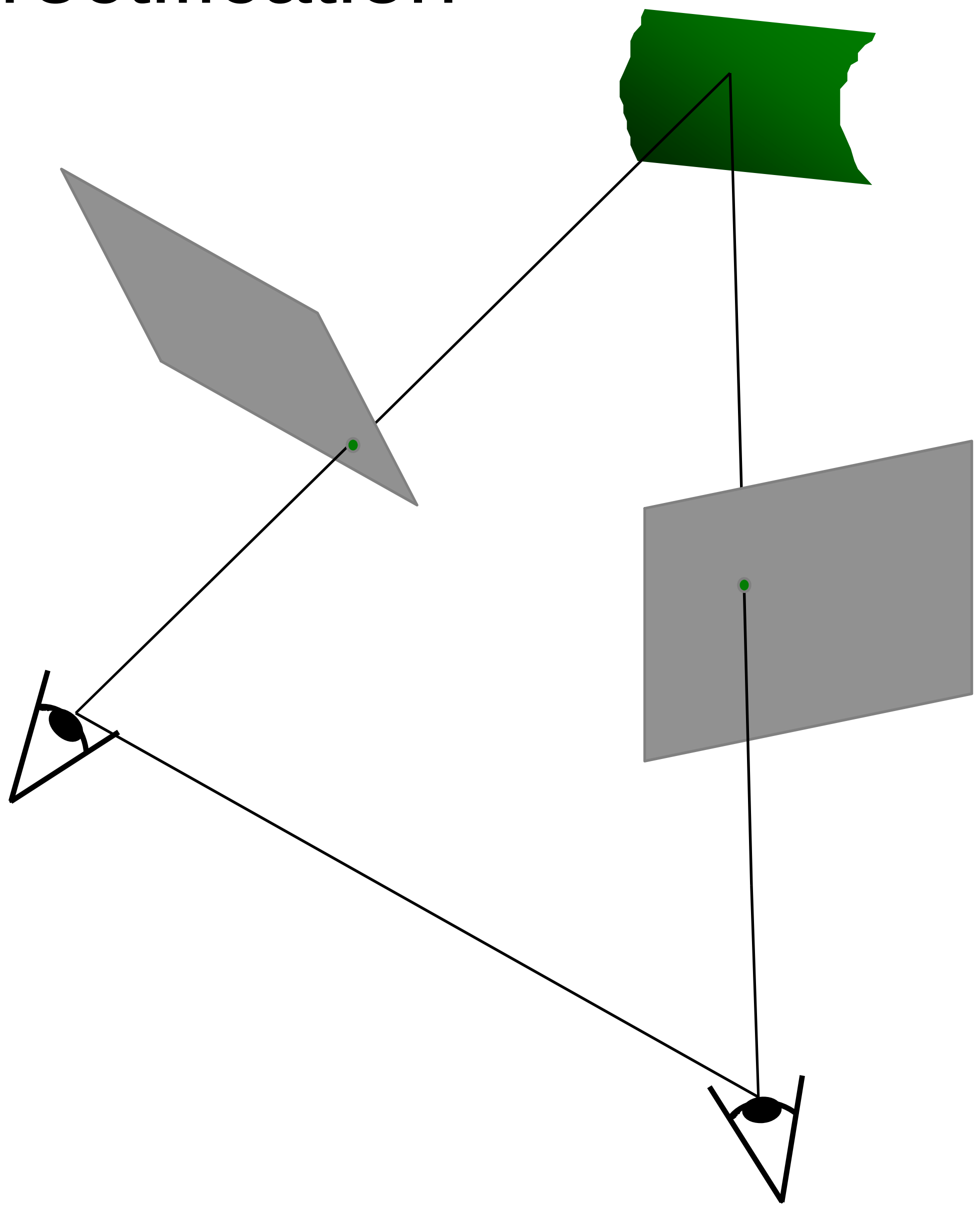
Depth from disparity



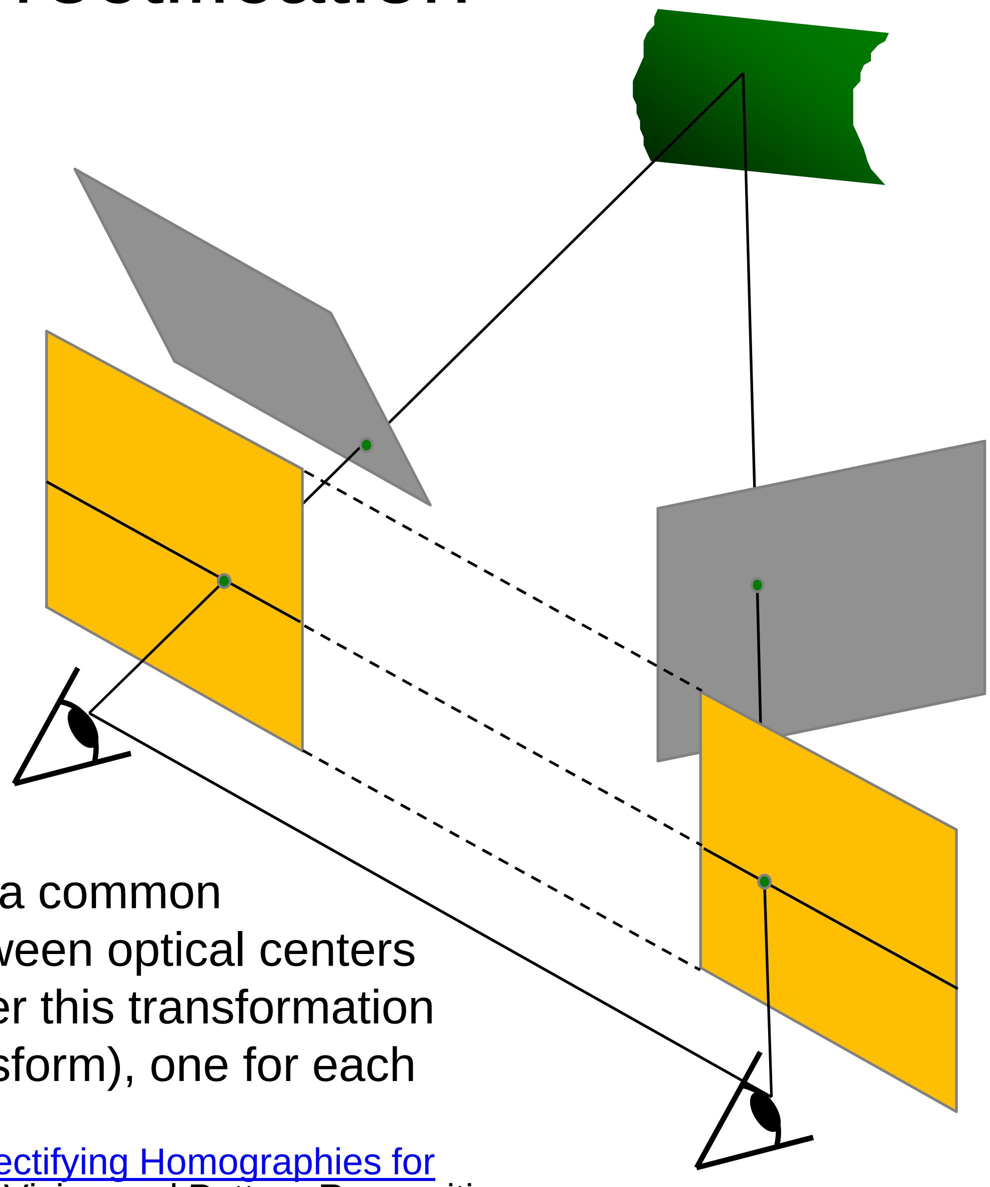
$$disparity = x - x' = \frac{B \cdot f}{z}$$

Disparity is inversely proportional to depth!

Stereo image rectification

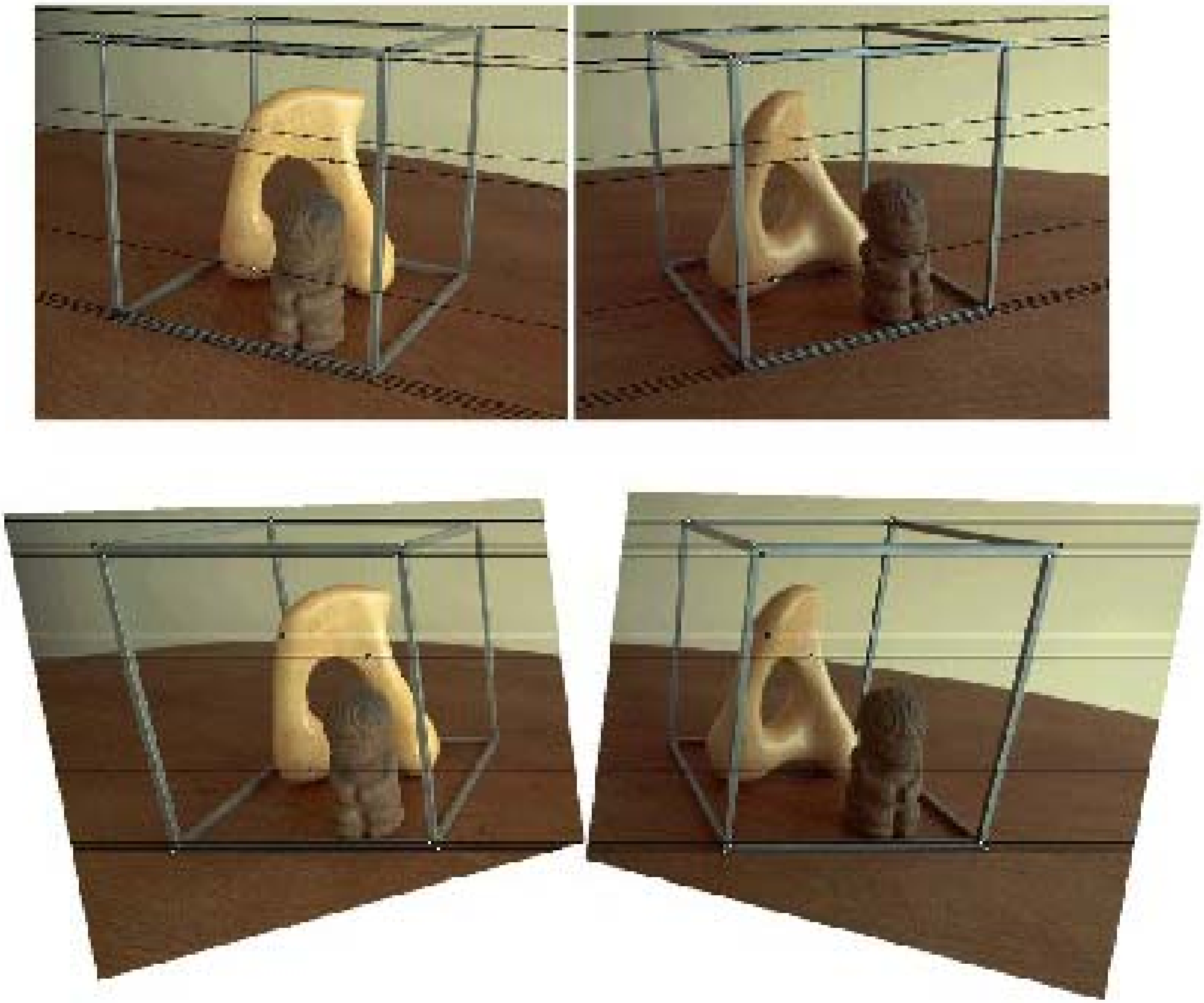


Stereo image rectification

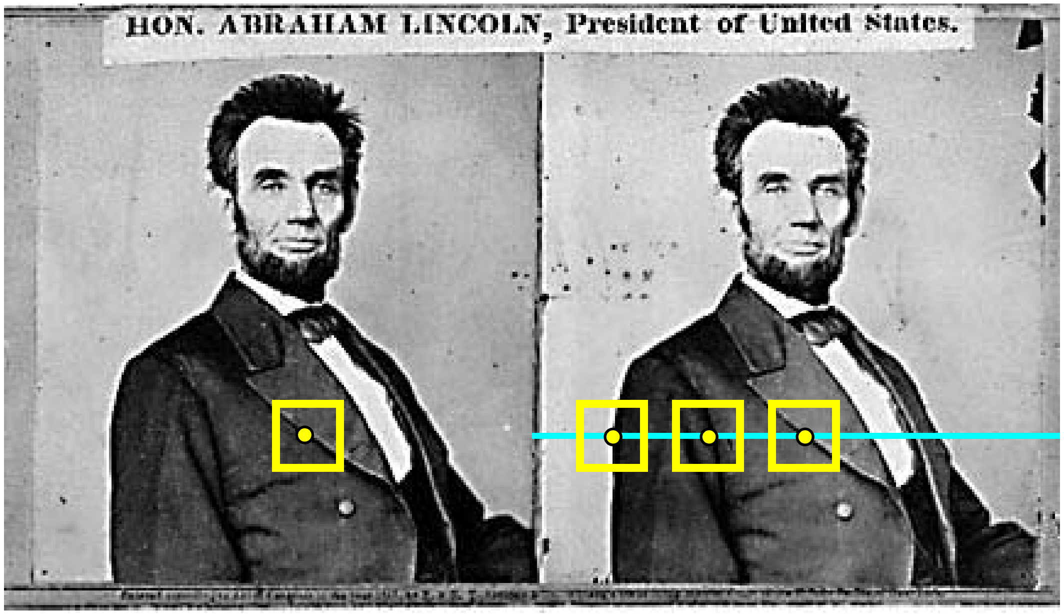


- reproject image planes onto a common plane parallel to the line between optical centers
- pixel motion is horizontal after this transformation
- two homographies (3x3 transform), one for each input image reprojection
  - C. Loop and Z. Zhang. [Computing Rectifying Homographies for Stereo Vision]().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

Rectification example

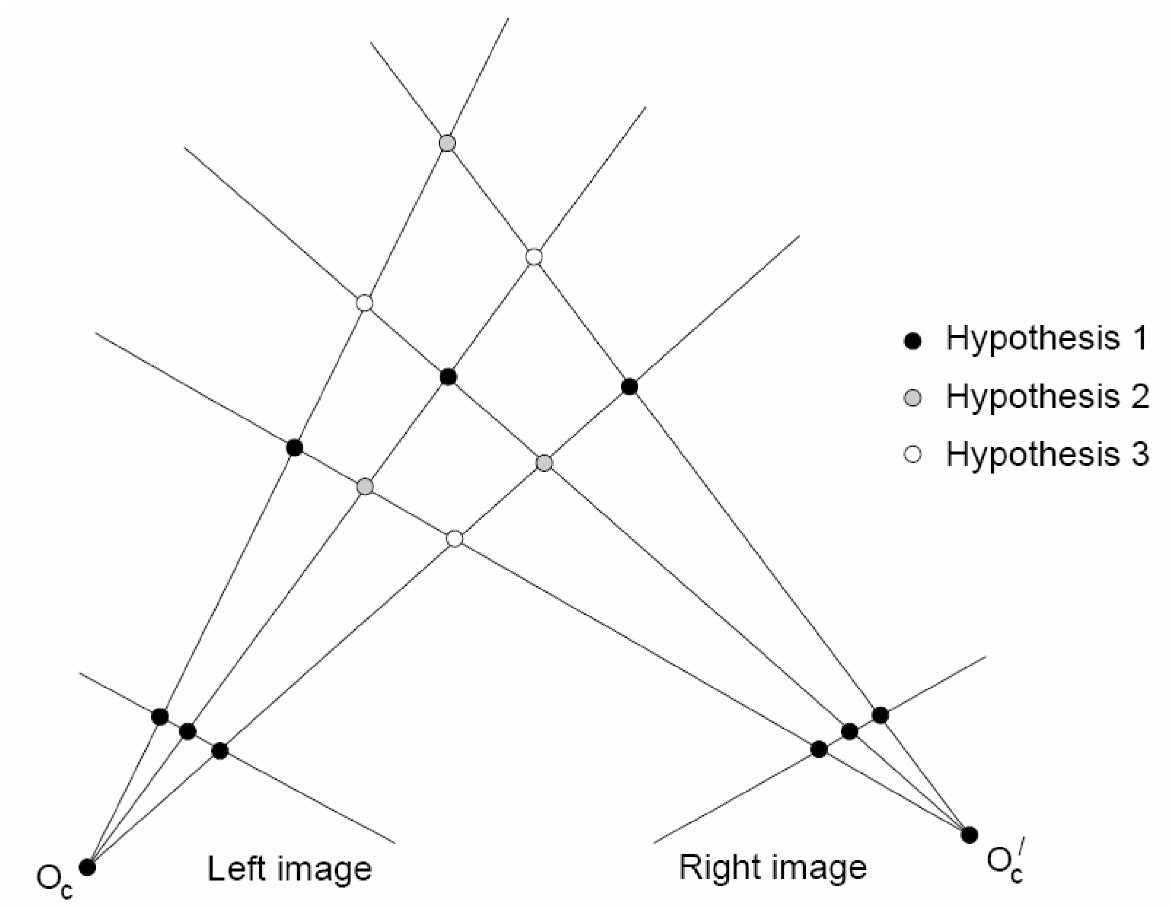


Basic stereo matching algorithm



- 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 depth(x) = 1/(x-x’)

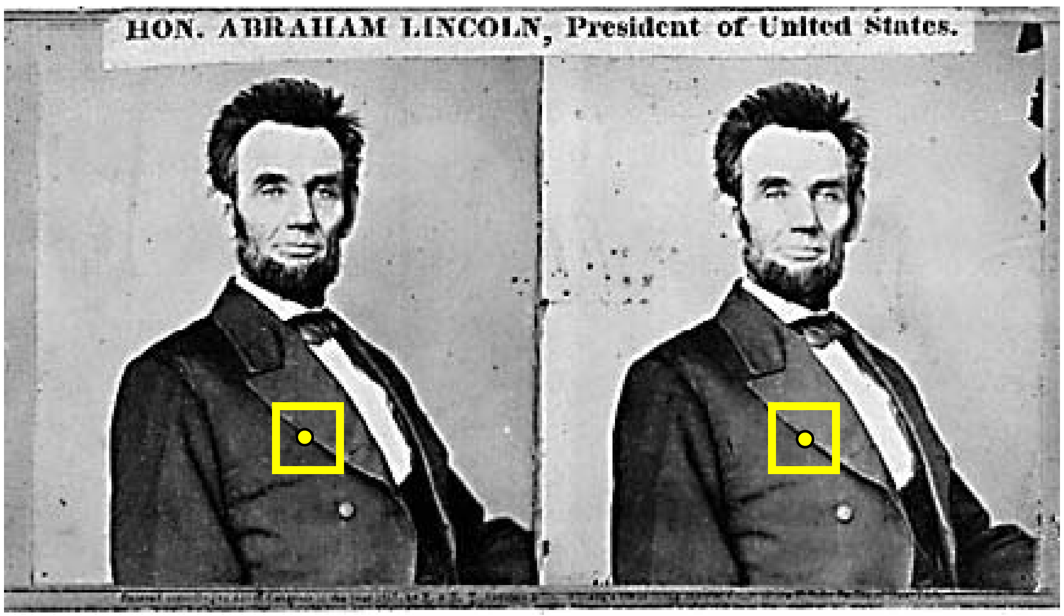
Correspondence problem



Multiple matching hypotheses satisfy the epipolar constraint, but which one is correct?

Correspondence problem

- Let's make some assumptions to simplify the matching problem
  - The baseline is relatively small (compared to the depth of scene points)
  - Then most scene points are visible in both views
  - Also, matching regions are similar in appearance

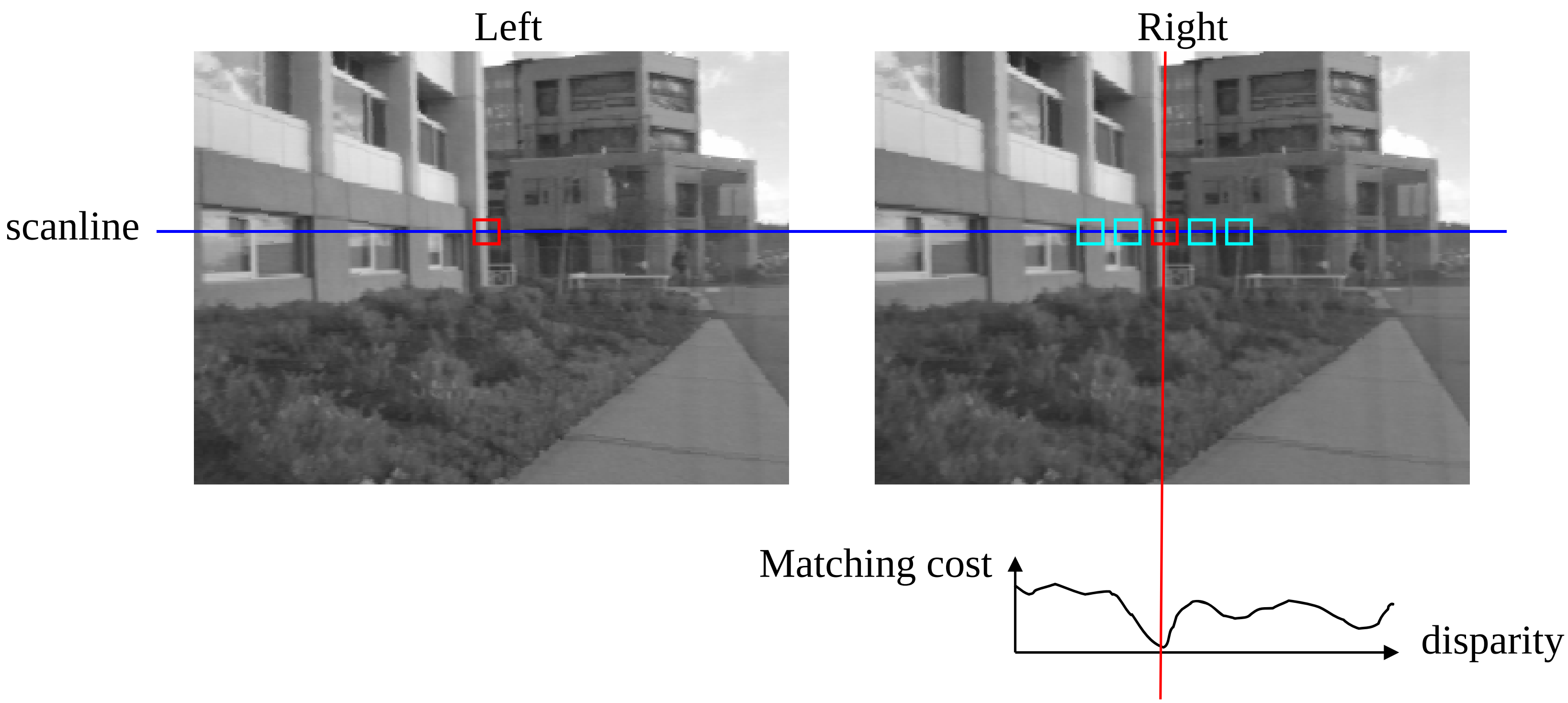


Correspondence problem

- Let's make some assumptions to simplify the matching problem
  - The baseline is relatively small (compared to the depth of scene points)
  - Then most scene points are visible in both views
  - Also, matching regions are similar in appearance

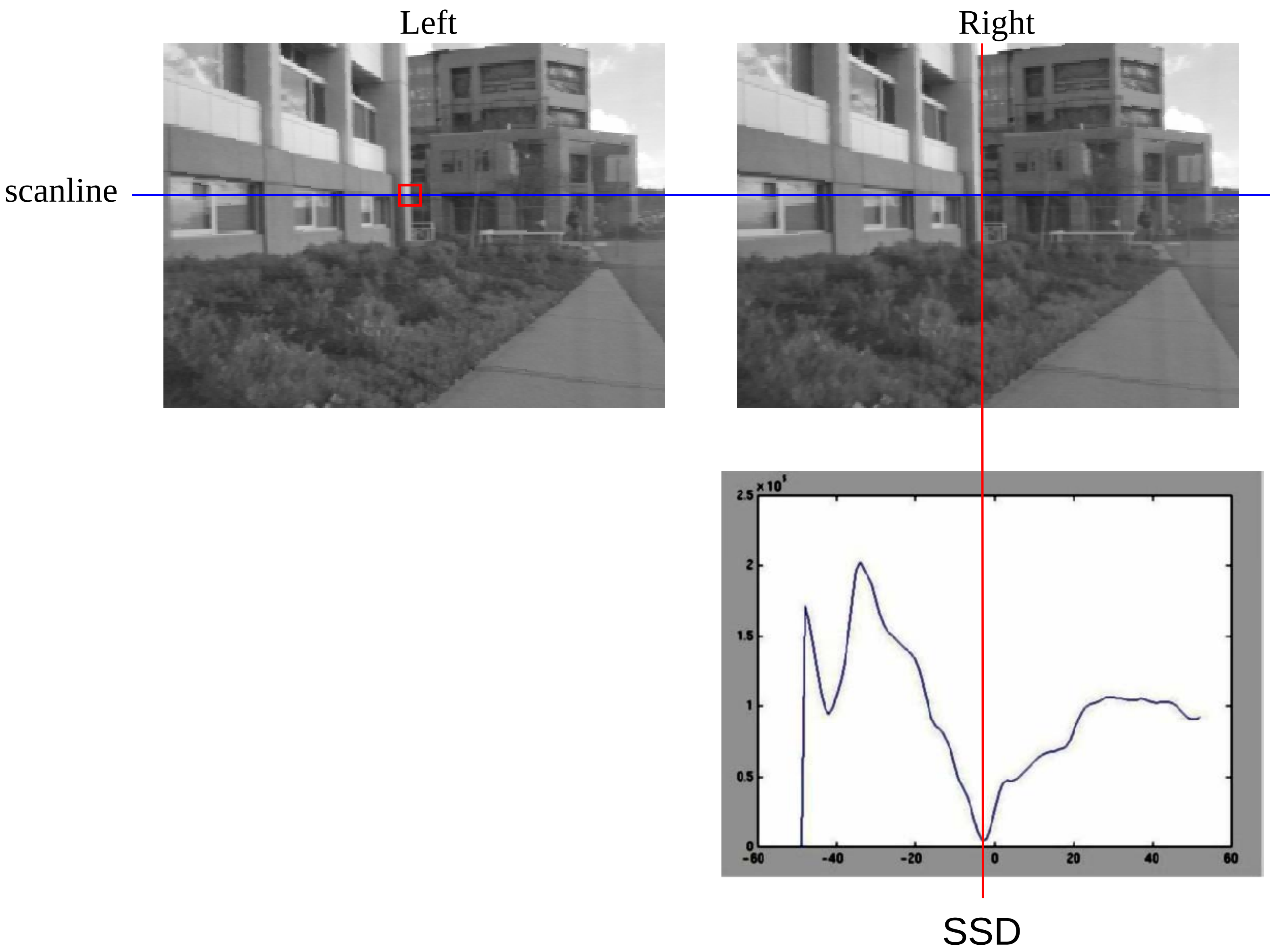
- Additional constraints
  - Uniqueness
  - Ordering
  - Continuity

Correspondence search with similarity constraint

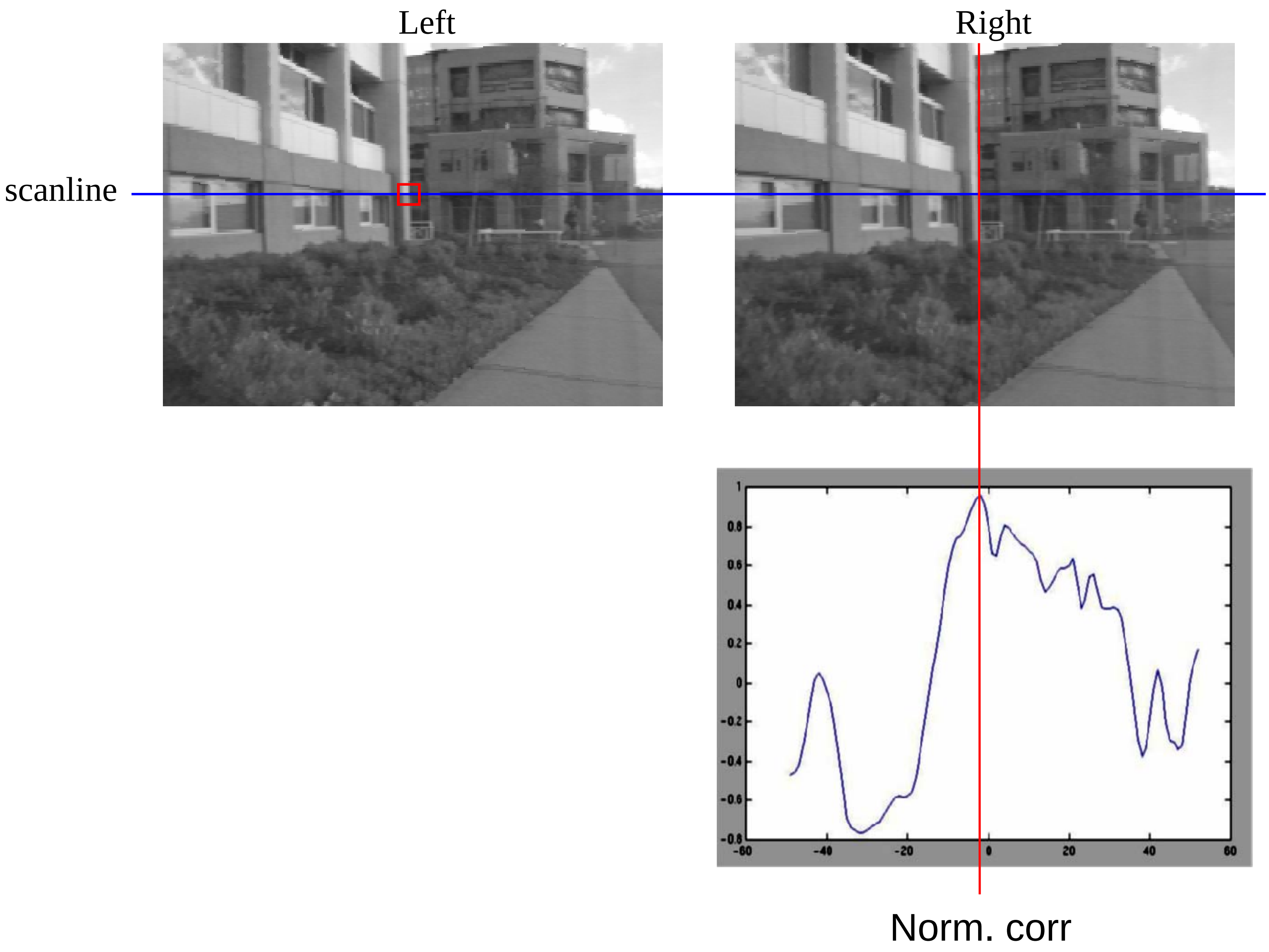


- Slide a window along the right scanline and compare contents of that window with the reference window in the left image
- Matching cost: SSD or normalized correlation

Correspondence search with similarity constraint

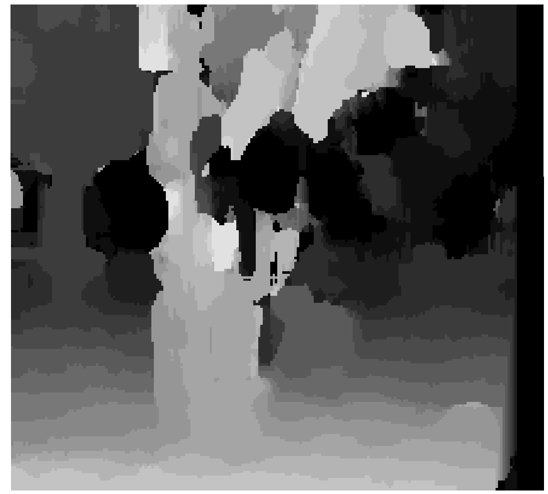


Correspondence search with similarity constraint



Effect of window size

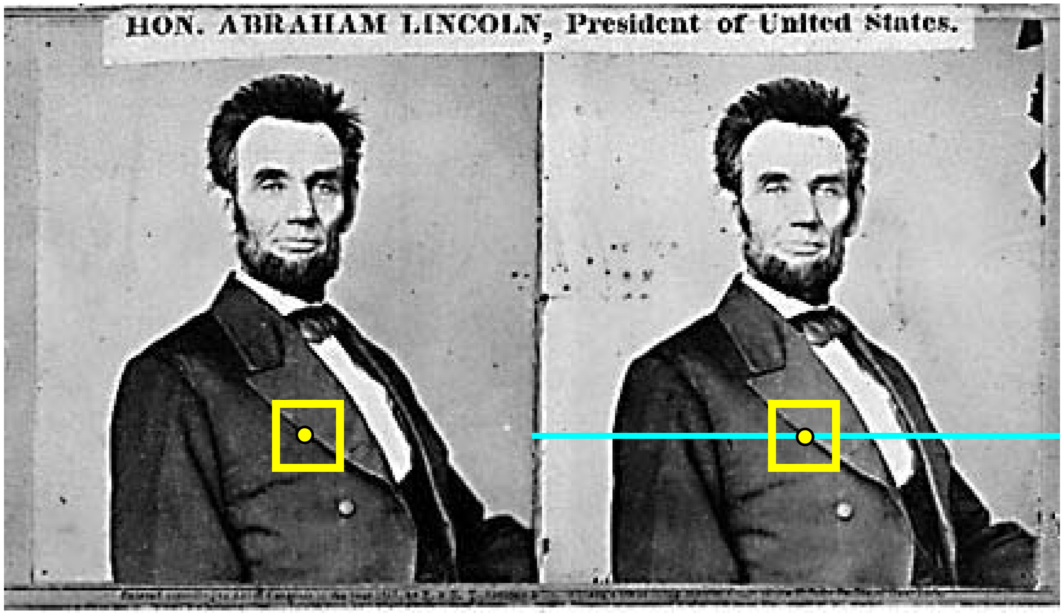
W = 3



W = 20

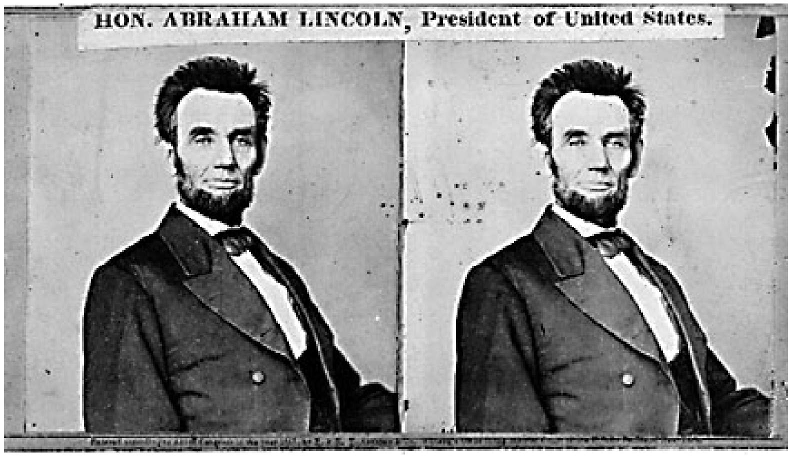
- Smaller window
  - \+ More detail
  - – More noise
- Larger window
  - \+ Smoother disparity maps
  - – Less detail

The similarity constraint



- Corresponding regions in two images should be similar in appearance
- …and non-corresponding regions should be different
- When will the similarity constraint fail?

Limitations of similarity constraint



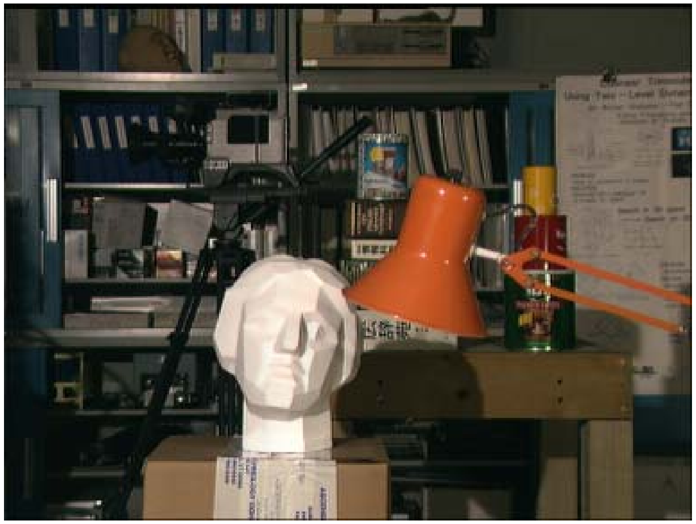
Textureless surfaces

Occlusions, repetition

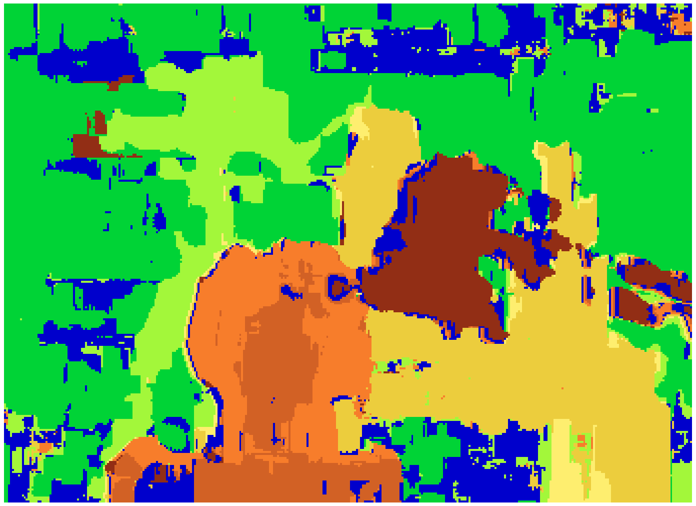
Non-Lambertian surfaces, specularities

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

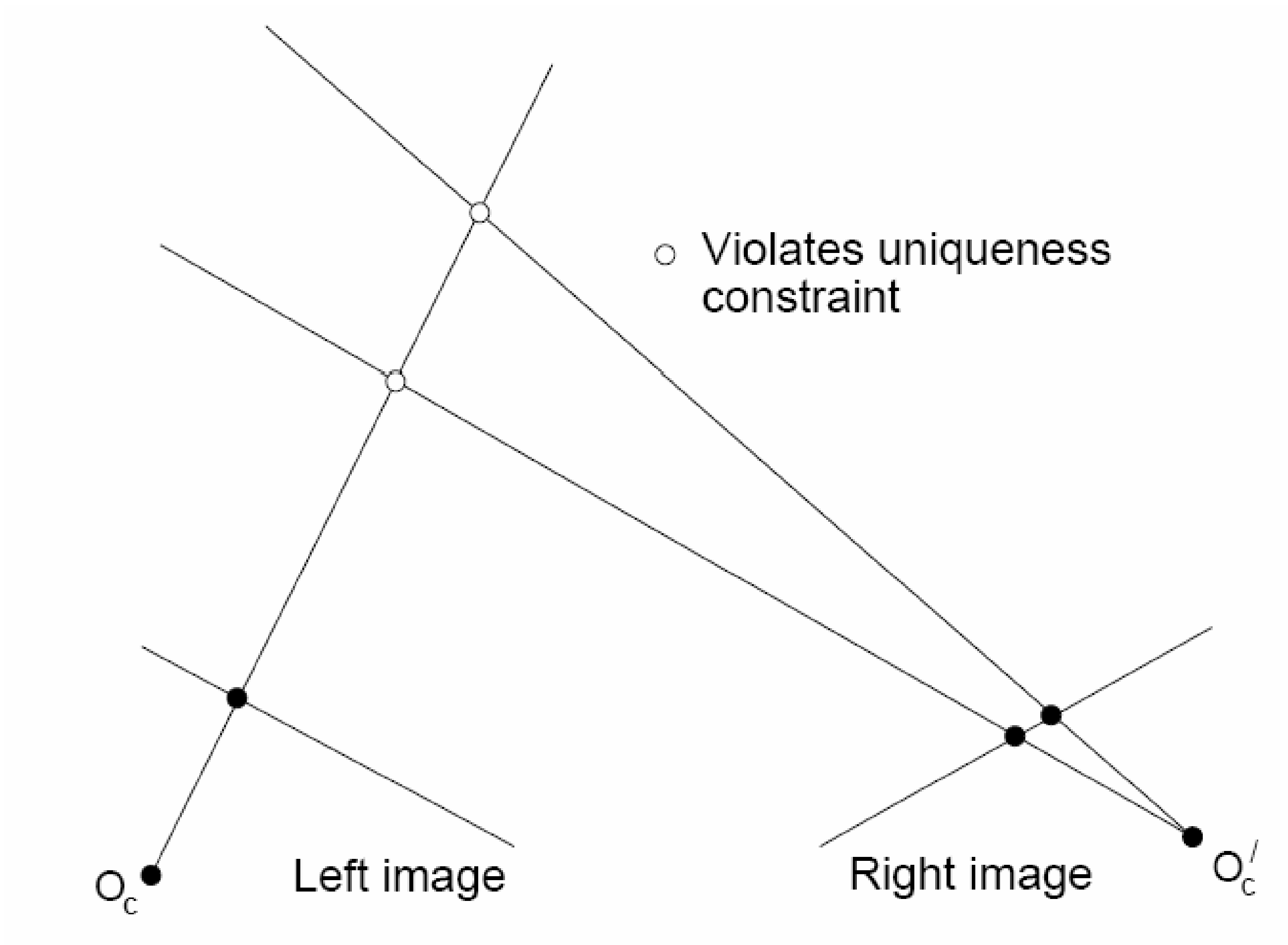
For the latest and greatest: [http://www.middlebury.edu/stereo/]()

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

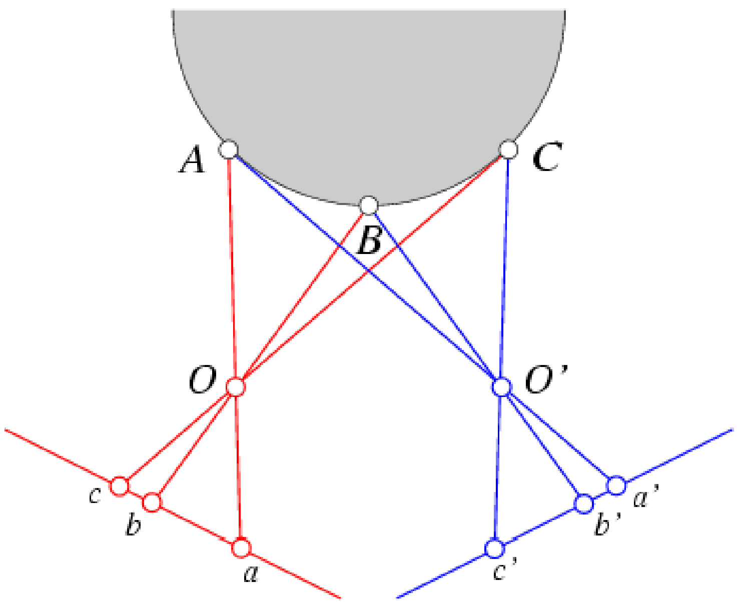
Non-local constraints

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



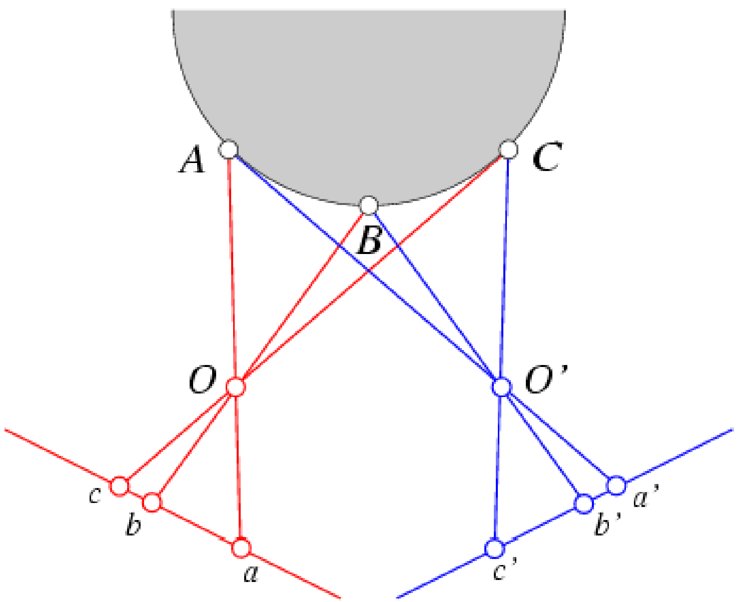
Non-local constraints

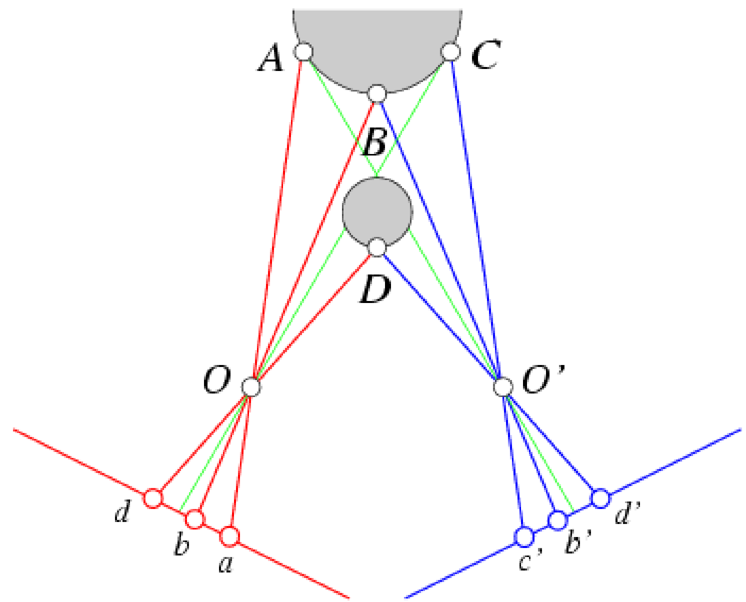
- Uniqueness
  - For any point in one image, there should be at most one matching point in the other image
- Ordering
  - Corresponding points should be in the same order in both views



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





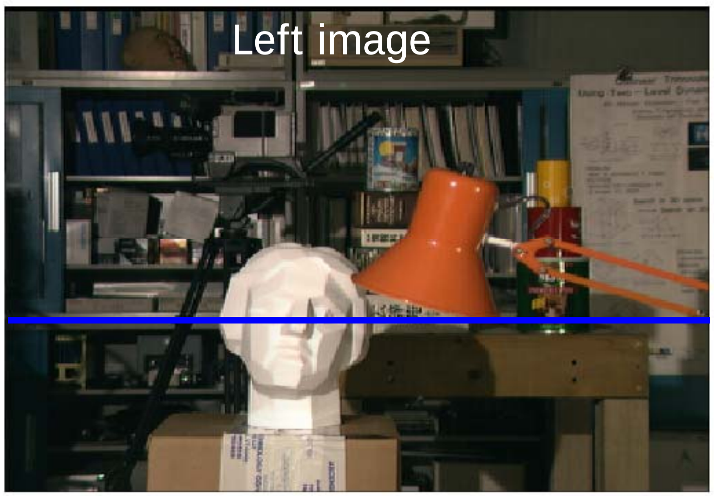
Ordering constraint doesn't hold

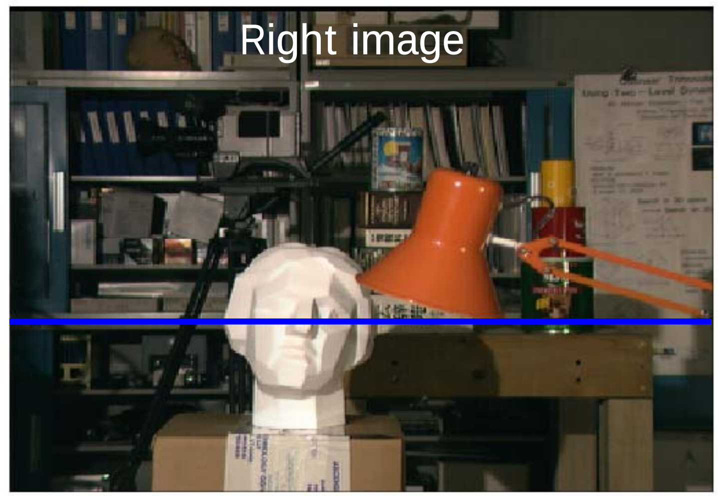
Non-local constraints

- 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)

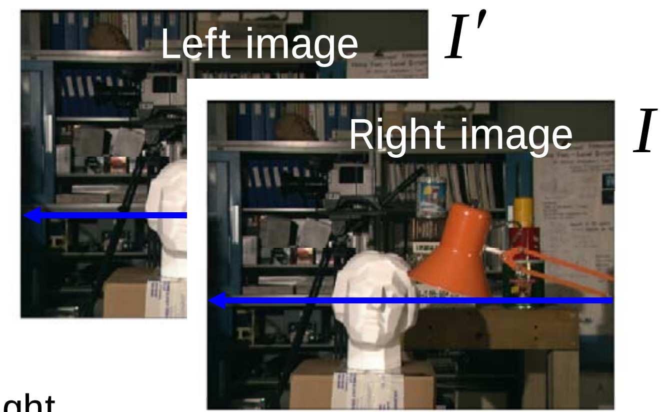
Scanline stereo

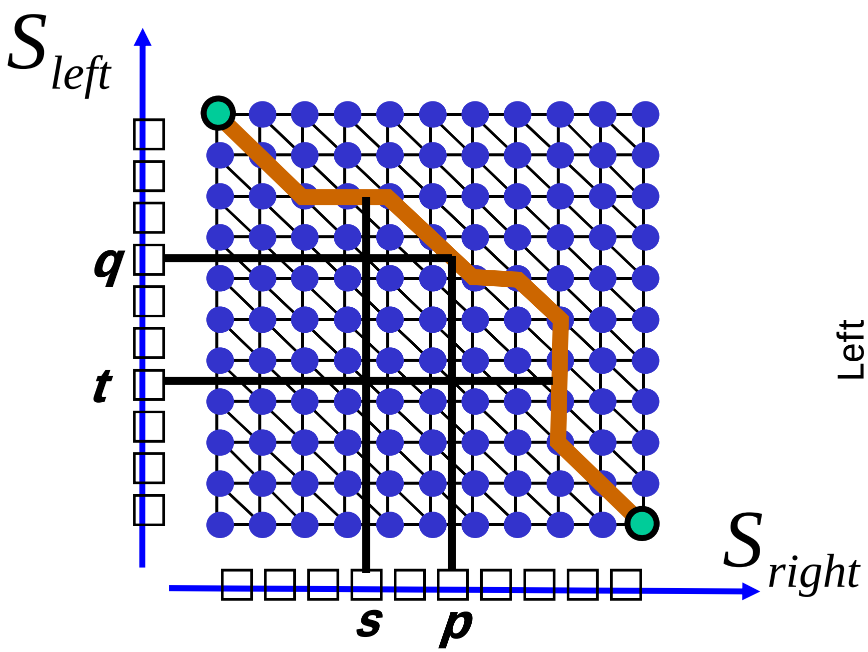
- Try to coherently match pixels on the entire scanline
- Different scanlines are still optimized independently

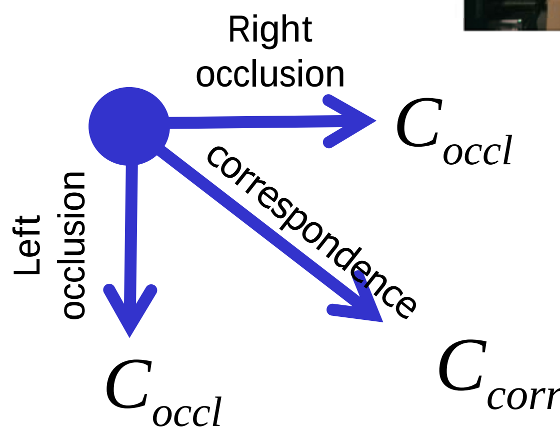




“Shortest paths” for scan-line stereo







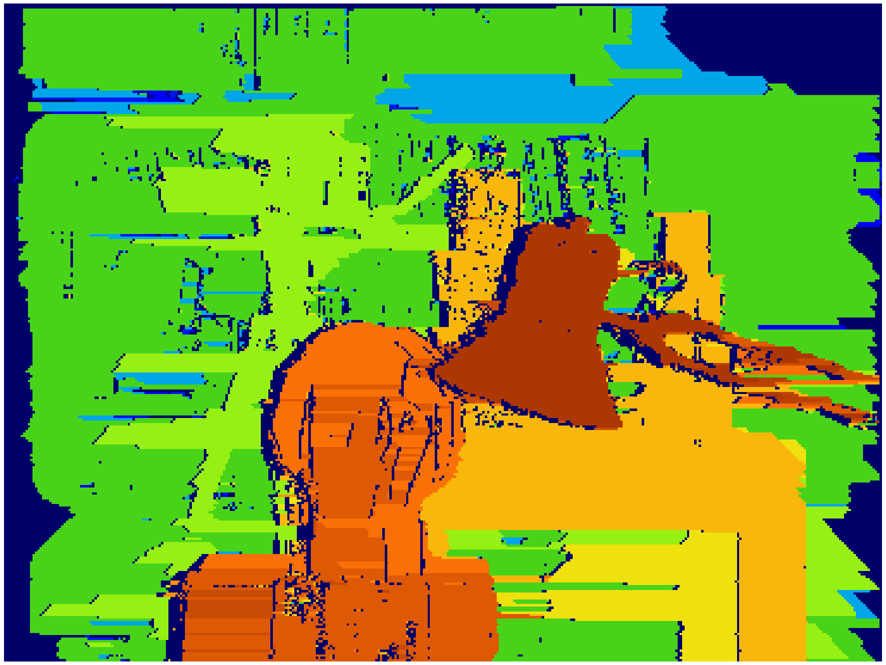
Can be implemented with dynamic programming

Ohta & Kanade ’85, Cox et al. ‘96

Slide credit: Y. Boykov

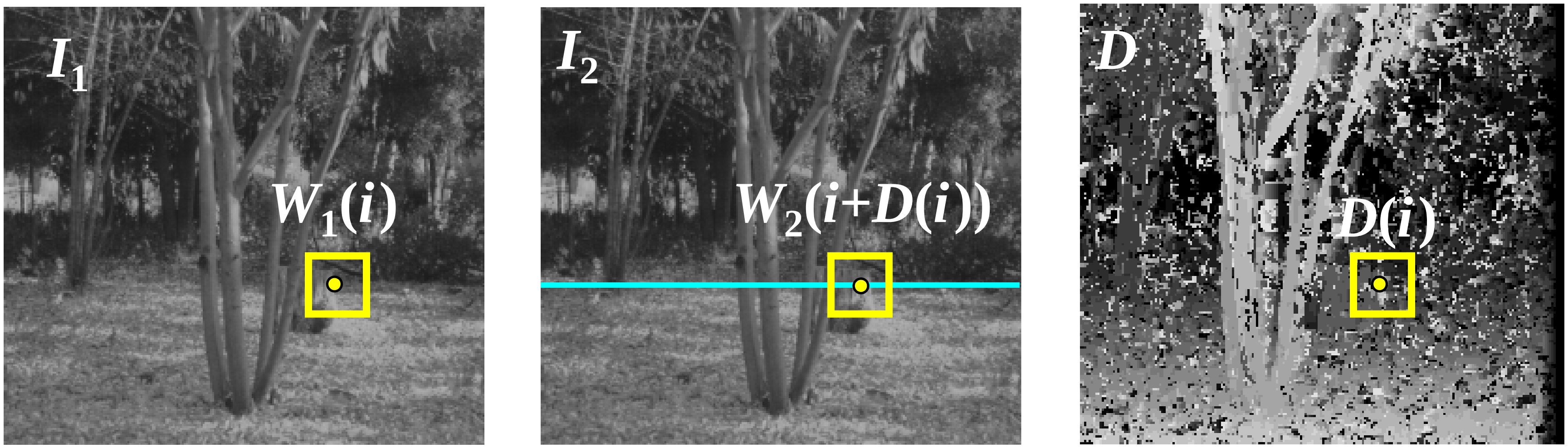
Coherent stereo on 2D grid

- 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



MAP estimate of disparity image *D*: $P(D \mid I_1, I_2) \propto P(I_1, I_2 \mid D) P(D)$

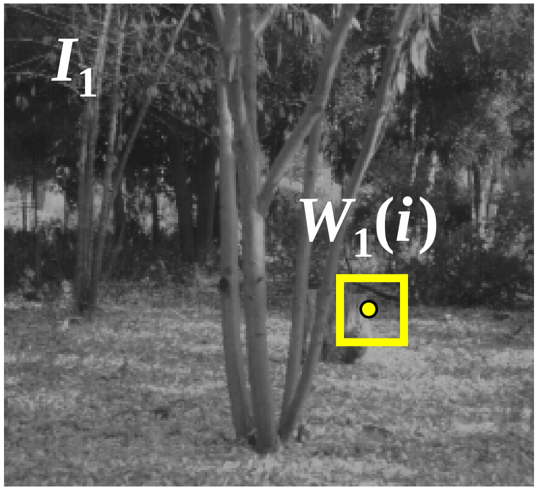
$$-\log P(D \mid I_1, I_2) \propto -\log P(I_1, I_2 \mid D) - \log P(D)$$

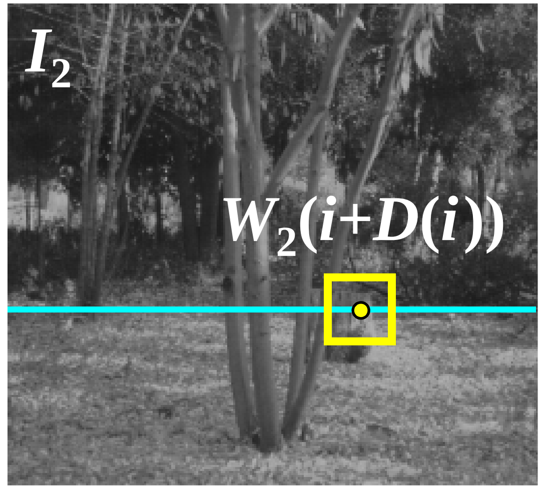
$$\boxed{E = \alpha E_{\text{data}}(I_1, I_2, D) + \beta E_{\text{smooth}}(D)}$$

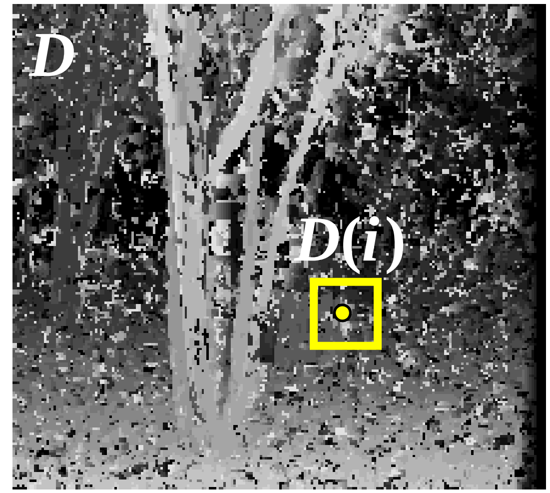
$$\boxed{E_{\text{data}} = \sum_i \left(W_1(i) - W_2(i + D(i))\right)^2}$$

$$\boxed{E_{\text{smooth}} = \sum_{\text{neighbors } i,j} \rho\left(D(i) - D(j)\right)}$$

Stereo matching as energy minimization







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

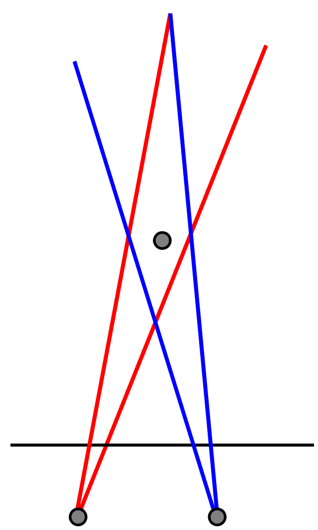
$$\boxed{E_{\text{data}} = \sum_i \left(W_1(i) - W_2(i + D(i))\right)^2}$$

$$\boxed{E_{\text{smooth}} = \sum_{\text{neighbors } i,j} \rho\left(D(i) - D(j)\right)}$$

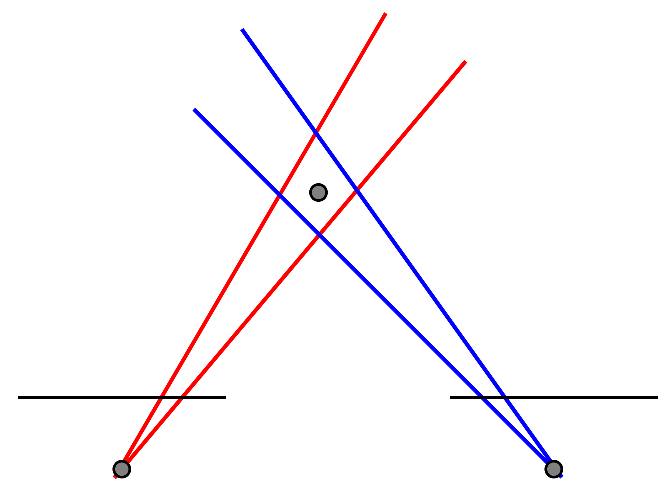
- Energy functions of this form can be minimized using *graph cuts*

Y. Boykov, O. Veksler, and R. Zabih, Fast Approximate Energy Minimization via Graph Cuts, PAMI 2001

The role of the baseline



Small Baseline



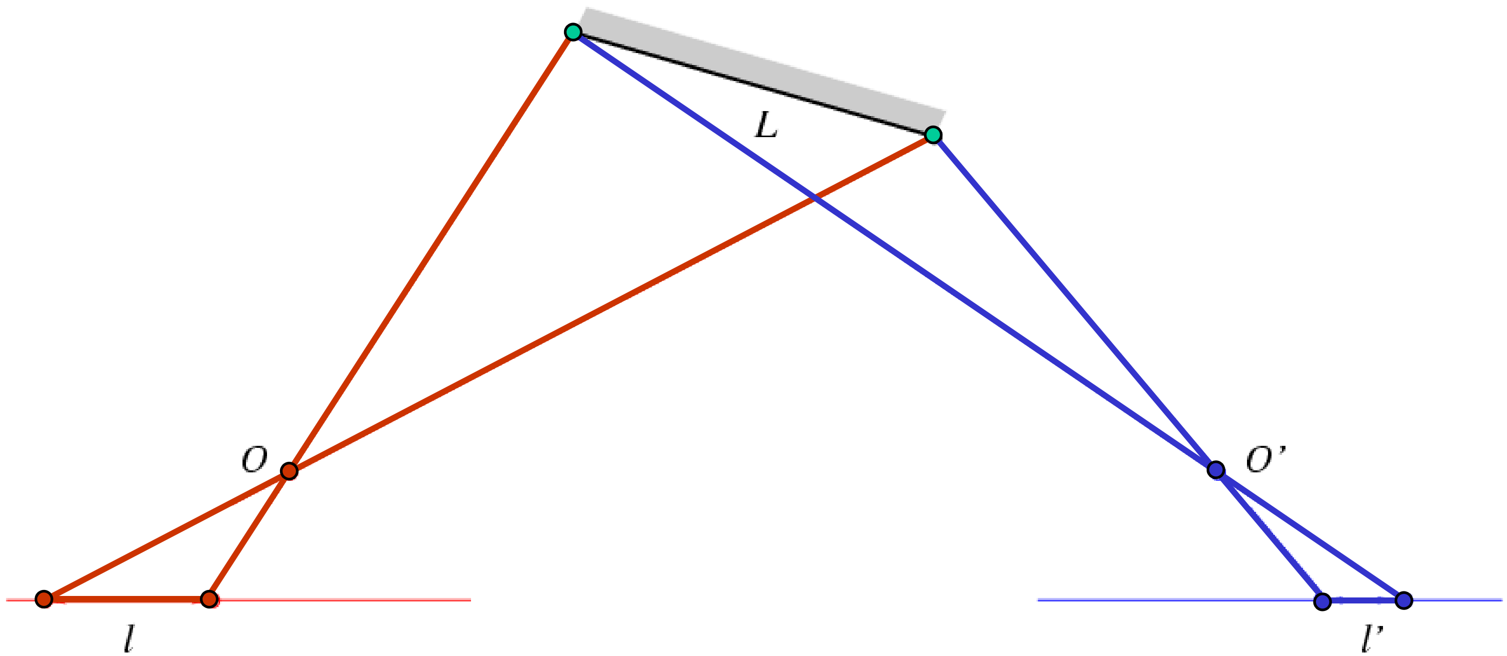
Large Baseline

Small baseline: large depth error

Large baseline: difficult search problem

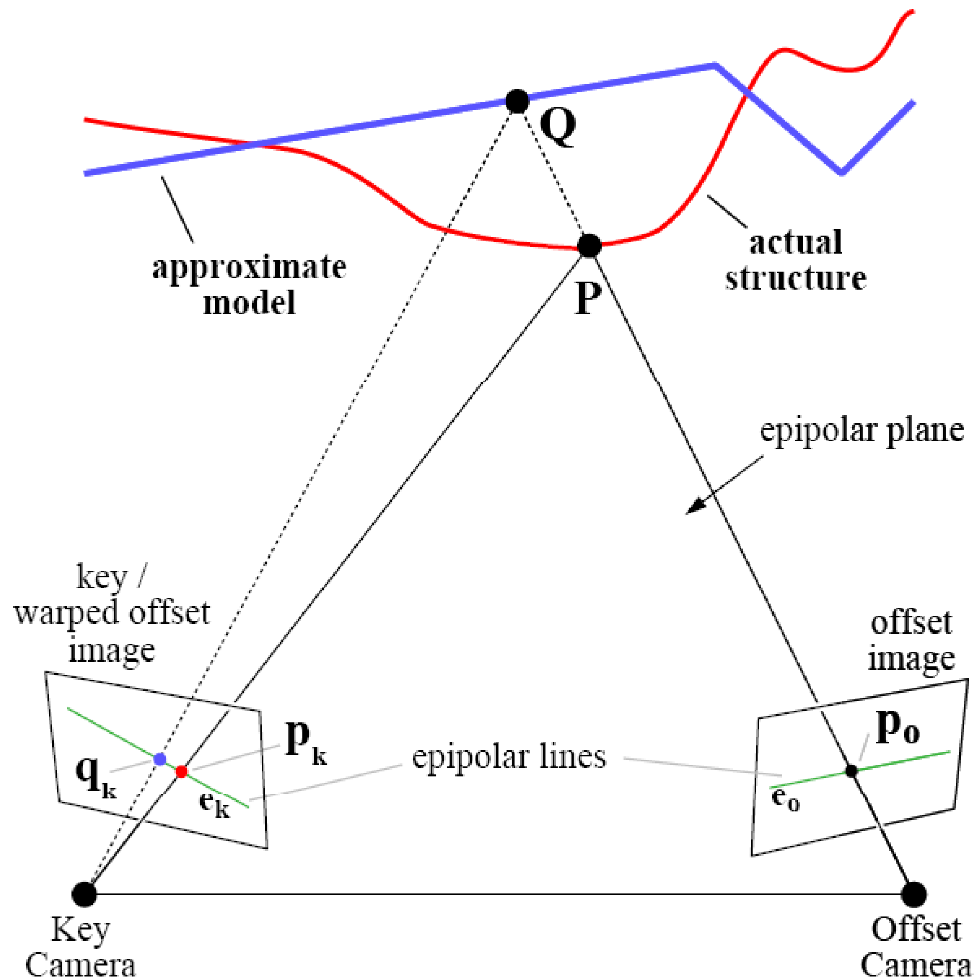
Source: S. Seitz

Problem for wide baselines: Foreshortening



- Matching with fixed-size windows will fail!
- Possible solution: adaptively vary window size
- Another solution: *model-based stereo*

Model-based stereo



Paul E. Debevec, Camillo J. Taylor, and Jitendra Malik. [Modeling and Rendering Architecture from Photographs.]() SIGGRAPH 1996.

Model-based stereo

key image

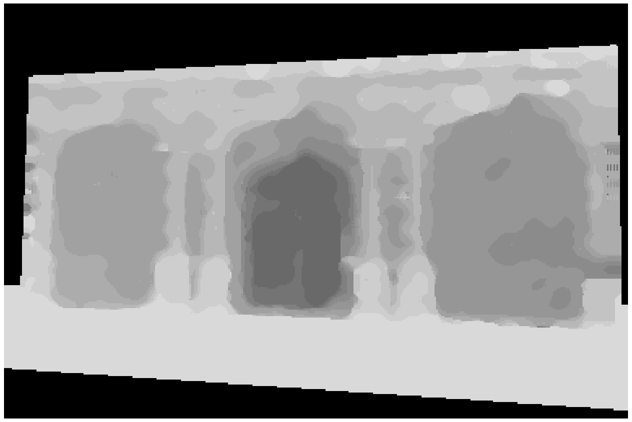
offset image

Paul E. Debevec, Camillo J. Taylor, and Jitendra Malik. Modeling and Rendering Architecture from Photographs. SIGGRAPH 1996.

Model-based stereo

key image

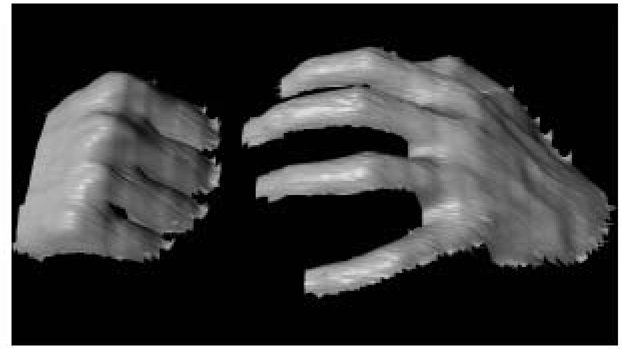
warped offset image



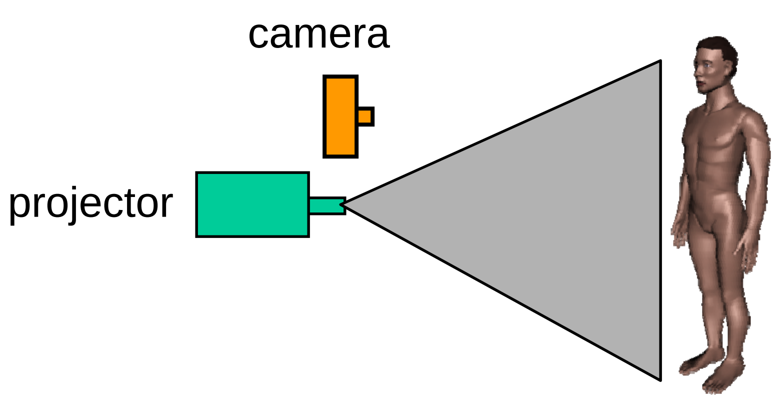
displacement map

Paul E. Debevec, Camillo J. Taylor, and Jitendra Malik. Modeling and Rendering Architecture from Photographs. SIGGRAPH 1996.

Active stereo with structured light

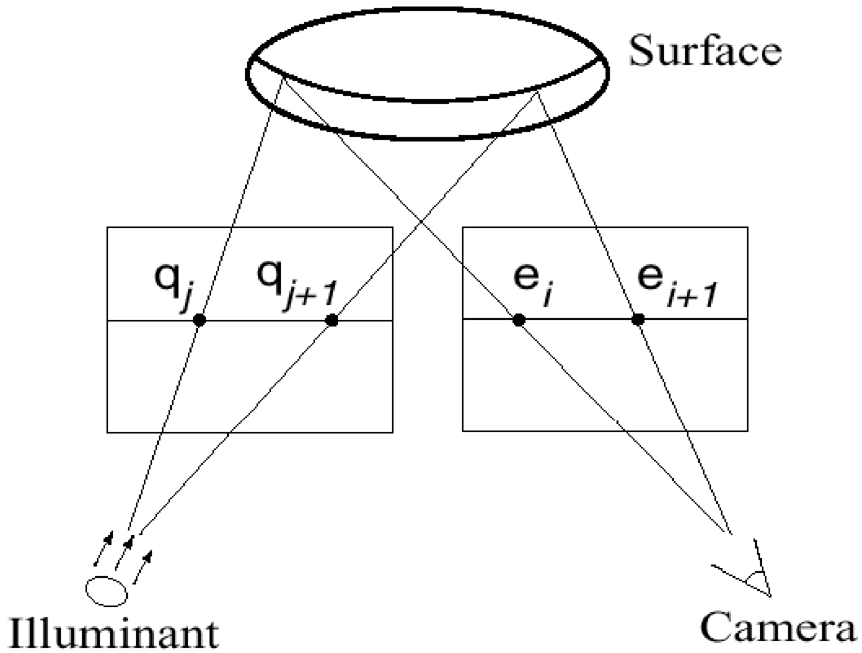


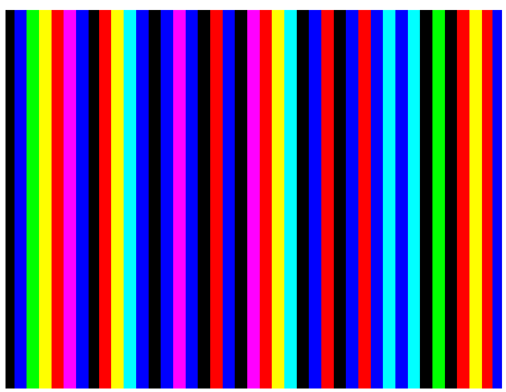
- Project "structured" light patterns onto the object
  - simplifies the correspondence problem
  - Allows us to use only one camera

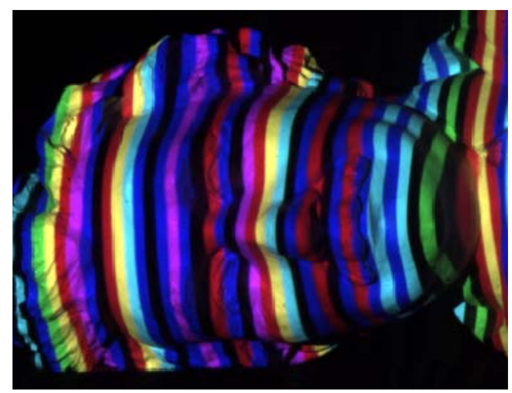


L. Zhang, B. Curless, and S. M. Seitz. [Rapid Shape Acquisition Using Color Structured Light and Multi-pass Dynamic Programming.]() *3DPVT* 2002

Active stereo with structured light

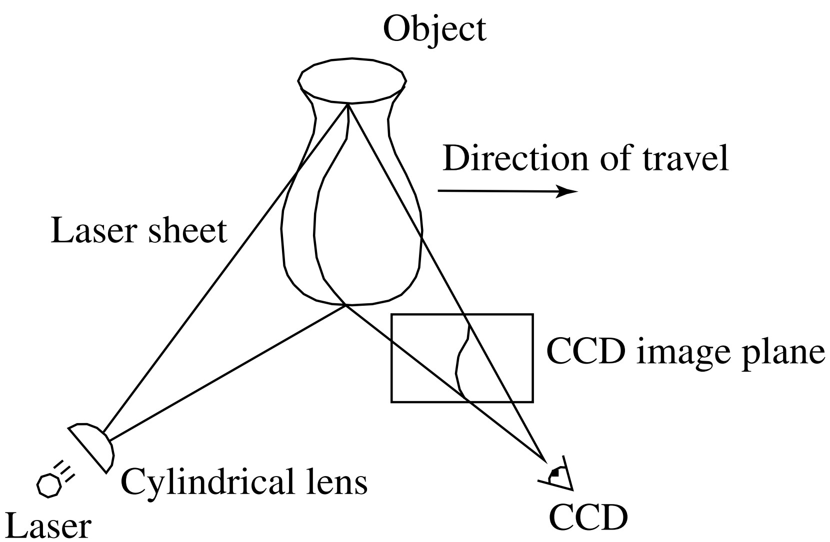


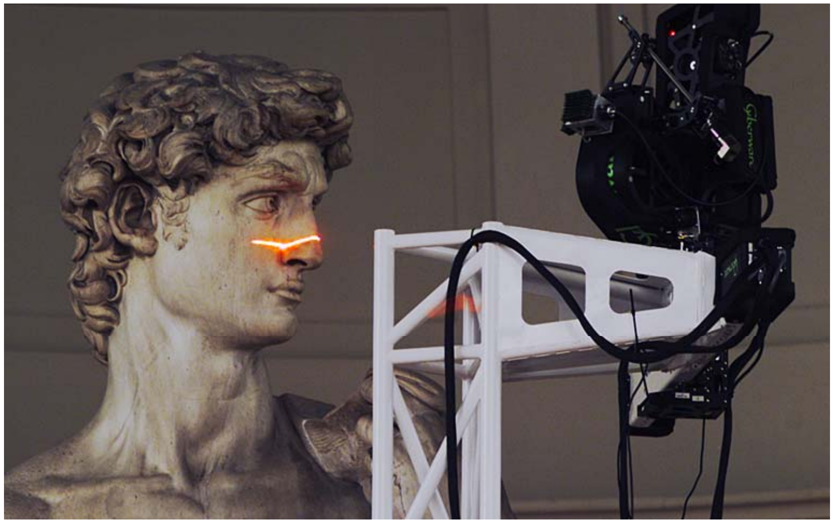




L. Zhang, B. Curless, and S. M. Seitz. [Rapid Shape Acquisition Using Color Structured Light and Multi-pass Dynamic Programming.] *3DPVT* 2002

Laser scanning





Digital Michelangelo Project
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

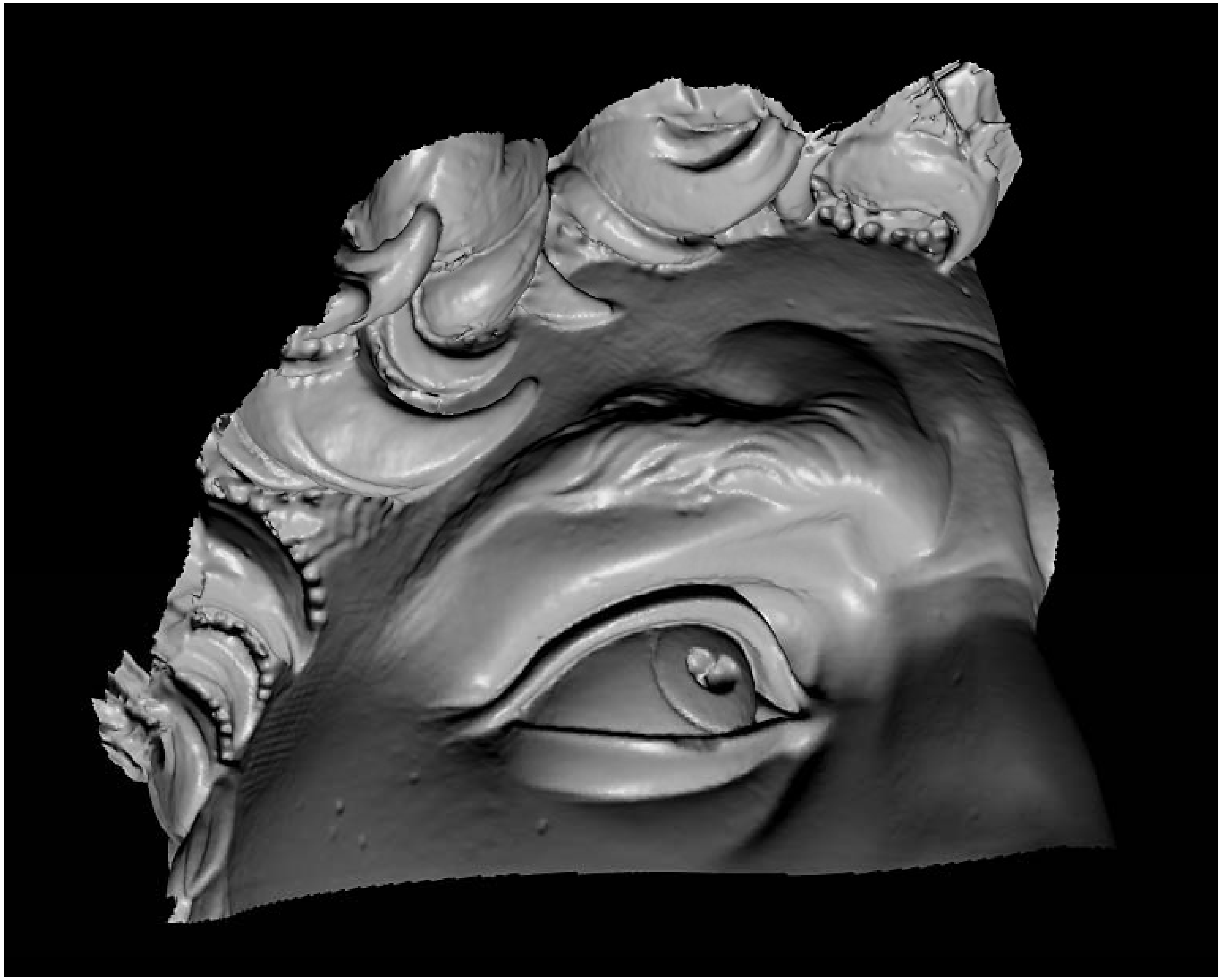
Source: S. Seitz

Laser scanned models

The Digital Michelangelo Project, Levoy et al.

Source: S. Seitz

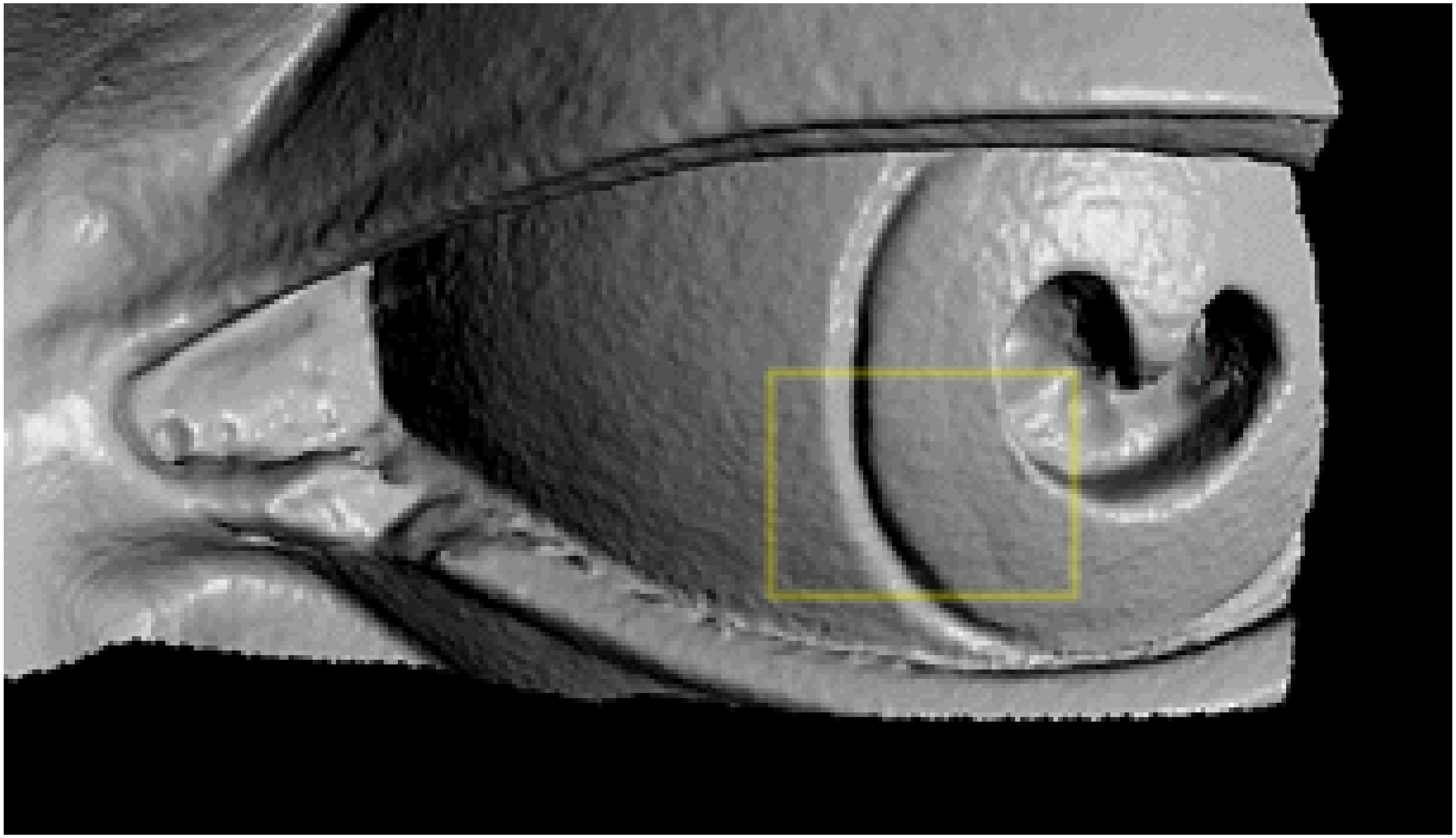
Laser scanned models



The Digital Michelangelo Project, Levoy et al.

Source: S. Seitz

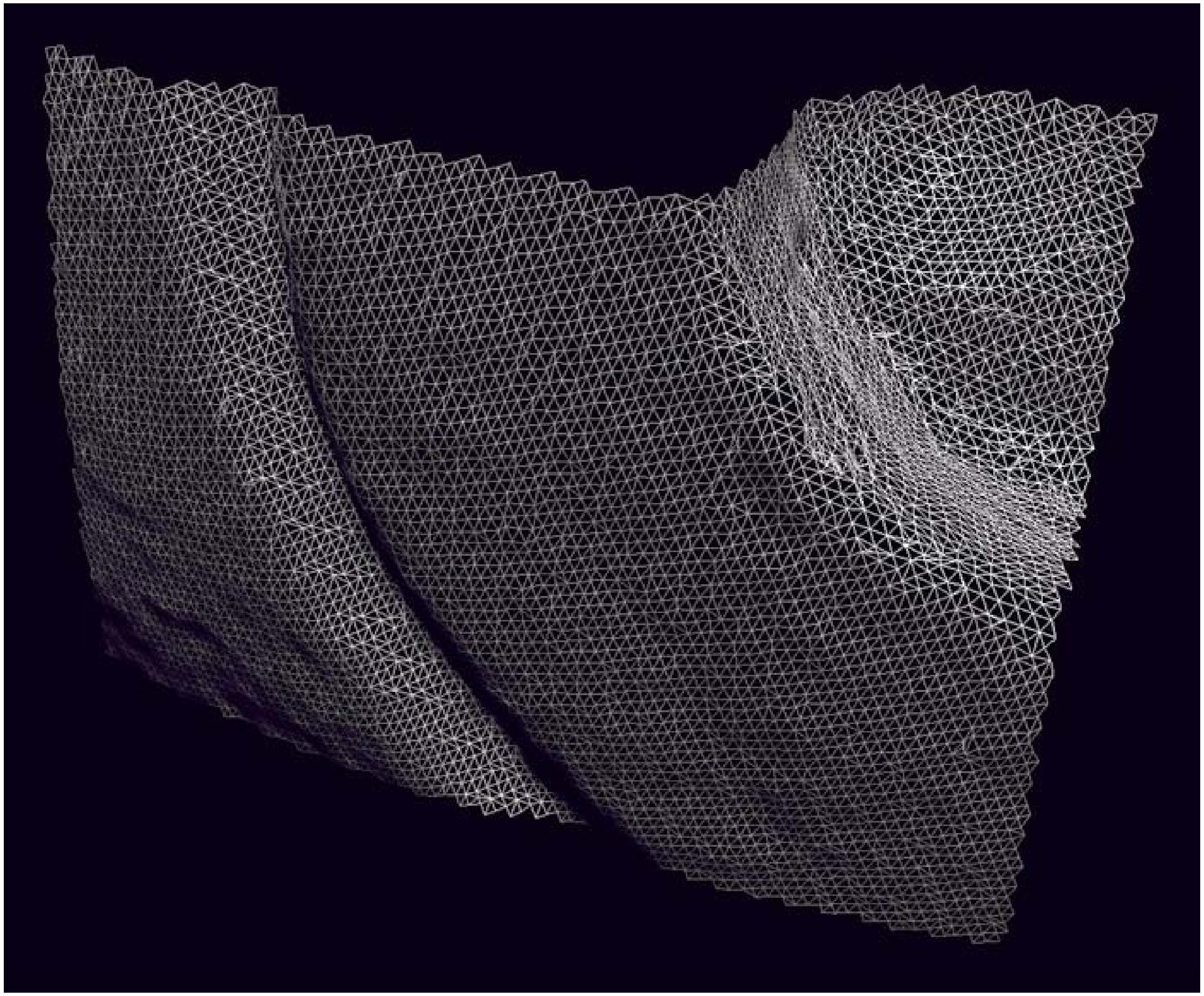
Laser scanned models



The Digital Michelangelo Project, Levoy et al.

Source: S. Seitz

Laser scanned models



The Digital Michelangelo Project, Levoy et al.

Source: S. Seitz