

Two-View Stereo



Many slides adapted from Steve Seitz

Problem formulation

- **Given:** stereo pair (assumed calibrated)
- **Wanted:** dense depth map



Outline

- Motivation and history
- Basic two-view stereo setup
- Local stereo matching algorithm
- Beyond local stereo matching
- Active stereo with structured light

Stereo vision and perception of depth

- What cues tell us about scene depth?



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

MAR 07, 2018

RON RISMAN

PetaPixel



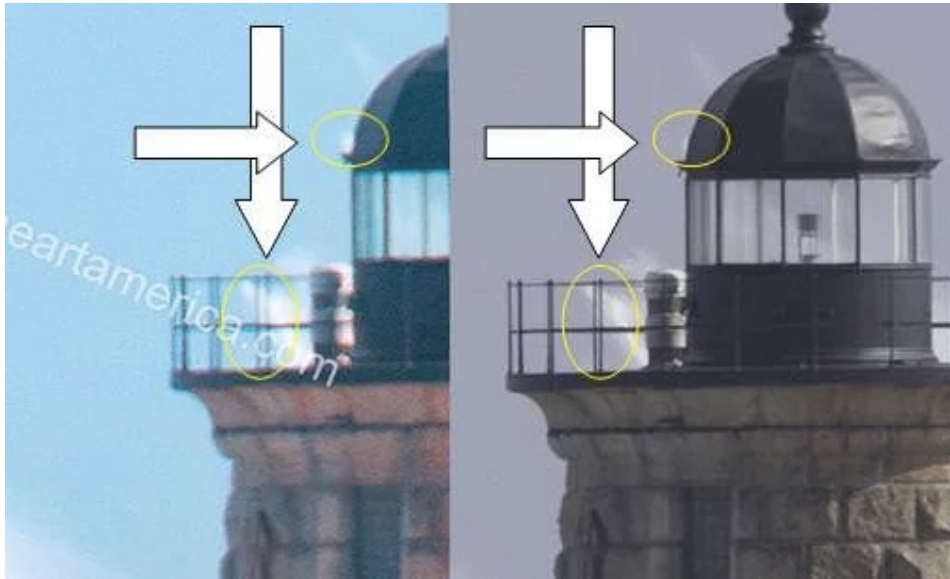
<https://petapixel.com/2018/03/07/two-photographers-unknowingly-shot-millisecond-time/>

How Two Photographers Unknowingly Shot the Same Millisecond in Time

MAR 07, 2018

RON RISMAN

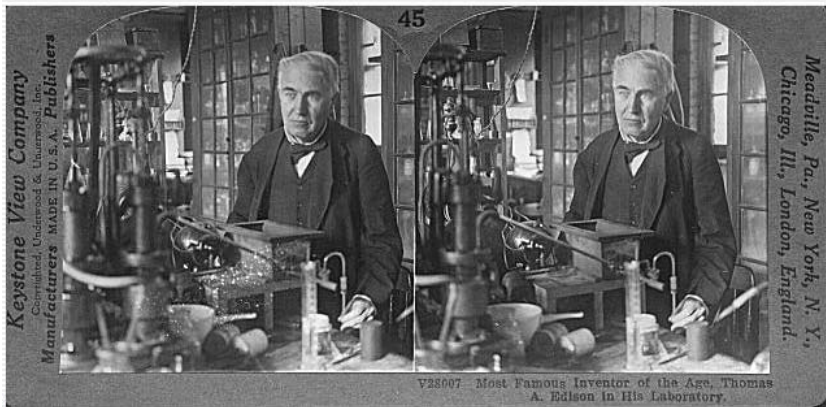
PetaPixel



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

History: Stereograms

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

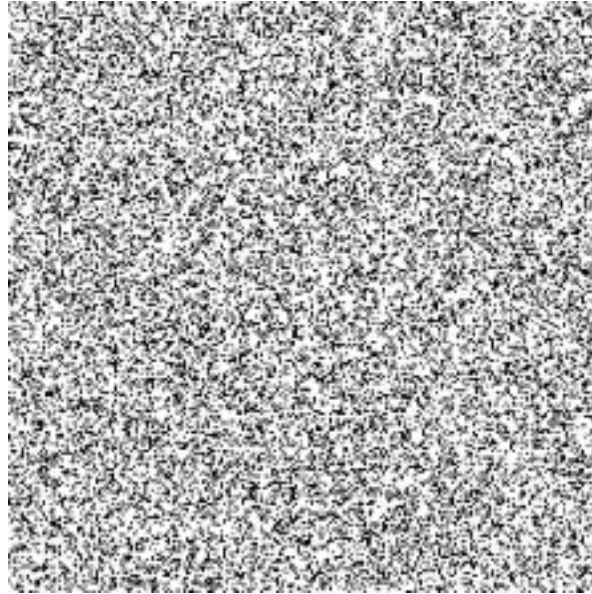


Stereograms: Invented by Sir Charles Wheatstone, 1838

<https://en.wikipedia.org/wiki/Stereoscopy>

History: Random dot stereograms

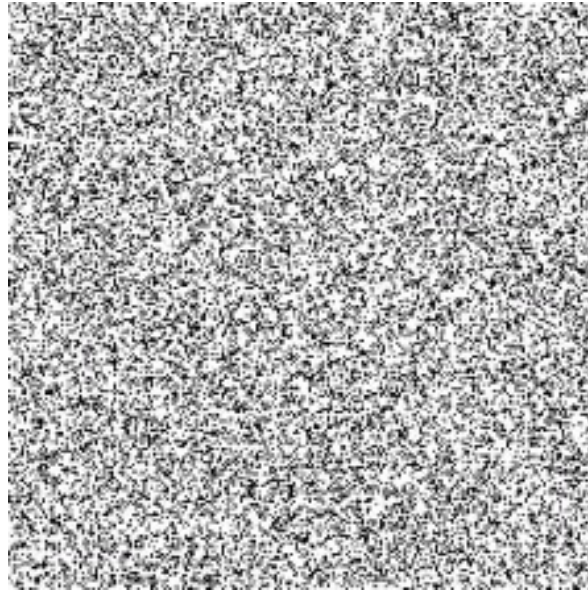
- Invented by [Bela Julesz](#) in the mid-20th century
- Demonstration that stereo perception can happen without any monocular cues



https://en.wikipedia.org/wiki/Random_dot_stereogram

History: Random dot stereograms

- Invented by [Bela Julesz](#) in the mid-20th century
- Demonstration that stereo perception can happen without any monocular cues



https://en.wikipedia.org/wiki/Random_dot_stereogram

Stereo: Outline

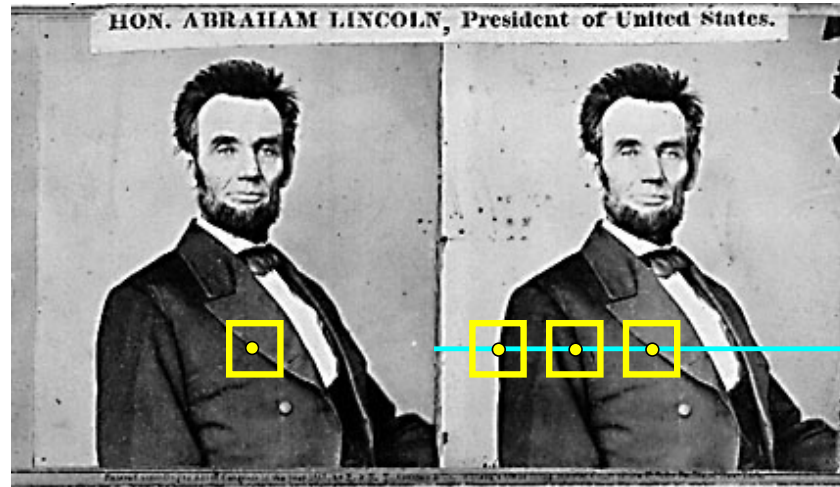
- Motivation and history
- Basic two-view stereo setup

Problem formulation

- **Given:** stereo pair (assumed calibrated)
- **Wanted:** dense depth map

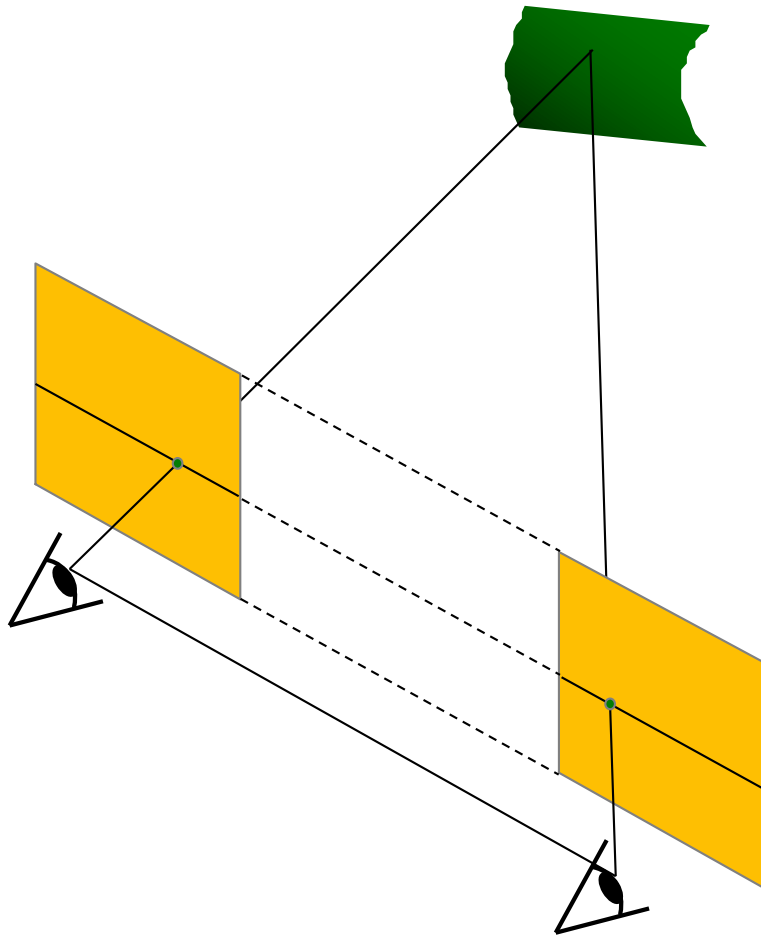


Basic stereo matching algorithm



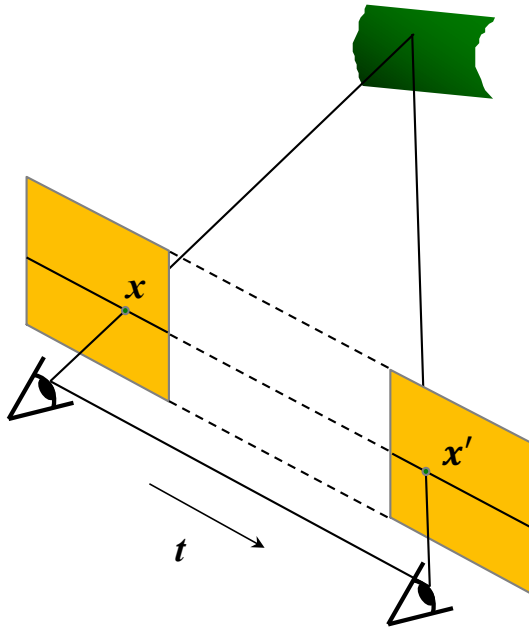
- For each pixel in the first image
 - Find corresponding epipolar line in the right image
 - Examine all pixels on the epipolar line and pick the best match
 - Triangulate the matches to get depth information
- Simplest case: epipolar lines are corresponding scanlines
 - When does this happen?

Parallel images



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

Essential matrix for parallel images



Epipolar constraint:

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

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

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

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

$$(u' \ v' \ 1) \begin{pmatrix} 0 \\ -t \\ tv \end{pmatrix} = 0$$

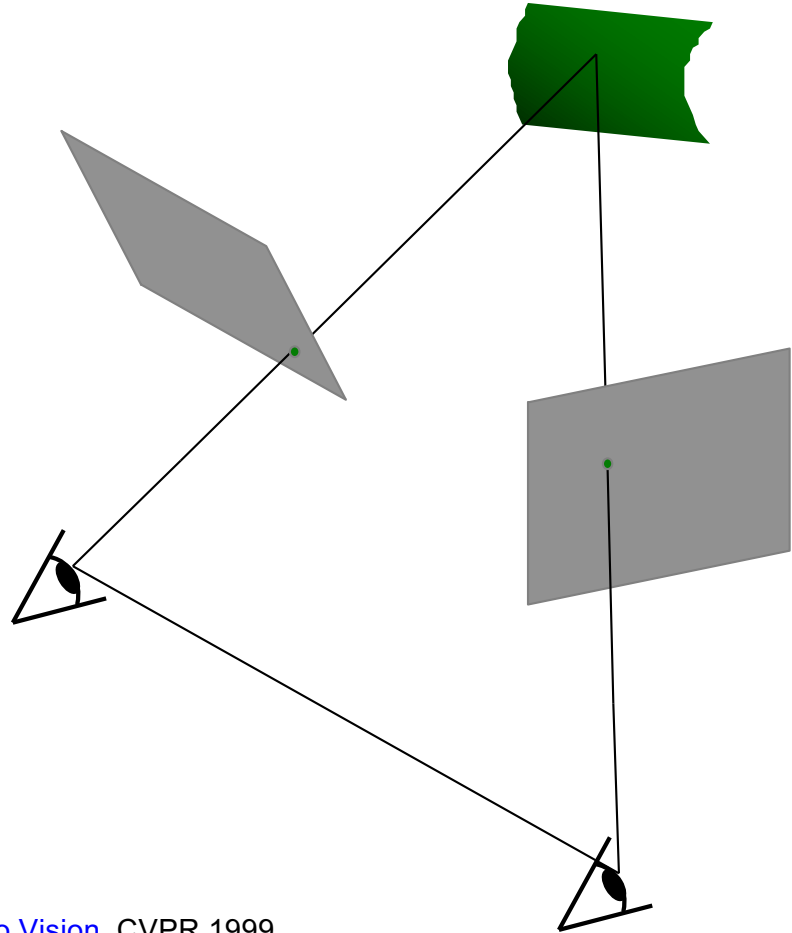
$$-tv + tv' = 0$$

$$v = v'$$

The y -coordinates of corresponding points are the same!

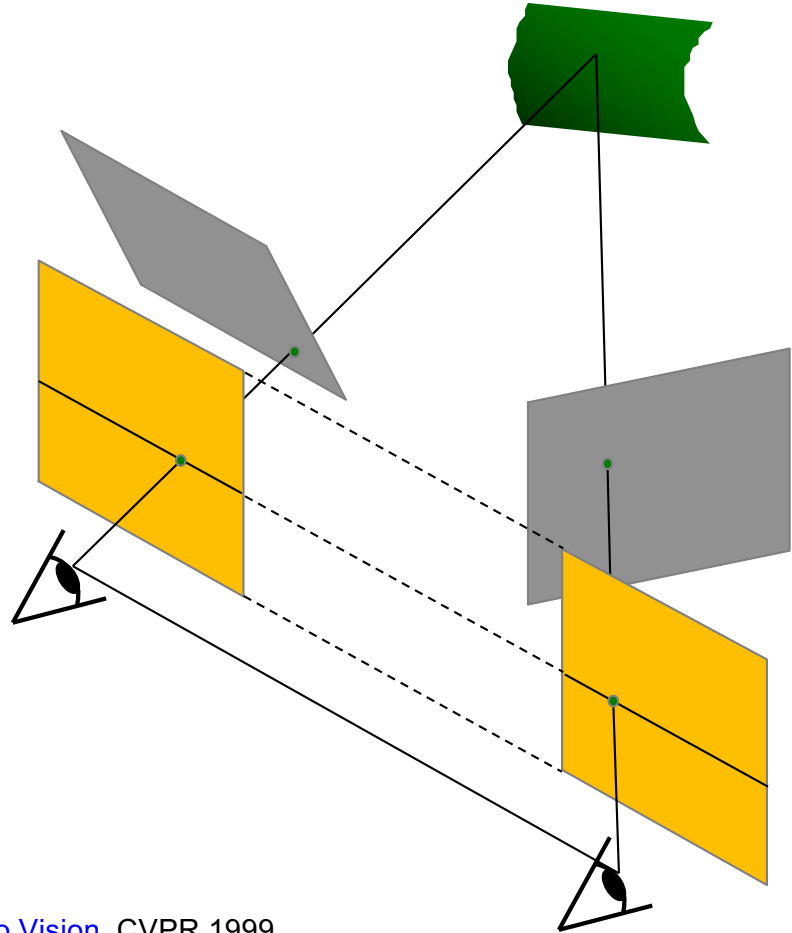
Stereo image rectification

- If the image planes are not parallel, we can find homographies to project each view onto a common plane parallel to the baseline



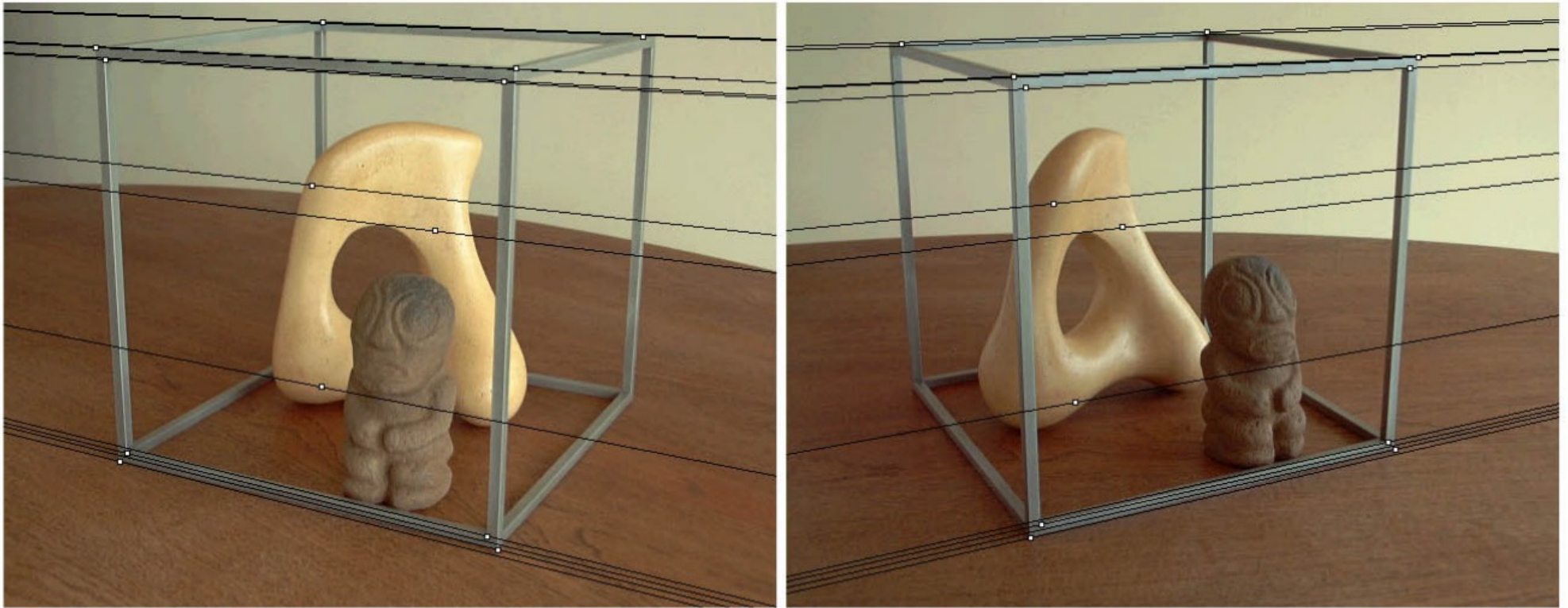
Stereo image rectification

- If the image planes are not parallel, we can find homographies to project each view onto a common plane parallel to the baseline



Stereo image rectification

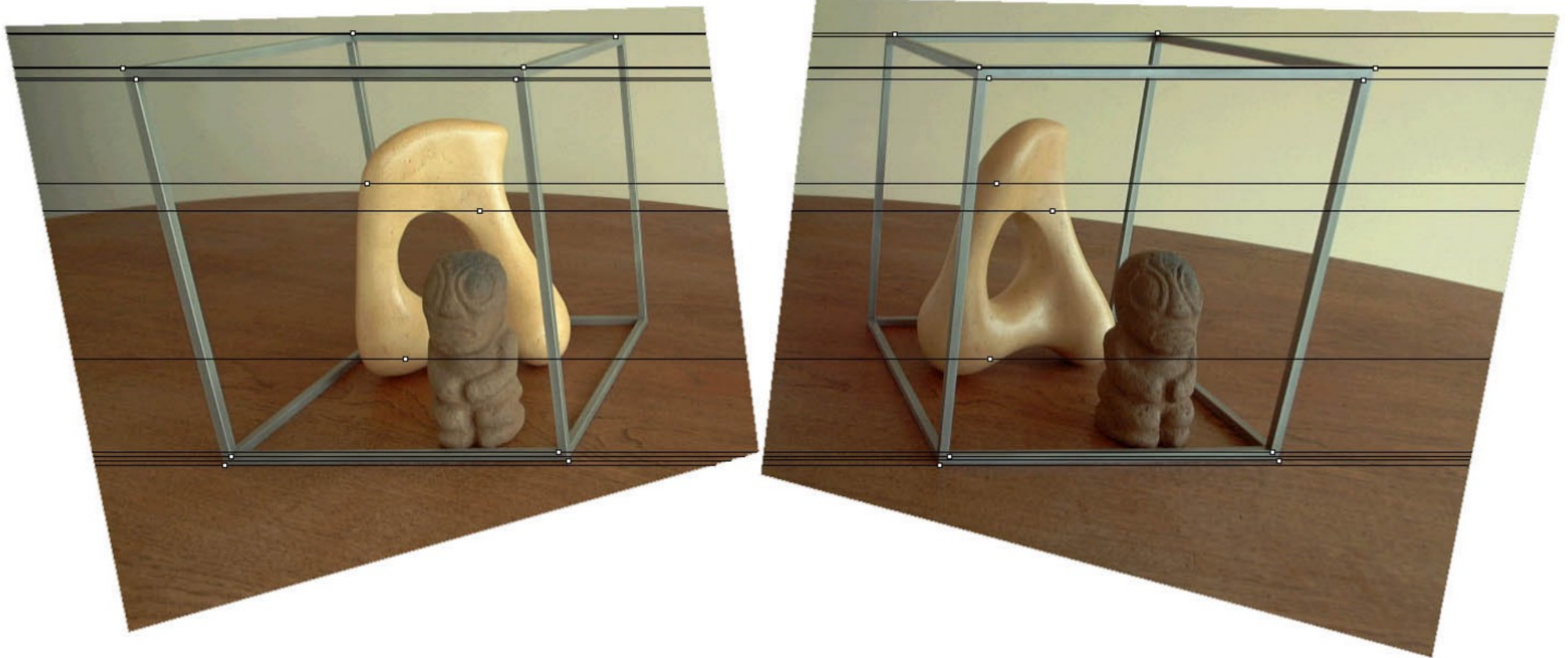
- Before rectification:



[Image source](#)

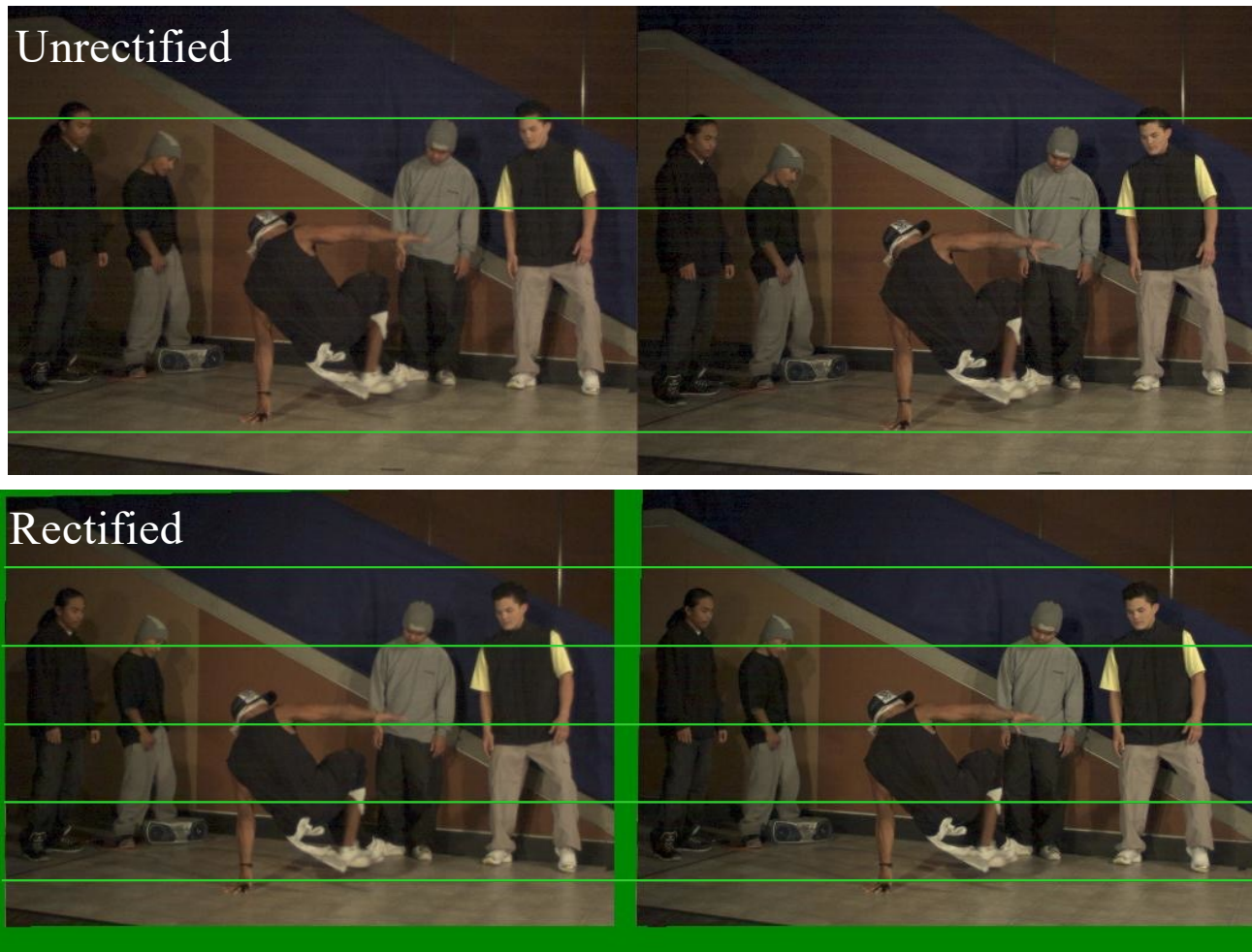
Stereo image rectification

- After rectification:

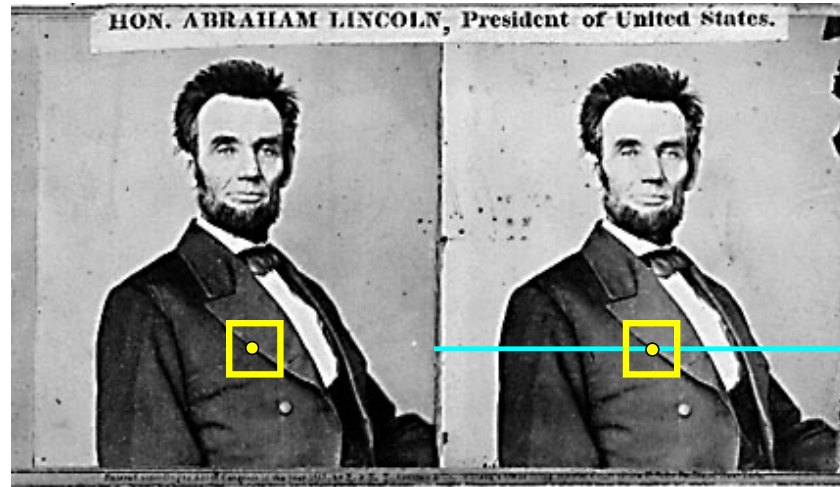


[Image source](#)

Another rectification example

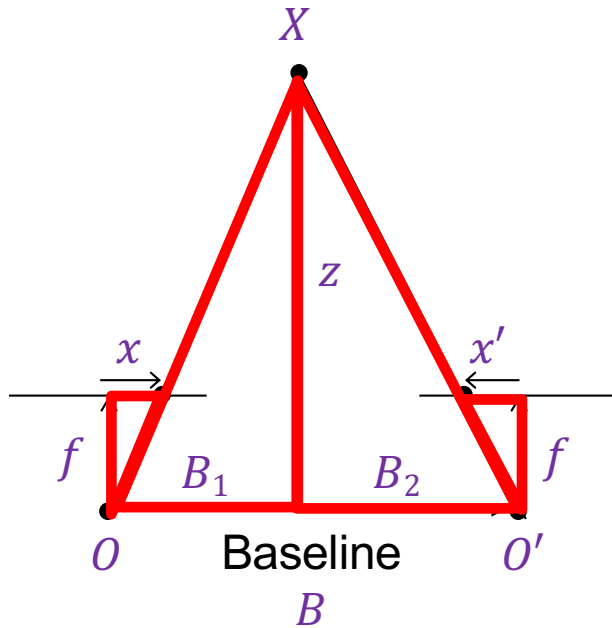


Basic stereo matching algorithm



- 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'
- Triangulate the matches to get depth information

Depth from disparity



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

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

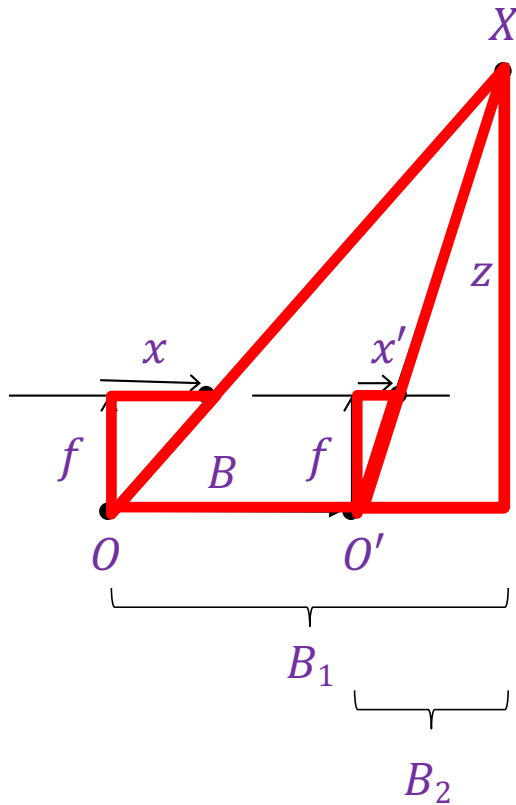
$$\frac{x - x'}{f} = \frac{B_1 + B_2}{z}$$

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

Disparity is
inversely
proportional to
depth!

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

Depth from disparity



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

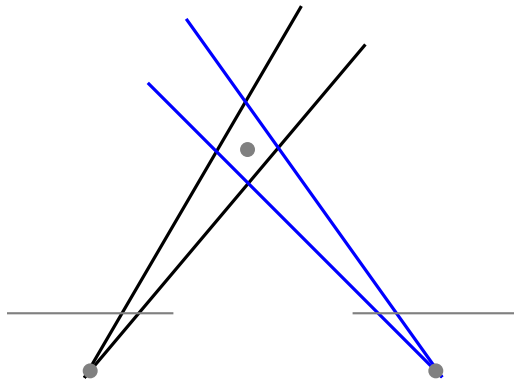
$$\frac{x'}{f} = \frac{B_2}{z}$$

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

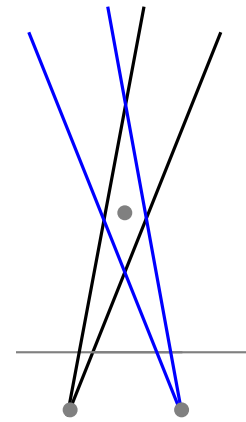
$$x - x' = \frac{fB}{z}$$

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

Effect of baseline on stereo results

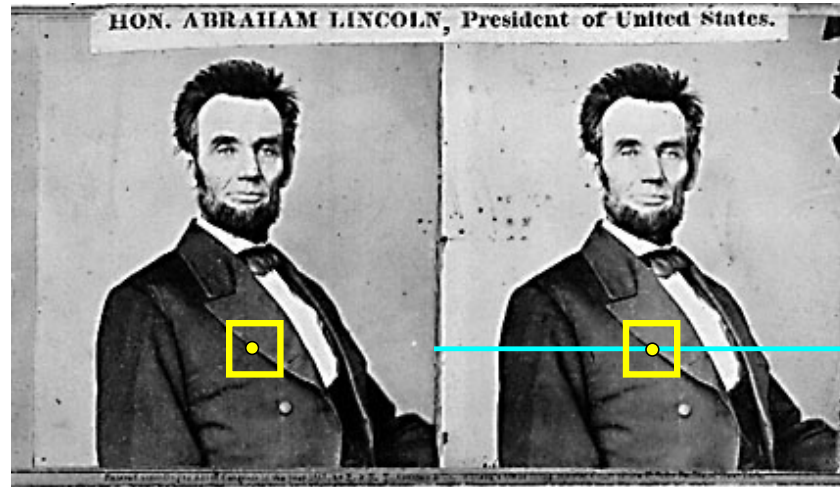


- Larger baseline
 - + Smaller triangulation error
 - Matching is more difficult



- Smaller baseline
 - Higher triangulation error
 - + Matching is easier

Basic stereo matching algorithm

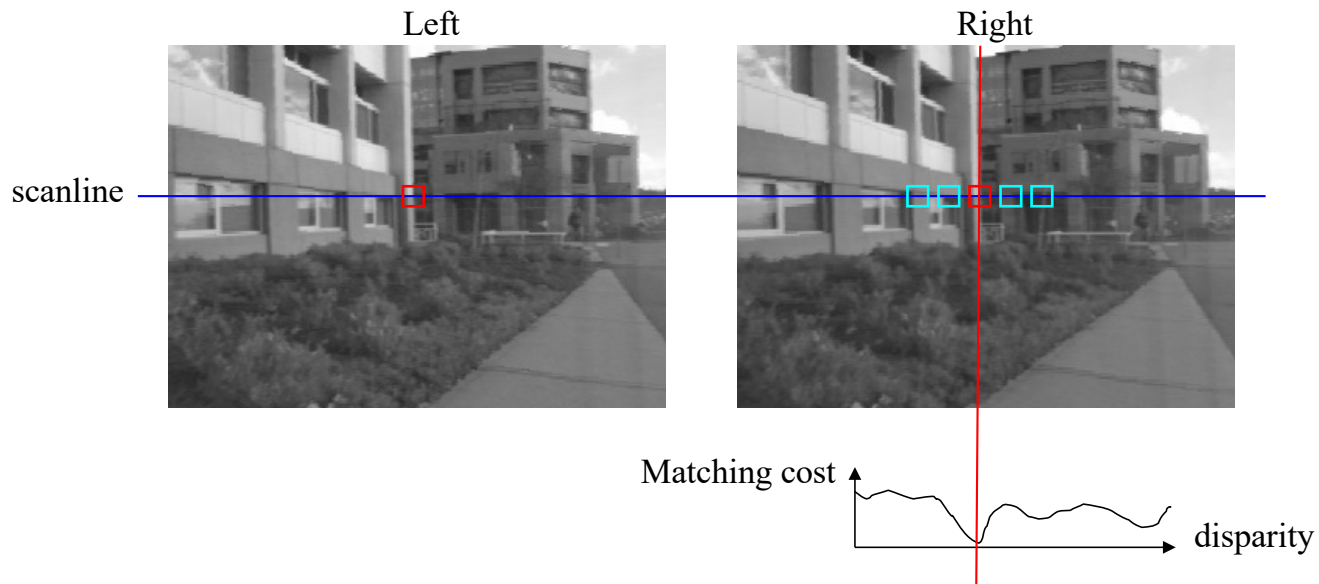


- If necessary, rectify the two stereo images to transform epipolar lines into scanlines
- For each pixel x in the first image
 - Find corresponding epipolar scanline in the right image
 - Examine all pixels on the scanline and pick the best match x'
 - Compute disparity $x - x'$ and set $\text{depth}(x)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

Outline

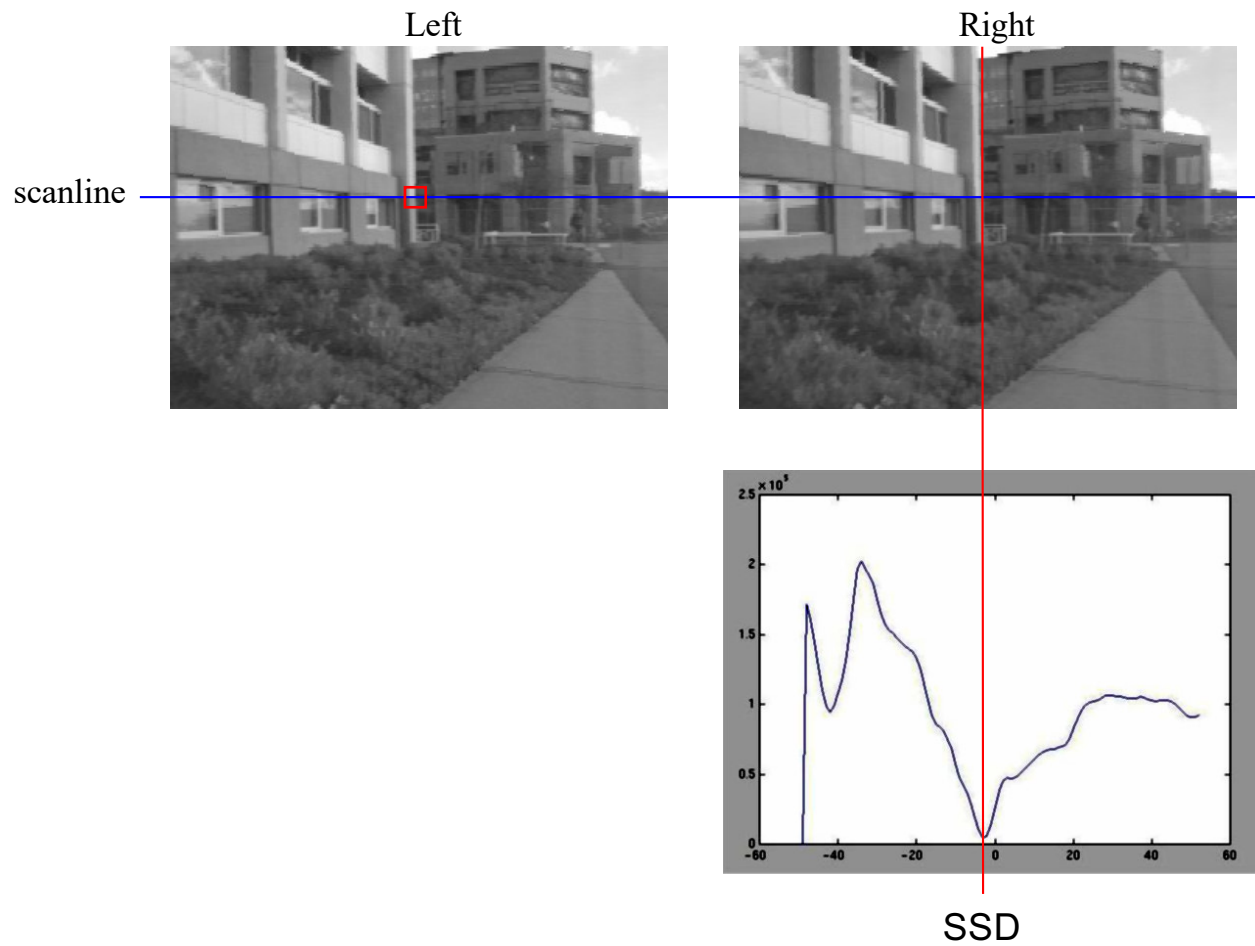
- Motivation and history
- Basic two-view stereo setup
- Local stereo matching algorithm

Correspondence search

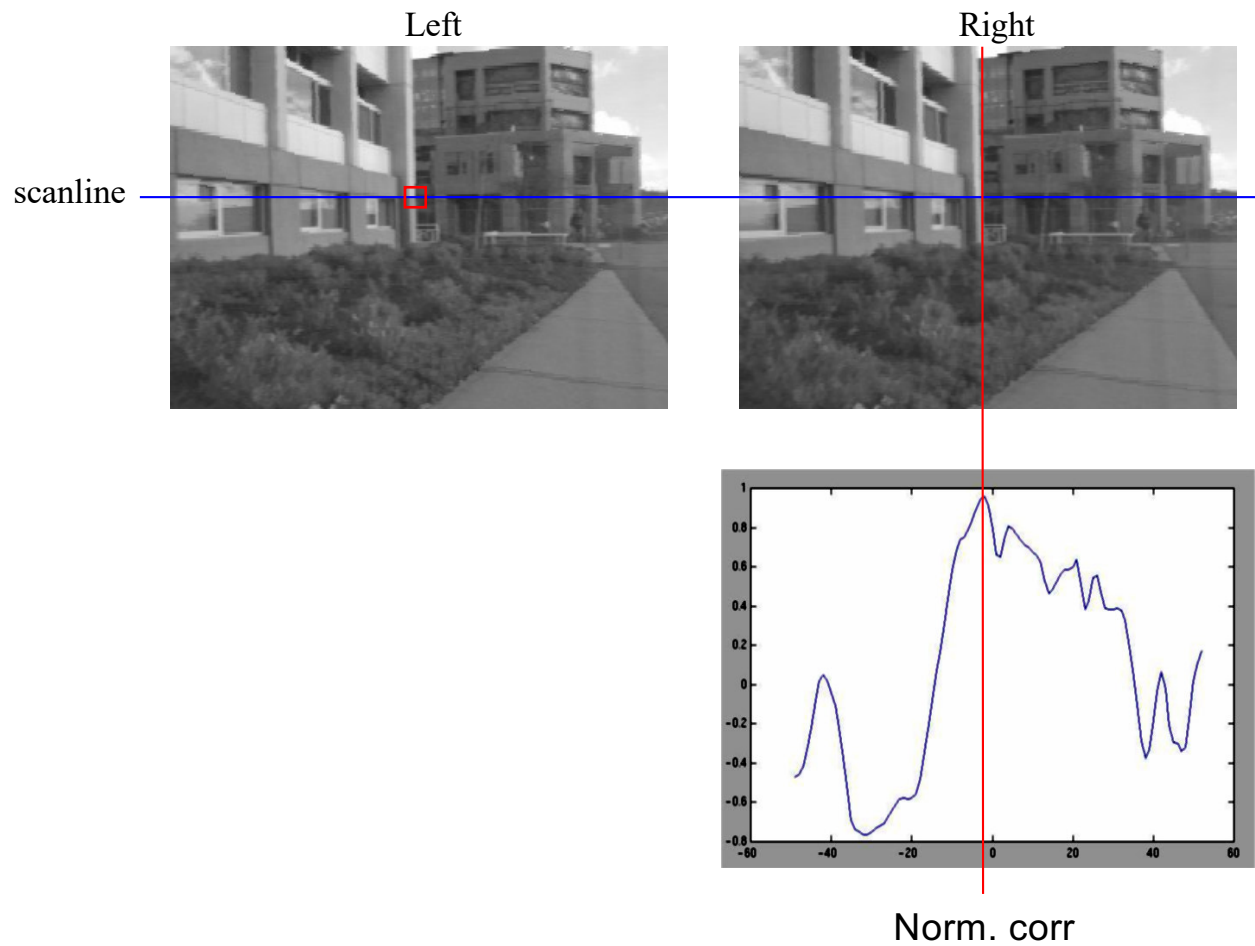


- Slide a window along the right scanline and compare contents of that window with the reference window in the left image
- Matching cost: SSD (or SAD) or normalized correlation

Correspondence search



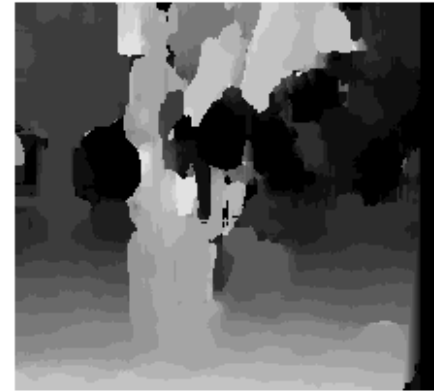
Correspondence search



Effect of window size on correspondence search



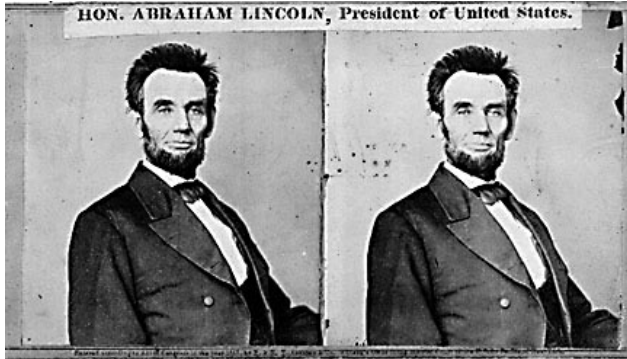
Window size 3



Window size 20

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

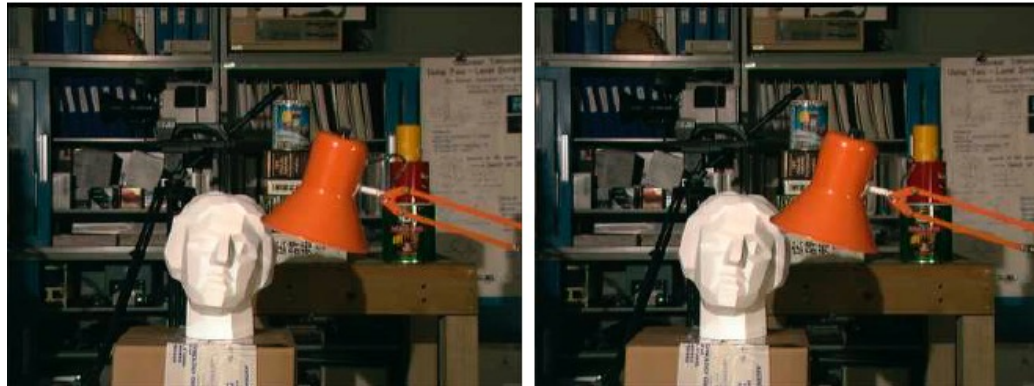
Where will basic window search fail?



Textureless surfaces



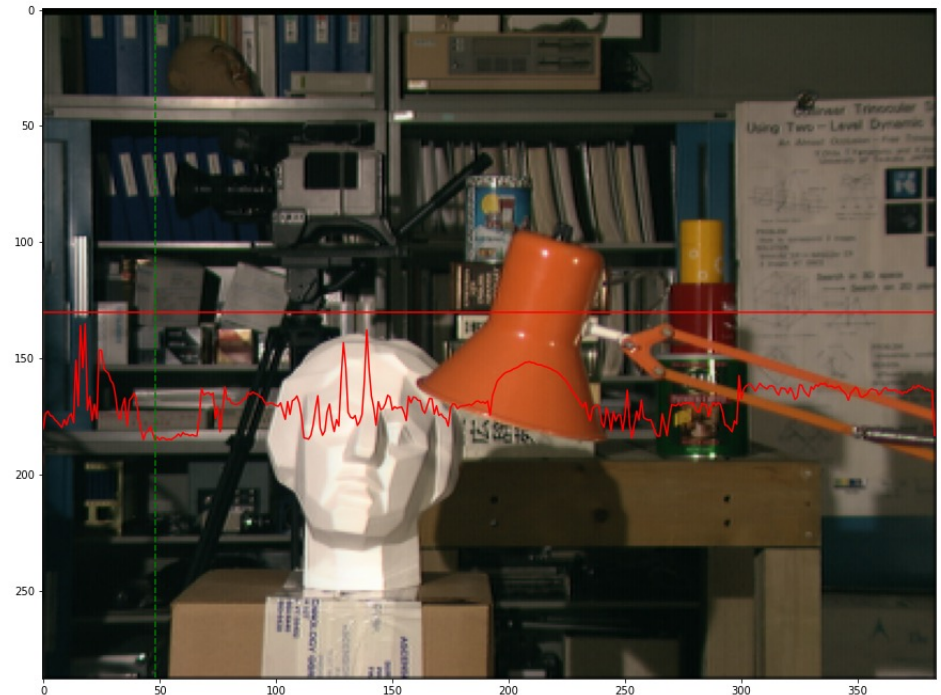
Occlusions, repetition



Non-Lambertian surfaces, specularities

Example: Textured neighborhood

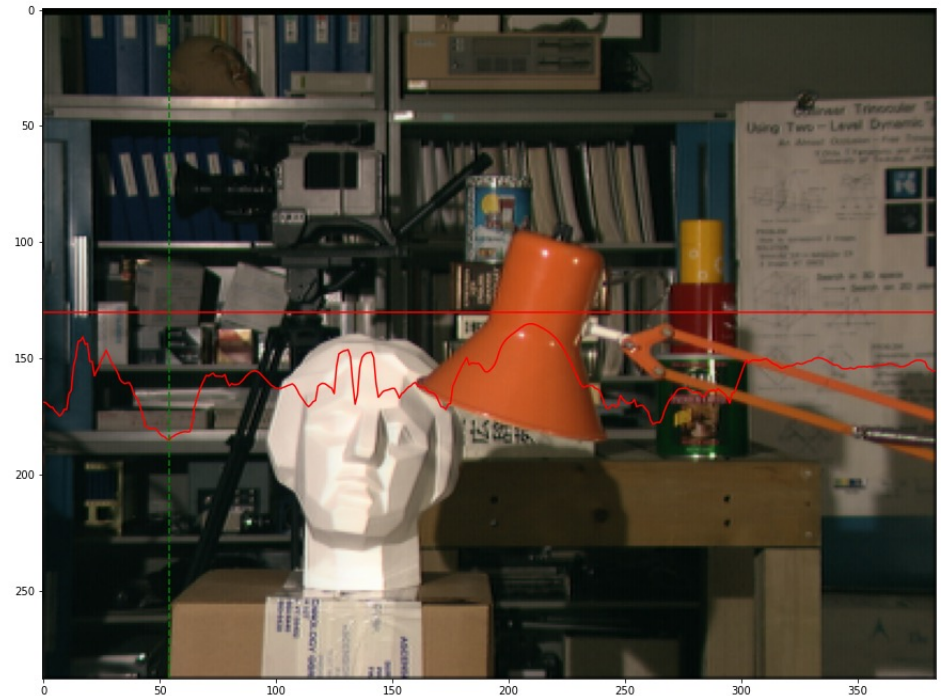
Window size: 1 pixel



Source: [D. Hoiem](#)

Example: Textured neighborhood

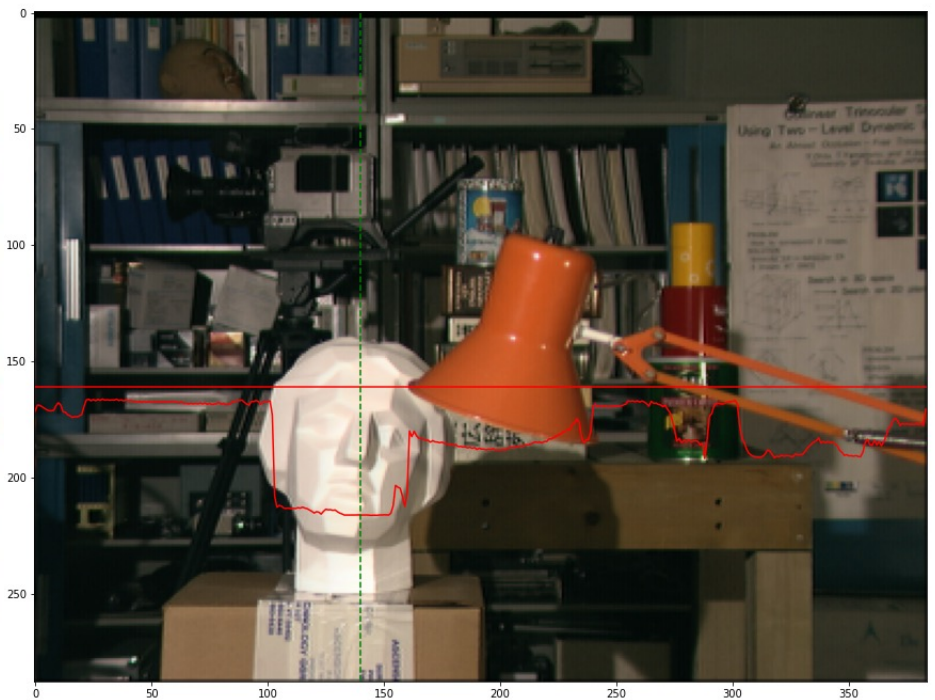
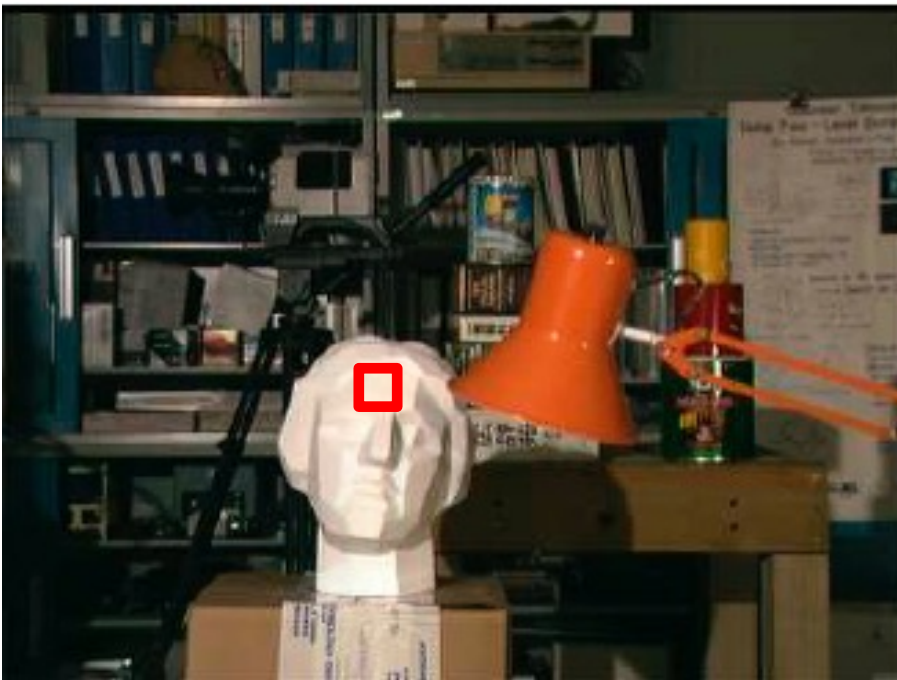
Window size: 7 pixels



Source: [D. Hoiem](#)

Example: Smooth neighborhood

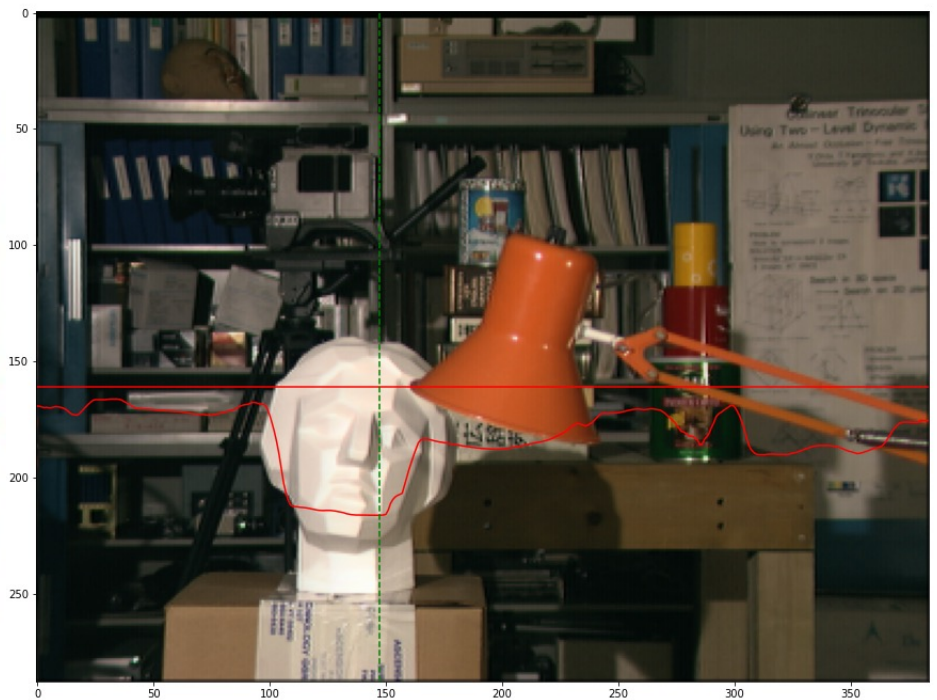
Window size: 1 pixel



Source: [D. Hoiem](#)

Example: Smooth neighborhood

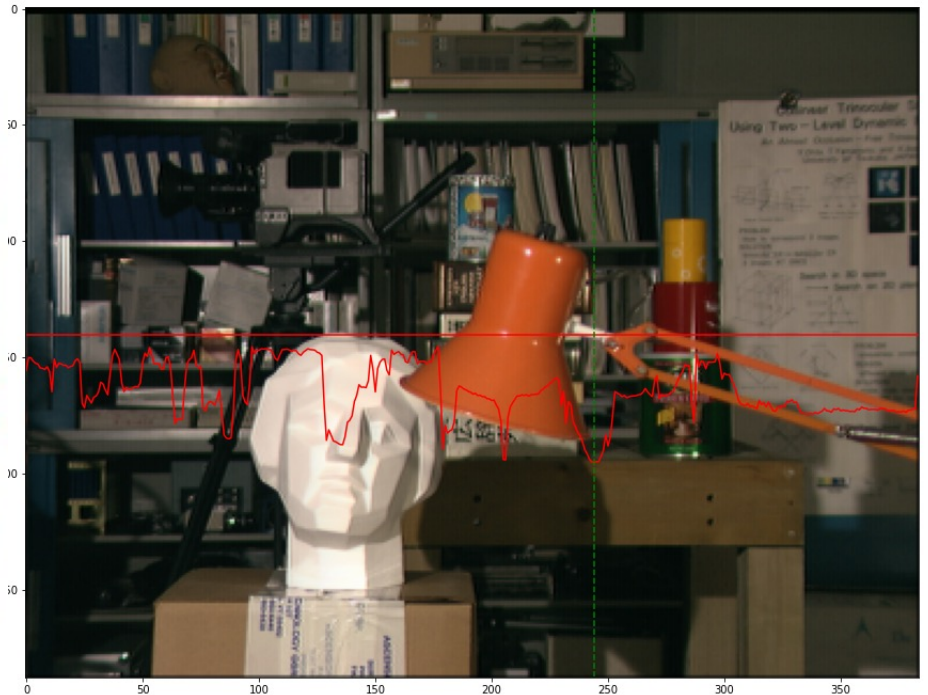
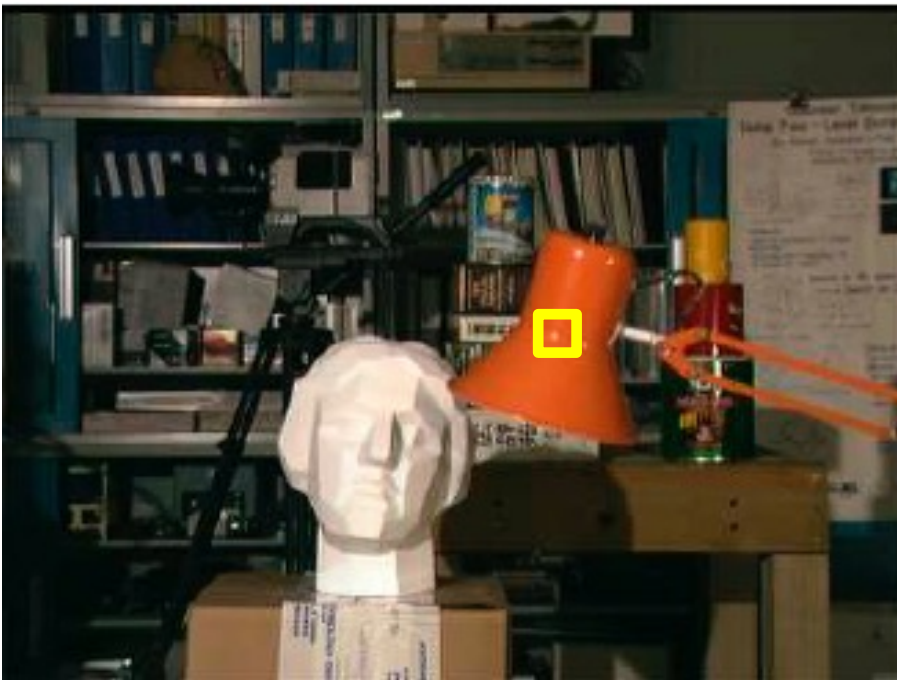
Window size: 7 pixels



Source: [D. Hoiem](#)

Example: Specular highlight

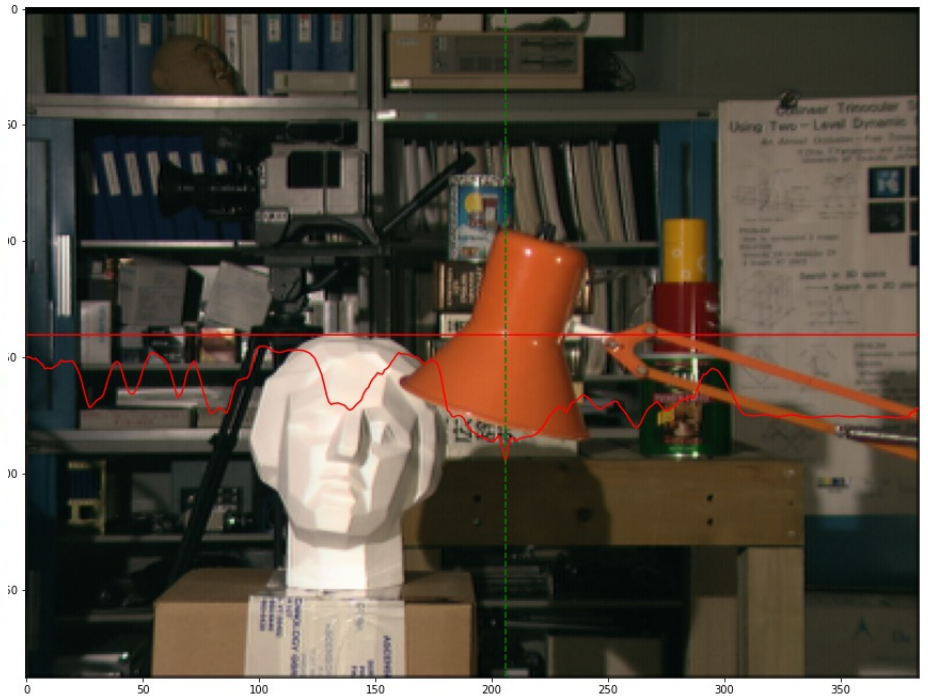
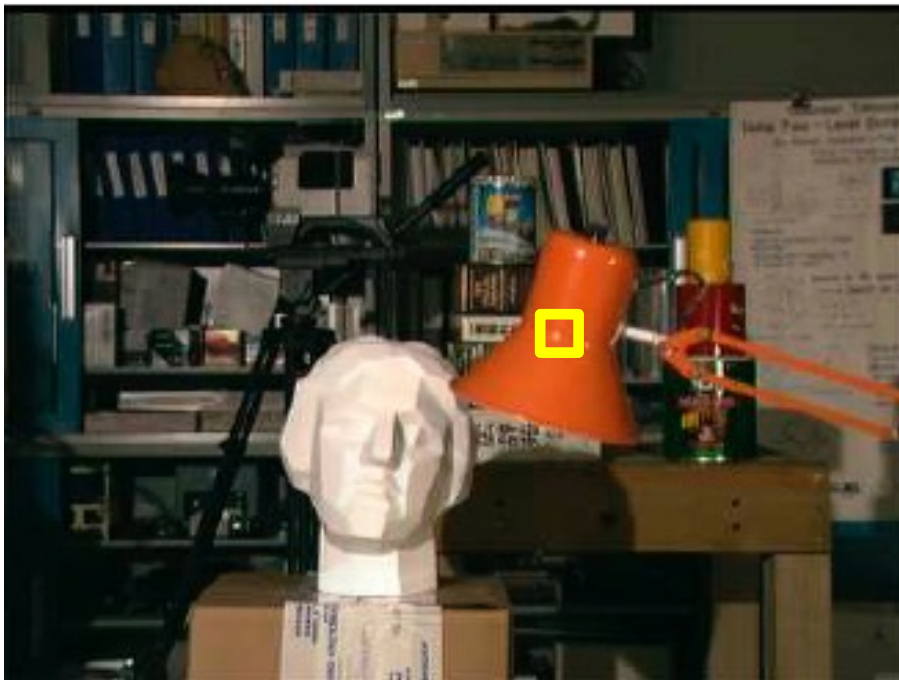
Window size: 1 pixel



Source: [D. Hoiem](#)

Example: Specular highlight

Window size: 7 pixels



Source: [D. Hoiem](#)

Outline

- Motivation and history
- Basic two-view stereo setup
- Local stereo matching algorithm
- Beyond local stereo matching

Stereo as optimization with non-local constraints

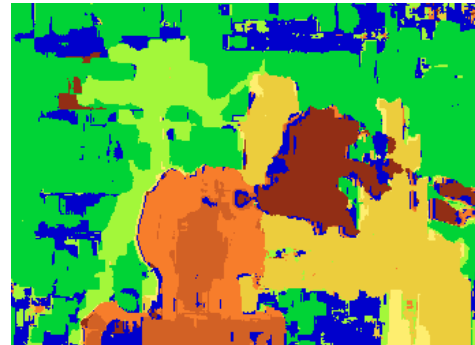
Data



Ground truth



Window-based matching



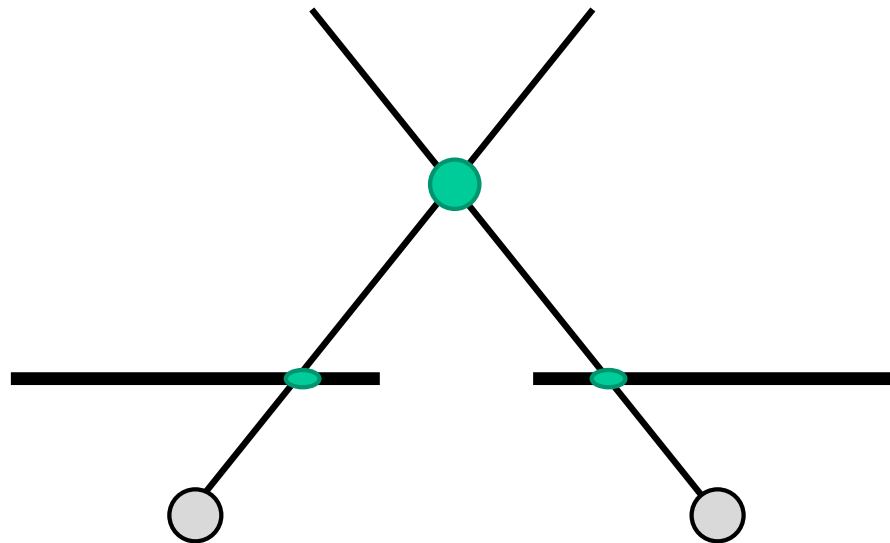
Global optimization
method (graph cuts)



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

Non-local constraint: Uniqueness

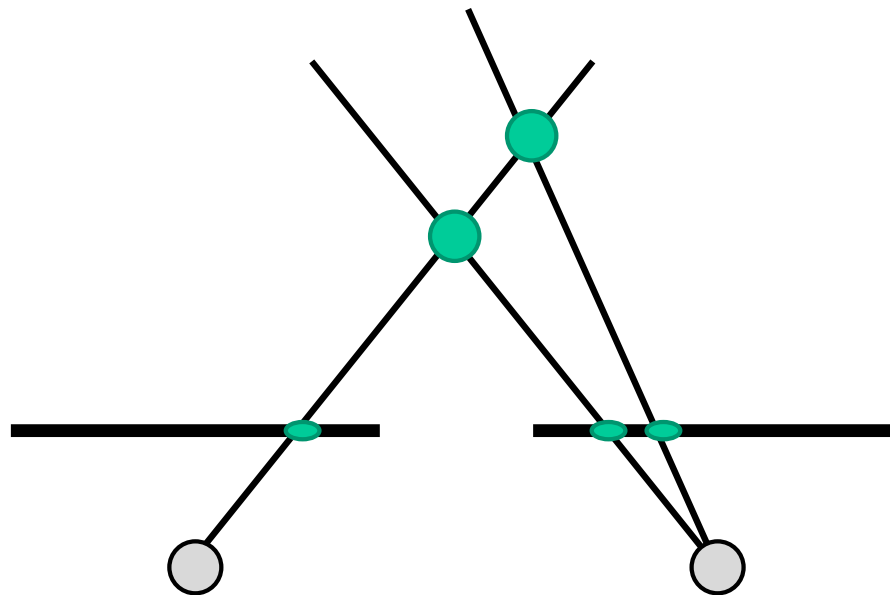
- Each point in one image should match at most one point in the other image
- Does uniqueness always hold in real life?



Source: [J. Johnson and D. Fouhey](#)

Non-local constraint: Uniqueness

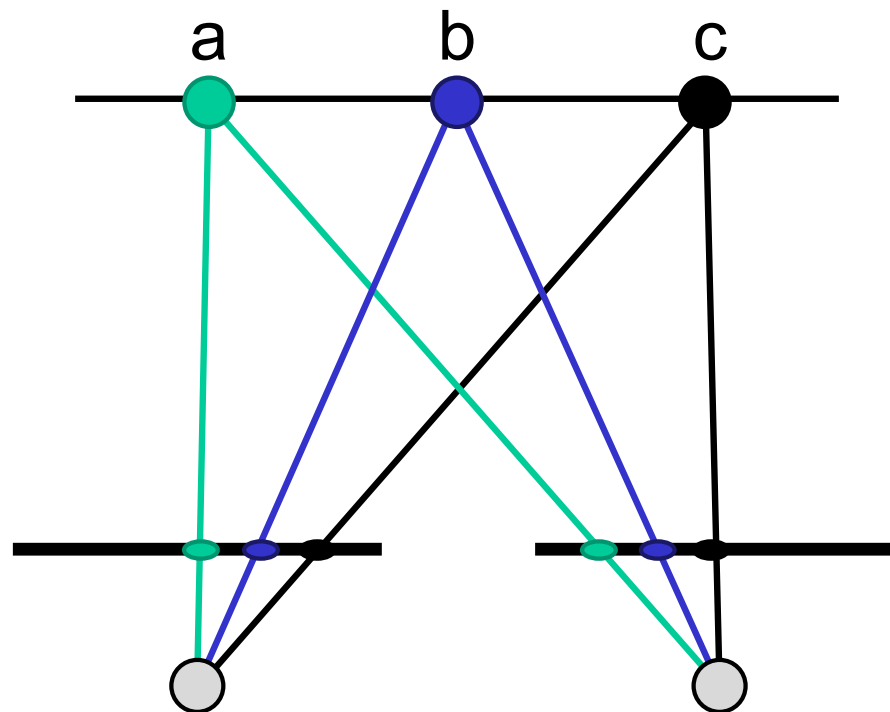
- Each point in one image should match at most one point in the other image
- Does uniqueness always hold in real life?



Source: [J. Johnson and D. Fouhey](#)

Non-local constraint: Ordering

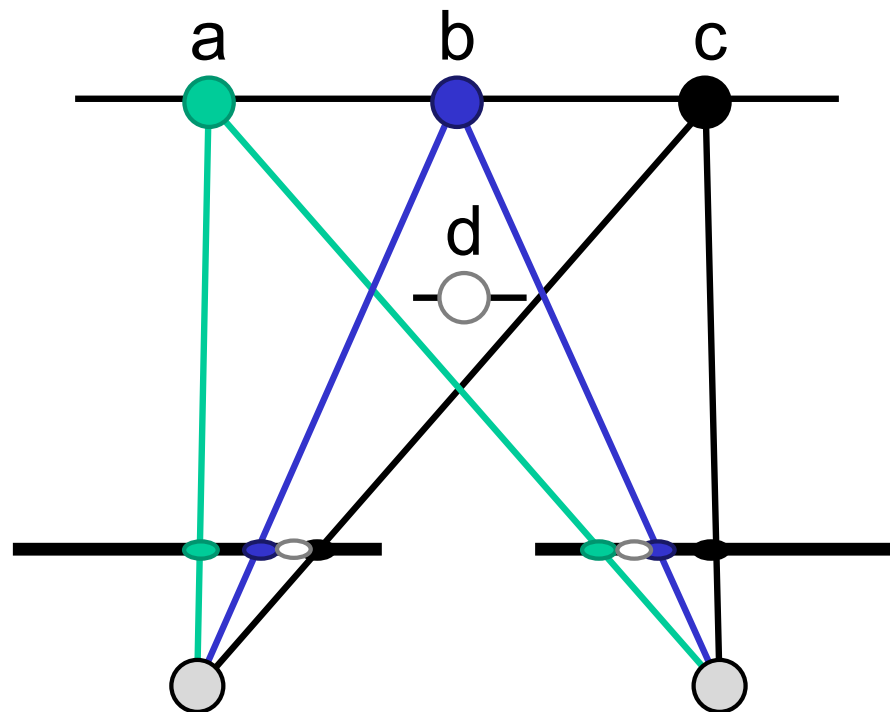
- Corresponding points should appear in the same order
- Is ordering always preserved in real life?



Source: [J. Johnson and D. Fouhey](#)

Non-local constraint: Ordering

- Corresponding points should appear in the same order
- Is ordering always preserved in real life?



Source: [J. Johnson and D. Fouhey](#)

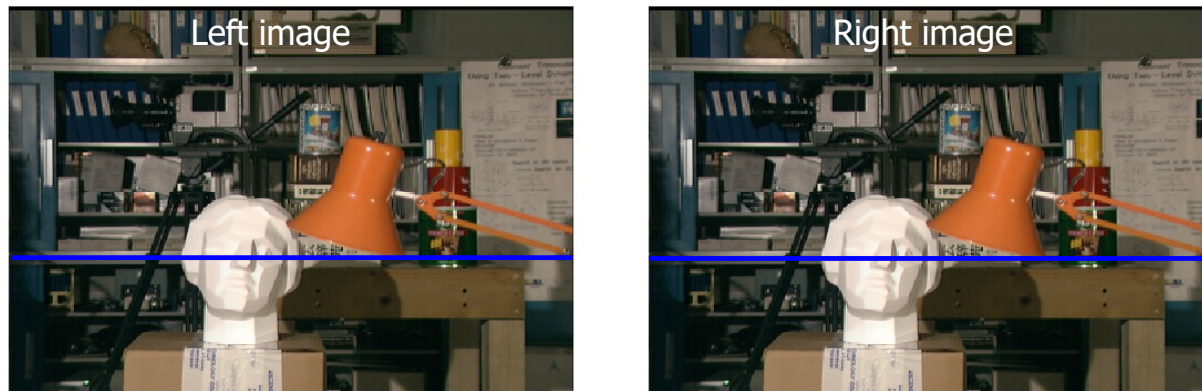
Non-local constraint: Smoothness

- We expect disparity values to change slowly (for the most part)

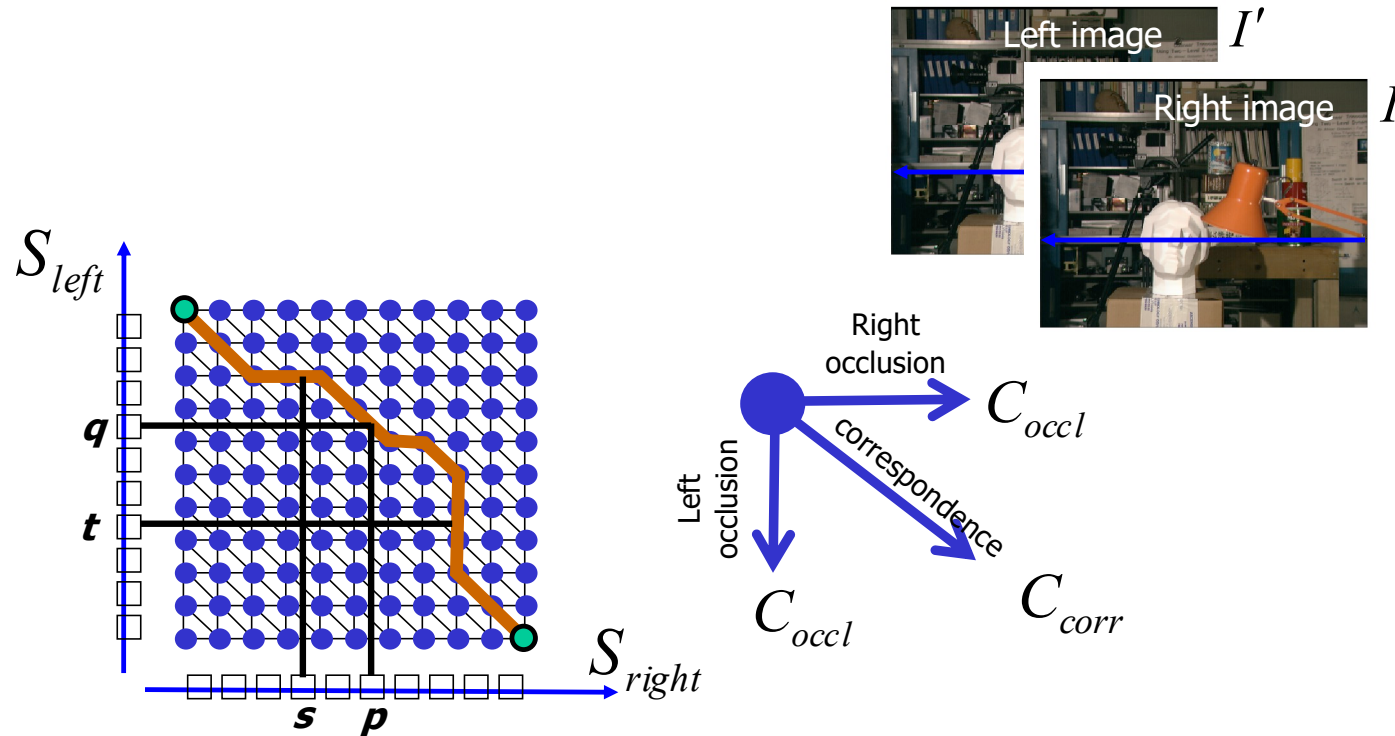


Scanline stereo by dynamic programming

- Match pixels along the entire scanline while preserving uniqueness and ordering
- Different scanlines are still optimized independently



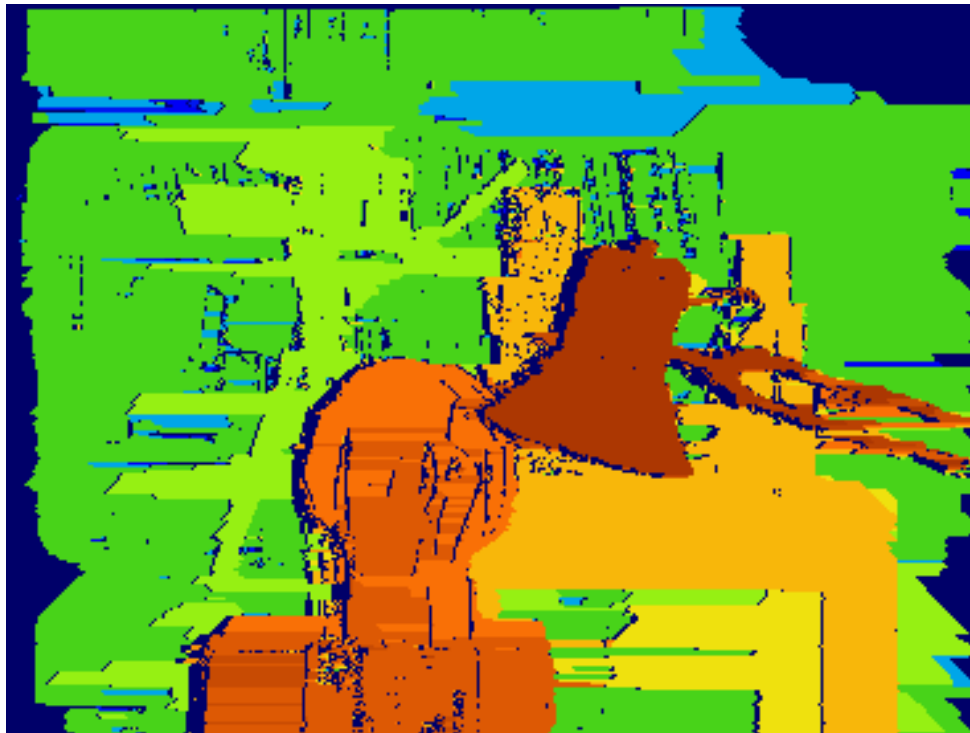
Scanline stereo by dynamic programming



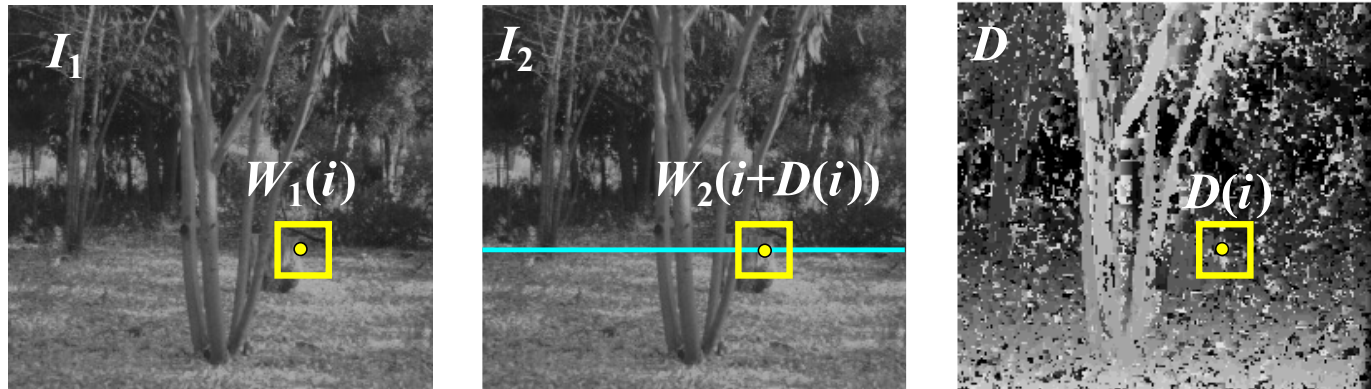
Source: Y. Boykov

Scanline stereo by dynamic programming

- Generates streaking artifacts!



Stereo matching as global optimization

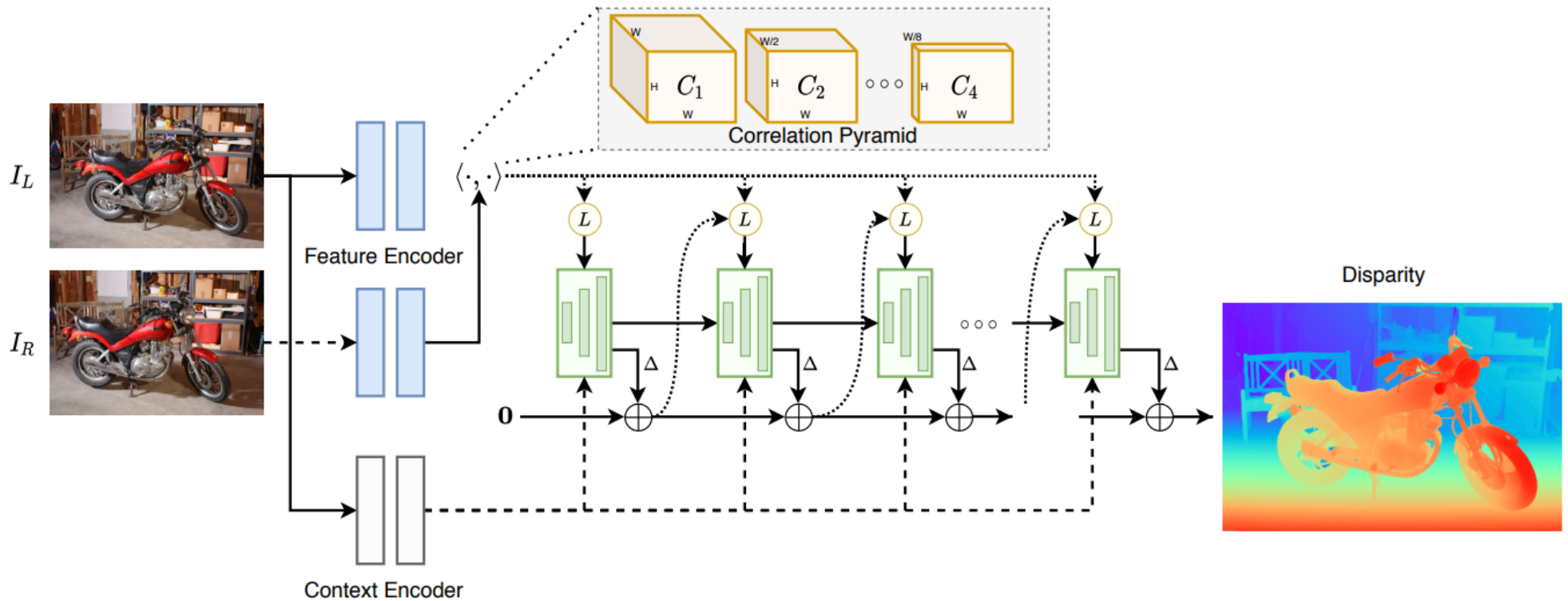


$$E(D) = \underbrace{\sum_i \left(W_1(i) - W_2(i + D(i)) \right)^2}_{\text{Data term}} + \lambda \underbrace{\sum_{(i,j) \in \mathcal{N}} \rho(D(i) - D(j))}_{\text{Neighborhood smoothness term}}$$

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

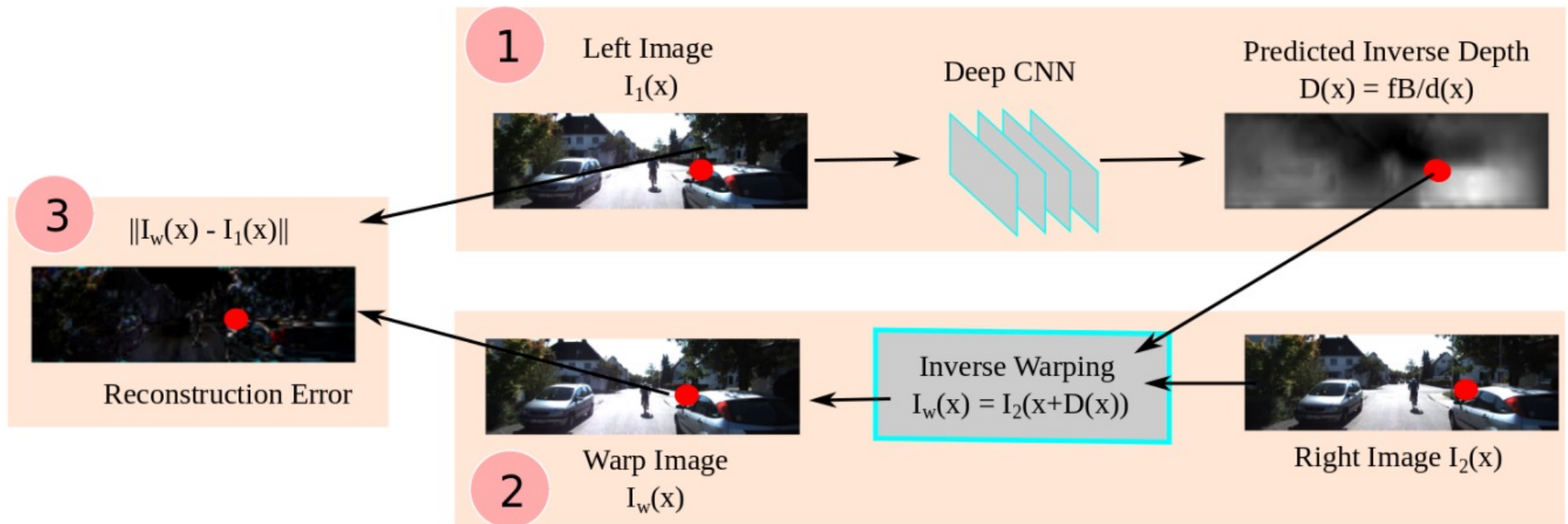
Y. Boykov, O. Veksler, and R. Zabih, [Fast Approximate Energy Minimization via Graph Cuts](#), PAMI 2001

Stereo matching with deep networks



L. Lipson et al. [RAFT-Stereo: Multilevel Recurrent Field Transforms for Stereo Matching](#). 3DV 2021

Self-supervised depth estimation



Cross-View Completion

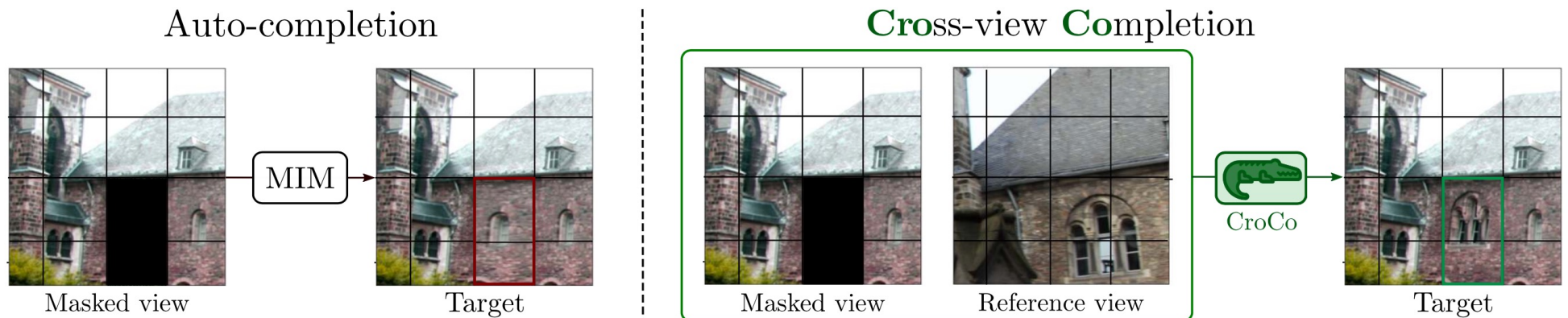


Figure 1: Auto-completion vs. Cross-view completion tasks. *Left:* given a masked image, a model trained for auto-completion can only leverage the visible context to fill-in the blanks and thus relies mainly on high-level semantic information. *Right:* given an additional view of the same scene, cross-view completion makes precise reconstruction possible, assuming that the model is able to understand both the scene geometry and the spatial relationship between the two images.

Cross-View Completion

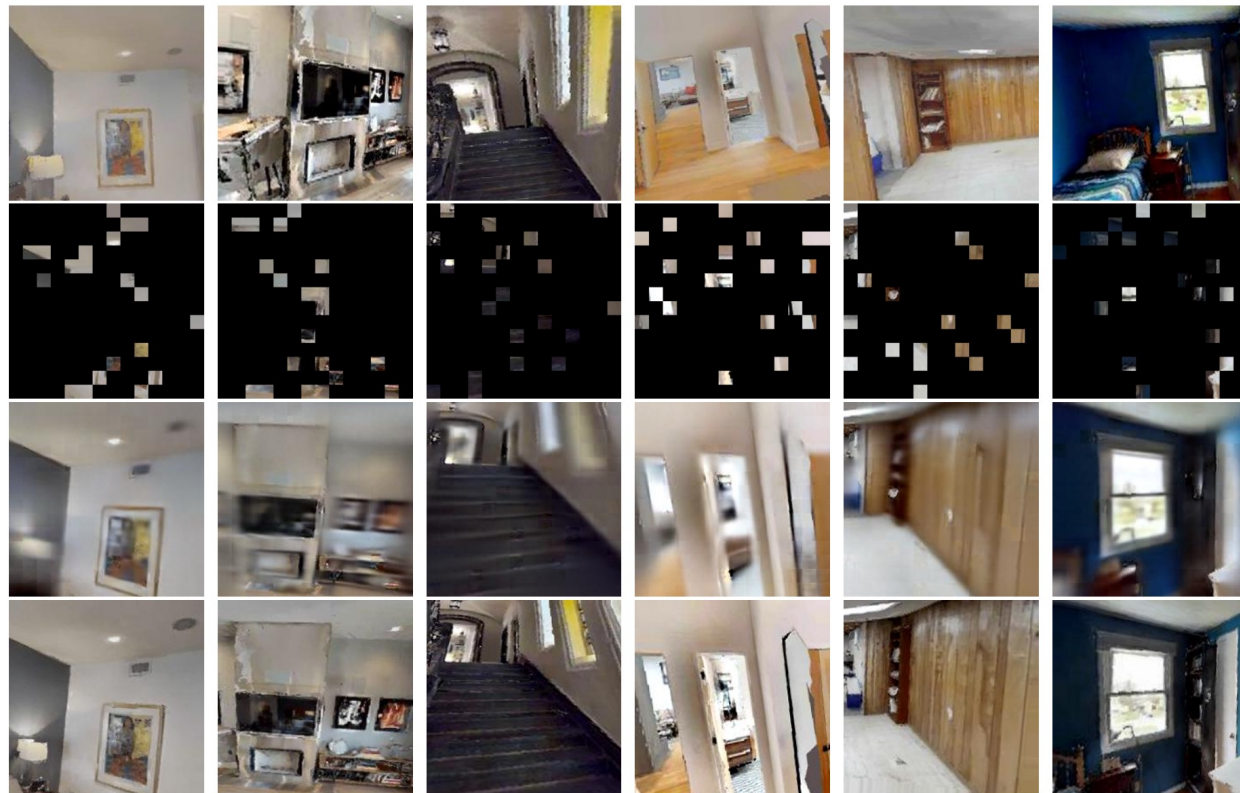


Figure 2: **Reconstruction examples from CroCo** on scenes unseen during training. From top to bottom: reference image (input), masked image (input), CroCo output, original target image.

Stereo datasets

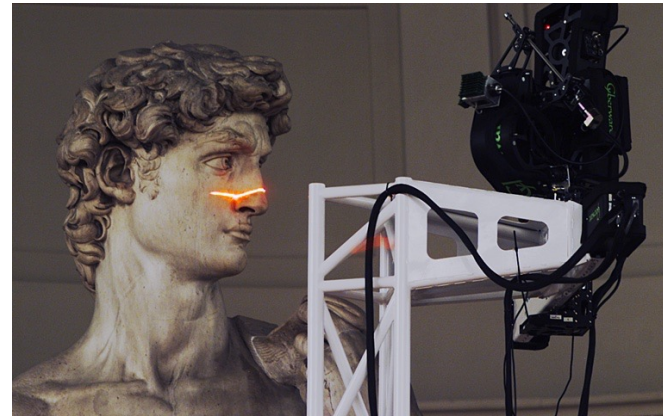
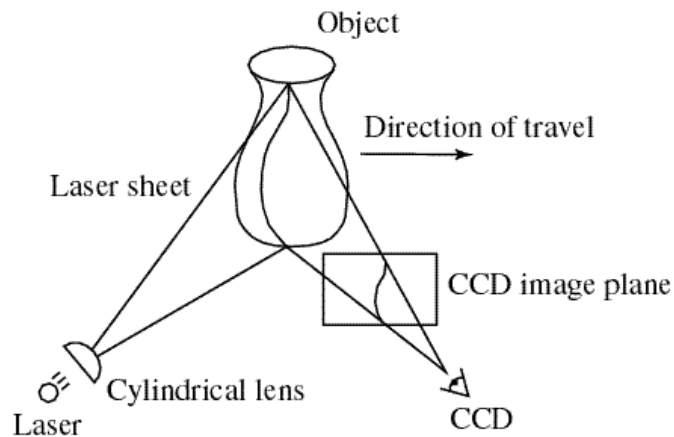
- [Middlebury stereo datasets](#)
- [KITTI](#)
- [Synthetic data](#)



Outline

- Motivation and history
- Basic two-view stereo setup
- Local stereo matching algorithm
- Stereo with non-local optimization
- **Active stereo with structured light**

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
 - For each point on the stripe, look for the displacement from reference location

Source: S. Seitz

Laser scanned models



The Digital Michelangelo Project, Levoy et al.

Source: S. Seitz

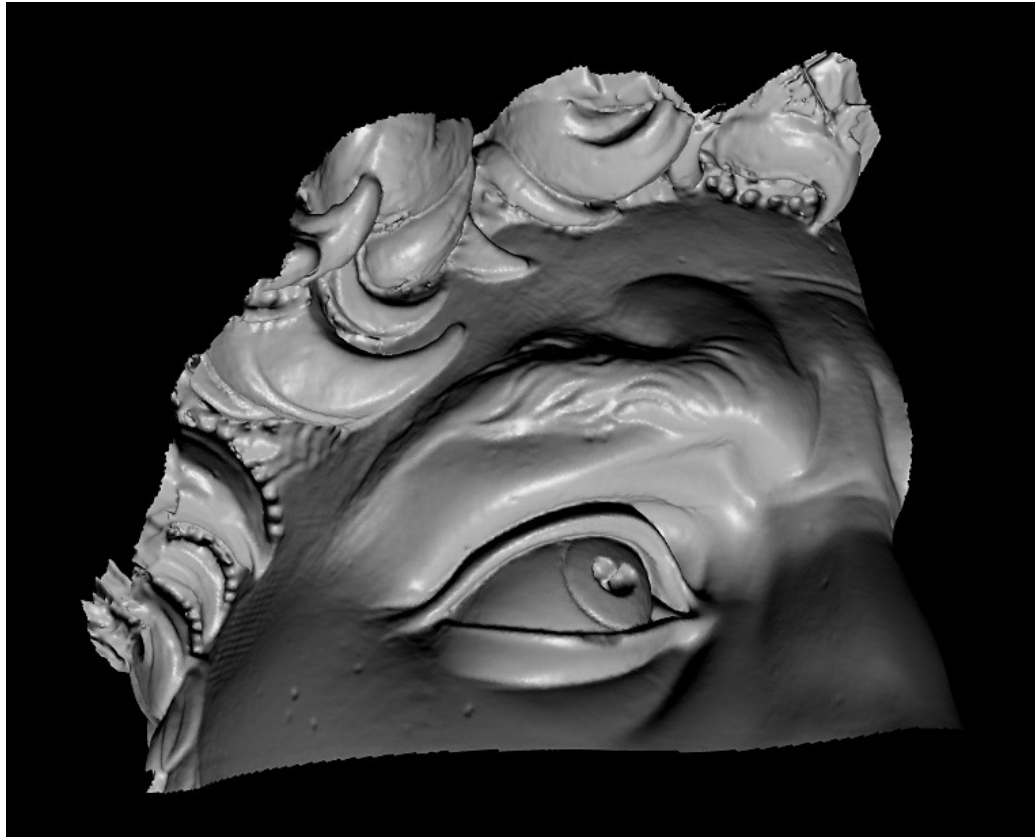
Laser scanned models



The Digital Michelangelo Project, Levoy et al.

Source: S. Seitz

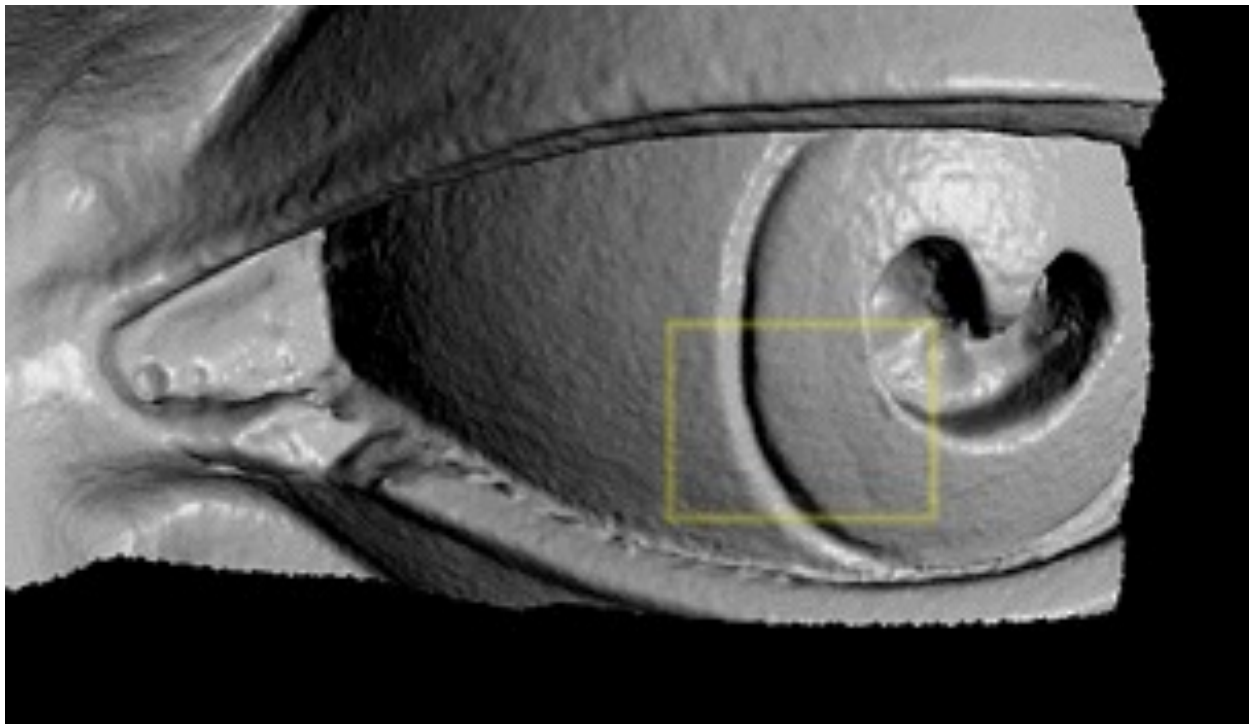
Laser scanned models



The Digital Michelangelo Project, Levoy et al.

Source: S. Seitz

Laser scanned models

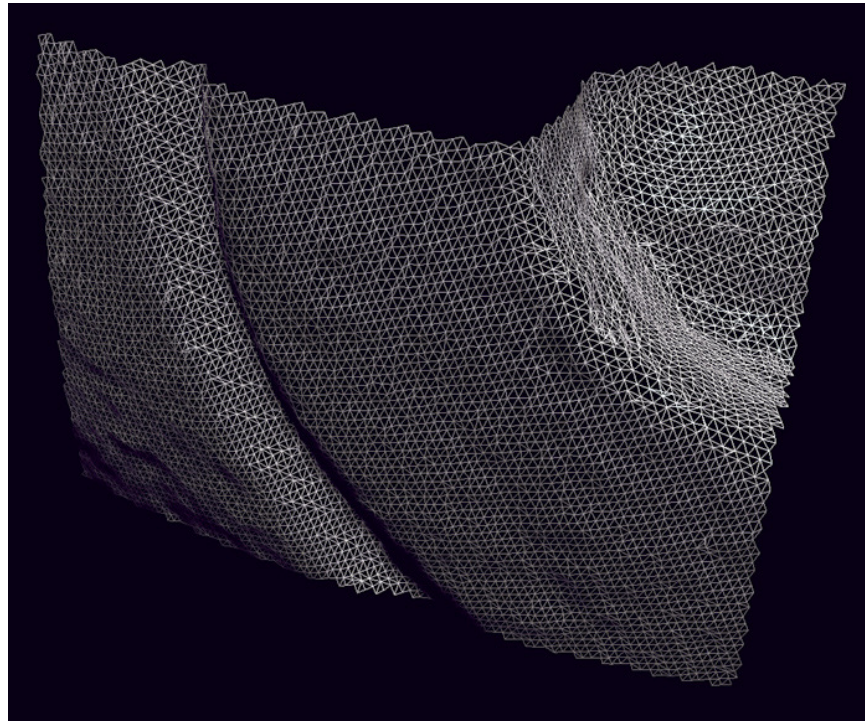


The Digital Michelangelo Project, Levoy et al.

Source: S. Seitz

Laser scanned models

1.0 mm resolution (56 million triangles)

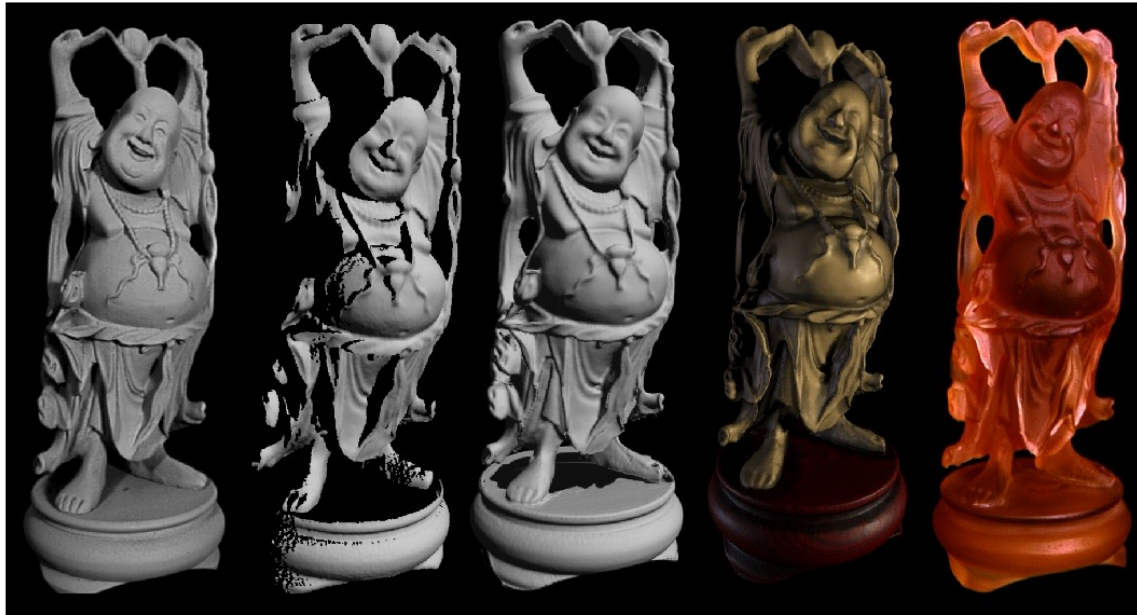


The Digital Michelangelo Project, Levoy et al.

Source: S. Seitz

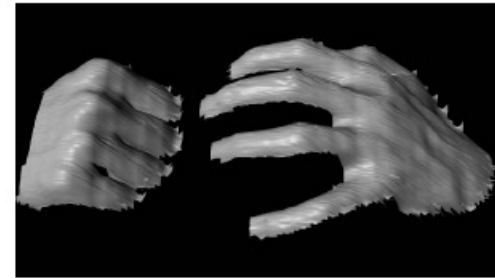
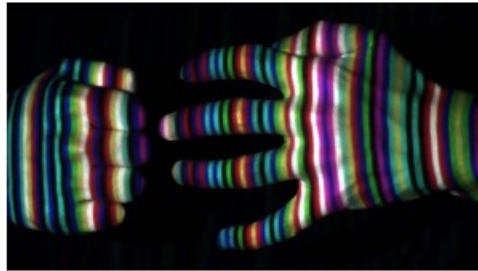
Aligning range images

- A single range scan is not sufficient to capture a complex surface
- Need techniques to register multiple range images (next lecture)

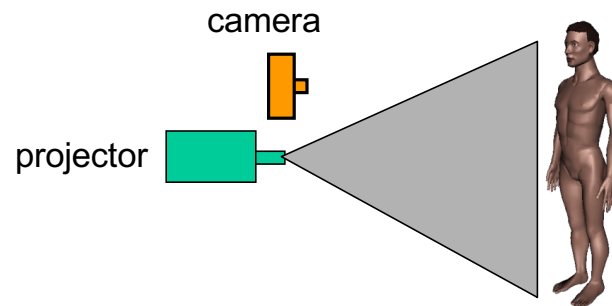


B. Curless and M. Levoy, [A Volumetric Method for Building Complex Models from Range Images.](#)
SIGGRAPH 1996

Active stereo with structured light

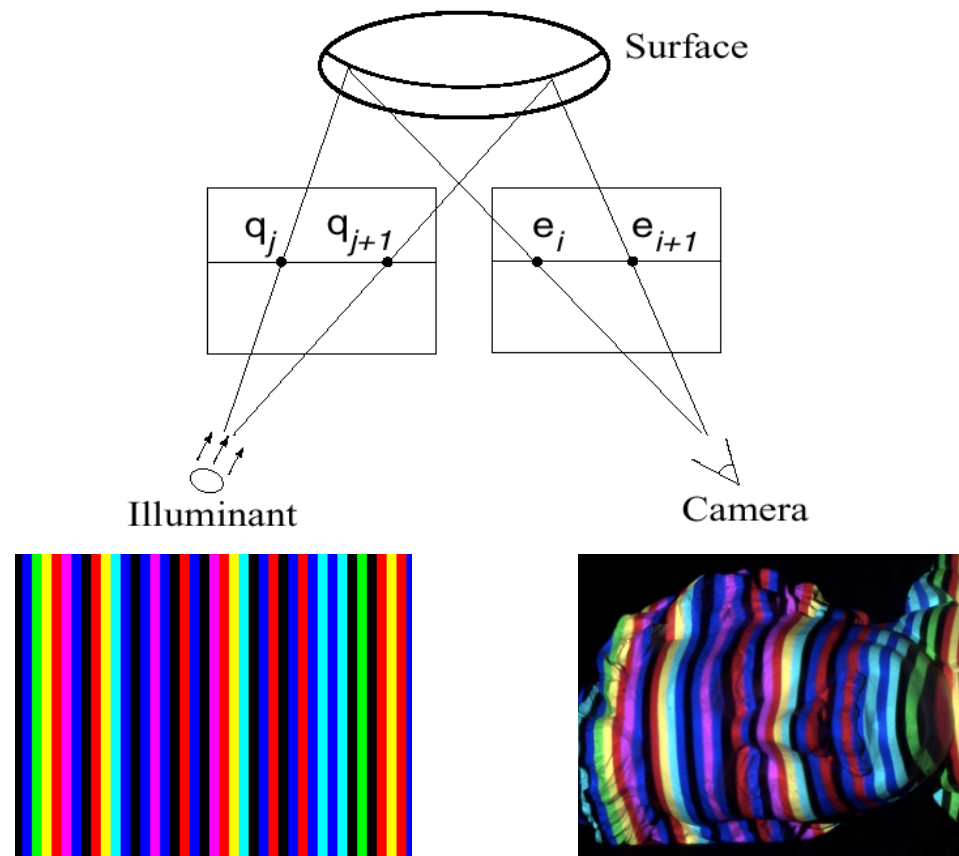


- Project “structured” light patterns onto the object
 - Simplifies the correspondence problem
 - Allows us to use only one camera



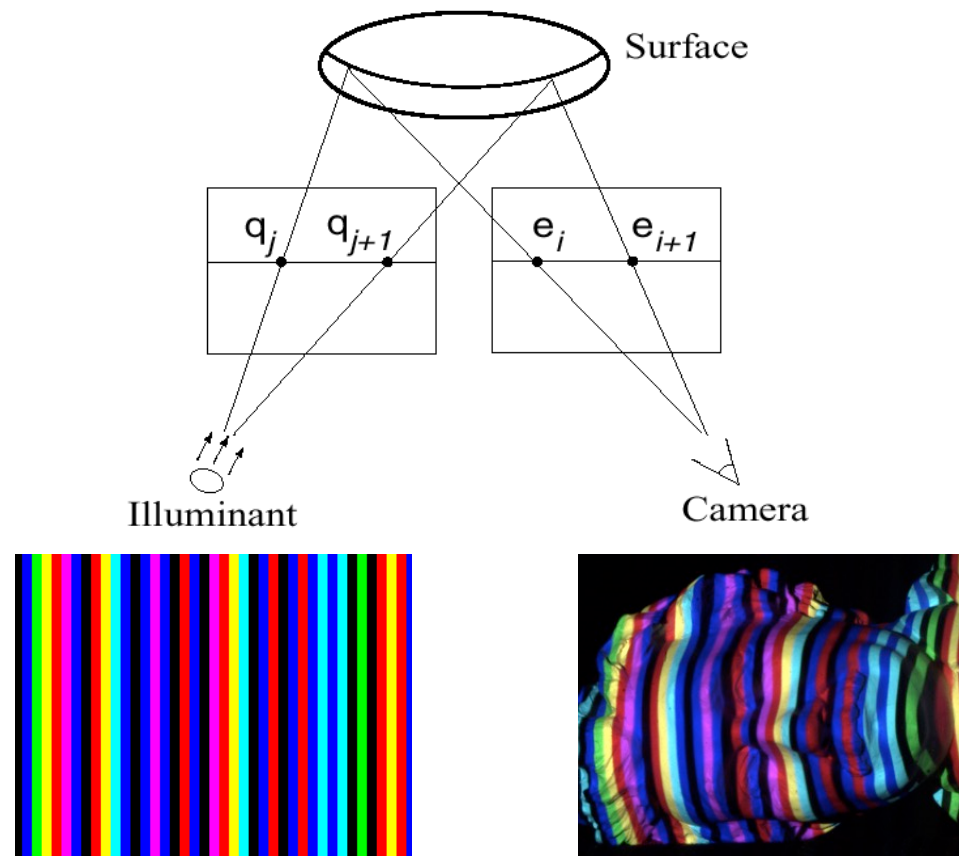
L. Zhang, B. Curless, and S. M. Seitz. [Rapid Shape Acquisition Using Color Structured Light and Multi-pass Dynamic Programming.](#) 3DPVT 2002

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

Active stereo with structured light



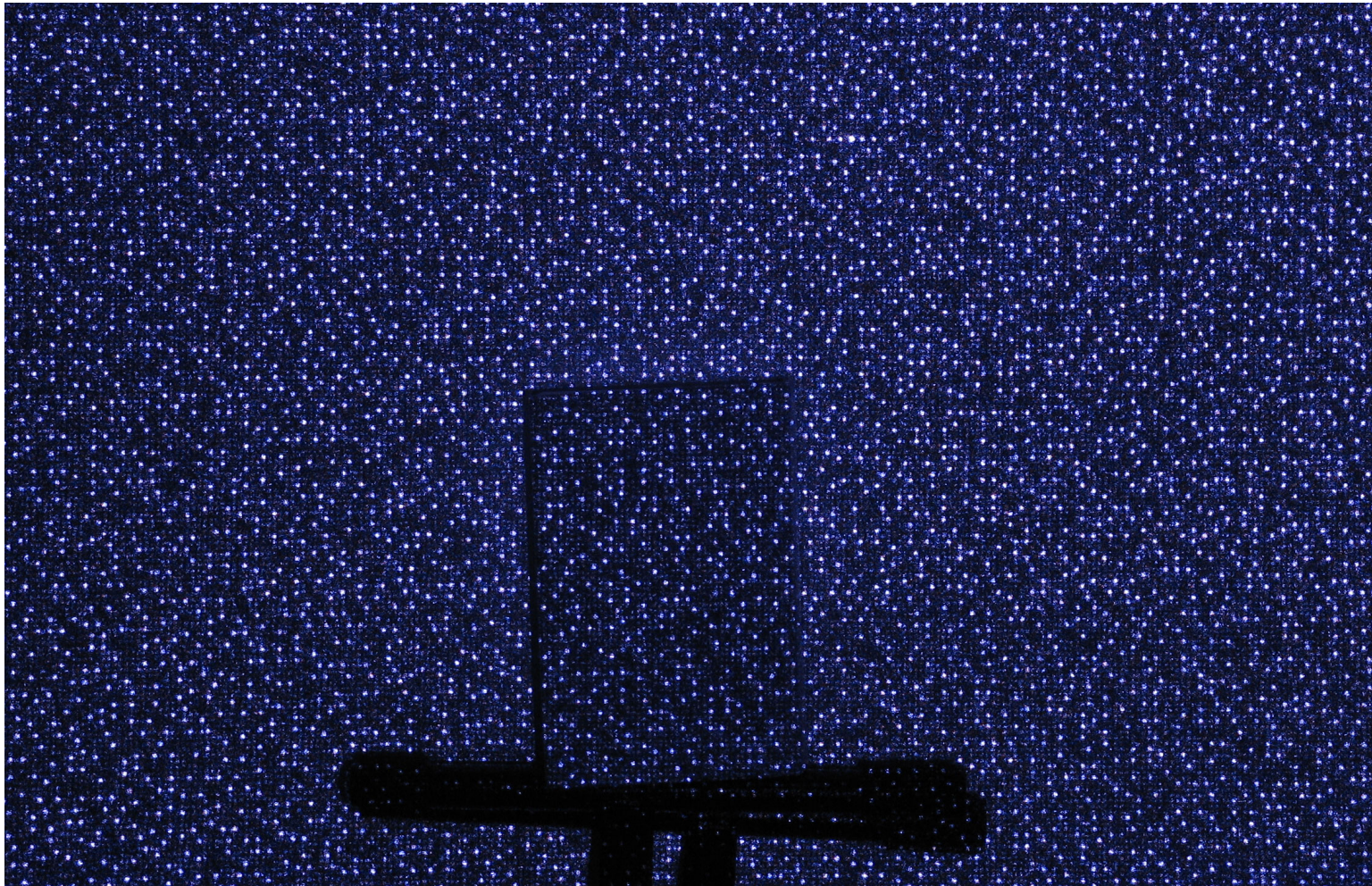
http://en.wikipedia.org/wiki/Structured-light_3D_scanner

Kinect: Structured infrared light



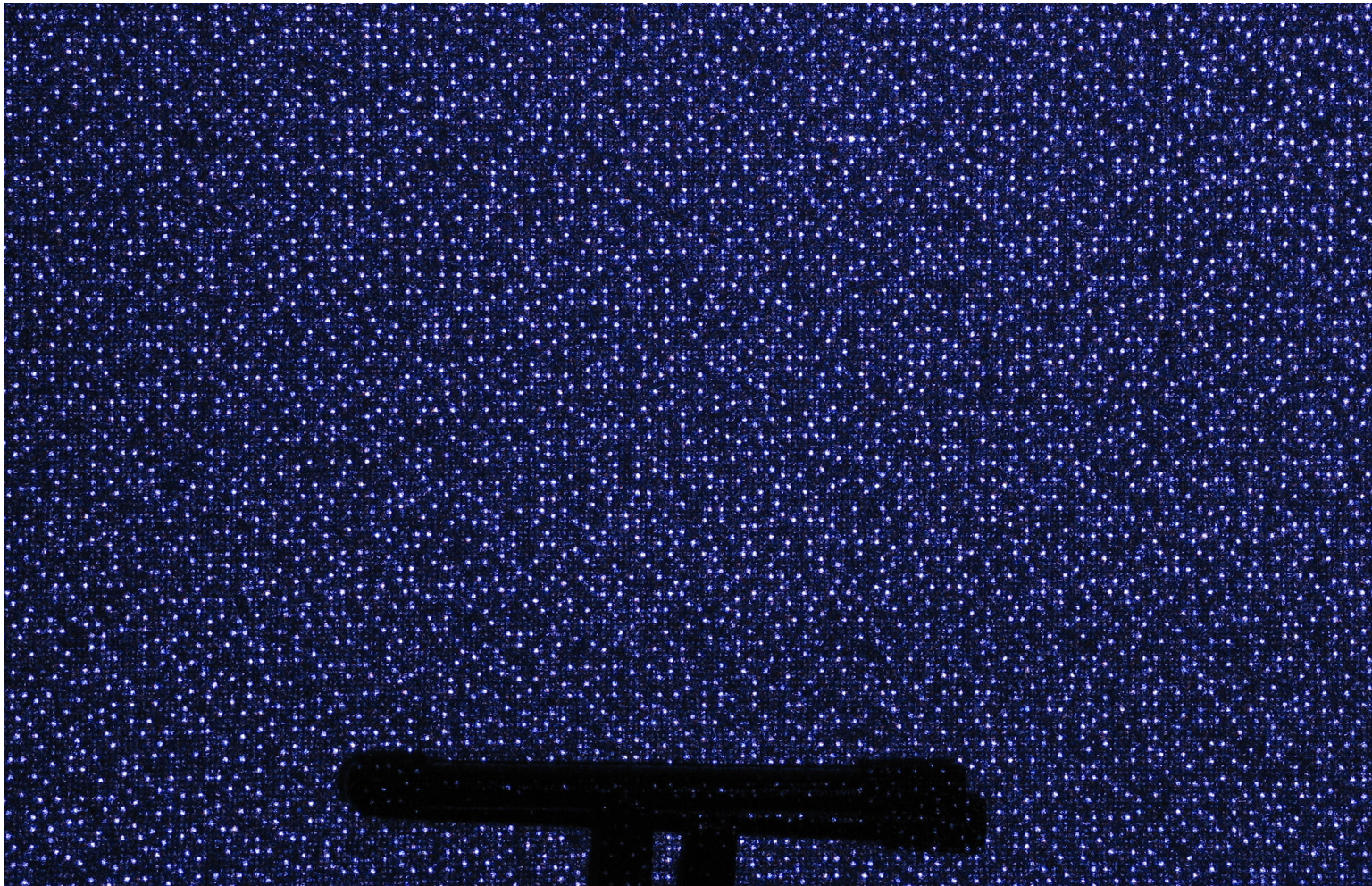
<http://bbzippo.wordpress.com/2010/11/28/kinect-in-infrared/>

Example: Book vs. No Book



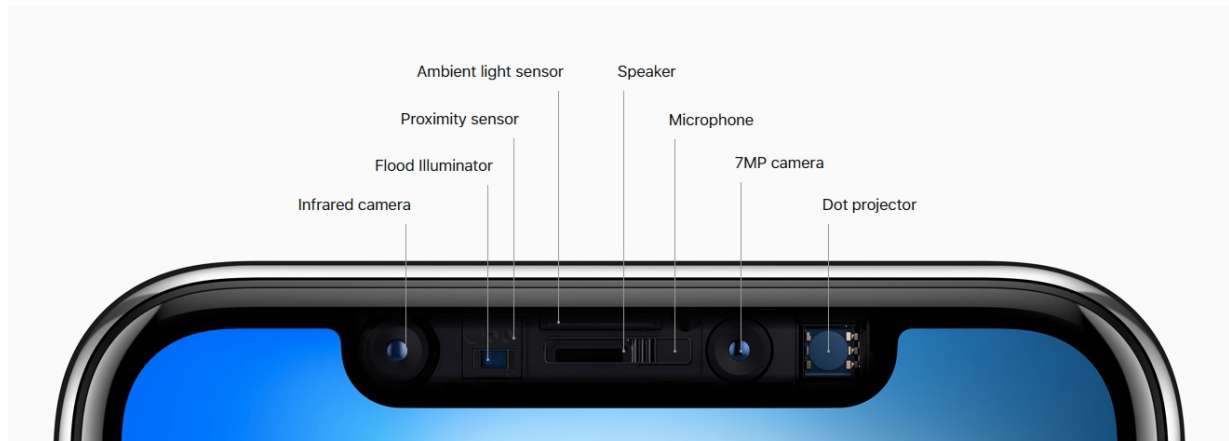
[Source](#) (via D. Hoiem)

Example: Book vs. No Book



[Source](#) (via D. Hoiem)

Apple TrueDepth



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

