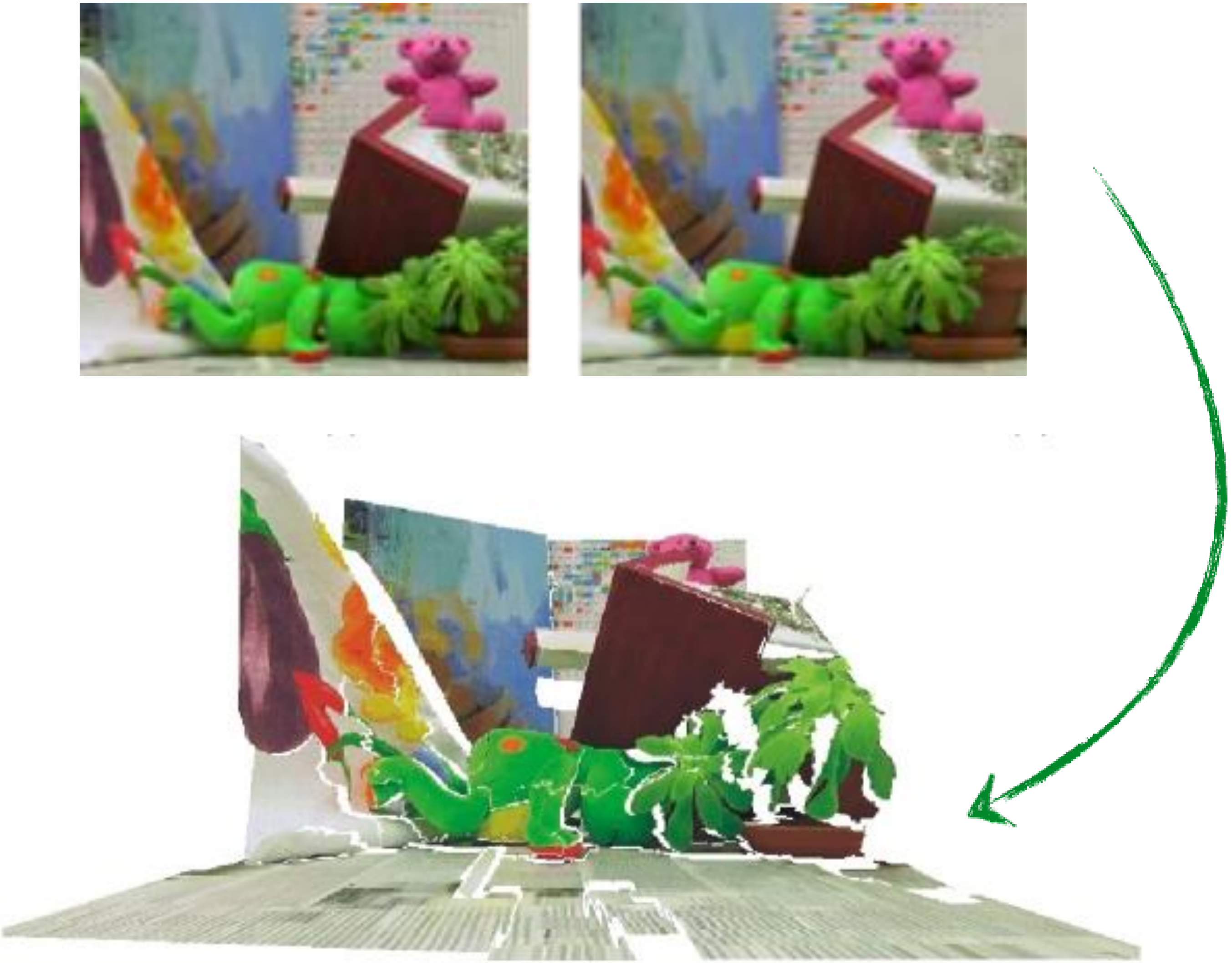
Stereo

Jiayuan Gu

gujy1@shanghaitech.edu.cn

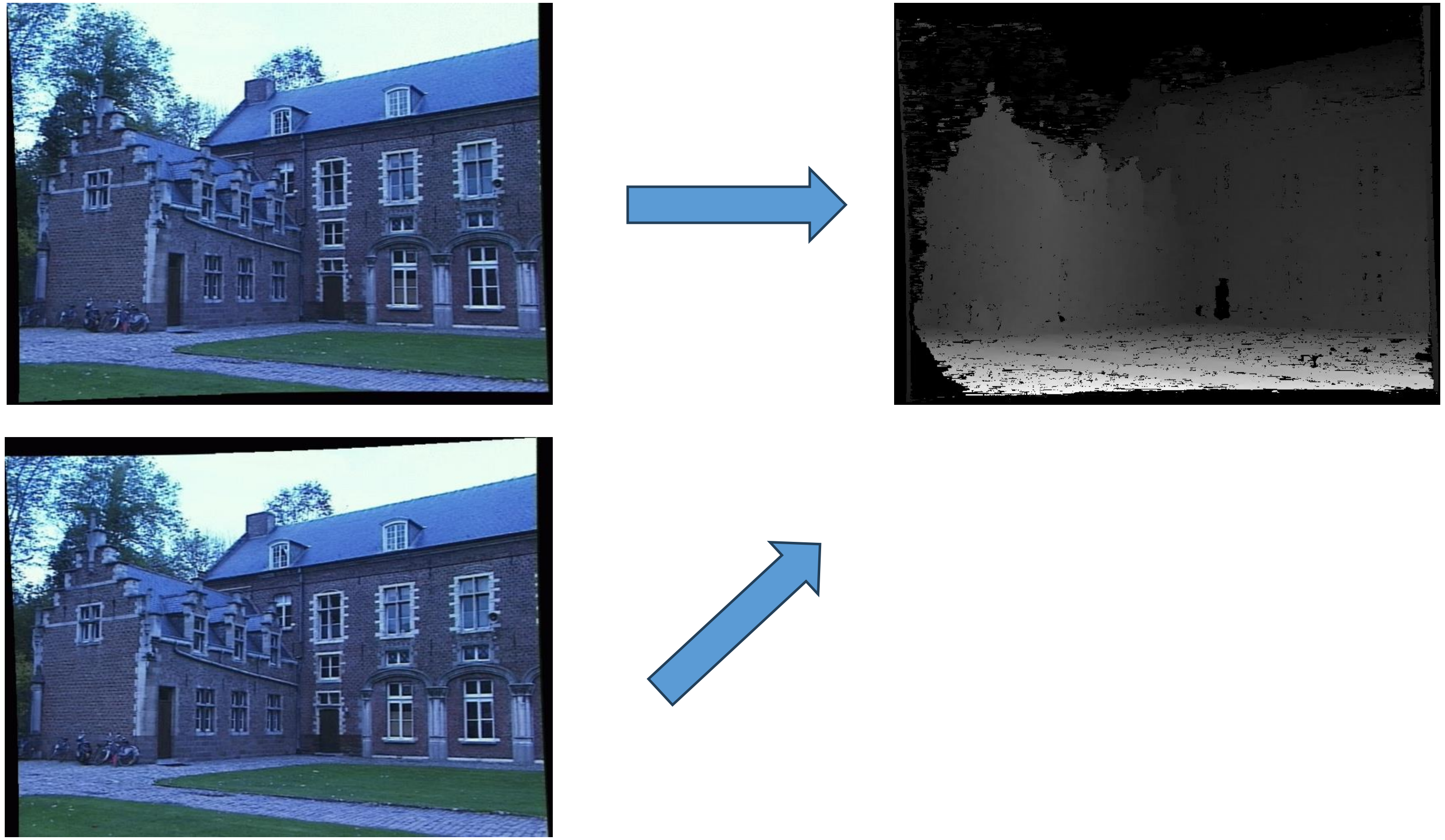
Many slides modified from Prof. Steve Seitz and Svetlana Lazebnik (CS543)

Two-View Stereo



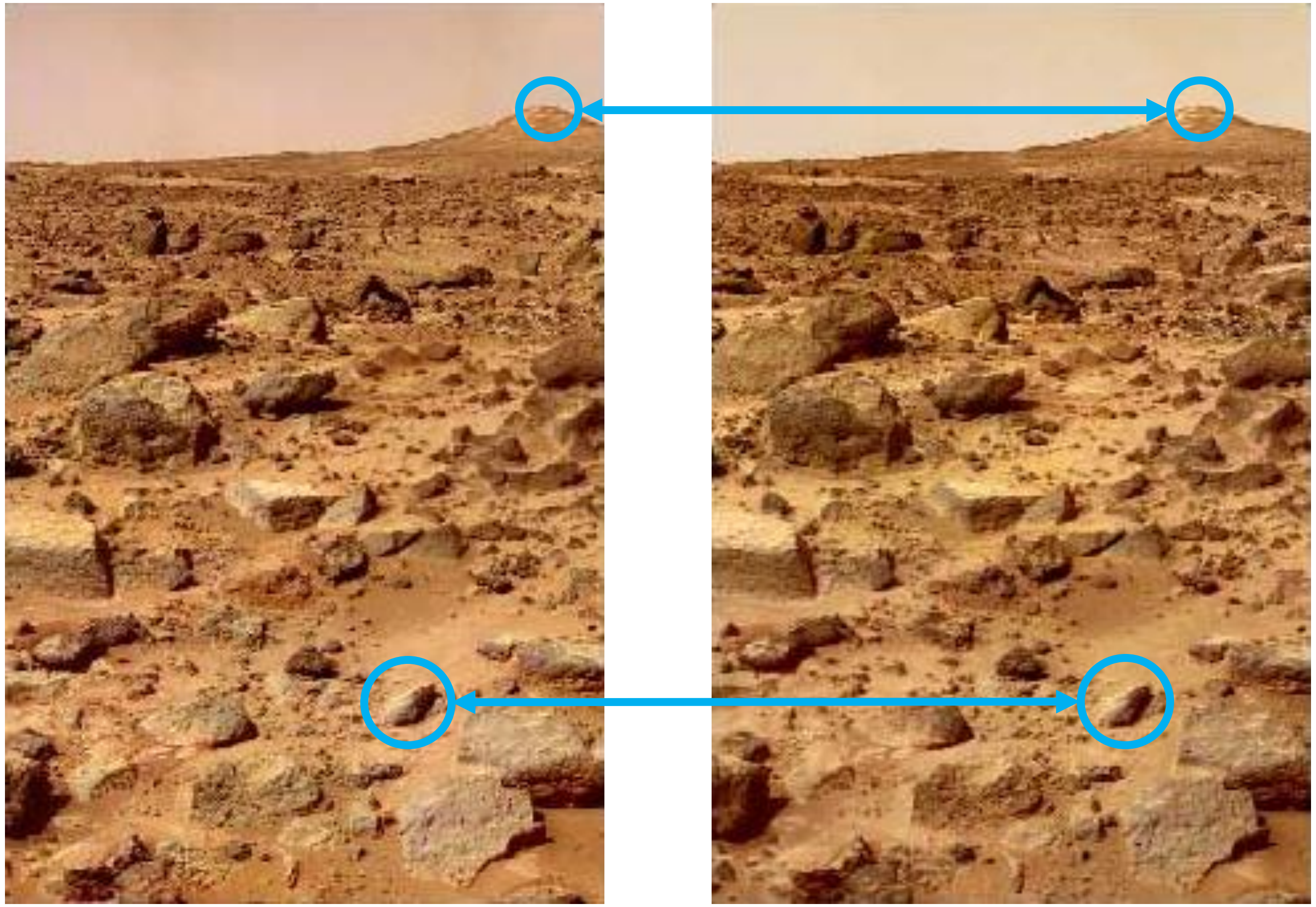
Problem Statement

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

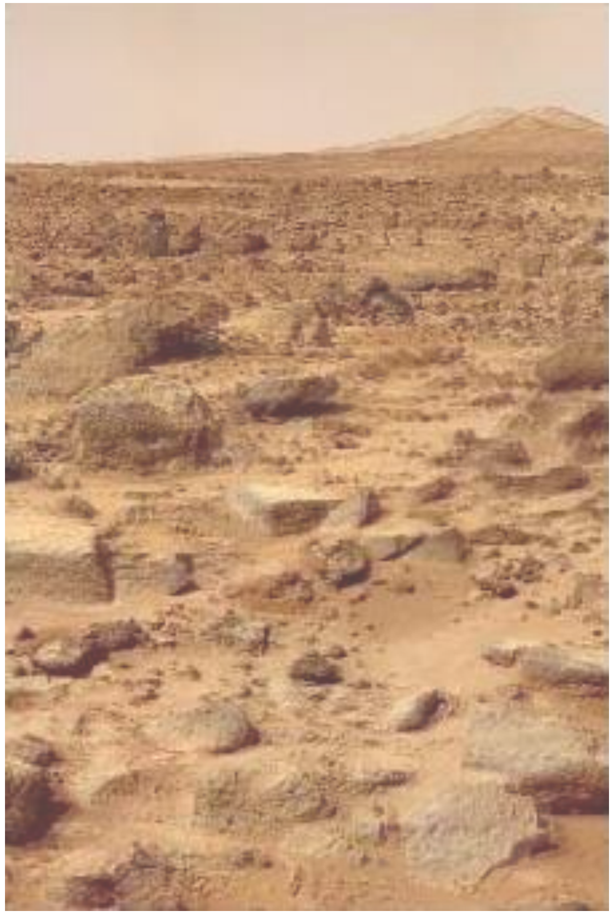


Stereo Vision and Perception of Depth

- What cues tell us about scene depth?



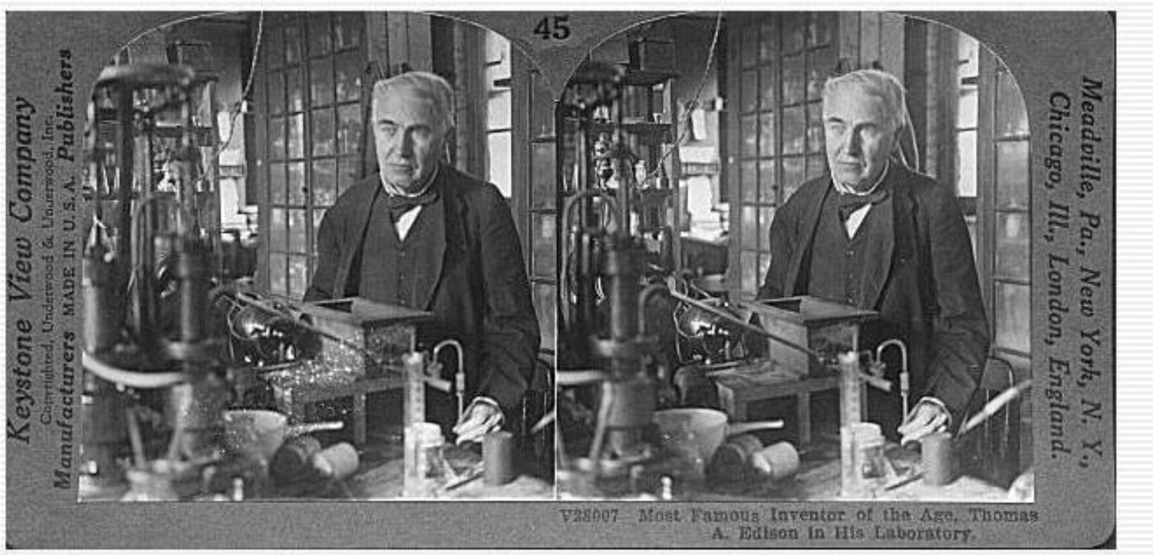
Stereo Vision and Perception of Depth



Notice the displacements of objects at different distances

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

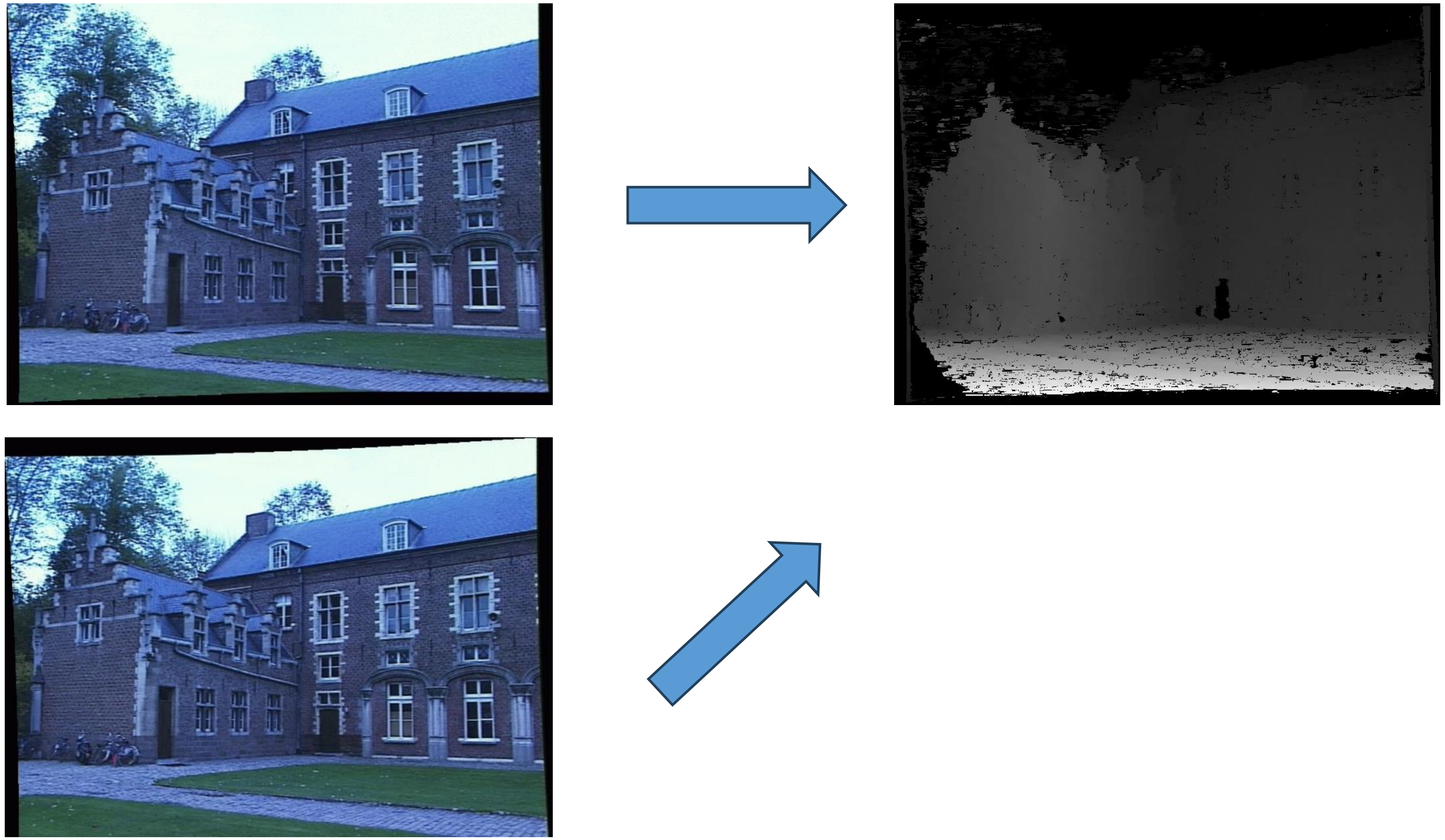
History: Stereograms

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

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

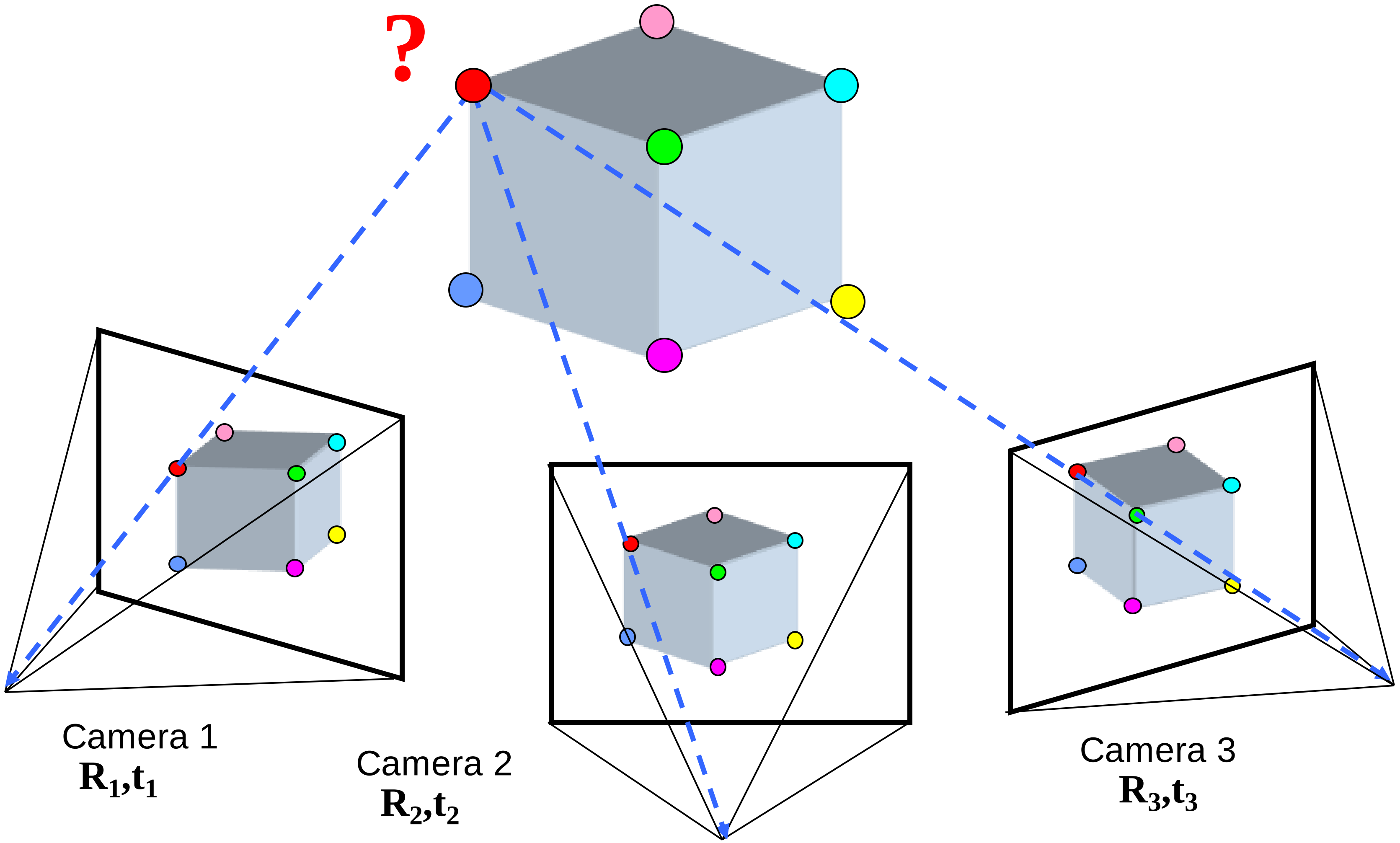
Stereo Matching

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



Review: Multi-view Geometry Problems

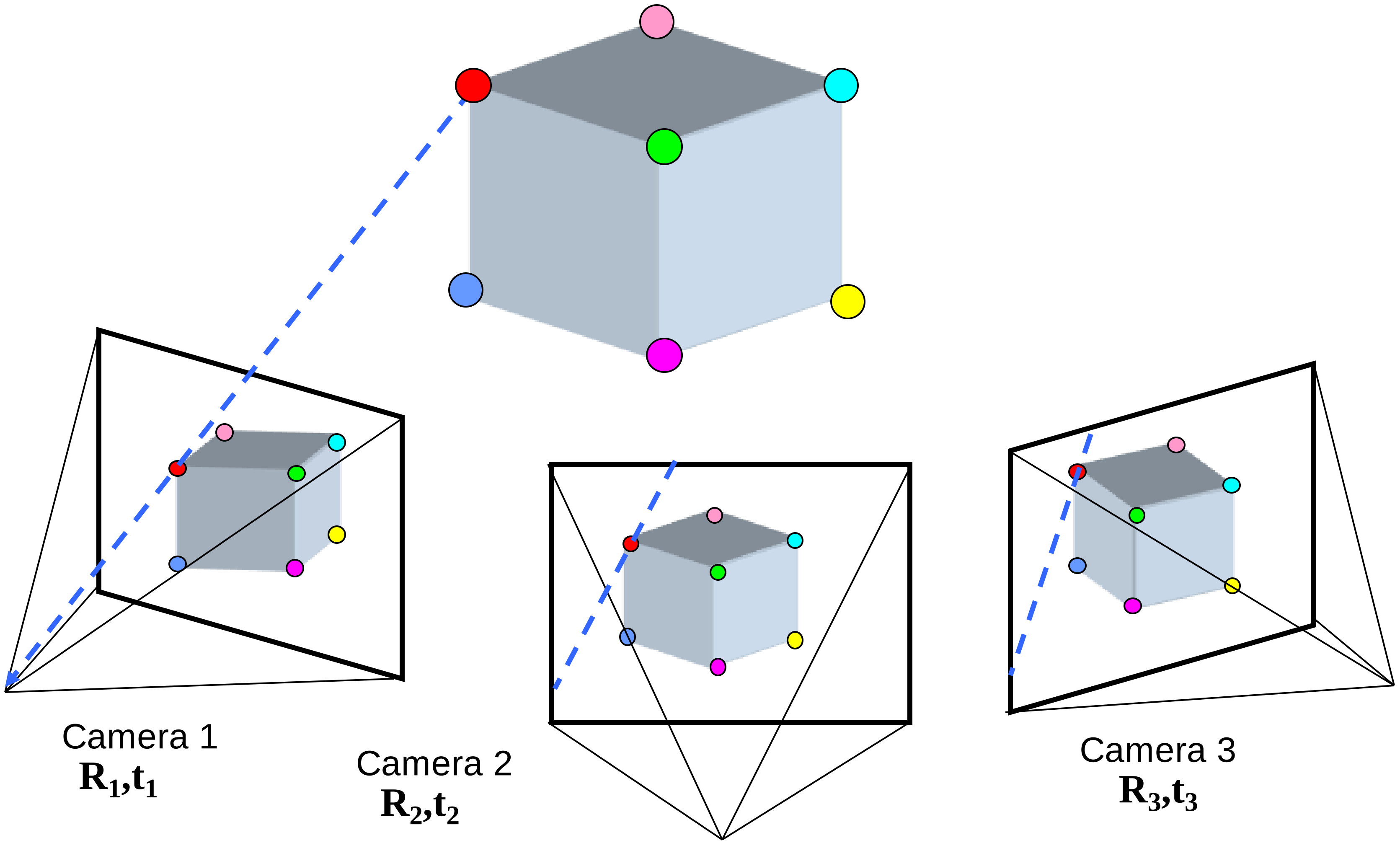
- **Structure:** Given projections of the same 3D point in two or more images, compute the 3D coordinates of that point



Slide credit: Noah Snavely

Review: Stereo Correspondence

- **Stereo correspondence:** Given a point in one of the images, where could its corresponding points be in the other images?



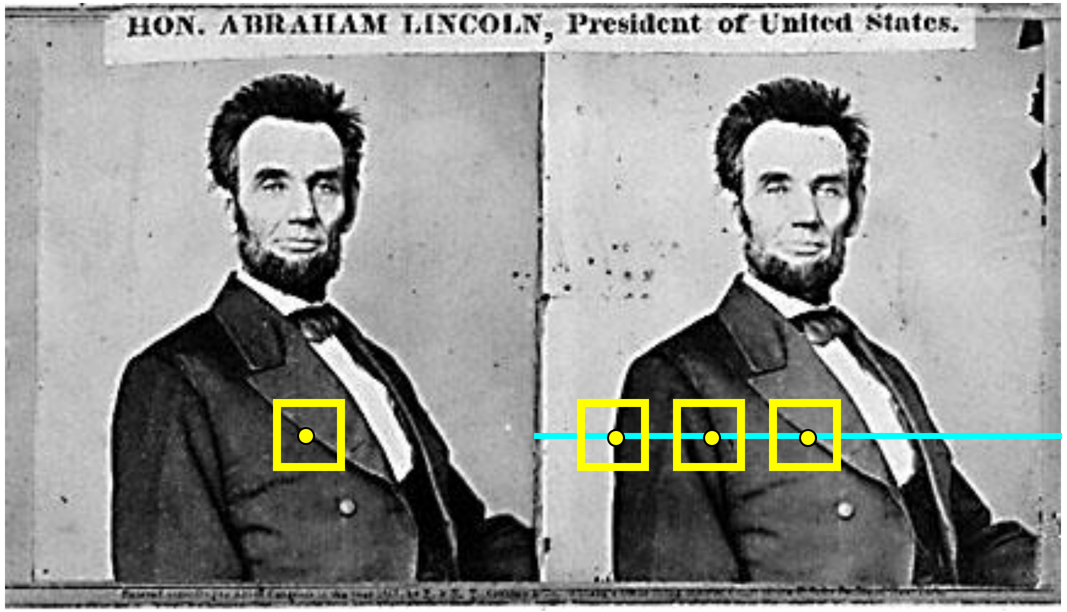
Camera 1
$\mathbf{R_1, t_1}$

Camera 2
$\mathbf{R_2, t_2}$

Camera 3
$\mathbf{R_3, t_3}$

Slide credit: Noah Snavely

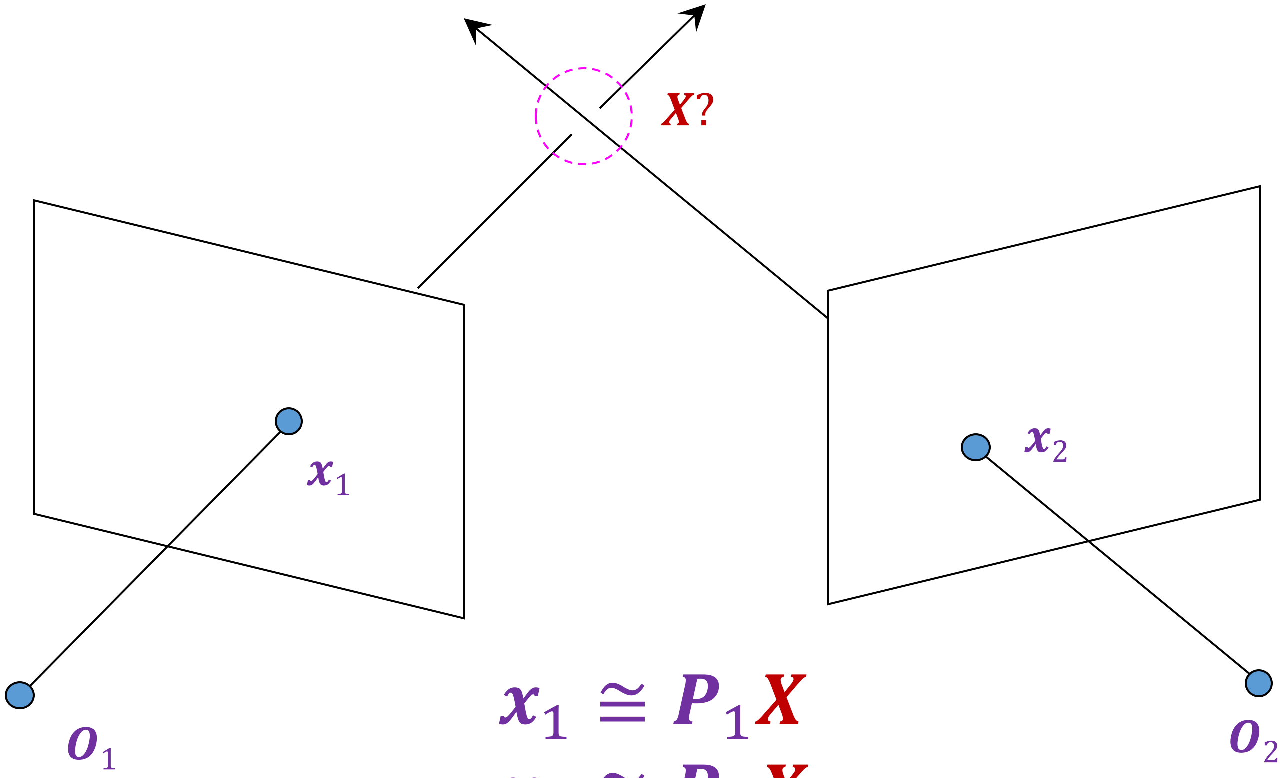
Basic Stereo Matching Algorithm



- 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?

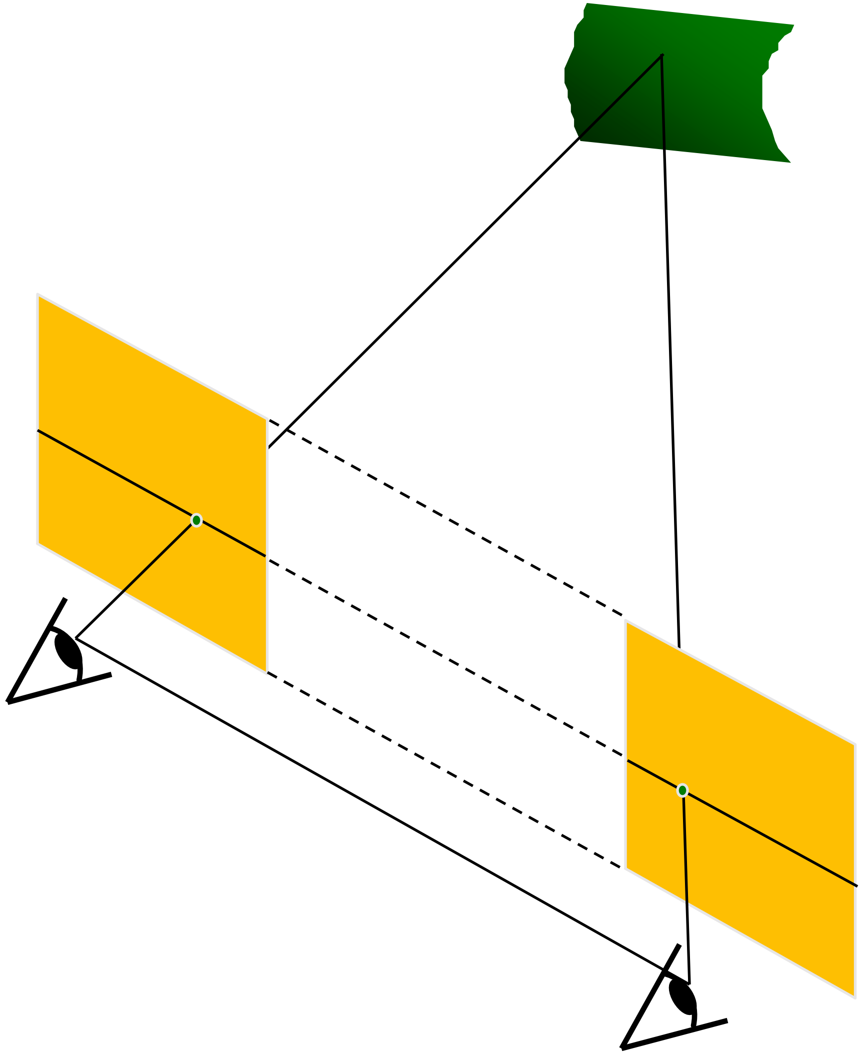
Triangulation

- Given projections of a 3D point in two or more images (with known camera matrices), find the coordinates of the point



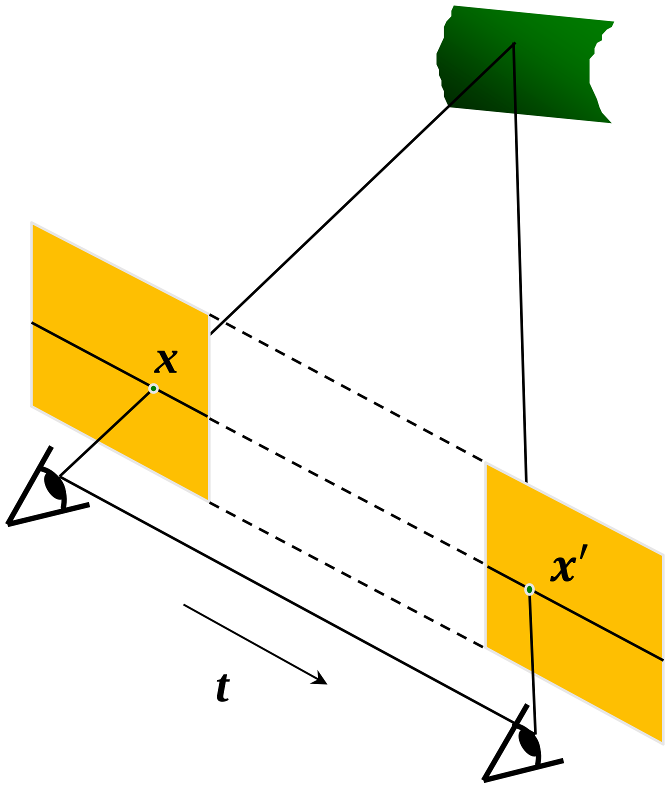
$$x_1 \cong P_1 X$$
$$x_2 \cong P_2 X$$

Simplest Case: 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

Simplest Case: Parallel Images



Let's show that $\boldsymbol{x}'$ has the same y coordinate as $\boldsymbol{x}$

Recall the essential matrix: $\boldsymbol{x}'^T \boldsymbol{E} \boldsymbol{x} = 0, \qquad \boldsymbol{E} = [\boldsymbol{t}_\times]\boldsymbol{R}$

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

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

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

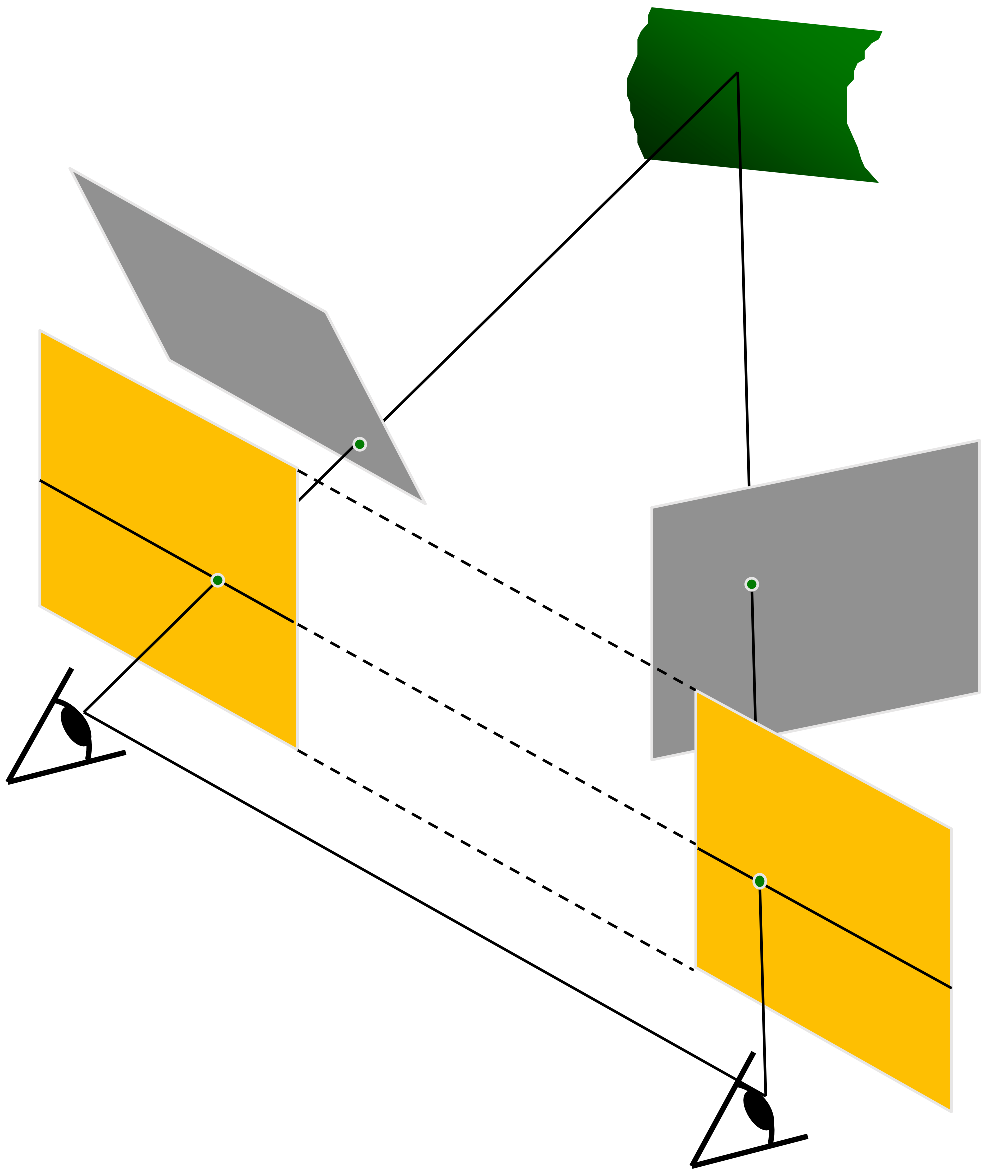
$$(u' \ v' \ 1) \begin{pmatrix} 0 \\ -t \\ tv \end{pmatrix} = 0$$

$$-tv + tv' = 0$$

$$v = v'$$

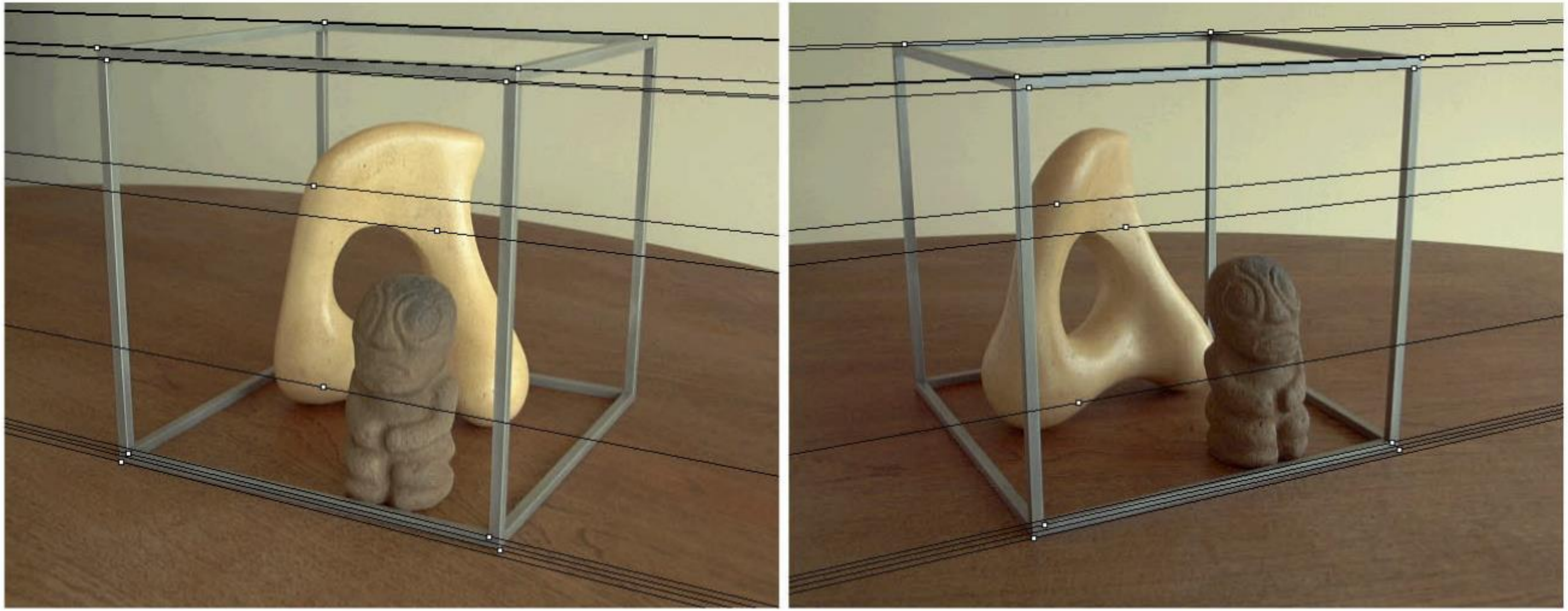
Stereo image rectification

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



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

Stereo Image Rectification



Before Rectification

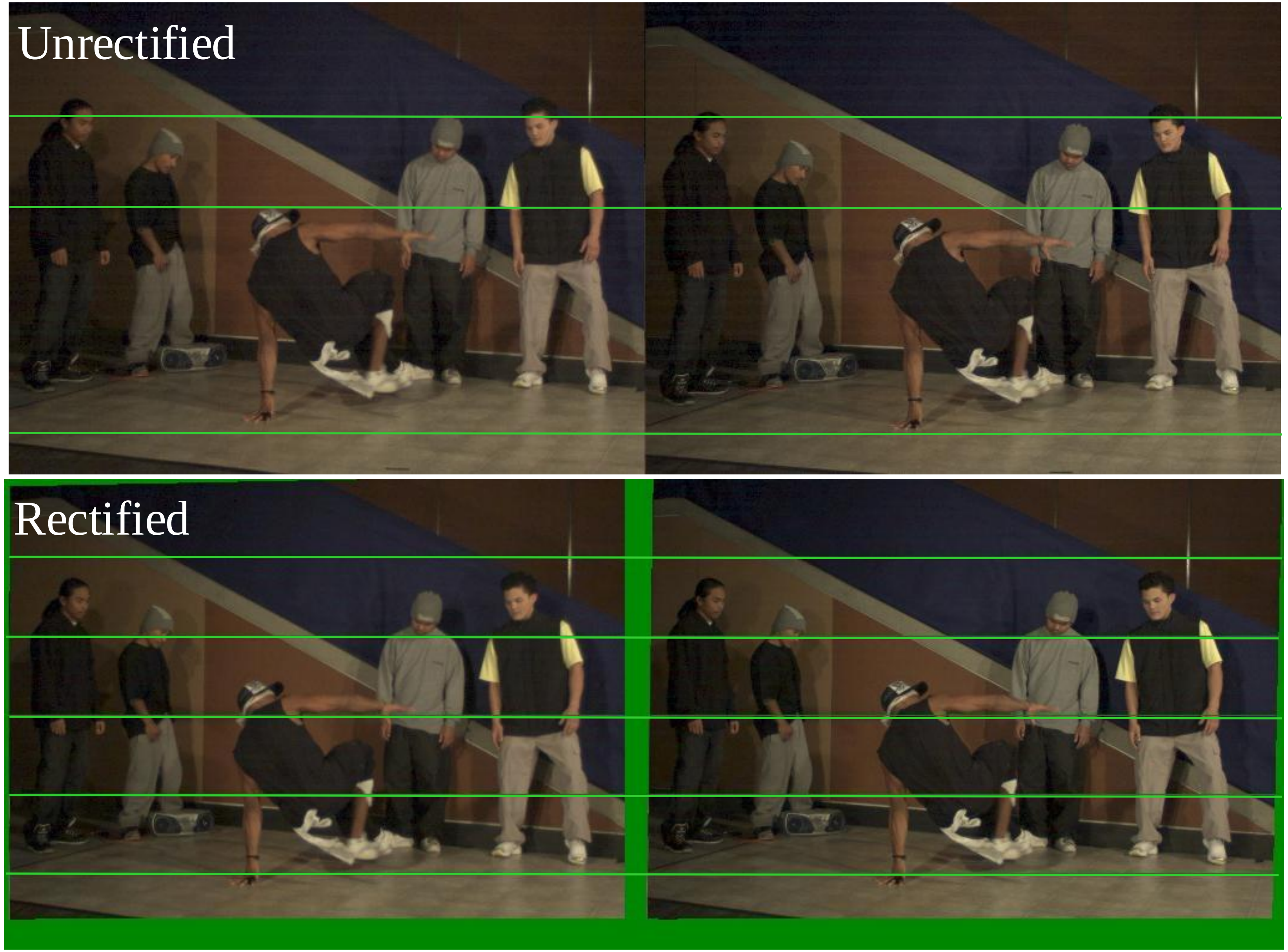
Image source

Stereo Image Rectification

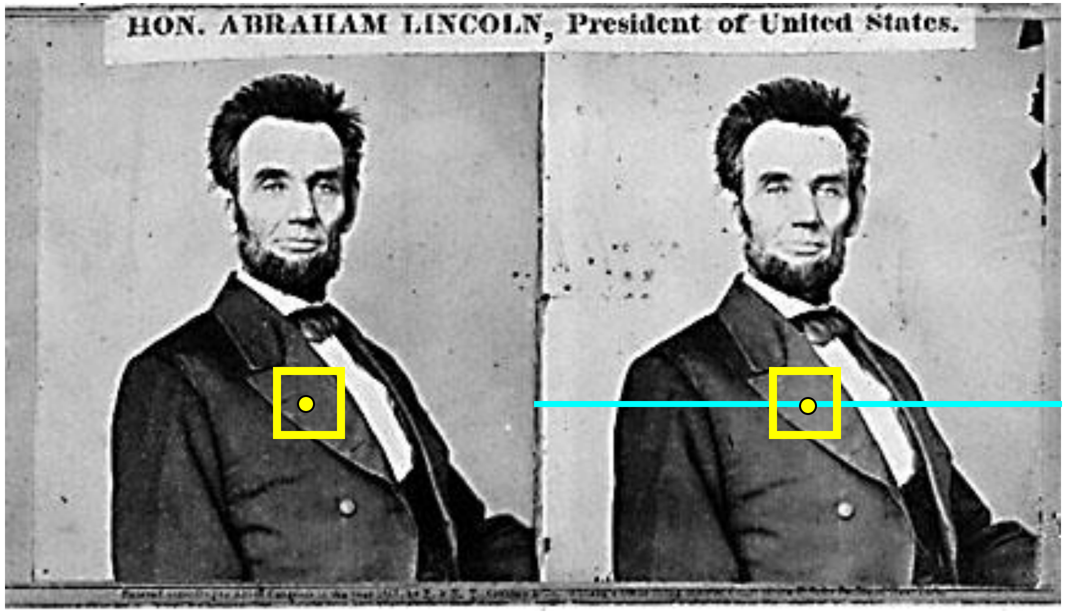
After Rectification

Image source

Another Example of Rectification

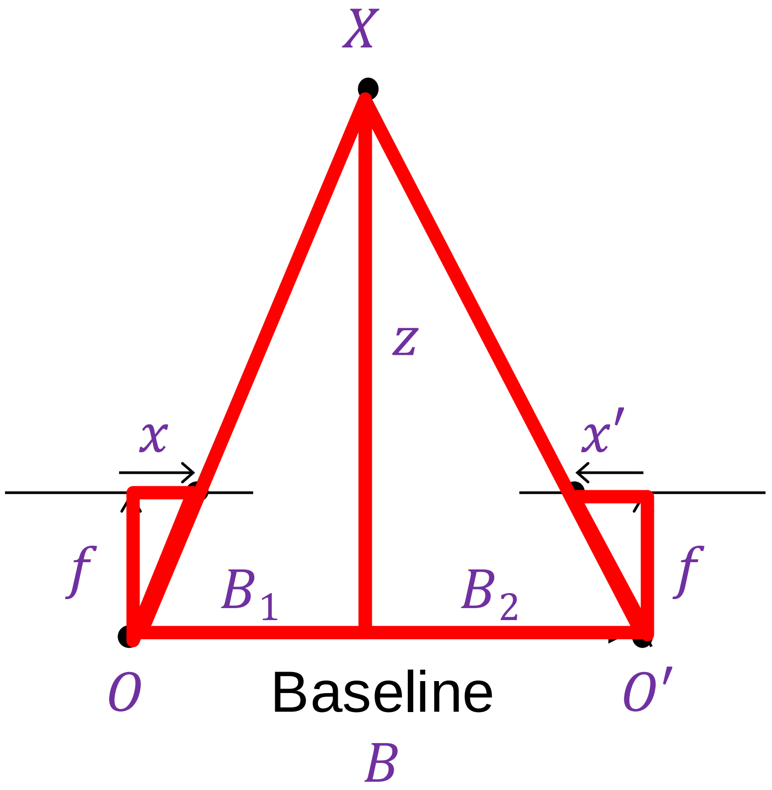


Basic stereo matching algorithm



- 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



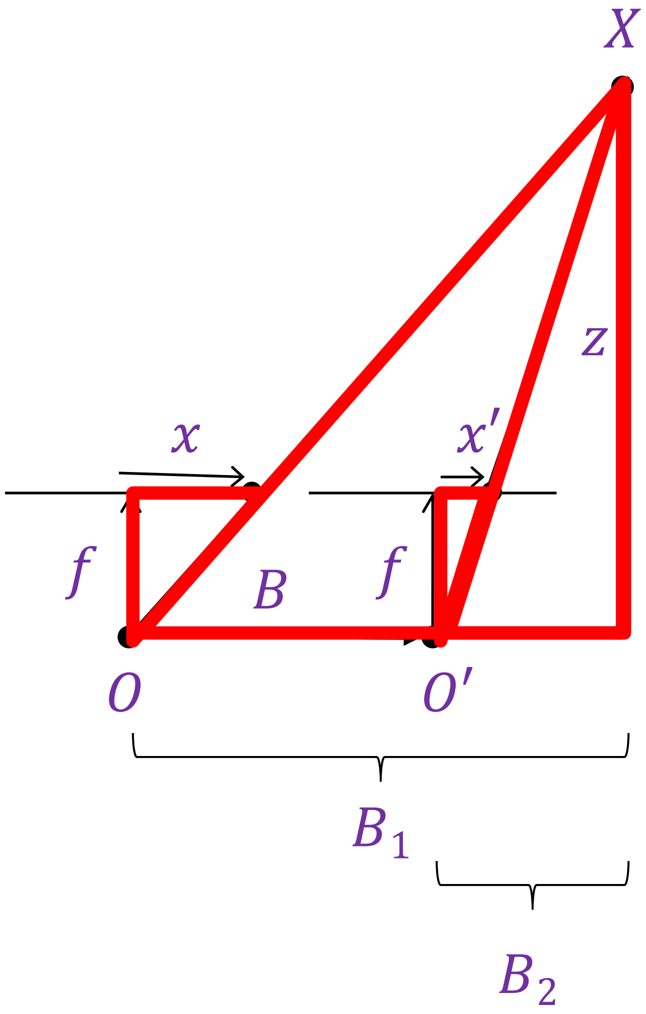
$$\frac{x}{f} = \frac{B_1}{z} \qquad \frac{-x'}{f} = \frac{B_2}{z}$$

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

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

Disparity is inversely proportional to depth!

Depth from Disparity



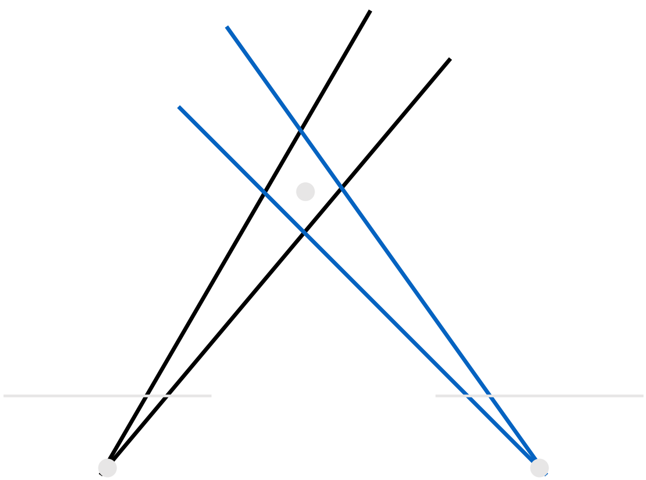
$$\frac{x}{f} = \frac{B_1}{z} \qquad \frac{x'}{f} = \frac{B_2}{z}$$

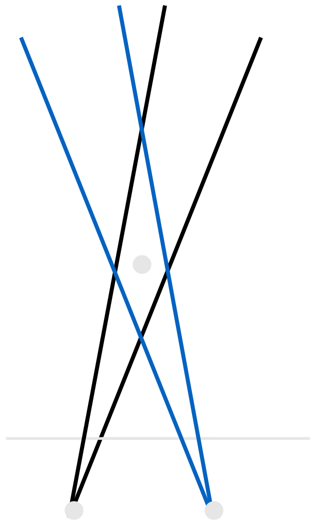
$$\frac{x - x'}{f} = \frac{B_1 - B_2}{z}$$

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

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

Effect of Baseline on Stereo Results

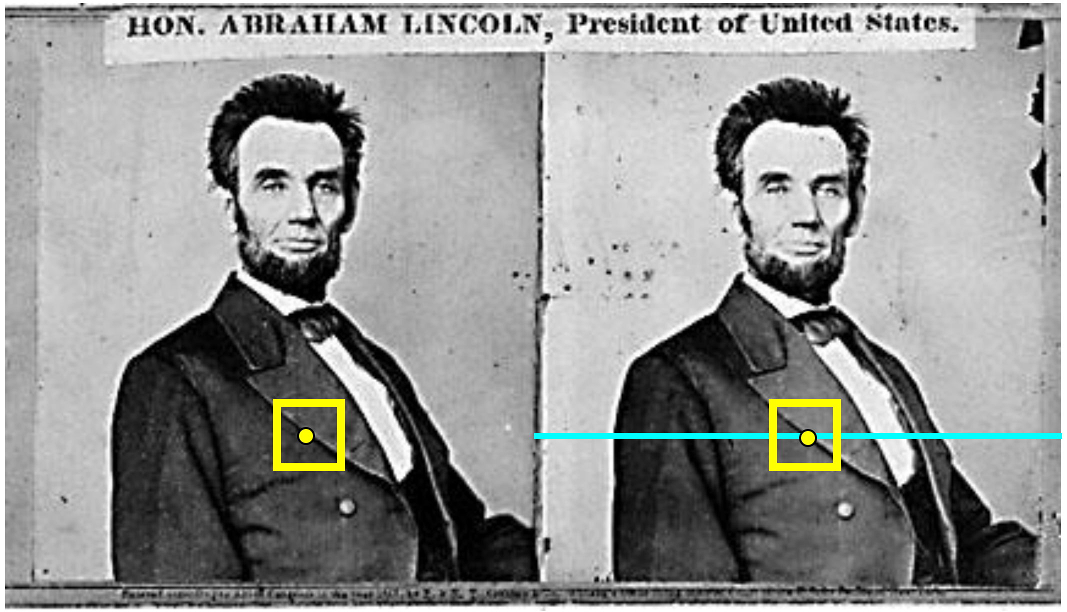




- Larger baseline
  + Smaller triangulation error
  – Matching is more difficult

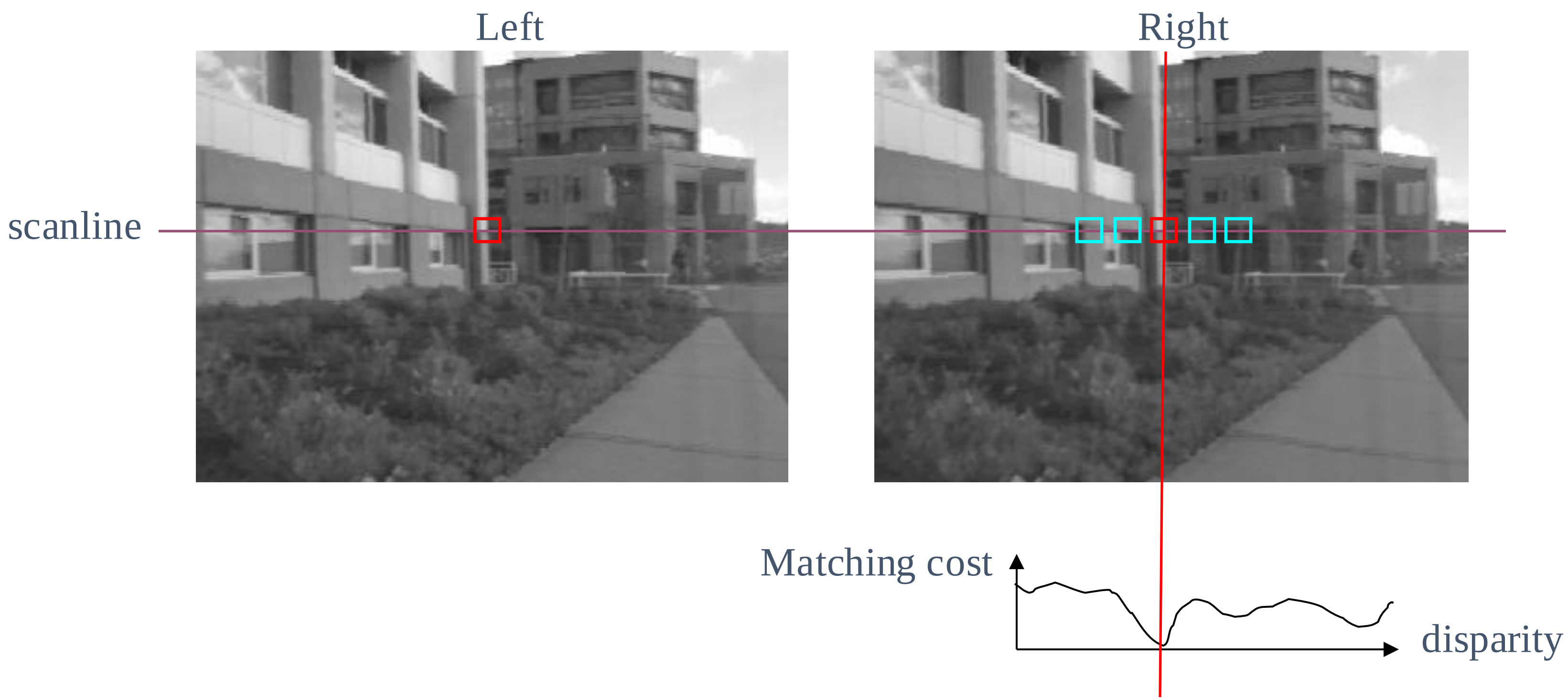
- Smaller baseline
  – Higher triangulation error
  + Matching is easier

Basic stereo matching algorithm



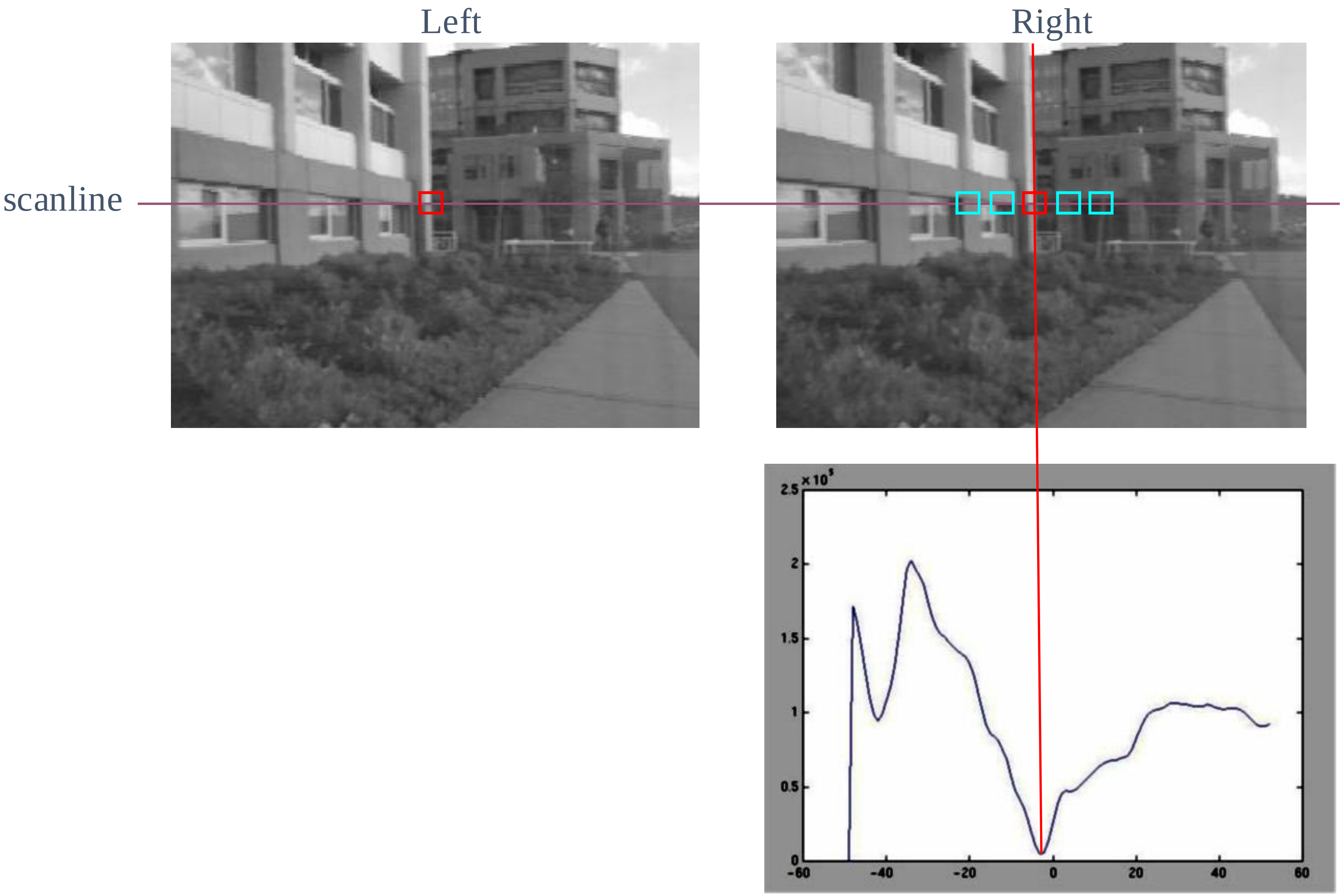
- If necessary, rectify the two stereo images to transform epipolar lines into scanlines
- For each pixel x in the first image
  - Find corresponding epipolar scanline in the right image
  - Examine all pixels on the scanline and pick the best match x'
  - Compute disparity $x - x'$ and set $\mathrm{depth}(x)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

Search Correspondence



- 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

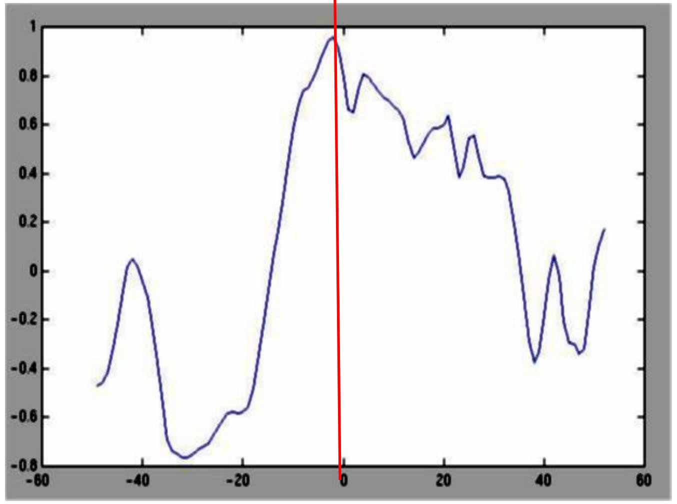
Search Correspondence



SSD: the sum of squared differences

Search Correspondence

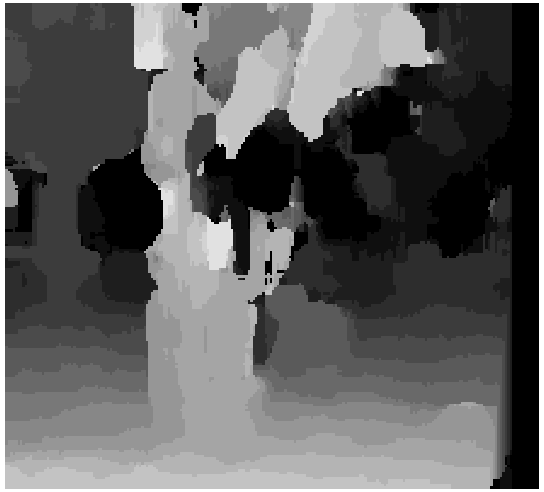
$$\gamma = \frac{\sum_x \sum_y (w'(x,y) - \bar{w}')(w(x,y) - \bar{w})}{\sqrt{\left(\sum_x \sum_y (w'(x,y) - \bar{w}')^2\right)\left(\sum_x \sum_y (w(x,y) - \bar{w})^2\right)}}$$



Normalized correlation

Effect of Window Size on Correspondence Search

Window size 3

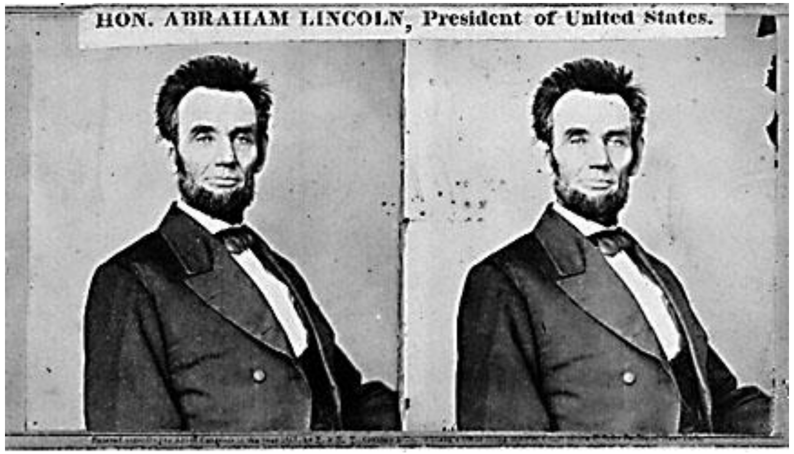


Window size 20

Smaller window:
+ More details
- More noise

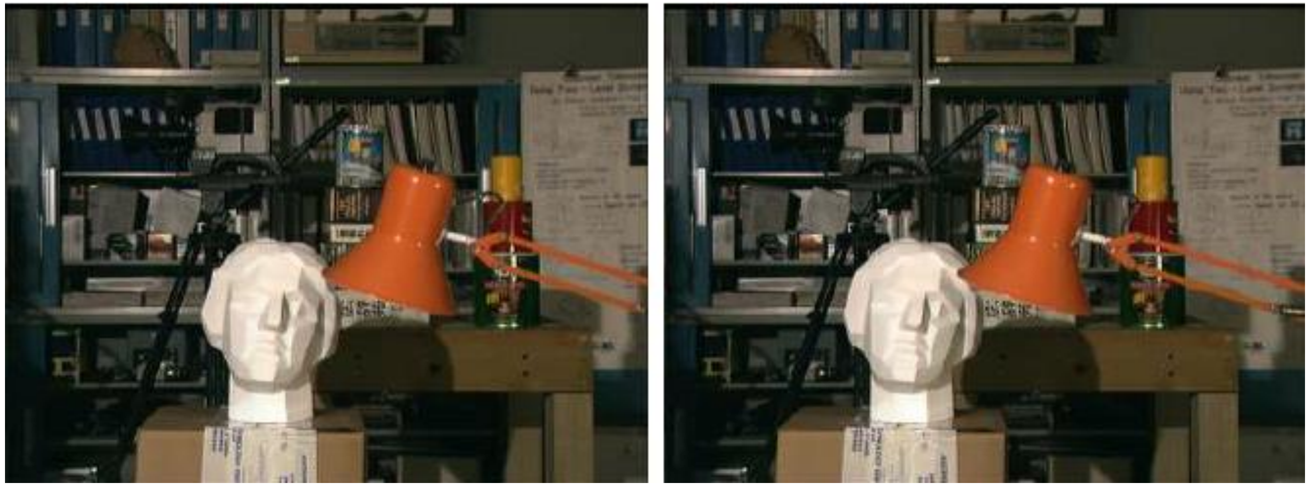
Larger window:
+ Smoother disparity maps
- Fewer details

When will Basic Window Search Fail?



Textureless surfaces

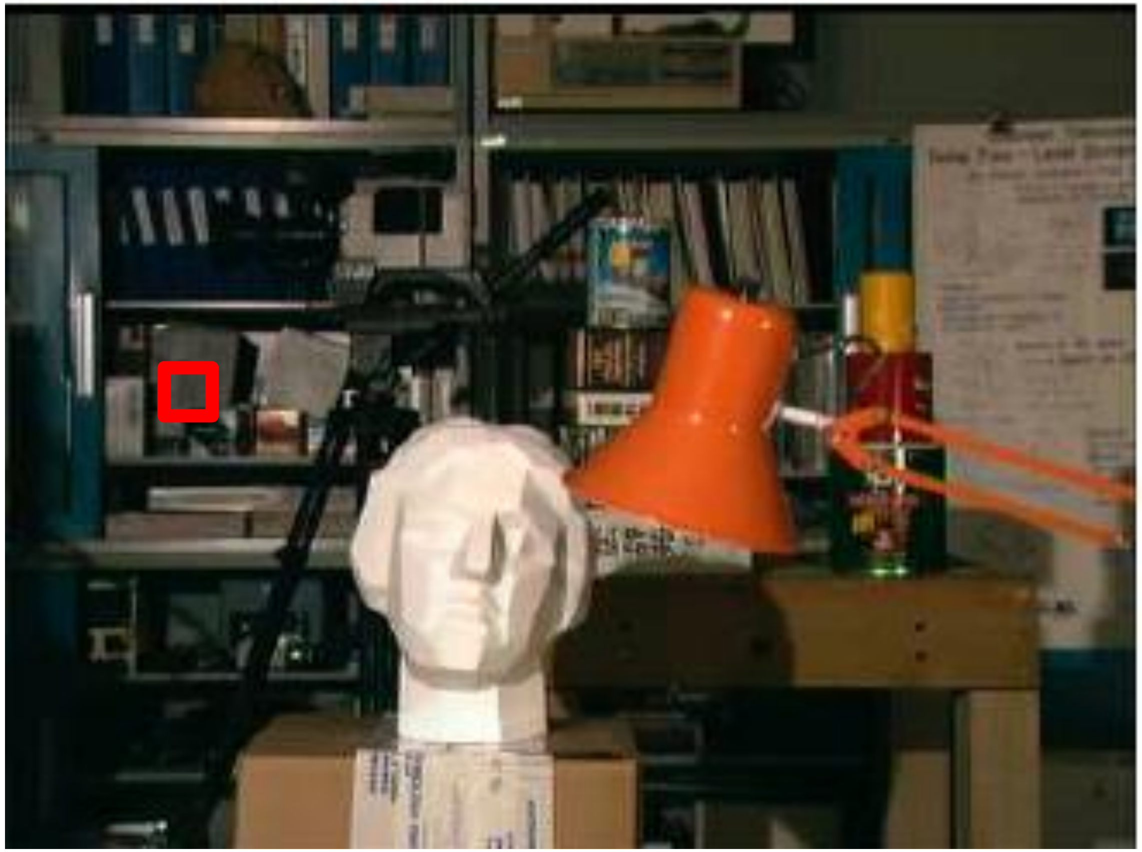
Occlusions, repetition

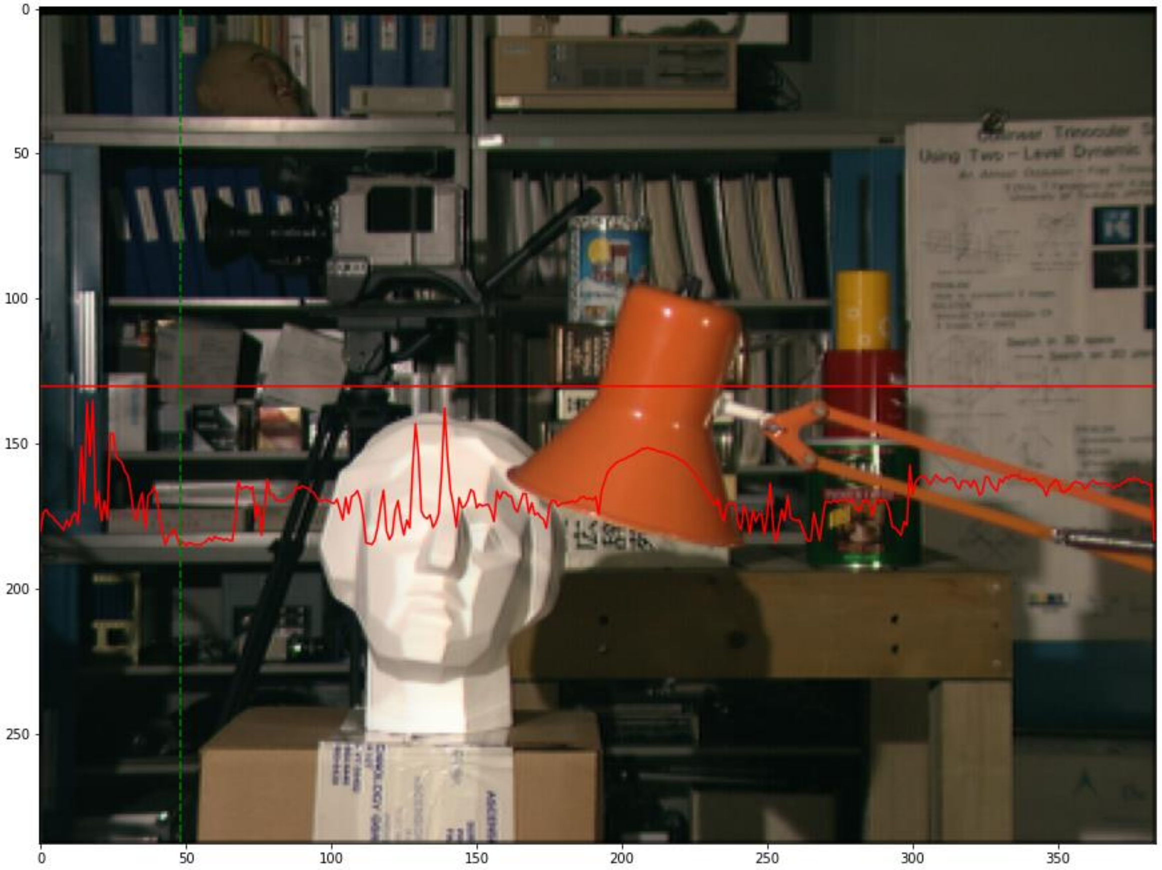


Non-Lambertian surfaces, specularities

Example: Textured Neighborhood

Window size: 1 pixel

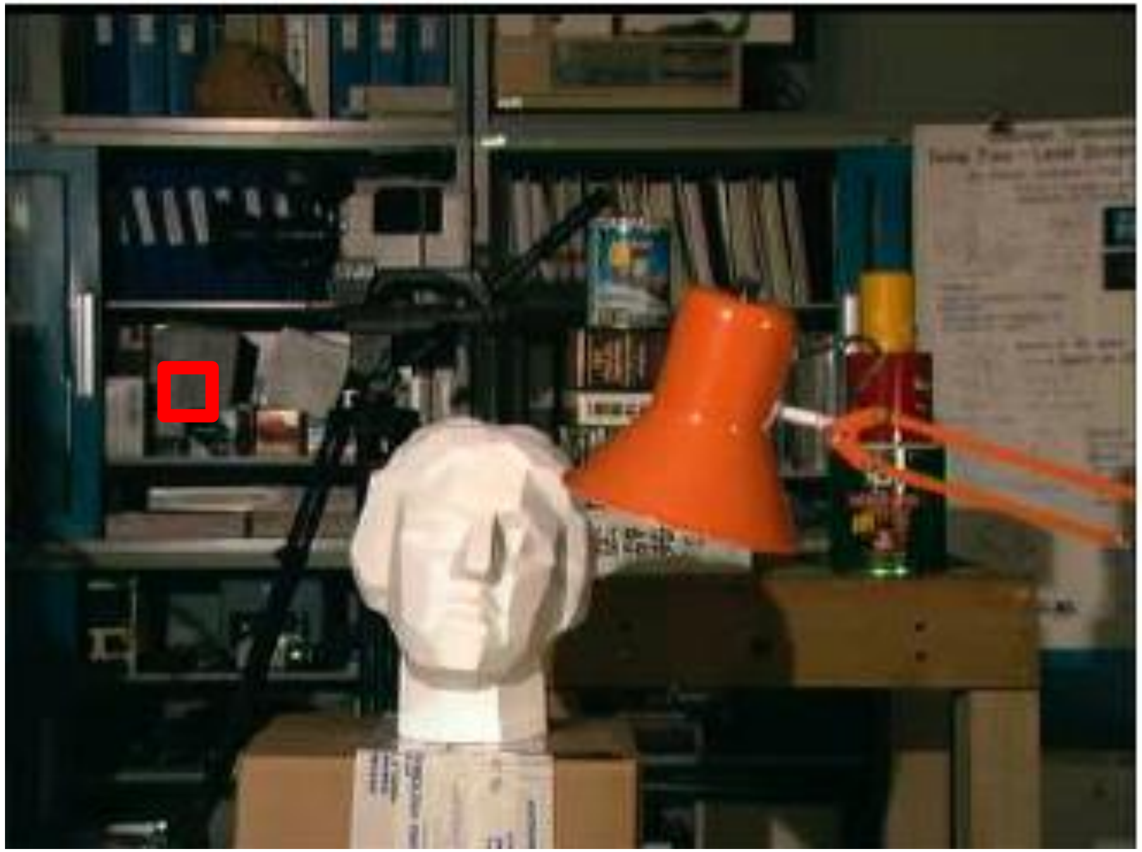




Source: [D. Hoiem]()

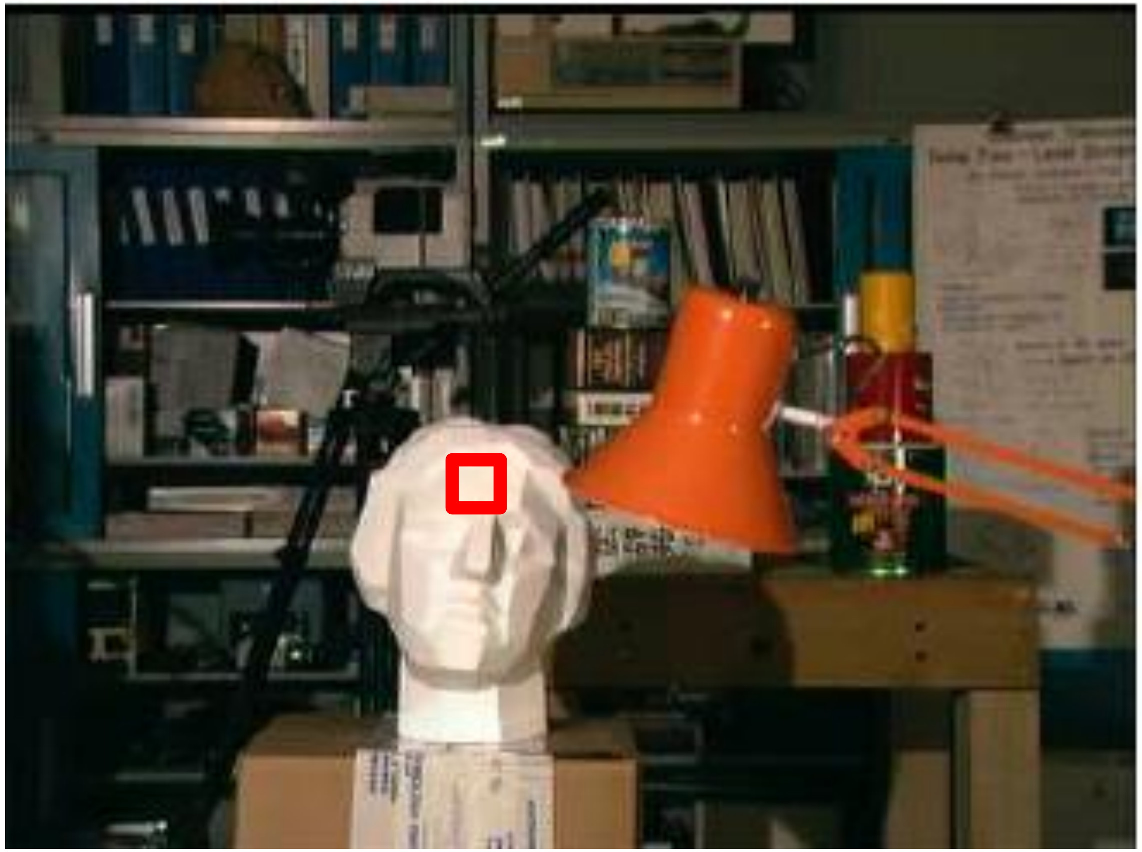
Example: Textured Neighborhood

Window size: 7 pixels



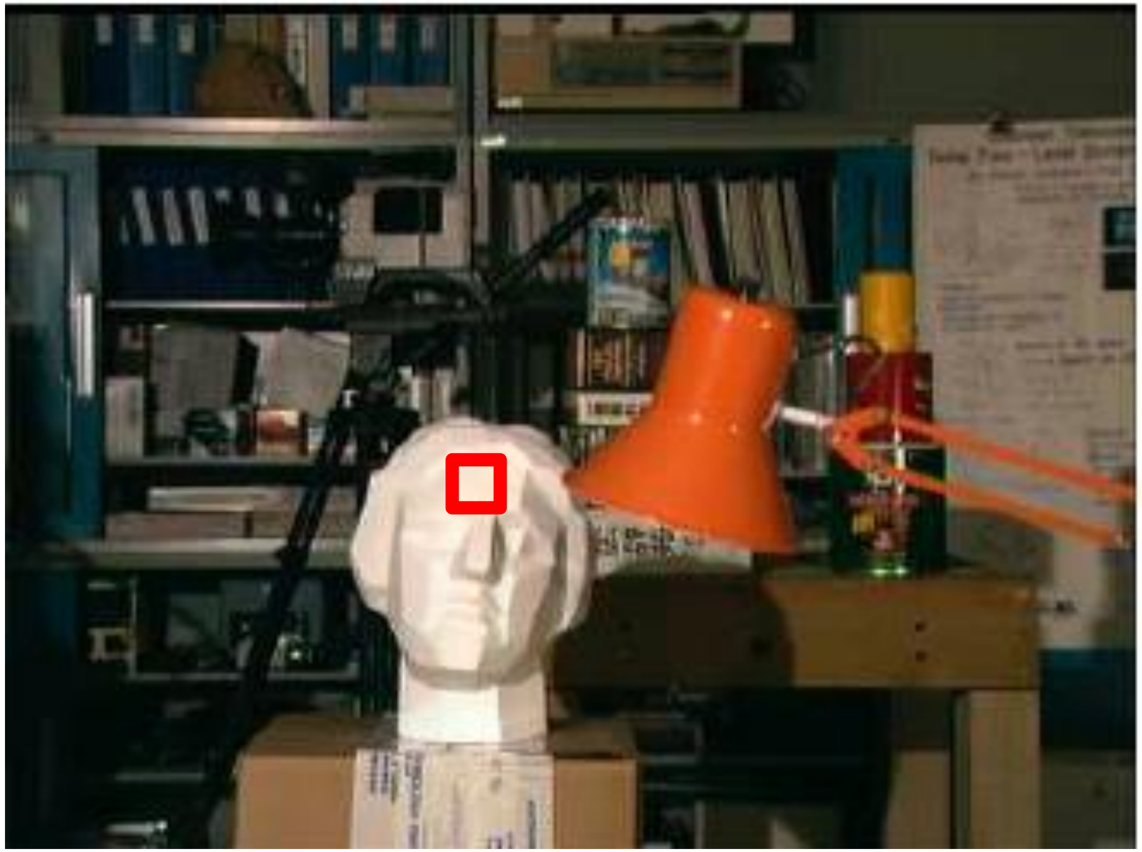
Example: Textureless Neighborhood

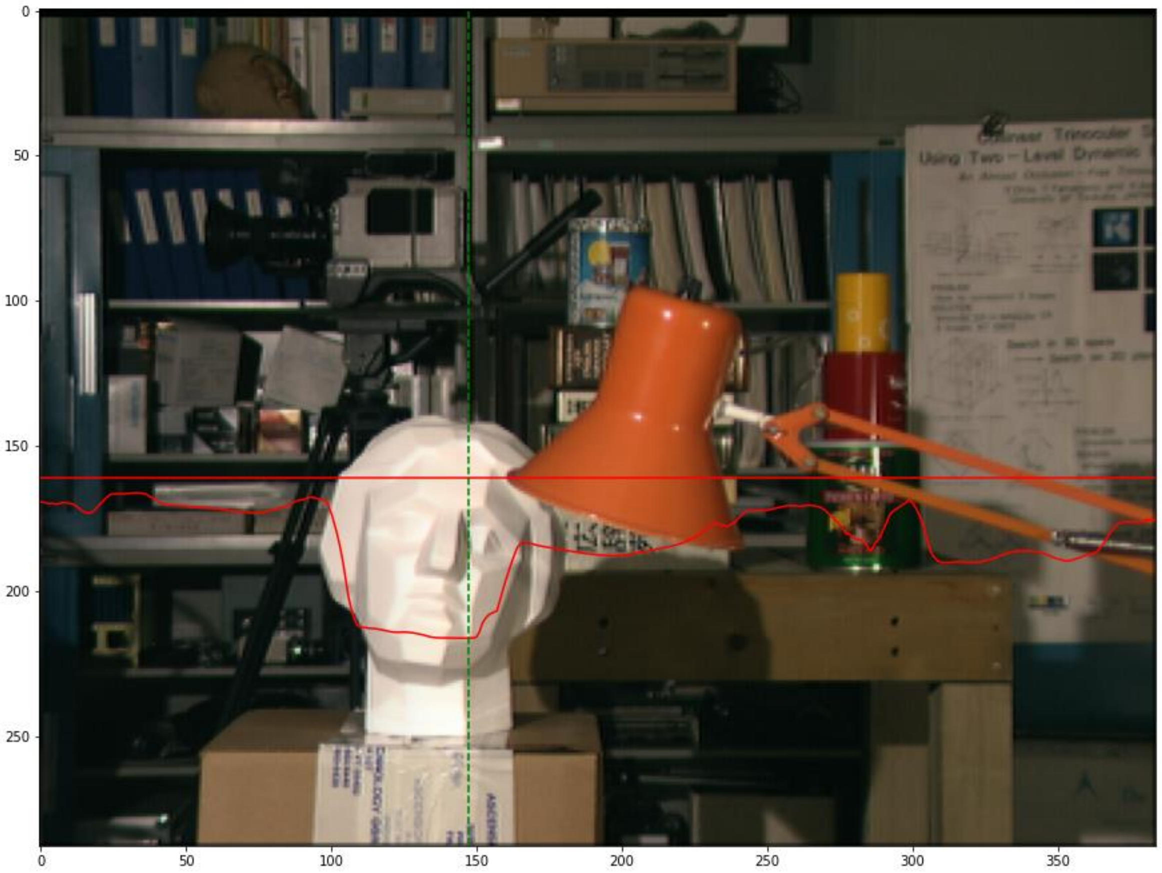
Window size: 1 pixel



Example: Textureless Neighborhood

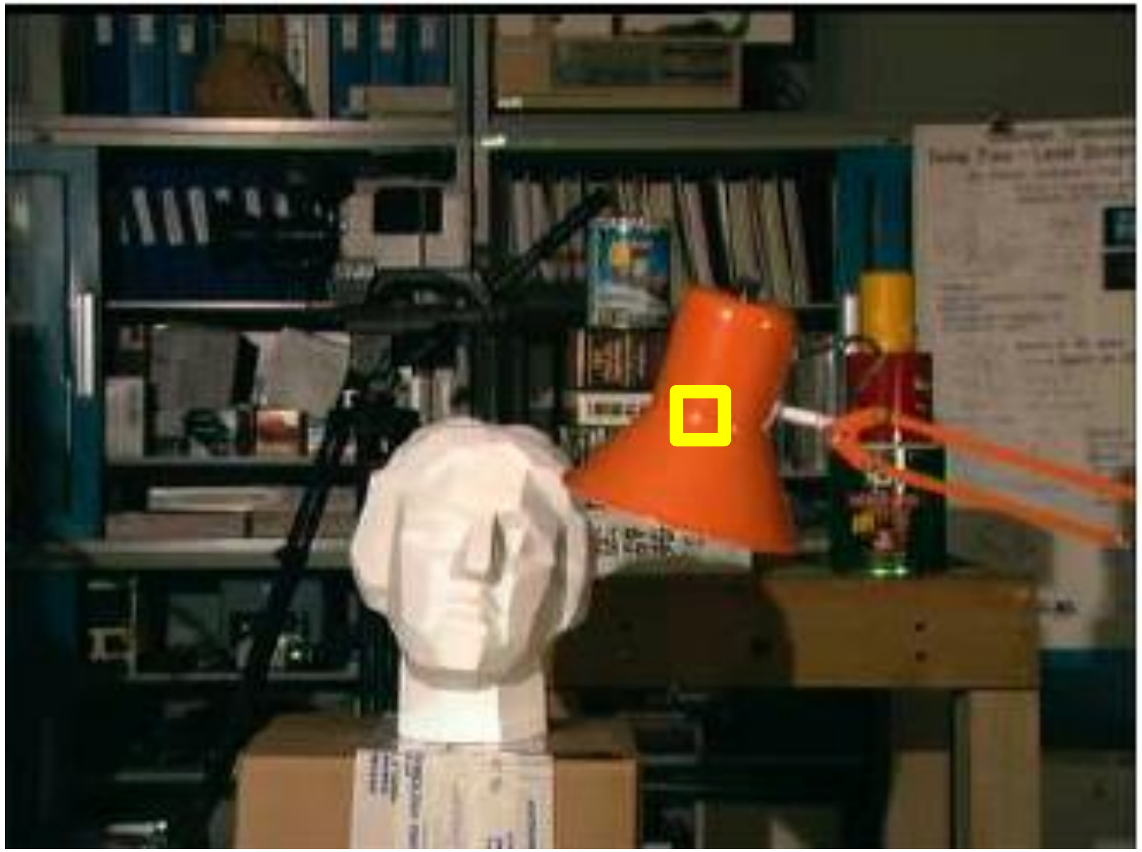
Window size: 7 pixels





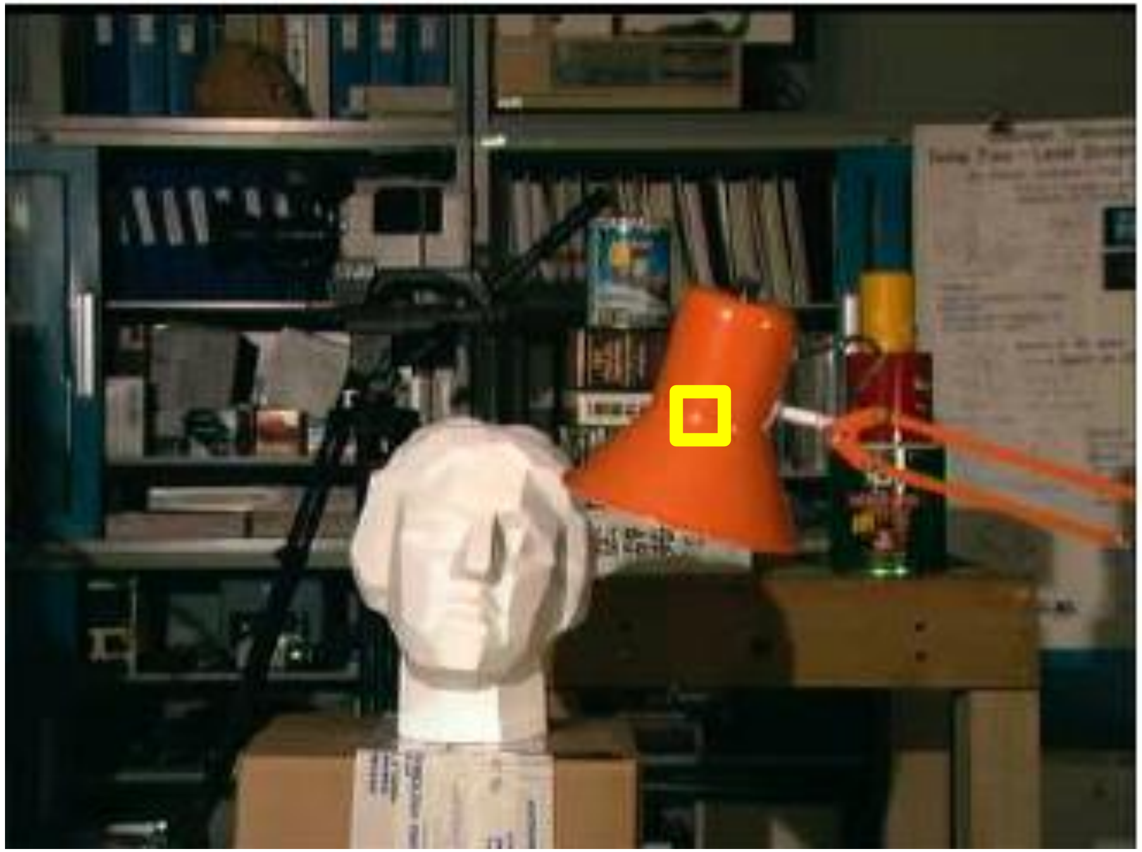
Example: Specular Highlight

Window size: 1 pixel



Example: Specular Highlight

Window size: 7 pixels

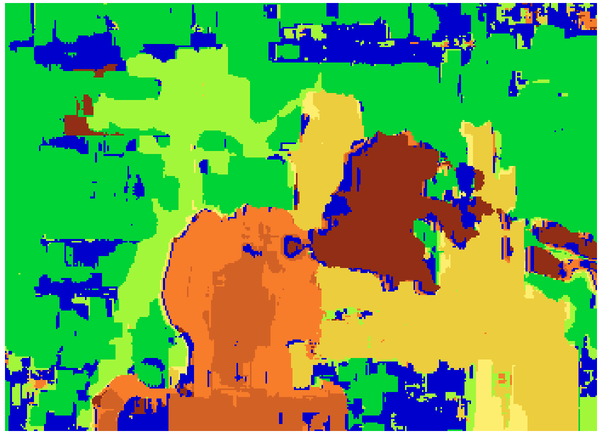


Stereo as Optimization with Non-Local Constraints

Data

Ground truth

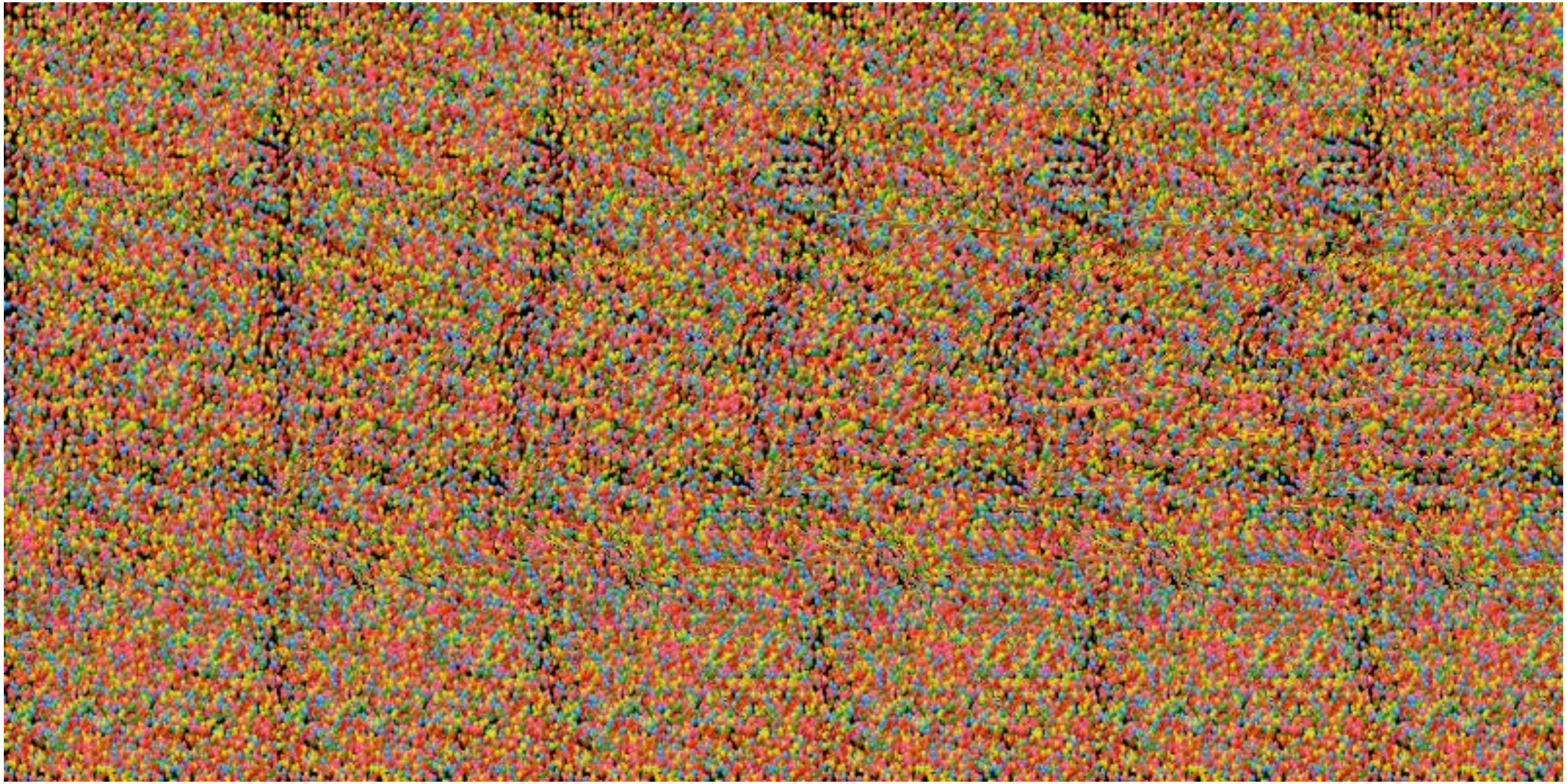
Window-based matching



Global optimization method (graph cuts)

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

Autosterograms



A random dot autostereogram encoding a 3D scene of a shark, which can be seen with proper viewing technique.

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

https://www.youtube.com/watch?v=v8O8Em_RPNg