Stereo II

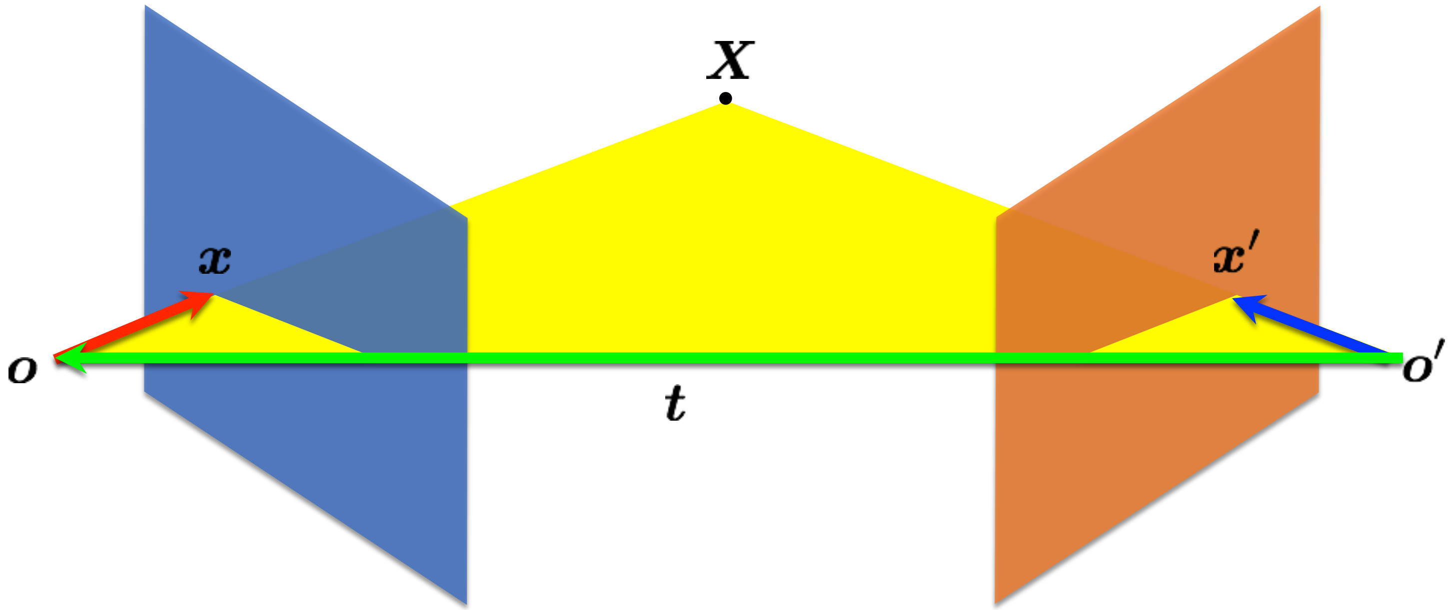
CSC420
David Lindell
University of Toronto
cs.toronto.edu/~lindell/teaching/420

Slide credit: Babak Taati ←Ahmed Ashraf ←Sanja Fidler, Yannis Gkioulekas, Kris Kitani, Srinivasa Narasimhan

Logistics

- A4 is out. Due date is March 28
- Final exam April 17th WB116/119 7pm–10pm
  - multiple choice, short answer, long answer

Recap



If these three vectors are coplanar $\boldsymbol{x}, \boldsymbol{t}, \boldsymbol{x}'$ then

$$(\boldsymbol{x} - \boldsymbol{t})^\top (\boldsymbol{t} \times \boldsymbol{x}) = 0$$

Essential Matrix

rigid motion

$$\boldsymbol{x}' = \mathbf{R}(\boldsymbol{x} - \boldsymbol{t})$$

coplanarity

$$(\boldsymbol{x} - \boldsymbol{t})^\top (\boldsymbol{t} \times \boldsymbol{x}) = 0$$

$$(\boldsymbol{x}'^\top \mathbf{R})(\boldsymbol{t} \times \boldsymbol{x}) = 0$$

$$(\boldsymbol{x}'^\top \mathbf{R})([\mathbf{t}_\times]\boldsymbol{x}) = 0$$

$$\boldsymbol{x}'^\top (\mathbf{R}[\mathbf{t}_\times])\boldsymbol{x} = 0$$

$$\boxed{\boldsymbol{x}'^\top \mathbf{E} \boldsymbol{x} = 0}$$

Essential Matrix
[Longuet-Higgins 1981]

$$\hat{\boldsymbol{x}}'^{\top}\mathbf{E}\hat{\boldsymbol{x}} = 0$$

The essential matrix operates on image points expressed in **2D coordinates** expressed in the camera coordinate system

$$\hat{\boldsymbol{x}}' = \mathbf{K}'^{-1}\boldsymbol{x}' \qquad \underset{\text{camera point}}{\hat{\boldsymbol{x}}} = \mathbf{K}^{-1}\underset{\text{image point}}{\boldsymbol{x}}$$

Writing out the epipolar constraint in terms of image coordinates

$$\mathbf{K}'^{-\top}\mathbf{E}\mathbf{K}^{-1}\boldsymbol{x} = 0$$

$$\boldsymbol{x}'^{\top}(\mathbf{K}'^{-\top}\mathbf{E}\mathbf{K}^{-1})\boldsymbol{x} = 0$$

$$\boldsymbol{x}'^{\top}\mathbf{F}\boldsymbol{x} = 0$$

properties of the ~~E~~ F matrix

Longuet-Higgins equation $\boldsymbol{x}'^{\top}\overset{F}{\mathbf{E}}\boldsymbol{x} = 0$

Epipolar lines

$\boldsymbol{x}^{\top}\boldsymbol{l} = 0$ $\quad$ $\boldsymbol{x}'^{\top}\boldsymbol{l}' = 0$

$\boldsymbol{l}' = \overset{F}{\mathbf{E}}\boldsymbol{x}$ $\quad$ $\boldsymbol{l} = \overset{F}{\mathbf{E}}^{T}\boldsymbol{x}'$

Epipoles $\boldsymbol{e}'^{\top}\overset{F}{\mathbf{E}} = \mathbf{0}$ $\quad$ $\overset{F}{\mathbf{E}}\boldsymbol{e} = \mathbf{0}$

(points in **image** coordinates)

Breaking down the fundamental matrix

$$\mathbf{F} = \mathbf{K}'^{-\top}\mathbf{E}\mathbf{K}^{-1}$$

$$\mathbf{F} = \mathbf{K}'^{-\top}[\mathbf{t}_{\times}]\mathbf{R}\mathbf{K}^{-1}$$

Depends on both intrinsic and extrinsic parameters

Breaking down the fundamental matrix

$$\mathbf{F} = \mathbf{K}'^{-\top}\mathbf{E}\mathbf{K}^{-1}$$

$$\mathbf{F} = \mathbf{K}'^{-\top}[\mathbf{t}_{\times}]\mathbf{R}\mathbf{K}^{-1}$$

Depends on both intrinsic and extrinsic parameters

How would you solve for F?

$$\boldsymbol{x}'^{\top}_{m}\mathbf{F}\boldsymbol{x}_{m} = 0$$

The 8-point algorithm

Assume you have ***M*** matched *image* points

$$\{\boldsymbol{x}_m, \boldsymbol{x}'_m\} \qquad m = 1, \ldots, M$$

Each correspondence should satisfy

$$\boldsymbol{x}'^{\top}_m \mathbf{F} \boldsymbol{x}_m = 0$$

How would you solve for the 3 x 3 ***F*** *matrix?*

Assume you have ***M*** matched *image* points

$$\{\boldsymbol{x}_m, \boldsymbol{x}'_m\} \qquad m = 1, \ldots, M$$

Each correspondence should satisfy

$$\boldsymbol{x}'^{\top}_m \mathbf{F} \boldsymbol{x}_m = 0$$

How would you solve for the 3 x 3 ***F*** *matrix?*

S V D

Assume you have ***M*** matched *image* points

$$\{\boldsymbol{x}_m, \boldsymbol{x}'_m\} \qquad m = 1, \ldots, M$$

Each correspondence should satisfy

$$\boldsymbol{x}'^{\top}_m \mathbf{F} \boldsymbol{x}_m = 0$$

How would you solve for the 3 x 3 ***F*** *matrix?*

Set up a homogeneous linear system with 9 unknowns

$$\boldsymbol{x}_m'^{\top}\mathbf{F}\boldsymbol{x}_m = 0$$

$$\begin{bmatrix} x_m' & y_m' & 1 \end{bmatrix} \begin{bmatrix} f_1 & f_2 & f_3 \\ f_4 & f_5 & f_6 \\ f_7 & f_8 & f_9 \end{bmatrix} \begin{bmatrix} x_m \\ y_m \\ 1 \end{bmatrix} = 0$$

How many equation do you get from one correspondence?

$$\begin{bmatrix} x'_m & y'_m & 1 \end{bmatrix} \begin{bmatrix} f_1 & f_2 & f_3 \\ f_4 & f_5 & f_6 \\ f_7 & f_8 & f_9 \end{bmatrix} \begin{bmatrix} x_m \\ y_m \\ 1 \end{bmatrix} = 0$$

ONE correspondence gives you ONE equation

$$\begin{aligned} x_m x'_m f_1 + x_m y'_m f_2 + x_m f_3 + \\ y_m x'_m f_4 + y_m y'_m f_5 + y_m f_6 + \\ x'_m f_7 + y'_m f_8 + f_9 = 0 \end{aligned}$$

$$\begin{bmatrix} x'_m & y'_m & 1 \end{bmatrix} \begin{bmatrix} f_1 & f_2 & f_3 \\ f_4 & f_5 & f_6 \\ f_7 & f_8 & f_9 \end{bmatrix} \begin{bmatrix} x_m \\ y_m \\ 1 \end{bmatrix} = 0$$

Set up a homogeneous linear system with 9 unknowns

$$\begin{bmatrix} x_1x'_1 & x_1y'_1 & x_1 & y_1x'_1 & y_1y'_1 & y_1 & x'_1 & y'_1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_Mx'_M & x_My'_M & x_M & y_Mx'_M & y_My'_M & y_M & x'_M & y'_M & 1 \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \\ f_3 \\ f_4 \\ f_5 \\ f_6 \\ f_7 \\ f_8 \\ f_9 \end{bmatrix} = \mathbf{0}$$

How many equations do you need?

Each point pair (according to epipolar constraint)
contributes only one <u>scalar</u> equation

$$\boldsymbol{x}'^{\top}_m \mathbf{F} \boldsymbol{x}_m = 0$$

Note: This is different from the Homography estimation
where each point pair contributes 2 equations.

We need at least 8 points

Hence, the 8 point algorithm!

How do you solve a homogeneous linear system?

$$\mathbf{A}\boldsymbol{X} = \mathbf{0}$$

How do you solve a homogeneous linear system?

$$\mathbf{A}\boldsymbol{X} = \mathbf{0}$$

Total Least Squares

minimize $\|\mathbf{A}\boldsymbol{x}\|^2$

subject to $\|\boldsymbol{x}\|^2 = 1$

How do you solve a homogeneous linear system?

$$\mathbf{A}\boldsymbol{X} = \mathbf{0}$$

Total Least Squares

minimize $\|\mathbf{A}\boldsymbol{x}\|^2$

subject to $\|\boldsymbol{x}\|^2 = 1$

SVD!

Eight-Point Algorithm

0. (Normalize points)
1. Construct the M x 9 matrix **A**
2. Find the SVD of **A**
3. Entries of **F** are the elements of column of **V** corresponding to the least singular value
4. (Enforce rank 2 constraint on F)
5. (Un-normalize F)

Eight-Point Algorithm

0. (Normalize points)
1. Construct the M x 9 matrix **A**
2. Find the SVD of **A**
3. Entries of **F** are the elements of column of **V** corresponding to the least singular value
4. (Enforce rank 2 constraint on F)
5. (Un-normalize F)

See Hartley-Zisserman for why we do this

Eight-Point Algorithm

0. (Normalize points)
1. Construct the M x 9 matrix **A**
2. Find the SVD of **A**
3. Entries of **F** are the elements of column of **V** corresponding to the least singular value
4. (Enforce rank 2 constraint on F)
5. (Un-normalize F)

How do we do this?

Eight-Point Algorithm

0. (Normalize points)
1. Construct the M x 9 matrix **A**
2. Find the SVD of **A**
3. Entries of **F** are the elements of column of **V** corresponding to the least singular value
4. (Enforce rank 2 constraint on F)
5. (Un-normalize F)

↖ How do we do this?

SVD!

Enforcing rank constraints

Problem: Given a matrix **F**, find the matrix **F’** of rank k that is closest to **F**,

$$\min_{\substack{F' \\ \text{rank}(F')=k}} \quad \|F - F'\|^2$$

Solution: Compute the singular value decomposition of **F**,

$$F = U\Sigma V^T$$

Form a matrix Σ’ by replacing all but the k largest singular values in Σ with 0.

Then the problem solution is the matrix **F’** formed as,

$$F' = U\Sigma' V^T$$

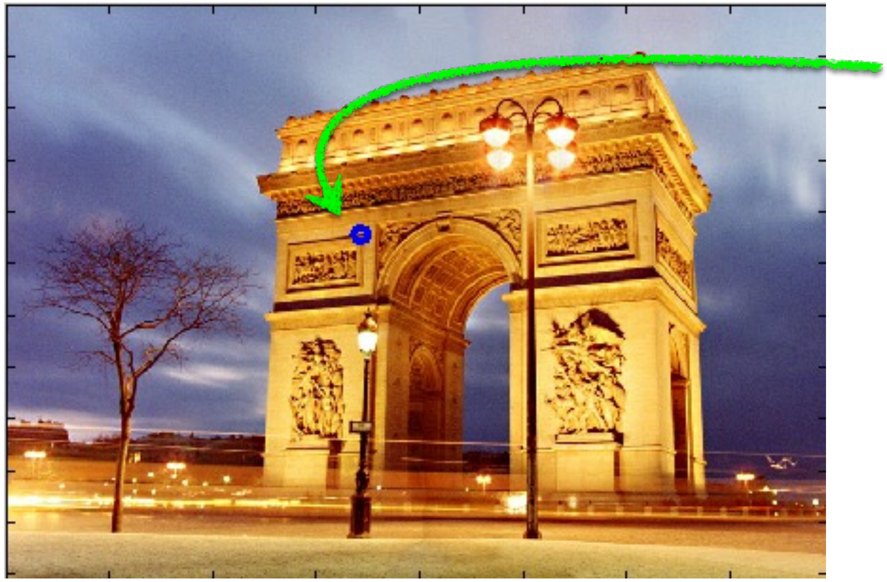
Eight-Point Algorithm

0. (Normalize points)
1. Construct the M x 9 matrix **A**
2. Find the SVD of **A**
3. Entries of **F** are the elements of column of **V** corresponding to the least singular value
4. (Enforce rank 2 constraint on F)
5. (Un-normalize F)

Example

epipolar lines

$$\mathbf{F} = \begin{bmatrix} -0.00310695 & -0.0025646 & 2.96584 \\ -0.028094 & -0.00771621 & 56.3813 \\ 13.1905 & -29.2007 & -9999.79 \end{bmatrix}$$



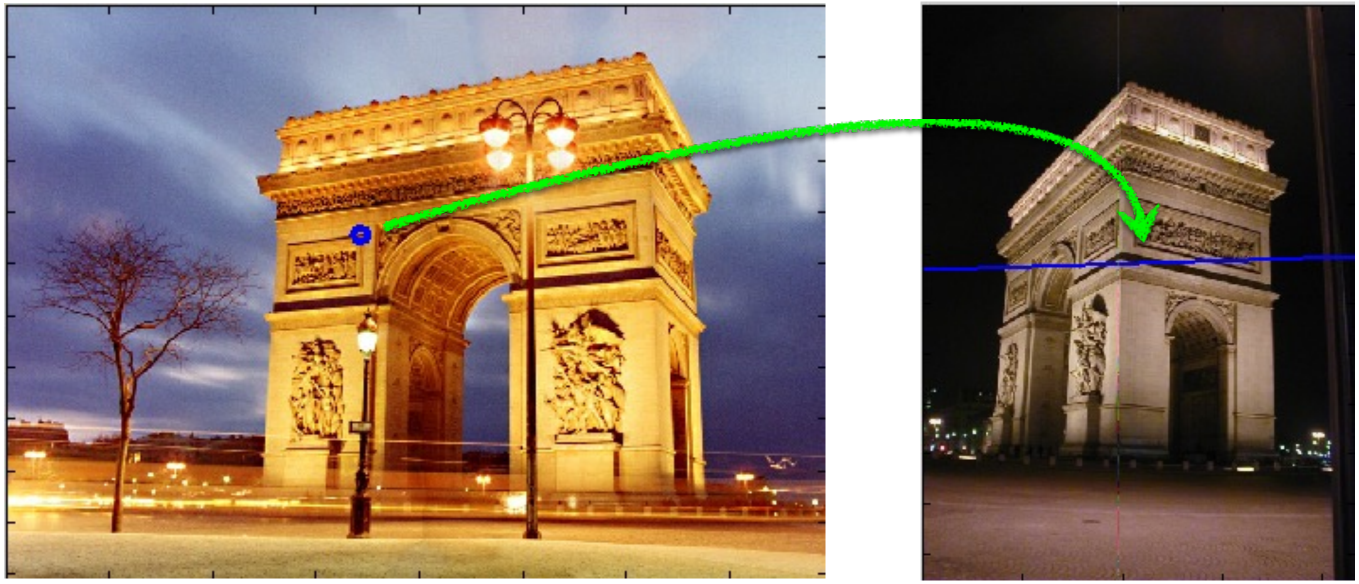
$$\boldsymbol{x} = \begin{bmatrix} 343.53 \\ 221.70 \\ 1.0 \end{bmatrix}$$

$$\boldsymbol{l}' = \mathbf{F}\boldsymbol{x}$$

$$= \begin{bmatrix} 0.0295 \\ 0.9996 \\ -265.1531 \end{bmatrix}$$

$$\boldsymbol{l}' = \mathbf{F}\boldsymbol{x}$$

$$= \begin{bmatrix} 0.0295 \\ 0.9996 \\ -265.1531 \end{bmatrix}$$



Where is the epipole?

How would you compute it?

$$\mathbf{F}\boldsymbol{e} = \mathbf{0}$$

The epipole is in the right null space of **F**

How would you solve for the epipole?

$$\mathbf{F}\boldsymbol{e} = \mathbf{0}$$

The epipole is in the right null space of **F**

How would you solve for the epipole?

SVD!

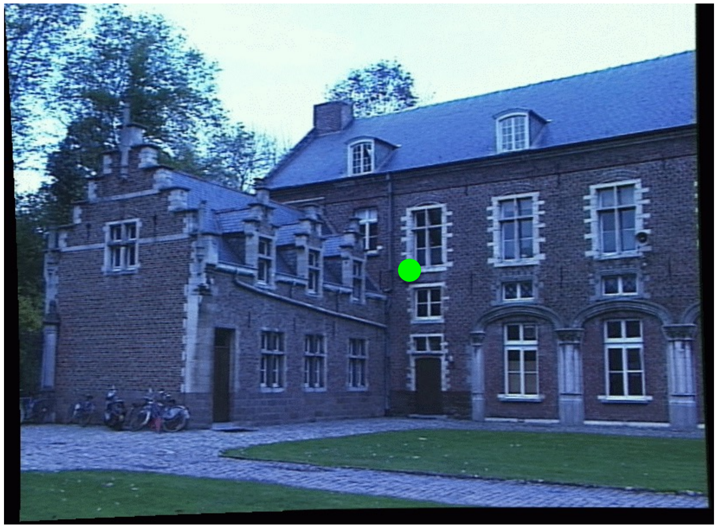
Revisiting triangulation

How would you reconstruct 3D points?

Left image

Right image

How would you reconstruct 3D points?

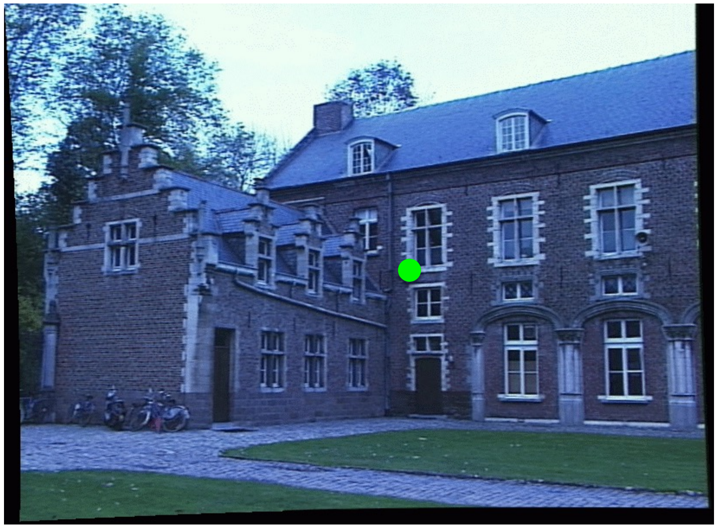


Left image

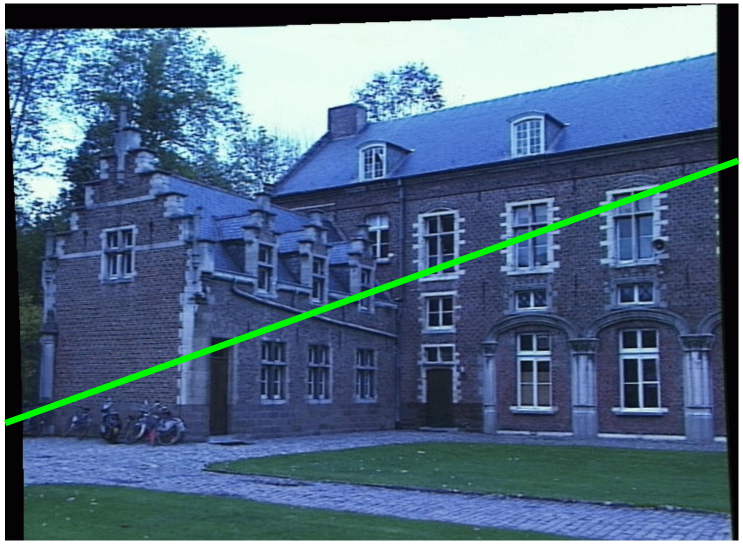
Right image

1. Select point in one image (how?)

How would you reconstruct 3D points?



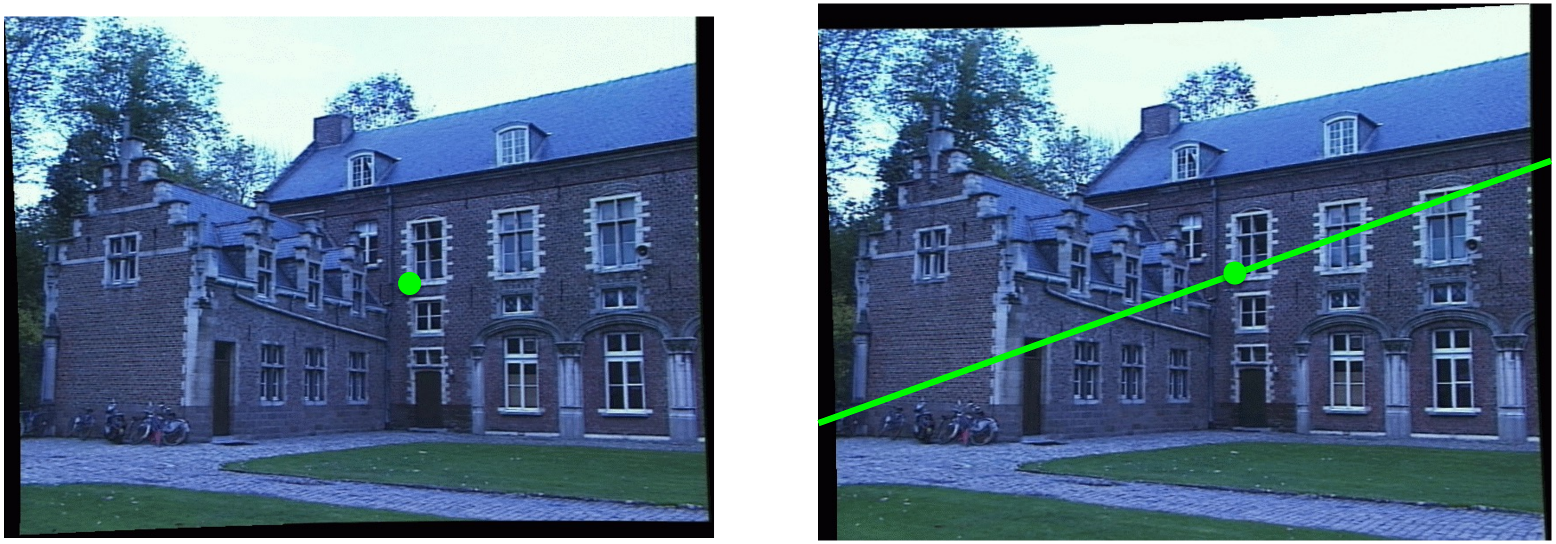
Left image



Right image

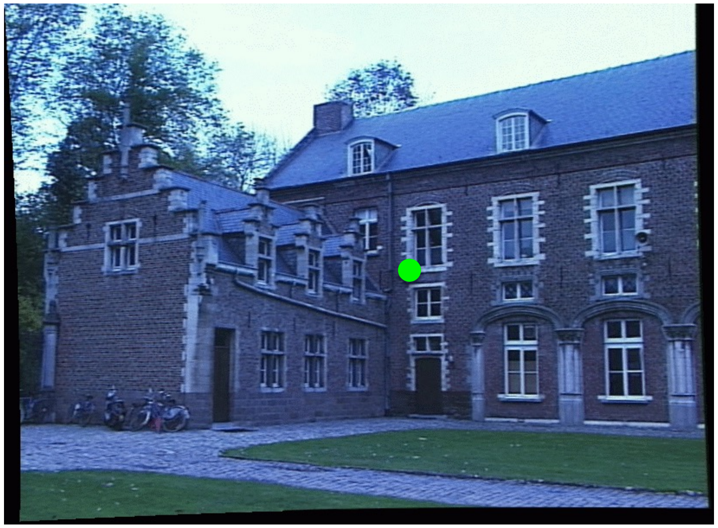
1. Select point in one image (how?)
2. Form epipolar line for that point in second image (how?)

How would you reconstruct 3D points?


Left image
Right image

1. Select point in one image (how?)
2. Form epipolar line for that point in second image (how?)
3. Find matching point along line (how?)

How would you reconstruct 3D points?



Left image

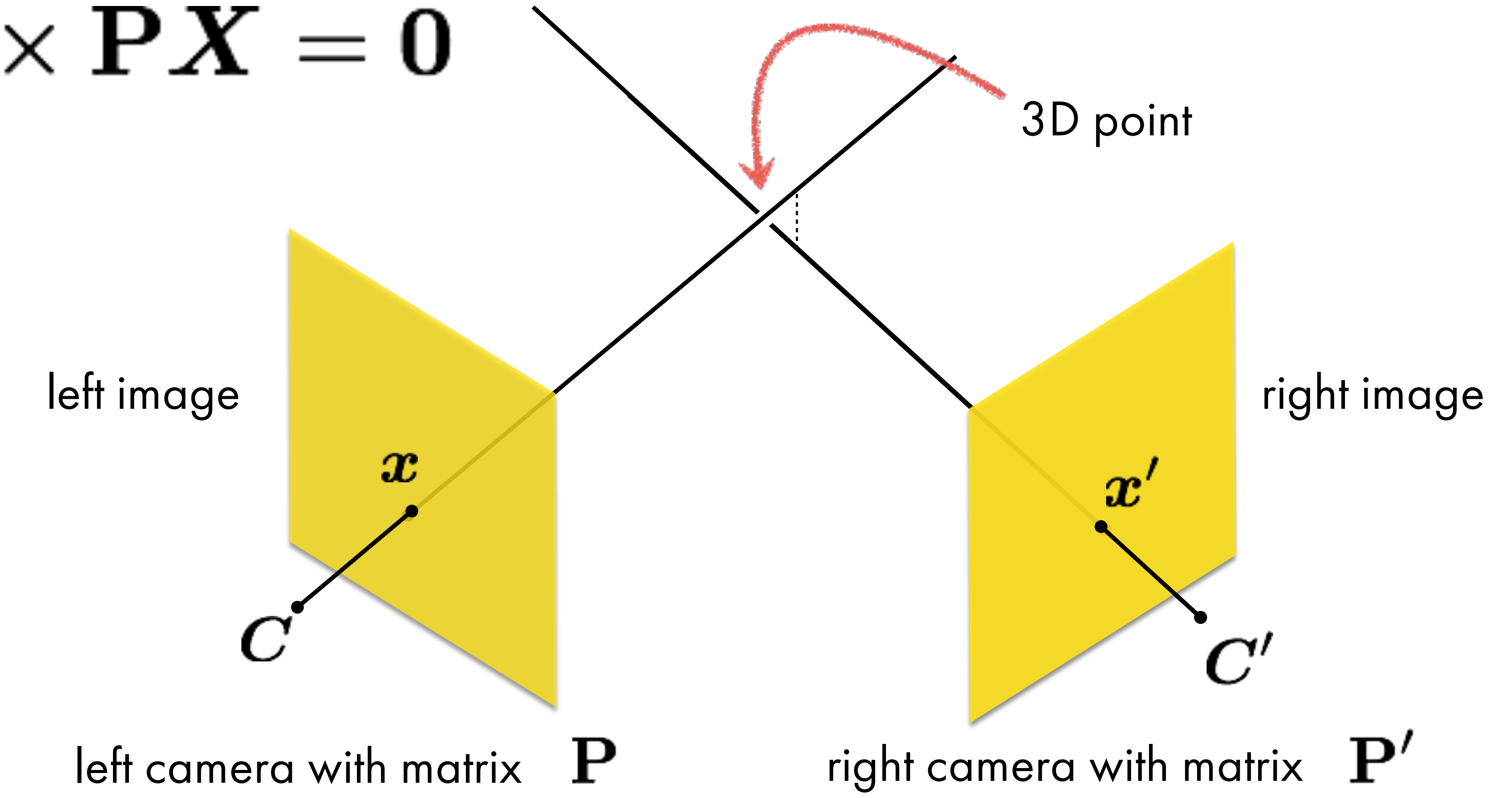
Right image

1. Select point in one image (how?)
2. Form epipolar line for that point in second image (how?)
3. Find matching point along line (how?)
4. Perform triangulation (how?)

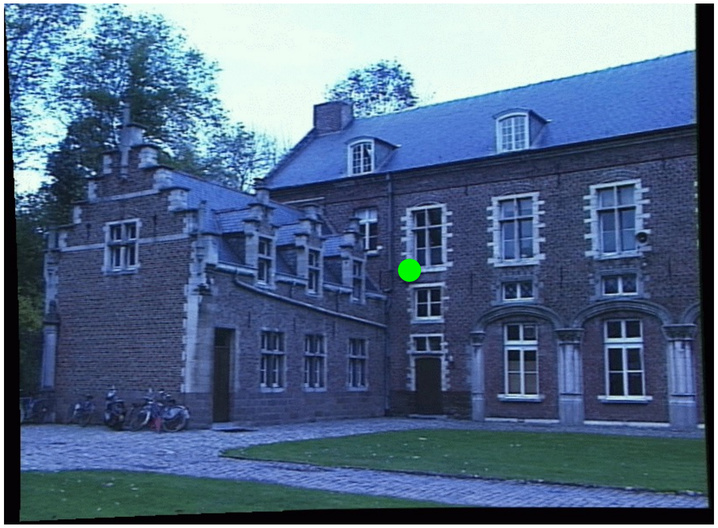
Triangulation

Using the fact that the cross product should be zero

$$\mathbf{x} \times \mathbf{P}\boldsymbol{X} = \mathbf{0}$$



How would you reconstruct 3D points?



Left image

Right image

1. Select point in one image (how?)
2. Form epipolar line for that point in second image (how?)
3. Find matching point along line (how?)
4. Perform triangulation (how?)

What are the disadvantages
of this procedure?

Stereo rectification

What's different between these two images?

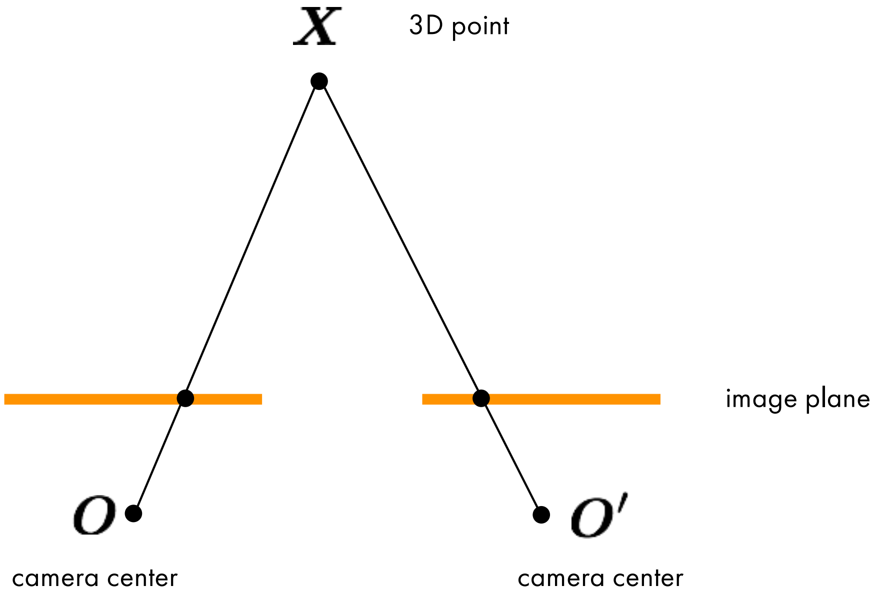
Objects that are close move more or less?

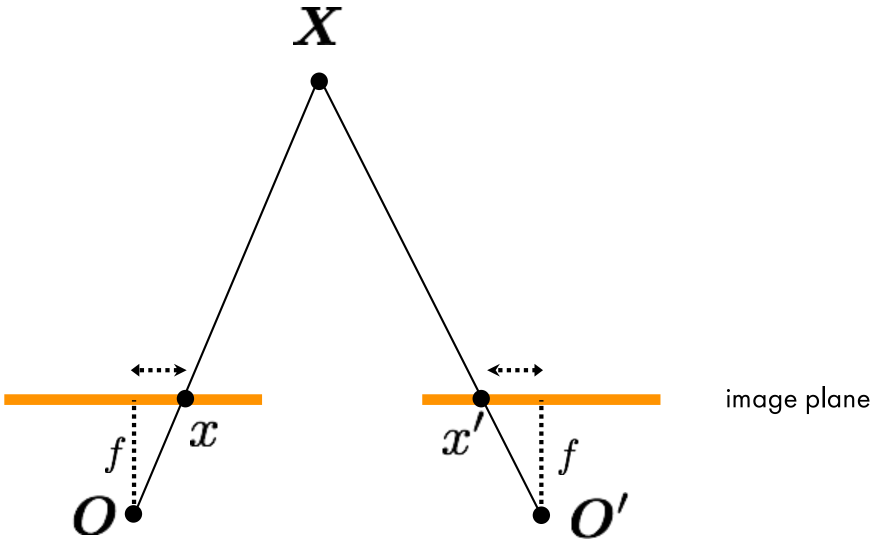
The amount of horizontal movement is inversely proportional to ...

The amount of horizontal movement is inversely proportional to ...

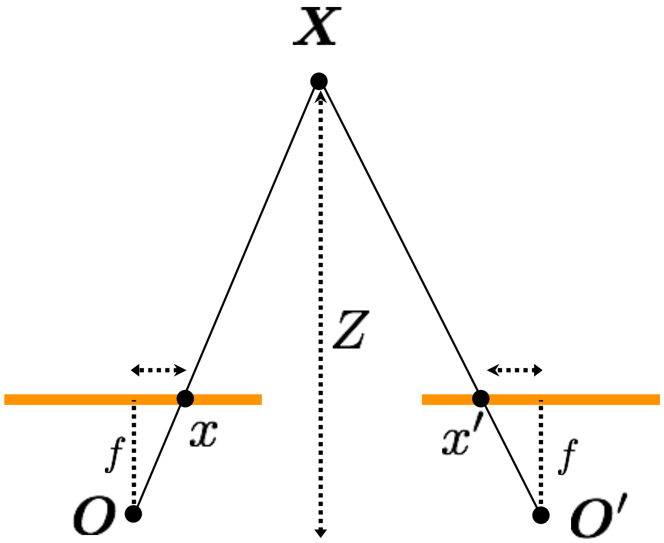
... the distance from the camera.

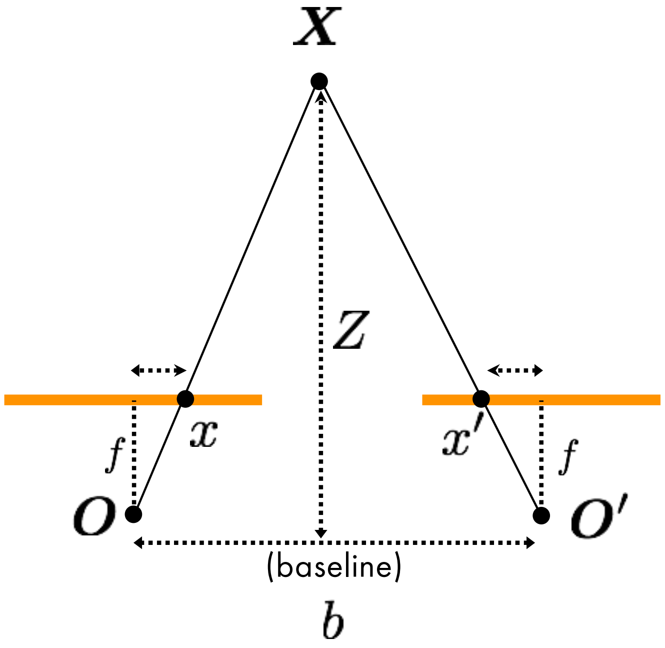
More formally...


X
3D point
image plane
O
camera center
O'
camera center

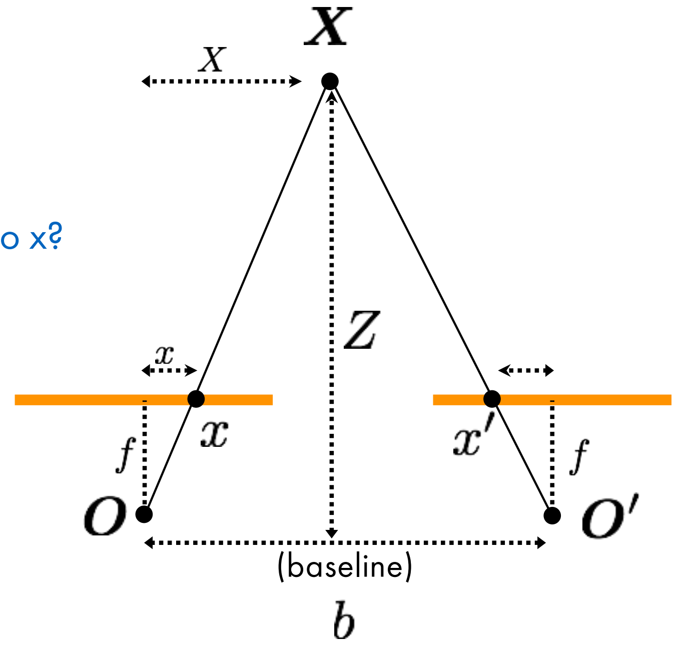


X
x
x'
f
f
O
O'
image plane

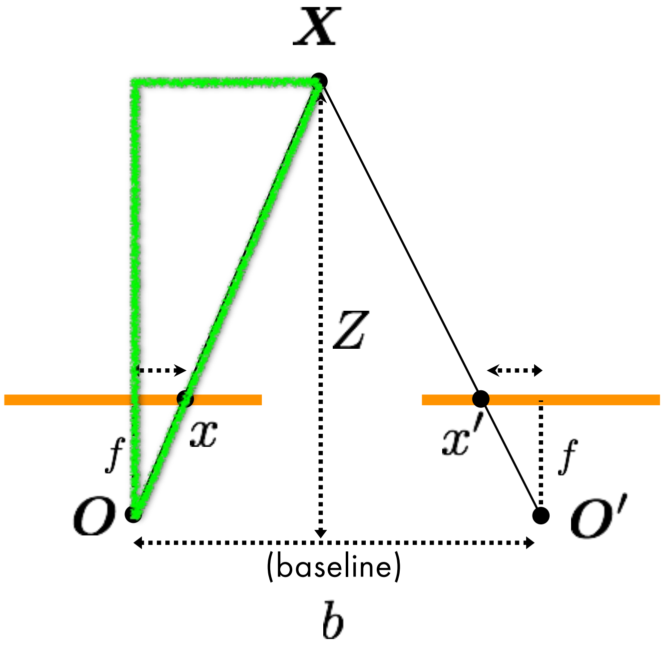

$\boldsymbol{X}$
Z
x
x'
f
f
$\boldsymbol{O}$
$\boldsymbol{O'}$


$\boldsymbol{X}$
Z
x
x'
f
f
$\boldsymbol{O}$
$\boldsymbol{O'}$
(baseline)
b

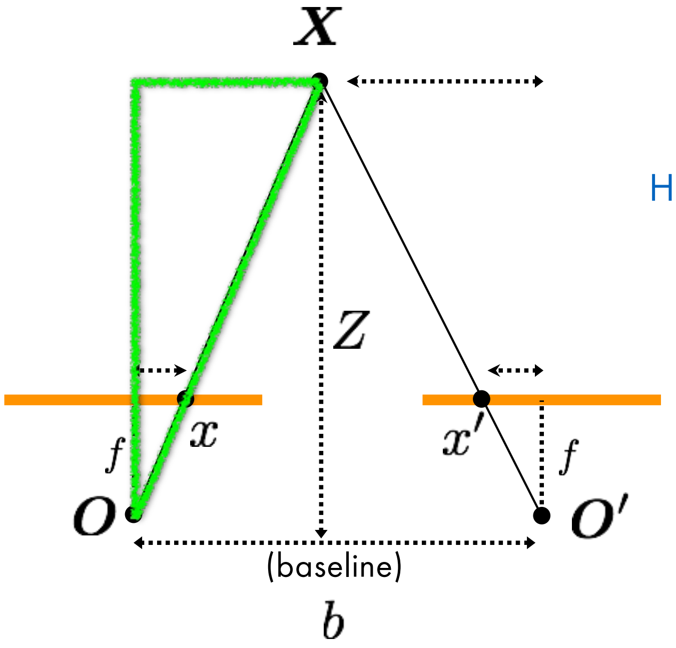
How is X related to x?



$$\frac{X}{Z} = \frac{x}{f}$$

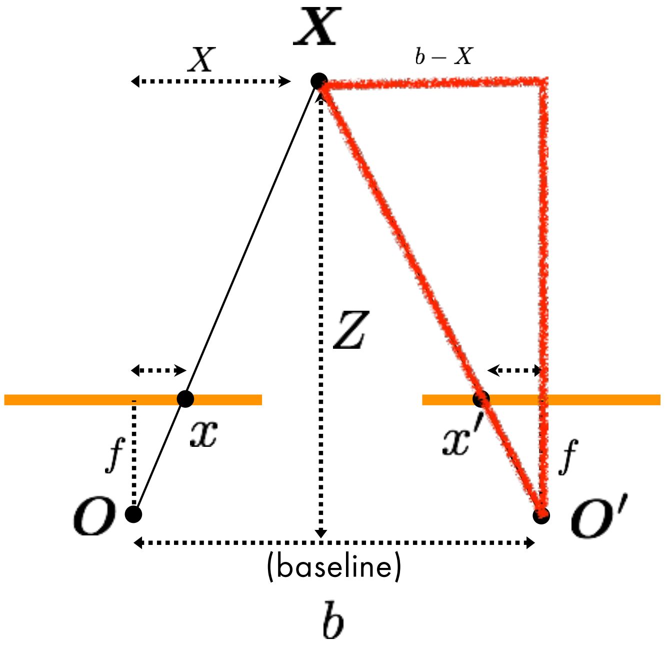


$$\frac{X}{Z} = \frac{x}{f}$$



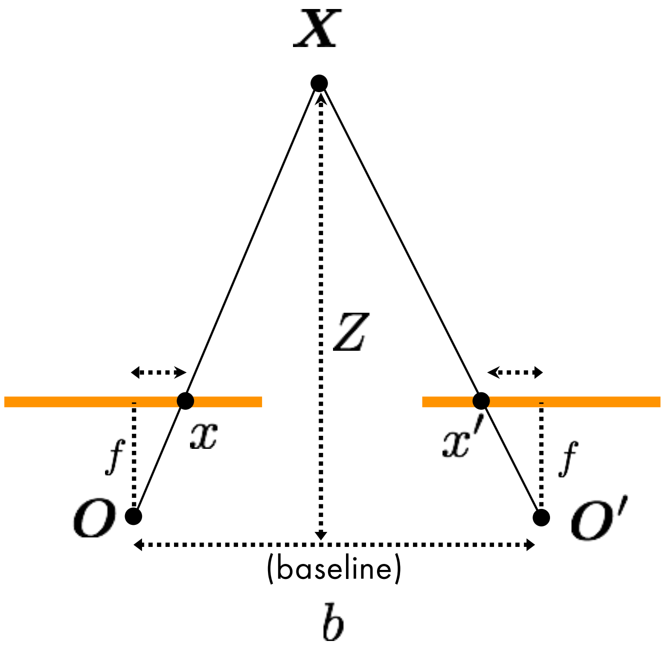
How is X related to x′?

$$\frac{X}{Z} = \frac{x}{f}$$



$$\frac{b - X}{Z} = \frac{x'}{f}$$

$$\frac{X}{Z} = \frac{x}{f}$$



$$\frac{b - X}{Z} = \frac{x'}{f}$$

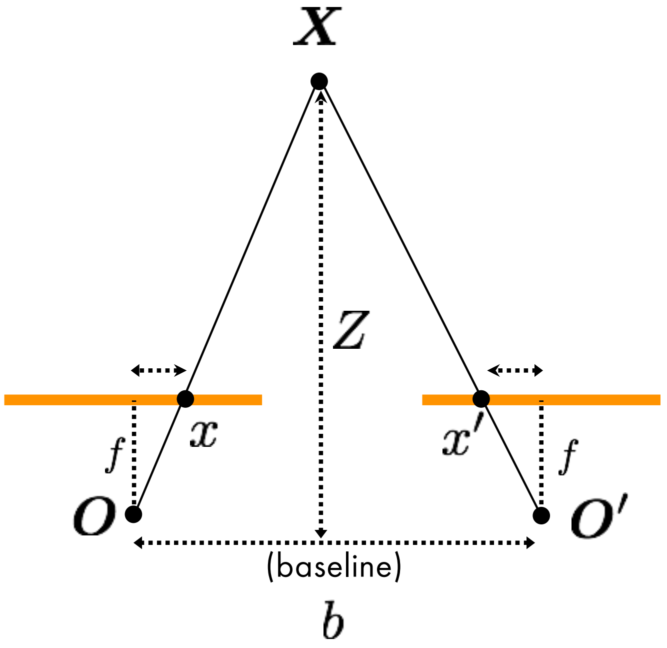
Disparity

$$d = x - x'$$

(wrt to camera origin of image plane)

$$= \frac{bf}{Z}$$

$$\frac{X}{Z} = \frac{x}{f}$$



$$\frac{b - X}{Z} = \frac{x'}{f}$$

Disparity

$$d = x - x'$$

$$= \frac{bf}{Z}$$

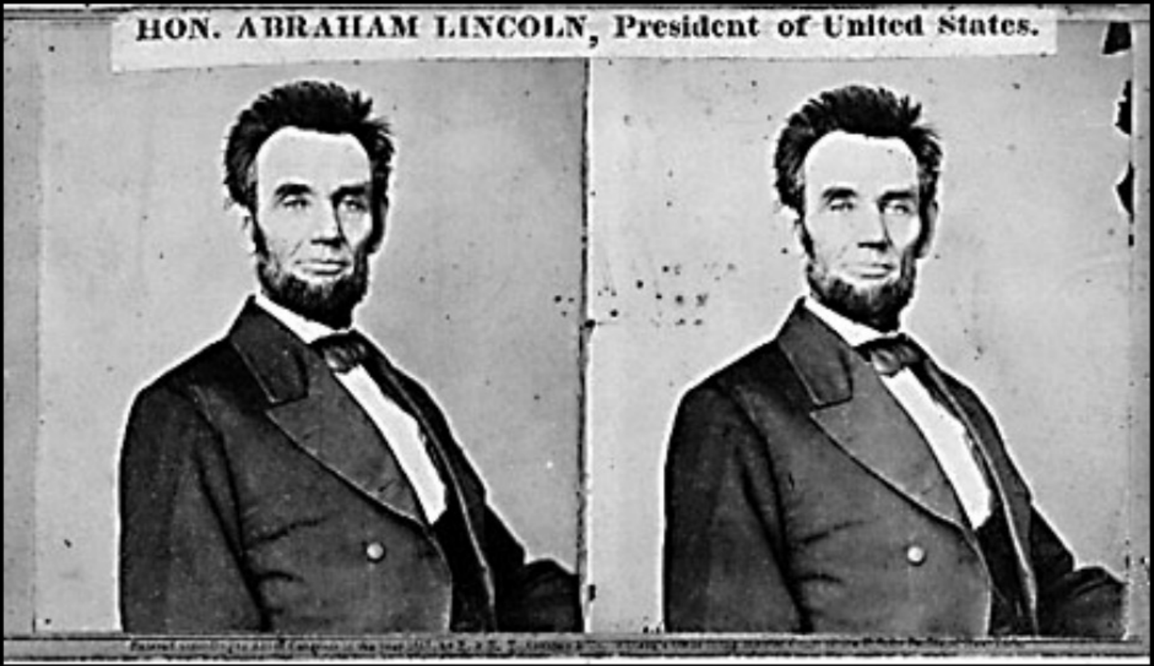
inversely proportional to depth

Subaru
Eyesight system

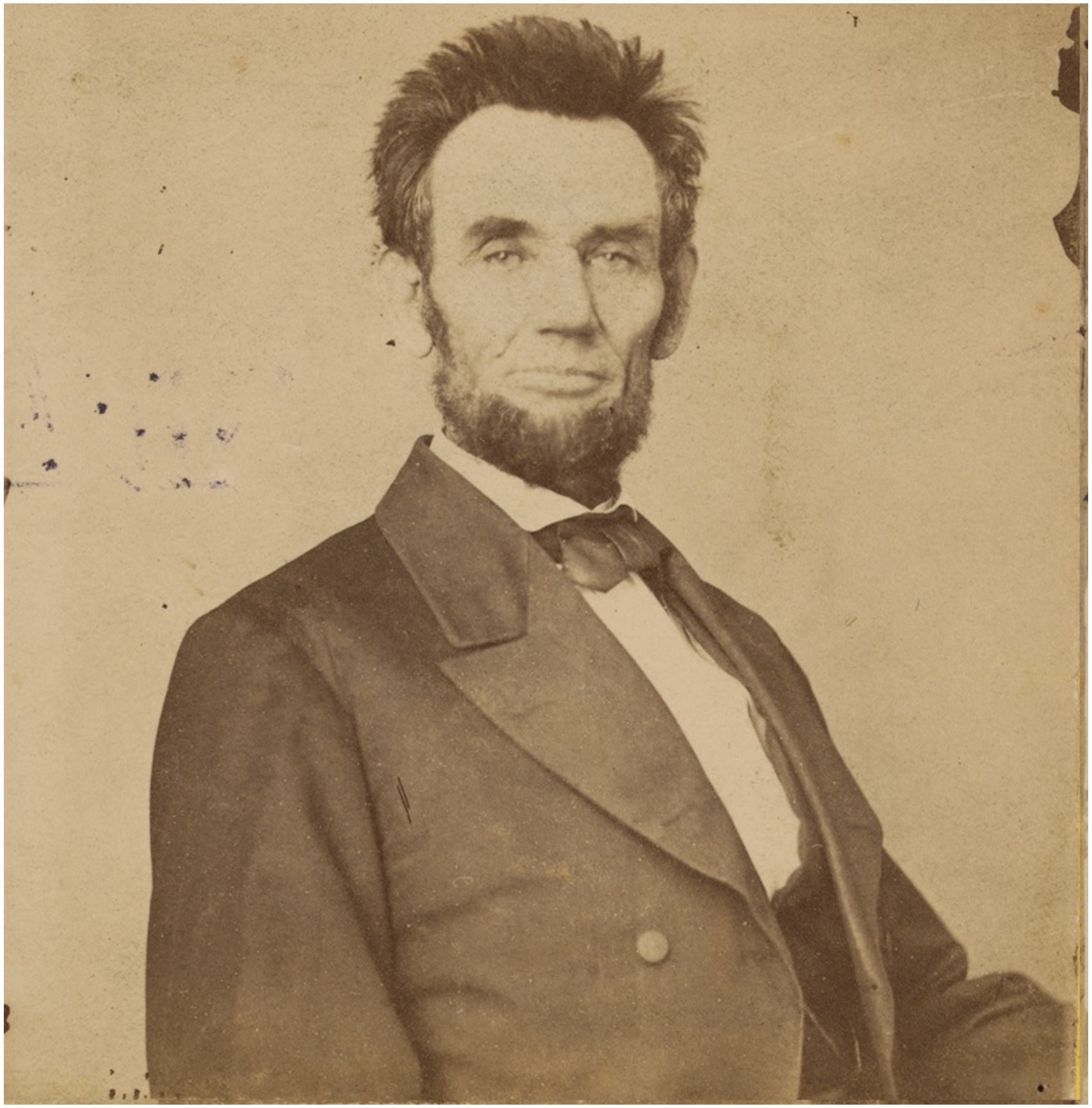
Pre-collision
braking

What other vision system uses disparity for depth sensing?

Stereoscopes: A 19th Century Pastime



HON. ABRAHAM LINCOLN, President of United States.

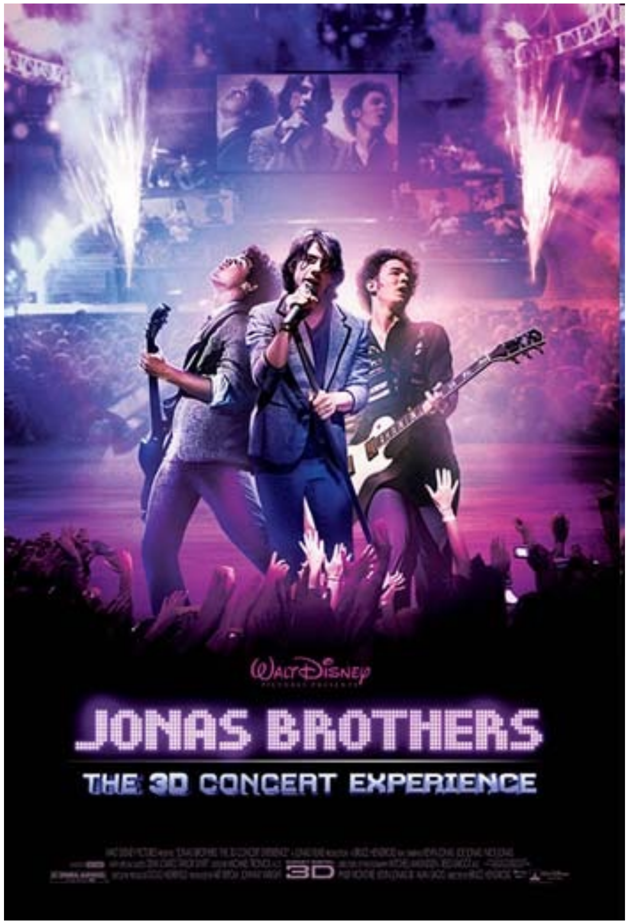


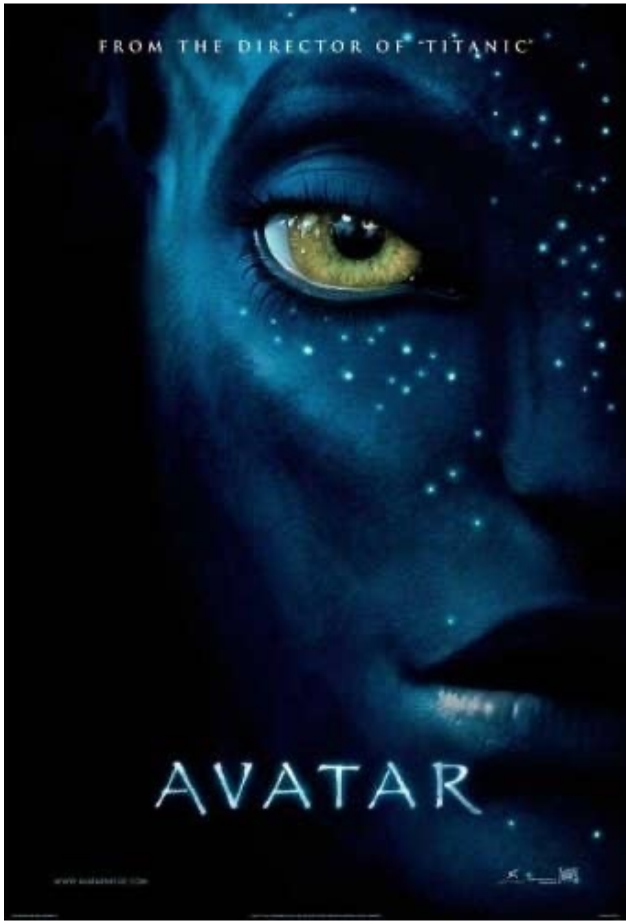
Public Library, Stereoscopic Looking Room, Chicago, by Phillips, 1923

Teesta suspension bridge-Darjeeling, India

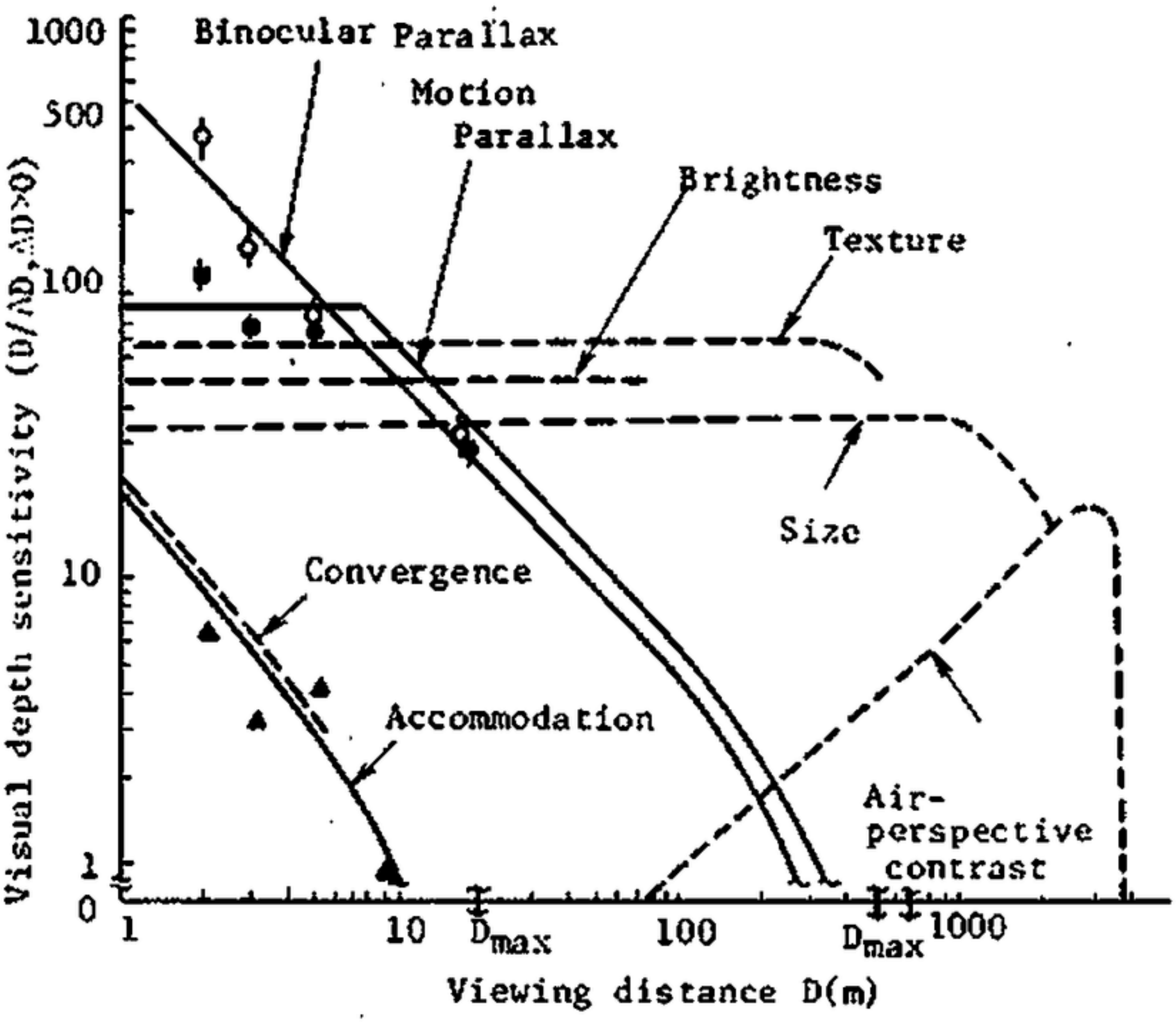
Mark Twain at Pool Table", no date, UCR Museum of Photography

This is how 3D movies work



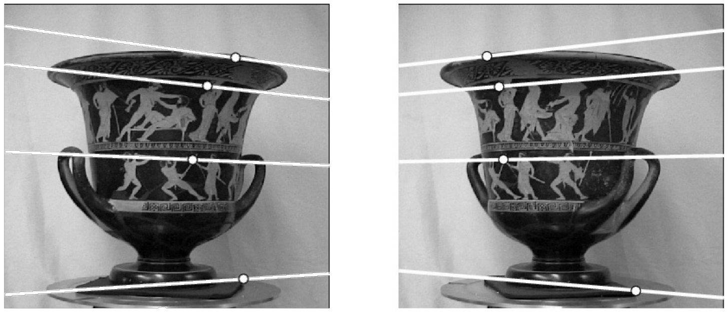


Is disparity the only depth cue the human visual system uses?

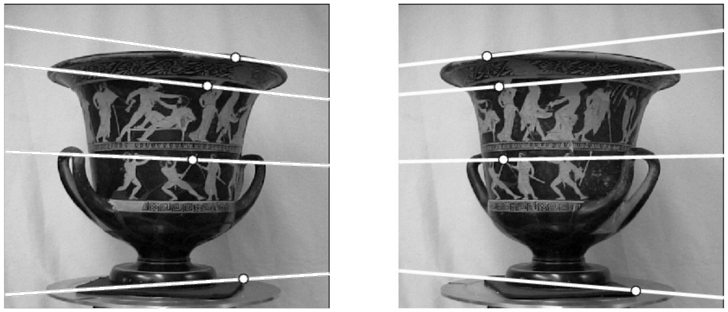

Binocular Parallax
Motion Parallax
Brightness
Texture
Size
Convergence
Accommodation
Air-perspective contrast
Visual depth sensitivity (D/ΔD, ΔD>0)
1000
500
100
10
1
0
1
10
Dmax
100
Dmax
1000
Viewing distance D(m)

Nagata '89

So can I compute depth from any two images of the same object?

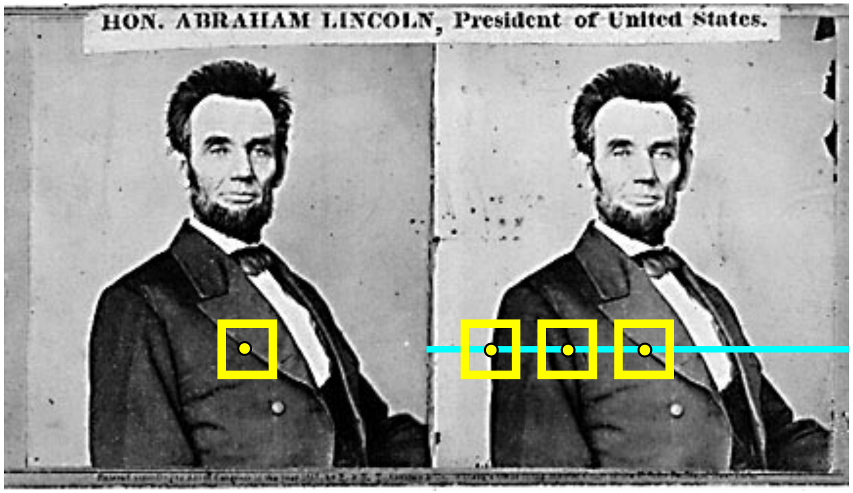


So can I compute depth from any two images of the same object?



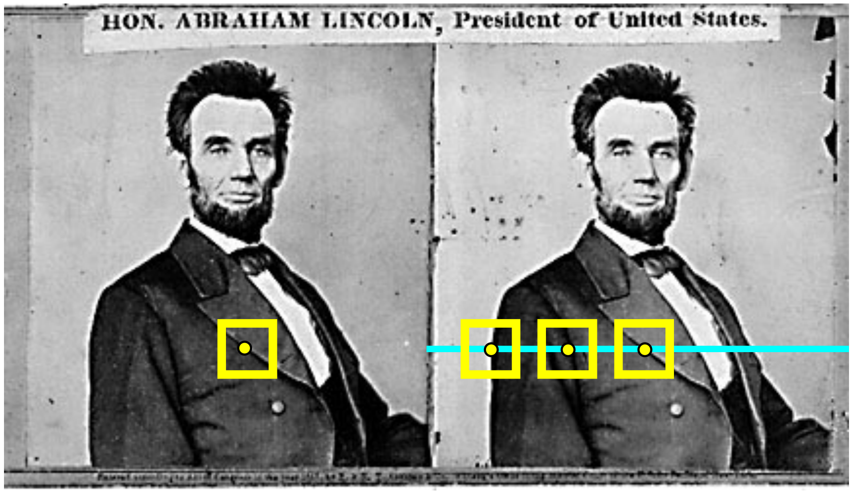
1. Need sufficient baseline

2. Images need to be ‘rectified’ first (make epipolar lines horizontal)

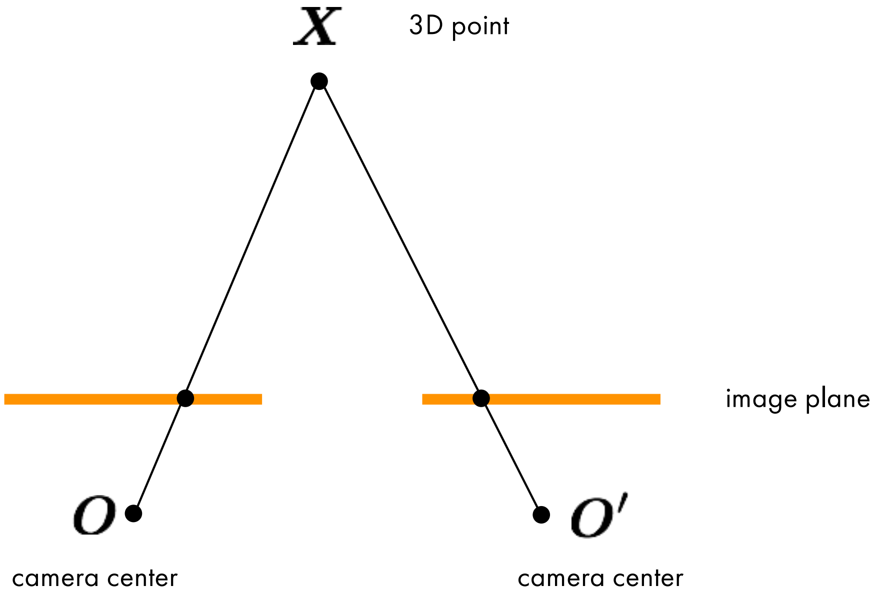


1. Rectify images
   (make epipolar lines horizontal)
2. For each pixel
   a. Find epipolar line
   b. Scan line for best match
   c. Compute depth from disparity

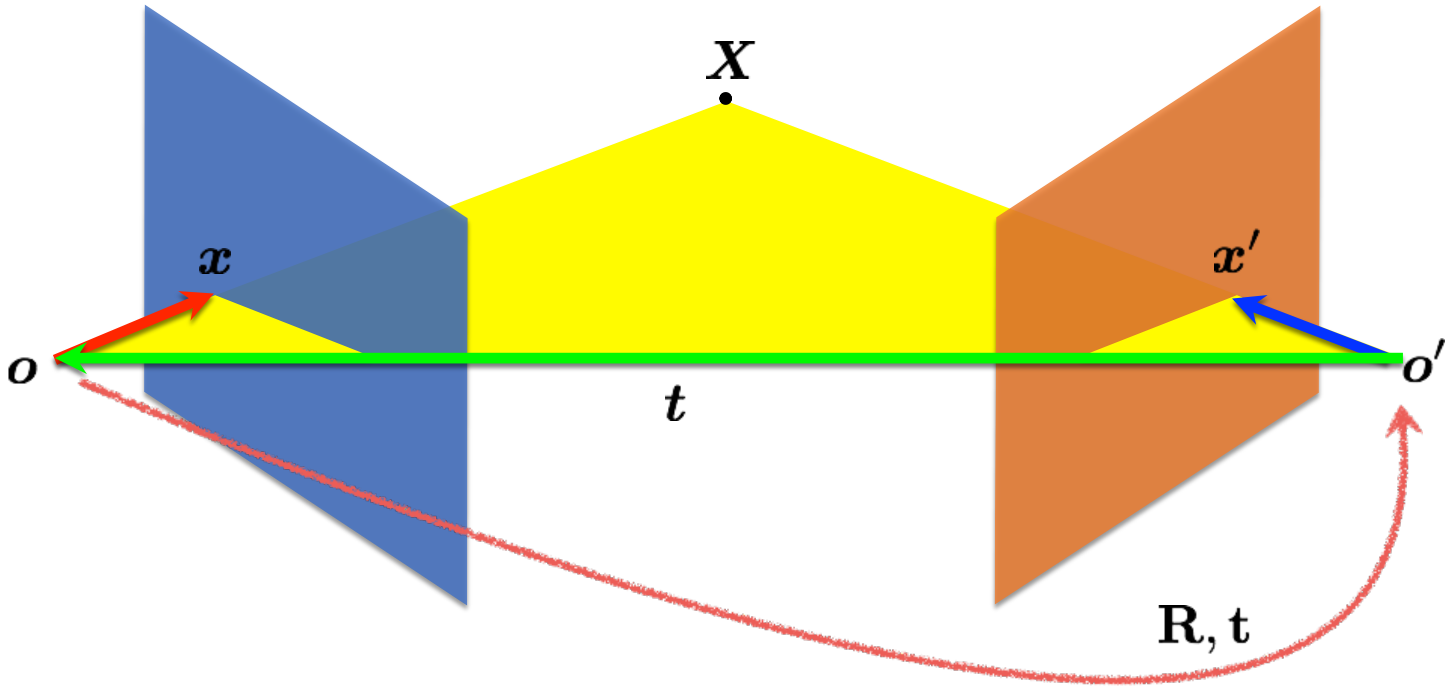
$$Z = \frac{bf}{d}$$



How can you make the epipolar lines horizontal?

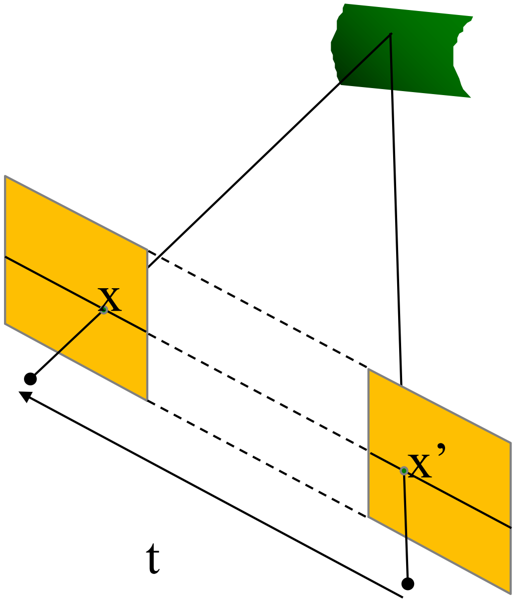


What's special about these two cameras?



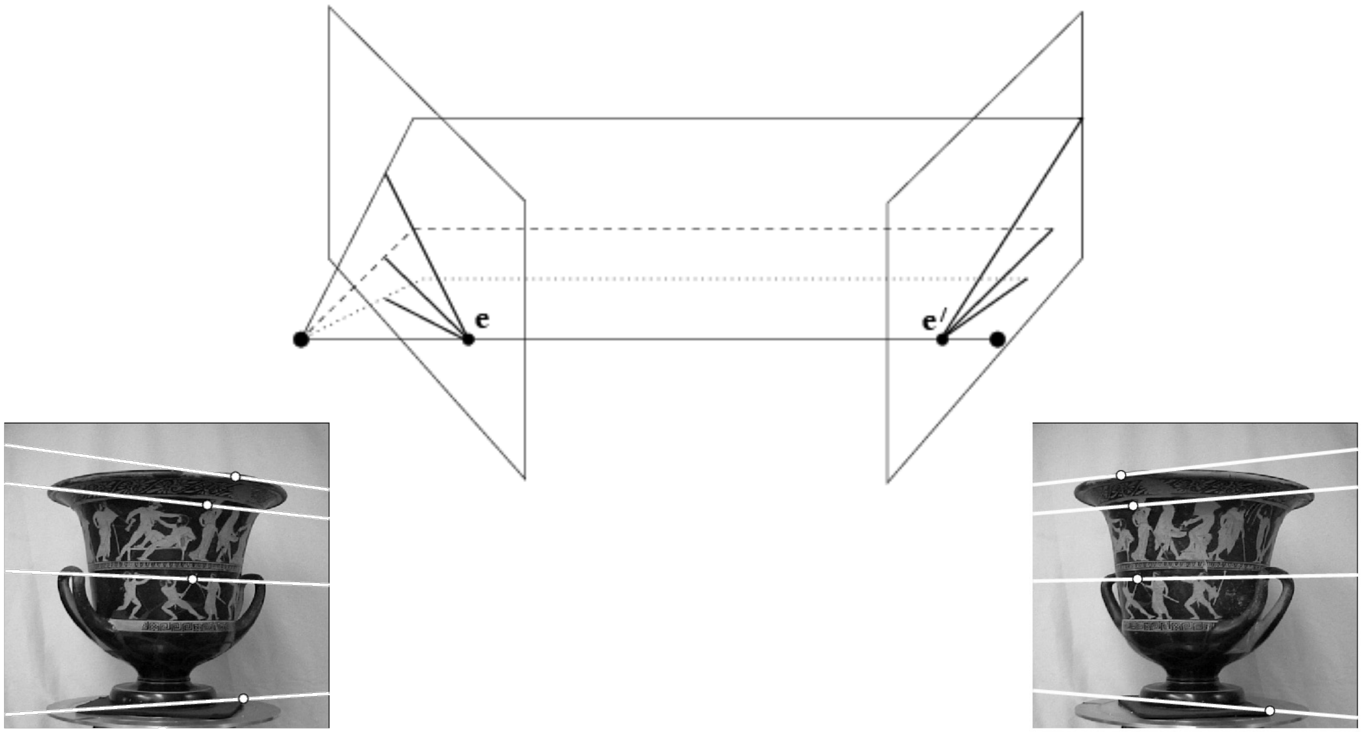
$$x' = \mathbf{R}(x - t)$$

When are epipolar lines horizontal?

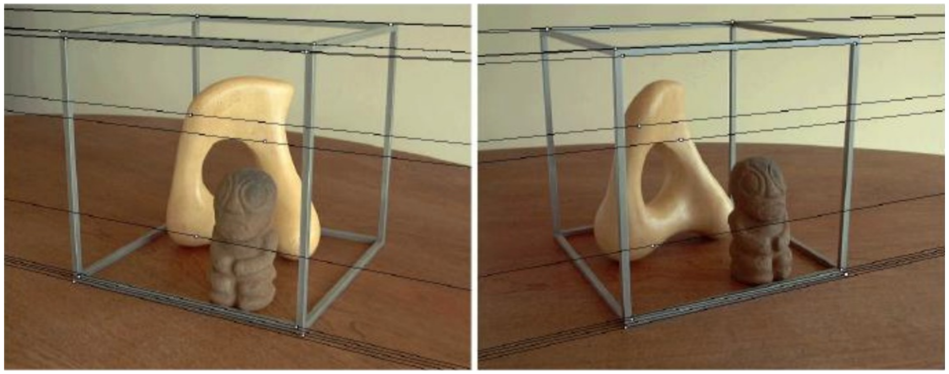


When this relationship holds:

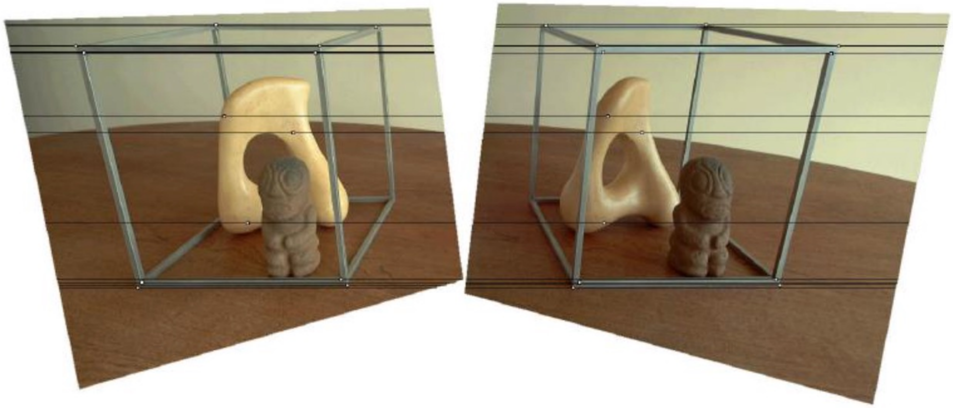
$$\boxed{R = I \qquad t = (T, 0, 0)}$$

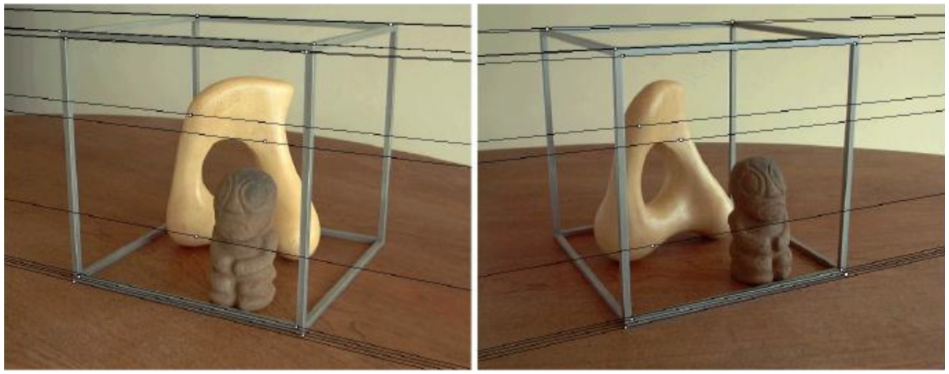


It’s hard to make the image planes exactly parallel

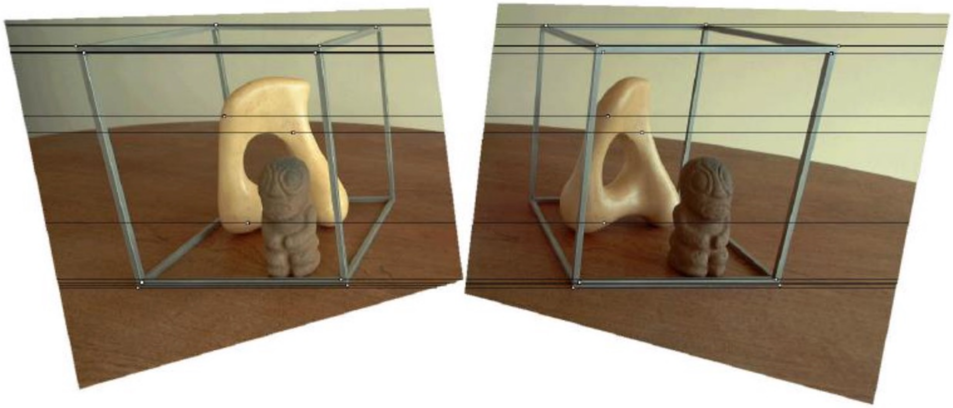


How can you make the epipolar lines horizontal?

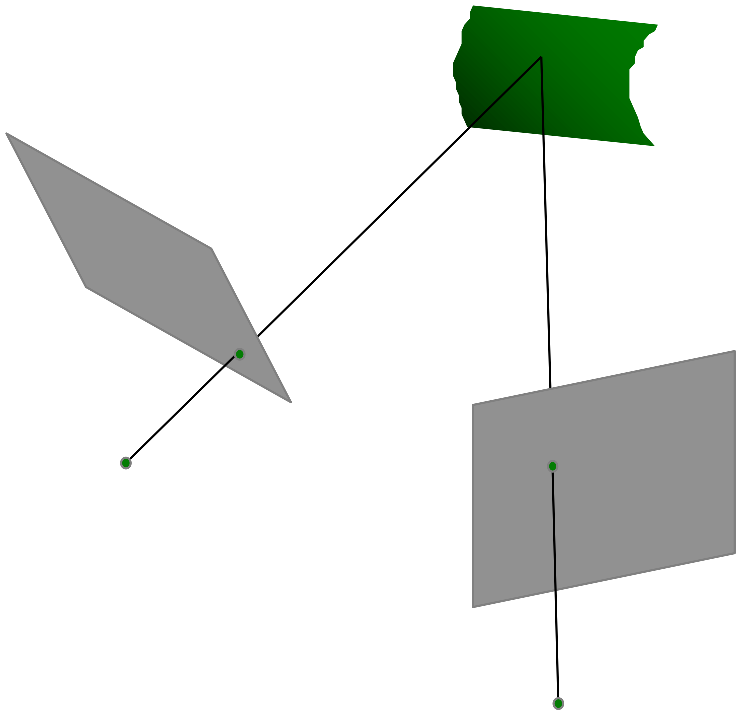




Use stereo rectification?

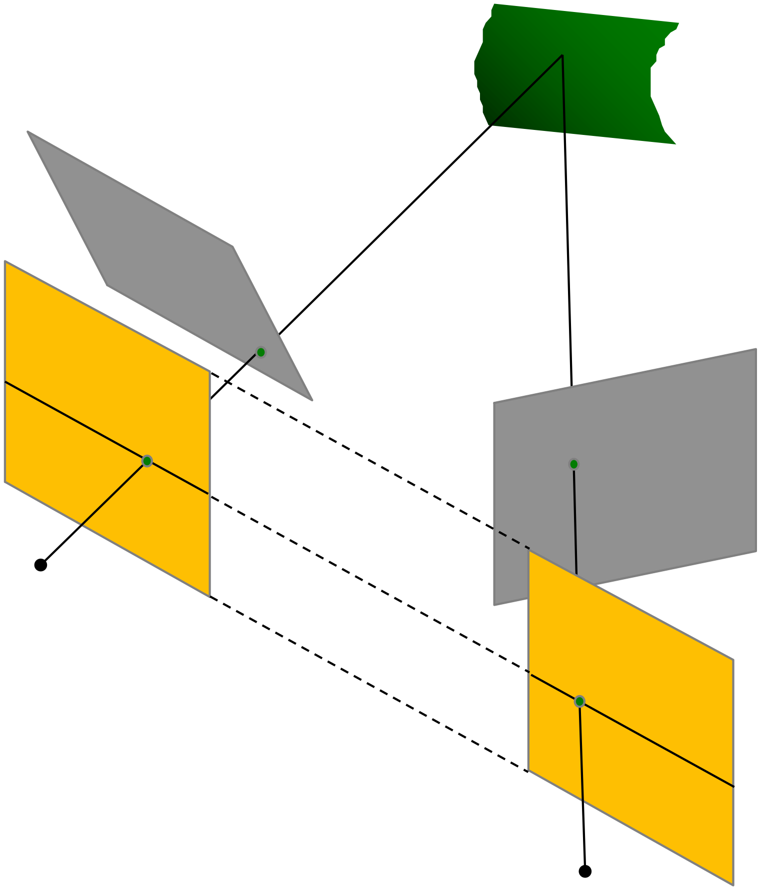


What is stereo rectification?



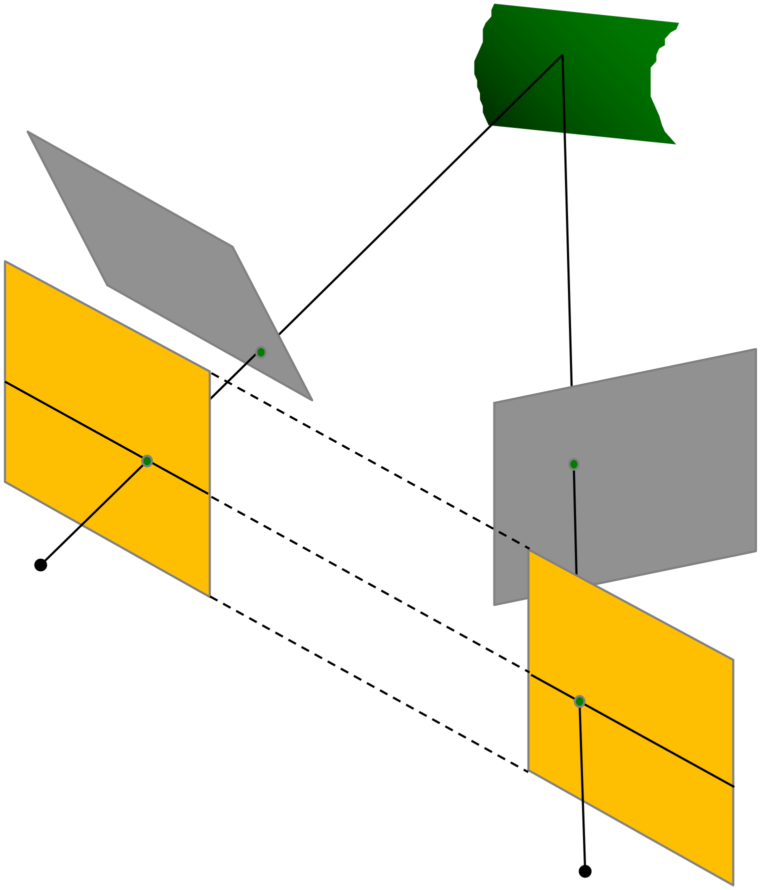
What is stereo rectification?

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



How can you do this?

What is stereo rectification?



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

Need two homographies (3x3 transform), one for each input image reprojection

C. Loop and Z. Zhang. Computing Rectifying Homographies for Stereo Vision.Computer Vision and Pattern Recognition, 1999.

Stereo Rectification

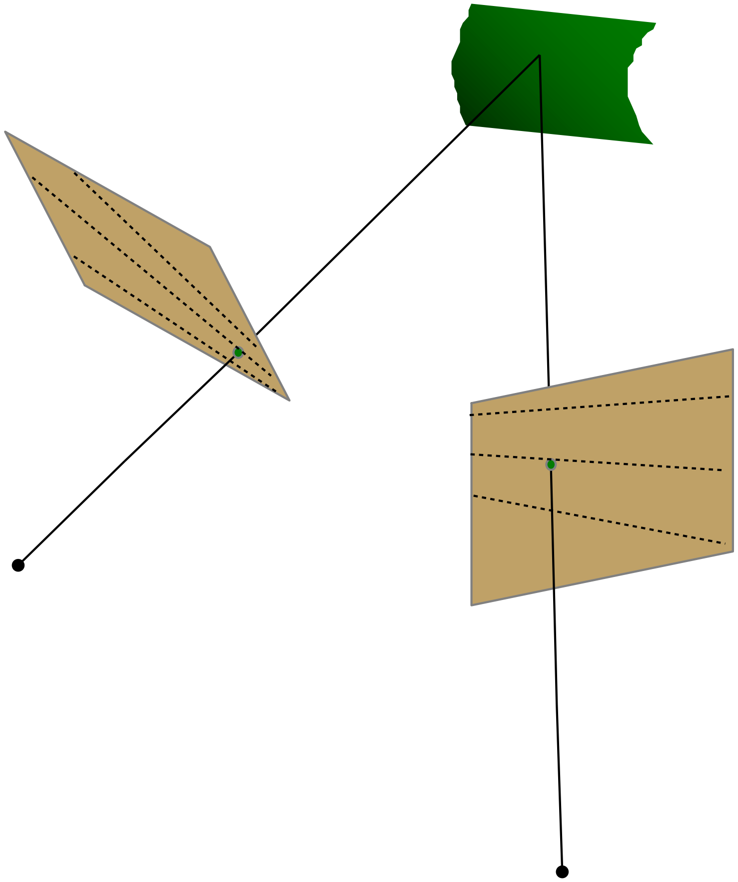
1. **Rotate** the right camera by **R**
   (aligns camera coordinate system orientation only)

2. Rotate (**rectify**) the left camera so that the epipole is at infinity

3. Rotate (**rectify**) the right camera so that the epipole is at infinity

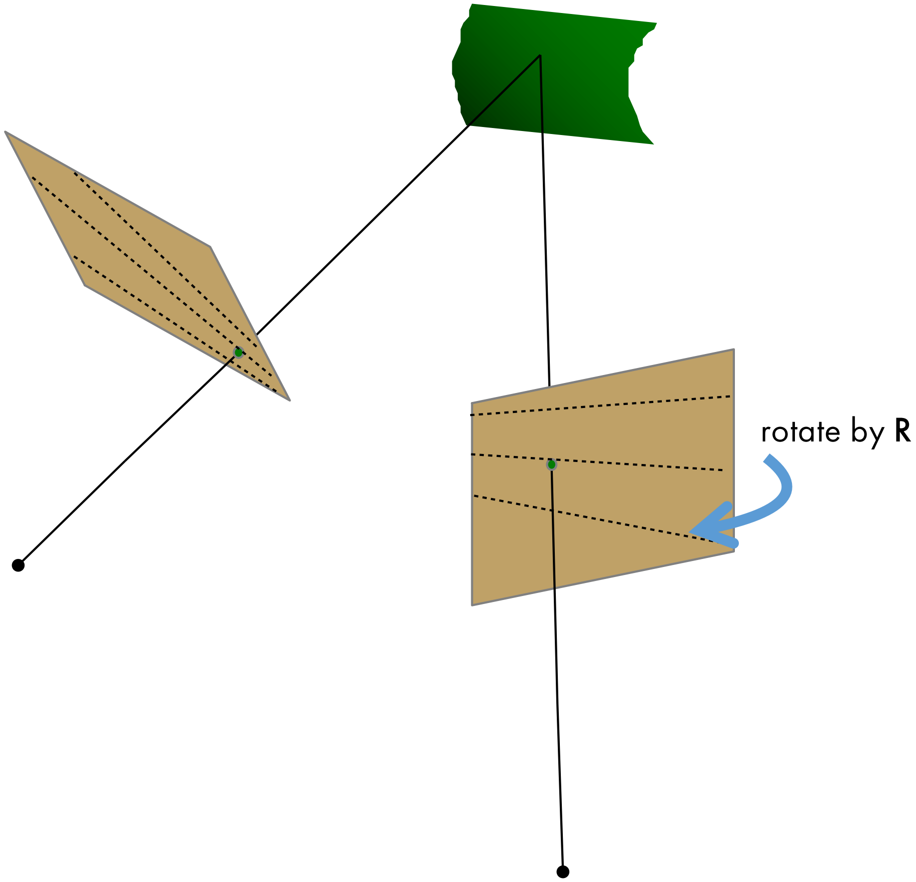
4. Adjust the **scale**

Stereo Rectification:



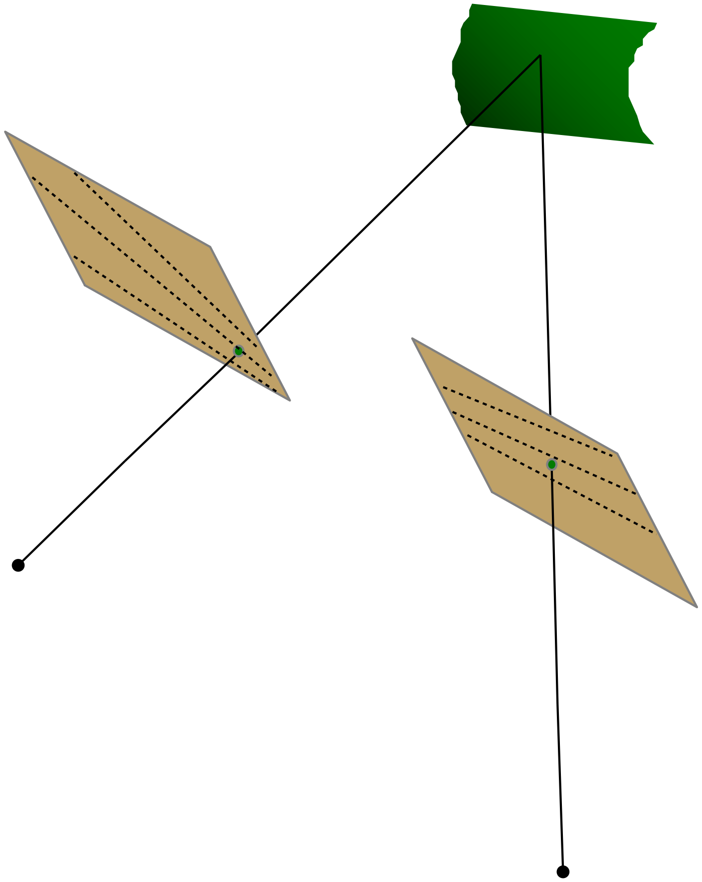
1. Compute **E** to get **R**
2. Rotate right image by **R**
3. Rotate both images by **R**rect
4. Scale both images by **H**

Stereo Rectification:



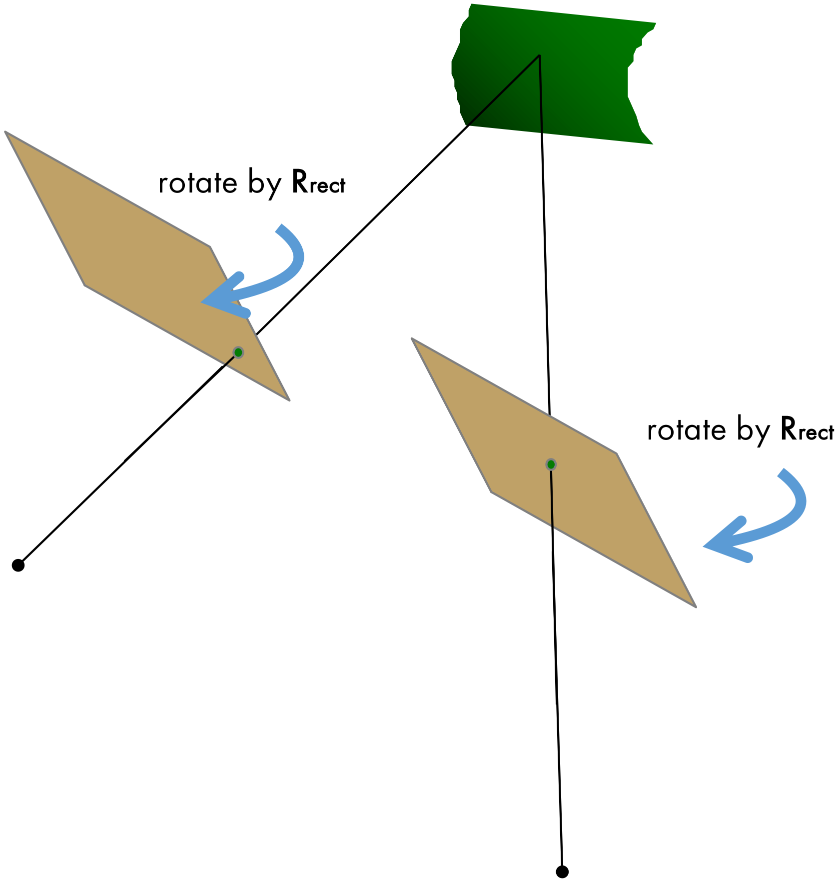
1. Compute **E** to get **R**
2. Rotate right image by **R**
3. Rotate both images by $\mathbf{R}_{\text{rect}}$
4. Scale both images by **H**

Stereo Rectification:



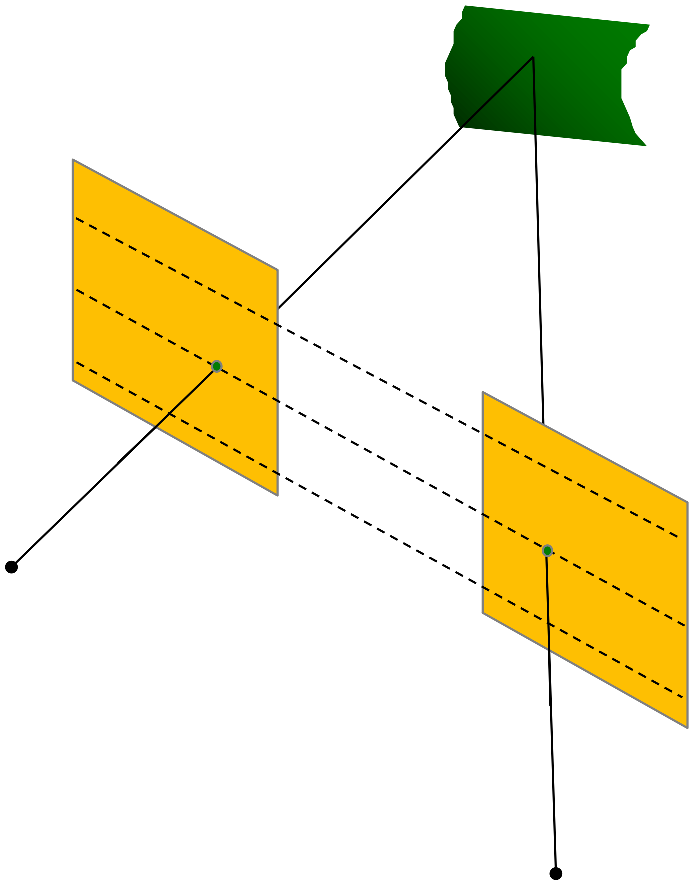
1. Compute **E** to get **R**
2. Rotate right image by **R**
3. Rotate both images by $\mathbf{R}_{rect}$
4. Scale both images by **H**

Stereo Rectification:



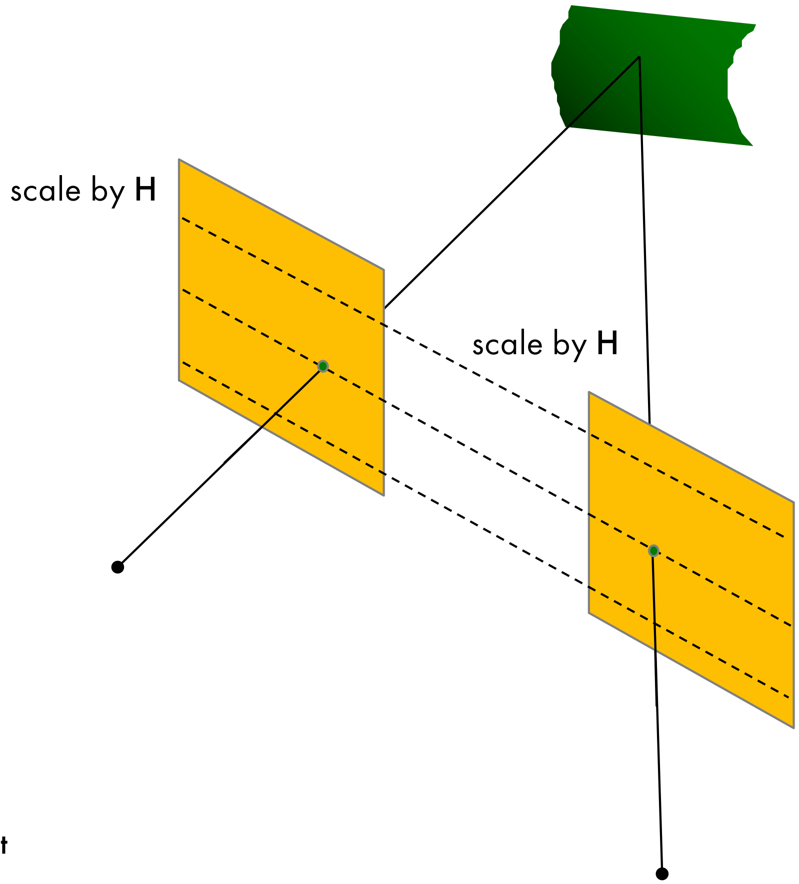
1. Compute **E** to get **R**
2. Rotate right image by **R**
3. Rotate both images by **R**rect
4. Scale both images by **H**

Stereo Rectification:



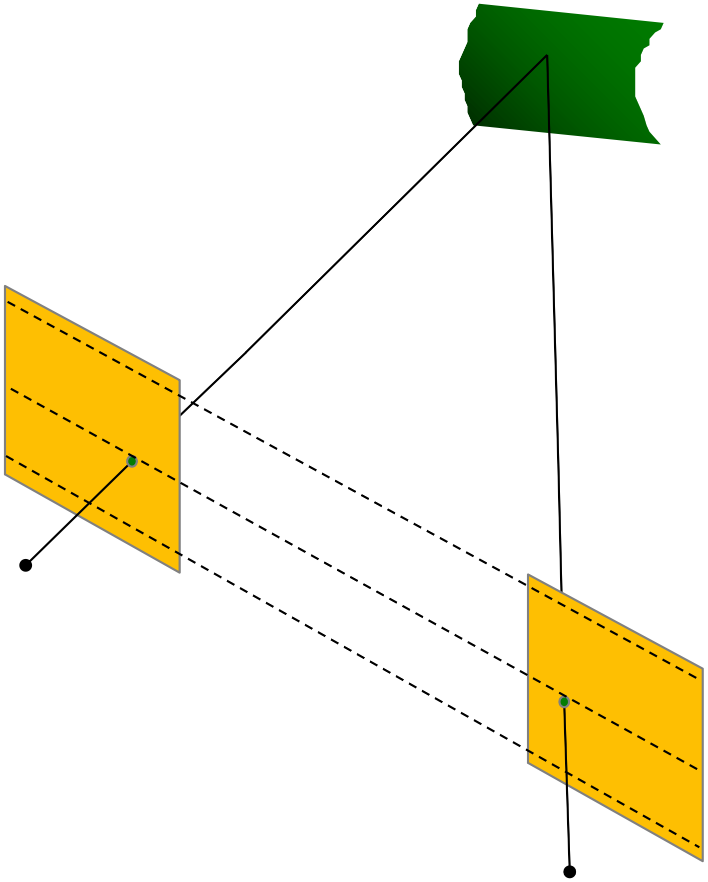
1. Compute **E** to get **R**
2. Rotate right image by **R**
3. Rotate both images by $\mathbf{R}_{\mathbf{rect}}$
4. Scale both images by **H**

Stereo Rectification:



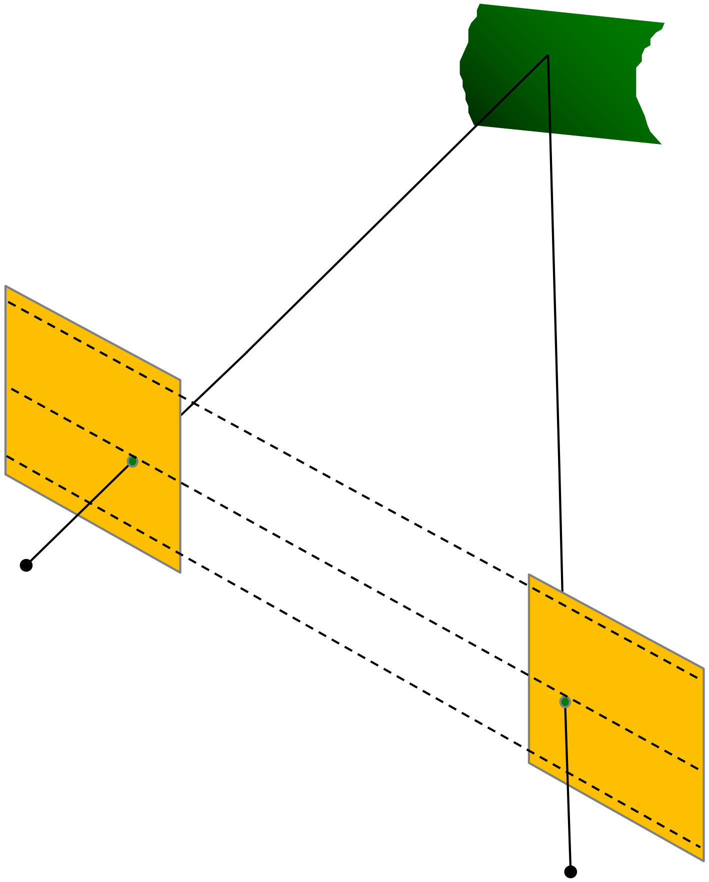
1. Compute **E** to get **R**
2. Rotate right image by **R**
3. Rotate both images by **R**$_{\text{rect}}$
4. Scale both images by **H**

Stereo Rectification:



1. Compute **E** to get **R**
2. Rotate right image by **R**
3. Rotate both images by $\mathbf{R}_{\text{rect}}$
4. Scale both images by **H**

Stereo Rectification:



1. Compute **E** to get **R**
2. Rotate right image by **R**
3. Rotate both images by $\mathbf{R}_{\text{rect}}$
4. Scale both images by **H**

Step 1: Compute E to get R

SVD: $\mathbf{E} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$ Let $\mathbf{W} = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

We get FOUR solutions:

$$\mathbf{E} = [\mathbf{T}]_\times \mathbf{R}$$

$\mathbf{R}_1 = \mathbf{U}\mathbf{W}\mathbf{V}^\top$ $\mathbf{R}_2 = \mathbf{U}\mathbf{W}^\top\mathbf{V}^\top$

two possible rotations

$\mathbf{T}_1 = U_3$ $\mathbf{T}_2 = -U_3$

two possible translations

Step 1: Compute E to get R

We get FOUR solutions:

$$\mathbf{E} = [\mathbf{T}]_\times \mathbf{R}$$

$$\mathbf{T}_1 = U_3 \qquad \mathbf{T}_2 = -U_3$$

two possible translations

$$\mathbf{T}^T [\mathbf{T}]_\times \mathbf{R} = 0$$

so T must be the left nullspace of E

Step 1: Compute **E** to get **R**

We get FOUR solutions:

$$\mathbf{E} = [\mathbf{T}]_\times \mathbf{R}$$

$$\mathbf{R}_1 = \mathbf{U}\mathbf{W}\mathbf{V}^\top \quad \mathbf{R}_2 = \mathbf{U}\mathbf{W}^\top\mathbf{V}^\top \qquad \mathbf{W} = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

two possible rotations

Step 1: Compute **E** to get **R**

We get FOUR solutions:

$$\mathbf{E} = [\mathbf{T}]_{\times}\mathbf{R}$$

$$\mathbf{R}_1 = \mathbf{U}\mathbf{W}\mathbf{V}^\top \quad \mathbf{R}_2 = \mathbf{U}\mathbf{W}^\top\mathbf{V}^\top \qquad \mathbf{W} = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

two possible rotations

$$[\mathbf{t}]_{\times}\,\mathbf{R} = \mathbf{U}\,\mathbf{W}\,\mathbf{\Sigma}\,\mathbf{U}^T\mathbf{U}\,\mathbf{W}^{-1}\,\mathbf{V}^T \;= \mathbf{U}\,\mathbf{\Sigma}\,\mathbf{V}^T = \mathbf{E}$$

Step 1: Compute E to get R

We get FOUR solutions:

$$\mathbf{E} = [\mathbf{T}]_\times \mathbf{R}$$

$$\mathbf{R}_1 = \mathbf{U}\mathbf{W}\mathbf{V}^\top \quad \mathbf{R}_2 = \mathbf{U}\mathbf{W}^\top\mathbf{V}^\top \qquad \mathbf{W} = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

two possible rotations

$$[\mathbf{t}]_\times \mathbf{R} = \underbrace{\mathbf{U}\,\mathbf{W}\,\mathbf{\Sigma}\,\mathbf{U}^T}_{\text{this extracts } U_3}\mathbf{U}\,\mathbf{W}^{-1}\,\mathbf{V}^T = \mathbf{U}\,\mathbf{\Sigma}\,\mathbf{V}^T = \mathbf{E}$$

Step 1: Compute E to get R

We get FOUR solutions:

$$\mathbf{E} = [\mathbf{T}]_\times \mathbf{R}$$

$$\mathbf{R}_1 = \mathbf{U}\mathbf{W}\mathbf{V}^\top \quad \mathbf{R}_2 = \mathbf{U}\mathbf{W}^\top\mathbf{V}^\top \qquad \mathbf{W} = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

two possible rotations

$$[\mathbf{t}]_\times \, \mathbf{R} = \mathbf{U}\,\mathbf{W}\,\mathbf{\Sigma}\,\mathbf{U}^T \underbrace{\mathbf{U}\,\mathbf{W}^{-1}\,\mathbf{V}^T} \; = \mathbf{U}\,\mathbf{\Sigma}\,\mathbf{V}^T = \mathbf{E}$$

all of these matrices are orthogonal, so the result is orthogonal—and a rotation matrix!

We get FOUR solutions:

$$\mathbf{R}_1 = \mathbf{U}\mathbf{W}\mathbf{V}^\top$$
$$\mathbf{T}_1 = U_3$$

$$\mathbf{R}_1 = \mathbf{U}\mathbf{W}\mathbf{V}^\top$$
$$\mathbf{T}_2 = -U_3$$

$$\mathbf{R}_2 = \mathbf{U}\mathbf{W}^\top\mathbf{V}^\top$$
$$\mathbf{T}_2 = -U_3$$

$$\mathbf{R}_2 = \mathbf{U}\mathbf{W}^\top\mathbf{V}^\top$$
$$\mathbf{T}_1 = U_3$$

Which one do we choose?

Compute determinant of R, valid solution must be equal to 1

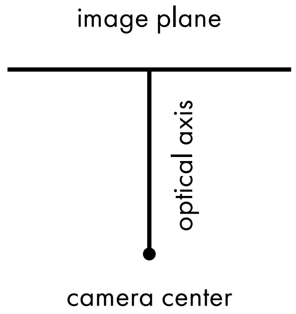
(note: det(R) = -1 means rotation and reflection)

Compute 3D point using triangulation, valid solution has positive Z value

(Note: negative Z means point is behind the camera)

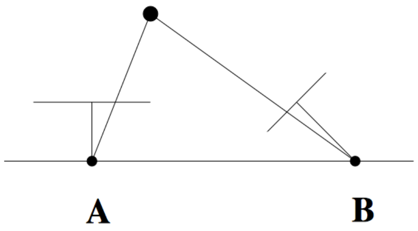
Let's visualize the four configurations...

Camera Icon

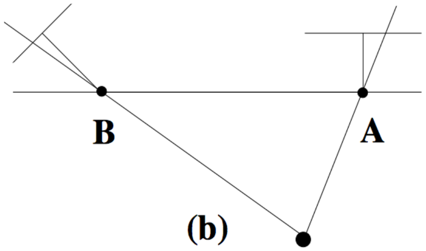


Find the configuration where the point is in front of both cameras

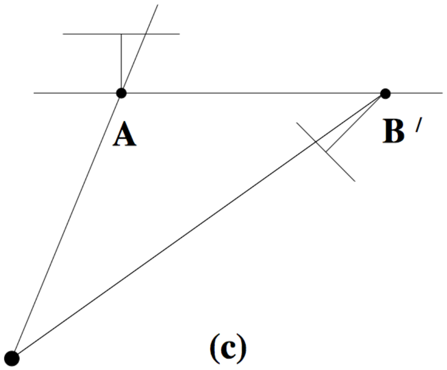
Find the configuration where the point is in front of both cameras



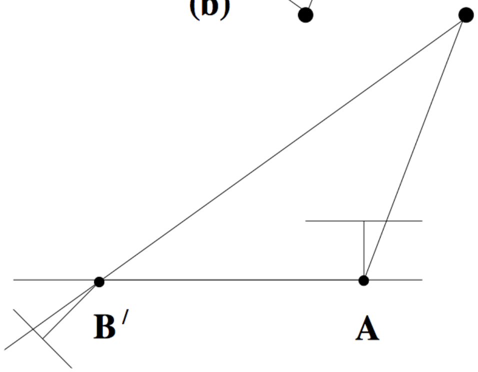
(a)



(b)

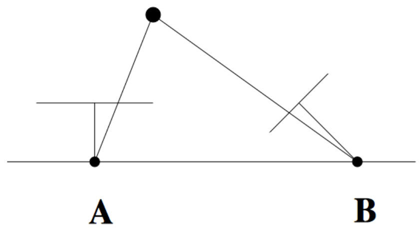


(c)

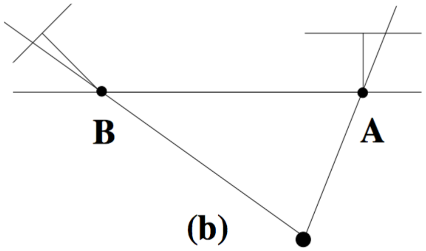


(d)

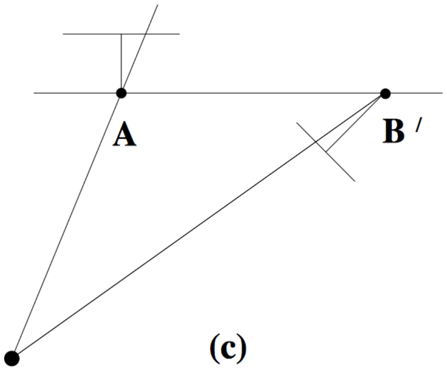
Find the configuration where the point is in front of both cameras



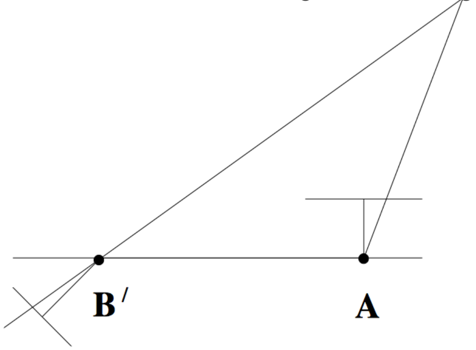
(a)



(b)

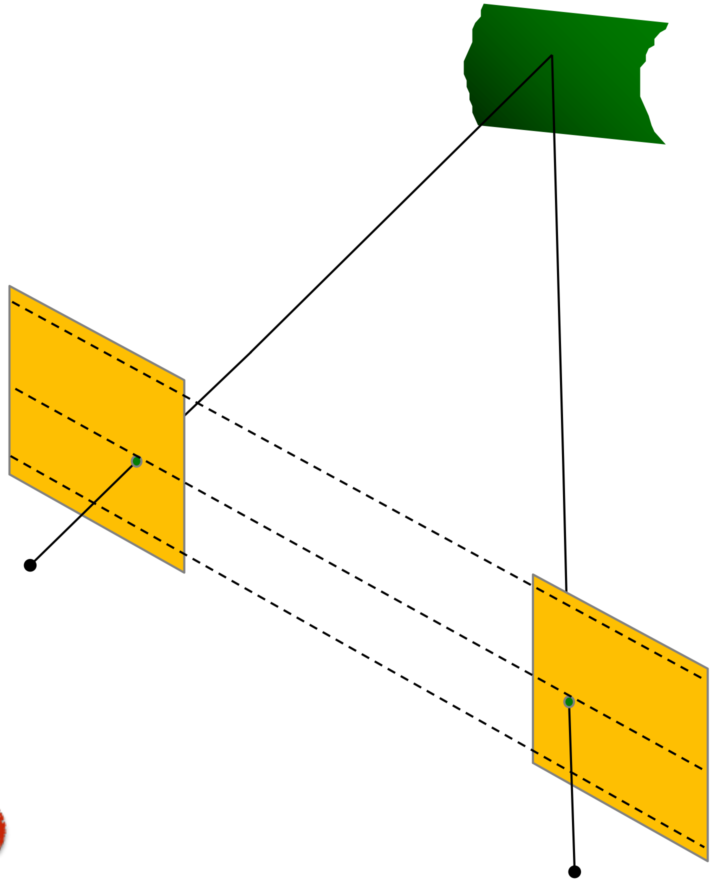


(c)



(d)

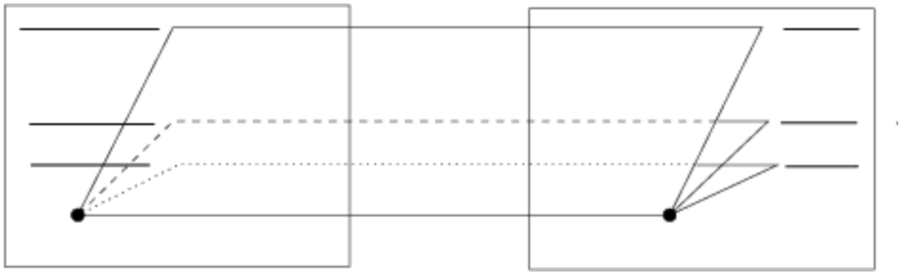
Stereo Rectification:

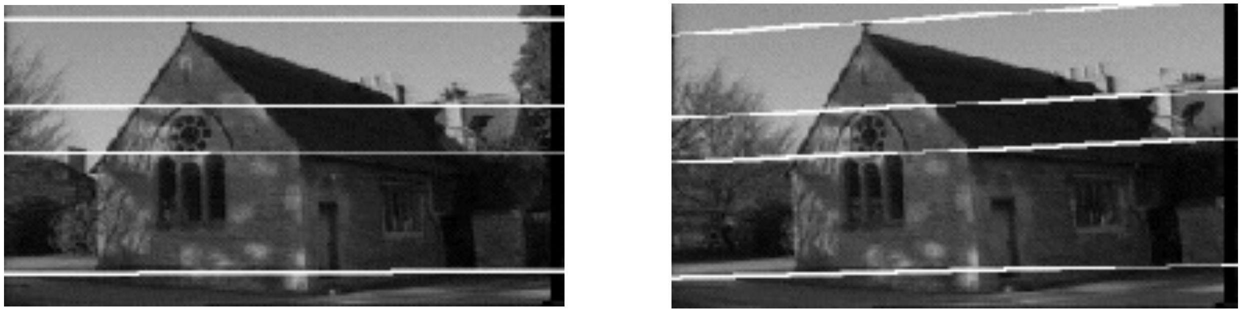


1. Compute E to get R
2. Rotate right image by R
3. Rotate both images by R_{rect}
4. Scale both images by H

When do epipolar lines become horizontal?

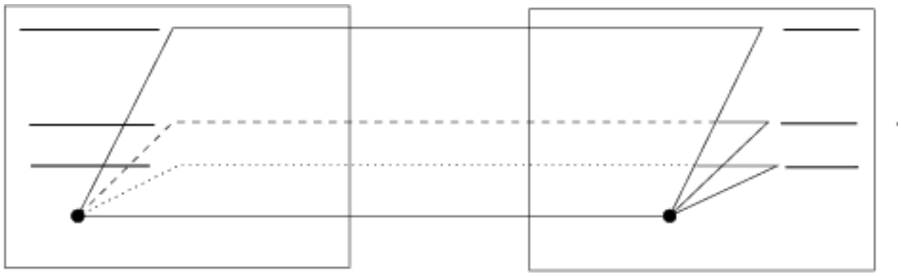
Parallel cameras

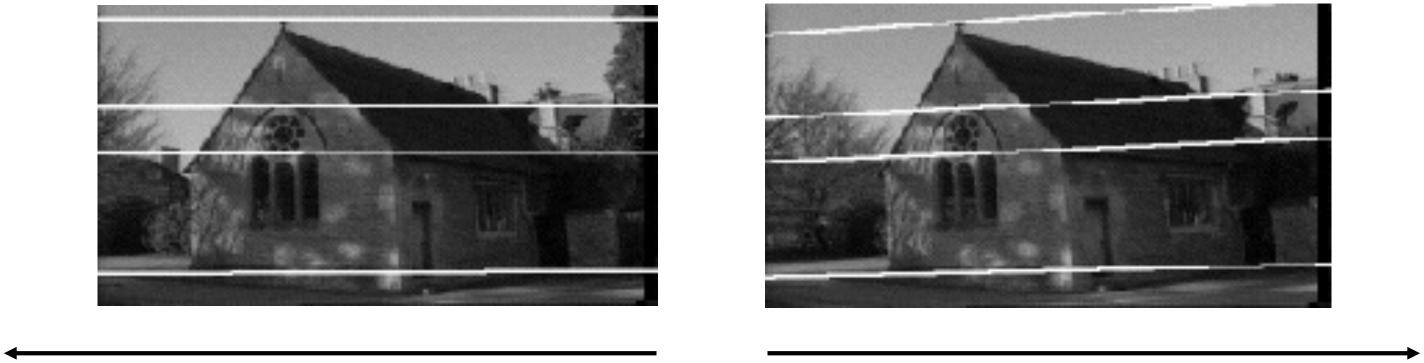




Where is the epipole?

Parallel cameras





epipole at infinity

Setting the epipole to infinity

(Building **R**rect from **E**)

Let $$R_{\text{rect}} = \begin{bmatrix} \boldsymbol{r}_1^\top \\ \boldsymbol{r}_2^\top \\ \boldsymbol{r}_3^\top \end{bmatrix}$$

Given:

epipole e
(using SVD on E)
(translation from E)

$$\boldsymbol{r}_1 = \boldsymbol{e}_1 = \frac{T}{||T||}$$

epipole coincides with translation vector

$$\boldsymbol{r}_2 = \frac{1}{\sqrt{T_x^2 + T_y^2}} \left[\begin{array}{ccc} -T_y & T_x & 0 \end{array} \right]$$

cross product of e and the direction vector of the optical axis [0, 0, 1]

$$\boldsymbol{r}_3 = \boldsymbol{r}_1 \times \boldsymbol{r}_2$$

orthogonal vector

If $\boldsymbol{r}_1 = \boldsymbol{e}_1 = \frac{T}{||T||}$ and $\boldsymbol{r}_2$ $\boldsymbol{r}_3$ orthogonal

then $R_{\text{rect}}\boldsymbol{e}_1 = \left[\begin{array}{c} \boldsymbol{r}_1^\top \boldsymbol{e}_1 \\ \boldsymbol{r}_2^\top \boldsymbol{e}_1 \\ \boldsymbol{r}_3^\top \boldsymbol{e}_1 \end{array}\right] = \left[\begin{array}{c} ? \\ ? \\ ? \end{array}\right]$

If $\boldsymbol{r}_1 = \boldsymbol{e}_1 = \frac{T}{||T||}$ and $\boldsymbol{r}_2$ $\boldsymbol{r}_3$ orthogonal

then $$R_{\text{rect}}\boldsymbol{e}_1 = \begin{bmatrix} \boldsymbol{r}_1^\top \boldsymbol{e}_1 \\ \boldsymbol{r}_2^\top \boldsymbol{e}_1 \\ \boldsymbol{r}_3^\top \boldsymbol{e}_1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

Where is this point located on the image plane?

If $\boldsymbol{r}_1 = \boldsymbol{e}_1 = \frac{T}{||T||}$ and $\boldsymbol{r}_2$ $\boldsymbol{r}_3$ orthogonal

then $R_{\text{rect}}\boldsymbol{e}_1 = \begin{bmatrix} \boldsymbol{r}_1^\top \boldsymbol{e}_1 \\ \boldsymbol{r}_2^\top \boldsymbol{e}_1 \\ \boldsymbol{r}_3^\top \boldsymbol{e}_1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$

Where is this point located on the image plane?

At x-infinity

Stereo Rectification Algorithm

1. Estimate **E** using the 8 point algorithm (SVD)

2. Estimate the epipole **e** (SVD of **E**)

3. Build $\mathbf{R}_{rect}$ from **e**

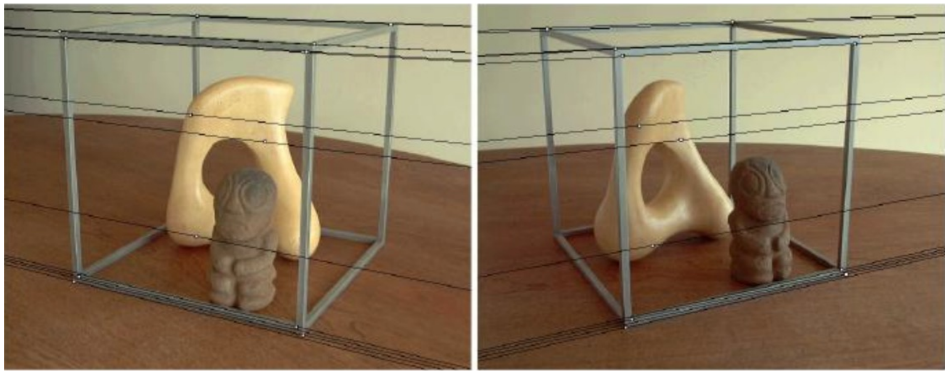
4. Decompose **E** into **R** and **T**

5. Set $\mathbf{R}_1 = \mathbf{R}_{rect}$ and $\mathbf{R}_2 = \mathbf{R}\mathbf{R}_{rect}$

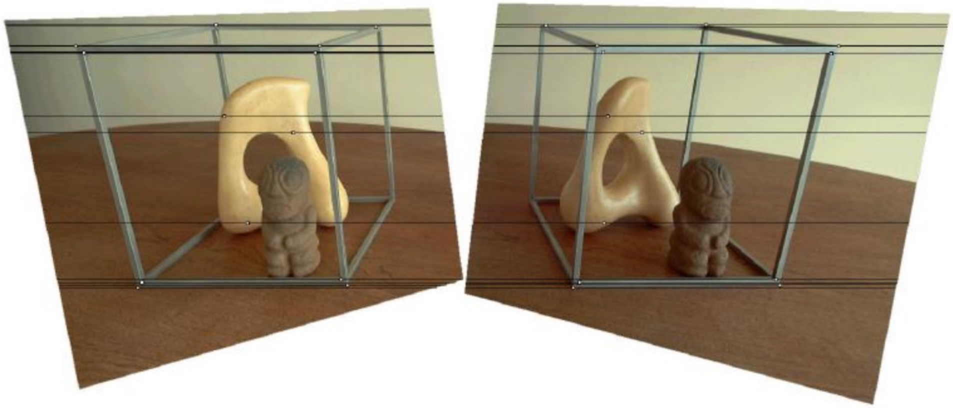
6. Rotate each left camera point (warp image) ← requires backprojection
   $[x' \ y' \ z'] = \mathbf{R}_1 [x \ y \ z]$

7. Rectified points as $\mathbf{p} = f/z'[x' \ y' \ z']$

8. Repeat 6 and 7 for right camera points using $\mathbf{R}_2$

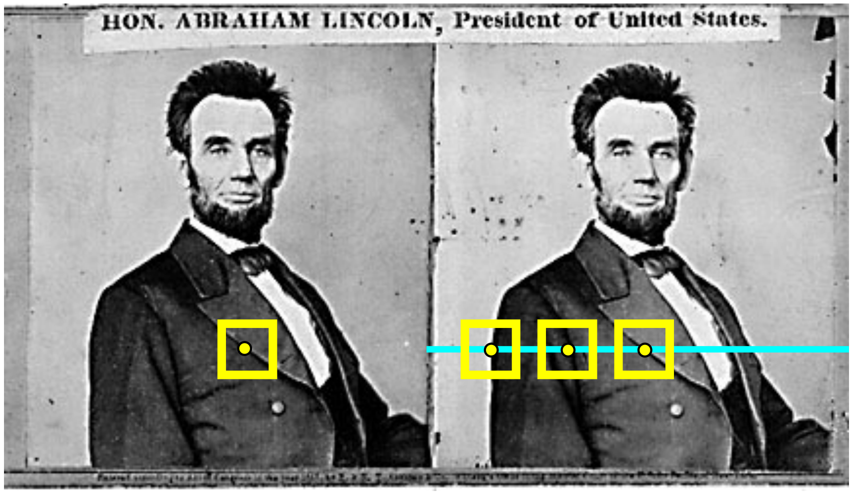


What can we do after rectification?



Stereo matching

Depth Estimation via Stereo Matching



1. Rectify images
   (make epipolar lines horizontal)
2. For each pixel
   a. Find epipolar line
   b. Scan line for best match
   c. Compute depth from disparity

$$Z = \frac{bf}{d}$$

How would you do this?

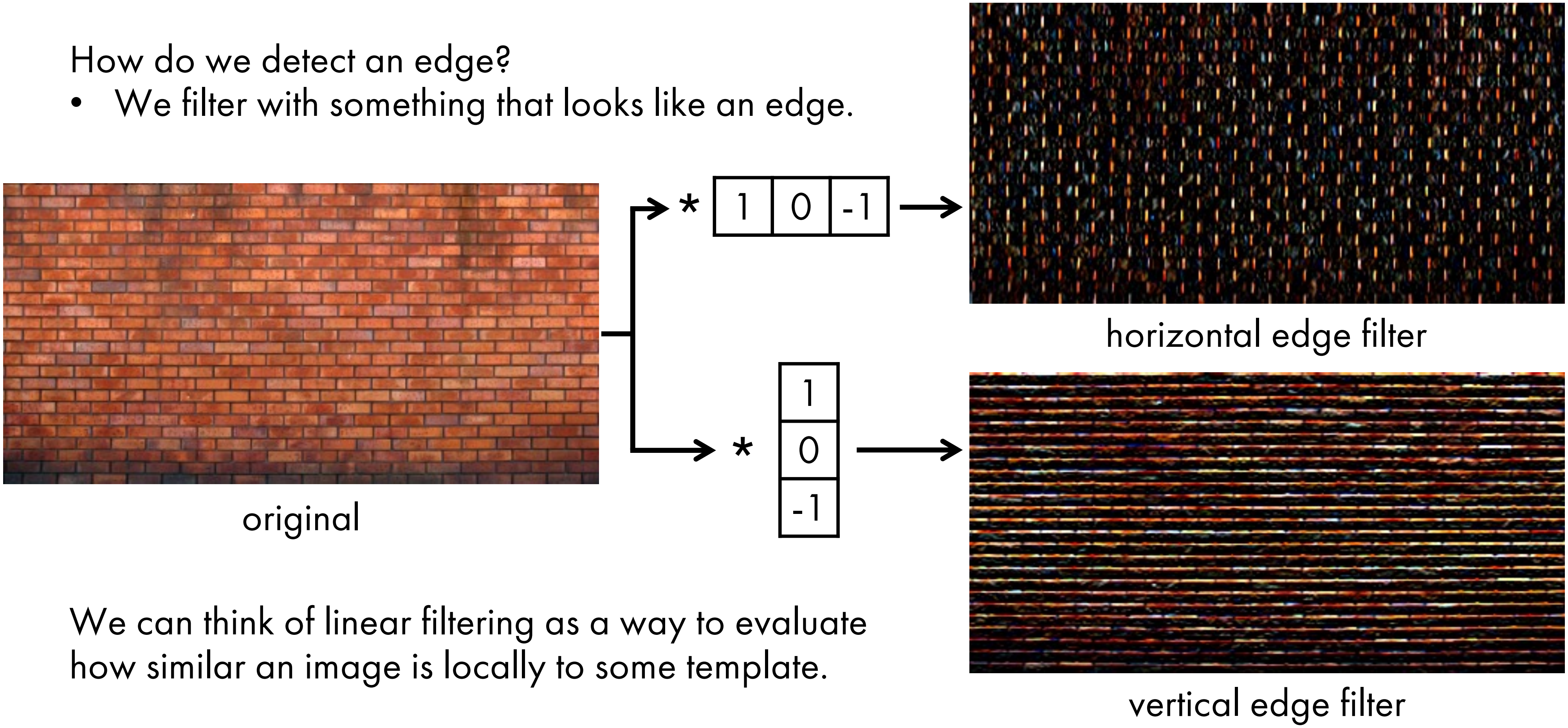
Reminder from filtering

How do we detect an edge?

Reminder from filtering

How do we detect an edge?

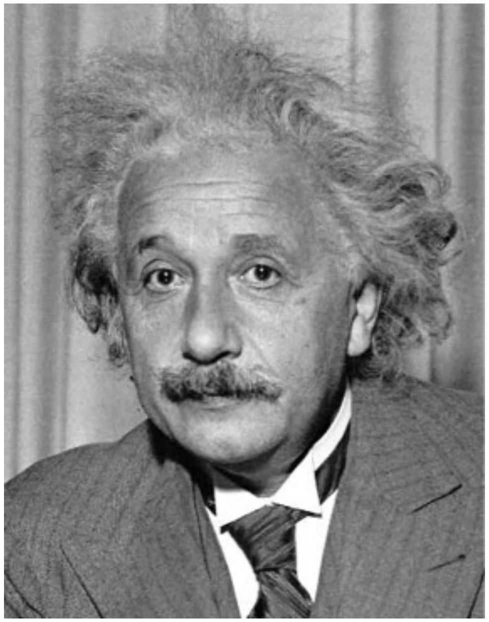
- We filter with something that looks like an edge.



We can think of linear filtering as a way to evaluate how similar an image is locally to some template.

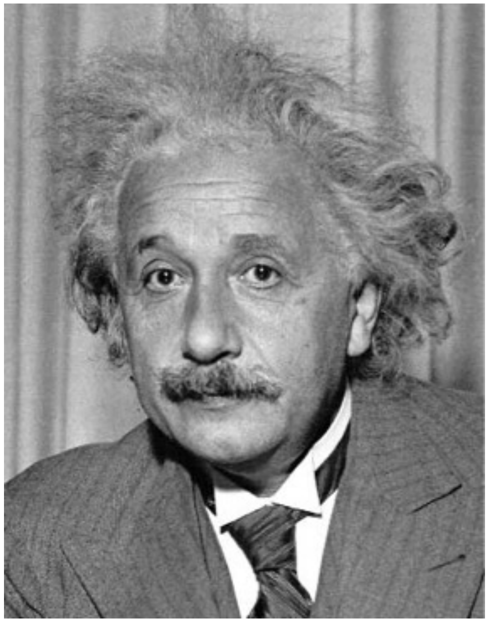
Find this template

How do we detect the template  n he following image?



Find this template

How do we detect the template n he following image?



filter 

output

$$h[m,n] = \sum_{k,j} g[k,l] f[m+k, n+l]$$

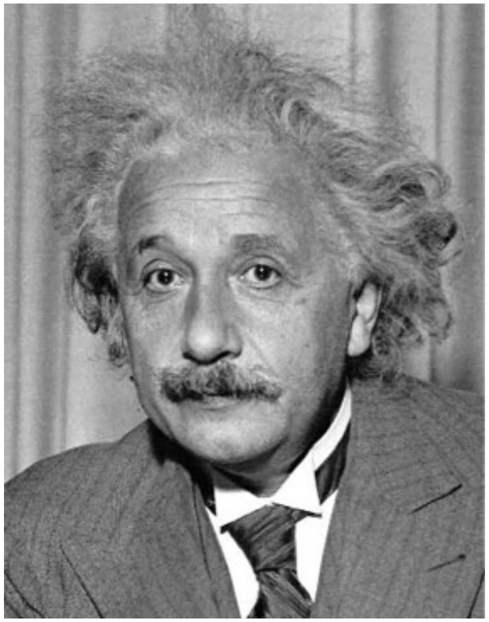
image

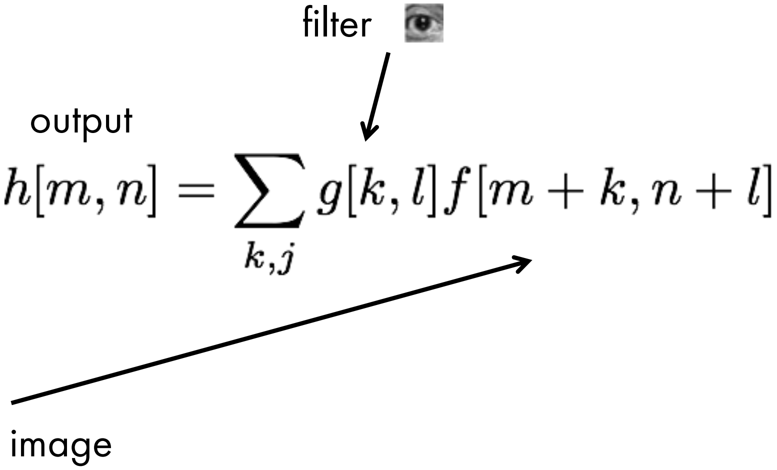
What will
the output
look like?

Solution 1: Filter the image using the template as filter kernel.

Find this template

How do we detect the template  n he following image?



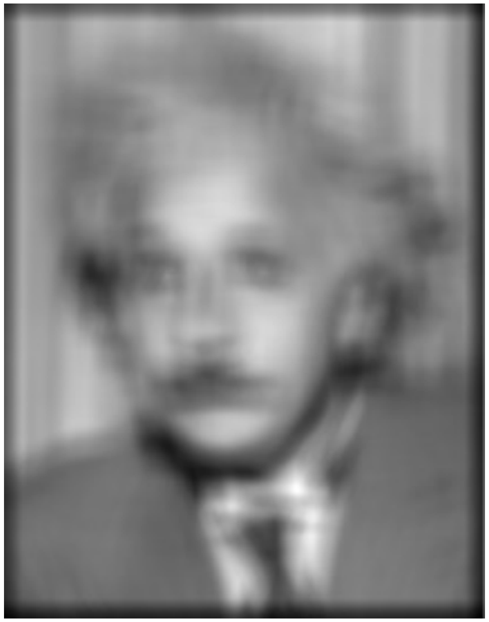


filter

output

$$h[m,n] = \sum_{k,j} g[k,l] f[m+k, n+l]$$

image

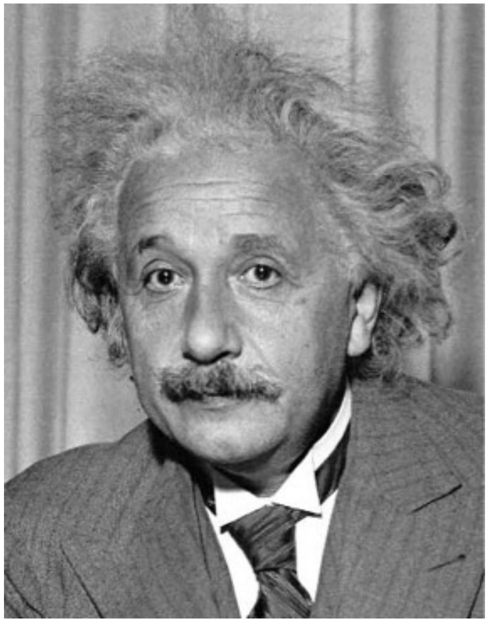


Solution 1: Filter the image using the template as filter kernel.

What went wrong?

Find this template

How do we detect the template n he following image?

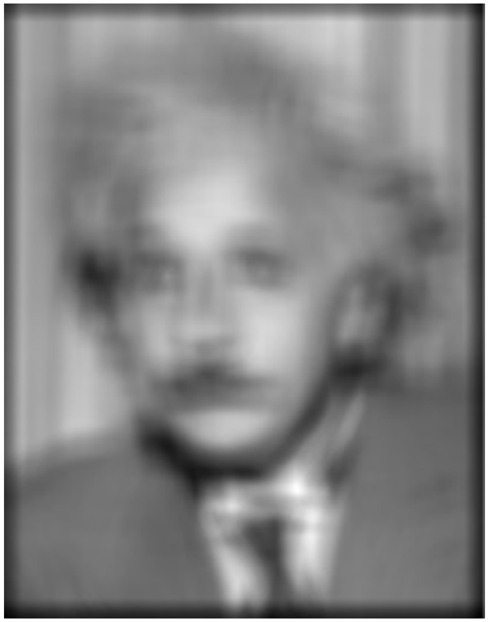


filter 

output

$$h[m,n] = \sum_{k,j} g[k,l] f[m+k, n+l]$$

image

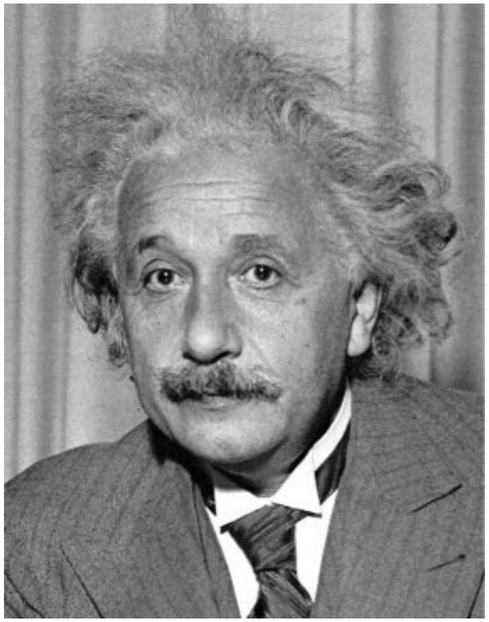


Increases for higher local intensities.

Solution 1: Filter the image using the template as filter kernel.

Find this template

How do we detect the template  n he following image?



filter  template mean

output

$$h[m,n] = \sum_{k,l} (g[k,l] - \bar{g}) f[m+k, n+l]$$

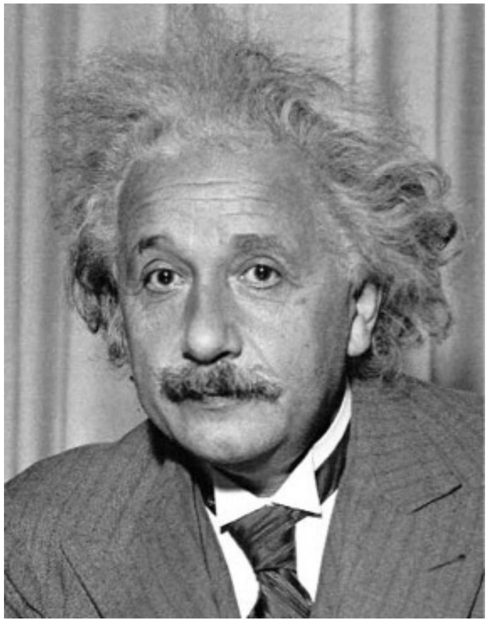
image

What will
the output
look like?

Solution 2: Filter the image using a zero-mean template.

Find this template

How do we detect the template  n he following image?



filter  template mean

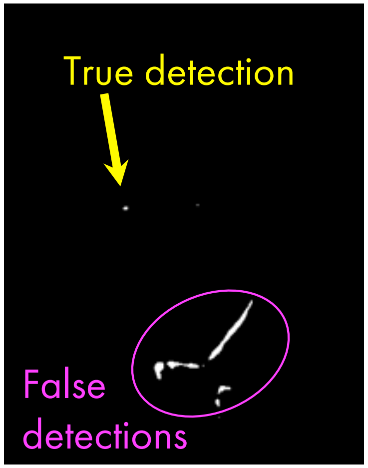
output

$$h[m,n] = \sum_{k,l} (g[k,l] - \bar{g}) f[m+k, n+l]$$

image

output

thresholding

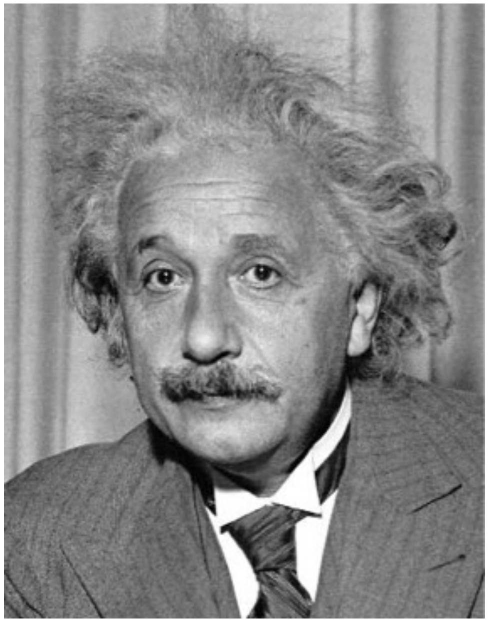


Solution 2: Filter the image using a zero-mean template.

What went wrong?

Find this template

How do we detect the template  n he following image?



filter 

template mean

output

$$h[m,n] = \sum_{k,l} (g[k,l] - \bar{g}) f[m+k, n+l]$$

image

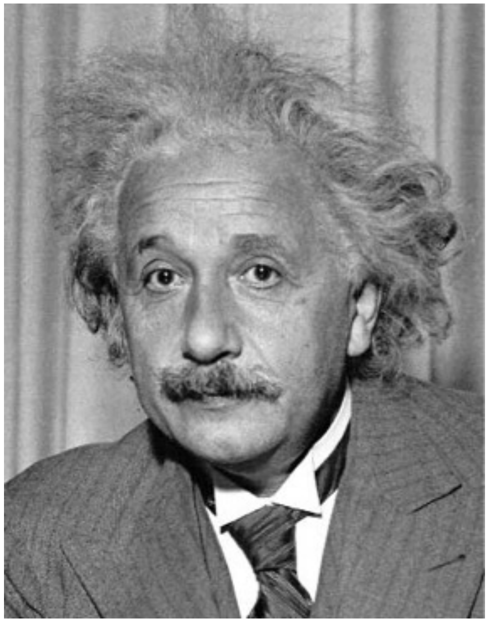
output

Not robust to high-contrast areas

Solution 2: Filter the image using a zero-mean template.

Find this template

How do we detect the template  n he following image?



filter 

output

$$h[m,n] = \sum_{k,l} (g[k,l] - f[m+k, n+l])^2$$

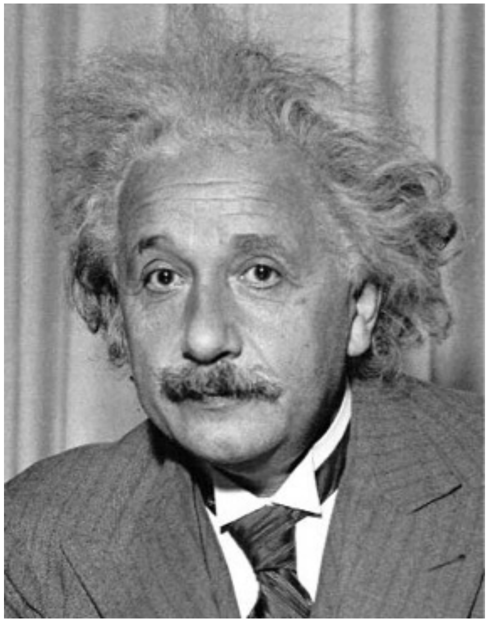
image

What will the output look like?

Solution 3: Use sum of squared differences (SSD).

Find this template

How do we detect the template  n he following image?



output

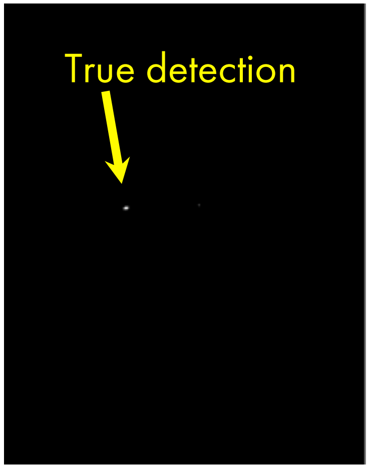
filter 

$$h[m,n] = \sum_{k,l} (g[k,l] - f[m+k,n+l])^2$$

image

1-output

thresholding

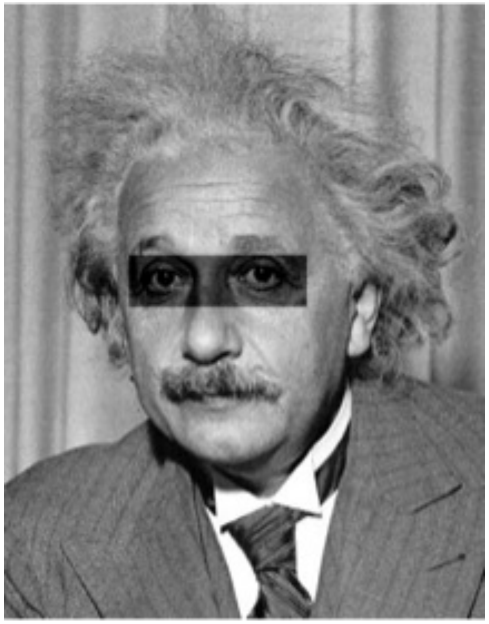


Solution 3: Use sum of squared differences (SSD).

What could go wrong?

Find this template

How do we detect the template  n he following image?



1-output

filter 

output

$$h[m,n] = \sum_{k,l} (g[k,l] - f[m+k, n+l])^2$$

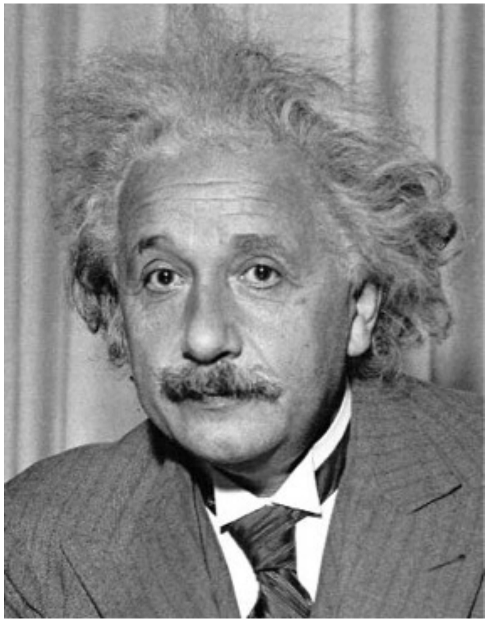
image

Not robust to local intensity changes

Solution 3: Use sum of squared differences (SSD).

Find this template

How do we detect the template n he following image?



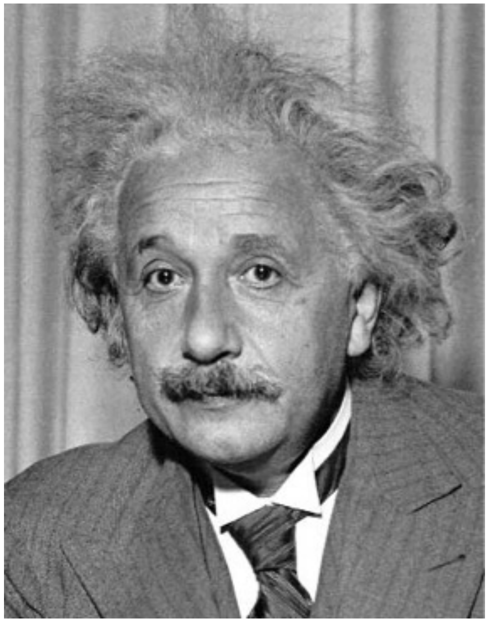
Observations so far:

- subtracting mean deals with brightness bias
- dividing by standard deviation removes contrast bias

Can we combine the two effects?

Find this template

How do we detect the template  n he following image?



What will the output look like?

filter 

template mean

output

$$h[m,n] = \frac{\sum_{k,l}(g[k,l]-\bar{g})(f[m+k,n+l]-\bar{f}_{m,n})}{\sqrt{(\sum_{k,l}(g[k,l]-\bar{g})^2\sum_{k,l}(f[m+k,n+l]-\bar{f}_{m,n})^2)}}$$

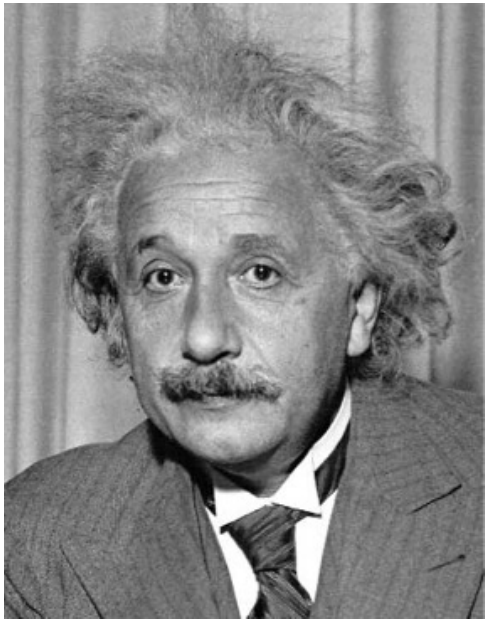
local patch mean

image

Solution 4: Normalized cross-correlation (NCC).

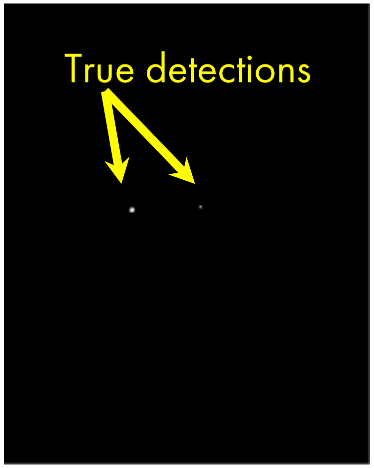
Find this template

How do we detect the template  n he following image?



1-output

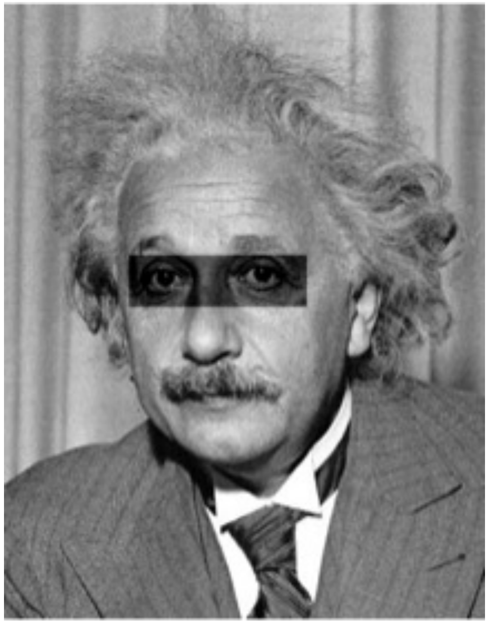
thresholding



Solution 4: Normalized cross-correlation (NCC).

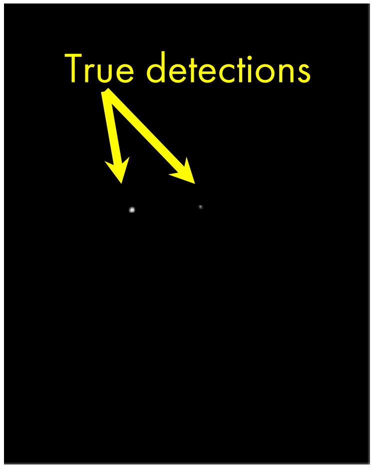
Find this template

How do we detect the template  n he following image?



1-output

thresholding



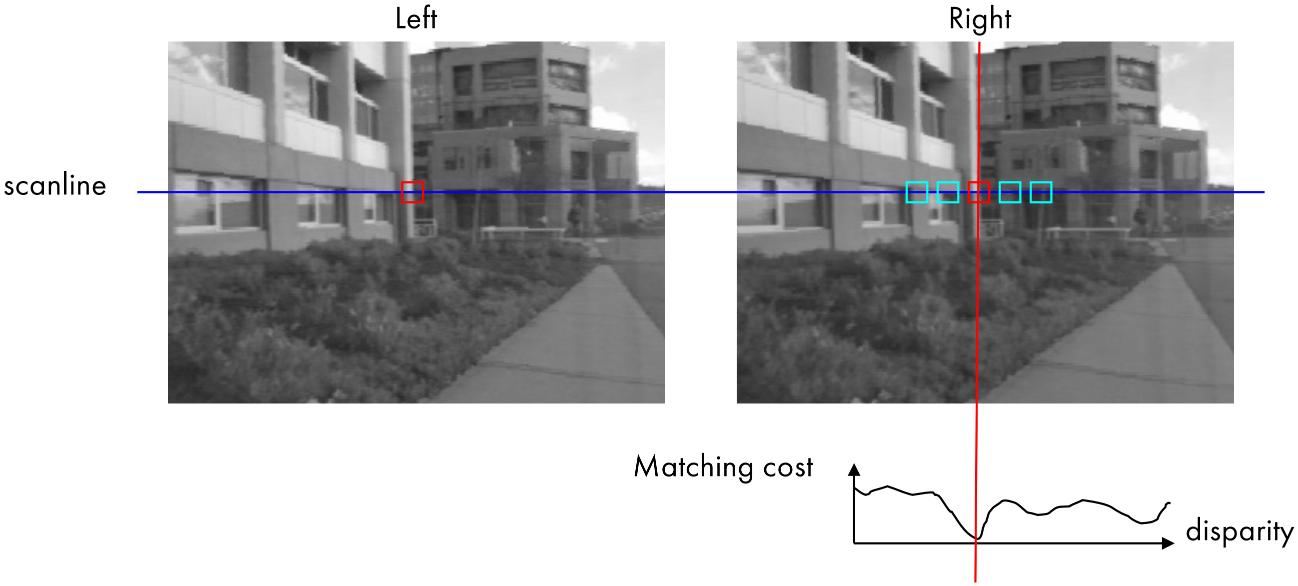
Solution 4: Normalized cross-correlation (NCC).

What is the best method?

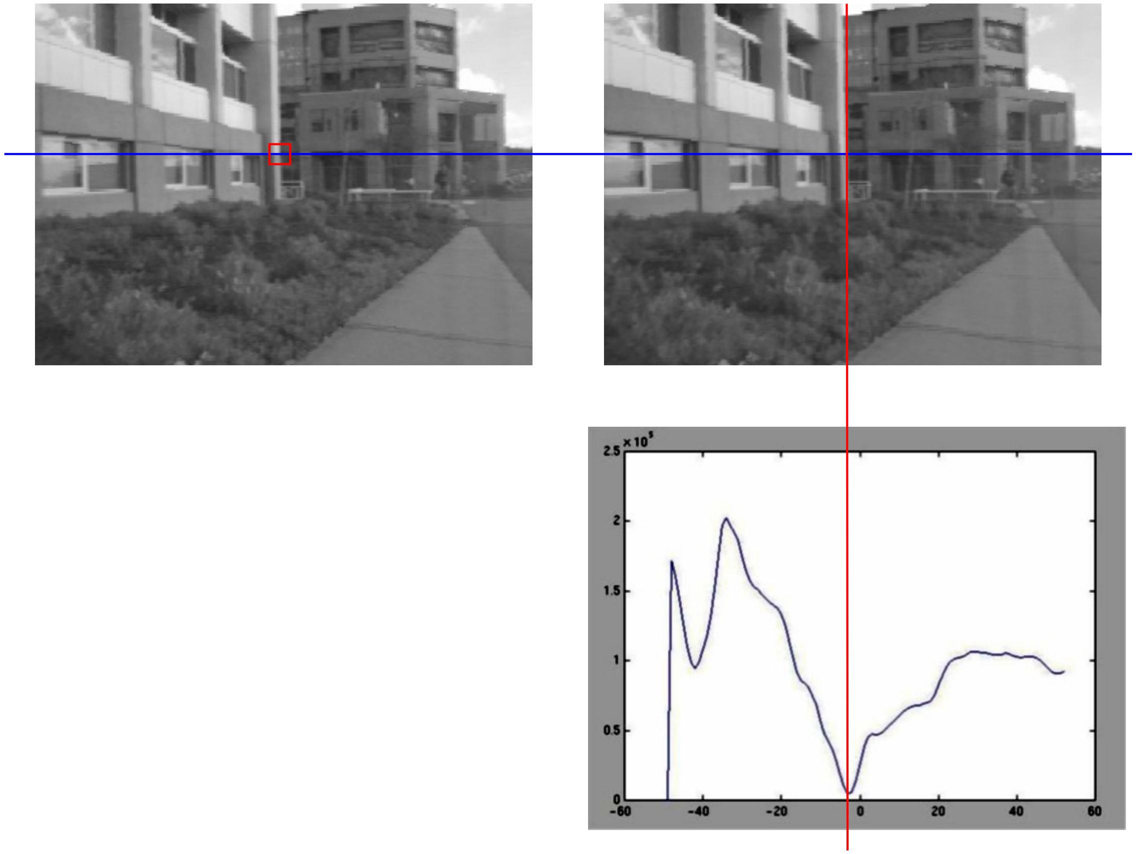
It depends on whether you care about speed or invariance.

- Zero-mean: Fastest, very sensitive to local intensity.
- Sum of squared differences: Medium speed, sensitive to intensity offsets.
- Normalized cross-correlation: Slowest, invariant to contrast and brightness.

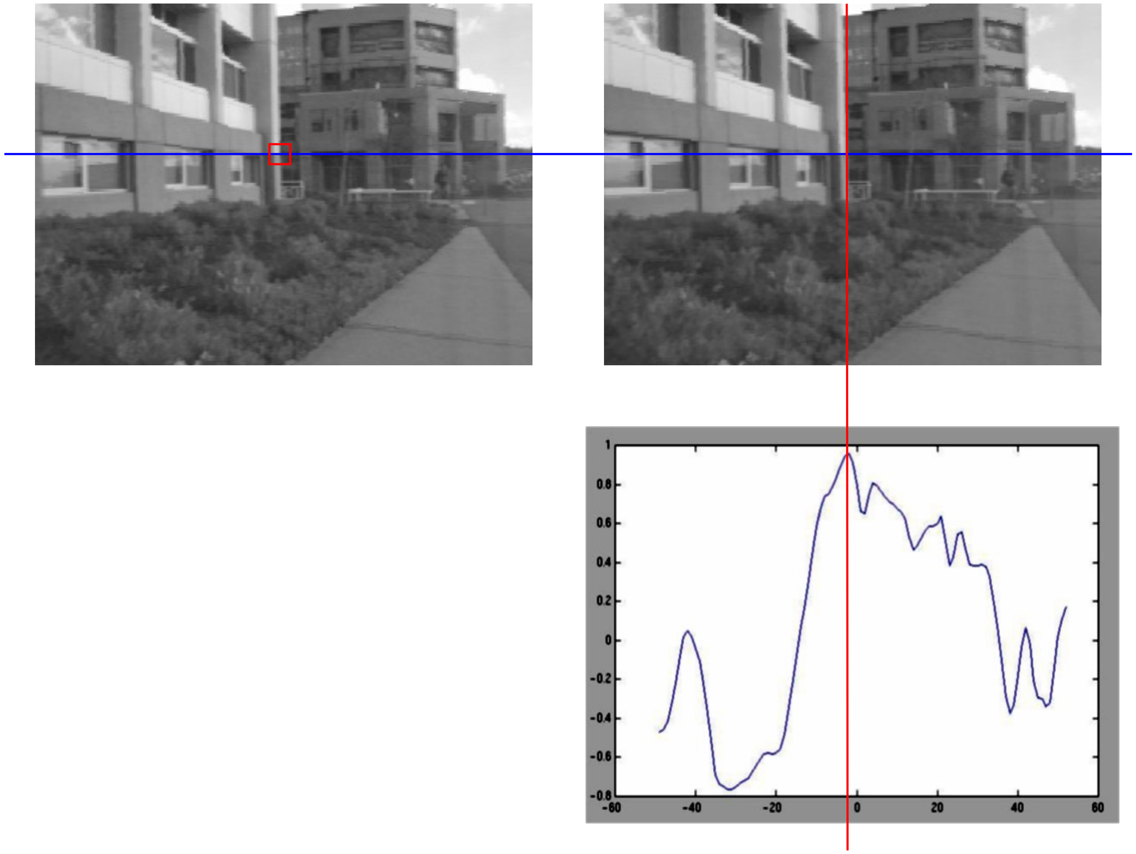
Stereo Block Matching



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

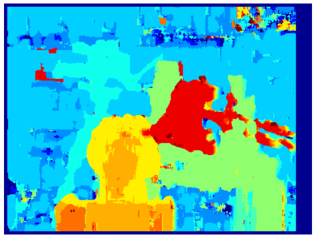


SSD

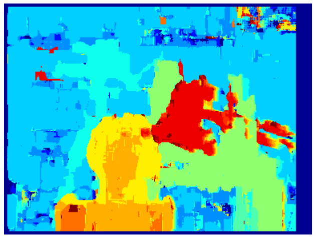


Normalized cross-correlation

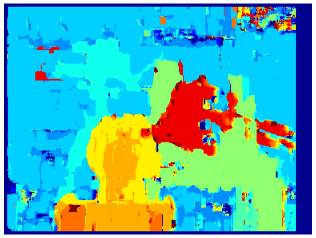
Similarity Measure	Formula
Sum of Absolute Differences (SAD)	$\sum_{(i,j)\in W} \lvert I_1(i,j) - I_2(x+i,y+j)\rvert$
Sum of Squared Differences (SSD)	$\sum_{(i,j)\in W} \left(I_1(i,j) - I_2(x+i,y+j)\right)^2$
Zero-mean SAD	$\sum_{(i,j)\in W} \lvert I_1(i,j) - \bar{I}_1(i,j) - I_2(x+i,y+j) + \bar{I}_2(x+i,y+j)\rvert$
Locally scaled SAD	$\sum_{(i,j)\in W} \lvert I_1(i,j) - \frac{\bar{I}_1(i,j)}{\bar{I}_2(x+i,y+j)} I_2(x+i,y+j)\rvert$
Normalized Cross Correlation (NCC)	$\frac{\sum_{(i,j)\in W} I_1(i,j).I_2(x+i,y+j)}{\sqrt[2]{\sum_{(i,j)\in W} I_1^2(i,j).\sum_{(i,j)\in W} I_2^2(x+i,y+j)}}$



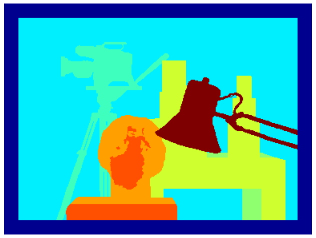
SAD



SSD



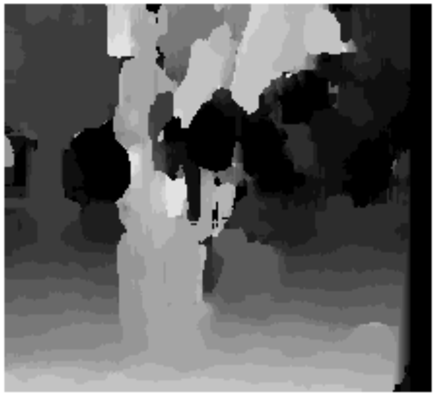
NCC



Ground truth

Effect of window size

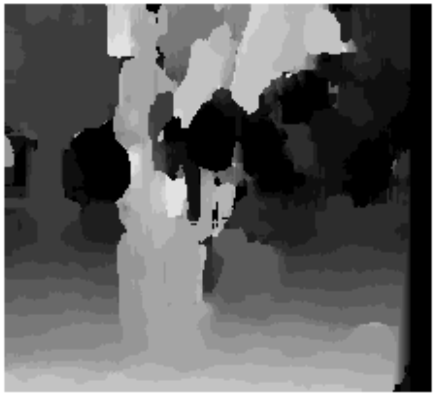
W = 3



W = 20

Effect of window size

W = 3

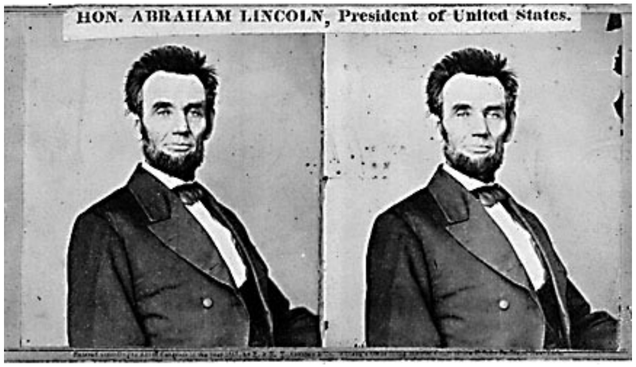


W = 20

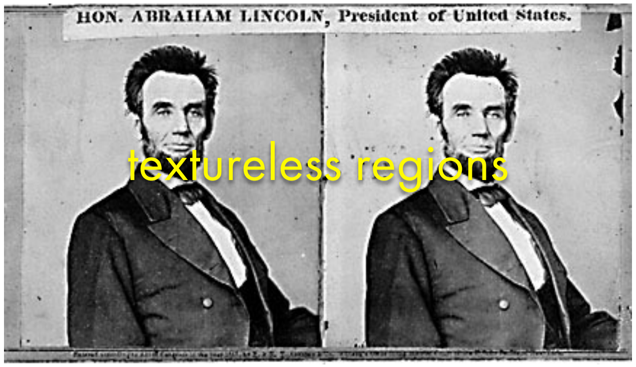
Smaller window
- \+ More detail
- \- More noise

Larger window
- \+ Smoother disparity maps
- \- Less detail
- \- Fails near boundaries

When will stereo block matching fail?

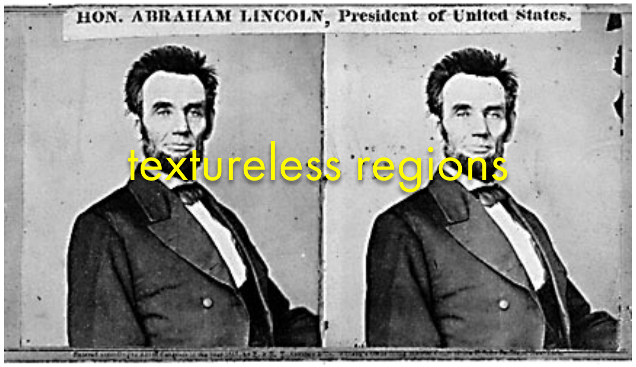


When will stereo block matching fail?

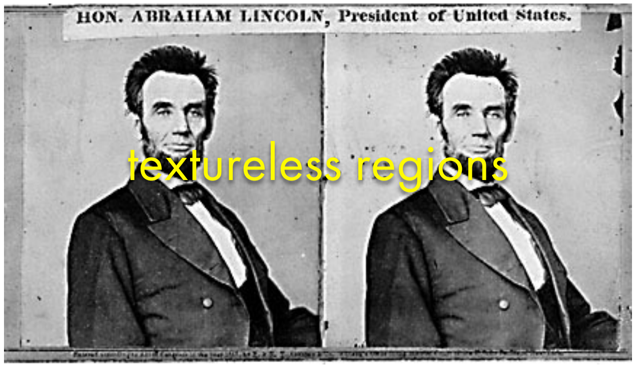


textureless regions

When will stereo block matching fail?



When will stereo block matching fail?



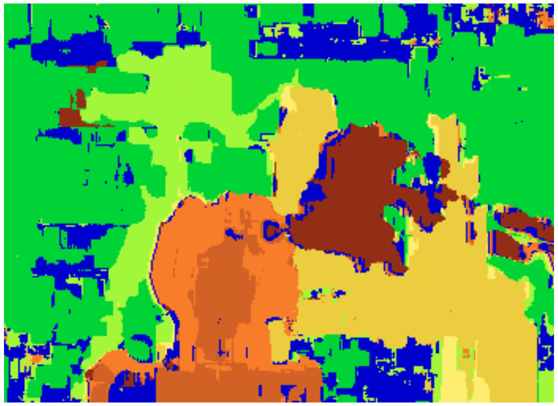
textureless regions

repeated patterns

specularities

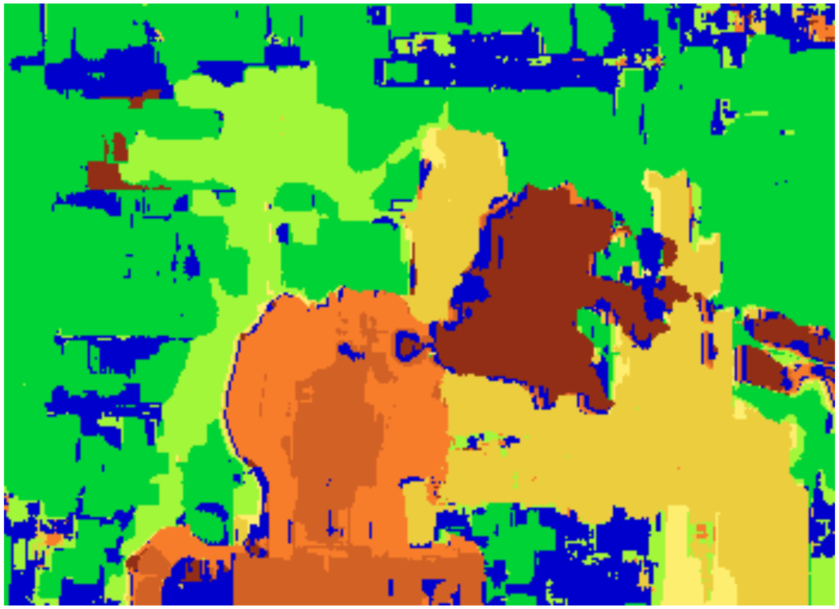
Improving stereo matching

Block matching

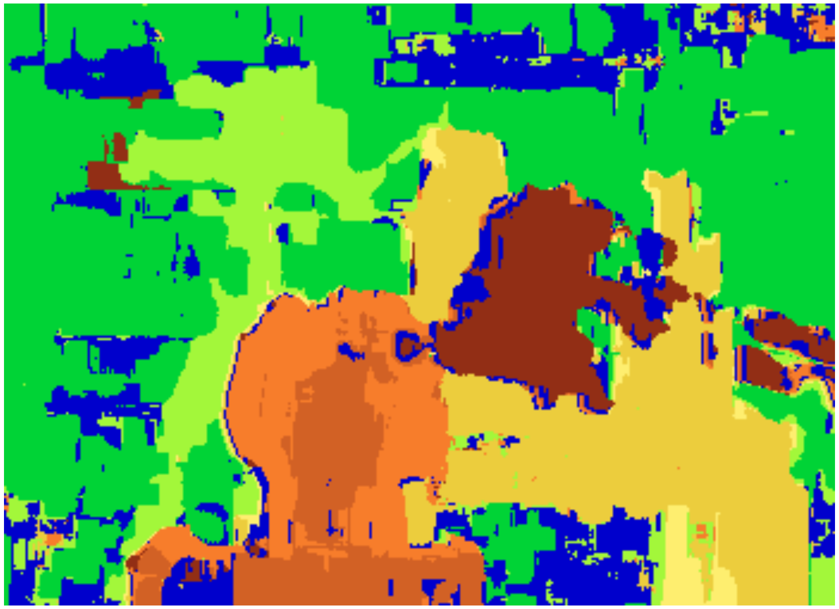


Ground truth

What are some problems with the result?



How can we improve depth estimation?



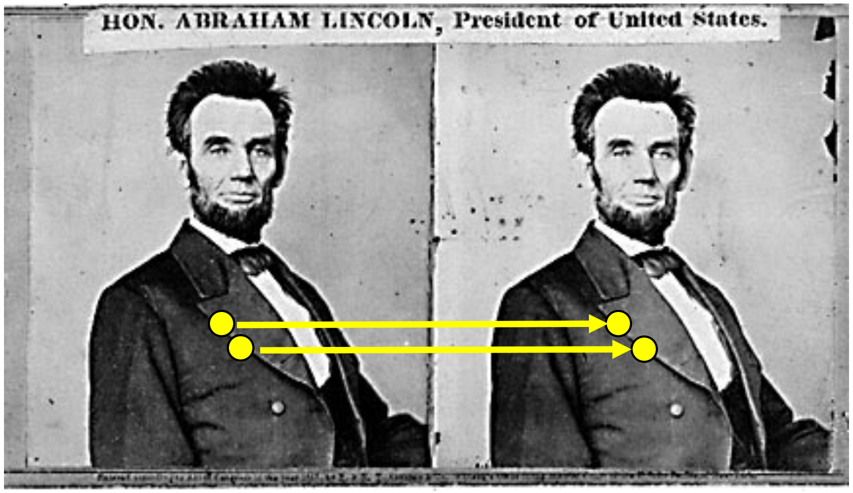
How can we improve depth estimation?

Too many discontinuities.
We expect disparity values to change slowly.

Let’s make an assumption:
depth should change smoothly

Stereo matching as ...

Energy Minimization



What defines a good stereo correspondence?

1. **Match quality**
   - Want each pixel to find a good match in the other image
2. **Smoothness**
   - If two pixels are adjacent, they should (usually) move about the same amount

energy function
(for one pixel)

$$E(d) = \underbrace{E_d(d)}_{\text{data term}} + \lambda \underbrace{E_s(d)}_{\text{smoothness term}}$$

Want each pixel to find a good match in the other image

(block matching result)

Adjacent pixels should (usually) move about the same amount

(smoothness function)

$$E(d) = \boxed{E_d(d)} + \lambda E_s(d)$$

$$E_d(d) = \sum_{(x,y)\in I} C(x, y, d(x, y))$$

data term

SSD distance between windows centered
at I(x, y) and J(x+ d(x,y), y)

$$E(d) = E_d(d) + \lambda \boxed{E_s(d)}$$

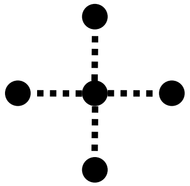
$$E_d(d) = \sum_{(x,y)\in I} C(x, y, d(x, y))$$

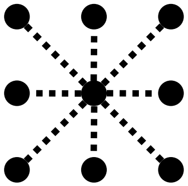
SSD distance between windows centered at I(x, y) and J(x+ d(x,y), y)

$$E_s(d) = \sum_{(p,q)\in \mathcal{E}} V(d_p, d_q)$$

smoothness term

$\mathcal{E}$: set of neighboring pixels


4-connected neighborhood

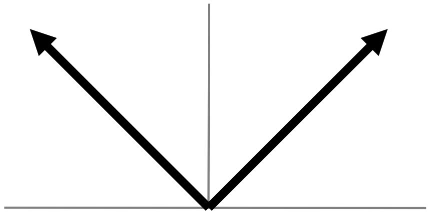

8-connected neighborhood

$$\boxed{E_s(d) = \sum_{(p,q)\in\mathcal{E}} V(d_p, d_q)}$$

smoothness term

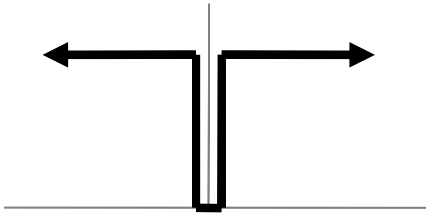
$$V(d_p, d_q) = |d_p - d_q|$$

L_1 distance



$$V(d_p, d_q) = \begin{cases} 0 & \text{if } d_p = d_q \\ 1 & \text{if } d_p \neq d_q \end{cases}$$

"Potts model"



One possible solution...

Dynamic Programming

$$E(d) = E_d(d) + \lambda E_s(d)$$

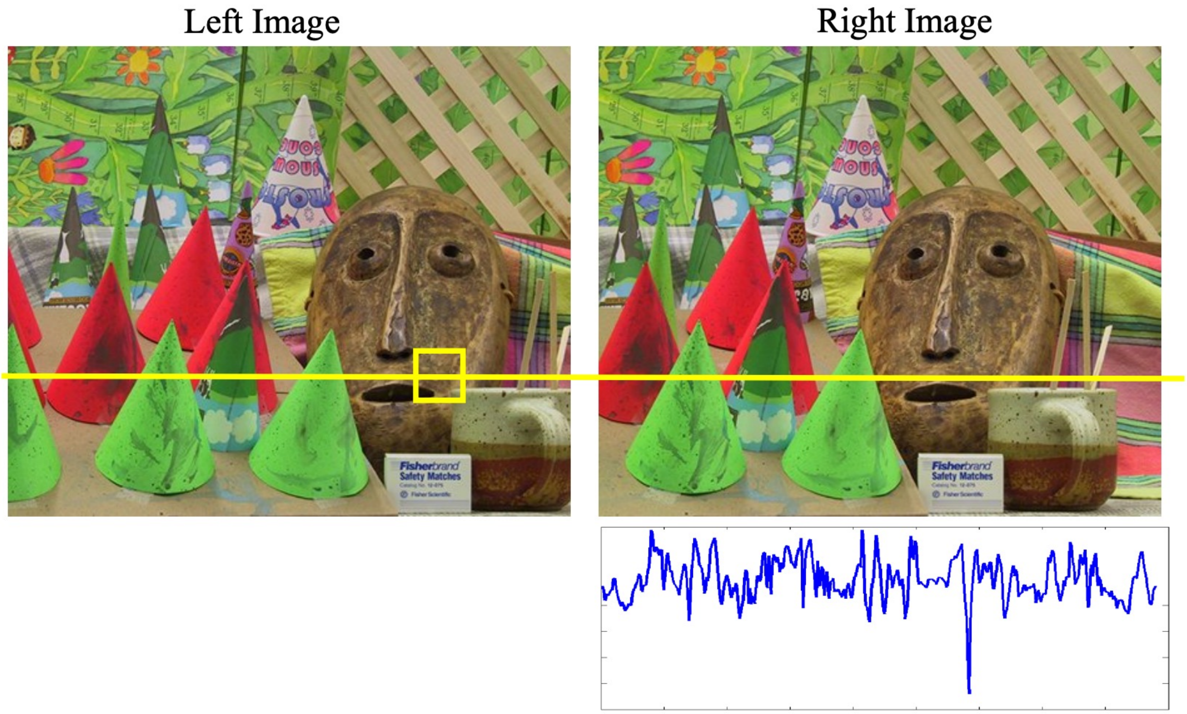
Can minimize this independently per scanline
using dynamic programming (DP)

$D(x, y, d)$: minimum cost of solution such that d(x,y) = d

$$D(x, y, d) = C(x, y, d) + \min_{d'} \left\{ D(x-1, y, d') + \lambda \left| d - d' \right| \right\}$$

One possible solution…

Dynamic Programming



Left Image

Right Image

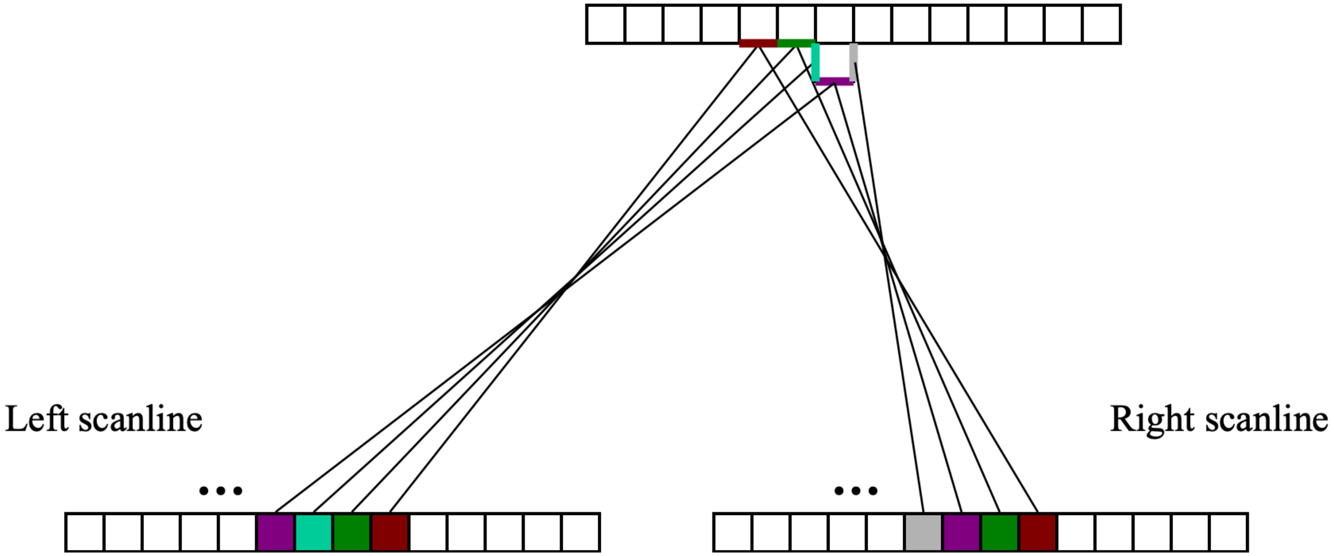
Dissimilarity Values

(1-NCC) or SSD

Slide: Robert Collins

One possible solution...

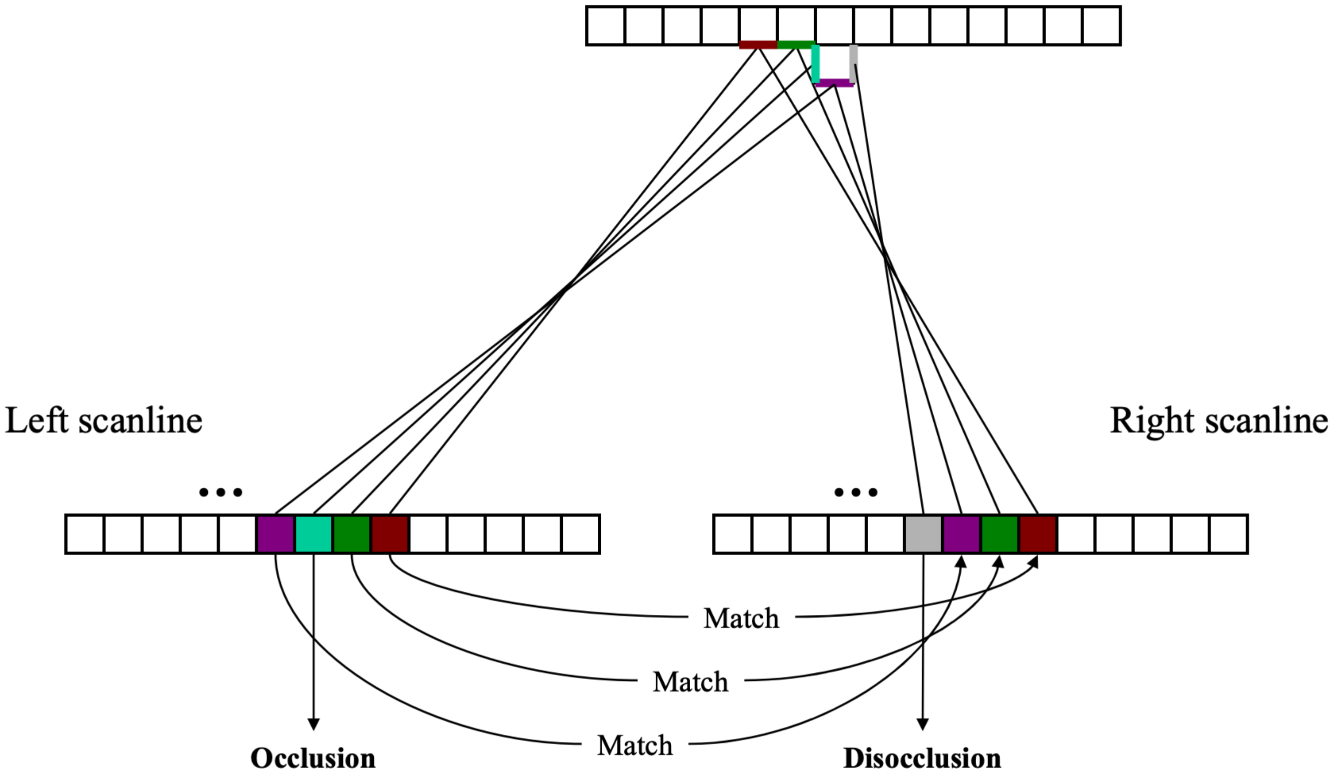
Dynamic Programming



Slide: Robert Collins

One possible solution...

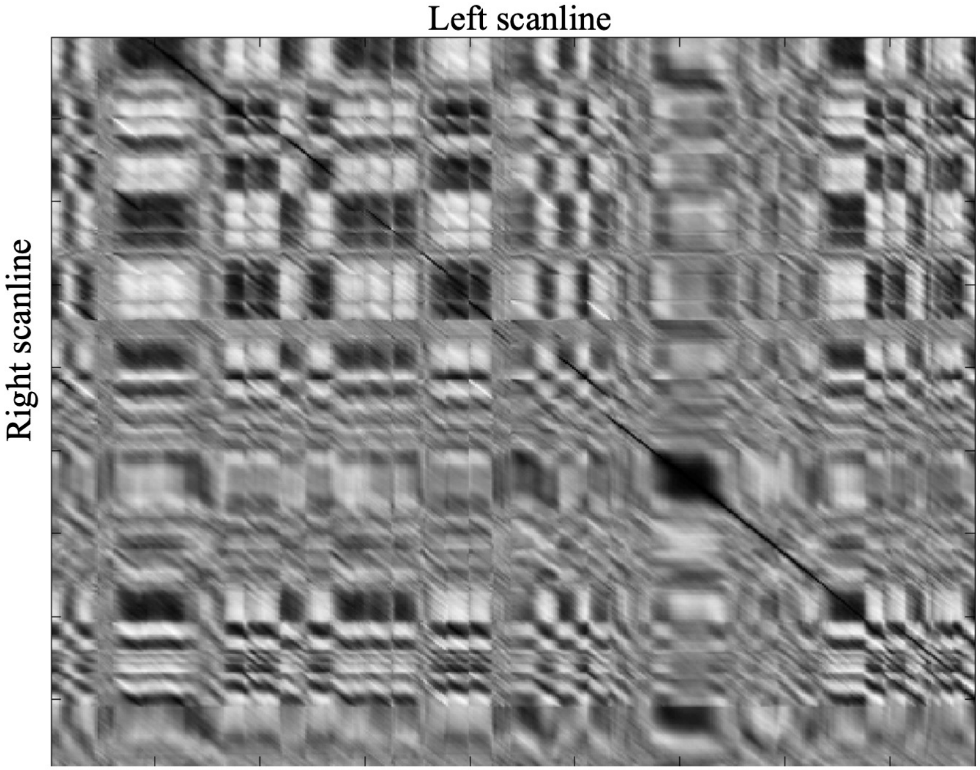
Dynamic Programming



Slide: Robert Collins

One possible solution...

Dynamic Programming

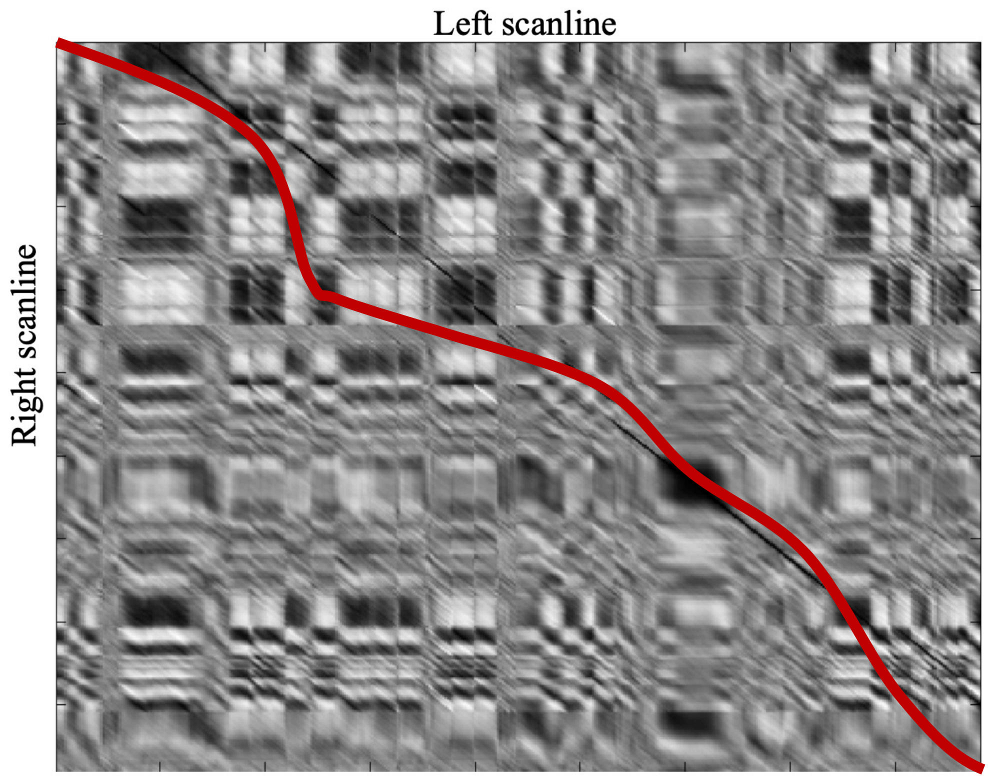


Disparity Space Image

Slide: Robert Collins

One possible solution...

Dynamic Programming

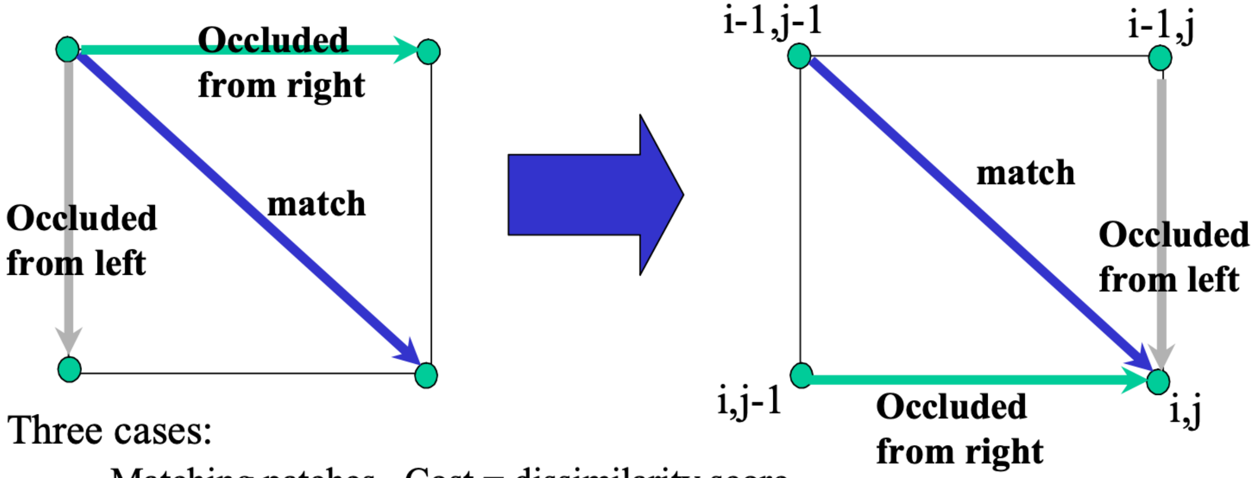


Disparity Space Image

Slide: Robert Collins

One possible solution...

Dynamic Programming



Three cases:

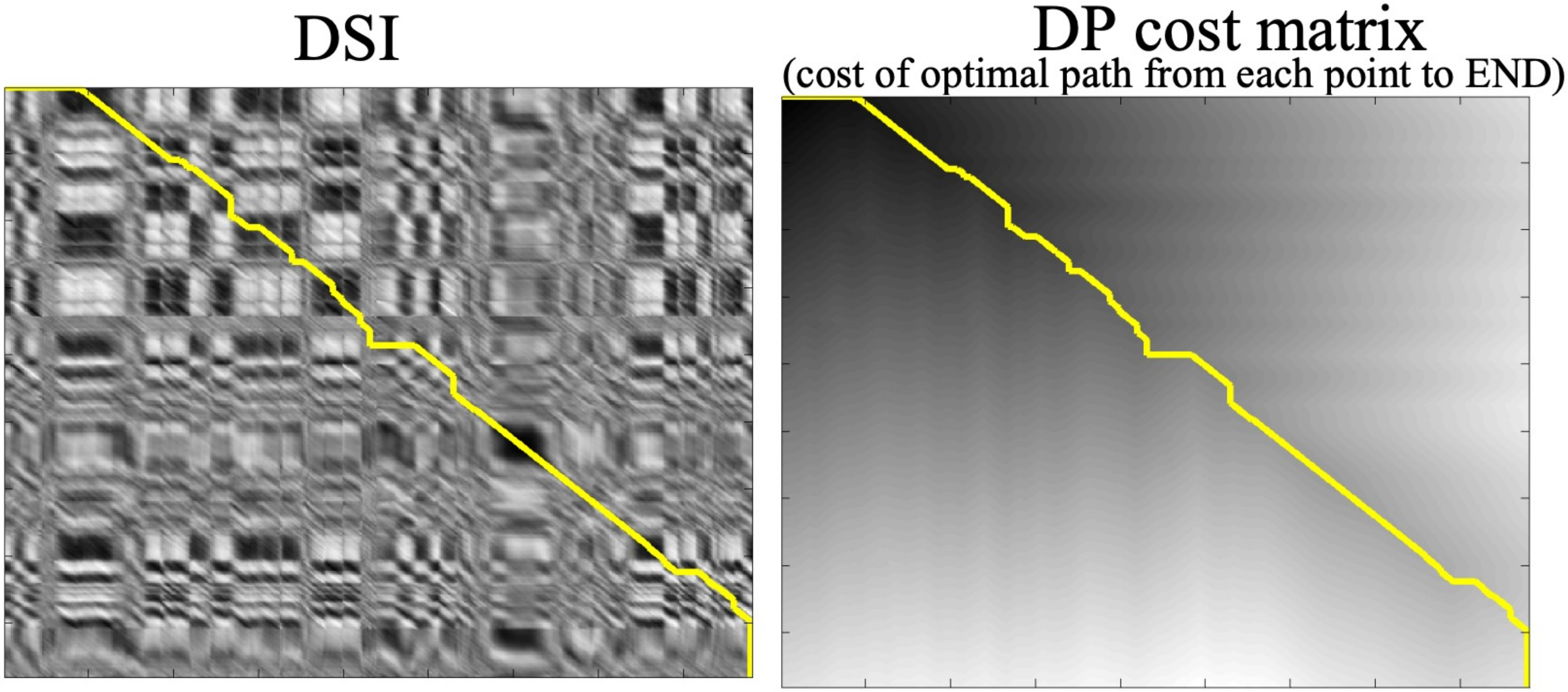
- Matching patches. Cost = dissimilarity score
- Occluded from right. Cost is some constant value.
- Occluded from left. Cost is some constant value.

$$
\boxed{\begin{aligned} C(i,j) = \min([\ & C(i-1,j-1) + \text{dissimilarity}(i,j) \\ & C(i-1,j) + \text{occlusionConstant}, \\ & C(i,j-1) + \text{occlusionConstant}]); \end{aligned}}
$$

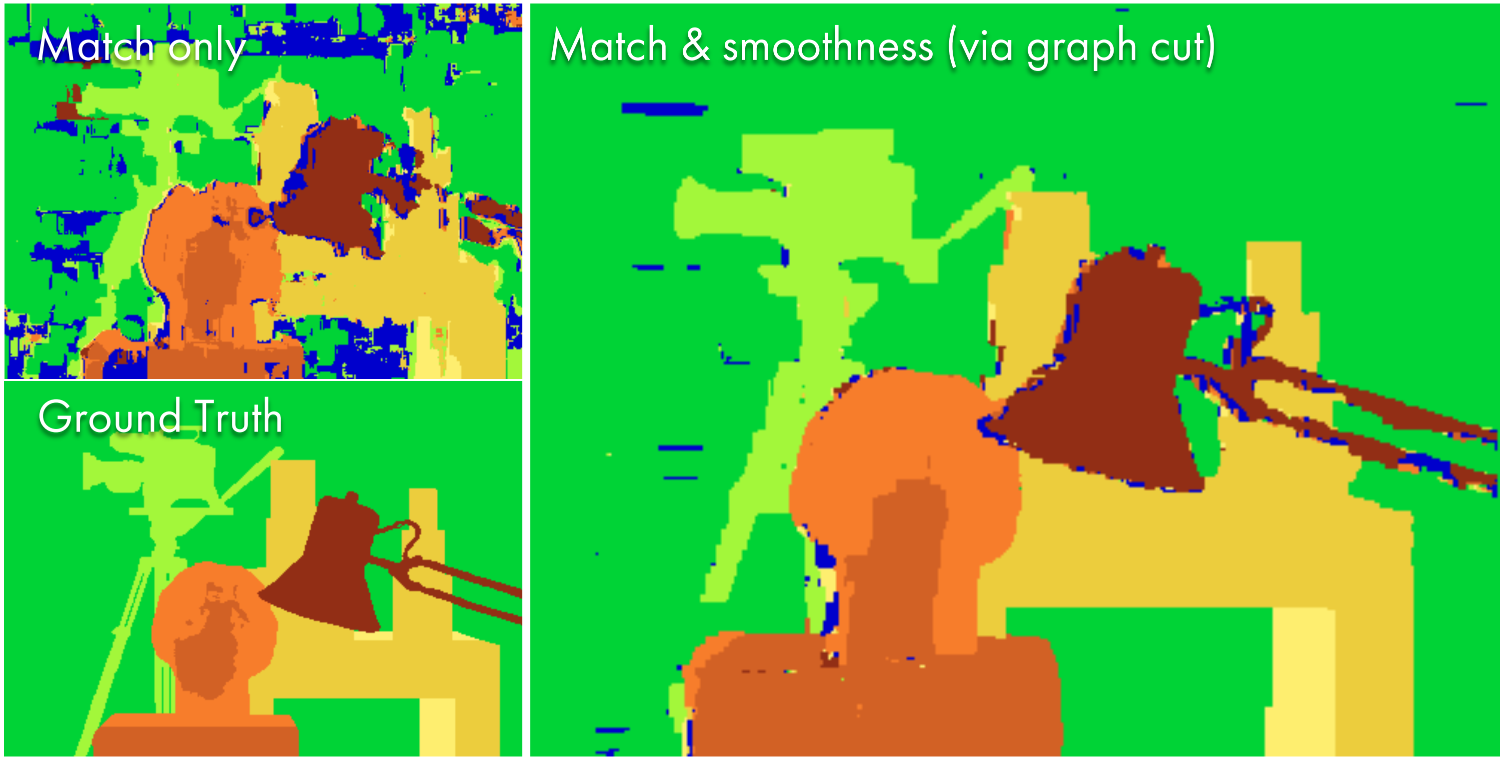
Slide: Robert Collins

One possible solution...

Dynamic Programming

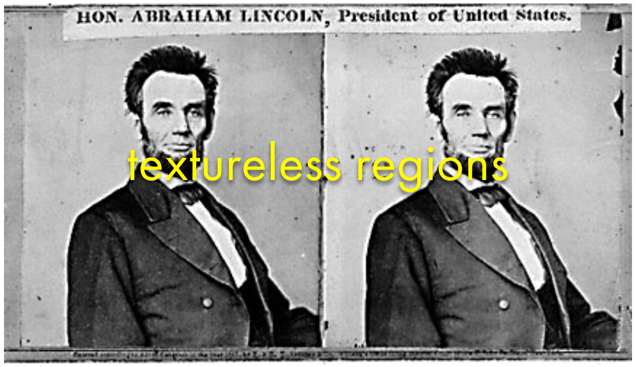


Slide: Robert Collins



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

All of these cases remain difficult, what can we do?

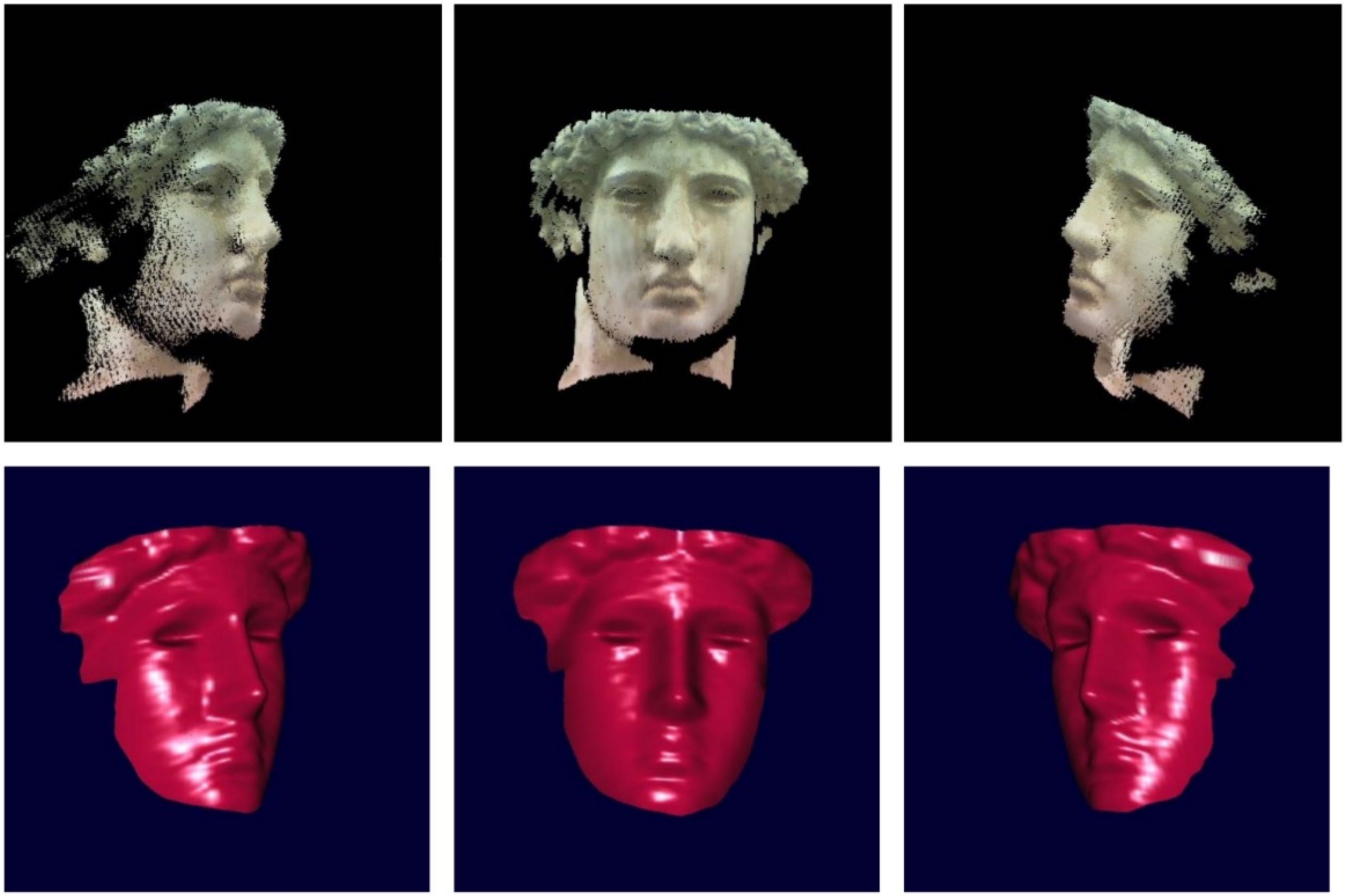


Structured light

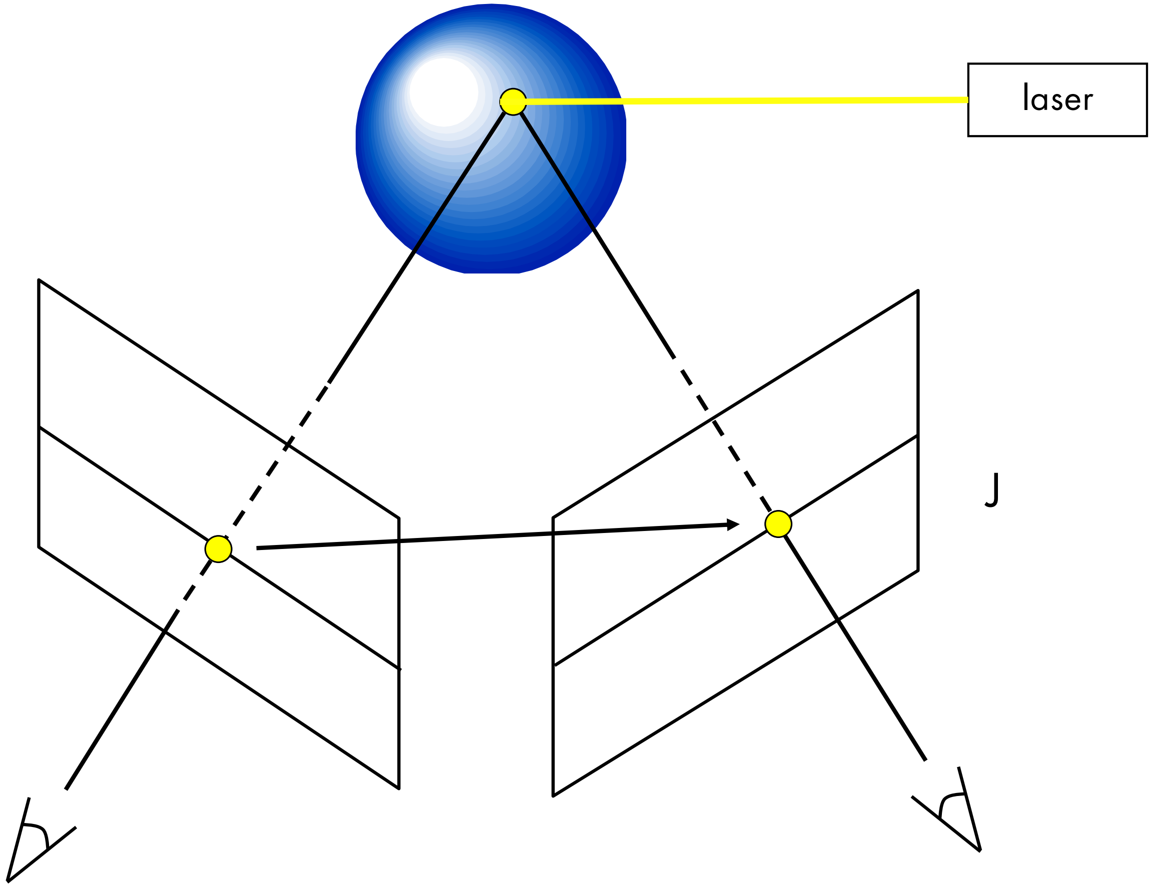
Use controlled ("structured") light to make correspondences easier

Disparity between laser points on the same scanline in the images determines the 3-D coordinates of the laser point on object

Use controlled ("structured") light to make correspondences easier

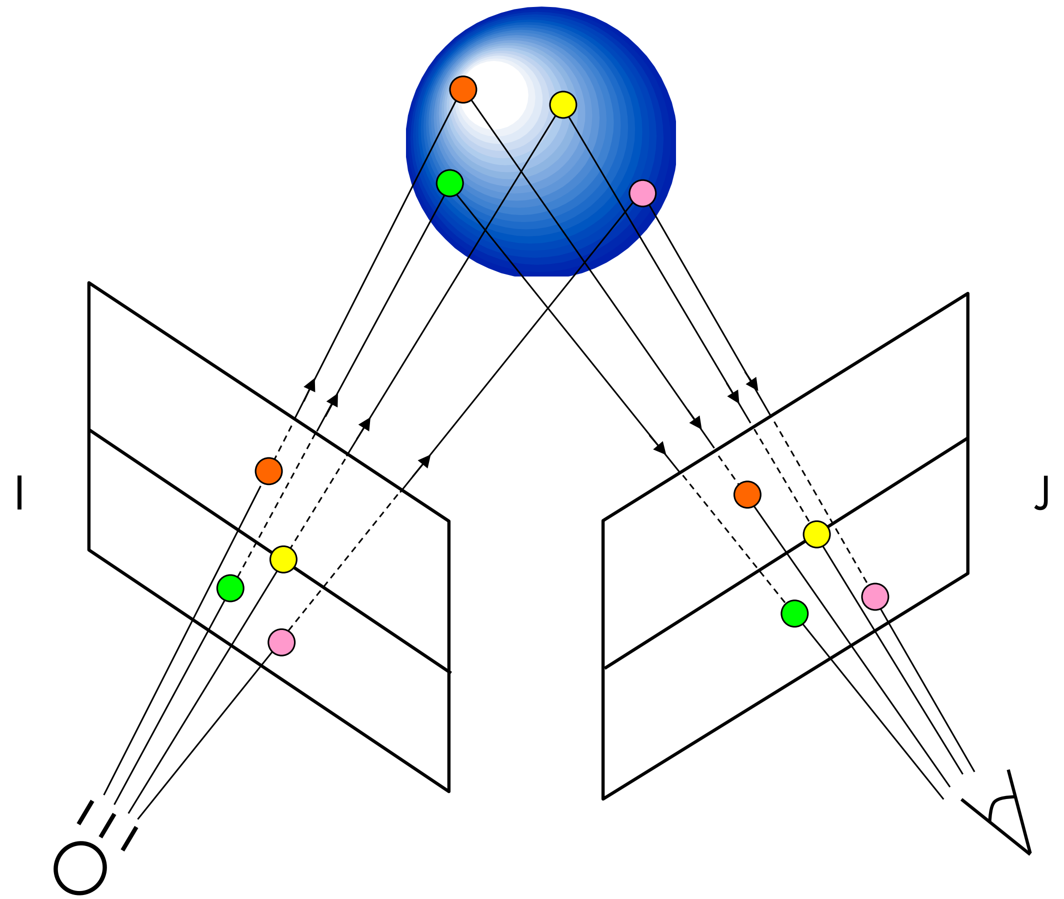


Structured light and two cameras

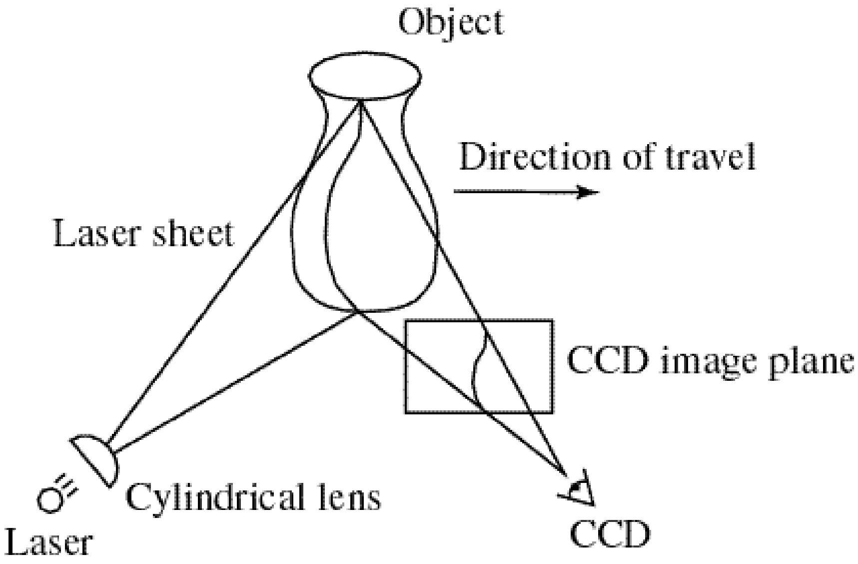


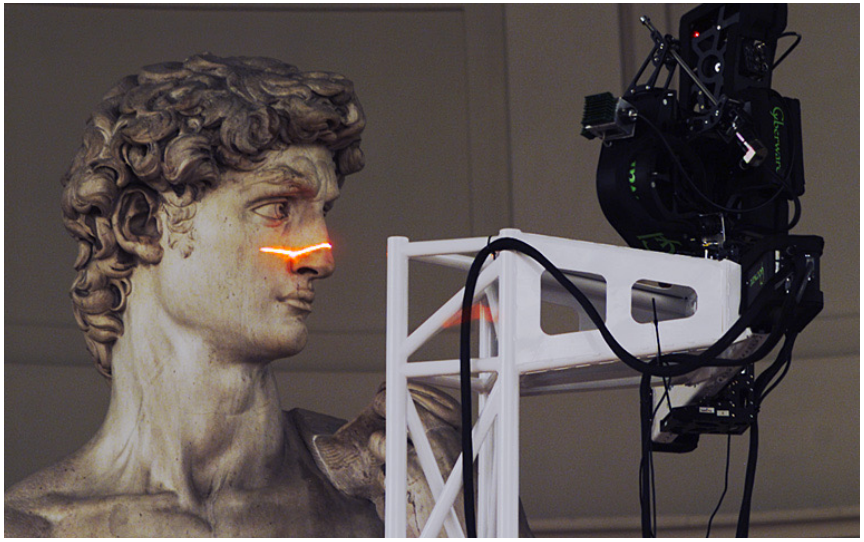
Structured light and one camera

Projector acts like
“reverse” camera



Example: Laser scanner

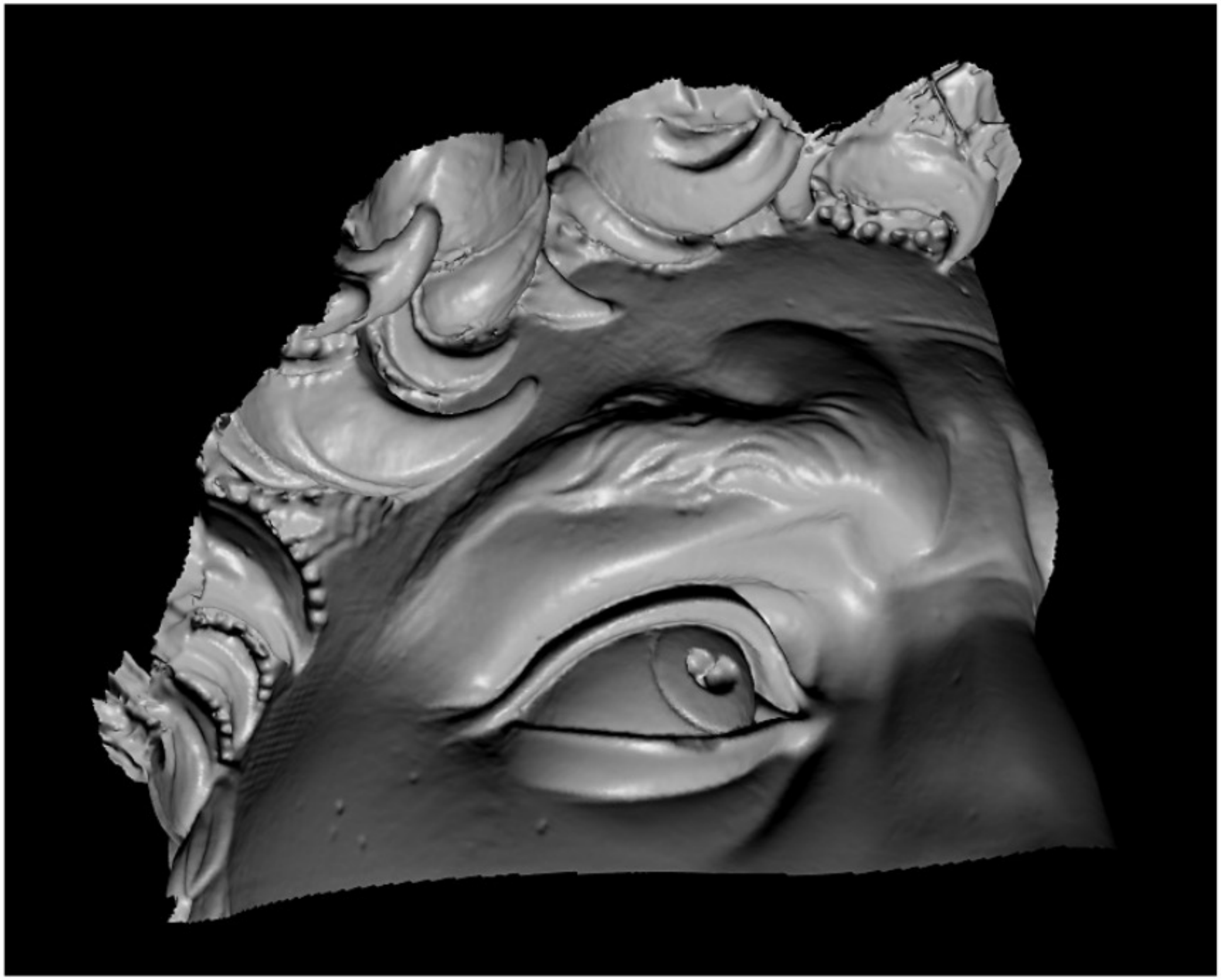




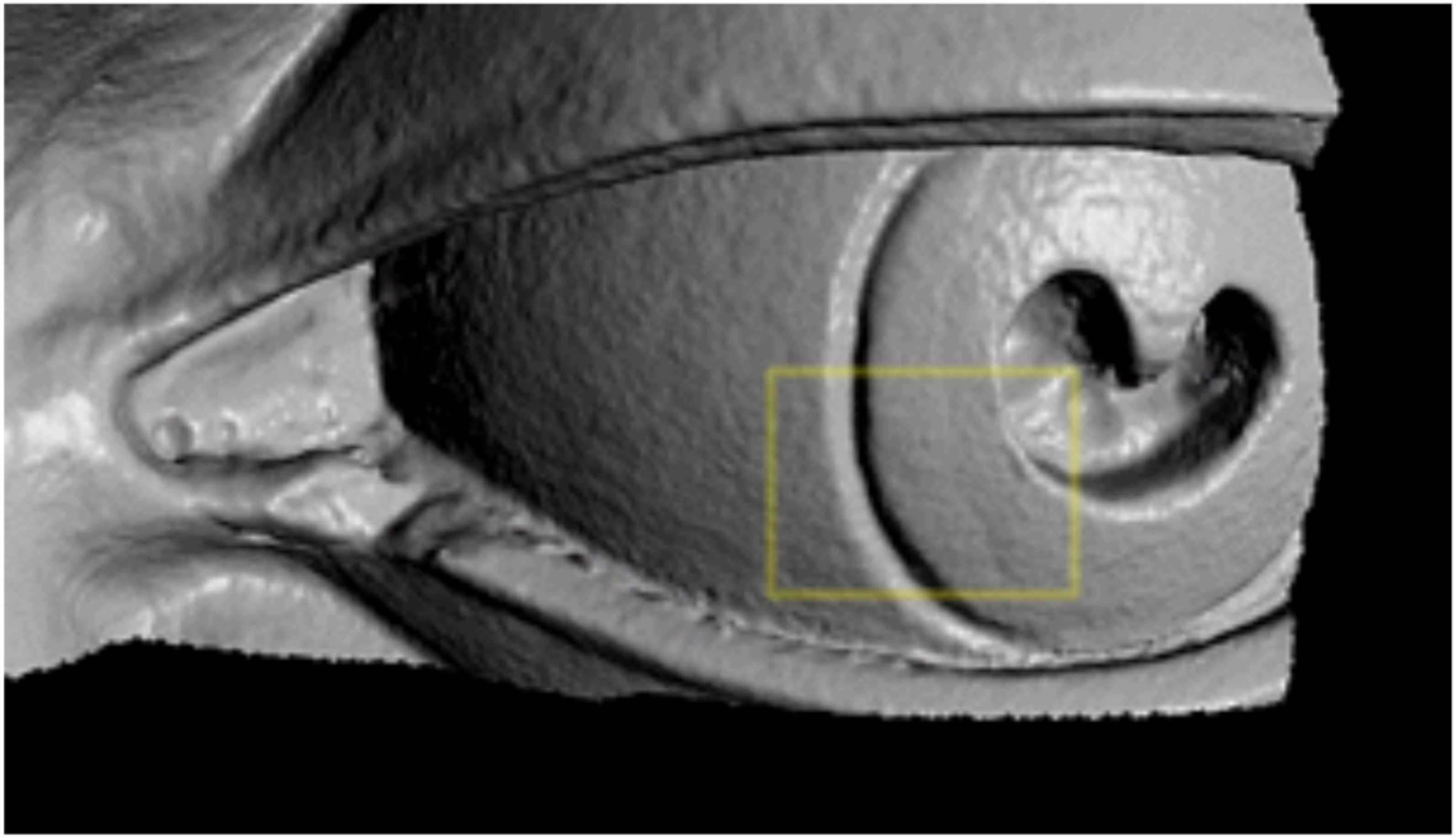
Digital Michelangelo Project
http://graphics.stanford.edu/projects/mich/

The Digital Michelangelo Project, Levoy et al.

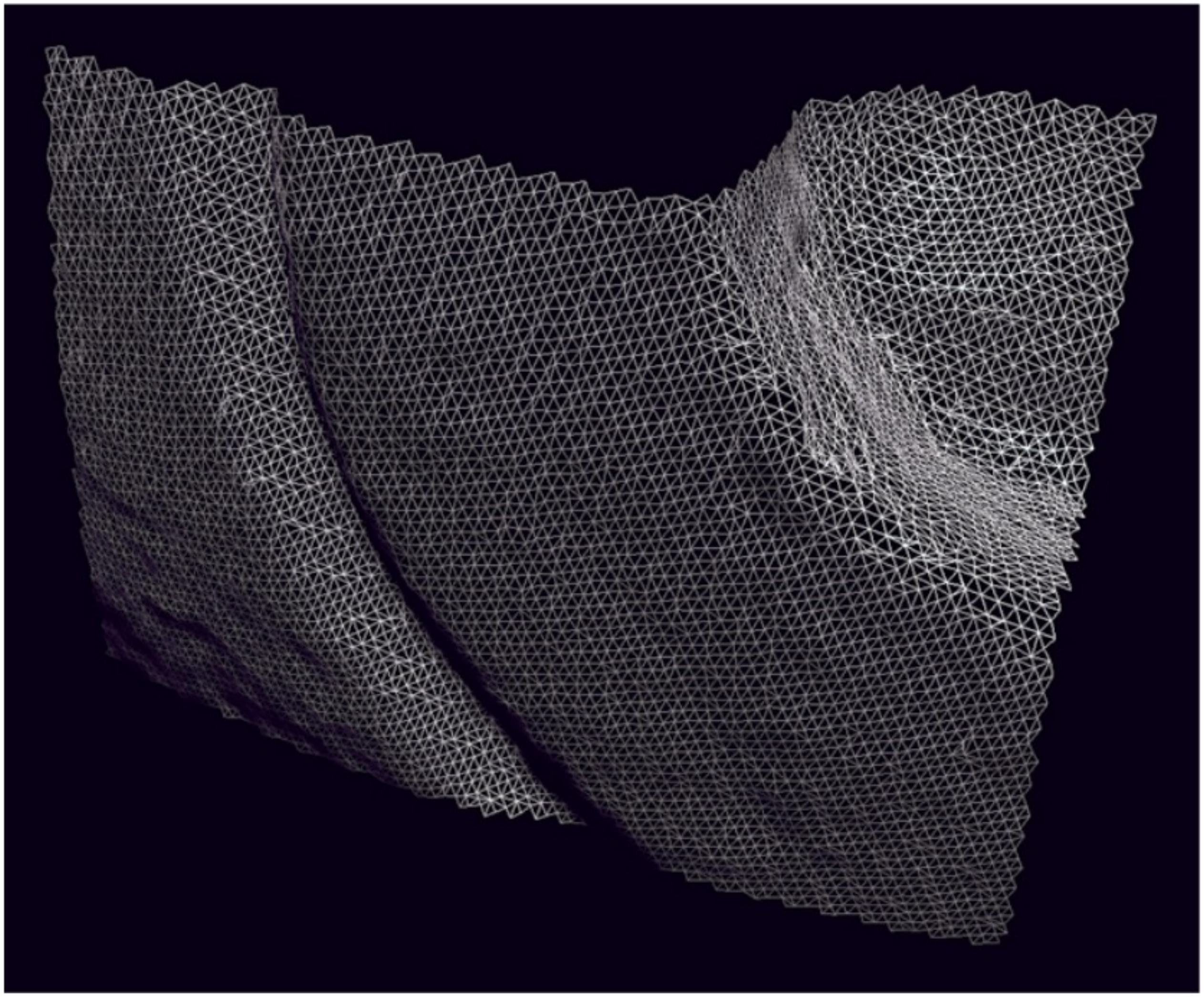
The Digital Michelangelo Project, Levoy et al.



The Digital Michelangelo Project, Levoy et al.



The Digital Michelangelo Project, Levoy et al.



The Digital Michelangelo Project, Levoy et al.

Summary – Stuff You Need To Know

Camera models

- intrinsic and extrinsic parameters
- camera matrix
- camera-to-world transformation/camera-to-camera transformation

Essential and Fundamental Matrices

- How **E** is derived and relates to **F**
- How to solve for **E** or **F** using the 8-point algorithm
- How to rectify both images to be parallel
- How are the epipoles computed?

Stereo Matching

- How does block matching work?
- How are 3D points computed once you have a rectified stereo camera?
- Given the intrinsics, disparity, and baseline, how is a 3D point computed?

References

Basic reading:

- Szeliski textbook, Section 8.1 (not 8.1.1-8.1.3), Chapter 11, Section 12.2.
- Hartley and Zisserman, Section 11.12.