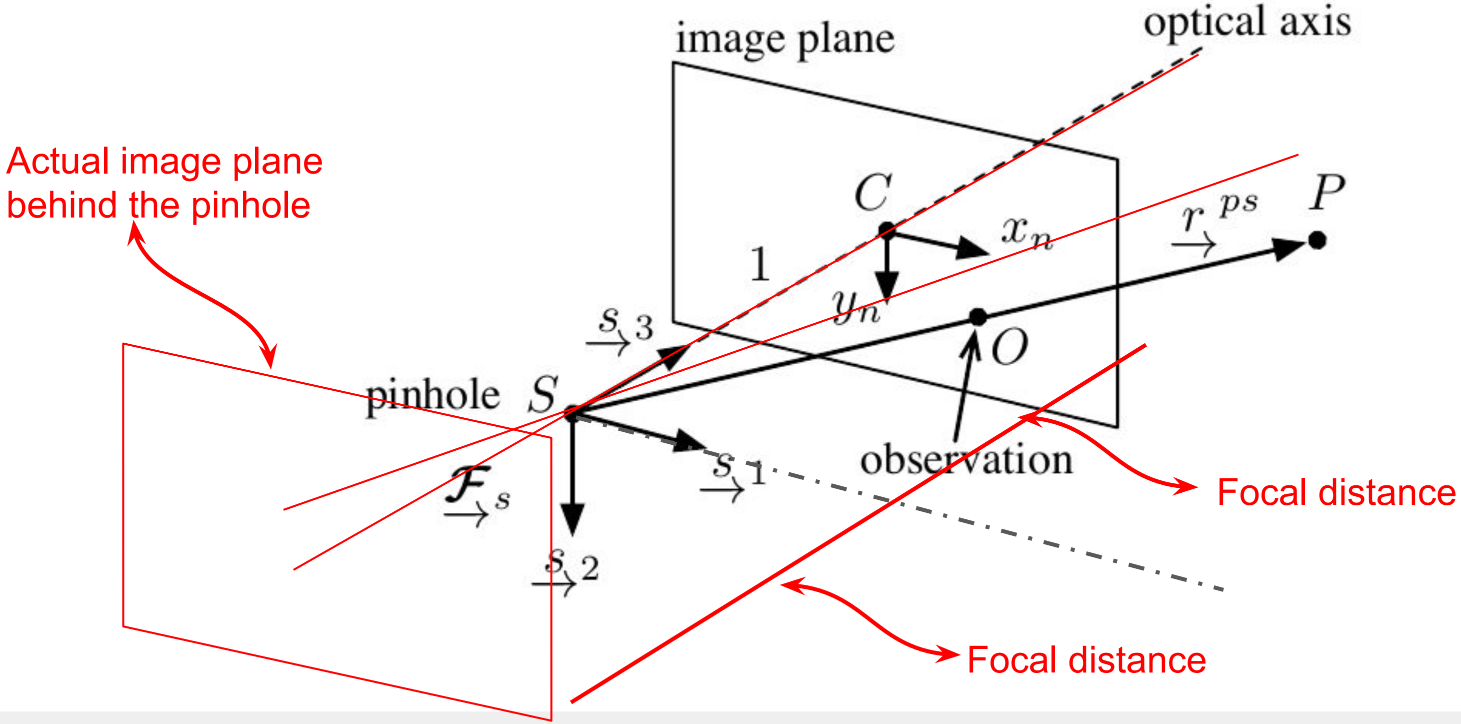
Introduction to Computer Vision for Robotics

AE640A Autonomous Navigation

11th March, 2019

… From the last class



… From the last class

K 0 R T

$$\underbrace{\begin{bmatrix} x' \\ y' \\ s \end{bmatrix}}_{p'} = \underbrace{\begin{bmatrix} f & 0 & c_x & 0 \\ 0 & f & c_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}}_{M = K\,[I\ 0]} \underbrace{\begin{bmatrix} R_{00} & R_{01} & R_{02} & T_x \\ R_{10} & R_{11} & R_{12} & T_y \\ R_{20} & R_{21} & R_{22} & T_z \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{H} \underbrace{\begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}}_{P}$$

2d homogeneous pixel (camera) coordinates

Focal length and focal center of camera

"Intrinsics"

Rigidly move every object in the world to simulate the camera's true position

"Extrinsics"

3d homogenous (world) coordinates

Lecture Outline

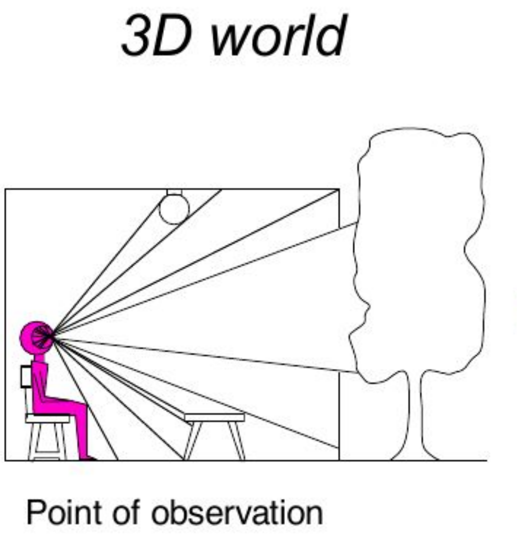
- Stereo Vision
  - Introduction to Stereo Vision
  - Epipolar Geometry
  - The correspondence problem
- Stereo Matching
  - Various methods for Stereo Matching
  - Stereo Block Matching
  - A look at SGBM

Stereo Vision

Introduction to stereo vision

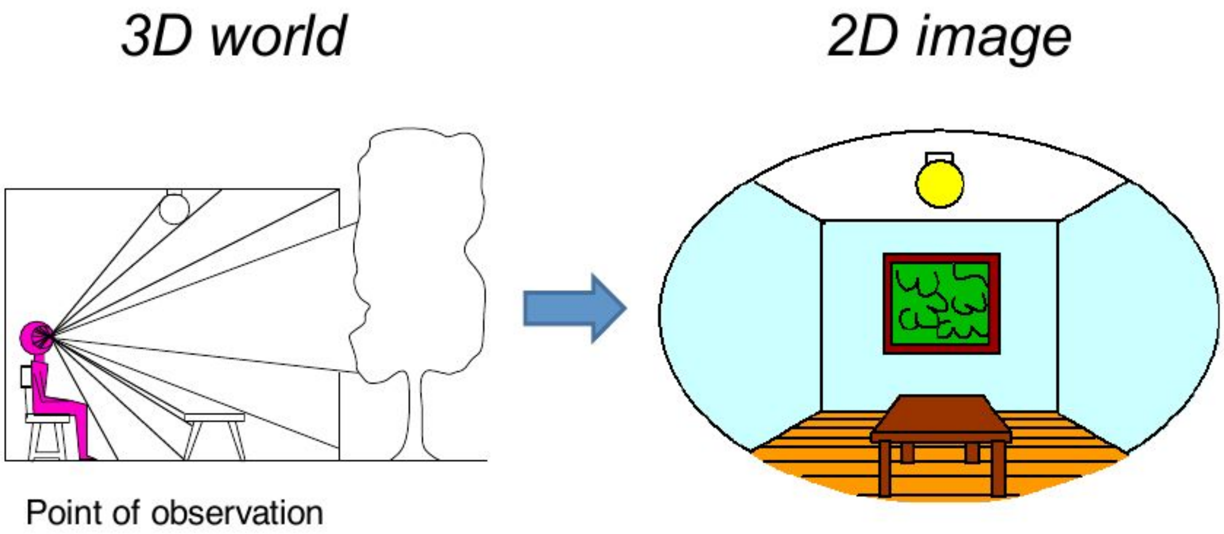
Credits: Kenji Hata, Silvio Savarese

Introduction to stereo vision



Credits: Fei Fei Li

Introduction to stereo vision



Credits: Fei Fei Li

Introduction to stereo vision

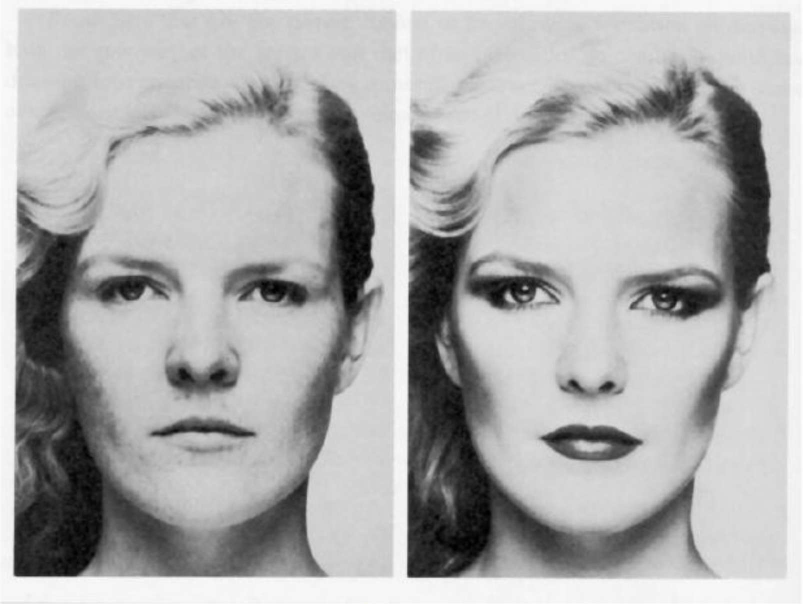
How do humans figure out 3D in 2D images?

Credits: Fei Fei Li

Introduction to stereo vision

How do humans figure out 3D in 2D images?

1. Shading



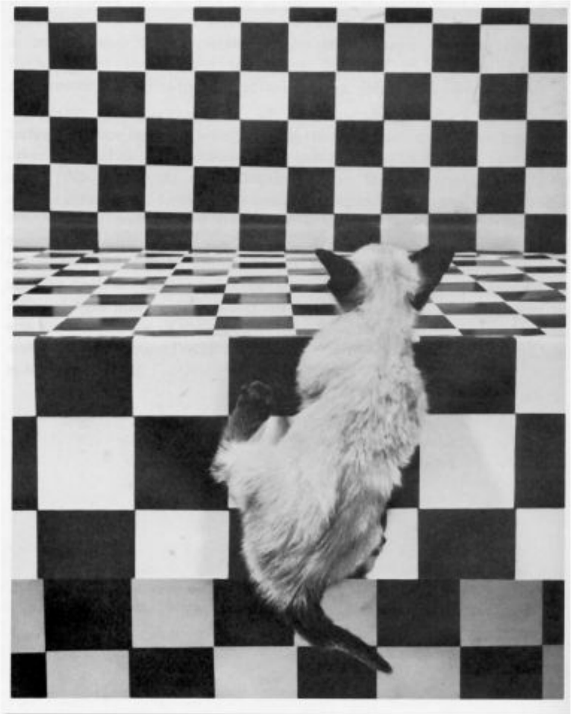
Merle Norman Cosmetics, Los Angeles

Credits: Fei Fei Li

Introduction to stereo vision

How do humans figure out 3D in 2D images?

1. Shading
2. Texture



The Visual Cliff, by William Vandivert, 1960

Credits: Fei Fei Li

Introduction to stereo vision

How do humans figure out 3D in 2D images?

1. Shading
2. Texture
3. Focus

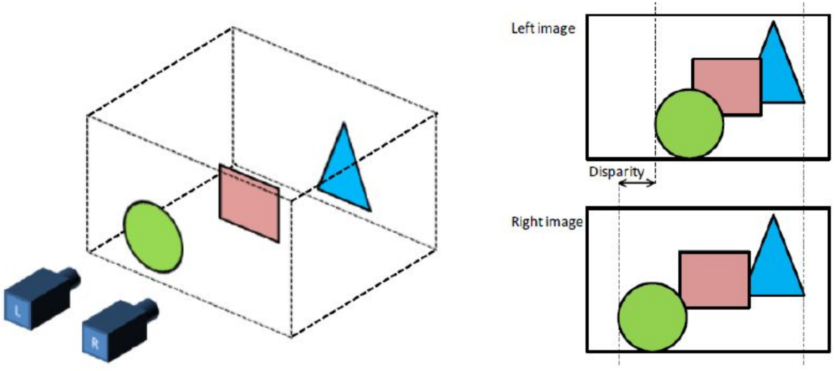
From *The Art of Photography*, Canon

Credits: Fei Fei Li

Introduction to stereo vision

The **stereo problem:**

- Nature Inspired approach to vision, i.e, 3D with two sensors.
- How to figure out the shape, more specifically the depth, of objects from a set of two or more images?



Credits: Gaurav Pandey, Ford

Introduction to stereo vision

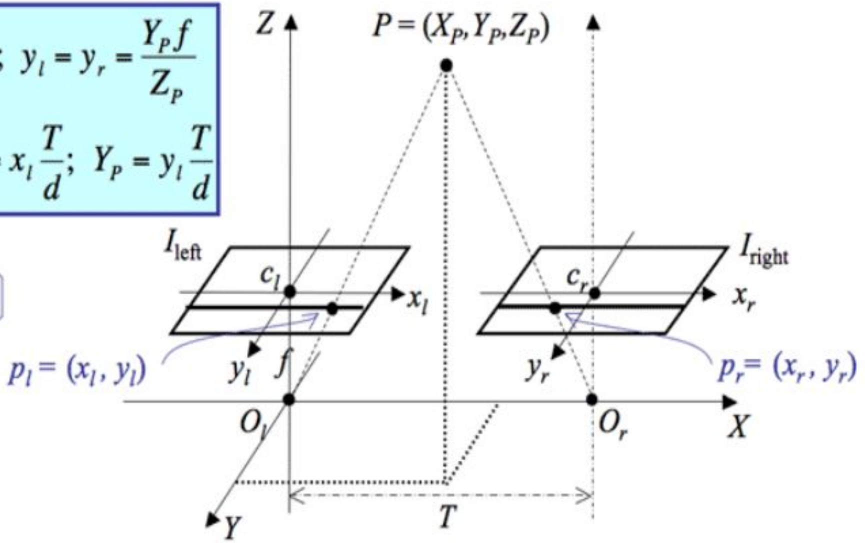
So, How do we go we go from Stereo Images to Depth Information ?

Introduction to stereo vision

$$x_l = \frac{X_P f}{Z_P};\; x_r = \frac{(X_P - T)f}{Z_P};\; y_l = y_r = \frac{Y_P f}{Z_P}$$

$$Z_P = f\frac{T}{x_l - x_r} \equiv f\frac{T}{d};\; X_P = x_l\frac{T}{d};\; Y_P = y_l\frac{T}{d}$$

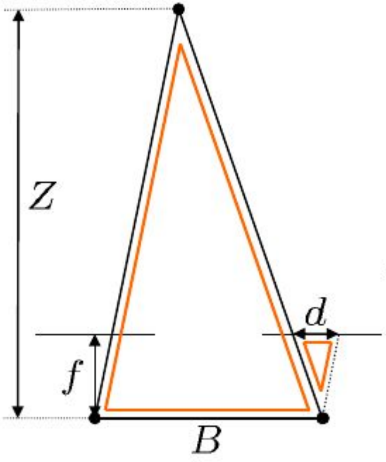
disparity $d = x_l - x_r$



Need to find corresponding point (xr,yr) for each (xl,yl) ⇒ Correspondence problem

Credits: Gaurav Pandey, Ford

Introduction to stereo vision

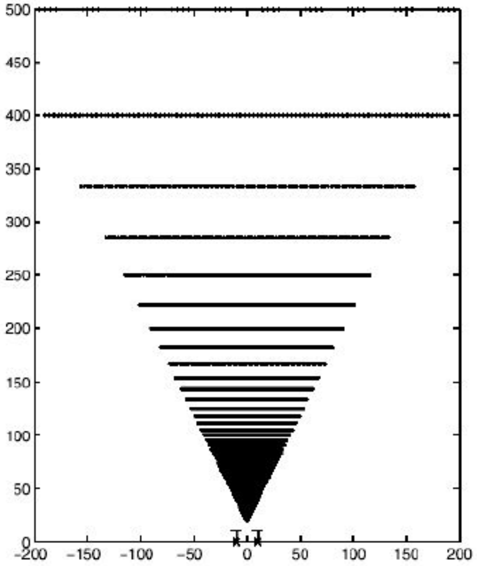


$$\frac{B}{Z} = \frac{d}{f}$$

$$d = -\frac{Bf}{Z}$$

$$\frac{dd}{dZ} = \frac{Bf}{Z^2}$$

$$\Delta Z = \frac{Z^2}{Bf}\Delta d$$

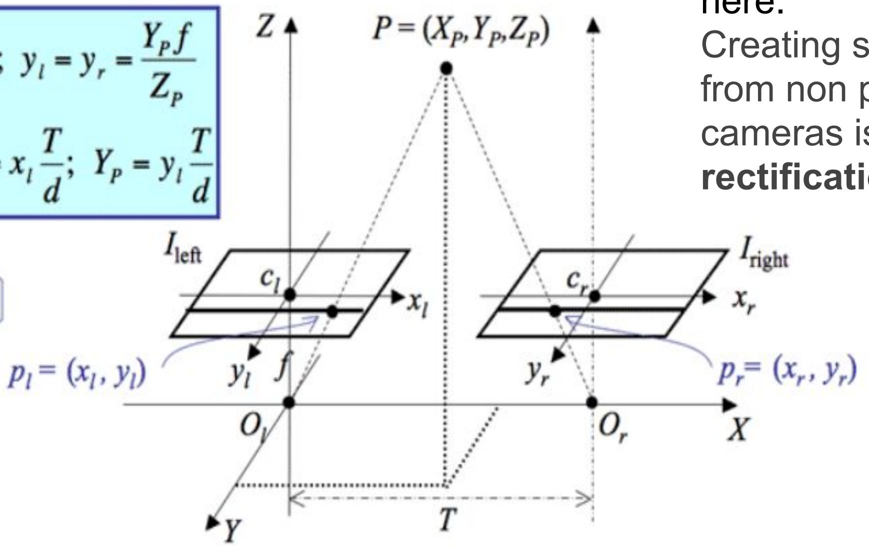


Introduction to stereo vision

$$x_l = \frac{X_P f}{Z_P};\ x_r = \frac{(X_P - T) f}{Z_P};\ y_l = y_r = \frac{Y_P f}{Z_P}$$

$$Z_P = f\frac{T}{x_l - x_r} = f\frac{T}{d};\ X_P = x_l\frac{T}{d};\ Y_P = y_l\frac{T}{d}$$

disparity $d = x_l - x_r$

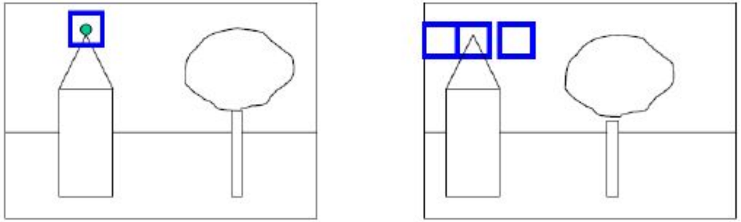


Need to find corresponding point (xr,yr) for each (xl,yl) $\Rightarrow$ Correspondence problem

Note: We have the image planes **parallel** here.
Creating such images from non parallel cameras is called **rectification**.

Introduction to stereo vision

1. Cross correlation or SSD using small windows.



dense

2. Symbolic feature matching, usually using segments/corners.

sparse

3. Use the newer interest operators, e.g., SIFT.

sparse

Credits: Gaurav Pandey, Ford

Introduction to stereo vision

Need to find corresponding point (xr,yr) for each (xl,yl) ⇒ Correspondence problem

Credits: Fei Fei Li

Introduction to stereo vision

Given this point how do you find the corresponding point on the other image?

Need to find corresponding point (xr,yr) for each (xl,yl) ⇒ Correspondence problem

Credits: Fei Fei Li

Introduction to stereo vision

Given this point how do you find the corresponding point on the other image?

Search the whole image?

Need to find corresponding point (xr,yr) for each (xl,yl) $\Rightarrow$ Correspondence problem

Credits: Fei Fei Li

Introduction to stereo vision

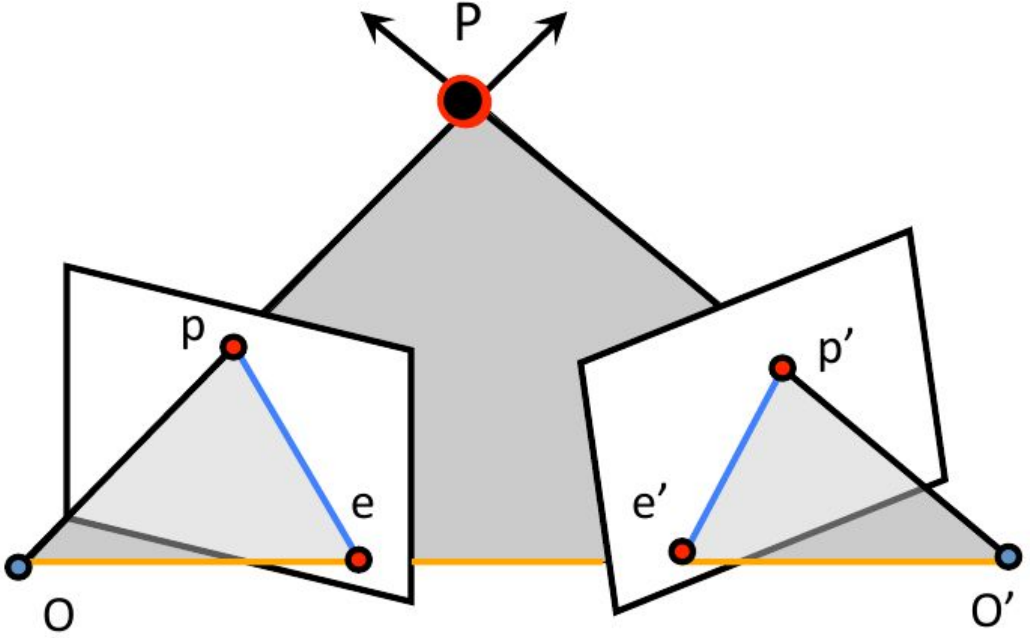
Given this point how do you find the corresponding point on the other image?

Search the whole image?

Difficult to solve accurately, very expensive without special methods

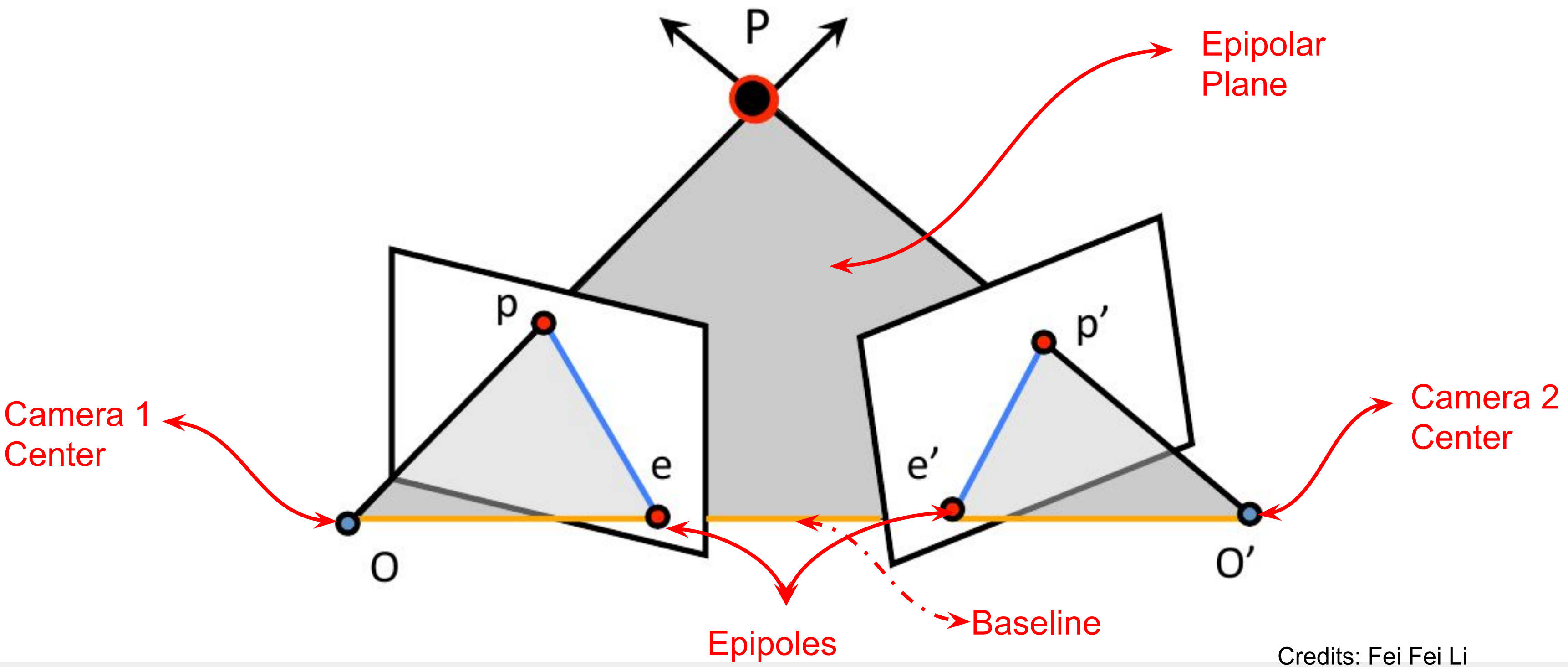
Need to find corresponding point (xr,yr) for each (xl,yl) $\Rightarrow$ $\boxed{\text{Correspondence problem}}$

Epipolar Geometry



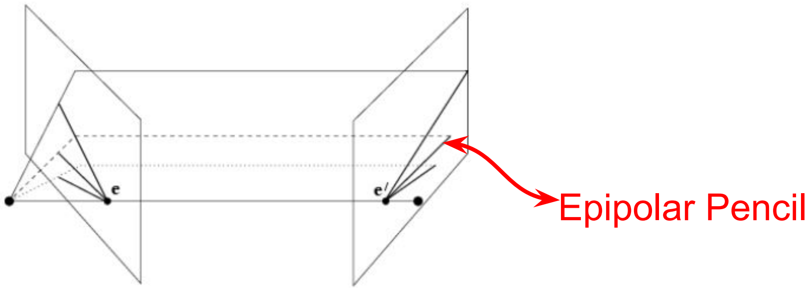
Credits: Fei Fei Li

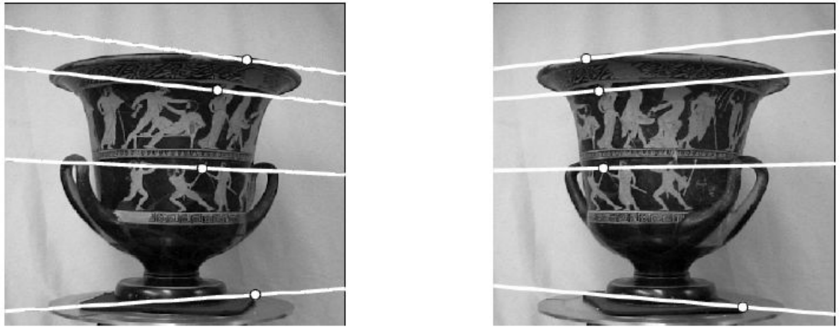
Epipolar Geometry



Credits: Fei Fei Li

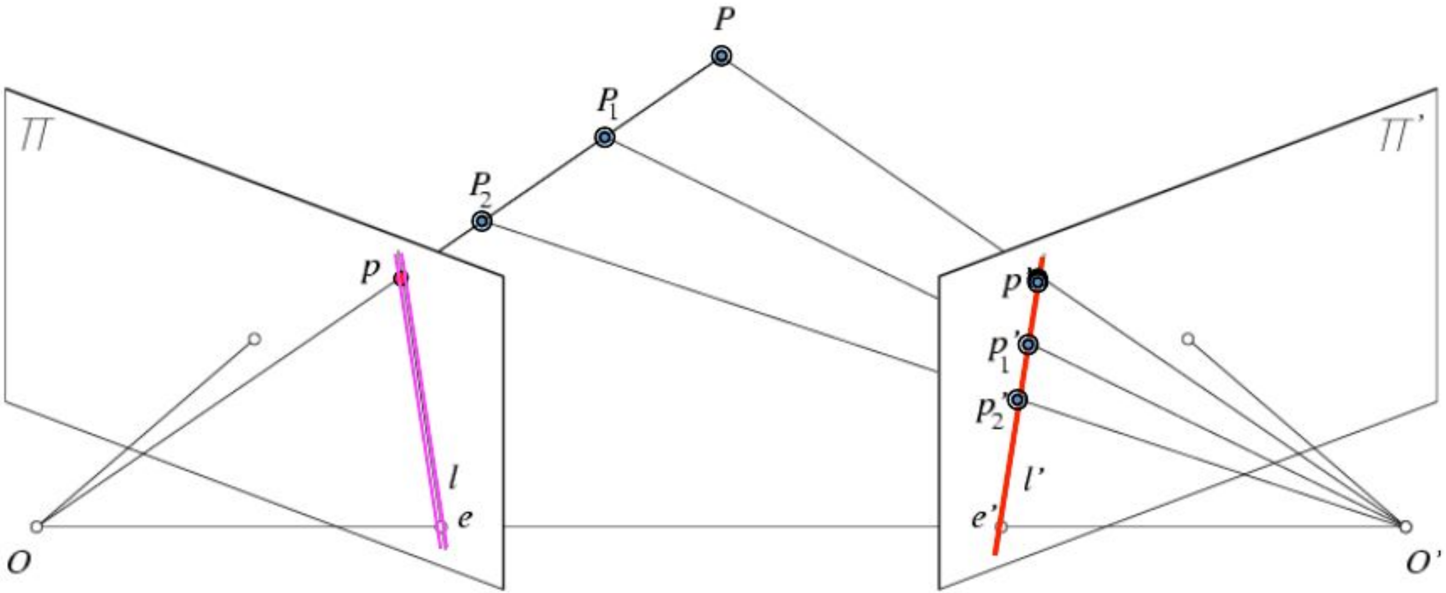
Epipolar Geometry





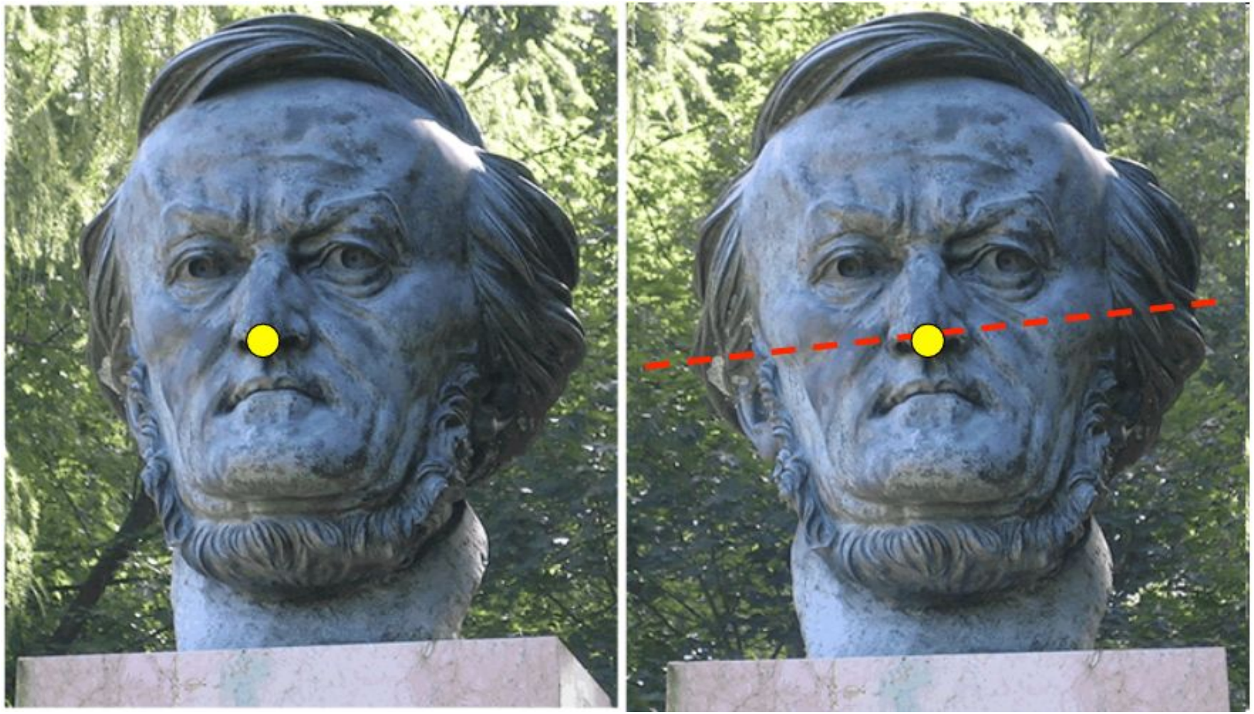
Credits: Richard Hartley, Andrew Zisserman

Epipolar Geometry



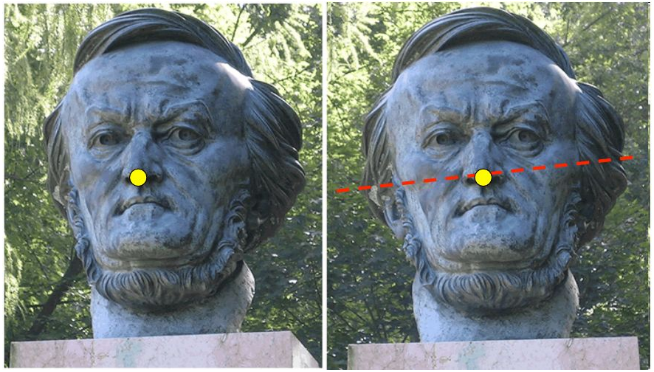
Credits: Fei Fei Li

Epipolar Geometry



Credits: Fei Fei Li

Epipolar Geometry



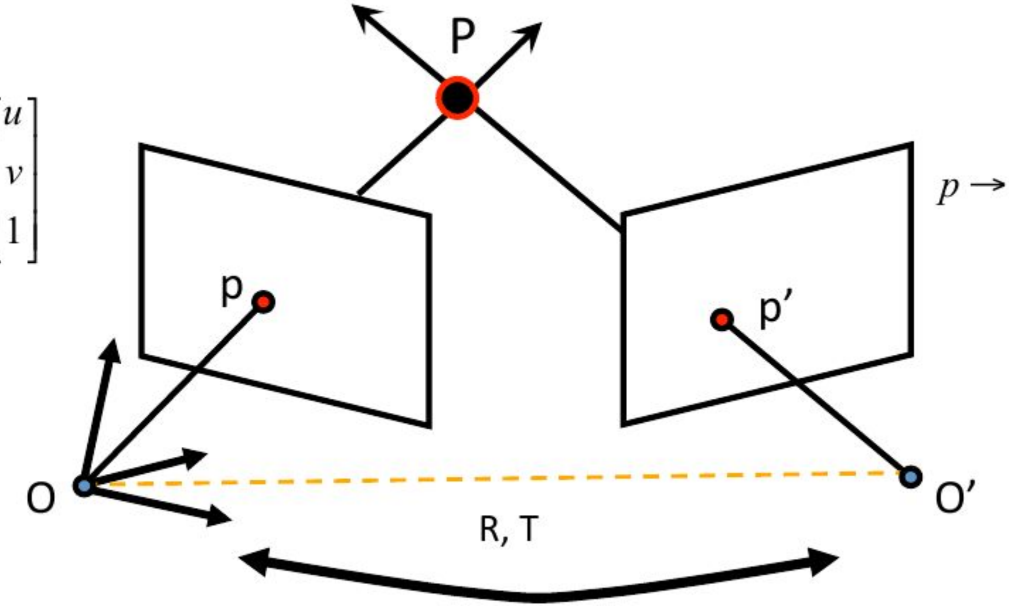
Search along this line for the closest point. Computationally way more efficient.

Easier to solve.
Can use simple SSD or similar methods.

Credits: Fei Fei Li

Epipolar Geometry

$$p \rightarrow M\,P = \begin{bmatrix} u \\ v \\ 1 \end{bmatrix}$$

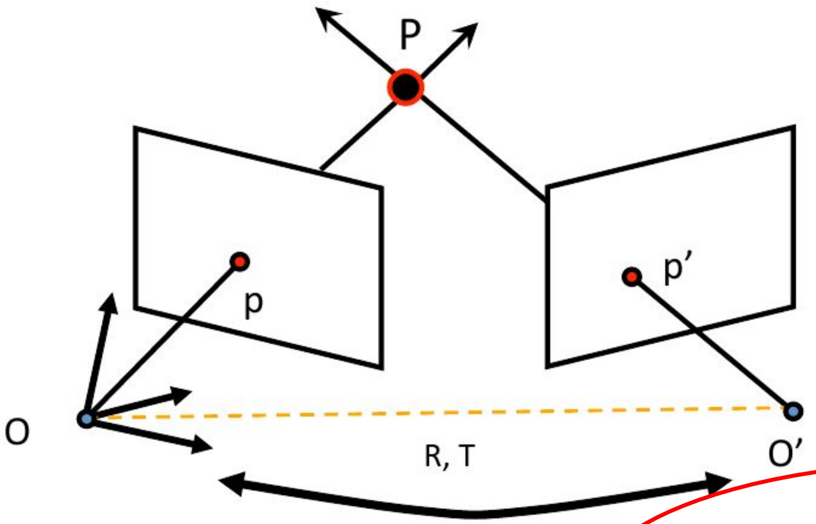


$$p \rightarrow M'\,P = \begin{bmatrix} u' \\ v' \\ 1 \end{bmatrix}$$

$$M = K\begin{bmatrix} I & 0 \end{bmatrix} \qquad M' = K'\begin{bmatrix} R & T \end{bmatrix}$$

Credits: Fei Fei Li

Epipolar Geometry



Assumed to be canonical camera

$$M = K\begin{bmatrix} I & 0 \end{bmatrix}$$

K$_1$ and K$_2$ are known (calibrated cameras)

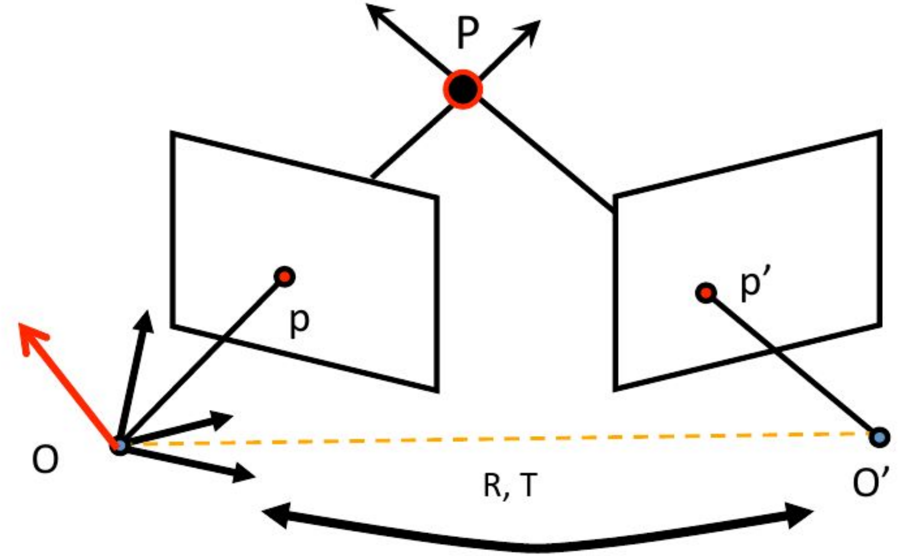
$$M' = K'\begin{bmatrix} R & T \end{bmatrix}$$

$$M = \begin{bmatrix} I & 0 \end{bmatrix}$$

$$M' = \begin{bmatrix} R & T \end{bmatrix}$$

Epipolar Geometry

$R^T p' - R^T T$ is p' in S_O
$R^T T$ also lies in plane
=> $R^T T \times (R^T p' - R^T T)$
is
perpendicular to
epipolar plane



Assumed to be canonical camera

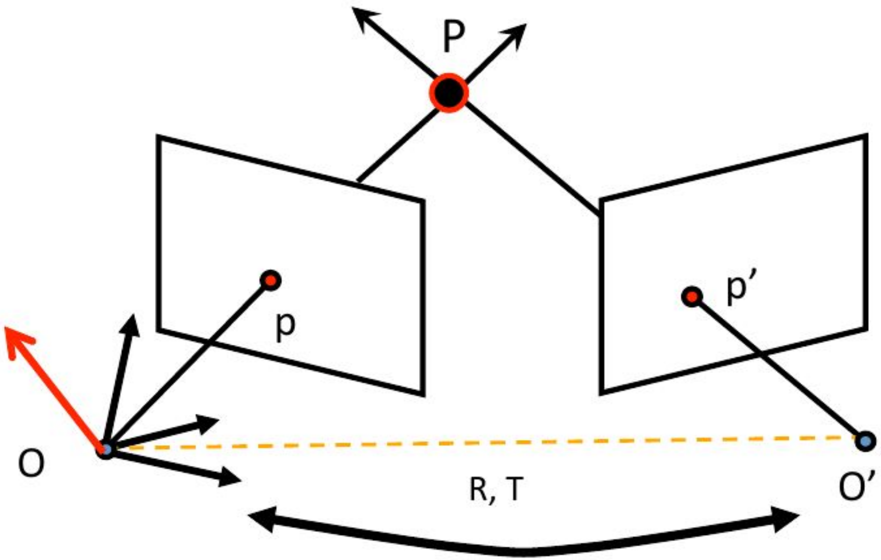
$$T \times (R\,p')$$

Perpendicular to epipolar plane

$$p^T \cdot \left[T \times (R\,p')\right] = 0$$

Epipolar Geometry

=> $R^T T \times (R^T p' - R^T T) =$
$R^T(T \times p')$ is
perpendicular to p
=> $(R^T(T \times p'))^T p = 0$
=> $(T \times p'^T)Rp = 0$



Assumed to be canonical camera

$$T \times (R\, p')$$

Perpendicular to epipolar plane

$$p^T \cdot [T \times (R\, p')] = 0$$

Epipolar Geometry

From Linear Algebra, the cross product of two vectors can be written as :

$$\mathbf{a} \times \mathbf{b} = \begin{bmatrix} 0 & -a_z & a_y \\ a_z & 0 & -a_x \\ -a_y & a_x & 0 \end{bmatrix} \begin{bmatrix} b_x \\ b_y \\ b_z \end{bmatrix}$$

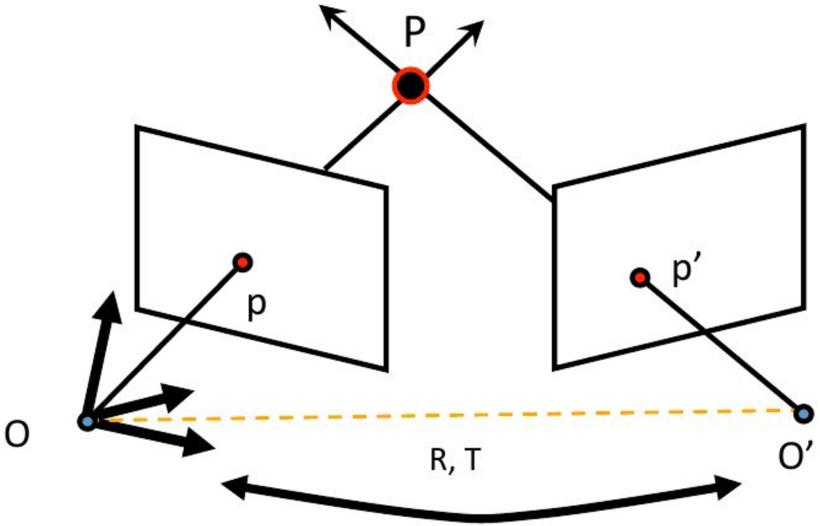
Epipolar Geometry

From Linear Algebra, the cross product of two vectors can be written as :

$$\mathbf{a} \times \mathbf{b} = \begin{bmatrix} 0 & -a_z & a_y \\ a_z & 0 & -a_x \\ -a_y & a_x & 0 \end{bmatrix} \begin{bmatrix} b_x \\ b_y \\ b_z \end{bmatrix}$$

$[a_x]$: skew symmetric

Epipolar Geometry

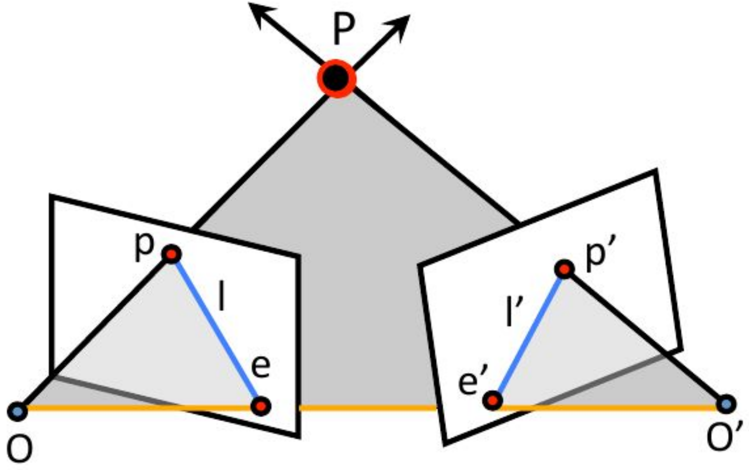


$$p^T \cdot [T \times (R\,p')] = 0 \rightarrow p^T \cdot [T_\times] \cdot R\,p' = 0$$

(Longuet-Higgins, 1981) E = essential matrix

Credits: Fei Fei Li

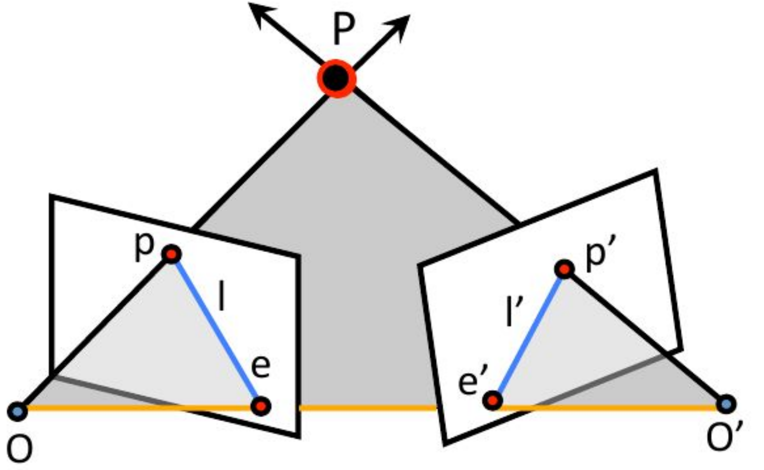
Epipolar Geometry: Essential Matrix (E)



- E p' is the epipolar line associated with p' ($l = E\,p'$)
- E^T p is the epipolar line associated with p ($l' = E^T p$)
- E is singular (rank two)
- $E\,e' = 0$ and $E^T e = 0$
- E is 3x3 matrix; 5 DOF

Credits: Fei Fei Li

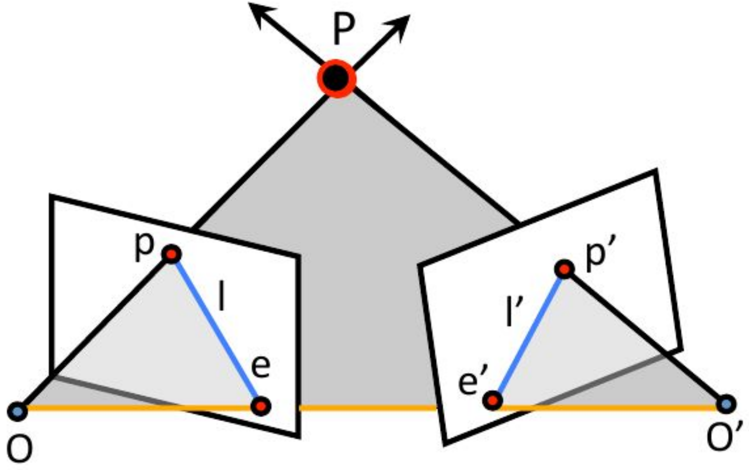
Epipolar Geometry



- $E\, p'$ is the epipolar line associated with p' ($l = E\, p'$)
- $E^T p$ is the epipolar line associated with p ($l' = E^T p$)
- E is singular (rank two)
- $E\, e' = 0$ and $E^T e = 0$
- E is 3x3 matrix; 5 DOF

Why?

Epipolar Geometry



- E p' is the epipolar line associated with p' (l = E p')
- E^T p is the epipolar line associated with p (l' = E^T p)
- E is singular (rank two)
- E e' = 0 and E^T e = 0
- E is 3x3 matrix; 5 DOF

ax + by + c = 0

i.e L = [a b c]^T represents a line in homogeneous coordinates.

=> $z^T L = 0$

where,
z = [x, y, 1]^T

Epipolar Geometry

$$M = K \begin{bmatrix} I & 0 \end{bmatrix} \qquad M' = K' \begin{bmatrix} R & T \end{bmatrix}$$

$$p'_c = K'^{-1} p'$$

$$p_c = K^{-1} p$$

Epipolar Geometry: Fundamental Matrix (F)

$$p'_c = K'^{-1} p' \qquad p_c = K^{-1} p$$

$$p_c'^T [T_\times] R p_c = 0$$

$$p'^T \underbrace{K'^{-T} [T_\times] R K^{-1}}_{\text{F: Fundamental Matrix}} p = 0$$

Epipolar Geometry: Properties of F

- F is a rank 2 homogeneous matrix with 7 degrees of freedom.
- **Point correspondence**: If $\mathbf{x}$ and $\mathbf{x}'$ are corresponding image points, then $\mathbf{x}'^{\mathsf{T}}\mathrm{F}\mathbf{x} = 0$.
- **Epipolar lines**:
  - $\mathbf{l}' = \mathrm{F}\mathbf{x}$ is the epipolar line corresponding to $\mathbf{x}$.
  - $\mathbf{l} = \mathrm{F}^{\mathsf{T}}\mathbf{x}'$ is the epipolar line corresponding to $\mathbf{x}'$.
- **Epipoles**:
  - $\mathrm{F}\mathbf{e} = \mathbf{0}$.
  - $\mathrm{F}^{\mathsf{T}}\mathbf{e}' = \mathbf{0}$.

Credits: Richard Hartley, Andrew Zisserman

Epipolar Geometry: Estimating F

- Assume that you have m correspondences
- Each correspondence satisfies:

$$\bar{p}_{r_i}^T F \bar{p}_{l_i} = 0 \quad i = 1, \ldots, m$$

- F is a 3x3 matrix (9 entries)
- Set up a **HOMOGENEOUS** linear system with 9 unknowns

Credits: Robert Collins, Penn State

Epipolar Geometry: Estimating F

$$\bar{p}_{l_i} = (x_i \;\; y_i \;\; 1)^T \quad \bar{p}_{r_i} = (x'_i \;\; y'_i \;\; 1)^T$$

$$\bar{p}_{r_i}^T F \bar{p}_{l_i} = 0 \quad i = 1, \ldots, m$$

$$\begin{bmatrix} x'_i & y'_i & 1 \end{bmatrix} \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix} \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} = 0$$

Credits: Robert Collins, Penn State

Epipolar Geometry: Estimating F

$$\begin{bmatrix} x'_i & y'_i & 1 \end{bmatrix} \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix} \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} = 0$$

Given m point correspondences…

$$\begin{bmatrix} x_1x'_1 & x_1y'_1 & x_1 & y_1x'_1 & y_1y'_1 & y_1 & x'_1 & y'_1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_mx'_m & x_my'_m & x_m & y_mx'_m & y_my'_m & y_m & x'_m & y'_m & 1 \end{bmatrix} \begin{bmatrix} f_{11} \\ f_{21} \\ f_{31} \\ f_{12} \\ f_{22} \\ f_{32} \\ f_{13} \\ f_{23} \\ f_{33} \end{bmatrix} = 0$$

Think: how many points do we need?

Credits: Robert Collins, Penn State

Epipolar Geometry: Estimating F

Assume that we need a non-trivial solution of:

$$Ax = 0$$

with m equations and n unknowns, m >= n – 1 and rank(A) = n-1

Since the norm of x (||x||) can be arbitrary we generally find the solution with norm equal to 1 in order to avoid the trivial solution.

Hence minimum 8 points are needed to solve the above equation. Therefore it is also called **8-point algorithm.**

Credits: Robert Collins, Penn State

Epipolar Geometry: Estimating F

Required Optimization:

$$\min_{\mathbf{x}} ||A\mathbf{x}||^2 \text{ s.t. } ||x||^2 = 1$$

$$||A\mathbf{x}||^2 = (A\mathbf{x})^T(A\mathbf{x}) = \mathbf{x}^T A^T A\mathbf{x}$$

$$||\mathbf{x}||^2 = \mathbf{x}^T\mathbf{x} = 1$$

Credits: Robert Collins, Penn State

Epipolar Geometry: Estimating F

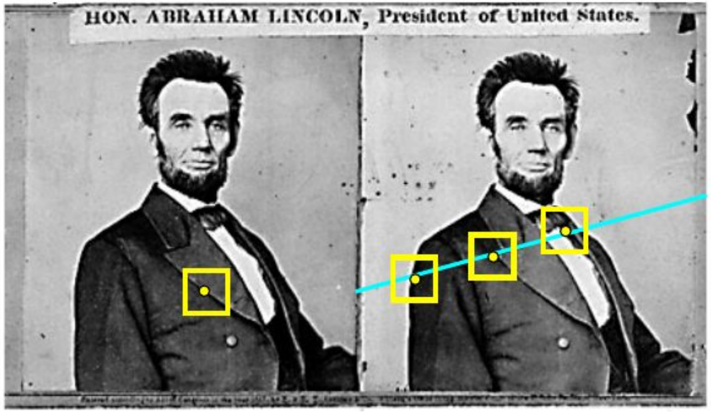
Solution:

- Construct the m x 9 matrix A
- Find the SVD of A: $A = UDV^T$
- The entries of F are the components of the column of V corresponding to the least s.v.

Credits: Robert Collins, Penn State

Stereo Matching

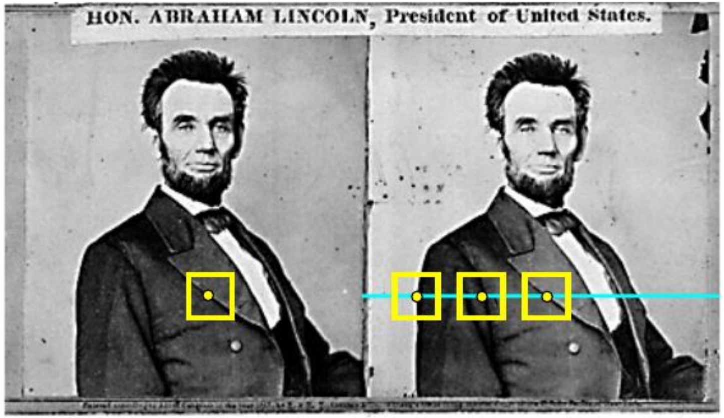
Various Methods for Stereo Matching



- 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?

Slide credit: J. Hayes

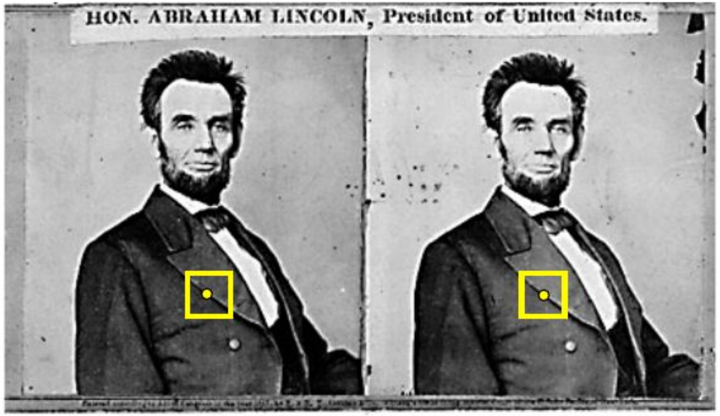
Various Methods for Stereo Matching



- 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')

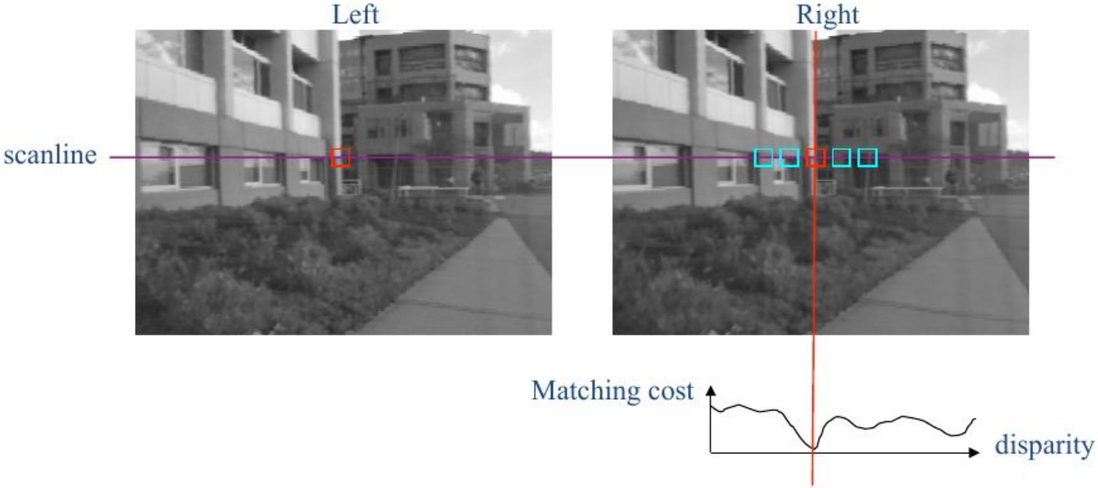
Various Methods for Stereo Matching

- 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



Slide credit: J. Hayes

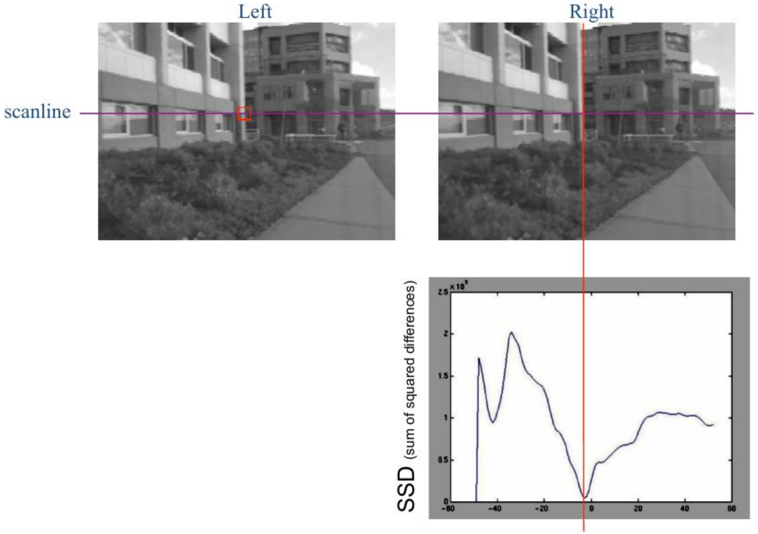
Various Methods for Stereo Matching

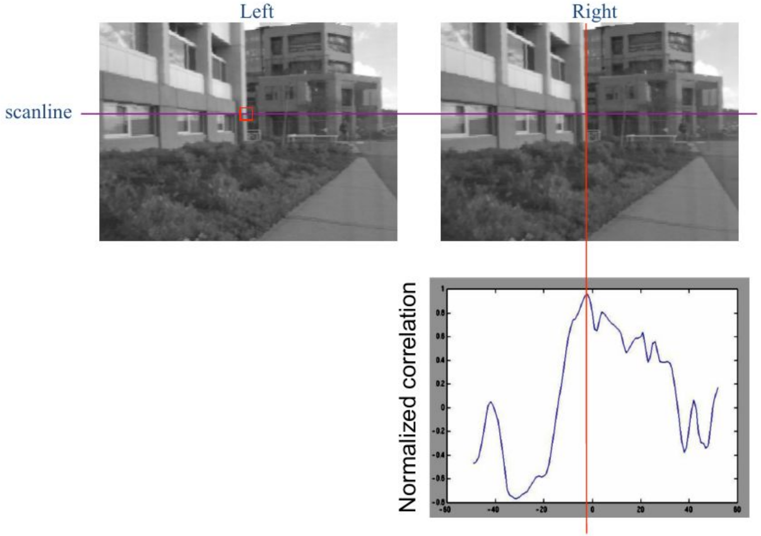


- Slide a window along the right scanline and compare contents of that window with the reference window in the left image
- Matching cost: SSD or normalized correlation

Slide credit: J. Hayes

Various Methods for Stereo Matching





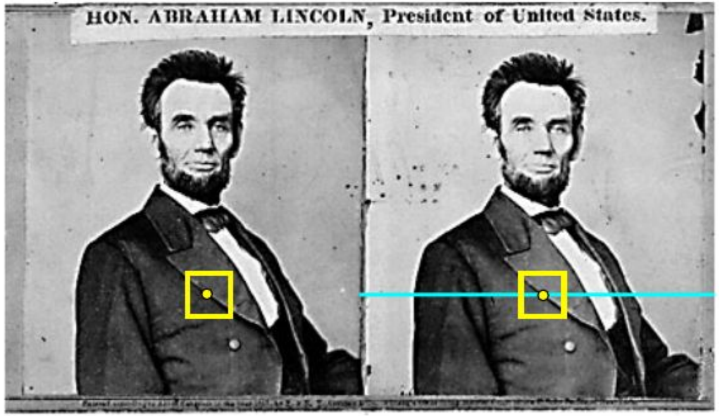
Various Methods for Stereo Matching

W = 3 W = 20

– Smaller window
  + More detail
  • More noise

– Larger window
  + Smoother disparity maps
  • Less detail

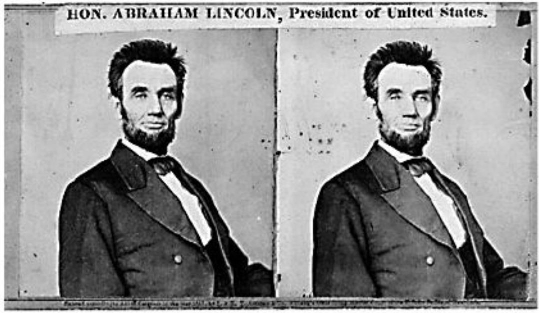
Various Methods for Stereo Matching



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

Slide credit: J. Hayes

Various Methods for Stereo Matching



Textureless surfaces

Occlusions, repetition

Specular surfaces

Slide credit: J. Hayes

Stereo Block Matching

- Similar to what we just saw in window sizes example.

- Idea is to instead of matching pixel values, match regions of image, this is done in order to increase robustness in the depth prediction.
- **Sparse Stereo Matching**: Use of key points or features to serve as corresponding points on the two images.
- **Dense Stereo Matching**: Match all pixels in a region along a scan line in pair of stereo rectified images.

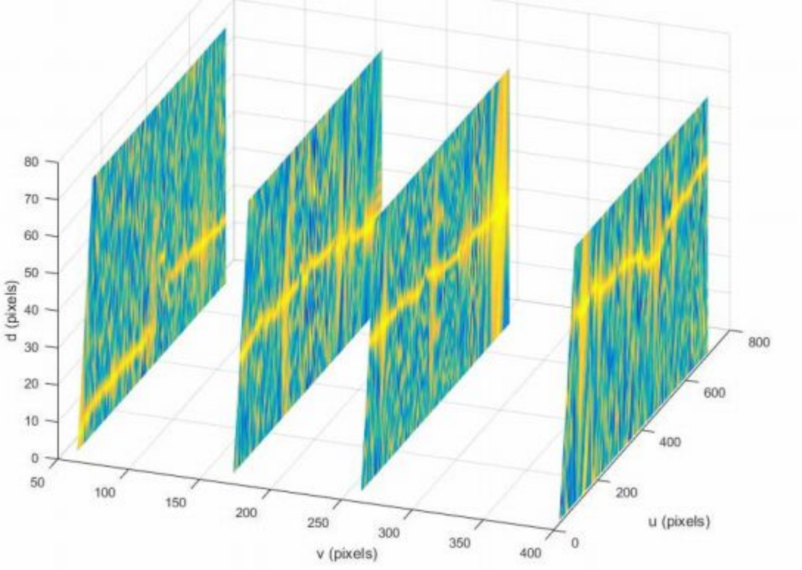
Stereo Block Matching

Credits: Trym Vegard Haavardsholm

Stereo Block Matching: Global Optimization

- Instead of finding best disparity for each pixel, find d so that global energy is minimum:

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



Credits: Trym Vegard Haavardsholm

Stereo Block Matching: Global Optimization

$$E(D) = \sum_{p} \left(C(p, D_p) + \sum_{q \in N_p} P\,T[\,|D_p - D_q| \geq 1] \right).$$

Minimize E over D to get D*

Cost of pixel wise matching

Penalty based on neighbours mismatches, I.e, penalty for neighbours having different disparity

Credits: HEIKO HIRSCHMÜLLER

Stereo Block Matching: Global Optimization

Guess the Drawbacks!!

$$E(D) = \sum_{p} \left(C(p, D_p) + \sum_{q \in N_p} PT[|D_p - D_q| \geq 1] \right).$$

Minimize E over D to get D*

Cost of pixel wise matching

Penalty based on neighbours mismatches, I.e, penalty for neighbours having different disparity

Stereo Block Matching: Global Optimization

Guess the Drawbacks!!

- Too Computationally Intensive
- NP Complete Problem

$$E(D) = \sum_{p} \left(C(p, D_p) + \sum_{q \in N_p} PT[|D_p - D_q| \geq 1] \right).$$

Minimize E over D to get D*

Cost of pixel wise matching

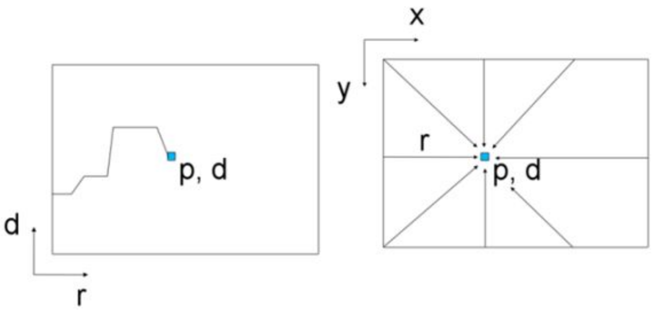
Penalty based on neighbours mismatches, I.e, penalty for neighbours having different disparity

Stereo Block Matching: Semi Global Matching

$$E(D) = \sum_{p} \left(C(p, D_p) + \sum_{q \in N_p} P_1 T[|D_p - D_q| = 1] + \sum_{q \in N_p} P_2 T[|D_p - D_q| > 1] \right).$$

Credits: HEIKO HIRSCHMÜLLER

Stereo Block Matching: Semi Global Matching



$$
\begin{aligned}
L_{\mathbf{r}}(\mathbf{p}, d) = C(\mathbf{p}, d) + \min(&L_{\mathbf{r}}(\mathbf{p} - \mathbf{r}, d), \\
&L_{\mathbf{r}}(\mathbf{p} - \mathbf{r}, d - 1) + P_1, \\
&L_{\mathbf{r}}(\mathbf{p} - \mathbf{r}, d + 1) + P_1, \\
&\min_i L_{\mathbf{r}}(\mathbf{p} - \mathbf{r}, i) + P_2) - \min_k L_{\mathbf{r}}(\mathbf{p} - \mathbf{r}, k).
\end{aligned}
$$

Credits: HEIKO HIRSCHMÜLLER