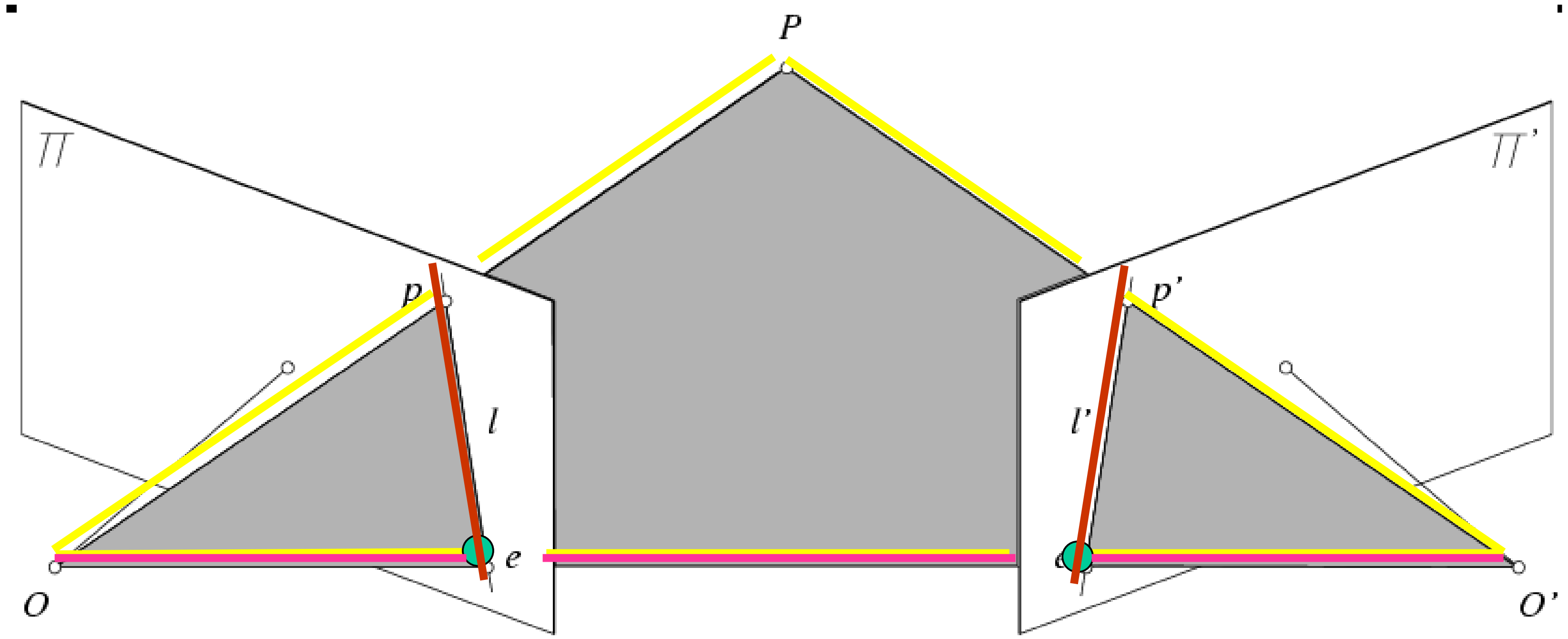
Unit 4
Stereo Vision

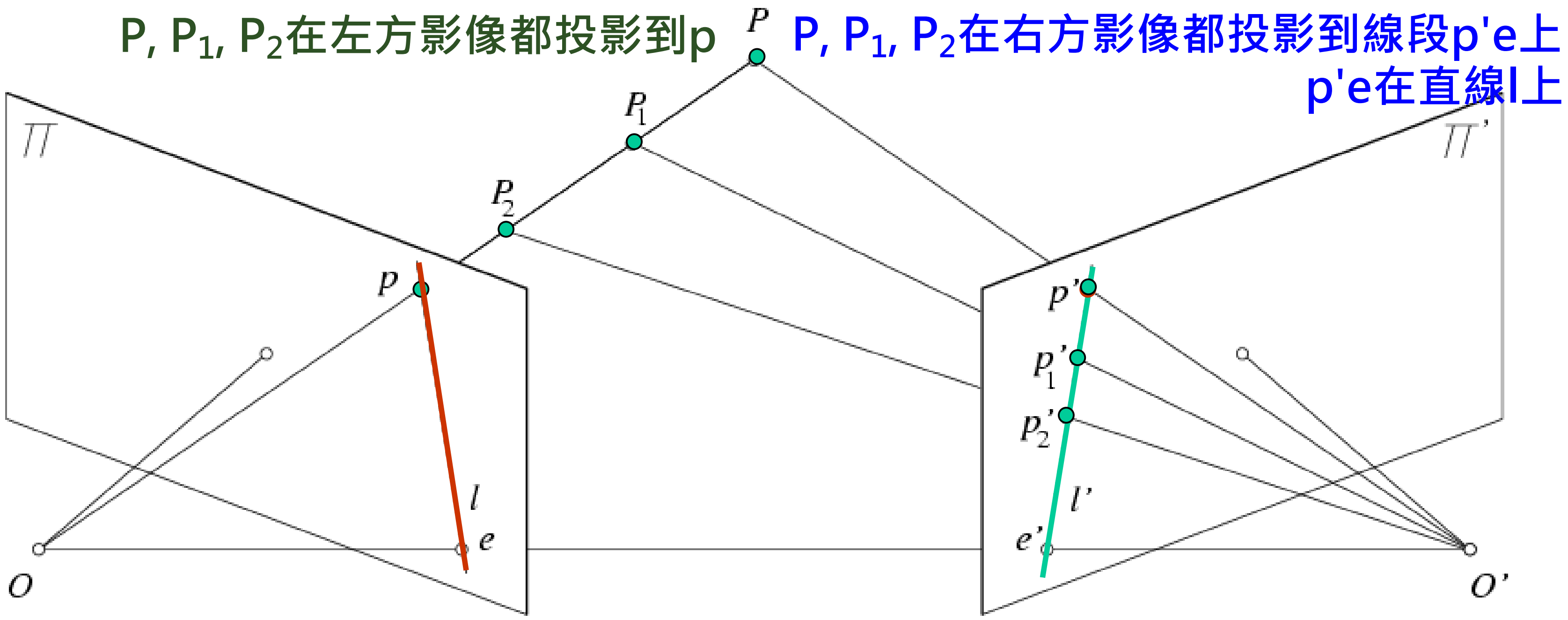
Ref: Szeliski, Sec. 6.2, 6.3, 7.1, 7.2

Epipolar Geometry



- Epipolar Plane
- Baseline 兩攝影機鏡心連線
- Epipoles 左方影像的Epipole對應到右方影像的Epipole Line 反之亦然
- Epipolar Lines

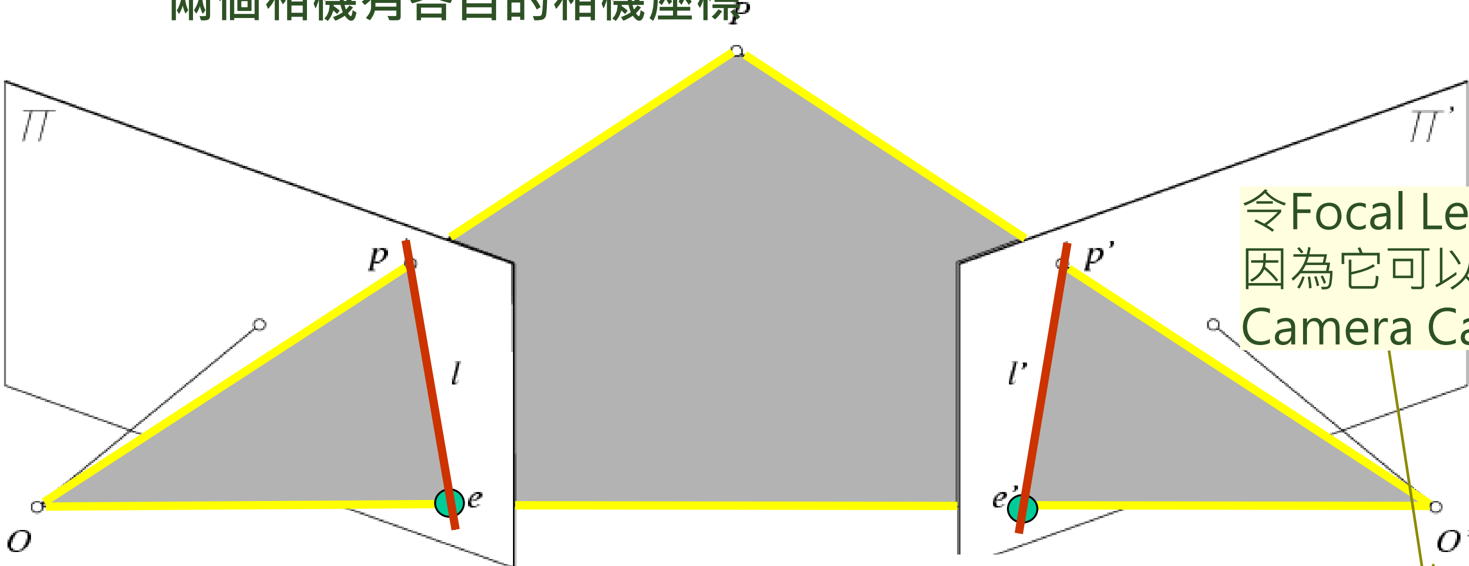
Epipolar Constraint



- Potential matches for *p* have to lie on the corresponding epipolar line *l′*.
- Potential matches for *p′* have to lie on the corresponding epipolar line *l*.

Epipolar Constraint: Calibrated Case

兩個相機有各自的相機座標



令Focal Length = 1
因為它可以塞進
Camera Calib Mat. K

$$\overrightarrow{Op} \cdot [\overrightarrow{OO'} \times \overrightarrow{O'p'}] = 0$$

平面上的
向量和平面
法向量垂直

外積得到平面
OO'P法向量

$$\boldsymbol{p} \cdot [\boldsymbol{t} \times (\mathcal{R}\boldsymbol{p}')] = 0 \quad \text{with} \begin{cases} \boldsymbol{p} = (u, v, 1)^T \\ \boldsymbol{p}' = (u', v', 1)^T \\ \mathcal{M} = (\mathrm{Id} \quad \mathbf{0}) \\ \mathcal{M}' = (\mathcal{R}^T, -\mathcal{R}^T\boldsymbol{t}) \end{cases}$$

右機 相對 左機
的Translation

$$\boldsymbol{p}^T \mathcal{E} \boldsymbol{p}' = 0 \quad \text{with} \quad \mathcal{E} = [\boldsymbol{t}_\times]\mathcal{R}$$

t×(Rp') = t_xR p' 外積寫成矩陣乘法

Essential Matrix
(Longuet-Higgins, 1981)

Cross Product

$$\mathbf{a} \times \mathbf{b} = \det \begin{bmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{bmatrix}.$$

外積的定義

$$\mathbf{a} \times \mathbf{b} = [\mathbf{a}]_\times \mathbf{b} = \begin{bmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

可以寫成這樣，意思相同

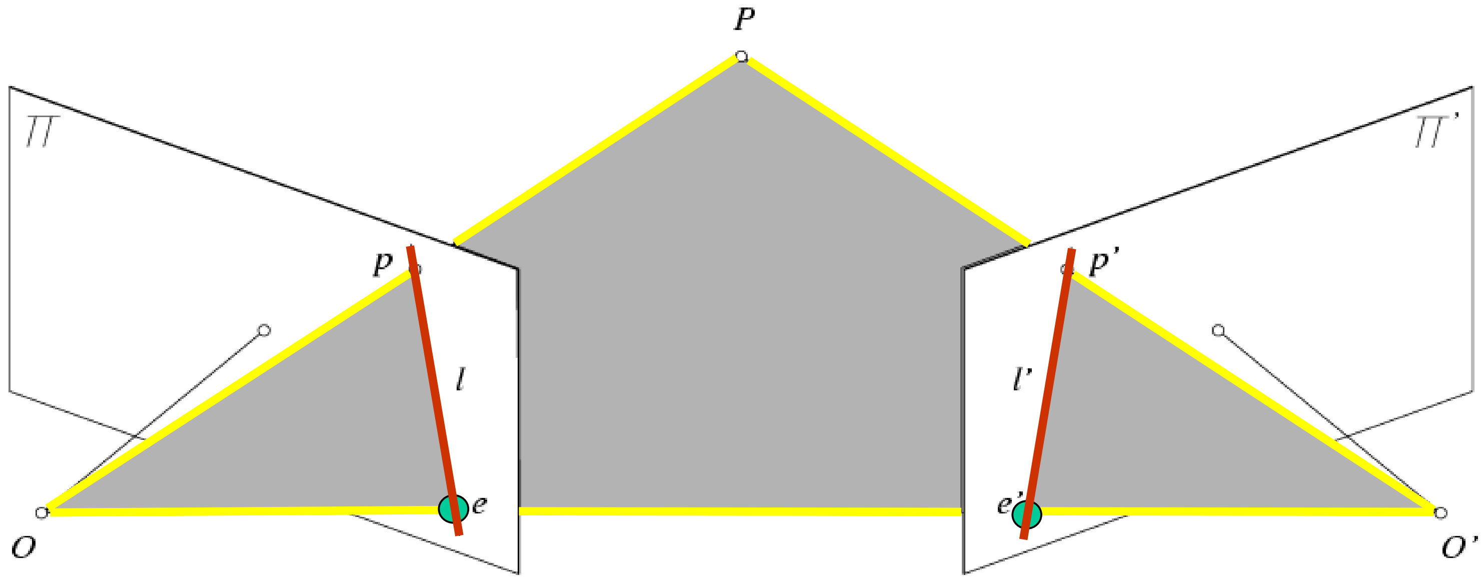
$$[\mathbf{a}]_\times \stackrel{\rm def}{=} \begin{bmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{bmatrix}.$$

Properties of the Essential Matrix

$$\boxed{\boldsymbol{p}^T \mathcal{E} \boldsymbol{p}' = 0 \quad \text{with} \quad \mathcal{E} = [\boldsymbol{t}_\times]\mathcal{R}}$$

- $\mathcal{E}$ p′ is the epipolar line associated with p′.
- $\mathcal{E}^T$p is the epipolar line associated with p.
- $\mathcal{E}$ e′=0 and $\mathcal{E}^T$e=0.
- $\mathcal{E}$ is singular.
- $\mathcal{E}$ has two equal non-zero singular values (Huang and Faugeras, 1989).

Epipolar Constraint: Uncalibrated Case



$$\hat{\boldsymbol{p}}^T \mathcal{E} \hat{\boldsymbol{p}}' = 0$$

$$\boldsymbol{p} = \mathcal{K} \hat{\boldsymbol{p}}$$

$$\boldsymbol{p}' = \mathcal{K}' \hat{\boldsymbol{p}}'$$

$$\boldsymbol{p}^T \mathcal{F} \boldsymbol{p}' = 0 \quad \text{with} \quad \mathcal{F} = \mathcal{K}^{-T} \mathcal{E} \mathcal{K}'^{-1}$$

Fundamental Matrix
(Faugeras and Luong, 1992)

Properties of the Fundamental Matrix

- $\mathcal{F}\,p'$ is the epipolar line associated with p'.
- $\mathcal{F}^{T}\,p$ is the epipolar line associated with p.
- $\mathcal{F}\,e'=0$ and $\mathcal{F}^{T}\,e=0$.
- $\mathcal{F}$ is singular.

The Eight-Point Algorithm (Longuet-Higgins, 1981)

$$(u, v, 1)\begin{pmatrix} F_{11} & F_{12} & F_{13} \\ F_{21} & F_{22} & F_{23} \\ F_{31} & F_{32} & F_{33} \end{pmatrix}\begin{pmatrix} u' \\ v' \\ 1 \end{pmatrix} = 0$$

$$(uu', uv', u, vu', vv', v, u', v', 1)\begin{pmatrix} F_{11} \\ F_{12} \\ F_{13} \\ F_{21} \\ F_{22} \\ F_{23} \\ F_{31} \\ F_{32} \\ F_{33} \end{pmatrix} = 0$$

$$\begin{pmatrix} u_1u'_1 & u_1v'_1 & u_1 & v_1u'_1 & v_1v'_1 & v_1 & u'_1 & v'_1 \\ u_2u'_2 & u_2v'_2 & u_2 & v_2u'_2 & v_2v'_2 & v_2 & u'_2 & v'_2 \\ u_3u'_3 & u_3v'_3 & u_3 & v_3u'_3 & v_3v'_3 & v_3 & u'_3 & v'_3 \\ u_4u'_4 & u_4v'_4 & u_4 & v_4u'_4 & v_4v'_4 & v_4 & u'_4 & v'_4 \\ u_5u'_5 & u_5v'_5 & u_5 & v_5u'_5 & v_5v'_5 & v_5 & u'_5 & v'_5 \\ u_6u'_6 & u_6v'_6 & u_6 & v_6u'_6 & v_6v'_6 & v_6 & u'_6 & v'_6 \\ u_7u'_7 & u_7v'_7 & u_7 & v_7u'_7 & v_7v'_7 & v_7 & u'_7 & v'_7 \\ u_8u'_8 & u_8v'_8 & u_8 & v_8u'_8 & v_8v'_8 & v_8 & u'_8 & v'_8 \end{pmatrix}\begin{pmatrix} F_{11} \\ F_{12} \\ F_{13} \\ F_{21} \\ F_{22} \\ F_{23} \\ F_{31} \\ F_{32} \end{pmatrix} = -\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

$$\sum_{i=1}^{n}(\boldsymbol{p}_i^T \mathcal{F} \boldsymbol{p}'_i)^2$$

under the constraint

$|\mathcal{F}|^2 = 1.$

Non-Linear Least-Squares Approach (Luong et al., 1993)

Minimize

$$\sum_{i=1}^{n} [\mathrm{d}^2(\boldsymbol{p}_i, \mathcal{F}\boldsymbol{p}_i') + \mathrm{d}^2(\boldsymbol{p}_i', \mathcal{F}^T\boldsymbol{p}_i)]$$

with respect to the coefficients of $\mathcal{F}$, using an appropriate rank-2 parameterization.

Problem with eight-point algorithm

250906.36	183269.57	921.81	200931.10	146766.13	738.21	272.19	198.81	1.00
2692.28	131633.03	176.27	6196.73	302975.59	405.71	15.27	746.79	1.00
416374.23	871684.30	935.47	408110.89	854384.92	916.90	445.10	931.81	1.00
191183.60	171759.40	410.27	416435.62	374125.90	893.65	465.99	418.65	1.00
48988.86	30401.76	57.89	298604.57	185309.58	352.87	846.22	525.15	1.00
164786.04	546559.67	813.17	1998.37	6628.15	9.86	202.65	672.14	1.00
116407.01	2727.75	138.89	169941.27	3982.21	202.77	838.12	19.64	1.00
135384.58	75411.13	198.72	411350.03	229127.78	603.79	681.28	379.48	1.00

$$\begin{pmatrix} F_{11} \\ F_{12} \\ F_{13} \\ F_{21} \\ F_{22} \\ F_{23} \\ F_{31} \\ F_{32} \\ F_{33} \end{pmatrix} = 0$$

linear least-squares:
unit norm vector F yielding smallest residual

What happens when there is noise?

The Normalized Eight-Point Algorithm (Hartley, 1995)

- Center the image data at the origin, and scale it so the mean squared distance between the origin and the data points is 2 pixels:

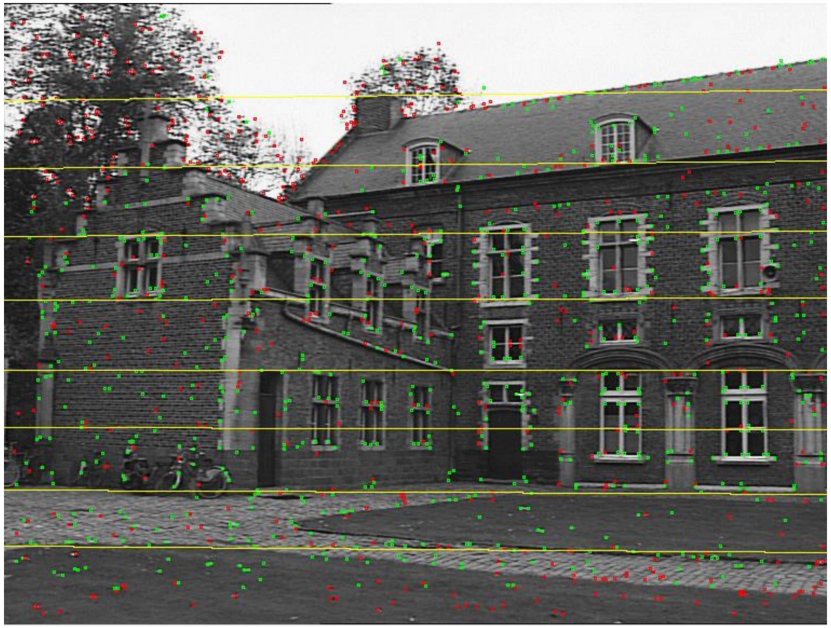
$$q_i = T\,p_i \quad , \quad q_i' = T'\,p_i'.$$

- Use the eight-point algorithm to compute $\mathcal{F}$ from the points q_i and q_i' .

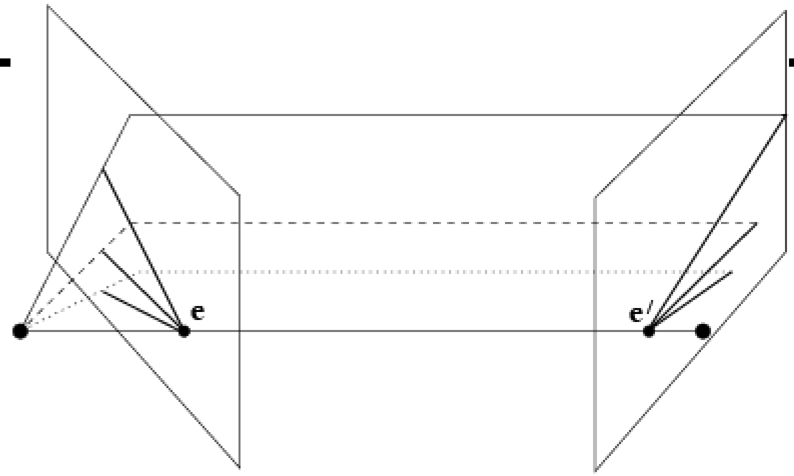
- Enforce the rank-2 constraint.

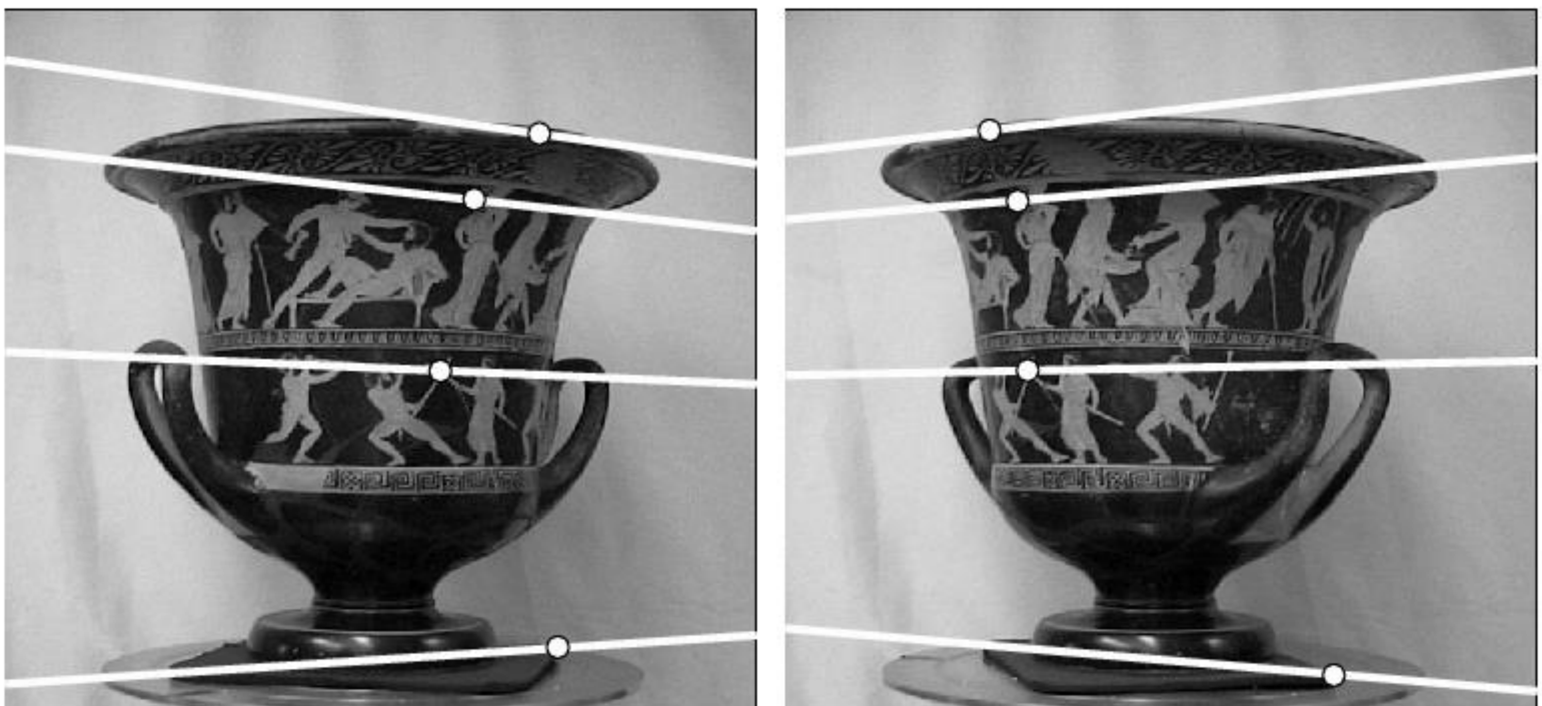
- Output $T^T \mathcal{F}\, T'$.

Epipolar geometry example



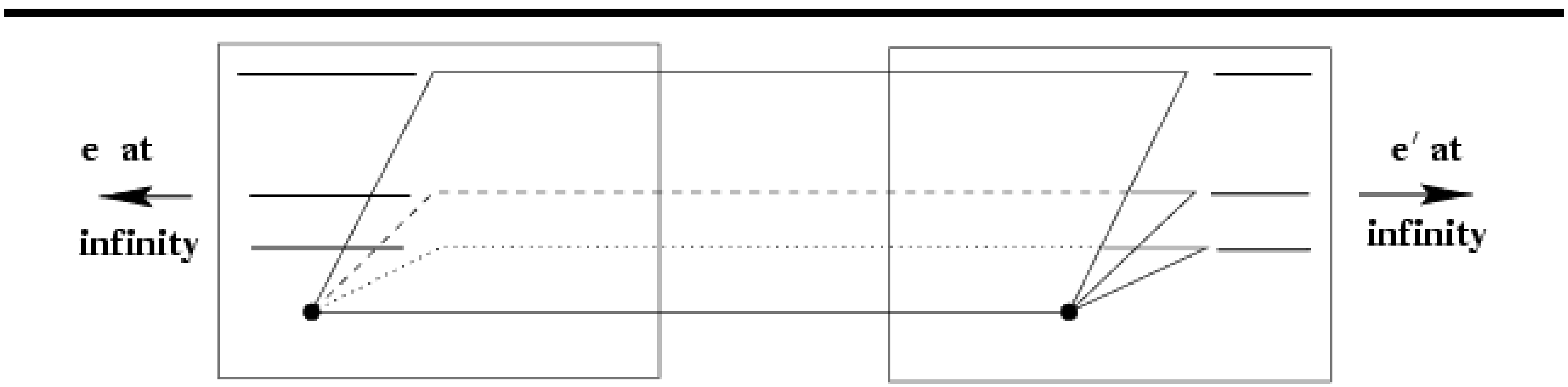
Example: converging cameras

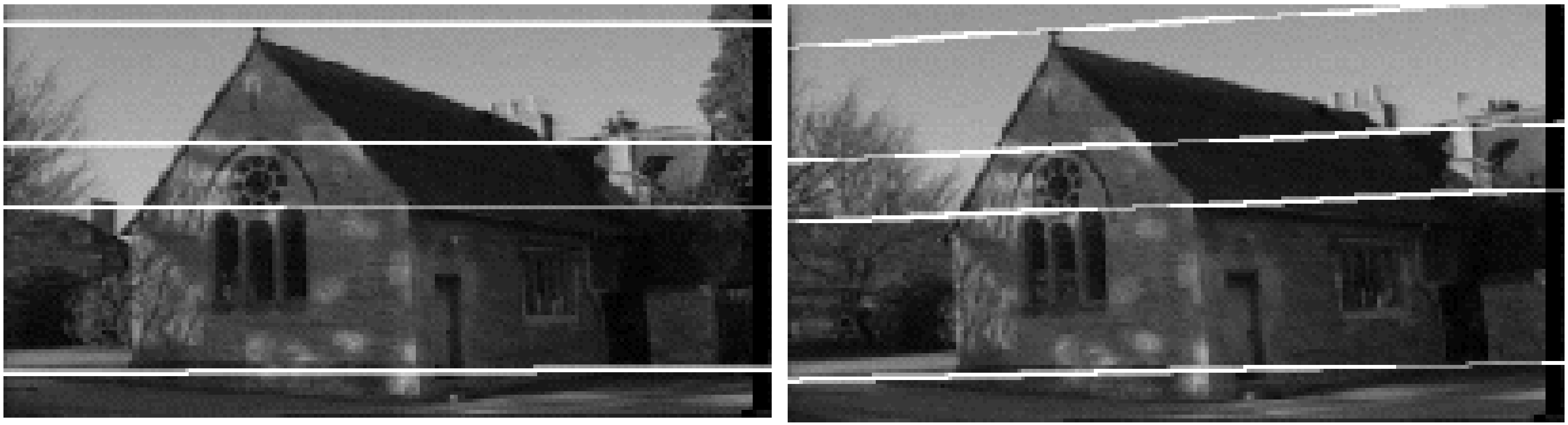




courtesy of Andrew Zisserman

Example: motion parallel with image plane

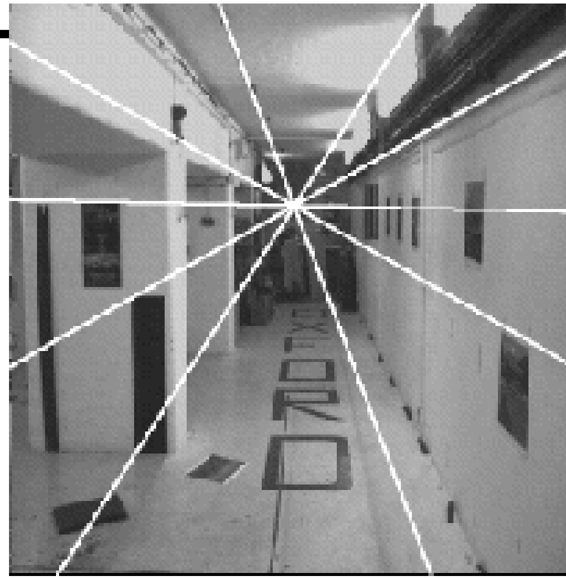


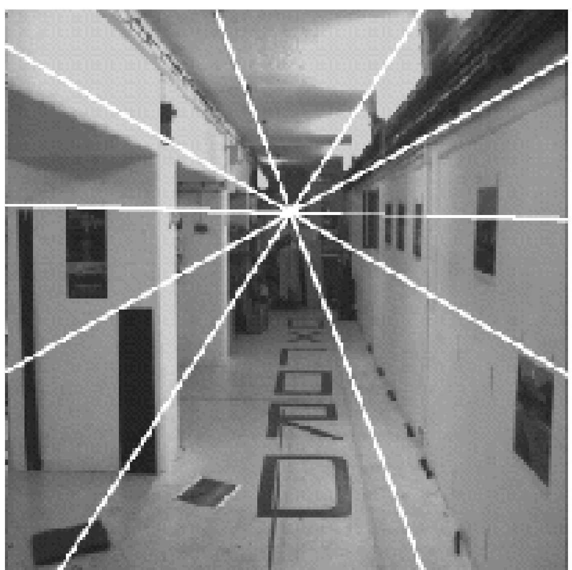


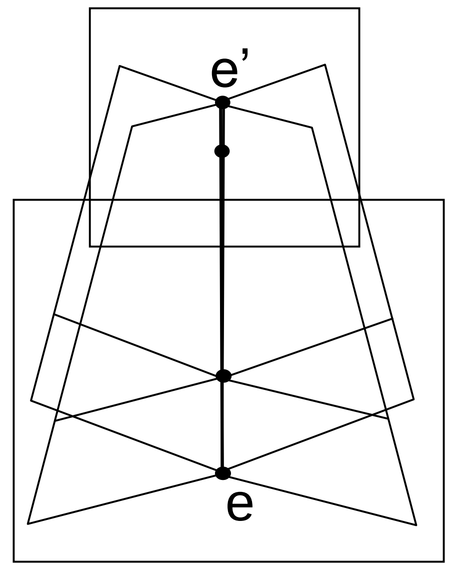
(simple for stereo → rectification)

courtesy of Andrew Zisserman

Example: forward motion





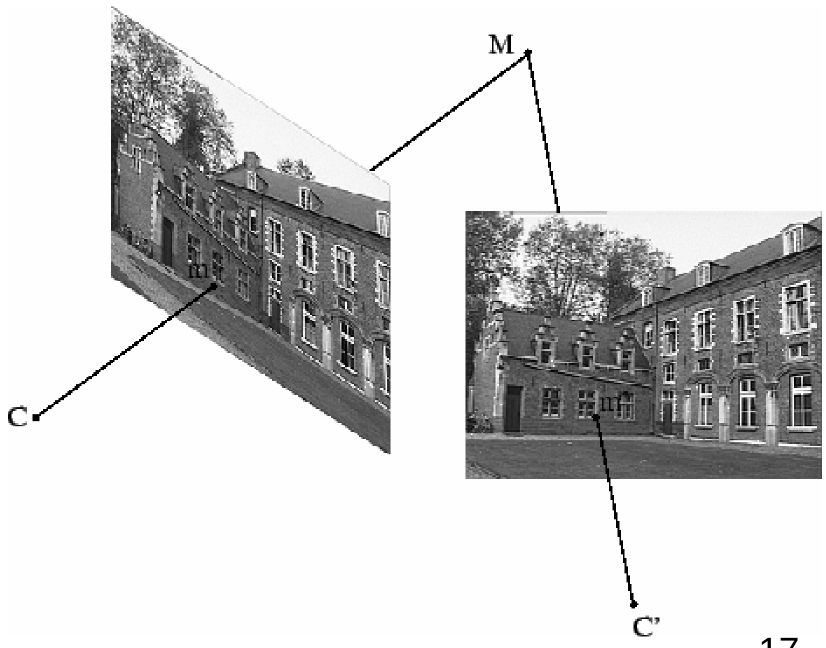


CS 6550

courtesy of Andrew Zisserman

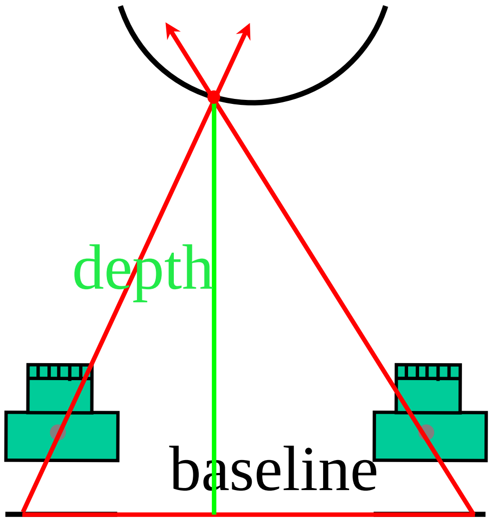
Stereo Reconstruction Problem

- Reconstruction of depth map from a pair of stereo images
- Triangulation
  - 3D point can be obtained as the intersection of the two line of sights



  – Requirements
    1. **Relative 3D camera poses and parameters for the stereo cameras (camera calibration)**
    2. **Pixel correspondences (stereo matching)**

Stereo Vision



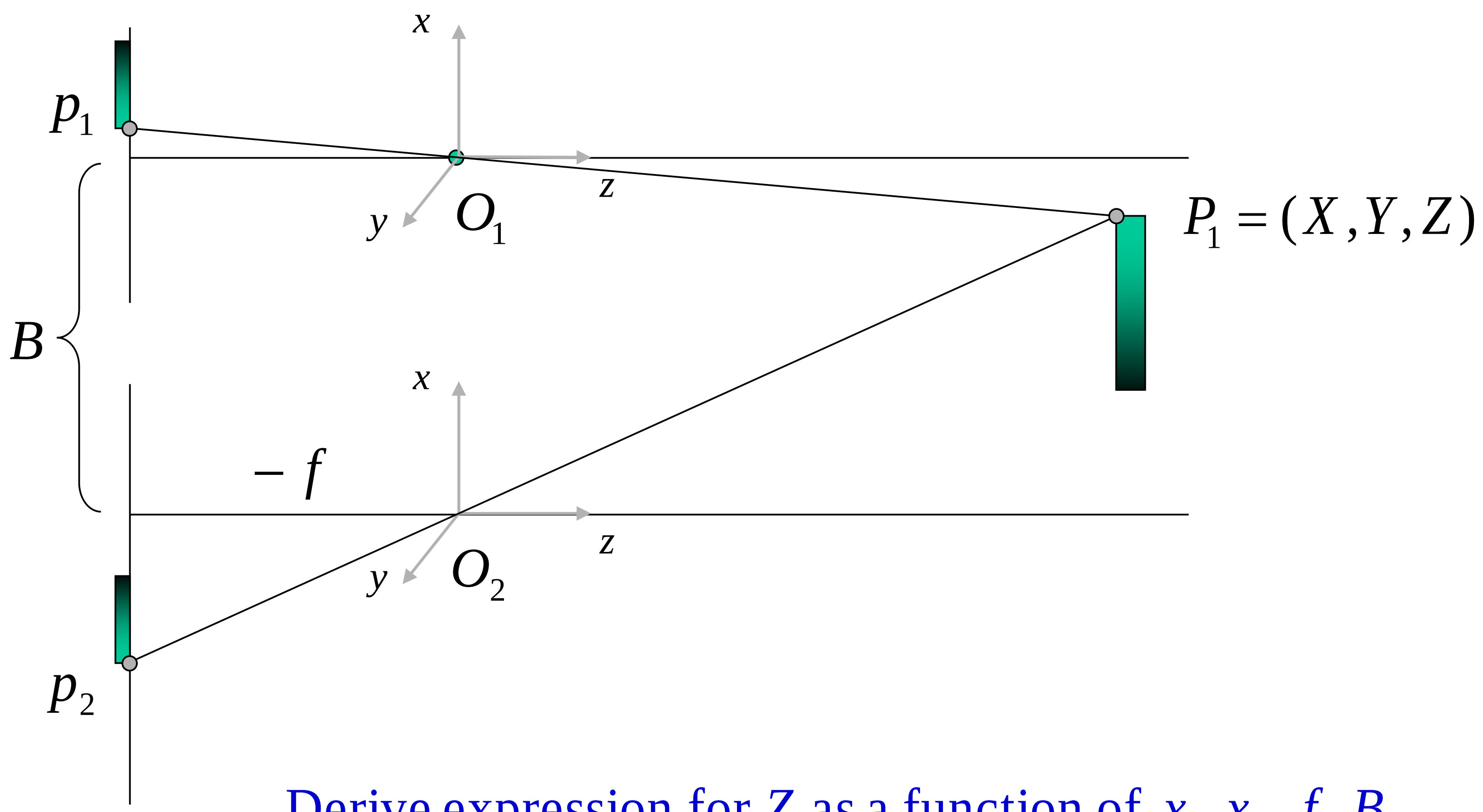
Triangulate on two images of the same point to recover depth.

- Feature matching across views
- Calibrated cameras

Matching correlation windows across scan lines.

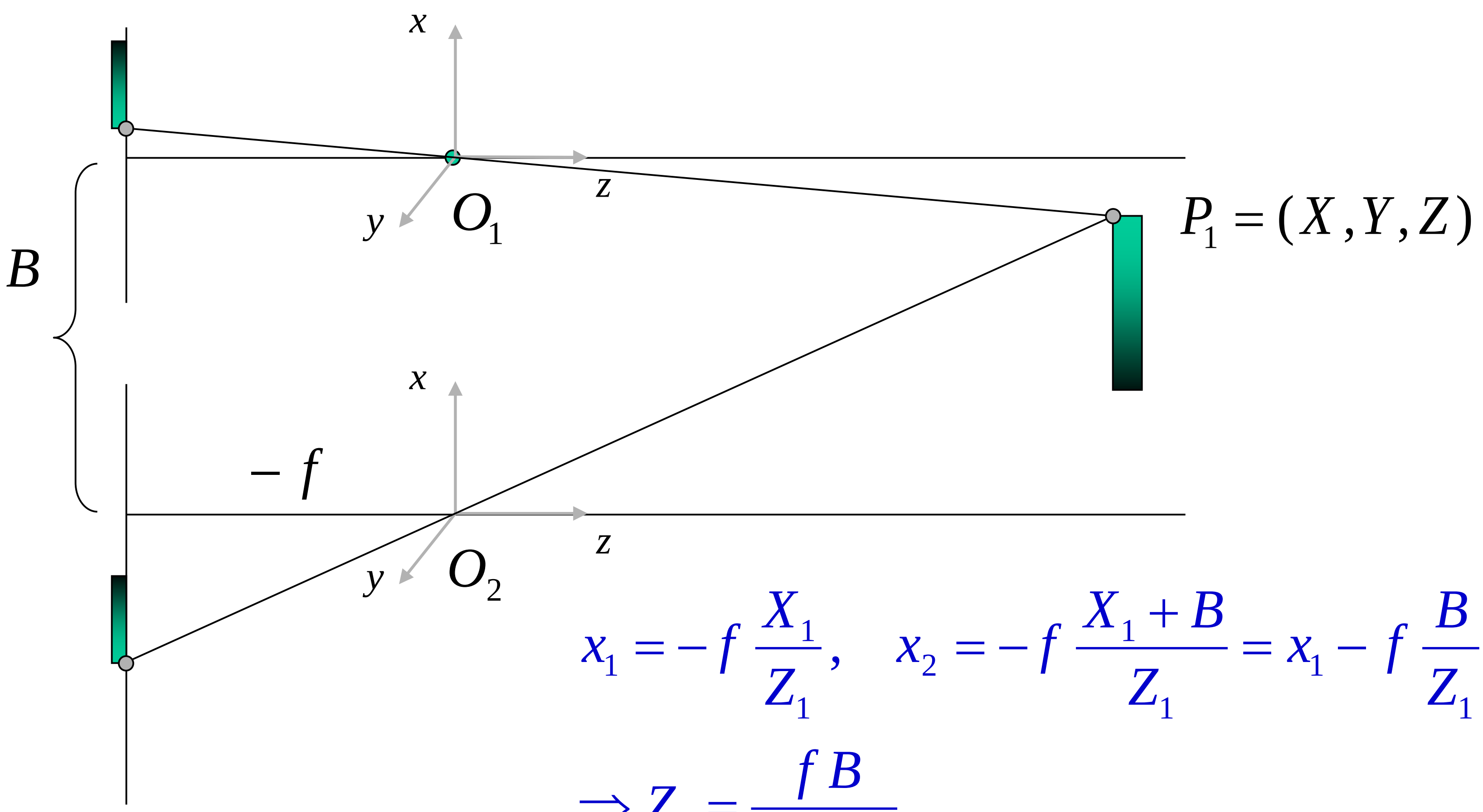
Disparity: deviation between horizontal positions of corresponding points in the calibrated stereo images, directly related to depth.

Basic Stereo Derivation



Derive expression for Z as a function of x_1, x_2, f, B

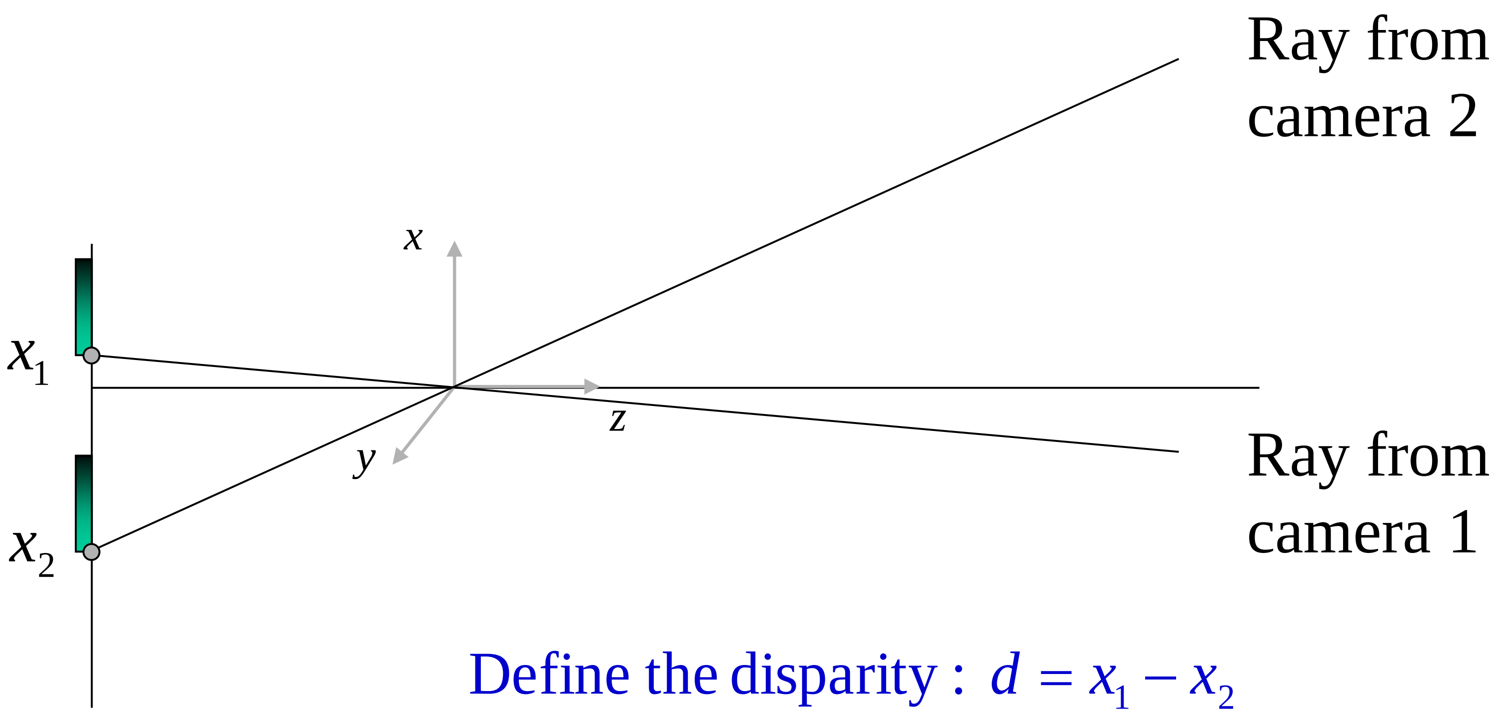
Basic Stereo Derivation



$$x_1 = -f\,\frac{X_1}{Z_1}, \quad x_2 = -f\,\frac{X_1 + B}{Z_1} = x_1 - f\,\frac{B}{Z_1}$$

$$\Rightarrow Z_1 = \frac{f\,B}{x_1 - x_2}$$

Basic Stereo Derivations

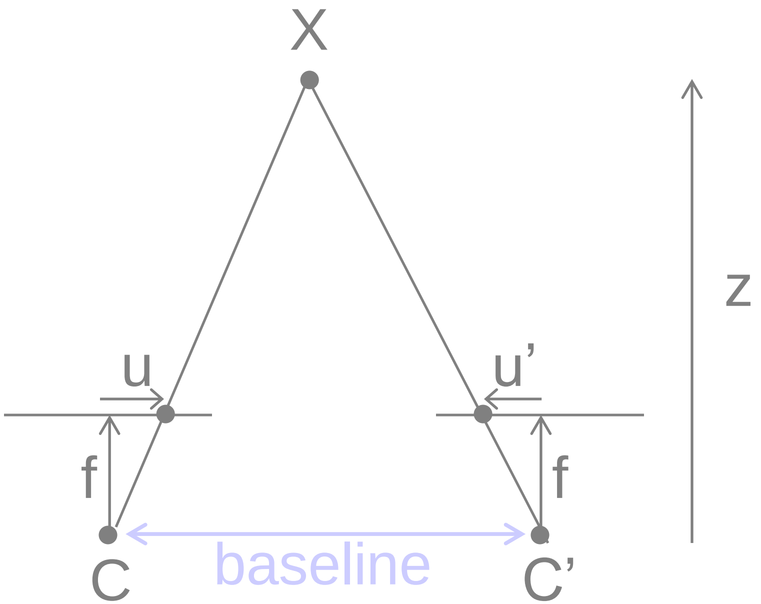


Define the disparity : $d = x_1 - x_2$

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

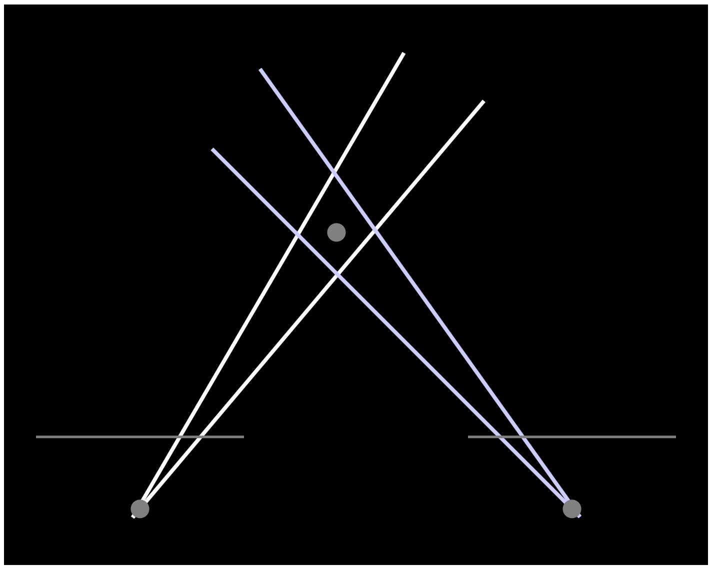
Stereo Reconstruction

- Steps
  - Calibrate cameras
  - Rectify images
  - Compute disparity
  - Estimate depth

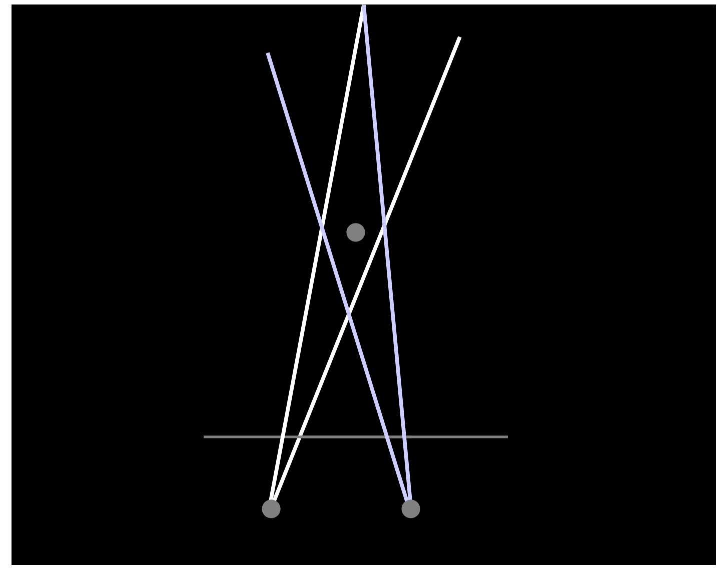


$$disparity = u - u' = \frac{baseline * f}{z}$$

Choosing the Baseline



Large Baseline



Small Baseline

- What's the optimal baseline?
  - Too small: large depth error
  - Too large: difficult search problem

Stereo vision

- Two processes:
  - the binocular fusion of features observed by the two eyes.
  - the reconstruction of their 3D preimage.
- The preimage of matching points can (in principle) be found at the intersection of the rays passing through these points and the associated pupil centers (or pinholes).
- Some methods must be devised to establish the correct correspondences and avoid erroneous depth measurements.
- Key components of stereo vision systems:
  - Camera calibration
  - Image rectification

Binocular Fusion – an example

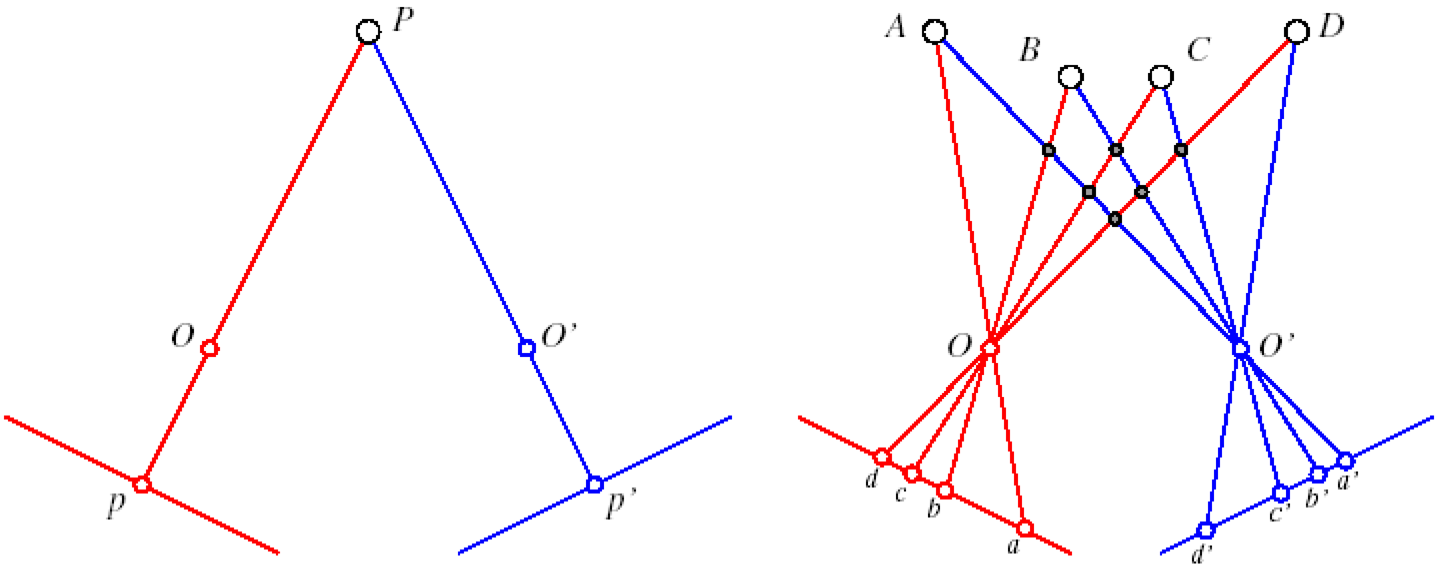


Figure 13.3. The binocular fusion problem: in the simple case of the diagram shown on the left, there is no ambiguity and stereo reconstruction is a simple matter. In the more usual case shown on the right, any of the four points in the left picture may, a priori, match any of the four points in the right one. Only four of these correspondences are correct, the other ones yielding the incorrect reconstructions shown as small grey discs.

Epipolar Geometry

- The epipolar geometry is the fundamental constraint in stereo.
- *Rectification* aligns epipolar lines with scanlines

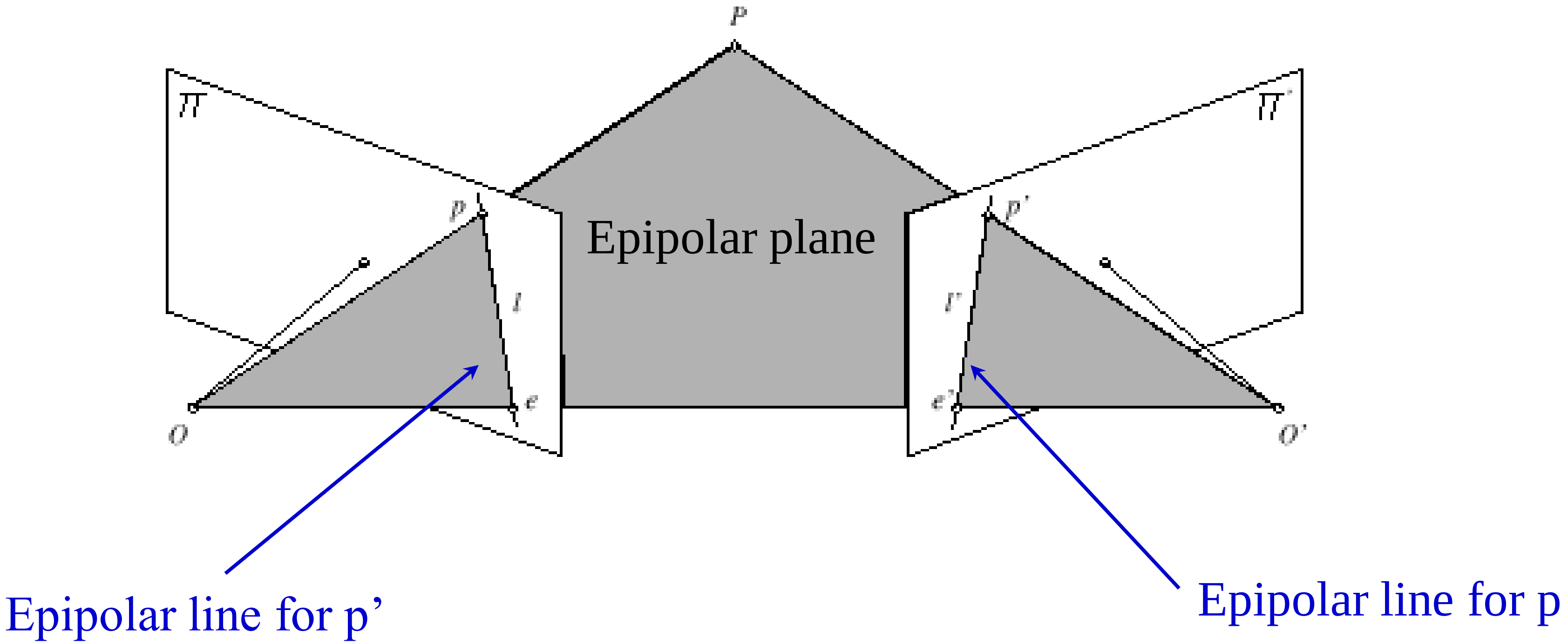
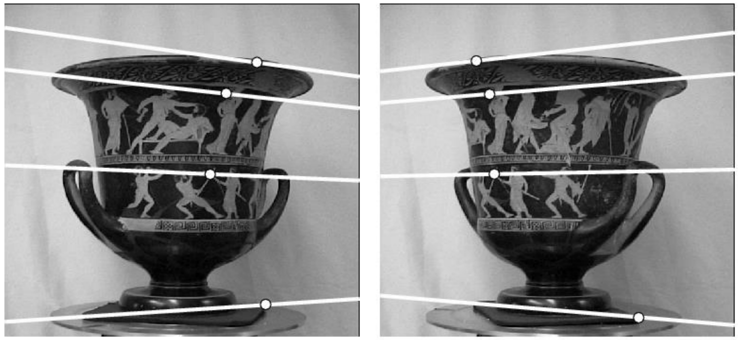


Image rectification

- *Rectification* aligns epipolar lines with scanlines
- Stereo algorithms are often considerably simplified when the images of interest have been rectified.
- Projecting the original pictures onto a common image plane parallel to the baseline joining the two optical centers. The rectified epipolar lines are scanlines of the new images, and they are also parallel to the baseline.
- Given two points p and p' located on the same scanline of the left and right images, with coordinates (u, v) and (u', v). The disparity is defined as the difference $d = u' - u$.

Image Rectification

- Once we have the camera information, the solution space of the matches between two views is restrained from 2D to 1D



- For further simplifying the searching problem, the stereo matching algorithm is performed on the rectified image pair

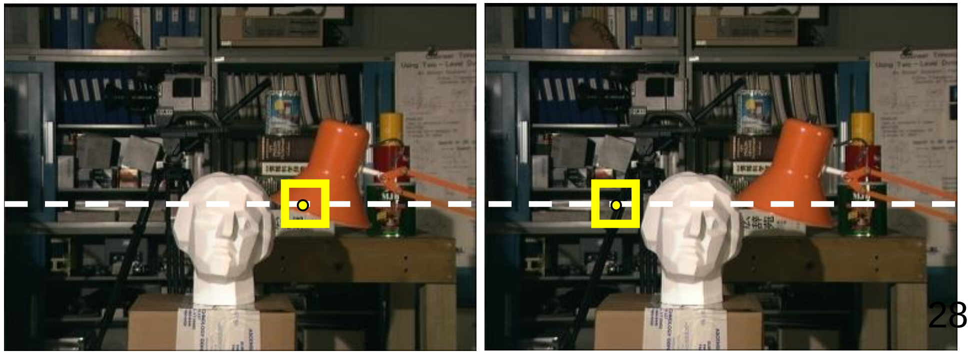
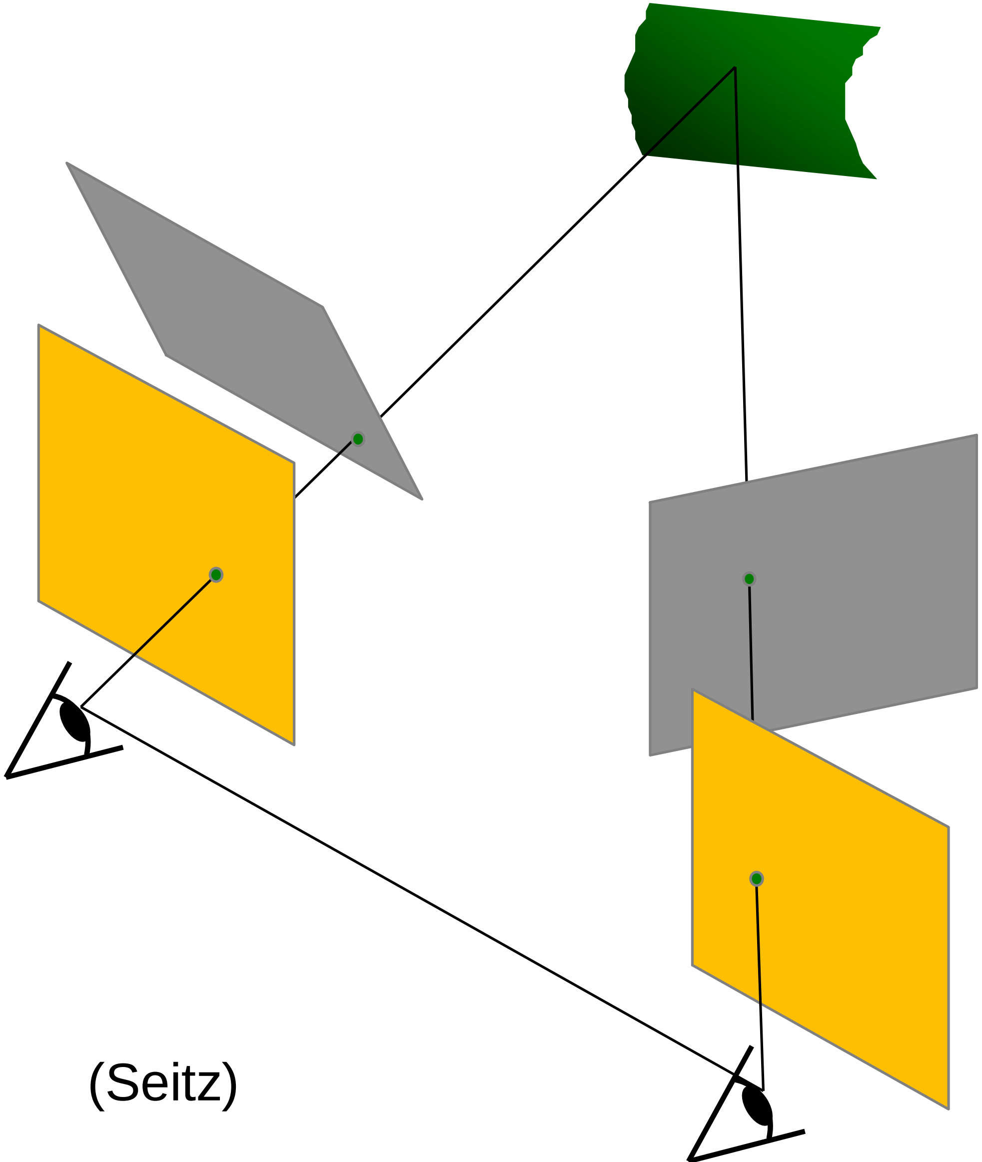


Image rectification

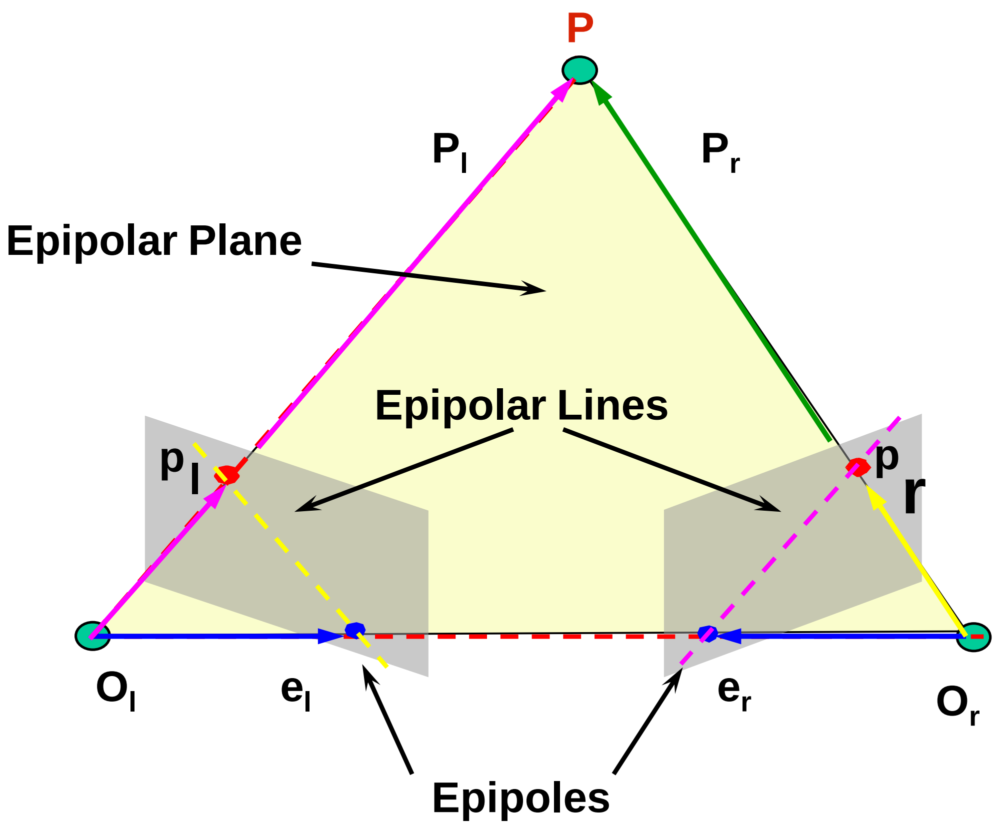
- Given general displacement how to warp the views
- Such that epipolar lines are parallel to each other
- How to warp it back to canonical configuration



(Seitz)

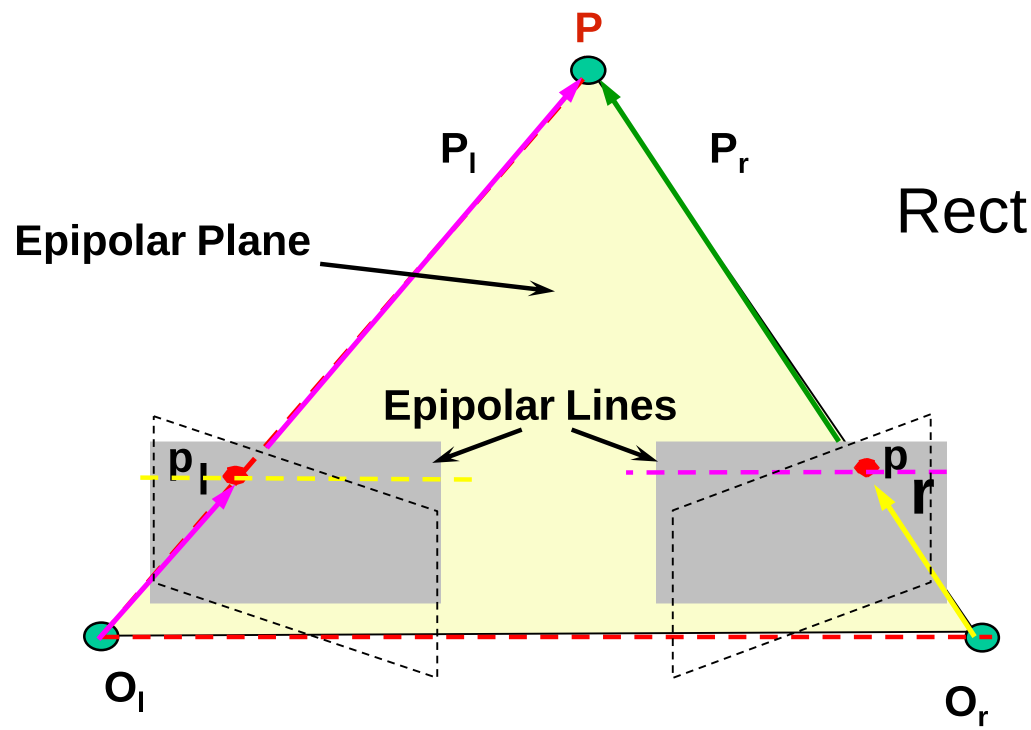
Rectification

- Problem: Epipolar lines not parallel to scan lines



Rectification

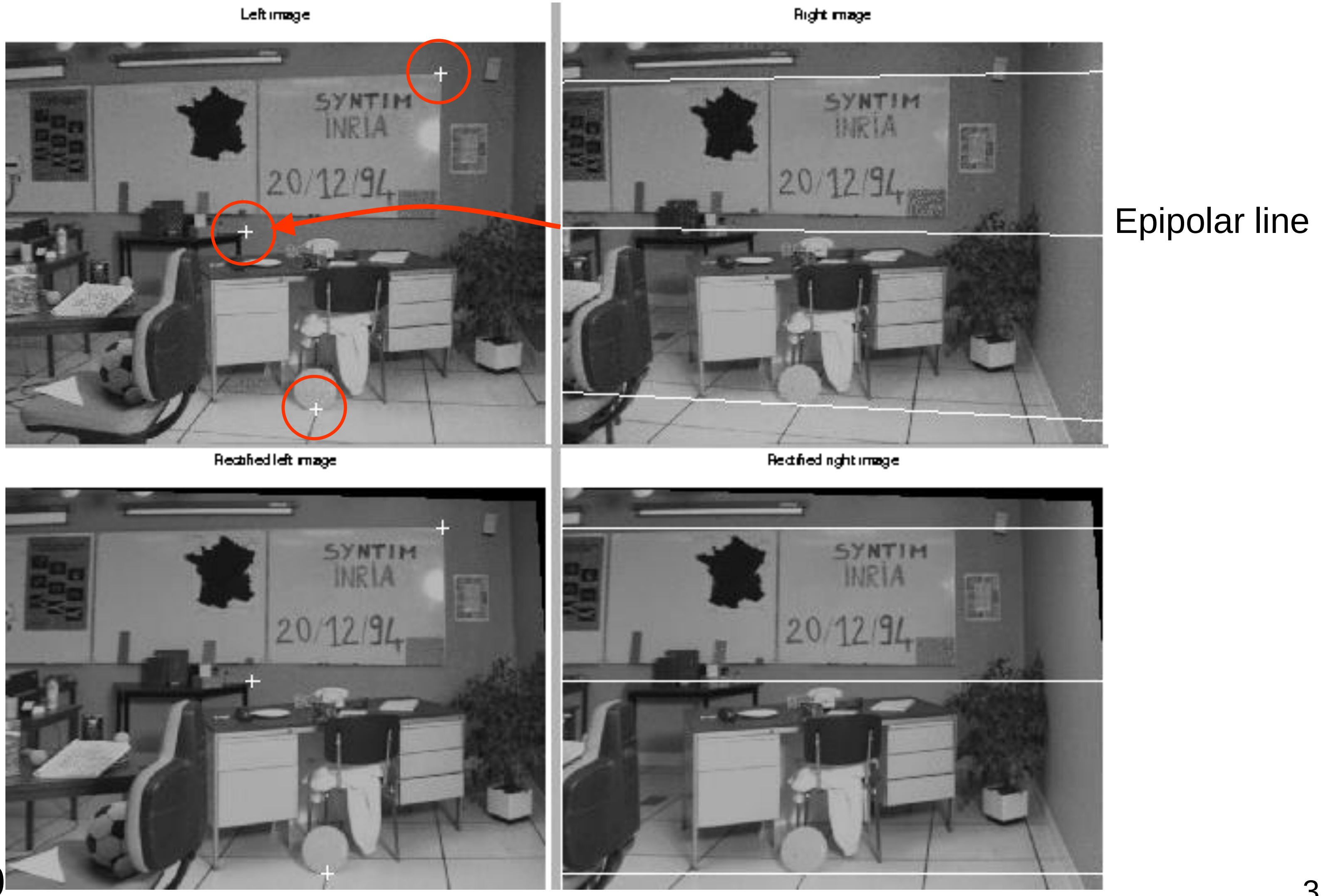
- Problem: Epipolar lines not parallel to scan lines



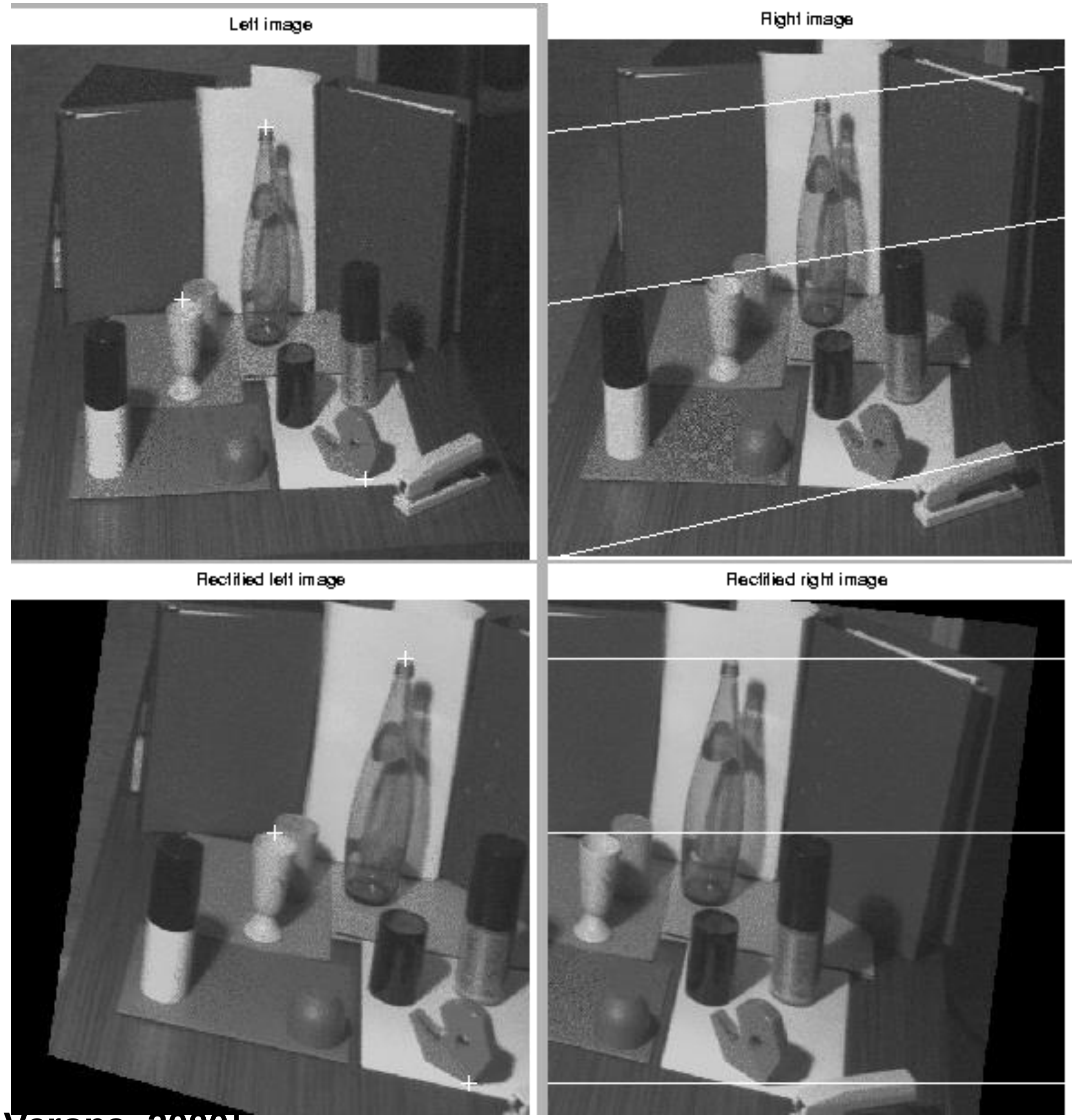
Rectified Images

Epipoles at ininity

Epipolar Rectified Stereo Images



Epipolar Rectified Images



Source: A. Fusiello, Verona, 2000]

Algorithm 11.1: Image rectification

1. *Epipoles are translated to infinity in both images.*

   Let $\mathbf{e}_L = [e_1, e_2, 1]^\top$ be the epipole in the left image and $e_1^2 + e_2^2 \neq 0$. This epipole is mapped to $\mathbf{e}^* \simeq [1, 0, 0]^\top$ as the rotation of the epipole $\mathbf{e}_L$ to the axis u and the projection

$$\hat{H}_L \simeq \begin{bmatrix} e_1 & e_2 & 0 \\ -e_2 & e_1 & 0 \\ -e_1 & -e_2 & e_1^2 + e_2^2 \end{bmatrix} . \qquad (11.50)$$

2. *Epipolar lines are unified to get a pair of elementary rectifying homographies.*

   Since $\mathbf{e}_R^* = [1, 0, 0]^\top$ is both left and right null space of $\hat{F}$, the modified fundamental matrix becomes

$$\hat{F} \simeq \begin{bmatrix} 0 & 0 & 0 \\ 0 & \alpha & \beta \\ 0 & \gamma & \delta \end{bmatrix} \qquad (11.51)$$

   and elementary rectifying homographies $\bar{H}_L$, $\bar{H}_R$ are chosen to make $\alpha = \delta = 0$ and $\beta = -\gamma$.

$$\bar{H}_L = H_S \hat{H}_L , \quad \bar{H}_R = \hat{H}_R , \quad \text{where } H_S = \begin{bmatrix} \alpha\delta - \beta\gamma & 0 & 0 \\ 0 & -\gamma & -\delta \\ 0 & \alpha & \beta \end{bmatrix} . \qquad (11.52)$$

   Then

$$F^* = \left(\hat{H}_R\right)^{-\top} F \left(H_S \hat{H}_L\right)^{-1} . \qquad (11.53)$$

3. *A pair of optimal homographies is selected from the class preserving the fundamental matrix* F^*.

Let $\bar{H}_L$, $\bar{H}_R$ be elementary rectifying homographies (or some other rectifying homographies). Homographies H_L, H_R are also rectifying homographies provided they obey equation $H_R F^* H_L^\top = \lambda F^*$, $\lambda \neq 0$, which guarantees that images are kept rectified.

The internal structure of H_L, H_R permits us to understand the meaning of free parameters in the class of rectifying homographies

$$H_L = \begin{bmatrix} l_1 & l_2 & l_3 \\ 0 & s & u_0 \\ 0 & q & 1 \end{bmatrix} \bar{H}_L \,, \qquad H_R = \begin{bmatrix} r_1 & r_2 & r_3 \\ 0 & s & u_0 \\ 0 & q & 1 \end{bmatrix} \bar{H}_R \,, \tag{11.54}$$

where $s \neq 0$ is a common vertical scale; u_0 is a common vertical shift; l_1, r_1 are left and right skews; l_2, r_2 are left and right horizontal scales; l_1, r_1 are left and right horizontal shifts and q is common perspective distortion.

This third step is necessary because elementary homographies may yield severely distorted images.

Summary

- Epipolar geometry
- Fundamental matrix estimation
  - Normalized 8-pont algorithm