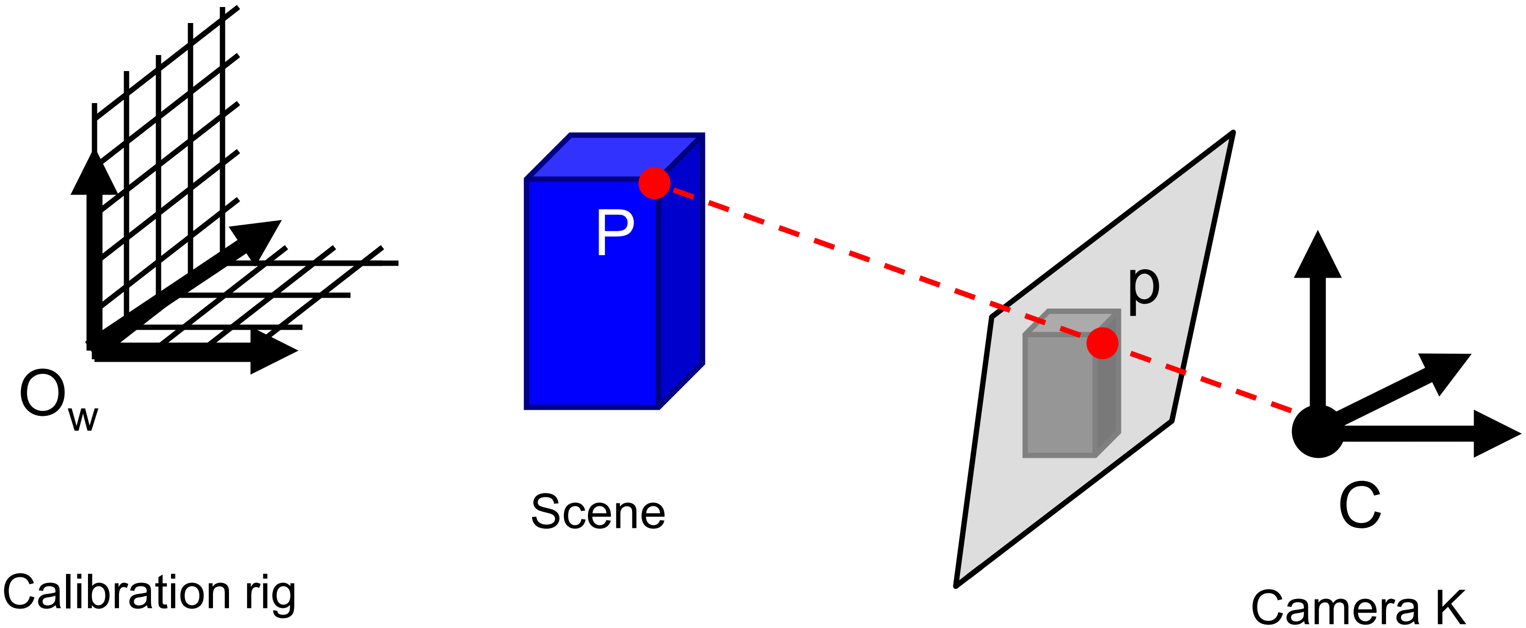
EPIPOLAR GEOMETRY

The slides are from several sources through James Hays (Brown);
Silvio Savarese (U. of Michigan); Svetlana Lazebnik (U. Illinois);
Bill Freeman and Antonio Torralba (MIT), including their own slides.

Recovering structure from a single view



Why is it so difficult?

Intrinsic ambiguity of the mapping from 3D to image in 2D.

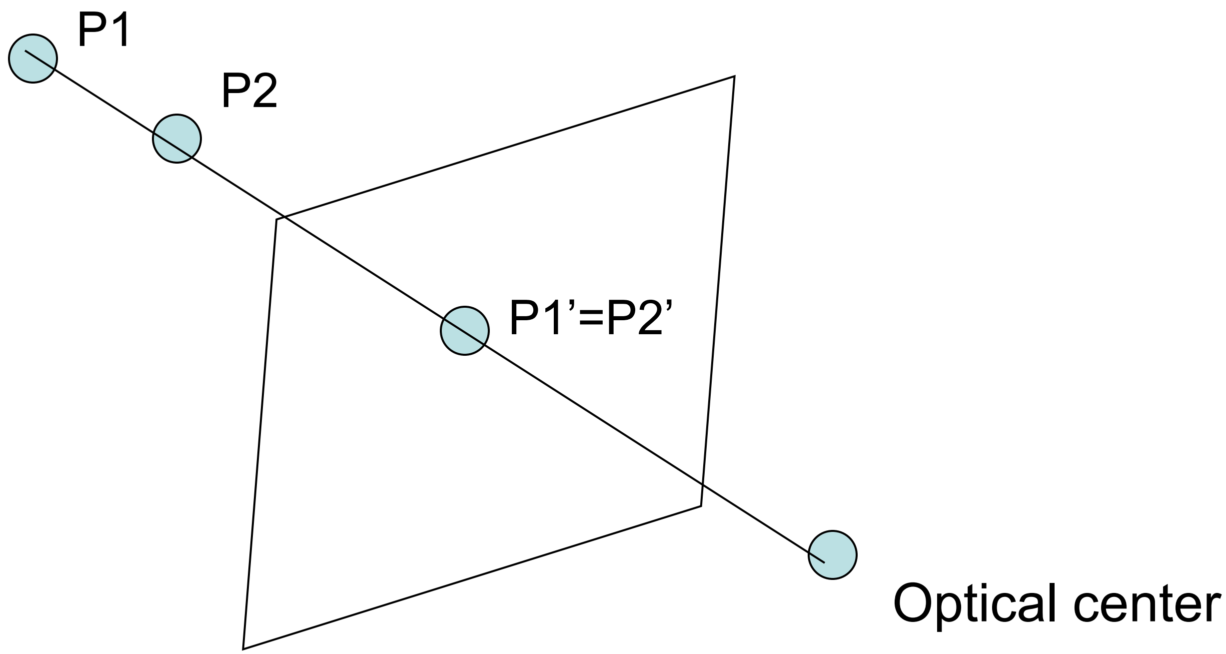
see example...

Is this an illusion of 3D to 2D?

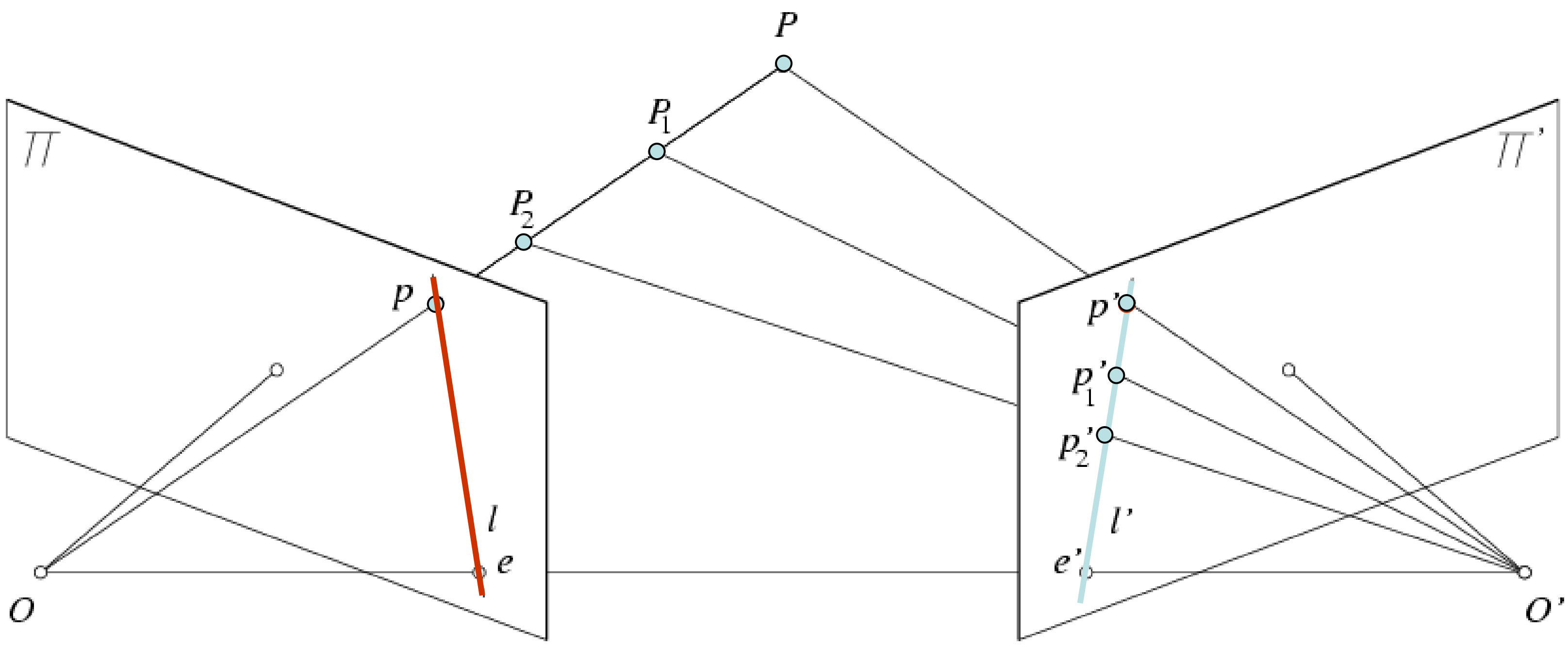
Courtesy slide S. Lazebnik

Why multiple views?

- Structure and depth are inherently ambiguous from single views.



Epipolar constraint

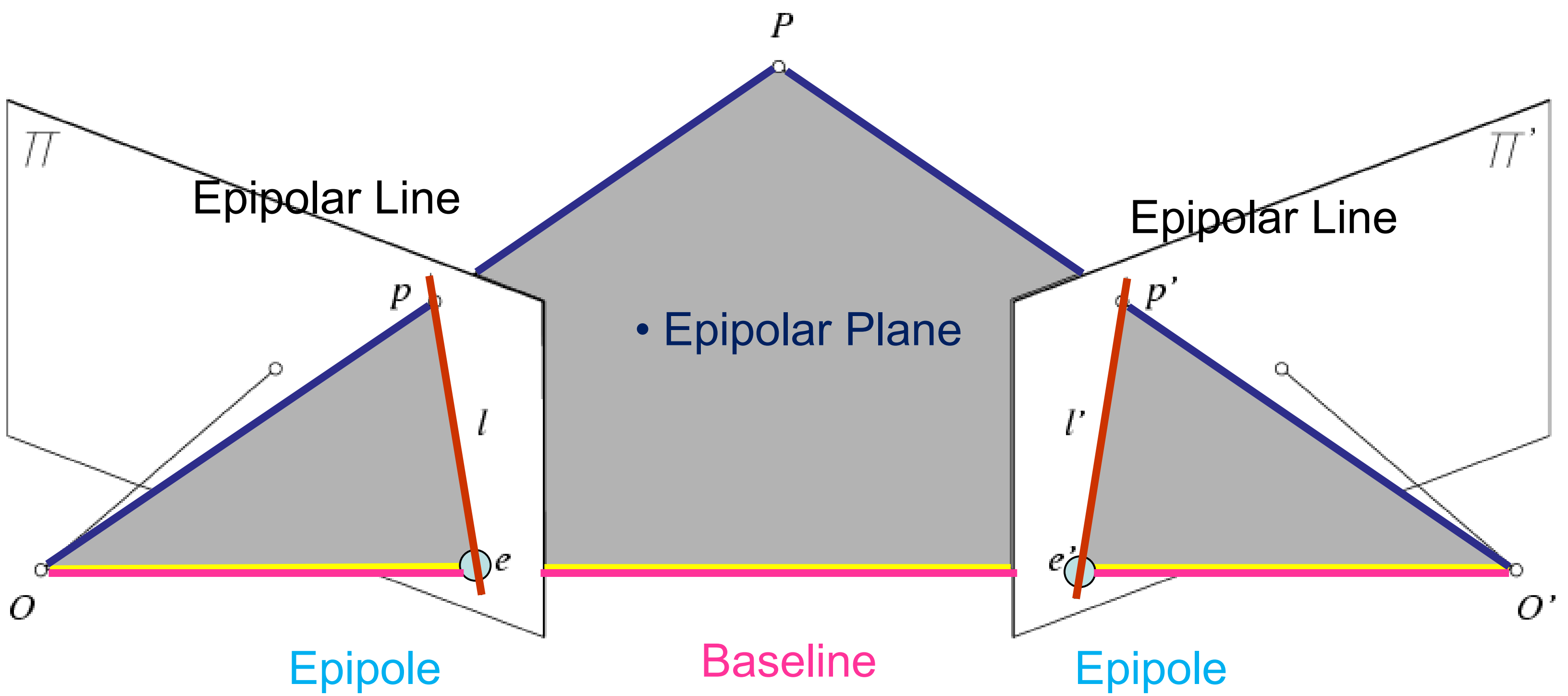


Geometry of two views constrains where the corresponding pixel for some image point in the first view *must occur*

in the second view.

It must be *on the line* carved out by a plane connecting the world point and the optical centers.

Epipolar geometry



Epipolar geometry: terms

- **Baseline**: line joining the camera centers.
- **Epipole**: point of intersection of baseline with image plane.
- **Epipolar plane**: plane containing baseline and world point.
- **Epipolar line**: intersection of epipolar plane with the image plane.

- All epipolar lines in an image intersect at the epipole... or, an epipolar plane intersects the left and right image planes in epipolar lines.

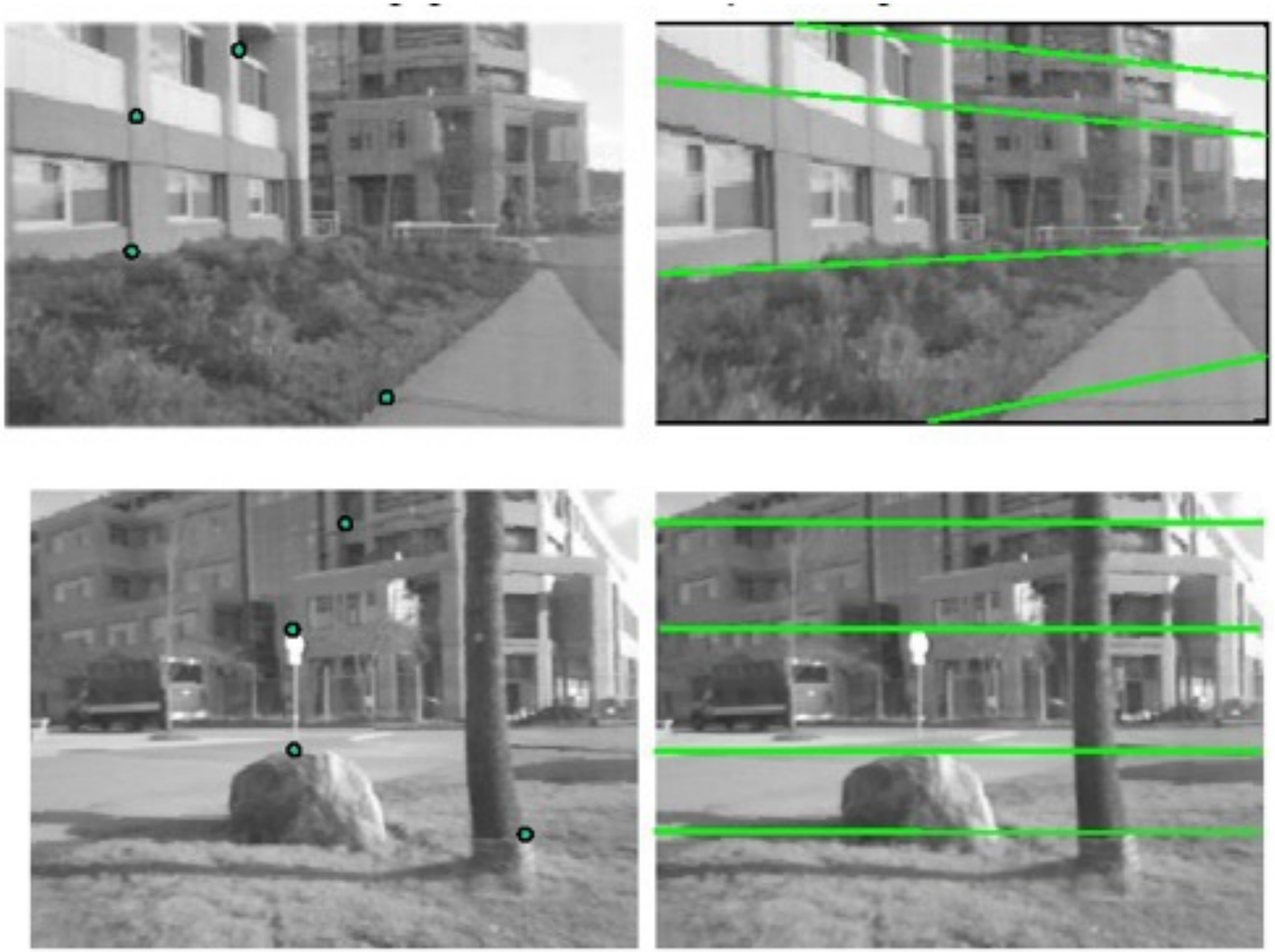
Why is the epipolar constraint useful?

Epipolar constraint

Reduces the correspondence problem to a 1D search in the second image along an epipolar line.

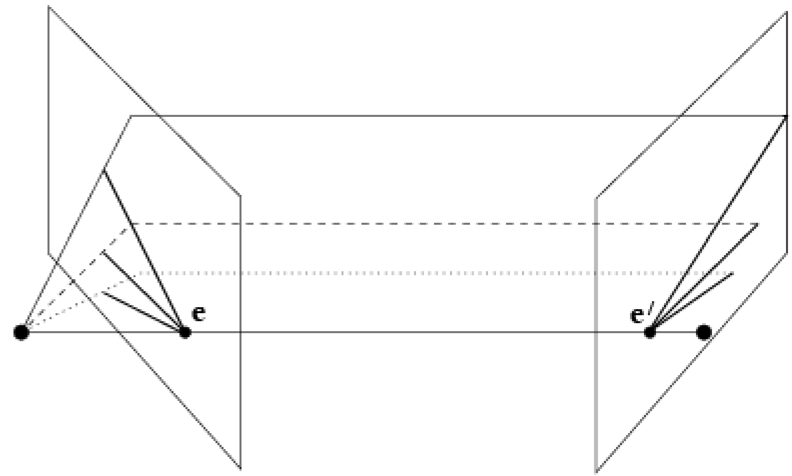
Image from Andrew Zisserman

Two examples:



Slide credit: Kristen Grauman

Converging cameras have finite epipoles.



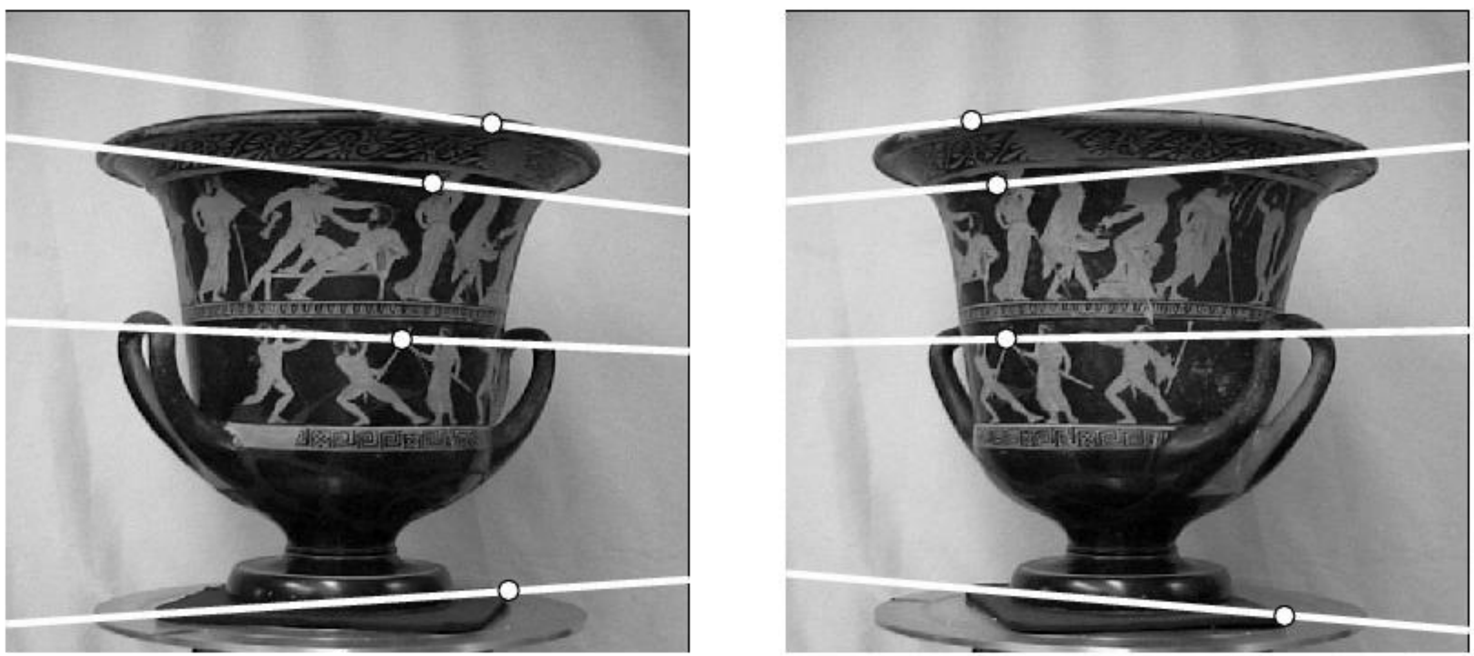
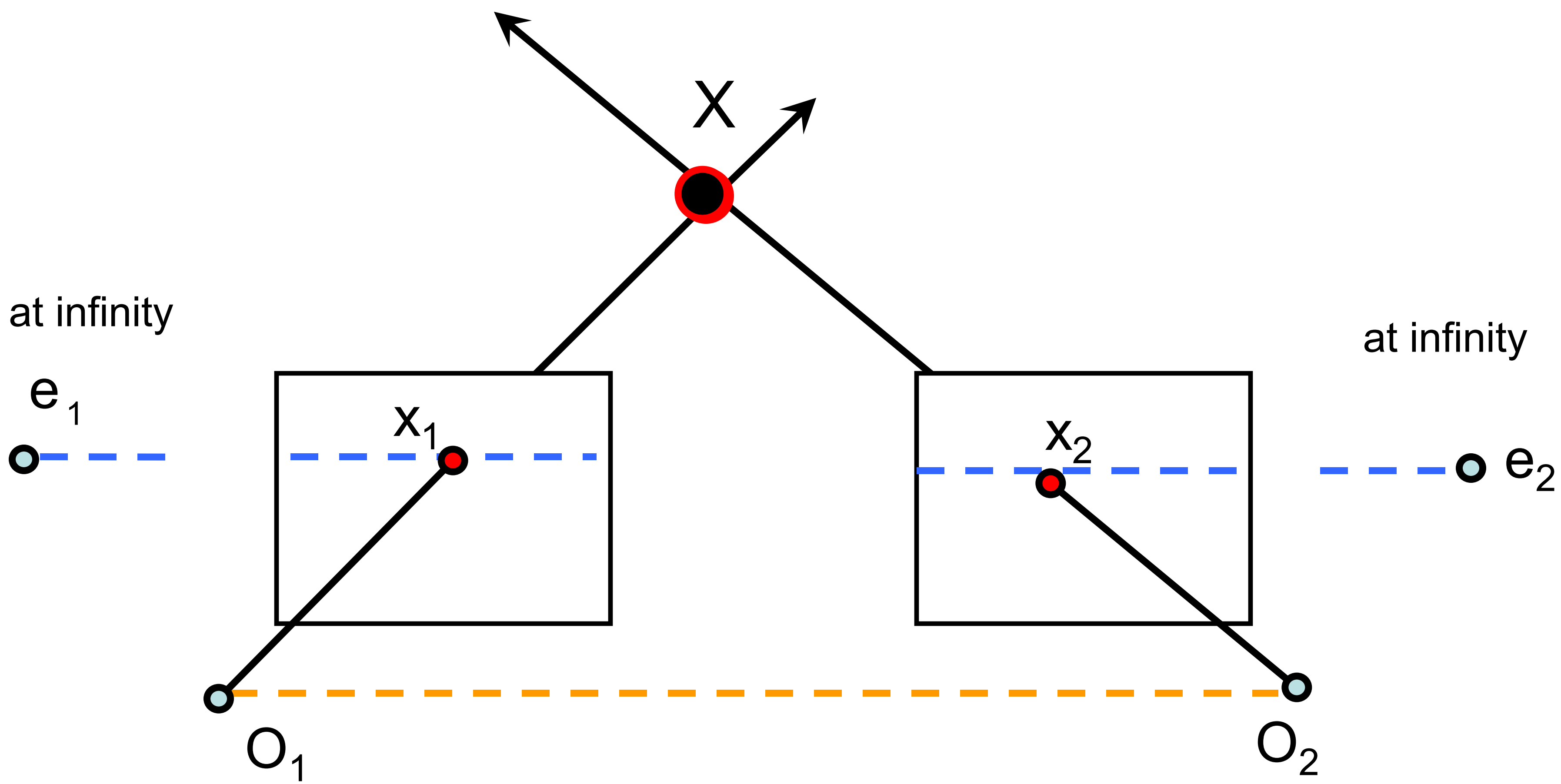


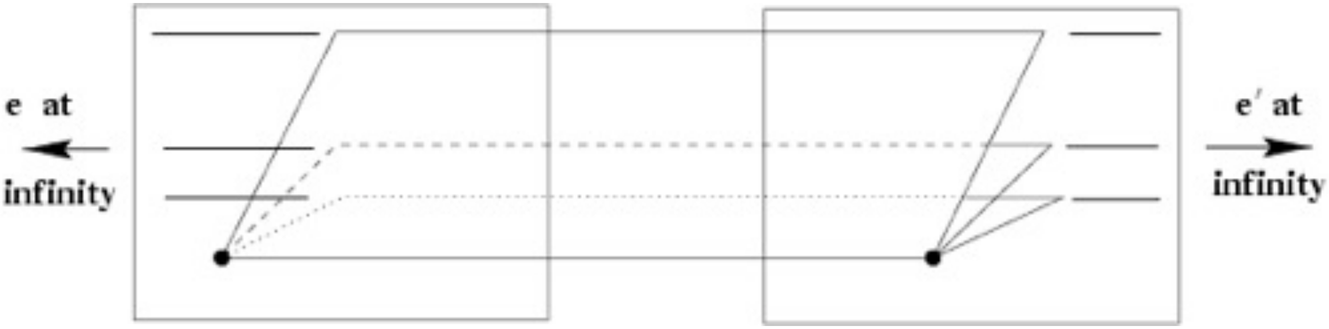
Figure from Hartley & Zisserman

Parallel cameras have epipoles at infinity.



- Baseline intersects the image plane at infinity.
- *Epipoles are at infinity.*
- Epipolar lines are parallel to x axis.

In parallel cameras search is only along x coord.



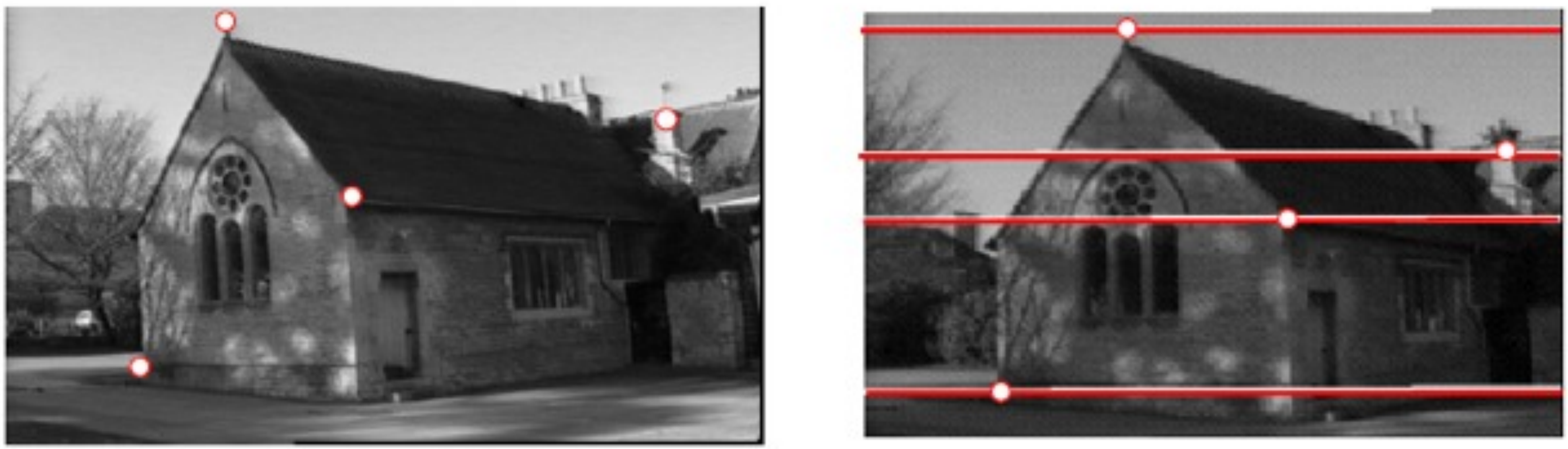


Figure from Hartley & Zisserman

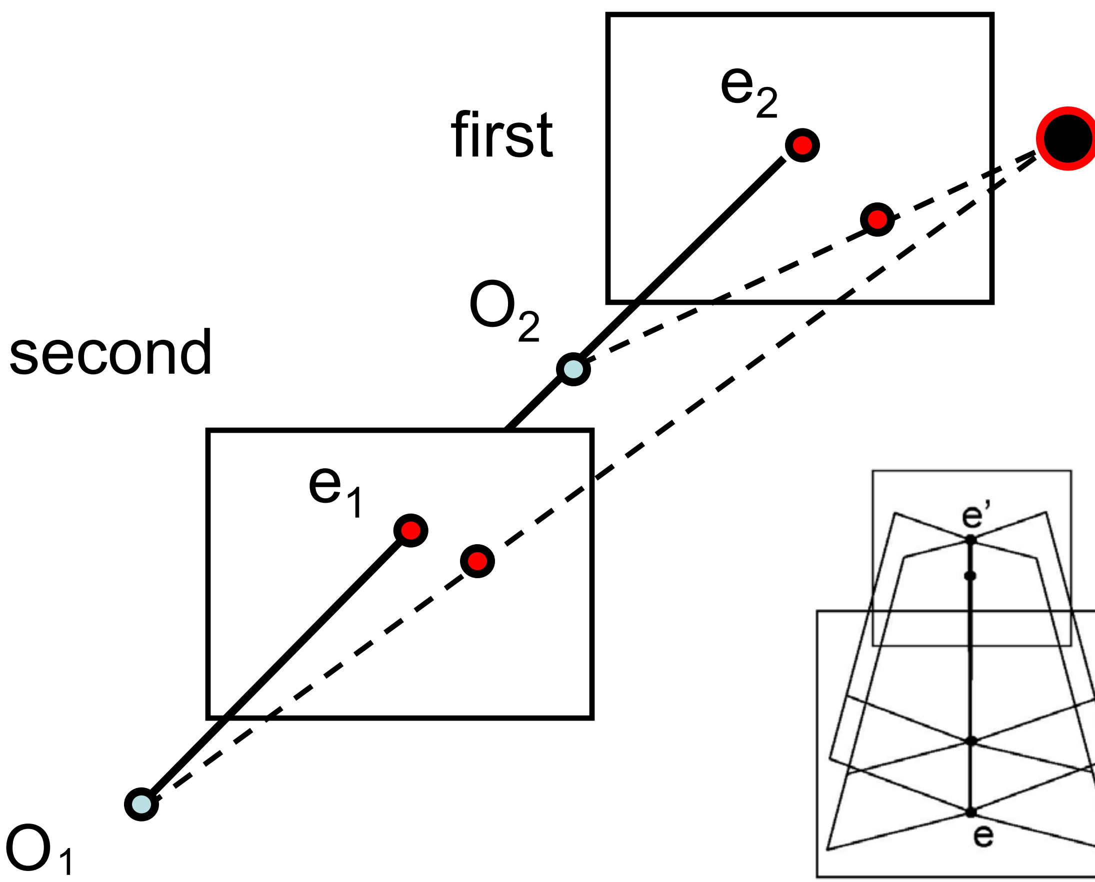
Slide credit: Kristen Grauman

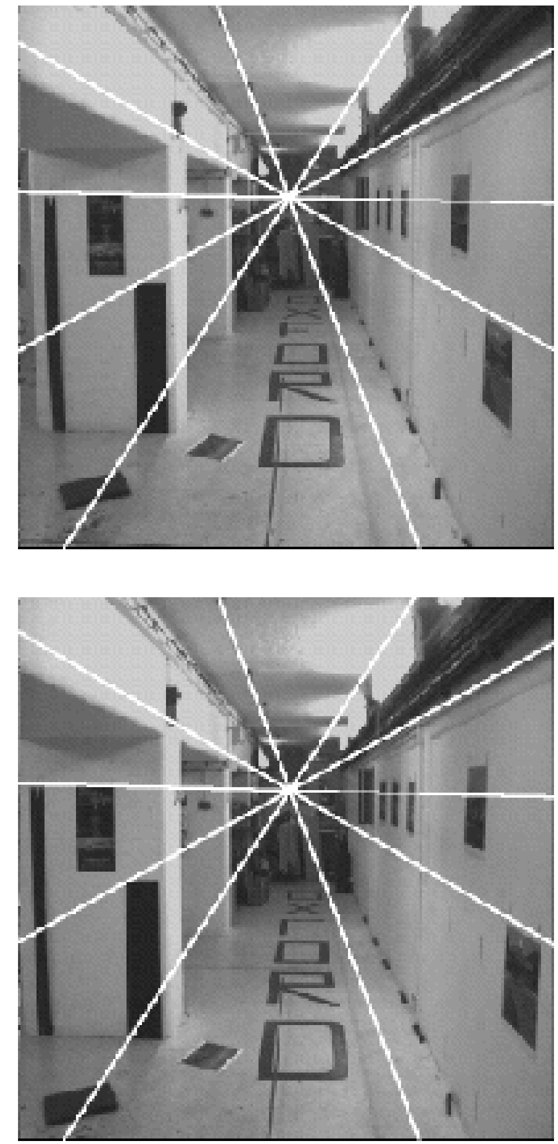
Motion perpendicular to image plane

Motion perpendicular to image plane

forward

Forward translation





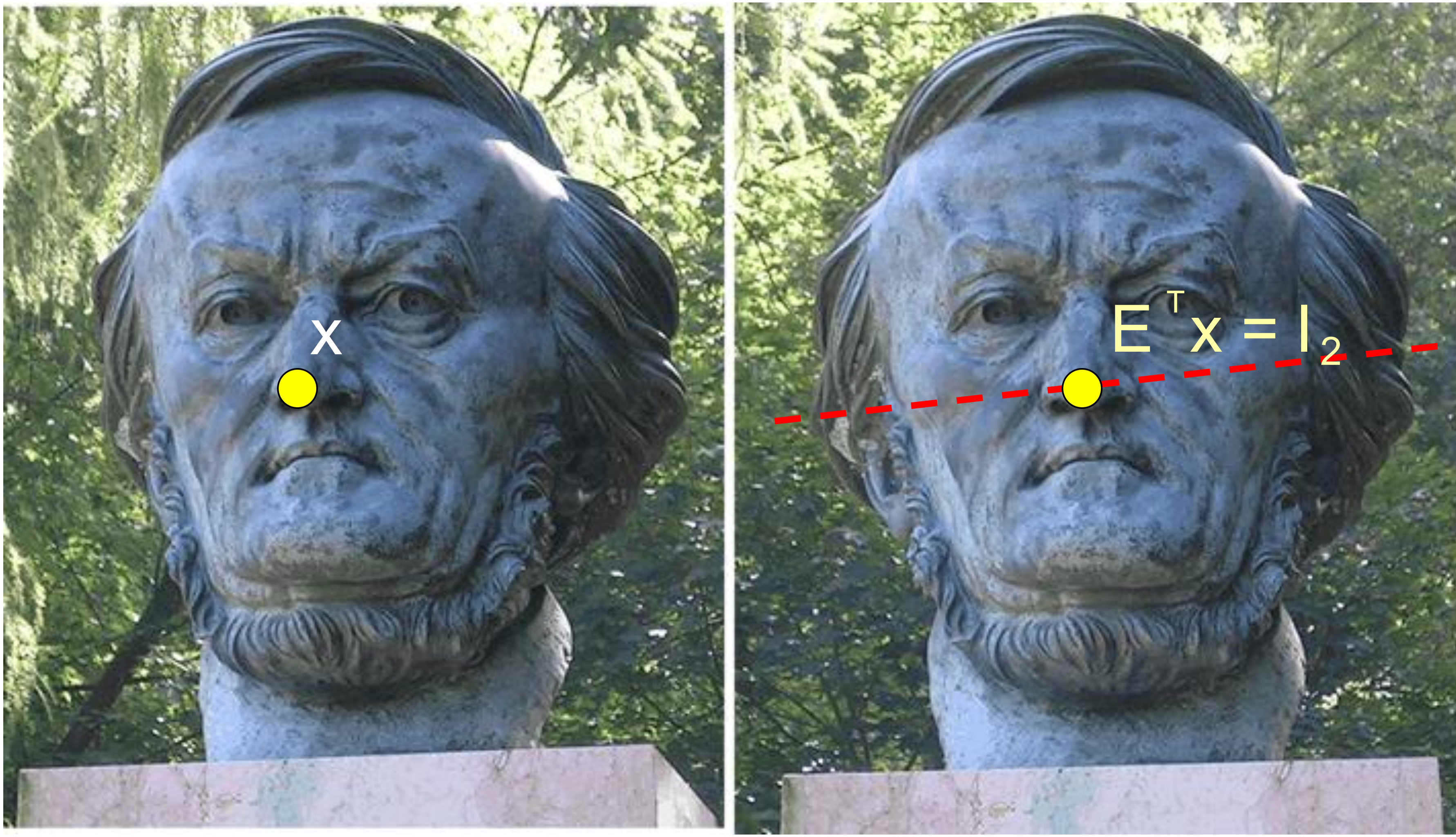
- The epipoles have same position in both images.
- Epipole here is called FOE (focus of expansion).

A 3x3 matrix connects the two 2D images.

This matrix is called

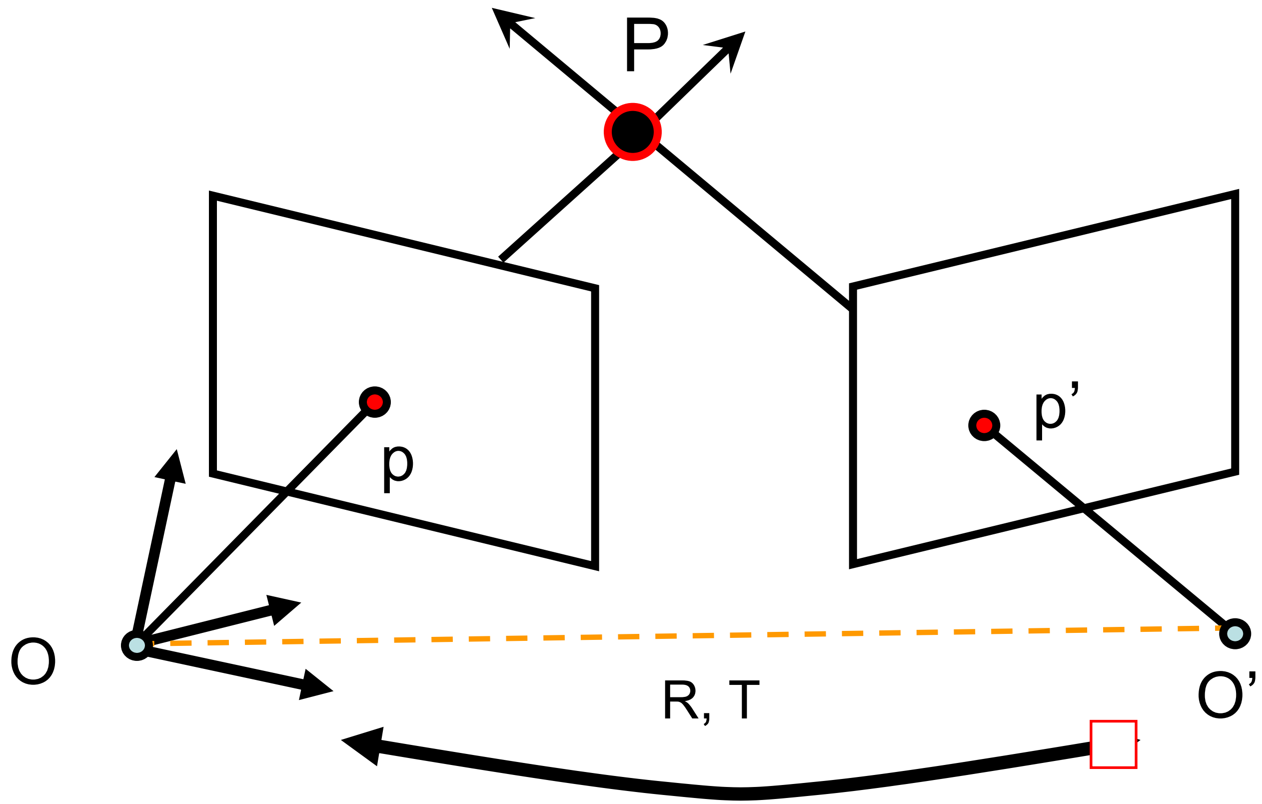
- the “Essential Matrix”, E
  - when image *intrinsic parameters* are known
- the “Fundamental Matrix”, F
  - more the general uncalibrated case

Essential matrix: E



- Two views of the same object
- Suppose we know the camera positions and camera matrices ==> E matrix
- Given a point on left image, how can we find the corresponding point on right image?

Epipolar Constraint - E matrix

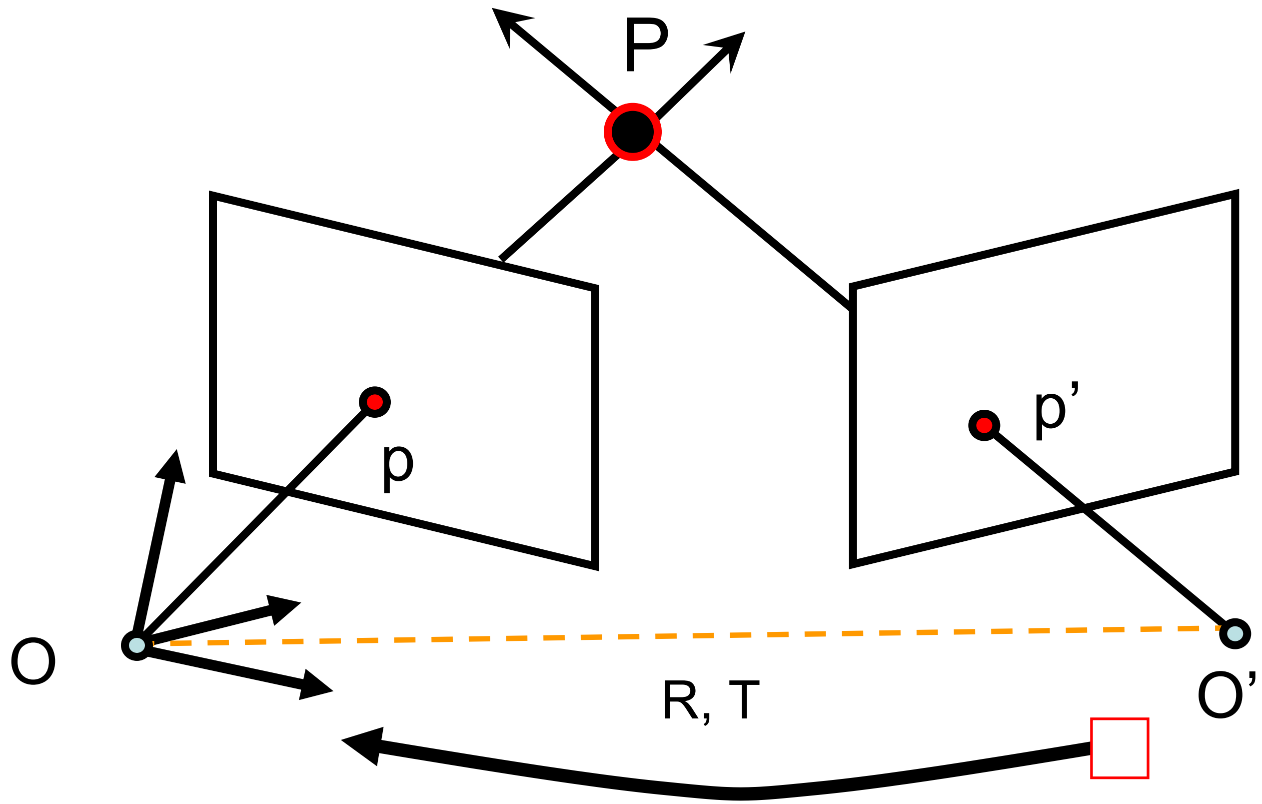


$$M = K\begin{bmatrix} I & 0 \end{bmatrix} \qquad M' = K'\begin{bmatrix} R & T \end{bmatrix}$$

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

homogeous coordinates

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



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

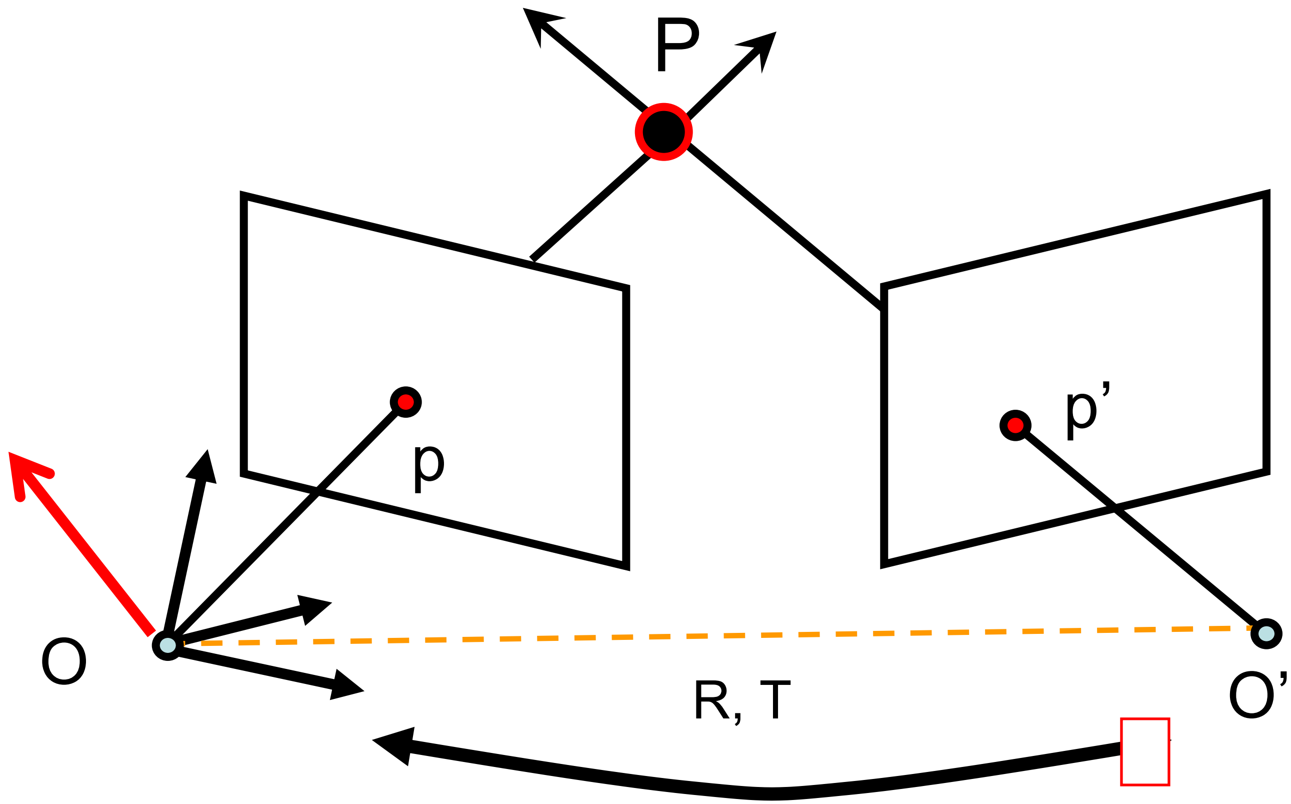
K and K' are known
(calibrated cameras)

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

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

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

In the epipolar plane we have



first camera coordinates

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

Perpendicular to epipolar plane

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

Cross products can be written as matrix multiplication.

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

...verify it

The matrix derived from **a** is skew-symmetric.

The matrix is rank 2. The null vector is along the vector **a**.

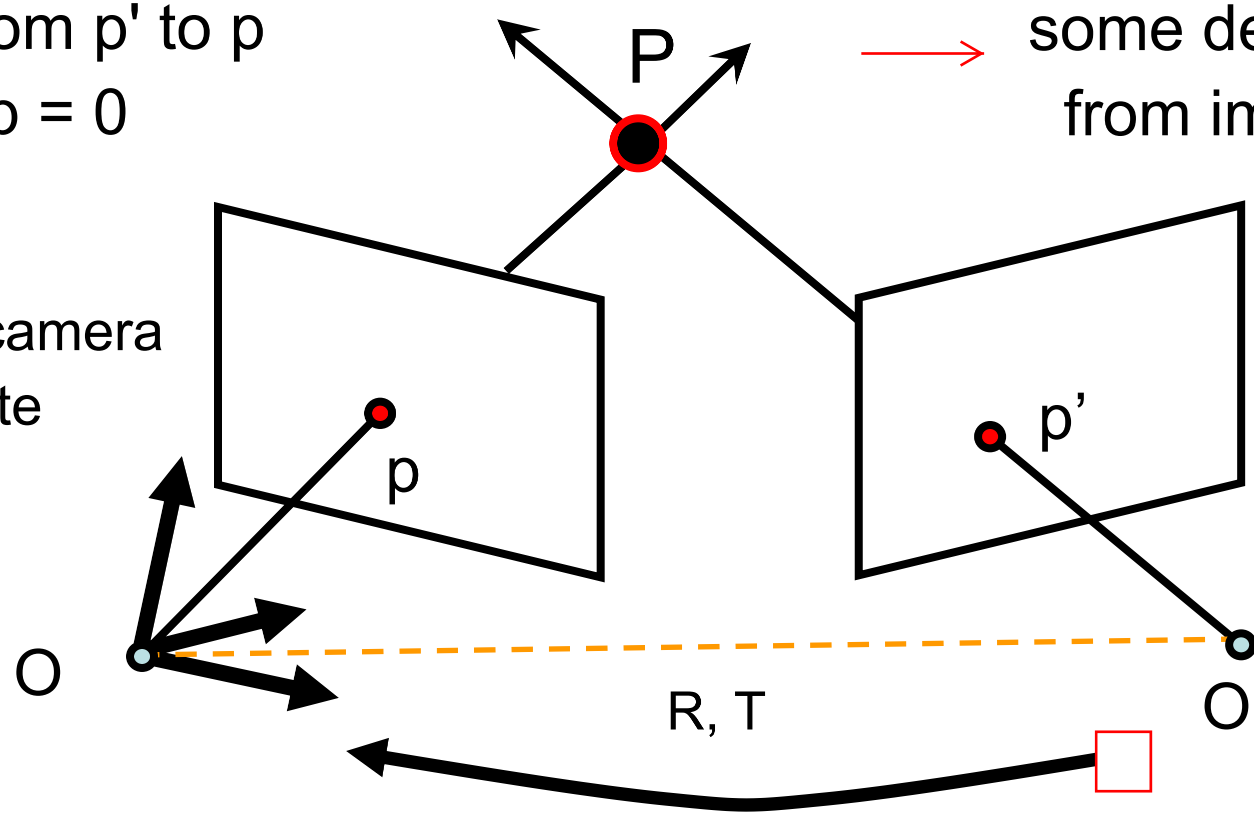
Essential matrix

E^T is from p' to p
$p'^T E^T p = 0$

second camera coordinate

some denote E as from image 2 to 1

then R and t are from image 1 to 2



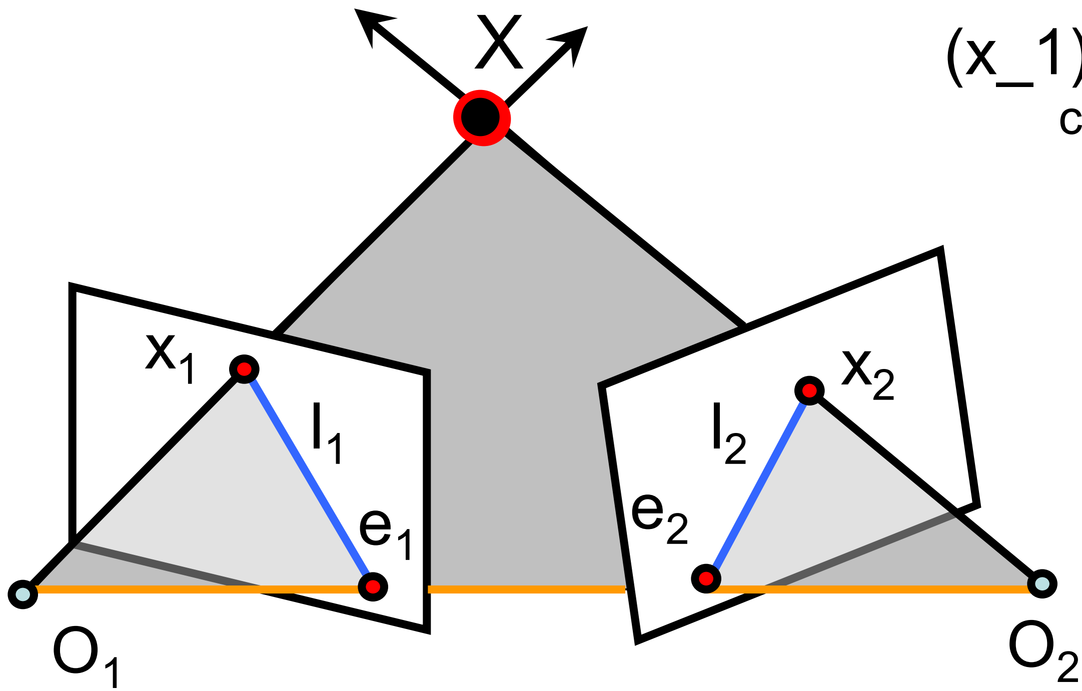
$$p^T \cdot [T \times (R\,p')] = 0 \rightarrow p^T \cdot \boxed{[T]_x R}\, p' = 0$$

E = essential matrix
(Longuet-Higgins, 1981)

E is a rank 2 matrix!

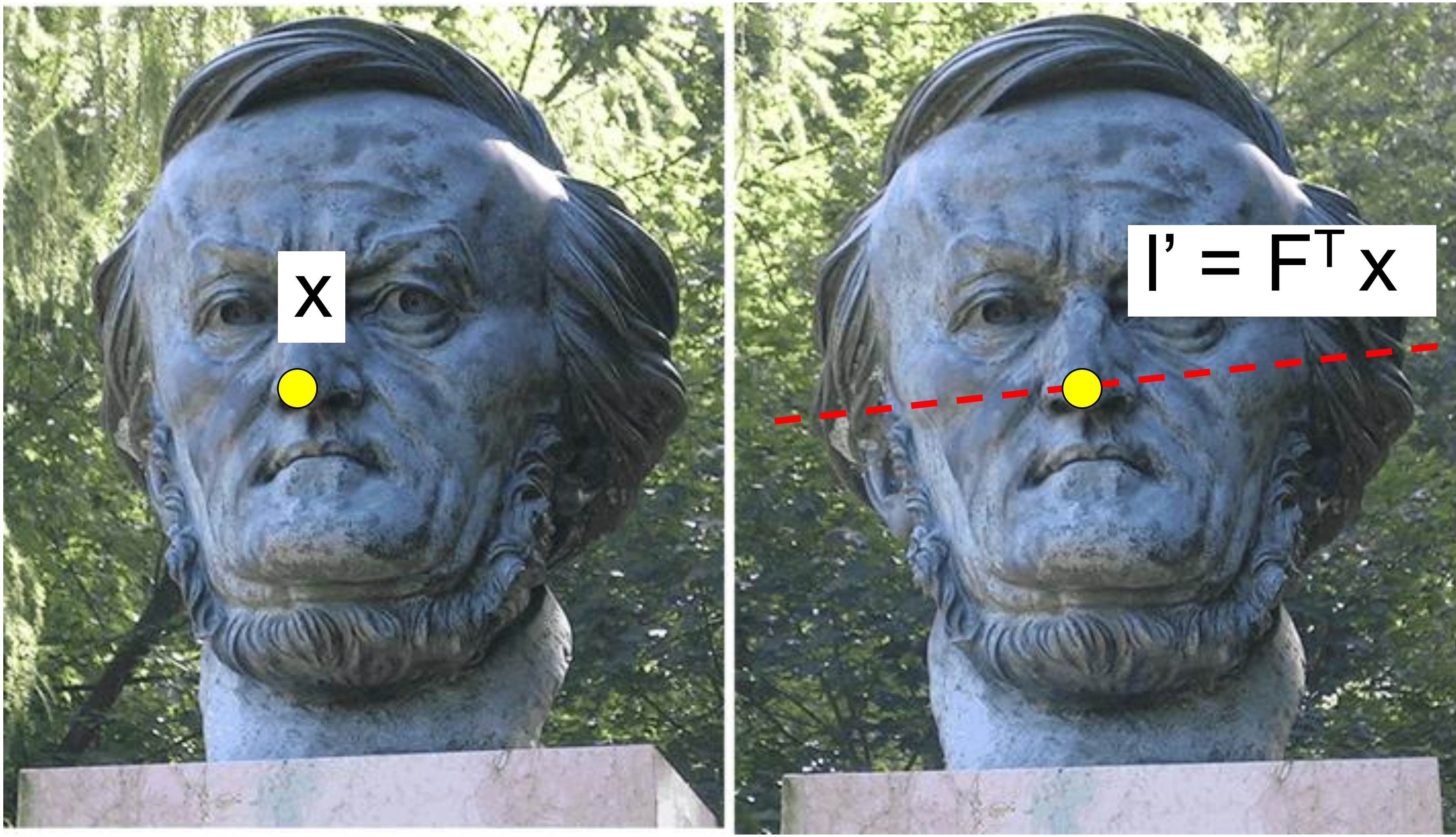
Essential matrix properties

$(x\_1)^T E\ x\_2 = 0$
calibrated



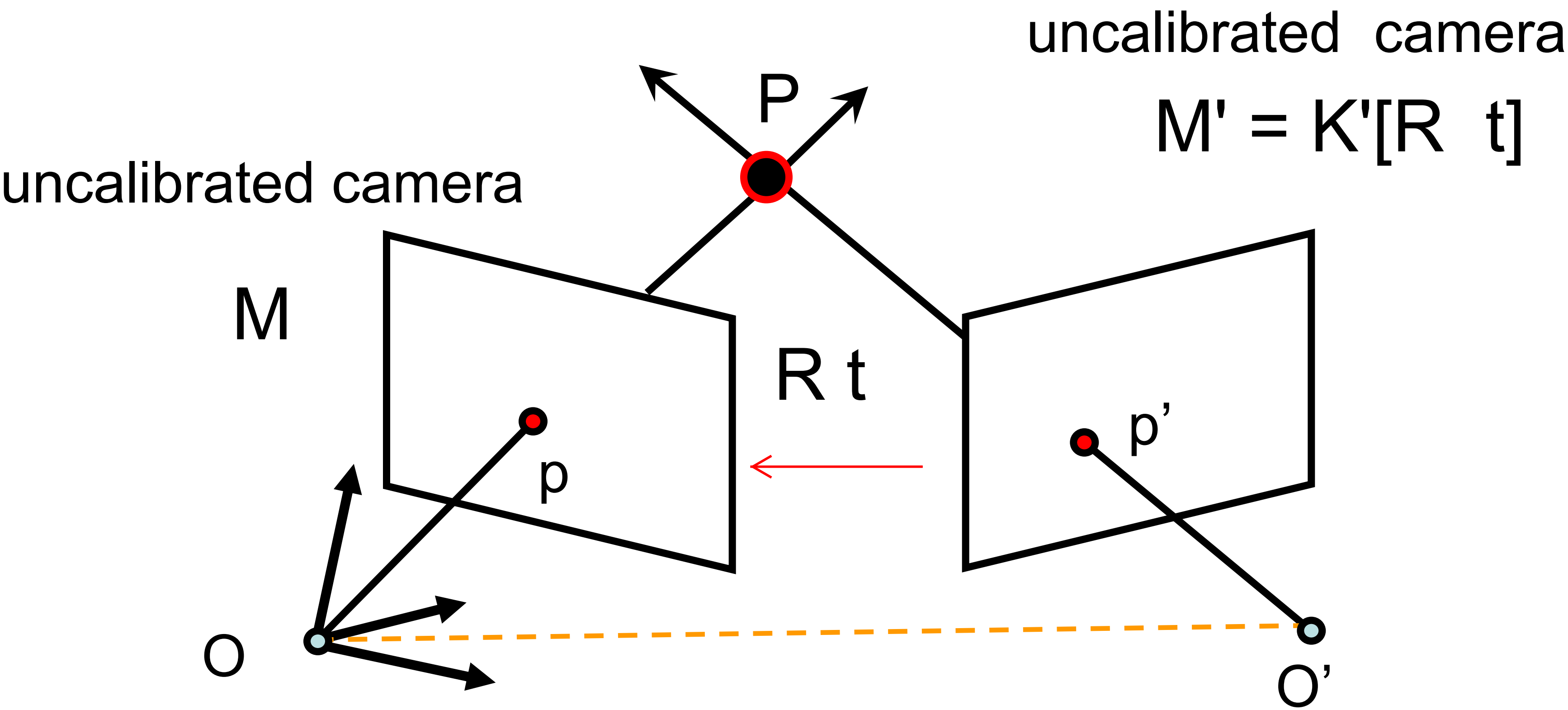
- $E\,x_2$ is the epipolar line associated with x_2 ($l_1 = E\,x_2$)
- $E^T x_1$ is the epipolar line associated with x_1 ($l_2 = E^T x_1$)
- E is *singular (rank two)* -- two equal singular values are one.
- $E\,e_2 = 0$ and $E^T e_1 = 0$ $\quad l_1^T e_1 = (E\,x_2)^T e_1 = 0$ valid for any x_2
- E is 3x3 matrix with 5 DOF: 3(R) + 3(t) -1(scale)

Fundamental matrix: F



- Uncalibrated cameras.
- No additional information about the scene and camera is given ==> F matrix
- Given a point on left image, how can I find the corresponding point on right image?

Epipolar Constraint - F matrix



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

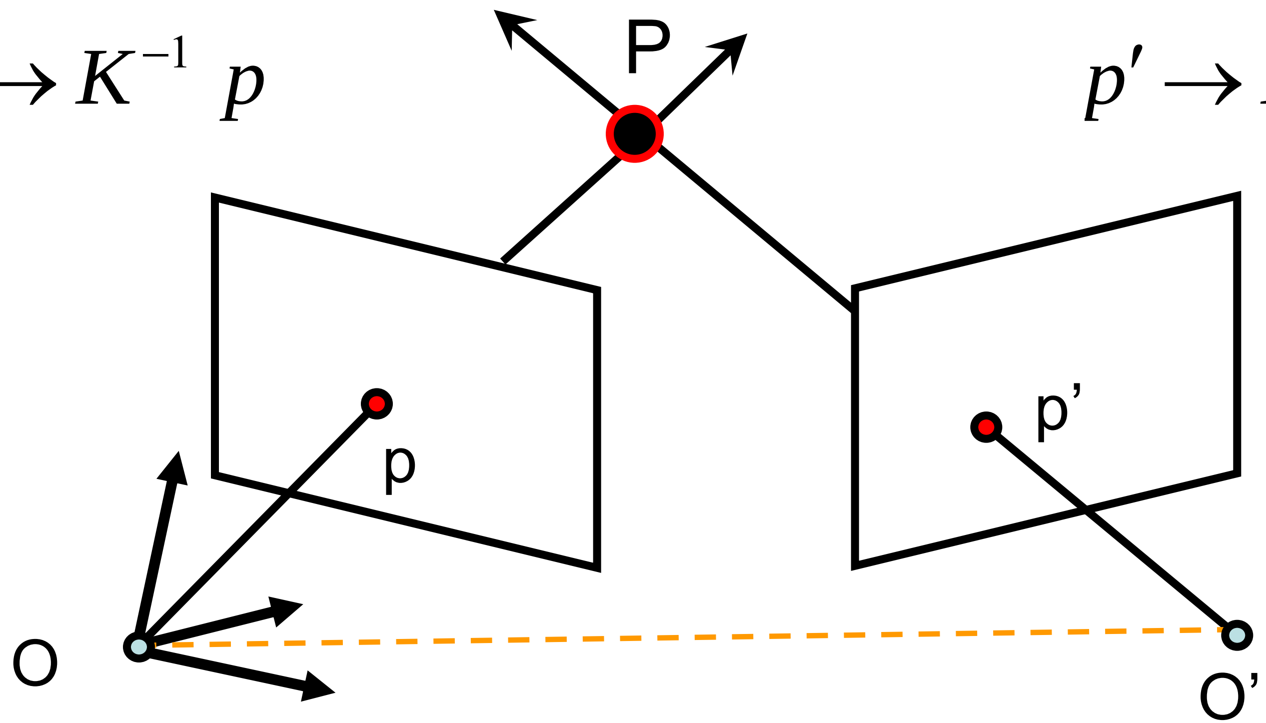
1 homogeous coord.

$$M = \boxed{K}\begin{bmatrix} I & 0 \end{bmatrix}$$ (3x4)

unknown

F matrix derived from E matrix

$$p \rightarrow K^{-1}\ p \qquad\qquad p' \rightarrow K'^{-1}\ p'$$



$[T]_x$!!

$$p^T \cdot \left[T_\times\right] \cdot R\, p' = 0 \rightarrow (K^{-1}\, p)^T \cdot \left[T_\times\right] \cdot R\, K'^{-1}\, p' = 0$$

$$p^T \boxed{K^{-T} \cdot \left[T_\times\right] \cdot R\, K'^{-1}}\, p' = 0 \rightarrow p^T\, \boxed{F}\, p' = 0$$

rank 2

Fundamental matrix

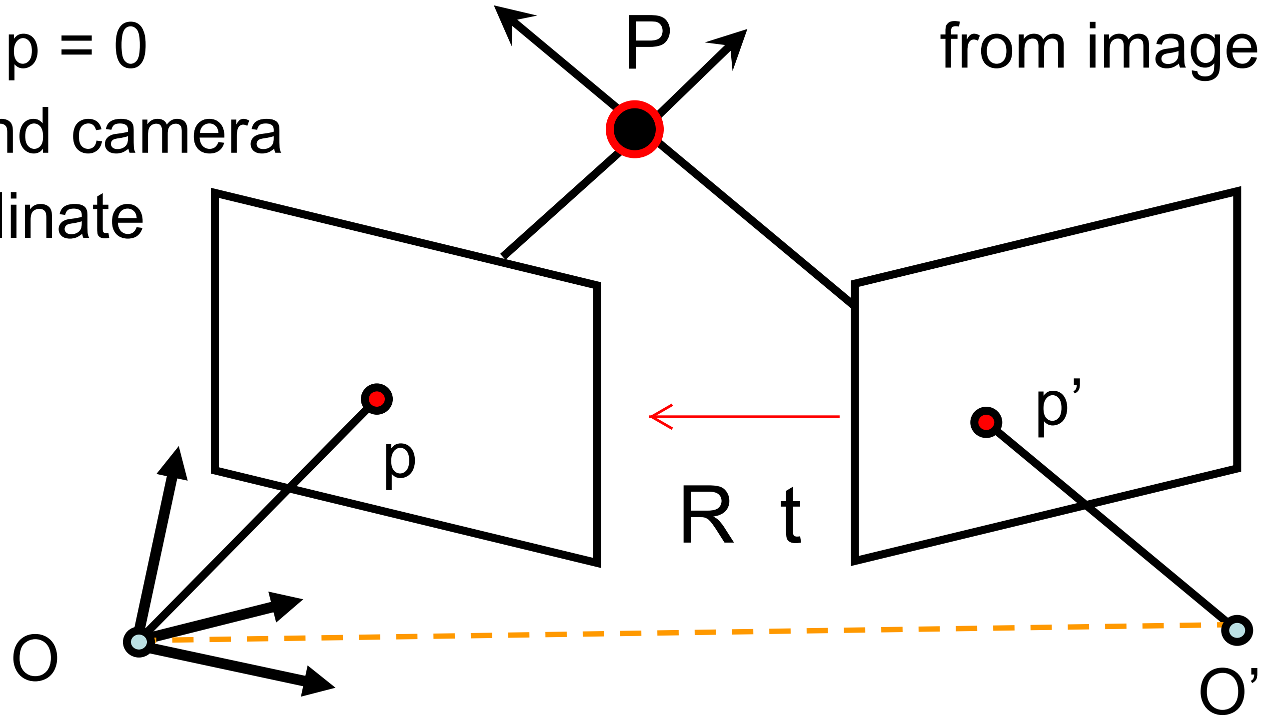
F^T is from p' to p

$p'^T F^T p = 0$

second camera coordinate

some denote F as from image 2 to 1

then R and t are from image 1 to 2



$$\boxed{p^T F p' = 0}$$ (Faugeras and Luong, 1992)

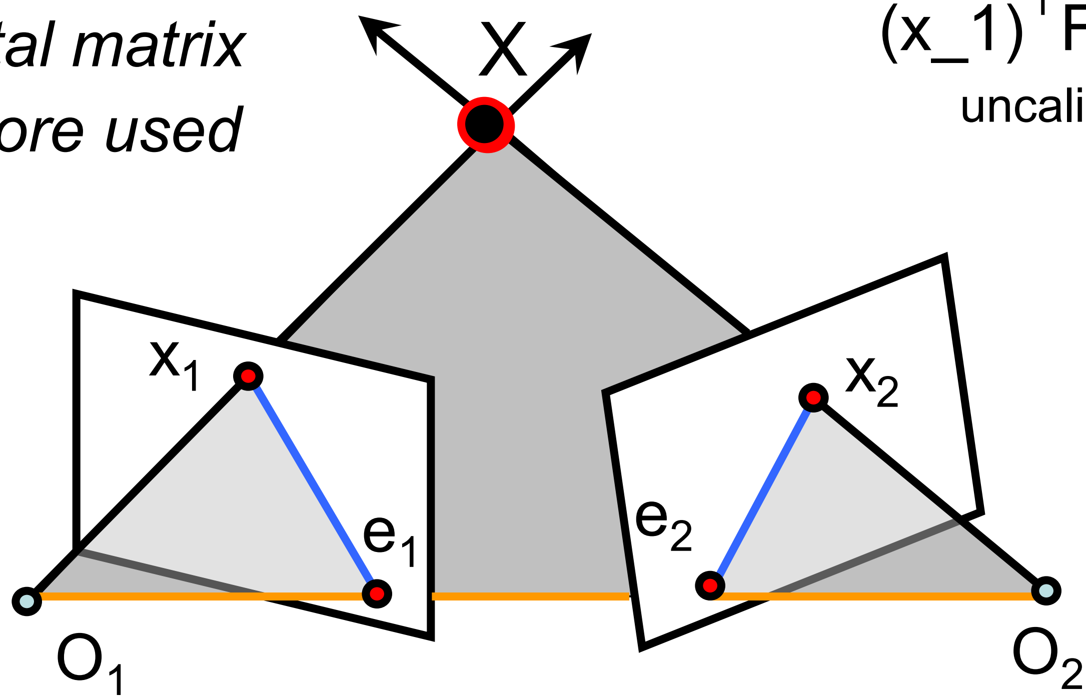
The fundamental matrix has a projective ambiguity. Two pairs of camera matrices (**P**, **P'**) and (**P̃**, **P̃'**) give the same **F** if **P̃** = **PH** and **P̃'** = **P'H** where **H** is a 4x4 nonsingular matrix.

Fundamental matrix properties

fundamental matrix is much more used

(x_1)^{T}F x_2 = 0

uncalibrated



- $F x_2$ is the epipolar line associated with x_2 ($l_1 = F x_2$)
- $F^T x_1$ is the epipolar line associated with x_1 ($l_2 = F^T x_1$)
- F is *singular (rank two)*
- $F e_2 = 0$ and $F^T e_1 = 0$
- F is 3x3 matrix with 7 DOF: 9 - 1(rank 2) - 1(scale)

(Hartley, 1995)

The eight-point algorithm of F (linear)

$\boldsymbol{x} = (u, v, 1)^T, \quad \boldsymbol{x'} = (u', v', 1)^T$

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

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

$$\begin{pmatrix} u_1u_1' & u_1v_1' & u_1 & v_1u_1' & v_1v_1' & v_1 & u_1' & v_1' \\ u_2u_2' & u_2v_2' & u_2 & v_2u_2' & v_2v_2' & v_2 & u_2' & v_2' \\ u_3u_3' & u_3v_3' & u_3 & v_3u_3' & v_3v_3' & v_3 & u_3' & v_3' \\ u_4u_4' & u_4v_4' & u_4 & v_4u_4' & v_4v_4' & v_4 & u_4' & v_4' \\ u_5u_5' & u_5v_5' & u_5 & v_5u_5' & v_5v_5' & v_5 & u_5' & v_5' \\ u_6u_6' & u_6v_6' & u_6 & v_6u_6' & v_6v_6' & v_6 & u_6' & v_6' \\ u_7u_7' & u_7v_7' & u_7 & v_7u_7' & v_7v_7' & v_7 & u_7' & v_7' \\ u_8u_8' & u_8v_8' & u_8 & v_8u_8' & v_8v_8' & v_8 & u_8' & v_8' \end{pmatrix}\begin{pmatrix} F_{11} \\ F_{12} \\ F_{13} \\ F_{21} \\ F_{22} \\ F_{23} \\ F_{31} \\ F_{32} \end{pmatrix} = -\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

Minimize:

$$\sum_{i=1}^{N} (x_i^T F x_i')^2$$

under the constraint

$$F_{33} = 1$$

can be an other F_ij constraint also

Estimating F

$$\underbrace{\begin{pmatrix} u_1u'_1 & u_1v'_1 & u_1 & v_1u'_1 & v_1v'_1 & v_1 & u'_1 & v'_1 & 1 \\ u_2u'_2 & u_2v'_2 & u_2 & v_2u'_2 & v_2v'_2 & v_2 & u'_2 & v'_2 & 1 \\ u_3u'_3 & u_3v'_3 & u_3 & v_3u'_3 & v_3v'_3 & v_3 & u'_3 & v'_3 & 1 \\ u_4u'_4 & u_4v'_4 & u_4 & v_4u'_4 & v_4v'_4 & v_4 & u'_4 & v'_4 & 1 \\ u_5u'_5 & u_5v'_5 & u_5 & v_5u'_5 & v_5v'_5 & v_5 & u'_5 & v'_5 & 1 \\ u_6u'_6 & u_6v'_6 & u_6 & v_6u'_6 & v_6v'_6 & v_6 & u'_6 & v'_6 & 1 \\ u_7u'_7 & u_7v'_7 & u_7 & v_7u'_7 & v_7v'_7 & v_7 & u'_7 & v'_7 & 1 \\ u_8u'_8 & u_8v'_8 & u_8 & v_8u'_8 & v_8v'_8 & v_8 & u'_8 & v'_8 & 1 \end{pmatrix}}_{\mathbf{W}} \underbrace{\begin{pmatrix} F_{11} \\ F_{12} \\ F_{13} \\ F_{21} \\ F_{22} \\ F_{23} \\ F_{31} \\ F_{32} \\ F_{33} \end{pmatrix}}_{\mathbf{f}} = 0$$

- Homogeneous system $\mathbf{W}\mathbf{f} = 0$
- Rank 8 $\longrightarrow$ A non-zero solution exists (unique)
- If N>8 $\longrightarrow$ Lsq. solution by SVD $\longrightarrow$ $\hat{\mathrm{F}}$

$\|\mathbf{f}\| = 1$

rank 3 solution

Taking into account the rank-2 constraint.

$$p^T \hat{F} p' = 0$$

The estimated $\hat{F}$ have full rank (det(F) $\neq \hat{0}$)
but F should have rank=2 instead.

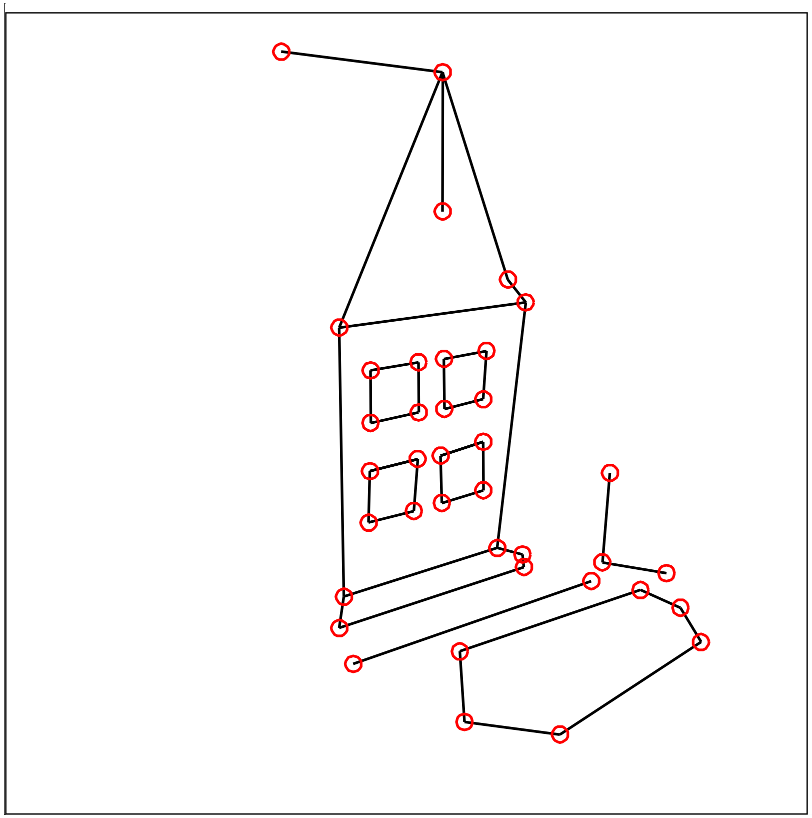
Find F that minimizes $\left\| F - \hat{F} \right\| = 0$

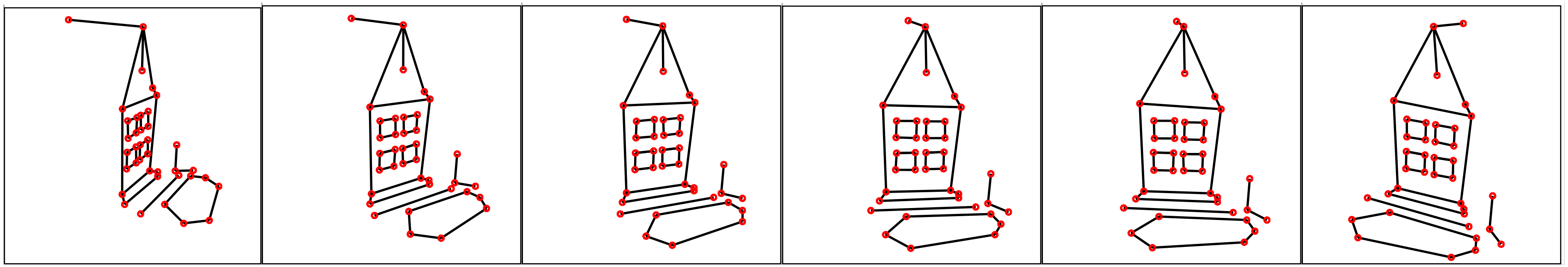
Frobenius norm (*)

subject to det(F)=0

Taking the first two s.v. and the three equal zero.

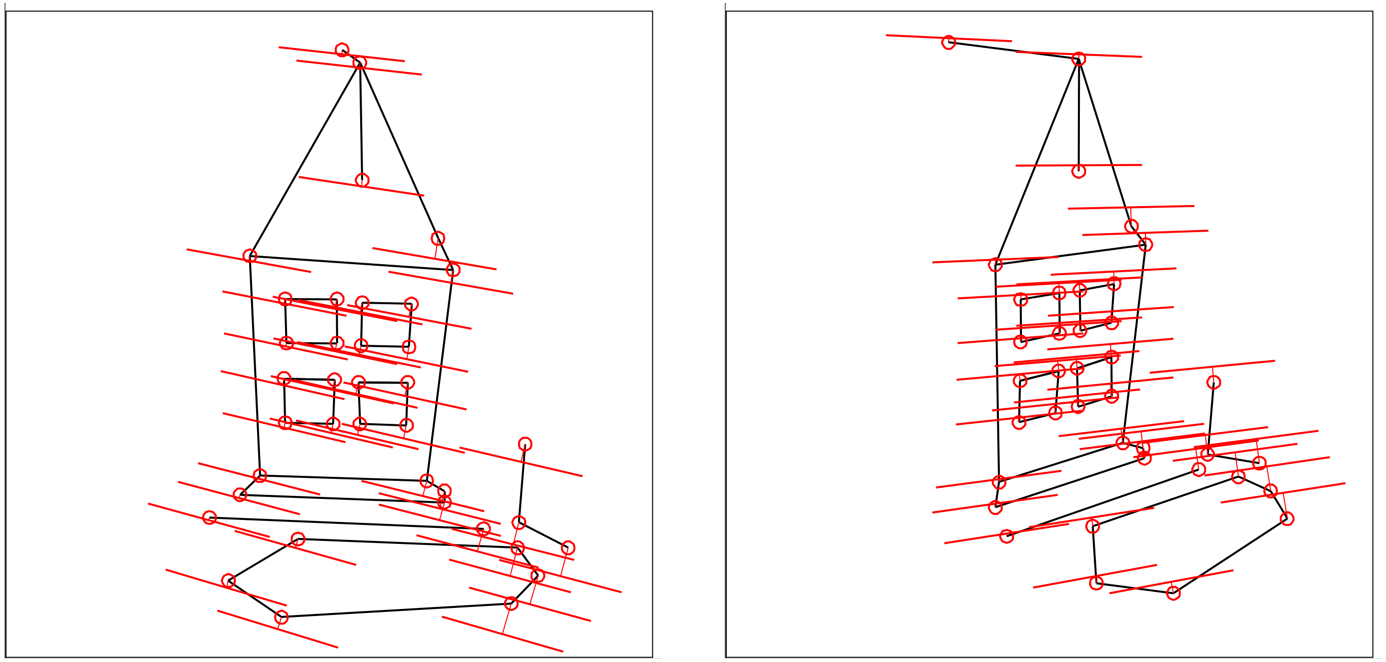
(*) Sqrt root of the sum of squares of all entries





Example of F recovery

Data courtesy of R. Mohr and B. Boufama.



Mean errors:
10.0pixel
and 9.1pixel

This are large errors...

The problem with eight-point algorithm

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

$$\begin{pmatrix} F_{11} \\ F_{12} \\ F_{13} \\ F_{21} \\ F_{22} \\ F_{23} \\ F_{31} \\ F_{32} \end{pmatrix} = - \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

Poor numerical conditioning.

Can be fixed by rescaling the data before estimation.

Can be used for any DLT type algorithm.

More sophisticated nonlinear methods after the 8-point algorithm exist, but we will not cover.

RESCALING BY NORMALIZATION

You have $i = 1, \ldots, n$ points $\mathbf{x}_i$. The mean of these points is

$$\bar{\mathbf{x}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i$$

which is translated to the origin $(0,\ 0)$ by the vector $-\bar{\mathbf{x}}$. The new coordinated of a point are $\tilde{\mathbf{x}}_i = \mathbf{x}_i - \bar{\mathbf{x}}$.

Compute the mean squared distance of the points from the center

$$a^2 = \frac{1}{n}\sum_{i=1}^{n}\tilde{\mathbf{x}}_i^\top\tilde{\mathbf{x}}_i$$

and move the square norm equal to 2 by multiplying the components of the original point $\sqrt{2}/a$. This is the scaling. The translation are the mean coordinates with opposite sign multiplied with $\sqrt{2}/a$.

In the homogeneous 2D coordinates

$$\mathbf{T} = \frac{\sqrt{2}}{a}\begin{bmatrix} 1 & 0 & -\bar{x}_1 \\ 0 & 1 & -\bar{x}_2 \\ 0 & 0 & \frac{a}{\sqrt{2}} \end{bmatrix} \qquad \mathbf{T}\mathbf{x}_i = \begin{bmatrix} \frac{\sqrt{2}}{a}(x_{i1} - \bar{x}_1) \\ \frac{\sqrt{2}}{a}(x_{i2} - \bar{x}_2) \\ 1 \end{bmatrix}$$

a 2D similarity transformation.

The normalized eight-point algorithm

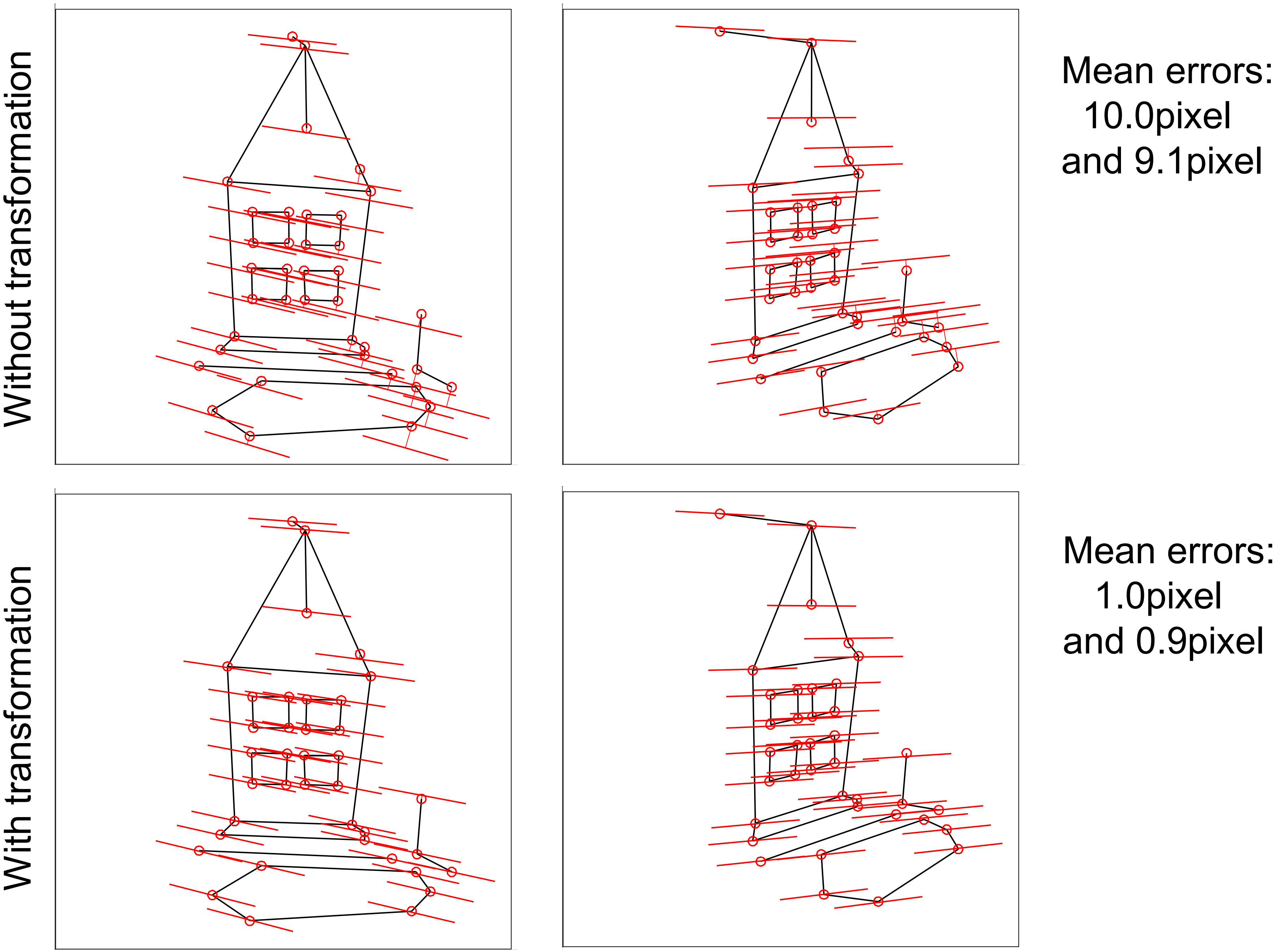
(Hartley, 1995)

- 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.
- Use the eight-point algorithm to compute *F* from the *normalized points,* n_1 and n_2.
- *Enforce the rank-2 constraint. F*or example, take SVD of *F* and throw out the smallest singular value.
- *Transform* fundamental matrix *back to original units:* if T and T' are the normalizing transformations in the two images, than the fundamental matrix in original coordinates is $T^T F T'$.

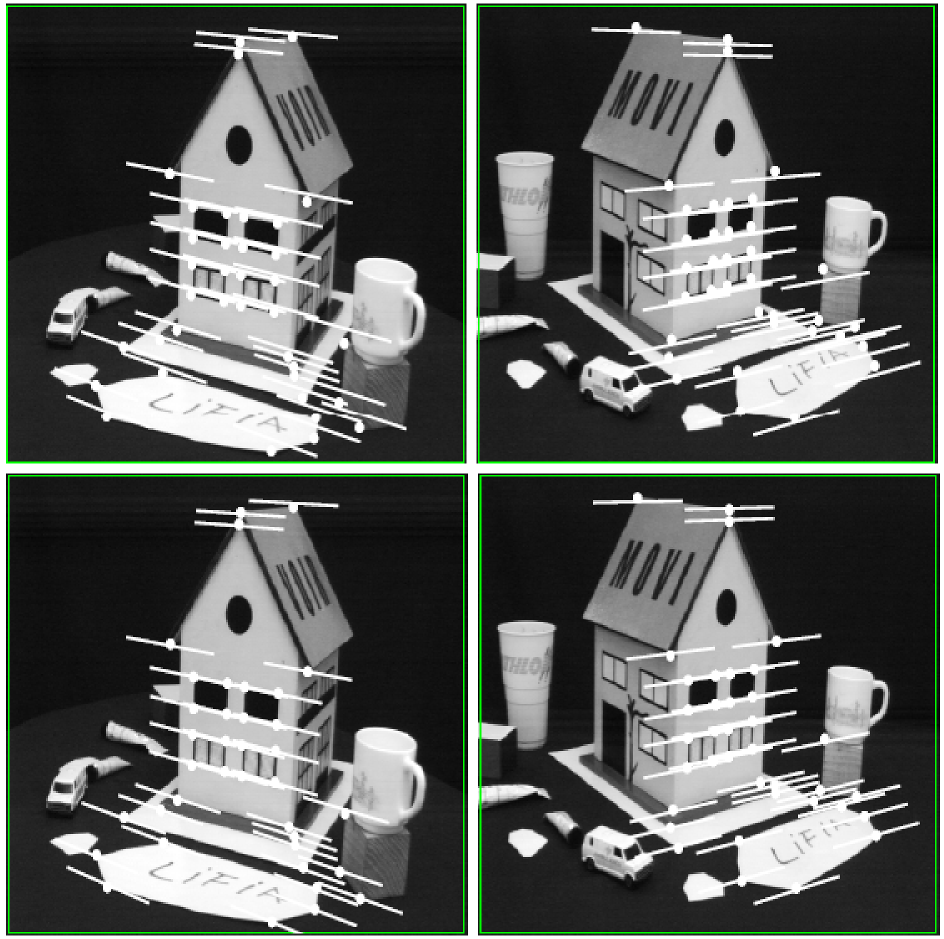
Isotropic translation (mean to origin) and scale in each image separately.

$x\_1 = T^{-1} n\_1 \qquad x\_2 = T'^{-1} n\_2$

$(n\_1)^T \underbrace{T^{-T} F T'^{-1}}_{\text{given}} n\_2 = 0 \quad \Rightarrow \text{final } F$


Without transformation
Mean errors:
10.0pixel
and 9.1pixel
With transformation
Mean errors:
1.0pixel
and 0.9pixel

Comparison of estimation algorithms

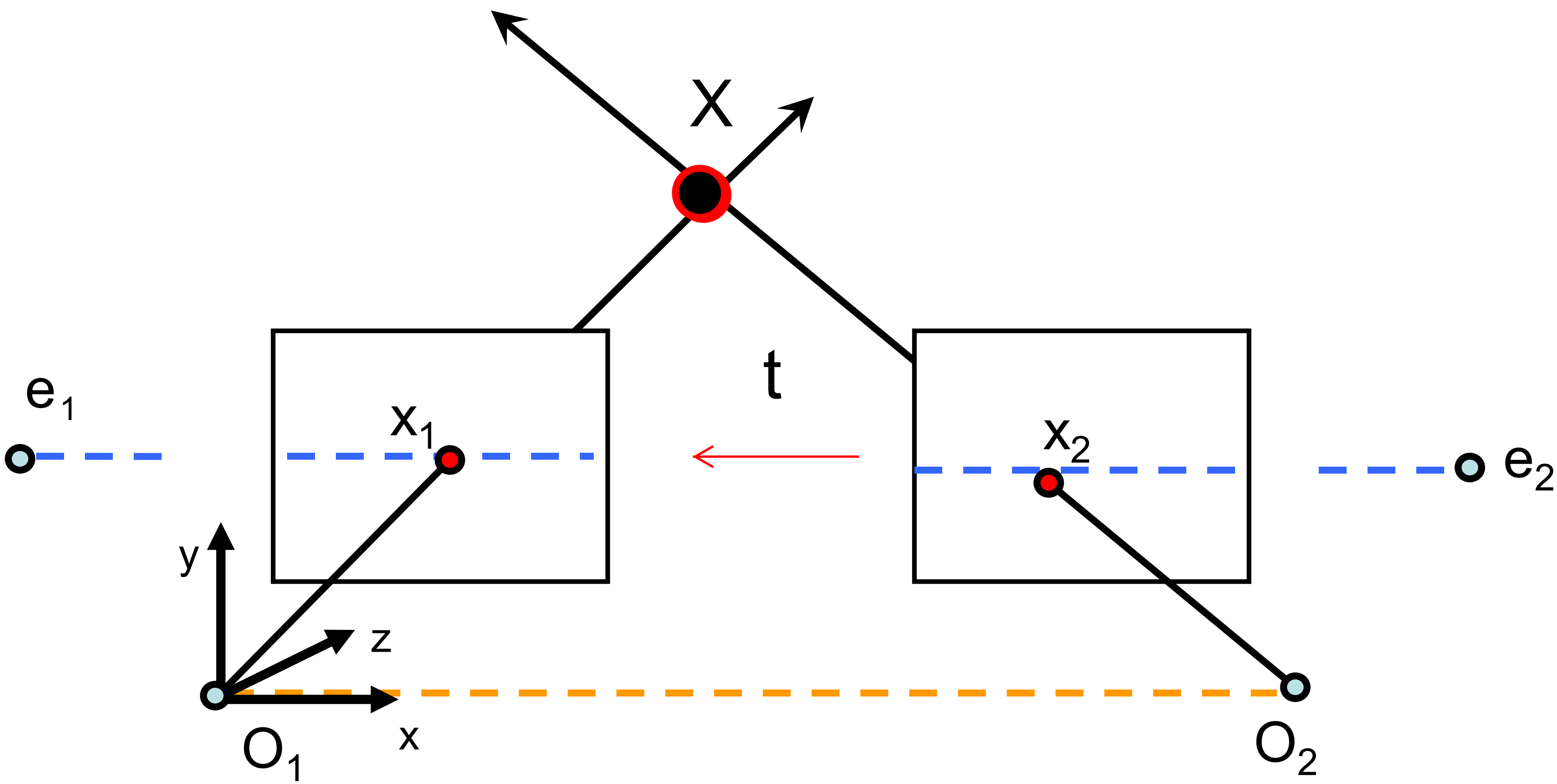


	8-point	Normalized 8-point	Nonlinear least squares
Av. Dist. 1	2.33 pixels	0.92 pixel	0.86 pixel
Av. Dist. 2	2.18 pixels	0.85 pixel	0.80 pixel

From epipolar geometry to camera calibration

- Estimating the fundamental matrix is known as “weak calibration”.
- If *we know* the calibration matrices of the two cameras, we can estimate the essential matrix: $E = K^T F K'$ (see F from E slide)
- The essential matrix can give us the relative rotation and translation between the cameras, with 5 point pairs.

Example: Parallel calibrated images



$K_1 = K_2$ = known

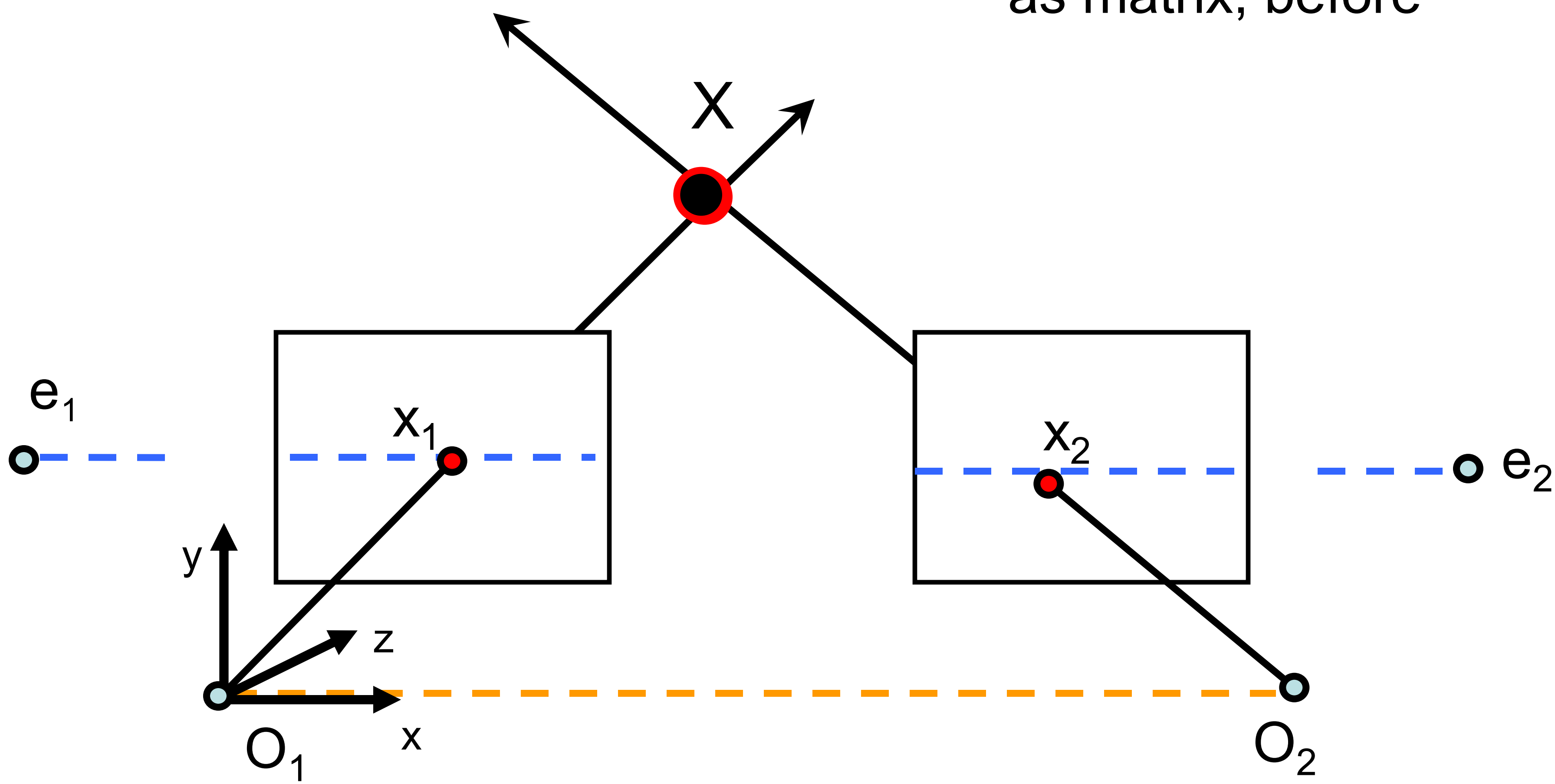
x parallel to O_1O_2

$$\boxed{E = ?}$$

Hint :

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

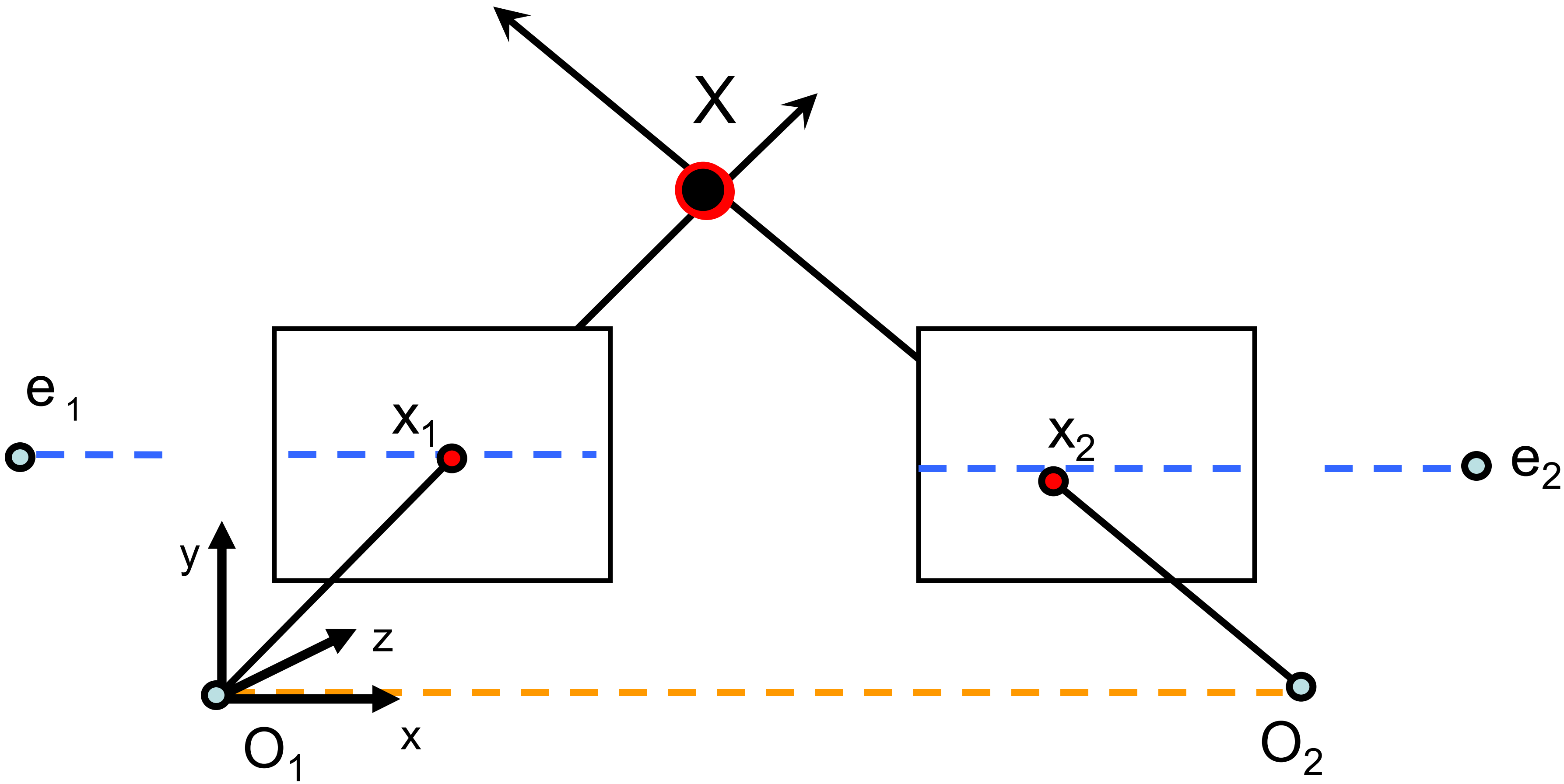
see cross-product
as matrix, before



K_1=K_2 = known
x parallel to O_1O_2

$$\boxed{E=?}$$

$$E=[t_\times]R=\begin{bmatrix}0&0&0\\0&0&-T\\0&T&0\end{bmatrix}$$



Epipolar constraint reduces to → y = y'

In stereo vision that will be a big help.