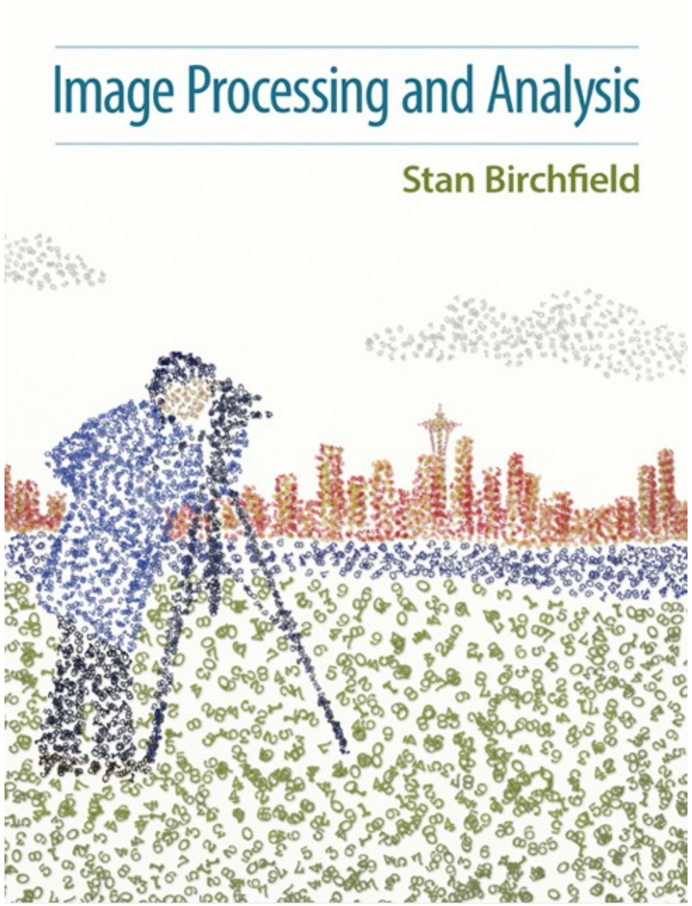
Prof. Kjersti Engan

# ELE510 Image processing and computer vision

Geometry of multiple views, ( chap 13.6 Birchfield) 2023

University of Stavanger

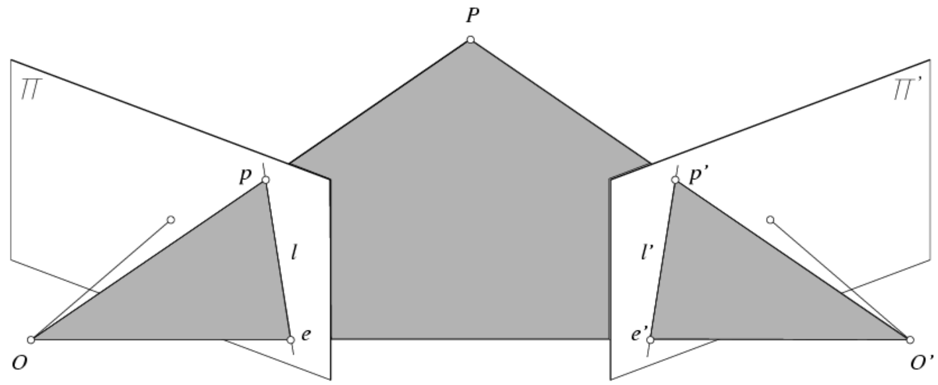



Parts in course presentations is material from Cengage learning. It can be used for teaching, and it can be share it with students on access controlled web-sites (Canvas) for use in THIS course. **But it should not be copied or distributed further in any way** (by you or anybody).

# Geometry of multiple views – 3D reconstruction

Three points from the topic:

1. Epipolar geometry help finding corresponding points
2. Fundamental and essential matrices
3. 3D Reconstruction from camera projection matrices

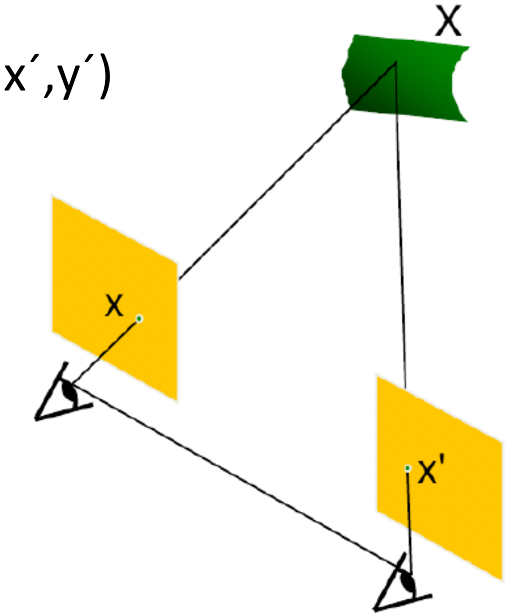


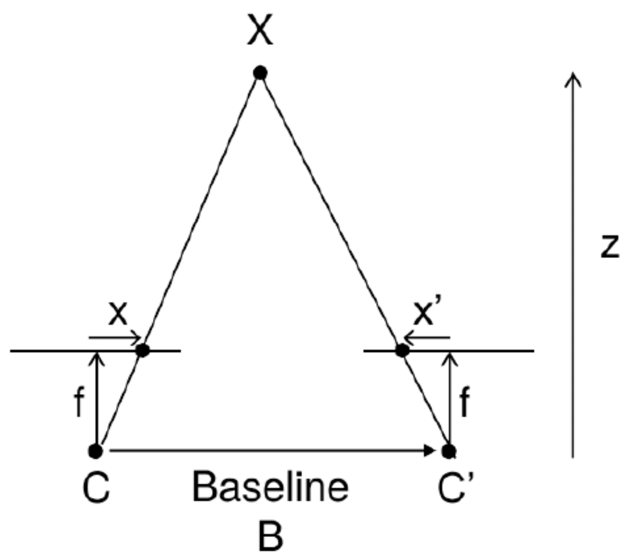
# Depth from Stereo

Goal: recover depth, Z

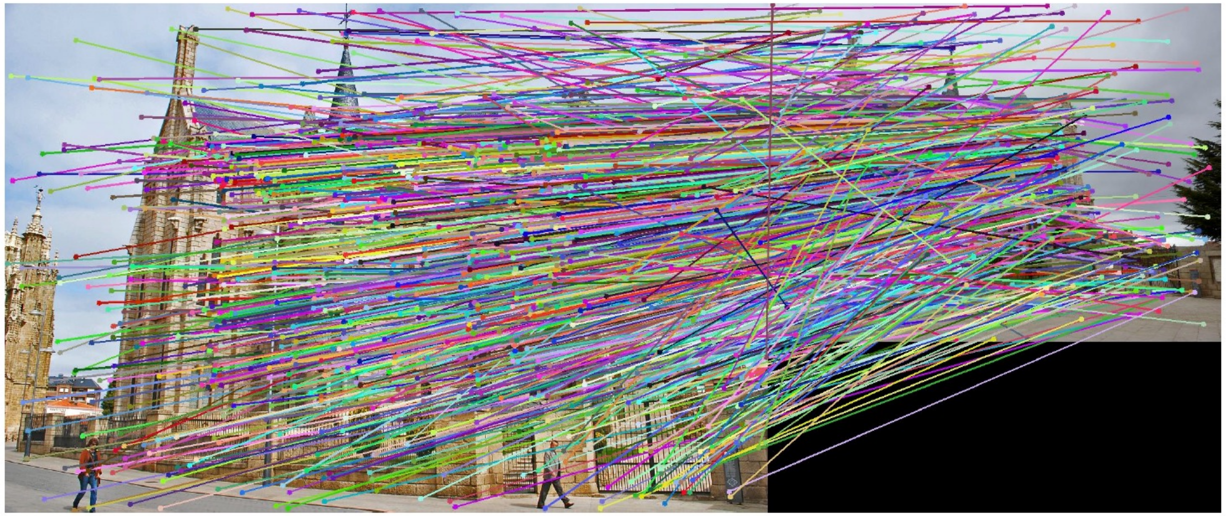
Subproblems:

- Calibration of cameras
- Correspondence - (x,y)<->(x´,y´)
- 3D Reconstruction

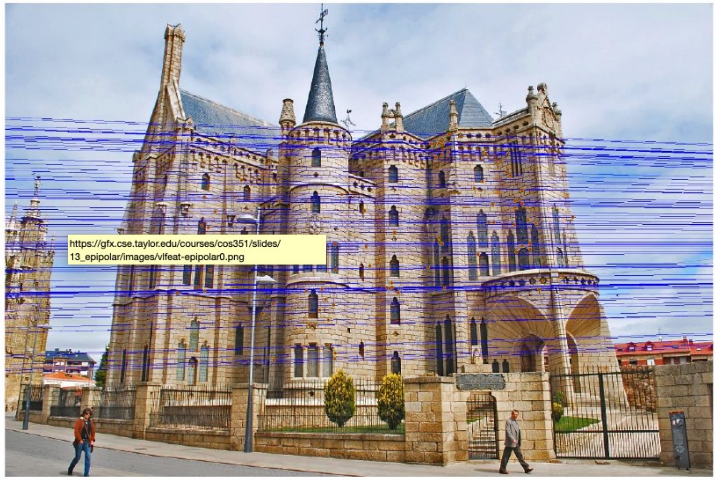


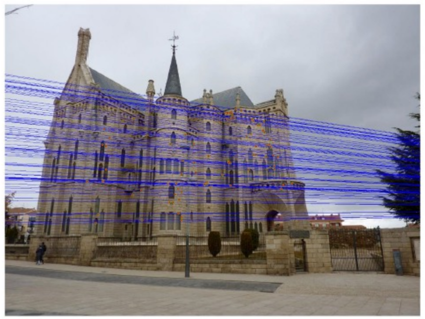



Recap correspondence -- Where are possible corresponding points – lies on the epipolar lines



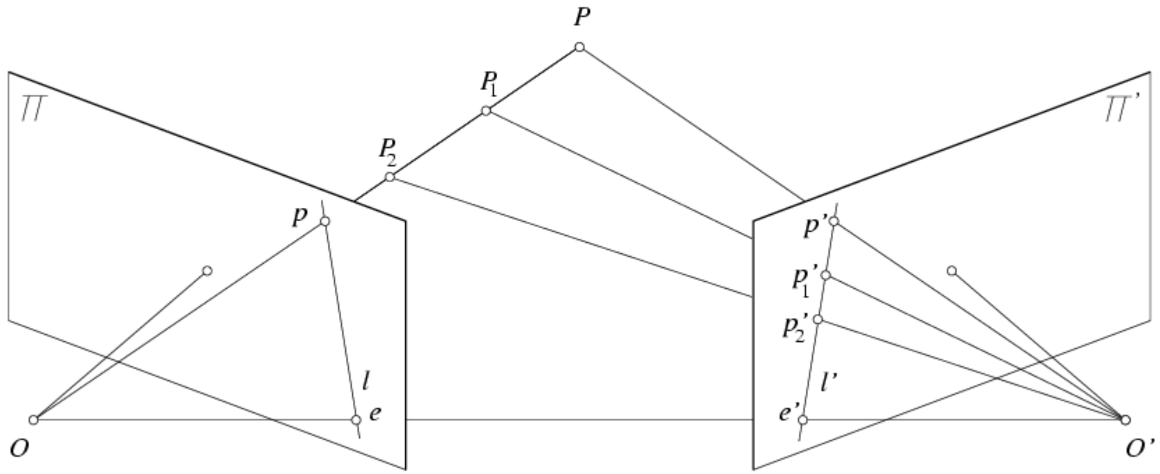
Remember rectified cameras: corresponding points are on the scan lines (y = y′).



# Epipolar Geometry

The *left camera coordinates are used as reference/ world coordinate system*




The *Epipolar Plane*: OO'P
*Epipoles*: e, e'
*Epipolar lines*: l, l'
Image points: p, p'

**Note!** There is a separate *epipolar plane* for each point in the scene.

Potential matches for ***p/ p'*** have to lie on the corresponding epipolar line ***l'/ l***

The optical center (=focal point for pinhole camera) is a distinct point on the other camera's image plane.

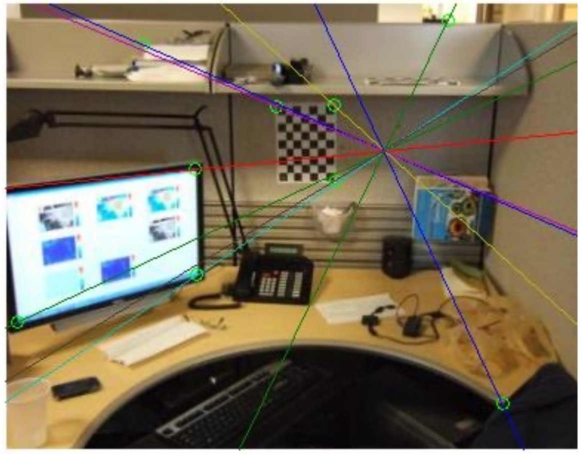
Denoted by **e** and **e´**, called *epipoles or epipolar points*.

Both epipoles **e** and **e´** in their respective image planes and both optical centers **O** and **O´** lie on a single 3D line.

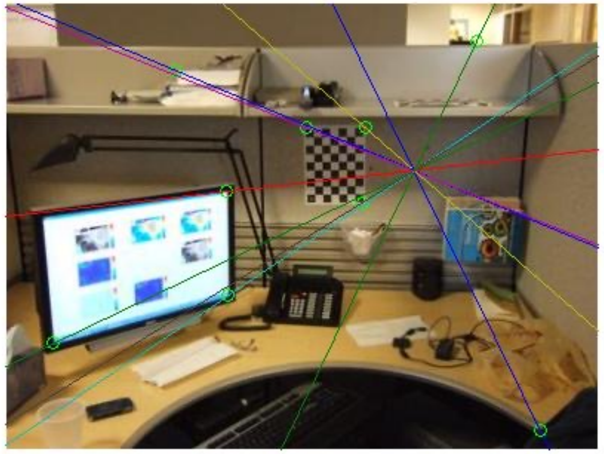
If this line does not cross the image plane, the epipoles are outside the image ( here: inside/visible in the image).

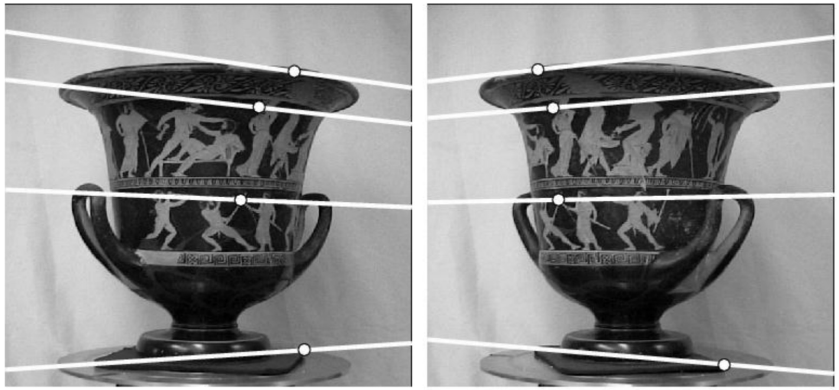
# Epipolar Lines

Inliers and Epipolar Lines in First Image



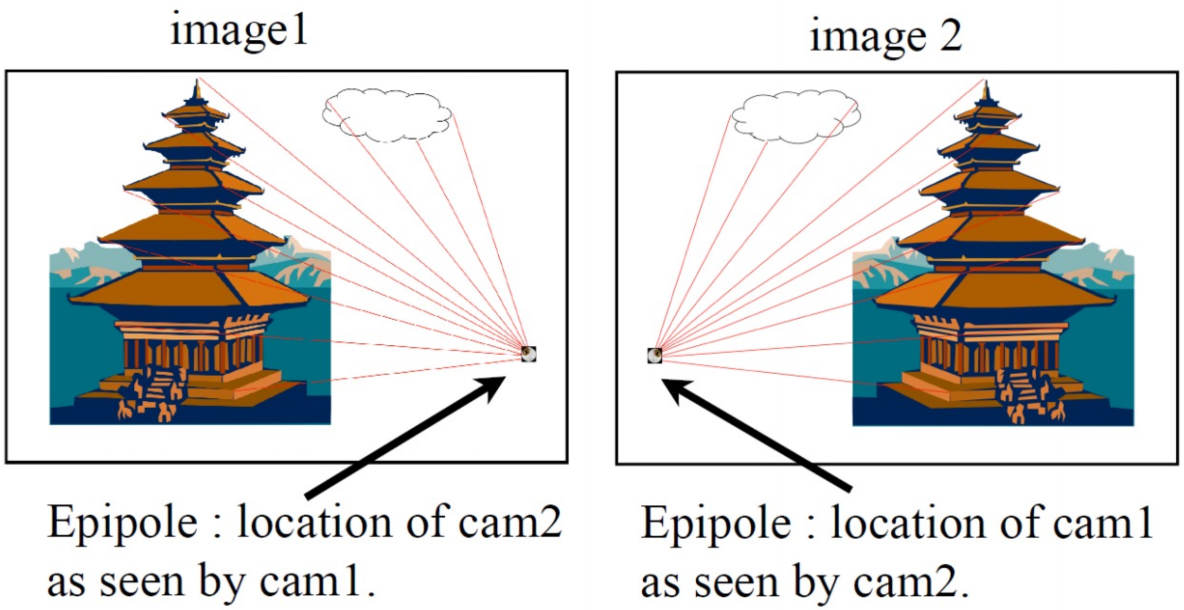
Inliers and Epipolar Lines in Second Image



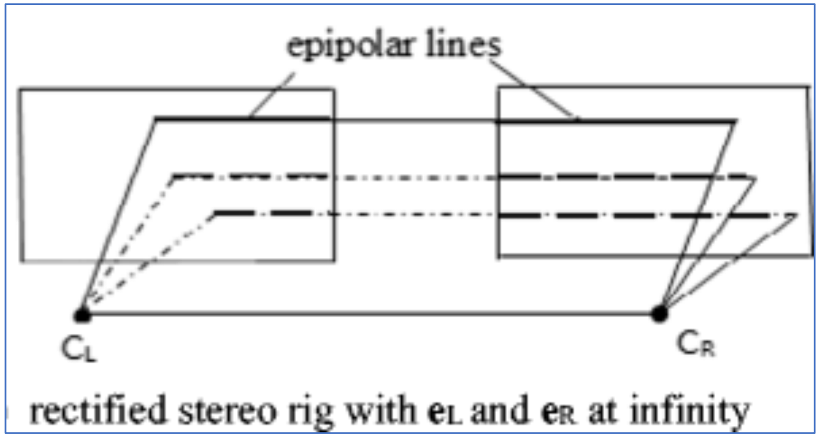


- Potential matches for ***p*** have to lie on the corresponding epipolar line ***l’***.
- Potential matches for ***p’*** have to lie on the corresponding epipolar line ***l***.

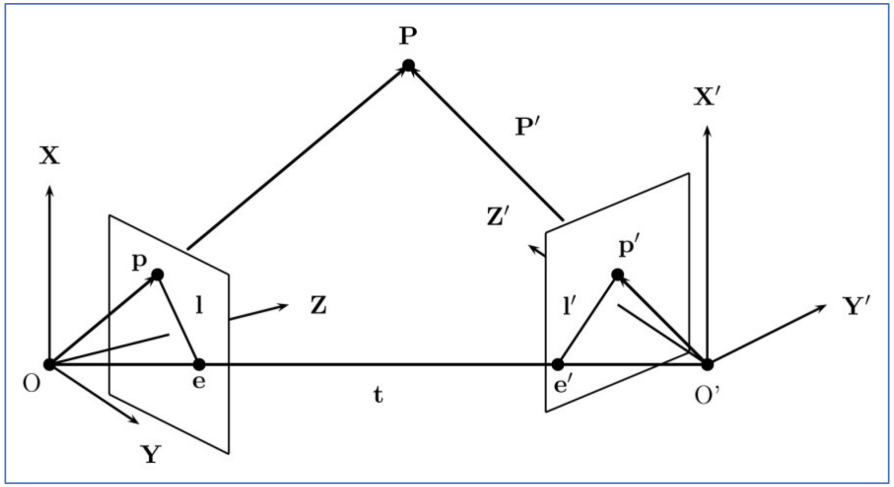
Epipolar plane and line is new for each physical point. Epipoles are constant - it is the focal point/origo of the other camera seen in the image, i.e. only one point. All epipolar lines goes through the epipole – i.e. intersection of epipolar lines revieals the epipoles.




Where is the epipoles for rectified cameras?



# Relationship between camera coordinates for the two cameras




The *left camera coordinates are used as world/reference*.

Right camera coordinates are denoted by primed symbols ‘

There is a rotation, **R**, and a translation, **t**, between the right and left coordinate systems (i.e. cameras). **R** is defined rotation to right camera described with left camera axis. Baseline translation **(t)** is given in left coordinates.
We get:

$$\mathbf{P} = \mathcal{R}\mathbf{P}' + \mathbf{t} \quad \Leftrightarrow \quad \mathbf{P}' = \mathcal{R}^{-1}(\mathbf{P} - \mathbf{t}) = \mathcal{R}^{T}\mathbf{P} - \mathcal{R}^{T}\mathbf{t}$$

Here P and P’ refers to the 3D world point described with left and right camera coordinate systems respectively.

# The Fundamental Matrix

- The fundamental matrix **F** capture the relative geometry between two cameras in pixel values. Let x and x' be corresponding points: $$(x_i, y_i) \Leftrightarrow (x'_i, y'_i)$$

$$\mathbf{x}^T \mathbf{l} = 0$$
$$\mathbf{x}'^T \mathbf{l}' = 0$$

Because the point lie on the epipolar lines

- How du we find the epipolar line *l'* in the second image associated with the point **x** in the first image? Has to be a linear function: $$\boxed{l' = \mathbf{F}\mathbf{x}}$$

We call F the fundamental matrix

# The Fundamental Matrix

$$(x_i, y_i) \Leftrightarrow (x'_i, y'_i)$$

- Given the fundamental matrix **F**, the epipolar line *l'* in the second image associated with the point **x** in the first image is given by:

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

- Similarly, the epipolar line l in the first image associated with the point **x'** in the second image is given by

$$\boxed{l = \mathbf{F}^T\mathbf{x}'}$$

$$\mathbf{x}^T\mathbf{l} = 0$$
$$\mathbf{x}'^T\mathbf{l}' = 0$$

Because the point lie on the epipolar lines

$$\Rightarrow \quad \mathbf{x}'^T\mathbf{F}\mathbf{x} = 0$$
$$\mathbf{x}^T\mathbf{F}^T\mathbf{x}' = 0$$

# Fundamental matrix – rectified cameras

For rectified cameras:   y - y' = 0

$$\mathbf{x'}^T \mathbf{F} \mathbf{x} = \begin{pmatrix} x' & y' & 1 \end{pmatrix} \mathbf{F} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = 0$$

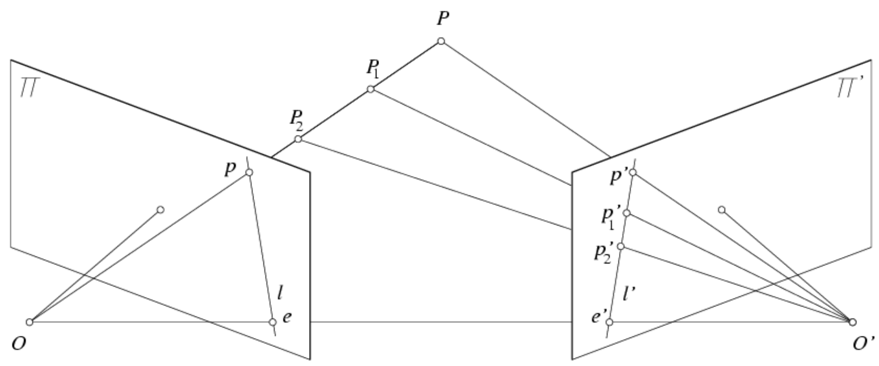
# Fundamental matrix and epipoles

The fundamental matrix also encodes the epipoles **e, e’.** since these are a point on the line

$$\left.\begin{aligned} \mathbf{l}^T\mathbf{e} &= 0 \\ \mathbf{l} &= \mathbf{F}^T\mathbf{x}' \end{aligned}\right\} \quad \mathbf{x}'^T\mathbf{F}\mathbf{e} = 0$$

This must be true for any **x**’. This means:

$$\mathbf{F}\mathbf{e} = 0 \qquad \mathbf{F}^T\mathbf{e}' = 0$$



# The Essential Matrix

- Like the fundamental matrix **F**, the essential matrix, **E**, captures the geometric relationship between the cameras in a compact matrix form.
- The fundamental matrix relates *uncalibrated* cameras (pixel values), while the essential matrix relates *calibrated* cameras (meter).
- If we know **F** or **E** we can find the other with the help of K and K' (the matrices of intrinsic camera parameters for the two cameras)
- Will show that the relationship between the two matrices is:

$$\boxed{\mathbf{F} = \mathbf{K}'^{-\top}\mathbf{E}\mathbf{K}^{-1}, \text{ or } \mathbf{E} = \mathbf{K}'^{\top}\mathbf{F}\mathbf{K}}$$

- E has 5 parameters: 3 for rotation and 2 for translation in direction between the cameras.

# The Essential Matrix

K and K' are the intrinsic calibration matrices for the two cameras

$\bar{\mathbf{x}} \text{ and } \bar{\mathbf{x}}'$ are metric coordinates $\mathbf{x} \text{ and } \mathbf{x}'$ are pixel coordinates

$$\mathbf{x} = \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$$

# Reconstruction

$$\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \propto \mathbf{P} \begin{pmatrix} x_w \\ y_w \\ z_w \\ 1 \end{pmatrix}$$

**Recap – perspective imaging**
**P** is camera projection matrix

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \propto \underbrace{\begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix}}_{\mathbf{P}_{\{3\times 4\}}} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix}$$

$$\mathbf{P}_{\{3\times 4\}} = \underbrace{\begin{bmatrix} \alpha & \gamma & u_0 \\ 0 & \beta & v_0 \\ 0 & 0 & 1 \end{bmatrix}}_{\mathbf{K}_{\{3\times 3\}}} [\mathbf{R}_{\{3\times 3\}} \ \mathbf{t}_{\{3\times 1\}}]$$

# Estimating the fundamental matrix

- If we let **x** and **x'** be points in the two images, then we can rewrite the fundamental matrix equation by explicitly listing the individual elements of the matrix and vectors as follows:

$$\boxed{\mathbf{x'}^T \mathbf{F} \mathbf{x} = 0} \qquad \mathbf{x'}^{\mathsf{T}} \mathbf{F} \mathbf{x} = [x' \ \ y' \ \ 1] \underbrace{\begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix}}_{\mathbf{F}} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = 0$$

- Rearranged we can collect the nine unknowns ( $f_{ij}$) in a vector:

$$\mathbf{A}_{\{n \times 9\}} \mathbf{f}_{\{9 \times 1\}} = \mathbf{0}_{\{9 \times 1\}}$$

- Where $n \geq 8$ is the number of corresponding pairs $\boxed{(x_i, y_i) \Leftrightarrow (x'_i, y'_i)}$
- Can solve for F, also need to make sure is a valid fundamental matrix

# Estimate and decomposing the Essential Matrix

- From **F** and **K, K'** we can find **E**: $\boxed{\mathbf{F} = \mathbf{K}'^{-\mathsf{T}}\mathbf{E}\mathbf{K}^{-1}, \text{ or } \mathbf{E} = \mathbf{K}'^{\mathsf{T}}\mathbf{F}\mathbf{K}}$
- Given an essential matrix **E**, we would like to be able to extract the translation vector **t** and rotation matrix **R**. This describes the rotation and translation between the two cameras.

- If we have the intrinsic parameters ( **K**, and **K´**) and **R** and **t**, we can find the camera projection matrices **P**, and **P´** $\boxed{\mathbf{P} = \mathbf{K}\,[\,\mathbf{R} \;\; \mathbf{t}]}$

- Decomposing **E to find R and t** can be done by computing its SVD and applying some properties. We will not dig into that – some details in the book.

# Finding the Camera projection Matrices

- The Essential matrix, E, can be decomposed into $[\mathbf{R}_{\{3\times3\}}\mathbf{t}_{\{3\times1\}}]$

describing rotation and translation between the two cameras.

- When that is done we can estimate the camera matrices. The left camera coordinate system is the world system, so no rotation and translation for P:

$$\mathbf{P} = \mathbf{K}[\mathbf{I}_{\{3\times3\}}\mathbf{0}_{\{3\times1\}}]$$

$$\mathbf{P}' = \mathbf{K}'[\mathbf{R}_{\{3\times3\}}\mathbf{t}_{\{3\times1\}}]$$

**ALGORITHM 13.15** Estimate the projection matrices of a pair of internally calibrated cameras

EstimateCameraProjectionMatrices$(\{\mathbf{x}_i \leftrightarrow \mathbf{x}'_i\}_{i=1}^{n}, \mathbf{K}, \mathbf{K}')$

**Input:** $n$ corresponding pairs of points $(x_i, y_i) \leftrightarrow (x'_i, y'_i)$ between two images intrinsic camera parameters $\mathbf{K}$ and $\mathbf{K}'$

**Output:** the $3 \times 4$ camera projection matrices $\mathbf{P}$ and $\mathbf{P}'$

$\mathbf{F} \leftarrow$ EightPointFundamental$(\{\mathbf{x}_i \leftrightarrow \mathbf{x}'_i\}_{i=1}^{n})$

$\mathbf{E} \leftarrow \mathbf{K}'^{\mathsf{T}}\mathbf{F}\mathbf{K}$

$\mathbf{R}, \mathbf{t} \leftarrow$ DecomposeEssentialMatrix$(\mathbf{E})$

$\mathbf{P} \leftarrow \mathbf{K}\,[\mathbf{I}_{\{3\times3\}} \quad \mathbf{0}_{\{3\times1\}}]$

$\mathbf{P}' \leftarrow \mathbf{K}'\,[\mathbf{R} \quad \mathbf{t}]$

**return** $\mathbf{P}, \mathbf{P}'$

# Calibration stereo - summary

Calibration for stereo setup means finding P and P' camera projection matrices:

1) Use for example Zhangs calibration algorithm to find K and K' (each camera alone). This gives us P. $\boxed{\mathbf{P} = \mathbf{K}[\mathbf{I}_{\{3\times3\}}\mathbf{0}_{\{3\times1\}}]}$

2) Need at least 5, typically 8 correspondence points x,x' to be able to estimate the fundamental matrix F

( need to search a larger space. Dont know the epipolar lines yet)

3) Find the essential matrix E from F, K and K'

4) Decompose E to get R and t – this will give us P´ $\boxed{\mathbf{P}' = \mathbf{K}'[\mathbf{R}_{\{3\times3\}}\mathbf{t}_{\{3\times1\}}]}$

# Doing 3D reconstruction

- When we have estimated F and the camera projection matrices P and P’, i.e. calibrated our stereo vision setup, we can use it to reconstruct 3D world:

1. Find corresponding points (**x, x**’). We know the fundamental matrix **F**, so this is a 1D search on the epipolar lines: $\boxed{l' = \mathbf{F}\mathbf{x}}$

2. Estimate the 3D world point **w** from (**x, x’, P** and **P’). How?**

# Estimating 3D Point Coordinates

- Given the camera projection matrices $\mathbf{P}_{\{3x4\}}$ and $\mathbf{P'}_{\{3x4\}}$, the 3D coordinates of a world point **w** can be estimated from its projections onto the two images.

$$\mathbf{P} = \begin{bmatrix} \mathbf{p}_1^\mathsf{T} \\ \mathbf{p}_2^\mathsf{T} \\ \mathbf{p}_3^\mathsf{T} \end{bmatrix} \quad \text{and} \quad \mathbf{P}' = \begin{bmatrix} \mathbf{p}_1'^\mathsf{T} \\ \mathbf{p}_2'^\mathsf{T} \\ \mathbf{p}_3'^\mathsf{T} \end{bmatrix}$$

- Let $\mathbf{w}_{\{4\times1\}}$ be the homogeneous coordinates of the world point. coordinates of the image point (in pixels) are then given by:

$$(x, y) = \left( \frac{\mathbf{p}_1^\mathsf{T}\mathbf{w}}{\mathbf{p}_3^\mathsf{T}\mathbf{w}}, \frac{\mathbf{p}_2^\mathsf{T}\mathbf{w}}{\mathbf{p}_3^\mathsf{T}\mathbf{w}} \right) \quad \text{and} \quad (x', y') = \left( \frac{\mathbf{p}_1'^\mathsf{T}\mathbf{w}}{\mathbf{p}_3'^\mathsf{T}\mathbf{w}}, \frac{\mathbf{p}_2'^\mathsf{T}\mathbf{w}}{\mathbf{p}_3'^\mathsf{T}\mathbf{w}} \right)$$

- Solve for the unknown (**w**), as shown on notes.

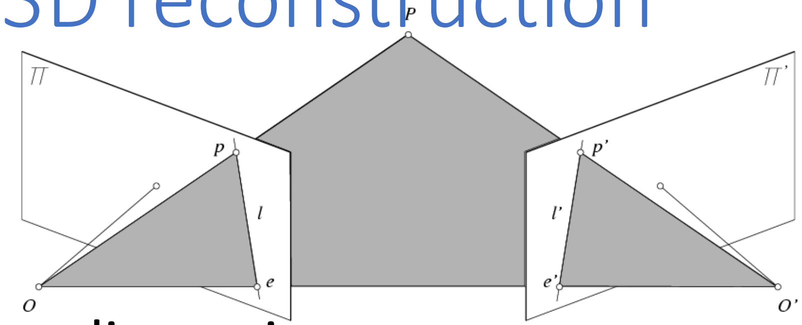
# 3D Reconstruction from Image Pairs - summary

- Find interest points (Harris-Stephens / SIFT )
- Match interest points
- Compute fundamental matrix F
- Calibrate cameras (Zhangs), find K, K´
- Estimate E from F, K and K´, find R, t
- Compute camera matrices P and P’ from K, K´, R and t
- For each matching image points x and x’, compute world point in scene using (x,y), (x´,y´) P and P´ (Use F to help find matching pairs)

Calibrating stereo setup

We can use the fundamental matrix **F** to limit correspondence to 1D search for general stereo camera positions in the same way as is possible for rectified stereo

# Geometry of multiple views – 3D reconstruction




Three points from the topic:

1. Epipolar geometry help finding corresponding points
   - ✓ Epipolar lines limits search space from 2D to 1D
2. Fundamental and essential matrices
   - ✓ Fundamental matrix (F) – relationship between the images ( image coordinates). Essential matrix € – relationship between cameras
3. 3D Reconstruction from camera projection matrices
   - ✓ Can find world point from P and P´ and (x,y)-(x´,y´)