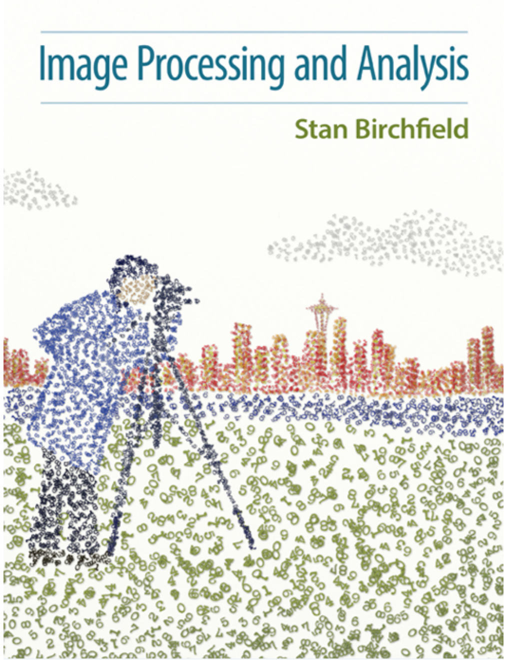
University
of Stavanger

Prof. Kjersti Engan

ELE510 Image processing and computer vision

Geometry of multiple views, (chap 13.6 Birchfield) 2020



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).

13.3 non-rectified cam.

$$\underset{\sim}{F} = \begin{bmatrix} 1 & 2 & 1 \\ -2 & -4 & -2 \\ 3 & 3 & 1 \end{bmatrix}$$

can $\underline{x} \Rightarrow (x, y) = (100, 50)$

$\underline{x}' \Rightarrow (x', y') = (78, 40)$

can these be corresponding points?

$$\underline{l}' = \underset{\sim}{F} \cdot \underline{x} = \begin{bmatrix} 1 & 2 & 1 \\ -2 & -4 & -2 \\ 3 & 3 & 1 \end{bmatrix} \cdot \begin{bmatrix} 100 \\ 50 \\ 1 \end{bmatrix} = \begin{bmatrix} 201 \\ -402 \\ 451 \end{bmatrix}$$

know $\underline{x}'^T \cdot \underline{l}' = 0$ point have to be on the line

test $\begin{bmatrix} 78 & 40 & 1 \end{bmatrix} \cdot \begin{bmatrix} 201 \\ -402 \\ 451 \end{bmatrix} = 49 \neq 0$

does not look like corresponding points

BUT need to normalize first

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)

- The relationship between the two matrices is: $\boxed{\mathbf{F} = \mathbf{K}'^{-\mathsf{T}}\mathbf{E}\mathbf{K}^{-1}, \text{ or } \mathbf{E} = \mathbf{K}'^{\mathsf{T}}\mathbf{F}\mathbf{K}}$
- E has 5 parameters: 3 for rotation and 2 for translation in direction between the cameras.

$$\underset{\sim}{K} = \begin{bmatrix} \alpha & \boxed{\gamma} & x_0 \\ 0 & \beta & y_0 \\ 0 & 0 & 1 \end{bmatrix}$$

$=0$

The Essential Matrix

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

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

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

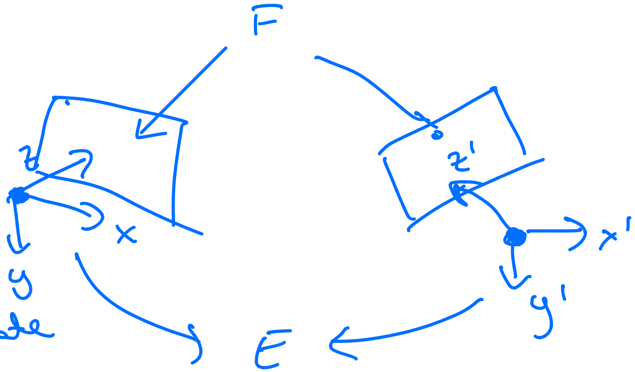
* $\underline{x} = \underset{\sim}{K} \cdot \underline{\bar{x}}$ $\qquad$ $\underline{x}' = \underset{\sim}{K}' \cdot \underline{\bar{x}}'$ $\Rightarrow$ $\underline{\bar{x}}' = \underset{\sim}{K}'^{-1} \cdot \underline{x}'$

Fundamental matrix $\quad \boxed{\underline{x}'^T \underset{\sim}{F}\, \underline{x} = 0}$

Essential matrix $\quad \underline{\bar{x}}'^T \underset{\sim}{E}\, \underline{\bar{x}} = 0$ $\qquad$ $\underline{\bar{x}} = \underset{\sim}{K}^{-1} \cdot \underline{x}$

$$\Rightarrow (\underset{\sim}{K}'^{-1} \underline{x}')^T \underset{\sim}{E}\, \underset{\sim}{K}^{-1} \underline{x} = \underline{x}'^T \underbrace{\underset{\sim}{K}'^{-T} \underset{\sim}{E}\, \underset{\sim}{K}^{-1}}_{\underset{\sim}{F}} \underline{x} = 0$$

$\underset{\sim}{F} = \underset{\sim}{K}'^{-T} \underset{\sim}{E}\, \underset{\sim}{K}^{-1}$ or $\underset{\sim}{E} = \underset{\sim}{K}'^{T} \underset{\sim}{F}\, \underset{\sim}{K}$



world coordinate system

is not world

$$\underset{\sim}{R} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\underline{t} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$\left.\begin{matrix} \underset{\sim}{R} \\ \underline{t} \end{matrix}\right\}$ describing right camera syst. relative to world. $\Rightarrow$ left camera.

$$E = \begin{bmatrix} \underset{\sim}{R} & \underline{t} \end{bmatrix}$$

Calibration and Reconstruction 3D world

- The general procedure to reconstruct the metric geometry of a scene from point correspondences taken by a pair of calibrated cameras:

calibrate stereo setup

  - (1) estimate the fundamental matrix from the correspondences; F
  - (2) construct the essential matrix using the intrinsic camera calibration parameters and the fundamental matrix; K, K' $\Rightarrow E$
  - (3) decompose the essential matrix into rotation and translation (up to scale); and
  - (4) estimate the 3D coordinates of world points. $\mapsto$

Reconstruction

$$\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \propto \underset{3\times 4}{\mathbf{P}} \begin{pmatrix} x_w \\ y_w \\ z_w \\ 1 \end{pmatrix}$$

P is camera projection matrix

$$\underset{\sim}{P} = \underset{\sim}{K} \cdot [\underset{\sim}{R} \quad \underline{t}]$$

intrinsic

m → pixels

extrinsic.
camera coords coord. syst. relative to the world.

$$\boxed{\begin{aligned} \underset{\sim}{P} &= \underline{\underset{\sim}{K}} \cdot [\underset{\sim}{I} \quad 0] \\ \underset{\sim}{P}' &= \underline{\underset{\sim}{K}}' \cdot \underbrace{[\underset{\sim}{R} \quad \underline{t}]}_{\text{cam 2 rel. to cam 1.}} \end{aligned}}$$

$$\underline{I} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

cam 2 rel. to
cam 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

Decomposing the Essential Matrix

- From **F** and **K, K'** we can find **E**: $\boxed{\mathbf{F} = \mathbf{K}'^{-\top}\mathbf{E}\mathbf{K}^{-1}, \text{ or } \mathbf{E} = \mathbf{K}'^{\top}\mathbf{F}\mathbf{K}}$
- 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** is 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 3×4 camera projection matrices $\mathbf{P}$ and $\mathbf{P}'$

```
F ← EightPointFundamental({x_i ↔ x'_i}_{i=1}^n)
E ← K'^T F K
R, t ← DecomposeEssentialMatrix(E)
P ← K [I_{3×3}  0_{3×1}]
P' ← K' [R  t]
return P, P'
```

Calibration stereo - summary

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

1) use Zhangs calibration algorithm to find K and K’ (each camera alone)
2) Need at least 5, typically 8 correspondence points x,x’ to be able to estimate F
   (need to search a larger space. Dont know the epipolar lines yet)
3) Find E from F, K and K’
4) Decompose E to get R and t

Doing 3D reconstruction

- When we have estimated the camera projection matrices P and P', i.e. calibrated our stereo vision setup, we can use it to reconstruct 3D world:
- Find corresponding points (**x, x'**). We know the fundamental matrix **F**, so this is a 1D search on the epipolar lines.
- Estimate the 3D world point **w** from (**x, x', P** and **P')**.

Computing 3D Point Coordinates

$$\begin{array}{c} \underline{x} \\ \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \end{array} \propto \underset{\sim}{P} \begin{array}{c} w \\ \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} \end{array}$$

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

$$\mathbf{P} = \begin{bmatrix} \mathbf{p}_1^\top \\ \mathbf{p}_2^\top \\ \mathbf{p}_3^\top \end{bmatrix} \quad \text{and} \quad \mathbf{P}' = \begin{bmatrix} \mathbf{p}_1'^\top \\ \mathbf{p}_2'^\top \\ \mathbf{p}_3'^\top \end{bmatrix}$$

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

$$(x, y) = \left(\frac{\mathbf{p}_1^\top \mathbf{w}}{\mathbf{p}_3^\top \mathbf{w}}, \frac{\mathbf{p}_2^\top \mathbf{w}}{\mathbf{p}_3^\top \mathbf{w}} \right) \quad \text{and} \quad (x', y') = \left(\frac{\mathbf{p}_1'^\top \mathbf{w}}{\mathbf{p}_3'^\top \mathbf{w}}, \frac{\mathbf{p}_2'^\top \mathbf{w}}{\mathbf{p}_3'^\top \mathbf{w}} \right)$$

$x \qquad y \qquad\qquad\qquad x' , y'$

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \propto \begin{bmatrix} - p_1^T - \\ p_2^T \\ - p_3^T - \end{bmatrix} \begin{bmatrix} xw \\ yw \\ zw \\ 1 \end{bmatrix}$$

$$\underline{x} \quad \propto \quad \underset{\sim}{p} \quad \underline{w}$$

$$\Rightarrow \quad \underline{x} \propto p_1^T \cdot \underline{w}$$

$$x = \frac{p_1^T \underline{w}}{p_3^T \cdot \underline{w}}$$

$$y \propto p_2^T \cdot \underline{w} \qquad y = \frac{p_2^T \cdot \underline{w}}{p_3^T \cdot \underline{w}}$$

$$1 = \frac{p_3^T \cdot w}{p_3^T \cdot w} = 1$$

$$x\, \underline{p}_3^T \cdot \underline{w} = p_1^T \cdot \underline{w} \quad \Rightarrow \quad x\, p_3^T \underline{w} - \underline{p}_1^T \underline{w} = 0$$

$$y\, \underline{p}_3^T \underline{w} = p_2^T \cdot \underline{w} \quad \Rightarrow \quad y\, p_3^T \underline{w} - p_2^T \underline{w} = 0$$

x', y' do the same using P'

$$(x\, \underline{p}_3^T - \underline{p}_1^T) \cdot \underline{w} = 0$$

$$(y\, p_3^T - p_2^T) \cdot \underline{w} = 0$$

Put together

$$\begin{bmatrix} x\underline{p}_3^T - \underline{p}_1^T \\ y\underline{p}_3^T - \underline{p}_2^T \\ x'\underline{p}_3'^T - \underline{p}_1'^T \\ y'\underline{p}_3'^T - \underline{p}_2'^T \end{bmatrix} \cdot \underline{w} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

$$\underline{w} = \begin{pmatrix} x_w \\ y_w \\ z_w \\ 1 \end{pmatrix}$$

Want to find.

using $(x, y) \leftrightarrow (x', y')$

P, P'

can estimate $\underline{w} \Rightarrow$ 3D, includes depth!

3D Reconstruction from Image Pairs - summary

- Find interest points
- Match interest points
- Compute fundamental matrix F
- Calibrate cameras (Zhangs), find K, K´
- Estimate E from F, K and K´
- Compute camera matrices P and P’ from E

Calibrating stereo setup

[R 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 find matching pairs)

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 simple stereo