Stereo view geometry

Epipolar geometry, triangulation

Morten R. Hannemose, mohan@dtu.dk

February 25, 2022

02504 Computer vision course lectures,
DTU Compute, Kgs. Lyngby 2800, Denmark

Learning objectives

After this lecture you should be able to:

- Derive and apply the fundamental matrix in computer vision
- Derive and apply the essential matrix in computer vision
- Implement the linear algorithm for triangulation
- Explain the pros and cons of using a linear algorithm

Presentation topics

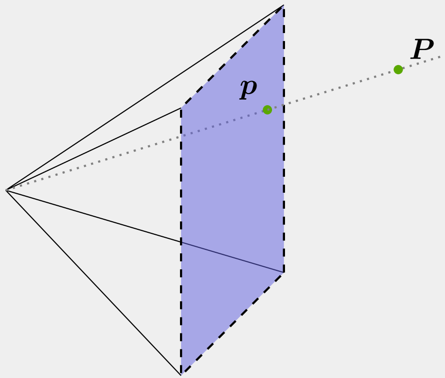
- Projection of points, lines and planes
- Epipolar planes and lines
- Essential and fundamental matrices
- Triangulation
    - Problems with linear algorithm

Projection of points, lines and planes

Point projections

Any 3D point $\boldsymbol{P}$ projects to a 2D point $\boldsymbol{p}$:

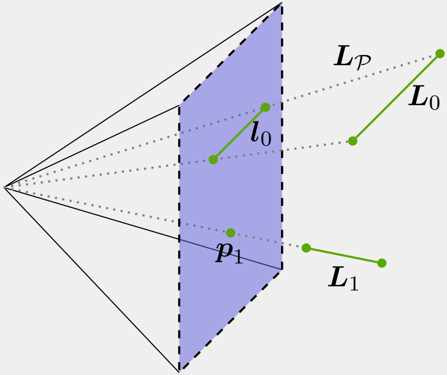
$$\boldsymbol{P} \xrightarrow{\text{projection}} \boldsymbol{p}\,.$$



Line projections

Any 3D line $\boldsymbol{L}$ projects to a 2D line $\boldsymbol{l}$, except if $\boldsymbol{L} \parallel \boldsymbol{L}_{\mathscr{P}}$:

$$\boldsymbol{L}_0 \xrightarrow{\text{projection}} \boldsymbol{l}_0 \,,$$
$$\boldsymbol{L}_1 \xrightarrow{\text{projection}} \boldsymbol{p}_1 \,.$$



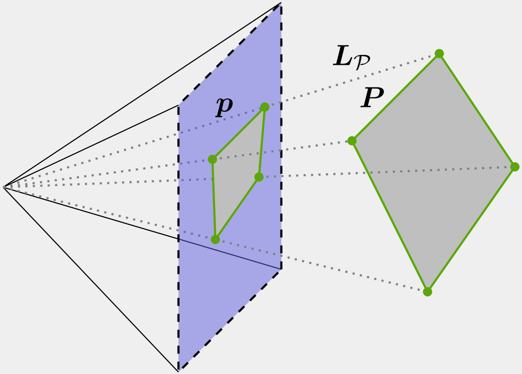
Plane projections

Any 3D plane $\boldsymbol{P}$ projects to the image plane,

$$\boldsymbol{P} \xrightarrow{\text{projection}} \boldsymbol{p}\,,$$

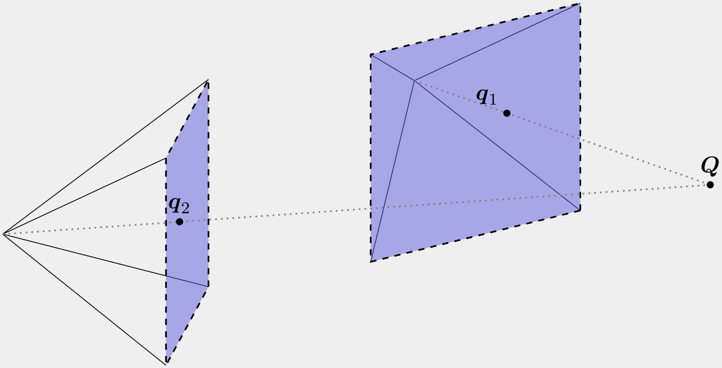
except if $\boldsymbol{P} \parallel \boldsymbol{L}_{\mathscr{P}}$

Where then?



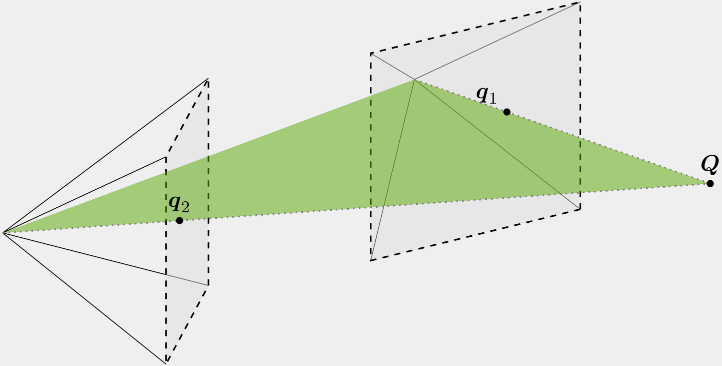
Epipolar planes and lines

Stereo view



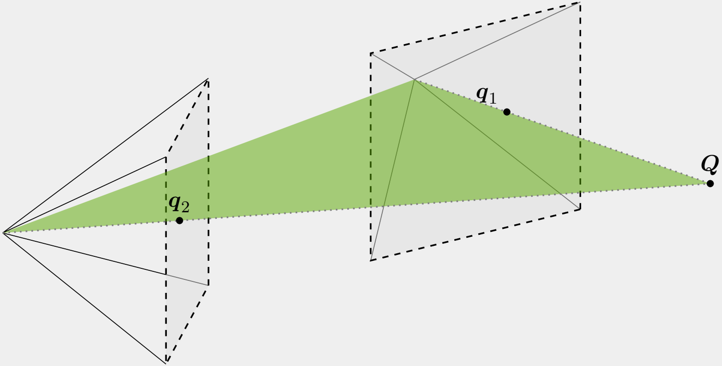
Assume stereo view

Epipolar planes



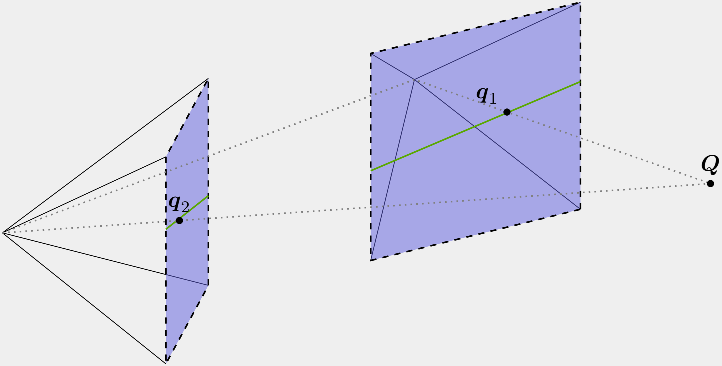
The camera centers and the 3D point Q form an epipolar plane!

Epipolar planes



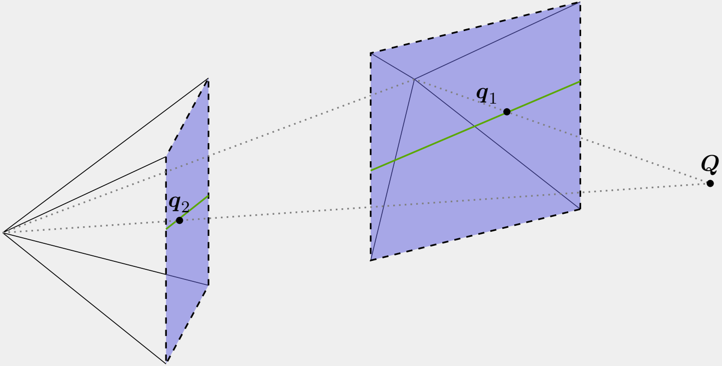
Which line in space is alway part of the epipolar planes?

Epipolar lines



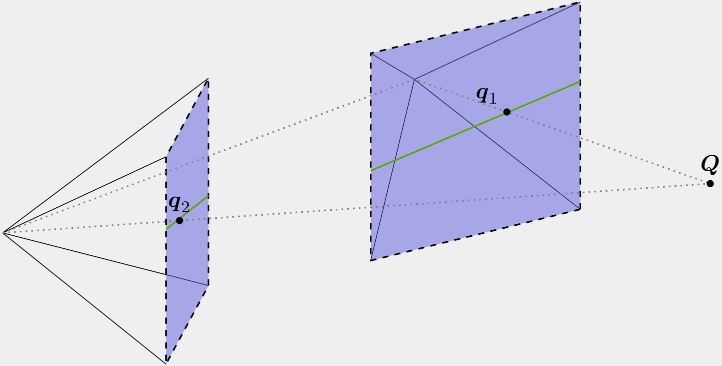
The epipolar lines intersect the epipolar plane and images!

Epipolar lines



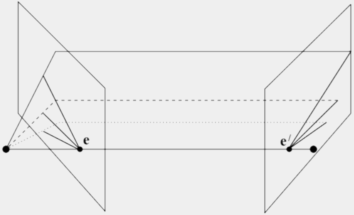
The relation between the line $\boldsymbol{q}_i\boldsymbol{Q}$ and the epipolar lines?

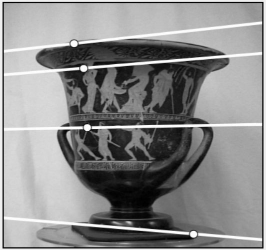
Epipolar lines



All epipolar lines intersect in the epipoles.

Epipoles





Where are the epipoles?

Stereo view is an epipolar geometry:
Each Q in 3D has an epipolar plane.
Pixels correspond if and only if
they lie in the same epipolar plane

Essential and fundamental matrices

Setting the scene

Let $\boldsymbol{q}$ refer to a 2D point in pixels and $\boldsymbol{p}$ refer to the same point in 3D in the reference frame of the camera.

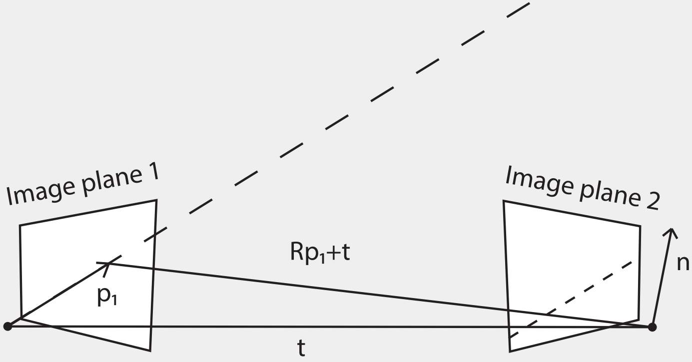
These are related as follows:

$$\boldsymbol{q} = \boldsymbol{K}\boldsymbol{p}$$
$$\boldsymbol{p} = \boldsymbol{K}^{-1}\boldsymbol{q}$$

$\boldsymbol{R}$, $\boldsymbol{t}$ maps from the reference frame of camera one to the reference frame of camera two (their relative transformation)

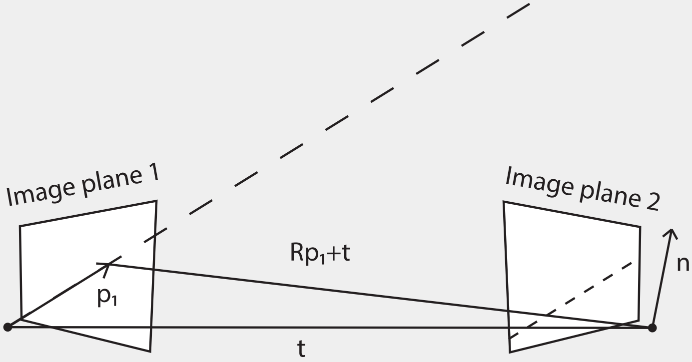
The essential matrix

Consider the epipolar plane given by $\boldsymbol{p}_1$



The essential matrix

Relative to camera 2, what is the normal of the epipolar plane?



The normal is orthogonal to $\boldsymbol{t}$ and $\boldsymbol{R}\boldsymbol{p}_1 + \boldsymbol{t}$

The essential matrix

The normal is then

$$
\begin{aligned}
\boldsymbol{n} &= \boldsymbol{t} \times (\boldsymbol{R}\boldsymbol{p}_1 + \boldsymbol{t}) \\
&= \boldsymbol{t} \times (\boldsymbol{R}\boldsymbol{p}_1) \\
&= \underbrace{[\boldsymbol{t}]_\times \boldsymbol{R}}_{\boldsymbol{E}} \boldsymbol{p}_1.
\end{aligned}
$$

$\boldsymbol{E}$ is called the essential matrix!

The essential matrix

Dot product of orthogonal vectors are zero.

For the corresponding $\boldsymbol{p}_2$ in the second camera

$$
\begin{aligned}
0 &= \boldsymbol{p}_2^\top \boldsymbol{n} \\
&= \boldsymbol{p}_2^\top \boldsymbol{E} \boldsymbol{p}_1
\end{aligned}
$$

The essential matrix imposes a constraint on corresponding points.

How to interpret this?

- $\boldsymbol{p} = \boldsymbol{K}^{-1}\boldsymbol{q}$.
- Are $\boldsymbol{p}_1$ and $\boldsymbol{p}_2$ in homogeneous or inhomogeneous coordinates?

How to interpret this?

- $\boldsymbol{p} = \boldsymbol{K}^{-1}\boldsymbol{q}$.
- Are $\boldsymbol{p}_1$ and $\boldsymbol{p}_2$ in homogeneous or inhomogeneous coordinates?
  - Yes!...
- There are two interpretations:

How to interpret this?

- $\boldsymbol{p} = \boldsymbol{K}^{-1}\boldsymbol{q}$.
- Are $\boldsymbol{p}_1$ and $\boldsymbol{p}_2$ in homogeneous or inhomogeneous coordinates?
  - Yes!...
- There are two interpretations:
  - $\boldsymbol{p}_1$ and $\boldsymbol{p}_2$ are 3D points and $\boldsymbol{n}$ is a vector in 3D. We use this in our derivations.
  - $\boldsymbol{p}_1$ and $\boldsymbol{p}_2$ are 2D points and $\boldsymbol{n} = \boldsymbol{E}\boldsymbol{p}_1$ is the epipolar line, both are in homogeneous coordinates.
- 🤯

The fundamental matrix

Recall that

$$\boldsymbol{p} = \boldsymbol{K}^{-1}\boldsymbol{q}.$$

Then

$$\begin{aligned} \boldsymbol{p}_2^\top \boldsymbol{E} \boldsymbol{p}_1 &= 0 \\ (\boldsymbol{K}_2^{-1}\boldsymbol{q}_2)^\top \boldsymbol{E} (\boldsymbol{K}_1^{-1}\boldsymbol{q}_1) &= 0 \\ \boldsymbol{q}_2^\top \underbrace{\boldsymbol{K}_2^{-\top}\boldsymbol{E}\boldsymbol{K}_1^{-1}}_{\boldsymbol{F}} \boldsymbol{q}_1 &= 0 \end{aligned}$$

where $\boldsymbol{F}$ is the fundamental matrix!

The essential and fundamental matrices form requirements for pixel correspondence:

$$\boldsymbol{p}_2^\intercal \boldsymbol{E} \boldsymbol{p}_1 = 0$$
$$\boldsymbol{q}_2^\intercal \boldsymbol{F} \boldsymbol{q}_1 = 0$$

Note on $\boldsymbol{R}$, $\boldsymbol{t}$

What if $\boldsymbol{R}$, $\boldsymbol{t}$ is not given, but you only know the pose of each camera in world coordinates ($\boldsymbol{R}_1$, $\boldsymbol{t}_1$, $\boldsymbol{R}_2$, $\boldsymbol{t}_2$)?

You can compute the relative transformation (which you will do in the exercise today)

Fundamental/essential matrix vs homography

What is the difference?

Fundamental/essential matrix vs homography

What is the difference?

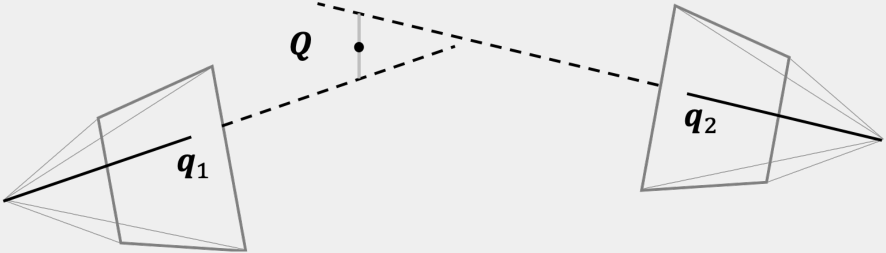
- The fundamental and essential matrices yield epipolar lines:
  - $0 = \boldsymbol{p}_2^\top \boldsymbol{E} \boldsymbol{p}_1$
  - $0 = \boldsymbol{q}_2^\top \boldsymbol{F} \boldsymbol{q}_1$
  - Corresponding points must lie on the epipolar line, no matter what is in the image.
- The homography establishes one-to-one correspondences:
  - $\boldsymbol{q}_2 = \boldsymbol{H} \boldsymbol{q}_1$
  - Only valid for planes.

Degrees of freedom of $\boldsymbol{F}$

- $\boldsymbol{F}$ has 9 numbers, and is scale invariant.
- $[\boldsymbol{t}]_\times$ has rank 2, and thus $\boldsymbol{F}$ has as well.
    - In other terms: $\det(\boldsymbol{F}) = 0$.
- You will estimate $\boldsymbol{F}$ in week 10.

Triangulation

Triangulation



We've seen the same point in many (known) cameras and want to find the point in 3D.

Because of noise, there is not always a solution without error.

Triangulation problem

Consider a projection matrix

$$\mathcal{P}_i = \begin{bmatrix} \boldsymbol{p}_i^{(1)} \\ \boldsymbol{p}_i^{(2)} \\ \boldsymbol{p}_i^{(3)} \end{bmatrix}$$

Triangulation equations

Projection gives the pixels

$$\boldsymbol{q}_i = \begin{bmatrix} s_i x_i \\ s_i y_i \\ s_i \end{bmatrix} = \mathcal{P}_i \boldsymbol{Q} = \begin{bmatrix} \boldsymbol{p}_i^{(1)} \boldsymbol{Q} \\ \boldsymbol{p}_i^{(2)} \boldsymbol{Q} \\ \boldsymbol{p}_i^{(3)} \boldsymbol{Q} \end{bmatrix}$$

This is two constraints (x_i, y_i) in three equations.

Triangulation equations

As $s_i = \boldsymbol{p}_i^{(3)}\boldsymbol{Q}$, we have

$$\left(\boldsymbol{p}_i^{(3)}\boldsymbol{Q}\right)\begin{bmatrix} x_i \\ y_i \end{bmatrix} = \begin{bmatrix} \boldsymbol{p}_i^{(1)}\boldsymbol{Q} \\ \boldsymbol{p}_i^{(2)}\boldsymbol{Q} \end{bmatrix}$$

rearranged into

$$\begin{aligned} \mathbf{0} &= \begin{bmatrix} \boldsymbol{p}_i^{(3)} x_i - \boldsymbol{p}_i^{(1)} \\ \boldsymbol{p}_i^{(3)} y_i - \boldsymbol{p}_i^{(2)} \end{bmatrix} \boldsymbol{Q} \\ &= \boldsymbol{B}^{(i)}\boldsymbol{Q} \end{aligned}$$

Triangulation solution

Define $\boldsymbol{B}$

$$\boldsymbol{B} = \begin{bmatrix} \boldsymbol{B}^{(1)} \\ \boldsymbol{B}^{(2)} \\ \vdots \\ \boldsymbol{B}^{(n)} \end{bmatrix} = \begin{bmatrix} \boldsymbol{p}_1^{(3)} x_1 - \boldsymbol{p}_1^{(1)} \\ \boldsymbol{p}_1^{(3)} y_1 - \boldsymbol{p}_1^{(2)} \\ \boldsymbol{p}_2^{(3)} x_2 - \boldsymbol{p}_2^{(1)} \\ \boldsymbol{p}_2^{(3)} y_2 - \boldsymbol{p}_2^{(2)} \\ \vdots \end{bmatrix}.$$

Use SVD to find $\arg\min_{\boldsymbol{Q}} \left\| \boldsymbol{B}\boldsymbol{Q} \right\|_2, \quad$ s.t. $\left\| \boldsymbol{Q} \right\|_2 = 1.$

Linear algorithm, hmm

- $\arg\min_{\boldsymbol{Q}} \|\boldsymbol{BQ}\|_2, \quad \text{s.t. } \|\boldsymbol{Q}\|_2 = 1.$
- This is a linear algorithm used to solve the problem.
- Which error would we actually like to minimize?

Linear algorithm, hmm

- $\arg\min_{\boldsymbol{Q}} \|\boldsymbol{BQ}\|_2\,, \quad \text{s.t. } \|\boldsymbol{Q}\|_2 = 1.$
- This is a linear algorithm used to solve the problem.
- Which error would we actually like to minimize?
    - Depends why we believe that there are errors
    - Errors in the observed pixel location

Triangulation: ideal error

- Let $\begin{bmatrix}\tilde{x}_i & \tilde{y}_i\end{bmatrix}$ and $\begin{bmatrix}x_i & y_i\end{bmatrix}$ refer to the observed and projected pixel coordinates, respectively.
- Which error would we like to minimize?

$$e_{\text{ideal}} = \sum_i \left\| \begin{bmatrix}\tilde{x}_i \\ \tilde{y}_i\end{bmatrix} - \begin{bmatrix}x_i \\ y_i\end{bmatrix} \right\|_2^2$$

Triangulation uh oh!

- Which error are actually we minimizing?

$$
\begin{aligned}
e_{\text{algebraic}} &= \textstyle\sum_i \|\boldsymbol{B}_i \boldsymbol{Q}\|_2^2 \\
&= \sum_i \left\| \begin{bmatrix} \boldsymbol{p}_i^{(3)} \tilde{x}_i - \boldsymbol{p}_i^{(1)} \\ \boldsymbol{p}_i^{(3)} \tilde{x}_i - \boldsymbol{p}_i^{(2)} \end{bmatrix} \boldsymbol{Q} \right\|_2^2 \\
&= \sum_i \left\| \underbrace{\boldsymbol{p}_i^{(3)} \boldsymbol{Q}}_{s_i} \begin{bmatrix} \tilde{x}_i \\ \tilde{x}_i \end{bmatrix} - \begin{bmatrix} \boldsymbol{p}_i^{(1)} \\ \boldsymbol{p}_i^{(2)} \end{bmatrix} \boldsymbol{Q} \right\|_2^2 \\
&= \sum_i \left\| s_i \begin{bmatrix} \tilde{x}_i \\ \tilde{y}_i \end{bmatrix} - s_i \begin{bmatrix} x_i \\ y_i \end{bmatrix} \right\|_2^2
\end{aligned}
$$

Error terms compared

- We can rewrite the terms slightly again

$$
\begin{aligned}
e_{\text{ideal}} &= \sum_i \quad (\tilde{x}_i - x_i)^2 + \quad (\tilde{y}_i - y_i)^2 \\
e_{\text{algebraic}} &= \sum_i s_i^2(\tilde{x}_i - x_i)^2 + s_i^2(\tilde{y}_i - y_i)^2
\end{aligned}
$$

- s_i is larger for cameras that are further from $\boldsymbol{Q}$

Linear vs non-linear algorithms

- e_{ideal} can be minimized using non-linear optimization
- Linear algorithms are very fast.
  - Only minimizes an algebraic error.
- We can estimate many things using linear algorithms
  - Triangulation, homography, fundamental matrix, projection matrix
  - They all have the problem that they don't minimize the exact error we desire
- Linear algorithms are acceptable in most cases.
- When high accuracy is desired initialize the non-linear optimization with the linear solution.

Learning objectives

After this lecture you should be able to:

- Derive and apply the fundamental matrix in computer vision
- Derive and apply the essential matrix in computer vision
- Implement the linear algorithm for triangulation
- Explain the pros and cons of using a linear algorithm

Exercise time!