Epipolar geometry
The fundamental matrix and the tensor

Robert Laganière
University of Ottawa
laganier@site.uottawa.ca
www.site.uottawa.ca/~laganier

Image formation through lenses

- The lens equation is: $$\frac{1}{f} = \frac{1}{do} + \frac{1}{di}$$

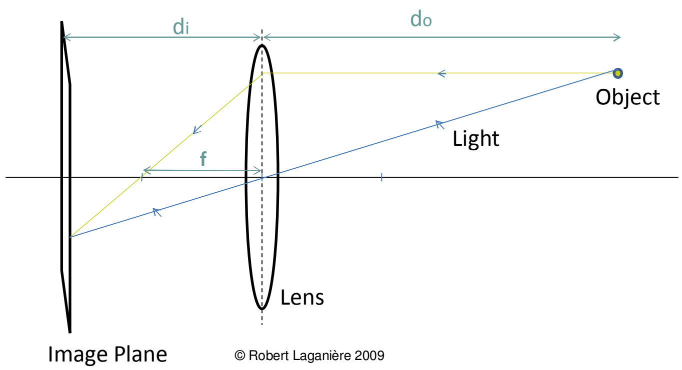
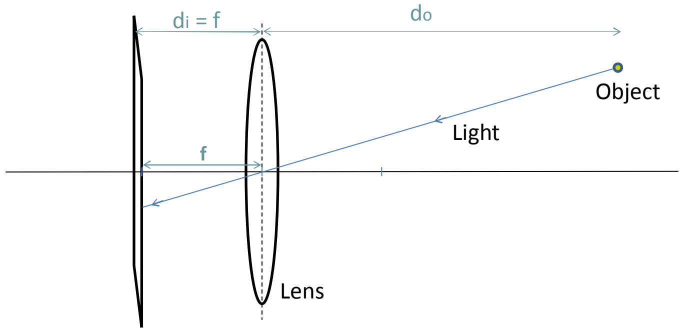


Image formation through lenses

- Most of the time we work at infinity i.e. do>>di
  - the image plane is at f
- We can assume a very narrow aperture
- here we are interested only in determining where the scene elements will project; intensity and color do not matter
- This is a pin-hole camera



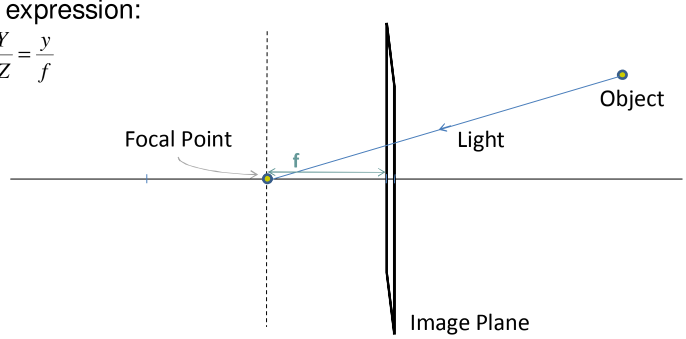
The pin-hole camera model

- With the reference frame located at the projection center, the basic projective equation is built from this simple expression:

$$\frac{Y}{Z} = \frac{y}{f}$$

- We therefore have:

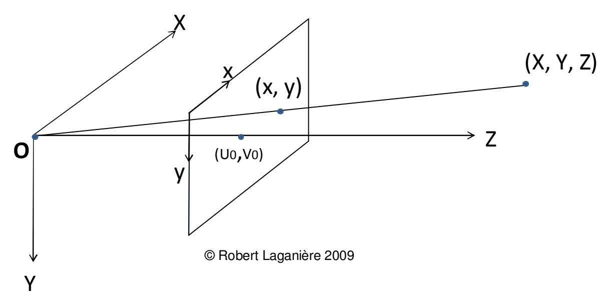
$$x = f\,\frac{X}{Z} \qquad y = f\,\frac{Y}{Z}$$



The pin-hole camera model

- On the image plane we work in pixel coordinates
  - If ℓ_x and ℓ_y are the horizontal and vertical pixel sizes and if (u0,v0) is the pixel coordinate of the image plane center, then the image of point X in pixels is:

$$u = \frac{f}{\ell_x}\frac{X}{Z} + uo = f_x \frac{X}{Z} + uo \qquad v = \frac{f}{\ell_y}\frac{X}{Z} + v_o = f_y \frac{X}{Z} + v_o$$



The pin-hole camera model

- If $\mathbf{X}$=[X,Y,Z,1]^T and $\mathbf{x}$=[u,v,1]^T then we have:

$$\mathbf{x} = \mathbf{K}[\mathbf{I} \mid \mathbf{0}]\mathbf{X} = \begin{bmatrix} f_x & 0 & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix}[\mathbf{I} \mid \mathbf{0}]\mathbf{X}$$

  - This equality is up to a scale factor since we used homogenous coordinates for **x**
  - **K** is the (invertible) calibration matrix
  - Here the world reference frame is at the camera focal point

(homogenous coordinates)

- **x**=$[x,y,1]^T$ is a 2D point
- **l**=$[a,b,c]^T$ defines a 2D line
- point **x** lies on line **l** if:

$$\mathbf{x}^T\mathbf{l} = 0$$

- The intersection of two lines is: $\mathbf{l} \times \mathbf{l}'$

- The line that passes through two points is given by:

$$\mathbf{x} \times \mathbf{x}'$$

The pin-hole camera model

- When the camera is at general position, we have:

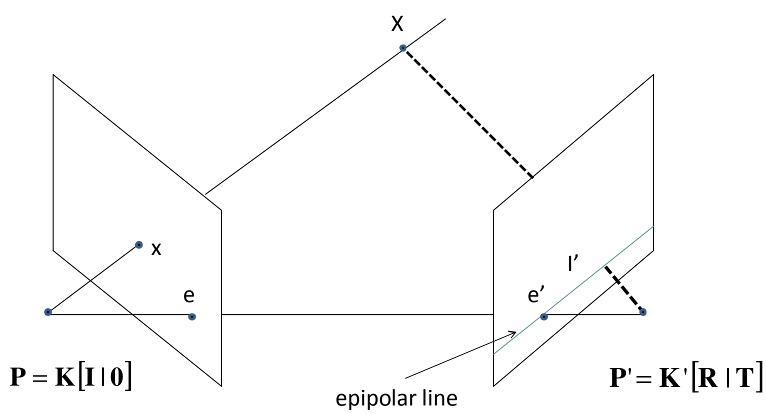
$$\mathbf{x} = \mathbf{K}[\mathbf{R} \mid \mathbf{T}]\mathbf{X} = \mathbf{P}\mathbf{X}$$

  - **R** and **T** are the rotation and the translation required to map a 3D point from world coordinates to camera coordinates
  - If **C** is the coordinate of the camera center then the translation **–RC**
  - **P** is the 3x4 projection matrix
    - It has 11 degrees of freedom (3 rotations, 3 translations, 2 principal points, 1 focal length, 1 pixel ratio, 1 skew)

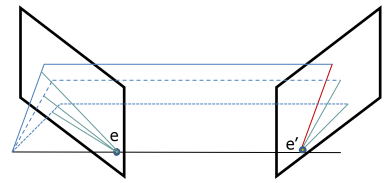
The pin-hole camera model

- The camera is said to be calibrated if we know the relation between world coordinates and pixel coordinates (up to a scale)
  - That is basically if we know the focal length in pixel units (and the location of the principal point)
  - A more accurate camera model would also include distortion parameters in the projection matrix

The geometry of two cameras

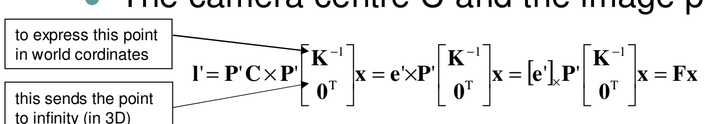


The geometry of two cameras



The fundamental matrix

- We can derive the algebraic expression of the epipolar line of image point **x** by projecting two points of the projection ray:
  - The camera centre C and the image point **x**



$$\mathbf{l'} = \mathbf{P'C} \times \mathbf{P'}\begin{bmatrix}\mathbf{K}^{-1} \\ \mathbf{0}^{\mathrm{T}}\end{bmatrix}\mathbf{x} = \mathbf{e'} \times \mathbf{P'}\begin{bmatrix}\mathbf{K}^{-1} \\ \mathbf{0}^{\mathrm{T}}\end{bmatrix}\mathbf{x} = [\mathbf{e'}]_\times \mathbf{P'}\begin{bmatrix}\mathbf{K}^{-1} \\ \mathbf{0}^{\mathrm{T}}\end{bmatrix}\mathbf{x} = \mathbf{Fx}$$

  - **F** is the fundamental matrix and because x’ lies on the epipolar line, then:

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

The fundamental matrix

- F is up to a scale factor and det(F) = 0
  - F is a mapping from a 2D plane to a 1D pencil of lines
  - It has two non-null eigenvalues
- F is of rank 2 with 7 degrees of freedom
- The epipoles correspond to the left and right null space of F because $\mathbf{x}'^T \mathbf{F}\mathbf{e} = 0$ for all **x'**

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

- The epipoles are also the projection of the **T** vector

$$\mathbf{e}' = \mathbf{K}'\mathbf{T} \qquad \mathbf{e} = \mathbf{K}\mathbf{R}^T\mathbf{T}$$

- The transpose of F is the fundamental matrix of (P',P)

Estimation of the fundamental matrix

- For each point correspondence, you can write the epipolar constraint
  - Which gives you a linear equation with the elements of F as unknown
  - Set the 9th element of F to 1 to fix the scale
- With 8 point correspondences you have enough linear equations to solve your matrix
  - But you will not get det(F)=0 necessarily
  - The epipolar lines will not intersect at a unique point
  - Solution: force the 3rd eigenvalue to 0

Estimation of the fundamental matrix

- The normalized 8-point algorithm:
  - Translate your points such that the centroid is at (0,0)
  - Scale your points such that the RMS distance is sqrt(2)
  - This way the system of equations is better conditioned
- If you have more points, it is just better
  - You perform an algebraic minimization of the over-determined system of equations
- A non-linear 7-point algorithm also exist
  - Very useful in robust estimation of F

Estimation of the fundamental matrix

- The previous algorithm consists in an algebraic minimization
- To optimally minimize the geometric error you have to use:

$$\sum_i dist(\mathbf{x}_i, \tilde{\mathbf{x}}_i)^2 + dist(\mathbf{x}'_i, \tilde{\mathbf{x}}'_i)^2$$

  - that minimizes the distance between the image point and their 'true' position on the epipolar line
  - This is the Gold Standard algorithm
  - Use algebraic minimization as an initial estimate

Estimation of the fundamental matrix

- You can also use an approximation of the geometric error:

$$\sum_i dist(\mathbf{x}_i, \mathbf{F}^{\mathrm{T}}\mathbf{x}'_i)^2 + dist(\mathbf{x}'_i, \mathbf{F}\mathbf{x}'_i)^2$$

  - where the distance between an image point and its epipolar line is given by:

$$dist(\mathbf{x}, \mathbf{F}^{\mathrm{T}}\mathbf{x}') = \mathbf{x}'^{\mathrm{T}}\,\mathbf{F}\mathbf{x} / ((\mathbf{F}\mathbf{x})_1^2 + (\mathbf{F}\mathbf{x})_2^2)$$

  - This distance is useful in measuring the support in a RANSAC scheme

The fundamental matrix

- We have, from previous derivations:

$$\mathbf{F} = [\mathbf{e}']_\times \mathbf{P}' \begin{bmatrix} \mathbf{K}^{-1} \\ \mathbf{0}^{\mathrm{T}} \end{bmatrix} = [\mathbf{K}'\mathbf{T}]_\times \mathbf{K}'\mathbf{R}\mathbf{K}^{-1} = \mathbf{K}'^{-\mathrm{T}} [\mathbf{T}]_\times \mathbf{R}\mathbf{K}^{-1}$$

- which allows us to express the epipolar constraint as:

$$\mathbf{x}'^{\mathrm{T}}\,\mathbf{K}'^{-\mathrm{T}} [\mathbf{T}]_\times \mathbf{R}\mathbf{K}^{-1}\mathbf{x} = 0$$

like if the image points would be expressed in world coordinates

The Essential matrix

- Two cameras are said to be calibrated if you can express the pixel coordinates of both cameras into common world coordinates
- E is a special case of F which expresses the same epipolar relation but with the K matrices removed

$$\mathbf{E} = \mathbf{K}'^{T}\mathbf{F}\mathbf{K} = [\mathbf{T}]_{\times}\mathbf{R}$$

- E is then built only from R and T
- Therefore if we know E we can obtain R and T

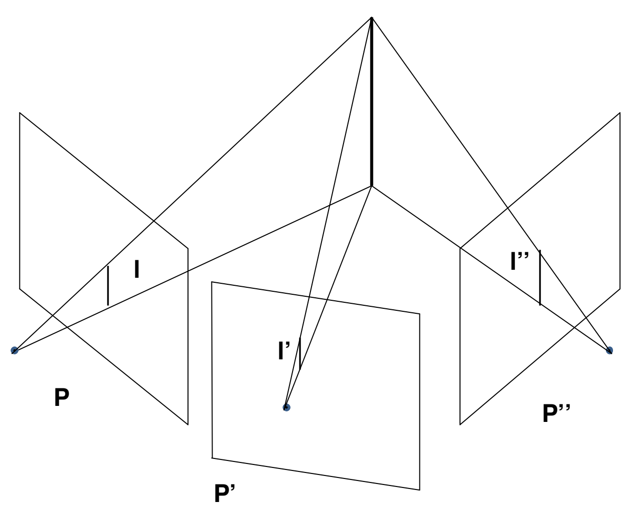
Estimation of the Essential matrix

- If your cameras are calibrated, then you can estimate E using the 8-point algorithm
- However, E has only 5 degrees of freedom
  - 3 from R and 3 from T minus 1 for the scale
  - We therefore have additional constraints on E
    - the first two eigenvalues must be equal (third one still null)
  - you have to impose this condition to make sure you obtain a valid E

The geometry of three cameras

- Suppose we have three cameras observing a line in space: $\mathbf{P} = [\mathbf{I} \mid \mathbf{0}]$ $\quad \mathbf{P}' = [\mathbf{A} \mid \mathbf{a}_4]$ $\quad \mathbf{P}'' = [\mathbf{B} \mid \mathbf{b}_4]$
- Each projective relation between this line and its image induces a 3D plane that is obtained by back-projection: $\mathbf{m} = \mathbf{P}^T \mathbf{l}$
- These three planes must intersect at a common line:
  - this is a geometric incidence

The geometry of three cameras



The geometry of three cameras

- We can then create a 4x3 matrix that expresses this line intersection constraint:

$$[\mathbf{m}, \mathbf{m}', \mathbf{m}'']^{\mathrm{T}} \mathbf{X} = \mathbf{M}^{\mathrm{T}} \mathbf{X} = \begin{bmatrix} \mathbf{l} & \mathbf{A}^{\mathrm{T}} \mathbf{l}' & \mathbf{B}^{\mathrm{T}} \mathbf{l}'' \\ 0 & \mathbf{a}_4^{\mathrm{T}} \mathbf{l}' & \mathbf{b}_4^{\mathrm{T}} \mathbf{l}'' \end{bmatrix}^{\mathrm{T}} \mathbf{X} = 0$$

  - for all **X** on the 3D line
- This matrix has rank 2 because if we fixed the first two planes (columns), then there remains only one degree of liberty for the last plane
  - or say otherwise, we can find two independent points for which $M^T X=0$ (i.e. 2-dimensional null space)
  - We therefore have a linear dependence:

$$k\mathbf{m} + k'\mathbf{m}' + k''\mathbf{m}'' = 0$$

The geometry of three cameras

- Considering the last row of **M** we have:

$$0 + k'\mathbf{a}_4^{\mathrm{T}}\mathbf{l}' + k''\mathbf{b}_4^{\mathrm{T}}\mathbf{l}'' = 0$$

$$k' = \mathbf{b}_4^{\mathrm{T}}\mathbf{l}'', k'' = -\mathbf{a}_4^{\mathrm{T}}\mathbf{l}'$$

- For the other columns we therefore have:

$$l_i = \mathbf{l}''^{\mathrm{T}}(\mathbf{b}_4\mathbf{a}_i^{\mathrm{T}})\mathbf{l}' - \mathbf{l}'^{\mathrm{T}}(\mathbf{a}_4\mathbf{b}_i^{\mathrm{T}})\mathbf{l}'' = \mathbf{l}'^{\mathrm{T}}(\mathbf{a}_i\mathbf{b}_4^{\mathrm{T}} - \mathbf{a}_4\mathbf{b}_i^{\mathrm{T}})\mathbf{l}'' = \mathbf{l}'\mathbf{T}_i\mathbf{l}''$$

- This is the line-line-line incidence relation which can be expressed in matrix form as:

$$\mathbf{l}^{\mathrm{T}} = \mathbf{l}'^{\mathrm{T}}[\mathbf{T}_1, \mathbf{T}_2, \mathbf{T}_3]\mathbf{l}''$$

  - These three matrices are known as the trifocal tensor

Incidence relations in three views

- Different incidence relations exist:
  - line-line-line (2 equations)
  - point-line-line (1 equation)
  - point-line-point (2 equations)
  - point-point-line (2 equations)
  - point-point-point (4 equations)
- All these equations are linear in the entries of the trifocal tensor

Estimation of the tensor

- A tensor is 3x3x3 matrix (27 elements)
- A tensor is geometrically valid if it is built from three projection matrices
  - It thus satisfies 8 algebraic constraints
  - A minimum of 6 points is required to solve
- Using the incidence relations it can be estimated linearly from 7 point triplets
  - Normalize your data
  - The resulting solution will not satisfy the constraints

Estimation of the tensor

- To obtain a valid tensor, you have to perform algebraic minimization using the relation:

$$\mathbf{T}_i = \mathbf{a}_i \mathbf{b}_4^{\mathrm{T}} - \mathbf{a}_4 \mathbf{b}_i^{\mathrm{T}}$$

- The optimal estimate is obtained using the Gold standard algorithm

$$\sum_i dist(\mathbf{x}_i, \tilde{\mathbf{x}}_i)^2 + dist(\mathbf{x}'_i, \tilde{\mathbf{x}}'_i)^2 + dist(\mathbf{x}''_i, \tilde{\mathbf{x}}''_i)^2$$

  - Which requires to extract the projection matrices from the tensor
    - from the epipoles

Image point transfer

- With the geometry between cameras known, it becomes possible to transfer two corresponding image points to a third view
- This can be done by intersecting the epipolar lines assuming you have the two F matrices:

$$\mathbf{x}_3 = (\mathbf{F}_{31}\mathbf{x}_1) \times (\mathbf{F}_{32}\mathbf{x}_2)$$

- But this is not very accurate when the epipolar lines are almost parallel
  - and not feasible for matches on the trifocal plane

Image point transfer

- It is also possible to use the incidence relations
  - Using point-point-point would work
  - But using point-line-point is simpler
- If $\mathbf{x}_1$ and $\mathbf{x}_2$ are known correspondences in views 1 and 2, then the position of $\mathbf{x}_3$ can be found as follows:
  - Use $\mathbf{F}_{21}$ to generate an image line in view 2 perpendicular to the epipolar line
    - Make sure your matches are consistent with this **F** and that your **F** is consistent with your tensor
  - The transferred point $\mathbf{x}_3$ is then given by the point-line-point incidence relations

The classic book

- Mutliple view geometry
  - Richard Hartley and Andrew Zisserman
  - Second edition
    - Cambridge University Press 2003