Geometry in Computer Vision

Spring 2010

Lecture 5A

Three-view geometry

Three-view geometry

- We take 3 images of the world

- How are corresponding points related?
- What relations between the cameras can be inferred from these correspondences?
- Are other correspondences than for points possible?

Epipolar geometry

- Epipolar geometry can be applied to pairs of cameras: (1,2) (2,3) (3,1)
- Gives fundamental matrices $\mathbf{F}_{12}$, $\mathbf{F}_{23}$, $\mathbf{F}_{31}$
- These, however, are not independent!
- If they are independently estimated, they may not be consistent (meaning what?)

Consistent three-view epipolar geometry

- From each of the fundamental matrices, we can derive a pair of *canonical cameras* (see lecture 2)
  - $\mathbf{F}_{12} \rightarrow \mathbf{C}_1, \mathbf{C}_2$
  - $\mathbf{F}_{23} \rightarrow \mathbf{C}'_2, \mathbf{C}_3$
  - $\mathbf{F}_{31} \rightarrow \mathbf{C}'_3, \mathbf{C}'_1$
- These are well-defined up to a 3D homography transformation
- If the fundamental matrices are mutually consistent, it must be possible to find such 3D homography transformations such that $\mathbf{C}_1=\mathbf{C}'_1$, $\mathbf{C}_2=\mathbf{C}'_2$, $\mathbf{C}_3=\mathbf{C}'_3$
- This will not be the case, in general, if the fundamental matrices are estimated independently!

Consistent three-view epipolar geometry

- A sufficient condition for consistent three-view epipolar geometry can be formulated as follows:

From $\mathbf{F}_{ij} \rightarrow$ epipoles $\mathbf{e}_{ij}$ and $\mathbf{e}_{ji}$ (how?)

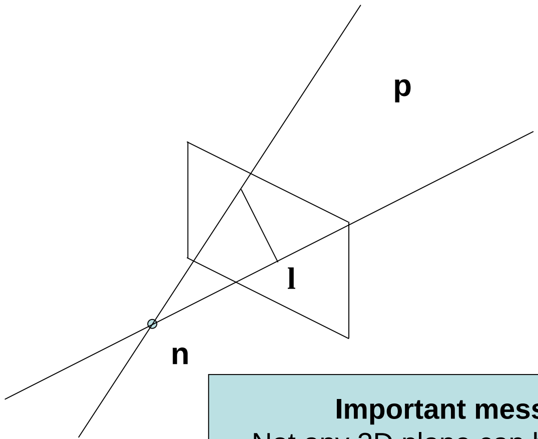
$$\mathbf{e}_{13}^T\mathbf{F}_{12}\mathbf{e}_{23} = \mathbf{e}_{21}^T\mathbf{F}_{23}\mathbf{e}_{31} = \mathbf{e}_{32}^T\mathbf{F}_{31}\mathbf{e}_{12} = 0$$

- How can we obtain such **F**'s?

2D lines and 3D planes

- Let $\mathbf{l}$ be the dual homogeneous coordinates of a line in an image that depicts the 3D world through camera matrix **C**
- If we project $\mathbf{l}$ out in the 3D world through the camera center **n**, we get a plane **p**

2D lines and 3D planes



It follows that

$$\mathbf{p} \sim \mathbf{C}^T\mathbf{l}$$
$$\mathbf{l} \sim \mathbf{C}^{+T}\mathbf{p}$$

(why?)

Important message:
Not any 3D plane can be projected onto a line in the image (which can?)

Line correspondences

- In three views, it turns out to be easer to start looking at line correspondences
- Let **L** be a 3D line that is projected into the three cameras as lines $\mathbf{l}_1$, $\mathbf{l}_2$, $\mathbf{l}_3$
- The three lines generate three planes:

$$\mathbf{p}_1 = \mathbf{C}_1^T\,\mathbf{l}_1$$
$$\mathbf{p}_2 = \mathbf{C}_2^T\,\mathbf{l}_2$$
$$\mathbf{p}_3 = \mathbf{C}_3^T\,\mathbf{l}_3$$

Line correspondences

- These planes must intersect at the line **L**
- The dual Plücker coordinates of **L** are given, e.g., by $\mathbf{p}_2\mathbf{p}_3^T - \mathbf{p}_3\mathbf{p}_2^T$ (lecture 1)
- Combined with $\mathbf{n}_1$ this gives a plane that projects into a line in image 1:

The plane = $(\mathbf{p}_2\mathbf{p}_3^T - \mathbf{p}_3\mathbf{p}_2^T)\mathbf{n}_1$
line in image 1 = $\mathbf{C}_1^{+T}(\mathbf{p}_2\mathbf{p}_3^T - \mathbf{p}_3\mathbf{p}_2^T)\ \mathbf{n}_1$

- This line must be $\mathbf{l}_1$!

Line correspondences

- We summarize

$$\mathbf{l}_1 \sim \mathbf{C}_1^{+T}(\mathbf{p}_2\mathbf{p}_3^T - \mathbf{p}_3\mathbf{p}_2^T)\ \mathbf{n}_1$$

$$\mathbf{l}_1 \sim \mathbf{C}_1^{+T}(\mathbf{C}_2^T\mathbf{l}_2)(\mathbf{l}_3^T\mathbf{C}_3\mathbf{n}_1) - \mathbf{C}_1^{+T}(\mathbf{C}_3^T\mathbf{l}_3)(\mathbf{l}_2^T\mathbf{C}_2\mathbf{n}_1)$$

Trifocal tensor

- Each element of $\mathbf{l}_1$ is a quadratic form in $\mathbf{l}_2$ and $\mathbf{l}_3$:

$$(\mathbf{l}_1)_i \sim \mathbf{l}_2^T\ \mathbf{T}_i\ \mathbf{l}_3$$

Here we mean that the r.h.s. is prop to the l.h.s. with the same scaling for all i=1, 2, 3

- The three 3×3 matrices $\mathbf{T}_i$ are given by $\mathbf{C}_2$, $\mathbf{C}_3$, and $\mathbf{n}_1$ (the last is derived from $\mathbf{C}_1$)
- Together they form a $3 \times 3 \times 3$ *trifocal tensor* $\mathcal{T}$

Trifocal tensor

- We write the last relation as

$$\mathbf{l}_1 \sim \mathbf{l}_2^T\ [\mathbf{T}_1, \mathbf{T}_2, \mathbf{T}_3]\ \mathbf{l}_3$$

- The trifocal tensor $\mathcal{T}$ is the three matrices

$$[\mathbf{T}_1, \mathbf{T}_2, \mathbf{T}_3]$$

- $\mathcal{T}$ is an element of a projective space (why?)

Trifocal tensors

- $\mathscr{T}$ is derived by considering how lines in the three images are related
- It is, however, not derived in a symmetric way:
  – it produces a line $\mathbf{l}_1$ specifically in image 1
- There must be three trifocal tensors:
  – one for each of the three images
- In the following: $\mathscr{T}$ refers to the one that produces $\mathbf{l}_1$ (unless stated otherwise)

Degrees of freedom and internal constraints

- $\mathscr{T}$ has 3^3=27 elements
  – It has 27-1=26 d.o.f. as a general projective element
- It is computed from $\mathbf{C}_1$, $\mathbf{C}_2$, $\mathbf{C}_3$
  – Each $\mathbf{C}_k$ has 11 degrees of freedom
  – In total $3 \times 11 = 33$ degrees of freedom
  - $\mathscr{T}$ is independent of the 3D coordinate system
    ⇒ invariant to any 3D homography transformation **H**
    - **H** has 15 degree of freedom
  - $\mathscr{T}$ has 33-15 = 18 d.o.f.
- $\mathscr{T}$ must satisfy 26-18 = 8 internal constraints to be properly related to 3 views

Point-line-line correspondence

- Let **x** be a point on **L**, projected into image 1 as $\mathbf{y}_1$
- $\mathbf{y}_1$ must lie on $\mathbf{l}_1$:

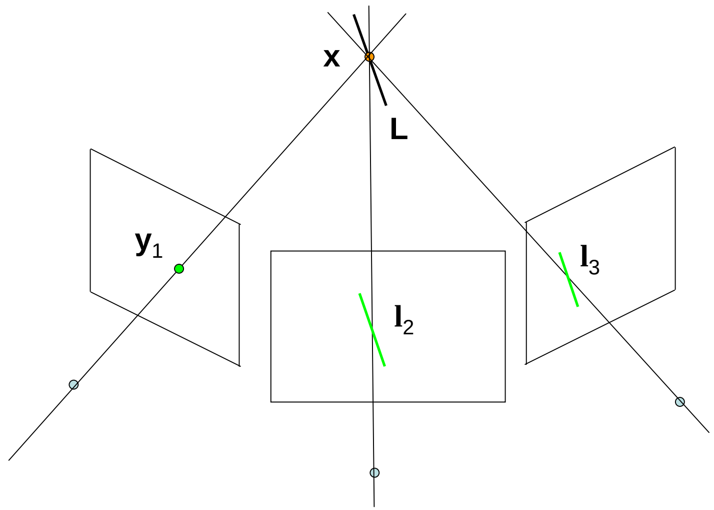
$$\mathbf{y}_1 \cdot \mathbf{l}_1 = 0$$

- With $\mathbf{y}_1 = (y_1, y_2, y_3)$ we get

$$0 = \mathbf{l}_2^T (\mathbf{T}_1 y_1 + \mathbf{T}_2 y_2 + \mathbf{T}_3 y_3)\, \mathbf{l}_3$$

- $\mathscr{T}$ gives a relation between a point in image 1 and *corresponding* lines in image 2 and 3

Point-line-line correspondence



Point-point-point correspondences

- Start with a 3D point $\mathbf{x}$, projected onto $\mathbf{y}_k$ in image k, k=1, 2, 3
- Consider the set of all 3D lines $\mathbf{L}$ that intersect $\mathbf{x}$
- $\mathbf{L}$ is projected onto lines $\mathbf{l}_2$ and $\mathbf{l}_3$ in images 2 and 3, respectively
- The set of all such $\mathbf{L}$ produces a set of lines $\mathbf{l}_2$ and a set of lines $\mathbf{l}_3$
- All lines $\mathbf{l}_2$ intersect $\mathbf{y}_2$ and all lines $\mathbf{l}_3$ intersect $\mathbf{y}_3$
- $\mathbf{l}_2 \sim [\mathbf{y}_2]_\times \mathbf{c}_2$ for all possible $\mathbf{c}_2 \in R^3$,
  $\mathbf{l}_3 \sim [\mathbf{y}_3]_\times \mathbf{c}_3$ for all possible $\mathbf{c}_3 \in R^3$

Point-point-point correspondences

- We summarize

$$0 = \mathbf{c}_2^T[\mathbf{y}_2]_\times^T(\mathbf{T}_1 y_1+\mathbf{T}_2 y_2+\mathbf{T}_3 y_3)\ [\mathbf{y}_3]_\times \mathbf{c}_3$$

  for all $\mathbf{c}_2, \mathbf{c}_3 \in R^3$

- This implies

$$\mathbf{0} = [\mathbf{y}_2]_\times^T(\mathbf{T}_1 y_1+\mathbf{T}_2 y_2+\mathbf{T}_3 y_3)\ [\mathbf{y}_3]_\times$$

Point-point-point correspondences

- For corresponding points in the three views, $\mathbf{y}_1$, $\mathbf{y}_2$, $\mathbf{y}_3$, we get 9 matching constraints

3 × 3 zero matrix

$$[\mathbf{y}_2]_\times^T(\mathbf{T}_1 y_1+\mathbf{T}_2 y_2+\mathbf{T}_3 y_3)\ [\mathbf{y}_3]_\times = \mathbf{0}$$

- But only 4 that are linearly independent (why?)

The trifocal tensor

Given that $\mathscr{T}$ is given

- It provides 4 linearly independent point-point-point constraints
- It provides 1 point-line-line constraint
- It provides 2 line-line-line constraints (how?)
- It provides 3 point-point-line constraints (how?)

The trifocal tensors

There are 3 trifocal tensors:

- Each gives a unique point-line-line constraint (with the point in a distinct view)
- They provide up to $3 \times 4 = 12$ linearly independent point-point-point constraints
- They provide up to $3 \times 2 = 6$ linearly independent line-line-line constraints
- There are, however, linearly dependence among the last two constraints, reducing them to smaller numbers

F and **C** from $\mathscr{T}$

Given $\mathscr{T}$ it possible to extract

- the three fundamental matrices $\mathbf{F}_{12}$, $\mathbf{F}_{23}$, $\mathbf{F}_{31}$
  - See HZ
  - From these **F**s we get all the epipoles
  - These **F**s are three-view consistent!
- the three camera matrices $\mathbf{C}_1$, $\mathbf{C}_2$ and $\mathbf{C}_3$
  - See HZ

Estimation of $\mathscr{T}$

Linear estimation:

- Each triplet of corresponding points provides 4 linear constraints in $\mathscr{T}$
- 7 triplets of corresponding points gives $7 \times 4 = 28$ linear constraints in $\mathscr{T}$
  - This is sufficient for determining $\mathscr{T}$ by solving a linear equation (why?)
- Remember: Hartley-normalization!
- This estimated $\mathscr{T}$ may not be a proper trifocal tensor

Estimation of $\mathscr{T}$

Non-linear estimation of $\mathscr{T}$:

- Find initial estimate of $\mathscr{T}$ using a linear method
- Reconstruct the three cameras
- Triangulate 3D points from corresponding image points
- Minimize the re-projection error in the images over the 3D points and the camera matrices (Levenberg-Marquardt)

F vs **T**

- Represents a point-point constraint
- Has 7 d.o.f.
- Uniquely represents the uncalibrated epipolar geometry
- Can be estimated linearly from 8 correspondences
- Internal constraint is trivial
- Relations to **C** and **e** are trivial
- $\mathbf{F}_{12}$ does not anything about $\mathbf{F}_{23}$ and $\mathbf{F}_{31}$

- Represents a point-line-line constraint, or 4 point-point-point constraints
- Has 18 d.o.f.
- Uniquely represents the uncalibrated three-view geometry
- Can be estimated linearly from 7 correspondences
- Internal constraints are non-trivial
- Relations to **C**, **F**, and **e** exist but are not straight-forward
- Relations between one trifocal tensor and the other two exists but are not stright-forward

General conclusions for the 3 view case

- The algebraic desciption of the three-view geometry is more complicated than the epipolar geometry
  – Internal constraints for $\mathscr{T}$?
  – How can they be enforced?
  – Simpler relations between $\mathscr{T}$ and other geometric objects?
  – Minimal parameterization of $\mathscr{T}$?

The key to three-view geometry

- [Nordberg, *A minimal parameterization of the trifocal tensor*, CVPR 2009]
- We know that F can be decomposed as

$\mathbf{F} = \mathbf{U}\,\mathbf{S}\,\mathbf{V}^T$

U and **V** are orthogonal
S is diagonal of rank 2

Important message:
⇐ This means that if we transform the two image spaces by means of **U** and **V**, respectively, then the fundamental matrix is simply **S**

- Can we find a similar decomposition of $\mathscr{T}$?

The key to three-view geometry

Main result:

- We can always find
  – (non-unique) orthogonal homography transformations of the image spaces
  – A general 3D homography transformation of the 3D space such that

$$\mathbf{C}'_1 = [\,\mathbf{I}\,|\,\mathbf{0}\,],$$

$$\mathbf{C}'_2 \sim \begin{pmatrix} c_0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ c_1 & 0 & c_2 & 0 \end{pmatrix}, \quad \mathbf{C}'_3 \sim \begin{pmatrix} c_3 & 0 & 0 & 0 \\ c_4 & c_5 & c_6 & 1 \\ c_7 & c_8 & c_9 & 0 \end{pmatrix}.$$

The key to three-view geometry

- Once these transformations have been applied it follows that

$$\mathbf{T}'_{1,1} = \mathbf{a}_1\mathbf{b}_4^T - \mathbf{a}_4\mathbf{b}_1^T = \begin{pmatrix} 0 & c_0 & 0 \\ -c_3 & -c_4 & -c_7 \\ 0 & c_1 & 0 \end{pmatrix},$$

$$\mathbf{T}'_{2,1} = \mathbf{a}_2\mathbf{b}_4^T - \mathbf{a}_4\mathbf{b}_2^T = \begin{pmatrix} 0 & 0 & 0 \\ 0 & -c_5 & -c_8 \\ 0 & 0 & 0 \end{pmatrix},$$

$$\mathbf{T}'_{3,1} = \mathbf{a}_3\mathbf{b}_4^T - \mathbf{a}_4\mathbf{b}_3^T = \begin{pmatrix} 0 & 0 & 0 \\ 0 & -c_6 & -c_9 \\ 0 & c_2 & 0 \end{pmatrix}.$$

Only 10 non-zero elements

$\mathscr{T}$ can be minimally parameterized by the 3 SO(3) transformations and the 10 non-zero elements

The key to three-view geometry

- The other two trifocal tensors are given by:

$$\mathbf{T}'_{1,2} \sim \begin{pmatrix} 0 & c_2c_5 & c_2c_8 \\ -c_2c_3 & c_1c_6 - c_2c_4 & c_1c_9 - c_2c_7 \\ 0 & -c_1c_5 & -c_1c_8 \end{pmatrix} \quad \mathbf{T}'_{1,3} \sim \begin{pmatrix} 0 & c_6c_8 - c_5c_9 & -c_2c_8 \\ -c_0c_9 & c_4c_9 - c_6c_7 & c_2c_7 - c_1c_9 \\ c_0c_8 & c_5c_7 - c_4c_8 & c_1c_8 \end{pmatrix},$$

$$\mathbf{T}'_{2,2} \sim \begin{pmatrix} 0 & 0 & 0 \\ 0 & -c_0c_2 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \mathbf{T}'_{2,3} \sim \begin{pmatrix} 0 & 0 & 0 \\ 0 & -c_3c_9 & 0 \\ 0 & c_3c_8 & 0 \end{pmatrix},$$

$$\mathbf{T}'_{3,2} \sim \begin{pmatrix} 0 & 0 & 0 \\ 0 & -c_0c_6 & -c_0c_9 \\ 0 & c_0c_5 & c_0c_8 \end{pmatrix}. \quad \mathbf{T}'_{3,3} \sim \begin{pmatrix} 0 & 0 & 0 \\ 0 & c_3c_6 & -c_2c_3 \\ 0 & -c_3c_5 & 0 \end{pmatrix}.$$

2nd order expressions in the canonical cameras or in $\mathscr{T}$

Note dissimilarity to $\mathscr{T}$

The key to three-view geometry

- The fundamental matrices become:

$$\mathbf{F}'_{12} = [\mathbf{e}'_{12}]_\times \mathbf{C}'_1 \mathbf{C}'^{+}_2 = \begin{pmatrix} c_1 & 0 & -c_0 \\ 0 & 0 & 0 \\ c_2 & 0 & 0 \end{pmatrix},$$

$$\mathbf{F}'_{13} = [\mathbf{e}'_{13}]_\times \mathbf{C}'_1 \mathbf{C}'^{+}_3 = \begin{pmatrix} c_7 & 0 & -c_3 \\ c_8 & 0 & 0 \\ c_9 & 0 & 0 \end{pmatrix}.$$

$$\mathbf{F}'_{23} = \begin{pmatrix} c_2c_4c_8 + c_1c_5c_9 - c_2c_5c_7 - c_1c_6c_8 & -c_2c_3c_8 & c_2c_3c_5 \\ c_0c_2c_8 & 0 & 0 \\ c_0c_6c_8 - c_0c_5c_9 & 0 & 0 \end{pmatrix}$$

linear expressions in the canonical cameras or in $\mathscr{T}$

3rd order expressions in the canonical cameras or in $\mathscr{T}$

Note lack of symmetries between the 3 fundamental matrices

The key to three-view geometry

- The paper suggest an algorithm for determining the orthogonal homography transformations of the coordinates for a general $\mathscr{T}$
  - e.g., on that is estimated from a linear method
- These transformations will always be able to set the “0-elements” in **T’**$_k$ to 0 if they are not at the corners
- Constraint enforcement can then be achieved by setting the corner element to 0 and re-transform

The key to three-view geometry

Summary

- Once the orthogonal homography transformatins on the image domains are applied:
  - Three-view geometry is a piece of cake!

- (How does an orthogonal homography transform images?)