Geometry in Computer Vision

Spring 2010

Lecture 6A

6-point geometry

Canonical 3D coordinates

- A set of 6 3D points (in homogeneous coordinates):

$$\begin{pmatrix}\mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_3 & \mathbf{x}_4 & \mathbf{x}_5 & \mathbf{x}_6\end{pmatrix} = \begin{pmatrix} x_{11} & x_{12} & x_{13} & x_{14} & x_{15} & x_{16} \\ x_{21} & x_{22} & x_{23} & x_{24} & x_{25} & x_{26} \\ x_{31} & x_{32} & x_{33} & x_{34} & x_{35} & x_{36} \\ x_{41} & x_{42} & x_{43} & x_{44} & x_{45} & x_{46} \end{pmatrix}$$

Canonical 3D coordinates

- We apply the 3D homography transformation

$$\mathbf{H}_1 = \begin{pmatrix} x_{11} & x_{12} & x_{13} & x_{14} \\ x_{21} & x_{22} & x_{23} & x_{24} \\ x_{31} & x_{32} & x_{33} & x_{34} \\ x_{41} & x_{42} & x_{43} & x_{44} \end{pmatrix}^{-1}$$

to get new 3D coordinates:

$$\begin{pmatrix}\bar{\mathbf{x}}_1 & \bar{\mathbf{x}}_2 & \bar{\mathbf{x}}_3 & \bar{\mathbf{x}}_4 & \bar{\mathbf{x}}_5 & \bar{\mathbf{x}}_6\end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & \bar{x}_{15} & \bar{x}_{16} \\ 0 & 1 & 0 & 0 & \bar{x}_{25} & \bar{x}_{26} \\ 0 & 0 & 1 & 0 & \bar{x}_{35} & \bar{x}_{36} \\ 0 & 0 & 0 & 1 & \bar{x}_{45} & \bar{x}_{46} \end{pmatrix}$$

Canonical 3D coordinates

- We apply another 3D homography transformations on these new coordinates

$$\mathbf{H}_2 = \begin{pmatrix} \bar{x}_{15} & 0 & 0 & 0 \\ 0 & \bar{x}_{25} & 0 & 0 \\ 0 & 0 & \bar{x}_{35} & 0 \\ 0 & 0 & 0 & \bar{x}_{45} \end{pmatrix}^{-1}$$

to get *canonical* 3D coordinates:

$$\begin{pmatrix}\hat{\mathbf{x}}_1 & \hat{\mathbf{x}}_2 & \hat{\mathbf{x}}_3 & \hat{\mathbf{x}}_4 & \hat{\mathbf{x}}_5 & \hat{\mathbf{x}}_6\end{pmatrix} = \begin{pmatrix} \hat{x}_{11} & 0 & 0 & 0 & 1 & \hat{x}_{16} \\ 0 & \hat{x}_{22} & 0 & 0 & 1 & \hat{x}_{26} \\ 0 & 0 & \hat{x}_{33} & 0 & 1 & \hat{x}_{36} \\ 0 & 0 & 0 & \hat{x}_{44} & 1 & \hat{x}_{46} \end{pmatrix}$$

Canonical 3D coordinates

- Since we are dealing with homogeneous coordinates, we can write

$$\begin{pmatrix}\hat{\mathbf{x}}_1 & \hat{\mathbf{x}}_2 & \hat{\mathbf{x}}_3 & \hat{\mathbf{x}}_4 & \hat{\mathbf{x}}_5 & \hat{\mathbf{x}}_6\end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & X \\ 0 & 1 & 0 & 0 & 1 & Y \\ 0 & 0 & 1 & 0 & 1 & Z \\ 0 & 0 & 0 & 1 & 1 & T \end{pmatrix}$$

- Summary: there exists a 3D homography transformation ($\mathbf{H}_2\mathbf{H}_1$) such that the resulting 3D coordinates are as above (always?)
- Note: $\mathbf{H}_2\mathbf{H}_1$ is data dependent
- We here interpret $\mathbf{H}_2\mathbf{H}_1$ as *transforming* coordinates rather than *moving* points

Canonical 2D coordinates

- Project the 6 3D points to a 2D image

$$\begin{pmatrix}\mathbf{y}_1 & \mathbf{y}_2 & \mathbf{y}_3 & \mathbf{y}_4 & \mathbf{y}_5 & \mathbf{y}_6\end{pmatrix} = \begin{pmatrix} y_{11} & y_{12} & y_{13} & y_{14} & y_{15} & y_{16} \\ y_{21} & y_{22} & y_{23} & y_{24} & y_{25} & y_{26} \\ y_{31} & y_{32} & y_{33} & y_{34} & y_{35} & y_{36} \end{pmatrix}$$

- We can do the corresponding coordinate transformation for the 2D coordinates
- We get *canonical* 2D coordinates:

$$\begin{pmatrix}\hat{\mathbf{y}}_1 & \hat{\mathbf{y}}_2 & \hat{\mathbf{y}}_3 & \hat{\mathbf{y}}_4 & \hat{\mathbf{y}}_5 & \hat{\mathbf{y}}_6\end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 1 & u_5 & u_6 \\ 0 & 1 & 0 & 1 & v_5 & v_6 \\ 0 & 0 & 1 & 1 & w_5 & w_6 \end{pmatrix}$$

The camera mapping

- After transformations of the 3D and 2D spaces, we have a camera matrix

$$\hat{\mathbf{C}} = \begin{pmatrix} \hat{c}_{11} & \hat{c}_{12} & \hat{c}_{13} & \hat{c}_{14} \\ \hat{c}_{21} & \hat{c}_{22} & \hat{c}_{23} & \hat{c}_{24} \\ \hat{c}_{31} & \hat{c}_{32} & \hat{c}_{33} & \hat{c}_{34} \end{pmatrix}$$

such that

$$\hat{\mathbf{y}}_k \sim \hat{\mathbf{C}}\,\hat{\mathbf{x}}_k, \quad k = 1, \ldots, 6$$

The camera mapping

- Using the last relation for *k*=1, 2, 3, 4 gives

$$\hat{\mathbf{C}} = \begin{pmatrix} \hat{c}_{11} & 0 & 0 & 1 \\ 0 & \hat{c}_{22} & 0 & 1 \\ 0 & 0 & \hat{c}_{33} & 1 \end{pmatrix}$$

- From *k*=5 and *k*=6 we get

$$\begin{pmatrix} u_5 \\ v_5 \\ w_5 \end{pmatrix} \sim \begin{pmatrix} \hat{c}_{11} + 1 \\ \hat{c}_{22} + 1 \\ \hat{c}_{33} + 1 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} u_6 \\ v_6 \\ w_6 \end{pmatrix} \sim \begin{pmatrix} X\hat{c}_{11} + T \\ Y\hat{c}_{22} + T \\ Z\hat{c}_{33} + T \end{pmatrix}$$

4 equations & 3 unkowns

- The last relation consists of 4 independent equations (why?)
- The last relation includes 3 variables that are unrelated to 3D and 2D coordinates:

$$\hat{c}_{11}, \quad \hat{c}_{22}, \quad \hat{c}_{33}$$

Quan's constraint (I)

- Solving for these "free" variables gives a constraint on the 3D and 2D coordinates:

$$i_1 I_1 + i_2 I_2 + i_3 I_3 + i_4 I_4 + i_5 I_5 + i_6 I_6 = 0$$

with

$$\begin{aligned} i_1 &= w_6(u_5 - v_5) & I_1 &= XY \\ i_2 &= v_6(w_5 - u_5) & I_2 &= XZ \\ i_3 &= u_5(v_6 - w_6) & I_3 &= XT \\ i_4 &= u_6(v_5 - w_5) & I_4 &= YZ \\ i_5 &= v_5(w_6 - u_6) & I_5 &= YT \\ i_6 &= w_5(u_6 - v_6) & I_6 &= ZT \end{aligned}$$

Quan's constraint (II)

- Quan notes that

$$i_1 + i_2 + i_3 + i_4 + i_5 + i_6 = 0$$

$\Rightarrow$ the constraint can be written as

$$i_1 \hat{I}_1 + i_2 \hat{I}_2 + i_3 \hat{I}_3 + i_4 \hat{I}_4 + i_5 \hat{I}_5 = 0$$

$$\begin{aligned} \hat{I}_1 &= XY - ZT \\ \hat{I}_2 &= XZ - ZT \\ \hat{I}_3 &= XT - ZT \\ \hat{I}_4 &= YZ - ZT \\ \hat{I}_5 &= YT - ZT \end{aligned}$$

This form of the constraint is not mentioned in Quan's paper!

Invariants

- Let's look closer at the scalars (*X*,*Y*,*Z*,*T*)
- They depend on the original 6 3D points
- They are, however, invariant to any 3D homography transformation of these points
  - If $\mathbf{H}_2\mathbf{H}_1$ transforms ($\mathbf{x}_1$ $\mathbf{x}_2$ $\mathbf{x}_3$ $\mathbf{x}_4$ $\mathbf{x}_5$ $\mathbf{x}_6$) to a canonical form $\Rightarrow$ gives a certain (*X*,*Y*,*Z*,*T*)
  - Then $\mathbf{H}_2\mathbf{H}_1\mathbf{H}^{-1}$ transforms ($\mathbf{Hx}_1$ $\mathbf{Hx}_2$ $\mathbf{Hx}_3$ $\mathbf{Hx}_4$ $\mathbf{Hx}_5$ $\mathbf{Hx}_6$) to the same canonical form $\Rightarrow$ gives same (*X*,*Y*,*Z*,*T*)

Configurations

- Two sets of 6 3D points $\mathbf{x}_k$ and $\mathbf{x}'_k$ represent the same *configuration* if there is a 3D homography **H** that transforms one set to the other

$$\mathbf{x}'_k \sim \mathbf{H}\,\mathbf{x}_k$$

$$k = 1, \ldots, 6$$

Configurations

- The 4 scalars (X,Y,Z,T) form a projective element (why?)
- Consequently, they have 3 d.o.f.
- A unique configuration of 6 3D points are represented by a unique projective element (X,Y,Z,T)
- $\Rightarrow$ The set of unique configurations have 3 degrees of freedom

Relative 3D invariants

- The scalars I_k (or $\hat{I}_k$) are functions of (X,Y,Z,T)

  $\Rightarrow$ they, too, are invariant to any homography transformations of the 3D space

- I_k (or $\hat{I}_k$) are *relative 3D invariants*

Relative 3D invariants

- We can form a 5-dimensional vector **s**:

$$\mathbf{s} = \begin{pmatrix} \hat{I}_1 \\ \hat{I}_2 \\ \hat{I}_3 \\ \hat{I}_4 \\ \hat{I}_5 \end{pmatrix}$$

- **s** is a relative 3D invariant: it is invariant to any homography transformation of the 3D space.
- **s** is a projective element

Relative 2D invariants

- In a similar way: (u_5,v_5,w_5) and (u_6,v_6,w_6) are invariant to any homography transformation of the image space
- Each triplet form a projective element (why?)
- The scalars i_k are invariant to any 2D homography transformation
- The scalars i_k form *2D relative invariants*

Relative 3D invariants

- We can form a 5-dimensional vector **z**:

$$\mathbf{z} = \begin{pmatrix} i_1 \\ i_2 \\ i_3 \\ i_4 \\ i_5 \end{pmatrix}$$

- **z** is a relative 2D invariant: it is invariant to any homography transformation of the image space.
- **z** is a projective element

Rigid transformations

- In practice we are interested in rigid transformations (rotation + translation) of 3D space
- This is a subset of the 3D homography transformations
- **s** is invariant to rigid transformations

Quan's constraint (III)

- Let **s** be computed from a particular configuration of 6 3D points
- Let **z** be computed from the projection of the 6 points onto the image
- Quan's constraint: $\mathbf{s} \cdot \mathbf{z} = 0$
- Make a rigid transformation of the 3D space
  - **s** is invariant to this transformation
  - **z** may or may not change
  - However, $\mathbf{s} \cdot \mathbf{z} = 0$ before and after the transformation

Quan's constraint (III)

For a given 3D configuration

- Any projection of the points into the image generates a relative 2D invariant **z** (a 5D vector)
- When the 3D points transform rigidly, **z** changes
- For a particular configuration, however, **z** is restricted to a 4D space
- This 4D space is orthogonal to **s**, the relative 3D invariant generated by the configuration
- Quan's constraint allows us to test if an observation of 6 image points is consistent with a certain configuration
  - Compare to the epipolar constraint
  - The points must be ordered in a specific way!

Internal constraint

- **s** has 4 d.o.f. as a general projective element
- However, **s** depends on (*X,Y,Z,T*) with 3 d.o.f.

  $\Rightarrow$ The elements of **s** must satisfy an internal constraint:

$$\hat{I}_1\hat{I}_2\hat{I}_5 - \hat{I}_1\hat{I}_3\hat{I}_4 + \hat{I}_2\hat{I}_3\hat{I}_4 - \hat{I}_2\hat{I}_3\hat{I}_5 - \hat{I}_2\hat{I}_4\hat{I}_5 + \hat{I}_3\hat{I}_4\hat{I}_5 = 0$$

Estimation of **s**

- **s** can be computed from a 3D configuration
- Alternatively:
  - Take 4 observations of **z** from the same configuration
  - Determine **s** from **s** · $\mathbf{z}_k$ = 0, k = 1, …, 4 (how?)
  - This **s** may not satisfy the int. const. in the case of noisy data
- Alternatively:
  - Take 3 observations of z from the same configuration
  - Determine **s** from **s** · $\mathbf{z}_k$ = 0, k = 1, …, 3 plus the int. constr. (how?)
  - This **s** is guaranteed to satisfy the int. constr.
  - Multiple solutions! (why?)
  - This is the method presented in Quan's paper
- What about Hartley-normalization?

6 points and 6 lines

- Quan's matching constraint can be expressed in terms of incidence relations between points and lines

- [Carlsson, *Duality of Reconstruction and Positioning from Projective Views*, WRVS, 1995]
- [Nordberg, *Single-view matching constraints*, ISVC, 2007]
- [Nordberg & Zografos, *Multibody motion classification using the geometry of 6 points in 2D images*, ICPR 2010]

6 points and 6 lines

- The computations from the image points to **z** are up to now implicit
- If we make them explicit, it turns out that

$$\mathbf{z} = \begin{pmatrix} D_{126} D_{354} \\ D_{136} D_{254} \\ D_{146} D_{253} \\ D_{145} D_{263} \\ D_{135} D_{246} \end{pmatrix} \qquad D_{ijk} = (\mathbf{y}_i \times \mathbf{y}_j) \cdot \mathbf{y}_k = \det(\mathbf{y}_i\, \mathbf{y}_j\, \mathbf{y}_k)$$

Important message:
Each index occurs exactly once in each element of **z**

6 points and 6 lines

- This means that we can rewrite **z**, e.g., as

$$\mathbf{z} = \begin{pmatrix} D_{354} (\mathbf{y}_2 \times \mathbf{y}_6)^T \\ D_{254} (\mathbf{y}_3 \times \mathbf{y}_6)^T \\ D_{253} (\mathbf{y}_4 \times \mathbf{y}_6)^T \\ D_{263} (\mathbf{y}_4 \times \mathbf{y}_5)^T \\ D_{246} (\mathbf{y}_3 \times \mathbf{y}_5)^T \end{pmatrix} \mathbf{y}_1$$

6 points and 6 lines

- Quan's constraint **s** · **z** = 0 then becomes

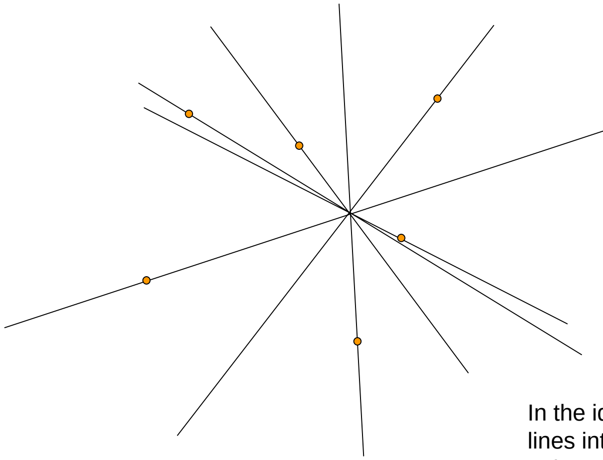
$$\mathbf{l}_1 \cdot \mathbf{y}_1 = 0$$

$$\begin{aligned} \mathbf{l}_1 = & \hat{I}_1 D_{354} (\mathbf{y}_2 \times \mathbf{y}_6)^T + \hat{I}_2 D_{254} (\mathbf{y}_3 \times \mathbf{y}_6)^T + \hat{I}_3 D_{253} (\mathbf{y}_4 \times \mathbf{y}_6)^T + \\ & + \hat{I}_4 D_{263} (\mathbf{y}_4 \times \mathbf{y}_5)^T + \hat{I}_5 D_{246} (\mathbf{y}_3 \times \mathbf{y}_5)^T \end{aligned}$$

6 points and 6 lines

- It makes sense to interpret $\mathbf{l}_1$ as the dual homogeneous coordinates of a line
- $\mathbf{l}_1$ depends on points $\mathbf{y}_2$, …, $\mathbf{y}_6$ and **s**
- Quan's constraint: point $\mathbf{y}_1$ must intersect line $\mathbf{l}_1$
- We can do the similar computations for the other points to get, in total, 6 lines
- Each point $\mathbf{y}_k$ must intersect its corresponding line $\mathbf{l}_k$
- Compare to epipolar lines

6 points and 6 lines



In the ideal case the 6 lines intersect at a single point

Why lines?

- In the practical situation, **s** · **z** may give a “large” value even for “good” correspondence. It is an *algebraic error*
- By describing the constraint in terms of a point-line incidence relation, we can quantify the constraint in terms of a *geometric error*, e.g.

$$\varepsilon_{GEO} = \sum_{k=1}^{6} d(\mathbf{y}_k, \mathbf{l}_k)^2$$

Applications

Motion segmentation:

- Basic idea:
  - Pick 6 points in the image
  - We can estimate **s** from 3 (or more) observations of these points
  - If they are on the same object (moving with the same rigid transformation):
    - The matching error between **s** and **z** should be small over many observations
  - If they are on different objects
    - The matching error between **s** and **z** should be large over many observations (not necessarily?)
- [Nordberg & Zografos, *Long title*, ICPR 2010]

Issues not covered here

- Degeneracies for **s**
- **s** can be linearly estimated even for degenerate cases