3D Reconstruction – Stereo

Jan 2004

Seong-Wook Joo

1. Projective Geometry (on 2D projective plane)

- **Homogeneous representation**

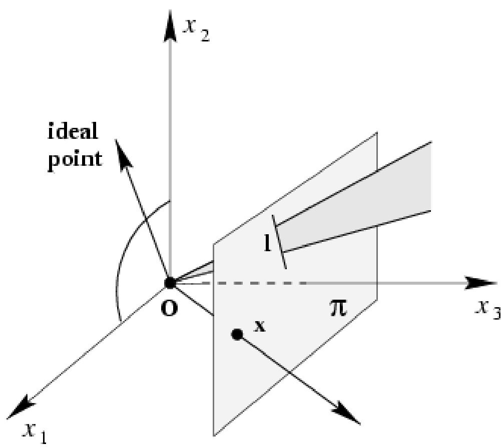
  Point: $\mathbf{x} = (x, y, 1)^T$

  Line: $\mathbf{l} = (a, b, c)^T$

  Point $\mathbf{x}$ is on line $\mathbf{l}$: $\mathbf{x}^T\mathbf{l} = 0$

  Intersection of $\mathbf{l}_1$ and $\mathbf{l}_2$: $\mathbf{x} = \mathbf{l}_1 \times \mathbf{l}_2$

  Line passing thru $\mathbf{x}_1$ and $\mathbf{x}_2$: $\mathbf{l} = \mathbf{x}_1 \times \mathbf{x}_2$ (dual of above)



- **Projective transformation (= homography)**

  Colinearity preserving transformation. Represented by a 3×3 invertible matrix "H"

  * 4pt correspondence needed to define H.

  * 2 views of planar object, 2 rotated views of any scene are related to each other by H

2. Camera Model: Projective Camera

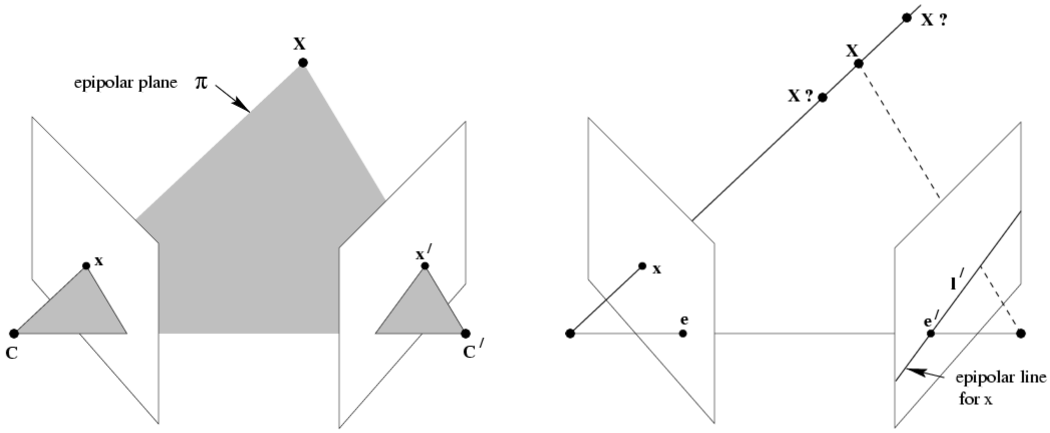
$\mathbf{x} = P\mathbf{X}$ ($\mathbf{x}$: image point, $\mathbf{X}$: world point, P: 3×4 camera projection matrix)

$$\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{bmatrix} f & & & 0 \\ & f & & 0 \\ & & 1 & 0 \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$ (simple case with aligned world and camera coordinates)

In general, $P = K_{3\times3}[R \mid \mathbf{t}]_{3\times4}$ (K: intrinsic calibration matrix, R,$\mathbf{t}$: extrinsic params)

3. Epipolar Geometry

Projective geometry between two views. Depends only on camera parameters (internal, external)



- **Fundamental matrix**

  Fundamental matrix F is the mapping of img point **x** to its epipolar line **l′**

$$\mathbf{l}' = \mathrm{F}\mathbf{x}$$

  It can be shown that

$$\mathrm{F} = [\mathbf{e}']_\times \mathrm{P}'\mathrm{P}^+$$

  where $\begin{bmatrix} e_1 \\ e_2 \\ e_3 \end{bmatrix}_\times = \begin{bmatrix} 0 & -e_3 & e_2 \\ e_3 & 0 & -e_1 \\ -e_2 & e_3 & 0 \end{bmatrix}$ is the matrix representation of the cross product.

  Also, for any corresponding point pairs **x**, **x′**

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

  F has 7 degrees of freedom: (3×3) – common scaling (1) – rank 2 constraint: det F=0 (1).

  * F is invariant under projective transformation H on the world space. i.e., even if $\mathbf{X} \rightarrow \mathrm{H}\mathbf{X}$, by letting $\mathrm{P} \rightarrow \mathrm{PH}^{-1}$, F remains unchanged. There is a projective ambiguity in P

- **Essential matrix**

  The fundamental matrix with the calibration matrices K, K′ removed. i.e., image points are normalized by $\hat{\mathbf{x}} = \mathrm{K}^{-1}\mathbf{x}$, $\hat{\mathbf{x}}' = \mathrm{K}'^{-1}\mathbf{x}'$. Letting $\mathrm{E} = \mathrm{K}'^{\mathrm{T}}\mathrm{F}\mathrm{K}$,

$$\hat{\mathbf{x}}'^{\mathrm{T}}\mathrm{E}\hat{\mathbf{x}} = 0$$

  For normalized cameras P = [I | **0**] and P′ = [R | **t**], the following holds.

$$\mathrm{E} = [\mathbf{t}]_\times \mathrm{R}$$

  E has 5 degrees of freedom: rotation (3) + translation (3) – common scale (1).

Given SVD of $E = U \operatorname{diag}(1, 1, 0) V^T$, and assuming first camera is $P = [I \mid \mathbf{0}]$, the second camera is one of the following

$P' = [UWV^T \mid +\mathbf{u}_3]$ or $P' = [UWV^T \mid -\mathbf{u}_3]$ or $P' = [UW^TV^T \mid +\mathbf{u}_3]$ or $P' = [UW^TV^T \mid -\mathbf{u}_3]$

where W = [0 –1 0; 1 0 0; 0 0 1] (in Matlab notation)
Only one of these is physically possible (positive depth from both cameras).

3. Computing the Fundamental Matrix

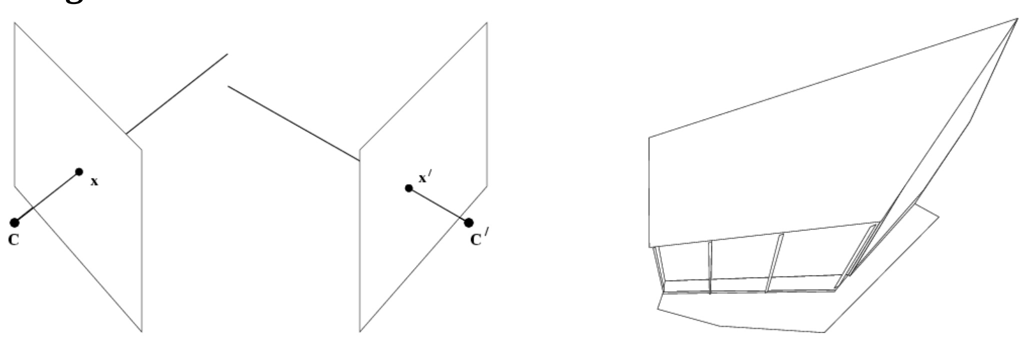
$$F = \begin{bmatrix} f_1 & f_2 & f_3 \\ f_4 & f_5 & f_6 \\ f_7 & f_8 & f_9 \end{bmatrix} \rightarrow \mathbf{f} = [\, f_1, f_2, f_3, f_4, f_5, f_6, f_7, f_8, f_9 \,]^T$$

Each corresponding point pairs $(x, y, 1)$ and $(x', y', 1)$ gives an equation

$$(\, x'x, \quad x'y, \quad x', \quad y'x, \quad y'y, \quad y', \quad x, \quad y, \quad 1 \,)\, \mathbf{f} = 0$$

Stacking n equations from n point correspondences gives linear system $A\mathbf{f} = 0$, where A is an $n \times 9$ matrix. If rank A = 8 then the solution is unique (up to scale) but in reality we seek a least-squares solution with $n \geq 8$. The LS solution is the last column of V in SVD of $A = UDV^T$ (last column corresponds to the smallest singular value)

4. Computing the Structure



Back-projected rays will not in general intersect. Minimum Euclidean distance is not suitable when the camera matrix P are known only up to a projective (or affine) transformation.

References

Richard Hartley and Andrew Zisserman, “Multiple View Geometry in Computer Vision”, Cambridge University Press, 2000.