521466S Machine Vision
Exercise #7
Transformations and Epipolar geometry

1. Homogeneous coordinates.

   a) The equation of a line in the plane is

$$ax + by + c = 0.$$

   Show that by using homogeneous coordinates this can be written as

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

   where $\mathbf{l} = \begin{pmatrix} a & b & c \end{pmatrix}^\top$.
   b) Show that the intersection of two lines $\mathbf{l}$ and $\mathbf{l}'$ is the point $\mathbf{x} = \mathbf{l} \times \mathbf{l}'$.
   c) Show that the line through two points $\mathbf{x}$ and $\mathbf{x}'$ is $\mathbf{l} = \mathbf{x} \times \mathbf{x}'$.
   d) Show that for all $\alpha \in \mathbb{R}$ the point $\mathbf{y} = \alpha\mathbf{x} + (1-\alpha)\mathbf{x}'$ lies on the line through points $\mathbf{x}$ and $\mathbf{x}'$.

Solution

The homogeneous coordinates of a point $(x, y) \in \mathbb{R}^2$ are obtained by adding a final coordinate of 1, i.e.

$$\mathbf{x} = \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}.$$

A homogeneous vector $\mathbf{x} = (x_1, x_2, x_3)^\top$, $x_3 \neq 0$, represents the point $(\frac{x_1}{x_3}, \frac{x_2}{x_3})$ in the inhomogeneous coordinates. In other words, all scaled 3-vectors $s\mathbf{x} = s(x, y, 1)^\top$, $s \neq 0$, represent the same planar point. Points with homogeneous coordinates $(x, y, 0)^\top$ do not correspond to any finite point in $\mathbb{R}^2$ and they are called *ideal points*. The homogeneous representation generalizes also to the three-dimensional case: the homogeneous coordinates of a point $(X, Y, Z) \in \mathbb{R}^3$ are $\mathbf{X} = (X, Y, Z, 1)^\top$.

a) The equation of a line can be directly written in the desired form:

$$0 = ax + by + c = \begin{pmatrix} x & y & 1 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \mathbf{x}^\top \mathbf{l}, \tag{1}$$

where $\mathbf{l} = (a, b, c)^\top$. Notice that this representation is independent of the scale of vectors $\mathbf{x}$ and $\mathbf{l}$ (both sides of the equation may be multiplied with a nonzero scalar and it still represents the same line). Thus, all vectors $s\mathbf{l} = s(a, b, c)^\top$, $s \neq 0$, represent the same line. So, also lines have homogeneous representation as 3-vectors. Hence, in terms of homogeneous coordinates we get:

The point $\mathbf{x}$ *lies on the line* $\mathbf{l}$ *if and only if* $\mathbf{x}^\top \mathbf{l} = 0$.

b) Given two lines $\mathbf{l} = (a, b, c)^\top$ and $\mathbf{l}' = (a', b', c')^\top$ we may define the vector $\mathbf{x} = \mathbf{l} \times \mathbf{l}'$. From the triple scalar product identity $\mathbf{l} \cdot (\mathbf{l} \times \mathbf{l}') = \mathbf{l}' \cdot (\mathbf{l} \times \mathbf{l}') = 0$ we see that $\mathbf{x}^\top \mathbf{l} = \mathbf{x}^\top \mathbf{l}' = 0$. Thus, if $\mathbf{x}$ is thought of as a homogeneous representation of a point, then $\mathbf{x}$ lies on both lines $\mathbf{l}$ and $\mathbf{l}'$ and hence is the intersection of the two lines.

c) This can be shown in a similar way as b). Given two points we may define the line $\mathbf{l} = \mathbf{x} \times \mathbf{x}'$. Then the triple scalar product identity gives that both points $\mathbf{x}$ and $\mathbf{x}'$ lie on $\mathbf{l}$.

d) The line through the points $\mathbf{x}$ and $\mathbf{x}'$ is $\mathbf{l} = \mathbf{x} \times \mathbf{x}'$. Then from the triple scalar product identity we get $\mathbf{y}^\top \mathbf{l} = \alpha \mathbf{x}^\top (\mathbf{x} \times \mathbf{x}') + (1 - \alpha) \mathbf{x}'^\top (\mathbf{x} \times \mathbf{x}') = 0$ so that $\mathbf{y}$ lies on the line $\mathbf{l}$.

2. Transformations in 2D.

a) Use homogeneous coordinates and give the matrix representations of the following transformation groups: translation, Euclidean transformation (rotation+translation), similarity transformation (scaling+rotation+translation), affine transformation, projective transformation.

b) What is the number of degrees of freedom in these transformations? How many point correspondences are required for estimating the parameters of the transformations?

Solution

Translation:

$$\begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = \begin{bmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$$

Euclidean transformation (rotation+translation):

$$\begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta & t_x \\ \sin\theta & \cos\theta & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$$

Similarity transformation (scaling+rotation+translation):

$$\begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = \begin{bmatrix} s\cos\theta & -s\sin\theta & t_x \\ s\sin\theta & s\cos\theta & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$$

Affine transformation:

$$\begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = \begin{bmatrix} a_{11} & a_{12} & t_x \\ a_{21} & a_{22} & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$$

Projective transformation:

$$\mathbf{x}' = \mathbf{H}\mathbf{x},$$

where $\mathbf{H}$ is invertible 3×3 matrix and $\mathbf{x}$ and $\mathbf{x}'$ are homogeneous representations of points in the plane.

Notice that translations are subgroup of Euclidean transformations, Euclidean transformations are subgroup of similarity transformations, similarity transformations are subgroup of affine transformations and affine transformations are subgroup of projective transformations. Most general is the group of projective transformations which covers all transformation groups above.

b) Translation has 2 degrees of freedom (t_x, t_y), Euclidean transformation has 3 (t_x, t_y, θ), similarity transformation has 4 (t_x, t_y, θ, s), affine transformation has 6 $(t_x, t_y, a_{11}, a_{12}, a_{21}, a_{22})$ and a general projective transformation has 8 degrees of freedom (a 3×3 matrix has 9 elements but the scale of the elements does not matter due to homogeneous representation).

Each point correspondence gives two scalar constraint equations for the transformation parameters. (In homogeneous coordinates the constraint is represented as $\mathbf{x}' \simeq \mathcal{T}(\mathbf{x})$ where

$\simeq$ denotes equality up to scale and $\mathcal{T}$ is the transformation.) In principle, the transformation can be estimated if the number of constraints is greater or equal than the number of degrees of freedom. Hence, the required number of point correspondences is 1 for translation, 2 for Euclidean transformation and similarity transformation, 3 for affine transformation and 4 for projective transformation.

Some properties of commonly occurring planar transformations are summarized in the following table.

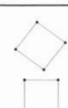
Group	Matrix	Distortion	Invariant properties
Projective 8 dof	$\begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix}$		Concurrency, collinearity, **order of contact**: intersection (1 pt contact); tangency (2 pt contact); inflections (3 pt contact with line); tangent discontinuities and cusps. cross ratio (ratio of ratio of lengths).
Affine 6 dof	$\begin{bmatrix} a_{11} & a_{12} & t_x \\ a_{21} & a_{22} & t_y \\ 0 & 0 & 1 \end{bmatrix}$		Parallelism, ratio of areas, ratio of lengths on collinear or parallel lines (e.g. midpoints), linear combinations of vectors (e.g. centroids). The line at infinity, $\mathbf{l}_\infty$.
Similarity 4 dof	$\begin{bmatrix} sr_{11} & sr_{12} & t_x \\ sr_{21} & sr_{22} & t_y \\ 0 & 0 & 1 \end{bmatrix}$		Ratio of lengths, angle. The circular points, **I**, **J** (see section 2.7.3).
Euclidean 3 dof	$\begin{bmatrix} r_{11} & r_{12} & t_x \\ r_{21} & r_{22} & t_y \\ 0 & 0 & 1 \end{bmatrix}$		Length, area

Table 2.1. **Geometric properties invariant to commonly occurring *planar* transformations.** *The matrix* $\mathtt{A} = [a_{ij}]$ *is an invertible* 2×2 *matrix,* $\mathtt{R} = [r_{ij}]$ *is a 2D rotation matrix, and* (t_x, t_y) *a 2D translation. The distortion column shows typical effects of the transformations on a square. Transformations higher in the table can produce all the actions of the ones below. These range from Euclidean, where only translations and rotations occur, to projective where the square can be transformed to any arbitrary quadrilateral (provided no three points are collinear).*

Figure 1: The table is from the book *Multiple View Geometry in Computer Vision* by Hartley and Zisserman (Cambridge, 2000).

3. Similarity transformation.

a) Describe a method for solving the parameters of a similarity transformation from two point correspondences.

b) Solve the transformation using the point correspondences $\{\mathbf{p}_1, \mathbf{p}_1'\} = \{(\frac{1}{2}, 0),\ (0,0)\}$ and $\{\mathbf{p}_2, \mathbf{p}_2'\} = \{(0, \frac{1}{2}),\ (-1,-1)\}$. Here $\mathbf{p}_1'$ and $\mathbf{p}_2'$ are the coordinates of the points *after* the transformation.

Solution

A similarity transformation can be written as a combination of rotation, scaling and translation:

$$\begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = \mathbf{D}_{t_x,t_y}\mathbf{S}_s\mathbf{R}_\theta \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{bmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} s & 0 & 0 \\ 0 & s & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$$
$$= \begin{bmatrix} s\,\cos\theta & -s\,\sin\theta & t_x \\ s\,\sin\theta & s\,\cos\theta & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{bmatrix} s\mathbf{R} & \mathbf{t} \\ \mathbf{0} & 1 \end{bmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix},$$

where $\mathbf{t} = (t_x, t_y)^\top$ is the translation, s is the scaling and $\mathbf{R}$ is the 2×2 rotation matrix. Given two correspondences $\{\mathbf{p}_1, \mathbf{p}_1'\}$ and $\{\mathbf{p}_2, \mathbf{p}_2'\}$ we get the equations

$$\mathbf{p}_1' = s\mathbf{R}\mathbf{p}_1 + \mathbf{t} \tag{2}$$
$$\mathbf{p}_2' = s\mathbf{R}\mathbf{p}_2 + \mathbf{t} \tag{3}$$

where we have used the usual inhomogeneous coordinates. Then, by subtracting the first equation from the second we get

$$(\mathbf{p}_2' - \mathbf{p}_1') = s\mathbf{R}(\mathbf{p}_2 - \mathbf{p}_1) \tag{4}$$

which states that the translation does not affect the vectors $\mathbf{v}' = \mathbf{p}_2' - \mathbf{p}_1'$ and $\mathbf{v} = \mathbf{p}_2 - \mathbf{p}_1$. Moreover, by defining the corresponding unit vectors $\mathbf{u}' = \mathbf{v}'/||\mathbf{v}'||$ and $\mathbf{u} = \mathbf{v}/||\mathbf{v}||$ we get from (4) that

$$\mathbf{v}' = s\mathbf{R}\mathbf{v} \quad \Rightarrow \quad \mathbf{u}' = \mathbf{R}\mathbf{u} \tag{5}$$

since $||\mathbf{R}\mathbf{v}|| = ||\mathbf{v}||$ when $\mathbf{R}$ is a rotation matrix, i.e. $\mathbf{R}^\top\mathbf{R} = \mathbf{I}$. Hence, by computing the angle α between the vectors $\mathbf{u}'$ and $\mathbf{u}$,

$$\alpha = \arccos(\mathbf{u}' \cdot \mathbf{u}),$$

we get the rotation angle θ which is either $+\alpha$ or $-\alpha$. The correct sign must be chosen such that $\mathbf{u}'$ is obtained from $\mathbf{u}$ by *counterclockwise* rotation through an angle θ. After θ is computed we get $\mathbf{R}$ from its definition

$$\mathbf{R} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}. \tag{6}$$

Then s is obtained by

$$s = \frac{||\mathbf{v}'||}{||\mathbf{v}||} \tag{7}$$

and $\mathbf{t}$ can be solved from (2) or (3).

In our example $\mathbf{p}_1 = (\frac{1}{2}, 0)^\top$, $\mathbf{p}'_1 = (0, 1)^\top$, $\mathbf{p}_2 = (0, \frac{1}{2})^\top$ and $\mathbf{p}'_2 = (-1, 0)^\top$. This gives $\mathbf{v} = (-\frac{1}{2}, \frac{1}{2})^\top$, $\mathbf{v}' = (-1, -1)^\top$ and $||\mathbf{v}|| = \sqrt{2}/2$, $||\mathbf{v}'|| = \sqrt{2}$. Further, $\mathbf{u} = \frac{1}{\sqrt{2}}(-1, 1)^\top$ and $\mathbf{u}' = \frac{1}{\sqrt{2}}(-1, -1)^\top$ so that $\theta = \alpha = \pi/2$. Then, $s = 2$ and $\mathbf{t} = (0, -1)^\top$. You can draw the points in order to illustrate the situation.

4. Epipolar geometry.

The camera projection matrices of two cameras (given in the coordinate system attached to the first camera) are

$$\mathbf{C} = \begin{bmatrix}\mathbf{I} & \mathbf{0}\end{bmatrix} \quad \text{and} \quad \mathbf{C}' = \begin{bmatrix}\mathbf{R} & \mathbf{t}\end{bmatrix},$$

where $\mathbf{R}$ is a rotation matrix and $\mathbf{t} = (t_1, t_2, t_3)^\top$ describes the translation between the cameras. Hence, the cameras have identical internal parameters and the image points are given in the normalized image coordinates (the origin of the image coordinate frame is at the principal point and the focal length is 1).

a) The epipolar constraint is illustrated in the figure below and it implies that if p and p' are corresponding image points then the vectors $\overrightarrow{Op}$, $\overrightarrow{O'p'}$ and $\overrightarrow{O'O}$ are coplanar, i.e.

$$\overrightarrow{O'p'} \cdot \left(\overrightarrow{O'O} \times \overrightarrow{Op}\right) = 0 \tag{8}$$

Let $\mathbf{p} = (x, y, 1)^\top$ and $\mathbf{p}' = (x', y', 1)^\top$ denote the homogeneous image coordinate vectors of p and p'. Show that the equation (8) can be written in the form

$$\mathbf{p}'^\top \mathbf{E} \mathbf{p} = 0, \tag{9}$$

where matrix $\mathbf{E}$ is the essential matrix

$$\mathbf{E} = [\mathbf{t}]_\times \mathbf{R}, \quad \text{where} \quad [\mathbf{t}]_\times = \begin{bmatrix} 0 & -t_3 & t_2 \\ t_3 & 0 & -t_1 \\ -t_2 & t_1 & 0 \end{bmatrix}.$$

b) How can the epipolar constraint be utilized when searching point correspondences between two views?

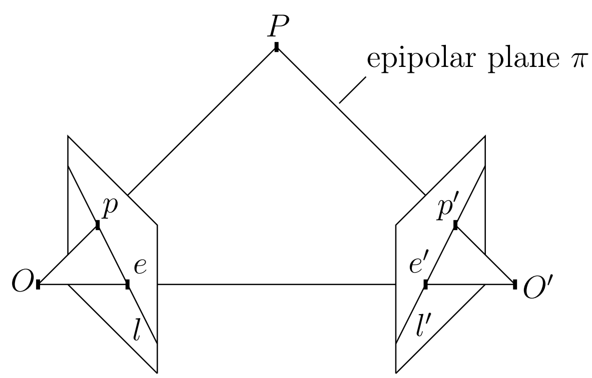


Figure 2: Epipolar geometry. Given a point p in the first image its corresponding point in the second image is constrained to lie on the line l' which is the epipolar line of p. Correspondingly, the line l is the epipolar line of p'. Points e and e' are the epipoles.

Solution

a) In the coordinate system of the second camera we have

$$\overrightarrow{O'p'} = \mathbf{p}', \quad \overrightarrow{O'O} = \mathbf{t}, \quad \overrightarrow{Op} = \mathbf{R}\mathbf{p}$$

and by substituting these into (8) we get

$$\mathbf{p}'^{\top}\left(\mathbf{t} \times \mathbf{R}\mathbf{p}\right) = 0. \tag{10}$$

The cross product $\mathbf{t} \times \mathbf{a}$, where $\mathbf{a}$ is an arbitrary 3-vector, may be written as

$$\mathbf{t} \times \mathbf{a} = \begin{vmatrix} i & j & k \\ t_1 & t_2 & t_3 \\ a_1 & a_2 & a_3 \end{vmatrix} = \begin{pmatrix} -t_3 a_2 + t_2 a_3 \\ t_3 a_1 - t_1 a_3 \\ -t_2 a_1 + t_1 a_2 \end{pmatrix} = \begin{bmatrix} 0 & -t_3 & t_2 \\ t_3 & 0 & -t_1 \\ -t_2 & t_1 & 0 \end{bmatrix} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = [\mathbf{t}]_{\times}\mathbf{a}$$

Hence, we may write (10) in the form (9), where $\mathbf{E} = [\mathbf{t}]_{\times}\mathbf{R}$. Geometrically this means that given the point $\mathbf{p}$ the corresponding point $\mathbf{p}'$ must lie on the line $\mathbf{l}' = \mathbf{E}\mathbf{p}$.

b) Given a point in the first image and the epipolar geometry one may directly compute the corresponding line in the second image. Hence, it is sufficient to search the corresponding point in the neighbourhood of this line and this significantly reduces the search area.

5. Epipolar line.

Let

$$\mathbf{R} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix}, \quad \mathbf{t} = (1, 1, 1)^\top$$

be the rotation and translation between two views. Compute the essential matrix and the epipolar line which corresponds to the principal point of the first image (i.e. the point $(0, 0)$ in the normalized image coordinate system).

Solution

Now

$$[\mathbf{t}]_\times = \begin{bmatrix} 0 & -1 & 1 \\ 1 & 0 & -1 \\ -1 & 1 & 0 \end{bmatrix}, \qquad \mathbf{R} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix}$$

and using the formula from the previous problem we get

$$\mathbf{E} = [\mathbf{t}]_\times \mathbf{R} = \begin{bmatrix} 0 & 1 & 1 \\ 1 & -1 & 0 \\ -1 & 0 & -1 \end{bmatrix}$$

The epipolar line $\mathbf{l}'$ corresponding to the point $(0, 0)$ is

$$\mathbf{l}' = \mathbf{E} \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \begin{bmatrix} 0 & 1 & 1 \\ 1 & -1 & 0 \\ -1 & 0 & -1 \end{bmatrix} \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$$
