Engineering Tripos Part IIB FOURTH YEAR

Module 4F12: Computer Vision and Robotics

Solutions to Examples Paper 3

1. *The stereo correspondence problem*

   The difficult task of matching features between left and right images can be simplified using the following constraints:

   Epipolar constraint: the stereo camera geometry constrains each point feature identified in one image to lie on a corresponding *epipolar line* in the other image. If the cameras are calibrated, then the equation of the epipolar line can be derived from the essential matrix. For uncalibrated cameras, it is possible to estimate the fundamental matrix from point correspondences and derive epipolar lines from the fundamental matrix. Epipolar lines meet at the *epipole*: this is the image of one camera's optical centre in the other camera's image plane. There are two epipoles, one for each image.

   Uniqueness: For scenes containing only opaque objects, each point in the left image has at most one match in the right image.

   Ordering: Corresponding points lying on the surface of an opaque object will be ordered identically in left and right images. The ordering constraint will not necessarily hold if the points do not lie on the surface of the same opaque object.

   Figural continuity: When distinguished points lie on image contours, we can sometimes use figural continuity as a matching constraint.

   Disparity gradient: If surfaces are smooth, then point disparities (differences in location between left and right images) must be locally smooth. So a further constraint comes from imposing a limit on the allowable spatial derivatives of disparity.

2. *Stereo and orthographic projection*

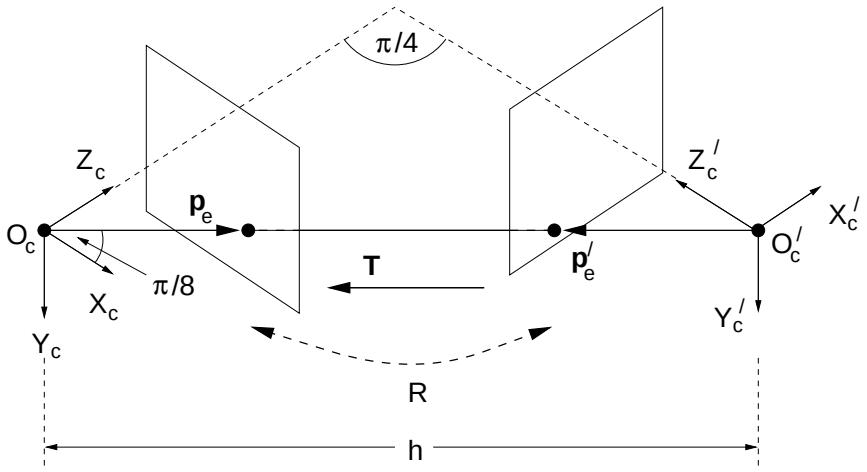
   The architecture student is clearly used to orthographic projections and not perspective projections. Let's assume that the left camera's coordinate system is aligned with the world coordinate system.

   Under perspective projection we use the left image to find a *ray* along which the point must lie, but this ray is not necessarily normal to the image plane and therefore does *not* fix the X and Y world coordinates of the point. We then use the right image to find where on the ray the point lies.

   Under orthographic projection the ray would be normal to the left image plane and would fix the X and Y world coordinates of the point. Observing the point in the right image would fix the Z coordinate.

3. *Epipolar geometry*

(a)



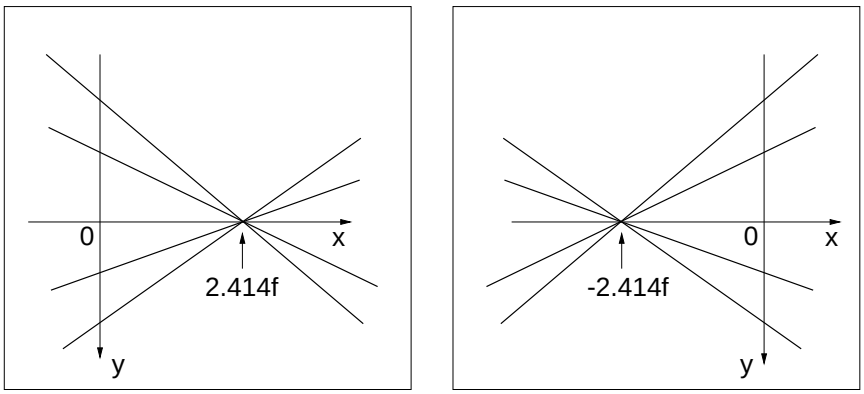
The epipolar lines will pass through the epipoles $\mathbf{p_e}$ and $\mathbf{p'_e}$, which are the images of the other camera's optical centre. We can find $\mathbf{p_e}$ and $\mathbf{p'_e}$ by simple trigonometry:

$$\hat{\mathbf{p_e}} = [\ \cos(\pi/8)\ \ 0\ \ \sin(\pi/8)\ ]^T \Rightarrow \mathbf{p_e} = (f/\sin(\pi/8))[\ \cos(\pi/8)\ \ 0\ \ \sin(\pi/8)\ ]^T$$
$$= [\ 2.414f\ \ 0\ \ f\ ]^T$$

Similarly, the epipole in the right image is

$$\mathbf{p'_e} = [\ -2.414f\ \ 0\ \ f\ ]^T$$

So the epipolar lines look like this:



Left image Right image

(b) By inspection of the figure in (a), the transformation between the left and right coordinate systems is $\mathbf{X'_C} = \mathrm{R}\mathbf{X_C} + \mathbf{T}$, where

$$\mathrm{R} = \begin{bmatrix} 1/\sqrt{2} & 0 & 1/\sqrt{2} \\ 0 & 1 & 0 \\ -1/\sqrt{2} & 0 & 1/\sqrt{2} \end{bmatrix} \quad \text{and} \quad \mathbf{T} = \begin{bmatrix} -h\cos(\pi/8) \\ 0 \\ h\sin(\pi/8) \end{bmatrix}$$

Adopting the shorthand $s \equiv \sin(\pi/8)$ and $c \equiv \cos(\pi/8)$, the essential matrix is

$$\begin{aligned} \mathrm{E} = \mathrm{T}_\times\mathrm{R} &= \begin{bmatrix} 0 & -hs & 0 \\ hs & 0 & hc \\ 0 & -hc & 0 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 0 & 1/\sqrt{2} \\ 0 & 1 & 0 \\ -1/\sqrt{2} & 0 & 1/\sqrt{2} \end{bmatrix} \\ &= \begin{bmatrix} 0 & -hs & 0 \\ (hs-hc)/\sqrt{2} & 0 & (hs+hc)/\sqrt{2} \\ 0 & -hc & 0 \end{bmatrix} \end{aligned}$$

and the epipolar constraints are

$$\begin{bmatrix} x' & y' & f \end{bmatrix} \begin{bmatrix} 0 & -s & 0 \\ (s-c)/\sqrt{2} & 0 & (s+c)/\sqrt{2} \\ 0 & -c & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ f \end{bmatrix} = 0$$

or equivalently

$$\sqrt{2}y(sx' + cf) = y'((s-c)x + (s+c)f)$$

This equation defines the epipolar lines. Given a point (x', y') in the right image, the equation describes a line in the left image on which the corresponding point (x, y) must lie, and vice-versa.

The epipolar lines in the left image will pass through the epipole $\mathbf{p_e}$, which lies in the null space of E. We can find the epipole as follows:

$$\mathrm{E}\,\mathbf{p_e} = \begin{bmatrix} 0 & -s & 0 \\ (s-c)/\sqrt{2} & 0 & (s+c)/\sqrt{2} \\ 0 & -c & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ f \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$$\Rightarrow y = 0\,, \quad x(c-s) = f(s+c)$$

Hence $\mathbf{p_e} = [\ f(s+c)/(c-s)\ \ 0\ \ f\ ]^T = [\ 2.414f\ \ 0\ \ f\ ]^T$.

Similarly, the epipolar lines in the right image will pass through the epipole $\mathbf{p'_e}$, which lies in the null space of E^T. By symmetry, we can conclude that $\mathbf{p'_e} = [\ -2.414f\ \ 0\ \ f\ ]^T$.

4. *Triangulation using rays*

Ray vectors $\mathbf{p}$ and world positions $\mathbf{X_C}$ are related via the unknown depth Z_c:

$$\mathbf{p} = \begin{bmatrix} x \\ y \\ f \end{bmatrix} = \begin{bmatrix} fX_c/Z_c \\ fY_c/Z_c \\ fZ_c/Z_c \end{bmatrix} = \frac{f}{Z_c}\mathbf{X_C}$$

Since $\mathbf{X'_C}$ and $\mathbf{p'}$ are parallel, we have

$$\begin{aligned} \mathbf{X'_C} \times \mathbf{p'} = \mathbf{0} \Leftrightarrow (\mathrm{R}\mathbf{X_C} + \mathbf{T}) \times \mathbf{p'} &= \mathbf{0} \\ \Leftrightarrow \left(\frac{Z_c}{f}\mathrm{R}\,\mathbf{p} + \mathbf{T}\right) \times \mathbf{p'} &= \mathbf{0} \end{aligned}$$

Let's consider the case when the image planes of the two cameras are aligned and the cameras have the same focal length:

$$\mathrm{R} = \mathrm{I}\,, \quad \mathbf{T} = [\,-d \;\; 0 \;\; 0\,]^T$$

The triangulation equations reduce to:

$$\begin{aligned} \frac{Z_c}{f}(\mathbf{p} \times \mathbf{p'}) &= -\mathbf{T} \times \mathbf{p'} \\ \Leftrightarrow Z_c(\mathbf{p} \times \mathbf{p'}) &= f \begin{bmatrix} d \\ 0 \\ 0 \end{bmatrix} \times \mathbf{p'} \\ \Leftrightarrow Z_c \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ x & y & f \\ x' & y' & f \end{vmatrix} &= f \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ d & 0 & 0 \\ x' & y' & f \end{vmatrix} \end{aligned}$$

Equating coefficients in $\mathbf{i}$, $\mathbf{j}$ and $\mathbf{k}$:

$$Z_c f(y - y') = 0 \tag{1}$$
$$Z_c f(x - x') = df^2 \tag{2}$$
$$Z_c(xy' - yx') = fdy' \tag{3}$$

(3) is not independent of (1) and (2). (2) allows us to recover the depth from the horizontal **disparity** $(x - x')$: $Z_c = df/(x - x')$. This result is intuitively correct: distant objects have smaller disparities than nearby objects.

5. *Uncalibrated stereo and the fundamental matrix*

Pixel coordinates and image plane coordinates are related by the CCD calibration matrix:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} k_u & 0 & u_0 \\ 0 & k_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

We can modify this to derive a relationship between pixel coordinates and rays:

$$\begin{bmatrix} fu \\ fv \\ f \end{bmatrix} = \begin{bmatrix} fk_u & 0 & u_0 \\ 0 & fk_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ f \end{bmatrix} = \mathrm{K} \begin{bmatrix} x \\ y \\ f \end{bmatrix}, \quad \text{say}$$

So $\tilde{\mathbf{w}} = \mathrm{K}\mathbf{p}$, and the epipolar constraint becomes

$$\begin{aligned} \mathbf{p}'^T \mathrm{E}\mathbf{p} &= 0 \Leftrightarrow \tilde{\mathbf{w}}'^T \mathrm{K}'^{-T} \mathrm{E} \mathrm{K}^{-1} \tilde{\mathbf{w}} = 0 \\ \Leftrightarrow \tilde{\mathbf{w}}'^T \mathrm{F} \tilde{\mathbf{w}} &= 0 \,, \quad \text{where } \mathrm{F} = \mathrm{K}'^{-T} \mathrm{E} \mathrm{K}^{-1} \end{aligned}$$

F is a 3×3 matrix known as the **fundamental matrix**. By decomposing E we can see how F is related to the *isometry* between the two camera coordinate systems and the *intrinsic* camera parameters:

$$\mathrm{F} = \mathrm{K}'^{-T} \mathrm{T}_\times \mathrm{R} \mathrm{K}^{-1}$$

The intrinsic parameters are represented by the matrix K and the isometry by R and $\mathbf{T}$.

The locations of the epipoles $\tilde{\mathbf{w}}_\mathrm{e}$ and $\tilde{\mathbf{w}}'_\mathrm{e}$ (in pixels) are given by

$$\begin{aligned} \mathrm{E}\mathbf{p_e} &= \mathbf{0} \Leftrightarrow \mathrm{E}\mathrm{K}^{-1}\tilde{\mathbf{w}}_\mathrm{e} = \mathbf{0} \\ \Leftrightarrow \mathrm{K}'^{-T}\mathrm{E}\mathrm{K}^{-1}\tilde{\mathbf{w}}_\mathrm{e} &= \mathbf{0} \Leftrightarrow \mathrm{F}\tilde{\mathbf{w}}_\mathrm{e} = \mathbf{0} \text{ and likewise } \mathrm{F}^T\tilde{\mathbf{w}}'_\mathrm{e} = \mathbf{0} \end{aligned}$$

It follows that F is not invertible (otherwise we could say $\tilde{\mathbf{w}}_\mathrm{e} = \mathrm{F}^{-1}\mathbf{0} = \mathbf{0}$, which is a contradiction) and therefore has maximum rank 2.

We can estimate F from point correspondences. Each correspondence $\tilde{\mathbf{w}} \leftrightarrow \tilde{\mathbf{w}}'$ generates one linear constraint on the elements of F:

$$\begin{bmatrix} u' & v' & 1 \end{bmatrix} \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = 0$$

Given 8 or more perfect correspondences (image points in *general* position, no noise), F can be determined uniquely up to scale. In practice, we may have more than 8 correspondences and the image measurements will be noisy. The system of equations is then solved by least squares (cf. camera calibration).

Note that we have not made any attempt to enforce the constraint that $\det \mathrm{F} = 0$. If the 8 image points are noisy, then we will find that our estimate of F does *not* have zero determinant and the epipolar lines do not meet at a point. Advanced nonlinear techniques exist to estimate F from 7 point correspondences, enforcing the rank 2 constraint.

6. *Affine fundamental matrix*

The weak perspective camera model is $\tilde{\mathbf{w}} = \mathrm{P}_{wp}\tilde{\mathbf{X}}$, where

$$\begin{aligned}\mathrm{P}_{wp} = \mathrm{P}_c\mathrm{P}_{pll}\mathrm{P}_r &= \begin{bmatrix} k_u & 0 & u_0 \\ 0 & k_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 0 & Z_c^{av} \end{bmatrix} \left[\begin{array}{cccc|c} & & & & \\ & & \mathrm{R} & & \mathbf{T} \\ & & & & \\ \hline 0 & 0 & 0 & & 1 \end{array}\right] \\ &= \begin{bmatrix} fk_ur_{11} & fk_ur_{12} & fk_ur_{13} & fk_uT_x + u_0Z_c^{av} \\ fk_vr_{21} & fk_vr_{22} & fk_vr_{23} & fk_vT_y + v_0Z_c^{av} \\ 0 & 0 & 0 & Z_c^{av} \end{bmatrix}\end{aligned}$$

If we assume, without loss of generality, that the left camera is aligned with the world coordinate system (so that R=I), then the camera matrix reduces to

$$\begin{bmatrix} fk_ur_{11} & 0 & 0 & fk_uT_x + u_0Z_c^{av} \\ 0 & fk_vr_{22} & 0 & fk_vT_y + v_0Z_c^{av} \\ 0 & 0 & 0 & Z_c^{av} \end{bmatrix}$$

Discarding the nonlinear constraints, we obtain affine models for the left and right cameras:

$$\textbf{Left:} \quad \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} p_{11} & 0 & 0 & p_{14} \\ 0 & p_{22} & 0 & p_{24} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

$$\textbf{Right:} \quad \begin{bmatrix} u' \\ v' \end{bmatrix} = \begin{bmatrix} p'_{11} & p'_{12} & p'_{13} & p'_{14} \\ p'_{21} & p'_{22} & p'_{23} & p'_{24} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

Eliminating X and Y from the above equations gives

$$\begin{aligned} u' &= p'_{11}\frac{(u - p_{14})}{p_{11}} + p'_{12}\frac{(v - p_{24})}{p_{22}} + p'_{13}Z + p'_{14} \\ v' &= p'_{21}\frac{(u - p_{14})}{p_{11}} + p'_{22}\frac{(v - p_{24})}{p_{22}} + p'_{23}Z + p'_{24} \end{aligned}$$

Eliminating Z we obtain

$$u' = p'_{11}\frac{(u - p_{14})}{p_{11}} + p'_{12}\frac{(v - p_{24})}{p_{22}} + p'_{14} + \frac{p'_{13}}{p'_{23}}\left(v' - p'_{21}\frac{(u - p_{14})}{p_{11}} - p'_{22}\frac{(v - p_{24})}{p_{22}} - p'_{24}\right)$$

or alternatively

$$au' + bv' + cu + dv + 1 = 0$$

We can rewrite this in matrix form:

$$[\, u' \;\; v' \;\; 1 \,] \, F_A \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = 0$$

where F_A is the affine fundamental matrix:

$$F_A = \begin{bmatrix} 0 & 0 & a \\ 0 & 0 & b \\ c & d & 1 \end{bmatrix}$$

By inspection, F_A has zero determinant and therefore maximum rank 2. The epipolar lines in the right image are given by

$$v' = -\frac{a}{b}u' - \frac{(cu + dv + 1)}{b}$$

Since the epipolar lines all have slope $-a/b$ they are parallel. A similar argument holds for the epipolar lines in the left image.

Roberto Cipolla
October 2020