Lecture 5: Epipolar Geometry and the Fundamental Matrix

1 Two-View Structure from Motion

In this lecture we will consider the two-view structure from motion problem. That is, given two images we want to compute both the camera matrices and the scene points such that the camera equations

$$\lambda_i \mathbf{x}_i = P_1 \mathbf{X}_i \tag{1}$$
$$\bar{\lambda}_i \bar{\mathbf{x}}_i = P_2 \mathbf{X}_i, \tag{2}$$

$i = 1, ..., n$, are fulfilled. In previous lectures we have considered sub-problems where either the camera matrices are known (the triangulation problem) or the scene points are known (the resection problem). Since the camera equations become linear in these two cases we could solve these directly by applying DLT. The situation becomes more complicated when both the scene points and camera matrices are unknown. The approach we will take in this lecture formulates a set of algebraic constraints that involve only the image points and the cameras, thereby eliminating the scene points. The resulting equations are linear and can be solved using SVD. Once the cameras are known the 3D points can be computed using triangulation.

1.1 Problem Formulation

Given two sets of corresponding points $\mathbf{x}_i$ and $\bar{\mathbf{x}}_i$, $i = 1, .., n$, our goal is to find camera matrices P_1 and P_2 such that (1)-(2) are solved. As we observed in lecture 3, if the cameras are uncalibrated the reconstruction can only be determined uniquely up to an unknown projective transformation. If the cameras are $P_1 = [A_1 \quad t_1]$ and $P_2 = [A_2 \quad t_2]$ then we can apply the transformation

$$H = \begin{bmatrix} A_1^{-1} & -A_1^{-1} t_1 \\ 0 & 1 \end{bmatrix} \tag{3}$$

to the camera equations (1)-(2). The camera matrix P_1 is then transformed to

$$P_1 H = \begin{bmatrix} A_1 & t_1 \end{bmatrix} \begin{bmatrix} A_1^{-1} & -A_1^{-1} t_1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} I & 0 \end{bmatrix}. \tag{4}$$

Therefore, we can always assume that there is a solution where $P_1 = [I \quad 0]$ and $P_2 = [A \quad t]$.

2 Epipolar Geometry

In this section we will derive the so called epipolar constraints. In the following sections we will use these constraints to find camera matrices that solve (1)-(2).

We first consider a point $\mathbf{x}$ in the first image. The scene points that can project to this image point all lie on a line (the viewing ray of $\mathbf{x}$) in 3D space, see Figure 1. If we assume that the $\mathbf{X} = \begin{bmatrix} X \\ 1 \end{bmatrix}$, where $X \in \mathbb{R}^3$, are the homogeneous coordinates of a scene point projecting to $\mathbf{x}$, then

$$\lambda \mathbf{x} = \begin{bmatrix} I & 0 \end{bmatrix} \begin{bmatrix} X \\ 1 \end{bmatrix} = X. \tag{5}$$

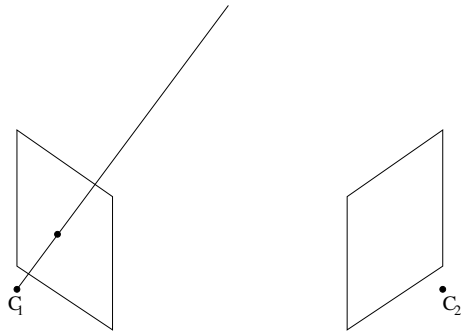

Figure 1: All scene points on the line project to the same point in the left camera.

Therefore the viewing ray of $\mathbf{x}$ can be parametrized by

$$\mathbf{X}(\lambda) = \begin{bmatrix} \lambda\mathbf{x} \\ 1 \end{bmatrix}. \tag{6}$$

The projection of this line into the second camera is

$$P_2\mathbf{X}(\lambda) = \begin{bmatrix} A & t \end{bmatrix} \begin{bmatrix} \lambda\mathbf{x} \\ 1 \end{bmatrix} = \lambda A\mathbf{x} + t. \tag{7}$$

This is a line in the second image, see Figure 2. We refer to this line as the **epipolar line** of $\mathbf{x}$. Since all scene points that can project to $\mathbf{x}$ are on the viewing ray, all points in the second image that can correspond $\mathbf{x}$ have to be on the epipolar line. This condition is called the **epipolar constraint**. For different points $\mathbf{x}$ in the first image we get different viewing rays that project to different epipolar lines. Since the viewing rays all pass through the camera center C_1 of the first camera the epipolar lines will all intersect each other in the projection $\mathbf{e}_2$ of the camera center C_1, see Figure 2. The projections of the camera centers $\mathbf{e}_1$ and $\mathbf{e}_2$ are called the **epipoles**.

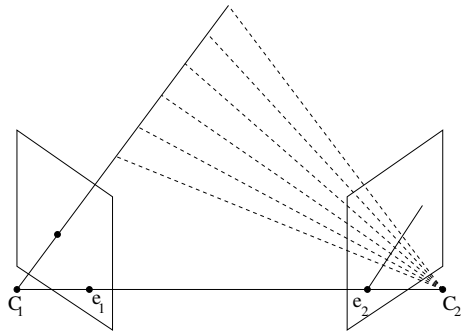

Figure 2: The projection of the viewing line into the second camera gives an epipolar line.

Exercise 1. Compute the epipoles for the camera pair $P_1 = \begin{bmatrix} I & 0 \end{bmatrix}$ and

$$P_2 = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}, \tag{8}$$

and verify that $\mathbf{e}_2$ lies on the epipolar line of $\mathbf{x} = (0, 1, 1)$.

For a general camera pair $P_1 = \begin{bmatrix} I & 0 \end{bmatrix}$, $P_1 = \begin{bmatrix} A & t \end{bmatrix}$, where A invertible we have that $\mathbf{C}_1 = (0, 0, 0, 1)$ and $\mathbf{C}_2 = \begin{bmatrix} -A^{-1}t \\ 1 \end{bmatrix}$ in homogeneous coordinates. Projecting into the two cameras gives $\mathbf{e}_2 \sim P_2\mathbf{C}_1 = t$. and $\mathbf{e}_1 \sim \begin{bmatrix} I & 0 \end{bmatrix} \begin{bmatrix} -A^{-1}t \\ 1 \end{bmatrix} = -A^{-1}t$.

The expression $\lambda A\mathbf{x} + t$ is a parametrization of the epipolar line of $\mathbf{x}$. Note that since $\mathbf{e}_2 = t$ we see that $\mathbf{e}_2$ is on the epipolar line by letting $\lambda = 0$. We know that a line in $\mathbb{P}^2$ can also be represented by a line equation $\mathbf{l}^T\mathbf{x} = 0$. To find the vector $\mathbf{l}$ we pick two points on the line, e.g. t and $A\mathbf{x} + t$ and insert into the line equation

$$\begin{cases} \mathbf{l}^T t = 0 \\ \mathbf{l}^T(A\mathbf{x} + t) = 0 \end{cases}. \tag{9}$$

These equations tells us that $\mathbf{l}$ has to be perpendicular to both t and $A\mathbf{x} + t$. We can find such an $\mathbf{l}$ using the vector product

$$\mathbf{l} = t \times (A\mathbf{x} + t) = t \times (A\mathbf{x}). \tag{10}$$

Since $t \sim \mathbf{e}_2$ we can also write this as $\mathbf{e}_2 \times (A\mathbf{x})$.

Exercise 2. If $P_1 = [I \quad 0]$ and

$$P_2 = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}, \tag{11}$$

which of the two points $\bar{\mathbf{x}}_1 = (1, 2, 1)$ and $\bar{\mathbf{x}}_2 = (1, 1, 1)$ in image 2 could correspond to $\mathbf{x} = (0, 1, 1)$ in image 1?

A cross pruduct $\mathbf{y} \times \mathbf{x}$ is a linear operation on $\mathbf{x}$ and can therefore be represented by a matrix which we denote $[\mathbf{y}]_\times$. If $\mathbf{x} = (x_1, x_2, x_3)$ and $\mathbf{y} = (y_1, y_2, y_3)$ then their cross product is

$$\mathbf{y} \times \mathbf{x} = (y_2x_3 - y_3x_2, y_3x_1 - y_1x_3, y_1x_2 - y_2x_1). \tag{12}$$

In matrix form we can write this

$$\underbrace{\begin{pmatrix} 0 & -y_3 & y_2 \\ y_3 & 0 & -y_1 \\ -y_2 & y_1 & 0 \end{pmatrix}}_{=[\mathbf{y}]_\times} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} y_2x_3 - y_3x_2 \\ y_3x_1 - y_1x_3 \\ y_1x_2 - y_2x_1 \end{pmatrix} \tag{13}$$

The matrix $[\mathbf{y}]_\times$ is skew symmetric, that is, $[\mathbf{y}]_\times = -[\mathbf{y}]_\times^T$. It is easy to see that for any 3×3 skew symmetric matrix S there is a vector $\mathbf{y}$ such that $S = [\mathbf{y}]_\times$.

With this notation the epipolar line can be written

$$\mathbf{l} = \mathbf{e}_2 \times (A\mathbf{x}) = [\mathbf{e}_2]_\times A\mathbf{x}. \tag{14}$$

The matrix $F = [\mathbf{e}_2]_\times A$ is called the fundamental matrix. It maps points in image 1 to lines in image 2. If $\bar{\mathbf{x}}$ corresponds to $\mathbf{x}$ then the epipolar constraint can be written

$$\bar{\mathbf{x}}^T\mathbf{l} = \bar{\mathbf{x}}^T F\mathbf{x} = 0. \tag{15}$$

Note that F only depends on the cameras and therefore the epipolar constraints only involves the image points and the camera P_2. We will use these constraints to compute F from a number of image correspondences. Once F has been determined the camera P_2 can be extracted.

Exercise 3. Let $P_1 = [I \quad 0]$ and

$$P_2 = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}. \tag{16}$$

Compute the fundamental matrix F and verify that $F\mathbf{e}_1 = 0$, $\mathbf{e}_2^T F = 0$ and $\det(F) = 0$.

3 Finding F: The Eight Point Algorithm

Recall that the objective of the two-view structure from motion problem is to find the scene points and the camera P_2. We will see in the next lecture that if the Fundamental matrix is known then P_2 can be extracted from it. We now present a simple algorithm for estimating F.

As we saw in the previous section, for each point correspondence $\mathbf{x}_i$,$\bar{\mathbf{x}}_i$ we get one epipolar constraint.

$$\bar{\mathbf{x}}_i^T F \mathbf{x}_i = 0. \tag{17}$$

If $\mathbf{x}_i \sim (x_i, y_i, z_i)$ and $\bar{\mathbf{x}} \sim (\bar{x}_i, \bar{y}_i, \bar{z}_i)$ then we can write this as

$$\begin{aligned} 0 = \bar{\mathbf{x}}_i^T F \mathbf{x}_i = \quad & F_{11}\bar{x}_i x_i + F_{12}\bar{x}_i y_i + F_{13}\bar{x}_i z_i \\ & + F_{21}\bar{y}_i x_i + F_{22}\bar{y}_i y_i + F_{23}\bar{y}_i z_i \\ & + F_{31}\bar{z}_i x_i + F_{32}\bar{z}_i y_i + F_{33}\bar{z}_i z_i. \end{aligned} \tag{18}$$

Therefore each correspondence gives one linear constraint on the entries of F. In matrix form we can write the resulting system as

$$\underbrace{\begin{pmatrix} \bar{x}_1 x_1 & \bar{x}_1 y_1 & \bar{x}_1 z_1 & \dots & \bar{z}_1 z_1 \\ \bar{x}_2 x_2 & \bar{x}_2 y_2 & \bar{x}_2 z_2 & \dots & \bar{z}_2 z_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \bar{x}_n x_n & \bar{x}_n y_n & \bar{x}_n z_n & \dots & \bar{z}_n z_n \end{pmatrix}}_{M} \begin{pmatrix} F_{11} \\ F_{12} \\ F_{13} \\ \vdots \\ F_{33} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}. \tag{19}$$

This is a linear homogeneous system which we can solve using SVD as in Lecture 3. The matrix F has 9 entries but the scale is arbitrary and the system therefore has 8 degrees of freedom. Each correspondence gives one new constraint on F and we therefore need 8 correspondences to solve this problem.

Note that it is in fact possible to solve the problem with only 7 point correspondences since we also have the constraint $\det(F) = 0$. However, this constraint is a polynomial of third order and we cannot use SVD to solve the resulting system. Therefore we use at least 8 correspondences.

Because of noise the resulting estimation $\tilde{F}$ of the fundamental matrix is not likely to have zero determinant. Therefore given this estimation we chose the matrix F that solves

$$\min_{\det(F)=0} \|\tilde{F} - F\| \tag{20}$$

(where the norm is the Frobenious/sum-of-squares norm). The solution to this problem is given by the SVD of $\tilde{F}$. If

$$USV^T = \tilde{F}, \tag{21}$$

where $S = \mathrm{diag}(\sigma_1, \sigma_2, \sigma_3)$. Then F can be found by setting the smallest singular value $\sigma_3 = 0$, that is,

$$F = U\mathrm{diag}(\sigma_1, \sigma_2, 0)V^T. \tag{22}$$

As was the case with the resection problem, normalization is important for numerical stability. If for example x_1 and $\bar{x}_1$ are both in the order of a 1000 pixels then $x_1\bar{x}_1 \approx 10^6$ while $z_1\bar{z}_1 = 1$. This makes the matrix $M^T M$ very poorly conditioned. To improve the numerics we can use the same normalization as in Lecture 3 (for both the cameras).

We summarize the different steps of the algorithm here:

- Extract at least 8 point correspondences.
- Normalize the coordinates (see Lecture 3).
- Form M and solve
$$\min_{||v||^2=1} ||Mv||^2,$$
using svd.
- Form the matrix F (and ensure that $\det(F) = 0$).
- Transform back to the original (un-normalized) coordinates.
- Compute P_2 from F (next lecture).
- Compute the scene points using triangulation (see Lecture 4).