Lecture 6: Camera Computation and the Essential Matrix

1   Computing Cameras From the Fundamental Matrix

In Lecture 5 we considered the two view structure from motion problem, that is, given a number of measured points in two images we want to compute both camera matrices and 3D points such that they project to the measurements. We showed that the 3D points can be eliminated from the problem by considering the fundamental matrix F. If $\mathbf{x}$ is an image point belonging to the fist image and $\bar{\mathbf{x}}$ belongs to the second then there is a 3D point that projects to to these if and only if the epipolar constraint

$$\bar{\mathbf{x}}^T F \mathbf{x} = 0 \tag{1}$$

is fulfilled. Using the projections of 8-scene points we can compute the fundamental matrix by solving a homogeneous least squares problem (the 8-point algorithm). What remains in order to find a solution to the two view structure from motion camera is to compute cameras from F and finally compute the 3D-points.

In general we may assume (see Lecture 5) that the cameras are of the form $P_1 = [I \quad 0]$ and $P_2 = [A \quad \mathbf{e}_2]$ where $\mathbf{e}_2$ is the epipole in the second image. Since we know that $F^T \mathbf{e}_2 = 0$ we can find $\mathbf{e}_2$ by computing the null space of F^T. A solution for the second camera is then given by

$$P_2 = [[\mathbf{e}_2]_\times F \quad \mathbf{e}_2]. \tag{2}$$

To see that these cameras have the fundamental matrix F we verify that their projections fulfill the epipolar constraint. If $\mathbf{X} = \begin{bmatrix} X \\ \rho \end{bmatrix}$, where $X \in \mathbb{R}^3$ and $\rho \in \mathbb{R}$ then the projections in the two cameras are given by

$$\mathbf{x} \quad \sim \begin{bmatrix} I & 0 \end{bmatrix} \begin{bmatrix} X \\ \rho \end{bmatrix} = X, \tag{3}$$

$$\bar{\mathbf{x}} \quad \sim \begin{bmatrix} [\mathbf{e}_2]_\times F & \mathbf{e}_2 \end{bmatrix} \begin{bmatrix} X \\ \rho \end{bmatrix} = [\mathbf{e}_2]_\times FX + \mathbf{e}_2 \rho = \mathbf{e}_2 \times (FX) + \mathbf{e}_2 \rho. \tag{4}$$

Therefore

$$\bar{\mathbf{x}}^T F \mathbf{x} = (\mathbf{e}_2 \times (FX) + \mathbf{e}_2 \rho)^T FX = (\mathbf{e}_2 \times (FX))^T FX + \rho \mathbf{e}_2^T FX. \tag{5}$$

Since $e_2 \times (FX)$ is a vector that is perpendicular to FX the term $(\mathbf{e}_2 \times (FX))^T FX = 0$. Furthermore, since $\mathbf{e}_2$ is in the null space of F^T the term $\rho \mathbf{e}_2^T FX = \rho (F^T \mathbf{e}_2)^T X = 0$. Therefore the projections of these camera matrices will fulfill the epipolar constraints.

Exercise 1. What is the camera center of $P_2 = [[\mathbf{e}_2]_\times F \quad \mathbf{e}_2]$? (Hint: Recall that $F\mathbf{e}_1 = 0$.)

Since there is a projective ambiguity there are many choices for P_2. Given that $P_1 = [I \quad 0]$ the general formula for P_2 is

$$P_2 = \begin{bmatrix} [\mathbf{e}_2]_\times F + \mathbf{e}_2 v^T & \lambda \mathbf{e}_2 \end{bmatrix}, \tag{6}$$

where v is some vector in $\mathbb{R}^3$ and λ is a non-zero scalar.

Exercise 2. Verify that the projections in $P_1 = [I \quad 0]$ and P_2 given by (6) fulfill the epipolar constraints for any $v \in \mathbb{R}^3$ and $\lambda \neq 0$.
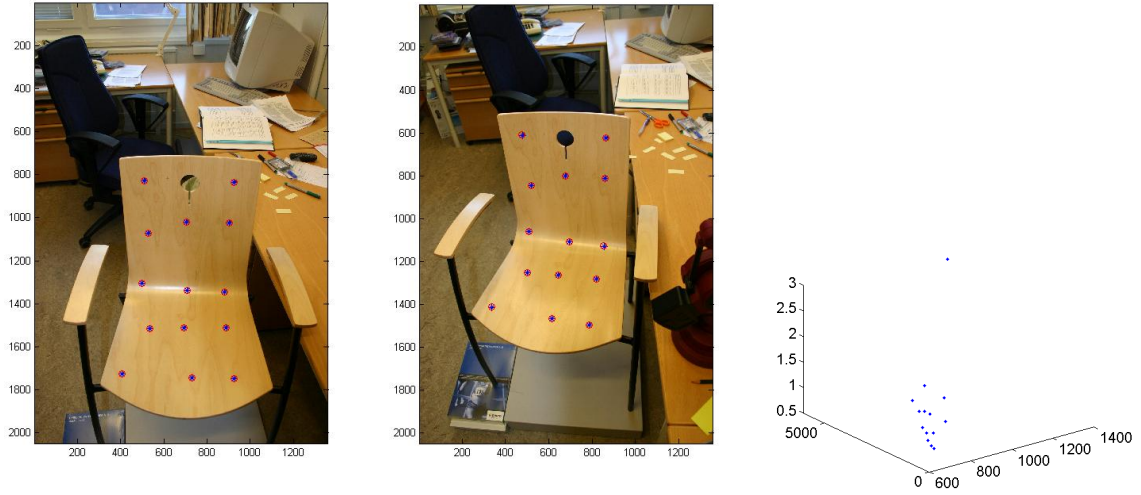
Figure 1: Two images of a chair with 14 known point correspondences. Blue ∗ are the image measurements and red *o* are the reprojections. The 3D points (to the right) look strange because of the projective ambiguity (note the difference in scale on the axes).

2   Relative Orientation: The Calibrated Case

When solving the relative orientation problem without camera calibration there is, as we saw in Lecture 3, an ambiguity. Basically any projective transformation can be applied to the 3D-points to give a new solution. Therefore the resulting constructions can often look strange even though the reprojections are correct (see Figure 1). To remove this ambiguity one has to add additional knowledge about the solution to the problem. For example, if we have some knowledge about the 3D scene, such as the distance between a few of the points, then we can apply a transform to the solution that make these distances correct.

Alternatively we can add knowledge about the cameras. If the inner parameters K_1 and K_2 are known we consider the calibrated two-view structure from motion problem. Given two sets of corresponding points $\mathbf{x}_i$ and $\bar{\mathbf{x}}_i$, $i = 1, ...n$ and inner parameters K_1 and K_2 our goal is to find $[R_1 \quad t_1]$, $[R_2 \quad t_2]$ and $\mathbf{X}_i$ such that

$$\mathbf{x}_i \sim K_1[R_1 \quad t_1]\mathbf{X}_i \tag{7}$$

$$\bar{\mathbf{x}}_i \sim K_2[R_2 \quad t_2]\mathbf{X}_i, \tag{8}$$

and R_1, R_2 are rotation matrices.

We can make two simplifications to the problem. First we normalize the cameras by multiplying equations (7) and (8) with K_1^{-1} and K_2^{-1} respectively. Furthermore, we apply the euclidean transformation

$$H = \begin{bmatrix} R_1^T & -R_1^T t_1 \\ 0 & 1 \end{bmatrix} \tag{9}$$

to the cameras (and H^{-1} to the 3D points). This gives us the new cameras

$$P_1 H = \begin{bmatrix} R_1 & t_1 \end{bmatrix} \begin{bmatrix} R_1^T & -R_1^T t_1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} I & 0 \end{bmatrix} \tag{10}$$

$$P_2 H = \begin{bmatrix} R_2 & t_2 \end{bmatrix} \begin{bmatrix} R_1^T & -R_1^T t_1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} \underbrace{R_2 R_1^T}_{=R} & \underbrace{-R_2 R_1^T t_1 - t_2}_{=t} \end{bmatrix}. \tag{11}$$

Therefore we search for a solution to the equations

$$\mathbf{y}_i \sim [I \quad 0]\mathbf{X}_i \tag{12}$$

$$\bar{\mathbf{y}}_i \sim [R \quad t]\mathbf{X}_i, \tag{13}$$

where $\mathbf{y}_i = K_1^{-1}\mathbf{x}_i$ and $\bar{\mathbf{y}}_i = K_1^{-1}\bar{\mathbf{x}}_i$ are the normalized image coordinates.

2.1   The Essential Matrix

The fundamental matrix for a pair of cameras of the form $[I \quad 0]$ and $[R \quad t]$ is given by

$$E = [t]_\times R, \tag{14}$$

and is called the **Essential matrix**. A rotation has 3 degrees of freedom and a translation 3. Since the scale of the essential matrix does not matter it therefore has 5 degrees of freedom. The reduction in freedom results in extra constraints on the singular values of E. In addition to having $\det(E) = 0$ the two non-zero singular values have to be equal. Furthermore, since the scale is arbitrary we can assume that these singular values are both 1. Therefore E has the SVD

$$E = U\mathrm{diag}([1 \quad 1 \quad 0])V^T. \tag{15}$$

The decomposition is not unique. We will assume that we have a singular value decomposition where $\det(UV^T) = 1$. It is easy to ensure this; If we have an SVD as in (15) with $\det(UV^T) = -1$ then we can simply switch the sign of the last column of V. Alternatively we can switch to $-E$ which then has the SVD

$$- E = U\mathrm{diag}([1 \quad 1 \quad 0])(-V)^T. \tag{16}$$

with $\det(U(-V)^T) = (-1)^3 \det(UV^T) = 1$. Note however that if we recompute the SVD for $-E$ we might get another decomposition since it is not unique.

To find the essential matrix we can use a slightly modified 8-point algorithm. From 8 points correspondences we form the M matrix (see Lecture 6) and solve the homogeneous least squares system

$$\min_{\|v\|^2=1} \|Mv\|^2 \tag{17}$$

using SVD. The resulting vector v can be used to form a matrix $\tilde{E}$ that does not necessarily have the right singular values $1, 1, 0$. We therefore compute the svd decomposition $\tilde{E} = USV^T$ and construct an essential matrix using $E = U\mathrm{diag}([1 \quad 1 \quad 0])V^T$.[1]

Since the essential matrix has only 5 degrees of freedom it is possible to find it using only 5 correspondences. However as in the case of the fundamental matrix the extra constraints are non-linear which makes estimation more difficult.

We summarize the steps of the modified 8-point algorithm here:

- Extract at least 8 point correspondences.

- Normalize the coordinates using K_1^{-1} and K_2^{-1} where K_1 and K_2 are the inner parameters of the cameras.

- Form M and solve

$$\min_{\|v\|^2=1} ||Mv||^2,$$

  using SVD.

- Form the matrix E (and ensure that E has the singular values $1, 1, 0$).

- Compute P_2 from E (next section).

- Compute the scene points using triangulation (see Lecture 4).

---

[1]Alternatively $E = U\mathrm{diag}([1 \quad 1 \quad 0])(-V)^T$ if $\det(UV^T) = -1$.

3   Computing Cameras from E

Once we have determined the essential matrix E we need extract cameras from it. Basically we want to decompose it into $E = SR$ where S is a skew symmetric matrix and R is a rotation. We will use the two matrices

$$W = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad Z = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{18}$$

The matrix W is a rotation and Z is skew symmetric. Furthermore, for these matrices we have that

$$ZW \quad = \text{diag}([1 \quad 1 \quad 0]) \tag{19}$$
$$ZW^T \quad = -\text{diag}([1 \quad 1 \quad 0]). \tag{20}$$

If E has the SVD from (15) we can now find two solutions; $E = S_1 R_1$, where

$$S_1 = -UZU^T, \quad R_1 = UW^T V^T \tag{21}$$

and $E = S_2 R_1$, where

$$S_2 = UZU^T, \quad R_2 = UWV^T. \tag{22}$$

To see that these are valid solutions we first verify that R_1 and R_2 are rotations. Since

$$R_1^T R_1 = (UW^T V^T)^T UW^T V^T = VWU^T UW^T V^T = I \tag{23}$$

R_1 is orthogonal. Furthermore,

$$\det(R_1) = \det(UW^T V^T) = \det(U)\det(W^T)\det(V^T) = \det(W)\det(UV^T) = 1, \tag{24}$$

and therefore R_1 is a rotation. (Note that if $\det(UV^T) = -1$ then the R_1 that we obtain is not a rotation but a rotation composed with a reflexion and therefore not a valid solution.) That S_1 is skew symmetric is easy to see since

$$-S_1^T = (UZU^T)^T = UZ^T U^T = -UZU^T = S_1. \tag{25}$$

Finally we see that

$$S_1 R_1 = -UZU^T UWV^T = -UZW^T V^T = -U(-\text{diag}([1 \quad 1 \quad 0]))V^T = E, \tag{26}$$

and therefore $S_1 R_1$ is a valid decomposition. $E = S_2 R_2$ can be verified similarly. Furthermore, it can be shown that these are the only two possibilities for each given E (see Hartley, Zisserman 2004).

When we have determined a decomposition $E = SR$ we need to compute a translation vector t from S such that $[t]_\times = S$. For such a t we have

$$St = [t]_\times t = t \times t = 0. \tag{27}$$

Therefore the vector t is in the null space of S. The null space of the two matrices S_1 and S_2 are the same and we can find it by looking in the third column of U.

Note that if t is in the null space of S then so is λt. In fact any non-zero λ gives a valid solution since $[\lambda t]_\times R = \lambda [t]_\times R = \lambda E$ which is also a valid essential matrix for the problem. Different λ corresponds to rescaling the solution and since there is a scale ambiguity we cannot determine a "true" value of λ. However the sign of λ is important since it determines whether points are in front of the cameras or not in the final reconstruction. To make sure that we can find a solution where the points are in front of both the cameras we therefore test $\lambda = \pm 1$.

If u_3 is the third column of U we get the four solutions

$$P_2 = \qquad [UWV^T \quad u_3] \quad \text{or} \quad [UW^T V^T \quad u_3] \qquad (\text{from } \lambda = 1) \tag{28}$$
$$\text{or} \quad [UWV^T \quad -u_3] \quad \text{or} \quad [UW^T V^T \quad -u_3] \qquad (\text{from } \lambda = -1) \tag{29}$$

When we have computed these four solutions we compute the 3D points using triangulation for all the choices of P_2 and select the one with where points are in front of both P_1 and P_2. Figure 2 shows the four calibrated reconstructions obtained using the images in Figure 1. Only one of them have all the points in front of both the cameras.
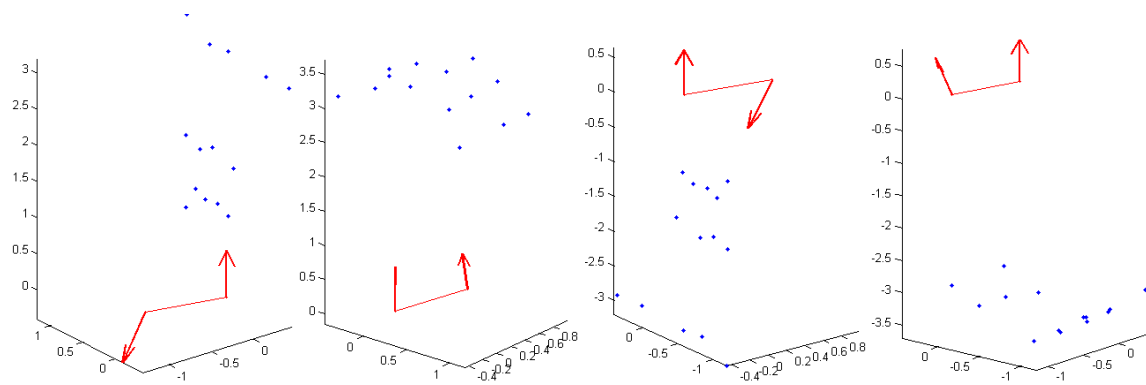
Figure 2: The 4 solutions when solving calibrated structure from motion for the chair image images in Figure 1. Only the second one have positive depths.