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. In what follows we will show that $A = [e_2]_\times F$ gives the correct epipolar geometry and therefore solution for the second camera is given by

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

The Fundamental matrix of the camera pair P_1 and P_2 is according to Lecture 5 given by $[t]_\times A = [\mathbf{e}_2]_\times [\mathbf{e}_2]_\times F$ we need to show that this expression reduces to F. The epipolar line of an arbitrary point $\mathbf{x}$ in image 1 is

$$[\mathbf{e}_2]_\times [\mathbf{e}_2]_\times F\mathbf{x} = \mathbf{e}_2 \times (\mathbf{e}_2 \times (F\mathbf{x}))). \tag{3}$$

Since e_2 is in the nullspace of F^T it is perpendicular to the columns of F and therefore also the vector $F\mathbf{x}$. This means $\mathbf{v}, \mathbf{e}_2, \mathbf{v} \times \mathbf{e}_2$, where $\mathbf{v} = \frac{1}{\|F\mathbf{x}\|} F\mathbf{x}$ forms a positive oriented orthonormal basis or $\mathbb{R}^3$. It is now easy to see that

$$\mathbf{e}_2 \times (\mathbf{e}_2 \times \mathbf{v}) = -\mathbf{e}_2 \times (\mathbf{v} \times \mathbf{e}_2) = -\mathbf{v}. \tag{4}$$

Therefore $\mathbf{e}_2 \times (\mathbf{e}_2 \times (F\mathbf{x}))) = F\mathbf{x}$ for all $\mathbf{x} \in \mathbb{R}^3$, which shows that $[\mathbf{e}_2]_\times [\mathbf{e}_2]_\times F = F$.

So in conclusion, if we have corresponding points that fulfill the epipolar constraints (1) then we can always find 3D points that project to these in the cameras $P_1 = \begin{bmatrix} I & 0 \end{bmatrix}$ and $P_2 = \begin{bmatrix} [\mathbf{e}_2]_\times F & \mathbf{e}_2 \end{bmatrix}$

Exercise 1. Find camera matrices P_1, P_2 such that

$$F = \begin{pmatrix} -1 & 0 & -1 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

and a 3D point X that projects to $x = (0, 1, 1)$ in P_1 and $x_2 = (1, 1, 1)$ in P_2.

The choice $P_2 = [[\mathbf{e}_2]_\times F \quad \mathbf{e}_2]$ may seem a little strange since the matrix $[\mathbf{e}_2]_\times F$ has the nullspace $\mathbf{e}_1$. Therefore we have

$$0 = [[\mathbf{e}_2]_\times F \quad \mathbf{e}_2] \begin{bmatrix} e_1 \\ 0 \end{bmatrix}, \tag{5}$$

which means that $P_2's$ camera center is a point at infinity. Since there is a projective ambiguity there are however many choices for P_2. Given that $P_1 = [I \quad 0]$ the general formula for P_2 is

$$P_2 = \left[[\mathbf{e}_2]_\times F + \mathbf{e}_2 v^T \quad \lambda \mathbf{e}_2\right], \tag{6}$$

where v is some vector in $\mathbb{R}^3$ and λ is a non-zero scalar. It is easy to verify that this camera pair gives the correct fundamental matrix similar to what we did previously.

Exercise 2. If F is as in Ex 1 is there v and λ such that

$$P_2 = \left[[e_2]_\times F + e_2 v^T \quad \lambda e_2\right] = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}?$$

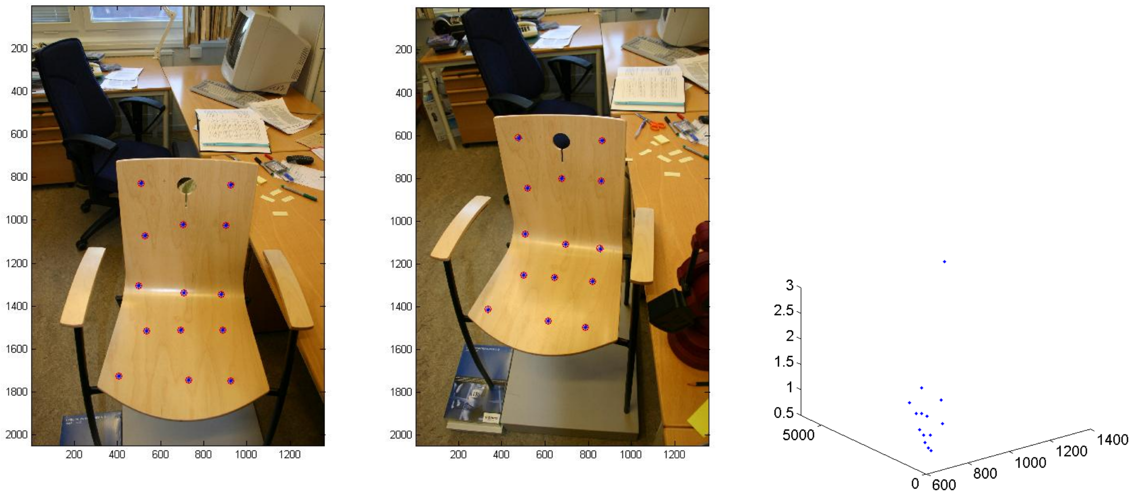

Figure 1: Two images of a chair with 14 known point correspondences. Blue $*$ are the image measurements and red o are the reprojections. The 3D points (to the right) look strange because of the projective ambiguity (note the difference in scale on the axes).

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 has 5 degrees of freedom. The reduction in freedom compared to F, 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

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

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

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

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 decomposition $\tilde{E} = USV^T$ and construct an essential matrix using $E = U\text{diag}([1 \quad 1 \quad 0])V^T$.[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 will consider this problem in Lecture 7.)

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

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

- 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).

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. For this purpose we will begin to find a decomposition of diag($\begin{bmatrix}1 & 1 & 0\end{bmatrix}$) $= ZW$, where Z is skew symmetric and W is a rotation. Since W is orthogonal we have diag($\begin{bmatrix}1 & 1 & 0\end{bmatrix}$) $= ZW \Leftrightarrow$ diag($\begin{bmatrix}1 & 1 & 0\end{bmatrix}$)$W^T = Z \Leftrightarrow$

$$\begin{pmatrix}1 & 0 & 0\\0 & 1 & 0\\0 & 0 & 0\end{pmatrix}\begin{pmatrix}w_{11} & w_{21} & w_{31}\\w_{12} & w_{22} & w_{32}\\w_{13} & w_{23} & w_{33}\end{pmatrix} = \begin{pmatrix}w_{11} & w_{21} & w_{31}\\w_{12} & w_{22} & w_{32}\\0 & 0 & 0\end{pmatrix} = \begin{pmatrix}0 & -z_3 & z_2\\z_3 & 0 & -z_1\\-z_2 & z_1 & 0\end{pmatrix}. \tag{18}$$

By inspecting the individual elements we see that $w_{11} = w_{22} = 0$, $w_{31} = z_2 = 0$, $w_{32} = -z_1 = 0$, and $w_{12} = z_3 = -w_{21}$. Since W is a rotation (with columns of length 1) it is clear that $w_{12} = \pm 1$. Choose $w_{12} = Z_3 = -1$ give $w_{21} = 1$ and because the third column of W is the vector product of the first two we get the solution

$$W = \begin{pmatrix}0 & -1 & 0\\1 & 0 & 0\\0 & 0 & 1\end{pmatrix} \quad \text{and} \quad Z = \begin{pmatrix}0 & 1 & 0\\-1 & 0 & 0\\0 & 0 & 0\end{pmatrix}. \tag{19}$$

Similarly if we chose $w_{12} = 1$, we see that second solution is given by W^T and Z^T.

To decompose E we now note that

$$E = U\text{diag}([1 \quad 1 \quad 0])V^T = UZWV^T = \underbrace{UZU^T}_{:=S_1}\underbrace{UWV^T}_{:=R_1} \tag{20}$$

and similarly

$$E = U\text{diag}([1 \quad 1 \quad 0])V^T = UZ^TW^TV^T = \underbrace{UZ^TU^T}_{:=S_2}\underbrace{UW^TV^T}_{:=R_2} \tag{21}$$

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

$$R_1^TR_1 = (UW^TV^T)^TUW^TV^T = VWU^TUW^TV^T = I \tag{22}$$

R_1 is orthogonal. Furthermore,

$$\det(R_1) = \det(UW^TV^T) = \det(U)\det(W^T)\det(V^T) = \det(W)\det(UV^T) = 1, \tag{23}$$

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

$$-S_1^T = (UZU^T)^T = UZ^TU^T = -UZU^T = S_1, \tag{24}$$

and therefore S_1R_1 is a valid decomposition. $E = S_2R_2$ can be verified similarly.

3.1 The Twisted Pair

To create the camera matrices corresponding to the solutions S_1R_1 and S_2R_2 we need to extract a translation vectors from the skew symmetric matrices S_1 and S_2. We note that since

$$S_1 = UZU^T = -UZ^TU^T = -S_2, \tag{25}$$

these are the same up to a scaling and it is therefore enough to determine t from S_1. Since $[t]_\times t = 0$ the vector t must be in the nullspace of S_1 which is the third column u_3 of U. We therefore obtain the solutions $P_1 = \begin{bmatrix} I & 0 \end{bmatrix}$

$$P_2 = \begin{bmatrix} UWV^T & u_3 \end{bmatrix} \text{ or } P_2' = \begin{bmatrix} UW^TV^T & u_3 \end{bmatrix}. \tag{26}$$

The two cameras P_2 and P_2' are called the twisted pair. The relative rotation between P_2 and P_2' is the rotation $R_1^TR_2 = VW^TU^TUW^TV^T = VW^TW^TV^T$. It is not hard to verify that if v_3 is the third column of V then $VW^TW^TV^Tv_3 = v_3$ and therefore v_3 is the rotation axis of $R_1^TR_2$. The rotation angle is given by

$$\cos(\phi) = \frac{\text{tr}(R_2^TR_1) - 1}{2} = \frac{\text{tr}(W^TW^T) - 1}{2} = -1, \tag{27}$$

which yields $\phi = \pi$. The camera centers of the two cameras P_2 and P_2' are

$$-R_1^Tu_3 = -VWU^Tu_3 = -VW\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = -V\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = -v_3 \tag{28}$$

and

$$-R_2^Tu_3 = -VW^TU^Tu_3 = -VW\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = -V\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = -v_3 \tag{29}$$

respectively. Hence in both solutions we are moving from the center of P_1 which is the origin in the direction $-v_3$. While P_2' is rotated $180°$ around this direction with respect to P_2.

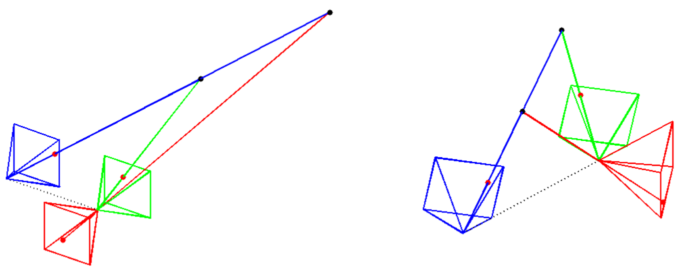

Figure 2: Two examples of twisted pair solution. Note that one of P_2 (green) and P_2' (red) has the reconstructed 3D behind itself.

3.2 Scale Ambiguity

Note that if $E = [t]_\times R$ then $\lambda E = [\lambda t]_\times R$ is also a valid solution. 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, see Figure 3. To make sure that we can find a solution where the points are in front of both the cameras we therefore test $\lambda = \pm 1$ and the twisted solution.

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

$$P_2 = \begin{bmatrix} UWV^T & u_3 \end{bmatrix} \text{ or } \begin{bmatrix} UW^TV^T & u_3 \end{bmatrix} \quad (\text{from } \lambda = 1) \tag{30}$$

$$\text{or } \begin{bmatrix} UWV^T & -u_3 \end{bmatrix} \text{ or } \begin{bmatrix} UW^TV^T & -u_3 \end{bmatrix} \quad (\text{from } \lambda = -1) \tag{31}$$

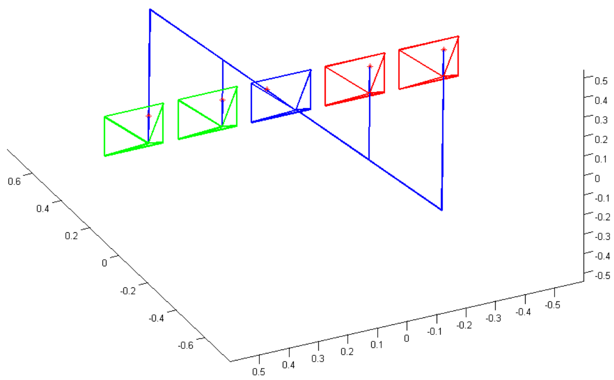

Figure 3: Reconstruction with different values of λ. Note that changing sign of λ moves the reconstructed points that are front of the camera to the rear of it and vice versa.

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 4 shows the four calibrated reconstructions obtained using the images in Figure 1. Only one of them has all the points in front of both the cameras.

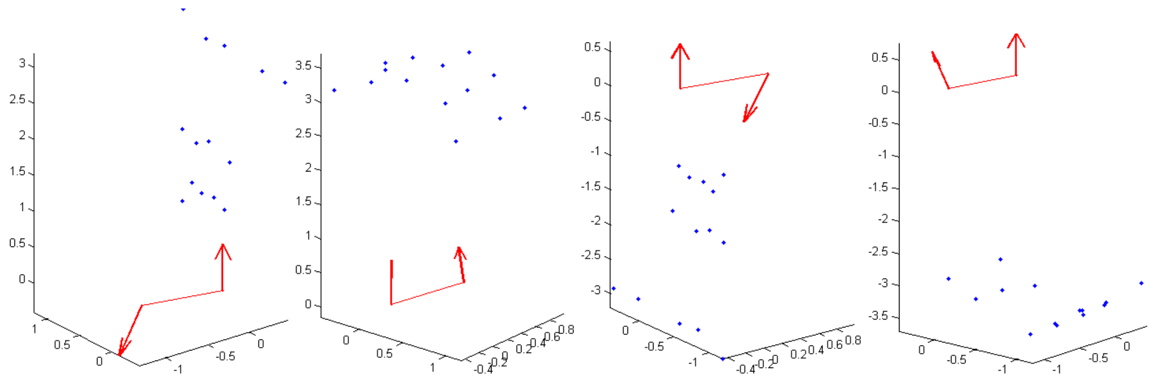

Figure 4: The 4 solutions when solving calibrated structure from motion for the chair image images in Figure 1. Only the second one has positive depths.