Epipolar geometry

April 28, 2020

We have seen how a rectified camera setup (where two cameras are oriented parallel to each other and are translated along the X axis) leads to a simplified correspondence search: the corresponding pixel is constrained to be on the same row. Does a similar constraint hold true in general? If it did, then we would not need to search the entire second image to find a correspondence for a given pixel in the first image.

It turns out that a similar constraint does indeed hold true in general. The reason is as follows (Figure 1, top left). Suppose we have two cameras in general position looking at a scene. Suppose we have a particular pixel $\vec{\mathbf{q}}_1$ in the first image. We know that the 3D point corresponding to this pixel must lie somewhere along the line connecting the pixel with the camera pinhole, and extending into the world. This line will still appear as a line in the second image. Thus the corresponding pixel in the second image must lie along a particular line, which is the image of this 3D line. This line in the second image is called the *epipolar line* in the second image corresponding to pixel $\vec{\mathbf{q}}_1$.

Thus for every pixel $\vec{\mathbf{q}}_1$ in image 1, there is a line in image 2 (called an *epipolar line*) on which the corresponding pixel is constrained to lie. This epipolar line is the *image* (in camera 2) of the ray connecting $\vec{\mathbf{q}}_1$ with the camera pinhole. It follows therefore that this epipolar line must pass through *the image of the first camera's pinhole in camera 2*. This is called the *epipole*. Thus, every epipolar line must pass through the epipole.

An alternate way of convincing yourself about this is as follows (Figure 1, top right). Suppose the two camera pinholes are $\mathbf{c}$ and $\mathbf{c}'$. Suppose $\mathbf{x}$ is a pixel in image 1. Consider the plane Π formed by $\mathbf{x}$, $\mathbf{c}$ and $\mathbf{c}'$. The world point $\mathbf{X}$ corresponding to $\mathbf{x}$ is constrained to lie on the line joining $\mathbf{c}$ and $\mathbf{x}$, so it must lie on this plane. The corresponding image location in image 2, $\mathbf{x}'$ lies on the line joining $\mathbf{X}$ and $\mathbf{c}'$, so it should also lie on this plane. At the same time it should lie on the image plane of camera 2. Therefore $\mathbf{x}'$ is constrained to lie on the intersection of the plane Π with the image plane. The intersection of two planes is a line: the *epipolar line*.

Figure 1 (bottom) shows an example of epipolar lines.

1 Deriving the epipolar constraint: the essential matrix

Let us derive the epipolar constraint mathematically. Let us assume, for now, that K for both cameras is I; we will remove this assumption in the next section.

Without loss of generality, we can set up our world coordinate system to be aligned with camera 1: the world origin is at the camera pinhole, and the world coordinate axes are aligned with the axes of the first camera. Then, the projection matrix of the first camera is $P^{(1)} = \begin{bmatrix} I & \mathbf{0} \end{bmatrix}$. We cannot say anything about the location and orientation of the second camera, so its projection matrix is $P^{(2)} = \begin{bmatrix} R & \mathbf{t} \end{bmatrix}$.

Suppose a world point $\mathbf{Q}$ is projected to $\mathbf{q}_1$ in image 1 and $\mathbf{q}_2$ in image 2. Then:

$$\vec{\mathbf{q}}_1 \equiv P^{(1)}\vec{\mathbf{Q}} \tag{1}$$

$$\equiv \begin{bmatrix} I & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{Q} \\ 1 \end{bmatrix} \tag{2}$$

$$\equiv \mathbf{Q} \tag{3}$$

$$\Rightarrow \lambda_1 \vec{\mathbf{q}}_1 = \mathbf{Q} \tag{4}$$

for some λ_1.

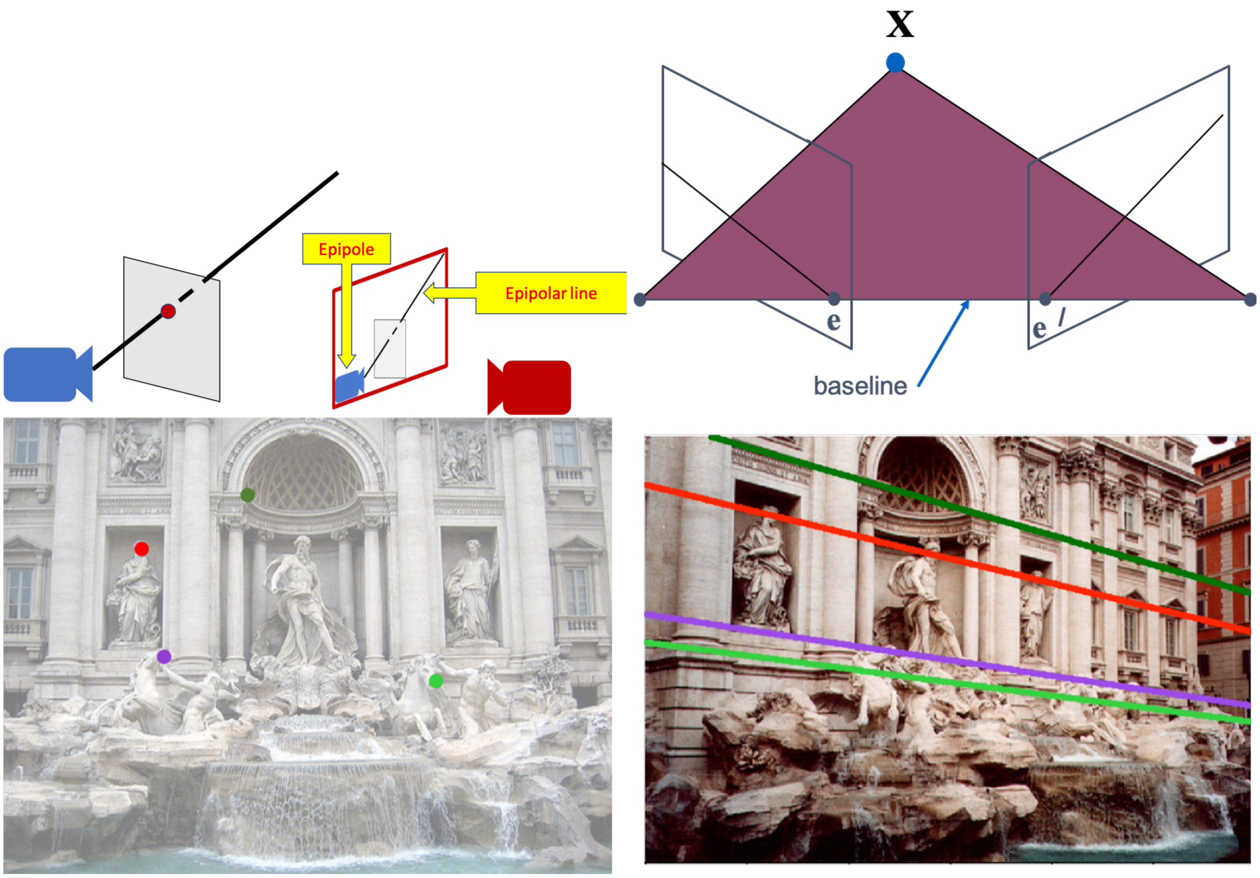

Figure 1: Top: Two explanations for epipolar lines. Bottom: example epipolar lines

Now going to the second image:

$$\vec{\mathbf{q}}_2 \equiv P^{(2)}\vec{\mathbf{Q}} \tag{5}$$

$$\equiv \begin{bmatrix} R & \mathbf{t} \end{bmatrix} \begin{bmatrix} \mathbf{Q} \\ 1 \end{bmatrix} \tag{6}$$

$$\equiv R\mathbf{Q} + \mathbf{t} \tag{7}$$

$$\Rightarrow \lambda_2 \vec{\mathbf{q}}_2 = R\mathbf{Q} + \mathbf{t} \tag{8}$$

Substituting equation (4) above, we get:

$$\lambda_2 \vec{\mathbf{q}}_2 = \lambda_1 R \vec{\mathbf{q}}_1 + \mathbf{t} \tag{9}$$

We now have an equation relating the two corresponding pixels, but it also includes the extraneous scalars λ_1 and λ_2. Let us try to remove them. First, we will take a cross product with $\mathbf{t}$:

$$\lambda_2 \mathbf{t} \times \vec{\mathbf{q}}_2 = \lambda_1 \mathbf{t} \times (R\vec{\mathbf{q}}_1) \tag{10}$$

Here we have used the fact that for any vector $\mathbf{a}$, $\mathbf{a} \times \mathbf{a} = 0$. Now, we take a dot product with $\vec{\mathbf{q}}_2$:

$$\lambda_2 \vec{\mathbf{q}}_2 \cdot (\mathbf{t} \times \vec{\mathbf{q}}_2) = \lambda_1 \vec{\mathbf{q}}_2 \cdot (\mathbf{t} \times (R\vec{\mathbf{q}}_1)) \tag{11}$$

$$\Rightarrow 0 = \lambda_1 \vec{\mathbf{q}}_2 \cdot (\mathbf{t} \times (R\vec{\mathbf{q}}_1)) \tag{12}$$

$$\Rightarrow 0 = \vec{\mathbf{q}}_2 \cdot (\mathbf{t} \times (R\vec{\mathbf{q}}_1)) \tag{13}$$

where we used the fact that for any 2 vectors $\mathbf{a}$ and $\mathbf{b}$, $\mathbf{a} \times \mathbf{b}$ is perpendicular to both, so $\mathbf{b} \cdot (\mathbf{a} \times \mathbf{b} = 0$.

To simplify this equation further, we convert all operations into matrix and vector products. First, it can be shown that if $\mathbf{a} = \begin{bmatrix} a_x \\ a_y \\ a_z \end{bmatrix}$, then $\mathbf{a} \times \mathbf{b} = [\mathbf{a}]_\times \mathbf{b}$, where $[\mathbf{a}]_\times$ is the following 3×3 matrix:

$$[\mathbf{a}]_\times = \begin{bmatrix} 0 & -a_z & a_y \\ a_z & 0 & -a_x \\ -a_y & a_x & 0 \end{bmatrix} \tag{14}$$

So we can write Equation (13) as:

$$\vec{\mathbf{q}}_2 \cdot ([\mathbf{t}]_\times R\vec{\mathbf{q}}_1) = 0 \tag{15}$$

Finally, it can be easily shown that $\mathbf{a} \cdot \mathbf{b} = \mathbf{a}^T \mathbf{b}$. Using this gives us:

$$\vec{\mathbf{q}}_2^T [\mathbf{t}]_\times R\vec{\mathbf{q}}_1 = 0 \tag{16}$$

The matrix $E = [\mathbf{t}]_\times R$ is called the *essential matrix*. Thus, if $\vec{\mathbf{q}}_1$ and $\vec{\mathbf{q}}_2$ are a pair of pixels, they must satisfy the following *epipolar constraint*:

$$\vec{\mathbf{q}}_2^T E \vec{\mathbf{q}}_1 = 0 \tag{17}$$

2 Epipolar lines and the epipole

How does the epipolar constraint yield epipolar lines? Consider a particular pixel $\vec{\mathbf{q}}_1$ in image 1. Then the corresponding pixel in image 2 must satisfy $\vec{\mathbf{q}}_2^T E \vec{\mathbf{q}}_1 = 0$. Since $\vec{\mathbf{q}}_1$ is known, we can write $E\vec{\mathbf{q}}_1$ as a single vector $\mathbf{l}$ of known coefficients. Then $\vec{\mathbf{q}}_2$ must satisfy: $\vec{\mathbf{q}}_2^T \mathbf{l} = 0$. This is the equation of a *line*: this is the epipolar line in image 2 corresponding to $\vec{\mathbf{q}}_1$. $\mathbf{l} = E\vec{\mathbf{q}}_1$ is the vector of coefficients of the line.

Similarly, we can show that if $\vec{\mathbf{q}}_2$ is a pixel in image 2, then the corresponding point $\vec{\mathbf{q}}_1$ in image 1 should satisfy $\mathbf{l}'^T \vec{\mathbf{q}}_1 = 0$, where $\mathbf{l}' = E^T \vec{\mathbf{q}}_2$. This is the epipolar line in image 1 corresponding to $\vec{\mathbf{q}}_2$.

What about the epipoles? Above, we said that the epipoles are the image of one camera's pinhole in the other camera. The first camera's pinhole is at the origin, $\mathbf{O} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$. It's image in image 2 is at

$$\vec{\mathbf{e}}_2 = P^{(2)} \vec{\mathbf{O}} \tag{18}$$

$$= \begin{bmatrix} R & \mathbf{t} \end{bmatrix} \begin{bmatrix} \mathbf{O} \\ 1 \end{bmatrix} \tag{19}$$

$$= \mathbf{t} \tag{20}$$

The claim is that all epipolar lines in image 2 pass through this epipole. As seen above, the equation of the epipolar line in image 2 corresponding to a pixel $\vec{\mathbf{q}}_1$ in image 1 is $\vec{\mathbf{q}}_2^T \mathbf{l} = 0$, where $\mathbf{l} = E\vec{\mathbf{q}}_1$. So we are claiming that for all such lines $\mathbf{l}$, $\vec{\mathbf{e}}_2^T \mathbf{l} = 0$. In other words, we are claiming that $\vec{\mathbf{e}}_2^T E \vec{\mathbf{q}}_1 = 0$ for all $\vec{\mathbf{q}}_1$. Let us see if this is true:

$$\vec{\mathbf{e}}_2^T E \vec{\mathbf{q}}_1 = \mathbf{t}^T E \vec{\mathbf{q}}_1 \tag{21}$$

$$= \mathbf{t}^T [\mathbf{t}]_\times R \vec{\mathbf{q}}_1 \tag{22}$$

$$= \mathbf{t} \cdot (\mathbf{t} \times (R\vec{\mathbf{q}}_1)) \tag{23}$$

$$= 0 \tag{24}$$

Thus, $\vec{\mathbf{e}}_2$ does indeed lie on every epipolar line in image 2.

What about the epipole in image 1? For that, we need to know the pinhole of camera 2. Recall from an earlier lecture that if the camera pinhole is at location $\mathbf{c}$, then $\mathbf{t} = -R\mathbf{c}$. We can invert this to get the location of the pinhole in terms of $\mathbf{t}$: $\mathbf{c} = -R^T \mathbf{t}$. The epipole in image 1 is thus:

$$\vec{\mathbf{e}}_1 = P^{(1)} \vec{\mathbf{c}} = \begin{bmatrix} I & \mathbf{0} \end{bmatrix} \begin{bmatrix} -R^T \mathbf{t} \\ 1 \end{bmatrix} = -R^T \mathbf{t} \tag{25}$$

We need to show that this lies on every epipolar line $\mathbf{l}' = E^T \vec{\mathbf{q}}_2$, which means that we need to show $\mathbf{l}'^T \vec{\mathbf{e}}_1 = \vec{\mathbf{q}}_2^T E \vec{\mathbf{e}}_1$ is 0 for all $\vec{\mathbf{q}}_2$.

$$\vec{\mathbf{q}}_2^T E \vec{\mathbf{e}}_1 = \vec{\mathbf{q}}_2^T [\mathbf{t}]_\times R \vec{\mathbf{e}}_1 \tag{26}$$

$$= -\vec{\mathbf{q}}_2^T [\mathbf{t}]_\times R R^T \mathbf{t} \tag{27}$$

$$= -\vec{\mathbf{q}}_2^T [\mathbf{t}]_\times \mathbf{t} \tag{28}$$

$$= -\vec{\mathbf{q}}_2^T \mathbf{t} \times \mathbf{t} \tag{29}$$

$$= 0 \tag{30}$$

Thus $\vec{\mathbf{e}}_1$ does indeed lie on every epipolar line in image 1.

Rank of the essential matrix As we have seen above,

$$\vec{\mathbf{q}}_2^T E \vec{\mathbf{e}}_1 = 0 \;\; \forall \vec{\mathbf{q}}_2 \tag{31}$$

$$\Rightarrow E\vec{\mathbf{e}}_1 = \mathbf{0} \tag{32}$$

Thus, there exists a non-zero vector $\vec{\mathbf{e}}_1$ for which $E\vec{\mathbf{e}}_1 = 0$. This means that E cannot be full rank: it must have rank less than 3. It turns out that E **has rank 2**.

3 Estimating the essential matrix

If we know R and $\mathbf{t}$ (which we would if we knew the location and orientation of the two cameras), then we can figure out E. But what if we don't know R and $\mathbf{t}$?

Again, we can *estimate* E from correspondences. Given a pair of corresponding pixels $\vec{\mathbf{q}}_1 \equiv \begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix}$ and $\vec{\mathbf{q}}_2 \equiv \begin{bmatrix} x_2 \\ y_2 \\ 1 \end{bmatrix}$, we can write down a constraint on $E = \begin{bmatrix} E_{11} & E_{12} & E_{13} \\ E_{21} & E_{22} & E_{23} \\ E_{31} & E_{32} & E_{33} \end{bmatrix}$:

$$\vec{\mathbf{q}}_2^T E \vec{\mathbf{q}}_1 = 0 \tag{33}$$

$$\Rightarrow \begin{bmatrix} x_2 & y_2 & 1 \end{bmatrix} \begin{bmatrix} E_{11} & E_{12} & E_{13} \\ E_{21} & E_{22} & E_{23} \\ E_{31} & E_{32} & E_{33} \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix} = 0 \tag{34}$$

$$\Rightarrow E_{11}x_1x_2 + E_{12}y_1x_2 + E_{13}x_2 + E_{21}x_1y_2 + E_{22}y_1y_2 + E_{23}y_2 + E_{31}x_1 + E_{32}y_1 + E_{33} = 0 \tag{35}$$

This is a linear equation in the entries of E. Thus one correspondence yields one linear equation.

If we have a set of correspondences, we can set up a system of linear equations. As with camera calibration, the equations take the form:

$$A\mathbf{e} = 0 \tag{36}$$

where $\mathbf{e} = \begin{bmatrix} E_{11} \\ E_{12} \\ \vdots \\ E_{33} \end{bmatrix}$ is the vector of unknowns. Again as with camera calibration, if E is a solution, so is αE. We therefore add an additional constraint that $\|E\|_F = 1$, and as before, solve the following minimization problem via SVD:

$$\begin{aligned} \min_{\mathbf{e}} A\mathbf{e} \\ \text{s.t} \\ \|\mathbf{e}\| = 1 \end{aligned} \tag{37}$$

However, this is not enough. Above, we noted that the essential matrix must be a rank 2 matrix. However, the solution method described above does not enforce that. How should we enforce this constraint?

Unfortunately, enforcing the rank constraint in the optimization problem above makes it intractable to optimize. Instead, we take the solution from the optimization problem above, say E_0, and we try to find the best rank-2 matrix that approximates E_0:

$$\begin{aligned} \min_{E} \|E - E_0\|_F \\ \text{s.t} \\ rank(E) = 2 \end{aligned} \tag{38}$$

This optimization problem is also solved using an SVD. We first compute an SVD of $E_0 : E_0 = U\Sigma V^T$. Σ is a diagonal 3×3 matrix. We find the diagonal entry that is the smallest in absolute value and set it to 0 to get Σ'. Then we compute E as $E = U\Sigma' V^T$.

How many correspondences do we need to estimate E? E has 9 entries. We get 1 equation from the constraint that $\|E\|_F = 1$, so we need 8 more. Each correspondence gives us a single equation, so we need at least 8 correspondences[1].

While we will not delve deeper into this, we can actually estimate $\mathbf{t}$ and R from E using an SVD.

4 The Fundamental matrix

In the above derivation of the essential matrix, we assumed that K is identity for both cameras. What if K is not identity?

[1] Actually, we only need 7, if we use the rank constraint to parametrize E differently

In this case, we will start with the following two equations:

$$\vec{\mathbf{q}}_1 \equiv K_1 \begin{bmatrix} I & \mathbf{0} \end{bmatrix} \vec{\mathbf{Q}} \tag{39}$$

$$\vec{\mathbf{q}}_2 \equiv K_2 \begin{bmatrix} R & \mathbf{t} \end{bmatrix} \vec{\mathbf{Q}} \tag{40}$$

Let us write $\vec{\mathbf{q}}_1' \equiv K_1^{-1}\vec{\mathbf{q}}_1$, and $\vec{\mathbf{q}}_2' \equiv K_2^{-1}\vec{\mathbf{q}}_2$. Then:

$$\vec{\mathbf{q}}_1' \equiv \begin{bmatrix} I & \mathbf{0} \end{bmatrix} \vec{\mathbf{Q}} \tag{41}$$

$$\vec{\mathbf{q}}_2' \equiv \begin{bmatrix} R & \mathbf{t} \end{bmatrix} \vec{\mathbf{Q}} \tag{42}$$

Following the derivation we did earlier in this documentation, we get:

$$\vec{\mathbf{q}}_2'^T [\mathbf{t}]_\times R \vec{\mathbf{q}}_1' = 0 \tag{43}$$

$$\Rightarrow \vec{\mathbf{q}}_2^T K_2^{-T} [\mathbf{t}]_\times R K_1^{-1} \vec{\mathbf{q}}_1 = 0 \tag{44}$$

$$\Rightarrow \vec{\mathbf{q}}_2^T F \vec{\mathbf{q}}_1 = 0 \tag{45}$$

Here $F = K_2^{-T}[\mathbf{t}]_\times R K_1^{-1}$ is the *Fundamental matrix.*

The fundamental matrix has the same properties as the essential matrix: it yields epipolar lines and epipoles, and it has rank 2. However, from the fundamental matrix we cannot estimate $\mathbf{t}$ and R from F.

5 Structure from motion

We now have everything we need for 3D reconstruction from a pair of cameras. We can imagine three scenarios:

1. If we can choose where to place cameras, then we can place the two cameras parallel to each other translated along the X axis (the rectified camera setup).
2. If not, but we know some (around 6) 3D points and their projections in the 2 images, we can estimate the projection matrices for both cameras. We can then compute E and use the epipolar constraint to clean up correspondences. The correspondences yield 3D points through triangulation.
3. Otherwise we can use 8 correspondences to estimate E, and from that estimate $\mathbf{t}$ and R (assuming K is I). We can then use correspondences to triangulate.

The problem of reconstructing a 3D scene from a pair of unknown cameras (route 3) is called *structure from motion.*