Epipolar Geometry and the Essential Matrix

Carlo Tomasi

The *epipolar geometry* of a pair of cameras expresses the fundamental relationship between any two corresponding points in the two image planes, and leads to a key constraint between the coordinates of these points that underlies visual reconstruction. The first Section below describes the epipolar geometry. The Section thereafter expresses its key constraint algebraically.

1 The Epipolar Geometry of a Pair of Cameras

Figure 1 shows the main elements of the epipolar geometry for a pair of cameras.

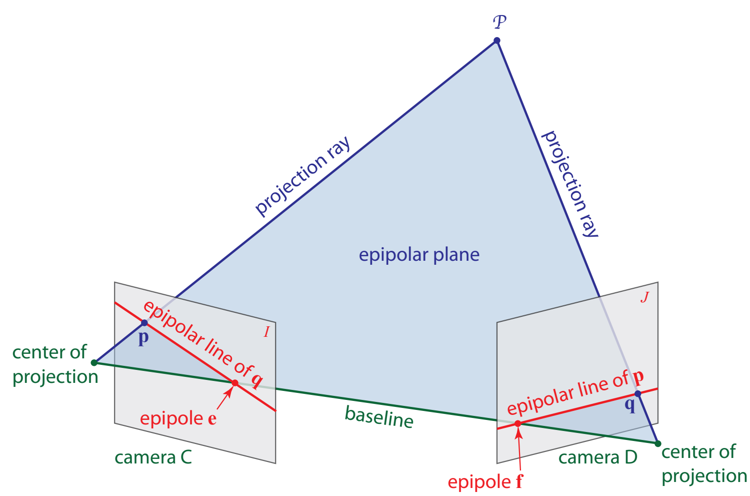

Figure 1: Essential elements of the epipolar geometry of a camera pair.

The world point $\mathcal{P}$ and the centers of projection of the two cameras identify a plane in space, the *epipolar plane* of point $\mathcal{P}$. The Figure shows a triangle of this plane, delimited by the two projection rays and by the *baseline* of the camera pair, that is, the line segment that connects the two centers of projection.[1]

If the image planes are thought of extending indefinitely, the baseline intersects the two image planes at two points called the *epipoles* of the two images. In particular, if the cameras are arranged so that the baseline is parallel to an image plane, then the corresponding epipole is a point at infinity.

The epipoles are fixed points for a given camera pair configuration. With cameras somewhat tilted towards each other, and with a sufficiently wide field of view, the epipoles would be image points. Epipole $\mathbf{e}$ in the image I taken by camera C would be literally the image of the center of projection of camera D in

[1] We use the term "baseline" for the line *segment*. However, this term is also often used for the entire *line* through the two centers of projection.

I, and *vice versa*. Even if the two cameras do not physically see each other, we maintain this description in an abstract sense: each epipole is the image of one camera in the other image.

The epipolar plane intersects the two image planes along the two *epipolar lines* of point $\mathcal{P}$, each of which passes by construction through one of the two projection points $\mathbf{p}$ and $\mathbf{q}$ and one of the two epipoles. Thus, epipolar lines come in corresponding pairs, and the correspondence is established by the single epipolar plane for the given point $\mathcal{P}$.

For a different world point $\mathcal{P}$, the epipolar plane changes, and with it do the image projections of $\mathcal{P}$ and the epipolar lines. However, all epipolar planes contain the baseline. Thus, the set of epipolar planes forms a *pencil* of planes supported by the line through the baseline, and the epipoles are fixed.

Suppose now that we are given the two images I and J taken by cameras C and D and a point $\mathbf{p}$ in I. We do not know where the corresponding point $\mathbf{q}$ is in the other image, nor where the world point $\mathcal{P}$ is, except that $\mathcal{P}$ must be somewhere along the projection ray of $\mathbf{p}$. However, *if we know the relative position and orientation of the two cameras*, we know where centers of projection are relative to each other. The two centers of projection and point $\mathbf{p}$ identify the epipolar plane, and this in turn determines the epipolar line of point $\mathbf{p}$ in image J. The point $\mathbf{q}$ must be somewhere on this line. This same construction holds for any other point $\mathbf{p}$ on the epipolar line in image I.

To understand what the epipolar constraint expresses, consider that the projection rays for two arbitrary points in the two images are generically two skew lines in space. The projection rays of two *corresponding* points, on the the other hand, are coplanar with each other and with the baseline. The epipolar geometry captures this key constraint, and pairs of point that do not satisfy the constraint cannot possibly correspond to each other.

2 The Essential Matrix

This section expresses the *epipolar constraint* described in the previous note algebraically. To emphasize the distinction between coordinate vectors in two different reference frames, coordinates in the second system have a prime.

The *standard reference system* for a camera C is a right-handed Cartesian coordinate system with its origin at the center of projection of C, its positive Z axis pointing towards the scene along the optical axis of the lens, and its X axis pointing to the right[2] along the rows of the camera sensor. As a consequence, the Y axis points downwards along the columns of the sensor. Coordinates in the standard reference system are measured in units of focal distance. Let

$$\mathbf{p} = \begin{bmatrix} p_1 \\ p_2 \\ f \end{bmatrix} \quad \text{and} \quad \mathbf{q}' = \begin{bmatrix} q_1' \\ q_2' \\ f' \end{bmatrix}$$

denote the image coordinates of the projections of the same 3D point $\mathcal{P}$ in the standard reference systems of two cameras C and D, and let

$$\mathbf{p}' = R(\mathbf{p} - \mathbf{t}) \tag{1}$$

be the rigid transformation between the two reference systems. Note that the coordinates of $\mathbf{p}$ are in reference system C and those of $\mathbf{q}'$ are in reference system D: An image point is measured in its own camera's reference system.

[2]When the camera is upside-up and viewed from behind it, as when looking through its viewfinder.

Solving equation (1) for $\mathbf{p}$ yields

$$\mathbf{p} = R^T \mathbf{p}' + \mathbf{t} = R^T(\mathbf{p}' + R\mathbf{t})$$

so that the reverse transformation is

$$\mathbf{p} = R'(\mathbf{p}' - \mathbf{s}') \quad \text{if we let} \quad R' = R^T \quad \text{and} \quad \mathbf{s}' = -R\mathbf{t} \; . \tag{2}$$

When expressed in the reference system of camera C, the directions of the projection rays through corresponding image points with coordinates $\mathbf{p}$ and $\mathbf{q}'$ are along the vectors

$$\mathbf{p} \quad \text{and} \quad R^T \mathbf{q}' \; ,$$

and the baseline in this reference system is along the translation vector with Euclidean coordinates $\mathbf{t}$. Coplanarity of these three vectors can be expressed by stating that their triple product is zero:

$$(R^T \mathbf{q}')^T (\mathbf{t} \times \mathbf{p}) = 0 \quad \text{that is,} \quad (\mathbf{q}')^T R \, (\mathbf{t} \times \mathbf{p}) = 0 \quad \text{or} \quad (\mathbf{q}')^T R \, [\mathbf{t}]_\times \mathbf{p} = 0$$

where $\mathbf{t} = \left[\begin{array}{ccc} t(1) & t(2) & t(3)]^T \end{array} \right]$ and

$$[\mathbf{t}]_\times = \left[\begin{array}{ccc} 0 & -t_3 & t_2 \\ t_3 & 0 & -t_1 \\ -t_2 & t_1 & 0 \end{array} \right]$$

is the matrix that expresses the cross-product of $\mathbf{t}$ with any other vector.

In summary, for corresponding points $\mathbf{p}$ and $\mathbf{q}'$ the following equation holds:

$$(\mathbf{q}')^T E \mathbf{p} = 0 \tag{3}$$

where

$$E = R \, [\mathbf{t}]_\times \; . \tag{4}$$

Equation (3) is called the *epipolar constraint* and the matrix E is called the *essential matrix*. Equation (3) expresses the coplanarity between *any* two points $\mathbf{p}$ and $\mathbf{q}'$ on the same epipolar plane for two fixed cameras. If $\mathbf{p}$ is fixed in image I, then any point $\mathbf{q}'$ in image J for which equation (3) holds must be on the epipolar line of $\mathbf{p}$. Formally, the product

$$\boldsymbol{\mu}'_{\mathbf{p}} = E\mathbf{p} \tag{5}$$

for fixed $\mathbf{p}$ is a column vector, and equation (3) becomes

$$(\mathbf{q}')^T \boldsymbol{\mu}'_{\mathbf{p}} = 0 \; .$$

This is the equation of the epipolar line in image J. Conversely, if $\mathbf{q}'$ is fixed in the second image, then the product

$$\boldsymbol{\lambda}^T_{\mathbf{q}'} = (\mathbf{q}')^T E \tag{6}$$

is a row vector, and equation (3) becomes

$$\boldsymbol{\lambda}^T_{\mathbf{q}'} \mathbf{p} = 0 \tag{7}$$

which is the equation of the epipolar line of $\mathbf{q}'$ in image I.

Since the epipole in image I belongs to all epipolar lines in I, the vector $\mathbf{e}$ of its homogeneous coordinates must satisfy equation (7) regardless of what point $\mathbf{q}'$ is used in the definition (6) of $\boldsymbol{\lambda}_{\mathbf{q}'}$. This can happen only if $\mathbf{e}$ is in the null space of E, which must therefore be degenerate.

Algebraically, it is easy to see that the rank of E is two for any nonzero $\mathbf{t}$. To this end, note first that the matrix $[\mathbf{t}]_\times$ has rank two if $\mathbf{t}$ is nonzero, because

$$[\mathbf{t}]_\times \mathbf{t} = \mathbf{t} \times \mathbf{t} = \mathbf{0}$$

so the null space of $[\mathbf{t}]_\times$ is the line through the origin and along $\mathbf{t}$. Since R is full rank, also the product $E = R\,[\mathbf{t}]_\times$ has rank 2 if $\mathbf{t} \neq \mathbf{0}$, and the null space of E is spanned by $\mathbf{t}$ as well.

Since $[\mathbf{t}]_\times$ is skew-symmetric, $[\mathbf{t}]_\times^T = -[\mathbf{t}]_\times$. So from equations (2) and (4) we obtain

$$E^T \mathbf{s}' = -[\mathbf{t}]_\times R^T(-R\mathbf{t}) = [\mathbf{t}]_\times \mathbf{t} = 0$$

so that $\mathbf{s}'$ spans the null space of E^T, that is, the left null space of E.

Since the left epipole $\mathbf{e}$ is along $\mathbf{t}$ and the right epipole $\mathbf{f}'$ is along $\mathbf{s}'$, we can summarize this discussion by stating that the epipole $\mathbf{e}$ in image I and the translation vector $\mathbf{t}$ are both in the null space of E:

$$E\mathbf{e} = E\mathbf{t} = \mathbf{0}$$

and the epipole $\mathbf{f}'$ in image J and the translation vector $\mathbf{s}'$ are both in the left null space of E:

$$(\mathbf{f}')^T E = (\mathbf{s}')^T E = \mathbf{0}^T \ .$$

The essential matrix E has special structure, as we now show. For any vector $\mathbf{v}$ orthogonal to $\mathbf{t}$, the definition of cross product yields

$$\|[\mathbf{t}]_\times \mathbf{v}\| = \|\mathbf{t}\|\,\|\mathbf{v}\| \ .$$

Because of this, the matrix $[\mathbf{t}]_\times$ maps all unit vectors in its row space into vectors of magnitude $\|\mathbf{t}\|$. In other words, the two nonzero singular values of $[\mathbf{t}]_\times$ are equal to each other (and to $\|\mathbf{t}\|$). Since multiplication by an orthogonal matrix does not change the matrix's singular values, we conclude that the essential matrix E has two nonzero singular values equal to each other, and a zero singular value. The left and right singular vectors $\mathbf{u}_3$ and $\mathbf{v}_3$ corresponding to the zero singular value of E are unit vectors along the epipoles and the translation vectors,

$$\mathbf{v}_3 \sim \mathbf{e} \sim \mathbf{t} \quad \text{and} \quad \mathbf{u}_3 \sim \mathbf{f}' \sim \mathbf{s}' \ . \tag{8}$$

In these expressions, the symbol '$\sim$' means "proportional to," or "equal up to a multiplicative constant." Since the other two singular values of E are equal to each other, the corresponding singular vectors are arbitrary, as long as they form orthonormal triples with $\mathbf{u}_3$ and $\mathbf{v}_3$. The preceding discussion is summarized in Table 1.

The epipolar constraint (3) is used in two different contexts. In stereo vision, R and $\mathbf{t}$ and therefore E are known. Given a point $\mathbf{p}$ in I, the epipolar constraint allows restricting the search for a corresponding point $\mathbf{q}$ to the epipolar line of $\mathbf{p}$. In visual reconstruction, several pairs $(\mathbf{p}_i, \mathbf{q}_i)$ of corresponding points are given, and equation (3) for each pair yields a linear equation in the entries of E. From this, E and then R and $\mathbf{t}$ can be found, as we will see in a later note.

For a camera pair (C, D) with nonzero baseline, let

$$\mathbf{p}' = R(\mathbf{p} - \mathbf{t}) \quad \text{and} \quad \mathbf{p} = R'(\mathbf{p}' - \mathbf{s}') \quad \text{with} \quad R' = R^T \quad \text{and} \quad \mathbf{s}' = -R\mathbf{t}$$

be the coordinate transformations between points $\mathbf{p}$ in C and points $\mathbf{p}'$ in D. The *essential matrix* of the camera pair is the matrix

$$E = R\,[\mathbf{t}]_\times \quad \text{where} \quad [\mathbf{t}]_\times = \begin{bmatrix} 0 & -t_3 & t_2 \\ t_3 & 0 & -t_1 \\ -t_2 & t_1 & 0 \end{bmatrix} .$$

The *epipole* $\mathbf{e}$ in the image I taken by C and that of $\mathbf{f}'$ in the image J taken by D satisfy

$$E\mathbf{e} = E^T\mathbf{f}' = \mathbf{0} \quad \text{and also} \quad E\mathbf{t} = E^T\mathbf{s}' = \mathbf{0} .$$

A point $\mathbf{p}$ in image I and its corresponding point $\mathbf{q}'$ in image J satisfy the *epipolar constraint*

$$(\mathbf{q}')^T E\mathbf{p} = 0 .$$

This equation can also be written as follows:

$$\boldsymbol{\lambda}_{\mathbf{q}'}^T\mathbf{p} = (\boldsymbol{\mu}'_{\mathbf{p}})^T\mathbf{q}' = 0$$

where

$$\boldsymbol{\lambda}_{\mathbf{q}'} = E^T\mathbf{q}' \quad \text{and} \quad \boldsymbol{\mu}'_{\mathbf{p}} = E\mathbf{p}$$

are the vectors of coefficients of the *epipolar line* of $\mathbf{q}'$ in image I and that of $\mathbf{p}$ in image J respectively.
The singular value decomposition of E is

$$E \sim U\Sigma V^T = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 \end{bmatrix} \text{diag}(1,1,0) \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 \end{bmatrix}^T \tag{9}$$

where $\mathbf{u}_1$, $\mathbf{u}_2$, $\mathbf{v}_1$, $\mathbf{v}_2$ are any vectors for which U and V become orthogonal and

$$\mathbf{v}_3 \sim \mathbf{e} \sim \mathbf{t} \quad \text{and} \quad \mathbf{u}_3 \sim \mathbf{f}' \sim \mathbf{s}'$$

Table 1: Definition and properties of the essential matrix.