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.

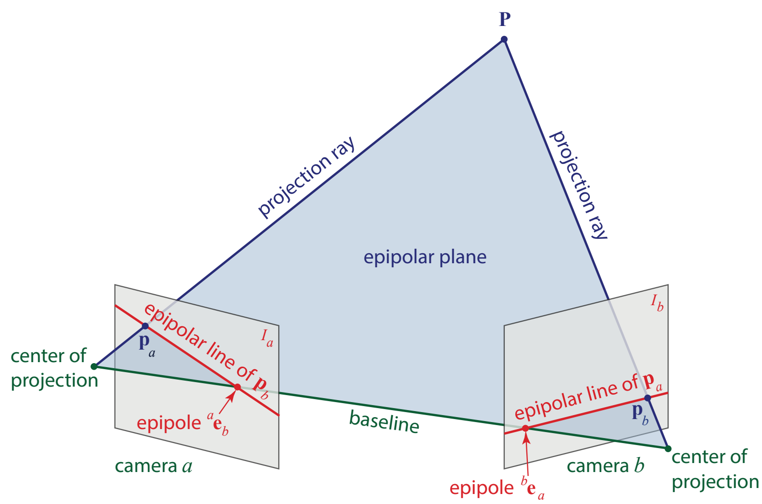

Figure 1: Essential elements of the epipolar geometry of a camera pair.

The world point $\mathbf{P}$ and the centers of projection of the two cameras identify a plane in space, the *epipolar plane* of point $\mathbf{P}$. 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 $\mathbf{e}_b$ in the image I_a taken by camera a would be literally the image of the center of projection of camera b 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_a, 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. Note that the epipole in image I_a is called $\mathbf{e}_b$, because it is the image *of* camera b from camera a. Similar considerations hold for $\mathbf{e}_a$.

The epipolar plane intersects the two image planes along the two *epipolar lines* of point $\mathbf{P}$, each of which passes by construction through one of the two projection points $\mathbf{p}_a$ and $\mathbf{p}_b$ 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 $\mathbf{P}$.

For a different world point $\mathbf{P}$, the epipolar plane typically changes, and with it do the image projections of $\mathbf{P}$ 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_a and I_b taken by cameras a and b and a point $\mathbf{p}_a$ in I_a. We do not know where the corresponding point $\mathbf{p}_b$ is in the other image, nor where the world point $\mathbf{P}$ is, except that $\mathbf{P}$ must be somewhere along the projection ray of $\mathbf{p}_a$. However, *if we know the relative position and orientation of the two cameras*, we know where the two centers of projection are relative to each other. The two centers of projection and point $\mathbf{p}_a$ identify the epipolar plane, and this in turn determines the epipolar line of point $\mathbf{p}_a$ in image I_b. The point $\mathbf{p}_b$ must be somewhere on this line. This same construction holds for any other point $\mathbf{p}_a$ on the epipolar line in image I_a.

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, because they belong to the same epipolar plane. 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 section algebraically.

Coordinate Systems. The standard reference system for camera a is a right-handed Cartesian coordinate system with its origin at the center of projection of a, 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. The standard reference system for camera b is defined similarly. Let

$$^a\mathbf{p}_a = \begin{bmatrix} {}^a x_a \\ {}^a y_a \\ f \end{bmatrix} \quad \text{and} \quad {}^b\mathbf{p}_b = \begin{bmatrix} {}^b x_b \\ {}^b y_b \\ f \end{bmatrix}$$

denote the coordinates, relative to each camera's canonical reference system, of the image points that are the projections of the same world point $\mathbf{P}$. Please pay attention to this definition: $^a\mathbf{p}_a$ is a point on the image plane, but is here viewed as a point in three-dimensional space. Like all points on the image plane of camera a, its third (Z) coordinate in the camera's reference system is f, the camera's focal distance. Similar considerations hold for $^b\mathbf{p}_b$. Also, since each point is observed in its own camera, the reference system (left superscript) is that of the camera the point appears in (right subscript).

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

Finally, let

$$ {}^{b}\mathbf{p} = {}^{a}R_{b}({}^{a}\mathbf{p} - {}^{a}\mathbf{t}_{b}) \tag{1} $$

be the rigid transformation between the two reference systems. As we know, the reverse transformation is

$$ {}^{a}\mathbf{p} = {}^{b}R_{a}({}^{b}\mathbf{p} - {}^{b}\mathbf{t}_{a}) \quad \text{where} \quad {}^{b}R_{a} = {}^{a}R_{b}^{T} \quad \text{and} \quad {}^{b}\mathbf{t}_{a} = -{}^{a}R_{b}{}^{a}\mathbf{t}_{b} \ . \tag{2} $$

The Essential Matrix. When expressed in the reference system of camera a, the directions of the projection rays through corresponding image points $\mathbf{p}_a$ and $\mathbf{p}_b$ are along the vectors

$$ {}^{a}\mathbf{p}_{a} \quad \text{and} \quad {}^{b}R_{a}{}^{b}\mathbf{p}_{b} \ , $$

and the baseline in this reference system is along the translation vector ${}^{a}\mathbf{t}_{b}$.

To simplify the notation in the manipulations that follows, we define

$$ \mathbf{a} = {}^{a}\mathbf{p}_{a} \quad , \quad \mathbf{b} = {}^{b}\mathbf{p}_{b} \quad , \quad R = {}^{a}R_{b} \quad , \quad \mathbf{t} = {}^{a}\mathbf{t}_{b} \quad , \quad \mathbf{e} = \mathbf{e}_{b} $$

to be the image measurements of the two corresponding points (each viewed as a three-dimensional point in its own camera's reference system), the parameters of the coordinate transformation from camera a to camera b, and the epipole. Then, the rotation and translation in the reverse direction are

$$ R^{T} = {}^{b}R_{a} \quad \text{and} \quad -R\mathbf{t} = {}^{b}\mathbf{t}_{a} \ . $$

Coplanarity of the projection-ray directions $\mathbf{a}$ and $R^T\mathbf{b}$ and baseline $\mathbf{t}$ can be expressed by stating that their triple product is zero:

$$ (R^{T}\mathbf{b})^{T}(\mathbf{t} \times \mathbf{a}) = 0 \quad \text{that is,} \quad \mathbf{b}^{T}R\,(\mathbf{t} \times \mathbf{a}) = 0 \quad \text{or} \quad \mathbf{b}^{T}R\,[\mathbf{t}]_{\times}\mathbf{a} = 0 $$

where $\mathbf{t} = (t_x, t_y, t_z)^T$ and

$$ [\mathbf{t}]_{\times} = \begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix} $$

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

In summary, for corresponding points $\mathbf{a}$ and $\mathbf{b}$ the following equation holds:

$$ \mathbf{b}^{T}E\,\mathbf{a} = 0 \tag{3} $$

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 $\mathbf{a}$ and $\mathbf{b}$ on the same epipolar plane for two fixed cameras.

If point $\mathbf{b}$ is fixed in image I_b, then the product

$$ \boldsymbol{\lambda}^{T} = \mathbf{b}^{T}E \tag{5} $$

is a row vector. If the fixed point $\mathbf{a}$ is replaced by a variable vector $\mathbf{x}$ in image I_a, then equation (3) can be written as follows:

$$ \boldsymbol{\lambda}^{T}\mathbf{x} = 0 \ . \tag{6} $$

This is a single linear equation in the coordinates of $\mathbf{x}$, and therefore represents a line in the image plane of I_a. The point $\mathbf{a}$ satisfies this equation by equation (3). Also the translation vector $\mathbf{t}$ satisfies equation (6), because

$$\boldsymbol{\lambda}^T \mathbf{t} = \mathbf{b}^T E \mathbf{t} = \mathbf{b}^T R [\mathbf{t}]_\times \mathbf{t} = 0$$

(recall that the cross product of a vector with itself is zero). The epipole $\mathbf{e}$ in image I_a is on the baseline, and therefore its coordinates in the reference frame of camera a are proportional to those of $\mathbf{t}$, so $\mathbf{e}$ satisfies equation (6) as well. Thus, this equation represents the line through $\mathbf{a}$ and $\mathbf{e}$, that is, the epipolar line of $\mathbf{b}$ in image I_a: If we knew the essential matrix E for a pair of cameras, then we could find the equation of the epipolar line for every point $\mathbf{b}$ in I_b.

This state of affairs must of course hold the other way around as well, when the roles of the two cameras are switched. Before seeing this in more detail, however, we explore the structure of the essential matrix E.

The Structure of E. First, this matrix cannot be full rank, as the following geometric argument proves: Since the epipole in image I_a belongs to *all* epipolar lines in I_a, not just one, the vector $\mathbf{e}$ of its coordinates must satisfy equation (6) *regardless of what point* $\mathbf{b}$ *is used* in the definition (5) of $\boldsymbol{\lambda}$. This can happen only if $\mathbf{e}$ is in the null space of E, so this matrix must be degenerate.

The degeneracy of E can also be shown algebraically. More specifically, 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

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

and the null space of $[\mathbf{t}]_\times$ is exactly 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 $\mathbf{t} \neq \mathbf{0}$. In addition, the null space of E and that of $[\mathbf{t}]_\times$ are the same, because the solutions to the two systems

$$[\mathbf{t}]_\times \mathbf{x} = 0 \quad \text{and} \quad E\,\mathbf{x} = 0$$

are the same, since R is full rank. Therefore, *the rank of E is 2 if* $\mathbf{t}$ *is nonzero, and the null space of E is the line spanned by* $\mathbf{t}$ *and* $\mathbf{e}$.

There is more to the structure of E. For any vector $\mathbf{v}$ orthogonal to $\mathbf{t}$, the definition of cross product yields

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

The vector $\mathbf{v}$ is orthogonal to $\mathbf{t}$ if it is in the row space of $[\mathbf{t}]_\times$, and the equation above then shows that the matrix $[\mathbf{t}]_\times$ maps all unit vectors ($\|\mathbf{v}\| = 1$) 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.[3] Since multiplication by an orthogonal matrix (R) 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 one zero singular value.* The right singular vector $\mathbf{v}_3$ corresponding to the zero singular value of E is a unit vector along the epipole and the translation vector,

$$\mathbf{v}_3 \sim \mathbf{e} \sim \mathbf{t} \; . \tag{7}$$

In these expressions, the symbol '$\sim$' means "proportional to," or "equal up to a multiplicative constant." Since the two nonzero singular values of E are equal to each other, the corresponding right singular vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ are arbitrary, as long as they form an orthonormal triple with $\mathbf{v}_3$.

[3] Since equation (3) is homogeneous, if E is an essential matrix then so is αE for any nonzero α. Therefore, the common magnitude of the two nonzero singular values is arbitrary.

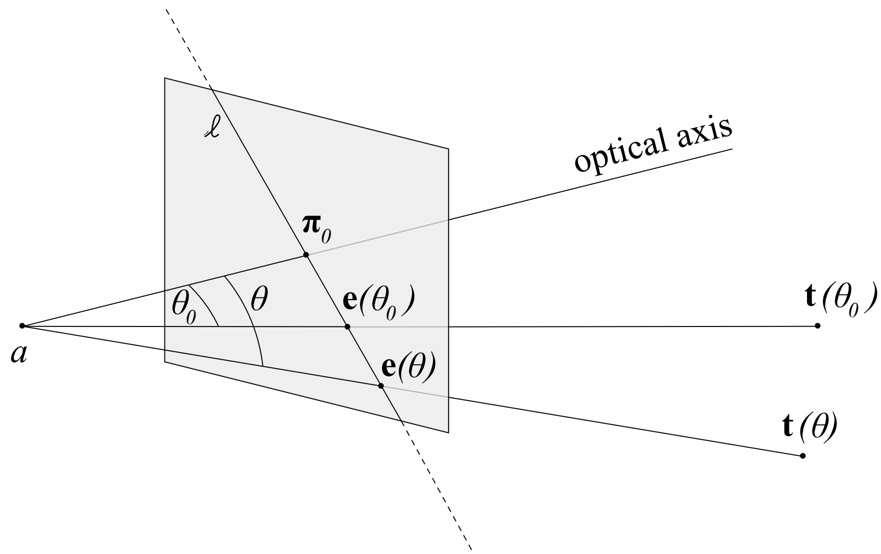

Figure 2: When the angle θ between the optical axis of camera a and the baseline approaches $\frac{\pi}{2}$, the baseline (that is, the line through a and along the translation vector $\mathbf{t}(\theta)$) becomes more and more parallel to the image plane of camera a, and the epipole $\mathbf{e}(\theta)$ tends to the point at infinity of the line ℓ through the principal point $\boldsymbol{\pi}_0$ and $\mathbf{e}(\theta_0)$.

Scale and Epipoles at Infinity. Since the systems involving the essential matrix E are all homogeneous, the translation vector $\mathbf{t}$ and the epipole $\mathbf{e}$ can only be found up to a scale factor. This limitation is consistent with the fact that cameras fundamentally measure angles between projection rays, and cannot measure lengths. For instance, if two images show a building, it is not possible to determine from image measurements alone whether the pictures are of a real building taken from two cameras, say, three meters apart, or they are images of a miniature building perhaps a hundred times smaller, taken from two cameras that are three centimeters apart. Scale is irretrievably lost in imaging, even if multiple cameras are used and as long as only the images are available—and not, say, the geometry of the camera arrangement.

While this loss of scale is generally a disadvantage of passive imaging with cameras at unknown positions, it has a positive consequence on the representation of epipoles and translation when the baseline is parallel to the image plane of either camera.

To understand this observation, consider a situation in which the angle $\theta = \theta_0$ between the optical axis of camera a and the baseline is less than 90 degrees, as illustrated in Figure 2. The orientation of camera b does not matter for this argument. Then, the baseline crosses the image plane of camera a at the epipole $\mathbf{e}$ of b in image I_a, and the translation vector from a to b is proportional to $\mathbf{e}$:

$$\mathbf{e} = \begin{bmatrix} e_{x0} \\ e_{y0} \\ 1 \end{bmatrix} \quad \text{and} \quad \mathbf{t} = c\,\mathbf{e}$$

where c is some constant.

Now gradually increase the angle θ beyond θ_0 by rotating the baseline away from the optical axis. For simplicity, think of this rotation occurring in the plane that contains the optical axis and $\mathbf{e}(\theta_0)$, so that the epipole $\mathbf{e}(\theta)$ moves along the line ℓ between the principal point $\boldsymbol{\pi}_0$ of a and $\mathbf{e}(\theta_0)$.

Since the epipole is always in the image, its third coordinate is 1, and we have

$$\mathbf{e}(\theta) = \begin{bmatrix} e_x(\theta) \\ e_y(\theta) \\ 1 \end{bmatrix} = \begin{bmatrix} h(\theta)e_{x0} \\ h(\theta)e_{y0} \\ 1 \end{bmatrix}$$

where $h(\theta)$ is an increasing function of θ. When θ tends to $\pi/2$, the baseline becomes parallel to the image plane of camera a. The scalar $h(\theta)$ tends to infinity, and the epipole moves infinitely far away from $\boldsymbol{\pi}_0$.

However, since the third left singular vector $\mathbf{v}_3(\theta)$ of the essential matrix has unit norm, it represents the epipole $\mathbf{e}(\theta)$—and the translation $\mathbf{t}(\theta)$—only up to a constant. More specifically,

$$\mathbf{v}_3(\theta) = \frac{\mathbf{e}(\theta)}{\|\mathbf{e}(\theta)\|} = \frac{1}{\sqrt{1 + h^2(\theta)\left(e_{x0}^2 + e_{y0}^2\right)}} \begin{bmatrix} h(\theta)e_{x0} \\ h(\theta)e_{y0} \\ 1 \end{bmatrix}$$

and we immediately see that

$$\lim_{\theta \to \pi/2} \mathbf{v}_3(\theta) = \frac{1}{\sqrt{e_{x0}^2 + e_{y0}^2}} \begin{bmatrix} e_{x0} \\ e_{y0} \\ 0 \end{bmatrix} ,$$

a unit-norm vector as expected.

Thus, a singular vector $\mathbf{v}_3$ that has a third component equal to zero can be viewed as pointing to an epipole $\mathbf{e}$ that is the point at infinity on the line ℓ. Since $\mathbf{t}$ is proportional to $\mathbf{v}_3$ as well, we see that $\mathbf{t}(\frac{\pi}{2})$ is also parallel to the image plane, consistently with the fact that for $\theta = \frac{\pi}{2}$ camera b is to the side of camera a, that is, in the plane $z = 0$ in the reference system of camera a.

In summary, the solution $\mathbf{e}$ or $\mathbf{t}$ provided by $\mathbf{v}_3$ is correct even when the baseline is parallel to the image plane, as long as the epipole $\mathbf{e}$ is interpreted as a point at infinity on the image plane of camera a.

Switching Cameras. Suppose now that we fix $\mathbf{a}$ in image I_a but replace $\mathbf{b}$ by a varying vector in I_b. Then we can repeat all the considerations above for the left null space and the left row space of E. In particular, the product $E\mathbf{a}$ for fixed $\mathbf{a}$ is a column vector, and equation (3) becomes the equation of the epipolar line in image I_b. The third *left* singular vector $\mathbf{u}_3$ of E is the direction of the epipole $\mathbf{e}_a$ in I_b *in the reference frame of camera* b. Rather than showing this through a separate argument, we prove that E^T is the essential matrix that would be obtained if the roles of cameras a and b were reversed.

To this end, Table 1 shows the results both ways using full subscripts, to make sure we do not confuse the two reference systems. To justify these results in the reverse direction, we then need to show that

$${}^aE_b^T = {}^bE_a ,$$

that is, that transposing one essential matrix yields the essential matrix in the opposite direction. This result is a straightforward consequence of the invariance of the cross product to rotation,

$$(R\mathbf{x}) \times (R\mathbf{y}) = R\,(\mathbf{x} \times \mathbf{y})$$

which can be restated as follows for cross-product matrices:

$$[R\mathbf{x}]_\times R = R\,[\mathbf{x}]_\times \ . \tag{8}$$

Because $[{}^a\mathbf{t}_b]_\times$ is skew-symmetric,

$$ {}^aE_b^T = ({}^aR_b\ [{}^a\mathbf{t}_b]_\times)^T = -[{}^a\mathbf{t}_b]_\times\ {}^aR_b^T\ . $$

From our discussion of rigid transformations, we also know that if

$$ {}^b\mathbf{p} = {}^aR_b({}^a\mathbf{p} - {}^a\mathbf{t}_b) $$

then

$$ {}^a\mathbf{p} = {}^bR_a({}^b\mathbf{p} - {}^b\mathbf{t}_a) \quad \text{where} \quad {}^aR_b = {}^bR_a^T \quad \text{and} \quad {}^a\mathbf{t}_b = -{}^bR_a\,{}^b\mathbf{t}_a\ . $$

Therefore,

$$ {}^aE_b^T = [{}^bR_a\,{}^b\mathbf{t}_a]_\times\ {}^bR_a $$

and from equation (8)

$$ {}^aE_b^T = {}^bR_a\ [{}^b\mathbf{t}_a]_\times = {}^bE_a $$

as promised.

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

For two cameras a and b with nonzero baseline, let

$$ {}^b\mathbf{p} = {}^aR_b\,({}^a\mathbf{p} - {}^a\mathbf{t}_b) $$

be the coordinate transformation between points ${}^a\mathbf{p}$ in a and points ${}^b\mathbf{p}$ in b, and let

$$ {}^a\mathbf{p} = {}^bR_a\,({}^b\mathbf{p} - {}^b\mathbf{t}_a) \quad \text{with} \quad {}^aR_b = {}^bR_a^T \quad \text{and} \quad {}^a\mathbf{t}_b = -{}^bR_a\,{}^b\mathbf{t}_a $$

be the transformation in the reverse direction.
The *essential matrix* of the camera pair (a, b) is the matrix

$$ {}^aE_b = {}^aR_b\,[{}^a\mathbf{t}_b]_\times \quad \text{where} \quad [\mathbf{t}]_\times = \begin{bmatrix} 0 & -t_3 & t_2 \\ t_3 & 0 & -t_1 \\ -t_2 & t_1 & 0 \end{bmatrix} $$

and the essential matrix of the camera pair (b, a) is

$$ {}^bE_a = {}^aE_b^T \ . $$

The *epipole* ${}^a\mathbf{e}_b$ is the image of the center of projection of camera b in image I_a and the epipole ${}^b\mathbf{e}_a$ is the image of the center of projection of camera a in image I_b. They satisfy

$$ {}^aE_b\,{}^a\mathbf{e}_b = {}^bE_a\,{}^b\mathbf{e}_a = \mathbf{0} \quad \text{and also} \quad {}^aE_b\,{}^a\mathbf{t}_b = {}^bE_a\,{}^b\mathbf{t}_a = \mathbf{0} \ . $$

A point ${}^a\mathbf{p}_a$ in image I_a and its corresponding point ${}^b\mathbf{p}_b$ in image I_b, both written as 3D vectors in their camera's standard reference system, satisfy the *epipolar constraint*

$$ {}^b\mathbf{p}_b^T\,{}^aE_b\,{}^a\mathbf{p}_a = 0 \ . $$

This equation can also be written as follows:

$$ \boldsymbol{\lambda}_b^T\,{}^a\mathbf{p}_a = \boldsymbol{\lambda}_a^T\,{}^b\mathbf{p}_b = 0 $$

where

$$ \boldsymbol{\lambda}_b = {}^bE_a\,{}^b\mathbf{p}_b \quad \text{and} \quad \boldsymbol{\lambda}_a = {}^aE_b\,{}^a\mathbf{p}_a $$

are the vectors of coefficients of the *epipolar line* of $\mathbf{p}_b$ in image I_a and that of $\mathbf{p}_a$ in image I_b respectively.
Up to a nonzero and otherwise arbitrary multiplicative constant, the singular value decomposition of ${}^aE_b$ is

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

where

$$ \mathbf{v}_3 \sim {}^a\mathbf{e}_b \sim {}^a\mathbf{t}_b \quad \text{and} \quad \mathbf{u}_3 \sim {}^b\mathbf{e}_a \sim {}^b\mathbf{t}_a $$

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

Table 1: Definition and properties of the essential matrix.