WIKIPEDIA
The Free Encyclopedia

Essential matrix

In computer vision, the **essential matrix** is a 3×3 matrix, $\mathbf{E}$ that relates corresponding points in stereo images assuming that the cameras satisfy the pinhole camera model.

Function

More specifically, if $\mathbf{y}$ and $\mathbf{y}'$ are homogeneous *normalized* image coordinates in image 1 and 2, respectively, then

$$(\mathbf{y}')^\top \, \mathbf{E} \, \mathbf{y} = 0$$

if $\mathbf{y}$ and $\mathbf{y}'$ correspond to the same 3D point in the scene.

The above relation which defines the essential matrix was published in 1981 by H. Christopher Longuet-Higgins, introducing the concept to the computer vision community. Richard Hartley and Andrew Zisserman's book reports that an analogous matrix appeared in photogrammetry long before that. Longuet-Higgins' paper includes an algorithm for estimating $\mathbf{E}$ from a set of corresponding normalized image coordinates as well as an algorithm for determining the relative position and orientation of the two cameras given that $\mathbf{E}$ is known. Finally, it shows how the 3D coordinates of the image points can be determined with the aid of the essential matrix.

Use

The essential matrix can be seen as a precursor to the *fundamental matrix*, $\mathbf{F}$. Both matrices can be used for establishing constraints between matching image points, but the fundamental matrix can only be used in relation to calibrated cameras since the inner camera parameters (matrices $\mathbf{K}$ and $\mathbf{K}'$) must be known in order to achieve the normalization. If, however, the cameras are calibrated the essential matrix can be useful for determining both the relative position and orientation between the cameras and the 3D position of corresponding image points. The essential matrix is related to the fundamental matrix with

$$\mathbf{E} = (\mathbf{K}')^\top \; \mathbf{F} \; \mathbf{K}.$$

Derivation and definition

This derivation follows the paper by Longuet-Higgins.

Two normalized cameras project the 3D world onto their respective image planes. Let the 3D coordinates of a point $\mathbf{P}$ be (x_1, x_2, x_3) and (x'_1, x'_2, x'_3) relative to each camera's coordinate system. Since the cameras are normalized, the corresponding image coordinates are

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \frac{1}{x_3} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} y'_1 \\ y'_2 \end{pmatrix} = \frac{1}{x'_3} \begin{pmatrix} x'_1 \\ x'_2 \end{pmatrix}$$

A homogeneous representation of the two image coordinates is then given by

$$\begin{pmatrix} y_1 \\ y_2 \\ 1 \end{pmatrix} = \frac{1}{x_3} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} y'_1 \\ y'_2 \\ 1 \end{pmatrix} = \frac{1}{x'_3} \begin{pmatrix} x'_1 \\ x'_2 \\ x'_3 \end{pmatrix}$$

which also can be written more compactly as

$$\mathbf{y} = \frac{1}{x_3}\,\tilde{\mathbf{x}} \quad \text{and} \quad \mathbf{y}' = \frac{1}{x'_3}\,\tilde{\mathbf{x}}'$$

where $\mathbf{y}$ and $\mathbf{y}'$ are homogeneous representations of the 2D image coordinates and $\tilde{\mathbf{x}}$ and $\tilde{\mathbf{x}}'$ are proper 3D coordinates but in two different coordinate systems.

Another consequence of the normalized cameras is that their respective coordinate systems are related by means of a translation and rotation. This implies that the two sets of 3D coordinates are related as

$$\tilde{\mathbf{x}}' = \mathbf{R}\,(\tilde{\mathbf{x}} - \mathbf{t})$$

where $\mathbf{R}$ is a 3×3 rotation matrix and $\mathbf{t}$ is a 3-dimensional translation vector.

The essential matrix is then defined as:

$$\boxed{\mathbf{E} = \mathbf{R}\,[\mathbf{t}]_{\times}}$$

where $[\mathbf{t}]_{\times}$ is the matrix representation of the cross product with $\mathbf{t}$. Note: Here, the transformation $[\mathbf{R}^T|\mathbf{t}]$ will transform points in the 2nd view to the 1st view.

For the definition of $\mathbf{E}$ we are only interested in the orientations of the normalized image coordinates [1] (See also: Triple product). As such we don't need the translational component when substituting image coordinates into the essential equation. To see that this definition of $\mathbf{E}$ describes a constraint on corresponding image coordinates multiply $\mathbf{E}$ from left and right with the 3D coordinates of point $\mathbf{P}$ in the two different coordinate systems:

$$\tilde{\mathbf{x}}'^T\,\mathbf{E}\,\tilde{\mathbf{x}} \overset{(1)}{=} \tilde{\mathbf{x}}^T\,\mathbf{R}^T\,\mathbf{R}\,[\mathbf{t}]_{\times}\,\tilde{\mathbf{x}} \overset{(2)}{=} \tilde{\mathbf{x}}^T\,[\mathbf{t}]_{\times}\,\tilde{\mathbf{x}} \overset{(3)}{=} 0$$

1. Insert the above relations between $\tilde{\mathbf{x}}'$ and $\tilde{\mathbf{x}}$ and the definition of $\mathbf{E}$ in terms of $\mathbf{R}$ and $\mathbf{t}$.
2. $\mathbf{R}^T\,\mathbf{R} = \mathbf{I}$ since $\mathbf{R}$ is a rotation matrix.
3. Properties of the matrix representation of the cross product.

Finally, it can be assumed that both x_3 and x'_3 are > 0, otherwise they are not visible in both cameras. This gives

$$0 = (\tilde{\mathbf{x}}')^T\,\mathbf{E}\,\tilde{\mathbf{x}} = \frac{1}{x'_3}(\tilde{\mathbf{x}}')^T\,\mathbf{E}\,\frac{1}{x_3}\tilde{\mathbf{x}} = (\mathbf{y}')^T\,\mathbf{E}\,\mathbf{y}$$

which is the constraint that the essential matrix defines between corresponding image points.

Properties

Not every arbitrary 3×3 matrix can be an essential matrix for some stereo cameras. To see this notice that it is defined as the matrix product of one rotation matrix and one skew-symmetric matrix, both 3×3. The skew-symmetric matrix must have two singular values which are equal and another which is zero. The multiplication of the rotation matrix does not change the singular values which means that also the essential matrix has two singular values which are equal and one which is zero. The properties described here are sometimes referred to as *internal constraints* of the essential matrix.

If the essential matrix $\mathbf{E}$ is multiplied by a non-zero scalar, the result is again an essential matrix which defines exactly the same constraint as $\mathbf{E}$ does. This means that $\mathbf{E}$ can be seen as an element of a projective space, that is, two such matrices are considered equivalent if one is a non-zero scalar multiplication of the other. This is a relevant position, for example, if $\mathbf{E}$ is estimated from image data. However, it is also possible to take the position that $\mathbf{E}$ is defined as

$$\mathbf{E} = [\tilde{\mathbf{t}}]_{\times} \, \mathbf{R}$$

where $\tilde{\mathbf{t}} = -\mathbf{R}\mathbf{t}$, and then $\mathbf{E}$ has a well-defined "scaling". It depends on the application which position is the more relevant.

The constraints can also be expressed as

$$\det \mathbf{E} = 0$$

and

$$2\mathbf{E}\mathbf{E}^T\mathbf{E} - \operatorname{tr}(\mathbf{E}\mathbf{E}^T)\mathbf{E} = 0.$$

Here, the last equation is a matrix constraint, which can be seen as 9 constraints, one for each matrix element. These constraints are often used for determining the essential matrix from five corresponding point pairs.

The essential matrix has five or six degrees of freedom, depending on whether or not it is seen as a projective element. The rotation matrix $\mathbf{R}$ and the translation vector $\mathbf{t}$ have three degrees of freedom each, in total six. If the essential matrix is considered as a projective element, however, one degree of freedom related to scalar multiplication must be subtracted leaving five degrees of freedom in total.

Estimation

Given a set of corresponding image points it is possible to estimate an essential matrix which satisfies the defining epipolar constraint for all the points in the set. However, if the image points are subject to noise, which is the common case in any practical situation, it is not possible to find an essential matrix which satisfies all constraints exactly.

Depending on how the error related to each constraint is measured, it is possible to determine or estimate an essential matrix which optimally satisfies the constraints for a given set of corresponding image points. The most straightforward approach is to set up a total least squares problem, commonly known as the eight-point algorithm.

Extracting rotation and translation

Given that the essential matrix has been determined for a stereo camera pair -- for example, using the estimation method above -- this information can be used for determining also the rotation $\mathbf{R}$ and translation $\mathbf{t}$ (up to a scaling) between the two camera's coordinate systems. In these derivations $\mathbf{E}$ is seen as a projective element rather than having a well-determined scaling.

Finding one solution

The following method for determining $\mathbf{R}$ and $\mathbf{t}$ is based on performing a SVD of $\mathbf{E}$, see Hartley & Zisserman's book.[2] It is also possible to determine $\mathbf{R}$ and $\mathbf{t}$ without an SVD, for example, following Longuet-Higgins' paper.

An SVD of $\mathbf{E}$ gives

$$\mathbf{E} = \mathbf{U}\,\mathbf{\Sigma}\,\mathbf{V}^T$$

where $\mathbf{U}$ and $\mathbf{V}$ are orthogonal 3×3 matrices and $\mathbf{\Sigma}$ is a 3×3 diagonal matrix with

$$\mathbf{\Sigma} = \begin{pmatrix} s & 0 & 0 \\ 0 & s & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

The diagonal entries of $\mathbf{\Sigma}$ are the singular values of $\mathbf{E}$ which, according to the internal constraints of the essential matrix, must consist of two identical and one zero value. Define

$$\mathbf{W} = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{with} \quad \mathbf{W}^{-1} = \mathbf{W}^T = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

and make the following ansatz

$$[\mathbf{t}]_\times = \mathbf{U}\,\mathbf{W}\,\mathbf{\Sigma}\,\mathbf{U}^T$$

$$\mathbf{R} = \mathbf{U}\,\mathbf{W}^{-1}\,\mathbf{V}^T$$

Since $\mathbf{\Sigma}$ may not completely fulfill the constraints when dealing with real world data (f.e. camera images), the alternative

$$[\mathbf{t}]_\times = \mathbf{U}\,\mathbf{Z}\,\mathbf{U}^T \quad \text{with} \quad \mathbf{Z} = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

may help.

Proof

First, these expressions for $\mathbf{R}$ and $[\mathbf{t}]_\times$ do satisfy the defining equation for the essential matrix

$$[\mathbf{t}]_\times \, \mathbf{R} = \mathbf{U}\,\mathbf{W}\,\mathbf{\Sigma}\,\mathbf{U}^T \mathbf{U}\,\mathbf{W}^{-1}\,\mathbf{V}^T \ = \mathbf{U}\,\mathbf{\Sigma}\,\mathbf{V}^T = \mathbf{E}$$

Second, it must be shown that this $[\mathbf{t}]_\times$ is a matrix representation of the cross product for some $\mathbf{t}$. Since

$$\mathbf{W}\,\mathbf{\Sigma} = \begin{pmatrix} 0 & -s & 0 \\ s & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

it is the case that $\mathbf{W}\,\mathbf{\Sigma}$ is skew-symmetric, i.e., $(\mathbf{W}\,\mathbf{\Sigma})^T = -\mathbf{W}\,\mathbf{\Sigma}$. This is also the case for our $[\mathbf{t}]_\times$, since

$$([\mathbf{t}]_\times)^T = \mathbf{U}\,(\mathbf{W}\,\mathbf{\Sigma})^T\,\mathbf{U}^T = -\mathbf{U}\,\mathbf{W}\,\mathbf{\Sigma}\,\mathbf{U}^T = -[\mathbf{t}]_\times$$

According to the general properties of the matrix representation of the cross product it then follows that $[\mathbf{t}]_\times$ must be the cross product operator of exactly one vector $\mathbf{t}$.

Third, it must also need to be shown that the above expression for $\mathbf{R}$ is a rotation matrix. It is the product of three matrices which all are orthogonal which means that $\mathbf{R}$, too, is orthogonal or $\det(\mathbf{R}) = \pm 1$. To be a proper rotation matrix it must also satisfy $\det(\mathbf{R}) = 1$. Since, in this case, $\mathbf{E}$ is seen as a projective element this can be accomplished by reversing the sign of $\mathbf{E}$ if necessary.

Finding all solutions

So far one possible solution for $\mathbf{R}$ and $\mathbf{t}$ has been established given $\mathbf{E}$. It is, however, not the only possible solution and it may not even be a valid solution from a practical point of view. To begin with, since the scaling of $\mathbf{E}$ is undefined, the scaling of $\mathbf{t}$ is also undefined. It must lie in the null space of $\mathbf{E}$ since

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

For the subsequent analysis of the solutions, however, the exact scaling of $\mathbf{t}$ is not so important as its "sign", i.e., in which direction it points. Let $\hat{\mathbf{t}}$ be normalized vector in the null space of $\mathbf{E}$. It is then the case that both $\hat{\mathbf{t}}$ and $-\hat{\mathbf{t}}$ are valid translation vectors relative $\mathbf{E}$. It is also possible to change $\mathbf{W}$ into $\mathbf{W}^{-1}$ in the derivations of $\mathbf{R}$ and $\mathbf{t}$ above. For the translation vector this only causes a change of sign, which has already been described as a possibility. For the rotation, on the other hand, this will produce a different transformation, at least in the general case.

To summarize, given $\mathbf{E}$ there are two opposite directions which are possible for $\mathbf{t}$ and two different rotations which are compatible with this essential matrix. In total this gives four classes of solutions for the rotation and translation between the two camera coordinate systems. On top of that, there is also an unknown scaling $s > 0$ for the chosen translation direction.

It turns out, however, that only one of the four classes of solutions can be realized in practice. Given a pair of corresponding image coordinates, three of the solutions will always produce a 3D point which lies *behind* at least one of the two cameras and therefore cannot be seen. Only one of the four classes will consistently produce 3D points which are in front of both cameras. This must then be the correct solution. Still, however, it has an undetermined positive scaling related to the translation component.

The above determination of $\mathbf{R}$ and $\mathbf{t}$ assumes that $\mathbf{E}$ satisfy the internal constraints of the essential matrix. If this is not the case which, for example, typically is the case if $\mathbf{E}$ has been estimated from real (and noisy) image data, it has to be assumed that it approximately satisfy the internal constraints. The vector $\hat{\mathbf{t}}$ is then chosen as right singular vector of $\mathbf{E}$ corresponding to the smallest singular value.

3D points from corresponding image points

Many methods exist for computing (x_1, x_2, x_3) given corresponding normalized image coordinates (y_1, y_2) and (y'_1, y'_2), if the essential matrix is known and the corresponding rotation and translation transformations have been determined.

See also

- Bundle adjustment
- Epipolar geometry
- Fundamental matrix
- Geometric camera calibration
- Triangulation (computer vision)
- Trifocal tensor

Toolboxes

- Essential Matrix Estimation (https://www.mathworks.com/matlabcentral/fileexchange/67580-essential-matrix-estimation) in MATLAB (Manolis Lourakis).

External links

- An Investigation of the Essential Matrix (http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.64.7518) by R.I. Hartley

References

1. *Photogrammetric Computer Vision: Statistics, Geometry, Orientation and Reconstruction* (1st ed.).
2. Hartley, Richard; Andrew Zisserman (2004). *Multiple view geometry in computer vision* (2nd ed.). Cambridge, UK. ISBN 978-0-511-18711-7. OCLC 171123855 (https://www.worldcat.org/oclc/171123855).

- David Nistér (June 2004). "An efficient solution to the five-point relative pose problem". *IEEE Transactions on Pattern Analysis and Machine Intelligence*. **26** (6): 756–777. doi:10.1109/TPAMI.2004.17 (https://doi.org/10.1109%2FTPAMI.2004.17). PMID 18579936 (https://pubmed.ncbi.nlm.nih.gov/18579936). S2CID 886598 (https://api.semanticscholar.org/CorpusID:886598).
- H. Stewénius and C. Engels and D. Nistér (June 2006). "Recent Developments on Direct Relative Orientation". *ISPRS Journal of Photogrammetry and Remote Sensing*. **60** (4): 284–294. Bibcode:2006JPRS...60..284S (https://ui.adsabs.harvard.edu/abs/2006JPRS...60..284S).

CiteSeerX 10.1.1.61.9329 (https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.61.9329). doi:10.1016/j.isprsjprs.2006.03.005 (https://doi.org/10.1016%2Fj.isprsjprs.2006.03.005).

- H. Christopher Longuet-Higgins (September 1981). "A computer algorithm for reconstructing a scene from two projections". *Nature*. **293** (5828): 133–135. Bibcode:1981Natur.293..133L (https://ui.adsabs.harvard.edu/abs/1981Natur.293..133L). doi:10.1038/293133a0 (https://doi.org/10.1038%2F293133a0). S2CID 4327732 (https://api.semanticscholar.org/CorpusID:4327732).
- Richard Hartley and Andrew Zisserman (2003). *Multiple View Geometry in computer vision*. Cambridge University Press. ISBN 978-0-521-54051-3.
- Yi Ma; Stefano Soatto; Jana Košecká; S. Shankar Sastry (2004). *An Invitation to 3-D Vision*. Springer. ISBN 978-0-387-00893-6.
- Gang Xu and Zhengyou Zhang (1996). *Epipolar geometry in Stereo, Motion and Object Recognition*. Kluwer Academic Publishers. ISBN 978-0-7923-4199-4.
- Förstner, Wolfgang and Wrobel, Bernhard P. (2016). *Photogrammetric Computer Vision: Statistics, Geometry, Orientation and Reconstruction* (1st ed.). Springer Publishing Company, Incorporated. ISBN 978-3319115498.

Retrieved from "https://en.wikipedia.org/w/index.php?title=Essential_matrix&oldid=1135334073"