# Does the Fundamental Matrix Define a One-to-One Relation between the Corresponding Image Points of a Scene?

Tayeb Basta
College of Engineering and Computing, Al-Ghurair University,
Academic city, Dubai, United Arab Emirates
Email: tayebasta@gmail.com

***Abstract*—In computer stereo vision, the fundamental matrix is the algebraic representation of the epipolar geometry that relates two images of a scene observed from two different viewpoints. The most important feature of the fundamental matrix is its independence of the scene structure. Different methods have been proposed to derive the fundamental matrix equation. This paper reviews one of these methods and reveals that it is based on flawed statements to conclude the existence of a homography between the points on the two images. This derivation of the fundamental matrix equation is based on the existence of a homography between the two images.**

***Index Terms*—computer vision, stereo vision, fundamental matrix, homography**

## I. Introduction

The main objective of stereo vision is to recover a 3D structure of a rigid scene which has been imaged from two different positions. For the purpose of solving such a problem, researchers have succeeded in defining the *epipolar geometry* that relates the points on two views of a rigid scene. In this context, the cameras that capture the views are characterized by intrinsic and extrinsic parameters. The intrinsic parameters include coordinates of the principal points, pixel aspect ratio, and focal lengths. The extrinsic parameters are the position and orientation of the camera with respect to the world coordinate system. The cameras are indicated by their centres $C_l$ and $C_r$, and their image planes $\pi_l$ and $\pi_r$ (Fig. 1). To each camera is associated a reference system. The motion between the two positions of the cameras is given by a translation vector $\boldsymbol{t}$ from $C_l$ to $C_r$ followed by a rotation matrix $R$.

In the classical route, the intrinsic camera parameters are known. Such knowledge is used to calculate the epipolar geometry by extracting the *essential matrix* $E$ [1]. When neither intrinsic nor extrinsic camera parameters are available, the problem is classified as *uncalibrated* and the *fundamental matrix* $F$ is the algebraic representation of the epipolar geometry.

Manuscript received February 6, 2013; revised April 16, 2013.

The two camera centres and a 3D point on the scene define an epipolar plane $\Pi$ (see Fig. 1).

For a world point on the scene $M=(X,Y,Z)$, the pair of points $(m_l, m_r)$ which are the left retinal image and the right retinal image of the point $M$, respectively, are called *corresponding* points.

The fundamental matrix encapsulates the parameters relating a world point to its images. The relation between a pair of corresponding points $(m_l, m_r)$ through the fundamental matrix is given by $m_r^T F m_l = 0$. Such a matrix can be calculated by providing eight or more pairs of corresponding points from the two views of the scene [2].

The eight-point algorithm is a frequently cited method for computing the fundamental matrix from a set of eight or more point matches [3].

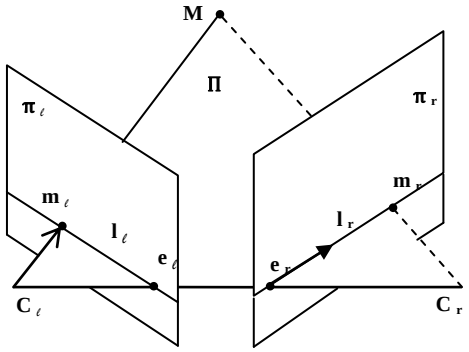


Figure 1. Epipolar geometry

Different methods have been proposed to derive the fundamental matrix [4], [2], [5], [6] and [7]. This paper reviews one of these methods. Such method asserts that the image points on the two views are projectively equivalent and concludes the existence of 2D homography mapping the corresponding points on the two views. The existence of a homography mapping the points $m_l$ and $m_r$ leads to the equation of the fundamental matrix, $m_r^T F m_l = 0$ [2].

The remainder of the paper is organized as follows. In section 2 we remind the derivation of the essential matrix. Section 3 introduces the derivation of $F$ that assumes the


doi: 10.12720/joig.1.3.125-128

existence of a homograph between the corresponding points on the two images. Section 4 reveals the flaws in the current derivation method. Finally, the paper concludes in section 5.

## II. Derivation of the Essential Matrix

Longuet-Higgins [1] introduced the essential matrix to the computer vision community and proposed the eight-point algorithm for its calculation. He defined the image coordinates $m_l$ and $m_r$ of the world point $M$ in the two cameras' coordinates systems as

$$\begin{cases}(x_l, y_l) = (X_l/Z_l, Y_l/Z_l) \\ (x_r, y_r) = (X_r/Z_r, Y_r/Z_r)\end{cases} \quad (1)$$

Given the translation vector of the right camera with respect to the left one $\boldsymbol{t} = [t_x \quad t_y \quad t_z]$ and the rotation matrix of the right camera coordinate system with respect to the left coordinate system $R$, the relation between the three-dimensional vectors representing the world point $M$ may be written as

$$\boldsymbol{M}_r = R(\boldsymbol{M}_l - \boldsymbol{t}) \quad (2)$$

The rotation $R$ satisfies the relation

$$RR^T = R^T R = 1 \text{ and } det(R) = 1 \quad (3)$$

The author [1] defines the essential matrix as

$$E = RS \quad (4)$$

where $S$ is the skew-symmetric matrix

$$S = \begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix} \quad (5)$$

He adopted the length of the vector $\boldsymbol{t}$ as the unit of distance

$$t^2 = t_x^2 + t_y^2 + t_z^2 = 1 \quad (6)$$

The author [1] then constructs the expression $\boldsymbol{M}_r^T E \boldsymbol{M}_l$ and used (2) to (6) to conclude the equation $\boldsymbol{M}_r^T E \boldsymbol{M}_l = 0$. He then divided by $Z_l Z_r$ to establish the equation of the essential matrix that relates the image points $m_l$ and $m_r$

$$m_r^T E m_l = 0 \quad (7)$$

While the essential matrix cannot be a transformation matrix, it is the product of a rotation matrix and a rank deficient matrix; $E = RS$ [1].

$$\boldsymbol{M}_r^T E \boldsymbol{M}_l = \boldsymbol{M}_r^T (RS) \boldsymbol{M}_l \quad (8)$$

The right hand side of (8) is the associative vector-matrix product of $\boldsymbol{M}_r^T$, $RS$, and $\boldsymbol{M}_l$. It is equivalent to $\left(\boldsymbol{M}_r^T R\right) \cdot S \cdot \boldsymbol{M}_l$.

By definition, the transpose of the rotation matrix is equal to its inverse $RR^T = I$. This means that if $R$ is the rotation of a left coordinate system to a right coordinate system, $R^T$ is the back rotation of the right coordinate system to the left one. Therefore, the term $\boldsymbol{M}_r^T R = \left(R^T \boldsymbol{M}_r\right)^T$ is the vector $\boldsymbol{M}_r$ expressed in the left coordinate system. Thus, both vectors $(\boldsymbol{M}_r^T R)$ and $\boldsymbol{M}_l$, involved in (8), are defined in the left coordinate system.

In addition, the matrix multiplication is associative,

$$\left(\boldsymbol{M}_r^T R\right) \cdot S \cdot \boldsymbol{M}_l = \boldsymbol{M}_r^T \cdot (R \cdot S) \cdot \boldsymbol{M}_l$$

Thus, the expression $\boldsymbol{M}_r^T E \boldsymbol{M}_l = 0$ is well defined and so are the equations of the essential and fundamental matrices $m_r^T E m_l = 0$ and $m_r^T F m_l = 0$.

## III. Introducing the New Derivation Method

The following terminology is used to describe the epipolar geometry [2]:

- The *epipole* is the point of intersection of the line joining the camera centers with the image plane. Equivalently, the epipole is the image in one view of the camera centre of the other view.
- An *epipolar line* is the intersection of an epipolar plane with the image plane. All epipolar lines intersect at the epipole. An epipolar plane intersects the left and right image planes in epipolar lines, and defines the correspondence between the lines.
- The *baseline* is the line joining the camera centers.
- The two camera centers and a given world point define the *epipolar plane* Π. All epipolar planes contain the baseline (Fig. 1).

Hartley and Zisserman [2] presented a geometric derivation of the fundamental matrix based on the existence of a homography between the two views: "The mapping from a point in one image to a corresponding epipolar line in the other image may be decomposed into two steps. In the first step, the point $m_l$ is mapped to some point $m_r$ in the other image lying on the epipolar line $l_r$. This point $m_r$ is a potential match for the point $m_l$.

In the second step, the epipolar line $l_r$ is obtained as the line joining $m_r$ to the epipole $e_r$.

Step 1: Point transfer via a plane. Consider a plane $\Pi$ in space not passing through either of the two camera centers. The ray through the first camera centre corresponding to the point $m_l$ meets the plane $\Pi$ in a point $M$. This point $M$ is then projected to a point $m_r$ in the second image. This procedure is known as transfer via the plane $\Pi$. Since $M$ lies on the ray corresponding to $m_l$, the projected point $m_r$ must lie on the epipolar line $l_r$ corresponding to the image of this ray.

The points $m_l$ and $m_r$ are both images of the 3D point $M$ lying on a plane. The set of all such points $m_l$ in the first image and the corresponding point $m_r$ in the second

image are projectively equivalent, since they are each projectively equivalent to the planar point set $M$ . Thus there is a 2D homography $H_\pi$ mapping each $m_l$ to $m_r$ .

Step 2: Constructing the epipolar line. Given the point $m_r$ the epipolar line $l_r$ passing through $m_r$ and the epipole $e_r$ can be written as $l_r = e_r \times m_r = [e_r]_\times m_r$ . Since $m_r$ may be written as $m_r = H_\pi m_l$ , we have $l_r = [e_r]_\times H_\pi m_l = F m_l$ where we define $F = [e_r]_\times H_\pi$ the fundamental matrix."

## IV. Flaws In the Current Derivation Method

In the current derivation method, Hartley and Zisserman [2] assert the existence of a homography mapping every pair of corresponding points on the two views. This assertion is the result of the statement: "The set of all such points $m_l$ in the first image and the corresponding point $m_r$ in the second image are projectively equivalent, since they are each projectively equivalent to the planar point set $M$ ."

The following points demonstrate the flaws of the authors [2] assertions that led to the current derivation of the fundamental matrix.

The first observation comes from the definition of a homography. It is a relation between two figures, such that to any point of the one corresponds one and only one point in the other, and vice versa [8].

An outstanding natural feature of a 3D scene is some of its parts hide some other parts. A 3D scene generally contains salient features. Consequently, some 3D points are exposed to one camera and are not seen by the other. Imaging 3D scenes cannot escape from occlusion. "Occluded regions are spatially coherent groups of pixels that can be seen in one image of a stereo pair but not in the other" [9]. The rounded rectangle of Fig. 2 shows points in the first image without correspondent points in the same region of the second image.

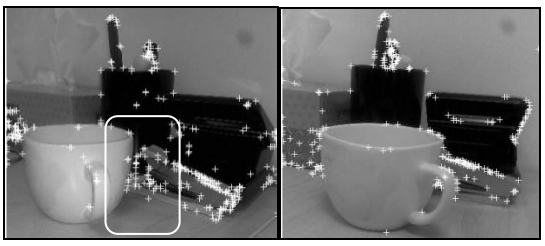

Figure 2. Points in one view without correspondent in the other

"The homography matrix is a corresponding matrix between two images, based on which the one-to-one relationship between the feature points of two images can be identified" [10]. This definition does not allow the existence of a homography between a selected set of points from the first image and a selected set of points from the second image while ignoring points that have no correspondents.

The authors [2] clearly deduce the existence of a homography that is mapping the projective points in the two views: "Thus there is a 2D homography $H_\pi$ mapping each $m_l$ to $m_r$ ." This one-to-one mapping is only valid when no-occlusion occurs which is a severe assumption that excludes all 3D scenes but concave ones, i.e. all points of the scene are exposed to the two cameras. The no-occlusion fact can be empirically (not theoretically) assured whenever the baseline is very small with respect to the distance between the camera and the scene. Fig. 3, Fig. 4, Fig. 5, and Fig. 6 from [11] and [12] are examples of such a case. They are carefully selected with small baseline with respect to the distance between the camera and the scene, i.e., the scene depth. These cases were used to assess the performance of estimation methods of the matrix $F$ .

Figure 3. View1 and view2 of scene1

Figure 4. View1 and view2 of scene2

Figure 5. View1 and view2 of scene4

Figure 6. View1 and view2 of scene5

A mapping $X \to Y$ is called injective if distinct elements of $X$ have distinct images in $Y$ . It is called surjective of every element of $Y$ is the image at least of one element in $X$ . A mapping which is simultaneously injective and surjective is called bijective [13].

Suppose that the camera is at the origin $(0,0,0)$. The ray from the origin represented by homogeneous coordinates $[x, y, z]$ is that passing through the 3D point $(x, y, z)$. The 3D point $\lambda \cdot (x, y, z) = (\lambda x, \lambda y, \lambda z)$, where $\lambda \neq 0$ also lies on (represents) the same ray [14]. So, all 3D points which belong to the same ray passing through the point $M = (X, Y, Z)$ and the origin $(0,0,0)$ are projected on the same projective point $(X/Z, Y/Z, 1)$. That class of 3D points which is projected onto one single point in one view will be projected onto more than one point in the other view. This means that the relation between the points in the two views is not injective and by definition it is not bijective. By consequence there is no homograph between the points of the two views.

The last point is the authors' assertion: "The set of all such points $m_l$ in the first image and the corresponding point $m_r$ in the second image are protectively equivalent."

Capturing a 3D scene from two viewpoints means exposing a set of world points $P_1$ of the scene to the left camera and exposing another set of world points $P_2$ to the right camera. In general, the two sets of points $P_1$ and $P_2$ are different from each other.

The set of image points $m_l$ are projectively equivalent to the set of world points $P_1$ that are visible to the left camera. And the set of image points $m_r$ are projectively equivalent to the set of world points $P_2$ that are visible to the right camera. And because the two sets $P_1$ and $P_2$ are not necessarily the same, the two sets of points $m_l$ and $m_r$ are not projectively equivalent. Therefore, there is no projectively equivalence between the image points on the two views and consequently there is no homograph that maps the points $m_l$ and $m_r$.

## V. Conclusion

The current derivation of the fundamental matrix is based on asserting misguided propositions:

- Stating that the world points of a 3D scene are planar.
- The image points on the two views as projectively equivalent.

Literally, the world points of 3D scenes are not planar unless the scene under analysis is a plane.

The set of image points on the left view are projectively equivalent to the world points exposed to the left camera. And the set of image points on the right view are projectively equivalent to the world points exposed to the right camera. The world points of a 3D scene exposed to one camera are not necessarily exposed to the other camera, consequently the two sets of image points on the two views are not projectively equivalent.

The projective equivalence between the points on the left view and the points on the right view is the condition of the existence of 2D homograph mapping the left view to the right one.

The estimation of the fundamental matrix was a flourishing theme, especially, in the nineties. The current work proves that, in general, there is no 2D homograph that maps the points on one image plane onto the points on the other image plane. Thus, the current derivation of the fundamental matrix is flawed.

## References


[1] H. C. Longuet-Higgins, "A computer algorithm for reconstructing a scene from two projections," *Nature*, vol. 293, pp. 133-135, 1981.

[2] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge University Press, 2004, ch. 9.

[3] R. Hartley, "In defense of the eight-point algorithm," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 6, pp. 580-593, 1997.

[4] O. D. Faugeras, "What can be seen in three dimensions with an uncelebrated stereo rig?" in *Proc. 2nd European Conference on Computer Vision*, G. Sandini Ed. Santa Margherita Ligure, Italy, Springer-Verlag, 1992, pp. 563-578.

[5] S. Ivekovic, A. Fusiello, and E. Trucco, "Fundamentals of multiple view geometry," in *3D Videocommunication: Algorithms, Concepts and Real-Time Systems in Human Centred Communication*, O. Schreer; P. Kauff, and T. Sikora Ed., John Wiley & Sons, 2005.

[6] Q. T. Luong and O. D. Faugeras, "Determining the fundamental matrix with planes: Instability and new algorithms," in *IEEE Press, CVPR*, 1993, pp. 489-494.

[7] Z. Y. Zhang, "Determining the epipolar geometry and its uncertainty - A review," *International Journal of Computer Vision*, vol. 27, no. 2, pp. 161-195, 1998.

[8] *Webster's Revised Unabridged Dictionary*, 1913.

[9] A. F. Bobick and S. S. Intille, "Large occlusion stereo," *International Journal of Computer Vision*, vol. 33, no. 3, pp. 181-200, 1999.

[10] P. K. Jain and C. V. Jawahar, "Homography estimation from planar contours, 3D data processing, visualization, and transmission," in *Third International Symposium, IEEE Computer Society*, June 2006, pp. 877-884.

[11] J. Salvi. (Feb. 2003). Fundamental Matrix Estimation Toolbox. [Online]. Available: http://eia.udg.es/~qsalvi/recerca.html

[12] Y. Wexler, A. W. Fitzgibbon, and A. Zisserman, "Learning epipolar geometry from image sequences," in *CVPR*, vol. 2, pp. 209-216, 2003.

[13] L. Mirsky, "An account of some aspects of combinatorial mathematics," *Transversal Theory*, Academic Press, 1971.

[14] M. Triggs. (Nov. 13, 1998). Homogeneous Coordinates. [Online]. Available:http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/MOHR_TRIGGS/node7.html



**Tayeb Basta** was born in Algeria, 1960. He received in 1983 a BSc degree in computer science from University of Annaba, Algeria. His PhD is in computer science from University of Manchester, UK in 1994. Tayeb is now associate professor at Al Ghurair University in Dubai, UAE.

Dr Basta is a senior member of IACSIT and SIE.