Examining the Validity of the Fundamental Matrix Equation

Tayeb Basta

***Abstract*—In stereo vision, two cameras are used to obtain two views of a scene from two different standpoints. The epipolar geometry describes the relation between the two views. When the intrinsic cameras parameters are known, the essential matrix is the algebraic representation of this geometry; otherwise the fundamental matrix is the representation of such geometry. A number of derivation methods of the essential and fundamental matrices are available in the computer vision literature.**

This paper questions the validity of the equations of these matrices and demonstrates that such equations are the result of an undefined vector operation.

***Index Terms*—Essential Matrix, fundamental matrix, stereo vision, epipolar geometry, dot product.**

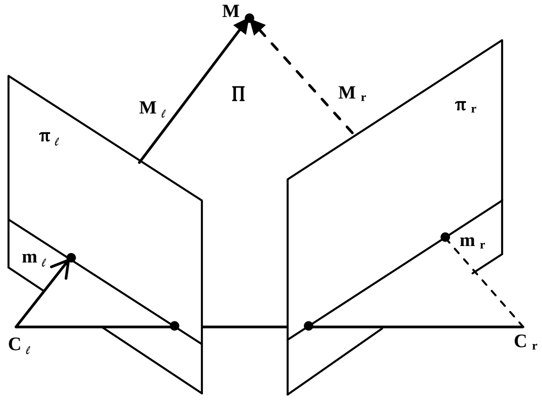

Fig. 1. The epipolar geometry.

I. Introduction

In stereo vision, two cameras are used to obtain two differing views (i.e., images) of a scene from two different standpoints. The cameras are supposed to satisfy the pinhole model assumption. In this context, a camera is described by intrinsic and extrinsic parameters. The former include coordinates of the principal points, pixel aspect ratio, and focal lengths. The latter are the position and orientation of the camera with respect to the world coordinate system. The *epipolar geometry* describes the relation between the two views. When the intrinsic cameras parameters are known, the *essential matrix* is the algebraic representation of this geometry. When none of the parameters are known, the *fundamental matrix* encapsulates all the information about the epipolar geometry. This geometry which is depicted in Figure 1 can be described as a world point $M=(X,Y,Z)$ defined in a world coordinate system and two cameras placed at two different positions C_l and C_r to capture the point. A coordinate system is defined for each camera. The points C_l and C_r are the origin of the theses two coordinate systems.

The fundamental matrix developed by Faugeras [4] is an improvement of the essential matrix which has been introduced by Longuet-Higgins [8] to compute the structure of a scene from two views. A number of derivation methods have been proposed to derive both the essential and the fundamental matrices.

For computing the fundamental matrix from a set of eight or more point matches [5]. Because there is no procedure that provide an accurate set of eight or more point matches, researchers focus on developing estimate methods of the fundamental matrix, rather than reviewing the epipolar geometry theory. In the last few years, several methods to estimate the fundamental matrix have been proposed, which can be classified into linear, iterative and robust methods. Linear and iterative methods can cope with bad point localization in the image plane due to noise in image segmentation. Robust methods can cope with both image noise and outliers, i.e. wrong matching between point correspondences in both image planes. All of these methods are based on solving a homogeneous system of equations which can be deduced from the fundamental matrix equation [2], [10].

This paper goes in a different vein to question the validity of the equations of the essential and fundamental matrices.

The rest of the paper is organized as follows: Section 2 presents Longuet-Higgins' derivation method of the essential matrix. Section 3 addresses the validity of the equations of the essential and fundamental matrices. This work concludes in section 4.

II. Longuet-Higgins' Derivation Method

Longuet-Higgins [8] defined the image coordinates m_l and m_r of the world point M in the two cameras' coordinate systems as

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

Manuscript received May 12, 2012; revised June 12, 2012

T. Basta is with the computing college of Al Ghurair University,PO Box 37374, Dubai, United Arab Emirates (e-mail：tayebasta@gmail.com; tel. (0)504446420)

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

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

The rotation R satisfies the relation

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

The author [8] defines the essential matrix as

$$E = RS \tag{4}$$

where S is the skew-symmetric matrix

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

He adopted the length of the vector t as the unit of distance

$$t^2 = t_x^2 + t_y^2 + t_z^2 = 1 \tag{6}$$

The author [8] 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

$$m_r^T E m_l = 0 \tag{7}$$

Based on the theory of the essential matrix established by Longuet-Higgins [8], other methods have been proposed to derive E [7], [9].

When the only information available is the pixel coordinates of points on the two views, the fundamental matrix F encapsulates the relation between the corresponding points on the two views through the following equation

$$m_r^T F m_l = 0 \tag{8}$$

In the same vein some different methods are proposed to derive the matrix F [4], [6], [7] , [10].

III. Validity of E and F Equations

In [3], Basta demonstrated that the development of the essential matrix of Longuet-Higgins [8] is flawed. He pointed out the origin of the flaw is the construction of the expression $\boldsymbol{M}_r^T E \boldsymbol{M}_l$. We will carry on in the same vein to examine the validity of the equations of the essential and fundamental matrices.

The epipolar geometry is defined in a 3D vector space where two coordinate systems are related by a translation $\boldsymbol{t}$ and a rotation R.

The Euclidean transformation matrix from the left to the right coordinate systems used in the epipolar geometry performs the translation first followed by the rotation [8]. So, as indicated by equation (2), the Euclidean 3D transformation matrix is:

$$T(\boldsymbol{v}) = R(\boldsymbol{v} - \boldsymbol{t}) \tag{9}$$

The rank of the 3D rotation matrix R is 3 as $\det(R) = 1$. So, the rank of T is 3 as well.

The projective transformation matrix from the left to the right coordinate systems is the 4×4 matrix which is the combination of a rotation matrix R and a translation vector $\boldsymbol{t}$ as: $P = \begin{bmatrix} R & \boldsymbol{t} \\ 0 & 1 \end{bmatrix}$, its rank is 4.

In all cases, the rank of the transformation matrix is greater than or equal to 3.

The rank of the essential and fundamental matrices is 2 [6]. Thus, the essential matrix E or the fundamental matrix F cannot be a 3D transformation matrix from any coordinate system to the other.

The points m_l and m_r are represented by the vectors $\boldsymbol{m}_l$ and $\boldsymbol{m}_r$. Thus, the equations of the matrices are equivalent to

$$\boldsymbol{m}_r^T E \boldsymbol{m}_l = 0 \text{ and } \boldsymbol{m}_r^T F \boldsymbol{m}_l = 0 \tag{10}$$

The product of a non-transformation matrix and a vector defined in a coordinate system is a vector defined in the same coordinate system.

Recall that the vector $\boldsymbol{m}_l$ is not defined in the right coordinate system, and $\boldsymbol{m}_r$ is not defined in the left coordinate system. Thus, the vector $\boldsymbol{m}_r^T F$ is defined in the right coordinate system and not defined in the left one. And the vector $F\boldsymbol{m}_l$ is defined in left the coordinate system and not defined in the right one.

The dot product of two vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ is simply the sum of products of their components [1]

$$\boldsymbol{u} \cdot \boldsymbol{v} = u_x v_x + u_y v_y + u_z v_z \tag{11}$$

The products $(\boldsymbol{m}_r^T F) \cdot \boldsymbol{m}_l$ and $\boldsymbol{m}_r^T \cdot (F\boldsymbol{m}_{l)})$ are operations between two vectors not defined in the same coordinate systems. These operations are not defined unless the two vectors (i.e. operands) are defined with respect to the same coordinate system.

The following example illustrated in Figure 2 validates this fact.

The two vectors $\boldsymbol{u} = [x, y, z]$ and $\boldsymbol{v}_1 = [a, b, c]$ are defined with respect to the left coordinate system, and the vector $\boldsymbol{v}_2 = [a, b, c]$ is defined with respect to the right coordinate system. The two vectors $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ have the same components.

Let us try answer the question: What is the value of the expression $\begin{bmatrix} x & y & z \end{bmatrix}\begin{bmatrix} a \\ b \\ c \end{bmatrix}$?

Is it equivalent to $\boldsymbol{u}^T\boldsymbol{v}_1$ or $\boldsymbol{u}^T\boldsymbol{v}_2$?

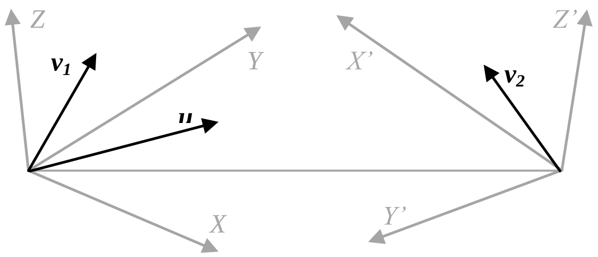

Fig. 2. Two vectors with same coordinates defined in two different coordinate systems.

As it is previously specified, the two vectors $\boldsymbol{u}$ and $\boldsymbol{v}_1$ are defined with respect to the same coordinate system, while the vectors $\boldsymbol{u}$ and $\boldsymbol{v}_2$ are **not** defined with respect to the same coordinate system.

It is easily concluded that $\begin{bmatrix} x & y & z \end{bmatrix}\begin{bmatrix} a \\ b \\ c \end{bmatrix} = \boldsymbol{u}^T\boldsymbol{v}_1$.

By contrast, the expression $\boldsymbol{u}^T\boldsymbol{v}_2$ is only defined when the vector $\boldsymbol{u}$ is transformed to the right coordinate system, or the vector $\boldsymbol{v}_2$ is transformed to the left coordinate system. Following equation (2), the coordinates of $\boldsymbol{u}$ in the right coordinate system will be

$$R\begin{bmatrix} x-t_x \\ y-t_y \\ z-t_z \end{bmatrix} \neq \begin{bmatrix} x \\ y \\ z \end{bmatrix} \tag{12}$$

Again following (2), the coordinates of $\boldsymbol{v}_2$ in the left coordinate system will be

$$R^{-1}\begin{bmatrix} a \\ b \\ c \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix} \neq \begin{bmatrix} a \\ b \\ c \end{bmatrix} \tag{13}$$

Therefore, in the right coordinate system:

$$\boldsymbol{uv}_2 = \left(R\begin{bmatrix} x-t_x \\ y-t_y \\ x-t_z \end{bmatrix} \right)^T \begin{bmatrix} a \\ b \\ c \end{bmatrix} \tag{14}$$

And in the left coordinate system:

$$\boldsymbol{uv}_2 = \begin{bmatrix} x & y & z \end{bmatrix}\left(R^{-1}\begin{bmatrix} a \\ b \\ c \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix} \right) \tag{15}$$

In all cases, $\boldsymbol{uv}_2 \neq \begin{bmatrix} x & y & z \end{bmatrix}\begin{bmatrix} a \\ b \\ c \end{bmatrix}$.

In terms of vector components, the second equation of (10) is equivalent to

$$\begin{bmatrix} x_r & y_r & z_r \end{bmatrix} F \begin{bmatrix} x_l \\ y_l \\ z_l \end{bmatrix} = 0 \tag{16}$$

The left hand side of (16) can be interpreted as the dot product of the two vectors $\begin{bmatrix} x_r & y_r & z_r \end{bmatrix}F$ and $\begin{bmatrix} x_l & y_l & z_l \end{bmatrix}^T$, or the dot product of the two vectors $\begin{bmatrix} x_r & y_r & z_r \end{bmatrix}$ and $F\begin{bmatrix} x_l & y_l & z_l \end{bmatrix}^T$.

In order to have the dot product of (16) defined; the two vectors must be defined in the same reference system. Which means either (x_r, y_r, z_r) are the components of a vector defined in the left coordinate system which is different of the vector $\boldsymbol{m}_r$ already defined in the right camera coordinate system, or (x_l, y_l, z_l) are the components of a vector defined in the right coordinate system which is different of the vector $\boldsymbol{m}_l$ defined in the left camera coordinate system.

Hence, whether the dot operation is performed in the left or in the right coordinate system, the equation $m_r^T F m_l = 0$ is either undefined or is not a relation between the points captured by the left camera and the points captured by the right camera. The same reasoning for F applies to E.

All these clarifications disclose the flaw in the theory of the epipolar geometry. Such a flaw is the result of performing the dot product on vectors not defined in the same coordinate system.

IV. Conclusion

In this paper, we recalled the following mathematical facts:

The product of a non-transformation matrix by a vector defined in a reference system is a vector defined in the same reference system.

The essential and fundamental matrices are not transformation matrices.

The dot product of two vectors is only defined when the two vectors are defined in the same reference system.

Based on these facts, we conclude that the equations of both essential and fundamental matrices are invalid. The reason for this is because these equations include an undefined operation on vectors.

References

[1] M. R. Adhikari, *Textbook of Linear Algebra an Introduction to Modern Algebra*, Allied Publishers Limited 2004.

[2] X. Armangué and J. Salvi, "Overall View Regarding Fundamental Matrix Estimation," *Image and Vision Computing*, vol. 21, no. 2, pp. 205-220, 2003.

[3] T. Basta, "Mathematical Flaws in the Essential Matrix Theory," in *Recent Advances in Signals and Systems, Proc. of the 9th WSEAS International Conference on Signal, Speech and Image Processing*. Budapest, Hungary: WSEAS Press, 2009, pp. 215-220.

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

[5] R. Hartley, "Defense of the Eight-Point algorithm," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 6, pp. 580-593, 1997.

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

[7] S. Ivekovic, A. Fusiello, and E. Trucco, "Fundamentals of multiple view geometry," In O. Schreer; P. Kauff; T. Sikora, *3D Videocommunication: Algorithms, concepts and real-time systems in human centered communication*, John Wiley & Sons, 2005.

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

[9] E. Trucco, A. Verri. *Introductory Techniques for 3-D Computer Vision*. Prentice Hall, 1998.

[10] 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.