Flaws in the Computer Algorithm for Reconstructing a Scene from Two Projections

Tayeb Basta, *Member, IACSIT*

***Abstract*—In 1981 Longuet-Higgins represented the world point by two vectors in the two camera reference frames and developed the essential matrix. Such a matrix is a relation between the corresponding image points on the two images of a world point on a rigid scene.**

The essential matrix is independent of the position and orientation of the cameras used to capture the two views.

The calculation of the essential matrix requires the knowledge of at least five accurate pairs of corresponding points. The unavailability of a procedure that fulfills such a requirement led researchers to focus their attention on developing estimation methods of the essential matrix without questioning the mathematical correctness of its derivation.

In this paper, we identify and expose flaws in Longuet-Higgins' derivation of the essential matrix. These flaws are the result of mixing up between the scalar product of vectors in a single reference frame and the transformation of vectors from one reference frame to another.

***Index Terms*—Dot product, essential matrix, epipolar geometry, Stereo vision.**

I. Introduction

Extracting a 3D shape from two images (views) captured of a rigid scene from two different standpoints is a basic component of the recovery of structure and motion from a sequence of camera images. Establishing a relation between the tokens of the two images was a challenge before the computer vision community.

Longuet-Higgins [3] was the pioneer in establishing a mathematical relation between the pairs of points on the two images; that is the essential matrix. Since then the extraction of 3D structure from two views became prominent in the computer vision literature. The main feature of the essential matrix is its independence of the scene structure, i.e. independent of the extrinsic camera parameters. The extrinsic camera parameters are the position and orientation of the camera with respect to the world coordinate system. Such a matrix encapsulates the relationships between pairs of corresponding points on the two views. Corresponding points are the projection of a world point on the two views.

Trucco and Verri [5] presented Higgins' derivation of the essential matrix in a slightly different manner. They formulated the two-view problem as the product of tree planar vectors, and then they extracted the fundamental matrix formula.

The work of Longuet-Higgins was supplemented by the work of Faugeras [2], who developed the essential matrix into the fundamental matrix which is independent of both extrinsic and intrinsic camera parameters. Intrinsic camera parameters include coordinates of the principal points, pixel aspect ratio, and focal lengths.

Since then a large number of publications appeared in the computer vision discipline to estimate the essential and fundamental matrices.

The contribution of this paper is to thoroughly reexamine the theory of Higgins' algorithm to disclose mathematical flaws in deriving the essential matrix.

The rest of the paper is as follows: Section 2 is a basic presentation of the epipolar geometry and the essential matrix. Section 3 introduces Longuet-Higgins' algorithm followed by identifying its flaw. In Section 4 presents Longuet-Higgins' derivation approach as presented by Trucco and Verri followed by clarifying the misconception that occurred in such a presentation. Finally, the current work concludes in Section 5.

II. Essential Matrix Derivation

From the standpoint of image analysis, the images of a scene are represented by sets of points with coordinates in the left and right camera coordinate frames. The epipolar geometry that represents the two-view extraction problem can be depicted as follows: a world point $M = (X, Y, Z)$ is defined in a world coordinate system. Two pinhole cameras are placed at two different positions C_l and C_r. The points C_l and C_r are the origin of the two coordinate frames of the two cameras. The right camera coordinate frame is obtained by a translation $\boldsymbol{t}$ and a rotation R of the left one.

The vectors $\boldsymbol{M}_l = (X_l, Y_l, Z_l)$ and represent the point M in the left and right camera coordinate frames, respectively. The point $m_l = (x_l, y_l)$ is the retinal image of the point M captured by the left camera; it belongs to the left camera plane π_l and it is defined in the left camera coordinate system. The point $m_r = (x_r, y_r)$ is the retinal image of the point M captured by the right camera; it belongs to the right camera plane π_r and it is defined in the right camera coordinate system (see Fig. 1).

Manuscript received April 17, 2012, revised May 21, 2012.

T. Basta is with Al Ghurair University, Dubai, UAE (e-mail: tayebasta@gmail.com).

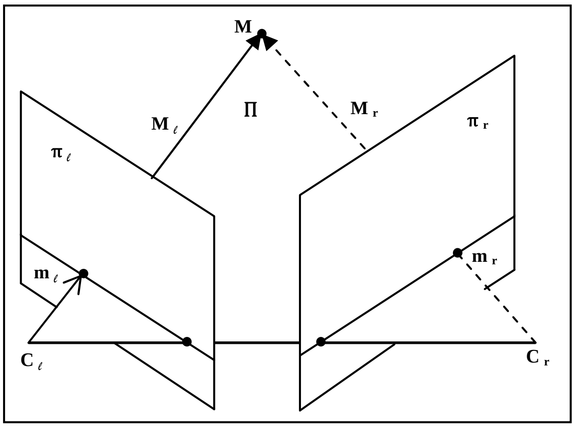

Fig. 1. The epipolar geometry.

The formula of the essential matrix is

$$\boldsymbol{M}_r^T E \boldsymbol{M}_l = 0 \quad (1)$$

The points m_l and m_r, the projections of the vectors $\boldsymbol{M}_l$ and $\boldsymbol{M}_r$ on the left and right camera frames, are obtained by dividing $\boldsymbol{M}_l = (X_l, Y_l, Z_l)$ by Z_l and $\boldsymbol{M}_r = (X_r, Y_r, Z_r)$ by Z_r, respectively. The relationship between a pair of corresponding points is then given by

$$m_r E m_l = 0 \quad (2)$$

The calculation of the essential matrix requires the knowledge of eight pairs of corresponding points [3]. The unavailability of a procedure that generates eight accurate pairs of corresponding points led researchers to focus their attention on developing new estimation methods and improving existing ones. Usually shortcomings of such methods are attributed to image noise, outliers (i.e. mismatches in correspondences), rounding errors, and the like. The theory of the essential matrix remains unquestionable.

III. Flaws in Longuet-Higgins' Derivation Approach

The right camera reference frame is obtained by a translation $\boldsymbol{t}$ to the right of the left camera position followed by a rotation R to direct the right camera towards the scene. Longuet-Higgins [3] created a matrix $E = RS$ and built the expression $\boldsymbol{M}_r^T E \boldsymbol{M}_l$ and after some arithmetic manipulations he established a relation between the vectors representing the point M in the two camera reference frames $\boldsymbol{M}_r^T E \boldsymbol{M}_l = 0$. He then divided by $Z_l Z_r$ to arrive at a relationship between m_l and m_r given by $m_r E m_l = 0$.

In the next points we analyze Longuet-Higgins' findings.

The entity E may be a transformation from the left reference frame to the right one or it may be any matrix.

1) The entity E is a transformation matrix. The value of the expression $\boldsymbol{M}_r^T E \boldsymbol{M}_l$ is independent of which operation $\boldsymbol{M}_r^T E$ or $E\boldsymbol{M}_l$ is performed first. We will consider the term $E\boldsymbol{M}_l$ first.
   a) "A vector is defined as an oriented segment with a certain magnitude and direction independent of any coordinate system. Vector transformation changes the coordinates without altering the magnitude and direction [1], [4]". The vector $\boldsymbol{V} = E\boldsymbol{M}_l$ is the transformation of the vector $\boldsymbol{M}_l$ from the left reference frame to the right reference frame. $\boldsymbol{V}$ is the vector $\boldsymbol{M}_l$ itself expressed in the right reference frame.
   b) The expression $\boldsymbol{M}_r^T E \boldsymbol{M}_l = \boldsymbol{M}_r^T \cdot \boldsymbol{V}$ is the dot product of the two vectors $\boldsymbol{M}_r$ and $\boldsymbol{V}$ which are defined in the same reference frame (i.e. the right frame). The dot product can be defined for two vectors $\boldsymbol{M}_r$ and $\boldsymbol{V}$ by $\boldsymbol{M}_r^T \cdot \boldsymbol{V} = |\boldsymbol{M}||\boldsymbol{V}|\cos\theta$, where θ is the angle between the vectors and $|X|$ is the norm. It follows immediately that $\boldsymbol{M}_r^T \cdot \boldsymbol{V} = 0$ if only $\boldsymbol{V}$ is perpendicular to $\boldsymbol{M}_r$ [6]. Remember that the vector $\boldsymbol{V}$ is nothing but $\boldsymbol{M}_l$ expressed in the right reference frame. As Fig. 1 illustrates, in the general case the vectors $\boldsymbol{M}_l$ and $\boldsymbol{M}_r$ are not perpendicular, in other words $\boldsymbol{M}_r^T \cdot \boldsymbol{V} \neq 0$. Thus, considering E as a transformation matrix contradicts Higgins' findings which is $\boldsymbol{M}_r^T E \boldsymbol{M}_l = \boldsymbol{M}_r^T \cdot \boldsymbol{V} = 0$.
2) The entity E is not a transformation matrix. The other possibility is that E is not a transformation matrix from the left reference frame to the right one. The vector $\boldsymbol{M}_l$ is defined with respect to the left reference frame; the vector $\boldsymbol{V} = E\boldsymbol{M}_l$ is defined in the left frame as well. The vector $\boldsymbol{V} = E\boldsymbol{M}_l$ has a magnitude and a direction different from the magnitude and/or the direction of $\boldsymbol{M}_l$ unless E is the identity matrix. $\boldsymbol{V}$ is not defined in the right frame. The vector $\boldsymbol{M}_r$ is defined in the right frame and not defined in the left frame. The expression $\boldsymbol{M}_r^T \cdot \boldsymbol{V}$ is not defined unless the two vectors are expressed in the same reference frame; consequently the expression $\boldsymbol{M}_r^T \cdot \boldsymbol{V} = \boldsymbol{M}_r^T E \boldsymbol{M}_l$ is undefined.

IV. Flaws in Trucco and Verri Derivation

Trucco and Verri [5] expressed Loguet-Higgins' approach by formulating the two-view problem as the product of three planar vectors in the world reference frame. They calculated the projections of the point M in the left and right reference frames as

$$m_l = {}^{f}\!/_{Z_l}\, \boldsymbol{M}_l \quad (3)$$

and

$$m_r = {}^{f}\!/_{Z_r}\, \boldsymbol{M}_r \quad (4)$$

Let $\boldsymbol{t} = \boldsymbol{C}_r - \boldsymbol{C}_l$ be the translation vector of the right camera with respect to the left frame. The relation between the vectors $\boldsymbol{M}_l$ and $\boldsymbol{M}_r$ is therefore

$$\boldsymbol{M}_r = R(\boldsymbol{M}_l - \boldsymbol{t}) \quad (5)$$

The three vectors $\boldsymbol{M}_l$, $\boldsymbol{t}$ and $\boldsymbol{M}_l - \boldsymbol{t}$ are coplanar, so they satisfy the coplanarity condition:

$$(\boldsymbol{M}_l - \boldsymbol{t})^T \boldsymbol{t} \times \boldsymbol{M}_l = 0 \tag{6}$$

Using (5), equation (6) can be written as

$$(R^T \boldsymbol{M}_r)^T \boldsymbol{t} \times \boldsymbol{M}_l = 0 \tag{7}$$

After some algebraic manipulations, the authors arrived at

$$\boldsymbol{M}_r^T E \boldsymbol{M}_l = 0 \tag{8}$$

Then they divided (8) by $Z_l Z_r$ and used (3) and (4) to establish the relationship between the image points as

$$m_r E m_l = 0 \tag{9}$$

In the next points we analyze Trucco and Verri findings.

1) Let us go back to equation (5) and examine the equality $\boldsymbol{M}_r = R(\boldsymbol{M}_l - \boldsymbol{t})$:

The matrices R and $\boldsymbol{t}$ are the rotation and translation of the right reference frame with respect to the left reference frame. The vectors $\boldsymbol{M}_l$ and $\boldsymbol{t}$ are defined in the left camera frame which coincides with the world coordinate system. Therefore, the vector $\boldsymbol{M}_l - \boldsymbol{t}$ is also defined in the left camera frame.

The production of a scalar by a 3D vector is another vector defined in the same coordinate system, and the production of a 3×3 scalar matrix by a 3D vector is another 3D vector defined in the same coordinate system. Thus, the vector $R(\boldsymbol{M}_l - \boldsymbol{t})$ is a vector defined in the left camera frame.

$\boldsymbol{M}_r$ is defined in the right camera frame.

The coordinates of the point M in a coordinate system is equal to the components of its position vector in that coordinate system.

Equation (5) uses the coordinates of the point M in the left coordinate system to calculate its coordinates in the right coordinate system. Therefore, the components of the vector $\boldsymbol{M}_r$ *in the right camera frame* are equal to the components of $R(\boldsymbol{M}_l - \boldsymbol{t})$ *in the left camera frame*. However, the equality in (5) does not mean that $\boldsymbol{M}_r$ and $R(\boldsymbol{M}_l - \boldsymbol{t})$ are the same vector.

The following example illustrates this idea.

Let us have left and right coordinate systems, and let $\boldsymbol{u} = [2,2,1]$ and $\boldsymbol{t} = [1,0,0]$ be two vectors defined in the left coordinate system, and $R = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \sqrt{3}/2 & 1/2 \\ 0 & -1/2 & \sqrt{3}/2 \end{bmatrix}$ is a rotation matrix.

The vector $R(\boldsymbol{u} - \boldsymbol{t}) = \left[1, 1/2 + \sqrt{3}, -1 + \sqrt{3}/2\right]$ is a 3D vector in the left coordinate system. At the same time, $\left[1, 1/2 + \sqrt{3}, -1 + \sqrt{3}/2\right]$ are the components of the vector $\boldsymbol{v} = R(\boldsymbol{u} - \boldsymbol{t})$ in the right coordinate system.

By definition we have $R^T R = 1$ where $R^T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \sqrt{3}/2 & -1/2 \\ 0 & 1/2 & \sqrt{3}/2 \end{bmatrix}$ the transpose of R. Multiplying $\boldsymbol{v} = R(\boldsymbol{u} - \boldsymbol{t})$ from the left by R^T produces the vectors: $R^T \boldsymbol{v} = [1,2,1]$ defined in the right coordinate system, and $\boldsymbol{u} - \boldsymbol{t} = [1,2,1]$ defined in the left coordinate system.

The vectors $R^T \boldsymbol{v}$ and $\boldsymbol{u} - \boldsymbol{t}$ have the same components but they are different as they are defined in two different coordinate systems (see Fig. 2).

In the current derivation method, the authors [5] obtained (6) by substituting in (5) the vector $R^T \boldsymbol{M}_r$ for the vector $\boldsymbol{M}_l - \boldsymbol{t}$. These two vectors have the same components but they are different from each other as they are defined in two different coordinate systems.

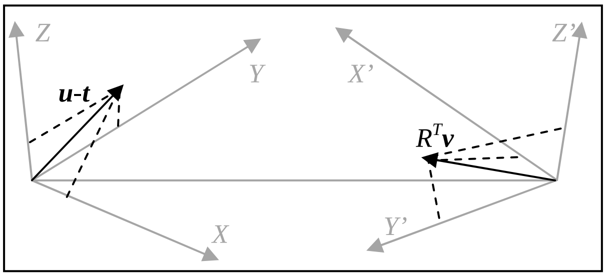

Fig. 2. Two vector in two references with same components

2) We can also approach the analysis of equation (5) in the following way:

The equality $\boldsymbol{M}_r = R(\boldsymbol{M}_l - \boldsymbol{t})$ implies that the vector $\boldsymbol{M}_r$ is the transformation of $\boldsymbol{M}_l$ from the left to the right frame. The transformation of a vector from one frame to another changes its coordinates and does not alter neither its magnitude nor its direction [1], [4]. Thus, $\boldsymbol{M}_l$ and $\boldsymbol{M}_r$ which satisfies the equality $\boldsymbol{M}_r = R(\boldsymbol{M}_l - \boldsymbol{t})$, are the same vector expressed in two different reference frames. This vector $\boldsymbol{M}_r$ is different of that of Fig. 1. However, as Fig. 1 illustrates the assumption of Trucco and Verri [5], $\boldsymbol{M}_l$ and $\boldsymbol{M}_r$ are two different vectors representing the same point M in two different reference frames.

V. Conclusion

Longuet-Higgins' approach to extract 3D structure from a two views of a rigid scene is considered a landmark in the history of computer vision discipline. The calculation of the essential matrix requires the determination of eight corresponding points. The unavailability of a procedure that provides eight pairs of accurate corresponding points led researchers to focus their attention on estimating the essential matrix.

Mixing up vector operations and vector transformation from a reference frame to another is the reason behind the flaw in Longuet-Higgins' derivation method. Such confusion became clearer in Trucco and Verri presentation of Longuet-Higgins' method when they explicitly specify that the relation between the vectors $\boldsymbol{M}_l$ and $\boldsymbol{M}_r$ as $\boldsymbol{M}_r = R(\boldsymbol{M}_l - \boldsymbol{t})$.

The equality $\boldsymbol{M}_r = R(\boldsymbol{M}_l - \boldsymbol{t})$ can be seen as the transformation of the point M from the left camera frame to the right camera frame. In such a case, it does mean that the components of the vector $\boldsymbol{M}_r$ in the right camera frame (i.e. the coordinates of the point M in the right camera frame) are equal to the components of the vector $R(\boldsymbol{M}_l - \boldsymbol{t})$ in the left camera frame, where $\boldsymbol{M}_l$ is the position vector of the point M in the left camera frame. However, it does not mean that $\boldsymbol{M}_r$ and $R(\boldsymbol{M}_l - \boldsymbol{t})$ are the same vector, as each of them is defined in a different coordinate system.

Equation $\boldsymbol{M}_r = R(\boldsymbol{M}_l - \boldsymbol{t})$ may also be considered a transformation of the vector $\boldsymbol{M}_l$ from the left camera frame to the right camera frame. In this case, the vector $\boldsymbol{M}_r$ is the same vector $\boldsymbol{M}_l$ expressed in the right frame; as the transformation of a vector from one frame to another changes its coordinates and does not alter neither its magnitude nor its direction. This option indicates that $\boldsymbol{M}_l$ and $\boldsymbol{M}_r$ is the same vector expressed in two different coordinate systems. Thus, the projections m_l and m_r of the vector $\boldsymbol{M}_l$ (i.e. $\boldsymbol{M}_r$) in the left and right coordinate systems are different of the points m_l and m_r the projections of the two different vectors $\boldsymbol{M}_l$ and $\boldsymbol{M}_r$ illustrated in Fig. 1. Thus, the equation $m_r E m_l = 0$ is not a relation between the projection of the point M on the left frame and the projection of M on the right frame.

REFERENCES

[1] Bate R., D. Mueller, and J. White, *Fundamentals of Astrodynamics*, Dover Publications, Inc., 1971.

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

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

[4] S. Pissanetzky, *Rigid Body Kinematics and C++ Code*, SciControls.com, 2005.

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

[6] W. Weisstein, Eric (2010). Dot Product. From MathWorld, a Wolfram Web Resource. [Online]. Available: http://mathworld.wolfram.com/DotProduct.html.

Tayeb Basta was born in Algeria in 1960. He received his BSc in computer science in 1983 from the University of Annaba in Algeria. His PhD is from the computation department of the University of Manchester in UK. He is now an assistant professor at Al Ghurair University in Dubai.