SHORTCOMINGS OF THE FUNDAMENTAL MATRIX EQUATION TO RECONSTRUCT 3D SCENES

Tayeb Basta

College of Engineering and Computing, Al Ghurair University, Dubai, UAE

ABSTRACT

In stereo vision, the epipolar geometry is the intrinsic projective geometry between the two views. The essential and fundamental matrices relate corresponding points in stereo images. The essential matrix describes the geometry when the used cameras are calibrated, and the fundamental matrix expresses the geometry when the cameras are uncalibrated. Since the nineties, researchers devoted a lot of effort to estimating the fundamental matrix. Although it is a landmark of computer vision, in the current work, three derivations of the essential and fundamental matrices have been revised. The Longuet-Higgins' derivation of the essential matrix where the author draws a mapping between the position vectors of a 3D point; however, the one-to-one feature of that mapping is lost when he changed it to a relation between the image points. In the two other derivations, we demonstrate that the authors established a mapping between the image points through the misuse of mathematics.

KEYWORDS

Fundamental Matrix, Essential Matrix, Stereo Vision, 3D Reconstruction.

1. INTRODUCTION

In computer stereo vision, the 3D object shape reconstruction from two 2d images can be defined as follows:

The object to be reconstructed is a set of 3D points M, it is depicted by two cameras from two different standpoints. Left and right coordinate systems are defined in each of these standpoints. And every 3D point is projected on the left and right images as two 2D points m_l and m_r, respectively.

The epipolar geometry is the intrinsic projective geometry between the two views. It is independent of scene structure, and only depends on the cameras' internal parameters and relative pose. The fundamental matrix F encapsulates this intrinsic geometry [1].

A 3D point M is represented in the left and right coordinate systems by two position vectors $M_l = [X_l \quad Y_l \quad Z_l]^T$ and $M_r = [X_r \quad Y_r \quad Z_r]^T$. And $m_l = [x_l \quad y_l]^T$ and $m_r = [x_r \quad y_r]^T$ are the position vectors of the projective points m_l and m_r in the left and right coordinate systems, respectively, as in Figure 1.

3D shape reconstruction is performed in the following steps [1]

1. Compute the fundamental matrix from point correspondences.

David C. Wyld et al. (Eds): NATP, ACSTY, CCCIOT, MLSC, ITCSS - 2022
pp. 77-86, 2022. CS & IT - CSCP 2022 DOI: 10.5121/csit.2022.120106

2. Compute the camera matrices from the fundamental matrix.
3. For each point correspondence $m_l \leftrightarrow m_r$, compute the point in space that projects to these two image points.

Thus, the first step is to compute the fundamental matrix and the eight-point algorithm is the most used method to do so. In practice the number of image points is large; so, the fundamental matrix can only be estimated rather calculated. Researchers keep developing methods that overcome previously devised ones in terms of accuracy and mitigating noise effects. Only few researchers thought that the bad performance of the eight-point algorithm would requires the revision of the projective geometry approach itself.

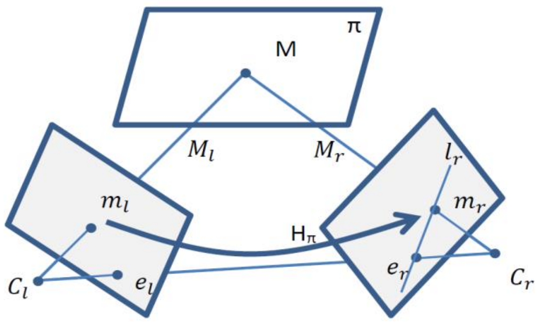

Figure1. The epipolar geometry. A point m_l in one image is transferred via the plane π to a matching point m_r in the second image. The epipolar line l_r through m_r is obtained by joining m_r to the epipole e_r.

The main objective of the current work is to revise the theory underpinning the derivation of the essential and fundamental matrices equations. Thus, clarify the reason behind the bad performance of the projective geometry application to 3D reconstruction from 2D views.

The rest of the paper is organized as follows: Section 2 introduces the motivation of addressing a classic problem like the fundamental matrix of stereo vision. Sections 3 exposes some related work. Section 4 demonstrates the shortcoming of the essential matrix equation. Section 5 shows the mathematical flaws of two derivations of the fundamental matrix. And the paper concludes in section 6.

2. WHY SHOULD WE ADDRESS SUCH A CLASSIC PROBLEM?

The epipolar geometry application in computer stereo vision represented by the fundamental matrix is still part of computer vision courses in most universities around the world. On top of that, researchers are still spending time to develop methods to estimating the fundamental matrix [2, 3, 4, 5, 6]. Table 1 shows a sample of outstanding universities with links to their computer vision courses that include at least one chapter on epipolar geometry and the fundamental matrix.

Table 1 Sample universities teaching the epipolar geometry to reconstruct 3D shape from two views.

University	Course Title	Course Link
Stanford University, USA	Computer Vision, From 3D Reconstruction to Recognition	http//web.stanford.edu/class/cs231a/syllabus.html
The University of Washington, USA	Computer Vision	https//courses.cs.washington.edu/courses/cse455/
MIT, USA	Computer Vision and Applications	www.ai.mit.edu/courses/6.891/lectnotes/lect8/lect8-slides.pdf

University College London, UK	Machine Vision	https//www.ucl.ac.uk/module-catalogue/modules/machine-vision/COMP0137
University of Toronto, Canada	Foundations of Computational Vision	http//www.cs.toronto.edu/~kyros/courses/2503
Tokyo Institute of Technology, Japan	Computer Vision	http//www.ocw.titech.ac.jp/index.php?module=General&action=T0300&JWC=201804591&lang=EN&vid=03
Sorbonne Université - Télécom Paris	Master Informatique - Parcours IMA	https//perso.telecom-paristech.fr/bloch/P6Image/VISION.html

3. RELATED WORK

Though the fundamental matrix theory is considered as a landmark achievement of computer vision, certain researchers called it into question [7, 8, 9,10, 11, 12]. In a series of research work, Basta demonstrated that many of the derivation methods of the essential and fundamental matrix equations are flawed [13, 14, 15, 16, 17, 18, 19].

In [17] and [19], the author presented extensive experimental results of two real images of a building captured from two standpoints. The building (Figure 2) is composed of two parts with different depths with respect to the camera lens. In [17], the author used a MATLAB Toolbox [20] that contains several methods for estimating the fundamental matrix using the eight-point algorithm. In [19], he implemented the solution in Python and used the findFundamentalMat() function of the cvonline package to estimate the fundamental matrix.

Figure 2. The building image used to estimate the fundamental matrix in [17] and [19].

In both works [17] and [19], the author estimated the fundamental matrix that satisfies the equation $m_r^T F m_l = 0$. Then, he calculated the values of the expression $m_r^T F m_l$ for several pairs of corresponding points (m_l, m_r). Such values are supposed to be equal to zero. The matrix F is calculated from different regions of the images (whole images, back part of the images, and front side of the images) and the pairs of corresponding points are selected arbitrarily from the images. Table 2 shows that the values of $m_r^T F m_l$ are sometimes very far away from 0; greater than 10 for some cases.

Table 2. the values of the expression $m_r^T F m_l$ calculated for selected points from the whole images, the back side, and the front side of the images. As it is apparent the image is composed of components with different depth with respect to the camera lens. This result is published in [19].

F matrix calculated from		
Whole	Back	Front
0.322	0.121	-0.504

0.084	1.496	0.557
-0.026	0.545	0.684
0.234	3.978	0.748
0.328	7.314	-0.726
0.135	16.158	-0.508
-0.165	9.001	-0.784
0.184	13.800	2.989
0.070	12.401	-0.109
0.135	10.794	-1.970

In the current work, three main publications where the essential and fundamental matrices are derived as a product of a skew matrix and a rotation transformation matrix are scrutinized. One of these is where the first time the essential matrix introduced to the computer vision community by Longuet-Higgins [21]. Next section shows how Longuet-Higgins succeeded in securing a one-to-one mapping between the position vectors of world points of a scene and that mapping is lost when he transformed it to a relation between the image points. In the other two derivations, the authors try to directly establish a one-to-one relation between the image points. Such a relation is represented by the fundamental matrix. The current work shows the mathematical flaws in these two derivations.

4. LONGUET-HIGGINS' DERIVATION OF THE ESSENTIAL MATRIX

4.1. The Equation Derivation

In [21], Longuet-Higgins created a matrix $Q = RS$ where $S = \begin{bmatrix} 0 & t_3 & -t_2 \\ -t_3 & 0 & t_1 \\ t_2 & -t_1 & 0 \end{bmatrix}$. The matrix R and the vector t are the rotation and translation of the right coordinate system with respect to the left coordinate system. M_l and M_r are the position vectors of a world point M on the left and right coordinate systems, respectively. The author formed the expression $M_r^T Q M_l$ and after some arithmetic manipulations he found out that

$$M_r^T Q M_l = 0 \quad (1)$$

For every 3D point there are exactly two position vectors; one represents that point in the left coordinate system and the other represents the point in the right coordinate system. Thus, Q in (1) is a one-to-one mapping between M_l and M_r.

In terms of coordinates, $M_l = (X_l, Y_l, Z_l)$ and $M_r = (X_r, Y_r, Z_r)$. And the coordinates of the projective points m_l and m_r of the point M in the left and right coordinate systems, respectively are

$$\begin{aligned} m_l &= (X_l/Z_l\,, Y_l/Z_l\,, 1) \\ m_r &= (X_r/Z_r\,, Y_r/Z_r\,, 1) \end{aligned} \quad (2)$$

Finally, the author divided the left-hand side of (1) by $Z_l Z_r$ to conclude the essential matrix equation

$$m_r^T E m_l = 0 \quad (3)$$

4.2. Shortcoming of Longuet-Higgins's derivation

Longuet-Higgins approached the problem from an algebraic perspective, he used matrix product as the main operation to derive the essential matrix equation. So, he has not been faced with the problem of transformation from one coordinate system to the other.

He formed the expression $M_r^T Q M_l$. And because the matrix product is an associative operation, the expression $M_r^T Q M_l$ is the product of 1×3 row matrix and a 3×3 matrix and a 3×1 column matrix which led to equation (1).

The problem of Longuet-Higgins' derivation started when he divided equation (1) by $Z_l Z_r$. As it is known, the position vector of a point is the unique vector from the origin of the coordinate system to the point itself. So, for every point M, equation (1) holds for exactly two position vectors M_l and M_r in the left and right coordinate systems, respectively. Once (1) is divided by $Z_l Z_r$ we will get the following equation

$$\frac{M_r^T}{Z_r} \cdot Q \cdot \frac{M_l}{Z_l} = 0 \qquad (4)$$

Where $m_l = \frac{M_l}{Z_l}$ and $m_r = \frac{M_r^T}{Z_r}$ are the projection of the vectors M_l and M_r on the left and right camera planes, respectively.

In projective geometry, m_l could be the projection of a single world point or multiple world points (Figure 3). It is the projection of all world points laying on the ray drawn from the camera lens centre to the point M.

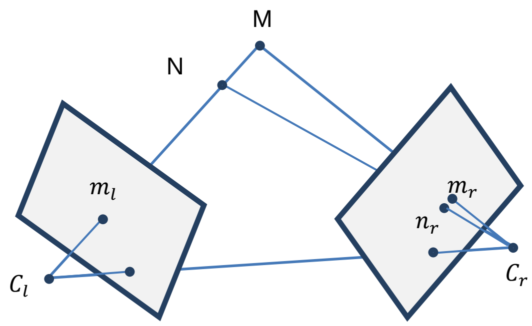

Figure 3. the image point m_l is the projection of two world points M and N. m_l is a corresponding point to two image points m_r and n_r.

Furthermore, there are world points visible to one camera and invisible to the other. This could be because these points are hidden by 3D objects in the scene. This is one of the characteristics of 3D scenes. So, they are projected on the first camera plane and does not have an image on the other camera. However, when you plug this image point into m_l or m_r and solve equation (3), you will get a false corresponding point.

Recall the 3D shape reconstruction as described in [1]

1. Compute the fundamental (essential) matrix from point correspondences.
2. Compute the camera matrices from the fundamental matrix.

3. For each point correspondence $m_l \leftrightarrow m_r$, compute the point in space that projects to these two image points.

Assuming the point $p = (1,2,1)$ is on the left camera plane (image). The matrix E is already calculated or estimated. To compute the corresponding point of p, we plug the value of p into equation (3).

$$[x_r \quad y_r \quad 1]E[1 \quad 2 \quad 1]^T = 0, E = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \tag{5}$$

Substituting for the matrix E, we will get the following equation

$$[x_r \quad y_r \quad 1]\begin{bmatrix} A_1 \\ A_2 \\ A_3 \end{bmatrix} = 0 \text{ , where} \begin{bmatrix} A_1 \\ A_2 \\ A_3 \end{bmatrix} = \begin{bmatrix} a_{11} + 2a_{12} + a_{13} \\ a_{21} + 2a_{22} + a_{23} \\ a_{31} + 2a_{32} + a_{33} \end{bmatrix} \tag{6}$$

which leads to the following equation

$$A_1 x_r + A_2 y_r + A_3 = 0 \tag{7}$$

There are infinite values of (x_r, y_r) satisfying equation (7). Geometrically, this means that any point p has many corresponding points. Which is incorrect; the certainty is each image point has at most one corresponding point in each other image except the case of occlusion when two different points have the same corresponding point.

Consequently, the essential (fundamental) matrix equation does not ensure the recovery of the right shapes of 3D scenes.

5. Establishing a Direct Mapping between the Image Points

Because the above essential matrix derivation suffers from the drawback of an image point can have unlimited number of corresponding points, computer vision researchers try to directly draw a mapping between the image points without passing through position vectors of the 3D point. The next sections explore the flaws of two well-known derivations of the essential and fundamental matrices equations.

5.1. Luong-Faugeras derivation of the essential matrix

In [22], Luong et al. assert that because the vector from the first camera optical centre to the first imaged point m_l, the vector from the second optical centre to the second imaged point m_r, and the vector from one optical center to the other t are all coplanar. In normalized coordinates, this constraint can be expressed simply as

$$m_r^T \, (t \times R m_l) = 0 \tag{8}$$

where R and t capture the rotation and translation of the right cameras coordinate system with respect to the left one. In [23], Birchfield explicitly stated that the multiplication by R is necessary to transform m_l into the second camera's coordinate system. The authors [22] defined $[t]_\times$ as the matrix such that $[t]_\times \, y = t \times y$ for any vector y, and they rewrite equation (8) as a linear equation

$$m_r^T \left([t]_\times R m_l\right) = m_r^T E m_l = 0, \tag{9}$$

Where $E = [t]_\times R$ is called the Essential matrix and $[t]_\times = \begin{bmatrix} 0 & -t_3 & t_2 \\ t_3 & 0 & -t_1 \\ -t_2 & t_1 & 0 \end{bmatrix}$.

5.2. The flaw in Luong-Faugeras derivation

Let us closely examine equation (8), $m_r^T\ (t \times Rm_l) = 0$.

We have the following facts. The point m_l is on the left image, so the vector m_l is defined in the left coordinate system and not defined in the right one. The point m_r is on the right image, then the vector m_r is defined in the right coordinate system and not defined in the left one. And the vector t, the translation of origin of the right coordinate system with respect to the left coordinate system; so, t is defined in the left coordinate system and not defined in the right one.

The left hand-side of (8) consists of three vector operations. The term inside the parenthesis is evaluated first which includes a vector product and a matrix product.

Let assume that Rm_l is to be evaluated first; it is the product of a rotation transformation matrix and a vector. So, $v = Rm_l$ is the vector m_l expressed in the right coordinate system. Therefore $t \times Rm_l = t \times v$ is the cross product of t defined in the left coordinate system and v defined in the right coordinate system. Thus, $t \times Rm_l$ is the cross product of two vectors not defined in the same coordinate systems; so, it is invalid.

Now, let us consider that the cross-product operation $t \times R$ is to be evaluated first.

DEFINITION

The cross product (or vector product) of two vectors

$x = \langle x_1, x_2, x_3 \rangle$ and $y = \langle y_1, y_2, y_3 \rangle$ in $\mathbb{R}^3$ is the vector $x \times y = \langle x_2y_3 - x_3y_2\ , x_3y_1 - x_1y_3\ , x_1y_2 - x_2y_1 \rangle$.

The cross product of two vectors x and y in $\mathbb{R}^3$ is a vector orthogonal to both x and y [24].

The cross product of a 3D vector and a 3×3 matrix is UNDEFINED [24].

Therefore, there is no operation called cross product of a vector and a matrix; therefore, the term $t \times R$ is undefined. Thus, equation (8) that is the premise of the current derivation of the essential matrix is invalid. And the current derivation of the essential matrix is flawed.

One could claim that Rm_l is a product of a matrix and a vector which produces a vector defined in the same coordinate system. Then the cross-product $t \times Rm_l$ is a vector defined in the left coordinate system. in this case, $m_r^T \cdot\ (t \times Rm_l)$ is a dot product of two vectors, one from the right coordinate system and the other from the left coordinate system. it is an undefined operation.

5.3. Hartley-Zisserman derivation of the fundamental matrix

In the geometric derivation of the fundamental matrix equation, the authors [1] assert the existence of 2D homography H_π mapping each point m_l from the left image to a point m_r on the right image, because the set of all such points m_l in the left image and the corresponding points m_r in the right image are projectively equivalent, since they are each projectively equivalent to the planar point set M (Figure 1). Thus, $F = [\mathrm{e}_r]_\times H_\pi$ that is a matrix product of a skew matrix and a transformation from left to right.

5.4. The flaw in Hartley-Zisserman derivation

The points M in the above statement are the world points of the 3D scene to be reconstructed from a pair of its images. If the 3D scene is planar, why are we constructing a planar scene from two of its planar images in the first place. Thus, the existence of a homography mapping points of the left image to points on the right image is on condition that the 3D scene is planar. Because typical 3D scenes might contain objects with different depths (i.e., distance from the camera centre). So, some points on these objects can be visible to one camera and hidden from the other. Therefore, some image points on the left camera plane will not have corresponding points on the right camera plane and points on the right image will not have corresponding points on the left image. Furthermore, researchers recognize the existence of the occlusion problem [25] where two 3D points or more are projected onto the same image point as in Figure 3. At the same time, they assert the existence of a homography between points of the left image and those on the right image. These facts, confirm that points on the left and right images are not projectively equivalent and no homography exists between them. In conclusion, the expression $F = [\mathrm{e}_r]_\times H_\pi$ where H_π is a homography is irrational.

6. CONCLUSION

In this work, we demonstrated that the first ever derivation of the essential matrix that has been introduced to the computer vision community is free of flaws; however, it does not ensure a one-to-one mapping between the image points of the two views. Later, researchers tried to address such shortcoming through deriving the essential and fundamental matrices equation as a mapping between the image points. We showed that two of the well-known of these derivations are mathematically flawed.

The current work establishes a rigorous scrutiny of a theory that claims to be mathematically founded. The trend for solving computer vision problems uses machine learning tools to obtain good solutions without requiring any mathematical basis.

REFERENCES

[1] Richard Hartley & Andrew Zisserman, (2004). *Multiple View Geometry in Computer Vision*, 2nd edn. Cambridge University Press, chapter 9.

[2] Omid Poursaeed, Guandao Yang, Aditya Prakash, Qiuren Fang, Hanqing Jiang, Bharath Hariharan & Serge Belongie, (2018) "Deep fundamental matrix estimation without correspondences", In *Proceedings of the European conference on computer vision workshop*, pp1–13.

[3] Daniel Barath, (2018) "Five-Point Fundamental Matrix Estimation for Uncalibrated Cameras", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (CVPR), pp235-243.

[4] Yi Zhou, Laurent Kneip & Hongdong Li, (2015) "A revisit of methods for determining the fundamental matrix with planes", In 2015 *International conference on digital image computing techniques and applications* (DICTA), pp 1–7. *IEEE*. http//ieeexplore.ieee.org/document/7371221/.

[5] Sushil Pratap Bharati, Feng Cen, Ajay Sharda & Guanghui Wang, (2018) "RES-Q Robust Outlier Detection Algorithm for Fundamental Matrix Estimation", *IEEE Access*, 6, 48664–48674.

[6] Nurollah Tatar & Hossein Arefi, (2019) "Stereo rectification of pushbroom satellite images by robustly estimating the fundamental matrix", *International Journal of Remote Sensing*, 40(23), 8879-8898.

[7] Andrew Zisserman & Stephen Maybank, (1993) "A Case Against Epipolar Geometry", In *2nd Europe-U.S. Workshop on Invariance*, pp 69–88. Ponta Delgada, Azores.

[8] Richard Hartley, (1997) "In Defense of the Eight-Point Algorithm", *IEEE Transaction on Pattern Recognition and Machine Intelligence*, 19(6), 580–593.

[9] Hugh Christopher Longuet-Higgins (1984) "The Reconstruction of a Scene from Two Projections- Configurations that Defeat the 8-Point Algorithm", In *Proceedings of 1st Conf. Artificial Intelligence Applications*, pp395-397.

[10] Quang-Tuan Luong & Olivier Faugeras, (1993) "Determining the Fundamental Matrix with Planes Instability and New Algorithms", *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, CVPR 93, pp489-494. New York.

[11] Thomas Marill, (1988) "A Counterexample to the Theory That Vision Recovers Three-Dimensional Scenes", *A. I. Working Paper, MIT Artificial Intelligence Laboratory*, 319.

[12] Berthold Horn, (1999) "Projective Geometry Considered Harmful", Copyright © people.csail.mit.edu/bkph/articles/Harmful.pdf, last accessed 2021/6/15.

[13] Tayeb Basta, (2009) "Mathematical flaws in the essential matrix theory", In *Proceedings of WSEAS International Conference*, in *Recent Advances in Computer Engineering*, 9, pp215-220. Budapest, Hungary.

[14] Tayeb Basta, (2010) "An Invalid Derivation of the Fundamental Matrix Based on the Equation of a Point Lying on a Line", Workshop on Frontiers of Computer Vision FCV 2010, pp83 – 88. Hiroshima, Japan.

[15] Tayeb Basta, (2012) "Does the Fundamental Matrix Define a One-to-One Relation between the Corresponding Image Points of a Scene?", *Journal of Image and Graphics*, 1(3), 125-128.

[16] Tayeb Basta, (2014) "Is the Fundamental Matrix Really Independent of the Scene Structure?", *International Journal of Signal Processing, Image Processing and Pattern Recognition*, 7(5), 149-168.

[17] Tayeb Basta, (2017) "The Eight-Point Algorithm is not in Need of Defense", *ARPN Journal of Engineering and Applied Sciences*, 12(3), 753-760.

[18] Tayeb Basta, (2019) "The Controversy Surrounding the Application of Projective Geometry to Stereo Vision", In *Proceedings of the 5th International Conference on Computer and Technology Applications*, ICCTA, pp 15–19. Istanbul.

[19] Tayeb Basta, (2020) "Experimental and Theoretical Scrutiny of the Geometric Derivation of the Fundamental Matrix", In *Proceedings of the 3rd International Conference on Artificial Intelligence and Pattern Recognition*, pp25-29. Huaqiao University, Xiamen, China.

[20] Joaquin Salvi, "Fundamental Matrix Estimation toolbox", http//eia.udg.es/~qsalvi/recerca.html, last accessed 3/3/2016.

[21] Hugh Christopher Longuet-Higgins, (1981) "A computer algorithm for reconstructing a scene from two projections", *Nature*, 293, 133-135.

[22] Quang-Tuan Luong & Olivier Faugeras, (1996) "The Fundamental matrix theory, algorithms, and stability analysis", *International Journal of Computer Vision*, 17(1), 43-76.

[23] Stan Birchfield, (1998) "An introduction to projective geometry (for computer vision)", Stanford university https//www.hhi.fraunhofer.de/fileadmin/Departments/VIT/IMC/Team/schreer/doc_pub/1998_Lecture_ProjGeometryIntro_Birchfield.pdf, last accessed 2021/04/12.

[24] Ron Larson, (2016) *Elementary Linear Algebra*. 8th edn. Cengage Learning.

[25] Jianguo Liu, Xuesong Li, Qin Qin & Hao Zhang, (2016) "Study of occlusion problem in stereo matching", *17th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing* (SNPD), pp195-199. Shanghai, China.

AUTHORS

Tayeb Basta graduated with a B. Eng. in computer science in 1983 from the University of Annaba in Algeria. In 1994, he obtained his PhD in Computer Science from the Victoria University of Manchester in UK. Basta is now an associate professor at Al Ghurair University in Dubai, UAE.