

## Is the Fundamental Matrix Really Independent of the Scene Structure?

Tayeb Basta

Associate Professor at Al Ghurair University, Dubai  
[tayebasta@gmail.com](mailto:tayebasta@gmail.com)

### Abstract

In stereo vision, two images of a 3D scene are acquired from two viewpoints. One of the objectives of stereo vision work is to recover the 3D structure of the scene. Epipolar geometry describes the relationship between the images, and the essential and fundamental matrices are the algebraic representations of this geometry. The most important feature of these matrices that is emphasized in the literature is that they are independent of the scene structure.

This article illustrates—empirically and theoretically—that the fundamental matrix depends on the scene structure and demonstrates that the matrix in  $m_r F m_l = 0$  not only represents a relationship between corresponding points of the two views but also represents a relationship between other non-corresponding points. Furthermore, we show empirically that the equation  $m_r F m_l = 0$  does not hold for any pair of corresponding points. In scenes with objects of different depths, the value of  $m_r F m_l$  depends on the depths of the 3D points and increases proportionally with an increasing baseline.

**Keywords:** Stereo vision, Epipolar geometry, Essential matrix, Fundamental matrix, Eight-point algorithm

### 1. Introduction

During the 1980s and 1990s, a central problem in computer vision involved inferring three-dimensional information from views (images) taken from two viewpoints. In solving such a problem, researchers succeeded in defining the *epipolar geometry* that describes the relationship between 3D points and their projections onto 2D images.

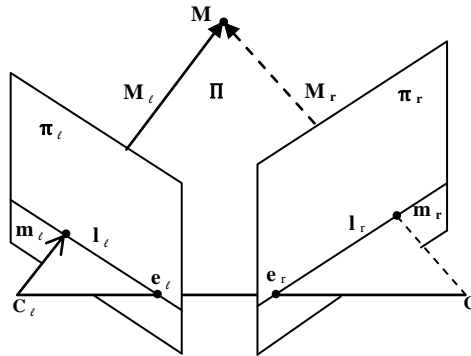
Cameras used to acquire these images are characterised by intrinsic and extrinsic parameters. The intrinsic parameters include the principal points' coordinates, the pixel aspect ratio, and focal lengths, whereas the extrinsic parameters include the position and orientation of the camera with respect to the world coordinate system.

In the classical method, the intrinsic camera parameters are known, and this knowledge is used to determine the *essential matrix*  $E$  and calculate the epipolar geometry. 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.

The epipolar geometry might be depicted as in Figure 1 and can be described 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  $c_r$ . Points  $c_l$  and  $c_r$  constitute the origins of the coordinate systems of the two cameras. The point  $m_l = (x_l, y_l)$  is the

retinal image of point  $M$  acquired by the left camera; it belongs to the left camera plane  $\pi_l$  and is defined in the left camera coordinate system. Similarly, 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  $\pi_r$  and it is defined in the right camera coordinate system. The points  $m_l$  and  $m_r$  are called *corresponding points*.



**Figure 1. Epipolar Geometry**

The following terminologies are used in epipolar geometry [1]:

1. The *baseline* is the line joining the camera centres.
2. The *epipole* is the intersection of the line joining the camera centres with the image plane. Thus, the epipole is the image in one view of the camera centre of the other view.
3. The *epipolar plane* is a plane containing the baseline.
4. 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.
5. The two camera centres and a given world point define the *epipolar plane*  $\Pi$  (Figure 1).

Longuet-Higgins [2] introduced the essential matrix to the computer vision community and proposed an eight-point algorithm for its calculation. Such a matrix associates the image points of the two views when the intrinsic parameters of the cameras are available. Faugeras [3] developed the fundamental matrix that encapsulates the geometric relations relating a world point to its images. The fundamental matrix is independent of the scene structure [1], *i.e.*, it is independent of the depth of the world points in different parts of the scene.

The relationship between points  $m_l$  and  $m_r$  through the essential matrix is expressed as  $m_r E m_l = 0$  [2]. The relationship between the fundamental matrix and essential matrix is expressed as  $F = K_r E K_l^{-1}$ , where  $K_l$  and  $K_r$  are the calibration matrices of the left and right cameras, respectively [1].

The availability of eight pairs of corresponding points from the two views of the scene ensures the calculation of the fundamental matrix. Once the fundamental matrix is determined, the availability of an image point in one view leads to a determination of its corresponding point in the other view through the fundamental matrix equation

$$m_r F m_l = 0 \quad (1)$$

The eight-point algorithm is frequently cited as a method for computing the fundamental matrix from a set of eight or more point matches [4].

This study reviews epipolar geometry theory and is motivated by potential applications of stereo vision. Some of these applications have limited or zero fault tolerance capability and include face recognition, robot-assisted surgical operations, and unmanned aerial vehicles. This study's contribution is to identify certain limitations of epipolar geometry theory and help avoid unpredicted consequences.

The remainder of this study is organised as follows. Section 2 presents previous work questioning the effectiveness of epipolar geometry. Section 3 presents the experimental results of estimating the fundamental matrix for scenes that include objects with different depths. Sections 4 and 5 analyse, algebraically and geometrically, the effectiveness of the fundamental matrix equation in extracting the shape of 3D scenes. In section 6, we attempt to identify the causes behind the poor performance of fundamental matrix theory. Finally, the study concludes in section 7.

## 2. Previous Work

This section begins by introducing Longuet-Higgins' method of deriving the essential matrix [2]. A world point is defined in the left camera coordinate system by  $M_l = (X_l, Y_l, Z_l)$  and in the right camera coordinate system by  $M_r = (X_r, Y_r, Z_r)$ .

Longuet-Higgins [2] defined the image points  $m_l$  and  $m_r$  of the world point  $M$  in the coordinate systems of the two cameras as

$$\begin{cases} (x_l, y_l) = (X_l/Z_l, Y_l/Z_l) \\ (x_r, y_r) = (X_r/Z_r, Y_r/Z_r) \end{cases} \quad (2)$$

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

$$\mathbf{M}_r = R(\mathbf{M}_l - \mathbf{t}) \quad (3)$$

The rotation  $R$  satisfies the relation

$$RR^T = R^T R = 1 \text{ and } \det(R) = 1 \quad (4)$$

Longuet-Higgins [2] defines the essential matrix as

$$E = RS \quad (5)$$

where  $S$  is the skew-symmetric matrix

$$S = \begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix} \quad (6)$$

and adopted the length of the vector  $\mathbf{t}$  as the unit of distance

$$t^2 = t_x^2 + t_y^2 + t_z^2 = 1 \quad (7)$$

Longuet-Higgins [2] then constructed the expression  $\mathbf{M}_r^T E \mathbf{M}_l$  and used (3) to (7) to conclude  $\mathbf{M}_r^T E \mathbf{M}_l = 0$ . Dividing by  $Z_l Z_r$  establishes the equation for the essential matrix that relates image points  $m_l$  and  $m_r$ ,

$$m_r^T E m_l = 0 \quad (8)$$

Three years later, Longuet-Higgins discovered configurations that defeat the eight-point algorithm [5]. Similarly, the work of Hartley [4] in defense of the eight-point algorithm is the result of criticising the performance of the eight-point algorithm as extremely susceptible to noise and virtually useless for most purposes. By preceding the algorithm with a very simple normalisation of the coordinates of the matched points, Hartley [4] challenged such criticisms and presented results that are comparable to the best iterative algorithms, claiming that the improved performance is justified by theory and verified by extensive experiments on real images.

Two years after introducing the fundamental matrix, Luong and Faugeras [6] discovered that the general methods to compute the fundamental matrix are unstable when the points lie near planes.

Zisserman, *et al.*, [7] demonstrated that it is not possible to recover the epipolar geometry for a number of configurations.

Marill [8] gave an example that should, at a minimum, cause scientists to consider recovering the three-dimensional scene that caused the image not as an axiom (as considered in vision literature) but as a theory that is subject to empirical verification or falsification. Marill further argued that if it held up under further examination, the example would be evidence that the theory is false.

Most of these reviews attribute the failure of the fundamental matrix theory to the performance of the eight-point algorithm.

### 3. Empirical Analysis

The main objective of the empirical analysis in the current study is to demonstrate the effect of the depth of 3D points on the value of the expression  $m_r F m_l$ . To that end, we used three real images of a scene acquired from three different positions (view 1, view 2, and view 3). The distinguishing feature of these images is that the baseline of (view 1, view 2) is smaller than the baseline of (view 1, view 3).

Figures 2, 3, and 4 show that the scene is composed of two building blocks that are positioned one behind the other. The three views show that the depth of the front block is smaller than the depth of the rear block.

In the current study, two experimental cases are considered. The first case consists of view 1 and view 2. The second case consists of view 1 and view 3, where the baseline is larger than in the former case. Four sets of data are created from these two cases. Each set contains the coordinates of points from a pair of views.

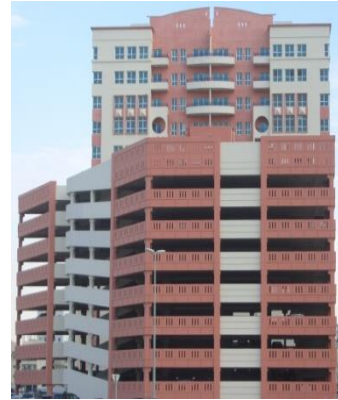
1. V21Rear contains coordinates of points from the rear block of the pair (view2, view1)
2. V31Rear contains coordinates of points from the rear block of the pair (view3, view1)
3. V21Mix contains coordinates of points from both blocks of the pair (view2, view1)
4. V31Mix contains coordinates of points from both blocks of the pair (view3, view1)



**Figure 2. View1 of the scene**



**Figure 3. View2 of the scene**



**Figure 4. View3 of the scene**

In experiments involving the current work, we used the eight-point algorithm to calculate the fundamental matrix. This algorithm is a frequently cited method for estimating the fundamental matrix from a set of eight or more point matches [4].

When assessing estimation methods for computing the fundamental matrix, scholars generally consider the following:

1. The mean and standard deviation of the discrepancy between points and epipolar lines;
2. Detecting outliers; and
3. Satisfying one of the characteristics of  $F$ , where  $\text{rank}(F)$  is equal to two [9].

The main focus of our experimental study is to check whether the fundamental matrix equation,  $m_r F m_l = 0$ , holds for any pair of corresponding points  $(m_l, m_r)$ , as claimed in [1- 4, 10-12].

To calculate the fundamental matrix, we selected points on the two views, first from the rear block and from both blocks in the second phase. To evaluate the effect of the depth of points—and the baseline—on the value of the expression  $m_r F m_l$ , we computed  $m_r F m_l$  for each pair of corresponding points in Tables 1 and 2. In these tables, points 1 to 10 are arbitrarily selected from the rear block, and points 11 to 20 are selected from the front block.

The program we used is coded in MATLAB and consists of the following components:

1. The Harris corner detection algorithm [13] to detect points of the views.
2. A piece of code to read the points detected by the Harris algorithm, which generates a matrix of four columns: two contain the coordinates of the right image points and two contain the coordinates of the left image points.
3. The MATLAB Toolbox [14] includes many different methods for estimating the fundamental matrix using the eight-point algorithm [9].

After generating the coordinates of corresponding points from the two views, we manually removed points that have no corresponding points in the opposite view. We run the program using the robust estimation method functions because they outperform the linear and iterative methods [9]. These robust methods include the following:

1. M-Estimator using least squares;
2. M-Estimator using least squares with an Eigen analysis;
3. M-Estimator proposed by Torr;

4. LMedS using least squares;
5. LMedS using least squares with an Eigen analysis;
6. RANSAC;
7. MLESAC implemented by Torr; and
8. MAPSAC implemented by Torr [9].

Each of these methods estimates a value for the matrix  $F$ . Tables 1 and 2 list the corresponding points from the two views that are used to calculate the values of  $m_r, Fm_l$  for each  $F$  obtained by an estimation method.

**Table 1. Selected Pairs of Points from View2-view1 to Evaluate  $m_r, Fm_l$**

point#	view2	view1	area
1	(618,313)	(833,331)	Rear
2	(636,369)	(853,389)	
3	(681,420)	(901,443)	
4	(952,448)	(1184,471)	
5	(953,341)	(1183,360)	
6	(1076,181)	(1306,193)	
7	(1078,555)	(1319,582)	
8	(1303,526)	(1551,548)	
9	(1577,263)	(1833,278)	
10	(1581,474)	(1841,497)	
11	(1010,925)	(1490,961)	Front
12	(1116,1035)	(1606,1079)	
13	(1145,1061)	(1637,1105)	
14	(1214,1197)	(1715,1248)	
15	(533,942)	(907,983)	
16	(392,905)	(727,948)	
17	(656,1014)	(1068,1114)	
18	(1470,1022)	(1983,1062)	
19	(1502,886)	(2016,915)	
20	(863,890)	(1328,924)	

**Table 2. Selected Pairs of Points from View3-view1 to Evaluate  $m_r, Fm_l$**

point#	view3	view1	area
1	(783,298)	(833,331)	Rear
2	(800,353)	(853,389)	
3	(844,403)	(901,443)	
4	(1114,429)	(1184,471)	
5	(1116,322)	(1183,360)	
6	(1241,160)	(1306,193)	
7	(1239,535)	(1319,582)	
8	(1465,504)	(1551,548)	
9	(1742,236)	(1833,278)	
10	(1746,449)	(1841,497)	
11	(1041,897)	(1490,961)	Front
12	(1146,1007)	(1606,1079)	

13	(1175,1033)	(1637,1105)
14	(1243,1170)	(1715,1248)
15	(607,917)	(907,983)
16	(484,882)	(727,948)
17	(712,1039)	(1068,1114)
18	(1503,993)	(1983,1062)
19	(1536,855)	(2016,915)
20	(897,862)	(1328,924)

First, in each case—(view2-view1) and (view3-view1)—we used nearly 450 pairs of corresponding points from the rear block to estimate the fundamental matrix. We then calculated  $m_r, Fm_l$  for the points found in Tables 1 and 2, and the results are shown in Tables 3- 6.

In the second phase, we combined points from both the rear and front blocks to estimate the fundamental matrix. We then calculated  $m_r, Fm_l$  for the points of Table 1 and Table 2. The results are shown in Tables 7-10. The latter phase is a response to requirements that specify that pairs of points used to estimate the fundamental matrix should be selected from all over the scene. Such a claim has no foundation in the theory of the fundamental matrix [2, 3], which specifically asserts that eight pairs of corresponding points are adequate to calculate the fundamental matrix without specifying from which region of the scene they come.

Along with the values of  $m_r, Fm_l$  presented in the tables mentioned above, Figures 5, 6, 7, and 8 present the mean and standard deviation of the discrepancy between points and epipolar lines for each case.

**Table 3. Values of  $m_r, Fm_l$  where the Corresponding Points are from the Rear Block of View2-view1 and the Fundamental Matrix is Estimated from the Rear Block**

Points Methods	1	2	3	4	5	6	7	8	9	10
1	0.012	0.004	0.004	0.025	0.021	0.002	0.031	0.015	0.016	0.012
2	0.008	0.002	0.005	0.015	0.010	0.010	0.025	0.009	0.008	0.005
3	0.007	0.001	0.005	0.015	0.010	0.006	0.028	0.010	0.007	0.005
4	0.013	0.005	0.004	0.029	0.025	0.001	0.034	0.017	0.019	0.015
5	0.008	0.002	0.005	0.015	0.010	0.011	0.024	0.008	0.009	0.006
6	0.008	0.002	0.005	0.015	0.010	0.011	0.024	0.008	0.009	0.006
7	0.007	0.002	0.004	0.014	0.010	0.005	0.026	0.009	0.007	0.005
8	0.007	0.002	0.005	0.015	0.010	0.005	0.027	0.009	0.008	0.007

**Table 4. Values of  $m_r Fm_l$  where the Corresponding Points are from the front Block of View2–view1 and the Fundamental Matrix is Estimated from the Rear Block**

Points Methods	11	12	13	14	15	16	17	18	19	20
1	0.946	0.925	0.916	0.876	0.565	0.417	0.682	0.917	0.957	0.923
2	1.124	1.128	1.126	1.125	0.720	0.560	0.867	1.135	1.152	1.092
3	1.161	1.175	1.175	1.190	0.753	0.588	0.911	1.180	1.185	1.125
4	0.887	0.858	0.847	0.796	0.519	0.375	0.626	0.842	0.888	0.867
5	1.110	1.110	1.107	1.099	0.705	0.547	0.848	1.118	1.139	1.079
6	1.110	1.110	1.107	1.099	0.705	0.547	0.848	1.118	1.139	1.079
7	1.172	1.187	1.188	1.204	0.752	0.584	0.913	1.197	1.202	1.134
8	1.164	1.178	1.178	1.194	0.752	0.586	0.911	1.183	1.187	1.128

**Table 5. Values of  $m_r Fm_l$  where the Corresponding Points are from the Rear Block of View3–view1 and the fundamental Matrix is Estimated from the Rear Block**

Points Methods	1	2	3	4	5	6	7	8	9	10
1	0.044	0.014	0.022	0.105	0.086	0.027	0.133	0.074	0.056	0.055
2	0.066	0.012	0.041	0.103	0.054	0.085	0.168	0.073	0.054	0.046
3	0.067	0.012	0.046	0.112	0.060	0.064	0.193	0.086	0.046	0.049
4	0.041	0.015	0.020	0.105	0.090	0.043	0.129	0.074	0.056	0.058
5	0.066	0.012	0.041	0.101	0.053	0.092	0.164	0.068	0.051	0.048
6	0.066	0.012	0.041	0.101	0.053	0.092	0.164	0.068	0.051	0.048
7	0.070	0.011	0.048	0.107	0.051	0.077	0.189	0.078	0.035	0.048
8	0.059	0.010	0.038	0.084	0.048	0.005	0.120	0.043	0.038	0.046

**Table 6. Values of  $m_r Fm_l$  where the Corresponding Points are from the front Block of View3–view1 and the Fundamental Matrix is Estimated from the Rear Block**

Points Methods	11	12	13	14	15	16	17	18	19	20
1	3.465	3.444	3.428	3.372	2.148	1.607	2.605	3.361	3.418	3.370
2	7.686	7.704	7.687	7.668	5.031	3.968	6.009	7.664	7.747	7.490
3	8.304	8.392	8.392	8.481	5.511	4.365	6.631	8.330	8.333	8.079
4	2.924	2.902	2.887	2.833	1.787	1.313	2.180	2.809	2.858	2.843
5	8.063	8.070	8.050	8.014	5.264	4.151	6.278	8.041	8.144	7.858
6	8.063	8.070	8.050	8.014	5.264	4.151	6.278	8.041	8.144	7.858
7	8.346	8.427	8.424	8.506	5.589	4.455	6.707	8.344	8.354	8.130
8	4.027	4.018	4.001	3.986	2.863	2.339	3.398	3.800	3.834	3.970



**Table 7. Values of  $m_r Fm_l$ , where the Corresponding Points are from the Rear Block of View2-view1 and the Fundamental Matrix is Estimated from both Blocks**

Points Methods	1	2	3	4	5	6	7	8	9	10
1	0.136	0.060	0.011	0.082	0.009	0.148	0.143	0.096	0.074	0.002
2	0.981	0.256	0.374	0.499	0.327	1.337	0.845	0.363	0.211	0.324
3	0.108	0.011	0.060	0.054	0.112	0.441	0.169	0.139	0.194	0.123
4	0.189	0.055	0.062	0.151	0.050	0.426	0.289	0.236	0.246	0.107
5	0.049	0.019	0.010	0.030	0.016	0.097	0.052	0.060	0.061	0.026
6	0.111	0.048	0.002	0.040	0.075	0.315	0.137	0.133	0.173	0.110
7	0.030	0.006	0.009	0.002	0.035	0.088	0.005	0.018	0.019	0.015
8	0.138	0.035	0.059	0.111	0.097	0.421	0.331	0.280	0.250	0.201

**Table 8. Values of  $m_r Fm_l$ , where the Corresponding Points are from the front Block of View2-view1 and the fundamental Matrix is Estimated from both Blocks**

Points Methods	11	12	13	14	15	16	17	18	19	20
1	0.039	0.035	0.062	0.182	0.246	0.262	0.237	0.298	0.262	0.098
2	2.705	2.580	2.630	2.501	2.746	3.934	3.066	6.010	6.804	1.935
3	0.383	0.258	0.224	0.005	0.468	0.532	0.258	0.278	0.423	0.430
4	0.021	0.113	0.146	0.352	0.173	0.221	0.125	0.287	0.230	0.026
5	0.340	0.220	0.192	0.069	0.196	0.161	0.058	0.259	0.414	0.372
6	0.367	0.280	0.259	0.091	0.188	0.159	0.041	0.388	0.506	0.362
7	0.275	0.185	0.161	0.058	0.075	0.015	0.071	0.224	0.337	0.285
8	1.615	1.874	1.934	2.246	1.374	1.216	1.654	1.950	1.661	1.501

**Table 9. Values of  $m_r Fm_l$ , where the Corresponding Points are from the Rear Block of View3-view1 and the fundamental Matrix is Estimated from both Blocks**

Points Methods	1	2	3	4	5	6	7	8	9	10
1	0.062	0.017	0.005	0.025	0.002	0.205	0.033	0.027	0.079	0.027
2	0.046	0.011	0.002	0.005	0.010	0.195	0.099	0.082	0.059	0.058
3	0.447	0.277	0.130	0.112	0.449	0.927	0.196	0.122	0.587	0.041
4	0.081	0.037	0.007	0.075	0.023	0.093	0.110	0.077	0.107	0.018
5	0.269	0.099	0.054	0.124	0.213	0.701	0.453	0.402	0.357	0.321
6	0.269	0.099	0.054	0.124	0.213	0.701	0.453	0.402	0.357	0.321
7	0.014	0.008	0.001	0.029	0.068	0.241	0.184	0.276	0.096	0.378
8	0.329	0.130	0.059	0.186	0.271	0.980	0.656	0.589	0.786	0.371

**Table 10. Values of  $m_r Fm_l$ , where the Corresponding Points are from the front Block of View3–view1 and the Fundamental Matrix is Estimated from both Blocks**

Points Methods	11	12	13	14	15	16	17	18	19	20
1	0.837	0.350	0.220	0.543	0.148	0.029	0.176	0.375	0.973	0.930
2	1.067	0.450	0.286	0.672	0.156	0.000	0.248	0.502	1.266	1.184
3	1.732	1.439	1.367	0.986	0.340	0.070	0.409	1.560	1.985	1.706
4	0.354	0.248	0.217	0.039	0.157	0.074	0.135	0.154	0.262	0.369
5	0.683	0.348	0.266	0.144	0.218	0.455	0.290	0.362	0.802	0.730
6	0.683	0.348	0.266	0.144	0.218	0.455	0.290	0.362	0.802	0.730
7	0.365	0.077	0.007	0.344	0.528	0.509	0.481	0.496	0.242	0.545
8	1.787	1.211	1.061	0.363	0.728	0.432	0.656	0.980	1.718	1.950

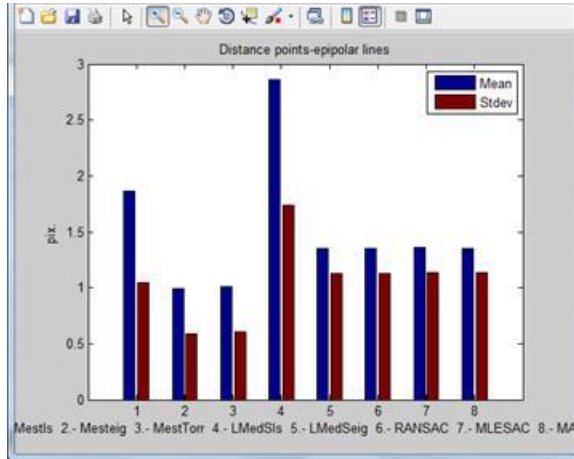
The values of  $m_r Fm_l$  in Table 3 are close to zero because the baseline in this case is quite small, and because the depths of the world points whose images are used to estimate the fundamental matrix and the points used in computing  $m_r Fm_l$  are roughly similar. The values of  $m_r Fm_l$  in Table 5 are ten times larger than the values in Table 3 because of the effect of the increased baseline, although the depth of the points used to estimate  $F$  and the points used to evaluate  $m_r Fm_l$  are in the same range. The increased baseline deepens the differences in depth.

The values of  $m_r Fm_l$  in Table 4 are in the range 0.4 to 1.19 because the points used to estimate the fundamental matrix are from the rear block, whereas the points used to compute  $m_r Fm_l$  are from the front block, and their depths are different.

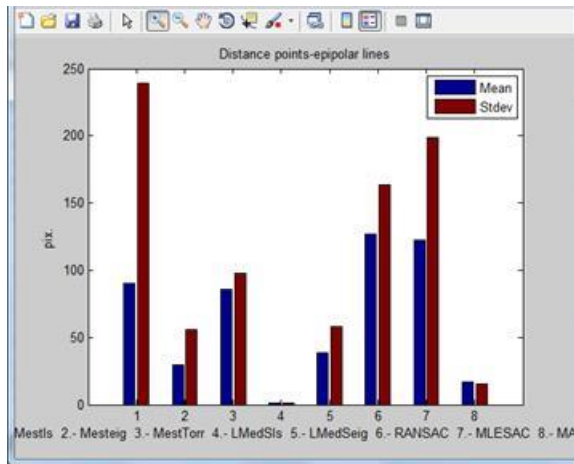
Similarly, the values in Table 6 are approximately 100 times the values in Table 5 because of the differences in the depths of the points.

Comparing Table 4 to Table 6, it appears that the latter values are roughly ten times the magnitude of the former, which is also due to the effect of increasing the baseline.

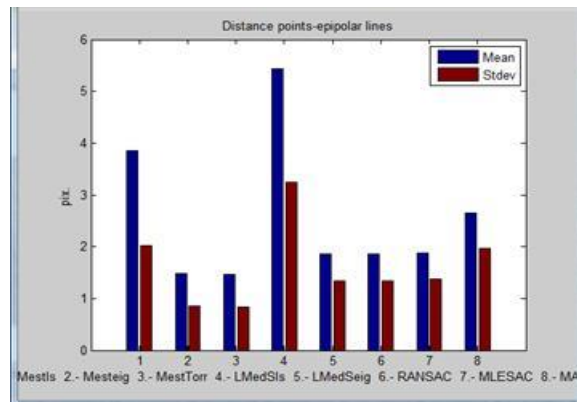
The values of  $m_r Fm_l$  are calculated from the same set of points in both Table 7 and Table 9, and on the same set of points in Table 8 and Table 10. However, the baseline in Table 7 and in Table 8 is smaller than the baseline in Table 9 and in Table 10. Comparing the values in Table 7 to Table 9, and the values in Table 8 to Table 10, it appears the values in the tables with smaller baselines are smaller than the values in the Tables with larger baselines.



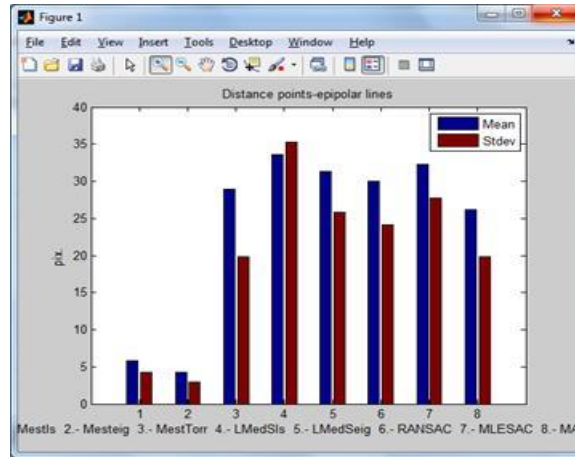
**Figure 5. The Mean and Standard Deviation of the Discrepancy between Points and Epipolar Lines of View2-view1; the Fundamental Matrix is Estimated from the Points of the Rear Block**



**Figure 6. The Mean and Standard Deviation of the Discrepancy between Points and Epipolar Lines of view2-view1; the Fundamental Matrix is Estimated from the Points of both Blocks**



**Figure 7. The Mean and Standard Deviation of the Discrepancy between Points and Epipolar Lines of View3-view1; the Fundamental Matrix is Estimated from the Points of the Rear Block**



**Figure 8. The Mean and Standard Deviation of the Discrepancy between Points and Epipolar Lines of View3-view1; the Fundamental Matrix is Estimated from the Points of both Blocks**

**Table 11. Outliers Computed Following [9]. In V21Rear and V31Rear  $F$  is Estimated from the Rear Block and in V21Mix and V31Mix  $F$  is Estimated from both Blocks**

Views Methods	V21Rear	V31Rear	V21Mix	V31Mix
1	46/442	27/453	130/672	191/690
2	67/442	51/453	215/672	236/690
3	61/442	58/453	98/672	229/690
4	0/442	0/453	230/672	237/690
5	0/442	0/453	144/672	142/690
6	0/442	0/453	137/672	46/690
7	14/442	25/453	257/672	222/690
8	1/442	39/453	236/672	172/690

The mean and standard deviation of the distances between points and epipolar lines in all the cases considered hereunder are shown in Figures 5-8.

In Figures 5 and 7, where the depths of all points involved are within a small range, the mean and standard deviation in most methods are bounded by 2. In Figures 6 and 8, where the corresponding points used to estimate  $F$  belong to two different categories regarding their depth; the mean and standard deviation are more than 20 for most methods and are above 100 for some methods.

Although all points involved in estimating the matrix  $F$  are given along with their correspondences, Table 11 indicates that the number of outliers increased dramatically for most methods when these points are selected from regions with different depths.

## 4. Algebraic Analysis of the Fundamental Matrix Equation

For each point  $m_l$  in the left retina, the epipolar constraint indicates that its corresponding point  $m_r$  on the right view lies on its epipolar line  $l_r$  [6], which reduces the search for the corresponding point from an entire image to merely a line.

Equation (1) is linear and homogeneous in the nine unknown coefficients of matrix  $F$ . Thus we know that, in general, if we are given eight matches we will be able to determine a unique solution for  $F$  [3].

The extraction of the 3D shape of the scene under study requires selecting a set of points from one view and computing their corresponding points in the other view. Once the fundamental matrix and an image point in one view are available, equation (1) generates its corresponding point in the other view. Let us ascertain whether (1) is adequate to determine the corresponding points.

Given an image point  $m_l = (x_l, y_l, 1)$  from the left view, substituting in (1) yields:

$$\begin{bmatrix} x_r & y_r & 1 \end{bmatrix} \begin{bmatrix} f_1 & f_2 & f_3 \\ f_4 & f_5 & f_6 \\ f_7 & f_8 & f_9 \end{bmatrix} \begin{bmatrix} x_l \\ y_l \\ 1 \end{bmatrix} = 0 \quad (9)$$

Equation (9) is equivalent to  $ax_r + by_r + c = 0$

where

$$a = x_l f_1 + y_l f_2 + f_3, \quad b = x_l f_4 + y_l f_5 + f_6, \quad \text{and} \quad c = x_l f_7 + y_l f_8 + f_9 \quad (10)$$

Equation (10) is a single equation in two variables,  $x_r$  and  $y_r$ , and has many solutions; for any value  $x_r$ , there is a value  $y_r$ .

Algebraically, knowledge of the fundamental matrix and an image point in one view generates unlimited corresponding points in the other view that satisfy (1). Thus, knowledge of the fundamental matrix does not ensure that an accurate 3D shape of the scene under study will be extracted.

This finding appraises the worth of the epipolar constraint. Reducing the search space of each corresponding point to a single line is not a major achievement, considering that a line may contain more than one corresponding point.

## 5. Geometric Analysis of the Fundamental Matrix Equation

### 5.1. Projective Points

By definition, a projective point  $[x, y, z]$  is considered to be equal to an equivalence class of 3D points that belong to the 3D line passing through the Cartesian point  $(X, Y, Z)$  and the origin  $(0,0,0)$  [15].

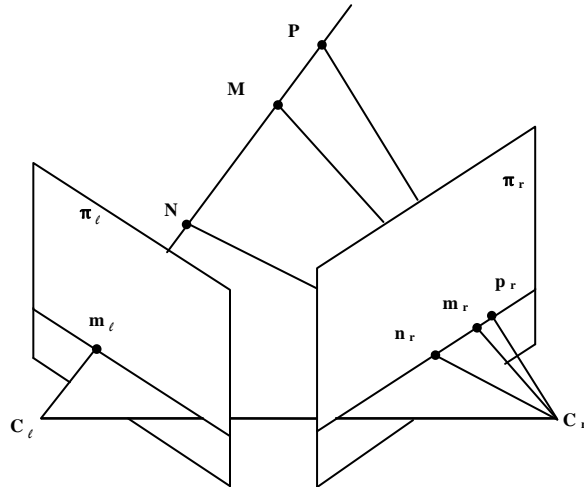
The projective point  $m_l$  in Figure 9 is equal to the class of 3D points  $\{N, M, P\}$ .

These 3D points are all projected (*i.e.*, imaged) onto the point  $m_l$  in the left camera plane. Simultaneously, these points are projected onto different points in the right camera plane. Thus, the following equalities hold simultaneously:

$$\begin{cases} m_l F n_r = 0 \\ m_l F m_r = 0 \\ m_l F p_r = 0 \end{cases}$$

Therefore, the knowledge of point  $m_l$  does not ensure the computation of its corresponding point. Although researchers and practitioners have named this

phenomenon the occlusion problem, *they do not exclude any type of scene and they do not report any conditions for occlusion when they set the epipolar geometry theory.*



**Figure 9. 3D Points Projected Onto One Point in the Left Camera Plane and Onto Many Points in the Right Camera Plane**

### 5.2. Coplanar World Points

Let us consider two world points,  $M$  and  $N$ , on the same epipolar plane  $\Pi$ , as in Figure 10.

Hartley and Zisserman [1] assert: “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 centres. 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  $l_r = e_r \times m_r = [e_r]_x m_r$ . Since  $m_r$  may be written as  $m_r = H_\pi m_l$ , we have  $l_r = [e_r]_x H_\pi m_l = F m_l$  where we define

$F = [e_r]_x H_\pi$  as the fundamental matrix. The point  $m_r$  lies on the epipolar line  $l_r = Fm_l, m_r Fm_l = 0$ ."

In Figure 10, the points  $n_l$  and  $n_r$  are the image points of world point  $N$ .

By definition, an epipolar line is the intersection of an epipolar plane with the image plane. The points  $N$  and  $M$  belong to the same epipolar plane  $\Pi$ . This epipolar plane intersects the left and right image planes in the epipolar lines  $l_l$  and  $l_r$ , respectively, and it follows that the points  $n_r$  and  $m_r$  lie on the same epipolar line,  $l_r = [e_r]_x H_\pi m_l = Fm_l$ . Simultaneously, we thus have  $m_r Fm_l = 0$ ,  $n_r Fm_l = 0$ ,  $n_r F n_l = 0$ , and  $m_r F n_l = 0$ .

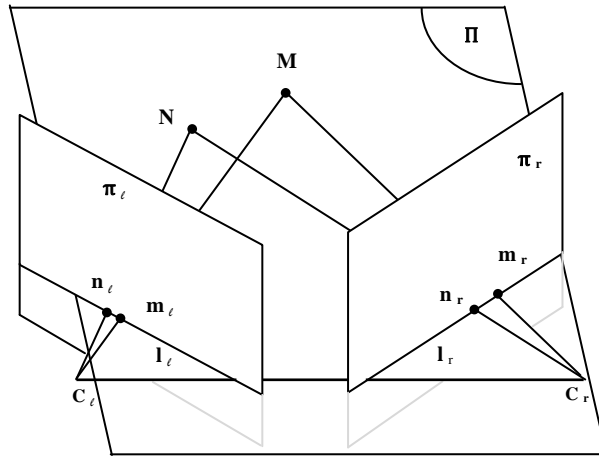


Figure 10. World Points on the Same Epipolar Plane

## 6. Pointing Out the Causes

The secret behind the poor performance of the estimation methods is not a flaw in the eight-point algorithm as some researchers anticipated.

Longuet-Higgins [2] divided  $M_r^T E M_l = 0$  by  $Z_l Z_r$  and used the relationships between the world points  $M$  and their projective counterparts  $m = M/Z$  to conclude the essential matrix equation  $m_r^T E m_l = 0$ .

1. By definition, a projective point  $m$  is considered to be equal to an equivalence class of 3D points that belong to the 3D line passing through points  $M$  and the origin of the coordinate system [15], excluding the origin. There is no one-to-one correspondence between world points  $M$  and projective points  $m$ , i.e.,  $M_r^T E M_l = 0$  implies  $m_r^T E m_l = 0$ , but  $m_r^T E m_l = 0$  does not imply  $M_r^T E M_l = 0$  because  $Z_l$  and  $Z_r$  are missing.
2. An epipolar plane intersects the left and right image planes in epipolar lines and defines the correspondence between the lines. Let  $M^i, i = 1, \dots, n$  be a set of world points belonging to the same epipolar plane. As we concluded in the previous section from the assertion in [1], the images of these 3D points  $(m_l^i, m_r^j), i = 1, \dots, n, j = 1, \dots, n$  satisfy  $m_r^j F m_l^i = 0$ .

Given a sufficient number of point matches (at least eight), the equation  $m_r^T F m_l = 0$  can be used to compute the unknown matrix  $F$  [4].

In particular, writing the image points in homogeneous coordinates  $m_l = (x_l, y_l, 1)$  and  $m_r = (x_r, y_r, 1)$ , each point match gives rise to one linear equation in the unknown entries of  $F$ . The coefficients of this equation are easily written in terms of the known coordinates  $m_l$  and  $m_r$

$$\begin{aligned} x_l x_r f_{11} + x_l y_r f_{21} + x_l f_{31} + y_l x_r f_{21} + y_l y_r f_{22} \\ + y_l f_{32} + x_r f_{13} + y_r f_{23} + f_{33} = 0 \end{aligned} \quad (11)$$

From all the point matches, we obtain a set of linear equations of the form

$$A f = 0 \quad (12)$$

where  $f$  is a nine-vector containing the entries of the matrix  $F$ , and  $A$  is the equation matrix.

It is possible to find a solution to the system (12) with as few as eight point matches. With more than eight point matches, we have an over specified system of equations [4].

Let us now consider eight world points,  $M_l^i = (X_l^i, Y_l^i, Z_l^i)$ ,  $i = 1, \dots, 8$ , in the left camera system; their coordinates in the right system are  $M_r^i = (X_r^i, Y_r^i, Z_r^i)$ ,  $i = 1, \dots, 8$ , and their image points in the left and right coordinate systems will be  $m_l^i = (x_l^i, y_l^i, 1)$ ,  $i = 1, \dots, 8$  and  $m_r^i = (x_r^i, y_r^i, 1)$ ,  $i = 1, \dots, 8$ , respectively.

Assuming that  $M_l^1$  and  $M_l^2$  are on the same epipolar plane,  $M_l^3$  is on the same arrow coming from the left camera centre toward point  $M_l^4$  (i.e.,  $m_l^3 = m_l^4$ ).

In this case, we have a set of exactly eight equations:

$$\mathbf{M}_r^i F \mathbf{M}_l^i = 0, \quad i = 1, \dots, 8 \quad (13)$$

However,  $m_r^i F m_l^i = 0$  will be equivalent to a set of ten equations were the single point  $m_l^3$  is related to two different points  $m_r^3$  and  $m_r^4$ :

$$\begin{cases} m_r^1 F m_l^1 = 0, & m_r^2 F m_l^2 = 0, & m_r^1 F m_l^2 = 0, & m_r^2 F m_l^1 = 0, \\ m_r^3 F m_l^3 = 0, & m_r^4 F m_l^3 = 0, \\ m_r^i F m_l^i = 0, & i = 5, \dots, 8 \end{cases} \quad (14)$$

The question now is whether the matrix  $F$ , calculated from (14), is the same as  $F$  obtained from (13). The answer is no; whereas  $F$  calculated from (13) is a one-to-one correspondence between the vectors representing the 3D point  $M$  in the left and right coordinate systems,  $F$  calculated from (14) is no longer only a relationship between pairs of corresponding points; it is also a relationship between other points that are not correspondents.

The equation  $\mathbf{M}_r^T E \mathbf{M}_l = 0$  is defined in the domain  $D = \{(M_r, M_l)\}$  where  $M_r$  is the vector representative of the point  $M$  in the right coordinate system and  $M_l$  is the vector representative of the point in the left coordinate system. However,  $m_r^T E m_l = 0$  is defined in a different domain  $d = \{(m_r, m_l)\}$  where there is no one-to-one correspondence between  $(M_r, M_l)$  and  $(m_r, m_l)$ .



When the 3D point's depths are different, it will frequently be the case that many points will belong to the same epipolar plane and, therefore,  $F$  will be a loose relationship between correspondent and non-correspondent points. This finding confirms that  $F$  is dependent on the scene structure, i.e., it depends on the depth of the world points.

To sum up, a 3D point  $M$  is represented in the two cameras coordinate systems by exactly two vectors  $M_r$  and  $M_l$ . The matrix  $F$  is calculated from the equation  $m_r^T F m_l = 0$  where the pair of points  $(m_r, m_l)$  is available and can be in one of the following cases:

1.  $m_r$  and  $m_l$  are corresponding points; the ideal case,
2.  $m_l$  is a correspondent to more than a point  $m_r$ ; occlusion in the left image,
3.  $m_r$  is a correspondent to more than a point  $m_l$ ; occlusion in the right image,
4.  $m_r$  and  $m_l$  are not corresponding points at all; they are images of different 3D points belonging to the same epipolar plane.

It is noteworthy to remember that one of the main uses of the fundamental matrix is the extraction of the 3D structure of the scene through the recovery of the 3D points. However, the matrix  $F$  calculated from  $m_r^T F m_l = 0$  does not always satisfy  $M_r^T F M_l = 0$ ; in other words, the availability of a pair of corresponding points  $(m_r, m_l)$  and  $F$  does not guarantee the recovery of the right 3D point.

## 7. Conclusions

Epipolar geometry consists of the relationships between the image points in two views of a scene. Simultaneously, it relates a world point from the scene to its pair of image points in the two views. The fundamental matrix  $F$  is the algebraic representation of such geometry. In addition, it is claimed that matrix  $F$  is independent of the scene structure [1].

The experimental results of the current study demonstrate that  $F$ , in fact, depends on the depth of the 3D points of the scene. Algebraic and geometric analyses emphasized that the fundamental matrix equation does not define a one-to-one correspondence between the corresponding points of a 3D point.

The fundamental matrix equation is linear and can be solved straightforwardly; however, it is misleading to assert that having knowledge of  $F$  and an image point in one view ensure that its corresponding image point can be determined in the other view.

An image point is not a scalar value; instead, it is a pair of two coordinates,  $m = (x, y)$ , i.e., a pair of scalars. Regardless of whether the matrix  $F$  is singular, the knowledge of one image point,  $m_l$ , and the matrix  $F$ ,  $m_r^T F m_l = 0$  is nothing more than a single linear equation in two unknowns,  $x_r$  and  $y_r$ . Therefore,  $m_r^T F m_l = 0$  holds for many values of  $x_r$  and  $y_r$ , i.e., for every point  $m_l$ ,  $m_r^T F m_l = 0$  is satisfied for more than one point  $m_r$ .

Geometrically, 3D points that lie on the same ray from one camera centre are projected to the same image point on the plane of such camera. The same 3D points are projected to different image points on the other camera plane. Furthermore, 3D points that belong to the same epipolar plane are projected to the left on the same left epipolar

line and projected to the right on the same right epipolar line. The right epipolar line is defined by any of the left image points,  $p_l$ , as  $Fp_l$ . All the right image points,  $p_r$ , lie on the same epipolar line  $Fp_l$ . Thus, all  $p_r$  satisfy  $p_r^T Fp_l = 0$  for all  $p_l$ , where  $p_l$  and  $p_r$  are the image points of any 3D point belonging on the same epipolar plane. Similarly all the left image points,  $p_l$  lie on the same epipolar line  $Fp_r$ . Thus, all  $p_l$  satisfy  $p_r^T Fp_l = 0$  for all  $p_r$ , where  $p_l$  and  $p_r$  are the image points of any 3D point belonging on the same epipolar plane.

On the one hand, the matrix  $F$  in  $M_r^T F M_l = 0$  is a one-to-one correspondence between the vectors  $M_r$  and  $M_l$ , which are the right and left vectors coordinates of the 3D point  $M$  in the right and left coordinate systems, respectively. However, the matrix  $F$  in  $m_r^T F m_l = 0$  is a relationship between image points on the two views that are not necessarily correspondents of one another.

The experiment reported in this study indicates that the matrix  $F$ , which is calculated from the equation  $m_r^T F m_l = 0$ , is not independent of the scene structure. In particular, such experiment is not an isolated case; instead, it represents a class of scenes that contain objects at many different depths.

## References

- [1] R. Hartley and A. Zisserman, "Multiple View Geometry in Computer Vision", Second Edition, Cambridge University Press, (2004).
- [2] H. C. Longuet-Higgins, "A Computer Program to Reconstruct a Scene From Two Projections", Nature, vol. 293, (1981), pp. 133-135.
- [3] 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, Springer-Verlag, Santa Margherita Ligure, Italy, (1992), pp. 563-578.
- [4] R. Hartley, "In Defense of the Eight-Point algorithm", IEEE Tran. on Pattern Analysis and Machine Intelligence, vol. 19, no. 6, (1997), pp. 580-593.
- [5] H. C. Longuet-Higgins, "The Reconstruction of a Scene from Two Projections-Configurations that Defeat the 8-Point Algorithm", Proc. 1st Conf. Artificial Intelligence Applications, (1984), pp. 395-397.
- [6] Q. T. Luong and O. D. Faugeras, "Determining the Fundamental Matrix with Planes: Instability and New Algorithms", IEEE Computer Society Conference on Computer Vision and Pattern Recognition, (CVPR '93), New York, (1993), pp. 489-494.
- [7] A. Zisserman and S. J. Maybank, "A Case Against Epipolar Geometry", In 2nd Europe-U.S. Workshop on Invariance, Ponta Delgada, Azores, (1993), pp. 69-88.
- [8] T. Marill, "A Counterexample to the Theory That Vision Recovers Three-Dimensional Scenes", A. I. Working Paper, MIT Artificial Intelligence Laboratory, vol. 319, (1988).
- [9] S. Ivezic, 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 Centred Communication, John Wiley & Sons, (2005).
- [10] X. Armangué and J. Salvi, "Overall View Regarding Fundamental Matrix Estimation", Image and Vision Computing vol. 21, no. 2, (2003), pp. 205-220.
- [11] E. Trucco and A. Verri, "Introductory Techniques for 3-D Computer Vision", Prentice Hall, (1998).
- [12] Z. Y. Zhang, "Determining The Epipolar Geometry And Its Uncertainty - A Review", International Journal of Computer Vision, vol. 27 no. 2, (1998), pp. 161-195.
- [13] C. Harris and M. Stephens, "A Combined Corner and Edge Detector", Proc. of the 4th Alvey Vision Conference, (1988), pp. 147-151.
- [14] J. Salvi, "Fundamental Matrix Estimation toolbox", (2003), <http://eia.udg.es/~qsalvi/recerca.html>.
- [15] "Wikis", Homogeneous Coordinates, (2010) [http://www.thefullwiki.org/Homogenous\\_coordinates#wikipedia\\_Addition\\_of\\_homogeneous\\_coordinates](http://www.thefullwiki.org/Homogenous_coordinates#wikipedia_Addition_of_homogeneous_coordinates).

## Authors



**Tayeb Basta**, received a BSc degree in Computer Science from University of Annaba (Algeria), and a PhD degree in Computer Science from University of Manchester (UK) in 1983 and 1994. Tayeb is now an associate professor at Al Ghurair University in Dubai, UAE.

