



Relative Orientation of Stereo-Pair Using 2D Projective Singular Correlation: Applications on Video Images and Scanned-Aerial Photographs

Gamal H. Seedahmed

*Department of Surveying Engineering, Faculty of Engineering, University of Khartoum
Khartoum, Sudan (E-mail: Gamal.Seedahmed@gmail.com)*

Abstract: Although the mathematical model of 2D projective singular correlation for relative orientation was introduced to the photogrammetric community more than fifty years ago, it does not penetrate the deep practice of analytical and digital photogrammetry. On the other hand, this model is widely accepted and used in the computer vision community, which is known as the fundamental matrix approach for relative orientation. The attractiveness of this model stems from two facts. First, it is a linear model. Second, it does not require any prior knowledge about the camera parameters. In other words, it can handle uncalibrated cameras. This paper offers a fresh look for the use of 2D projective singular correlation in the relative orientation of stereo-pair. Moreover, this paper argues the case for the use of this model in the daily practice of photogrammetry; and this is through the practical demonstration on rich data sets. In particular, this paper provides critical evaluations and new insights for this model in terms of its computational procedures, least-squares random errors modeling, external accuracy checking, and practical issues of implementation. For example, it presents a general formulation for external accuracy checking that uses the epipolar distance as a metric for the quality of the solution. This model is tested on three stereo-pairs that were obtained from handheld video camera, aerial video camera, and scanned aerial photographs. In all tests, subpixel accuracy was achieved from external checking of the average distances between the epipolar-lines and their corresponding conjugate points.

Keywords: *Relative Orientation; Stereo-Pair; 2D Projective Singular Correlation; Fundamental Matrix; Video Images; Scanned-Aerial Photographs.*

1. INTRODUCTION

Relative orientation of stereo-pair is a fundamental problem in photogrammetry. It is the mechanism behind 3D visualization of stereo-pairs; 3D point coordinates computation in the model-space; and the reduction of the search space for image matching; and this is through the epipolar-line geometry [1]. Computationally, the relative orientation of a stereo-pair can be accomplished by the knowledge of five conjugate points, which will lead to the determination of five parameters. These parameters could be a combination of rotation angles from the stereo-pair (independent relative orientation) or a combination of rotation angles and translation parameters of a single image and holding the parameters of the other image fixed (dependent relative orientation). Mathematically, the relative orientation problem can be solved through the collinearity model or the coplanarity condition equation or model. Both models are non-linear and require initial approximations for their unknown parameters. The coplanarity model is superior to the collinearity model in terms of eliminating the need for initial approximations for the 3D point coordinates in the model-space. In other words, the 3D point coordinates in the model-space are not part of the unknown parameters in the

coplanarity model. Approximate linear models for the relative orientation can be obtained for near-vertical images by rearrangement of the collinearity model [2]. In general, the problem of relative orientation is considered solved; and therefore it received little attention in modern photogrammetric literature. Yes, it is solved for the aerial photographic mapping, but this is not the case for close range photogrammetric applications. The main debate here refers to the practical issue of the non-linearity of the photogrammetric models and the difficulties associated with obtaining the values of initial approximation for the configurations of images in close range applications. In aerial mapping, this non-linearity was solved by near-vertical image configuration and later by GPS-INS (GPS stands for Global Positioning System and INS stands for Inertial Navigation System) integration. On the other hand, this non-linearity remains to be an issue in close-range applications, which will dominate the practice of photogrammetry in the near future [3]. Moreover, new digital sensors such as video cameras require flexible and generic models for image orientation; otherwise the tapped values in their images cannot be easily exploited. Therefore, the issue of non-linearity should receive more attention to push the photogrammetric practice into new fronts.

The 2D projective singular correlation or relative linear transform was introduced to the photogrammetric community in 1959 by Thompson [4]. Nevertheless, it does not penetrate the deep practice of analytical and digital photogrammetry. This is may be explained by the following reasons. Historically, optical-mechanical stereo-plotting instruments were used to measure photo-coordinates, trace contour lines, and draw planimetric maps. Those instruments were large, expensive, bulky, and required well trained personnel. Let us put this reasoning or argument in a more general context. Prior to the availability of digital imaging and computer software, performing such tasks as image coordinates measurement was not easily accessible a wider community of general users. Clearly, the basis of computational or analytical photogrammetry methods, such as 2D projective singular correlation or the fundamental matrix, lie in the acceptance and dominance of digital sensor technologies for image acquisition and the availability of easy to develop software for image processing and implementation of photogrammetric algorithms. By now, digital technology is the mainstream in the photogrammetric development and applications. Nevertheless, modern textbooks of photogrammetry slightly touch the relative orientation through the fundamental matrix [2-5]. On the other hand, the fifth edition of the manual of photogrammetry provides a thorough explanation and analysis for the fundamental matrix and its application to relative orientation [6]. However, there is an objection against the use of the fundamental matrix for relative orientation [7,8]. This objection is mainly due to the inability of the fundamental matrix approach to estimate the parameters of the relative orientation from coplanar points. It is argued in this paper that there is a wide range of applications in which the coplanar points can be avoided and this objection should not hampered or restrict the usefulness of the fundamental matrix for relative orientation. In fact, a similar argument was held against the Direct Linear Transformation (DLT) since it cannot handle coplanar points, but a remedy to this problem was found [9]. The underlying ideas of the fundamental matrix were introduced to the computer vision communities by Longuet-Higgins in 1981 [10]. And since then this approach for relative orientation was subjected to intensive and extensive research, which was addressed in several textbooks [11-13].

The attractiveness of 2D projective singular correlation or the fundamental matrix approach stems from two main facts. First, it is a linear model. Indeed, there are non-linear versions of this model that were proposed to improve its accuracy [11], but this non-linearity is not an issue since there are good methods that can be used to obtain the initial approximations. As is well known, the non-linearity of the photogrammetric orientation procedures is one of the major issues that slow the acceptance of photogrammetry as a generic measurement technology [3]. Second, it does not require any prior knowledge of the camera calibration parameters or the interior orientation of the metric camera. In other words, it can handle uncalibrated cameras and cameras with unknown interior orientation parameters such as video images and scanned aerial photographs with unknown or missing camera calibration certificate.

This paper offers a fresh look for the use of 2D projective singular correlation in the relative orientation of stereo-pair. Moreover, it argues the case for the use of this model in the daily practice of photogrammetry; and this is through the practical demonstration on rich data sets. In particular, this paper provides critical evaluations for this model in terms of its computational procedures, least-squares random errors modeling, external accuracy checking, and practical issues of implementation such as the representation of the image coordinates. This model is tested on three stereo-pairs that were obtained from handheld video camera, aerial video camera, and scanned aerial photographs.

This paper is organized as follows. The next section provides the geometric principles and the mathematical formulation of the fundamental matrix and its direct connection with the classical representation of relative orientation in photogrammetry using the coplanarity model. Then followed by the data sets and results and analysis sections. The last section concludes the paper.

2. GEOMETRIC PRINCIPLES AND MATHEMATICAL FORMULATION

This section presents the geometric principles and the mathematical formulation of the 2D projective singular correlation or the fundamental matrix. In particular, it exploits the direct connection between the fundamental matrix and the classical representation of relative orientation in photogrammetry using the coplanarity model.

Fig.1 shows the geometry of relative orientation between two images. Suppose that a point A in the 3D object space is imaged in two images and at two locations a in the left image and a' in the right image. The pair (a, a') defines conjugate points. As shown in Fig. 1, the conjugate points, the object space point A, and the camera centers (C1, C2) are coplanar. Mathematically, this relationship is the basis of the coplanarity model; and it can be defined by the following equation:

$$\vec{b} \cdot (\vec{a}_1 \times \vec{a}_2) = 0 \quad (1)$$

where:

\vec{b} : is the vector between by C1 and C2. This vector is called the baseline vector.

\vec{a}_1 : is the vector that connects the image point a and the camera perspective center in the left image.

\vec{a}_2 : is the vector that connects the image point a' and the camera perspective center in the right image.

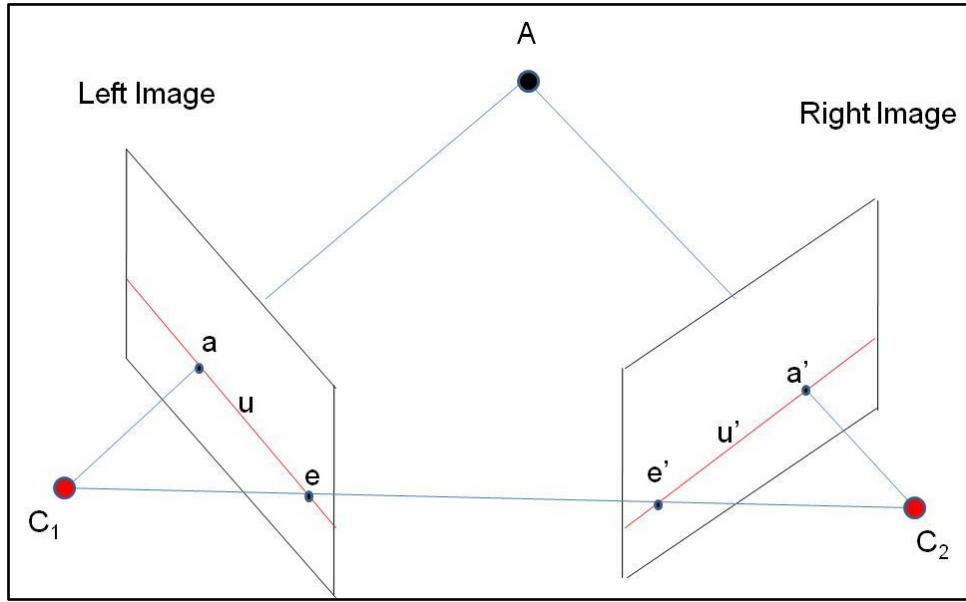


Fig. 1. Geometric principles of coplanarity condition equation.

[1] Before delving further, let us define the geometric entities of the relative orientation as illustrated in Fig. 1:

- The epipole is the point of intersection of the line joining the two camera centers, (C1, C2), with the image plane. e and e' are the epipoles of the left and right image respectively (see Fig. 1).
- An epipolar plane is a plane containing the baseline. For example, point A and the two camera centers (C1, C2) are defining an epipolar plane (see Fig. 1).
- An epipolar line is the intersection of an epipolar plane with the image plane. All epipolar lines intersect at the epipole. u and u' (red lines in the left and right images in Fig. 1 are epipolar lines.

By using matrix notation, the content of the vectors shown in equation (1) can be written explicitly as follows:

$$\vec{b} = \begin{bmatrix} b_x \\ b_y \\ b_z \end{bmatrix} \quad (2)$$

$$\vec{a}_1 = R_1^T \begin{bmatrix} x_1 - x_p \\ y_1 - y_p \\ -f \end{bmatrix} = R_1^T \underbrace{\begin{bmatrix} 1 & 0 & -x_p \\ 0 & 1 & -y_p \\ 0 & 0 & -f \end{bmatrix}}_K \begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix} = R_1^T K \begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix} \quad (3)$$

$$\vec{a}_2 = R_2^T \begin{bmatrix} x_2 - x_p \\ y_2 - y_p \\ -f \end{bmatrix} = R_2^T \underbrace{\begin{bmatrix} 1 & 0 & -x_p \\ 0 & 1 & -y_p \\ 0 & 0 & -f \end{bmatrix}}_K \begin{bmatrix} x_2 \\ y_2 \\ 1 \end{bmatrix} = R_2^T K \begin{bmatrix} x_2 \\ y_2 \\ 1 \end{bmatrix} \quad (4)$$

where:

K: is known as the calibration matrix.

f: is the focal length of the camera.

x_p and y_p: are the coordinates of the principal point of the camera.

R₁ and R₂: are the rotation matrices for the left and the right image shown in Fig. 1.

x₁, y₁: image coordinates in the left image.

x₂, y₂: image coordinates in the right image.

By inserting equations (2), (3), and (4) in equation (1) and after rearrangement, we get the coplanarity condition equation or model:

$$\begin{bmatrix} x_1 & y_1 & 1 \end{bmatrix} R_1 K^T \begin{bmatrix} 0 & b_z & -b_y \\ -b_z & 0 & b_x \\ b_y & -b_x & 0 \end{bmatrix} R_2^T K \begin{bmatrix} x_2 \\ y_2 \\ 1 \end{bmatrix} = 0 \quad (5)$$

Now, we are at the position to define the singular correlation matrix for the 2D projective transformation or the fundamental matrix F as follows:

$$F = R_1 K^T \underbrace{\begin{bmatrix} 0 & b_z & -b_y \\ -b_z & 0 & b_x \\ b_y & -b_x & 0 \end{bmatrix}}_{|b|=0} R_2^T K = \begin{bmatrix} f_1 & f_2 & f_3 \\ f_4 & f_5 & f_6 \\ f_7 & f_8 & f_9 \end{bmatrix} \quad (6)$$

Notice that the b matrix is a skew-symmetric matrix since its determinant is equal to zero and $b^T = -b$ [14]. In fact, equation (6) provides an analytical proof for equation (1). This property (skew-symmetric) of the b matrix will induce the rank or the singularity constraint, which will be defined and exploited in equation (14). In light of equation (6), equation (5) can be written compactly as follows:

$$\begin{bmatrix} x_1 & y_1 & 1 \end{bmatrix} F \begin{bmatrix} x_2 \\ y_2 \\ 1 \end{bmatrix} = \begin{bmatrix} x_1 & y_1 & 1 \end{bmatrix} \begin{bmatrix} f_1 & f_2 & f_3 \\ f_4 & f_5 & f_6 \\ f_7 & f_8 & f_9 \end{bmatrix} \begin{bmatrix} x_2 \\ y_2 \\ 1 \end{bmatrix} = 0 \quad (7)$$

Therefore, equation (7) is another way to parameterize the coplanarity condition equation or the epipolar constraint. This constraint is encoded by the 3 x 3 matrix F . This parameterization brings new properties for equation (6) that will be discussed in the sequel of this section. Equation (7) is a homogenous equation and is typically scaled by setting one of its unknown equal to 1 (here: $f_9 = 1$). More importantly, equation (7) is linear relationship for the coplanarity condition equation or the epipolar constraint. Terms-wise, the observation equation of this relationship can be written as follows:

$$x_{2i}x_{1i}f_1 + x_{2i}y_{1i}f_2 + x_{2i}f_3 + y_{2i}x_{1i}f_4 + y_{2i}y_{1i}f_5 + y_{2i}f_6 + x_{1i}f_7 + y_{1i}f_8 = -1 + e_i \quad (8)$$

where: (x_{1i}, y_{1i}) and (x_{2i}, y_{2i}) are the image coordinates for a conjugate point.

f_1, \dots, f_8 : are the unknown elements (parameters) of the F matrix.

e_i : is the random error associated with each observation equation. Geometrically, this error can be interpreted as a distance between the epipolar-line and one of the conjugate points in one of the images from the stereo-pair.

Each conjugate point will generate one observation equation. Geometrically, each observation equation will generate one epipolar constraint. Therefore, a minimum of 8 conjugate points are required to solve equation (8). Equation (8) is typically called the eight-point algorithm [15]. For best results, the image coordinates in a stereo-pair are normalized to avoid the problem of large numbers during the inversion of the F matrix. Large numbers in the F matrix will be induced by the bilinear coefficients of equation (8). Accordingly, two normalization procedures were used and tested in this paper. In the first one, the image coordinates are reduced to the physical center of the image or the sensor of the camera. In the second one, the image coordinates are reduced to the location of their statistical average.

More than 8 points are typically treated by a least-squares solution to handle the inconsistency of the observations and to estimate the most probable values of the unknowns. The following target function is minimized to find the optimal

solution for the unknown parameters of the fundamental matrix F :

$$e^T P e + 2\lambda_1^T (A\xi - \tau - e) = \min(e, \xi, \lambda_1) \quad (9)$$

where:

e is the vector of the random errors associated with the observation equations.

P : is a weight matrix.

λ_1 : A vector of Lagrange multipliers for the epipolar constraint.

A : is the coefficient or the design matrix of the unknown parameters shown in equation (8).

ξ : is the vector of the unknown parameters shown in equation (8).

τ : A vector of constant values and each value is equal to -1.

The solution of the target function shown in equation (9) will lead to the following set of equations:

$$\hat{\xi} = (A^T P A)^{-1} A^T P \tau \quad (10)$$

$$D\{\hat{\xi}\} = \sigma_o^2 (A^T P A)^{-1} \quad (11)$$

$$\tilde{e} = \tau - A \hat{\xi} \quad (12)$$

$$\sigma_o^2 = \frac{\tilde{e}^T P \tilde{e}}{n - r} \quad (13)$$

where:

$\hat{\xi}$: Are the estimated unknown parameters of the F matrix shown in equation (8).

$D\{\hat{\xi}\}$: Dispersion matrix of the unknown parameters.

\tilde{e} : A vector of the predicted residuals.

σ_o^2 : The estimated variance component.

n : Number of observation equation.

r : Number of unknown parameters (here: 8 parameters).

As shown in equation (6), a real non-zero 3 x 3 matrix F is a fundamental matrix if it satisfies the following constraint:

$$\det(F) = |F| = 0 \quad (14)$$

Equation (14) states that the determinant of F should equal to zero, which was shown in equation (6). This is also known as the rank constraint for the fundamental matrix. As stated, this is a direct consequence of the skew symmetric matrix shown in equation (6). Terms-wise the rank constraint can be written as follows:

$$\phi = f_1(f_5 - f_6 f_8) - f_2(f_4 - f_6 f_7) + f_3(f_4 f_8 - f_5 f_7) = 0 \quad (15)$$

The rank constraint shown in equation (15) is imposed on equation (8); and this is after its linearization. Indeed, this constraint will lead to a non-linear least squares estimation solution that requires Taylor's series expansion and initial approximations. Therefore, the target function shown in equation (9) will be modified to include the rank or the singularity constraint as a fixed-constraint during the minimization process as follows:

$$e^T P e + 2\lambda_1^T (A\Delta\xi - \Delta\tau - e) + 2\lambda_2^T (H\Delta\xi - \Delta\kappa_o) = \min(e, \Delta\xi, \lambda_1, \lambda_2) \quad (16)$$

where:

$\Delta\xi$: is the correction vector for the unknown parameters of the fundamental matrix F.

λ_1 : is the vector of Lagrange multiplier for the epipolar constraint.

λ_2 : is the vector of the Lagrange multipliers for the rank constraint.

$\Delta\tau$: is the vector of discrepancies of the epipolar constraints.

H : is a row vector for the partial derivatives of equation (15).

$\Delta\kappa_o$: is the vector of discrepancy for the rank constraint equation.

The provision of initial values or approximations for the non-linear solution shown in equation (16) is not an issue since very good approximations can be obtained by the linear solution shown in equations (8) and (9).

Tsai and Huang [16] proposed another approach to enforce the rank constraint. This approach replaces the matrix F found by equation (8) by a singular matrix F' . This new matrix minimizes the Frobenius norm $\|F-F'\|$. This step can be implemented by using a singular value decomposition and setting the smallest eigen value to zero as follows:

$$F = U\Lambda V^T \quad (17)$$

$$\Lambda = \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \end{bmatrix} \quad (18)$$

where:

U and V are 3 x 3 orthogonal matrices.

Λ : is a diagonal matrix of the eigen values.

$\sigma_1 \geq \sigma_2 \geq \sigma_3 \geq 0$: are the ordered eigen values from higher

to lower. The smallest eigen (σ_3) value is set to zero and the new matrix is:

$$F' = U \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & 0 \end{bmatrix} V^T \quad (19)$$

In general, the 2D projective singular correlation or the fundamental matrix can be formulated as a two-step solution:

- Linear solution using equation (8).
- Constraint enforcement, which is a non-linear solution.

After estimating the elements of the fundamental matrix F, the following quantities can be derived:

- The left and right epipolar lines.
- The left and right epipoles.
- The distances between the left and right epipolar lines and the two coordinate of the conjugate points in the left and right images or a stereo-pair.
- The root-mean-square-errors (rmse) of the check points.

The general equation of the epipolar line is:

$$ax + by + c = 0 \quad (20)$$

Where a, b, and c are the parameters of the straight line shown in equation (20). For a stereo-pair, each conjugate point will generate 2 epipolar lines. One line in the right image and another one in the left image.

The parameters (a, b, c) of the right epipolar lines (ELright) for a specific conjugate point are:

$$EL_{right} = \begin{bmatrix} a_{1i} \\ b_{1i} \\ c_{1i} \end{bmatrix} = F^T \begin{bmatrix} x_{1i} \\ y_{1i} \\ 1 \end{bmatrix} \quad (21)$$

Similarly, the parameters (a, b, c) of the left epipolar lines (ELleft) for a specific conjugate point are:

$$EL_{left} = \begin{bmatrix} a_{2i} \\ b_{2i} \\ c_{2i} \end{bmatrix} = F \begin{bmatrix} x_{2i} \\ y_{2i} \\ 1 \end{bmatrix} \quad (22)$$

left epipole (e_{left}^T), which is the left null-space of the F matrix, can be computed as follows:

$$e_{left}^T F = e_{left}^T U\Lambda V^T = 0 \quad (23)$$

e_{left}^T is the column of U that corresponds to the zero elements of the eigen values.

The right epipole (e_{right}), which is the right null-space of the F matrix, can be computed as follows:

$$F e_{right} = U\Lambda V^T e_{right} = 0 \quad (24)$$

e_{right} is the column of V that corresponds to the zero elements of the eigen values.

The signed distance (d_i) between the epipolar line in the right image and its conjugate point in the same image is:

$$d_i = [x_{2i} \quad y_{2i} \quad 1] F^T \begin{bmatrix} x_{1i} \\ y_{1i} \\ 1 \end{bmatrix} \quad (25)$$

Similar relationship can be written for the left image. In ideal image coordinates measurement, (d_i) should equal to zero, but this is not the case since random noise and the lack of camera calibration will make the (d_i) values deviate from zero. Equation (25) will be part of a measure for external accuracy checking in the sense of root-mean-square errors (rmse). In particular, some of the conjugate points will not be used for estimation of the unknown parameters of the fundamental matrix (let us call them check points); and those points will be substituted into equation (25) to compute the rmse values. Specifically, the equation of the external accuracy checking is:

$$rmse = \sqrt{\frac{\sum_{i=1}^{i=N} d_i^2}{N}} \quad (24)$$

where: N is the number of check points.

Geometrically, equation (26) represents the average distance between the epipolar lines and the conjugate points, which are also, can be interpreted as the rmse for the check points. It is very important to stress that equation (26) provide a common ground to compare the performance of the F matrix under different normalizations and formulations.

In light of equations (9), (16), (19), and (1), the following two key facts can be deduced:

- The linear solution of the fundamental matrix, shown in equation (9), minimizes the distance between the epipolar lines and the conjugate points. In other words, it is an image-based constraint and it does not include the model-space information, shown in equation (1), to the solution.
- The rank constraint solution of the fundamental matrix, shown in equation (16) or (19), minimizes the distance between the epipolar lines and the conjugate points as well as the constraint shown in equation (1), which is a model-space constraint for 3D point intersection. In other words, this solution minimizes simultaneously the image space information and the model-space information shown in equation (1). Therefore, this solution captures the implicit error of 3D point reconstruction or intersection in the model space.

Therefore, the mathematical construct of the target functions shown in equations (9) and (16) gave a very

clear picture for the above deduction and an interesting insight of the constraint and unconstraint solutions.

3. DATA SETS

Three stereo-pairs were used in this study. In all pairs, no camera parameters were utilized. In particular, the following data sets were used:

- A stereo-pair from a handheld video camera (see Fig. 2). Twenty two (22) conjugate points were measured in this pair (see Fig. 3); and four of them were reserved as check points.
- A stereo-pair from an aerial video camera (see Fig. 4). Twenty two (22) conjugate points were measured in this pair (see Fig. 5) ; and four of them were reserved as check points.
- A stereo-pair from an aerial film-based camera (see Fig. 6). This pair is scanned at a resolution of 400 dots per inch (400 dpi). Twenty six (26) conjugate points were measured in this pair (see Fig. 7) ; and four of them were reserved as check points.

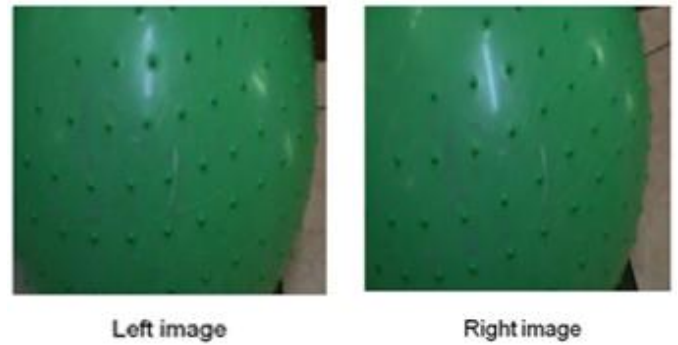


Fig. 2. A stereo-pair from a handheld video camera (an image of a plastic ball).

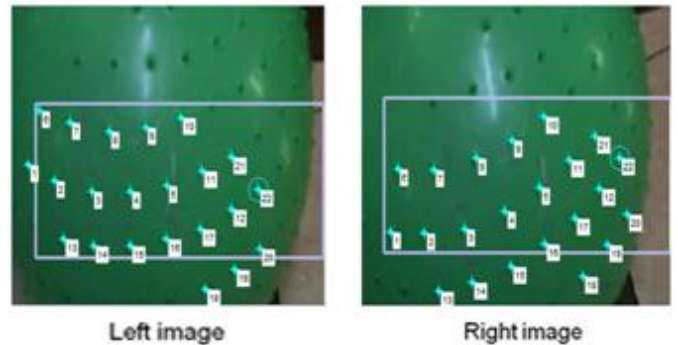


Fig. 3. Conjugate points from the stereo-pair of the handheld video camera.



Left image



Right image

Fig. 4. Stereo-pair from an aerial video camera.



Left image

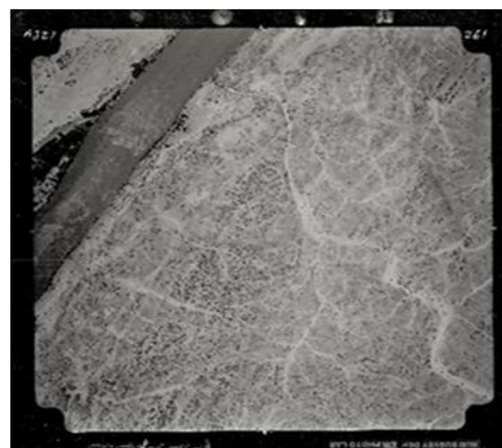


Right image

Fig. 5. Conjugate points from the stereo-pair of the aerial video camera.



Left Image



Right Image

Fig. 6. Stereo-pair from scanned aerial photographs

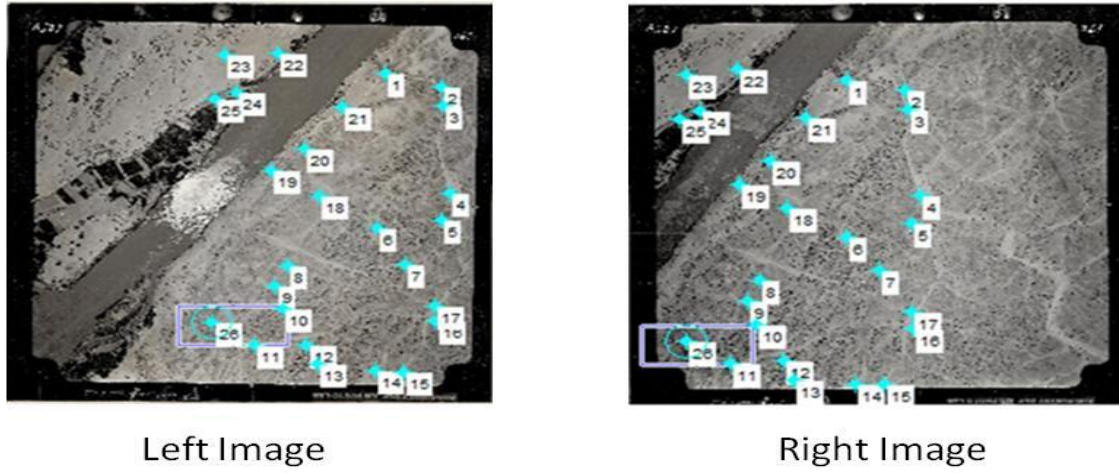


Fig. 7. Conjugate points from the stereo-pair of the scanned aerial photographs.

4. RESULTS AND DISCUSSION

MATLAB-based prototype software was developed to implement the presented work in this paper. For each stereo-pair listed in the data sets section, the following types of experiment were conducted:

- Linear estimation of the fundamental matrix (F1) using equation (9). The image coordinates are normalized by reduction to the physical center of the image. This experiment will be called case no. 1.
- Non-linear estimation of the fundamental matrix (F2) using equation (16). The image coordinates are normalized by reduction to the physical center of the image. This experiment will be called case no. 2.
- Estimation of the fundamental matrix (F3) using Frobenius norm shown in equation (19). The image coordinates are normalized by reduction to the physical

center of the image. This experiment will be called case no. 3.

- Estimation of the fundamental matrix (F4) using Frobenius norm using equation (19), but the image coordinates are normalized by reduction to the statistical average of the image coordinates. This experiment will be called case no. 4.
- In all cases, four (4) conjugate points were reserved for external accuracy checking in the form of rmse.

Table 1 shows the image coordinates of the conjugate points (22 points) that were measured from the stereo-pair shown in Fig. 2 (handheld video images). The locations of these conjugate points were shown in Fig. 3. These coordinates were measured with the left upper corner of the image as an origin. This origin was transformed or normalized; and this is in accordance to the specifications of the four cases listed before.

Table 1: Coordinates for the left and right images from the handheld video camera.

Point ID	x ₁ (pixels)	y ₁ (pixels)	x ₂ (pixels)	y ₂ (pixels)	Point ID	x ₁ (pixels)	y ₁ (pixels)	x ₂ (pixels)	y ₂ (pixels)
1	36	310	69	435	12	507	394	553	352
2	97	339	147	438	13	117	451	180	549
3	183	361	241	431	14	188	465	258	532
4	272	361	334	396	15	273	465	348	502
5	358	348	415	351	16	355	451	425	459
6	62	205	86	315	17	435	432	495	409
7	133	229	169	317	18	447	547	515	520
8	223	245	267	295	19	514	509	571	460
9	309	237	350	261	20	573	470	616	402
10	390	218	423	215	21	507	294	543	252
11	439	321	483	297	22	569	357	599	295

The fundamental matrices of the four cases for the stereo-pair that was obtained from the handheld camera are listed below. The first two fundamental matrices (F_1 and F_2) that were estimated from the linear solution of the fundamental matrix shown in equation (9) and the constrained one depicted by equation (16) are identical. On the other hand, there are some changes in the elements of the third fundamental matrix (F_3), which was obtained by the singular value decomposition. For example, $F_3(2,2)$ is less than the corresponding elements in F_1 and F_2 by an order of magnitude. The fourth fundamental matrix (F_4) is completely different from the other three ones in terms of the numerical values since it uses a different normalization procedure for the image coordinates.

$$F_1 = \begin{bmatrix} -8.1538\text{e-}006 & -3.9317\text{e-}006 & -0.020961 \\ 2.4277\text{e-}006 & -3.0736\text{e-}006 & 0.0025608 \\ 0.018953 & -0.0057545 & 1 \end{bmatrix}$$

$$F_2 = \begin{bmatrix} -8.1538\text{e-}006 & -3.9317\text{e-}006 & -0.020961 \\ 2.4277\text{e-}006 & -3.0736\text{e-}006 & 0.0025608 \\ 0.018953 & -0.0057545 & 1 \end{bmatrix}$$

$$F_3 = \begin{bmatrix} -8.0511\text{e-}006 & -3.6097\text{e-}006 & -0.020961 \\ 3.3122\text{e-}006 & -3.0225\text{e-}007 & 0.0025608 \\ 0.018953 & -0.0057545 & 1 \end{bmatrix}$$

$$F_4 = \begin{bmatrix} -2.3196\text{e-}005 & -2.1702\text{e-}005 & -0.090445 \\ 2.5329\text{e-}005 & 4.2155\text{e-}006 & -0.0048565 \\ 0.08546 & 0.017492 & 1 \end{bmatrix}$$

Table 2 shows the derived information from the four (4) fundamental matrices of the handheld video images. From this Table, the following observations can be made:

- The left and right coordinates of the epipoles for the first three cases are identical. On the other hand, these coordinates are different for the fourth case and this is due to the use of different normalization procedure for the image coordinates.

- The variance components $\hat{\sigma}_o^2$ of the first three cases are less by an order of magnitude than the fourth case. This will suggest that the normalization of the image coordinates by the physical center of the image is better than the normalization by the reduction to the statistical average of the image coordinates.
- The rmse for the first two cases are identical and this is can be explained by the equality of their fundamental matrices. In other words, the determinant constraint does not change the values of the elements of fundamental matrix for the second case.
- The rmse for the third case is bigger by an order of magnitude than the first two cases. In other words, the Frobenius norm increases the rmse. More precisely, there are intersection inaccuracies or errors that were brought implicitly by conjugate points; and this inaccuracy is captured by the increase in the rmse.
- The rmse of the fourth case is bigger than the three previous cases. This large value can be explained by two arguments. First, there is intersection inaccuracy. Second, there is a bias that was introduced by the normalization procedure of the image coordinates (here: reduction to the statistical average of the image coordinates).
- Regardless of the normalization procedures, sub-pixel accuracy was achieved in all cases. In other words, the *rmse* for the four cases is less than one pixel (check the second column in Table 2).

Table 3 shows the image coordinates of the conjugate points (22 points) that were measured from the stereo-pair shown in Fig. 4 (aerial video images). The locations of these conjugate points were shown in Fig. 5. These coordinates were measured with the left upper corner of the image as an origin. This origin was transformed or normalized; and this is in accordance to the specifications of the four cases listed before.

Table 2: The derived information from the four fundamental matrices for the handheld video images

Case no.	rmse (pixels)	$\hat{\sigma}_o^2$ (pxels ²)	Left epipole (pixels)	Right epipole (pixels)
1	0.0761	0.000498	$[0.11541 \ 0.99332 \ -0.0001245]^T$	$[-0.30406 \ -0.95265 \ 0.0002808]^T$
2	0.0761	0.000442	$[0.11541 \ 0.99332 \ -0.0001245]^T$	$[-0.30406 \ -0.95265 \ 0.0002808]^T$
3	0.13069	0.000498	$[0.11541 \ 0.99332 \ -0.0001245]^T$	$[-0.30406 \ -0.95265 \ 0.0002808]^T$
4	0.40453	0.00869	$[0.057051 \ -0.99837 \ 0.00031138]^T$	$[-0.19845 \ 0.98011 \ -0.00018428]^T$

Table 3: Coordinates for the left and right images from the aerial video camera.

Point ID	x ₁ (pixels)	y ₁ (pixels)	x ₂ (pixels)	y ₂ (pixels)	Point ID	x ₁ (pixels)	y ₁ (pixels)	x ₂ (pixels)	y ₂ (pixels)
1	1000	372	903	494	12	1370	612	1294	810
2	1212	478	1124	634	13	1228	616	1127	811
3	1478	612	1414	815	14	1550	688	1490	909
4	751	481	603	623	15	1429	789	1345	1036
5	1066	787	922	1028	16	1341	741	1245	975
6	1025	127	954	199	17	1111	762	977	995
7	1619	310	1593	425	18	1069	721	938	941
8	1448	171	1414	258	19	1235	784	1116	1034
9	958	777	799	1019	20	1055	677	924	879
10	578	640	398	818	21	1059	270	978	368
11	1619	655	1571	874	22	1301	191	1253	279

The four fundamental matrices for the stereo-pair that were obtained from the aerial camera and for the four cases are listed below. The first three fundamental matrices (F1, F2, and F3) that were obtained from the linear solution of the fundamental matrix shown in equation (9), the constrained one depicted by equation (16), and Frobenius norm solution that was obtained from equation (19) respectively are not exactly identical, but they are very close to each other. The fourth fundamental matrix (F4) is completely different from the other three ones in terms of the numerical values since it uses a different normalization procedure for the image coordinates.

$$F_1 = \begin{bmatrix} -2.8865e-007 & 1.0465e-005 & 0.00041448 \\ -9.2257e-006 & -5.2757e-007 & 0.0057605 \\ -0.0018254 & -0.0061675 & 1 \end{bmatrix}$$

$$F_2 = \begin{bmatrix} -2.7836e-007 & 1.0465e-005 & 0.00041448 \\ -9.2295e-006 & -5.2743e-007 & 0.0057605 \\ -0.0018254 & -0.0061675 & 1 \end{bmatrix}$$

$$F_3 = \begin{bmatrix} -2.7836e-007 & 1.0465e-005 & 0.00041448 \\ -9.2295e-006 & -5.2743e-007 & 0.0057605 \\ -0.0018254 & -0.0061675 & 1 \end{bmatrix}$$

$$F_4 = \begin{bmatrix} 2.3034e-006 & 0.00020352 & 0.056695 \\ -0.00018535 & 4.4502e-006 & 0.079801 \\ -0.067425 & -0.097294 & 1 \end{bmatrix}$$

Table 4 shows the derived information from the fundamental matrix for the aerial video images. From this Table the following observations can be made:

- The left and right coordinates of the epipoles for the first three cases are identical. On the other hand, these coordinates are different for the fourth case and this is

due to the use of different normalization procedure for the image coordinates.

- The variance components $\hat{\sigma}_o^2$ of the first three cases are less by an order of magnitude than the fourth case. This suggests that the normalization of the image coordinates by the physical center of the image is better than the normalization by the reduction to the statistical average of the image coordinates.
- The rmse for the first two cases are identical and this can be explained by the equality of their fundamental matrices. In other words, the determinant constraint does not change the values of the elements of fundamental matrix for the second case.
- The rmse for the third case is slightly bigger than the first two cases. In other words, the Frobenius norm made a minor change in the rmse value. More precisely, there are slight intersection inaccuracies that could be induced by errors in the conjugate points identification; and this inaccuracy is captured by the slight increase in the rmse.
- The rmse of the fourth case is bigger than the three previous cases. This large value can be explained, once again, by two arguments. First, there is intersection inaccuracy, which is very slight as shown in the previous analysis. Second, there is a little bit large bias that was introduced by the normalization procedure of the image coordinates normalization procedure (here: reduction to the statistical average of the image coordinates).
- Regardless of the normalization procedures, sub-pixel accuracy was achieved in all cases. In other words, the rmse for the four cases is less than one pixel (check the second column in Table 4).

Table 5 shows the image coordinates of the conjugate points (26 points) that were measured from the stereo-pair shown in Fig. 6 (scanned aerial images). The locations of these conjugate points were shown in Fig. 7. These coordinates were measured with the left upper corner of the image as an origin. This origin was transformed or normalized; and this in accordance to the specifications of the four cases listed before.

Table 4: The derived information from the four fundamental matrices for the aerial video images

Case no.	Rmse (pixels)	$\hat{\sigma}_o^2$ (pixels ²)	Left epipole (pixels)			Right epipole (Pixels)		
1	0.01917	0.000238	[-0.93724	0.34867	-0.001620] ^T	[0.99932	-0.036702	0.0015978] ^T
2	0.01917	0.000212	[-0.93724	0.34867	-0.001620] ^T	[0.99932	-0.036702	0.0015978] ^T
3	0.01938	0.000238	[-0.93724	0.34867	-0.001620] ^T	[0.99932	-0.036702	0.0015978] ^T
4	0.21768	0.30615	[0.80528	-0.59289	0.0016573] ^T	[-0.83125	0.5559	-0.0019617] ^T

Table 5: Coordinates for the left and right images from the scanned aerial images.

Point ID	x ₁ (pixels)	y ₁ (pixels)	x ₂ (pixels)	y ₂ (pixels)	Point ID	x ₁ (pixels)	y ₁ (pixels)	x ₂ (pixels)	y ₂ (pixels)
1	905	423	1607	373	14	942	1918	1568	1870
2	1135	469	1838	441	15	1055	1918	1679	1877
3	1149	564	1846	541	16	1175	1646	1813	1622
4	1201	990	1872	978	17	1169	1567	1808	1546
5	1167	1124	1837	1112	18	666	1050	1335	994
6	905	1197	1572	1153	19	473	938	1140	865
7	1034	1360	1690	1333	20	598	819	1274	754
8	553	1406	1207	1343	21	742	605	1431	546
9	509	1513	1157	1446	22	468	368	1166	272
10	543	1628	1189	1562	23	263	393	949	278
11	437	1820	1073	1742	24	318	570	998	470
12	649	1805	1286	1742	25	236	613	910	508
13	696	1898	1329	1834	26	263	1712	900	1624

The four fundamental matrices for the stereo-pair that was obtained from the scanned aerial images and for the four cases are listed below. The first three fundamental matrices (F1, F2, and F3) that were obtained from the linear solution of the fundamental matrix shown in equation (9), the constrained one depicted by equation (16), and Frobenius norm solution that was obtained from equation (19) respectively are not exactly identical, but they are very close to each other. The fourth fundamental matrix (F4) is completely different from the other three ones in terms of the numerical values since it uses a different normalization procedure for the image coordinates.

$$F_1 = \begin{bmatrix} 7.9133\text{e-}008 & -1.6195\text{e-}006 & -0.0015142 \\ 1.6667\text{e-}006 & 7.5003\text{e-}008 & 0.00019774 \\ 0.0014498 & 0.00094755 & 1 \end{bmatrix}$$

$$F_2 = \begin{bmatrix} 7.9133\text{e-}008 & -1.6195\text{e-}006 & -0.0015142 \\ 1.6667\text{e-}006 & 7.5003\text{e-}008 & 0.00019774 \\ 0.0014498 & 0.00094755 & 1 \end{bmatrix}$$

$$F_3 = \begin{bmatrix} 7.9122\text{e-}008 & -1.6196\text{e-}006 & -0.0015142 \\ 1.6667\text{e-}006 & 7.5206\text{e-}008 & 0.00019774 \\ 0.0014498 & 0.00094755 & 1 \end{bmatrix}$$

$$F_4 = \begin{bmatrix} -1.4375\text{e-}005 & -0.00017148 & -0.039252 \\ 0.0001691 & -1.046\text{e-}005 & -0.092587 \\ 0.045349 & 0.091438 & 1 \end{bmatrix}$$

Table 6 shows the derived information from the fundamental matrix for the scanned aerial images. From this Table the following observations can be made:

- The left and right coordinates of the epipoles for the first three cases are identical. On the other hand, these coordinates are different for the fourth case and this is due to the normalization procedure of the image coordinates.

- The variance components $\hat{\sigma}_o^2$ of the first three cases are less by four orders of magnitude than the fourth case. This will suggest that the normalization of the image coordinates by the physical center of the image is far better than the normalization of the image coordinates by the reduction to the statistical average of the image coordinates. Therefore, proper normalization for the image coordinates is very critical for aerial images.

- The rmse for the first two cases are identical and this is can be explained by the equality of their fundamental matrices. In other words, the determinant constraint does not change the values of the elements of fundamental matrix for the second case.

Table 6: The derived information from the fundamental matrix for the scanned aerial images

No.	Rmse (pixels)	$\hat{\sigma}_o^2$ (pixel ²)	Left epipole (pixels)			Right epipole (pixels)		
1	0.020279	5.8065e-006	[-0.51875	0.85493	-0.00095454] ^T	[-0.081005	-0.99671	0.0010619] ^T
2	0.020279	5.3598e-006	[-0.51875	0.85493	-0.00095454] ^T	[-0.081005	-0.99671	0.0010619] ^T
3	0.020216	5.8065e-006	[-0.51875	0.85493	-0.00095454] ^T	[-0.081005	-0.99671	0.0010619] ^T
4	0.36374	0.0378	[0.92713	-0.37475	0.0016949] ^T	[0.88846	-0.45895	0.0016752] ^T

- The rmse for the third case is slightly bigger than the first two cases. In other words, the Frobenius norm made a minor change in the rmse value. More precisely, there are slight intersection inaccuracies that were induced by errors in conjugate points identification; and this inaccuracy is captured by the slight increase in the rmse.
- The rmse of the fourth case is bigger than the three previous cases. This large value can be explained, once again, by two arguments. First, there is intersection inaccuracy, which is very slight as shown in the previous analysis. Second, there is a little bit large bias that was introduced by the normalization procedure of the image coordinates (here: reduction to the statistical average of the image coordinates).
- Regardless of the normalization procedures, sub-pixel accuracy was achieved in all cases. In other words, the rmse for the four cases is less than one pixel (check the second column in Table 6).

In all experiments, the first three cases produce identical epipole coordinates. The epipole coordinates for the fourth case are different and this is due to the use of different normalization procedure for the image coordinates (see the last two columns in Tables 2, 4, and 6).

5. CONCLUSIONS

This paper offers a fresh look for the use of 2D projective singular correlation in the relative orientation of stereo-pair. Moreover, this paper argues the case for the use of this model in the daily practice of photogrammetry; and this is through the practical demonstration on rich data sets. In particular, this paper provides critical evaluations for this model in terms of its computational procedures, least-squares random errors modeling, external accuracy checking, and practical issues of implementation. This model is tested on three stereo-pairs that were obtained from handheld video camera, aerial video camera, and scanned aerial photographs. In all tests, subpixel accuracy was obtained from external checking of the average distances between the epipolar-lines and their conjugate points.

This paper presents a general or universal formulation for external accuracy checking that uses the epipolar distance. The formulation is invariant to the constraint or unconstrained solutions as well as the chosen normalization procedure. The mathematical construct of the target functions, used in this research, reveals that the rank constraint solution minimizes simultaneously the image space information as well as model-space information. Practical experiments demonstrate that this solution captures the implicit error of 3D point reconstruction or intersection in the model space. This error is induced by the incorrect identification of the conjugate points in the image space.

It is not clear why the determinant does not impose the rank constraint as did by the singular value decomposition. Future work should investigate this issue.

Indeed, future work should extend this work to automatic relative orientation as well as 3D point reconstruction in the model and object spaces. Moreover, it should investigate the issue of the use of coplanar points in the 2D projective singularity correlation or the fundamental matrix.

REFERENCES

- [1] Schenk, T., 1999. Digital Photogrammetry. Terra Science, 428 pages.
- [2] Kraus, K., 2007. Photogrammetry: Geometry from Images and Laser Scans. 2nd edition. Walter de Gruyter GmbH & Co., 459 pages.
- [3] Seedahmed G., 2008. Autonomous Initialization of Exterior Orientation Parameters using a Collinearity Search-Based Solution. The Photogrammetric Record, 23(121): 90-108.
- [4] Thompson, E. H., 1959. A Rational Algebraic Formulation of the Problem of Relative Orientation. Photogrammetric Record, Vol. 3, pp. 152-159.
- [5] Mikhail, E., Bethel J., and McGlone J., 2001. Introduction to Modern Photogrammetry. John Wiley and Sons, Inc., 479 pages.
- [6] McGlone J., (Ed), 2004. Manual of Photogrammetry. 5th edition. Published by the American Society of Photogrammetry and Remote Sensing, 1151 pages.
- [7] Horn, B., 1990. Relative Orientation. International Journal of Computer Vision. Vol. 4 (1), pp: 59-78.
- [8] Simon, C., Fraser C., and Hanley H., 2006. Automated Metric Calibration of Colour Digital Cameras. The Photogrammetric Record, 21 (116), pp:355-372.
- [9] Seedahmed, G., (2006). Direct Retrieval of Exterior Orientation Parameters Using A 2-D Projective Transformation. The Photogrammetric Record, 21(115): 211-231.
- [10] Longuet-Higgins, H., 1981. A Computer Algorithm for Reconstructing a Scene from Two Projections. Nature, Vol. 239, pp:133-135.
- [11] Faugeras, O., Lung Q., and Papadopoulos T., 2001. The Geometry of Multiple Images. The MIT Press, 644 pages
- [12] Hartley, R. and Zisserman A., 2003. Multiple View Geometry. 2nd edition. Cambridge University Press, 655 pages
- [13] Ma, Y., Soatto S., Kosecka J., and Sastry S., 2004. An Invitation to 3-D Vision: From Images to Geometric Models. Springer, 526 pages.
- [14] Strang, G., 2006. Linear algebra and its applications. 4th edition. Thomson Academic Inc., 487 pages
- [15] Hartley, R., 1997. In defense of the eight-point algorithm. IEEE Transaction on Pattern Analysis and Machine Intelligence. Vol. 19 (6), pp: 580-593.
- [16] Tsai, R., and Huang, T., 1984. Uniqueness and estimation of Three-Dimensional Motion Parameters of Rigid Objects with Curved Surfaces. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 6(1), pp: 13-26.