DIGITAL APPROACH TO PHOTOGRAMMETRY: A FAILURE CONFIGURATION CASE

G.M. Cortelazzo
Professor at D.E.I.
Università degli Studi di Padova - Italy -
e-mail: corte@dei.unipd.it

A. Vettore
Professor at Istituto di Topografia
Università degli Studi di Padova - Italy -
e-mail: vettoan@ux1.unipd.it

F. Zanon
Engineer at Istituto di Topografia
Università degli Studi di Padova -Italy -
e-mail: fzano@tin.it

Commission VI, Working group 3

KEY WORDS: Digital Photogrammetry, 3D Structure Recovery, Singular Configuration.

ABSTRACT

When a classical photogrammetry approach is used to recover a 3D scene structure the pictures are carefully taken and usually a small angle intersect situation is avoided. This is due to the fact that when two, or more, pictures are taken in such a way a bad parallax minimization is achieved and with consequent invalid results.
In digital photogrammetry the same problem has not been solved yet and a great number of programs fail if this situation is encountered. For example in a commercial software of EOS SYSTEM © big ray intersect angles are recommended.
We want to investigate this problem testing a well established computer vision algorithm, the *Longuet-Higgins* one, over a bad camera set configuration. We will illustrate that if the camera set is close to such a singular configuration the algorithm fails as it is in the classical surveying techniques.

1. Classical approach to photogrammetry

Let's assume to have two pictures of the same scene, taken from two different positions. It is possible to recover the real 3D coordinates of a point in the scene referred to a system located in any position in the real world if at least one dimension of a scene-object is known.
The classical approach is generally divided in two phases: the first one is the *inner orientation* (or *relative orientation*) solution in which camera position are solved one related to the other; the second is the *absolute orientation* in which if some fixed points in the (real) *world* are known the real 3D coordinates refereed to a (real world) system are calculated.
The basis are the so called *collinearity equation*:

$$\begin{cases} x = -f \cdot \dfrac{r_1(X-X_0)+r_4(Y-Y_0)+r_7(Z-Z_0)}{r_3(X-X_0)+r_6(Y-Y_0)+r_9(Z-Z_0)} \\ y = -f \cdot \dfrac{r_2(X-X_0)+r_5(Y-Y_0)+r_8(Z-Z_0)}{r_3(X-X_0)+r_6(Y-Y_0)+r_9(Z-Z_0)} \end{cases} \qquad (1)$$

where $\begin{pmatrix} X_0 \\ Y_0 \\ Z_0 \end{pmatrix}$ is a camera reference system origin, $\begin{pmatrix} X \\ Y \\ Z \end{pmatrix}$ are the 3D point coordinates, while $R = [r_i]_{i=1\div 9}$ are the rotation parameters between the world and the camera reference systems. Usually the rotation matrix is expressed in a way which depends on the particular fields of the surveys: for aero-photogrammetry the angles of Cardano are used, while for terrestrial surveys Eulero's angles are chosen.
Axis orientation depends on the photogrammetry convention: the z generally points backward the (real world) object. The origin is chosen on the *focal point* of the instrument.
Equations *(1)* can be easily derived noting that the projection of a 3D point on a plane far f [1] from the origin of the system needs first to be related to the camera system via a roto-traslation and then by a projective law. Generally homogeneous coordinates are used so that a 2D projection of a 3D point has three coordinates: the third one, in this case, is (minus) the focal length:

$$\begin{pmatrix} x \\ y \\ -f \end{pmatrix} = \alpha \cdot R \cdot \begin{pmatrix} X-X_0 \\ Y-Y_0 \\ Z-Z_0 \end{pmatrix} \qquad (2)$$

with α an *unknown* scale factor.

2. The parallax and the inner orientation

An important geometric concept in projective geometry and in photogrammetry, is the *parallax*.
Parallax vector is simply defined as the vector connecting the two projections of a 3D point, while parallax is the measure of the same vector projected on a plane which is parallel to an image plane. Figure *(1)* gives an idea of the geometrical meaning of parallax. Also note that while moving toward the original 3D point the parallax decreases and it completely vanishes when such a point is reached.

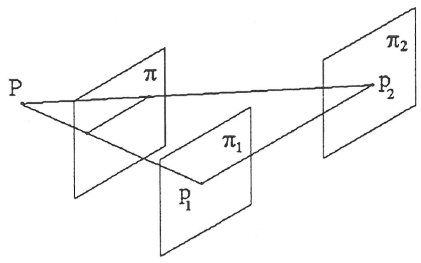

Figure 1: Geometric interpretation of the parallax. π_1 and π_2 are the two image planes, while π is the one where the parallax is evaluated.

The last statement can be reconsidered starting from the end: if we have two projections of the same point we can recover it as

[1] The focal length of the instrument.

the point where the parallax is zero or, in the case of presence of noise, as any point of the minimum parallax vector. This is the most common approach used by optical and mechanical photogrammetry machinery, and also by the most of the software available.
Solving for a zero value of the parallax determines the inner orientation but it does not give a metric reconstruction of the scene. For a metric reconstruction the absolute orientation must be resolved and some extra information is needed: for example some points in the real world and well located in the two photos.
This approach involves the investigation of twelve parameters: three angles for each of the two rotation matrices and three components for each of the two translation vectors. Five of them are solved in the first phase, the inner orientation solution, while the remaining are solved in the absolute orientation solution.

3. The Longuet-Higgins algorithm

For each of the two cameras we will use the *pin-hole* model: starting from a 3D point its projection is obtained by a simple projective geometry law as formalized in section *1*:

$$p = \frac{P}{Z} \cdot f \qquad (3)$$

The focal point is the point where all the projective rays intersect while the *principal point* is its projection on the image plane. The reference system is characterized by the x axis pointing down, the z axis pointing toward the objects and the y axis derived from the previous ones using the right-hand rule, with the origin on the focal point.
We can easily refer a point on the first system to the second one using the following (obvious) relation:

$$P_2 = R \cdot (P_1 - t) \qquad (4)$$

where R is the rotation matrix while t is the translation vector.
Due to the fact that the translation vector and the two projective rays are co-planar we obtain:

$$p_1' \cdot Q \cdot p_2 = 0 \qquad (5)$$

Equation *(5)* is called the *epipolar constraint* while matrix Q is called essential matrix [2]. Such a matrix, containing all the information regarding the rigid motion between the two camera stations, is defined as $R \cdot (t\wedge)$ where:

$$(t\wedge) = \begin{pmatrix} 0 & t_z & -t_y \\ -t_z & 0 & t_x \\ t_y & -t_x & 0 \end{pmatrix}$$

is the skew matrix version of vector t.
What we need to calculate are the translation vector and the rotation matrix from a set of epipolar constraints related to different features in the two images. A very simple way to do this is rewriting the set of epipolar constraints as a set of linear equations and then solving the resulting system:

$$\Pi \cdot q = \begin{pmatrix} \pi_1 \\ \vdots \\ \pi_n \end{pmatrix} \cdot q = 0 \qquad (6)$$

where n is the number of feature pairs and $\pi_{i=1\div n}$ is a rearrangement of feature couple coordinates. It is important to note that no more than eight equations are useful because the essential matrix is defined up to a scale factor [3], i.e. you can divide left hand side of equation *(5)* by any number different from zero.
Usually noisy images are assumed so a *least square estimation* is performed and more than eight pairs of features are used. Via *singular value decomposition* (*SVD*) we can rewrite matrix $\Pi' \cdot \Pi$ as:

$$\Pi' \cdot \Pi = U \cdot \Lambda \cdot V \qquad (7)$$

with U and V orthogonal matrices while Λ is 9×9, diagonal and positive defined, whose elements are decreasing ordered.
Due to noise not all the epipolar constraints may be simultaneously satisfied so the following *minimum problem* needs to be solved:

$$min\left(\left\|\Pi \cdot q\right\|\right)_{\|q\|=1} \qquad (8)$$

The last *singular value*, i. e. the 9^{th} element of matrix Λ, is the measure of the goodness of the solution: the closer it is to zero the better the solution is. Not only: the bigger is the ratio between the 8^{th} and the 9^{th} *singular value* the less distorted is the solution. The best estimate of matrix Q is the *singular vector* associated to the 9^{th} singular value, i.e. $q = V_9$, because it selects the 9^{th} eigenvalue which is the smallest one.
To extract the rotation matrix R and the translation vector t from the vector q we need first to reorder the elements of it.
It is important to observe that four solutions are possible depending on the sign of t and on the rotation direction. By a second SVD derived from the geometrical interpretation of the roto-traslation *[3]* it is very simple to determine each components of the motion. The right couple (R,t) can be recovered noting that all depths are positive according to the camera reference system. Depths may be easily calculated by the following relationships:

$$\begin{cases} Z_1 = (n_1' \cdot m_1) / \|m_1\|^2 \cdot f \\ m_1 = (R \cdot p_1) \wedge p_2 \\ n_1 = (R \cdot t) \wedge p_2 \end{cases} \qquad (9.a)$$

$$\begin{cases} Z_2 = (n_2' \cdot m_2) / \|m_2\|^2 \cdot f \\ m_2 = p_1 \wedge t \\ n_2 = (R' \cdot t) \wedge p_1 \end{cases} \qquad (9.b)$$

4. Parallax and epipolar constraint

[2] Or fundamental matrix in the case of known *calibration parameters*.

[3] Situation which is very similar to the one previously encountered while analyzing the classical approach to photogrammetry.

The Longuet-Higgins approach to the problem is strictly related to the classical approach because, as we will focus in this section in a very intuitive way, an intimate relation exists between the minimum problem *(8)* and the parallax minimization.
This relation derives from the geometrical interpretation of the epipolar constraint: it states that the volume of the 3D solid delimited by the two corresponding projective rays and the translation vector t is zero due to their co-planarity, see figure *(2)*. So, parallax for each couple of features needs to be zero to satisfy such a situation. As explained in section the projection of the vector connecting point p_1 to point p_2 vanishes moving toward the 3D (original) point.
Minimizing the (sum of the) parallaxes and minimizing the norm of the vector $\Pi \cdot q$ as is in problem *(8)*, is just the same thing, and even if the two quantities are geometrically different one from the other, they both involve the same value in the optimal case. Not only. The more parallel are the projective rays the worse the situation is because bigger parallaxes exist and no good (simultaneously) vanishing is achieved. This problem persists in each approach and is strictly connected to the camera set: the best solution is taking shoots avoiding big ray angle intersect.

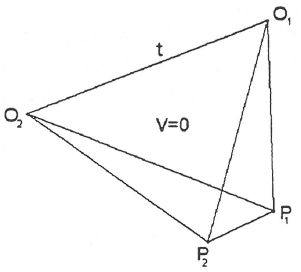

Figure 2: Geometric interpretation of the epipolar constraint.

5. Comment about the experimental results on the failure configuration

The experiment aimed to recover the 3D coordinates recover via the Longuet-Higgins algorithm of a few points located over some boards reproducing a terrain: figures *(3)* and *(4)* show the photos which have been used for the test.
Using the classical approach unreliable results have been obtained for the 3D points and so an alternative solution for the surveys was investigated.

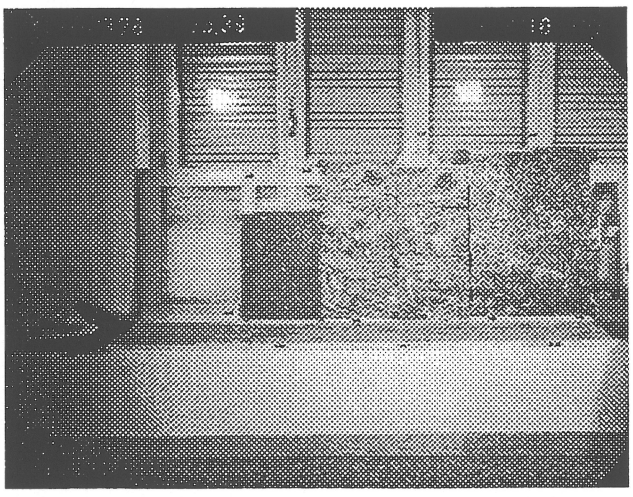

Figure 3: The right picture reproducing the boards to be measured.

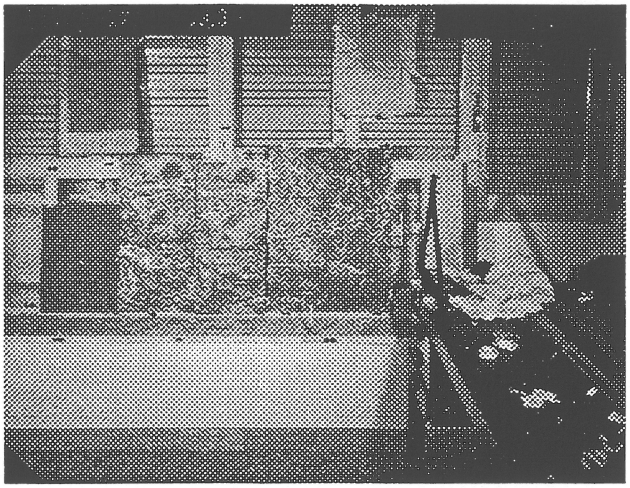

Figure 4: The left picture of the same scene.

The alternative method gave a very unpleasant result: depending on the set of features chosen in the two pictures the 3D coordinates of the target points changed a lot.
Ten trials were run but the results obtained were very unreliable. In one case the algorithm completely failed and no positive depths were calculated.
The reason of the unreliability was found on the very low values of the ratio between the 8^{th} and the 9^{th} singular values. A good solution is achieved when such a ratio is of the order of thousand, while in this experiment the biggest value was of the unity order.
Figure *(5)* reports a graphic interpretation of the results which well illustrates the estimation bias: Σ_0 is a *test* reference system while Σ_1 and Σ_2 are the corresponding ones obtained by the application of the rigid motion using two different estimated couples $(R.t)$. Precisely Σ_1 is obtained using the result of the 1st trial while Σ_2 is the obtained from the parameters of the 7th trial.

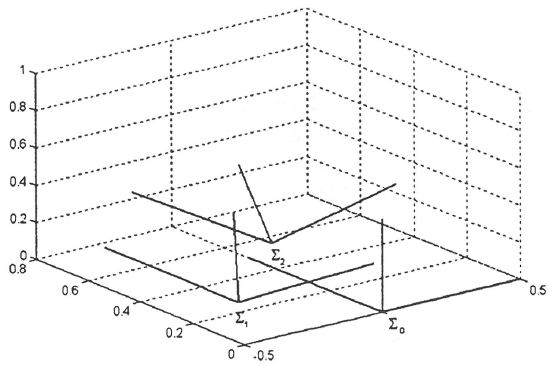

Figure 5: Graphic comparison of two different estimated motion parameter couples $(R.t)$ obtained in two different trials using different set of features. Unit is meters and translation vectors are normalized.

In order to justify this the scene and the camera set have been investigated. As mentioned in the original article of Longuet-Higgins some singular scene configurations exist:

- when at least seven features lie over the same plane;
- when at least four features are located on the same straight line;
- when the 3D surface can be described with a quadratic equation.

The first case was the most similar to the one under investigation and the failure was originally attributed to the *quasi* planar situation of the scene. But it was not right due to the fact that a lot of points have been selected outside the boards and at least three targets are located on the outside part of the marble wall.
Another thing was the camera set: the angle between the two camera is very low and they have *quasi* parallel projective rays.

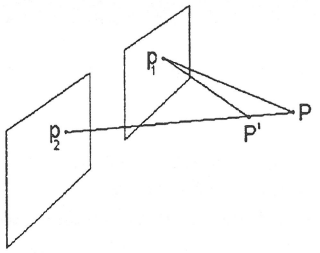

Figure 6: Big angle rays intersection.

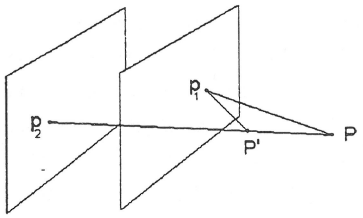

Figure 7: Low angle rays intersection. As you can notice the error on the identification of the re-projected point is bigger than in the previous figure.

Figures *(6)* and *(7)* well illustrate the situation. Here the bias is modeled as a wrong angle between two correspondent projective rays: obviously a better model should include a biased localization of the 2D feature in the image plane.
The low angle intersection involves a worst approximation of the real 3D point compared to the approximation resulting from the situation of bigger angle.

6. Conclusions

Digital approaches to photogrammetry offer the possibility of automating most of the operations for aerial and terrestrial surveys but they do not eliminate the presence of some singular configurations in the camera set and in the scene objects geometry too.
Here we only considered a well established computer vision algorithm in a very unpleasant situation but commercially available software does not correctly handle this situation.
The problem depends on the geometry of the scene which cannot be predicted. The only solution is avoiding such configurations both in the camera set and in the object-scene set.

7. Acknowledgement

This work was partially supported by the MURST (40%) funds for the year 1997.

BIBLIOGRAPHY

1. G. GOLUB, C.F. Van LOAN, 1989. Matrix Computation (2nd edition). *The Johns Hopkins University Press, Baltimore and London.*
2. H.C. LONGUET-HIGGINS, 1981. A computer vision algorithm for reconstructing a scene from two projections. *Nature*, pp. 133-135.
3. S. SOATTO, P. PERONA, 1993. Motion and structure for reconstructing a scene from uncalibrated image sequences. *CIT-CDS Technical Report, n. 23.*
4. S. SOATTO, P. PERONA, 1994. Recursive estimation of camera motion from uncalibrated image sequences. *CIT-CDS Technical Report, 94-005.*
5. G. FANGI, 1995. Note di fotogrammetria con elementi di trattamento delle misure. *Clua Edizioni Ancona.*
6. K. KRAUS, 1994. Fotogrammetria, teoria ed applicazioni. *Libreria Universitaria Levrotto e Bella, Torino..*
7. F. ZANON, 1996. Tecniche automatiche di fotogrammetria. *Tesi di Laurea, Università degli Studi di Padova.*
8. EOS SYSTEM©, 1996. PhotoModeler® Pro 2.1 User Manual.