# ESTIMATION OF CAMERA PARAMETERS FROM STEREO PAIRS WITH NO EXTERNAL CONTROL INFORMATION

I. Kalisperakis, G. Karras, E. Petsa*
Department of Surveying, National Technical University of Athens (NTUA), GR-15780 Athens, Greece
* Department of Surveying, Technological Educational Institute of Athens (TEI−A), GR-12210 Athens Greece E-mail:
ilias_k@central.ntua.gr, gkarras@central.ntua.gr, petsa@teiath.gr

**KEY WORDS:** Calibration, Orientation, Bundle, Lens Distortion, Precision

**ABSTRACT**

Leaving planar rectification aside, in most practical cases of architectural photogrammetry – where mostly non-metric cameras are now used – the camera interior orientation must be known. Of course, it is infeasible to anticipate every possible case and calibrate in advance all available (including auto-focus and zoom) lenses. Therefore, tools are required for the on-the-job estimation of interior orientation parameters. This paper focuses on the simplest instances, namely when no ground control is available. To this end, there exist both single-image methods (for images with clearly defined vanishing points) as well as a multi-image approaches (self-calibrating bundle adjustment which, in principle, also allows estimating the camera interior parameters without exterior information). Most applications, however, still rely on the stereo pair. Thus, the question of determining camera parameters solely from point correspondences on two images is addressed here. Indeed, it is known that simple point homologies in uncalibrated stereo pairs can impose the epipolar constraint, expressed by the 'fundamental matrix' (well documented in computer vision literature), and even permit *partial* self-calibration. For instance, the camera constant of normal cameras – even if different for the two images – may be recovered if the principal point is assumed to be known. Several non-iterative algorithms have also been presented for computing one or two camera constants (the principal point can, too, be localized for given camera constants). The pertinent critical configurations have also been studied. In this contribution, the specific question of estimating a common camera constant is treated. After using a linear algorithm for obtaining initial values, bundle adjustment of the stereo pair is performed for the estimation of relative orientation together with the camera constant as well as radial lens distortion. Relying on extensive real data, the performance of this approach for partial self-calibration without external information is assessed against full self-calibrating bundle adjustment with control points. A comparison of the accuracy of 3D reconstructions with the two methods indicates that the investigated approach is capable of providing comparable results. Further tests are still needed, however, focusing on issues such as proximity to critical geometries or the effect of noise.

## 1. INTRODUCTION

Most of us active in architectural or archaeological photogrammetry are aware that the primary question regarding camera interior orientation can be safely bypassed when direct projective approaches are applicable. Basically, this refers to either planar rectification or the DLT approach (which, however, necessitates sufficient 3D control information). In all other cases the issue of camera calibration – gaining in significance with the wider use of non-metric digital cameras – must be coped with. Evidently, an attempt to anticipate the possible practical situations and pre-calibrate available lenses for different focusing distances would be unrealistic – let alone auto-focusing or zoom lenses. Hence, tools are needed for on-the-job camera calibration. The question motivating the present contribution concerns the available alternatives in cases without sufficient ground control (or configurations unsuitable for single-image calibration).

A possible answer would be to use the 'nominal' values for the camera elements, namely to ignore the principal point and adopt the nominal focal length as camera constant. It has been indicated that, if no exaggerated object relief is present, this choice is generally valid, provided that lens distortion has been taken into consideration (Karras & Mavromati, 2001). Further alternatives can include single-image camera calibration (if suitable vanishing points exist) or, at the other extreme, multi-image self-calibrating bundle adjustment which, in principle, allows estimating all camera parameters without any control information. But as most applications apparently still rely on stereo pairs, the specific aspect addressed here is in fact the recovery of camera parameters from simple point correspondences on two images.

Typically, the task of relative orientation (RO) is defined as that of establishing the relative position of two homologue bundles of rays in model space. Strictly speaking, however, this is not the whole truth: in order to perform a relative orientation, one is not compelled to assume an actual existence of formed bundles, i.e. full knowledge of camera interior orientation (IO). Recovery of relative orientation is indeed possible together with a partial estimation of IO.

To put the issue in a wider framework, one should point out that formulating alternative mathematical models for RO has been a rather early task in photogrammetry, mainly in search for algorithms with no need for initial values. Thus, Thompson (1959) has given an algebraic formulation of the coplanarity condition, in which the RO parameters are grouped in a 3×3 matrix, in fact representing the 'essential matrix' **E** (as it came to be known in the computer vision literature). In other formulations, the equation for computing **E** from ≥8 homologue points is linear, with its 9 elements being reduced to 8 by fixing the scale. From its 8 elements the RO parameters – two relative base components and three angles forming a rotation matrix – may be extracted (Stefanovic, 1973; Khlebnikova, 1983; Shih, 1992). However, it has also been demonstrated that the 2D epipolar geometry of image pairs may still be established even with unknown IO (Faugeras, 1992; Hartley, 1992). The 'fundamental matrix' **F** (having 7 independent parameters, found from simple point homologies) establishes the epipolar constraint on the uncalibrated pair and actually allows a projectively distorted (non Euclidean) 3D reconstruction. It is noted that Zhang (1996) was first to deal with the simultaneous determination of **F** and the polynomial of radial lens distortion.

Furthermore, the calibration potential of simple image correspondences has also been addressed. Chang (1986) gave an early illustration of the possibility to calculate the IO parameters with the simultaneous adjustment of independent pairs from the same camera. But it was Faugeras et al. (1992) who showed that the assumption of common IO in an image pair produces two independent conditions among the elements of **F** and the IO parameters. The resulting equations necessitate ≥3 images to give a full solution for IO solely from point homologies. However, if some camera elements are considered as fixed, partial self-calibration

is feasible from a stereo pair, too. By fixing the principal point, for instance, and disregarding image skewness and aspect ratio, one may recover the camera constant, even if it varies between the two views (Hartley, 1992). Several non-iterative algorithms have been reported in literature for obtaining one (c) or two (c1, c2) camera constants from the fundamental matrix.
For varying c values, Pan et al. (1995) were initially led to a 3rd degree equation in $c^2$. Next, they presented a linear system in $c^2$ for the cases of identical and different camera constants of the two views (Newsam et al., 1996). They also pointed at two critical configurations, which do not allow computation of varying c values from **F**: the camera axes are coplanar with the base (a situation of practical importance); one camera axis is perpendicular to the plane defined by the other camera axis and the base. Bougnoux (1998) has presented an equivalent equation. Hartley & Kaucic (2002) have studied the effect of a wrong assumption about the principal point position on the determination of different c values for the pair. Finally, Sturm (2001) and Sturm et al. (2005) formulated three different equations (one quadratic, two linear) for the determination of a common c from **F**. Studying critical geometries, they demonstrated that c may be calculated even when the camera axes are coplanar, as long as they do not run parallel or their point of intersection is not equidistant from the two projection centres. Closing, it is noted that an additional limitation of the approach is object planarity, in which case the fundamental matrix itself cannot be estimated.
Generally, one may adopt the approximation that the principal point coincides with the image centre. On the contrary, the focal length may be totally unknown or the camera constant may well refer to some unknown zoom factor or focusing distance. Undocumented historic images also fall into this category; in such a case, for instance, the authors were able to obtain reliable initial values only thanks to existing vanishing points (Kalisperakis et al., 2003). Thus, in this first experimentation the focus is on the determination of a common camera constant from image correspondences on two images. After the application of linear algorithms from the literature on real data for drawing initial values, bundle adjustment simultaneously estimates relative orientation along with the camera constant and lens distortion. Using these IO and RO data, 3D reconstructions are finally evaluated against both standard bundle adjustment with self calibration and available ground truth.

## 2. OUTLINE OF THE APPROACH

The central projection of an object point onto the image plane is described by the well-known collinearity equations. Assuming a normal camera (unit aspect ratio, zero skewness), this object-to-image relation involves the image rotation matrix **R**, the translation vector **t** of the projective centre and the interior orientation (xo, yo, c) of the camera. Coefficients (k1, k2) of radial lens distortion can optionally also be included in the model. In the case of relative orientation, **R** obviously refers to the 3 relative rotations of the second image and **t** denotes the relative components of the stereo-base (a total of 5 unknowns).
As mentioned above, in the case of the pair the camera constant c can be extracted along with the RO parameters if the location of the principal point is known, or assumed to be at the image centre (Hartley, 1992). Estimation of the radial distortion coefficients is also feasible. Here, the computation is carried out via a bundle adjustment (with no external control) using the collinearity equations, for whose non-linear least-squares minimization initial values are needed. For the automatic initialization of the algorithm above-mentioned closed-form solutions, based on the fundamental matrix **F**, are employed. The steps of the algorithm applied here are as follows:

1. Computation of **F** and subsequent estimation of the camera constant c (or c1, c2 for each image).
2. Given the above results, computation of the essential matrix **E**, from which the RO parameters **R** and **t** are extracted.
3. Using the latter as initial values, execution of self-calibrating bundle adjustment for finding optimal estimations for c (or c1, c2), **R** and **t**.

*Fundamental matrix and camera constant***.** The coplanarity condition in a pair with unknown interior orientation is generalized via **F** (Hartley & Zisserman, 2000) as

$$x'^T Fx = 0 \qquad (1)$$

whereby **F** is a 3×3 matrix of rank 2, while **x** and **x′** are two homologue image points in homogenous coordinates. In order to compute **F**, the '8-point algorithm' was applied (Hartley, 1997). With ≥8 corresponding points Eq. (1) is expanded in a linear set of equations

$$\mathbf{Af} = 0 \qquad (2)$$

where $\mathbf{f} = [f11,\ldots,f33]^T$ is the vector of the 9 elements of **F**. To enhance numerical stability, image coordinates are first normalized by transferring their origin to the centre of gravity of the
tance of image points from their origin equals is solved for **f** using the singular value decomposition (SVD) of **A** under the constraint ||f|| = 1. This generally yields a matrix **F** of full rank; through SVD factorization and by setting the third singular value to zero, the final rank-deficient **F** is obtained. Coefficients of radial lens distortion are introduced into **F** in a way similar to that of Zhang (1996) and recovered along with the 7 independent parameters of **F**. The algorithm, which is no longer linear, is based on a Levenberg-Marquardt minimization of the distances from the corresponding epipolar lines.
The algorithm described in Newsam et al. (1996) was used to estimate a common c from **F**. Fixing the principal point at the image center, $c^2$ can be found from the roots of the following 2nd degree polynomial:

$$\begin{aligned}&\left[(u_1^T i_3)(v_1^T i_3)\sigma_1 - (u_2^T i_3)(v_2^T i_3)\sigma_2\right] f_{33} m^2 + \\ &\left[\left((u_1^T i_3)^2 + (v_1^T i_3)^2\right)\sigma_1^2 - \left((u_2^T i_3)^2 (v_2^T i_3)^2\right)\sigma_2^2\right] m + \\ &\sigma_1^2 - \sigma_2^2 = 0\end{aligned} \qquad (3)$$

In Eq. (3) the factorization is expressed as $\mathbf{F} = \mathbf{UDV}^T$, ui and vi are the $i^{th}$ column of **U** and **V**, respectively; **D** = diag(σ1,σ2,0) with σ1>σ2>0 being the singular values of **F**; ii is the $i^{th}$ column of the 3×3 identity matrix; and $m = c^2 - 1$. Thus, Eq. (3) is first solved for m, from which c is finally computed.
Sturm (2001) proposed a similar algorithm for c, also based on the SVD factorization of **F**. He found three equations in $c^2$, two linear and one quadratic. The invalid solutions of the latter may be discarded using the linear equations.
In case of c varying between the two images, the most compact formula for estimating the values of c1 and c2 was given by Bougnoux (1998). Newsam et al. (1996) have proposed a different formulation leading to the same results. Here, however, results only for a common c will be given.
*2. Estimation of RO parameters*. Once **F** and c are computed, the essential matrix **E** (known IO) can be determined as

$$E = K^T F K, \qquad K = \begin{bmatrix} -c & 0 & 0 \\ 0 & -c & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{4}$$

ignoring image skewness, aspect ratio and principal point (Hartley & Zisserman, 2000). To ensure that **E** can be factorized into a rotation matrix and a translation vector (‘rigidity constraint’) one forces the two non-zero singular values of **E** to be equal to their mean. The question of the extraction of the RO parameters from the linear expression of the coplanarity condition has been addressed in the photogrammetric literature (Stefanovic, 1973; Khlebnikova, 1983; Shih, 1992). Probably the easiest way for drawing the RO parameters from **E** is through its SVD factorization (Hartley & Zisserman, 2000). Of the four possible solutions only one gives intersected points in front of both cameras (one point suffices for finding the correct solution).

*3. Bundle adjustment*. The obvious merit of the above approach is its linearity. Regarding accuracy, however, its results are approximate and it is generally admitted that they must be essentially regarded as initial values for a proper bundle adjustment. Thus, the approximate values for the RO parameters and c are introduced here into the collinearity equations within a bundle adjustment, incorporating estimation of radial lens distortion, to obtain the final refined values.

A direct way to assess these results, obtained from simple image homologies, is by comparing them to a standard bundle adjustment approach of the stereo pair with full self-calibration (c, xo, yo, k1, k2) using sufficient 3D control information. However, even more eloquent than the direct comparison of values for the parameter c themselves (which are correlated with the RO parameters) is probably the assessment of 3D reconstructions from the two approaches against existing external information.

## 3. EXPERIMENTAL TESTS

Clearly, simulations have to be carried out for a more thorough evaluation of this algorithm, particularly regarding closeness to critical configurations and noise level. Results from tests with simulated and real data have been given in literature (notably in Sturm et al., 2005), but these refer exclusively to closed-form solutions. Here, it is rather the general potential of the stereo pair as regards calibration from point homologies which is of more interest. In this context, results from bundle adjustments of real data will be given. As a general experience from the tests, however, one can observe that in about 75% of the cases the closedform algorithms provided initial values which allowed the bundle adjustment to converge; in the remaining cases, the closedform algorithms either did not supply solution, or their outcome could not serve as approximations leading to convergence. Furthermore, 3D reconstruction directly from the closed-form algorithms exhibited an accuracy which was inferior, by a factor of 5−10, compared to the rigorous approach. In some cases, the introduction of radial lens distortion into the computation of **F** resulted in a certain accuracy improvement.

All image groups used here, whose characteristics are shown in Table 1, had been acquired with the same camera on four occasions for different purposes (Figure 1 shows a typical pair). In all cases, ground control was at hand (signalised in three cases, detail points in the fourth). This allowed to weaken the effect of noisy measurements of homologue points, but also to assess the actual accuracy of reconstruction. As illustrated at the example of Figure 2, object points were well-distributed in 3D (not close to planarity); referring to Table 1, z shows the mean ratio of the maximum difference in depth to the imaging distance.

On the other hand, the image axes of the stereo pairs were generally not far from coplanarity; in Table 1, s represents the mean skewness of image axes expressed as % of the mean distance of the projective centres from the point of ‘best intersection’ of the camera axes (i.e. the mid-point of their shortest distance). The mean ratio of these distances is d. It is clear that configurations are basically close to a coplanarity of image axes, while the two perspective centres are not far from being equidistant from the ‘common’ point of the camera axes.

Table 1. Characteristics of the 4 image groupsσo: mean standard error of relative orientation z: maximum extension in depth as % of imaging distance s: skewness of axes as % of the imaging distance d: ratio of the two imaging distances

| group | number of points | σo(pixel) | z (%) | s (%) | d |
|---|---|---|---|---|---|
| 1 | 25 | 0.20 | 35 | 1.6 | 0.93 |
| 2 | 25 | 0.22 | 20 | 3.1 | 0.95 |
| 3 | 35 | 0.26 | 40 | 3.1 | 0.85 |
| 4 | 16 | 0.19 | 40 | 3.0 | 0.95 |

In all cases focusing was fixed (yet slightly different in the first group). The groups included a variety of geometries, particularly regarding the convergence of image axes. Out of a total of 57 tested stereo pairs, 6 provided no solution, mainly due to critical geometry (e.g. axis convergence less than 1°, projective centres practically equidistant from axes intersection). For the remaining 51 stereo pairs, the mean c values along with their standard deviations are given in Table 2, for the cases without and with use of geodetic control (cR and cC, respectively). The values for c estimated without control clearly exhibit a considerably wider dispersion about their mean, as also illustrated in Figure 3.

Figure 1. A typical stereo pair of the first image group.

Figure 2. Imaging configuration of the first image group.

This is more marked in the fourth case (which generally appears as less precise, probably because of involving fewer points and non-signalised control), with a standard deviation of 3% (as op-

posed to 1.5% in the other cases). This is due to the presence of two extreme values for c, whose omission reduces the standard
camera constant (pixel)
deviation to the half). On the other hand, the differences of the mean values for c are indeed quite small (they nowhere exceed 0.5%). This is an indication that, in the absence of control information, simple image correspondences might provide acceptable estimations for the camera constant, though with an understandably higher uncertainty. It is also noted that the lens distortion curves from the two approaches were practically identical.

Table 2. Bundle adjustment with and without control cR: mean camera constant from adjustments without control cC: mean camera constant from adjustments with control rR: mean RMS of residuals of 3D similarity transformation rC: mean RMS deviations from control points

| group | number of pairs | cR (pixel) | cC (pixel) | rR (mm) | rC (mm) |
|---|---|---|---|---|---|
| 1 | 16 | 2589 ± 39 | 2579 ± 2 | 0.9 | 0.1 |
| 2 | 13 | 2564 ± 37 | 2573 ± 4 | 1.2 | 0.1 |
| 3 | 6 | 2566 ± 39 | 2568 ± 4 | 2.1 | 1.0 |
| 4 | 16 | 2558 ± 81 | 2571 ± 6 | 3.5 | 0.8 |

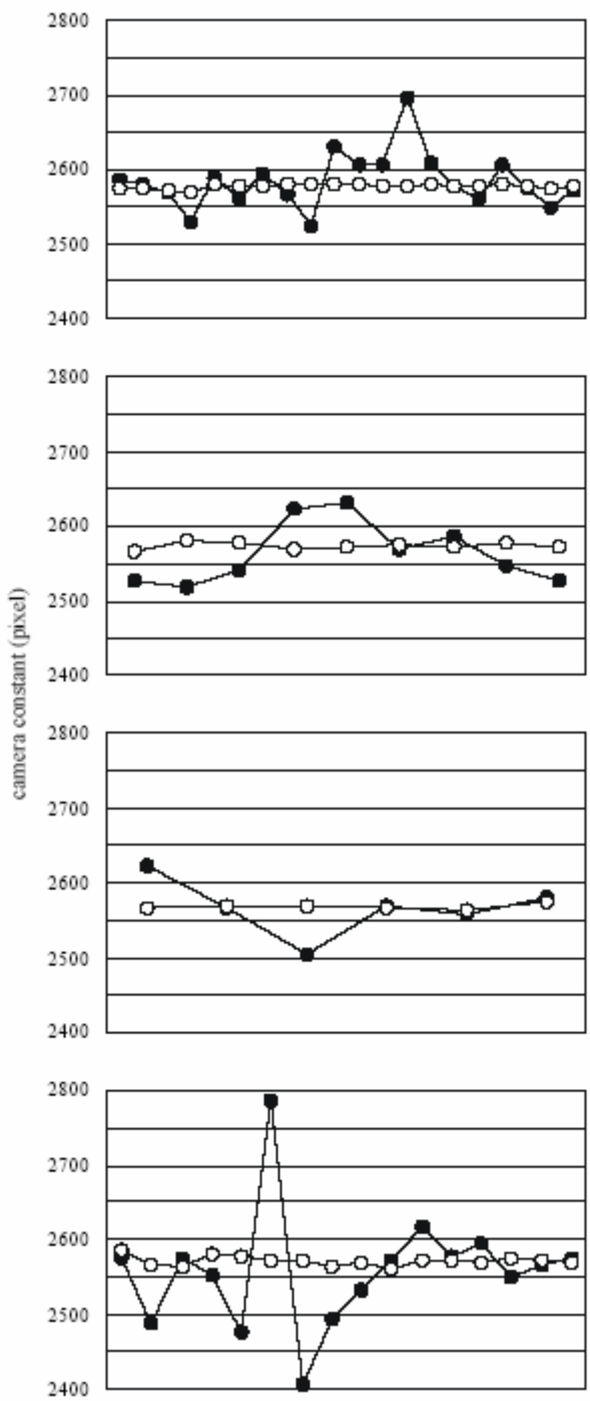


Figure 3. Results for the camera constant from stereo pairs without (black dots) and with control points.

As already pointed out, however, a direct comparison of values originating from different adjustments is not always clarifying. Besides, one cannot generally assert within which exact limits a difference in the value of the camera constant is indeed of significance or not. In this spirit, the results have also been evaluated against the control points. Thus, rC in Table 2 denotes the mean RMS differences of points which were estimated photogrammetrically, with full bundle adjustments, from their known 3D geodetic coordinates. On the contrary, when no external control information was used the reconstructed models were subject to 3D similarity transformation. The mean RMS coordinate differ-

ences R of points thus transformed from their geodetically measured values are also given in Table 2.
The results incorporated in rR have been obtained independently of geodetically measured control. Hence, they obviously cannot be expected to be as close to the control points as the points reconstructed in a procedure, which actually compels homologue image rays to intersect at given geodetic coordinates. Further, in the first instance the principal point has been ignored (its average displacement from the image centre was ~10 pixel in both directions). In addition, the depth extension of objects was large and this enhances the effect of IO errors. In view of the above, this reconstruction, which makes no use of external information and relies on partly uncalibrated stereo pairs, is regarded as indeed satisfactory. The image scales represented mean pixel dimensions <3 mm on the object, i.e. they were suitable for products of scale 1:20. The rR values given above are also compatible with this requirement.

## 4. CONCLUDING REMARKS

This contribution has addressed the question of relative orientation of stereo pairs, focusing on the specific issue of unknown camera constant c (and radial lens distortion) in the absence of object control. This aspect is not only of theoretical but also of practical interest. Generally, one might fix the principal point at the image centre, but in several instances the camera constant c cannot be reliably assumed, e.g. when using a zooming lens or historic images. The published closed-form algorithms applied here gave results which, as regards c (and subsequent 3D reconstruction), are understandably far from being correct; however, in most cases they were indeed capable of automatically providing useable approximate values. Based on these, bundle adjustment without external control allowed to drastically refining the estimation of c and 3D model reconstruction. The experimental results, involving 51 stereo pairs, are from the same camera but cover a large range of configurations. They indicated that c, although its mean values tend towards those from rigorous bundle adjustment with control points, is determined with considerable uncertainty; yet, it does not appear to be directly influenced by closeness to certain critical geometries. Radial lens distortion is estimated reliably. What is probably more significant, however, is the indication that generally acceptable reconstructions might indeed be expected from partly uncalibrated stereo pairs.
Of course, further tests are required to assess the potential of the simple stereo pair as regards reconstruction and partial calibration. For instance, here all point sets were far from being planar, while the noise level of image measurements was indeed low; these aspects should be further studied, along with a more systematic consideration of configurations close to being critical. Finally, initial examples with real data in cases with two different camera constants have given rather satisfactory results; this investigation, too, has to be further pursued.

## REFERENCES


Chang, B., 1986. The formulae of the relative orientation for non-metric camera. Int. Archives of Photogrammetry & Remote Sensing, 26(5):14-22.

Bougnoux, S., 1998. From projective to Euclidean space under any practical situation, a criticism of self-calibration. IEEE Int. Conference on Computer Vision, Bombay, pp. 790–796.

Faugeras, O.D., 1992. What can be seen in three dimensions with an uncalibrated stereo-rig? European Conference on Computer Vision, Springer, pp. 563-578.

Faugeras, O.D., Luong Q.T., Maybank S. J., 1992. Camera self-calibration: theory and experiments. European Conference on Computer Vision, Springer, pp. 321–334.

Hartley, R., 1992. Estimation of relative camera positions for uncalibrated cameras. European Conference on Computer Vision, Springer, pp. 579-587.

Hartley, R., 1997. In defence of the eight-point algorithm. IEEE Trans. Pattern Analysis & Machine Intelligence, 19(6):580-593.

Hartley, R., Zisserman, A., 2000. Multiple View Geometry in Computer Vision. Cambridge University Press.

Hartley, R., Kaucic, R., 2002. Sensitivity of calibration to principal point position. European Conference on Computer Vision, Springer, vol. 2, pp. 433-446.

Kalisperakis, I., Rova, M., Petsa, E., Karras, G., 2003. On multiimage reconstruction from historic photographs. XIX CIPA International Symposium, Antalya, pp. 216-219.

Karras, G., Mavromati, D., 2001. Simple calibration techniques for non-metric cameras. XVIII CIPA International Symposium, Potsdam, pp. 39-46.

Khlebnikova, T., 1983. Determining relative orientation angles of oblique aerial photographs. Mapping Science & Remote Sensing, 21(1):95-100.

Newsam, G.N., Huynh, D.Q., Brooks, M.J., Pan H.P., 1996. Recovering unknown focal lengths in self-calibration: an essentially linear algorithm and degenerate configurations. Int. Archives of Photogrammetry & Remote Sensing, 31(3):575-580.

Pan H.P., Brooks M. J., Newsam G.N., 1995. Image resituation: initial theory. Videometrics IV, Proc. SPIE, 2598:162-173.
Shih, T.Y., 1994. RLT: a closed form solution for relative orientation. Int. Archives Photogrammetry & Remote Sensing, 30(5), pp. 357-363.

Stefanovic, P., 1973. Relative orientation – a new approach. ITC Journal, pp. 417-447.

Sturm, P., 2001. On focal length calibration from two views. IEEE Int. Conference on Computer Vision & Pattern Recognition, pp. 145–150.

Sturm, P., Cheng, Z.L., Chen, P.C.Y., Poo A.N., 2005. Focal length calibration from two views: method and analysis of singular cases. Computer Vision & Image Understanding (in print).

Thompson, E. H., 1959. A rational algebraic formulation of the problem of relative orientation. The Photogrammetric Record, 3(14):152-159.

Zhang, Z., 1996. On the epipolar geometry between two images with lens distortion. IEEE International Conference on Pattern Recognition, vol. 1, pp. 407–411.