IS IT REALISTIC TO GENERATE CONTROL POINTS FROM A STEREO PAIR?

G. E. Karras
Laboratory of Photogrammetry, Department of Surveying,
National Technical University of Athens (NTUA), GR-15780 Athens, Greece
E-mail: gkarras@central.ntua.gr

KEY WORDS: Convergent, Geometry, Orientation, Precision, Rectification

ABSTRACT

As a rule, photogrammetric applications in cultural heritage conservation rely on geodetically measured control points for establishing image and model registration into some object coordinate system. But for relatively modest tasks, such as digital rectification of building façades, geodetic work indeed represents a disproportionately large part of the project. Besides other alternatives, the possibility of generating control points in a purely photogrammetric fashion may be considered. In lack of any control data, bundle adjustment may still create a 3D model, properly scaled if a distance is known, which could be then oriented in a local system based on inherent object geometry. Here, the simplest option is addressed, namely the generation of control points from non-metric stereo pairs. Given the camera parameters and a known object dimension, the powerful geometry of relative orientation allows generating a 3D point set in an arbitrarily oriented but properly scaled model space. If – as in typical cases of building rectification – certain reconstructed points belong to a vertical object plane, plane-fitting allows extracting two rotations which restore verticality, while existing vertical or horizontal lines allow fully fixing the rotational part of absolute orientation. Transformed model points may now serve as control in an object-referenced system. Stereo pairs, taken with three digital cameras (2, 5, 8 MPixel) in various configurations were used in order to evaluate the normalised sets of reconstructed points against geodetic measurements. Reduced onto the image plane, the RMS in-plane discrepancies did not exceed 1 pixel. This indicates that, in principle, the title can be answered in the affirmative, provided, of course, that the 'groundel' size of the images is suitably chosen to meet the accuracy requirements of the final scale.

1. INTRODUCTION

It is known that for some – rather secondary – tasks, for which scale and accuracy are not important (e. g. simple visualizations to be viewed with an Internet browser), photogrammetrists may well dispense with geodetic control. Furthermore, instances also emerge where establishment of control data is indeed out of the question, as in cases of historic photos of demolished buildings, which may only be handled by exploiting object geometry, e. g. through the use of vanishing points (Karras at al., 1993). Nevertheless, the bulk of photogrammetric activities in cultural heritage conservation rely on sets of geodetically measured control points intended to establish image orientation, thus inserting the product into a specific space coordinate system.

It is common knowledge to everyone involved in a conservation project that geodetic field work consumes a considerable part of resources, in terms of both time and cost. In case laser scanning is also involved, extraction of control information from point clouds – or even the intensity images (Forkuo & King, 2004) – may be considered. However, the problem gains in significance in the most straightforward tasks, those in particular concerning simple rectification of building façades (Kager et al., 1985). Indeed, geodetic field work represents a most unwelcome burden when added to the free-hand acquisition of only a few images – or even just one! Establishment of control through direct tape measurements of suitable distances on a planar façade is clearly an option. Yet, the object must be fully accessible; besides, only few control points may be thus established (small redundancy).

A more general option – considered here with specific reference to the requirements of planar rectification – is the possibility of generating control points in a purely photogrammetric fashion.

It is known that, in the absence of exterior control information, multi-image bundle adjustment may, in principle, generate a 3D point set, which represents an arbitrary orientation, position and scale of object shape, namely a 3D model. Although the camera interior orientation parameters can also be estimated in this procedure, it is much simpler to assume a priori knowledge of camera geometry. Given a known distance to accommodate scale, the model may then be transformed into a suitable system based on inherent object features (verticality or horizontality of planes and lines) – if available. Model points thus established could, in principle, furnish the missing control for rectification.

Here the focus is on the simplest alternative, namely the possibility of acquiring control points just from a relative orientation of non-metric stereo pairs. Although *partial* camera calibration is in principle also feasible in this process (a question addressed in Kalisparakis et al., 2005), camera geometry, including radial lens distortion, is considered here to be known. With one object dimension at hand, the powerful tool of relative orientation may lead to a 3D point set in an arbitrarily oriented and located, yet properly scaled, model space. Assuming that (as it is the case in typical building rectifications) some reconstructed model points belong to a vertical plane, the coefficients of plane-fitting provide the two rotations Ω and Φ, which render this model plane indeed vertical (Z = constant). Finally, available vertical and/or horizontal lines provide the third rotation (K), which produces a horizontal X-axis and a vertical Y-axis.

With the rotation matrix of absolute orientation thus estimated, transformed points may now serve as ground control for rectification in this object-referenced 2D system. The general idea is clearly sound, yet its practical potential and limits of application must be investigated. To evaluate the approach, stereo pairs of different imaging configurations have been taken with three digital cameras of different resolution; the photogrammetrically generated control was assessed against geodetic measurements.

It may be mentioned here that Kager et al. (1985), based on the theoretical elaboration of Wunderlich (1982), have presented an approach for planar rectification without control using a stereo pair with given camera geometry. Making explicit use of the 2D projectivity between images, this closed-form solution (actually equivalent to a relative orientation conditioned by a planarity constraint imposed on the model) does not require initial values. However, linear formulations for relative orientation also exist (e.g. Stefanovic, 1973). Furthermore, the approach adopted here is not limited to one single plane.

2. OUTLINE OF THE APPROACH

The process is schematically illustrated in Figure 1. The model, expressed in the system xyz of the left image and scaled via the arbitrary size b of the basis, is transformed in the object system XYZ, scaled through dimension D and oriented through rotation matrix **R**, thanks to the assumed façade geometry. Optionally, a translation vector **t** transfers the origin at a chosen object point.

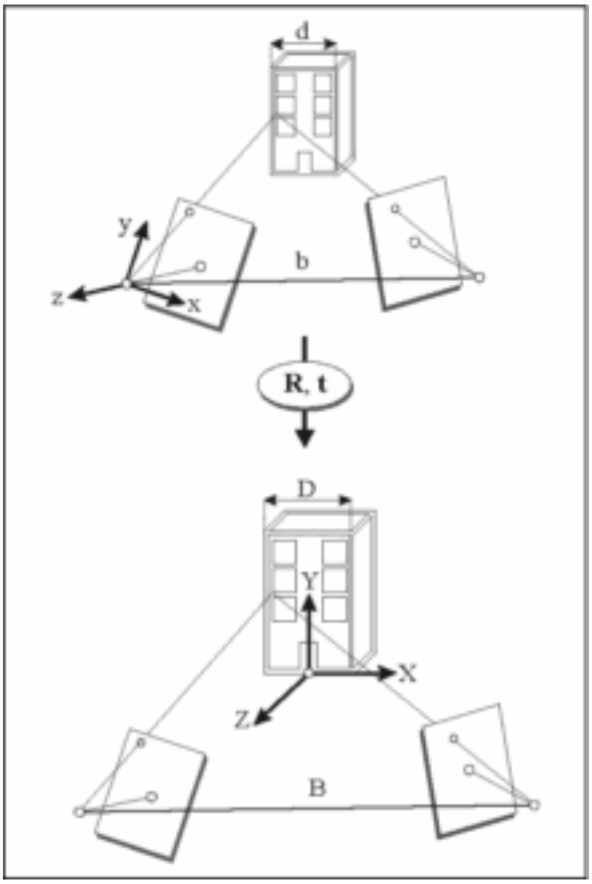

Figure 1. Using geometric properties of the façade and a known dimension D, the initial model system xyz is transformed to the object-referenced system XYZ through a rotational matrix **R** (and an optional translation **t**).

Of course, the above formulation refers to a much more generic problem – 3D object reconstruction with no control points and with only minimal object information – whose practical metric potential needs to be fully investigated. Here only its particular aspect of control point extraction for rectification from a stereo pair is addressed. In general, different solutions may be adopted to this end, namely a one step approach (bundle adjustment) or a succession of simpler steps (relative orientation; space intersection; scaling; recovery of model rotations).

2.1 Bundle adjustment

Applied to stereo pairs, bundle adjustment means here employment of the collinearity equations for simultaneously estimating image exterior orientations and the 3D space coordinates of tie points. If no control is utilized, the 7 parameters of 3D similarity transformation need to be fixed; the model may be subsequently oriented in an object-referenced system, as outlined above. Yet, a one-step solution is also possible. Our bundle adjustment software (Kalisperakis et al., 2003), for instance, can accommodate control points with only one or two known object space coordinates; the remaining ones can be estimated in the adjustment as 'partial' tie points. This particular feature is, indeed, very useful in several architectural applications where control might not be available, but the regular geometry of the object can be exploited instead. Referring to the example of Figure 2, one sees how a horizontal dimension D on a planar façade XY allows establishing, besides the two full control points 1 and 4, also points with known X,Z (points 5, 8), points with known Y,Z (points 2,3) or also points simply belonging on the plane (points 6,7).

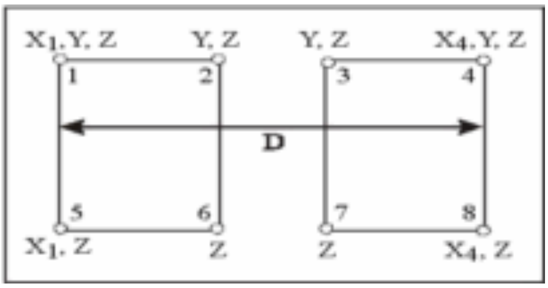

Figure 2. Definition of 'partial' control points by using a known dimension D and exploiting object planarity and symmetry.

This approach is, in principle, more rigorous since it makes use of all available information in one single adjustment. Besides, it can provide additional 'control points' through the estimated 3D object coordinates of tie points. Here, however, the step-to-step approach has been adopted, since it appears as algorithmically simpler, while the assessment of individual steps allows a better control over the process and a direct detection of error sources.

2.2 Step-by-step approach

This procedure consists of the following five separate steps.

1. *Relative orientation*. Using the coplanarity condition, the 5 parameters of relative orientation of the stereo pair are found; typically, these are the three rotations (ω,φ,κ) of the right image with respect to the left, along with the two relative basis components βy = By/Bx, βz = Bz/Bx (which thus correspond to βx = 1). Besides being well distributed, homologue points on the images should include all points which are expected to serve later for control in rectification, along with the endpoints of the known dimension. The standard error σoR of relative orientation, which is the criterion of precision, has to be a small fraction of a pixel. As mentioned above, camera geometry is assumed to be known. Regarding analogue cameras, Karras & Mavromati (2003) indicate that, for most practical purposes, one might consider using the nominal focal length (for focusing at infinity) and disregard the principal point; apparently, use of a digital camera is more demanding. Needless to stress, however, that – particularly if using a wide-angle lens – it is indispensable to correct for radial lens distortion (which may be adequately estimated in a number of simple ways, as outlined in Karras & Mavromati, 2003).
2. *Model generation*. Using the above orientation data, space intersections allow determining the xyz model coordinates of all participating points (evidently, the mean standard error of intersections σoI will be very close to σoR of relative orientation).
3. *Scaling*. The known dimension allows computing the scale factor λ which gives the model its true size: (xyz)′ = $\lambda\times$(xyz).
4. *Adjustment of verticality*. It has been assumed that several object points belong to a plane, which is here supposed

to be a vertical façade (all points of Figure 2 fall into this category). In order to restore verticality, plane-fitting to all involved model points (xyz)′ is performed using the basic observation equation:

$$z' = Ax' + By' + C \tag{1}$$

Now the model points (xyz)′have to be transformed, according to the following equation, to a new system (XYZ)′ such that the plane equation will be Z′ = constant:

$$\begin{bmatrix} X' \\ Y' \\ Z' \end{bmatrix} = R^T \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} \tag{2}$$

The rotation matrix $\mathbf{R} = \mathbf{R}_\Phi \mathbf{R}_\Omega$ is the standard rotation matrix used in photogrammetry and contains the rotations Ω about the X-axis and Φ about the Y-axis which render the plane vertical. From Eqs. (1) and (2), these rotations can be extracted from the coefficients A and B of the fitted plane as

$$\Phi = \arctan(A) \qquad \Omega = -\arctan(B\cos\Phi) \tag{3}$$

while the equation of the plane now becomes: $Z' = C\cos\Phi\cos\Omega$ (assuming that all points involved are indeed coplanar, the RMS deviation in the Z′direction from the fitted mathematical plane indicates the precision of reconstruction). From Eq. (2), through Eq. (3), the new coordinates (XYZ)′ of model points are found.

5. *In-plane rotation*. Finally, a horizontal or vertical line can provide the angle K of rotation about the Z′-axis using the ΔX′ and ΔY′differences of the endpoints (if more lines are available the mean K-value is used). Thus, the final XYZ object point coordinates (whereby, of course, Z=Z′) are obtained. The origin is optionally moved to a selected object point.

2.3 Accuracy considerations

The accuracy of this generation of 'control points' is obviously subject to a number of error sources. First, the inner orientation parameters are assumed to be known. As mentioned already, the interest focuses here chiefly on radial-symmetric lens distortion Δr (indeed, for one of the lenses used in the tests disregarding of distortion increased inaccuracy by a factor of up to 6).

Generally, of course, the geometry of the stereo pair also affects accuracy, but this refers primarily to depth estimation (which is not all that important in the present context). However, in order to 'open' the bundle of imaging rays one may prefer to use convergent stereo pairs. Abdel-Aziz & Karara (1974) have studied error propagation in the symmetric convergent case, in which a part of the error in depth Z is transferred to planimetry. Yet, for relatively limited angles of convergence one might be content with the approximation that on a XY plane the measuring error is propagated, basically, through the image scale factor k.

As regards image scale, it is clear that digital images cannot be treated in the conventional way of analogue cameras, since here the matter of resolution (pixel size) has a primary role. With all image data, including camera constant, being generally in pixel dimensions, image scale 1:k = c:H is in dimensions 'pixel/m', and hence refers directly to the projection of the pixel size onto the object (mean 'groundel' size). Planning of image acquisition should primarily take this into account; in this sense, it appears as more reasonable to express accuracy measures in pixel units.

3. EXPERIMENTAL APPLICATION

In order to check the approach under practical conditions, three digital cameras of different resolutions (2, 5, 8 Megapixel) were used, for which the interior orientation parameters were at hand from previous bundle adjustments:

- Agfa 1800×1600 (c = 1854; negligible Δr);
- Sony 2592×1944 (c = 2574; Δr reaching 15 pixels near the image corners);
- Canon 3264×2448 (c = 3586; Δr reaching 10 pixels near the image corners).

Two planar building façades were recorded, one using the first two cameras and one using the third camera, resulting in three models (a1−a3) for the Agfa camera with a mean scale of about 140 pixel/m, four models (s1−s4) for the Sony camera having a mean scale of ~200 pixel/m and three models (c1−c3) for the last camera with a mean scale of ~280 pixel/m. The range of the angles of convergence of the stereo pairs was 10°−45°. Figure 3 presents a pair from each of the different cameras. The standard error of relative orientations was in the range of 0.2−0.5 pixel.

Figure 3. Stereo pairs from the three digital cameras.

In order to evaluate the results, points were also measured geodetically. For each model, the corresponding geodetic and photogrammetric 2D point sets were first normalized (transferred to their centres of gravity), and subsequently the RMS differences δX, δY were computed and reduced on the image plane through the corresponding mean image scale. Table 1 shows the results.

Table 1. RMS differences δX, δY between geodetic and photogrammetric points

model	points	δX (pixel)	δY (pixel)
a1		1.0	1.1
a2	19	0.7	0.8
a3		0.8	1.1
s1		1.0	1.2
s2 s3	27	1.11.1	1.4 1.3
s4		1.0	1.3
c1		1.2	1.6
c2	12	1.0	1.2
c3		1.4	1.3

In order to check the approach under practical conditions, three digital cameras of different resolutions (2, 5, 8 Megapixel) were The first observation is that, with one or two exceptions, these results from cameras of different resolutions show a consistency on the image plane in the range of 1.0−1.4 pixels (with values in Y being somewhat larger). Furthermore, the differences must not be fully attributed to the photogrammetric process, since the geodetic measurements (particularly of detail points which need a certain 'interpretation') also introduce errors. This is also indicated by the fact that the obtained differences are much larger than the errors of relative orientation, which essentially express the 'internal' precision of the photogrammetric process.

Thus, for an independent estimate of precision, several images have been used in 2D projective transformation (Christodoulou, 2005). Table 2 shows the standard errors (σXY) when using geodetic control (g) as well as photogrammetric 'control', derived from stereo pairs with no participation of the particular images.

Table 2. Accuracy σ_{XY} (pixel) of projective transformations using control points obtained geodetically (g) and photogrammetrically (a, s, c)

g	*a1*	*a2*	*a3*	
1.2	0.5	–	–	
1.2	–	0.9	–	
1.1	–	–	0.8	
	s1	*s2*	*s3*	*s4*
1.2	0.9	0.9	1.1	–
1.2	0.8	0.7	–	1.0
1.1	–	0.6	0.8	0.8
1.5	0.9	–	1.1	1.2
1.2	0.9	–	–	–
1.2	–	0.8	0.7	0.8
	c1	*c2*	*c3*	
1.8	1.6	–	–	
1.3	–	1.2	–	
0.6	–	–	1.0	

It is seen that the 2D points derived photogrammetrically from independent imagery actually appear as fitting somewhat closer to the images than geodetic points, which hints at differences in identification when one observes geodetically and on the image. Since two of the cameras depicted the same object, it was also possible to compare to each other point sets reconstructed from altogether different cameras (Christodoulou, 2005). The RMS differences of 0.6−1.4 pixel do not contradict the above results.

Thus, one may assume that the photogrammetric accuracy σP is roughly equal to that of the geodetic measurements (σ_G). This is to say that for the RMS differences δ it holds

$$(\delta)^2 = (\sigma_G)^2 + (\sigma_P)^2 = 2(\sigma_P)^2 \qquad (4)$$

Thus, from the mean value of Table 1 (~1.4 pixel), the accuracy of photogrammetric 'control points' may be finally estimated as σ_P = 1 pixel. As regards practical applications, the scale of rectification 1:k_R generally implies that an error $\sigma_R = k_R \times 0.2$ mm can be tolerated on the object plane. If one assumes that control points should be at least twice as accurate, it is concluded that

$$\sigma_P \leq \sigma_R/2 \Rightarrow 1 \text{ pixel} \leq k_R \times 0.1 \text{ mm}$$

But 100 μm is in fact the pixel dimension usually used in image transformations (rectification, orthoimaging); thus, $k_R \times 0.1$ mm also represents the 'groundel' of the final result. Source images generally have a 'groundel' size which is by 2−3 times smaller than this value, to allow high quality resampling. Consequently, one can conclude that – as regards resolution and scale – source imagery suitable for rectification is in principle also suitable for the photogrammetric generation of control.

By way of example, it is noted that the mean RMS differences δ_{XY} for each camera were 6.3 mm, 5.7 mm and 4.6 mm, giving respective photogrammetric accuracy estimations σP as 4.5 mm, 4.0 mm and 3.3 mm. This means that the stereo pairs used here could be regarded as suitable for a rectification scale of 1:50.

4. CONCLUDING REMARKS

The task of 3D object reconstruction with no control points and minimal external information represents a fundamental question in photogrammetry. Its simplest aspect has been addressed here, namely the generation from non-metric stereo pairs of a number of points, sufficiently accurate to serve as control for digital rectifications tasks. Obviously, this possibility could be of interest in several practical situations. The results presented here using different digital cameras indicate that an accuracy of 1 pixel can be reasonably expected; hence, one may plan image acquisition according to the scale of the final product. Reliable information on camera geometry is clearly of primary importance in such an exclusively image-based approach.

It is believed that – after certain indispensable checks to verify the performance of such a modest approach – one might indeed 'trust' the potential of non-metric digital imagery. However, the limitations of the stereo pair are evident, particularly as regards object size in relation to the required resolution. Such instances would call for more images, namely for bundle adjustment with minimal control. This next step towards a fuller metric exploitation of the 'internal' photogrammetric precision is worth taking.

REFERENCES

Abdel-Aziz, Y. I., Karara, H. M., 1974. Expected accuracy of convergent photos. Photogrammetric Engineering, vol. 40, no. 11, pp. 1341-1346.

Christodoulou C., 2005. Generation of Control Points from Nonmetric Stereo Pairs. Diploma Thesis, NTUA, pp. 59.

Forkuo, E. K., King, B., 2004. Registration of photogrammetric imagery and laser scanner point clouds. Proc. of ASPRS Annual Conference, Denver, Colorado (in CD).

Kager, H., Kraus, K., Novak, K., 1985. Entzerrung ohne Paßpunkte. Bildmessung und Luftbildwesen, 53(2), pp. 43-53.

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

Kalisperakis, I., Karras, G., Petsa, E., 2005. Estimation of camera parameters from stereo pairs with no external control information. In this volume.

Karras, G. E., Patias, P., Petsa, E., 1993. Experiences with rectification of non-metric digital images when ground control is not available. XV International CIPA Symposium, Bucharest.

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

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

Wunderlich, W., 1982. Rechnerische Rekonstruktion eines ebenen Objekts aus zwei Photographien. Mitteilungen der geodätischen Institute der TU Graz, Folge 40, pp. 365-377.