ROBUST LEAST-SQUARES ADJUSTMENT BASED ORIENTATION AND AUTO-CALIBRATION OF WIDE-BASELINE IMAGE SEQUENCES

Helmut Mayer
Institute for Photogrammetry and Cartography, Bundeswehr University Munich, D-85577 Neubiberg, Germany
Helmut.Mayer@unibw.de

ABSTRACT

In this paper we propose a strategy for the orientation and auto-calibration of wide-baseline image sequences. Our particular contribution lies in demonstrating, that by means of robust least-squares adjustment in the form of bundle adjustment as well as least-squares matching (LSM), one can obtain highly precise and reliable results. To deal with large image sizes, we make use of image pyramids. We do not need approximate values, neither for orientation nor calibration, because we use direct solutions and robust algorithms, particularly fundamental matrices $\mathbf{F}$, trifocal tensors $\mathcal{T}$, random sample consensus (RANSAC), and auto-calibration based on the image of the dual absolute quadric. We describe our strategy from end to end, and demonstrate its potential by means of examples, showing also one way for evaluation. The latter is based on imaging a cylindrical object (advertisement column), taking the last to be the first image, but without employing the closedness constraint. We finally summarize our findings and point to further directions of research.

1 INTRODUCTION

(Hartley and Zisserman, 2000) has transformed the art of producing a Euclidean model from basically nothing into text-book knowledge. As can be seen from recent examples such as (Nistér, 2004, Pollefeys et al., 2004, Lhuillier and Quan, 2005) a very high level has been reached.

We also head into this direction, making it possible to generate a Euclidean three-dimensional (3D) relative model (no scale, translation, and rotation known, i.e., seven degrees of freedoms undefined) from not much more than the images and the knowledge, that the images are perspective and sufficiently overlapping. Besides the latter, we make two in many practical cases reasonable assumptions, namely, that the camera is not too strongly (below about 15°) rotated around its optical axis between consecutive images and that all images are taken with one set of calibration (interior) parameters. The latter has to be true only approximately. While we cannot deal with zooming, we found empirically, that we can handle focusing.

The strategy, that we propose, particularly focuses on robust least-squares adjustment (Mikhail et al., 2001) in the form of bundle adjustment and least-squares matching (LSM). By means of affine LSM, we obtain highly precise conjugate points. Together with bundle adjustment, which we use for the computation of every fundamental matrix $\mathbf{F}$ as well as trifocal tensor $\mathcal{T}$, and after linking triplets via 3D projective transformation, we obtain highly precise and at the same time reliable solutions. This is demonstrated by means of two examples, in one of which a cylindrical object (advertisement column) was imaged with 28 images. Even though the information, that for the last image the first has been taken, has not been used in the adjustment, the cylinder is preserved very well.

Basically, our strategy rests on extracting points which we match highly precisely with LSM (cf. Section 2). Section 3 explains how hypothesis for conjugate points undergo rigorous geometric checks by projective reconstruction via computing $\mathbf{F}$ and $\mathcal{T}$, robustified by means of random sample consensus (RANSAC), as well as linking triplets via 3D projective transformation. All, including intermediate results of projective reconstruction are improved via robust bundle adjustment, important issues for which we explain in Section 4. As we deal with images of several Mega pixels, we employ image pyramids including tracking points via LSM through the pyramid (cf. Section 5). The projective reconstruction is upgraded to Euclidean via auto-calibration, described in Section 6. In Section 7 we demonstrate the potential of our strategy, particularly the high geometric precision and reliability achievable by means of LSM and bundle adjustment by means of an experiment specifically designed to evaluate the precision of the 3D reconstruction. Finally, we present a summary and directions for further research.

2 POINT EXTRACTION AND LEAST-SQUARES MATCHING

We start by extracting Förstner (Förstner and Gülch, 1987) points. An even distribution of the conjugate points on the image is enforced if possible by regional non-maximum suppression in the reference image of a particular matching step. No suppression is employed in the other images, because due to noise and occlusions the regionally strongest points in two images do not have to be the conjugate points.

Contrary to most approaches, we do not use the coordinates of the points for the conjugate points directly, but we determine relative coordinates by selecting one image and determining the relative shift of image patches around the points in the other images via LSM. This has the big advantage, that we obtain an estimate of the precision of the match.

To be able to deal with large baseline scenarios, we use as search space the size of the image. This naturally leads to a large number of hypotheses. As LSM is computational expensive, we first sort out unlikely candidates for conjugate points by means of normalized cross correlation. We

particularly have found that correlating in red, green, and blue and combining the outcome by means of multiplication is a good choice for making use of color information. We employ a relatively low threshold of 0.7^3 to keep most of the correct points. Experiments with color spaces have not been successful as we found the color information to be mostly noisy, leading to bad correlation in the chrominance, etc., band.

As color information has already been used, we do not make use of it for LSM. For it, we employ affine geometric transformation, because the parameters for a projective transformation cannot be reliably determined for image patches in the range of 11×11 pixels. Additionally to the the six affine geometric parameters, we determine a bias and a drift (contrast) parameter for the brightness. For two images we just match the second to the first. For three and more images we determine an average image in the geometry of the reference image. Matching against it, we avoid the bias by a radiometrically badly selected reference image (e.g., distorted by occlusion).

The result of this step are highly precise image coordinates for the conjugate points including an estimate of the precision. This value is mostly over optimistic (one often obtains standard deviations in the range of one hundredth of a pixel), but they still give a good hint on the relative quality of the solution obtained.

3 ROBUST PROJECTIVE RECONSTRUCTION

The conjugate points of the preceding section are input for projective reconstruction. Basically, the goal is reconstruction of the whole sequence. Because of the inherent noise and due to problems with similar and repeating structures as well as occlusions, the strategy needs to be rather robust, and at the same time efficient.

We have decided to use triplets as the basic building block of our strategy. This is due to the fact, that by means of the intersection of three image rays one can sort out wrong matches, i.e., outliers, highly reliably. Opposed to this, one cannot check the depth for image pairs, as the only constraint is, that a point has to lie on the epipolar line. Even though using triplets as basic building block, combinatorics suggests to actually start with image pairs, restricting the search space via epipolar lines. For the actual estimation of the relations of pairs and triplets we employ $\mathbf{F}$ and $\mathcal{T}$ (Hartley and Zisserman, 2003). Triplets are computed sequentially and are linked by means of projecting points of the preceding triplet via the new $\mathcal{T}$ into the new last image resulting into (n+1)-fold points as well as computing the projection matrix of the last image via 3D projective transformation for the first and second but last images. (Projection matrices for $\mathbf{F}$ and $\mathcal{T}$ can be obtained with the standard algorithms explained in (Hartley and Zisserman, 2003).) Finally, points not yet seen are added.

Of extreme importance for the feasibility of our strategy is the use of robust means, particularly RANSAC (Fischler and Bolles, 1981), that we use for the computation of $\mathbf{F}$ and $\mathcal{T}$. As we are dealing with a relatively large number of outliers in the range of up to 80%, RANSAC becomes especially for the computation of $\mathcal{T}$ extremely slow. This is mostly due to the fact, that for reliably estimating $\mathcal{T}$, it is necessary to compute a point-wise bundle adjustment. We use a modified version of RANSAC speeding up the computation by more than one order of magnitude for high noise levels, where as shown in (Tordoff and Murray, 2002), often much larger numbers of iterations are needed to obtain a correct result than predicted by the standard formula given in (Hartley and Zisserman, 2003).

4 ROBUST BUNDLE ADJUSTMENT

Bundle adjustment is at the core of our strategy. We have found, that only by adjusting virtually all results, we obtain a high precision, but also reliability. The latter stems from the fact, that by enforcing highly precise results for a large number of points, one can guarantee with a very high likelihood, that the solution is not random.

Basically this means, that when estimating $\mathbf{F}$ and $\mathcal{T}$, we compute the optimum RANSAC solution for junks of several hundreds of iterations and then we run a projective bundle adjustment on it. This is done a larger number of times (we have found empirically five to be the minimum number), as the bundle adjustment solution is partly much better than the RANSAC solution and its result can vary a lot. But having several instances of bundle solutions, there is nearly always one which is sufficiently precise and representing the correct solution.

We employ projective as well as Euclidean bundle adjustment, both including radial distortion $ds = 1. + k_2 * (r^2 - r_0^2) + k_4 * (r^4 - r_0^4)$ with r the distance of a point to the principal point (or its estimate) and r_0 the distance where ds is 0. $r_0 = 0.5$ is used as recommended in literature and empirically verified. We have found by a larger number of experiments, that it is important to employ radial distortion only after outlier removal. It is not used at all for the determination of $\mathbf{F}$ or $\mathcal{T}$, but only after we have tracked down points to the original image resolution (cf. below).

We originally wanted to employ standard least-squares adjustment without Levenberg Marquardt stabilization (Hartley and Zisserman, 2003), to avoid a bias during estimation. Therefore, we are using the SVD-based minimal parameterization proposed in (Bartoli and Sturm, 2001) for the first camera for projective bundle adjustment. Yet, we have found, that only by means of a Levenberg Marquardt stabilization we can deal with the large initial distortions of the solution caused by outliers. Particularly, this means, that we multiply the elements of the diagonal of the normal equations with $1 + stab$, the stabilization parameter $stab$ being adaptively determined by means of varying it with a factor of 10 between 1.e-5 and 1.

We base the robustness of bundle adjustment on standardized residuals $\bar{v}_i = v_i / \sigma_{v_i}$ involving the standard deviations σ_{v_i} of the residuals, i.e., the differences between observed and predicted values. As a first means we employ

reweighting with $w_i = \sqrt{2 + \bar{v}_i^2}$ (McGlone et al., 2004). Additionally, having obtained a stable solution concerning reweighting, outliers are characterized by $\bar{v}_i$ exceeding a threshold, which we have set to 4, in accordance with theoretical derivations and empirical findings, eliminating the outliers for the next iteration.

For bundle adjustment, efficient solutions are extremely important. E.g., a 29 image sequence as the one presented below leads to more than thirty thousand unknowns, making straightforward computation impossible. We therefore follow (Mikhail et al., 2001) and reduce the normal equations in two steps: First, we reduce the points. Secondly, we also reduce parameters which are common to all, or at least sets of images, namely the calibration and / or (radial) distortion parameters. This results into a tremendous reduction in computation time and storage requirements, even when computing also σ_{v_i}.

5 HIERARCHICAL PROCESSING VIA PYRAMIDS

As we deal with relatively large images in the range of 5 Mega pixels or above and we assume at the same time, that we do not know the percentage or direction of overlap of the images, only a hierarchical scheme allows for an adequate performance. We particularly compute image pyramids with a reduction factor of 2. For the highest level we found that a size of about 100×100 pixels is sufficient in most cases. On this level we compute $\mathbf{F}$. $\mathcal{T}$ are computed on the second highest and for images with a size of more than 1000×1000 pixels also on the third highest level.

We do not compute $\mathcal{T}$ on the fourth highest or lower levels, firstly due to the complexity of the matching and secondly because already on the second or third highest level we obtain for most sequences hundredth of points, more than enough for a stable and precise solution. To still use the information from the original resolution, we track the points via LSM down to the original resolution once the sequence has been oriented completely on the second or third highest level. This is rather efficient also for images of several Mega pixels. As reference image we use for every point the image, where the point is closest to the center of the image, assuming that there the perspective distortion of the patches around the points is minimum on average. After tracking, a final robust projective bundle adjustment is employed, at this time including radial distortion.

6 AUTO-CALIBRATION

To proceed from projective to Euclidean space, one needs to estimate the position of the plane at infinity π_∞ as well as the calibration matrix

$$\boldsymbol{K} = \begin{bmatrix} c & c \cdot s & x_0 \\ & c \cdot (1+m) & y_0 \\ & & 1 \end{bmatrix}$$

with c the principal distance, m the scale factor between x- and y-axis, needed, e.g., for video cameras with rectangular instead of quadratic pixels, x_0 and y_0 the coordinates of the principal point in x- and y-direction, and finally s the sheer, i.e., the deviation of a 90° angle between the x- and the y-axis. The latter can safely be assumed to be zero for digital cameras.

To compute $\boldsymbol{K}$ and a transform to upgrade our projective to a Euclidean configuration, we use the approach of Pollefeys (Pollefeys et al., 2002, Pollefeys et al., 2004). It is based on the image of the dual absolute quadric

$$\omega^* \approx \mathbf{K}\mathbf{K}^\top \approx \mathbf{P}\Omega^*\mathbf{P}^\top$$

which is related to the calibration matrix multiplied with any scalar $\neq 0$ ($\mathbf{K}$) and the dual absolute quadric Ω^*, projected by the projection matrices $\mathbf{P}$. (Pollefeys et al., 2002, Pollefeys et al., 2004) employ knowledge about meaningful values and their standard deviations for the parameters of $\mathbf{K}$ to constrain the computation of Ω^* such as, that the principal distance is one with a standard deviation of nine and all other parameters are zero with standard deviations of 0.1 for the principal point and m and 0.01 for s. The result is a transformation matrix from projective to Euclidean space and one $\boldsymbol{K}$ for every image.

We have experienced, that the resulting Euclidean configuration can be some way off the final result, especially for longer sequences. I.e., for the sequence of 29 images below, the estimated principal distance, known to be constant, varied between 0.3 and 3. To avoid this problem, we have found it to be sufficient to compute the calibration for the first few images and transform the rest of the sequence accordingly. Though this has worked for our experiments, a better way might be to define a number of images n, say three or five, and compute the calibration, which is of very low computational complexity, for all subsequent n images. Finally, the solution should be taken with the smallest summed up standard deviation of all parameters for the average $\boldsymbol{K}$.

As demonstrated, e.g., by the experiments below, robust bundle adjustment including radial distortion is an absolute must after calibration. We start with configurations where the back projection errors can be in the range of several hundred pixels. This stems from the fact, that the calibration procedure produces locally varying $\boldsymbol{K}$ (cf. above). Using Levenberg Marquardt stabilization, it is possible to bring down theses large values to fractions of a pixel. In the beginning the multiplication factor for the elements on the main diagonal can be as high as two, i.e., $stab = 1$.

Because also after projective bundle adjustment there still can be a large number of outliers, also the strategy for bundle adjustment was found to be very important. This is due to the fact, that we accepted sound configurations in projective space, which yet can imply relatively different $\boldsymbol{K}$. Optimizing all parameters of an average $\boldsymbol{K}$ simultaneously can lead to initially very wrong values for x_0, y_0, and s. It was therefore found to be very important to first optimize only c and $c \cdot (1+m)$, and to optimize the rest of the parameters only when this adjustment has converged. Optimizing c and $c \cdot (1+m)$ independently makes the whole procedure less stable on one hand, but allows on the other hand to check the quality of the result by comparing both.

7 EXPERIMENTS AND EVALUATION

In this section we report about results for the proposed strategy and propose one means to evaluate results. All images used in the experiments shown here have been acquired with the same camera, namely a Sony P100 5 Mega pixel camera with Zeiss objective using the smallest possible focal length / principal distance to optimize the geometry of the intersections. To guarantee sharp images (and to make the experiments more difficult), the camera was allowed to auto-focus, leading to slightly varying principal distances. We first present the result for one example out of tens, namely the scene yard, for which our strategy works reliably using the same set of parameters. I.e., one acquires the images, runs the program implementing the strategy and obtains the result consisting of 3D points, camera translations and rotations as well as the calibration, all including standard deviations.

Additionally, we report about one experiment we have devised to evaluate the quality of the solution. For it we acquired 28 images of an advertisement column, which is close to a perfect cylinder. The images have been taken walking unconstrained, so there is some flexibility in the orientation. Though, by always trying to be able to see the whole width of the column, there was a strong constraint to actually take the images from positions on a circle.

The scene yard consists of eight images taken in a backyard. The first three images and the last image are given in Figure 1. Figure 2 shows a view on the resulting VRML model. For the sequence we have obtained 426 threefold points, i.e., points which could be matched in three images, 377 fourfold, 228 fivefold, 103 sixfold, and 20 sevenfold points resulting in an uncalibrated back projection error σ_0 of 0.39 pixels and a σ_0 of 0.3 pixels after calibration. Further parameters such as the calibration matrix $\boldsymbol{K}$ can be found in Table 1.

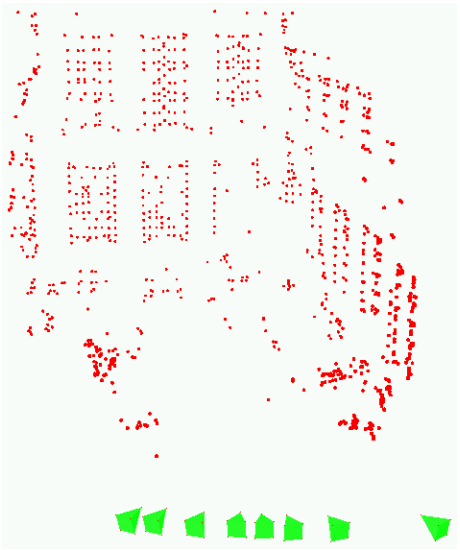

Figure 2: Visualization of points (red) and cameras (green pyramids) of model yard

number images	8
σ_0 projective / Euclidean	0.39 / 0.30 pixel
$\boldsymbol{K}$	$\begin{matrix} 1.247 & -0.001 & -0.004 \\ & 1.251 & 0.0024 \\ & & 1 \end{matrix}$
k_2 / k_4 (radial distortion)	-0.041 / -0.069

Table 1: Results for sequence yard

Of the 28 approximately evenly spaced images of the advertisement column / cylinder, the first three and the fifth are shown in Figure 3. Four other images, showing the variety of texture found on the column, are given in Figure 4.

For the evaluation we have devised three experiments. The first is with the original resolution of 2592×1944 pixels, the second with the resolution reduced by a factor of three, i.e., 864×648 pixels, and for the last experiment we have reduced the resolution by a factor of three and the number of images, wherever there is enough texture, by a factor of two. I.e., we have taken the first, third, and fifth image, etc., as shown in Figure 4.

On the original resolution we obtained 2498 threefold, 3387 fourfold, 2559 fivefold, 1085 sixfold, 309 sevenfold, and 45 eightfold points, as well as one ninefold point resulting in a back projection error of $\sigma_0 = 0.1$ pixels on the third highest pyramid level and of $\sigma_0 = 0.29$ pixels after tracking down to the original resolution. Auto-calibration resulted into estimated $c = 1.04$ and $c \cdot (1 + m) = 1.05$. The resulting configuration is given in Figure 5 left. The back projection error has been in the range of 500 pixels before bundle-adjustment. Bundle adjustment reduced it to 0.19 pixels. The final result is very close to a perfect cylinder as proven by Figure 5 right.

Table 2 shows a comparison of the results. They are rather similar for the original and the reduced resolution sequence. This suggests, that probably because of the relatively small pixel size of the employed mid-end Sony P 100 consumer camera, the original resolution does not convey much more information than the reduced resolution. Similar findings have been made for other sequences. On the other hand, the results for the sequence with the reduced number of images are rather different. This probably stems from the fact, that the overlap between the images is small and the view angles on the surface are partly rather large. For large areas of weak or no texture, such as in image thirteen (cf. Figure 4), we even had to use the original configuration. One can see this, e.g., as a hole in the upper right of the cylinder in Figure 5, right. The comparison of Tables 2 and 1 shows, that even though the time between acquiring the cylinder and the yard sequence was about one year, all the parameters including the distortion are rather similar, if enough images were used for the cylinder sequence. (Please remember, that the same camera has been used.)

For the evaluation of the different versions of the cylinder sequence, we have taken the first image to be the last image of the sequence as well. Instead of using this information in the bundle adjustment, we employ it for evaluation by

Figure 1: First three images and image eight, i.e., last image, of sequence yard

Figure 3: First three images and fifth image of the original sequence cylinder with 28 images

Figure 4: Images eight, thirteen, eighteen and twenty three of the original sequence cylinder

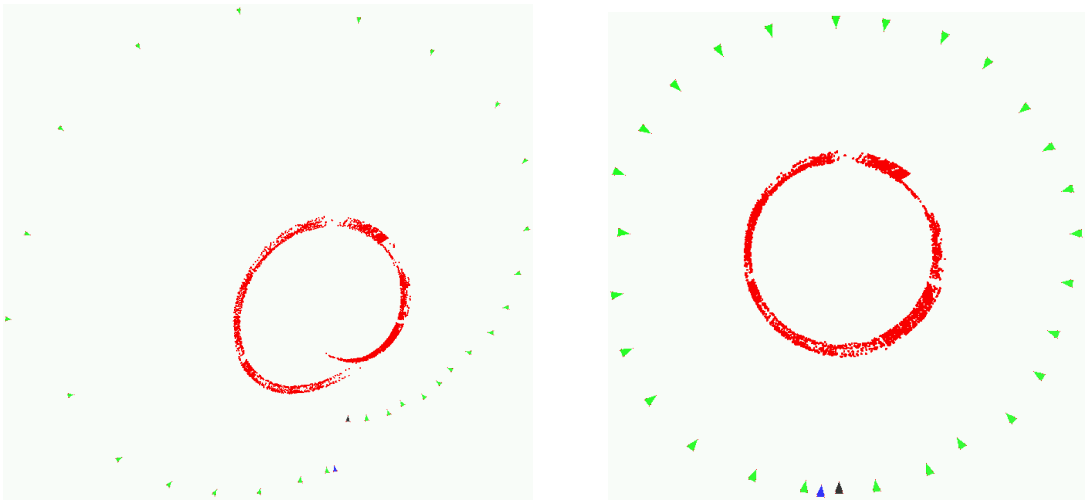

Figure 5: Result for the original sequence cylinder before (left) and after (right) robust Euclidean bundle adjustment. The first and the last camera are marked as black and blue and the rest of the cameras as green pyramids. Points are shown in red.

comparing the parameters of the first and the last camera, which ideally should be the same. Table 3 gives two different types of descriptions, namely the translation in x-, y-, and z-direction of the first $=$ last camera in relation to the radius of the circle constructed by all cameras, as well as the difference in rotation (this is the rotation angle of an axis-angle representation), the latter also in terms of a single image. One can see, that the difference is rather small for the original as well as for the sequence with reduced resolution. Only for the sequence with the reduced number of images there is a significant reduction of the quality.

8 SUMMARY AND CONCLUSIONS

We have shown, that via least-squares adjustment based techniques, particularly least-squares matching and bundle adjustment, highly precise and at the same time reliable results can be obtained. This has been demonstrated by means of a cylindrical object, for which it was shown, that the ring of cameras closes very well and for which at least visually also the shape is preserved extremely well. By means of enlarging the distance between the cameras, we have shown difficulties of the strategy when using a weaker

	original	resolution reduced by 3	reduced number images
number images	29	29	22
σ_0 projective / Euclidean	0.29 / 0.19 pixel	0.12 / 0.08 pixel	0.24 / 0.13 pixel
$\boldsymbol{K}$	$\begin{matrix} 1.239 & 0.0002 & 0.002 \\ & 1.241 & 0.0001 \\ & & 1 \end{matrix}$	$\begin{matrix} 1.242 & 0.0001 & 0.003 \\ & 1.241 & -0.0003 \\ & & 1 \end{matrix}$	$\begin{matrix} 1.168 & -0.0006 & -0.0015 \\ & 1.179 & -0.0062 \\ & & 1 \end{matrix}$
k_2 / k_4 (radial distortion)	-0.040 / -0.060	-0.043 / -0.053	-0.041 / -0.069

Table 2: Results for sequence cylinder

	original	resolution reduced by 3	reduced number images
dx / dy / dz in % of radius circle images	3.5 / -0.36 / 0.74	3.8 / -0.81 / 0.8	7.1 / -1. / 1.1
$d\phi$ global / $d\phi$ per image	$5°$ / $0.18°$	$5.8°$ / $0.21°$	$8.7°$ / $0.41°$

Table 3: Differences in translation and rotation of the parameters of the first = last image of sequence circle. dx, dy, and dz are given in relation to the approximate radius of the circle constructed by the camera positions.

geometry.

A first issue for further research is a more quantitative evaluation of the shape of the given object. This could be done in our case by fitting a cylinder to the object and determining the distances from this cylinder. Though the object is not an ideal cylinder, it should be rather close to it.

Calibration is a further issue. Here the approach of (Nistér, 2004) based on the cheirality constraint seems to be extremely promising. We also still need to deal with planar parts of the sequence. For this we want to follow (Pollefeys et al., 2002), though we note that we have found the issue of model selection (homography versus $\mathbf{F}$ or $\mathcal{T}$) rather tricky.

Finally, an issue that we see as particularly important to achieve the goal of being able to orient also traditional photogrammetric close range image setups is matching which is more invariant with respect to strong geometric distortion. For it we find especially (Georgescu and Meer, 2004) and (Lowe, 2004) very interesting.

REFERENCES

Bartoli, A. and Sturm, P., 2001. Three New Algorithms for Projective Bundle Adjustment with Minimum Parameters. Rapport de Recherche 4236, INRIA, Sophia Antipolis, France.

Fischler, M. and Bolles, R., 1981. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. Communications of the ACM 24(6), pp. 381–395.

Förstner, W. and Gülch, E., 1987. A Fast Operator for Detection and Precise Location of Distinct Points, Corners and Centres of Circular Features. In: ISPRS Intercommission Conference on Fast Processing of Photogrammetric Data, Interlaken, Switzerland, pp. 281–305.

Georgescu, B. and Meer, P., 2004. Point Matching Under Large Image Deformations and Illumination Changes. IEEE Transactions on Pattern Analysis and Machine Intelligence 26(6), pp. 674–688.

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

Hartley, R. and Zisserman, A., 2003. Multiple View Geometry in Computer Vision – Second Edition. Cambridge University Press, Cambridge, UK.

Lhuillier, M. and Quan, L., 2005. A Qasi-Dense Approach to Surface Reconstruction from Uncalibrated Images. IEEE Transactions on Pattern Analysis and Machine Intelligence 27(3), pp. 418–433.

Lowe, D., 2004. Distintive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision 60(2), pp. 91–110.

McGlone, J., Bethel, J. and Mikhail, E. (eds), 2004. Manual of Photogrammetry. American Society of Photogrammetry and Remote Sensing, Bethesda, USA.

Mikhail, E., Bethel, J. and McGlone, J., 2001. Introduction to Modern Photogrammetry. John Wiley & Sons, Inc, New York, USA.

Nistér, D., 2004. Untwisting a Projective Reconstruction. International Journal of Computer Vision 60(2), pp. 165–183.

Pollefeys, M., Van Gool, L., Vergauwen, M., Verbiest, F., Cornelis, K. and Tops, J., 2004. Visual Modeling with a Hand-Held Camera. International Journal of Computer Vision 59(3), pp. 207–232.

Pollefeys, M., Verbiest, F. and Van Gool, L., 2002. Surviving Dominant Planes in Uncalibrated Structure and Motion Recovery. In: Seventh European Conference on Computer Vision, Vol. II, pp. 837–851.

Tordoff, B. and Murray, D., 2002. Guided Sampling and Consensus for Motion Estimation. In: Seventh European Conference on Computer Vision, Vol. I, pp. 82–96.