Global Trifocal Adjustment

Patrik Persson[(✉)] and Kalle Åström[(✉)]

Centre for Mathematical Sciences, Lund University, Lund, Sweden
{patrik.persson,karl.astrom}@math.lth.se

Abstract. In this paper we introduce a fast and robust structure-less alternative to full bundle adjustment. The method is based on optimizing algebraic errors for trilinear constraints from triplets of views. It is shown that the error generated by a triplet of views can be described by a fixed triangular 27×27 matrix regardless of the number of feature correspondences between the views. The method has been evaluated on various real and synthetic datasets and shows good convergence properties with a large convergence basin and solutions that are close to the optimal solution. The method has been compared to Global Epipolar Adjustment, GEA, which is based on the bilinear constraint. It will be shown that the method can handle the degenerate configurations of GEA.

Keywords: Structure-less bundle adjustment · SFM · Optimization

1 Introduction

Structure from motion [10], is the problem of estimating the parameters of scene structure and of camera motion using image features. Assuming that feature position errors are zero-mean Gaussian, the maximum likelihood estimate is that of minimizing the sum of squares of the residuals. The theory and methods go back to the developments of Gauss and Legendre, cf. [8,15]. In the photogrammetry and computer vision literature this process is denoted bundle adjustment, where 'bundle' refers to the bundle of light rays connecting each camera with each 3D point. For an overview of the literature and theory see [25]. For examples of bundle adjustment in photogrammetry see [9].

Bundle adjustment is almost always used both as a final step for the estimation of the parameters, and as an intermediary steps e.g. during tracking, where it has been demonstrated to reduce failures, [6,14]. There are several software packages for bundle adjustment, e.g. SBA [18], Bundler [23,24], and general optimization packages such as Ceres [1]. Many of these take advantage of recent developments in making large scale bundle adjustment fast using e.g. the sparsity and the structure of the problem, [3,4]. This has made it possible to solve large scale bundle adjustment problems as shown in [2,5,20,23]. It has also been shown how bundle adjustment can be used in real-time as a component of SLAM systems using autonomous vehicles.

Despite these development, many of these algorithms still need to go through all data at each step. This limits how fast the bundle adjustment can be. As an alternative there are structure less approaches, which although approximate can give satisfactory results at a much increased speed. One example of this is the approach of Global Epipolar Adjustment [17,22], in which a simplified error metric based on the linear constraint on the epipolar constraints for each pair of images. Another is incremental light bundle adjustment, iBLA, [13] in which an error metric based on a combination of epipolar constraints and a variant of the trifocal constraint is used.

In this paper we introduce a method based on the trifocal constraint from each image triplet. One argument for this is the fact that bilinearities do not restrict the image correspondences fully as shown in [12], whereas trilinearities are sufficient. Another argument is that bilinearites does not restrict camera motions, e.g. if the camera motion is linear. As will be demonstrated in the paper it turns out that using trilinearities gives a larger basin of convergence as compared to using bilinearities. Contrary to iBLA, the proposed method can significantly reduce the number of residuals that need to be evaluated to a fix 27 residuals per triplet of views regardless of the number of feature correspondences. In iBLA each feature correspondence in three views would result in a residual. This makes the computational cost dependent on the number of correspondences which typically is in the order of hundreds.

2   The Bundle Adjustment Problem

A widely accepted model of image formation is the pinhole camera model

$$\lambda \mathbf{x} = \begin{bmatrix} \gamma f & s & x_0 \\ 0 & f & y_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} R| - RT \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \iff \tag{1}$$

$$\lambda \mathbf{x} = K \begin{bmatrix} R| - RT \end{bmatrix} \mathbf{X} = P\mathbf{X}, \tag{2}$$

where $\mathbf{X} = \begin{bmatrix} X & Y & Z & 1 \end{bmatrix}^T$ denotes homogeneous object coordinates, $\mathbf{x} = \begin{bmatrix} x & y & 1 \end{bmatrix}^T$ homogeneous image coordinates, λ depth, rotation matrix R and translation vector T are the extrinsic parameters and K contains the intrinsic parameters.

Given an initial set of cameras P_i, object points $\mathbf{X}_p$ and observations of these $\mathbf{x}_{ip}$, bundle adjustment seeks to minimize the total re-projection error in all views

$$e = \sum_i \sum_p \|x_{ip} - \frac{1}{\lambda_{ip}} P_i \mathbf{X}_p\|^2 \tag{3}$$

by optimizing over the cameras and object points. This error function can be minimized using iterative algorithms. Several fast algorithms exists that exploit the sparse structure of the problem.

Typically the number of object points greatly exceed the number of cameras. To avoid having to include all the object points in the bundle adjustment, the global epipolar adjustment method [17,22], GEA, exploits the epipolar, also called bilinear, constraint between camera pairs. This constraint only depends on the camera parameters and the observed features and do not include object points. This allows the adjustment method to only optimize over the camera parameters allowing for structure-less bundle adjustment.

The epipolar constraint between view i and j is expressed as

$$x_{ip}^T E_{ij} x_{jp} = x_{ip}^T R_i \left[\frac{T_j - T_i}{\|T_j - T_i\|} \right]_\times R_j^T x_{jp} = 0, \tag{4}$$

where R_i, R_j represents the rotations and T_i, T_j the camera center of cameras i and j. Here x_{ip} and x_{jp} represents the observations of object point p in camera i and j. By collecting all the epipolar constraints between possible pairs of views

$$C_e = \sum_{i,j} \sum_p (x_{ip}^T E_{ij} x_{jp})^2 \tag{5}$$

the cost function used in GEA is found. This cost function can be rewritten as

$$C_e = \sum_{i \neq j} e_{ij}^T M_{ij}^T M_{ij} e_{ij}, \tag{6}$$

where e_{ij} is a vectorized column matrix of E_{ij} and M_{ij} is the Jacobian of the constraints between view i and j with respect to e_{ij}. The matrix M_{ij} is of size $n_{ij} \times 9$ where n_{ij} is the number of corresponding points between the views. The matrix can be reduced to an equivalent 9×9 matrix $\tilde{M}_{ij}$ in the sense that the error $e_{ij}^T M_{ij}^T M_{ij} e_{ij} = e_{ij}^T \tilde{M}_{ij}^T \tilde{M}_{ij} e_{ij}$ will be the same for all e_{ij}. $\tilde{M}$ can be calculated using QR factorization of M_{ij}. This greatly reduces the number of residuals that need to be evaluated at each iteration.

The matrices M_{ij} and $\tilde{M}_{ij}$ are independent of the camera parameters and can be computed once and then be reused in future iterations.

3   The Trifocal Tensor

While GEA uses the bilinear constraint to perform optimization, GTA uses the trilinear constraint involving three cameras instead of two. The constraint can be expressed as [11]

$$\frac{1}{2} \epsilon_{ii'i''} \epsilon_{jj'j''} \epsilon_{kk'k''} \det \begin{bmatrix} A^{i'} \\ A^{i''} \\ B^j \\ C^k \end{bmatrix} x_A^i x_B^{j'} x_C^{k'} = 0_{j''k''}, \tag{7}$$

where A, B and C denote the three camera matrices and x_A, x_B and x_C the observations of an object point. A^i corresponds to the i:th row of A and

$\mathbf{x}_A = \begin{bmatrix} x_A^1 & x_A^2 & x_A^3 \end{bmatrix}^T$. Here $\epsilon_{ii'i''}$ denotes the permutation symbol, j'' and k'' denotes free indices and can be selected from $\in [1, 2, 3]$. From these 9 constraints only 4 are linear independent.

Using the trifocal tensor

$$
{}_I^{JK}T_i^{jk} = \frac{1}{2}\epsilon_{ii'i''} \det \begin{bmatrix} P_I^{i'} \\ P_I^{i''} \\ P_J^{j} \\ P_K^{k} \end{bmatrix}
\tag{8}
$$

the constraint can be expressed as

$$
{}_I^{JK}T_i^{jk}\mathbf{x}_I^i\epsilon_{jj'j''}\mathbf{x}_J^{j'}\epsilon_{kk'k''}\mathbf{x}_K^{k'} = 0_{j''k''}.
\tag{9}
$$

This constraint is linear in the components of the trifocal tensor, which makes it possible to estimate it using linear methods. By vectorizing the tensor ${}_I^{JK}T_i^{jk}$ into a 27×1 column-vector $\mathbf{t}$ and forming the Jacobian matrix A of the trilinear constraints with respect to t, the equation system

$$
A\mathbf{t} = 0
\tag{10}
$$

can be formed. The tensor can now be found by solving this equation system. Next, the camera matrices can be extracted from the tensor. For a more comprehensive description of how to form A, see [21]. In the rest of the paper, A will be called cost matrix.

Besides the trilinear constraint on point correspondences, there exists constraints involving line correspondences and combinations of lines and points, see [11] and [21].

4   Global Trifocal Adjustment

One issue with using linear methods to solve for the trifocal tensor is, that due to noise in the feature correspondences, the solution may not result in a valid tensor. Instead of directly solving for the trifocal tensors and then extracting the camera matrices, we propose to parameterize the trifocal tensors with the rotation and translation of the cameras and solving the minimization problem

$$
\Theta = \underset{(\theta_1, \cdots, \theta_n)}{\arg\min} \sum_{(i,j,k)\in Q} W_{ijk}^2 \|A_{ijk}\mathbf{t}_{ijk}(\theta_i, \theta_j, \theta_k)\|^2,
\tag{11}
$$

$$
W_{ijk} = \frac{1}{\|T_i - T_j\|} + \frac{1}{\|T_i - T_k\|} + \frac{1}{\|T_k - T_j\|}
\tag{12}
$$

using an iterative scheme such as Levenberg-Marquardt (LM) [16, 19]. We use Q to denote the set of triplets of cameras, A_{ijk} is the cost matrix for the triplet and $\mathbf{t}_{ijk}$ is the vectorized trifocal tensor of the triplet. Here $\theta_i = (R_i, T_i)$ parameterize the rotation and camera center of camera i respectively. The first camera in

the set of all the cameras defines the coordinate system and the distance to the second camera is constrained to one to fix the scale. Using this parameterization, we can ensure that all estimated trifocal tensors are valid at every iteration. W_{ijk} has been added to prevent the system form converging to the trivial solution where all camera centres coincide.

It is important to notice that several cameras are optimized jointly in (11) and Q is the set of triplets of these cameras, where one camera can belong to several triplets. The set Q can be formed in several ways, ranging from forming all possible triplets of the involved cameras to a minimum of connected triplets of cameras.

4.1 Reducing the Number of Residuals

Each cost matrix A_{ijk} is of size $N_{ijk} \times 27$. Here N_{ijk} depends on the number of corresponding features between the three views and the number of linear independent constraints that can be generated from each correspondence, e.g. for a point correspondence there exists 4 independent constraints. Similar to the procedure in GEA, the A_{ijk} matrix can be replaced by an equivalent reduced matrix $\tilde{A}_{ijk}$, this time of dimensions 27×27.

Considering the error of one triplet of cameras

$$e = \|A_{ijk}\mathbf{t}_{ijk}\|^2 \tag{13}$$

$$= \mathbf{t}_{ijk}^T A_{ijk}^T A_{ijk}\mathbf{t}_{ijk}, \tag{14}$$

QR decomposition of $A_{ijk} = QR$ can be performed. Substituting in the above expression we get

$$e = \mathbf{t}_{ijk}^T A_{ijk}^T A_{ijk}\mathbf{t}_{ijk} = \mathbf{t}_{ijk}^T R^T Q^T Q R \mathbf{t}_{ijk} \tag{15}$$

$$= \mathbf{t}_{ijk}^T R^T R \mathbf{t}_{ijk} = \mathbf{t}_{ijk}^T \begin{bmatrix} R_1^T & 0 \end{bmatrix} \begin{bmatrix} R_1 \\ 0 \end{bmatrix} \mathbf{t}_{ijk} \tag{16}$$

$$= \mathbf{t}_{ijk}^T R_1^T R_1 \mathbf{t}_{ijk}. \tag{17}$$

Here R_1 is of dimension 27×27. The reduced matrix can thus be selected as $\tilde{A}_{ijk} = R_1$. This matrix will be equivalent to A_{ijk} in the sense that they will generate the same error $\mathbf{t}_{ijk}^T A_{ijk}^T A_{ijk}\mathbf{t}_{ijk} = \mathbf{t}_{ijk}^T \tilde{A}_{ijk}^T \tilde{A}_{ijk}\mathbf{t}_{ijk}$ for all $\mathbf{t}_{ijk}$.

The large reduction in the dimensions of the cost matrix greatly reduces the amount of computations needed in each iteration. Since neither A_{ijk} nor $\tilde{A}_{ijk}$ depend on the parameters of the cameras i, j and k, they need only be evaluated once and can be reused in the next iterations. This is of benefit in incremental structure form motion, since only a handful of new cost matrices need to be calculated at the insert of a new keyframe. The rest can be reused.

5   Experimental Validation

The method has been evaluated using both real and synthetic datasets. Each dataset contains initial camera poses, object points and a ground "truth". For the synthetic datasets, the ground truths consist of the simulated tracks and for the real datasets, the ground truths are generated by bundle adjustment. The termination criterion used to determine convergence is the norm of the gradient and the norm of the delta update should be smaller than selected thresholds. The thresholds are the same for all of the methods.

Besides investigating and comparing GTA and GEA, a combination of the epipolar and trifocal constraints called Global Epipolar and Trifocal Adjustment (GETA) will also be investigated. In this version the cost-matrices in GEA and GTA are used simultaneously.

5.1   Real Datasets

The real datasets consist of the publicly available VGG datasets [26] *Corridor* and *Model House*, and a two new datasets *Long Corridor* and *Dining Hall*, illustrated in Fig. 1.

Fig. 1. Sample image from the four datasets: top-left *Corridor*, top-left *Model House*, bottom-left *Long Corridor*, bottom-right *Dining Hall*.

VGG Datasets *Corridor* and *Model House*. The VGG datasets provide point and line correspondences, initial camera matrices and 3d structure. However, only the point correspondences and the calibration matrices retrieved from the initial camera matrices are used.

Datasets *Long Corridor* and *Dining Hall*. Two longer datasets *Long Corridor* and *Dining Hall* were collected. The *Long Corridor* dataset captures motion along a long straight corridor, while *Dining Hall* captures a more general motion. The datasets were collected using a Samsung Galaxy s7 mobile phone camera with a resolution of 1512×2016 pixels and variable focus. A structure from motion system based on [7] was used to find the initial structure of the datasets.

The program assumes fixed calibration and to account for varying focal length, bundle adjustment was applied.

Constraint History. The average error of the solutions when varying the number of previous cameras a camera is allowed to form constraints with, is illustrated in Fig. 2. This will be called the constraint history. From the plot it can be seen that GEA requires a longer constraint history to achieve a low position error compared to GTA and GETA. The latter methods achieve a low error even for the shortest constraint history. Furthermore, a clear minimum error can be observed where additional constraints degrade the solution. The rotation error decreases for both of the methods as the constraints increase.
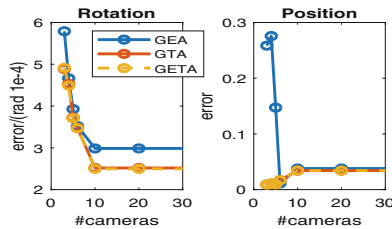

Fig. 2. The error on the *corridor* dataset relative to the optimal solution varying to the number of previous cameras a camera is allowed to form constraints with, starting at 3. The rotation error is small regardless of constraint history, while GTA and GETA both have much smaller position error compared to GEA for short histories.

Rotational and Positional Error. The position and rotation error of the methods relative to the bundle adjustment solution is displayed in Figs. 4. It can be seen that GTA and GETA has a lower position error compared to GEA in both datasets. Figure 3 illustrates how close GEA and GTA converges to the bundle adjustment solution in the *Model House* dataset. A significant overlap can be seen indicating close convergence to the bundle adjustment solution.

We studied the number of iterations needed for convergence and similar to the previous dataset, few iterations are required and the three methods worked equally well in this respect.

Sensitivity to Noise. The average error when increasing the observation noise is displayed in Fig. 5. It can be seen that the performance of GEA deteriorate significantly compared to GTA, even at low noise levels, when subject motion along a corridor.
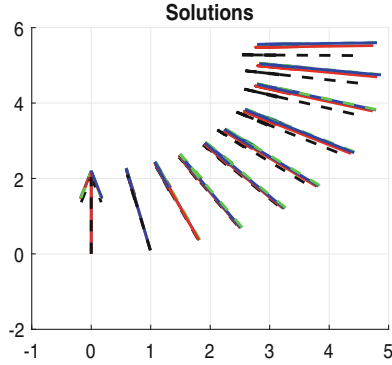
Fig. 3. Superimposed solutions of the *Model House* dataset using GEA, GTA and bundle adjustment. Initial pose (black), bundle adjustment (green), GEA (red), GTA (blue) (Color figure online)
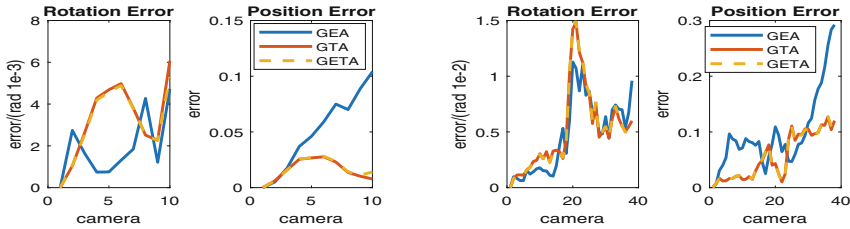

Fig. 4. Error relative to bundle adjustment solution in the *Model House* and *Dining Hall* datasets. GEA has a larger position error compared to the others.

Convergence Basin. If the initial guess is not very close to the optimum GEA has problems converging, which is illustrated in the convergence histograms in Fig. 6. The histograms shows the percentage of times GEA and GTA converged when initialized with various perturbation of the optimal solution. The convergence at each combination of rotation and translation perturbation is tested independently 10 times, such that spurious convergence and divergence can be seen.

From the histogram it can be seen that GEA quickly becomes unstable when the initial guess is perturbed from the optimal solution. In contrast, GTA has a much larger convergence region and does not suffer from the same instability issues on the dataset.

6   Synthetic Validation

To study the behavior of GEA and GTA in the case of collinear and near collinear movement, synthetic rooms and trajectories have been generated. The trajectory is generated by starting in one end of a room with a constant velocity and
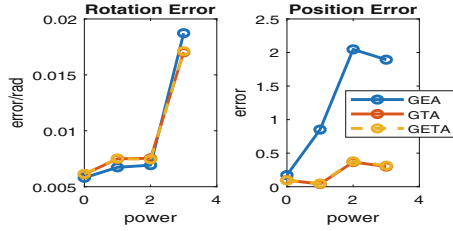
Fig. 5. The average error relative to the optimal solution at different noise level on the *Long Corridor* datasets. The resolution has been divided with 2^x. Highest resolution 1512×2016, lowest 189×252. Here it can be seen that the GEA solution quickly deteriorates when the noise increases, while GTA and GETA are largely unaffected.
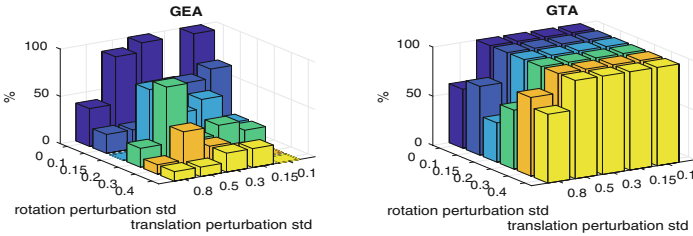

Fig. 6. A histogram over the number of times GEA (top) and GTA (bottom) converged when the solution is perturbed different amounts in rotation and translation from the ground truth for the *Long Corridor* dataset. It can be seen that GEA quickly fails when perturbed from the optimal solution, whereas GTA converges for a wider range of perturbations.

applying an constant acceleration acting perpendicular to the initial velocity direction. The amplitude of the acceleration is determined by a parameter called curvature. This generates curves similar to those in the *Corridor* and *Long Corridor* datasets, where a smaller value will generate a straighter trajectory while larger values generate a more curved trajectory. Several such datasets have been generated with varying curvature, observation noise and perturbations form the ground truth. GEA and GTA have then been applied to the datasets. The results can be seen in Fig. 7 where the median of the mean position error has been plotted as a function of the curvature and observation noise.

Here the position error for GEA grows large as the curvature decreases and the noise increases. For small curvatures the error grows significantly when the noise is increased. As the curvature grows larger the dependency on noise become less prominent. In contrast to GEA, the error of GTA is shows no large dependency of the curvature and depends only on the observation noise.
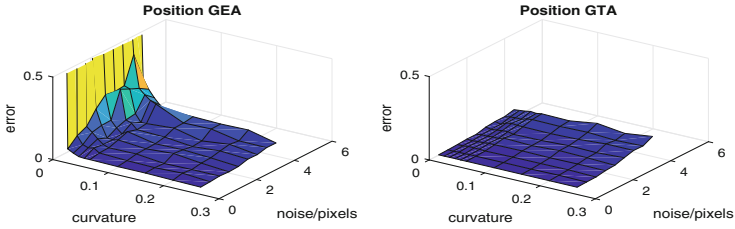
Fig. 7. The average error of GEA (left) and GTA (right) on the syntactic dataset with varying degrees of collinearity and observation noise. A large dependency on the curvature parameter can be seen for GEA, whereas GTA is largely independent of the curvature parameter.

7   Discussion

7.1   General Motion

In the setting of general motion, both GEA and GTA have a large convergence basin and tend to converge close to the bundle adjustment solution. Moreover the methods typically only require few iterations to convergence, despite poor initial guesses and large amounts of cameras. A pattern can be observed where GTA tends to have a lower position error compared to GEA.

7.2   Near Collinear Motion

GEA showed a significant performance decrease in the near collinear case. If the observation noise is sufficiently low, the constraint history is long enough and that the initial guess is very close to the optimal solution, GEA could converge. However, if these conditions are not met, the solution quickly deteriorates or diverge. This issue was described in [22] where isolated chains of cameras with collinear camera centers could gain additional degrees of freedom.

This is expected behavior since if only the bifocal constraint is used and the cameras have moved on a line, the camera translations can theoretically not be resolved [12]. In contrast, this is no restriction when using the trifocal constraint, which can resolve even collinear cameras. This explains why GTA performed better in the near collinear case compared to GEA.

7.3   Relation to Constraint History

Considering the relationship between the errors and the constraint history it can be seen that GEA consistently has a larger position error for short constraint histories compared to GTA, which is largely unaffected by the constraint history.

If the constraint history is restricted, the local camera configuration will become more collinear. If the constraint history is progressively restricted, the cameras may eventually become too collinear in relation to the noise for GEA

to accurately estimate the translations. As stated before, GTA does not suffer from the collinearity problem and performs equally well in the exact collinear case as in the case of more general case. This explain why GTA is not affected as much by the constraint history.

8   Conclusions

In this paper a novel extension to the Global Epipolar Adjustment method has been presented called Global Trifocal Adjustment, which is based on the well known trifocal constraint. It has been shown that the method has a large convergence basin, can handle the degenerate configurations of GEA, is less sensitive to the constraint history compared to GEA and in general only require few iterations to converge close to the bundle adjustment solution. Furthermore GTA can, in addition to points, also use lines and combinations of linens and points to form the cost matrices, while GEA is restricted to using only points.

The method is completely structure-less leading to a vast reduction in the optimization parameters compared to bundle adjustment. Furthermore the number of residuals can be reduced to 27 residuals per trifocal constraint irrespective of the number of feature correspondences. The cost matrices are independent of the camera parameters and need only be evaluated once and can be reused in future computations. The combination of decreased number of parameters, the large reduction in residuals and the constant cost matrices, leads to a potentially fast method well suited to real-time slam systems.

References

1. Agarwal, S., Mierle, K., et al.: Ceres solver. http://ceres-solver.org
2. Agarwal, S., Snavely, N., Simon, I., Seitz, S.M., Szeliski, R.: Building rome in a day. In: Proceedings 12th International Conference on Computer Vision, Kyoto, Japan (2009)
3. Byröd, M., Åström, K.: Efficient bundle adjustment using conjugate gradients with multiscale preconditioning. In: British Machine Vision Conference (2009)
4. Byröd, M., Åström, K.: Conjugate gradient bundle adjustment. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010. LNCS, vol. 6312, pp. 114–127. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-15552-9_9
5. Cornelis, N., Leibe, B., Cornelis, K., Van Gool, L.: 3D urban scene modeling integrating recognition and reconstruction. Int. J. Comput. Vis. (IJCV) **78**(2–3), 121–141 (2008)
6. Engels, C., Stewenius, H., Nister, D.: Bundle adjustment rules. In: PCV 2006 (2006)
7. Enqvist, O., Kahl, F., Olsson, C.: Non-sequential structure from motion (2011)

8. Gauss, C.F.: Theoria Motus Corporum Coelestium in sectionibus conicis solem ambientium. Göttingen (1809)
9. Granshaw, S.I.: Bundle adjustment methods in engineering photogrammetry. Photogram. Rec. **10**(56), 181–207 (1980)
10. Hartley, R., Zisserman, A.: Multiple View Geometry in Computer Vision, Second Edn. Cambridge University Press, Cambridge (2004)
11. Heyden, A.: Tensorial properties of multiple view constraints. Math. Methods Appl. Sci. **23**(2), 169–202 (2000)
12. Heyden, A., Åström, K.: Algebraic properties of multilinear constraints. Math. Methods Appl. Sci. **20**(13), 1135–1162 (1997)
13. Indelman, V., Roberts, R., Beall, C., Dellaert, F.: Incremental light bundle adjustment, 01 2012
14. Klein, G., Murray, D.: Parallel tracking and mapping for small AR workspaces. In: 6th IEEE and ACM International Symposium on Mixed and Augmented Reality, ISMAR 2007, pp. 225–234. IEEE (2007)
15. Legendre, A.M.: Nouvelles méthodes pour la détermination des orbites des comètes. F. Didot (1805)
16. Levenberg, K.: A method for the solution of certain non-linear problems in least squares. Q. Appl. Math. **2**(2), 164–168 (1944)
17. Rodrıguez López, A.L.: Algebraic epipolar constraints for efficient structureless multiview motion estimation. Ph.D. thesis, Universidad de Murcia. Facultad de Informática. Departamento de Ingeniería y Tecnología de Computadores (2013)
18. Lourakis, M.I.A., Argyros, A.A.: SBA: a software package for generic sparse bundle adjustment. ACM Trans. Math. Softw. **36**(1), 1–30 (2009)
19. Marquardt, D.W.: An algorithm for least-squares estimation of nonlinear parameters. J. Soc. Ind. Appl. Math. **11**(2), 431–441 (1963)
20. Akbarzadeh, A., Mordohai, P.: Towards urban 3D reconstruction from video (2006)
21. Ressl, C.: Geometry, constraints and computation of the trifocal tensor. Ph.D. thesis, Technische Universität Wien, Austria (2003)
22. Rodríguez, A.L., López-de Teruel, P.E., Ruiz, A.: Reduced epipolar cost for accelerated incremental SfM. In: 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3097–3104. IEEE (2011)
23. Snavely, N., Seitz, S.M., Szeliski, R.: Modeling the world from internet photo collections. Int. J. Comput. Vis. **80**(2), 189–210 (2007)
24. Snavely, N., Seitz, S.M., Szeliski, R.: Photo tourism: exploring photo collections in 3D. In: ACM SIGGRAPH 2006 Papers, SIGGRAPH 2006, pp. 835–846. ACM, New York (2006)
25. Triggs, B., McLauchlan, P.F., Hartley, R.I., Fitzgibbon, A.W.: Bundle adjustment—a modern synthesis. In: Triggs, B., Zisserman, A., Szeliski, R. (eds.) IWVA 1999. LNCS, vol. 1883, pp. 298–372. Springer, Heidelberg (2000). https://doi.org/10.1007/3-540-44480-7_21
26. VGG: Visual geometry group - data (2004). http://www.robots.ox.ac.uk/~vgg/data/data-mview.html. Accessed 12 July 2018