Noname manuscript No.
(will be inserted by the editor)

On Camera Calibration with Linear Programming and Loop Constraint Linearization

Jérôme Courchay · Arnak Dalalyan · Renaud Keriven · Peter Sturm

Received: date / Accepted: date

Abstract A technique for calibrating a network of perspective cameras based on their graph of trifocal tensors is presented. After estimating a set of reliable epipolar geometries, a parameterization of the graph of trifocal tensors is proposed in which each trifocal tensor is linearly encoded by a 4-vector. The strength of this parameterization is that the homographies relating two adjacent trifocal tensors, as well as the projection matrices depend linearly on the parameters. Two methods for estimating these parameters in a global way taking into account loops in the graph are developed. Both methods are based on sequential linear programming: the first relies on a locally linear approximation of the polynomials involved in the loop constraints whereas the second uses alternating minimization. Both methods have the advantage of being non-incremental and of uniformly distributing the error across all the cameras. Experiments carried out on several real data sets demonstrate the accuracy of the proposed approach and its efficiency in distributing errors over the whole set of cameras.

Keywords structure from motion · camera calibration · sequential linear programming

Jérôme Courchay, Arnak Dalalyan, Renaud Keriven
IMAGINE, LIGM, Université Paris-Est
Tel.: (+33) 1 64 15 21 72
Fax: (+33) 1 64 15 21 99
E-mail: courchay,dalalyan,keriven@imagine.enpc.fr

Peter Sturm
Laboratoire Jean Kuntzmann,
INRIA Grenoble Rhône-Alpes
E-mail: Peter.Sturm@inrialpes.fr

1 Introduction

Camera calibration from images of a 3-dimensional scene has always been a central issue in Computer Vision. The success of textbooks like [1,2] attests this interest. In recent years, many methods for calibration have been proposed. Most of these works either rely on known or partially known internal calibrations [3–10] or deal with an ordered sequence of cameras [11,12,14,15]. In many practical situations, however, the internal parameters of cameras are unavailable or available but very inaccurate. The absence of an order in the set of cameras is also very common when processing, for instance, Internet images.

In this paper, we deal with the problem of calibrating a network of cameras from a set of unordered images, the main emphasis being on the accuracy of the projective reconstruction of camera matrices. Traditionally, this situation is handled by factorizing the measurement matrix [16,17], which may be subject to missing data [18,19] because of occlusions. The methodology adopted in the present work is substantially different and is based on the notion of the graph of trifocal tensors rather than on the factorization. The experiments on real data sets show that our approach leads to highly competitive results that furnish a good initialization to the bundle adjustment (BA) algorithm [20].

Even in the case of calibrated cameras, most of the aforementioned methods are based on a graph of cameras (in which the edges are the epipolar geometries) which is made acyclic by discarding several edges. On the other hand, a number of recent studies, oriented toward city modeling from car or aerial sequences, point out the benefits of enforcing loop constraints. Considering loops in the graph of cameras has the advantage of reducing the drift due to errors induced while processing the trajectory sequentially (cf. Fig. 1). [21]

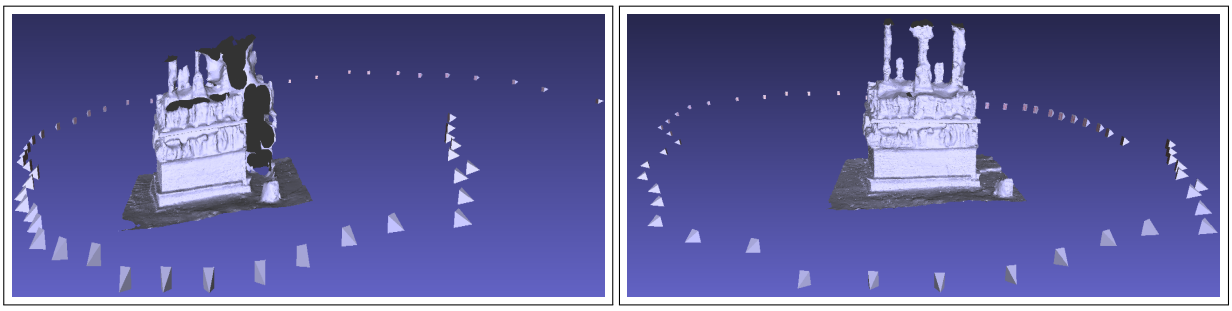


Fig. 1 Multi-view stereo reconstruction [26] using cameras calibrated without (left) and with (right) using the loop constraint. When the loop constraint is not enforced, the accumulation of errors results in an extremely poor reconstruction.

merges partial reconstructions, [22] constrains consistent rotations for loops and planar motion. Adapted to their specific input, these papers often rely on trajectory regularization or dense matching [23,24]. [25] is a notable exception, where loop constraints are added to sparse Structure from Motion (SfM), yet taking as input an ordered omnidirectional sequence and assuming known internal parameters.

Using the graph of triplets instead of the graph of images or image pairs confers a number of important advantages: the strong geometric trifocal constraint means that very few false correspondences are encountered, there is no need to distinguish between feasible and infeasible paths as it is done in [43] and the risk of having problems with critical surfaces is lessened [40].

The method proposed in the present work consists of the following points:

1. Our starting point is a set of unknown cameras linked by estimated epipolar geometries (EG). These are computed using a state-of-the-art version of RANSAC [27], followed by a maximum likelihood improvement described in [14]. We assume that along with the estimated fundamental matrices, reliable epipolar correspondences are known. These correspondences are made robust by simultaneously considering several camera pairs, like in [3]. This produces a set of three-view correspondences that will be used in the sequel.
2. We group views into triplets. Three views (i, j, k) are considered as a valid triplet if (a) the EGs between i and j as well as between j and k have been successfully computed at the previous step and (b) there are at least 4 three-view correspondences in these images. To reduce the number of nodes, some of the estimated epipolar geometries are ignored, so that inside a triplet, only two of the three fundamental matrices are considered known. The advantage of this strategy is that we do not need to enforce the coherence of fundamental matrices. At first sight, this can be seen as a loss of information. However, this information is actually recovered via trifocal tensors.
3. We define a graph having as nodes valid camera triplets. Therefore, there are two fundamental matrices available for each node. Two nodes are connected by an edge if they share a fundamental matrix. We demonstrate that for each node there exists a 4-vector such that all the entries of the three camera matrices are affine functions with known coefficients of this 4-vector. Moreover, the homographies that allow the registration of two adjacent nodes ν and ν' are also affine functions depending on 4 out of the 8 unknown parameters corresponding to ν and ν'. To speed-up the computations, for each node only 50 (or less) three-view correspondences that are the most compatible with the EGs are used.
4. If the graph of triplets is acyclic, the equations of three-view correspondences for all nodes lead to a linear estimate of all the cameras. In case the graph of triplets contains one or several loops, each loop is encoded as a (non-linear) constraint on the unknowns. Starting from an initial value computed as a solution to the unconstrained least squares, we sequentially linearize the loop constraints and solve the resulting problem by (sparse) linear programming. This can be efficiently done even for very large graphs. It converges very rapidly, but the loop constraints are fulfilled only approximately.
5. In the case where the loop constraints are not satisfied exactly, we proceed by homography registration and estimation of camera matrices by linear least-squares under a norm constraint. This is done exactly via a singular value decomposition producing as output all cameras in a projective space. To provide a qualitative evaluation, we recover the metric space using an implementation of [28], and a single Euclidean bundle adjustment that refines the metric space and camera positions.
6. As an alternative to steps 4 and 5, we also propose a second approach based on alternating minimization. In all our experiments, this method converges to a solution for which the loop constraints are satisfied to very high precision. The main idea underlying this method is that when, in a loop, all but 4 trifocal tensors are given, the

loop constraints can be written as linear equations. A precise formulation and the proof of this are presented in Section 4 and the Appendix, respectively. While the method described above in steps 1.-5. appeared in our ECCV paper [13], this part is new and leads to improved results in terms of the achievability of loop constraints.

Thus, we propose a method that accurately recovers geometries, without any sequential process, and attempts to enforce the compatibility of cameras within loops in the early stages of the procedure. An important advantage conferred by our approach is that the number of unknown parameters is kept fairly small, since we consider only the cameras (four unknowns for each triplet) and not the 3D points. Our reconstruction is further refined by bundle adjustment. Taking loops into account and avoiding error accumulation, the proposed solution is less prone to get stuck in local minima.

After submission of our manuscript, two papers, [41] and [42], related to the present work have been published. In [41], the authors aim at exploiting loops in the graph of images in order to detect and to identify incorrect geometric relations between images. The main common feature with our approach is that the loop consistency statistics are defined by chaining transformations over loops. Thus, there are analogies between their techniques and the ours but these techniques are applied to two different problems. On the other hand, the problem studied in [42] is the same as in the present paper, but the methodology is different. In fact, the authors of [42] propose a structure from motion pipeline—based on a hierarchical representation of the graph of images—capable to deal with large-scale data sets but bound to trees of images.

The remainder of the paper is organized as follows. Section 2 presents the background theory and terminology. Our algorithms are thoroughly described in Sections 3 and 4. The results of numerical experiments conducted on several real data sets as well as a comparison to state-of-the-art software are provided in Section 5.

2 Background

In this work, we consider a network of N uncalibrated cameras and assume that for some pairs of cameras (i, j), where $i, j = 1, \ldots, N$, $i \neq j$, an estimation of the fundamental matrix, denoted by $\mathtt{F}^{i,j}$, is available. Let us denote by $\mathbf{e}^{i,j}$ the unit norm epipole in view j of camera center i. Recall that the fundamental matrix leads to a projective reconstruction of camera matrices $(\mathtt{P}^i, \mathtt{P}^j)$, which is unique up to a homography.

The geometry of three views i, j and k is described by the Trifocal Tensor, hereafter denoted by $\mathscr{T}^{i,j,k}$. It consists of three 3×3 matrices: $\mathtt{T}_1^{i,j,k}$,$\mathtt{T}_2^{i,j,k}$ and $\mathtt{T}_3^{i,j,k}$ and provides a particularly elegant description of point-line-line correspondences in terms of linear equations

$$\mathbf{p}_i^\mathsf{T} \begin{bmatrix} \mathbf{l}_j^\mathsf{T} \mathtt{T}_1^{i,j,k} \\ \mathbf{l}_j^\mathsf{T} \mathtt{T}_2^{i,j,k} \\ \mathbf{l}_j^\mathsf{T} \mathtt{T}_3^{i,j,k} \end{bmatrix} \mathbf{l}_k = 0, \tag{1}$$

where $\mathbf{p}_i$ is a point in image i (seen as a point in projective space $\mathbb{P}^2$) which is in correspondence with the line $\mathbf{l}_j$ in image j and with the line $\mathbf{l}_k$ in image k. Considering the entries of $\mathscr{T}^{i,j,k}$ as unknowns, we get thus one linear equation for each point-line-line correspondence. Therefore, one point-point-point correspondence $\mathbf{p}_i \longleftrightarrow \mathbf{p}_j \longleftrightarrow \mathbf{p}_k$ leads to 4 independent linear equations by combining an independent pair of lines passing through $\mathbf{p}_j$ in image j with an independent pair of lines passing through $\mathbf{p}_k$ in image k.

Since a Trifocal Tensor has 27 entries, the previous argument shows that 7 point-point-point correspondences suffice for recovering the Trifocal Tensor as a solution of an overdetermined system of linear equations. Recall however that the Trifocal Tensor has only 18 degrees of freedom. Most algorithms estimating a Trifocal Tensor from noisy point-point-point correspondences compute an approximate solution to the linear system by a least squares estimator (LSE) and then perform a post-processing in order to get a valid Trifocal Tensor. An alternative approach consists in using a minimal solution that determines the three-view geometry from six points [29,30].

2.1 Parameterization of the set of trifocal tensors

Let us describe now two elementary results that represent the building blocks of our approach, relying on the fact that when two out of three fundamental matrices are known, the Trifocal Tensor has exactly 4 degrees of freedom.

Proposition 1 *For three views i, j and k, given two fundamental matrices $\mathtt{F}^{i,j}$ and $\mathtt{F}^{i,k}$, there exists a 4-vector $\gamma = [\gamma_0, \gamma_1, \gamma_2, \gamma_3]$ such that $\mathscr{T}^{i,j,k}$ is given by:*

$$T_t^{i,j,k} = A_t^{i,j} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & \gamma_0 \\ 0 & 1 & \gamma_t \end{bmatrix} (A_t^{i,k})^\mathsf{T} \tag{2}$$

for every $t = 1,2,3$, where $A_t^{i,s} = \left[(F_{t,1:3}^{i,s})^\mathsf{T} , (F_{t,1:3}^{i,s})^\mathsf{T} \times \mathbf{e}^{i,s} , \mathbf{e}^{i,s} \right]$, for $s = j,k$. Moreover, $\mathscr{T}^{i,j,k}$ is geometrically valid, i.e. , there exist 3 camera matrices $\mathtt{P}^i$, $\mathtt{P}^j$ and $\mathtt{P}^k$ compatible with $\mathtt{F}^{i,j}$ and $\mathtt{F}^{i,k}$ and having $\mathscr{T}^{i,j,k}$ as the Trifocal Tensor.

The proof of this result is deferred to the appendix. It is noteworthy that the claims of Proposition 1 hold true under full

generality, even if the centers of three cameras are collinear. In view of [1], the camera matrices parameterized by γ that are compatible with the fundamental matrices $\mathtt{F}^{i,j}$ and $\mathtt{F}^{i,k}$ as well as with the Trifocal Tensor defined by Eq. (2) are given by (up to a projective homography)

$$\begin{array}{l} \mathtt{P}^i = [\mathtt{I}_{3\times 3}\,|\,\mathbf{0}_{3\times 1}], \qquad \mathtt{P}^k = [[\mathbf{e}^{i,k}]_\times \mathtt{F}^{k,i}\,|\,\mathbf{e}^{i,k}], \\ \mathtt{P}^j = \mathrm{kron}\big([\gamma_{1:3},1];\mathbf{e}^{i,j}\big) - \gamma_0[[\mathbf{e}^{i,j}]_\times \mathtt{F}^{j,i}\,|\,\mathbf{0}_{3\times 1}], \end{array} \tag{3}$$

where $\mathrm{kron}(\cdot,\cdot)$ stands for the Kronecker product of two matrices. These cameras are obtained directly using results in [1, Eq. 15.1].

In the noiseless setting, Proposition 1 offers a minimal way of computing the 4 remaining unknowns from point-point-point correspondences. One could think that one point-point-point correspondence leading to 4 equations is enough for retrieving the 4 unknowns. However, since two EGs are known, only one equation brings new information from one point-point-point correspondence. So we need at least 4 point-point-point correspondences to compute the Trifocal Tensor compatible with the two given fundamental matrices. In the noisy case, if we use all 4 equations associated to point-point-point correspondences, the system becomes overdetermined and one usually proceeds by computing the LSE.

The second ingredient in our approach is the parameterization of the homography that bridges two camera triplets having one fundamental matrix in common. Let i, j, k and ℓ be four views such that (a) for views i and k we have successfully estimated the fundamental matrix $\mathtt{F}^{i,k}$ and (b) for each triplet (i,j,k) and (k,i,ℓ) the estimates of two fundamental matrices are available. Thus, the triplets (i,j,k) and (k,i,ℓ) share the same fundamental matrix $\mathtt{F}^{i,k}$. Using equations (3), one obtains two projective reconstructions of camera matrices of views i and k based on two 4-vectors γ and $\underline{\gamma}$. Let us denote the reconstruction from the triplet (i,j,k) (resp. (k,i,ℓ)) by $\mathtt{P}^i_\gamma$ and $\mathtt{P}^k_\gamma$ (resp. $\mathtt{P}^i_{\underline{\gamma}}$ and $\mathtt{P}^k_{\underline{\gamma}}$). If the centers of cameras i and k differ, then there is a unique homography $\underline{\mathtt{H}}$ such that

$$\mathtt{P}^i_\gamma \underline{\mathtt{H}} \cong \mathtt{P}^i_{\underline{\gamma}}, \quad \mathtt{P}^k_\gamma \underline{\mathtt{H}} \cong \mathtt{P}^k_{\underline{\gamma}}, \tag{4}$$

where $\cong$ denotes equality up to a scale factor. Considering the camera matrices as known, one can solve equations (4) w.r.t. $\underline{\mathtt{H}}$ and then compute its inverse $\mathtt{H}$. Setting $\lambda = -\frac{1}{2}\,\mathrm{tr}([\mathbf{e}^{i,k}]_\times \mathtt{F}^{k,i}[\mathbf{e}^{k,i}]_\times \mathtt{F}^{i,k})$, one readily checks that [1]

$$\underline{\mathtt{H}} = \left[\begin{array}{c|c} \mathrm{kron}(\underline{\gamma}_{1:3}, \mathbf{e}^{k,i}) - \underline{\gamma}_0[\mathbf{e}^{k,i}]_\times \mathtt{F}^{i,k} & \mathbf{e}^{k,i} \\ \hline \lambda\,(\mathbf{e}^{i,k})^\mathsf{T} & 0 \end{array}\right], \tag{5}$$

[1] See Appendix B for more details

and

$$\mathtt{H} = \left[\begin{array}{c|c} [\mathbf{e}^{i,k}]_\times \mathtt{F}^{k,i} & \mathbf{e}^{i,k} \\ \hline \underline{\gamma}_0 \lambda\,(\mathbf{e}^{k,i})^\mathsf{T} - \underline{\gamma}_{1:3}^\mathsf{T}[\mathbf{e}^{i,k}]_\times \mathtt{F}^{k,i} & -\underline{\gamma}_{1:3}\mathbf{e}^{i,k} \end{array}\right]. \tag{6}$$

To sum up this section, let us stress that the main message to retain from all these formulas is that the homographies $\mathtt{H}$ and $\underline{\mathtt{H}}$, as well as the camera matrices given by (3) are linear in $(\gamma,\underline{\gamma})$.

2.2 Graph of trifocal tensors and loop constraints

This section contains the core of our contribution which relies on a graph-based representation of the set of camera triplets. This is closely related to the framework developed in [5], where the graph of camera pairs is considered. The advantage of operating with triplets instead of pairs is that there is no need to distinguish between feasible and infeasible paths.

The starting point for our algorithm is a set of estimated EGs that allows us to define a graph $\mathscr{G}_{\text{cam}}$ so that (a) $\mathscr{G}_{\text{cam}}$ has N nodes corresponding to the N cameras and (b) two nodes of $\mathscr{G}_{\text{cam}}$ are connected by an edge if a reliable estimation of the corresponding epipolar geometry is available. Then, a triplet of nodes i,j,k of $\mathscr{G}_{\text{cam}}$ is called valid if

- there is a sufficient number of three-view correspondences between i, j and k,
- at least two out of three pairs of nodes are adjacent in $\mathscr{G}_{\text{cam}}$.

If for some valid triplet all three EGs are available, we remove the least reliable one (according to the Geometric Robust Information Criterion score [32]) and define the graph $\mathscr{G}_{\text{triplet}} = (\mathscr{V}_{\text{triplet}}, \mathscr{E}_{\text{triplet}})$ having as nodes valid triplets of cameras and as edges the pairs of triplets that have one fundamental matrix in common. In view of Proposition 1, the global calibration of the network is equivalent to the estimation of a 4-vector for each triplet of cameras. Thus, to each node v of the graph of triplets we associate a vector $\gamma^v \in \mathbb{R}^4$. The large vector $\Gamma = (\gamma^v : v \in \mathscr{V}_{\text{triplet}})$ is the parameter of interest in our framework.

2.2.1 Tree of Trifocal Tensors

If, by some chance, it turns out that the graph of triplets is acyclic, then the problem of estimating Γ reduces to estimating $N_V = \mathrm{Card}(\mathscr{V}_{\text{triplet}})$ independent vectors γ^v. This task can be effectively accomplished using point-point-point correspondences and the equation (1). As explained in Section 2,

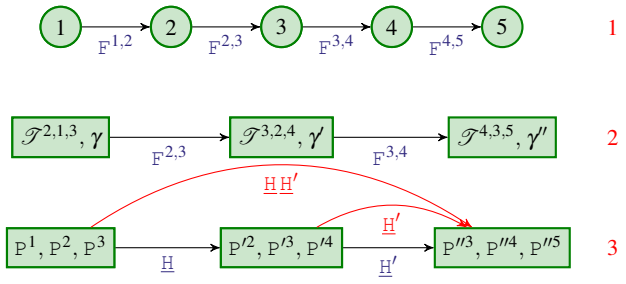


Fig. 2 The first row shows the graph of cameras with estimated fundamental matrices. Based on this graph, we construct the graph of camera triplets (as shown on the second row) and estimate the unknowns γ by RANSAC. The third row illustrates that we transfer all the cameras to the same projective space (that of $\mathscr{T}^{4,3,5}$).

a few point-point-point correspondences suffice for computing an estimator of γ^v by least squares. In our implementation, we use RANSAC with a minimal configuration of four 3-view correspondences in order to perform robust estimation.

Once the unknowns γ^v are recovered, the set of associated cameras can be reconstructed in a unique manner. As shown in Fig. 2, one can transport all the camera triplets to the same projective space by applying the homographies computed using Equation 5. This can be seen by considering the degrees of freedom (DF) of a network of cameras such that the corresponding graph of trifocal tensors is acyclic. This network has $11N-15$ DF since each one of N cameras has 11 DF and they are defined up to an overall 4×4 homography, having 15 DF. On the other hand, our parametrization—in the case of an acyclic graph of trifocal tensors—is composed of $N-1$ fundamental matrices and $N-2$ vectors $\gamma\in\mathbb{R}^4$. Thus, the total number of parameters is $7(N-1)+4(N-2)=11N-15$, which coincides with the number of DF.

2.2.2 *Loop of Trifocal Tensors*

However, acyclic graphs are the exception rather than the rule. Even if the camera graph is acyclic, the resulting triplet graph may contain loops. If there is a loop, then one has 15 loop constraints, *i.e.* one additional fundamental matrix and two additional trifocal tensors as shown in Fig. 3. These constraints translate the fact that the camera triplet of Trifocal Tensor $\mathscr{T}^{2,1,3}$ obtained by iteratively transferring it via the homographies $\underline{\mathrm{H}}$ along the loop matches the initial camera triplet. Denoting by $\{\underline{\mathrm{H}}^{1,2},\underline{\mathrm{H}}^{2,3},\ldots,\underline{\mathrm{H}}^{N,1}\}$ the homographies that connect the nodes of the loop of Trifocal Tensors, the constraints can be written in the form:

$$\prod_{i=1}^{N}\underline{\mathrm{H}}^{i,i+1}\cong\mathtt{I}. \tag{7}$$

Equation (7) defines a set of 15 polynomial constraints on

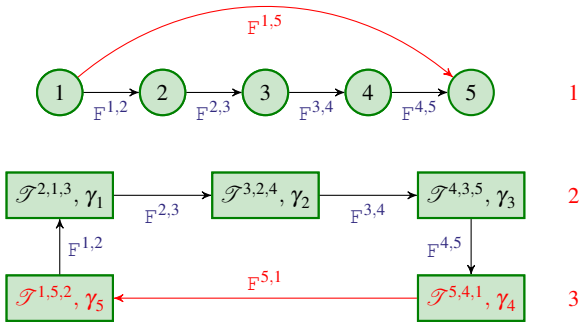


Fig. 3 The first row shows an example of a graph of cameras forming a loop; if we compare with the tree presented in Fig. 2, there is one additional fundamental matrix leading to 7 additional constraints. The second and the third rows illustrate the loop of camera triplets. There are two additional reduced tensors (cf. row 3) giving raise to 8 new constraints.

the unknown vector Γ. If the triplet graph contains N_{loop} loops, then we end up with $15N_{\text{loop}}$ constraints. To give more details, let us remark that every loop constraint (7) can be rewritten as $f_j(\Gamma)=0$, $j=1,\ldots,15$, for some polynomial functions f_j. Gathering these constraints for all N_{loop} loops, we get

$$f_j(\Gamma)=0,\quad j=1,\ldots,15N_{\text{loop}}. \tag{8}$$

On the other hand, in view of (1) and (2), the point-point-point correspondences can be expressed as an inhomogeneous linear equation system in Γ

$$\mathtt{M}\Gamma=\mathbf{m}, \tag{9}$$

where $\mathtt{M}$ is a $4N_{3\text{-corr}}\times 4N$ matrix and $\mathbf{m}$ is a $4N_{3\text{-corr}}$ vector with $N_{3\text{-corr}}$ being the number of correspondences across three views. The matrix $\mathtt{M}$ and the vector $\mathbf{m}$ are computed using the known fundamental matrices. Since in practice these matrices are estimated from available data, the system (9) need not be satisfied exactly. Then, it is natural to estimate the parameter-vector Γ by solving the problem

$$\begin{aligned}&\min \|\mathtt{M}\Gamma-\mathbf{m}\|_q^q\\ \text{subject to}\quad &f_j(\Gamma)=0,\ \forall j=1,\ldots,15N_{\text{loop}},\end{aligned} \tag{10}$$

for some $q\geq 1$. Unfortunately, this problem is non-convex and, therefore, it is very hard to solve. To cope with this issue, we propose two strategies based on local linearization and alternating minimization.

The main advantage of this approach is that if a solution to the proposed optimization problem is found, it is guaranteed to be consistent w.r.t. the loops, meaning that each camera matrix will be uniquely determined up to a scale factor and an overall homography ambiguity.

3 First approach : sequential linear programming and homography registration

We present now in full details the first approach to solving for projective structure from motion. The second approach, which differs from the first one only in the way the loop constraints are dealt with, will be described in the next section.

3.1 Constraint linearization

Instead of solving the optimization problem that is obtained by combining the LSE with the loop constraints, we propose here to replace it by a linear program.

We start by computing an initial estimator of Γ, *e.g.*, by solving the unconstrained (convex) problem with some $q \geq 1$. In our implementation, we use RANSAC with $q = 2$ for ensuring robustness to erroneous three-view correspondences. Then, given an initial estimator Γ_0, we define the sequence Γ_k by the following recursive relation: Γ_{k+1} is the solution to the linear program (LP)

$$\begin{aligned} &\min \|\mathtt{M}\Gamma - \mathbf{m}\|_1 \\ \text{subject to} \quad &|f_j(\Gamma_k) + \nabla f_j(\Gamma_k)(\Gamma - \Gamma_k)| \leq \varepsilon, \end{aligned} \tag{11}$$

where ε is a small parameter (we use $\varepsilon = 10^{-6}$).

There are many softwares, such as GLPK, SeDuMi, SDP3, for solving problem (11) with highly attractive execution times even for thousands of constraints and variables. Furthermore, empirical experience shows that the sequence Γ_k converges very rapidly. Typically, a solution with satisfactory accuracy is obtained after five to ten iterations. Note that a similar approach for imposing loop constraints has been proposed in [38], where several other techniques of relaxation of non-linear constraints in (10) are considered.

Remark 1 Since our parameterization of the network of cameras relies on a set of estimated fundamental matrices, LP (11) might have no solution, *i.e.*, there is no vector Γ satisfying all the constraints in (11). We did not observe such a situation in our experiments, since in all our examples the fundamental matrices have been estimated very accurately and the number of unknowns in (11) was much larger than the number of constraints. However, if for some data set the LP is infeasible, it is possible to increase the tuning parameter ε so that the resulting LP becomes feasible.

3.2 Accounting for heteroscedasticity

The goal now is to make the energy that we minimize in (11), which is purely algebraic, meaningful from a statistical viewpoint. Assume equations (9) are satisfied up to an additive random noise: $\mathtt{M}\Gamma = \mathbf{m} + \xi$, where the random vector ξ has independent coordinates drawn from the centered Laplace distribution with constant scale. Then the energy in (11) is proportional to the negative log-likelihood. The constancy of the scale factor means that the errors are homoscedastic, which is a very strong hypothesis. We observed that all three-view correspondences recorded by a fixed triplet have nearly the same scale for the errors, while the scales for different triplets are highly variable (cf. Fig. 4).

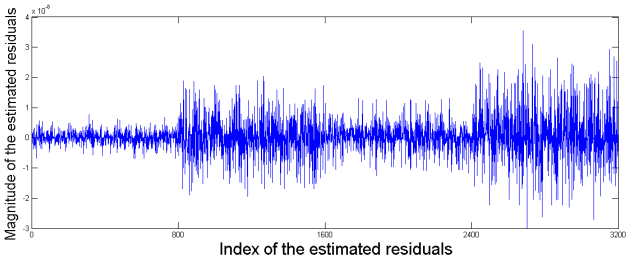


Fig. 4 This figure illustrates the heteroscedasticity of the noise in Eq. (9) when the matrix M and the vector **m** are computed using estimated EGs. The estimated residuals for 4 nodes of the triplet graph are plotted, with 800 measurements available for each node.

To account for this heteroscedasticity of the noise, we use the initial estimator of Γ to estimate one scale parameter σ_v per node $v \in \mathscr{V}_{\text{triplet}}$. This is done by computing the standard deviation of the estimated residuals. Using $\{\sigma_v\}$, the energy in problem (11) is replaced by $\sum_v \|\mathtt{M}_v\Gamma - \mathbf{m}_v\|_1/\sigma_v$. Here, $\mathtt{M}_v$ is the submatrix of M containing only those rows that are obtained from three-view correspondences recorded by v. The vector $\mathbf{m}_v$ is obtained from $\mathbf{m}$ in the same way.

3.3 Homography registration and estimation of projection matrices

Assume that we have a graph of trifocal tensors, $\mathscr{G}_{\text{triplet}}$, the nodes of which are denoted by $v_1, v_2, \ldots, v_n$. In the previous step, we have determined parameters $\gamma_1, \ldots, \gamma_n$, such that γ_i characterizes the trifocal tensor represented by v_i. A naive strategy for estimating camera matrices is to set one of the cameras equal to $[\mathtt{I}_{3\times 3} \,|\, \mathbf{0}_{3\times 1}]$ and to recover the other cameras by successive applications of the homographies $\underline{\mathtt{H}}$ to the camera matrices reconstructed according to (3).

However, in general situations, the vector Γ computed by sequential linear programming as described in the previous section satisfies the loop constraints up to a small error. Therefore, the aforementioned naive strategy has the drawback of increasing the error of estimation for cameras computed using many homographies $\underline{\mathtt{H}}$.

In order to avoid this and to uniformly distribute the estimation error over the set of camera matrices, we propose

a method based on homography registration by SVD. Thus, the input for the method described in this section is a vector Γ for which the loop constraints are satisfied up to a small estimation error.

3.3.1 The case of a single loop

We assume in this subsection that $\mathscr{G}_{\text{triplet}}$ reduces to one loop, that is each node v_i has exactly two neighbors v_{i-1} and v_{i+1} with standard convention and $v_{n+i} = v_i$ for all i. (This applies to all the indices in this subsection.) For each node v_i representing three views, we have already computed a version of the projection matrices $\mathtt{P}^{1,\gamma_i}, \mathtt{P}^{2,\gamma_i}, \mathtt{P}^{3,\gamma_i}$. Furthermore, for two neighboring nodes v_i and v_{i+1} we have computed a homography $\mathtt{H}^{i,i+1}$ so that

$$\mathtt{P}^{j,\gamma_{i+1}} \cong \mathtt{P}^{j+1,\gamma_i}\mathtt{H}^{i,i+1}, \quad j \in \{1,2\}. \tag{12}$$

Based on the relative homographies $\{\mathtt{H}^{i,i+1}\}$ we want to recover absolute homographies $\mathtt{H}^{v_i}$ that allow to represent all the matrices $\mathtt{P}^{j,\gamma_i}$ in a common projective frame. In other terms, in the ideal case where there is no estimation error, the matrices $\mathtt{H}^{v_i}$ should satisfy

$$\mathtt{P}^{j,\gamma_i}\,\mathtt{H}^{v_i} \cong \mathtt{P}^{j+i-1,*}, \qquad j \in \{1,2,3\}. \tag{13}$$

Obviously, the set $\{\mathtt{H}^{v_i}\}$ can only be determined up to an overall projective homography.

Proposition 2 *If for some $i = 1,\ldots,n$, the cameras $\mathtt{P}^{i+1,*}$ and $\mathtt{P}^{i+2,*}$ have different centers, then $\mathtt{H}^{v_i} \cong \mathtt{H}^{i,i+1}\mathtt{H}^{v_{i+1}}$. Furthermore, if the centers of each pair of consecutive cameras are different, then one can find a projective coordinate frame so that*

i) $\mathtt{H}^{v_i} = \mathtt{H}^{i,i+1}\mathtt{H}^{v_{i+1}}, \qquad \forall i = 1,\ldots,n-1,$

ii) $\alpha\,\mathtt{H}^{v_n} = \mathtt{H}^{n,1}\mathtt{H}^{v_1}$, *where* $\alpha = \frac{1}{4}\text{Trace}(\prod_{i=1}^{n}\mathtt{H}^{i,i+1})$,

iii) Let $\bar{\mathtt{H}}$ *be the* $(4n)\times 4$ *matrix resulting from the vertical concatenation of matrices* $\mathtt{H}^{v_i}$. *The four columns of* $\bar{\mathtt{H}}$ *are orthonormal.*

This result, the proof of which is presented in the Appendix, allows us to define the following algorithm for estimating the matrices $\{\mathtt{H}^{v_i}\}$. Given the relative homographies $\{\mathtt{H}^{i,i+1}\}$, we first compute α according to the formula in ii) and then minimize the cost function

$$\sum_{i=1}^{n-1}\frac{\|\mathtt{H}^{v_i} - \mathtt{H}^{i,i+1}\mathtt{H}^{v_{i+1}}\|_2^2}{\max(\sigma_{v_i}^2,\sigma_{v_{i+1}}^2)} + \frac{\|\alpha\,\mathtt{H}^{v_n} - \mathtt{H}^{n,1}\mathtt{H}^{v_1}\|_2^2}{\max(\sigma_{v_1}^2,\sigma_{v_n}^2)} \tag{14}$$

w.r.t. $\{\mathtt{H}^{v_i}\}$, subject to the orthonormality of the columns of $\bar{\mathtt{H}}$. Here, $\|\cdot\|_2$ is the Frobenius norm. The exact solution of this (non-convex) optimization problem can be computed using the singular value decomposition of a matrix of size $4n\times 4n$ constructed from α and $\{\mathtt{H}^{i,i+1}\}$. Since this is quite standard (based on the Courant-Fischer minimax theorem [31, Thm. 8.1.2]), we do not present more details here.

Fig. 5 This figure illustrates the improvement achieved at each step of our algorithm. If the cameras are reconstructed without imposing loop constraints, the epipolar lines between the first and the last frames are extremely inaccurate (1st row). They become much more accurate when the constrained optimization is performed (2nd row). Finally, the result is almost perfect once the homography registration is done (3rd row). Figure can be best viewed under pdf magnification.

Remark 2 An alternative to the approach we have just described consists in considering the homographies as elements of a Lie group and in estimating the absolute homographies by Lie averaging of relative motions as described in [39].

3.3.2 The case of several loops

Assume now that we have identified several loops in the graph of trifocal tensors. Let N_{loop} be the number of these loops. We apply to each loop the method of the previous section and get a homography for every node of the loop. In general, one node of $\mathscr{G}_{\text{triplet}}$ may lie in several loops, in which case we will have several homographies for that node. It is then necessary to enforce the coherence of these homographies. To this end, we define the graph $\mathscr{G}_{\text{loop}}$ having N_{loop} nodes, each node representing a loop. Two nodes of $\mathscr{G}_{\text{loop}}$ are linked by an edge, if the corresponding loops have non-empty intersection. We will assume that the graph $\mathscr{G}_{\text{loop}}$ is connected, since otherwise it is impossible to simultaneously calibrate different connected components.

The next step consists in determining a minimal depth spanning tree $\mathscr{T}_{\text{loop}}$ of $\mathscr{G}_{\text{loop}}$. Since the number of loops is assumed small, this step will not be time consuming. Let $(\mathscr{L},\mathscr{L}')$ be a pair of adjacent nodes of $\mathscr{T}_{\text{loop}}$. By an argument analogous to that of Proposition 2, one can show that there exists a 4×4 homography $\mathtt{H}^{\mathscr{L},\mathscr{L}'}$ such that $\mathtt{H}^{v,\mathscr{L}} \cong \mathtt{H}^{v,\mathscr{L}'}\mathtt{H}^{\mathscr{L}',\mathscr{L}}$ up to an estimation error, for every triplet of cameras $v \in \mathscr{L}\cap\mathscr{L}'$. Here, $\mathtt{H}^{v,\mathscr{L}}$ (resp. $\mathtt{H}^{v,\mathscr{L}'}$) stands for the homography assigned (*cf.* previous subsection) to the triplet v as a part of the loop $\mathscr{L}$ (resp. $\mathscr{L}'$). The homography $\mathtt{H}^{\mathscr{L}',\mathscr{L}}$ can be estimated by minimizing the objective function

$$\sum_{v\in\mathscr{L}\cap\mathscr{L}'}\|\alpha_v\mathtt{H}^{v,\mathscr{L}} - \mathtt{H}^{v,\mathscr{L}'}\mathtt{H}^{\mathscr{L}',\mathscr{L}}\|_2^2/\sigma_v^2 \tag{15}$$

w.r.t. the matrix $\mathtt{H}^{\mathscr{L}',\mathscr{L}}$ and parameters $\{\alpha_v\}$ subject to

$$\|\mathtt{H}^{\mathscr{L}',\mathscr{L}}\|_2^2 + \sum_{v\in\mathscr{L}\cap\mathscr{L}'} \alpha_v^2 = 1. \tag{16}$$

Once again, this minimization can be carried out by computing the eigenvector corresponding to the smallest singular value of a suitably defined matrix.

Finally, to enforce the coherence of absolute homographies computed using different loops, we proceed as follows. We do not modify the homographies computed within the loop $\mathscr{L}_0$ constituting the root of the minimal depth spanning tree $\mathscr{T}_{\text{loop}}$. For any other loop $\mathscr{L}$, let $\mathscr{L}_0 \to \mathscr{L}_1 \to \ldots \to \mathscr{L}_k \to \mathscr{L}$ be the (unique) path joining $\mathscr{L}$ to the root. Then, every absolute homography $\mathtt{H}^{v,\mathscr{L}}$, $v \in \mathscr{L}$, computed within the loop $\mathscr{L}$ using the method of the previous subsection is replaced by $\mathtt{H}^{v,\mathscr{L}}\mathtt{H}^{\mathscr{L},\mathscr{L}_k}\cdots\mathtt{H}^{\mathscr{L}_1,\mathscr{L}_0}$. After this modification, the images by $\mathtt{H}^{v,\mathscr{L}}$ of the projection matrices $\mathtt{P}^{j,\gamma_v}$ ($j = 1,2,3$) will all lie in nearly the same projective space. This makes it possible to recover the final projection matrices $\mathtt{P}^i$ by a simple computation presented in the next subsection.

3.3.3 Estimating projection matrices

Once the set of absolute homographies estimated, we turn to the estimation of camera matrices $\{\mathtt{P}^{j,*}\}$. Due to the estimates computed in previous steps, each projection matrix $\mathtt{P}^{j,*}$ can be estimated independently of the others. To ease notation and since there is no loss of generality, let us focus on the estimation of $\mathtt{P}^{1,*}$. We start by determining the nodes in $\mathscr{G}_{\text{triplet}}$ that contain the first view. Let $\mathscr{V}_1$ denote the set of these nodes. To each node $v \in \mathscr{V}_1$ corresponds one estimator of $\mathtt{P}^{1,*}$, denoted by $\mathtt{P}^{1,\gamma_v}$. Furthermore, we have a set of estimated homographies $\mathtt{H}^{v,\mathscr{L}}$ that satisfy, up to an estimation error, the relation $\mathtt{P}^{1,v}\mathtt{H}^{v,\mathscr{L}} \cong \mathtt{P}^{1,*}$. This is equivalent to

$$\alpha_{v,\mathscr{L}}\mathtt{P}^{1,v}\mathtt{H}^{v,\mathscr{L}} = \mathtt{P}^{1,*}, \qquad \forall v \in \mathscr{V}_1,\ \forall \mathscr{L} \supset \{v\} \tag{17}$$

for some $\alpha_{v,\mathscr{L}} \in \mathbb{R}$. In Eq. (17), the unknowns are the scalars $\alpha_{v,\mathscr{L}}$ and the matrix $\mathtt{P}^{1,*}$. Since this matrix should be of rank 3, it has nonzero Frobenius norm. Therefore, we estimate $\mathtt{P}^{1,*}$ by $\mathtt{P}^1$ defined as a solution to

$$\arg\min_{\mathtt{P}} \min_{\|\mathtt{P}\|_2^2+\|\alpha\|_2^2=1} \sum_{\mathscr{L}} \sum_{v\in\mathscr{L}\cap\mathscr{V}_1} \frac{\|\alpha_{v,\mathscr{L}}\mathtt{P}^{1,v}\mathtt{H}^{v,\mathscr{L}} - \mathtt{P}\|_2^2}{\sigma_v^2}, \tag{18}$$

where α stands for the vector having as coordinates the numbers $\alpha_{v,\mathscr{L}}$. Once again, the problem (18) can be explicitly solved using the SVD of an appropriate matrix.

3.4 Summary of the first approach

Prior to switching to the second approach, let us provide a concise description of the pipeline presented in this section.

1. Use robust estimators of fundamental matrices for constructing the matrix $\mathtt{M}$, the vector $\mathbf{m}$ and the polynomial functions f_j involved in the optimization problem (10).
2. Find an initial estimator of Γ by solving the unconstrained problem using robust least squares. Estimate the scale parameters using this initial estimate of Γ (cf. Subsection 3.2).
3. Iteratively solve the version of LP (11) described in Subsection 3.2 until convergence is observed. This gives a vector Γ.
4. For each loop in the graph of triplets:
   a. Compute relative homographies by (5) and (6).
   b. Compute loop-specific homographies solving (14).
5. Determine absolute homographies as described in Subsection 3.3.2.
6. Solve (18) to get the final camera matrices.

4 Second approach: alternating linear minimization

We describe now a second approach that is also tailored to the graphs of triplets constituting a loop, or a chain of a few overlapping loops. It follows the pipeline of the first approach summarized in Subsection 3.4 except Step 3, which is replaced by alternating minimization.

4.1 Background

We start this section by stating some basic results, the proofs of which are deferred to the Appendix. The following proposition states that, with our Trifocal Tensor parameterization, the third fundamental matrix can be computed explicitly and is affine in $\boldsymbol{\gamma}$.

Proposition 3 *Let $G^{j,i} = [\mathbf{e}^{i,j}]_\times F^{j,i}$ and let us consider the parameterization of the Trifocal Tensor $\mathscr{T}^{i,j,k}$ described in Proposition 1. Then, the fundamental matrix $F^{j,k}$ is an affine function of $\boldsymbol{\gamma}$ and can be computed by the formula:*

$$F^{j,k} = [G^{j,i}\mathbf{e}^{k,i}]_\times \left[\left(\mathbf{e}^{i,j}\boldsymbol{\gamma}_{1:3}^\mathsf{T} - \boldsymbol{\gamma}_0 G^{j,i}\right) G^{i,k} - \boldsymbol{\lambda}\,\mathbf{e}^{i,j}(\mathbf{e}^{i,k})^\mathsf{T}\right] + (\boldsymbol{\gamma}_{1:3}\mathbf{e}^{k,i})[\mathbf{e}^{i,j}]_\times G^{j,i} G^{i,k}.$$

Proposition 4 *Assume that we are given a chain of Trifocal Tensors with the associated parameterization as described in Proposition 1:*

$$\left[\mathscr{T}^{1,0,2}, \boldsymbol{\gamma}^1\right] \ldots \left[\mathscr{T}^{i,i-1,i+1}, \boldsymbol{\gamma}^i\right] \ldots \left[\mathscr{T}^{\ell,\ell-1,\ell+1}, \boldsymbol{\gamma}^\ell\right]. \tag{19}$$

The Trifocal Tensor $\mathcal{T}^{i,1,\ell}$ can be computed by formula (2) using the fundamental matrices $F^{i,1}$, $F^{i,\ell}$ and a 4-vector $\gamma^{i,1,\ell}$, satisfying the following properties:

i) The fundamental matrices $F^{i,1}$ and $F^{i,\ell}$ do not depend on $(\gamma^1,\gamma^i,\gamma^\ell)$.

ii) The parameter $\gamma^{i,1,\ell}$ can be computed as an affine function of γ^i (independent of γ^1 and γ^ℓ),

$$\gamma^{i,1,\ell} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ a'_1 & -1 & 0 & 0 \\ a'_2 & 0 & -1 & 0 \\ a'_3 & 0 & 0 & -1 \end{bmatrix} \gamma^i + \begin{bmatrix} 0 \\ a_1 \\ a_2 \\ a_3 \end{bmatrix}. \tag{20}$$

Assume now that the graph of camera triplets is a loop with N nodes. Let us choose k distinct nodes, $v_{i_1},\ldots,v_{i_k}$, in this graph. According to the previous proposition, one can compute the trifocal tensor $\mathcal{T}^{i_s,i_{s-1},i_{s+1}}$ as a function of γ_{i_s}. This tensor will be affine in γ_{i_s} and independent of $\{\gamma_{i_t} : t \neq s\}$. In such a way, one can compute the k tensors

$$\mathcal{T}^{i_2,i_1,i_3},\ldots\mathcal{T}^{i_k,i_{k-1},i_{k+1}},\mathcal{T}^{i_{k+1},i_k,i_{k+2}},$$

hereafter referred to as extended tensors. Considering extended tensors instead of the original ones allows us to reduce the graph to a loop with k nodes. For our purposes, the case $k=4$ is of particular interest.

As already mentioned, each loop gives rise to 15 loop constraints, that have been formulated so far as "the product of transformation matrices is proportional to the identity matrix". Now we will formulate these loop constraints in a different manner. The point is that given a graph of four (extended) trifocal tensors $\left[\mathcal{T}^{i,j,k},\gamma^1\right]$, $\left[\mathcal{T}^{k,i,\ell},\gamma^2\right]$, $\left[\mathcal{T}^{\ell,k,j},\gamma^3\right]$ and $\left[\mathcal{T}^{j,\ell,i},\gamma^4\right]$ forming a loop, the fundamental matrix $\mathrm{F}^{j,k}$ can be computed by the formula presented in Proposition 3 based on either $\mathcal{T}^{i,j,k}$ or $\mathcal{T}^{\ell,k,j}$. This leads to two different expressions of $\mathrm{F}^{j,k}$; the first one depends only on γ^1 while the second one depends only on γ^3. Since these dependencies are affine, the equality of two expressions for $\mathrm{F}^{j,k}$ leads to a set of linear constraints. Using similar kind of arguments one can prove the following result (see Appendix F).

Proposition 5 *Let us consider a loop of 4 parameterized, possibly extended, Trifocal Tensors, denoted by $[\mathcal{T}^{i,j,k},\gamma^1]$, $[\mathcal{T}^{k,i,\ell},\gamma^2]$, $[\mathcal{T}^{\ell,k,j},\gamma^3]$ and $[\mathcal{T}^{j,\ell,i},\gamma^4]$. If we define the vector $\Gamma = \left(\gamma^1,\gamma^2,\gamma^3,\gamma^4\right)^\mathsf{T} \in \mathbb{R}^{16}$, the loop constraints can be written as a linear system $\mathtt{C}\Gamma = \mathbf{d}$, where $\mathtt{C}$ is a 15×16 full rank matrix and $\mathbf{d}\in\mathbb{R}^{15}$.*

If the loop contains more than N triplets and we have an initial estimate for the parameters γ of every node, we can pick (according to some rule) a set of four nodes and look for an update of the parameters of these nodes, leaving the parameters of the other nodes unchanged. This amounts to solving, w.r.t. the vector Γ, the following problem

$$\begin{aligned} &\min \|\mathtt{M}\Gamma - \mathbf{m}\|_2 \\ &\text{subject to } \mathtt{C}\Gamma = \mathbf{d}. \end{aligned} \tag{21}$$

To this end, one can compute a particular solution Γ_{sol} of the system $\mathtt{C}\Gamma=\mathbf{d}$ along with the null vector of $\mathtt{C}$ and find the value of α minimizing the squared error $\|\mathtt{M}\Gamma_{\text{sol}} + \alpha\mathtt{M}\Gamma_{\text{null}} - \mathbf{m}\|_2^2$. This gives $\alpha = \Gamma_{\text{null}}^\mathsf{T}\mathtt{M}^\mathsf{T}(\mathbf{m}-\mathtt{M}\Gamma_{\text{sol}})/\|\mathtt{M}\Gamma_{\text{null}}\|_2^2$. As one can see in the case of 4 trifocal tensors the constrained optimization can be carried out without resorting to any optimization package.

Algorithm 1 2nd approach: alternating linear minimization

Require: 1 loop, N Tensors, $\Gamma_{\text{init}} = \{\gamma^i_{\text{init}}\}$, $\mathtt{M}$, $\mathbf{m}$, and $1 < K < 20$.
1: bool *Ended* =**false**.
2: $\Gamma = \Gamma_{\text{init}}$, $\gamma^i = \gamma^i_{\text{init}}$.
3: **while** (*Ended* ==**false**) **do**
4: Compute the vector of errors $\mathbf{E}$, $\mathbf{E}[i] = \frac{\|\mathtt{M}_i\gamma^i - \mathbf{m}_i\|}{\|\mathtt{M}_i\gamma^i_{\text{init}} - \mathbf{m}_i\|}$.
5: Choose the 4 Trifocal Tensors with the smallest errors.
6: Determine the graph of the 4 extended Trifocal Tensors along with the matrix $\mathtt{C}$ and the vector $\mathbf{d}$ (see Proposition 5 and Appendix F).
7: Compute Γ_{sol} as solution to $\mathtt{C}\Gamma=\mathbf{d}$.
8: Compute Γ_{null}, the null vector of $\mathtt{C}$.
9: Set $\Gamma_s = \Gamma_{\text{sol}} + \alpha\Gamma_{\text{null}}$ with $\alpha = \Gamma_{\text{null}}^\mathsf{T}\mathtt{M}^\mathsf{T}(\mathbf{m}-\mathtt{M}\Gamma_{\text{sol}})/\|\mathtt{M}\Gamma_{\text{null}}\|^2$.
10: Set $\beta = \frac{\|\Gamma\|}{\|\Gamma-\Gamma_s\|KN} \wedge 1$.
11: Update the vector $\Gamma : \Gamma = \Gamma - \beta(\Gamma - \Gamma_s)$.
12: **if** ($\beta == 1$) **then**
13: *Ended* = **true**
14: **end if**
15: **end while**
16: **return** Γ.

4.2 Algorithm

The main idea behind the algorithm—largely inspired by the well-known coordinate descent technique of optimization (see, e.g., [37])—is to iteratively choose 4 Trifocal Tensors among N and to update the parameters of these tensors by minimizing the error $\|\mathtt{M}\Gamma-\mathbf{m}\|_2$ subject to the (linear) loop constraints. This approach guarantees that the error decreases along the iterations. Indeed, at each iteration, the current vector Γ is a feasible solution to the optimization problem at hand, therefore the minimizer will necessarily have a smaller error than the current vector Γ. While this has the advantage of ensuring the convergence of the algorithm and the loop constraints, our empirical results revealed that it gives too much importance to the constraints and does not distribute the error over the set of parameters.

To cope with this, we propose the following "intermediate" solution. Instead of replacing the current Γ by the solution Γ_s of the constrained minimization problem (21),

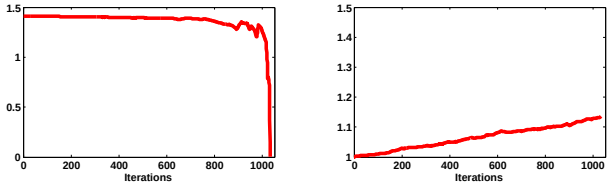


Fig. 6 The left panel shows the evolution of the cyclicity error as a function of the number of iterations. One can observe that it converges to zero after nearly 1000 iterations. In the right panel we plotted the error $\|\mathbb{M}\Gamma - \mathbf{m}\|_2$ against the number of iterations. One can observe that the attainability of the loop constraint does not deteriorate the quality of explanation of point correspondences. In fact, the minimal error is multiplied by a factor less than 1.15. Both left and right panels correspond to the *dinosaur* data set.

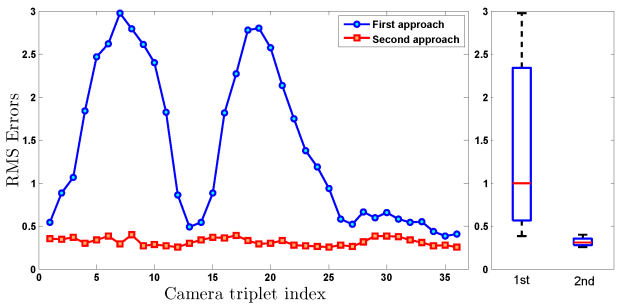


Fig. 7 For the 36 camera triplets of the *dinosaur* data set, we compare the reprojection errors obtained by the first and the second approaches. On the left, one can observe that the second approach leads to a much smaller Root Mean Square Error (RMSE) than the first approach. Moreover, the inter-image variability of the RMSE is drastically smaller for the second approach.

we suggest to do a small displacement in the direction of Γ_s: $\Gamma = \Gamma - \beta(\Gamma - \Gamma_s)$ with $\beta = \frac{\|\Gamma\|}{\|\Gamma - \Gamma_s\| KN} \wedge 1$, where $K > 0$ is a tuning parameter, permitting to achieve a reasonable trade-off between the uniformity of the distribution of error over the parameters γ^i and the attainability of loop constraints.

While the previous system allows to get interesting theoretical guarantees, we propose an algorithm allowing to regularly spread the error among all the Trifocal Tensors. As extreme cases, we retrieve the unconstrained LSE for $K = \infty$ and a vector satisfying the constraints for K very close to zero. To summarize, the larger the parameter K is, the slower the algorithm converges but the better the error distribution between the tensors is. Our numerical experiments showed that nice trade-off is obtained for $K \in (1, 20)$.

At this stage, it is important to present an automatic rule for choosing the 4 tensors for which the parameters γ^i are updated. We propose a purely heuristical rule, consisting in choosing the tensors minimizing the error

$$\mathbf{E}_\gamma = \frac{\|\mathbb{M}\gamma - \mathbf{m}\|}{\|\mathbb{M}\gamma_{\text{init}} - \mathbf{m}\|}. \tag{22}$$

This choice allows us to spread the errors over the entire set of cameras.

4.3 Numerical results and comparison to the first approach

Since many steps of the second approach are largely heuristical and are based on empirical arguments, we present here some numerical results that aim to demonstrate the convergence of the proposed procedure as well as the fact that the goal of uniform distribution of the error is achieved with high accuracy.

For the well-known *dinosaur* sequence containing 36 images, the second approach described above allows to converge—after nearly 1000 iterations—to a solution satisfying the loop constraints (see Fig. 6, left panel). In this and subsequent experiments, we have chosen $K = 15$. The measure of attainability of the loop constraint, represented in Fig. 6 by the y-axis, is the Frobenius norm of the difference between the identity matrix and the normalized product of all homographies. On the right panel of Fig. 6, we plotted the error $\|\mathbb{M}\Gamma - \mathbf{m}\|_2$ against the number of iterations. As expected, this function has an increasing trend. It is however noteworthy that this trend has a very moderate slope. Indeed, the final error is within 15% of the initial one.

It is also interesting to have a look at the quality of the final reconstruction obtained by 2 versions of our procedure. While the accuracy of the estimated cameras is of the same order, there is a clear difference between the quality in terms of the reprojection error. This can be seen in Fig. 7, where the x-axis corresponds to the index of the camera triplet (the triplet i being made of cameras i, $i+1$ and $i+2$) and the y-axis represents the root mean squared error (in pxl) for the 3-view correspondences belonging to the relevant triplet.

We see that the Root Mean Square Error (RMSE) of the second approach is uniformly smaller than that of the first approach and, more importantly, the RMSE for the second approach is uniformly distributed over the camera triplets. A similar experiment has been conducted for 3 other benchmark data sets and the results are summarized in Table 1. One can observe that in all experiments the second approach succeeds in finding a feasible solution (*i.e.*, a vector satisfying the constraints). In terms of the RMSE, the second approach outperforms significantly the first one in the case of the Dinosaur and the Detenice data sets. The results of both methods for the Temple are quite similar, whereas on the Calvary data set the first approach is significantly better.

Fig. 8 Three frames of each data set used to test our methodology. From left to right: dinosaur, temple, fountain P11, Herz-Jesu R23 [36], Calvary, Detenice fountain.

Dataset	**Method**	**# Iterations** Until CV	**RMSE**	**St. Dev.**
Dinosaur	1st	-	1.01	1.31
	2nd	1027	0.31	0.71
Temple	1st	-	0.10	1.21
	2nd	340	0.09	1.92
Detenice	1st	-	2.28	1.49
	2nd	1393	1.27	0.41
Calvary	1st	-	1.02	1.22
	2nd	1286	1.76	1.11

Table 1 For classical data sets we compare the Root Mean Square Error (RMSE), measured in pixels, of the first with that of the second approach. We observe that the RMSE is often smaller for the second approach. The standard deviations of squared errors are presented in the last column.

5 Experiments

5.1 Implementation

In order to apply the methodology we have just described, we extract and match SIFT [33] descriptors from all the images. Then, epipolar geometries are estimated by DEGENSAC [34]. Note that some speed-up in this step can be achieved by using one of the recent versions of RANSAC [27,35]. Estimated EGs allow us to identify and remove wrong correspondences as well as to create feature tracks. Using these tracks and EGs as input for our algorithm, we compute as output the projection matrices of all the cameras. In order to be able to visually assess the reconstruction quality, all cameras and the 3D structure are upgraded to Euclidean [28].

5.2 Datasets

We tested our methodology on six data sets: the *dinosaur* (36 frames), the *temple* (45 frames), the *fountain P11* (11 frames), the *Herz-Jesu R23* (23 frames), the *Detenice fountain* (34 frames) and the *calvary* (52 frames) sequences. For the first three data sets, the ground truth of camera matrices is available on the Internet, see respectively
`http://www.robots.ox.ac.uk/~vgg/data1.html`,
`http://vision.middlebury.edu/mview/data/`,
`http://cvlab.epfl.ch/~strecha/multiview/`.

5.3 Quality measures

Since the main contribution of the present paper concerns the projective reconstruction, it is natural to assess the quality of the proposed approach using the distance:

$$d_{proj}(\{\mathtt{P}^j\},\{\mathtt{P}^{j,*}\}) = \inf_{\alpha,\mathtt{H}} \sum_{j=1}^{n} \|\alpha_j \mathtt{P}^j \mathtt{H} - \mathtt{P}^{j,*}\|_2^2, \tag{23}$$

where $\mathtt{P}^j$ and $\mathtt{P}^{j,*}$ are respectively the reconstructed and the true camera projection matrices, $\alpha = (\alpha_1,\ldots,\alpha_n)$ is a vector of real numbers and $\mathtt{H}$ is a 3D-homography. Naturally, this measure can be used only on sequences for which the ground truth is available. Note also that the computation

Dataset	#frames	resolution	# image points		RMSE (pxl)	
			Our	Bundler	Our	Bundler
Dinosaur	36	576 × 720	45,250	37,860	0.27	0.25
Temple	45	640 × 480	26,535	23,761	0.08	0.11
Fountain P11	11	2048 × 3072	57,547	23,648	0.16	0.13
Herz-Jesu R23	23	2048 × 3072	129,803	–	0.41	–
Detenice	34	1536 × 2048	30,200	–	0.15	–
Calvary	52	2624 × 3972	54,798	–	0.51	–

Table 2 Characteristics of the data sets used for the experimental validation. From left to right: number of frames in each sequence, the resolution of each image (all images of a data set have the same resolution), the number of 2D image points used for the final BA for our method and for bundler [5], the RMSE for both methods. The numbers in the four last columns are reported to show that both methods work well and *not* to compare them.

of the infimum in (23) is a non-convex optimization problem. We solve it by first computing the unit-norm solution to the least squares problem $\min_{\alpha,\mathtt{H}} \sum_{j=1}^{n} \|\mathrm{P}^j\mathtt{H} - \alpha_j^{-1}\mathrm{P}^{j,*}\|_2^2$, and then use this solution as a starting point for an alternating minimization. For the examples considered here, this converges very rapidly and, since the results are good, we believe that the local minimum we find is in fact a global minimum, or at least not too far from it.

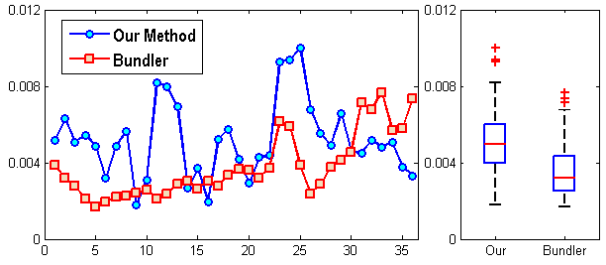
5.4 Results

For the dinosaur, temple and fountain P11 sequences, since ground truth exists, we compared our results with those of bundler [5], which is a state-of-the-art calibration software. The ground truth was normalized so that the Frobenius norm of all the cameras is one. For both reconstructions (ours and bundler), we computed numbers α_j and a homography $\mathtt{H}$ by minimizing (23). This allows us to define the per-camera error as $\|\alpha_j\mathrm{P}^j\mathtt{H} - \mathrm{P}^{j,*}\|_2^2$ for the jth camera. For our method, the results of the first approach are shown, those of the second approach being of the same order.

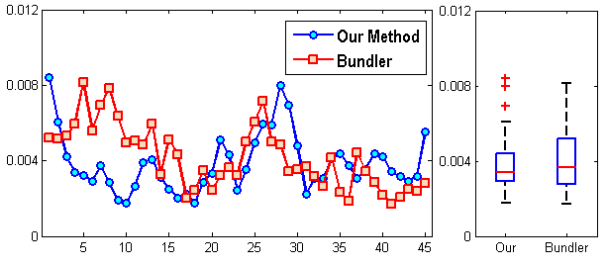
As shown in Fig. 9, not only the errors are small, but also our results are quite comparable to those of bundler despite the fact that our method does not perform intermediate BAs and does not assume that the principal point is in the center and the skew is zero. One can also note that the error is well distributed over the whole sequence of cameras due to the fact that both methods operate on the closed sequence. Furthermore, the results reported for fountain P11 are achieved without final BA, proving that the method we proposed furnishes a good starting point for the non-linear optimization.

As for the data sets where no ground truth is known, we have chosen to use as measure of evaluation the multiview stereo reconstruction of the scene based on the method of [26][2]. The results are shown in Fig. 1 (right) for the cal-

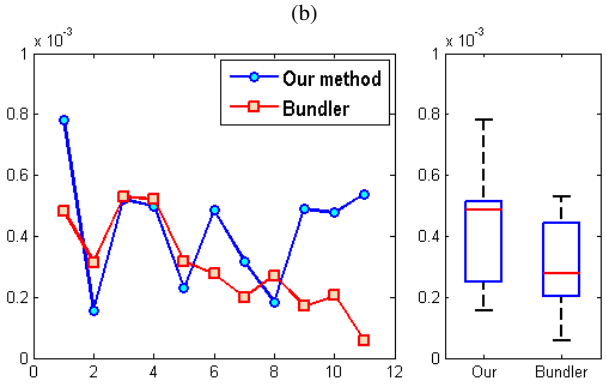
[2] Since multiview stereo reconstruction is not the purpose of the paper and is only used for illustration, the results shown in Fig. 1 and 10 are obtained without the final mesh refinement.



(a)



(b)



(c)

Fig. 9 This figure shows the errors in estimated camera matrices for our method and for bundler [5]. The per-camera errors and their boxplots for the dinosaur sequence (a), the temple sequence (b) and for the fountain P11 sequence (c). One can remark that our method achieves the same level of accuracy as that of bundler, despite the fact that we do not use any information on the internal parameters, while bundler assumes that the skew is zero and the principal point is in the center, hypotheses that are close to reality in these data sets.

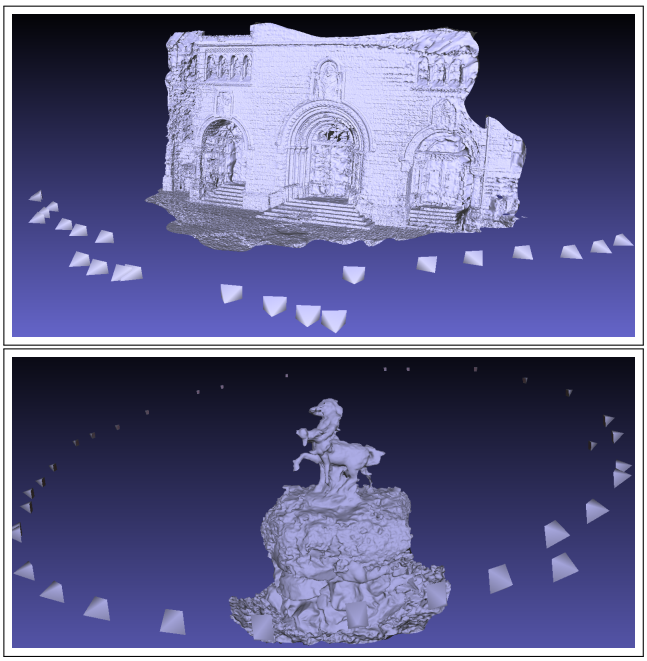


Fig. 10 Multi-view stereo reconstruction using the camera matrices estimated by our method for the Herz-Jesu R23 and Detenice fountain data sets. For these data, the ground truth is unavailable but the quality of the scene reconstruction demonstrates the accuracy of estimated cameras.

vary sequence and in Fig. 10 for the Herz-Jesu R23 and the Detenice fountain sequences. In the aim of comparing our results with other approaches, let us recall that (as reported in [36]) on the Herz-Jesu R23 data the ARC3D software succeeded to calibrate four of the 23 cameras, while the method proposed in [4] calibrated all the cameras with a relatively large error for cameras 6-11. Although we are unable to quantitatively compare our reconstruction to that of [4], the accuracy of the 3D scene reconstruction makes us believe that the estimated cameras are very close to the true ones.

5.5 Run times

To give a rough idea of running times of the pipeline we proposed, note that on the calvary dataset consisting of one loop of 52 images of size $2624x3972$ and on an Intel Xeon 2,33 GHz CPU, it took

- 231 seconds to load images and to compute interest points,
- 1791 seconds to compute the fundamental matrices with classical RANSAC and to look for additional matches,
- 1308 seconds to determine a first estimator of Γ with 4 points RANSAC applied to each camera triplet,
- 9 seconds to perform one iteration of the constrained optimization with linearized loop constraints (the LP is solved using SeDuMi). In this example, we used 15 iterations.

For a Matlab implementation of the second approach, each iteration took 4 seconds and 1286 iterations were needed for the convergence. Thus, the first approach has very attractive execution times while the second approach is significantly more time consuming. It is, however, very likely that this drawback can be partly overcome by using C implementation instead of Matlab.

6 Summary and outlook

In this work, we have presented a new pipeline for solving the uncalibrated structure from motion problem of multi-view geometry. To this end, we have introduced a new parameterization of the trifocal tensor which has the advantage of being linear with respect to the unknown parameters, provided that two fundamental matrices out of three are known. We proposed to use this parameterization in order to estimate the projection matrices of cameras from the graph of trifocal tensors, a set of estimated fundamental matrices and a collection of three-view correspondences. The main emphasis is put on the ability of enforcing loop constraints.

Two variants of the pipeline are considered, differing by the way of handling loop constraints. Experiments carried out on standard data sets and reported in this work have shown that the second variant outperforms the first one in terms of distributing estimation error uniformly across all the cameras. The downside of the second variant is the higher run-time since, typically, it requires many more iterations to converge than the first variant.

Although the pipeline we have presented can deal with a network of cameras containing several loops, the current implementation of the pipeline is limited to at most one loop in the graph of trifocal tensors, or to several loops predefined by the user. Next in our agenda is to combine this pipeline with an algorithm performing automatic loop detection in a graph. This is a delicate issue since the loops need to satisfy several properties (covering nearly all the nodes of the graph, avoiding edges corresponding to poorly estimated fundamental matrices, etc.) which may often be mutually contradictory.

It should also be mentioned that, in our pipeline, the computations for different loops are completely independent and are parallelizable. Therefore, we envisage in the future to implement the pipeline or a part of it on GPU in order to decrease the running times. On a related note, making the pipeline scaleable to data sets containing thousands of images is another important avenue for future research.

Appendix

A Proof of Proposition 1

We begin by considering the case where all the 3 rows of the fundamental matrices $\mathtt{F}^{i,j}$ and $\mathtt{F}^{i,k}$ are different from the zero vector of $\mathbb{R}^3$. This implies that the columns of the matrices $\mathtt{A}_t^{i,j}$ and $\mathtt{A}_t^{i,j}$ form two orthogonal bases of $\mathbb{R}^3$. (Indeed, it is well-known that the epipole $\mathbf{e}^{i,j}$ is orthogonal to the rows of $\mathtt{F}^{i,j}$, while $(\mathtt{F}_{t,1:3}^{i,j})^\mathsf{T} \times \mathbf{e}^{i,j}$ is orthogonal to $\mathtt{F}_{t,1:3}^{i,j}$ and to $\mathbf{e}^{i,j}$ by virtue of the definition of the vector product.) Therefore, $\mathtt{A}_t^{i,k}$ and $\mathtt{A}_t^{i,j}$ are invertible. Let us define

$$\begin{bmatrix} a_t & b_t & c_t \\ d_t & e_t & f_t \\ g_t & h_t & i_t \end{bmatrix} = (\mathtt{A}_t^{i,j})^{-1} \mathtt{T}_t^{i,j,k} (\mathtt{A}_t^{i,k})^{-\mathsf{T}}. \tag{24}$$

Let us show that $a_t = b_t = c_t = 0$. Recall that the matrix $\mathtt{T}_t^{i,j,k}$ relates a point $\mathbf{p} = (p_1, p_2, p_3) \in \mathbb{P}^2$ to its epipolar line $\mathbf{l} = \mathtt{F}^{i,j\mathsf{T}}\mathbf{p}$ through the equation $\mathbf{l}^\mathsf{T} \sum_{s=1}^3 p_s \mathtt{T}_s^{i,j,k} = \mathbf{0}^\mathsf{T}$ [1, p. 373]. Choosing as $\mathbf{p}$ the vector $(\delta_{t1}, \delta_{t2}, \delta_{t3})^\mathsf{T}$, where $\delta_{t\ell}$ stands for the Kronecker symbol that equals one if $t = \ell$ and zero otherwise, we get $\mathtt{F}_{t,1:3}^{i,j} \mathtt{T}_t^{i,j,k} = \mathbf{0}^\mathsf{T}$. This equation, in conjunction with (24), the definition of $\mathtt{A}_t^{i,j}$ and the invertibility of $\mathtt{A}_t^{i,k}$ entails that

$$\begin{bmatrix} a_t & d_t & g_t \\ b_t & e_t & h_t \\ c_t & f_t & i_t \end{bmatrix} \begin{bmatrix} \|(\mathtt{F}_{t,1:3}^{i,j})^\mathsf{T}\|^2 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

This yields $a_t = b_t = c_t = 0$. By a symmetric argument, we also check that $d_t = g_t = 0$. Thus, the Trifocal Tensor necessarily reduces to the form:

$$\mathtt{T}_t^{i,j,k} = \mathtt{A}_t^{i,j} \begin{bmatrix} 0 & 0 & 0 \\ 0 & e_t & f_t \\ 0 & h_t & i_t \end{bmatrix} (\mathtt{A}_t^{i,k})^\mathsf{T} \tag{25}$$

Since the rank of any fundamental matrix is equal to two, there exists an index $t' \in \{1,2,3\}$ such that $\mathtt{F}_{t',1:3}^{i,j}$ and $\mathtt{F}_{t,1:3}^{i,j}$ are not collinear. We have already checked that $\mathtt{F}_{t,1:3}^{i,j} \mathtt{T}_t^{i,j,k} = \mathbf{0}^\mathsf{T}$ and, similarly, $\mathtt{F}_{t',1:3}^{i,j} \mathtt{T}_{t'}^{i,j,k} = \mathbf{0}^\mathsf{T}$. Therefore, substituting $\mathbf{p} = (\delta_{t1}, \delta_{t2}, \delta_{t3})^\mathsf{T} + (\delta_{t'1}, \delta_{t'2}, \delta_{t'3})^\mathsf{T}$ in the equation $\mathbf{p}^\mathsf{T} \mathtt{F}^{i,j} \sum_{s=1}^3 p_s \mathtt{T}_s^{i,j,k} = 0$, we get

$$\mathtt{F}_{t',1:3}^{i,j} \mathtt{T}_t^{i,j,k} + \mathtt{F}_{t,1:3}^{i,j} \mathtt{T}_{t'}^{i,j,k} = \mathbf{0}^\mathsf{T}. \tag{26}$$

Now, let us observe that

$$(\mathtt{F}_{t',1:3}^{i,j})^\mathsf{T} ((\mathtt{F}_{t,1:3}^{i,j})^\mathsf{T} \times \mathbf{e}^{i,j}) = -(\mathtt{F}_{t,1:3}^{i,j})^\mathsf{T} ((\mathtt{F}_{t',1:3}^{i,j})^\mathsf{T} \times \mathbf{e}^{i,j}) = \beta_{t,t'}. \tag{27}$$

Moreover, $\beta_{t,t'} \neq \mathbf{0}$ since the vectors $\mathtt{F}_{t,1:3}^{i,j}$ and $\mathtt{F}_{t',1:3}^{i,j}$ are linearly independent and orthogonal to $\mathbf{e}^{i,j}$. This observation together with (25) and (26) leads to

$$\mathtt{A}_t^{i,k} \begin{bmatrix} 0 & 0 & 0 \\ 0 & e_t & h_t \\ 0 & f_t & i_t \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \\ 0 \end{bmatrix} + \mathtt{A}_{t'}^{i,k} \begin{bmatrix} 0 & 0 & 0 \\ 0 & e_{t'} & h_{t'} \\ 0 & f_{t'} & i_{t'} \end{bmatrix} \begin{bmatrix} \alpha \\ -\beta \\ 0 \end{bmatrix} = \mathbf{0},$$

where we have used the shorthands $\alpha = \alpha_{t,t'}$ and $\beta = \beta_{t,t'}$. Matrix multiplication yields

$$\beta_{t,t'} \big(\begin{bmatrix} 0, & e_t, & f_t \end{bmatrix} (\mathtt{A}_t^{i,k})^\mathsf{T} - \begin{bmatrix} 0, & e_{t'}, & f_{t'} \end{bmatrix} (\mathtt{A}_{t'}^{i,k})^\mathsf{T} \big) = \mathbf{0}^\mathsf{T}. \tag{28}$$

The last equality is equivalent to

$$(e_t \mathtt{F}_{t,1:3}^{i,k} - e_{t'} \mathtt{F}_{t',1:3}^{i,k})^\mathsf{T} \times \mathbf{e}^{i,k} + (f_t - f_{t'}) \mathbf{e}^{i,k} = \mathbf{0}, \tag{29}$$

which is possible if and only if $f_t = f_{t'}$ and $e_t \mathtt{F}_{t,1:3}^{i,k} - e_{t'} \mathtt{F}_{t',1:3}^{i,k} = \mathbf{0}$. Since $\mathtt{F}_{t,1:3}^{i,k}$ and $\mathtt{F}_{t',1:3}^{i,k}$ are linearly independent, we conclude that $e_t = e_{t'} = 0$. In addition, using a symmetric argument, we get $h_t = h_{t'}$ and thus

$$\mathtt{T}_t^{i,j,k} = \mathtt{A}_t^{i,j} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & f \\ 0 & h & i_t \end{bmatrix} (\mathtt{A}_t^{i,k})^\mathsf{T}, \; \mathtt{T}_{t'}^{i,j,k} = \mathtt{A}_{t'}^{i,j} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & f \\ 0 & h & i_{t'} \end{bmatrix} (\mathtt{A}_{t'}^{i,k})^\mathsf{T}. \tag{30}$$

Let now t'' be the element of the index set $\{1,2,3\}$ that is different from t and t'. Repeating the same arguments with t replaced by t'' leads to a formula similar to (30) in which t' is replaced by t'' everywhere. The first claim of the proposition follows by dividing all the entries of $[\mathtt{T}^{i,j,k}]$ by h, which is $\neq 0$ since otherwise the fundamental matrix computed from this trifocal tensor is of rank < 2, which is impossible. So in the inhomogeneous formulation there are only 4 unknowns corresponding to the remaining degrees of freedom, and in the homogeneous formulation there are 5 unknowns.

In the case where the matrix $\mathtt{F}^{i,j}$ or $\mathtt{F}^{i,k}$ contains zero rows, one can merely use the fact that Eq. (3) (see page 4 of the paper) characterize all the triplets of camera matrices (up to a projective homography) which are compatible with the fundamental matrices $\mathtt{F}^{i,j}$ and $\mathtt{F}^{i,k}$. In view of [1, Eq. 15.1], the trifocal tensor corresponding to these camera matrices coincides with the one defined in the statement of the proposition. This completes the proof.

B Proofs of Equations (5) and (6)

Let us start with the proof of Equation (5). Let $\mathtt{P}^i$ and $\mathtt{P}^k$ be the camera matrices computed from the triplet (i,j,k) with the parameter γ and let $\underline{\mathtt{P}}^i$ and $\underline{\mathtt{P}}^k$ be those computed from the triplet (k,i,ℓ) with the parameter $\underline{\gamma}$. To simplify the exposition, we denote $\mathtt{G}^{k,i} = [\mathbf{e}^{i,k}]_\times \mathtt{F}^{k,i}$ and $\mathtt{G}^{i,k} = [\mathbf{e}^{k,i}]_\times \mathtt{F}^{i,k}$. We are looking for a homography $\underline{\mathtt{H}}$ such that $\mathtt{P}\underline{\mathtt{H}} \cong \underline{\mathtt{P}}$. This amounts to solving:

$$[\mathtt{I}_{3\times3}|\mathbf{0}]\underline{\mathtt{H}} \cong \mathrm{kron}([\underline{\gamma}_{1:3}, 1], \mathbf{e}^{k,i}) - \underline{\gamma}_0 [\mathtt{G}^{i,k}|\mathbf{0}], \quad [\mathtt{G}^{k,i}|\mathbf{e}^{i,k}]\underline{\mathtt{H}} \cong [\mathtt{I}_{3\times3}|\mathbf{0}]. \tag{31}$$

The first equation in (31) yields

$$\underline{\mathtt{H}}_{1:3,1:4} \cong \mathrm{kron}([\underline{\gamma}_{1:3}, 1], \mathbf{e}^{k,i}) - \underline{\gamma}_0 [\mathtt{G}^{i,k}|\mathbf{0}]. \tag{32}$$

Inserting this in (31) and using $\mathtt{G}^{k,i}\mathbf{e}^{k,i} = \mathbf{0}$, we get

$$-\underline{\gamma}_0 \mathtt{G}^{k,i}\mathtt{G}^{i,k} + \mathbf{e}^{i,k}\underline{\mathtt{H}}_{4,1:3} = \alpha \mathtt{I}_{3\times3}, \tag{33}$$

and $\mathbf{e}^{i,k}\underline{\mathtt{H}}_{4,4} = 0$. This implies that $\underline{\mathtt{H}}_{4,4} = 0$. Furthermore, multiplying both sides of (33) by $(\mathbf{e}^{i,k})^\mathsf{T}$, we get $\underline{\mathtt{H}}_{4,1:3} = \alpha(\mathbf{e}^{i,k})^\mathsf{T}$. To complete the proof, it remains to determine the value of α. This is done by computing the trace of both sides in (33).

Now we prove the formula for $\mathtt{H}_{\gamma,\underline{\gamma}} := \mathtt{H}$ given in Eq. (6). The relation $\mathtt{P} \cong \underline{\mathtt{P}}\mathtt{H}$ implies that:

$$[\mathtt{I}_{3\times3}|\mathbf{0}] \cong \Big(\mathrm{kron}([\underline{\gamma}_{1:3}, 1], \mathbf{e}^{k,i}) - \underline{\gamma}_0 [\mathtt{G}^{i,k}|\mathbf{0}]\Big)\mathtt{H}, \; [\mathtt{G}^{k,i}|\mathbf{e}^{i,k}] \cong [\mathtt{I}_{3\times3}|\mathbf{0}]\mathtt{H}, \tag{34}$$

The second equation yields $\mathtt{H}_{1:3,1:4} \cong [\mathtt{G}^{k,i}|\mathbf{e}^{i,k}]$. Inserting this in the first equation of (34), we get:

$$\underline{\gamma}_{1:3}\mathbf{e}^{i,k}\mathbf{e}^{k,i} + \mathtt{H}_{4,4}\mathbf{e}^{k,i} = 0 \iff \mathtt{H}_{4,4} = -\underline{\gamma}_{1:3} \cdot \mathbf{e}^{i,k}.$$
$$-\underline{\gamma}_0 \mathtt{G}^{i,k}\mathtt{G}^{k,i} + \mathbf{e}^{k,i}\underline{\gamma}_{1:3}^\mathsf{T}\mathtt{G}^{k,i} + \mathbf{e}^{k,i}\mathtt{H}_{4,1:3} = \lambda \underline{\gamma}_0 \mathtt{I}_{3\times3}. \tag{35}$$

Multiplying both sides of the last equation by $(\mathbf{e}^{k,i})^\mathsf{T}$ we get $\mathtt{H}_{4,1:3} = \lambda\underline{\gamma}_0(\mathbf{e}^{k,i})^\mathsf{T} - (\underline{\gamma}_{1:3})^\mathsf{T}\mathtt{G}^{k,i}$. After inserting this value of $\mathtt{H}_{4,1:3}$ into Eq. (35) and computing the trace of both sides, we get the equality $2\lambda = -\mathrm{tr}(\mathtt{G}^{i,k}\mathtt{G}^{k,i})$ and the desired result follows.

C Proof of Proposition 2

In view of Eq. (12) of the paper, it holds

$$\mathtt{P}^{2,\gamma_i}\mathtt{H}^{v_i} \cong \mathtt{P}^{i+1,*}, \qquad \mathtt{P}^{3,\gamma_i}\mathtt{H}^{v_i} \cong \mathtt{P}^{i+2,*} \tag{36}$$

$$\mathtt{P}^{1,\gamma i+1}\mathtt{H}^{v_{i+1}} \cong \mathtt{P}^{i+1,*}, \qquad \mathtt{P}^{2,\gamma i+1}\mathtt{H}^{v_{i+1}} \cong \mathtt{P}^{i+2,*}. \tag{37}$$

Therefore,

$$\mathtt{P}^{2,\gamma_i}\mathtt{H}^{v_i} \cong \mathtt{P}^{1,\gamma i+1}\mathtt{H}^{v_{i+1}}, \qquad \mathtt{P}^{3,\gamma_i}\mathtt{H}^{v_i} \cong \mathtt{P}^{2,\gamma i+1}\mathtt{H}^{v_{i+1}}. \tag{38}$$

Furthermore, by virtue of Eq. (13) of the paper,

$$\mathtt{P}^{1,\gamma_{i+1}} \cong \mathtt{P}^{2,\gamma_i}\mathtt{H}^{i,i+1}, \quad \mathtt{P}^{2,\gamma_{i+1}} \cong \mathtt{P}^{3,\gamma_i}\mathtt{H}^{i,i+1}. \tag{39}$$

Substituting (39) in (38), we get

$$\mathtt{P}^{2,\gamma_i}\mathtt{H}^{v_i} \cong \mathtt{P}^{2,\gamma_i}\mathtt{H}^{i,i+1}\mathtt{H}^{v_{i+1}}, \mathtt{P}^{3,\gamma_i}\mathtt{H}^{v_i} \cong \mathtt{P}^{3,\gamma_i}\mathtt{H}^{i,i+1}\mathtt{H}^{v_{i+1}}. \tag{40}$$

If the centers of $\mathtt{P}^{2,\gamma_i}$ and $\mathtt{P}^{3,\gamma_i}$ differ, which is equivalent to $\mathtt{P}^{i+1,*}, \mathtt{P}^{i+2,*}$ having different centers, then Eq. (40) can be satisfied if and only if $\mathtt{H}^{v_i} \cong \mathtt{H}^{i,i+1}\mathtt{H}^{v_{i+1}}$.

To prove i) and ii), we need to transform the relations of proportionality into the relations of equality. Since $\mathtt{H}^{v_i} = \mathtt{H}^{i,i+1}\mathtt{H}^{v_{i+1}}$, there exists a real number $\alpha_i \neq 0$ such that $\mathtt{H}^{v_i} = \alpha_i \mathtt{H}^{i,i+1}\mathtt{H}^{v_{i+1}}$. Thus, we have

$$\begin{aligned}
&\mathtt{H}^{v_1} = \alpha_1 \mathtt{H}^{1,2}\mathtt{H}^{v_2},\\
&\mathtt{H}^{v_2} = \alpha_2 \mathtt{H}^{2,3}\mathtt{H}^{v_3},\\
&\quad\vdots\\
&\mathtt{H}^{v_n} = \alpha_n \mathtt{H}^{n,1}\mathtt{H}^{v_1}.
\end{aligned}$$

Let us denote $\tilde{\mathtt{H}}^{v_1} = \mathtt{H}^{v_1}$, $\tilde{\mathtt{H}}^{v_2} = \alpha_1\mathtt{H}^{v_2}$, $\tilde{\mathtt{H}}^{v_3} = \alpha_1\alpha_2\mathtt{H}^{v_3}$, …, $\tilde{\mathtt{H}}^{v_n} = \alpha_1 \times \ldots \times \alpha_{n-1}\mathtt{H}^{v_n}$. Each $\tilde{\mathtt{H}}^{v_i}$ being proportional to $\mathtt{H}^{v_i}$, satisfies Eq. (12) of the paper. This leads to the assertion i) of Proposition 2, since $\tilde{\mathtt{H}}^{v_i} = \mathtt{H}^{i,i+1}\tilde{\mathtt{H}}^{v_{i+1}}$ for $i = 1,\ldots,n-1$. Moreover, we have

$$\tilde{\mathtt{H}}^{v_n} = \Big(\prod_{i=1}^{n}\alpha_i\Big)\mathtt{H}^{n,1}\tilde{\mathtt{H}}^{v_1} \overset{\text{notation}}{=} \alpha^{-1}\mathtt{H}^{n,1}\tilde{\mathtt{H}}^{v_1}. \tag{41}$$

This implies that

$$\prod_{i=1}^{n}\tilde{\mathtt{H}}^{v_i} = \Big(\prod_{i=1}^{n-1}\mathtt{H}^{i,i+1}\tilde{\mathtt{H}}^{v_{i+1}}\Big)\Big(\prod_{i=1}^{n}\alpha_i\Big)\mathtt{H}^{n,1}\tilde{\mathtt{H}}^{v_1} = \alpha^{-1}\Big(\prod_{i=1}^{n}\mathtt{H}^{i,i+1}\Big)\prod_{i=1}^{n}\tilde{\mathtt{H}}^{v_i},$$

which is equivalent to $\alpha\mathtt{I}_{4\times4} = \prod_{i=1}^{n}\mathtt{H}^{i,i+1}$. Taking the trace of both sides, we get the desired expression for α, completing thus the proof of ii).

To prove iii), we simply remark that all the homographies $\mathtt{H}^{v_i}$ are defined up to an overall homography ambiguity. In other terms, we can replace all $\mathtt{H}^{v_i}$ by $\mathtt{H}^{v_i}\mathtt{Q}$, where $\mathtt{Q}$ is an invertible 4×4 matrix. Let $\tilde{\mathtt{H}}$ be the $4n\times4$ matrix resulting from the vertical concatenation of matrices $\mathtt{H}^{v_i}$ (satisfying conditions i) and ii) of the proposition). Let $\tilde{\mathtt{H}}^\top\tilde{\mathtt{H}} = \mathtt{U}\mathtt{L}\mathtt{U}^\top$ be the SVD of $\tilde{\mathtt{H}}$. Thus, $\mathtt{L}$ is a 4×4 diagonal matrix with strictly positive diagonal entries and $\mathtt{U}$ is a 4×4 orthogonal matrix. Therefore, $\mathtt{U}$ is a homography. Setting $\mathtt{Q} = \mathtt{U}$, we get a new version of homographies $\mathtt{H}^{v_i}$ that satisfy all the conditions of Proposition 2.

D Proof of Proposition 3

The camera triplet corresponding to $\mathscr{T}^{i,j,k}$ is:

$$\begin{aligned}
&\mathtt{P}^i = [\mathtt{I}_{3\times3}\,|\,\mathbf{0}_{3\times1}], \qquad \mathtt{P}^k = [[\mathbf{e}^{i,k}]_\times\mathtt{F}^{k,i}\,|\,\mathbf{e}^{i,k}],\\
&\mathtt{P}^j = \mathrm{kron}\big([\gamma_{1:3},1];\mathbf{e}^{i,j}\big) - \gamma_0[[\mathbf{e}^{i,j}]_\times\mathtt{F}^{j,i}\,|\,\mathbf{0}_{3\times1}].
\end{aligned} \tag{42}$$

There exists a projective transform such that $\mathtt{P}'^i = \mathtt{P}^i\mathtt{H}$, $\mathtt{P}'^k = \mathtt{P}^k\mathtt{H}$ and:

$$\begin{aligned}
&\mathtt{P}'^i = [[\mathbf{e}^{k,i}]_\times\mathtt{F}^{i,k}\,|\,\mathbf{e}^{k,i}], \qquad \mathtt{P}'^k = [\mathtt{I}_{3\times3}\,|\,\mathbf{0}_{3\times1}],\\
&\mathtt{P}'^j = \big(\mathrm{kron}\big([\gamma_{1:3},1];\mathbf{e}^{i,j}\big) - \gamma_0[[\mathbf{e}^{i,j}]_\times\mathtt{F}^{j,i}\,|\,\mathbf{0}_{3\times1}]\big)\,\mathtt{H}.
\end{aligned}$$

The equation $[[\mathbf{e}^{k,i}]_\times\mathtt{F}^{i,k}\,|\,\mathbf{e}^{k,i}] = [\mathtt{I}_{3\times3}\,|\,\mathbf{0}_{3\times1}]\mathtt{H}$ can be rewritten in the form $\mathtt{H}_{1:3,1:4} = \big[[\mathbf{e}^{k,i}]_\times\mathtt{F}^{i,k}|\mathbf{e}^{k,i}\big]$. The other equation is:

$$\beta[\mathtt{I}_{3\times3}\,|\,\mathbf{0}_{3\times1}] = [[\mathbf{e}^{i,k}]_\times\mathtt{F}^{k,i}\,|\,\mathbf{e}^{i,k}]\begin{bmatrix}[\mathbf{e}^{k,i}]_\times\mathtt{F}^{i,k} & \mid & \mathbf{e}^{k,i}\\ \mathbf{v}_{1\times3} & \mid & v_4\end{bmatrix}. \tag{43}$$

By denoting $\mathtt{G}^{k,i} = [\mathbf{e}^{i,k}]_\times\mathtt{F}^{k,i}$ and $\mathtt{G}^{i,k} = [\mathbf{e}^{k,i}]_\times\mathtt{F}^{i,k}$, the equality of the last columns leads to $v_4 = 0$, while that of the first 3 columns is equivalent to

$$\beta\,\mathtt{I}_{3\times3} = \mathtt{G}^{k,i}\mathtt{G}^{i,k} + \mathbf{e}^{i,k}\mathbf{v}_{1\times3}. \tag{44}$$

Multiplying from left by $(\mathbf{e}^{i,k})^\mathsf{T}$ one gets $\mathbf{v}_{1\times3} = \beta(\mathbf{e}^{i,k})^\mathsf{T}$. This expression of $\mathbf{v}_{1\times3}$ in conjunction with Eq. (44) yields $\beta\,\mathtt{I}_{3\times3} = \mathtt{G}^{k,i}\mathtt{G}^{i,k} + \beta\mathbf{e}^{i,k}(\mathbf{e}^{i,k})^\mathsf{T}$. Computing the trace of both sides of this equation, we get $\beta = \frac{1}{2}\mathrm{trace}\big(\mathtt{G}^{k,i}\mathtt{G}^{i,k}\big) = -\lambda$. Therefore, the camera triplet can be written as: $\mathtt{P}'^i = [\mathtt{G}^{i,k}\,|\,\mathbf{e}^{k,i}]$, $\mathtt{P}'^k = [\mathtt{I}_{3\times3}\,|\,\mathbf{0}_{3\times1}]$ and

$$\begin{aligned}
\mathtt{P}'^j &= \big(\mathbf{e}^{i,j}[\gamma_{1:3};1]^\mathsf{T} - \gamma_0[\mathtt{G}^{j,i}\,|\,\mathbf{0}_{3\times1}]\big)\begin{bmatrix}\mathtt{G}^{i,k} & \mid & \mathbf{e}^{k,i}\\ -\lambda(\mathbf{e}^{i,k})^\mathsf{T} & \mid & 0\end{bmatrix}\\
&= \big[\mathbf{e}^{i,j}(\gamma_{1:3}\mathtt{G}^{i,k} - \lambda(\mathbf{e}^{i,k})^\mathsf{T}) - \gamma_0'\mathtt{G}^{j,i}\mathtt{G}^{i,k} \;\mid\; (\gamma_{1:3}\mathbf{e}^{k,i})\mathbf{e}^{i,j} - \gamma_0'\mathtt{G}^{j,i}\mathbf{e}^{k,i}\big].
\end{aligned} \tag{45}$$

The projection matrix $\mathtt{P}'^j$ having the form $[\mathtt{M}_{3\times3}\,|\,\mathbf{m}_{3\times1}]$, one can compute the fundamental matrix $\mathtt{F}^{j,k} \cong [\mathbf{m}_{3\times1}]_\times\mathtt{M}_{3\times3}$. Now, simple algebra yields the claim of Proposition 3.

E Proof of Proposition 4

If one considers a set of connected and consistent Trifocal Tensors on a chain, with the parametrization proposed in Proposition 1:

$$\Big(\big[\mathscr{T}^{1,0,2},\boldsymbol{\gamma}^1\big],\ldots,\big[\mathscr{T}^{i,i-1,i+1},\boldsymbol{\gamma}^i\big],\ldots,\big[\mathscr{T}^{\ell;\ell-1,\ell+1},\boldsymbol{\gamma}^\ell\big]\Big) \tag{46}$$

We can compute the camera triplet $\mathtt{P}^i$, $\mathtt{P}^1$, $\mathtt{P}^\ell$ in the projective space of the Trifocal Tensor $\big[\mathscr{T}^{i,i-1,i+1},\boldsymbol{\gamma}^i\big]$. It is obvious that $\mathtt{P}^i = [\mathtt{I}_{3\times3}\,|\,\mathbf{0}_{3\times1}]$ (see equation (3)). Using the formula for the projective transform between two consecutive camera triplets, one easily derives that:

$$\mathtt{P}^\ell = [[\mathbf{e}^{\ell-1,\ell}]_\times\mathtt{F}^{\ell,\ell-1}\,|\,\mathbf{e}^{\ell-1,\ell}]\prod_{j=\ell-2}^{i}\mathtt{H}_{j,j+1}, \tag{47}$$

$$\mathtt{P}^1 = [\mathtt{I}_{3\times3}\,|\,\mathbf{0}_{3\times1}]\prod_{j=1}^{i-1}\underline{\mathtt{H}}_{j,j+1}. \tag{48}$$

The formula for $\mathtt{P}^\ell$ can be retrieved using the expression of the Trifocal Tensor associated to $\boldsymbol{\gamma}^{\ell-1}$ and transferring it to the projective space of $\big[\mathscr{T}^{i,i-1,i+1},\boldsymbol{\gamma}^i\big]$, with the help of the homographies defined by equation (6). Similarly, one retrieves $\mathtt{P}^1$ from the Trifocal Tensor $\big[\mathscr{T}^{102},\boldsymbol{\gamma}^1\big]$ by transferring to the space of $\big[\mathscr{T}^{i,i-1,i+1},\boldsymbol{\gamma}^i\big]$, with the help of the inverse homography defined by equation (5). As one can see, $\mathtt{P}^\ell$ does not depend on $\boldsymbol{\gamma}^1$, $\boldsymbol{\gamma}^i$ and $\boldsymbol{\gamma}^\ell$. So this matrix will be denoted by $\mathtt{P}^\ell = [\mathtt{M}_{3\times3}\,|\,\mathbf{m}_{3\times1}]$. We normalize it so that $\|\mathbf{m}\| = 1$. One can also notice that $[\mathtt{I}_{3\times3}\,|\,\mathbf{0}_{3\times1}]\prod_{j=1}^{i-2}\underline{\mathtt{H}}_{j,j+1}$ does not depend on $\boldsymbol{\gamma}^1$, $\boldsymbol{\gamma}^i$ and $\boldsymbol{\gamma}^\ell$. So it will be written as $[\mathtt{M}'_{3\times3}\,|\,\mathbf{m}'_{3\times1}]$. Thus, we have the triplet of cameras:

$$\mathtt{P}^i = [\mathtt{I}_{3\times3}\,|\,\mathbf{0}_{3\times1}],\ \mathtt{P}^\ell = [\mathtt{M}_{3\times3}\,|\,\mathbf{m}_{3\times1}],\ \mathtt{P}^1 = [\mathtt{M}'_{3\times3}\,|\,\mathbf{m}'_{3\times1}]\underline{\mathtt{H}}_{i-1,i}. \tag{49}$$

As stated in [1], the fundamental matrix associated to $\mathtt{P}^i$ and $\mathtt{P}^\ell$ is $\mathtt{F}^{\ell,i} = [\mathbf{m}]_\times \mathtt{M}$ and the epipole is $\mathbf{e}^{i\ell} = \mathbf{m}$.

Recall that the object we are interested in is the Trifocal Tensor $\mathscr{T}^{i,1\ell}$. We can write this trifocal tensor in a projective space such that: $\mathtt{P}'^i = [\mathtt{I}_{3\times3} \,|\, \mathbf{0}_{3\times1}]$ and $\mathtt{P}'^\ell = [[\mathbf{e}^{i,\ell}]_\times \mathtt{F}^{\ell,i} \,|\, \mathbf{e}^{i,\ell}]$, with $\mathtt{P}'^i = \mathtt{P}^i\mathtt{H}$ and $\mathtt{P}'^\ell = \mathtt{P}^\ell\mathtt{H}$. The first of these equations implies that $\mathtt{H}$ reduces to the simple form:

$$\mathtt{H} = \begin{bmatrix} \mathtt{I}_{3\times3} & \mathbf{0}_{3\times1} \\ \mathbf{a}_{1:3} & a_4 \end{bmatrix}. \tag{50}$$

Using the notation $\mathtt{G}^{\ell,i} = [\mathbf{e}^{i,\ell}]_\times \mathtt{F}^{\ell,i}$, the second equation becomes

$$[\mathtt{G}^{\ell,i} \,|\, \mathbf{e}^{i,\ell}] = [[\mathbf{m}]_\times [\mathbf{m}]_\times \mathtt{M} \,|\, \mathbf{m}] \cong [\mathtt{M}_{3\times3} \,|\, \mathbf{m}_{3\times1}]\mathtt{H}. \tag{51}$$

Since the vector $\mathbf{m}$ is of unit norm, using the triple cross product equation

$$[\mathbf{m}]_\times [\mathbf{m}]_\times \mathtt{M}_{:,i} = \mathbf{m}\,(\mathbf{m}\cdot\mathtt{M}_{:,i}) - \mathtt{M}_{:,i}, \tag{52}$$

one obviously finds that $a_4 = -1$ and $\mathbf{a}_{1:3} = -\mathbf{m}^\mathsf{T}\mathtt{M}$. Therefore, the camera $\mathtt{P}'^1$ can be written as

$$\mathtt{P}'^1 = [\mathtt{M}'_{3\times3} \,|\, \mathbf{m}'_{3\times1}]\, \underline{\mathtt{H}}_{i-1,i} \begin{bmatrix} \mathtt{I}_{3\times3} & \mathbf{0}_{3\times1} \\ -\mathbf{m}^\mathsf{T}\mathtt{M} & -1 \end{bmatrix}. \tag{53}$$

We will now retrieve the fundamental matrix $\mathtt{F}^{1,i}$ and show that is does not depend on $\boldsymbol{\gamma}^1$, $\boldsymbol{\gamma}^i$ and $\boldsymbol{\gamma}^\ell$. This fundamental matrix corresponds to the camera pair:

$$\mathtt{P}^i = [\mathtt{I}_{3\times3} \,|\, \mathbf{0}_{3\times1}], \qquad \mathtt{P}^1 = [\mathtt{M}'_{3\times3} \,|\, \mathbf{m}'_{3\times1}]\, \underline{\mathtt{H}}_{i-1,i} \tag{54}$$

That is, $\mathtt{P}^i = [\mathtt{I}_{3\times3} \,|\, \mathbf{0}_{3\times1}]$ and

$$\mathtt{P}^1 = [\mathtt{M}'_{3\times3} \,|\, \mathbf{m}'_{3\times1}] \begin{bmatrix} [\mathbf{e}^{i,i-1}]_\times \mathtt{F}^{i-1,i} & \mathbf{e}^{i,i-1} \\ -\mu\mathbf{e}^{i-1,i} & 0 \end{bmatrix} \begin{bmatrix} -\gamma^i_0 \mathtt{I}_{3\times3} & \mathbf{0}_{3\times1} \\ \boldsymbol{\gamma}^i_{1:3} & 1 \end{bmatrix}. \tag{55}$$

The two left matrices of the camera decomposition $\mathtt{P}^1$ do not depend on $\boldsymbol{\gamma}^1$, $\boldsymbol{\gamma}^i$ and $\boldsymbol{\gamma}^\ell$, so one can merge these two left matrices under the fixed form $[\mathtt{M}''_{3\times3} \,|\, \mathbf{m}''_{3\times1}]$. We deduce that the camera triplet in equation (53) can be written as $\mathtt{P}'^i = [\mathtt{I}_{3\times3} \,|\, \mathbf{0}_{3\times1}]$, $\mathtt{P}'^\ell = [[\mathbf{e}^{i,\ell}]_\times \mathtt{F}^{\ell,i} \,|\, \mathbf{e}^{i,\ell}]$ and

$$\mathtt{P}'^1 = [\mathtt{M}''_{3\times3} \,|\, \mathbf{m}''_{3\times1}] \begin{bmatrix} -\gamma^i_0 \mathtt{I}_{3\times3} & \mathbf{0}_{3\times1} \\ \boldsymbol{\gamma}^i_{1:3} & 1 \end{bmatrix} \begin{bmatrix} \mathtt{I}_{3\times3} & \mathbf{0}_{3\times1} \\ -\mathbf{m}^\mathsf{T}\mathtt{M} & -1 \end{bmatrix}. \tag{56}$$

Moreover if we normalize $[\mathtt{M}''_{3\times3} \,|\, \mathbf{m}''_{3\times1}]$ so that $\|\mathbf{m}''\| = 1$, we recover the fundamental matrix $\mathtt{F}^{1,i} = [\mathbf{m}'']_\times \mathtt{M}''$ and the unit norm epipole $\mathbf{e}^{i,1} = \mathbf{m}''$. Furthermore, using computations similar to those leading to (50), and noticing that $\mathtt{H}^{-1} = \mathtt{H}$, we can write the camera triplet:

$$\begin{aligned} \mathtt{P}'^i &= [\mathtt{I}_{3\times3} \,|\, \mathbf{0}_{3\times1}], \qquad \mathtt{P}'^\ell = [\mathtt{G}^{\ell,i} \,|\, \mathbf{e}^{i,\ell}], \\ \mathtt{P}'^1 &= [\mathtt{G}^{1,i} \,|\, \mathbf{e}^{i,1}] \begin{bmatrix} \mathtt{I}_{3\times3} & \mathbf{0} \\ \mathbf{a}'_{1\times3} & a'_4 \end{bmatrix} \begin{bmatrix} -\gamma^i_0 \mathtt{I}_{3\times3} & \mathbf{0} \\ \boldsymbol{\gamma}^i_{1:3} & 1 \end{bmatrix} \begin{bmatrix} \mathtt{I}_{3\times3} & \mathbf{0} \\ \mathbf{a}_{1\times3} & a_4 \end{bmatrix} \end{aligned} \tag{57}$$

with $a'_4 = -1$ and $\mathbf{a}'_{1\times3} = -\mathbf{m}''^\mathsf{T}\mathtt{M}''$. Finally we want to express, by identification, this triplet in a projective space matching the formulation in Proposition 1. The tensor $\mathscr{T}^{i,1,\ell}$ should be equivalent to the camera triplet:

$$\begin{aligned} &\mathtt{P}^i = [\mathtt{I}_{3\times3} \,|\, \mathbf{0}_{3\times1}], \qquad \mathtt{P}^\ell = [[\mathbf{e}^{i,\ell}]_\times \mathtt{F}^{\ell,i} \,|\, \mathbf{e}^{i,\ell}], \\ &\mathtt{P}^1 = \mathrm{kron}\big([\boldsymbol{\gamma}^s_{1:3}, 1]; \mathbf{e}^{i,1}\big) - \gamma^s_0 [[\mathbf{e}^{i,1}]_\times \mathtt{F}^{1,i} \,|\, \mathbf{0}_{3\times1}], \end{aligned} \tag{58}$$

One has just to solve,

$$\begin{bmatrix} -\gamma^s_0 \mathtt{I} & \mathbf{0} \\ \boldsymbol{\gamma}^s_{1:3} & 1 \end{bmatrix} \cong \begin{bmatrix} \mathtt{I} & \mathbf{0} \\ \mathbf{a}'_{1\times3} & -1 \end{bmatrix} \begin{bmatrix} -\gamma^i_0 \mathtt{I} & \mathbf{0} \\ \boldsymbol{\gamma}^i_{1:3} & 1 \end{bmatrix} \begin{bmatrix} \mathtt{I} & \mathbf{0} \\ \mathbf{a}_{1\times3} & -1 \end{bmatrix}. \tag{59}$$

Solving this linear system, we retrieve the coefficients of the affine transformation in equation (20) and the desired result follows.

F Proof of Proposition 5

Let us consider the loop made of four Trifocal Tensors, $\big[\mathscr{T}^{i,j,k}, \boldsymbol{\gamma}^1\big]$, $\big[\mathscr{T}^{k,i,\ell}, \boldsymbol{\gamma}^2\big]$, $\big[\mathscr{T}^{\ell,k,j}, \boldsymbol{\gamma}^3\big]$ and $\big[\mathscr{T}^{j,\ell,i}, \boldsymbol{\gamma}^4\big]$. A fundamental matrix having 7 degrees of freedom, the fact that the tensors $\mathscr{T}^{i,j,k}$ and $\mathscr{T}^{\ell,k,j}$ share the matrix $\mathtt{F}^{j,k}$ leads to 7 "coherence" constraints. By a symmetric argument, $\mathscr{T}^{k,i,\ell}$ and $\mathscr{T}^{j,\ell,i}$ share the fundamental matrix $\mathtt{F}^{i\ell}$, giving raise to additional 7 constraints. According to Proposition 3, all these constraints are linear. So we have 14 independent linear "coherence" constraints.

In view of Proposition 3, the fundamental matrix $\mathtt{F}^{j,k}$ (resp. $\mathtt{F}^{kj}$) can be deduced from the Trifocal Tensor $\mathscr{T}^{i,j,k}$ (resp. $\mathscr{T}^{\ell,k,j}$) as follows:

$$\begin{aligned} \mathtt{F}^{j,k} = &\Big[\mathtt{G}^{j,i}\mathbf{e}^{k,i}\Big]_\times \Big[\Big(\mathbf{e}^{i,j}{\boldsymbol{\gamma}^1_{1:3}}^\mathsf{T} - \gamma^1_0 \mathtt{G}^{j,i}\Big)\mathtt{G}^{i,k} - \boldsymbol{\lambda}^{i,k}\mathbf{e}^{i,j}{\mathbf{e}^{i,k}}^\mathsf{T}\Big] \\ &+ \Big(\boldsymbol{\gamma}^1_{1:3}\cdot\mathbf{e}^{k,i}\Big)[\mathbf{e}^{i,j}]_\times \mathtt{G}^{j,i}\mathtt{G}^{i,k}. \end{aligned} \tag{60}$$

$$\begin{aligned} \mathtt{F}^{kj} = &\Big[\mathtt{G}^{k,\ell}\mathbf{e}^{j,\ell}\Big]_\times \Big[\Big(\mathbf{e}^{\ell,k}{\boldsymbol{\gamma}^3_{1:3}}^\mathsf{T} - \gamma^3_0 \mathtt{G}^{k,\ell}\Big)\mathtt{G}^{\ell,j} - \boldsymbol{\lambda}^{\ell,j}\mathbf{e}^{\ell,k}{\mathbf{e}^{\ell,j}}^\mathsf{T}\Big] \\ &+ \Big(\boldsymbol{\gamma}^3_{1:3}\cdot\mathbf{e}^{j,\ell}\Big)[\mathbf{e}^{\ell,k}]_\times \mathtt{G}^{k,\ell}\mathtt{G}^{\ell,j}, \end{aligned} \tag{61}$$

where we used the notation $\boldsymbol{\lambda}^{i,k} = -\frac{1}{2}\,\mathrm{tr}([\mathbf{e}^{i,k}]_\times \mathtt{F}^{k,i}[\mathbf{e}^{k,i}]_\times \mathtt{F}^{i,k})$ and $\boldsymbol{\lambda}^{\ell j} = -\frac{1}{2}\,\mathrm{tr}([\mathbf{e}^{\ell j}]_\times \mathtt{F}^{j,\ell}[\mathbf{e}^{j,\ell}]_\times \mathtt{F}^{\ell j})$. A set of 7 coherence equations is thus obtained from $\eta_1 \mathtt{F}^{j,k} = \mathtt{F}^{kj\mathsf{T}}$, where η_1 is a real number. In the same way, we get seven other coherence equations $\eta_2 \mathtt{F}^{i,\ell} = (\mathtt{F}^{\ell,i})^\mathsf{T}$. Multiplying this equation from right by $\mathbf{e}^{i,k}$ and from left by $(\mathbf{e}^{\ell,j})^\mathsf{T}$, and using equations (60) and (61), we get:

$$\eta_1\beta\mathbf{e}^{\ell,j\mathsf{T}}\Big[\mathtt{G}^{j,i}\mathbf{e}^{k,i}\Big]_\times \mathbf{e}^{i,j} = \boldsymbol{\lambda}^{\ell,j}\mathbf{e}^{\ell,k\mathsf{T}}\Big[\mathtt{G}^{k,\ell}\mathbf{e}^{j,\ell}\Big]_\times{}^\mathsf{T}\,\mathbf{e}^{i,k}. \tag{62}$$

This allows us to compute the scale factor η_1. In the way we retrieve the scale factor η_2. Hence, we get a system of $9\times2 = 18$ equations of rank 14. It remains to find one more (independent) coherence constraint.

Since $\mathscr{T}^{i,j,k}$ and $\mathscr{T}^{\ell,k,j}$ have the fundamental matrix $\mathtt{F}^{j,k}$ in common, it exists a projective homography mapping one camera triplet to the other camera triplet. Therefore, one can derive the form of $\mathscr{T}^{k,i,\ell}$. Indeed, in equation (45), Section D, we have seen that the camera triplet corresponding to $\mathscr{T}^{i,j,k}$ can be written in one projective space:

$$\begin{aligned} &\mathtt{P}'^i = [\mathtt{G}^{i,k} \,|\, \mathbf{e}^{k,i}], \qquad \mathtt{P}'^k = [\mathtt{I}_{3\times3} \,|\, \mathbf{0}_{3\times1}], \\ &\mathtt{P}'^j = [\mathtt{G}^{j,i} \,|\, \mathbf{e}^{i,j}] \begin{bmatrix} -\gamma^1_0 \mathtt{I}_{3\times3} & \mathbf{0} \\ \boldsymbol{\gamma}^1_{1:3} & 1 \end{bmatrix} \begin{bmatrix} \mathtt{G}^{i,k} & \mathbf{e}^{k,i} \\ -\boldsymbol{\lambda}\mathbf{e}^{i,k\mathsf{T}} & 0 \end{bmatrix}, \end{aligned} \tag{63}$$

Now, if we consider the tensor $\mathscr{T}^{k,i,\ell}$, it corresponds to the camera triplet:

$$\mathtt{P}^k = [\mathtt{I} \,|\, \mathbf{0}],\ \mathtt{P}^\ell = [\mathtt{G}^{\ell,k} \,|\, \mathbf{e}^{k,\ell}],\ \mathtt{P}^i = [\mathtt{G}^{i,k} \,|\, \mathbf{e}^{k,i}] \begin{bmatrix} -\gamma^2_0 \mathtt{I}_{3\times3} & \mathbf{0} \\ \boldsymbol{\gamma}^2_{1:3} & 1 \end{bmatrix}, \tag{64}$$

Considering the cameras k and i in equations (63) and (64), one can see that the projective transform between these camera triplets is

$$\mathtt{H} = \begin{bmatrix} -\gamma^2_0 \mathtt{I}_{3\times3} & \mathbf{0} \\ \boldsymbol{\gamma}^2_{1:3} & 1 \end{bmatrix}. \tag{65}$$

Finally, the tensor $\mathscr{T}^{i,j,k}$ in the projective space of $\mathscr{T}^{k,i,\ell}$ is:

$$\begin{aligned} &\mathtt{P}'^i = [\mathtt{G}^{i,k} \,|\, \mathbf{e}^{k,i}] \begin{bmatrix} -\gamma^2_0 \mathtt{I} & \mathbf{0} \\ \boldsymbol{\gamma}^2_{1:3} & 1 \end{bmatrix}, \quad \mathtt{P}'^k = [\mathtt{I} \,|\, \mathbf{0}], \\ &\mathtt{P}'^j = [\mathtt{G}^{j,i} \,|\, \mathbf{e}^{i,j}] \begin{bmatrix} -\gamma^1_0 \mathtt{I} & \mathbf{0} \\ \boldsymbol{\gamma}^1_{1:3} & 1 \end{bmatrix} \begin{bmatrix} \mathtt{G}^{i,k} & \mathbf{e}^{k,i} \\ -\boldsymbol{\lambda}\mathbf{e}^{i,k\mathsf{T}} & 0 \end{bmatrix} \begin{bmatrix} -\gamma^2_0 \mathtt{I} & \mathbf{0} \\ \boldsymbol{\gamma}^2_{1:3} & 1 \end{bmatrix}. \end{aligned} \tag{66}$$

We will also compute the tensor $\mathscr{T}^{\ell,k,j}$ in the projective space of $\mathscr{T}^{k,i,\ell}$. The camera triplet corresponding to $\mathscr{T}^{\ell,k,j}$ is:

$$\begin{array}{l} \mathtt{P}^{\ell} = [\mathtt{I}_{3\times 3}\,|\,\mathbf{0}_{3\times 1}], \qquad \mathtt{P}^{j} = [[\mathbf{e}^{\ell,j}]_{\times}\mathtt{F}^{j,\ell}\,|\,\mathbf{e}^{\ell,j}], \\ \mathtt{P}^{k} = \mathrm{kron}\big([\gamma^3_{1:3},1];\mathbf{e}^{\ell,k}\big) - \gamma^3_0[[\mathbf{e}^{\ell,k}]_{\times}\mathtt{F}^{k,\ell}\,|\,\mathbf{0}_{3\times 1}]. \end{array} \tag{67}$$

We look for a projective transform $\mathtt{H}$ such that : $\mathtt{P}''^{\ell} = \mathtt{P}^{\ell}\mathtt{H} = [\mathtt{G}^{\ell,k}\,|\,\mathbf{e}^{k,\ell}]$ and $\mathtt{P}''^{k} = \mu\mathtt{P}^{k}\mathtt{H} = [\mathtt{I}\,|\,\mathbf{0}]$. The first equation implies that $\mathtt{H}_{1:3,1:4} = [[\mathbf{e}^{k,\ell}]_{\times}\mathtt{F}^{\ell,k}\,|\,\mathbf{e}^{k,\ell}]$. Inserting that equality in the second equation we get:

$$\mathbf{e}^{\ell,k}\left(\gamma^3_{1:3}\cdot\mathbf{e}^{k,\ell}\right) + \mathtt{H}_{4,4}\mathbf{e}^{\ell,k} = \mathbf{0}_{3\times 1}, \tag{68}$$

$$\mu\,\mathtt{I} = \mathbf{e}^{\ell,k}{\gamma^3_{1:3}}^{\mathsf{T}}\mathtt{G}^{\ell,k} - \gamma^3_0\mathtt{G}^{k,\ell}\mathtt{G}^{\ell,k} + \mathbf{e}^{\ell,k}\mathtt{H}_{4,1:3}. \tag{69}$$

Left-multiplying the second equation by $[\mathbf{e}^{\ell,k}]_{\times}[\mathbf{e}^{\ell,k}]_{\times}$ and computing the trace, we get

$$\mu = \frac{1}{2}\mathrm{trace}\left([\mathbf{e}^{\ell,k}]_{\times}[\mathbf{e}^{\ell,k}]_{\times}\mathtt{G}^{k,\ell}\mathtt{G}^{\ell,k}\right). \tag{70}$$

One can find $\mathtt{H}_{4,1:3}$ by left-multiplying the second equation by ${\mathbf{e}^{\ell,k}}^{\mathsf{T}}$:

$$\mathtt{H} = \begin{bmatrix} \mathtt{G}^{\ell,k} & | & \mathbf{e}^{k,\ell} \\ \mu\gamma^3_0{\mathbf{e}^{\ell,k}}^{\mathsf{T}} - {\gamma^3_{1:3}}^{\mathsf{T}}\mathtt{G}^{\ell,k} & | & -\gamma^3_{1:3}\cdot\mathbf{e}^{k,\ell} \end{bmatrix}. \tag{71}$$

So, the tensor $\mathscr{T}^{\ell,k,j}$ admits the following representation in the projective space of $\mathscr{T}^{k,i,\ell}$:

$$\begin{array}{l} \mathtt{P}''^{\ell} = [\mathtt{G}^{\ell,k}\,|\,\mathbf{e}^{k,\ell}], \qquad \mathtt{P}''^{k} = [\mathtt{I}_{3\times 3}\,|\,\mathbf{0}_{3\times 1}], \\ \mathtt{P}''^{j} = [\mathtt{G}^{j,\ell}\,|\,\mathbf{e}^{\ell,j}]\begin{bmatrix} \mathtt{G}^{\ell,k} & | & \mathbf{e}^{k,\ell} \\ \mu\gamma^3_0{\mathbf{e}^{\ell,k}}^{\mathsf{T}} - {\gamma^3_{1:3}}^{\mathsf{T}}\mathtt{G}^{\ell,k} & | & -\gamma^3_{1:3}\cdot\mathbf{e}^{k,\ell} \end{bmatrix}, \end{array} \tag{72}$$

As seen from equations (66) and (72), one has camera triplets associated to tensors $\mathscr{T}^{i,j,k}$ and $\mathscr{T}^{\ell,k,j}$ in the same projective space as that of $\mathscr{T}^{k,i,\ell}$. So one can write the last constraints as being $\mathtt{P}''^{j} \cong \mathtt{P}'^{j}$, or more precisely $[\mathbf{e}^{i,j}]_{\times}\mathtt{P}''^{j} = \eta_3[\mathbf{e}^{i,j}]_{\times}\mathtt{P}'^{j}$. We aim at computing the scale factor η_3. Using the last column and left-multiplying by $\left[[\mathbf{e}^{i,j}]_{\times}\mathbf{e}^{\ell,j}\right]_{\times} = \mathtt{X}$ one gets:

$$\mathtt{X}[\mathbf{e}^{i,j}]_{\times}\mathtt{G}^{j,\ell}\mathbf{e}^{k,\ell} = -\eta_3\gamma^1_0\mathtt{X}[\mathbf{e}^{i,j}]_{\times}\mathtt{G}^{j,i}\mathbf{e}^{k,i}. \tag{73}$$

This leads to the value of $\eta_3\gamma^1_0$, and the equation

$$[\mathbf{e}^{i,j}]_{\times}\mathtt{P}''^{j} = \eta_3[\mathbf{e}^{i,j}]_{\times}\mathtt{P}'^{j}, \tag{74}$$

becomes simple and linear. We get 12 more constraints, which are however redundant with the 14 previous ones, the underlying rank being equal to 15.

Acknowledgements

The authors acknowledge support by the Agence Nationale de la Recherche (ANR) under grant Callisto (ANR-09-CORD-003).

References

1. Hartley, R., Zisserman., A.: Multiple view geometry in computer vision. Cambridge University, 2nd edition, (2003)
2. Faugeras, O., Luong, Q.T., Papadopoulou, T.: The Geometry of Multiple Images: The Laws That Govern The Formation of Images of A Scene and Some of Their Applications. MIT Press, Cambridge, MA, USA (2001)
3. Snavely, N., Seitz, S.M., Szeliski, R.: Photo tourism: Exploring photo collections in 3D. ACM Press, New York, NY, USA (2006)
4. Martinec, D., Pajdla, T.: Robust rotation and translation estimation in multiview reconstruction. In: CVPR. (2007)
5. Snavely, N., Seitz, S.M., Szeliski, R.: Modeling the world from Internet photo collections. Int. J. Comput. Vision **80** (2008) 189–210
6. Furukawa, Y., Ponce, J.: Accurate camera calibration from multi-view stereo and bundle adjustment. International Journal of Computer Vision **84** (2009) 257–268
7. Bujnak, M., Kukelova, Z., Pajdla, T.: 3D reconstruction from image collections with a single known focal length. In: ICCV. (2009) 351–358
8. Agarwal, S., Snavely, N., Simon, I., Seitz, S.M., Szeliski, R.: Building Rome in a day. In: ICCV. (2009)
9. Havlena, M., Torii, A., Knopp, J., Pajdla, T.: Randomized structure from motion based on atomic 3d models from camera triplets. In: CVPR. (2009)
10. Mouragnon, E., Lhuillier, M., Dhome, M., Dekeyser, F., Sayd, P.: Generic and real-time sfm using local bundle adjustment. Image Vision Comput. **27** (2009) 1178–1193
11. Fitzgibbon, A.W., Zisserman, A.: Automatic camera recovery for closed or open image sequences. In: ECCV. (1998) 311–326
12. Avidan, S., Shashua, A.: Threading fundamental matrices. IEEE Trans. Pattern Anal. Mach. Intell. **23** (2001) 73–77
13. Courchay, J., Dalalyan, A.S., Keriven, R. and Sturm, P. Exploiting loops in the graph of trifocal tensors for calibrating a network of cameras. In: ECCV (2010).
14. Pollefeys, M., Van Gool, L., Vergauwen, M., Verbiest, F., Cornelis, K., Tops, J., Koch, R.: Visual modeling with a hand-held camera. Int. J. Comput. Vision **59** (2004) 207–232
15. Sinha, S.N., Pollefeys, M., McMillan, L.: Camera network calibration from dynamic silhouettes. In: CVPR. (2004)
16. Tomasi, C., Kanade, T.: Shape and motion from image streams under orthography: a factorization method. Int. J. Comput. Vision **9** (1992) 137–154
17. Sturm, P., Triggs, B.: A factorization based algorithm for multi-image projective structure and motion. In: ECCV (2). (1996) 709–720
18. Jacobs, D.: Linear fitting with missing data: Applications to structure-from-motion and to characterizing intensity images. In: CVPR. (1997) 206
19. Martinec, D., Pajdla, T.: 3D reconstruction by fitting low-rank matrices with missing data. In: CVPR. (2005) I: 198–205
20. Triggs, B., McLauchlan, P., Hartley, R., Fitzgibbon, A.: Bundle adjustment - a modern synthesis. In: Workshop on Vision Algorithms. (1999) 298–372
21. Klopschitz, M., Zach, C., Irschara, A., Schmalstieg, D.: Generalized detection and merging of loop closures for video sequences. In: 3DPVT. (2008)
22. Scaramuzza, D., Fraundorfer, F., Pollefeys, M.: Closing the loop in appearance-guided omnidirectional visual odometry by using vocabulary trees. Robot. Auton. Syst. **58**(2010) 820-827
23. Cornelis, N., Cornelis, K., Van Gool, L.: Fast compact city modeling for navigation pre-visualization. In: CVPR. (2006)
24. Tardif, J., Pavlidis, Y., Daniilidis, K.: Monocular visual odometry in urban environments using an omnidirectional camera. In: IROS. (2008) 2531–2538
25. Torii, A., Havlena, M., Pajdla, T.: From google street view to 3d city models. In: OMNIVIS. (2009)
26. Vu, H., Keriven, R., Labatut, P., Pons, J.P.: Towards high-resolution large-scale multi-view stereo. In: CVPR. (2009)
27. Chum, O., Matas, J.: Matching with PROSAC: Progressive sample consensus. In: CVPR. (2005) I: 220–226
28. Ponce, J., McHenry, K., Papadopoulo, T., Teillaud, M., Triggs, B.: On the absolute quadratic complex and its application to autocalibration. In: CVPR. (2005) 780–787

29. Quan, L.: Invariants of six points and projective reconstruction from three uncalibrated images. IEEE Trans. Pattern Anal. Mach. Intell. **17** (1995) 34–46
30. Schaffalitzky, F., Zisserman, A., Hartley, R.I., Torr, P.H.S.: A six point solution for structure and motion. In: ECCV, London, UK, Springer-Verlag (2000) 632–648
31. Golub, G.H., Van Loan, C.F.: Matrix computations. Third edn. Johns Hopkins Studies in the Mathematical Sciences. Johns Hopkins University Press, Baltimore, MD (1996)
32. Torr, P.H.S.: An Assessment of Information Criteria for Motion Model Selection. In CVPR (1997) 47–53
33. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. IJCV **60** (2004) 91–110
34. Chum, O., Werner, T., Matas, J.: Two-view geometry estimation unaffected by a dominant plane. In: CVPR. (2005) I: 772–779
35. Sattler, T., Leibe, B., Kobbelt, L.: Scramsac: Improving ransac's efficiency with a spatial consistency filter. In: ICCV. (2009)
36. Strecha, C., von Hansen, W., Van Gool, L., Fua, P., Thoennessen, U.: On Benchmarking Camera Calibration and Multi-View Stereo for High Resolution Imagery. In: CVPR. (2008)
37. Bertsekas, D.: Nonlinear Programming. 2nd edition, Athena Scientific, Belmont. (1999).
38. Estrada, C., Neira, J. and Tardós, J. D.: Hierarchical SLAM: real-time accurate mapping of large environments. IEEE Transactions on Robotics, **21** (2005), 588–596
39. Govindu, M.: Robustness in Motion Averaging. In ACCV (2006) 457–466
40. Maybank, S. J. and Shashua, A.: Ambiguity in reconstruction from images of six points. In ICCV (1998) 703–708
41. Zach, C., Klopschitz, M., Pollefeys, M.: Disambiguating visual relations using loop constraints. In CVPR (2010) 1426–1433
42. Gherardi, R., Farenzena, M. and Fusiello, A.: Improving the efficiency of hierarchical structure-and-motion. In CVPR (2010) 1594-1600
43. Snavely, N., Seitz, S. and Szeliski, R.: Skeletal Sets for Efficient Structure from Motion. In CVPR (2008) 1–8