Computer Vision and Image Understanding 210 (2021) 103223

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

journal homepage: www.elsevier.com/locate/cviu

A linear method for camera pair self-calibration

Nikos Melanitis [*,1], Petros Maragos [1]

National Technical University of Athens, 9, Iroon Polytechniou Str., 15780 Zografos Athens, Greece

ARTICLE INFO

Communicated by Nikos Paragios

Keywords:
Self-calibration
Multi-view reconstruction
Rotation averaging
Multi-view geometry
Structure from motion
Absolute conic

ABSTRACT

We examine 3D reconstruction in unordered sets of uncalibrated images. We introduce a linear method to self-calibrate and find the metric reconstruction of a camera pair. We assume unknown and varying focal lengths but otherwise known internal camera parameters and a known projective reconstruction of the camera pair. We recover two possible camera configurations in space and use the Cheirality condition, that all 3D scene points are in front of both cameras, to disambiguate the solution. Towards identifying camera configurations that would perplex solution disambiguation, we show in two Theorems, first that the two solutions are in mirror positions and then the relations between their viewing directions. We validate our approach in synthetic and real scenes. In camera pair self-calibration and metric reconstruction, our method performs on par (median rotation error $\Delta R = 3.49°$) with the standard approach of Kruppa equations followed by 5P algorithm ($\Delta R = 3.77°$). We get realistic multi-view reconstructions, using numerous camera pair metric reconstructions generated by our linear method, rotation-averaging algorithms and a novel method to average focal length estimates.

1. Introduction

Multi-view geometry (mvg) is a Computer Vision (CV) subfield that attempts to understand the structure of the 3D world given a collection of its images (Hartley and Zisserman, 2004). As the binocular human vision is naturally 3D, the same underlying principles allow the recovery of the 3D world structure in mvg reconstruction methods. However, a prerequisite is to have calibrated cameras, an assumption that is often violated in unordered image sets, in which we only have images that are obtained from various sources (e.g. found on the internet). In this paper we focus on self-calibration and multi-view reconstruction using relations between camera pairs.

Assuming a camera pair with unknown and different focal lengths as the only unknown internal parameters, a standard robust approach to self-calibration and metric reconstruction first applies the 7-point algorithm (Hartley and Zisserman, 2004) inside a RANSAC (Fischler and Bolles, 1981) procedure to find the fundamental matrix. In this projective framework, the Kruppa equations (Hartley, 1997) are used to determine the unknown focal lengths. Next, applying the 5-point algorithm inside a RANSAC procedure (Nistér, 2004), leads to a metric reconstruction. Since focal lengths are recovered in a projective framework, only epipolar geometry constraints (a point in one image must lie on the corresponding epipolar line in another image) may be used to check the solution plausibility. Solving self-calibration and metric reconstruction problems simultaneously, permits application of more intuitive and restrictive geometric arguments, i.e the Cheirality condition, to solidly verify both extrinsic and intrinsic camera parameters.

Self-calibration methods are derived from relations on the Dual Absolute Conic (DAC, also appears as 'absolute quadric' in the literature) Q_∞^* and the Dual Image of the Absolute Conic (DIAC) ω_∞^* (Pollefeys et al., 1999; Seo et al., 2001; Hartley, 1997). However, existing methods require three or four images to provide a solution (Seo et al., 2001), use numerical methods to determine DAC when more than two views are used (Pollefeys et al., 1999), provide an initial DAC estimate that violates the rank-2 condition (Pollefeys et al., 1999; Seo et al., 2001) and do not examine the relations between the recovered putative solutions (Pollefeys et al., 1999). A self-calibration method for two views has been proposed which arrives at a one-dimensional space of solutions that can be reduced to four solutions by imposing the rank-2 condition (Pollefeys et al., 1999). In mvg reconstructions additional assumptions have been made to determine focal lengths, as availability of EXIF tags (Olsson and Enqvist, 2011; Snavely et al., 2010), equality of focal lengths across all images (Martinec and Pajdla, 2007; Stewénius et al., 2005) and vanishing points correspondences (Sinha et al., 2010).

Towards a multi-view reconstruction, camera pairs have been utilized in previous approaches. Different estimates for a rotation matrix R can be combined with a rotation averaging algorithm (Hartley et al., 2011) and reconstructions of pairs of images can be combined with rotation registration methods (Govindu, 2001, 2004; Hartley et al.,

* Corresponding author.
E-mail addresses: nimepar@gmail.com (N. Melanitis), maragos@cs.ntua.gr (P. Maragos).
[1] School of Electrical and Computer Engineering, National Technical University of Athens, Athens, Greece.

https://doi.org/10.1016/j.cviu.2021.103223
Received 27 January 2020; Received in revised form 17 May 2021; Accepted 18 May 2021
Available online 27 May 2021

2013) to initialize an instance of Structure-from-Motion with known rotations (Kahl and Hartley, 2008; Olsson and Enqvist, 2011) and produce a multiple-view reconstruction.

Concerning robustness, erroneous solutions in mvg problems are directly caused by erroneous or noisy image correspondences. Two complementary general approaches, applying RANSAC procedures to repeatedly sample minimal point sets and verifying the initial point correspondences, have been utilized in different problems to improve the validity of the recovered solution (Chum and Matas, 2012; Snavely et al., 2006).

In this paper, we assume that the only available input are the images and use image pair relations and averaging methods to robustly reconstruct the depicted scene. More specifically, we derive a linear method for the self-calibration and metric reconstruction of camera pairs with unknown and different focal lengths, unifying two problems that were previously solved independently, to a single system of equations. We further examine the two solutions recovered by our method through the derivation of two theorems about the solutions' positions and viewing directions. We integrate our aforementioned methods to a multi-view reconstruction pipeline, utilizing L_∞-norm algorithms and introducing a method to average different estimates of a single focal length f_i, which uses the structure of the problem, specifically that each estimate for f_i comes from a pair of images i, j and is so paired with a second estimate f_j. Our main innovations and contributions are[2]:

- We introduce a novel linear method for the self-calibration and metric reconstruction of an image pair which recovers the minimum possible number of solutions (two). Solution disambiguation is simplified as the ambiguity in the recovered solutions is minimized.
- We extend camera-pair self-calibration and metric reconstruction theory with a method, based on DIAC, which acquires a closed-form solution. Our equations and solutions may be further analyzed in the future to quantify the effect of noise or special camera-pair configurations.
- We exploit the underlying geometry, specifically that all reconstructed world points must lie in front of the cameras (Cheirality condition), to disambiguate the metric reconstruction solutions. We also apply the Cheirality condition to reject pair-based solutions in the practical case of outlier ridden image correspondences. We show that heuristics can be avoided when simple geometric arguments are applied.
- We describe, in two theorems, the relative orientation and positions of the two recovered metric-reconstruction solutions. We identify (in Appendix) critical camera configurations. This theory can be used to determine if a given camera configuration can be disambiguated easily.
- In multi-view reconstruction, we shift the focus from getting a single accurate solution to averaging multiple solutions. We integrate our methods to a multi-view reconstruction pipeline that is based in averaging methods for relative rotations, registered rotations and focal lengths. We recover numerous image pair metric reconstructions in minimal correspondences sets (eight-point) and average the recovered solutions. We show that our methods for fast camera-pair self-calibration and metric reconstruction can be integrated in the framework of optimal algorithms in mvg under the L_1 (e.g. rotation averaging) and L_∞ (e.g. structure from motion with known rotations) norms (Hartley and Kahl, 2007).
- We introduce a focal length averaging method and a measure ('Joint confidence count') to evaluate fit of focal length estimates. Our focal length averaging relies on the introduced image pair self-calibration method and utilizes the pair-based recovery of focal lengths.

[2] Implementations available at https://github.com/nmelan/Multi-view-Geometry.

The rest of this article is organized as follows: Section 2 briefly outlines mvg background and discusses related work. Section 3 introduces our method for self-calibration and metric reconstruction. In Section 4, we integrate our methods to a reconstruction pipeline. In doing so, we develop novel averaging methods for f estimates recovered from image pairs. Results for camera pair reconstruction, focal length averaging and multi-view reconstruction are given in Section 5. Section 6 contains our conclusions.

2. Background & related work

In the following bold font (e.g. $\mathbf{v}$) is used for vectors and capital case normal font (e.g. K) is reserved for matrices.

2.1. Elements of multiple view geometry

In this section we summarize basic notions about the projection of 3D scenes to 2D planes (Hartley and Zisserman, 2004; Faugeras et al., 2004). In a metric reconstruction parallel world lines converge at the plane at infinity $\boldsymbol{\pi}_\infty$: $\begin{pmatrix}0 & 0 & 0 & 1\end{pmatrix}^T$. The absolute conic Ω_∞ is a conic on $\boldsymbol{\pi}_\infty$ which satisfies $x_1^2 + x_2^2 + x_3^2 = 1$, $x_4 = 0$, where $\mathbf{X} = \begin{pmatrix}x_1 & x_2 & x_3 & x_4\end{pmatrix}^T$ is the homogeneous representation of world points.

By taking all the planes tangent to Ω_∞, we construct Q_∞^*, which is the dual surface of Ω_∞. Q_∞^* is described in a metric reconstruction by the 4×4 matrix

$$Q_\infty^* = \begin{bmatrix} I_{3\times3} & \mathbf{0_3} \\ \mathbf{0_3^T} & 0 \end{bmatrix} \tag{1}$$

Now, considering projective reconstructions of 3-space (a 3D scene) and projections to image plane we have the following Results (Hartley and Zisserman, 2004):

Result 1. The projection of Q^* by projection matrix P in the image plane is the dual conic

$$D^* = PQ^*P^T \tag{2}$$

Result 2. If the 3-space is transformed by homography H, that is $\mathbf{X}' = H\mathbf{X}$, then planes of 3-space are transformed according to

$$\boldsymbol{\pi}' = H^{-T}\boldsymbol{\pi} \tag{3}$$

Result 3. If H is a 4×4 matrix representing a projective transformation of 3-space, then the fundamental matrices corresponding to the pairs of camera matrices $\{P, P'\}$, $\{PH, P'H\}$ are the same.

Result 4. Suppose the $\operatorname{rank} 2$ matrix F can be decomposed in two different ways as

$$F = [\mathbf{a}]_x A \tag{4}$$

$$F = [\hat{\mathbf{a}}]_x \hat{A} \tag{5}$$

then

$$\hat{\mathbf{a}} = \kappa\mathbf{a} \tag{6}$$

$$\hat{A} = \kappa^{-1}(A + \mathbf{a}\mathbf{v}^\mathbf{T}) \tag{7}$$

for some non-zero constant κ and 3-vector $\mathbf{v}$

Using the preceding Results, we formulate the equations to solve the camera self-calibration problem and to determine $\boldsymbol{\pi}_\infty$ position in a projective reconstruction.

We summarize our notation in Table 1. We use subscripts $(1, 2, i, j \ldots)$ to refer to different cameras and superscripts to refer to different solutions in the metric reconstruction of the camera pair. We use K to denote internal calibration matrices, R for rotation matrices and $\mathbf{C}$ for camera center of projection. In Fig. 1 we vizualize camera projection and geometric entities $(Q_\infty^*, \omega_\infty, \pi_\infty)$ we use to derive our self-calibration and metric reconstruction method.

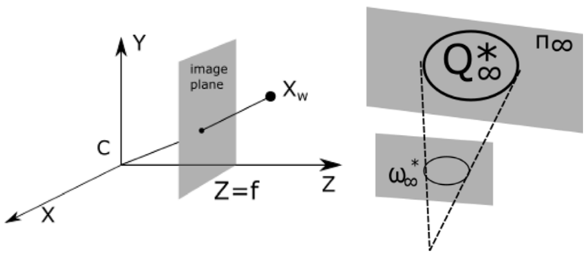

Fig. 1. Camera geometry. **Left:** Pinhole camera projection. Let the center of projection $\mathbf{C}$ be at the origin of a Euclidean coordinate system and the plane $Z = f$ be the image plane. A world point $\mathbf{X_w}$ is projected to the image by a ray joining $\mathbf{C}$ with $\mathbf{X_w}$. In general, the camera and world coordinate frames are related by a Euclidean (Rotation R and translation $\mathbf{t}$) transformation. **Right:** The dual absolute conic Q^*_∞ projects to the dual image of the absolute conic ω^*_∞. The projection equation provides the constraints we use in this paper to solve came pair self-calibration and metric reconstruction problem.

Table 1
A summary of notation, with references to uses in text.

subscripts	
M	Metric Reconstruction, e.g. P_M
P	Projective Reconstruction, e.g. P_P
i	Refers to Camera i ($i = 1, 2$) in camera pair, e.g. P_{P1}
GT	Ground Truth
superscripts	
i	Refers to solution i ($i = 1, 2$) for the second camera, e.g. $\mathbf{x^1}$
Accents, as in $\mathbf{p}'$ and $\mathbf{p}$	Discriminate between the two solutions for camera 2
P matrix representations	
$[KR \quad -KRC]$	Metric reconstruction
$[P_i \quad \mathbf{a}]$	Metric reconstruction
$[KR \quad \mathbf{a}]$	Metric reconstruction
$\mathbf{m_i}$	i-row vector of left 3×3 P matrix block
$[[\mathbf{a}]_x F \quad \mathbf{a}]$ with $\mathbf{a} : F^T\mathbf{a} = \mathbf{0}$	Projective Reconstruction in canonical form
Simplifications	
$[\mathbf{a}]_x \triangleq \begin{bmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{bmatrix}$ with $\mathbf{a} = \begin{pmatrix} a_1 & a_2 & a_3 \end{pmatrix}^T$	$[\mathbf{a}]_x$ satisfies $\mathbf{a} \times \mathbf{b} = [\mathbf{a}]_x\mathbf{b}$

2.2. Focal length estimation in the general setting

Different approaches have been followed to estimate focal lengths in related mvg problems. In Perspective-n-Points (PnP) problem, Grobner basis methods have been recently employed (Wu, 2015; Jiang et al., 2014). A new camera parametrization has been introduced to solve the minimal 3.5-correspondences PnP leading to ten solutions (Wu, 2015). Steering to a different direction, a sampling scheme to randomly sample promising f values and then solve camera pose with known focal length has been developed (Sattler et al., 2014). Finally, concerning a moving camera with constant parameters, branch and prune paradigm has been applied to estimate Dual Absolute Quadric, DIAC (Paudel and Van Gool, 2018; Habed et al., 2014).

2.3. Approaches to multiple view reconstruction

In a reconstruction pipeline, initially Structure from Motion (SfM) is solved to get $P, \mathbf{X}$, assuming image point correspondences and self-calibrated cameras. The fundamental method to solve SfM is Bundle Adjustment (BA) (Lourakis and Argyros, 2009),[3] an iterative, numerical algorithm to minimize the reprojection error of the recovered solution.

In standard approaches to SfM a sequence of SfM sub-problems are solved (sequential SfM) (Snavely et al., 2010, 2006; Wu et al., 2011). In each iteration, more, possibly uncalibrated, cameras and world points are added to the SfM problem which is solved using BA. Such methods are sensitive to the initial camera pair selection, solve a large number of optimization problems numerically and optimize an objective function with possibly multiple local minima. An optimized sequential SfM pipeline that addresses robustness, accuracy, reconstruction completeness and efficiency has been made publicly available (Schonberger and Frahm, 2016).

A different approach has been developed to solve the SfM-with-known-rotations problem within the framework of optimal algorithms in multiple-view geometry (mvg) and L_∞ mvg algorithms (Dalalyan and Keriven, 2009; Hartley and Kahl, 2007; Kahl and Hartley, 2008; Olsson and Enqvist, 2011; Olsson and Kahl, 2010; Zach and Pollefeys, 2010). In this formulation, the camera rotation matrices R_i are initially determined or are assumed known. Then SfM-with-known-rotations is formulated as a convex-optimization problem, for which a unique global minimum exists. Rotation averaging has become a key sub-problem in SfM and has been studied in depth both theoretically, to identify when rotation averaging becomes hard (Wilson et al., 2016), and in practice by benchmarking methods with respect to performance and robustness to outliers (Tron et al., 2016).

For the actual solution of SfM-with-known-rotations, either a sequence of Second-Order Cone Programs (SOCP) are solved to arrive at an exact solution, or approximate solutions are recovered by solving SOCP or linear programs (Martinec and Pajdla, 2007; Enqvist et al., 2011; Olsson and Enqvist, 2011; Sinha et al., 2010). BA may still be applied as a last fine-tuning of the solution.

A SfM solution, allows the reconstruction of a low number of 3D points (sparse point cloud), limited by the number of image points correspondences. *Multi-view stereo* (mvs) algorithms can be used at this point to produce a *dense point cloud*, which contains a much larger number of 3D points (Furukawa and Ponce, 2010). Finally, surface reconstruction algorithms can be used to produce a 3D surface (Kazhdan et al., 2006).

3. A method for metric reconstruction in pairs of uncalibrated images

In this section we examine a camera pair i, j described by a projective reconstruction and recover the pair's metric reconstruction: $f_i, f_j, R_{ij}, \mathbf{C_i}, \mathbf{C_j}$. We formulate the problem's equations and then propose a linear solution. The spatial arrangement of the two recovered solutions is described in two theorems (proofs are provided in Appendix). The solutions are disambiguated using the Cheirality condition. To simplify the exposition of our method's derivation we disregard the constants involved in projective identities. We show in Section 3.4 that, nevertheless, the correctness of our method is preserved.

3.1. Formulation of system equations

Let us consider two pinhole cameras P_1, P_2 and further-to simplify the equations and matrices — that P_1 coordinate system is aligned with the world coordinate system ($\mathbf{C_1} = \mathbf{0}$, $z = f$ is the image plane in a xyz Cartesian coordinate system). Let us further assume, that the corresponding image coordinate systems are selected so that the internal parameters of each camera K_i can be written as

$$K_i = \begin{bmatrix} f_i & 0 & 0 \\ 0 & f_i & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{8}$$

[3] Implementation available at http://users.ics.forth.gr/~lourakis/sba/index.html.

where f_i is the focal length. We assumed square camera pixels and zero skew. Principal point is set to zero assuming that it lies at the image center and setting the origin of image coordinates at the principal point. We disregard nonlinear lens distortions, e.g. as in fisheye lenses, that diverge from the pinhole camera model. The previous assumptions are routinely employed in multiple view geometry and are thoroughly discussed in the literature (Hartley and Zisserman, 2004).

We start from a projective reconstruction of the two cameras, given by P_{P1}, P_{P2}, which is related to the metric reconstruction by a world (3D) homography H as in

$$\begin{aligned} P_{M1} &= P_{P1} H \\ P_{M2} &= P_{P2} H \end{aligned} \tag{9}$$

Using Result 1, Eq. (1), we project Q^*_∞ to the image plane of camera 2. For this projection, ω^*_2 we have:

$$\begin{bmatrix} f_2^2 & 0 & 0 \\ 0 & f_2^2 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \omega^*_2 = P_{P2} H Q^*_\infty H^T P_{P2}^T \tag{10}$$

To introduce the unknowns in Eq. (10), we use the canonical representation of the projective reconstruction, so that $P_{P1} = \begin{bmatrix} I & \mathbf{0} \end{bmatrix}$. From Eq. (9), we have for the homography

$$H = \begin{bmatrix} K_1 & \mathbf{0} \\ \mathbf{v^T} & \sigma \end{bmatrix} \tag{11}$$

where $\mathbf{v}$ is yet undetermined and the scale factor σ can be ignored ($\sigma = 1$).

To fully determine H, we turn to the plane at infinity

$$\boldsymbol{\pi}_{\infty,\mathbf{P}} \triangleq \begin{pmatrix} \mathbf{p^T} & 1 \end{pmatrix}^T \triangleq \begin{pmatrix} p_1 & p_2 & p_3 & 1 \end{pmatrix}^T \tag{12}$$

Using Result 2 we arrive at

$$H = \begin{bmatrix} K_1 & \mathbf{0} \\ -\mathbf{p^T} K_1 & 1 \end{bmatrix} \tag{13}$$

Substituting H from Eq. (13) to Eq. (10) we get

$$\omega^*_2 = P_{P2} \begin{bmatrix} K_1 K_1^T & -K_1 K_1^T \mathbf{p} \\ -\mathbf{p^T} K_1 K_1^T & \mathbf{p^T} K_1 K_1^T \mathbf{p} \end{bmatrix} P_{P2}^T \tag{14}$$

Eq. (14) comprise a non-linear system with respect to the five unknowns (plane at infinity coordinates and focal lengths) we want to determine to acquire a metric reconstruction of the scene. We note that ω^*_∞ is symmetric by definition, and is also homogeneous, thus it provides five independent equations.

3.2. Linearization

In Eq. (14), we substitute

$$P_{P2} \triangleq \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix} \tag{15}$$

We define an indeterminate vector $\mathbf{x_o}$ using polynomials in f_1, f_2, p_1, p_2, p_3

$$\mathbf{x_o} \triangleq \begin{pmatrix} f_1^2 \\ f_2^2 \\ f_1^2 p_1^2 + f_1^2 p_2^2 + p_3^2 \\ p_3 \\ f_1^2 p_1 \\ f_1^2 p_2 \end{pmatrix} \tag{16}$$

The augmented matrix $[A_s|\mathbf{b}]$ for the linear system

$$A_s \mathbf{x_o} = \mathbf{b} \tag{17}$$

is then given by Eq. (18) in Box I.

We derived the above equations (in order of appearance) from elements $\omega^*_2(2,2)$, $\omega^*_2(2,3)$, $\omega^*_2(1,3)$, $\omega^*_2(1,1)$, $\omega^*_2(1,2)$, $\omega^*_2(3,3)$ of ω^*_2. In the following, we use the first five equations as explained in Section 3.4.

The matrix of Eq. (18) is rank deficient. Thus, we presented a linear system of five (in the best case) linearly-independent equations, in six unknowns. To solve it, we turn to the polynomial relations between the coordinates of $\mathbf{x_o}$.

3.3. Recovering the solutions

Taking five of Eqs. (17) (see Section 3.4 about which equations to take) we have the linear system

$$A_5 \mathbf{x_o} = \mathbf{b_5} \tag{19}$$

Applying Gaussian elimination to (19), we bring the augmented matrix to the form

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & \mathbf{0} & b_1 \\ 0 & 1 & 0 & 0 & 0 & \mathbf{0} & b_2 \\ 0 & 0 & 1 & 0 & 0 & 0 & b_3 \\ 0 & 0 & 0 & 1 & 0 & c & b_4 \\ 0 & 0 & 0 & 0 & 1 & d & b_5 \end{bmatrix} \tag{20}$$

where:

1. The elements in default font, are in the usual form expected when we apply Gaussian elimination in the general case
2. The elements in bold font, are a result of the problem's structure, that is of the special relations in Eq. (19)
3. Finally, the element in slanted font (third row, sixth column), is as given when we use the canonical representation for the projective reconstruction, which is:

$$P_{P1} = \begin{bmatrix} I & \mathbf{0} \end{bmatrix}, P_{P2} = \begin{bmatrix} [\mathbf{a}]_x F & \mathbf{a} \end{bmatrix} \tag{21}$$

Where $\mathbf{a}$ is the left null vector of F: $F^T \mathbf{a} = \mathbf{0}$. By using the canonical pair, the leftmost 3×3 block in P_{P2} is rank 2, and consequently has linearly-dependent row-vectors

The derivation of Eq. (20) is given in Appendix.

To solve for the focal lengths (f_1, f_2) and $\boldsymbol{\pi}_\infty$ (p_1, p_2, p_3), we now have from (20)

$$f_1^2 = b_1 \tag{22}$$

$$f_2^2 = b_2 \tag{23}$$

$$f_1^2 p_1^2 + f_1^2 p_2^2 + p_3^2 = b_3 \tag{24}$$

$$p_3 + c f_1^2 p_2 = b_4 \tag{25}$$

$$f_1^2 p_1 + d f_1^2 p_1 = b_5 \tag{26}$$

We substitute p_1, from (26), and p_3, from (25), to Eq. (24) and obtain a quadratic equation in p_2. Thus, we determine f_1, f_2 uniquely and p_1, p_2, p_3 with a two-way ambiguity. We refer to those two solutions as

$$\begin{aligned} \mathbf{x_o^1} &= \begin{pmatrix} b_1 & b_2 & b_3 & p_3 & f_1^2 p_1 & f_1^2 p_2 \end{pmatrix}^T \\ \mathbf{x_o^2} &= \begin{pmatrix} b_1 & b_2 & b_3 & p_3' & f_1^2 p_1' & f_1^2 p_2' \end{pmatrix}^T \end{aligned} \tag{27}$$

3.4. The effect of homogeneous representation on the derived equations

In this section we investigate the effect of projective identities on the metric reconstruction method we introduced, specifically on the formulation of Eq. (17). We show that the consideration of the proper constants involved in Eq. (14) does not cancel the linearity of Eq. (17), owing to the assumed diagonal K matrices and provided we choose specific ω^*_2 elements to write Eq. (17).

Let

$$\{P'_{GT1}, P'_{GT2}\} \triangleq \{\begin{bmatrix} I & \mathbf{0} \end{bmatrix}, \begin{bmatrix} A & \mathbf{a} \end{bmatrix}\} \tag{28}$$

be the ground truth camera matrices we aim to recover, and

$$\{P_{P1}, P_{P2}\} \triangleq \{\begin{bmatrix} I & \mathbf{0} \end{bmatrix}, \begin{bmatrix} \hat{A} & \hat{\mathbf{a}} \end{bmatrix}\} \tag{29}$$

$$\begin{pmatrix} p_{21}^2+p_{22}^2 & -1 & p_{24}^2 & -2\,p_{23}p_{24} & -2\,p_{21}p_{24} & -2\,p_{22}p_{24} & -p_{23}^2 \\ p_{21}p_{31}+p_{22}p_{32} & 0 & p_{24}p_{34} & -p_{23}p_{34}-p_{24}p_{33} & -p_{21}p_{34}-p_{24}p_{31} & -p_{22}p_{34}-p_{24}p_{32} & -p_{23}p_{33} \\ p_{11}p_{31}+p_{12}p_{32} & 0 & p_{14}p_{34} & -p_{13}p_{34}-p_{14}p_{33} & -p_{11}p_{34}-p_{14}p_{31} & -p_{12}p_{34}-p_{14}p_{32} & -p_{13}p_{33} \\ p_{11}^2+p_{12}^2 & -1 & p_{14}^2 & -2\,p_{13}p_{14} & -2\,p_{11}p_{14} & -2\,p_{12}p_{14} & -p_{13}^2 \\ p_{11}p_{21}+p_{12}p_{22} & 0 & p_{14}p_{24} & -p_{13}p_{24}-p_{14}p_{23} & -p_{11}p_{24}-p_{14}p_{21} & -p_{12}p_{24}-p_{14}p_{22} & -p_{13}p_{23} \\ p_{31}^2+p_{32}^2 & 0 & p_{34}^2 & -2\,p_{33}p_{34} & -2\,p_{31}p_{34} & -2\,p_{32}p_{34} & -p_{33}^2+1 \end{pmatrix} \tag{18}$$

Box I.

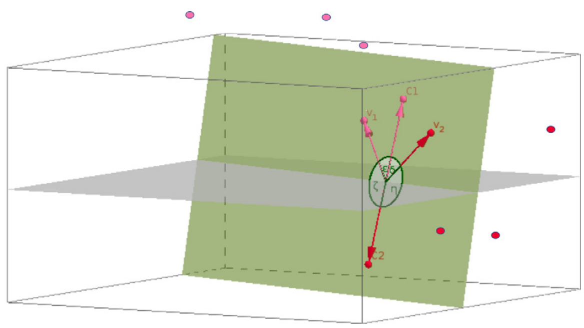

Fig. 2. We visualize the geometric relations of Theorems 1,2. In the graph, we display the centers of projection ($\mathbf{C^1}, \mathbf{C^2}$) and viewing directions ($\mathbf{v^1}, \mathbf{v^2}$), for each of the two solutions of Eq. (27). In pink, we display solution 1 and in red solution 2. The common plane of $\mathbf{C^1}, \mathbf{C^2}, \mathbf{v^1}, \mathbf{v^2}$ is highlighted. The angles satisfy $\epsilon = \delta$, $\zeta = \eta$. Pink and red world-points demonstrate points in front of the respective cameras-solutions. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

be the starting projective reconstruction.

P'_{GT1} is related to the recovered solution P_{GT1} (applying the method we introduced) by the homography

$$H_k = \begin{bmatrix} K_1^{-1} & \mathbf{0} \\ \mathbf{0^T} & 1 \end{bmatrix} \tag{30}$$

To recap, we get from P_P to P_{GT} by homography H, from P_P to P'_{GT} by H' and from P_{GT} to P'_{GT} by H_k.

For the camera pairs (28), (29), we have the respective Fundamental matrices

$$F'_{GT} = \left[[\hat{\mathbf{a}}]_x \hat{A}\right], \text{ for (28)} \tag{31}$$

$$F_P = \left[[\mathbf{a}]_x A\right], \text{ for (29)} \tag{32}$$

From Result 3, since P'_{GT}, P_P reconstructions are related by H', the reconstructions share a common fundamental matrix. Since fundamental matrices are homogeneous entities, we have

$$F_P = \epsilon F'_{GT} \tag{33}$$

Now, we turn to Result 4 and get

$$\hat{\mathbf{a}} = \kappa \mathbf{a} \tag{34}$$

$$\hat{A} = \epsilon^{-1}\kappa^{-1}(A + \mathbf{a}\mathbf{v^T}) \tag{35}$$

We write the previous equations in matrix form to get the projective transformation H'

$$H' \triangleq \begin{bmatrix} \kappa^{-1}\epsilon^{-1} I & \mathbf{0} \\ \kappa^{-1}\epsilon^{-1}\mathbf{v^T} & \kappa \end{bmatrix} \tag{36}$$

H' satisfies

$$\begin{aligned} \kappa^{-1}\epsilon^{-1} P'_{GT1} &= P_{P1} H' \\ P'_{GT2} &= P_{P1} H' \end{aligned} \tag{37}$$

Now, we get H from H', H_k^{-1}

$$\begin{bmatrix} \kappa^{-1}\epsilon^{-1} K_1 & \mathbf{0} \\ \kappa^{-1}\epsilon^{-1}\mathbf{v^T} K_1 & \kappa \end{bmatrix} \tag{38}$$

We set the bottom-right element to 1, as we disregard the true scale of the reconstruction, and get the final form of H

$$H = \begin{bmatrix} \kappa^{-1}\epsilon^{-1} K_1 & \mathbf{0} \\ \kappa^{-1}\epsilon^{-1}\mathbf{v^T} K_1 & 1 \end{bmatrix} \tag{39}$$

One should compare the homographies of Eqs. (13) and (39). Using Eq. (13) instead of Eq. (39) in Eq. (9), we get from P_P to

$$\begin{aligned} P_{M1} &= \begin{bmatrix} K_1 & \mathbf{0} \end{bmatrix} \\ P_{M2} &= \begin{bmatrix} \mu K_2 R_2 & \mathbf{a} \end{bmatrix} \end{aligned} \tag{40}$$

We observe that the translation direction $\mathbf{a}$ is correct but the left-most 3×3 block of camera 2 is multiplied by a constant μ.

To see how the constants in Eq. (39) affect Eq. (17), we substitute Eq. (39) in Eq. (10) and get for ω_2^* the expression

$$\omega_2^* = P_{P2} \begin{bmatrix} (\kappa\epsilon)^{-2} K_1 K_1^T & -(\kappa\epsilon)^{-2} K_1 K_1^T \mathbf{p} \\ -(\kappa\epsilon)^{-2} \mathbf{p^T} K_1 K_1^T & (\kappa\epsilon)^{-2} \mathbf{p^T} K_1 K_1^T \mathbf{p} \end{bmatrix} P_{P2}^T \tag{41}$$

We encourage to compare the corrected ω_2^* equation-Eq. (41)- to the simplified ω_2^* equation-Eq. (14).

To avoid the determination of additional unknowns in Eq. (17), we have:

- All equations derived from ω_2^* elements off the diagonal are of the form $\mathbf{a}\mathbf{x_0} = \mathbf{0}$, thus the constant $(\kappa\epsilon)^{-1}$ can be eliminated
- The equation derived from element $\omega_2^*(3,3)$ cannot be used without determining additional constants. So, excluding $\omega_2^*(3,3)$ we may only use the other five of the six original equations of Eq. (17)
- Using equations involving f_2 ($\omega_2^*(1,1), \omega_2^*(2,2)$), we can only determine f_2 up to a multiplicative constant. So we may redefine $\mathbf{x_0}$ (Eq. (16)) and substitute f_2^2 with cf_2^2 (c a properly defined constant)

3.5. The final self-calibration and metric reconstruction method

The complete method to solve the metric reconstruction and self calibration problem follows:

1. We solve the system (17), keeping five equations and discarding the equation derived from $\omega_2^*(3,3)$.
2. In the previous step (1), we have recovered cf_2^2 (c a constant). To fully determine f_2, many different approaches are possible. We propose to repeat step 1, putting camera 2 at the origin of the coordinate system (in place of camera 1). This can be done by transposing the fundamental matrix F for the camera pair. Following this approach, we may additionally determine the constant $\kappa\epsilon$ in Eq. (41).
3. Using the homography of Eq. (13) or Eq. (39), we recover the metric reconstruction P_{M1}, P_{M2} from Eq. (9). Depending on which homography we have used, one camera matrix (P_{M2} for Eq. (13) or P_{M1} for Eq. (39)) will have the left-most 3×3 block multiplied by a constant, as in Eq. (40). This has no effect on the correctness of the representation, and the image points are the same in each case.
4. The P_{Mi} matrices are factored as $P_{Mi} = \begin{bmatrix} K_i R_{ij} & \mathbf{t_{ij}} \end{bmatrix}$ to recover $R_{ij}, \mathbf{t_{ij}}$.

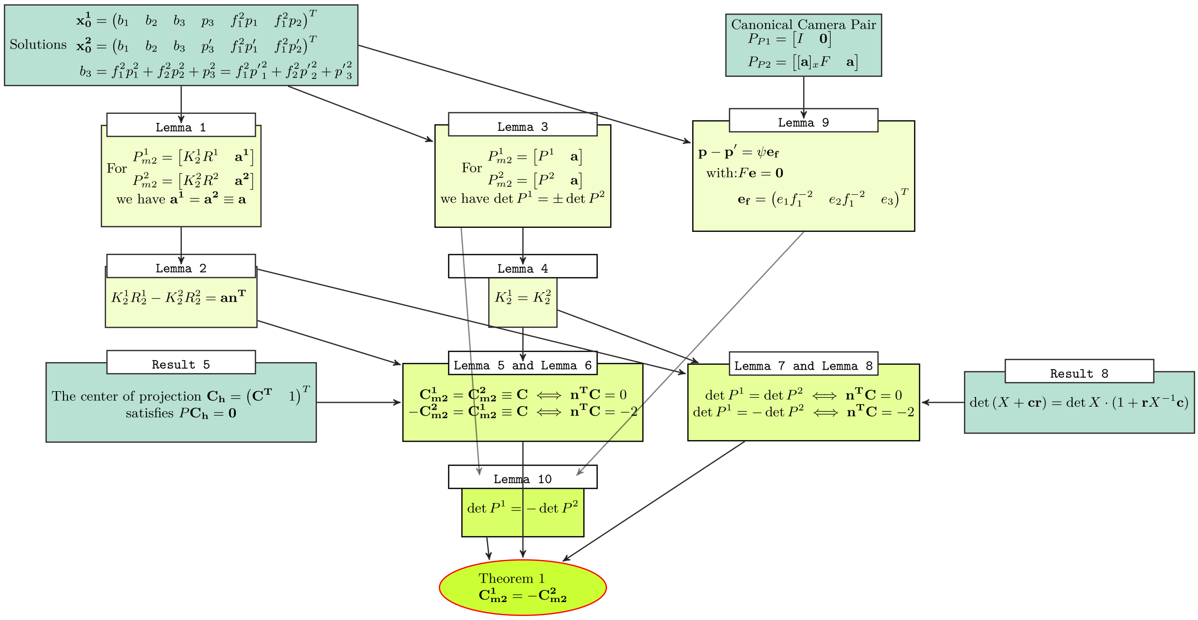

Fig. 3. An illustration of the intermediate results leading to Theorem 1. An arrow is drawn whenever the equation at its tail is used to prove the equation at its tip. The full proof is given in Appendix.

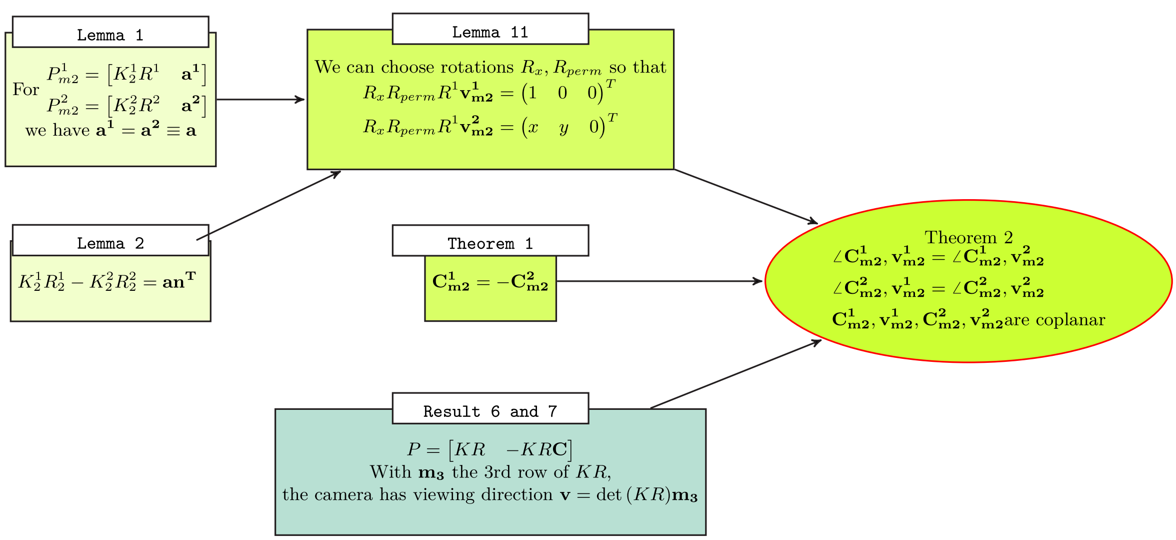

Fig. 4. An illustration of the intermediate results leading to Theorem 2. An arrow is drawn whenever the equation at its tail is used to prove the equation at its tip. The full proof is given in Appendix.

3.6. Solution disambiguation and geometric relations of the two solutions

We use the Cheirality condition to determine the valid solution of Eq. (17): The correct one of solutions (27) can be identified by requiring all world points that are visible from camera 2 to be in the space in front of camera 2 (see Fig. 2).

Whenever the two recovered solutions represent cameras with divergent viewing directions, Cheirality condition is more likely to identify the valid solution. We explore in Theorems 1 and 2 the geometric relations between the two solutions, aiming to vizualize solutions' relations and disambiguation. Proofs of Theorems 1 and 2 are outlined in Figs. 3 and 4. Full proofs are given in Appendix.

Theorem 1. *Let*

$$\{P_{m1}^1, P_{m2}^1\}, \{P_{m1}^2, P_{m2}^2\} \tag{42}$$

denote the reconstructions derived from Eq. (27). Then, cameras P_{m2}^1, P_{m2}^2 are in mirror positions with respect to the origin (position of P_{m1}^1, P_{m1}^2). The centers of projection $\mathbf{C^1_{m2}}, \mathbf{C^2_{m2}}$ satisfy

$$\mathbf{C^1_{m2}} = -\mathbf{C^2_{m2}} \tag{43}$$

Theorem 2. *Let camera 1 be positioned on the origin of the world coordinate system, with a viewing direction aligned to z axis. We denote $\mathbf{v^1_{m2}}, \mathbf{v^2_{m2}}$ the viewing directions of P_{m2}^1, P_{m2}^2 and $\mathbf{C^1_{m2}}, \mathbf{C^2_{m2}}$ the position vectors of the corresponding centers of projection. Then, $\mathbf{C^1_{m2}}, \mathbf{C^2_{m2}}$ bisect the angles formed by $\mathbf{v^1_{m2}}, \mathbf{v^2_{m2}}$, in the plane defined by $\mathbf{v^1_{m2}}, \mathbf{v^2_{m2}}$. Thus, we have:*

$$\angle \mathbf{C^1_{m2}}, \mathbf{v^1_{m2}} = \angle \mathbf{C^1_{m2}}, \mathbf{v^2_{m2}} \tag{44}$$

$$\angle \mathbf{C^2_{m2}}, \mathbf{v^1_{m2}} = \angle \mathbf{C^2_{m2}}, \mathbf{v^2_{m2}} \tag{45}$$

From Theorems 1,2, we easily deduce that

$$\angle \mathbf{C^1_{m2}}, \mathbf{v^j_{m2}} + \angle \mathbf{C^2_{m2}}, \mathbf{v^j_{m2}} = 180° \tag{46}$$

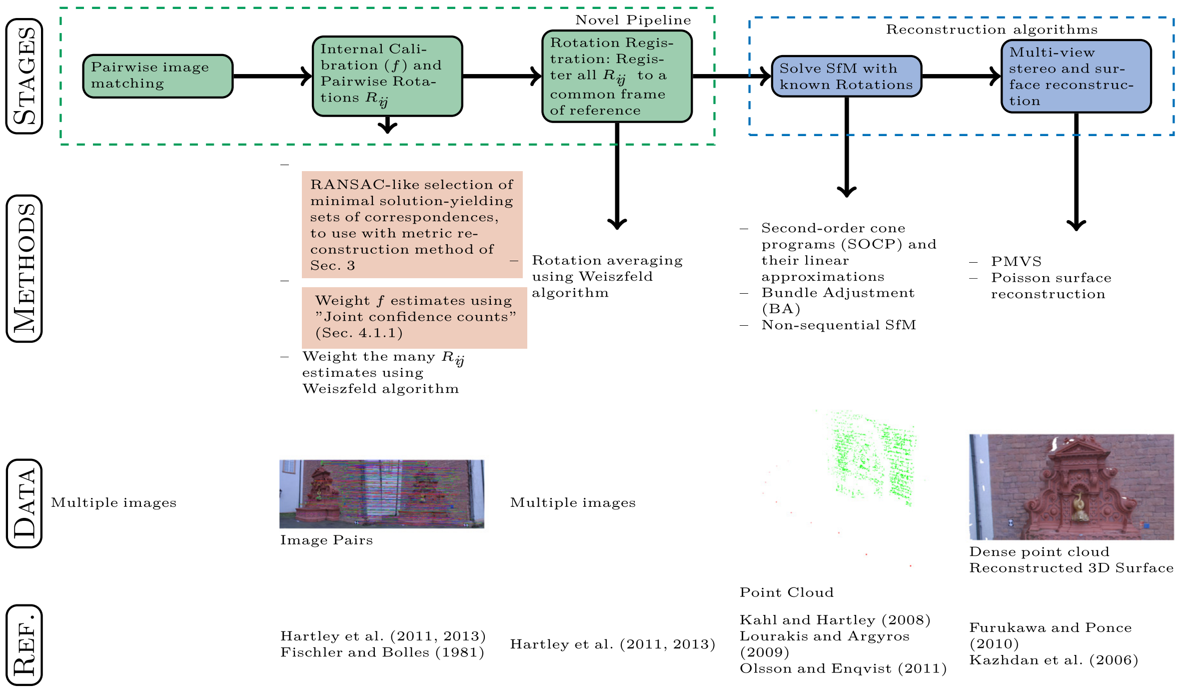

Fig. 5. Our pipeline to reconstruct a 3D scene from an unordered set of 2D photographs. In the first row, we display a flow diagram of the algorithm **stages**. Novel parts are displayed in green. The second row outlines the core **methods** we use. We highlight methods we introduced in this paper. The third row contains a visualization of **data type**. In the last row we list the most important **references** per stage.

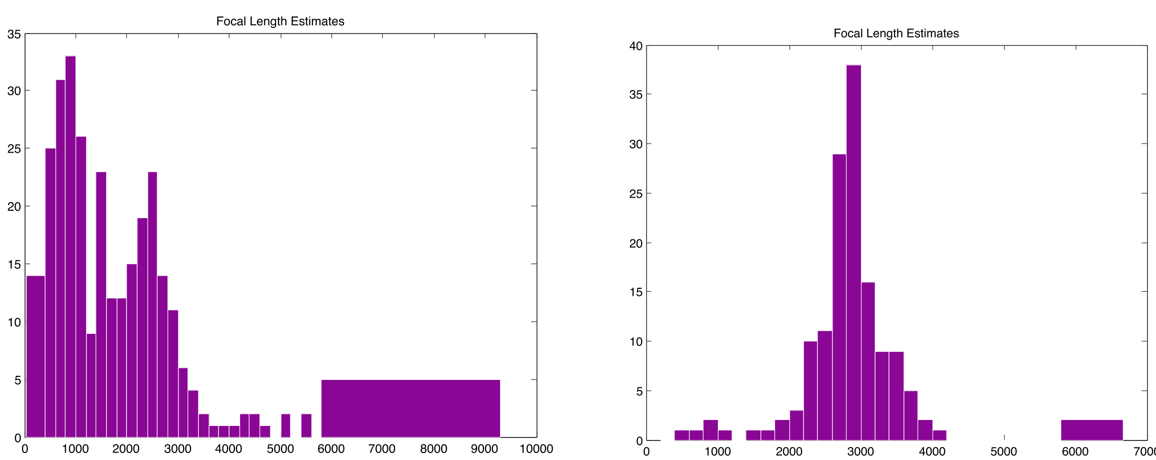

Fig. 6. Distribution of f_i estimates for camera i, obtained by camera pairs' reconstructions. f estimates were collected from all available camera pairs. We observe that for some cameras (right), focal length can be readily determined. The opposite holds for other cameras (left).
Source: Data from castle-P30 (Strecha et al., 2008a).

The larger $\angle \mathbf{v}^1_{\mathbf{m2}}, \mathbf{v}^2_{\mathbf{m2}}$ is, the easier it would be to disambiguate the solutions by the Cheirality condition. From Theorem 2, we have that $\angle \mathbf{v}^1_{\mathbf{m2}}, \mathbf{v}^2_{\mathbf{m2}} = 2\angle \mathbf{v}^1_{\mathbf{m2}}, \mathbf{C}^1_{\mathbf{m2}}$ and that $\angle \mathbf{v}^1_{\mathbf{m2}}, \mathbf{C}^1_{\mathbf{m2}}$ decreases when the translation of camera 2 (with respect to camera 1) is aligned with camera 2 viewing direction (Eq. (112)).

4. An application to the multiple-view reconstruction problem

We integrate our method for the pair-based estimation of R, f in existing pipelines to solve the multiple-view reconstruction problem and produce the 3D-model of a scene.

Our approach is outlined in Fig. 5. We start by pairwise matching of SIFT features. To reduce the outliers, we validate the initial correspondences using a custom verification method which we have successfully tested (i.e. high precision) on small sets of photographs of architectural scenes (buildings). Pairwise metric reconstructions (Section 4.1) are acquired by the methods of Section 3. Focal lengths estimates are averaged (Section 4.1.1). R_{ij} estimates are also averaged and then registered in a global coordinate system (Section 4.1.2). Then final reconstruction is done using the non-sequential SfM-with-known-rotations formulation of Olsson and Enqvist (2011), which we have modified extensively, using the methods of the preceding sections as well as the f, R averaging algorithms (Section 4.1).

4.1. Averaging pair-based solutions for f, R

In this paper we introduced computationally efficient methods for R_{ij}, f_i estimation, which we apply in randomly sampled minimal correspondences sets, in a way that resembles RANSAC procedures (Fischler and Bolles, 1981). Each sample yields a f_i, R_{ij} solution which is validated by the Cheirality condition. The multiple f_i, R_{ij} estimates, one

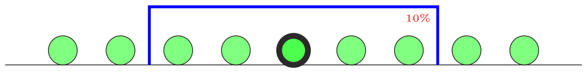

Fig. 7. A demonstration of confidence count computation. Each disk represents a f_i estimate. We compute the cc for the central value, depicted here with a bold border. This cc depends on the number of estimates within a $\beta = 10\%$ range, depicted with a blue rectangle in the picture.

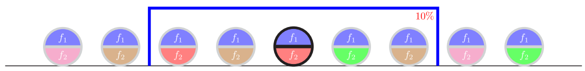

Fig. 8. Joint confidence count computation for a f_1 estimate. Each disk represents a f_1 estimate. We compute the Jcc for the central value, depicted here with a bold border. This time, each disk is divided in half, to demonstrate that each f_1 estimate is paired with one f_2 estimate, each f_2 estimate originating from a different $1, j$ pair. We use different disk colors for different cameras. Jcc depends on the sum of elements within range (inside the blue rectangle). In contrast with cc computation, each element contributes a different amount to the sum, depending on cc of f_2.

from each minimal sample, are then averaged, to produce the final solutions.

The Cheirality condition is applied to disambiguate the solutions. In practical situations erroneous matches corrupt the data. We reject a solution if there exists a point in the sampled minimal set that is behind a camera in both of the two recovered solutions.

In f_i, case, we introduce a novel averaging method. In the case of pairwise rotations R_{ij}, we apply the Weiszfeld algorithm (Hartley et al., 2011, 2013), which converges to the median (L_1-average) rotation. We also use a form of the Weiszfeld algorithm (multiple rotation averaging) in the rotation registration problem to get the final camera rotation matrices R_i (Section 4.1.2).

4.1.1. Focal length estimates

The distribution of f_i estimates for camera i collected from all the reconstructed image pairs i, j can be skewed or multimodal (Fig. 6), in which case the mean or median estimate will not correctly determine f_i value.

We introduce new measures to evaluate the fit of focal length estimates. We initially introduce the Confidence count (cc) and then modify cc using the problem structure to introduce the Joint confidence count (Jcc). We assume that in each image pair that contains image i, we receive a number of correct and a number of erroneous estimates for f_i, and that erroneous estimates originating from different i, j image pairs vary significantly in value, whereas correct ones aggregate.

We visualize cc computation in Fig. 7. Simplifying aspects of the computation, we can describe it as a binning procedure, where the bin range is adapted to contain all estimates within $\beta\%$ deviation:

1. We collect all f_i^n estimates of f_i, originating from all the different images we have matched with image i.
2. For each f_i^n, we count the number of estimates, f_i^k, within a $\beta\%$ error range. This sum is the confidence count cc_i^n for estimate f_i^n.
3. We normalize cc_i^n values to $0 \dots 1$ range. This step is critical for Jcc computation.

We estimate f_i by the f_i^n value with maximum cc_i^n.

To further improve the f estimation, we introduce Jcc (Fig. 8). Since each estimate f_i^n is paired with some estimate f_j^n (the estimates were computed in an image pair), we expect that if f_i^n is a good estimate then f_j^n will be accurate too. To compute Jcc, we follow a similar to cc procedure, but this time each estimate f_i^k in $\beta\%$ range contributes a different amount to Jcc sum. This amount is proportional to cc_j^k of estimate f_j^k that is paired with f_i^k. Good f_j^k estimates have higher confidence counts, and contribute more to Jcc.

In greater detail, to compute the Jcc_i^n of estimate f_i^n about image i, we have:

1. Let $k = 1 \dots m$ be the m images we matched with image i. For each image k we have:
   - From all estimates within $\beta\%$ range of f_i^n, we pick the L ones that originate from pair i, k
   - Since every f_i estimate originating from i, k pair is matched to an f_k estimate, from the L estimates of the previous step we get the corresponding L estimates of f_k
   - For each of the L estimates of f_k, we have a confidence count cc_k^n. We get their mean. We do not use the direct sum, to diminish the influence of a large sum (large L) of low cc's
2. Jcc_i^n is the sum of the previous m mean values.

We estimate f_i by the f_i^n value with maximum Jcc_i^n.

In all our experiments we set $\beta = 10\%$, as a reasonable error range around f estimates. In synthetic scenes (data not shown) we have found that our focal length estimates were repeatedly within a 10% error margin, in the majority ($\approx 80\%$) of the cases for reasonable noise levels corrupting the correspondences (Gaussian noise with standard deviation equal to 1% image size).

4.1.2. Rotation estimates

In this section, we summarize rotation averaging using the Weiszfeld algorithm (Hartley et al., 2011, 2013). Weiszfeld algorithm returns the L_1-mean in a set of points in space $\mathcal{R}^n$. Many different metrics have been defined for rotation matrices (Hartley et al., 2013). We limit our analysis here to

$$d_{geometric}(R, S) \triangleq \text{angle of rotation } RS^{-1} \tag{47}$$

Weiszfeld algorithm is a gradient-descent method and is guaranteed to converge to the true L_1-mean in the case of single rotation averaging, as averaging of pairwise rotation estimates R_{ij}.

The L_1-mean of R_i estimates of a single rotation is the rotation R_y that minimizes:

$$\sum_{i=1}^{n} d_{geometric}(R_i, R_y) \tag{48}$$

In the case of rotation registration, the convergence of Weiszfeld algorithm is not guaranteed.

We applied Weiszfeld algorithm to weight the estimates R_{ij} of the pairwise rotations we acquired through random sampling of minimal point sets (8 points) yielding a R_{ij} solution.

In the rotation registration problem we applied the Weiszfeld algorithm in the following manner:

1. We construct the rotations graph which has a node for every image and an edge e_{ij} between nodes i, j if we have the relative rotation R_{ij} between the respective images. We take a spanning tree in this graph, and using $R_j = R_{ij} R_i$ we get the initial R_j estimates.
2. For every node i in the graph, we use all available estimates R_{ij} to get inconsistent estimates R_i^k, $k = 1, 2, \dots$ through $R_i = R_{ji} R_j$. We average estimates R_i^k with one iteration of Weiszfeld algorithm.
3. We repeat the previous step n times ($n = 20$).

5. Results & discussion

5.1. Metric reconstruction in pairs of images

Our method for self-calibration in image pairs may be used instead of the established Kruppa equations approach, as both methods are based on ω_∞^* and provide the exact same focal length estimates. However our formulation in Eqs. (18) allows to additionally recover a metric reconstruction.

Table 2
Median errors in focal length (f), relative rotation (R) and relative translation ($\mathbf{t}$) in pairwise reconstructions obtained by our method. Synthetic scenes were created, corrupted by Gaussian noise. Noise level σ is given as a percentage of image size. We created two thousand scenes per noise level.

Noise level σ (%)	Median Δf	ΔR (°)	$\Delta \mathbf{t}$ (°)
0.2	0.0069	0.04	0.53
0.4	0.0179	0.08	1.17
0.6	0.0200	0.12	1.65
0.8	0.0360	0.16	2.42
1	0.0412	0.21	2.98

We compare the two methods in noise-corrupted synthetic scenes. We put one camera at the origin of the coordinate system, oriented towards z-axis and sample the second camera position uniformly on the unit sphere. To orient the second camera we sample rotation axis coordinates from a uniform distribution in range $[0, 1]$ and randomly set a rotation angle less than 30° (Pernek and Hajder, 2013). World (3-D) points are uniformly sampled and all points not visible by either one of the cameras are filtered out in the final synthetic scene. Each camera has an image diagonal of 1 and a focal length uniformly sampled in $[0.5, 1.5]$. We added Gaussian noise to the image points positions, and not to world points or other entities, to simulate actual noisy correspondences. The noise standard deviation ranges from 0 to 1% of image size (Chandraker et al., 2007a,b; Gherardi and Fusiello, 2010; Kukelova et al., 2008).

To quantify the error in f estimation we use Δf (Chandraker et al., 2007a; Gherardi and Fusiello, 2010; Kukelova et al., 2008):

$$\Delta f \triangleq \left| \frac{\hat{f}}{f} - 1 \right|, \quad \text{where } \hat{f} \text{ is an estimate of } f \tag{49}$$

To quantify the reconstruction error, we use (i)the angle (ΔR) between the relative rotation estimate and the true relative rotation R_{ij} (between two paired views i, j) and (ii)the angle ($\Delta\mathbf{t}$) between the translation estimate and the true translation $\mathbf{t}$.

We observed that our method (Section 3) and Kruppa equations produce identical f_i estimates. In rare cases with extremely noise-corrupted correspondences, our method failed — i.e. provided $f_i^2 < 0$ solutions, and the Kruppa method produced largely erroneous focal length estimates, with $\Delta f > 1.5$ on average.

In Table 2 we provide $\Delta f, \Delta R, \Delta\mathbf{t}$ errors in synthetic scenes, for noise levels with standard deviation up to 1% of image size. Our method has low $\Delta f, \Delta R$ errors but larger $\Delta\mathbf{t}$ error. In our multi-view reconstruction pipeline (Fig. 5) we improve the f, R estimates we obtain from camera-pair reconstructions through averaging methods and re-estimate from scratch $\mathbf{t}$ in the subsequent SfM-with-known-rotations stage.

Next we focus on relative rotation R_{ij} evaluation, since R_{ij} estimates are used in our reconstruction pipeline (Section 4) and thus the accuracy of R_{ij} estimation is important. We compare our approach to the 5-Point (5P) algorithm for relative pose estimation (Nistér, 2004). In order to fairly compare the two approaches, we used the same noisy point correspondences as input, which were obtained through SIFT feature matching. Also, the same internal camera parameters were used in both approaches and were provided by the method described in Section 3. The dataset used was the multi-view dataset of Strecha et al. (2008b). The two methods we compared were:

- Randomly sample n_1 minimal subsets of correspondences and acquire R_{ij} estimates using the method of Section 3. Acquire the final R_{ij} estimate by rotation averaging (Hartley et al., 2011) as detailed in Section 4.1.2. We sampled $n_1 = 200$ sets, after experimenting with different n_1 values in various datasets. The used n_1 value is much lower than the maximum allowed RANSAC iterations in available F-matrix estimation implementations (e.g 1000 iterations in 7-point F matrix algorithm (Kovesi, 2000))

Table 3
Performance of the 5P algorithm and our method in recovering the relative rotations R_{ij}. We initialize BA with each of the aforementioned methods and do 20 BA iterations. In BA, the internal parameter matrices K_i are held constant. Dataset **castle-P30** (Strecha et al., 2008b). Total number of pairs is 186. Castle-P30 was selected as the worst performing in multi-view reconstruction amongst (Strecha et al., 2008b) sets (data not shown)

	Median ΔR(°)	$\Delta R < 10°$ (pairs)	$\Delta R < 5°$ (pairs)
Proposed Method	3.49	123	106
5P	3.77	128	104

- Use the 5P algorithm inside a RANSAC procedure to acquire R_{ij} estimate. The implementation we used does 50 RANSAC iterations (Olsson and Enqvist, 2011). Note that because only five (compared to seven or eight points) are sampled, less RANSAC iterations are required (Fischler and Bolles, 1981)

In each of the two methods above, R_{ij} estimates were fine-tuned by running few (twenty) BA iterations for each reconstructed pair i, j (Lourakis and Argyros, 2009).

The initialization of BA is important, to improve convergence and to reduce the computational cost. The results show that 5P and our method can both be used as BA initializations with similar performance (Table 3). In our multi-view reconstruction pipeline (Fig. 5), pairwise rotation inaccuracies are reduced through rotation averaging in rotation registration step and BA.

The pairwise rotations results (Table 3) imply that to further reduce the reconstruction error, we should turn to other problem parameters as image correspondences and focal length estimates.

We have argued that the metric reconstruction framework allows for more solid validation. Indeed, we observed that defining as inliers the points satisfying the Cheirality condition and then simply selecting the maximum-inliers solution (i.e. RANSAC-like, no averaging yet), not only greatly simplifies the problem of best solution selection, but also already improves the reconstruction's accuracy. In data not shown, an approach based on epipolar geometry (i.e. robust F matrix computation) has been extensively investigated. Competing (1)F matrix estimation approaches (RANSAC, MAPSAC (Torr, 2002), least median squares (Rousseeuw, 1984), M-estimator (Torr and Murray, 1997), Levenberg–Marquardt optimization, orthogonal least squares (Armangué and Salvi, 2003), see Armangué and Salvi (2003) for F estimation methods) and (2)various heuristics on defining inliers using thresholds on reprojection error and using the number of inliers to choose the reliable reconstructions, were applied. Both steps, robust F estimation and using inliers to reject reconstructions, affected the reconstruction's accuracy, yet all these methods underperformed applying the Cheirality condition to select the best (maximum-inliers) solution.

Furthermore, through Theorems 1, 2 we get insight on the configuration in space of the two acquired solutions and the way the actual camera pair configuration determines the two solutions' configuration.

The metric reconstruction formulation of Sections 3.1–3.4 allows to explore important aspects including the critical camera configurations (see Appendix).

Finally, concerning the optimality of our method, it is known that in the case of two views – as examined in this paper — π_∞ can only be determined up to a two-solutions ambiguity (Kahl, 1999; Soatto et al., 2003). Thus we have introduced a linear method that requires a minimal number of point correspondences (required to acquire a projective transform) and is optimal in the number of recovered solutions.

5.2. Improving focal length estimation in multi-view reconstructions

We show in Table 4 the improved f estimates we get with cc. Further improvement is achieved by Jcc measure. Specifically, we see that using the exact same f estimates but altering the way we acquire the final f value can greatly reduce mean Δf (over all images in the

Table 4
Focal length averaging on **castle-P30** (Strecha et al., 2008a). We did 300 RANSAC iterations and used the 8-point algorithm for fundamental matrix computation. Then we applied the method of Section 3 for f estimation. Correspondences were obtained through SIFT features detection and matching.

Method	Median	Confidence count	Joint confidence count
Mean Δf error	0.28	0.17	0.07

Table 5
Median errors in $\mathbf{t}, R, f$ estimation in the final reconstruction. Correspondences were obtained through SIFT features detection and matching.
Source: Datasets from (Strecha et al., 2008a).

Dataset	$\Delta\mathbf{t}(°)$	$\Delta R(°)$	Δf
castle-P30	3.10	1.06	0.0389
castle-P19	7.35	4.17	0.0586
entry-P10	4.62	4.67	0.2118
Herz-Jesu-P8	1.00	0.68	0.0266
Herz-Jesu-P25	0.41	0.31	0.0049
Fountain-P11	0.44	0.41	0.0095

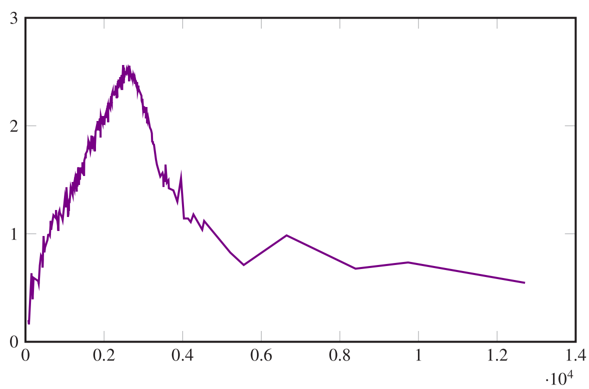

Fig. 9. Joint confidence count distribution, which displays a clear peak near correct f value.
Source: Data from **castle-P30** (Strecha et al., 2008a).

dataset) from 0.28, achieved using the median of f_i estimates, down to 0.07, achieved using Jcc method.

In Fig. 9 we present an non-ambiguous Jcc distribution from which f can be correctly determined, in contrast to ambiguous f estimate distributions displayed in Fig. 6.

5.3. Multi-view reconstruction in unordered image sets

Multi-view reconstructions demonstrate the validity of our approach. In Table 5, we provide quantitative performance measures for multi-view reconstructions that were acquired applying the proposed pipeline (Fig. 5), on unordered image datasets and with no other input apart from the scene photographs. In Fig. 10 we qualitatively display the results of the proposed reconstruction pipeline. The results in Table 5 and Fig. 10 demonstrate that the introduced methods can be used in unordered image sets to produce quality reconstructions of the photographed scenes.

Using the introduced (Section 3) self-calibration and metric reconstruction method combined with f and rotation averaging methods, allowed us to shift the focus from robust F estimation (projective reconstruction) to robust averaging of pairwise metric reconstruction obtained from minimal point correspondences sets. Each pairwise solution is solidly verified using metric reconstruction arguments (Cheirality condition). Furthermore, instead of relying on a single, accurate solution, e.g. robust F estimation, we propose to average multiple estimates of an unknown parameter (f_i, R_{ij}, R_i) to reach a robust final estimate.

Thus, we utilized the fast linear pairwise metric reconstruction method (Section 3) to sample numerous minimal solution-yielding sets in all the available camera pairs. Then, we diverged from choosing the best f_i, R_{ij} solution, e.g. by maximizing the number of inliers or minimizing the reconstruction error, and instead used f_i, R_{ij} averaging methods (Section 4.1). In f_i estimation, grouping all f_i estimates that originate in different i, j camera pairs (in Jcc estimation) allowed us to further improve estimation accuracy.

Averaging pairwise estimates has been explored in the literature enabling the integration of our self-calibration and metric reconstruction method to optimized SfM algorithms (Wilson and Snavely, 2014; Cui and Tan, 2015; Chatterjee and Govindu, 2018) which focus on aspects as translation averaging (Wilson and Snavely, 2014), scale, rotation and translation (similarity) averaging through initial depth-map computation (Cui and Tan, 2015) or fast and accurate relative rotation averaging (Chatterjee and Govindu, 2018).

5.4. Contributions to linear self-calibration and metric reconstruction

We discuss our contributions to linear self-calibration of image pairs and to multi-view reconstruction, in the context of the closely related approach of Pollefeys et al. (1999). Concerning the two aforementioned methods, our approach and Pollefeys et al. (1999), differences emerge on linearization, number of solutions and solution disambiguation. Both methods are based on ω^* expression in Eq. (14). However, while our approach solves directly for f_1, f_2, p_1, p_2, p_3 and directly imposes the rank degeneracy of the dual absolute quadric, the approach of Pollefeys et al. (1999) solves for $HQ^*_\infty H^T$ (Eq. (10)) elements in Eq. (14), and then imposes the degeneracy constraint through SVD decomposition. In the 2-views case, this approach leads to four solutions, however disambiguation using the Cheirality condition was not discussed and the geometric arrangement of the four solutions in space was left unexplored (Pollefeys et al., 1999). Furthermore, while our approach only discusses image pairs, the approach of Pollefeys et al. (1999) extends Eq. (14) to n-views ($n > 2$).

More importantly, while we validate our methods for two views, the authors in Pollefeys et al. (1999) steer their attention to simultaneous n-view self-calibration, omitting any experiments on two views. Specifically, simulations with synthetic data explored error with respect to noise-level (for six views) and error with respect to number of views (for $n > 5$ views).

Of equal importance are differences in integration to multi-view reconstruction. While in our approach we start from pair reconstructions, benefiting from the combinatorial increase in the number of pair reconstructions compared to the number of images, the authors of Pollefeys et al. (1999) upgrade a projective n-view reconstruction to a metric n-view reconstruction in one-step, and always consider all n-views simultaneously. Both approaches provide images of reconstructed geometry as a demonstration of faithful, to the human eye, n-view reconstructions.

Consequently, we see that both approaches validate n-view reconstruction similarly, yet two-view validation is only investigated in our approach.

Overall, while both methods apply to two-view reconstruction, method (Pollefeys et al., 1999) is validated only in n-view case ($n > 2$). In contrast, we explore pair-based reconstructions, provide data on Δf and ΔR_{ij} errors (ΔR_{ij} is not discussed in Pollefeys et al. (1999)) and construct a pair-based multi-view reconstruction pipeline that aims at robustness through solution averaging methods.

6. Conclusions

Using the DIAC, we developed a linear self-calibration and metric reconstruction method. We recover the optimal number of solutions (two) for the case of two cameras, assuming unknown and varying focal lengths but otherwise known internal parameters. Two theorems

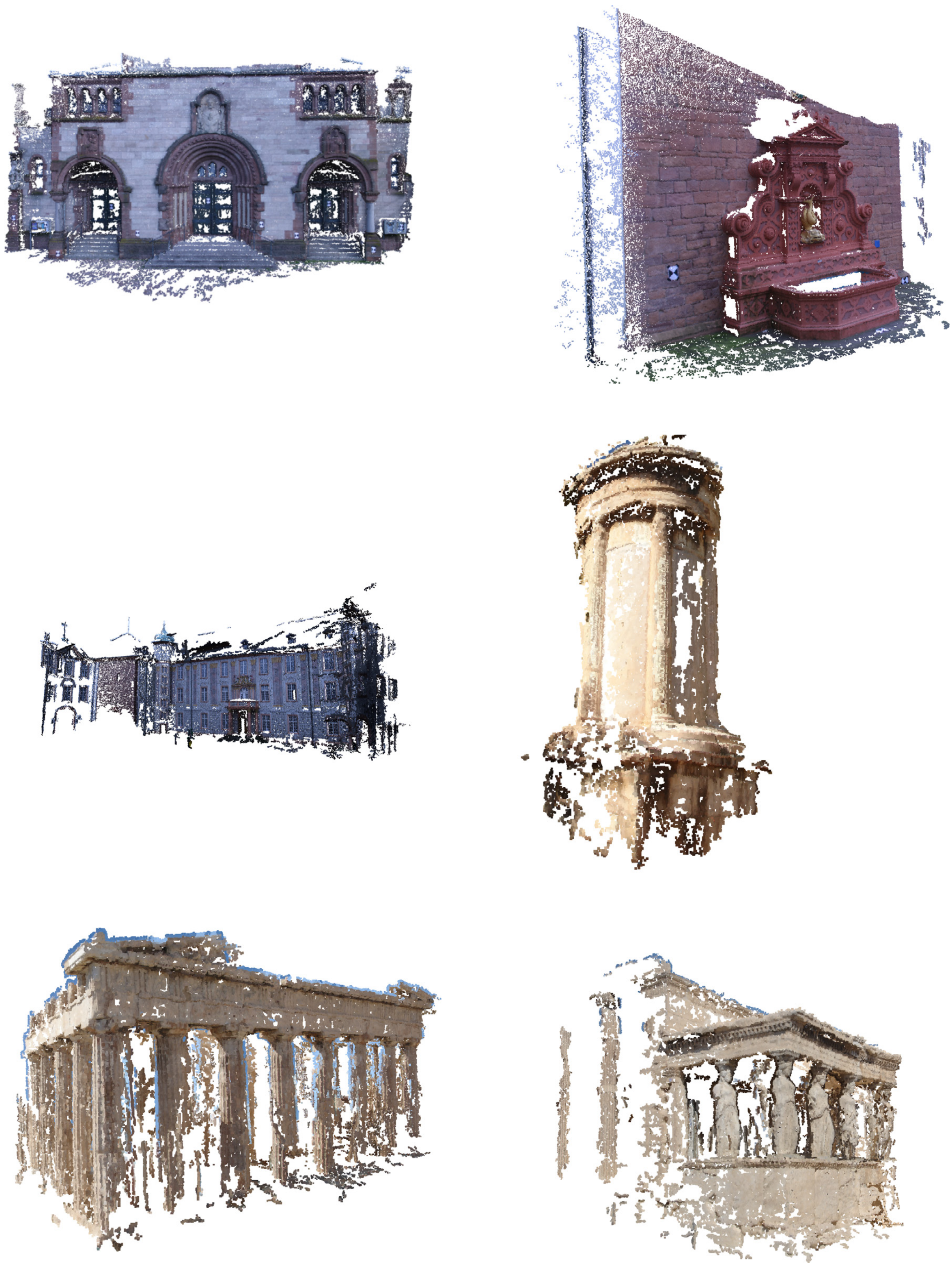

Fig. 10. 3D reconstruction results, as dense point clouds. **Top Row:** Datasets (Strecha et al., 2008a), Herz-Jesu-P25 (left) and Fountain-P11. **Middle Row:** Dataset castle-P19 (Strecha et al., 2008a) (left) and photo set of Monument of Lysicrates, Athens (right). **Bottom Row:** Photo sets of locations in Athens, Parthenon (left) and Karyatids (right). The scenes in Athens were photographed by the authors with a simple compact camera.

describe the relative configuration of the two recovered solutions and provide support to use the Cheirality condition for solution disambiguation. We demonstrate the validity of our approach using both synthetic and real data. Comparisons to Kruppa equations and the 5P algorithm revealed that our method performs similarly to these standard approaches. Subsequently we show that the large number of f, R estimates that are produced by our self-calibration and metric reconstruction method can be utilized through averaging methods, shifting our focus from choosing the best solution, e.g. as in optimized and robust F estimation prior to self-calibration, to finding the best solution averaging method. All our methods were integrated to a full multiple-view reconstruction pipeline to produce visually high-quality reconstructions on both standard datasets and image sets we shot using a conventional camera. Multi-view reconstructions were obtained combining camera pair reconstructions using rotation averaging algorithms and a novel approach to average focal length estimates.

CRediT authorship contribution statement

Nikos Melanitis: Conceptualization, Methodology, Software, Writing - original draft, Writing - review & editing. **Petros Maragos:** Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

N. Melanitis is funded by General Secretariat for Research and Technology (GSRT), Greece and Hellenic Foundation for Research and Innovation (HFRI), Greece.

Appendix A. Gaussian elimination in self calibration and metric reconstruction equations: The general case and critical configurations

To simplify the expressions, we introduce the notation

$\mathbf{P_i^\gamma}$: row vector produced from ith row of $[\mathbf{a}]_x F$

and permute $\mathbf{x_0}$ elements with the permutation

$$4 \leftrightarrow 6 \tag{50}$$

We denote the permuted vector by $\mathbf{x}$ and the corresponding system matrix by A_{pr}. Using this notation, we write A_{pr} as

$$\begin{bmatrix} \mathbf{r_1} & \phi_1^1\mathbf{P_2^\gamma} & \psi_1 \\ \mathbf{r_2} & \phi_2^1\mathbf{P_2^\gamma}+\phi_2^2\mathbf{P_3^\gamma} & \psi_2 \\ \mathbf{r_3} & \phi_3^1\mathbf{P_1^\gamma}+\phi_3^2\mathbf{P_3^\gamma} & \psi_3 \\ \mathbf{r_4} & \phi_4^1\mathbf{P_1^\gamma} & \psi_4 \\ \mathbf{r_5} & \phi_5^1\mathbf{P_1^\gamma}+\phi_5^2\mathbf{P_2^\gamma} & \psi_5 \end{bmatrix} \tag{51}$$

where $\mathbf{r_i}$ are 1×3 vectors and ϕ, ψ are appropriate constants of no special structure.

We aim to eliminate the elements in the rows 1–3 and columns 4–6 of A_{pr}, which we refer to as A_n, and then to apply regular Gaussian elimination. This is generally possible (see below), owing to the structure of A_n rows in (51), which are linear combinations of $\mathbf{P_i^\gamma}$ vectors and, also, using the canonical projective reconstruction allows us to substitute

$$\mathbf{P_3^\gamma} = d_1\mathbf{P_1^\gamma} + d_2\mathbf{P_2^\gamma} \tag{52}$$

Thus, the elimination of A_n elements is now straightforward by applying row-operations to matrix A_{pr}. We then apply ordinary Gaussian elimination to reduce A_{pr} to the form of (20).

Next, we identify critical camera configurations in which the linear system we examine becomes degenerate.

First, we have camera configurations in which f_1, f_2, π_∞ cannot be determined, following any approach:

- no rotation (Strecha et al., 2006)
- translation along the viewing direction of camera and rotation around the viewing direction (Kahl, 1999; Sturm, 2002)

Additionally, we have critical configurations for specific self-calibration and metric reconstruction methods. We searched for critical configurations in the literature and checked, reproducing the critical configurations in synthetic scenes, if the degeneracies arise using the method we introduced. The following camera configurations were found to be critical for our method:

- relative position of two cameras can be described by planar motion (Brooks et al., 1996)
- the first camera's viewing direction, the baseline (line connecting the two camera centers) and the vector perpendicular to the baseline and to the second camera's viewing direction, are all coplanar (Brooks and Pan, 1996)
- camera centers are positioned on a sphere and their viewing directions are radiuses of that sphere (Brooks et al., 1996)
- camera relative rotation is around an axis that is parallel to camera translation vector and rotation angle is 90° (Ma et al., 2000)

Appendix B. Geometric relations between the two recovered solutions for metric reconstruction of a camera pair

We proceed with the proofs of Theorems 1 and 2.

Result 5. Let P denote a projection matrix. The center of projection $\mathbf{C_P}$ has no image, as it is projected to point $\mathbf{0}$. Equivalently, $\mathbf{C_P} = \begin{pmatrix}\mathbf{C^T} & 1\end{pmatrix}^T$ is a right null-vector of P.

Result 6. Let P denote a projection matrix. P can be decomposed as

$$P = \begin{bmatrix} KR & -KR\mathbf{C} \end{bmatrix} \tag{53}$$

Results 5, 6 describe properties of the camera position $\mathbf{C}$. The following Result is concerned with the camera direction

Result 7. Assume a projection matrix

$$P = \begin{bmatrix} M & \mathbf{p} \end{bmatrix} \tag{54}$$

Let the vector $\mathbf{m_3^T}$ denote the third row of M. Then the vector

$$\mathbf{v} = \det(M)\mathbf{m_3} \tag{55}$$

is in the direction of the principal axis (the viewing direction) of P and is directed towards the front of the camera.

The next two lemmas describe properties of metric reconstructions P_{m2}^1, P_{m2}^2 derived from Eq. (27)

Lemma 1. *Let*

$$\begin{aligned} P_{m2}^1 &= \begin{bmatrix} K_2^1 R^1 & \mathbf{a^1} \end{bmatrix} \\ P_{m2}^2 &= \begin{bmatrix} K_2^2 R^2 & \mathbf{a^2} \end{bmatrix} \end{aligned} \tag{56}$$

be the projection matrices for camera 2 derived from Eq. (27). Then

$$\mathbf{a^1} = \mathbf{a^2} \triangleq \mathbf{a} \tag{57}$$

Proof. Considering:

1. The form of homography (13)
2. Eq. (9): $P_{m2}^i = P_{P2}H^i$ where H^i denotes the homography obtained by substituting the ith solution of Eq. (27)

the lemma is readily deduced □

Lemma 2. *Let P_{m2}^1, P_{m2}^2 as in Lemma 1. We have:*

$$K_2^1 R_2^1 - K_2^2 R_2^2 = \mathbf{a}\mathbf{n^T} \tag{58}$$

where $\mathbf{n}$ is an appropriate vector.

Proof. As in proof of Lemma 1, by observing that H^i for different i values differ only in $\mathbf{v} \triangleq -\mathbf{p^T}K$. □

We note that Lemma 2 is a general result, independent of (27). However, we omit this proof now.

Lemma 3. *For the reconstructions* $P^1_{m2} = \begin{bmatrix} P_1 & \mathbf{a} \end{bmatrix}$, $P^2_{m2} = \begin{bmatrix} P_2 & \mathbf{a} \end{bmatrix}$ *we have*

$$\det P_1 = \pm \det P_2 \tag{59}$$

Proof. We formed P^1_{m2}, P^2_{m2} from solutions of Eq. (14), so the projection matrices have the same ω^*. Using now Eq. (1) we have:

$$\begin{aligned}\omega_1^* &= P^1_{m2} Q^*_\infty {P^1_{m2}}^T \\ &= P_1 P_1^T \\ &= \omega_2^* \\ &= P_2 P_2^T\end{aligned} \tag{60}$$

Using known properties of determinants we have

$$\begin{aligned}&\det P_1 P_1^T = \det P_2 P_2^T \Rightarrow \\ &\det {P_1}^2 = \det {P_2}^2 \Rightarrow \\ &\det P_1 = \pm \det P_2 \quad \square\end{aligned} \tag{61}$$

From this last proof, the next Lemma becomes apparent

Lemma 4. *Concerning the reconstructions* P^1_{m2}, P^2_{m2} *of Lemma 3, we have*

$$K^1_{m2} = K^2_{m2} \tag{62}$$

Proof. From the equality of ω_1^*, ω_2^* and the diagonality of the internal calibration matrices K^1_{m2}, K^2_{m2} we prove the Lemma. $\square$

In the following we introduce the simplified notation: $K_2 \triangleq K^1_{m2} = K^2_{m2}$. We next refine Lemma 3, to lift the sign ambiguity in Eq. (59).

Lemma 5. *For the reconstructions* P^1_{m2}, P^2_{m2} *we have*

$$\mathbf{C^1_{m2}} = \mathbf{C^2_{m2}} \triangleq \mathbf{C} \iff \mathbf{n^{TC}} = 0 \tag{63}$$

Proof.

$$\begin{aligned}&P^1_{m2}\begin{pmatrix}\mathbf{C}\\1\end{pmatrix} = \mathbf{0} \\ &\iff \mathbf{a} = -K_2 R^1_2 \mathbf{C}\end{aligned} \tag{64}$$

Similarly for P^2_{m2}

$$\begin{aligned}&P^2_{m2}\begin{pmatrix}\mathbf{C}\\1\end{pmatrix} = \mathbf{0} \\ &\iff K_2 R^1_2 \mathbf{C} + \mathbf{an^{TC}} + \mathbf{a} = \mathbf{0} \\ &\iff K_2 R^1_2 \mathbf{C} + \mathbf{an^{TC}} - K_2 R^1_2 \mathbf{C} = \mathbf{0}, \text{holds from Eq. (64)} \\ &\iff \mathbf{an^{TC}} = \mathbf{0} \\ &\iff \mathbf{n^{TC}} = 0 \quad \square\end{aligned} \tag{65}$$

Lemma 6. *For the reconstructions* P^1_{m2}, P^2_{m2} *we have*

$$\mathbf{C^1_{m1}} = -\mathbf{C^2_{m2}} \triangleq \mathbf{C} \iff \mathbf{n^{TC}} = -2 \tag{66}$$

Proof. As in the proof of Lemma 5

$$\begin{aligned}&P^1_{m2}\begin{pmatrix}\mathbf{C}\\1\end{pmatrix} = \mathbf{0} \\ &\iff \mathbf{a} = -K_2 R^1_2 \mathbf{C}\end{aligned} \tag{67}$$

Similarly, from matrix P^2_{m2} we have

$$\begin{aligned}&P^2_{m2}\begin{pmatrix}-\mathbf{C}\\1\end{pmatrix} = \mathbf{0} \\ &\iff -K_2 R^1_2 \mathbf{C} + \mathbf{an^{TC}} + \mathbf{a} = \mathbf{0} \\ &\iff \mathbf{a} + \mathbf{an^{TC}} + \mathbf{a} = \mathbf{0}, \text{from (67)} \\ &\iff \mathbf{a}(2 + \mathbf{n^{TC}}) = \mathbf{0} \\ &\iff \mathbf{n^{TC}} = -2, \text{provided } \mathbf{a} \neq \mathbf{0} \quad \square\end{aligned} \tag{68}$$

We complement each of Lemmas 5,6, with Lemma 7 and 8 respectively. To prove the last two Lemmas, we use Eq. (69)

Result 8. For each square, invertible matrix X, column-vector $\mathbf{c}$ and row-vector $\mathbf{r}$ we have

$$\det(X + \mathbf{cr}) = \det X \cdot \det(1 + \mathbf{r}X^{-1}\mathbf{c}) \tag{69}$$

Lemma 7. *For the reconstructions* P^1_{m2}, P^2_{m2} *we have*

$$\begin{aligned}&\det P_1 = \det P_2 \iff \\ &\mathbf{n^{TC_1}} = 0\end{aligned} \tag{70}$$

Proof. We use Result 8, for which we note:

1. P_1 is a full-rank matrix (rank 3) for every projection matrix. The exception, referred to in the literature as "camera at infinity", is out of our scope. Remember we are handling a metric reconstruction.
2. P_2 can be expressed in terms of P_1, $\mathbf{n}, \mathbf{a}$, thus permitting the application of Eq. (69) to determine $\det P_2$.

Now applying the previous points, we have

$$\begin{aligned}&\det P_2 = \det P_1 \\ &\iff 1 - \mathbf{n^T} {R^1_2}^T K_2^{-1}\mathbf{a} = 1 \\ &\iff \mathbf{n^T} {R^1_2}^T K_2^{-1}\mathbf{a} = 0 \\ &\iff -\mathbf{n^T} {R^1_2}^T K_2^{-1} K_2 R^1_2 \mathbf{C_1} = 0, \\ &\text{from (64): } \mathbf{a} = -K_2 R^1_2 \mathbf{C_1} \\ &\iff \mathbf{n^{TC_1}} = 0, \\ &\text{as } RR^T = I \text{ for rotation matrices } R \quad \square\end{aligned} \tag{71}$$

Lemma 8. *For the reconstructions* P^1_{m2}, P^2_{m2} *we have*

$$\begin{aligned}&\det P_1 = -\det P_2 \iff \\ &\mathbf{n^{TC_1}} = -2\end{aligned} \tag{72}$$

Proof. As in the proof of Lemma 7,we have:

$$\begin{aligned}&\det P_2 = -\det P_1 \\ &\iff 1 - \mathbf{n^T} {R^1_2}^T K_2^{-1}\mathbf{a} = -1 \\ &\iff \mathbf{n^T} {R^1_2}^T K_2^{-1}\mathbf{a} = 2 \\ &\iff -\mathbf{n^T} {R^1_2}^T K_2^{-1} K_2 R^1_2 \mathbf{C_1} = 2, \\ &\text{from Eq. (64): } \mathbf{a} = -K_2 R^1_2 \mathbf{C_1} \\ &\iff \mathbf{n^{TC_1}} = -2, \\ &\text{as } RR^T = I \text{ for rotation matrices } R \quad \square\end{aligned} \tag{73}$$

Now, we show that the case of same-sign determinants ($\det P^1 = \det P^2$) produces a contradiction, and is so rejected. Regarding the notation in the following, we clarify that:

1. The projective reconstruction P_{P2} is in the canonical representation form
$$\begin{bmatrix} [\mathbf{a}]_x F & \mathbf{a} \end{bmatrix} \tag{74}$$
with $F^T\mathbf{a} = \mathbf{0}$
2. $[\mathbf{a}]_x$ denotes the anti-symmetric matrix defined to compute outer product with vector $\mathbf{a}$
$$[\mathbf{a}]_x \mathbf{v} = \mathbf{a} \times \mathbf{v} \tag{75}$$
3. $\mathbf{e}$ denotes the right null vector of F,
$$F\mathbf{e} = \mathbf{0} \tag{76}$$

Lemma 9. *Let*

$$P_{P2} = \begin{bmatrix} A & \mathbf{a} \end{bmatrix} = \begin{bmatrix} [\mathbf{a}]_x F & \mathbf{a} \end{bmatrix} \tag{77}$$

denote the projection matrix for camera 2 in the projective reconstruction and $\mathbf{p}, \mathbf{p}'$ *the solutions for* $\boldsymbol{\pi}_\infty$ *acquired from Eq.* (27)

$$\begin{aligned} \mathbf{p}^{\mathbf{T}} &= \begin{pmatrix} p_1 & p_2 & p_3 \end{pmatrix} \\ \mathbf{p}'^{\mathbf{T}} &= \begin{pmatrix} p_1' & p_2' & p_3' \end{pmatrix} \end{aligned} \tag{78}$$

Then

$$\mathbf{p} - \mathbf{p}' = \psi \mathbf{e_f} \tag{79}$$

where

$$\mathbf{e_f} = \begin{pmatrix} e_1/f_1^2 \\ e_2/f_1^2 \\ e_3 \end{pmatrix} \tag{80}$$

Proof. From Eq. (14) and because solutions (27) share the same f_1 value, we have

$$\begin{gathered} \omega_1^* = \omega_2^* \iff \\ P_{P2} \begin{bmatrix} K_1 K_1^T & -K_1 K_1^T \mathbf{p} \\ -\mathbf{p}^{\mathbf{T}} K_1 K_1^T & \mathbf{p}^{\mathbf{T}} K_1 K_1^T \mathbf{p} \end{bmatrix} P_{P2}^T = \\ P_{P2} \begin{bmatrix} K_1 K_1^T & -K_1 K_1^T \mathbf{p}' \\ -\mathbf{p}'^{\mathbf{T}} K_1 K_1^T & \mathbf{p}'^{\mathbf{T}} K_1 K_1^T \mathbf{p}' \end{bmatrix} P_{P2}^T \iff \\ A K_1 K_1^T A^T - A K_1 K_1^T \mathbf{p} \mathbf{a}^{\mathbf{T}} - \mathbf{a} \mathbf{p}^{\mathbf{T}} K_1 K_1 A^T + \mathbf{a} \mathbf{p}^{\mathbf{T}} K_1 K_1^T \mathbf{p} \mathbf{a}^{\mathbf{T}} = \\ A K_1 K_1^T A^T - A K_1 K_1^T \mathbf{p}' \mathbf{a}^{\mathbf{T}} - \mathbf{a} \mathbf{p}'^{\mathbf{T}} K_1 K_1 A^T + \mathbf{a} \mathbf{p}'^{\mathbf{T}} K_1 K_1^T \mathbf{p}' \mathbf{a}^{\mathbf{T}} \end{gathered} \tag{81}$$

From Eqs. (27), (24) we have

$$f_1^2 p_1^2 + f_2^2 p_2^2 + p_3^2 = f_1^2 p'^2_1 + f_2^2 p'^2_2 + p'^2_3 = b_3 \tag{82}$$

and so

$$(f_1^2 p_1^2 + f_1^2 p_2^2 + p_3^2) \mathbf{a} \mathbf{a}^{\mathbf{T}} = \mathbf{a} \mathbf{p}^{\mathbf{T}} K_1 K_1^T \mathbf{p} \mathbf{a}^{\mathbf{T}} = \mathbf{a} \mathbf{p}'^{\mathbf{T}} K_1 K_1^T \mathbf{p}' \mathbf{a}^{\mathbf{T}} \tag{83}$$

By eliminating the common terms (Eq. (83) and $A K_1 K_1 A^T$) we continue the computations and arrive at

$$\begin{gathered} A K_1 K_1^T ((\mathbf{p} - \mathbf{p}') \mathbf{a}^{\mathbf{T}}) + (\mathbf{a}(\mathbf{p}^{\mathbf{T}} - \mathbf{p}'^{\mathbf{T}})) K_1 K_1^T A^T = 0 \iff \\ Q + Q^T = 0 \end{gathered} \tag{84}$$

In Eq. (84) we defined

$$Q \triangleq A K_1 K_1^T \left(\left(\mathbf{p} - \mathbf{p}' \right) \mathbf{a}^{\mathbf{T}} \right) \tag{85}$$

We write Q as

$$\begin{aligned} Q &= A K_1 K_1^T ((\mathbf{p} - \mathbf{p}') \mathbf{a}^{\mathbf{T}}) \\ &= A K_1 K_1^T \begin{pmatrix} (p_1 - p_1') \mathbf{a}^{\mathbf{T}} \\ (p_2 - p_2') \mathbf{a}^{\mathbf{T}} \\ (p_3 - p_3') \mathbf{a}^{\mathbf{T}} \end{pmatrix} \\ &= A \begin{pmatrix} f_1^2 (p_1 - p_1') \mathbf{a}^{\mathbf{T}} \\ f_1^2 (p_2 - p_2') \mathbf{a}^{\mathbf{T}} \\ (p_3 - p_3') \mathbf{a}^{\mathbf{T}} \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{A^1}^T \\ \mathbf{A^2}^T \\ \mathbf{A^3}^T \end{pmatrix} \begin{pmatrix} \Delta_f \alpha_1 & \Delta_f \alpha_2 & \Delta_f \alpha_3 \end{pmatrix} \end{aligned} \tag{86}$$

where in Eq. (86) we defined

$$\Delta_f \triangleq \begin{pmatrix} f_1^2 (p_1 - p_1') \\ f_1^2 (p_2 - p_2') \\ (p_3 - p_3') \end{pmatrix} \tag{87}$$

From Eq. (84), matrix Q is anti-symmetric and so has a zero diagonal. Imposing the last condition on expression (86), we extract the following relations

$$\mathbf{A^1}^T \Delta_f = 0 \tag{88}$$

$$\mathbf{A^2}^T \Delta_f = 0 \tag{89}$$

$$\mathbf{A^3}^T \Delta_f = 0 \tag{90}$$

We substitute A in Eqs. (88),(89),(90), using the canonical representation assumption

$$P_{P2} = \begin{bmatrix} A & \mathbf{a} \end{bmatrix} = \begin{bmatrix} [\mathbf{a}]_x F & \mathbf{a} \end{bmatrix} \tag{91}$$

and write the three resulting equations in matrix form to get

$$[\mathbf{a}]_x F \Delta_f = \mathbf{0} \tag{92}$$

From Eq. (92) and because $[\mathbf{a}]_x F$ has the null vector $\mathbf{e}$, we get

$$\Delta_f = \psi \mathbf{e} \tag{93}$$

where ψ is a constant.

Now, from Eq. (93), with simple manipulations we obtain:

$$\begin{pmatrix} p_1 - p_1' \\ p_2 - p_2' \\ p_3 - p_3' \end{pmatrix} = \psi \mathbf{e_f} \quad \square \tag{94}$$

Lemma 10. *With the assumptions and notation of Lemma 9, we have*

$$\det P_1 = -\det P_2 \tag{95}$$

Proof. We assume that

$$\det P_1 = \det P_2 \tag{96}$$

and produce a contradiction.

From Lemma 7, we get Eq. (70) and equivalently require that:

$$\mathbf{n}^{\mathbf{T} \mathbf{C}_1} = 0 \tag{97}$$

To specify $\mathbf{n}$ in Eq. (97), we use

1. The definition of $\mathbf{n}$ in Eq. (58)
2. The relation between P_{P2}, P_{M2}, H (Eqs. (9),(13)) and the notation for P matrix of Lemma 9

and have

$$\begin{aligned} P_1 &= A K_1 - \mathbf{a} \mathbf{p}^{\mathbf{T}} K_1 \\ P_2 &= A K_1 - \mathbf{a} \mathbf{p}'^{\mathbf{T}} K_1 \\ \iff P_2 &= P_1 + \mathbf{a} (\mathbf{p} - \mathbf{p}')^{\mathbf{T}} K_1 \triangleq P_1 - \mathbf{a} \mathbf{n}^{\mathbf{T}} \end{aligned} \tag{98}$$

Now, we can rewrite Eq. (97) as

$$(\mathbf{p} - \mathbf{p}')^{\mathbf{T}} \mathbf{K_1} \mathbf{C_1} = 0 \tag{99}$$

We next have

$$\begin{gathered} P_{M2}^1 \begin{pmatrix} \mathbf{C^1} \\ 1 \end{pmatrix} = 0 \iff \\ P_{P2} H^1 \begin{pmatrix} \mathbf{C^1} \\ 1 \end{pmatrix} = 0 \iff \\ P_{P2} \begin{pmatrix} K_1 \mathbf{C^1} \\ -\mathbf{p}^{\mathbf{T}} K_1 \mathbf{C^1} + 1 \end{pmatrix} = 0 \end{gathered} \tag{100}$$

From the assumption that P_{P2} is in the canonical form (Eq. (74)), it has a null vector (Eq. (76)) that is written as

$$\begin{pmatrix} \mathbf{e} \\ 0 \end{pmatrix} \tag{101}$$

So we have:

$$P_{P2} \begin{pmatrix} K_1 \mathbf{C^1} \\ -\mathbf{p}^{\mathbf{T}} K_1 \mathbf{C^1} + 1 \end{pmatrix} = 0 \iff$$

$$K_1 \mathbf{C^1} = \psi \mathbf{e} \text{ ,where } \psi \text{ is a constant} \tag{102}$$

$$-\mathbf{p}^{\mathbf{T}} K_1 \mathbf{C^1} + 1 = 0 \tag{103}$$

From Lemma 9 (Eq. (79)) and the previous Eqs. (99), (102) we get:

$$\mathbf{e}_{\mathbf{F}}^{\mathbf{Te}} = 0 \iff$$

$$e_1^2/f_1^2 + e_2^2/f_1^2 + e_3^2 = 0 \tag{104}$$

Since Eq. (104) has no solutions ($\mathbf{e} \neq \mathbf{0}$), we produced a contradiction.

Thus, from Lemma 3, we have proved that

$$\det P_1 = -\det P_2 \quad \square \tag{105}$$

From the preceding Lemmas, we can now readily obtain Theorem 1

Proof of Theorem 1. From Lemma 10

$$\det P_1 = -\det P_2 \tag{106}$$

From Lemmas 6,8 we obtain the equivalent relation

$$\mathbf{C^{1}}_{\mathbf{m2}} = -\mathbf{C}^{\mathbf{2}}_{\mathbf{m2}} \quad \square \tag{107}$$

Next, we prove Theorem 2. To avoid a lengthy proof, we settle the coplanarity of $\mathbf{v}^{\mathbf{1,2}}_{\mathbf{m2}}, \mathbf{C}^{\mathbf{1,2}}_{\mathbf{m2}}$ with Lemma 11, which follows the main proof.

Let us first summarize some notation

1. We denote $\mathbf{v}^{\mathbf{1}}_{\mathbf{m2}}, \mathbf{v}^{\mathbf{2}}_{\mathbf{m2}}$ the vectors that point along the viewing directions of cameras P^1_{m2}, P^2_{m2} respectively
2. For P^1_{m2} we assume

$$\det P_1 > 0 \tag{108}$$

$$\mathbf{C} \triangleq \mathbf{C^{1}}_{\mathbf{m2}} \tag{109}$$

Proof of Theorem 2. From Results 6, 7, Lemma 1, Theorem 1 we have for P^1_{m2}:

$$K_2 R^1 \mathbf{C} = -\mathbf{a} \iff$$

$$\begin{pmatrix} f_2 \mathbf{R}_{\mathbf{1}}^T \\ f_2 \mathbf{R}_{\mathbf{2}}^T \\ \mathbf{R}_{\mathbf{3}}^T \end{pmatrix} \mathbf{C} = -\mathbf{a} \iff$$

$$\mathbf{R}_{\mathbf{3}}^T \mathbf{C} = -a_3 \tag{110}$$

We have $\det P^1 = \det K_2 R_1 > 0$ and so

$$\mathbf{v}^{\mathbf{1}}_{\mathbf{2m}} = \mathbf{R}_{\mathbf{3}} \tag{111}$$

Consequently, from Eq. (110), we have

$$\mathbf{v}^{\mathbf{1}}_{\mathbf{2m}}{}^T \mathbf{C} = \|\mathbf{v}^{\mathbf{1}}_{\mathbf{2m}}\| \|\mathbf{C}\| \cos \angle \mathbf{C}, \mathbf{v}^{\mathbf{1}}_{\mathbf{2m}} = -a_3 \tag{112}$$

In Eq. (112),

$\|\mathbf{R}_{\mathbf{3}}^{\mathbf{T}}\| = 1$, since R^1 is orthogonal as a rotation matrix

We can normalize $\mathbf{C}$ to unitary by satisfying the condition

$$\|K_2^{-1}\mathbf{a}\| = 1 \tag{113}$$

since rotations leave vectors' measure unchanged.

We can now write Eq. (112) as

$$\cos \angle \mathbf{C}, \mathbf{v}^{\mathbf{1}}_{\mathbf{2m}} = -a_3 \tag{114}$$

Similarly, using

$$\mathbf{C}^{\mathbf{2}}_{\mathbf{m2}} = -\mathbf{C} \tag{115}$$

$$\det P^2 < 0 \tag{116}$$

$$\mathbf{a^1} = \mathbf{a^2} \tag{117}$$

we have

$$\cos \angle \mathbf{C}, \mathbf{v}^{\mathbf{2}}_{\mathbf{m2}} = -a_3 \tag{118}$$

and the remaining relations required for the proof:

$$\cos \angle \mathbf{C^2}, \mathbf{v}^{\mathbf{1}}_{\mathbf{m2}} = a_3 \tag{119}$$

$$\cos \angle \mathbf{C^2}, \mathbf{v}^{\mathbf{2}}_{\mathbf{m2}} = a_3 \tag{120}$$

$$\angle \mathbf{C}, \mathbf{v}^{\mathbf{1}}_{\mathbf{m2}} + \angle \mathbf{C^2}, \mathbf{v}^{\mathbf{1}}_{\mathbf{m2}} = 180° \tag{121}$$

$$\angle \mathbf{C}, \mathbf{v}^{\mathbf{2}}_{\mathbf{m2}} + \angle \mathbf{C^2}, \mathbf{v}^{\mathbf{2}}_{\mathbf{m2}} = 180° \tag{122}$$

To complete the proof, we show that $\mathbf{v}^{\mathbf{1,2}}_{\mathbf{m2}}, \mathbf{C}^{\mathbf{1,2}}_{\mathbf{m2}}$ are coplanar. We provide a constructive proof in Lemma 11. $\square$

Lemma 11. *There exist rotation matrices R_x, R_{perm} so that*

$$R_x R_{perm} R^1 \mathbf{v}^{\mathbf{1}}_{\mathbf{m2}} = \begin{pmatrix} 1 & 0 & 0 \end{pmatrix}^T \tag{123}$$

$$R_x R_{perm} R^1 \mathbf{v}^{\mathbf{2}}_{\mathbf{m2}} = \begin{pmatrix} x & y & 0 \end{pmatrix}^T \tag{124}$$

Proof. In this proof, we apply to the 3D space similarity transforms, that do not alter angles. The aim is to transform the space so that the resulting coordinate system simplifies the relations of the entities we examine.

We visualize this process as placing and orienting a "virtual" camera, so that the camera primary plane is the plane on which $\mathbf{v}^{\mathbf{1,2}}_{\mathbf{m2}}, \mathbf{C}^{\mathbf{1,2}}_{\mathbf{m2}}$ lie. We first do some hypotheses, without loss of generality, to simplify the notation in the proof:

- Let P^1_{m2} denote the correct representation of P_{m2} and P^2_{m2} the erroneous one
- Let

$$\text{sign}(\det P_1) > 0 \tag{125}$$

so that we can simplify the expression for camera viewing direction

We apply to space the rotations

$$R_x R_{perm} R^1 \tag{126}$$

where

R^1 : rotation matrix of P^1_{m2}

R_{perm} : rotation to transpose x_1, x_3 of a vector: $\begin{pmatrix} x_1 & x_2 & x_3 \end{pmatrix}^T$

R_x :rotation to place $\mathbf{C^1}$ in the desired plane

Applying R^1, using orthogonality of R^1 and Result 7, we have for the viewing direction of camera 2

$$R^1 \mathbf{v}^{\mathbf{1}}_{\mathbf{m2}} = \begin{pmatrix} 0 & 0 & 1 \end{pmatrix}^T \tag{127}$$

We then apply R_{perm}, to help with the visualization of this proof

$$R_{perm} = R_y(90°) = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ -1 & 0 & 0 \end{bmatrix} \tag{128}$$

We define R_x, a rotation around x-axis, to place $\mathbf{C^1}$ on z−plane and at the same time leave $\mathbf{v}^{\mathbf{1}}_{\mathbf{m2}}$ unchanged. We have

$$R_x = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_x & -\sin\theta_x \\ 0 & \sin\theta_x & \cos\theta_x \end{bmatrix} \tag{129}$$

We transformed $\mathbf{C^1}$ to:

$$K_2 R^1 \mathbf{C^1} = -\mathbf{a} \iff$$

$$R^1 \mathbf{C^1} = -K_2^{-1}\mathbf{a} \iff$$

$$R_{perm} R^1 \mathbf{C^1} = -R_{perm}\left(-K_2^{-1}\right)\mathbf{a} = \begin{pmatrix} -a_3 \\ -f_2 a_2 \\ f_2 a_1 \end{pmatrix} \tag{130}$$

Then, applying R_x we have:

$$R_x R_{perm} R^1 \mathbf{C^1} = R_x \begin{pmatrix} -a_3 \\ -f_2 a_2 \\ f_2 a_1 \end{pmatrix}$$

$$= \begin{pmatrix} x_1 \\ x_2 \\ -f_2 a_2 \sin\theta_x + f_2 a_1 \cos\theta_x \end{pmatrix} \tag{131}$$

To satisfy the condition ($x_3 = 0$) for $\mathbf{C^1}$, we get for R_x

$$-f_2 a_2 \sin\theta_x + f_2 a_1 \cos\theta_x = 0 \iff$$
$$\tan\theta_x = {}^{a_1}/_{a_2} \tag{132}$$

Now, we show that the R_x we specified previously, also places $\mathbf{v^2_{m2}}$ on the $z-$plane of the virtual camera. Let

$$\mathbf{m^i_j} : j\text{−row of } R^i \tag{133}$$

We have

$$\begin{aligned} R_{perm} R^1 \mathbf{v^2_{m2}} &= R_{perm} R^1 (-\mathbf{m^2_3}^T) \\ &= \left(\mathbf{m^1_3}(-\mathbf{m^2_3}^T) \quad \mathbf{m^1_2}(-\mathbf{m^2_3}^T) \quad -\mathbf{m^1_1}(-\mathbf{m^2_3}^T)\right)^T \\ &\triangleq (x_1 \quad x_2 \quad x_3)^T \end{aligned} \tag{134}$$

where we used that $\det P^2 < 0$. Now, it suffices to show

$$\begin{aligned} R_x (x_1 \quad x_2 \quad x_3)^T = (x'_1 \quad x'_2 \quad 0) &\iff \\ \sin\theta_x \mathbf{m^1_2}\mathbf{m^2_3}^T = \cos\theta_x \mathbf{m^1_1}\mathbf{m^2_3}^T &\iff \\ a_1(\mathbf{m^2_2} + f_2^{-1} a_2 \mathbf{n^T})\mathbf{m^2_3}^T = a_2(\mathbf{m^2_1} + f_2^{-1} a_1 \mathbf{n^T})\mathbf{m^2_3}^T &\iff \\ a_1 f_2^{-1} a_2 \mathbf{n^T}\mathbf{m^2_3}^T = a_2 f_2^{-1} a_1 \mathbf{n^T}\mathbf{m^2_3}^T & \end{aligned} \tag{135}$$

where we used Lemma 2, diagonal form of K_2, the orthogonality of rotation matrices and that $\mathbf{a_1} = \mathbf{a_2}$.

Thus, we proved that $\mathbf{v^1_{m2}}, \mathbf{v^2_{m2}}, \mathbf{C^1}, \mathbf{C^2}$ lie in the plane $z = 0$ of the transformed world coordinate system, which is the primary plane of the "virtual" camera. □

References

Armangué, X., Salvi, J., 2003. Overall view regarding fundamental matrix estimation. Image Vis. Comput. 21 (2), 205–220.

Brooks, M.J., de Agapito, L., Huynh, D.Q., Baumela, L., 1996. Direct methods for self-calibration of a moving stereo head. In: European Conference on Computer Vision. Springer, pp. 413–426.

Brooks, G.N.D.H.M., Pan, H., 1996. Recovering unknown focal lengths in self-calibration: An essentially linear algorithm and degenerate configurations. In: Proc. ISPRS-Congress, vol. 31. Citeseer, pp. 575–580.

Chandraker, M., Agarwal, S., Kahl, F., Nistér, D., Kriegman, D., 2007a. Autocalibration via rank-constrained estimation of the absolute quadric. In: Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on. IEEE, pp. 1–8.

Chandraker, M., Agarwal, S., Kriegman, D., Belongie, S., 2007b. Globally optimal affine and metric upgrades in stratified autocalibration. In: 2007 IEEE 11th International Conference on Computer Vision. IEEE, pp. 1–8.

Chatterjee, A., Govindu, V.M., 2018. Robust relative rotation averaging. IEEE Trans. Pattern Anal. Mach. Intell. 40 (4), 958–972. http://dx.doi.org/10.1109/TPAMI.2017.2693984.

Chum, O., Matas, J., 2012. Homography estimation from correspondences of local elliptical features. In: Pattern Recognition (ICPR), 2012 21st International Conference on. IEEE, pp. 3236–3239.

Cui, Z., Tan, P., 2015. Global structure-from-motion by similarity averaging. In: The IEEE International Conference on Computer Vision (ICCV). pp. 864–872.

Dalalyan, A., Keriven, R., 2009. L_1-Penalized robust estimation for a class of inverse problems arising in multiview geometry. In: Advances in Neural Information Processing Systems. pp. 441–449.

Enqvist, O., Kahl, F., Olsson, C., 2011. Non-sequential structure from motion. In: Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on. IEEE, pp. 264–271.

Faugeras, O., Luong, Q.-T., Papadopoulo, T., 2004. The Geometry of Multiple Images: The Laws that Govern the Formation of Multiple Images of a Scene and Some of their Applications. MIT press.

Fischler, M.A., Bolles, R.C., 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun. ACM 24 (6), 381–395.

Furukawa, Y., Ponce, J., 2010. Accurate, dense, and robust multiview stereopsis. IEEE Trans. Pattern Anal. Mach. Intell. 32 (8), 1362–1376.

Gherardi, R., Fusiello, A., 2010. Practical autocalibration. In: Computer Vision–ECCV 2010. Springer, pp. 790–801.

Govindu, V.M., 2001. Combining two-view constraints for motion estimation. In: Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on. 2, IEEE, pp. II–218.

Govindu, V.M., 2004. Lie-Algebraic averaging for globally consistent motion estimation. In: Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on. 1, IEEE, pp. I–684.

Habed, A., Pani Paudel, D., Demonceaux, C., Fofi, D., 2014. Efficient pruning lmi conditions for branch-and-prune rank and chirality-constrained estimation of the dual absolute quadric. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 493–500.

Hartley, R.I., 1997. Kruppa's equations derived from the fundamental matrix. IEEE Trans. Pattern Anal. Mach. Intell. 19 (2), 133–135.

Hartley, R., Aftab, K., Trumpf, J., 2011. L1 rotation averaging using the weiszfeld algorithm. In: Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on. IEEE, pp. 3041–3048.

Hartley, R., Kahl, F., 2007. Optimal algorithms in multiview geometry. In: Computer Vision–ACCV 2007. Springer, pp. 13–34.

Hartley, R., Trumpf, J., Dai, Y., Li, H., 2013. Rotation averaging. Int. J. Comput. Vis. 1–39.

Hartley, R.I., Zisserman, A., 2004. Multiple View Geometry in Computer Vision, second ed. Cambridge University Press.

Jiang, F., Kuang, Y., Solem, J.E., Åström, K., 2014. A minimal solution to relative pose with unknown focal length and radial distortion. In: Asian Conference on Computer Vision. Springer, pp. 443–456.

Kahl, F., 1999. Critical motions and ambiguous euclidean reconstructions in auto-calibration. In: Proceedings of the Seventh IEEE International Conference on Computer Vision, vol. 1. IEEE, pp. 469–475.

Kahl, F., Hartley, R., 2008. Multiple-view geometry under the $\{L\_infty\}$-norm. IEEE Trans. Pattern Anal. Mach. Intell. 30 (9), 1603–1617.

Kazhdan, M., Bolitho, M., Hoppe, H., 2006. Poisson surface reconstruction. In: Proceedings of the Fourth Eurographics Symposium on Geometry Processing.

Kovesi, P.D., 2000. MATLAB and Octave Functions for Computer Vision and Image Processing, Vol. 147. Centre for Exploration Targeting, School of Earth and Environment, The University of Western Australia, p. 230, available from: http://www.csse.uwa.edu.au/~pk/research/matlabfns.

Kukelova, Z., Bujnak, M., Pajdla, T., 2008. Polynomial eigenvalue solutions to the 5-pt and 6-pt relative pose problems. In: BMVC. pp. 1–10.

Lourakis, M.I.A., Argyros, A.A., 2009. SBA: a software package for generic sparse bundle adjustment. ACM Trans. Math. Software 1–30.

Ma, Y., Vidal, R., Košecká, J., Sastry, S., 2000. Kruppa equation revisited: Its renormalization and degeneracy. In: European Conference on Computer Vision. Springer, pp. 561–577.

Martinec, D., Pajdla, T., 2007. Robust rotation and translation estimation in multiview reconstruction. In: Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on. IEEE, pp. 1–8.

Nistér, D., 2004. An efficient solution to the five-point relative pose problem. IEEE Trans. Pattern Anal. Mach. Intell. 26 (6), 756–770.

Olsson, C., Enqvist, O., 2011. Stable structure from motion for unordered image collections. In: Image Analysis. Springer, pp. 524–535.

Olsson, C., Kahl, F., 2010. Generalized convexity in multiple view geometry. J. Math. Imaging Vision 38 (1), 35–51.

Paudel, D.P., Van Gool, L., 2018. Sampling algebraic varieties for robust camera autocalibration. In: European Conference on Computer Vision. Springer, pp. 275–292.

Pernek, Á., Hajder, L., 2013. Automatic focal length estimation as an eigenvalue problem. Pattern Recogn. Lett. 34 (9), 1108–1117.

Pollefeys, M., Koch, R., Van Gool, L., 1999. Self-calibration and metric reconstruction inspite of varying and unknown intrinsic camera parameters. Int. J. Comput. Vis. 32 (1), 7–25.

Rousseeuw, P.J., 1984. Least median of squares regression. J. Amer. Statist. Assoc. 79 (388), 871–880.

Sattler, T., Sweeney, C., Pollefeys, M., 2014. On sampling focal length values to solve the absolute pose problem. In: European Conference on Computer Vision. Springer, pp. 828–843.

Schonberger, J.L., Frahm, J.-M., 2016. Structure-from-motion revisited. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 4104–4113.

Seo, Y., Heyden, A., Cipolla, R., 2001. A linear iterative method for auto-calibration using the dac equation. In: Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on. 1, IEEE, pp. I–880.

Sinha, S.N., Steedly, D., Szeliski, R., 2010. A multi-stage linear approach to structure from motion. In: ECCV 2010 Workshop on Reconstruction and Modeling of Large-Scale 3D Virtual Environments, vol. 3002, pp. 3003–3005.

Snavely, N., Seitz, S.M., Szeliski, R., 2006. Photo tourism: exploring photo collections in 3D. In: ACM Transactions on Graphics (TOG), vol. 25. (3), ACM, pp. 835–846.

Snavely, N., et al., 2010. Bundler: Structure from motion (SFM) for unordered image collections. Available online: phototour. cs. washington. edu/bundler/ (Accessed on 12 July 2013).

Soatto, S., Yezzi, A.J., Jin, H., 2003. Tales of shape and radiance in multi-view stereo. In: Proceedings of the Ninth IEEE International Conference on Computer Vision - Volume 2. In: ICCV '03, IEEE Computer Society, USA, p. 974.

Stewénius, H., Nistér, D., Kahl, F., Schaffalitzky, F., 2005. A minimal solution for relative pose with unknown focal length. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), vol. 2. IEEE, pp. 789–794.
Strecha, C., Fransens, R., Van Gool, L., 2006. Combined depth and outlier estimation in multi-view stereo. In: 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06), vol. 2. IEEE, pp. 2394–2401.
Strecha, C., Von Hansen, W., Van Gool, L., Fua, P., Thoennessen, U., 2008a. On benchmarking camera calibration and multi-view stereo for high resolution imagery. In: Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on. IEEE, pp. 1–8.
Strecha, C., Von Hansen, W., Van Gool, L., Fua, P., Thoennessen, U., 2008b. On benchmarking camera calibration and multi-view stereo for high resolution imagery. In: Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on. IEEE, pp. 1–8.
Sturm, P., 2002. Critical motion sequences for the self-calibration of cameras and stereo systems with variable focal length. Image Vis. Comput. 20 (5–6), 415–426.
Torr, P.H.S., 2002. Bayesian Model estimation and selection for epipolar geometry and generic manifold fitting. Int. J. Comput. Vis. 50 (1), 35–61.
Torr, P.H., Murray, D.W., 1997. The development and comparison of robust methods for estimating the fundamental matrix. Int. J. Comput. Vis. 24 (3), 271–300.
Tron, R., Zhou, X., Daniilidis, K., 2016. A survey on rotation optimization in structure from motion. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). pp. 1032–1040. http://dx.doi.org/10.1109/CVPRW.2016.133.
Wilson, K., Bindel, D., Snavely, N., 2016. When is rotations averaging hard? In: European Conference on Computer Vision. Springer, pp. 255–270.
Wilson, K., Snavely, N., 2014. Robust global translations with 1dsfm. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (Eds.), Computer Vision – ECCV 2014. Springer International Publishing, Cham, pp. 61–75.
Wu, C., 2015. P3. 5p: Pose estimation with unknown focal length. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 2440–2448.
Wu, C., Agarwal, S., Curless, B., Seitz, S.M., 2011. Multicore bundle adjustment. In: Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on. IEEE, pp. 3057–3064.
Zach, C., Pollefeys, M., 2010. Practical methods for convex multi-view reconstruction. In: Computer Vision–ECCV 2010. Springer, pp. 354–367.