A linear method for camera pair self-calibration

Nikos Melanitis, Petros Maragos
National Technical University of Athens, 9, Iroon Polytechniou Str., 15780 Zografos Athens, Greece

A R T I C L E   I N F O

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

A B S T R A C T

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 $\Delta R = 3.49^\circ$) with the standard approach of Kruppa equations followed by 5P algorithm ($\Delta R = 3.77^\circ$). 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) $Q_\infty$ and the Dual Image of the Absolute Conic (DIAC) $\omega_\infty^*$ (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; Snively 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., 2011).
estimate $f$ for a method to average different estimates of a single focal length $f$. To further examine the two solutions recovered by our method, we need to determine $L$ 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 $L_\infty$-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:\n
- 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 DIA2C, 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 $L_\infty$ (e.g. rotation averaging) and $L_\infty$ (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.

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. $v$) 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 $\mathbf{v}_\infty$: $[0\ 0\ 0\ 1]^T$. The absolute conic $\mathbf{Q}_\infty$ is a conic on $\mathbf{v}_\infty$ which satisfies $x_1^2 + x_2^2 + x_3^2 = 1$, $x_4 = 0$, where $X = (x_1\ x_2\ x_3\ x_4)^T$ is the homogeneous representation of world points.

By taking all the planes tangent to $\mathbf{Q}_\infty$, we construct $Q^*_\infty$, which is the dual surface of $\mathbf{Q}_\infty$. $Q^*_\infty$ is described in a metric reconstruction by the $4 \times 4$ matrix

$$Q^*_\infty = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

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 $D^* = PQ^*P^T$.

**Result 2.** If the 3-space is transformed by homography $H$, that is $X' = HX$, then planes of 3-space are transformed according to $\mathbf{x}' = H^T\mathbf{x}$.

**Result 3.** If $H$ is a $4 \times 4$ 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 rank 2 matrix $F$ can be decomposed in two different ways as

$$F = [\hat{a}]_\times A$$

then

$$\hat{a} = \kappa a$$

$$A = \kappa^{-1} (A + \kappa v \times)$$

for some non-zero constant $\kappa$ and 3-vector $v$.

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

We summarize our notation in Table 1. We use subscripts $(1,2,i,j,...)$ 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 $C$ for camera center of projection. In Fig. 1 we visualize camera projection and geometric entities $(Q^*_\infty, \mathbf{v}_\infty, \mathbf{x}_\infty)$ we use to derive our self-calibration and metric reconstruction method.
Fig. 1. Camera geometry. Left: Pinhole camera projection. Let the center of projection $C$ be at the origin of a Euclidean coordinate system and the plane $Z = f$ be the image plane. A world point $X_w$ is projected to the image by a ray joining $C$ with $X_w$. In general, the camera and world coordinate frames are related by a Euclidean (Rotation $R$ and translation $t$) transformation. Right: The dual absolute conic $Q_\infty$ and metric reconstruction.

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

<table>
<thead>
<tr>
<th>Subscripts</th>
<th>Superscripts</th>
<th>Notation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M$</td>
<td>$i$</td>
<td>$P$</td>
<td>Metric Reconstruction, e.g. $P_{ij}$</td>
</tr>
<tr>
<td>$i$</td>
<td>$i$</td>
<td>$P$</td>
<td>Projective Reconstruction, e.g. $P_{ij}$</td>
</tr>
<tr>
<td>GT</td>
<td>$i$</td>
<td>$a$</td>
<td>Ground Truth</td>
</tr>
</tbody>
</table>

**Table 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, Groebner 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 $P, X$, assuming image point correspondences and self-calibrated cameras. The fundamental method to solve SfM is Bundle Adjustment (BA) (Lourakis and Argyros, 2009), 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 $L_\infty$ mvg algorithms (Dalalyan and Keriven, 2009; Hartley and Kahl, 2007; Kahl and Hartley, 2006; 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: $f_i, f_j, R_{ij}, C_i, C_j$. 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 $P_i, P_j$ and further—to simplify the equations and matrices — that $P_i$ coordinate system is aligned with the world coordinate system ($C_i = 0$, $z = f$ 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:

$$K_i = \begin{bmatrix} f_i & 0 & 0 \\ 0 & f_i & 0 \\ 0 & 0 & 1 \end{bmatrix}$$
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 \( \mathbb{P}_1, \mathbb{P}_2 \), which is related to the metric reconstruction by a world (3D) homography \( H \) as in

\[
P_{M1} = \mathbb{P}_1 H
\]

(9)

\[
P_{M2} = \mathbb{P}_2 H
\]

Using Result 1, Eq. (1), we project \( Q^* \) to the image plane of camera 2. For this projection, \( \omega_2^* \) we have:

\[
\begin{bmatrix}
    f_2^* & 0 & 0 \\
    0 & f_2^* & 0 \\
    0 & 0 & 1
\end{bmatrix} = \omega_2^* = \mathbb{P}_2 H Q^*_o H^T P^T
\]

(10)

To introduce the unknowns in Eq. (10), we use the canonical representation of the projective reconstruction, so that \( \mathbb{P}_1 = [I \ 0] \). From Eq. (9), we have for the homography

\[
H = \begin{bmatrix}
    K_1 & 0 \\
    v^T & \sigma
\end{bmatrix}
\]

where \( v \) is yet undetermined and the scale factor \( \sigma \) can be ignored \((\sigma = 1)\).

To fully determine \( H \), we turn to the plane at infinity \( \kappa_{\infty, p} \equiv (p^T 1)^T \). In Eq. (1), we substitute \( \mathbb{P}_1, \mathbb{P}_2 \)

\[
\begin{bmatrix}
    p_1 \\
    p_2 \\
    p_3 \\
    1
\end{bmatrix}
\]

Using Result 2 we arrive at

\[
H = \begin{bmatrix}
    K_1 & 0 & 0 \\
    v^T & \sigma
\end{bmatrix}
\]

(11)

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

\[
\omega_2^* = \mathbb{P}_2 \begin{bmatrix}
    K_1 K_1^T & -K_1 K_1^T p \\
    -p^T K_1 & p^T K_1 K_1^T p
\end{bmatrix} p^T
\]

(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 \( \omega_2^* \) is symmetric by definition, and is also homogeneous, thus it provides five independent equations.

3.2. Linearization

In Eq. (14), we substitute

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

(15)

We define an indeterminate vector \( x_o \) using polynomials in \( f_1, f_2, p_1, p_2, p_3 \): 

\[
x_o \equiv \begin{bmatrix}
    f_1^2 \\
    f_2^2 \\
    f_1^2 p_1^2 + f_2^2 p_2^2 + p_3^2 \\
    f_1^2 p_1 \\
    f_1^2 p_2
\end{bmatrix}
\]

(16)

The augmented matrix \( [A, \ b] \) for the linear system

\[
A x_o = b
\]

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 \( \omega_2^* \). 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 \( x_o \).

3.3. Recovering the solutions

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

\[
A x_o = b
\]

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

\[
\begin{bmatrix}
    1 & 0 & 0 & 0 & 0 & 0 \\
    0 & 1 & 0 & 0 & 0 & 0 \\
    0 & 0 & 1 & 0 & 0 & 0 \\
    0 & 0 & 0 & 1 & c & b_4 \\
    0 & 0 & 0 & 0 & 1 & d
\end{bmatrix}
\]

(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:

\[
P_{P1} = \begin{bmatrix} I \ 0 \end{bmatrix}, \ P_{P2} = [a_1, F \ a]
\]

(21)

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

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

\[
f_1^2 = b_1
\]

(22)

\[
f_2^2 = b_2
\]

(23)

\[
f_1^2 p_1^2 + f_2^2 p_2^2 + p_3^2 = b_3
\]

(24)

\[
p_1 + c f_1^2 p_2 = b_4
\]

(25)

\[
f_1^2 p_1 + d f_2^2 p_1 = b_5
\]

(26)

We substitute \( p_1 \) from (26), and \( p_2 \), 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

\[
\begin{bmatrix}
    f_1^2 p_1^2 + f_2^2 p_2^2 + p_3^2 \\
    f_1^2 p_1 \\
    f_1^2 p_2
\end{bmatrix} = \begin{bmatrix}
    b_3 \\
    b_4 \\
    b_5
\end{bmatrix}
\]

(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 \( \omega_2^* \) elements to write Eq. (17).

Let

\[
\begin{bmatrix}
    P'_{GT1} \ P'_{GT2}
\end{bmatrix} \triangleq \begin{bmatrix} I \ 0 \end{bmatrix}, [\hat{A} \ \hat{a}]
\]

(28)

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

\[
[\mathbb{P}_1, \mathbb{P}_2] \triangleq \begin{bmatrix} I \ 0 \end{bmatrix}, [\hat{A} \ \hat{a}]
\]

(29)
We write the previous equations in matrix form to get the projective transformation $H'$:

$$H' \triangleq \begin{bmatrix} 
\kappa^{-1}e^{-1}I & 0 \\
\kappa^{-1}e^{-1}v^T & \kappa
\end{bmatrix}$$

(36)

$H'$ satisfies

$$\kappa^{-1}e^{-1}P'_{GT} = P_H H'$$

(37)

Now, we get $H$ from $H', H'_k^{-1}$:

$$\begin{bmatrix} 
\kappa^{-1}e^{-1}K_1 & 0 \\
\kappa^{-1}e^{-1}v^TK_1 & \kappa
\end{bmatrix}$$

(38)

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

$$H = \begin{bmatrix} 
\kappa^{-1}e^{-1}K_1 & 0 \\
\kappa^{-1}e^{-1}v^TK_1 & 1
\end{bmatrix}$$

(39)

We propose to repeat step 1, putting camera 2 at the origin $c_f$. In the previous step (1), we have recovered $c_f$ to the left-most $3 \times 3$ block of camera 2 is multiplied by a constant $\mu$.

To see how the constants in Eq. (39) affect Eq. (17), we substitute Eq. (39) in Eq. (10) and get for $\omega_2^*$ the expression

$$\omega_2^* = P_2' \begin{bmatrix} 
(\kappa e)^{-2}K_1 K_1^T & - (\kappa e)^{-2}K_1 K_1^TP p^T \\
-(\kappa e)^{-2}p^TK_1 K_1^T & (\kappa e)^{-2}p^TK_1 K_1^TP p^T
\end{bmatrix} p^T$$

(41)

We encourage to compare the corrected $\omega_2^*$ equation-Eq. (41) to the simplified $\omega_2^*$ equation-Eq. (14).

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

- All equations derived from $\omega_2^*$ elements off the diagonal are of the form $a_{ij} = 0$, thus the constant $(\kappa e)^{-1}$ 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 can redefine $x_2$ (Eq. (16)) and substitute $f_2^*$ with $c f_2^*$ ($c$ 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 $c f_2^*$ ($c$ 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 $\kappa e$ in Eq. (41).
3. Using the homography of Eq. (13) or Eq. (39), we recover the metric reconstruction $P_{M1, M2}$ from Eq. (9). Depending on which homography we have used, one camera matrix ($P_{M1}$) for Eq. (13) or $P_{M1}$ for Eq. (39)) will have the left-most $3 \times 3$ 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 $P_{M1}$ matrices are factored as $P_{M1} = [K_{M1} R_{M1}]$ to recover $R_{M1}, t_{M1}$.
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 visualize 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
\[
\{P^{1}_{m1}, p^{1}_{m1}\}, \{P^{2}_{m2}, p^{2}_{m2}\}
\]
denote the reconstructions derived from Eq. (27). Then, cameras \(P^{1}_{m1}, P^{2}_{m2}\) are in mirror positions with respect to the origin (position of \(P^{1}_{m1}, P^{2}_{m2}\)). The centers of projection \(C^{1}_{m2}, C^{2}_{m2}\) satisfy
\[
C^{1}_{m2} = -C^{2}_{m2}
\] (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 \(v^{1}_{m2}, v^{2}_{m2}\) the viewing directions of \(P^{1}_{m2}, P^{2}_{m2}\) and \(C^{1}_{m2}, C^{2}_{m2}\) the position vectors of the corresponding centers of projection. Then, \(C^{1}_{m2}, C^{2}_{m2}\) bisect the angles formed by \(v^{1}_{m2}, v^{2}_{m2}\), in the plane defined by \(v^{1}_{m2}, v^{2}_{m2}\). Thus, we have:
\[
\angle C^{1}_{m2}, v^{1}_{m2} = \angle C^{1}_{m2}, v^{2}_{m2}
\] (44)
\[
\angle C^{2}_{m2}, v^{1}_{m2} = \angle C^{2}_{m2}, v^{2}_{m2}
\] (45)

From Theorems 1,2, we easily deduce that
\[
\angle C^{1}_{m2}, v^{1}_{m2} + \angle C^{2}_{m2}, v^{2}_{m2} = 180^\circ
\] (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. $f_i$ 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 $\angle v_1 m_2, v_2 m_2$ is, the easier it would be to disambiguate the solutions by the Cheirality condition. From Theorem 2, we have that $\angle v_1 m_2, v_2 m_2 = 2 \angle v_1 m_2, C_1 m_2$ and that $\angle v_1 m_2, C_1 m_2$ 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 length 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
We estimate the mean or median estimate will not correctly determine the solutions. 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 each estimate.

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 $f_i$ estimate. Each disk represents a $f_i$ 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 $f_i$ estimate is paired with one $f_j$ estimate, each $f_i$ estimate originating from a different $i, j$ 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 each $f_i$.

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 matrix $R_i$ (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, $f_i^{n+}$, 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…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 $f_i$ 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 $f_i^n$ in $\beta$% range contributes a different amount to Jcc sum. This amount is proportional to $cc_i^n$ of estimate $f_i^n$ that is paired with $f_j^n$. Good $f_i^n$ 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 $k = 1 … m$ 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 $f_j$ 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 $R^n$. Many different metrics have been defined for rotation matrices (Hartley et al., 2013). We limit our analysis here to

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

(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:

$$
\sum_{i=1}^{n} d_{\text{geo}}(R_i, R_y)
$$

(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 $R_y$ 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 $R_i = R_{ij}R_j$ we get the initial $R_i$ estimates.
2. For every node $i$ in the graph, we use all available estimates $R_{ij}$ to get inconsistent estimates $R_i^k, k = 1, 2, …$ through $R_i = R_{ij}R_j$. 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 $cc^n$ and provide the exact same focal length estimates. However our formulation in Eqs. (18) allows to additionally recover a metric reconstruction.
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 \( \Delta f \) (Chandraker et al., 2007a; Gherardi and Fusiello, 2010; Kukelova et al., 2008):

\[
\Delta f \equiv \left| \frac{1}{f} - 1 \right|
\]  \( (49) \)

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

We observed that our method (Section 3) and Kruppa equations produce identical \( f \) estimates. In rare cases with extremely noise-corrupted correspondences, our method failed to acquire \( f \) estimates at all (20 or more point correspondences, required to acquire a projective reconstruction pipeline). In each of the two methods above, \( f \) 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 (Armanquè and Salvi, 2003), see Armanquè 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 – \( x_{\infty} \) 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 \( \Delta f \) (over all images in the
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 SIM 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 \( \omega \) expression in Eq. (14). However, while our approach solves directly for \( f_i, R_{ij}, p_i, p_j \) and directly imposes the rank degeneracy of the dual absolute quadric, the approach of Pollefeys et al. (1999) solves for \( HQ_*H^T \) (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 \( \Delta f \) and \( \Delta R_{ij} \) errors (\( \Delta R_{ij} \) 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

---

**Table 4**

<table>
<thead>
<tr>
<th>Method</th>
<th>Model (Mean)</th>
<th>Confidence count</th>
<th>Joint confidence count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean ( f ) error</td>
<td>0.28</td>
<td>0.17</td>
<td>0.07</td>
</tr>
</tbody>
</table>

**Table 5**

<table>
<thead>
<tr>
<th>Dataset</th>
<th>( \Delta t ) (°)</th>
<th>( \Delta R ) (°)</th>
<th>( \Delta f )</th>
</tr>
</thead>
<tbody>
<tr>
<td>castle-P30</td>
<td>3.10</td>
<td>1.06</td>
<td>0.0389</td>
</tr>
<tr>
<td>castle-P19</td>
<td>7.35</td>
<td>4.17</td>
<td>0.0586</td>
</tr>
<tr>
<td>entry-P10</td>
<td>4.62</td>
<td>4.67</td>
<td>0.2118</td>
</tr>
<tr>
<td>Herz-Jesse-P8</td>
<td>1.00</td>
<td>0.68</td>
<td>0.0266</td>
</tr>
<tr>
<td>Herz-Jesse-P25</td>
<td>0.41</td>
<td>0.31</td>
<td>0.0049</td>
</tr>
<tr>
<td>Fountain-P11</td>
<td>0.44</td>
<td>0.41</td>
<td>0.0095</td>
</tr>
</tbody>
</table>

**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 \( f \) 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 \( R \) averaging solutions, 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.
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 $P^i_1$: row vector produced from $i$th row of $[a]$, $F$ and permute $x_6$ elements with the permutation

$$4 \leftrightarrow 6$$

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

$$
\begin{bmatrix}
1
\phi_1 P_1^2 \\
\phi_2 P_2^2 + \phi_3 P_3^2 \\
\phi_4 P_4^2 + \phi_5 P_5^2 \\
\end{bmatrix}
$$

(51)

where $r_i$ are $1 \times 3$ vectors and $\psi, \phi$ 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 $A$, and then to apply regular Gaussian elimination. This is generally possible (see below), owing to the structure of $A_{pr}$ rows in (51), which are linear combinations of $P_i^2$ vectors and, also, using the canonical projective reconstruction allows us to substitute

$$P_i^2 = d_1 P_1^2 + d_2 P_2^2$$

(52)

Thus, the elimination of $A_{pr}$ 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 $f_1, f_2, \kappa_\infty$ 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 radiiuses 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 $90^\circ$ (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 $C_P$ has no image, as it is projected to point $0$. Equivalently, $C_P = (C^T 1)^T$ is a right null-vector of $P$.

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

$$P = [KR \ -KR]$$

(53)

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

Result 7. Assume a projection matrix

$$P = [M \ p]$$

Let the vector $m_3^T$ denote the third row of $M$. Then the vector

$$v = \det (M)m_3$$

(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 $P_{m1}^1, P_{m2}^2$ derived from Eq. (27)

Lemma 1. Let

$$P_{m1}^1 = [K_1^2 R_1^1 \ a^1]$$

$$P_{m2}^2 = [K_2^2 R_2^2 \ a^2]$$

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

$$a^1 = a^2 \perp a$$

(57)

Proof. Considering:

1. The form of homography (13)
2. Eq. (9): $P_{m2}^2 = P_{m1}^1 H^r$ where $H^r$ denotes the homography obtained by substituting the $i$th solution of Eq. (27)

the lemma is readily deduced

Lemma 2. Let $P_{m1}^1, P_{m2}^2$ as in Lemma 1. We have:

$$K_1^2 R_1^1 - K_2^2 R_2^2 = an^T$$

(58)

where $n$ is an appropriate vector.

Proof. As in proof of Lemma 1, by observing that $H^r$ for different $i$ values differ only in $v \perp -p^T K$.

We note that Lemma 2 is a general result, independent of (27). However, we omit this proof now.
Lemma 3. For the reconstructions $P_{m_1}^1 = [P_1 \ a]$, $P_{m_2}^1 = [P_2 \ a]$ we have
\[ \det P_1 = \pm \det P_2 \] (59)

Proof. We formed $P_{m_1}^1$, $P_{m_2}^1$ from solutions of Eq. (14), so the projection matrices have the same $o_1^*$. Using now Eq. (1) we have:
\[ o_1^* = P_{m_1}^1 Q_{m_1}^T P_{m_2}^1 \]
\[ = P_1^{-T} \]
\[ = o_2^* \]
\[ = P_2^{-T} \] (60)
Using known properties of determinants we have
\[ \det P_1 P_1^{-T} = \det P_2 P_2^{-T} \]
\[ \Rightarrow \det P_1 = \pm \det P_2 \] (61)

From this last proof, the next Lemma becomes apparent

Lemma 4. Concerning the reconstructions $P_{m_2}^1$, $P_{m_2}^2$ of Lemma 3, we have
\[ K_{m_2}^1 = K_{m_2}^2 \] (62)

Proof. From the equality of $o_1^*$, $o_2^*$ and the diagonality of the internal calibration matrices $K_{m_2}^1$, $K_{m_2}^2$ we prove the Lemma.

In the following we introduce the simplified notation: $K_{m_2} \triangleq K_{m_2}^1$. We next refine Lemma 3, to lift the sign ambiguity in Eq. (59).

Lemma 5. For the reconstructions $P_{m_2}^1$, $P_{m_2}^2$ we have
\[ C_{m_2}^1 = C_{m_2}^2 \triangleq C \iff \mathbf{n}^T \mathbf{c} = 0 \] (63)

Proof. \[ P_{m_2}^1 \left( \begin{array}{c} C \\ 1 \end{array} \right) = 0 \]
\[ \iff \mathbf{a} = -K_2 R_2^1 \mathbf{c} \] (64)
Similarly for $P_{m_2}^2$
\[ P_{m_2}^2 \left( \begin{array}{c} C \\ 1 \end{array} \right) = 0 \]
\[ \iff K_2 R_2^2 \mathbf{c} + \mathbf{a} = 0 \]
\[ \iff K_2 R_2^2 \mathbf{c} + \mathbf{a} = 0 \]
\[ \iff \mathbf{a} = -K_2 R_2^1 \mathbf{c} \]
\[ \iff \mathbf{n}^T \mathbf{c} = 0 \] (65)

Lemma 6. For the reconstructions $P_{m_2}^1$, $P_{m_2}^2$ we have
\[ C_{m_1}^1 = -C_{m_2}^2 \triangleq C \iff \mathbf{n}^T \mathbf{c} = -2 \] (66)

Proof. As in the proof of Lemma 5
\[ P_{m_2}^1 \left( \begin{array}{c} C \\ 1 \end{array} \right) = 0 \]
\[ \iff \mathbf{a} = -K_2 R_2^1 \mathbf{c} \] (67)
Similarly, from matrix $P_{m_2}^2$ we have
\[ P_{m_2}^2 \left( \begin{array}{c} -C \\ 1 \end{array} \right) = 0 \]
\[ \iff -K_2 R_2^2 \mathbf{c} + \mathbf{a} = 0 \]
\[ \iff \mathbf{a} + \mathbf{n}^T \mathbf{c} + \mathbf{a} = 0 \]
\[ \iff \mathbf{a} + \mathbf{n}^T \mathbf{c} = 0 \]
\[ \iff \mathbf{n}^T \mathbf{c} = -2 \] provided $\mathbf{a} \neq 0$ (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 $e$ and row-vector $r$ we have
\[ \det (X + er) = \det X \cdot \det (1 + r X^{-1} e) \] (69)

Lemma 7. For the reconstructions $P_{m_2}^1$, $P_{m_2}^2$ we have
\[ \det P_1 = \det P_2 \iff \mathbf{n}^T \mathbf{c}_1 = 0 \] (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
\[ \det P_2 = \det P_1 \]
\[ \iff 1 - \mathbf{n}^T R_2^1 K_2^{-1} \mathbf{a} = 1 \]
\[ \iff \mathbf{n}^T R_2^1 K_2^{-1} \mathbf{a} = 0 \]
\[ \iff -\mathbf{n}^T R_2^1 K_2^{-1} K_2 R_2^1 \mathbf{c}_1 = 0, \]
from (64): $\mathbf{a} = -K_2 R_2^1 \mathbf{c}_1$
\[ \iff \mathbf{n}^T \mathbf{c}_1 = 0, \]
as $RR^T = I$ for rotation matrices $R$ (71)

Lemma 8. For the reconstructions $P_{m_2}^1$, $P_{m_2}^2$ we have
\[ \det P_2 = -\det P_2 \iff \mathbf{n}^T \mathbf{c}_1 = -2 \] (72)

Proof. As in the proof of Lemma 7, we have:
\[ \det P_2 = -\det P_1 \]
\[ \iff 1 - \mathbf{n}^T R_2^1 K_2^{-1} \mathbf{a} = -1 \]
\[ \iff \mathbf{n}^T R_2^1 K_2^{-1} \mathbf{a} = 2 \]
\[ \iff -\mathbf{n}^T R_2^1 K_2^{-1} K_2 R_2^1 \mathbf{c}_1 = 2, \]
from Eq. (64): $\mathbf{a} = -K_2 R_2^1 \mathbf{c}_1$
\[ \iff \mathbf{n}^T \mathbf{c}_1 = -2, \]
as $RR^T = I$ for rotation matrices $R$ (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 $P_{p_2}$ is in the canonical representation form
\[ \begin{bmatrix} \mathbf{a} \end{bmatrix} \begin{bmatrix} F & \mathbf{a} \end{bmatrix} \]
with $F^T \mathbf{a} = 0$
2. $[\mathbf{a}]$, denotes the anti-symmetric matrix defined to compute outer product with vector $\mathbf{a}$
\[ [\mathbf{a}] \mathbf{v} = \mathbf{a} \times \mathbf{v} \] (75)
3. $e$ denotes the right null vector of $F$,
\[ Fe = 0 \] (76)
Lemma 9. Let

\[ P_{p_2} = [A, a] = [a], F, a] \] (77)

denote the projection matrix for camera \( \varpi \) in the projective reconstruction and \( p, p' \) the solutions for \( \varpi \), acquired from Eq. (27)

\[ p^T = (p_1, p_2, p_3) \] (78)

\[ p'^T = (p'_1, p'_2, p'_3) \] (79)

Then

\[ p - p' = \psi e_t \]

where

\[ e_t = \begin{pmatrix} e_1 \\ e_2 \\ e_3 \end{pmatrix} \] (80)

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

\[ \omega_i = \omega_2 \iff P_{p_2} = P_{p_2} \]

We write

\[ A \equiv \begin{pmatrix} K_1^T & -K_1^Tp \end{pmatrix} \]

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

\[ n^{TC} = 0 \]

To specify \( n \) in Eq. (97), we use

1. The definition of \( n \) in Eq. (58)
2. The relation between \( P_{p_2}, P_{M_2}, H \) (Eqs. (9),(13)) and the notation for \( P \) matrix of Lemma 9

and have

\[ P_1 = AK_1 - ap^T K_1 \]

\[ P_2 = AK_1 - a^T K_1 \]

\[ \psi \]

Now, we can rewrite Eq. (97) as

\[ (p - p')^T K_1 C_1 = 0 \] (99)

We next have

\[ P_{p_2}^T \begin{pmatrix} C_1^T \\ 1 \end{pmatrix} = 0 \iff \]

\[ P_{p_2} H_1 \begin{pmatrix} C_1^T \\ 1 \end{pmatrix} = 0 \iff \]

\[ P_{p_2} \begin{pmatrix} K_1^T C_1 \\ -p^T K_1 C_1 + 1 \end{pmatrix} = 0 \] (100)

From the assumption that \( P_{p_2} \) is in the canonical form (Eq. (74)), it has a null vector (Eq. (76)) that is written as

\[ \begin{pmatrix} e \\ 0 \end{pmatrix} \] (101)

So we have:

\[ P_{p_2} \begin{pmatrix} K_1^T C_1 \\ -p^T K_1 C_1 + 1 \end{pmatrix} = 0 \iff \]

\[ K_1^T C_1 = \psi e \text{, where } \psi \text{ is a constant} \] (102)

\[ -p^T K_1 C_1 + 1 = 0 \] (103)
From Lemma 9 (Eq. (79)) and the previous Eqs. (99), (102) we get:
\[ e_{\text{Te}}^2 = 0 \iff e_t^2 / e_t^1 + e_t^1 / e_t^2 = 0 \] (104)

Since Eq. (104) has no solutions (\( e \neq 0 \)), we produced a contradiction. Thus, from Lemma 3, we have proved that
\[ \det P_2 = - \det P_1 \] (105)

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

**Proof of Theorem 1.** From Lemma 10
\[ \det P_1 = - \det P_2 \] (106)

From Lemmas 6,8 we obtain the equivalent relation
\[ C_{1m_2} = - C_{2m_2}^2 \] (107)

Next, we prove Theorem 2. To avoid a lengthy proof, we settle the coplanarity of \( v_{m_2}^{12}, C_{1m_2}^{12} \) with Lemma 11, which follows the main proof.

Let us first summarize some notation
1. We denote \( v_{m_2}^{12}, v_{m_2}^{21} \) the vectors that point along the viewing directions of cameras \( P_{m_2}, P_{m_2}^2 \), respectively.
2. For \( P_{m_2}^1 \) we assume
\[ \det P_1 > 0 \] (108)
\[ C \triangleq C_{1m_2} \] (109)

**Proof of Theorem 2.** From Results 6, 7, Lemma 1, Theorem 1 we have for \( P_{m_2}^1 \):
\[ K_2 R_1^1 C = -a \iff \begin{cases} f_1 R_1^T \frac{f_1}{f_2} R_2^T \frac{f_2}{R_3^T} \end{cases} C = -a \iff \begin{cases} R_3^2 C = -a_3 \end{cases} \] (110)

We have \( \det P_1 = \det K_2 R_1 > 0 \) and so
\[ v_{m_2}^{21} = R_3 \] (111)

Consequently, from Eq. (110), we have
\[ v_{m_2}^{21}^T C = \| v_{m_2}^{21} || C \| \cos \angle C, v_{m_2}^{21} = -a_3 \] (112)

In Eq. (112),
\[ \| R_3^2 \| = 1 \], since \( R^1 \) is orthogonal as a rotation matrix

We can normalize \( C \) to unitary by satisfying the condition
\[ \| K_2^{-1} a \| = 1 \] (113)

since rotations leave vectors' measure unchanged.

We can now write Eq. (112) as
\[ \cos \angle C, v_{m_2}^{21} = -a_3 \] (114)

Similarly, using
\[ C_{m_2}^{21} = -C \] (115)
\[ \det P^2 < 0 \] (116)
\[ a_1 = a_1 \] (117)

we have
\[ \cos \angle C, v_{m_2}^{12} = -a_3 \] (118)

and the remaining relations required for the proof:
\[ \cos \angle C, v_{m_2}^{12} = -a_3 \] (119)

To complete the proof, we show that \( v_{m_2}^{12}, C_{m_2}^{12} \) are coplanar. We provide a constructive proof in Lemma 11.

**Lemma 11.** There exist rotation matrices \( R_x, R_{perm} \) so that
\[ R_x R_{perm} R_1^{1m_2} = (1 \ 0 \ 0)^T \] (123)
\[ R_x R_{perm} R_2^{1m_2} = (x \ y \ 0)^T \] (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 \( v_{m_2}^{12}, C_{m_2}^{12} \) lie. We first do some hypotheses, without loss of generality, to simplify the notation in the proof:

- Let \( P_{m_2}^1 \) denote the correct representation of \( P_{m_2} \) and \( P_{m_2}^2 \) the erroneous one.
- Let
\[ \text{sign(det } P_1 ) > 0 \] (125)

so that we can simplify the expression for camera viewing direction

We apply to space the rotations
\[ R_x R_{perm} R_1^1 \] (126)

where
\[ R_1^1 : \text{rotation matrix of } P_{m_2}^1 \]
\[ R_{perm} : \text{rotation to transpose } x_1, x_3 \]
\[ \text{of a vector: } (x_1 \ x_2 \ x_3)^T \]
\[ R_x : \text{rotation to place } C^1 \text{ in the desired plane} \]

Applying \( R_1^1 \), using orthogonality of \( R_1^1 \) and Result 7, we have for the viewing direction of camera 2
\[ R_1^1 v_{m_2}^{12} = (0 \ 0 \ 1)^T \] (127)

We then apply \( R_{perm} \), to help with the visualization of this proof
\[ R_{perm} = R_1^1 (90^\circ) = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -1 & 0 & 0 \end{bmatrix} \] (128)

We define \( R_x \), a rotation around \( x \)-axis, to place \( C^1 \) on \( z \)-plane and at the same time leave \( v_{m_2}^{12} \) unchanged. We have
\[ R_x = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \theta_x & -\sin \theta_x \\ 0 & \sin \theta_x & \cos \theta_x \end{bmatrix} \] (129)

We transformed \( C^1 \) to:
\[ K_2 R_1^1 C^1 = -a \iff R_1^1 C^1 = -K_2^{-1} a \iff \begin{cases} R_{perm} R_1^1 C^1 = -R_{perm} (-K_2^{-1}) a = \begin{bmatrix} -a_3 \\ -f_2 a_1 \\ f_2 a_3 \end{bmatrix} \end{cases} \] (130)

Then, applying \( R_x \) we have:
\[ R_x R_{perm} R_1^1 C^1 = R_x \begin{bmatrix} -a_3 \\ -f_2 a_1 \\ f_2 a_3 \end{bmatrix} \]


