Simplified Mirror-Based Camera Pose Computation via Rotation Averaging

Gucan Long[*1], Laurent Kneip[2], Xin Li[1], Xiaohu Zhang[1], and Qifeng Yu[1]

[1]College of Aerospace Science and Engineering
National University of Defense Technology, P.R. China
[2]Research School of Engineering, Australian National University

Abstract

We propose a novel approach to compute the camera pose with respect to a reference object given only mirrored views. The latter originate from a planar mirror at different unknown poses. This problem is highly relevant in several extrinsic camera calibration scenarios, where the camera cannot see the reference object directly. In contrast to numerous existing methods, our approach does not employ the fixed axis rotation constraint, but represents a more elegant formulation as a rotation averaging problem. Our theoretical contribution extends the applicability of rotation averaging to a more general case, and enables mirror-based pose estimation in closed-form under the chordal L2-metric, or in an outlier-robust way by employing iterative L1-norm averaging. We demonstrate the advantages of our approach on both synthetic and real data, and show how the method can be applied to calibrate the non-overlapping pair of cameras of a common smart phone.
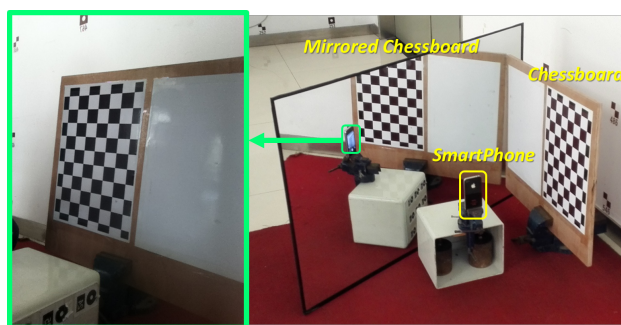
Figure 1. Extrinsic calibration of a smart phone's camera-pair: The green rectangle contains a mirrored scene captured by the smart phone's front camera. As can be observed in the right figure, both front and back camera can see the chessboard only via mirror reflections. Mirror based pose estimation consists of finding both cameras' absolute pose with respect to the target object given only (at least 3) mirrored views of the object (i.e. the mirror's position is changed and unknown). By combining the absolute poses, we can then easily retrieve the relative pose between the non-overlapping pair of cameras.

1. Introduction

The present paper discusses the calibration of extrinsic camera parameters, meaning a Euclidean transformation that defines the camera position and orientation with respect to another frame. For instance, in virtual reality we ought to know the transformation between a camera and a reference frame, thus enabling virtual objects to be added onto real scene video. While this can be recovered by well-known relative or absolute pose computation algorithms, there remain several important applications where the pose computation can no longer rely on basic geometric constraints:

- A camera that is custom-mounted in the front or back of a car. The corresponding calibration task is commonly denoted eye-body calibration. Despite the fact that a 3D model of the vehicle might be readily available, it is not visible inside the camera's field of view, thus blocking a calibration based on object pose estimation. The problem is typically solved by computing the egomotion of the camera and aligning it with dynamic sensor information, for instance given by an odometer or an inertial measurement unit.

- Multiple cameras pointing into different directions such that their fields of view are no longer overlapping. Deriving the relative pose from the absolute poses with respect to a commonly observed calibration target hence becomes impossible. The problem of calibrating non-overlapping cameras is often approached by again computing the trajectory of each camera under motion, and then aligning the trajectories by solving the so-called hand-eye calibration problem.

---
[*]*gucan.long@nudt.edu.cn*

A daily-life example of a non-overlapping camera pair is given by smart-phone devices equipped with a front-back camera pair (cf. Figure 1).

Many cases deny the above calibration methods because they constrain the camera motion and thus render certain parameters unobservable (e.g. a car exerting planar motion, or a static setup such as surveillance cameras). An important stream of extrinsic camera calibration therefore is given by employing mirrors in the calibration process, allowing (co-)visibility of known objects or calibration targets. Mirror-based camera pose computation consists of using planar mirrors to observe the reflections of the reference object and then computing the extrinsic parameters along with the location and orientation of planar mirrors (cf. Figure 1). It is the main subject of this paper.

Most of the existing solutions to the problem are derived from the *fixed-axis rotation constraint* discovered in [5]. This work proves that the relative transformation between two mirrored views is a pure rotation about the intersection line of the corresponding mirror planes. Various forms of this constraint [14, 8, 15, 13] allow us to recover the direction vectors of intersection lines, from which the mirror and camera poses can then be derived.

The present paper introduces a novel and certainly more elegant solution to the problem: We show that the combination of an arbitrary number of mirrored images of a calibration target—seen from a camera with constant pose within a reference frame—turns out to be a generalization of the rotation averaging problem introduced in [7]. This notably allows us to solve for all *virtual poses*[1] individually, and then combine them within a single fusion stage. Extended from [7], our rotation averaging formalism allows us to devise different strategies:

- A closed-form solution based on singular value decomposition of the sum of all virtual cameras' rotation matrices. This formulation minimizes the chordal L2-norm of the rotation residuals, resulting in state-of-the-art noise resilience under elegant linear complexity.

- L1-averaging as presented in [7]. This variant is able to transparently handle real world situations where part of the mirrored images are captured wrongly, and without depending on Ransac [4].

The paper is organized as follows. Section 2 describes related prior techniques. Section 3 briefly introduces the geometry of planar mirror reflections. Section 4 introduces our rotation-averaging approach. Section 5 shows comparative results on both synthetic and real world data. We furthermore show a successful application of the approach

---

[1] A virtual pose is the pose of a (left-handed) camera frame that returns the same observation than a real camera, without employing a mirror.

to find the extrinsic parameters of a smart phone's non-overlapping pair of cameras. Section 6 concludes the work.

2. Related work

The work of Sturm and Bonfort [14] is the first to introduce the problem of estimating the pose of an object in the absence of a direct view (i.e. using mirror reflections). It proves that a unique determination of mirror positions and camera pose requires at least 3 virtual views, and discusses singular configurations. The approach is constructive in that—for a certain virtual view—it first computes the fixed-axes of rotation in combination with the other two virtual views in closed form. It then determines the mirror position by fitting a plane to the estimated 3D lines, and finally recovers the real camera pose by reflecting the virtual view with respect to the mirror. The method works for non-parallel mirror planes exclusively, thus ensuring that the intersection of fixed-axes of rotation is feasible.

Rodrigues et al. [13] build on top of this result by presenting a linear reformulation of the fixed-axis rotation constraint. The method leads to improved performance over [14] since it utilizes an arbitrary number of fixed-axis constraints for each mirror pose, and computes the corresponding mirror plane parameters in a single shot rather than recovering it from 3D lines. It also enables a transparent handling of parallel mirror planes. Hesch et al. [8] and Takahashi et al. [15] study further minimal solutions to the problem that exploit the relative pose between virtual cameras. The latter work in particular uses the 3D points directly rather than the relative motion parameters, thus providing an orthogonality constraint that is more beneficial for estimating the orientation of the mirror. Kumar et al. [10] provide a linear estimator that does not rely on the fixed-axis rotation constraint. However, the algorithm requires at least five mirrored views and—as typical for highly linearized solvers—the resulting algebraic error metric leads to reduced noise resilience. More recently, Agrawal and Ramalingam [1] present an approach that utilizes spherical mirror reflections, thus avoiding the problem of parallel mirror configurations. However, planar mirrors are much easier to find, thus ensuring that planar mirror based pose estimation remains a practically relevant problem.

In conclusion, all solutions to mirror-based pose estimation (except the less accurate one presented in [10]) rely on the fixed-axis rotation constraint, and thus utilize at least pair-wise combinations of mirror poses. This leads to quadratic complexity. In contrast, we show that mirror-based pose estimation can be solved in an easier way by first finding all virtual poses individually, and then the camera orientation through a rotation-averaging step.

3. Geometry of planar mirror based pose estimation

This section summarizes a number of geometric concepts substantial to mirror-based camera pose estimation. We will see the notations used throughout this paper along with the basic geometric formulation of the problem. The section then goes into further geometric details, such as the algebraic formulation of planar mirror reflections, the modified camera projection model that takes mirror reflections into account, and finally the concept of a *virtual camera* and what we call an *improper rotation matrix*.

3.1. Problem formulation

The geometry of the problem is described in Figure 2. Let $\mathcal{F}_c$ be a camera frame without a direct view onto 3D points $\{\mathbf{p}_1, \ldots, \mathbf{p}_m\}$ defined inside a reference frame $\mathcal{F}_r$. Let $\{\pi_1, \ldots, \pi_n\}$ be different mirror positions rendering the 3D points visible inside the camera's field of view. Let $\mathbf{T}$ describe the pose of the camera with respect to $\mathcal{F}_r$. Mirror-based camera pose estimation consists of computing $\mathbf{T}$ while the coordinates of the 3D points are known, and the parameters of all π_i are unknown.

3.2. Planar mirror reflections

A planar mirror can be described by the plane parameters $\pi = \{\mathbf{n}, d\}$ within a given reference frame. The unit vector $\mathbf{n}$ denotes the normal vector of the mirror plane, and d represents the orthogonal distance between the origin and the plane. A point $\mathbf{x}$ lies on the plane if it satisfies the constraint $\mathbf{n}^T \mathbf{x} = d$. The relation between the point $\mathbf{p}$ and its mirrored point $\tilde{\mathbf{p}}$ reflected by π is given by

$$\begin{bmatrix} \tilde{\mathbf{p}} \\ 1 \end{bmatrix} = \mathbf{S} \begin{bmatrix} \mathbf{p} \\ 1 \end{bmatrix}, \text{ where } \mathbf{S} = \begin{bmatrix} \mathbf{I} - 2\mathbf{nn}^T & 2d\mathbf{n} \\ \mathbf{0} & 1 \end{bmatrix} \quad (1)$$

denotes the symmetric transformation induced by π. Note that $\mathbf{S} = \mathbf{S}^{-1}$, and $(\mathbf{I} - 2\mathbf{nn}^T)$ is a Householder matrix.

3.3. Mirrored camera projection model

The rigid transformation that transforms points from the reference to the camera frame is given by $\mathbf{T} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0} & 1 \end{bmatrix}$, where $\mathbf{t}$ denotes the origin expressed inside the camera frame, and $\mathbf{R}$ denotes the rotation that rotates points from the reference into the camera frame. We ignore camera intrinsics and adopt a normalized perspective camera model where points are projected onto the plane $z = 1$. Taking the mirror reflection into account, the 3D point corresponding to a normalized image point actually corresponds to its mirrored point. Concatenating the camera model with the mirror reflection, the mirrored camera projection model becomes

$$\mathbf{v} \sim \begin{bmatrix} \mathbf{I} & \mathbf{0} \end{bmatrix} \mathbf{T} \begin{bmatrix} \tilde{\mathbf{P}} \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \end{bmatrix} \mathbf{TS} \begin{bmatrix} \mathbf{P} \\ 1 \end{bmatrix}, \quad (2)$$
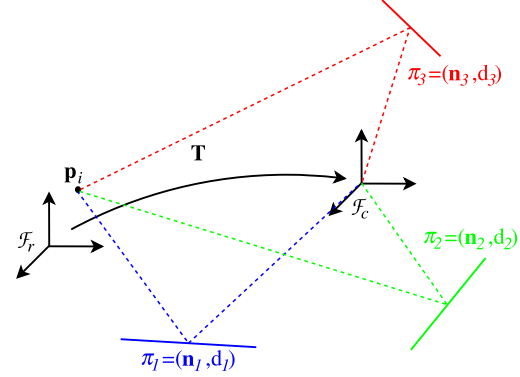

Figure 2. The geometry of the mirror-based camera pose estimation problem.

where $\mathbf{v}$ denotes a normalized image point expressed in homogeneous coordinates, and $\sim$ means equality up to a non-zero scale factor.

3.4. Virtual cameras and improper rotations

Let $\tilde{\mathbf{T}} = \mathbf{TS}$. The transformation $\tilde{\mathbf{T}}$ includes a reflection and a rigid transformation, and is given by

$$\tilde{\mathbf{T}} = \begin{bmatrix} \tilde{\mathbf{R}} & \tilde{\mathbf{t}} \\ \mathbf{0} & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{R}(\mathbf{I} - 2\mathbf{nn}^T) & 2d\mathbf{Rn} + \mathbf{t} \\ \mathbf{0} & 1 \end{bmatrix} \quad (3)$$

The projection of a mirrored point by a real camera can be regarded as the projection of a real point by a mirrored camera. While $\mathbf{T}$ denotes the pose of the real camera, $\tilde{\mathbf{T}}$ denotes the pose of the mirrored camera, which is also called a *virtual camera*.

$\tilde{\mathbf{R}}$ is the reflection of $\mathbf{R}$ by the normal vector $\mathbf{n}$ and denotes an improper rotation matrix satisfying $\tilde{\mathbf{R}}^T\tilde{\mathbf{R}} = \tilde{\mathbf{R}}\tilde{\mathbf{R}}^T = \mathbf{I}$ and $\det(\tilde{\mathbf{R}}) = -1$. As introduced in [3], improper rotation matrices hold a number of additional properties that are key to our rotation averaging formulation:

- (a) Improper rotation matrices must have one eigenvalue equal to -1. This property is key to our closed-form chordal L2 algorithm. The proof is given in [12].

- (b) Improper rotation matrices can be uniquely decomposed into a product of a planar reflection and a rotation around the corresponding normal vector. The decomposition is given by

$$\tilde{\mathbf{R}} = \mathbf{R}(\mathbf{e}, \theta)(\mathbf{I} - 2\mathbf{ee}^T), \quad (4)$$

where $\mathbf{e}$ is the eigenvector of $\tilde{\mathbf{R}}$ corresponding to the eigenvalue of -1, and $\mathbf{R}(\mathbf{e}, \theta)$ is a rigid rotation by an angle θ around the axis $\mathbf{e}$, with $\theta = \cos^{-1}\left(\frac{tr(\tilde{\mathbf{R}})+1}{2}\right)$. The proof can be found in [3].

- (c) Property (b) further means that there does not exist a rotation axis for which the decomposition would lead

to a smaller rotation angle about the chosen axis of $\mathbf{e}$. The proof consists of first multiplying both sides of (4) by -1, thus turning improper rotations into proper ones: $-\tilde{\mathbf{R}} = \mathbf{R}(\theta, \mathbf{e})(2\mathbf{e}\mathbf{e}^T - \mathbf{I})$. $(2\mathbf{e}\mathbf{e}^T - \mathbf{I})$ can be regarded as a rotation by π about the axis $\mathbf{e}$. We obtain

$$
\begin{aligned}
-\tilde{\mathbf{R}} &= \mathbf{R}(\theta, \mathbf{e})(2\mathbf{e}\mathbf{e}^T - \mathbf{I}) = \mathbf{R}(\theta, \mathbf{e})\mathbf{R}(\pi, \mathbf{e}) \\
&= \mathbf{R}(\theta + \pi, \mathbf{e}).
\end{aligned} \tag{5}
$$

$(\theta + \pi, \mathbf{e})$ is an axis-angle representation of the rotation $-\tilde{\mathbf{R}}$. From Euler's rotation theorem, we know that axis and angle are unique. This in turn means that it is impossible to obtain a smaller rotation angle than θ.[2]

4. Rotation-averaging for mirror-based pose estimation

The present section describes how rotation averaging can be extended from proper to improper rotations. It then concludes with practical solutions to the problem formulation, namely a closed-form solution finding the L2-mean based on the chordal distance between rotation matrices, and an iterative L1-norm minimization scheme. Both schemes are effective in solving mirror-based pose estimation, a problem in which improper rotation matrices naturally appear.

4.1. Rotation-averaging with improper rotation matrices

Let $\left\{ \tilde{\mathbf{T}}_1, \ldots, \tilde{\mathbf{T}}_n \right\}$ be a set of (improper) transformation matrices obtained by a modified absolute pose algorithm applied to the mirrored projections of known 3D world points defined in a certain reference frame. Let the mirror pose be the only difference between the various mirrored views. The problem of mirror-based absolute pose estimation consists of finding the pose $\mathbf{T}$ of the camera directly with respect to the reference frame, as well as the plane parameters $\{\pi_1, \ldots, \pi_n\}$ of every mirror pose. We first solve a sub-problem in which we only seek the proper absolute rotation $\mathbf{R}$ as well as the mirror plane normal vectors $\mathbf{n}_i$ such that they minimize the cost function

$$
C(\mathbf{R}, \mathbf{n}_i) = \sum_i^n \varepsilon(\tilde{\mathbf{R}}_i(\mathbf{I} - 2\mathbf{n}_i\mathbf{n}_i^T), \mathbf{R})^p, \tag{6}
$$

where ε represents the geometric, chordal, or quaternion metric, and $p = 1$ or $p = 2$.

The problem has similarities with the rotation averaging formulation of [7]. The difference lies in the existence of improper rotation matrices generated by mirror reflections that are expressed by the Householder matrices $(\mathbf{I} - 2\mathbf{n}_i\mathbf{n}_i^T)$. The additional unknown planar normal vectors $\mathbf{n}_i$ render

the problem a more difficult one compared to [7], which is based on proper rotation matrices only. However, as shown in the remainder of this section, the exploitation of the properties of improper rotation matrices still allows us to solve the problem in similar ways.

4.2. Chordal L2-mean algorithm

We first consider rotation averaging under the L2 chordal metric ($p = 2$) which—in analogy to the original case with proper rotation matrices—leads to a closed-form solution for finding the global minimum of the objective function[3]. (6) will notably appear as

$$
E(\mathbf{R}, \mathbf{n}_i) = \sum_{i=1}^n \|\tilde{\mathbf{R}}_i(\mathbf{I} - 2\mathbf{n}_i\mathbf{n}_i^T) - \mathbf{R}\|_F^2. \tag{7}
$$

If $\langle \cdot, \cdot \rangle$ represents the Frobenius inner product— i.e. the sum of the element-wise products of two matrices—, (7) can be rewritten as

$$
\begin{aligned}
E(\mathbf{R}, \mathbf{n}_i) &= \sum_{i=1}^n \left\langle \tilde{\mathbf{R}}_i(\mathbf{I} - 2\mathbf{n}_i\mathbf{n}_i^T) - \mathbf{R}, \tilde{\mathbf{R}}_i(\mathbf{I} - 2\mathbf{n}_i\mathbf{n}_i^T) - \mathbf{R} \right\rangle \\
&= \sum_{i=1}^n \Big(\left\langle \tilde{\mathbf{R}}_i(\mathbf{I} - 2\mathbf{n}_i\mathbf{n}_i^T), \tilde{\mathbf{R}}_i(\mathbf{I} - 2\mathbf{n}_i\mathbf{n}_i^T) \right\rangle \\
&\quad -2 \left\langle \tilde{\mathbf{R}}_i(\mathbf{I} - 2\mathbf{n}_i\mathbf{n}_i^T), \mathbf{R} \right\rangle + \langle \mathbf{R}, \mathbf{R} \rangle \Big) \\
&= 6n - 2\sum_{i=1}^n \left\langle \tilde{\mathbf{R}}_i(\mathbf{I} - 2\mathbf{n}_i\mathbf{n}_i^T), \mathbf{R} \right\rangle \\
&= 6n - 2\sum_{i=1}^n \left\langle \tilde{\mathbf{R}}_i, \mathbf{R} \right\rangle + 4\sum_{i=1}^n \left\langle \tilde{\mathbf{R}}_i\mathbf{n}_i\mathbf{n}_i^T, \mathbf{R} \right\rangle \quad (8)
\end{aligned}
$$

Let

$$
E_1 = \sum_{i=1}^n \left\langle \tilde{\mathbf{R}}_i, \mathbf{R} \right\rangle = \left\langle \sum_{i=1}^n \tilde{\mathbf{R}}_i, \mathbf{R} \right\rangle \tag{9}
$$

$$
E_2 = \sum_{i=1}^n \left\langle \tilde{\mathbf{R}}_i\mathbf{n}_i\mathbf{n}_i^T, \mathbf{R} \right\rangle = \sum_{i=1}^n \mathbf{n}_i^T \tilde{\mathbf{R}}_i^T \mathbf{R}\mathbf{n}_i. \tag{10}
$$

It is easy to see that minimizing E is equivalent to a simultaneous maximization of E_1 and minimization of E_2 over $\mathbf{R}$ and all $\mathbf{n}_i$'s.

Minimization of E_2: Consider that $\mathbf{n}_i$ is a unitary vector and $\tilde{\mathbf{R}}_i^T\mathbf{R}$ must be an improper rotation matrix with respect to any true rotation $\mathbf{R}$. It is obvious that the minimum value of $\mathbf{n}_i^T\tilde{\mathbf{R}}_i^T\mathbf{R}\mathbf{n}_i$ is -1: Recall property (a) in Section 3.4, stating that an improper rotation $\tilde{\mathbf{R}}_i^T\mathbf{R}$ must have an eigenvalue -1. This means that $\mathbf{n}_i^T\tilde{\mathbf{R}}_i^T\mathbf{R}\mathbf{n}_i$ will always reach its minimum value of -1 if $\mathbf{n}_i$ is computed as the eigenvector corresponding to the eigenvalue -1 of $\tilde{\mathbf{R}}_i^T\mathbf{R}$. This statement is true for any rotation $\mathbf{R}$, and the problem is reduced to maximizing E_1.

---

[2]These statements are true if bounding the rotation angle to the interval $[0, \pi)$, and ignoring the special case where it is 0 anyway.

[3]The induced metrics of the chordal and the geodesic distance differ only by a scale factor of $\sqrt{2}$. For small residuals, the chordal L2-norm is very close to the geodesic L2-norm. For large residuals (up to and including the maximum residual), the chordal L2-norm can be regarded as a robust alternative to the geodesic L2-norm. Please refer to [7] for further details.

Maximization of E_1: From the trace-maximization problem[9], it is well known that

$$\operatorname*{argmax}_{\mathbf{R}\in SO(3)}\langle\mathbf{G},\mathbf{R}\rangle = \operatorname*{argmin}_{\mathbf{R}\in SO(3)}\|\mathbf{G}-\mathbf{R}\|_F^2. \quad (11)$$

Maximizing E_1—and thus minimizing (7)—therefore is equivalent to finding the rotation $\mathbf{R}$ that is closest to $\mathbf{G}=\sum_{i=1}^n\tilde{\mathbf{R}}_i$ under the Frobenius norm, which indeed—as shown in [9]—has the closed-form solution

$$\hat{\mathbf{R}} = \mathbf{U}\mathbf{S}\mathbf{V}^T, \quad (12)$$

where $\mathbf{U}$ and $\mathbf{V}$ are given by the singular value decomposition $\text{SVD}(\mathbf{G})=\mathbf{U}\Sigma\mathbf{V}^T$, and $\mathbf{S}=\text{diag}(1,1,\det(\mathbf{U}\mathbf{V}^T))$.

Uniqueness conditions of the solution: The problem of rotation averaging with improper rotations needs at least 3 rotations to find a unique solution. This stands in contrast with the original rotation averaging with proper rotations which only needs one rotation. More specifically, let $\mathbf{F}=[\mathbf{n}_1\ \ \mathbf{n}_2\ \ \cdots\ \ \mathbf{n}_n]$. The presented chordal L2-mean algorithm can obtain a unique solution *iff* $rank(\mathbf{F})=3$. The detailed proof is provided in Appendix A. Degenerate configurations have already been observed by both Sturm et al. [14] and Rodrigues et al. [13]. Our condition means that the presented L2-averaging algorithm will only work if the normal vectors of the mirror planes are not coplanar.

4.3. Geometric L1-mean algorithm

We now proceed to the derviations of our L1 averaging scheme. The goal is to reduce the geometric distances (i.e. geodesic distance) between $\mathbf{R}$ and $\tilde{\mathbf{R}}_i(\mathbf{I}-2\mathbf{n}_i\mathbf{n}_i^T)$. This in turn means that $\mathbf{R}^T\tilde{\mathbf{R}}_i(\mathbf{I}-2\mathbf{n}_i\mathbf{n}_i^T)$ needs to be close to identity, and indeed represents our residual. $\mathbf{R}^T\tilde{\mathbf{R}}_i$ is an improper rotation matrix that differs from the proper residual rotation only by the reflection $(\mathbf{I}-2\mathbf{n}_i\mathbf{n}_i^T)$. According to property (c) in Section 3.4, $\mathbf{n}_i$ can be obtained by eigendecomposition of $\mathbf{R}^T\tilde{\mathbf{R}}_i$, and this $\mathbf{n}_i$ is optimal in the sense of characterizing the reflection that leads to a minimal residual rotation. Note that $\mathbf{n}$ is optimal regardless of ε and p.

The entire algorithm is inspired by the practical solution to L1-mean averaging presented in [7]. Each iteration proceeds by deriving the sum of axes of residual rotations $\mathbf{r}$ obtained via the log map. This is followed by an update of $\mathbf{R}$: We seek an ideal rotation angle $s^\star$ about $\mathbf{r}$, and thus perform a 1 dimensional search for an optimal correction. The cost function is evaluated with $p=1$, ensuring L1-norm averaging. After each update, all $\mathbf{n}_i$'s have to be recomputed via eigen-decomposition. Algorithm 1 provides a description in form of pseudo-code. For a convergence analysis of L1 rotation averaging, the reader is referred to [7].

4.4. Recovering camera position and mirror distances

Looking at the translational part of (3), we can easily observe that $\tilde{\mathbf{t}}_i = 2d_i\mathbf{R}\mathbf{n}_i + \mathbf{t}$. After rotation averaging,

---

Input: $\tilde{\mathbf{R}}_i, \mathbf{R}_{\text{initial}}, \epsilon > 0$
Result: Optimal $\mathbf{R}$
$\mathbf{R} = \mathbf{R}_{\text{initial}}$;
repeat

$\quad \mathbf{n}_i = \mathbf{V}(:,c)$, with $\begin{cases} [\mathbf{V},\mathbf{D}] = \text{eig}(\mathbf{R}^T\tilde{\mathbf{R}}_i) \\ c \text{ such that } \mathbf{D}[c] = -1 \end{cases}$;

$\quad \mathbf{r} = \sum_{i=1}^n \frac{\log(\mathbf{R}^T\tilde{\mathbf{R}}_i(\mathbf{I}-2\mathbf{n}_i\mathbf{n}_i^T))}{\|\log(\mathbf{R}^T\tilde{\mathbf{R}}_i(\mathbf{I}-2\mathbf{n}_i\mathbf{n}_i^T))\|}$;

$\quad s^* = \operatorname*{argmin}_{s\geq 0} C_{p=1}(\mathbf{R}\exp(s^*\mathbf{r}))$;

$\quad \mathbf{R} = \mathbf{R}\exp(s^*r)$;
until $\|s^*\mathbf{r} < \epsilon\|$;

Algorithm 1: Practical solution to L1-mean averaging presented in [7] and adapted to the case of improper rotations.

only $\mathbf{t}$ and all d_i's remain unknown. It is easy to recognize that the separation

$$\begin{bmatrix} 2\mathbf{R}\mathbf{n}_1 & & & \mathbf{I} \\ & \cdot & & \cdot \\ & & \cdot & \cdot \\ & & 2\mathbf{R}\mathbf{n}_n & \mathbf{I} \end{bmatrix}\begin{bmatrix} d_1 \\ \cdot \\ \cdot \\ d_n \\ \mathbf{t} \end{bmatrix} = \begin{bmatrix} \tilde{\mathbf{t}}_1 \\ \cdot \\ \cdot \\ \tilde{\mathbf{t}}_n \end{bmatrix} \quad (13)$$

allows for a linear computation of $\mathbf{t}$ and all d_i. The solution can be found in linear time by computing the pseudo-inverse using the Schur-complement trick and block-wise matrix inversion.

5. Experimental evaluation

This section presents both synthesized and real experiments to demonstrate the advantages of the proposed rotation averaging solutions with respect to alternative algorithms.

5.1. Performance in terms of accuracy

We first compare our chordal L2 averaging method with the state-of-the-art closed-form solutions presented in [14, 15, 13]. The simulation experiments are generated as follows: We assume a camera with an image size of 1000×1000 pixels, and a field of view of 45 degrees. 3D reference points are randomly generated within the volume $[-25, 25] \times [-25, 25] \times [-25, 25]$. A set of virtual mirror planes are randomly generated ensuring that all 3D reference points can be observed by the camera. For every mirrored image, the corresponding virtual pose is computed by using the modified left-handed PnP algorithm [11][4].

We compare our algorithms against three other methods, which are given by the algorithms of Sturm [14], Ro-

---

[4]Any PnP algorithm can be employed: Let [2D-points, -3D-points] be the input of the PnP algorithm, and $[\mathbf{t}, \mathbf{R}]$ be the result. The corresponding virtual pose is then given by $[\mathbf{t}, -\mathbf{R}]$
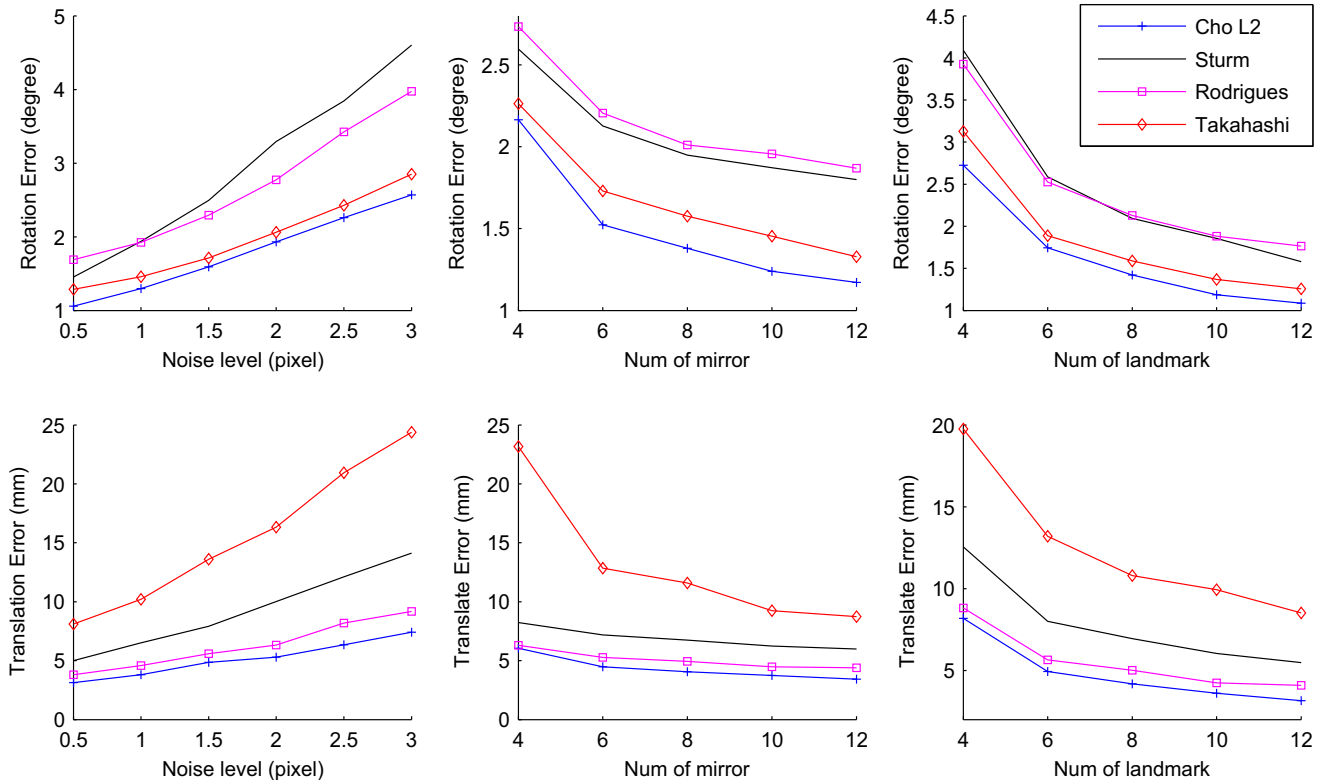
Figure 3. Rotation and translation errors in synthesized experiments on accuracy. Each point represents median values over 1000 trials. **First column:** Median error in terms of varying pixel noise. The number of mirrors and landmarks is 9. **Second column:** Median error in terms of different numbers of mirrors. The pixel noise level is 1.0, and the number of landmarks is 9. **Third column:** Median error in terms of different numbers of landmarks. The pixel noise level is 1.0, and the number of mirrors is 9.

drigues[5] [13], and Takahashi [15]. We did not include the algorithms presented by Kumar et al. [10] and Hesch et al. [8] since their inferior accuracy has already been demonstrated in previous works [13, 15]. Note that Takahashi's original algorithm [15] served as a minimal solver to the problem of mirror-based pose estimation. We modified their published implementation such that it can handle the case of multiple mirrors and points. The modification is simply based on encoding more mirrors and points within their large system of linear equations.

The performance of the algorithms is evaluated in terms of different levels of pixel noise, different numbers of mirrors, and different numbers of landmarks. For each test the simulation performs 1000 runs and returns the median error. The results are presented in Figure 3.

Sturm's algorithm is more accurate than Rodrigues' in case the virtual cameras' poses have small errors (i.e. they are estimated with less pixel noise or more landmarks), which confirms the results in [13]. Most importantly, we can see that the presented chordal L2 rotation averaging

algorithm consistently outperforms the state-of-the-art. It shows highest accuracy for all analyzed noise levels and numbers of mirrors and landmarks. We also compared estimates of the mirror plane parameters, and our algorithm outperforms equally well in this regard.

5.2. Performance in terms of robustness

This section evaluates the performance of the presented geometric L1 averaging method in terms of robustness. We only compare it against the chordal L2 averaging method, directly applied to simulated improper rotation matrices. To setup our experiments, we first generate a random axis-angle rotation $\mathbf{r}$, and translate it into our proper rotation matrix $\mathbf{R}$. We then generate a group of random mirror planes and derive the corresponding improper rotations $\tilde{\mathbf{R}}_i$ based on (3). Noise is added by left-multiplying each improper rotation with a rotation sampled from another normally distributed axis-angle rotation. Outliers are simulated similarly, however using normally distributed noise with a standard deviation of 50 degrees. We compare the L1 and L2 averaging algorithms under different noise conditions and different outlier levels. Note that the L1 averaging is an

---
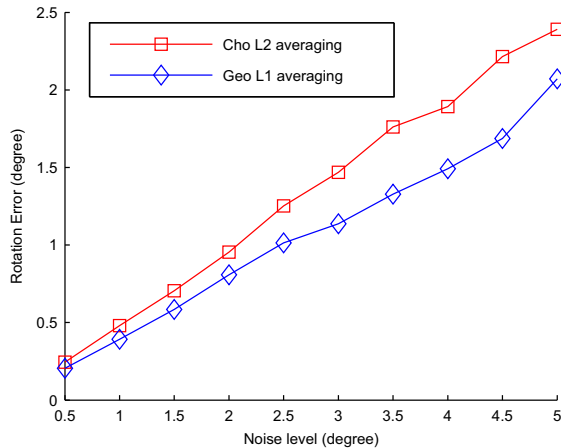[5]We adopt method 2 from their paper.

Figure 4. The median errors of the geometric L1 averaging algorithm compared with the chordal L2 averaging approach for varying noise parameters. The total number of averaged improper rotations is 20. The noise level represents the standard deviation of the angle of the disturbance rotation. Each point represents median values over 1000 trials.
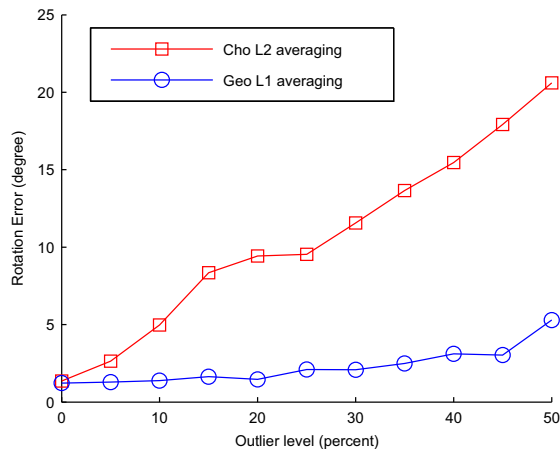

Figure 5. The median errors of geometric L1 averaging compared with chordal L2 averaging for varying levels of outliers. The total number of averaged improper rotations is 20. Outliers are simulated by sampling rotation matrices with normally distributed angles having a standard deviation of 50 degrees (noise: 3 degrees std. dev.). Each point represents the median over 1000 trials.

iterative algorithm initialized with the output of L2 averaging. For each test the simulation performs 1000 runs and the median error is recorded.

The results are shown in Figures 4 and 5. It can be observed that L1 averaging with improper rotations is more robust than L2 averaging, especially in the presence of outliers. The observation is consistent with the findings in [2, 7].

Note: The reader might wonder about the practical usefulness of the L1 averaging scheme. The L1 averaging scheme enables an alternative way of dealing with outliers besides the traditional Ransac approach. Its practical benefits are demonstrated in the following real world experiment. Furthermore, as argued in [6], the optimality of a solution to a geometric problem is not clearly definable. The L1 metric always has to be considered a valid alternative with respect to the L2 metric. For further discussion on this topic, the reader in kindly referred to [6].

5.3. Real world experiment

In this experiment, we employ mirror-based pose estimation to perform the extrinsic calibration between an iPhone 4's front and back camera. As shown in Figure 1, a calibration chessboard is placed besides the iPhone such that both the front and back cameras can observe the chessboard only via mirror refections. Multiple images are captured by each camera for individual mirror poses. Each camera's pose is then related to the chessboard by applying the modified PnP algorithm presented in [11] and using the presented mirror-based rotation averaging methods. Finally, we can easily derive the relative pose between the two non-overlapping

cameras.

The MLE (Maximum Likelihood Estimate) is computed by non-linear minimization of the re-projection errors over both cameras' extrinsic and intrinsic parameters as well as the poses of all mirrors. The relative rotation between the MLEs of the front and back cameras is [-179.46, -1.80, -0.71] degrees in terms of Euler angles, and the relative translation is [3.43, -34.8, -11.5] mm. This value for the relative pose between the front and back camera frames of an iPhone 4 agrees with our visual observation.

We compared the outputs of the algorithms of Sturm [14], Rodrigues [13] and Takahashi [15] with the presented Chordal L2-mean and geometric L1-mean algorithms for the back camera's mirror-based pose estimation. Table 1 shows the errors computed by using the MLE as ground truth. As we can see, the Chordal L2 algorithm got the closest result to the MLE among the closed-form algorithms. This is consistent with the conclusions from the simulation experiments. Furthermore, it should be noted that the geometric L1 averaging algorithm yields more accurate results than the L2 algorithm. We consider this a possible outcome because in a real-life scenario, the virtual poses may be estimated with different levels of accuracy since noise characteristics may be different for every image. This means that the L2 mean may not be the optimal metric as previously discussed.

To further prove the practical advantages of L1 averaging, we added 3 mirrored outlier images to our computations. The outlier images are captured with varied (i.e. wrong) relative displacement between the camera and the chessboard. This displacement should remain fixed during

the entire calibration process. Table 2 shows the resulting error with respect to the MLE result from the previous outlier-free experiment. As we can see, the L1 averaging algorithm clearly outperforms other algorithms in the presence of outliers. Please note that the non-linear minimization of re-projection errors cannot do its job in this situation. We could certainly run Ransac followed by non-linear minimization, but our L1 averaging is a much easier alternative.

	Sturm	Rod˜	Taka˜	Cho L2	Geo L1
R (deg)	1.121	1.524	1.241	1.106	0.821
T (mm)	12.15	23.56	13.38	12.11	11.85

Table 1. Rotation and translation errors with respect to the MLE on a real world experiment.

	Sturm	Rod˜	Taka˜	Cho L2	Geo L1
R (deg)	8.01	8.68	4.69	7.87	1.08
T (mm)	175.66	148.97	91.80	210.28	42.25

Table 2. Rotation and translation errors in an outlier-affected case and with respect to the MLE obtained from the previous outlier-free experiment.

6. Discussion

We presented a novel solution to mirror-based camera pose estimation that does not rely on the commonly employed fixed-axis rotation constraint. Instead, we show how the problem can be tackled by an elegant generalization of rotation averaging able to handle the improper rotations of mirrored views. Our theoretical contribution also lies in the extension of the applicability of rotation averaging to a more general case. The practical usefulness of our formulation is supported by the quality of our results. We outperform the state-of-the-art in terms of accuracy, computational complexity, and robustness with respect to outlier images.

Perhaps the major drawback of our approach is that the matrix of all mirror plane normal vectors is required to have full rank. This is partly due to the ignorance of translations in the averaging stage, which leads to a loss of information. Our follow-up research direction therefore consists of pursuing complete motion averaging strategies as opposed to averaging rotations only. In most practical applications, however, degenerate situations hardly ever appear, as the mirror poses can be chosen accordingly.

Appendix A

Proof that (12) is unique *iff* $\text{rank}(\mathbf{F}) = 3$, where $\mathbf{F} = [\mathbf{n}_1 \quad \mathbf{n}_2 \quad \cdots \quad \mathbf{n}_n]$.

Theorem 1 (given in [9]): If $\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T$ is the SVD of $\mathbf{G}$, $\hat{\mathbf{R}} = \mathbf{U}\,\text{diag}(1, 1, \det(\mathbf{U}\mathbf{V}^T))\mathbf{V}^T$ is unique if

- $\text{rank}(\mathbf{G}) > 1$ and $\det(\mathbf{U}\mathbf{V}^T) = 1$.

- $\text{rank}(\mathbf{G}) > 1$ and the minimum singular value of $\mathbf{G}$ is a simple root.

We now give a group of definitions and conclusions which are useful for the subsequent proof:

- Let $\mathbf{J} = \sum\limits_{i=1}^{n} (\mathbf{I} - 2\mathbf{n}_i\mathbf{n}^T_i)$. Let $\lambda_1 \geq \lambda_2 \geq \lambda_3$ be the eigenvalues of $\mathbf{J}$, and $\sigma_1 \geq \sigma_2 \geq \sigma_3 \geq 0$ the singular values of $\mathbf{F}$. It is obvious that $\mathbf{J} = \mathbf{nI} - 2\mathbf{FF}^T$, and we have $\lambda_1 = n - 2\sigma_3^2, \lambda_2 = n - 2\sigma_2^2, \lambda_3 = n - 2\sigma_1^2$.

- Considering that $\lambda_1 = \mathbf{n} - 2\sigma_3^2 \leq n$ and $\sum\limits_{i=1}^{3} \lambda_i = \text{tr}(J) = n$, we have $\lambda_2 + \lambda_3 \geq 0$. Recalling that $\lambda_1 \geq \lambda_2 \geq \lambda_3$, we obtain $|\lambda_1| = \lambda_1, \ |\lambda_2| = \lambda_2$.

- Furthermore, considering that $\mathbf{J} = \mathbf{J^T}$, we know that the singular values of $\mathbf{J}$ are $\{\lambda_1, \lambda_2, |\lambda_3|\}$.

- Since $\mathbf{J} = \mathbf{R}^T\mathbf{G}$, we have $\det(\mathbf{G}) = \det(\mathbf{J}) = \prod\limits_{i=1}^{3} \lambda_i$, and the singular values of $\mathbf{G}$ are the same as those of $\mathbf{J}$, namely $\{\lambda_1, \lambda_2, |\lambda_3|\}$.

Proof of Sufficiency: If $\text{rank}(\mathbf{F}) = 3$, $\sigma_3 > 0$, and $\lambda_1 = n - 2\sigma_3^2 < n$. Recall that $\text{tr}(\mathbf{J}) = n = \sum\limits_{i=1}^{3} \lambda_i$, and that $\lambda_2 + \lambda_3 > 0$. This means that there are 3 cases:

- **Case 1** : $\lambda_2 > 0, \lambda_3 > 0$. In this case $\det(\mathbf{G}) > 0$, and therefore $\det(\mathbf{U}\mathbf{V}^T) = 1$. $\hat{\mathbf{R}}$ is unique according to Theorem 1.

- **Case 2** : $\lambda_2 > 0, \lambda_3 < 0$. In this case $\det(\mathbf{G}) < 0$, and $\lambda_2 > -\lambda_3$. Recalling that $\lambda_2 \geq \lambda_3$, we have $|\lambda_2| \neq |\lambda_3|$, which means that the minimum singular value is a simple root. $\hat{\mathbf{R}}$ is unique according to Theorem 1.

- **Case 3** : $\lambda_2 > 0, \lambda_3 = 0$. In this case $\text{rank}(\mathbf{G}) = 2$, and the minimum singular value is 0 and thus a simple root. $\hat{\mathbf{R}}$ is unique according to Theorem 1.

Proof of Necessity: We only need to prove that $\hat{\mathbf{R}}$ is not unique in the case of $\text{rank}(\mathbf{F}) = 1$ or $\text{rank}(\mathbf{F}) = 2$.

In case $\text{rank}(\mathbf{F}) = 1$, we have $\sigma_2 = 0$ and $\sigma_3 = 0$. It follows that $\{\lambda_1 = n, \lambda_2 = n, \text{ and } \lambda_3 = -n\}$. Thus $\det(\mathbf{G}) = -n^3 < 0$, and all of the 3 singular values of $\mathbf{G}$ are the same. $\hat{\mathbf{R}}$ is not unique according to Theorem 1.

In case $\text{rank}(\mathbf{F}) = 2$, we have $\sigma_3 = 0$, and thus $\lambda_1 = n$ and $\lambda_2 + \lambda_3 = 0$. There are two cases:

- If $\lambda_2 = \lambda_3 = 0$, then $\text{rank}(\mathbf{G}) = 1$, and $\hat{\mathbf{R}}$ is not unique according to Theorem 1.

- If $\lambda_2 = -\lambda_3 \neq 0$, then $\det(\mathbf{G}) = \lambda_1\lambda_2\lambda_3 < 0$, and the minimum singular value $\lambda_2 = |\lambda_3|$ is not a simple root. $\hat{\mathbf{R}}$ is not unique according to Theorem 1.

References

[1] A. Agrawal and S. Ramalingam. Single image calibration of multi-axial imaging systems. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1399–1406, Portland, USA, 2013. 2

[2] Y. Dai, J. Trumpf, H. Li, N. Barnes, and R. Hartley. Rotation averaging with application to camera-rig calibration. In *Computer Vision–ACCV 2009*, pages 335–346. Springer, 2010. 7

[3] J. Fillmore. A note on rotation matrices. *IEEE Computer Graphics and Applications*, 4(2):30–33, 1984. 3

[4] M. Fischler and R. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981. 2

[5] J. Gluckmann and S. K. Nayar. Catadioptric stereo using planar mirrors. *International Journal of Computer Vision (IJCV)*, 44(1):65–79, 2001. 2

[6] R. Hartley and F. Kahl. Optimal algorithms in multiview geometry. In *Computer Vision–ACCV 2007*, pages 13–34. Springer, 2007. 7

[7] R. Hartley, J. Trumpf, Y. Dai, and H. Li. Rotation averaging. *International Journal of Computer Vision (IJCV)*, 103(3):267–305, 2013. 2, 4, 5, 7

[8] J. A. Hesch, A. I. Mourikis, and S. I. Roumeliotis. *Springer Tracts in Advanced Robotics: Algorithmic foundations of robotics VIII*. Springer, 2009. 2, 6

[9] K. Kanatani. Analysis of 3-d rotation fitting. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 16(5):543–549, 1994. 5, 8

[10] R. Kumar, A. Ilie, J.-M. Frahm, and M. Pollefeys. Simple Calibration of Non-Overlapping Cameras with a Mirror. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Miami, FL, USA, 2008. 2, 6

[11] S. Li, C. Xu, and M. Xie. A robust o (n) solution to the perspective-n-point problem. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 34(7):1444–1450, 2012. 5, 7

[12] A. Morawiec. *Orientations and rotations*. Springer, 2003. 3

[13] R. Rodrigues, P. Barreto, and U. Nunes. Camera pose estimation using images of planar mirror reflections. In *Proceedings of the European Conference on Computer Vision (ECCV)*, Crete, Greece, 2010. 2, 5, 6, 7

[14] P. Sturm and T. Bonfort. How to compute the pose of an object without a direct view? In *Proceedings of the Asian Conference on Computer Vision (ACCV)*, pages 21–31, Hyderabad, India, 2006. 2, 5, 7

[15] K. Takahashi, S. Nobuhara, and T. Matsuyama. A new mirror-based extrinsic camera calibration using an orthogonality constraint. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Providence, USA, 2012. 2, 5, 6, 7