

Multicamera calibration from visible and mirrored epipoles

Andrey Bushnevskiy, Lorenzo Sorgi
Technicolor Research & Innovation
Karl-Wiechert Allee 74, 30625
Hannover, Germany

Bodo Rosenhahn
Leibniz University Hannover
Appelstr 9A, 30169
Hannover, Germany

Abstract

Multicamera rigs are used in a large number of 3D Vision applications, such as 3D modeling, motion capture or telepresence and a robust calibration is of utmost importance in order to achieve a high accuracy results. In many practical configurations the cameras in a rig are arranged in such a way, that they can observe each other, in other words a number of epipoles correspond to the real image points. In this paper we propose a solution for the automatic recovery of the external calibration of a multicamera system by enforcing only simple geometrical constraints, arising from the epipole visibility, without using any calibration object, such as checkerboards, laser pointers or similar. Additionally, we introduce an extension of the method that handles the case of epipoles being visible in the reflection of a planar mirror, which makes the algorithm suitable for the calibration of any multicamera system, irrespective of the number of cameras and their actual mutual visibility, and furthermore we remark that it requires only one or a few images per camera and therefore features a high speed and usability. We produce an evidence of the algorithm effectiveness by presenting a wide set of tests performed on synthetic as well as real datasets and we compare the results with those obtained using a traditional LED-based algorithm. The real datasets have been captured using a multicamera Virtual Reality (VR) rig and a spherical dome configuration for 3D reconstruction.

1. Introduction

Camera calibration is the specification of a mathematical model describing the image formation process realized by an imaging system. For a single camera it comprises two steps, the photometric and the internal calibration that specifically addresses the estimation of the camera response function [6] and the system lens-sensor (and optionally -mirror) geometry [18, 17]. When multiple cameras are used simultaneously, an additional calibration step is required. It copes with the estimation of the camera cluster geome-

try, namely the set of camera poses in a common reference frame (rf.) [8, 4]; which is generally referred to as external calibration. In many calibration algorithms and tools this task is blended in a joint process together with the internal calibration of the cameras, using a calibration object [2]. This solution, however, requires a simultaneous visibility of the pattern in many or even in all calibration images, which in practice is not always feasible and does not scale well with the increasing number of cameras and complexity of the cluster geometry. Eventually, in the recent years the advantage of decoupling these two operations into two different calibration phases has become clear [10]. A different approach actually neglects the external calibration and postpones it to the further operation of the system, that is the video capture in an uncontrolled environment. In [5] the multicamera geometry is extracted only enforcing the rigidity constraint, which is the constraint of the cameras having a fixed orientations and translations between each other. The elegant formulation based on the definition of a virtual camera makes the system suitable also for image sequences with non-overlapping content, however it requires the multicamera system to undergo a consistent motion and relies on the performance of the feature tracker, which is itself dependent on the image content. Therefore this is not the optimal solution for the calibration of an indoor immovable systems. A more flexible alternative is provided by a single-point calibration techniques, namely the algorithms based on the projection of a single moving point in the parallel video sequences. The latter is usually provided by a laser pointer, which can be easily detected in each image [16, 1]. These techniques however, despite being very powerful, are still designed as a joint framework for external and internal calibration which limits the flexibility in case of variation of the cluster geometry, removal of cameras or addition of the new ones.

In this paper we present a novel approach for the multicamera systems calibration, based only on the geometrical constraints arising from the visible epipoles and featuring important advantages. The size of the calibration dataset is highly reduced, therefore is less error-prone and easy to col-

lect, automatize and review for possible outliers in the set of detected epipoles. Therefore, the system becomes flexible enough to enable a frequent re-arrangement of the camera geometry with a small amount extra work required for re-calibration, making it a good candidate for an integration into a really practical, handy and user-friendly calibration tool. The system also has some limitations, which, however, can be easily circumvented. The accuracy of the recovered geometry is high enough to provide a reliable initialization of any 3D reconstruction system, but needs to be further refined using a bundle adjustment (BA) together with the reconstructed model, which is anyway a common approach in any 3D reconstruction architecture. A second issue is a possible lack of a visible epipoles, which may happen for some camera configurations as well as for cameras with a narrow field-of views lenses. For this reason, inspired by the work proposed in [13], we have designed an extension, based on a planar mirror, which enables the joint exploitation of mirrored and directly visible epipoles for the multicamera camera pose recovery and eventually makes the system flexible enough for any camera cluster configuration.

1.1. Notation

In order to ease the paper understanding, we first introduce the mathematical notation. We will consider a multicamera system comprised of N cameras, which projection centers are denoted by $\{C_i\}_{i=1,\dots,N}$. The absolute reference frame will be identified with the index w and the Euclidean transformation converting the vector representation of a 3D point from w to the rf. of the i -th camera will be denoted as $T_{wi} = [R_{wi} | t_{wi}] : R_{wi} \in SO(3), t_{wi} \in \mathbb{R}^3$, where $SO(3)$ is the group of 3D rotation matrices, (Fig. 1) and $I_{3 \times 3} \in SO(3)$ is the null rotation, namely the identity matrix. The action of the transformation T_{wi} on a vector \mathbf{X} will be synthetically expressed as $T_{wi} \circ \mathbf{X} = R_{wi} \mathbf{X} + t_{wi}$.

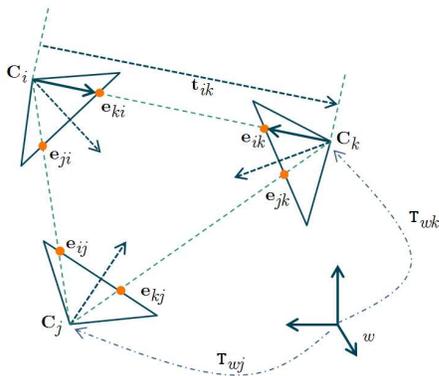


Figure 1: Multicamera model with visible epipoles.

Inner product and cross product between 3D vectors will be denoted as $\mathbf{v}_a^T \mathbf{v}_b$ and $[\mathbf{v}_a]_{\times} \mathbf{v}_b$, where $[\mathbf{v}]_{\times}$ is the skew

symmetric matrix built using the vector \mathbf{v} .

We assume working in the calibrated camera conditions, which implies that the image points can be represented as a unit-norm vectors, that is points on unit sphere S^2 or equivalently 3D directions in the 3D space. Occasionally, the projection of a 3D vector on S^2 will be explicitly expressed using the notation $\mathcal{N}\{\mathbf{v}\} = \frac{\mathbf{v}}{\sqrt{\mathbf{v}^T \mathbf{v}}}$.

The epipoles of the multicamera systems will identified by the letter e , and specifically e_{ij} denotes the projection of the point C_i on the j -th camera; epipoles will be represented as unit-norm vectors as well.

Our aim will be to estimate the set of camera poses, namely the set of transformations $\{T_{wi}\}_{i=1,\dots,N}$ and the estimation will be formulated as the minimization of an objective function built using the geometrical constraints between the visible epipoles. The minimization will be performed by means of the Levenberg-Marquardt iterative algorithm (LM). By the formalization of the LM iterations, we will use the a^+ in order to identify the update of the variable a . We will also use the notations \tilde{a} and \hat{a} in order to distinguish the actual measurement from the estimate of a . The working dataset will consist of the set of visible epipoles, denoted as $\mathcal{V} = \{(i, j) : \exists \tilde{e}_{ij}\}$. The epipoles measurements in the set \mathcal{V} will be also grouped pairwise in a second set \mathcal{M} , the set of pairs of mutually visible epipoles, $\mathcal{M} = \{(i, j) : \exists \tilde{e}_{ij} \wedge \exists \tilde{e}_{ji}\}$.

2. Calibration of the multicamera geometry

Without loss of generality we assume that the rf. w is aligned with the first camera, that is $T_{w1} = [I_{3 \times 3} | \mathbf{0}]$. The calibration process consequently reduces to the estimation of the transformation set $\{T_{wi}\}_{i=2,\dots,N}$. In the next sections we show how this task can decoupled in two consecutive steps aimed at estimation of rotational and translational components of the camera poses, namely $\{R_{wi}\}_{i=2,\dots,N}$ and $\{t_{wi}\}_{i=2,\dots,N}$.

2.1. Recovery of the rotational components

Let us consider two cameras i and j that can observe each other, as in the model in Fig.1. By applying the coordinate transformation between the i -th camera rf. and w , the epipole e_{ji} can be then expressed as

$$e_{ji} \sim T_{wi} \circ C_j = R_{wi}^T (C_j - C_i) . \quad (1)$$

A similar relation can be established also for the corresponding epipole e_{ij} , which together with (1) allows for the removal of the dependency from the (unknown) camera centers:

$$e_{ji} = -R_{wi} R_{wj}^T e_{ij} . \quad (2)$$

Equation (2) could have been derived also by considering the epipoles to be the unit norm vectors, representation of two points on S^2 .

Each pair of epipoles in the set \mathcal{M} can be used to build an equation in a form (2), which stacked together lead to the formulation of the estimation of the multicamera geometry rotational components as an optimisation problem

$$\{\hat{\mathbf{R}}_{wi}\}_{i=2,\dots,N} = \arg \min_{\mathbf{R}_{wi} \in SO(3)} \left\{ \sum_{(i,j) \in \mathcal{M}} \|\boldsymbol{\rho}_{i,j}\|^2 \right\}, \quad (3)$$

where each error contribution is defined as:

$$\boldsymbol{\rho}_{i,j} = \tilde{\mathbf{e}}_{ij} + \mathbf{R}_{wj} \mathbf{R}_{wi}^\top \tilde{\mathbf{e}}_{ji} \quad (4)$$

A closed form solution for single rotation estimation given a set of corresponding vectors is a well known result of matrix analysis [11]; however we are not aware of similar solutions for the simultaneous recovery of multiple rotations as in (3). In many approaches rotations and translations are estimated in a non-iterative manner, however those approaches rely on the external feature points, which are not available in our calibration case scenario [14, 7, 15, 3, 15, 12]. The same observation applies to the other state of the art works addressing the translation estimation under the assumption of known rotation [9]. Therefore, we attempt to solve the problem within an iterative minimization context using the LM algorithm.

For this purpose we parameterize a differential rotation with a 3D vector $\mathbf{d}\boldsymbol{\omega}$, which is used to compute the corresponding unit quaternion and then the rotation update in the matrix representation:

$$\begin{cases} \mathbf{d}\mathbf{q} = \left[\sqrt{(1 - \mathbf{d}\boldsymbol{\omega}^\top \mathbf{d}\boldsymbol{\omega})} \mid \mathbf{d}\boldsymbol{\omega}^\top \right] \\ \mathbf{d}\mathbf{R} = \mathbf{d}\mathbf{R}(\mathbf{d}\mathbf{q}) = \mathbf{I}_{3 \times 3} + 2[\mathbf{d}\boldsymbol{\omega}]_\times \\ \mathbf{R}^+ = \mathbf{R} \cdot \mathbf{d}\mathbf{R} \end{cases} \quad (5)$$

Using equations (5) in the definition (4) in place of both the camera rotations \mathbf{R}_{wj} and \mathbf{R}_{wi} , and discarding the second order terms in $\mathbf{d}\mathbf{p}$, one obtains:

$$\boldsymbol{\rho}_{i,j}^+ = \tilde{\mathbf{e}}_{ij} + \mathbf{R}_{wj}^+ \mathbf{R}_{wi}^{+\top} \tilde{\mathbf{e}}_{ji} = \boldsymbol{\rho}_{i,j}^+ + \mathbf{J}_\rho \mathbf{d}\mathbf{p} \quad (6)$$

where $\mathbf{d}\mathbf{p}$ is the update vector collecting all the rotation update parameters stacked in a single vector, $\mathbf{d}\mathbf{p} = [\mathbf{d}\boldsymbol{\omega}_2^\top, \dots, \mathbf{d}\boldsymbol{\omega}_k^\top, \dots, \mathbf{d}\boldsymbol{\omega}_N^\top]^\top$. After a few mathematical manipulations the Jacobian \mathbf{J}_ρ can be arranged in a compact matrix form as

$$\begin{cases} \mathbf{J}_\rho = [0\dots, \mathbf{J}_b, \dots 0\dots, -\mathbf{J}_b, \dots 0] \\ \mathbf{J}_b = 2\mathbf{R}_{wj} [\mathbf{R}_{wi}^\top \tilde{\mathbf{e}}_{ji}]_\times \end{cases} \quad (7)$$

The update equations (7) and (6) are then used within the LM optimization framework to build the normal equation that is solved in each LM iteration, with the trivial initialization $\mathbf{R}_{wj} = \mathbf{I}_{3 \times 3}$ for each camera. As each camera

pose has three degrees of freedom and each vector equation (2) provides only two independent linear equations, one can infer that the number of visible epipoles must satisfy the condition

$$2|\mathcal{M}| \geq 3(N-1), \quad (8)$$

where $|\mathcal{M}|$ is the cardinality of the set \mathcal{M} . As N cameras define a maximum number of $\frac{N^2 N}{2}$ pairs of mutually visible epipoles, one can also conclude that a necessary, but not sufficient condition to meet the constraint (8) is $N \geq 3$.

2.2. Recovery of the translational components

Once the set of camera orientations $\{\hat{\mathbf{R}}_{wi}\}_{i=2,\dots,N}$ is available, one can use it as an additional input to tackle the estimation of the camera translational components. Let us rewrite equation (1), or rather its complementary equation for the epipole \mathbf{e}_{ij} , explicitly considering the normalization factor

$$\mathbf{e}_{ij} = \mathbf{R}_{wj} \cdot \mathcal{N}\{\mathbf{C}_i - \mathbf{C}_j\}. \quad (9)$$

Now the camera orientation \mathbf{R}_{wj} can be assumed to be known, therefore (9) becomes a non-linear equation in the unknowns \mathbf{C}_i and \mathbf{C}_j . As each visible epipole provides an equation in a form of (9), one can formulate the estimation of the camera translation components as the solution of the optimization problem

$$\{\hat{\mathbf{C}}_i\}_{i=2,\dots,N} = \arg \min_{\mathbf{C}_i \in \mathbb{R}^3} \left\{ \sum_{\mathcal{V}} \|\boldsymbol{\tau}_{i,j}\|^2 \right\}, \quad (10)$$

where each error contribution is defined as:

$$\boldsymbol{\tau}_{i,j} = \tilde{\mathbf{e}}_{ij} - \hat{\mathbf{R}}_{wj} \cdot \mathcal{N}\{\mathbf{C}_i - \mathbf{C}_j\}. \quad (11)$$

Similarly to the rotational component estimation, we solve the problem (10) using the LM algorithm. For a comfortable formulation of the LM iterations, still keeping a minimal dimensionality of the optimization space, we parameterize the projection center of each camera using the corresponding *epipolar ray* on the reference camera,

$$\mathbf{C}_i = \lambda_i \mathbf{e}_{i1}, \quad (12)$$

and we express the parameters update as

$$\begin{cases} \lambda_i^+ = \lambda_i (1 + \delta\lambda_i) \\ \mathbf{e}_{i1}^+ = \mathbf{e}_{i1} + \mathbf{B}_i \delta\mathbf{e}_i, \end{cases} \quad (13)$$

where \mathbf{B}_i is the 3×2 matrix representing any basis for the plane tangent to the unit sphere at the epipole \mathbf{e}_{i1} , $\delta\mathbf{e}_i$ is a 2D vector and $\delta\lambda_i$ is a scalar. Notice that the definition of the epipoles as unit norm vectors justifies the update equation (13), which describes a local update of each epipole \mathbf{e}_{i1} on S^2 (Fig.2). We omit the description of a basis retrieval for a given 3D plane for brevity.

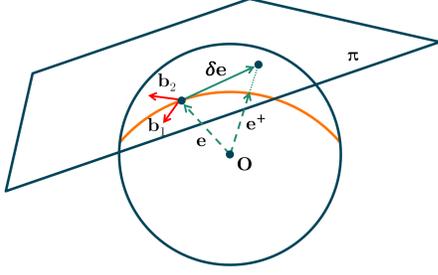


Figure 2: Epipole update parameterization.

Using equations (13) and (12), we can rewrite the update of the error term (11) as a function of the parameter vector $\mathbf{dp} = [\delta e_2^\top, \delta \lambda_2, \dots, \delta e_i^\top, \delta \lambda_i, \dots, \delta e_N^\top, \delta \lambda_N]^\top$, collecting the parameters of each camera center stacked in a single vector. After a few mathematical manipulations one can arrange a linear equation in the form

$$\tau_{i,j}^+ = \tau_{i,j} + J_\tau \mathbf{dp}, \quad (14)$$

where similarly to (7) the Jacobian matrix J_τ can be expressed in a relatively compact matrix form by applying the basic derivation rules for the square root and quotient of functions. We omit this derivation in order to keep the notation reasonably light.

Similarly to (2), also equation (9) provides only two independent linear equations in the unknowns C_i and C_j ; consequently one can show that the number of visible epipoles must satisfy the condition

$$2|\mathcal{V}| \geq 3(N - 1), \quad (15)$$

where $|\mathcal{V}|$ is the cardinality of the set \mathcal{V} . The latter is anyway always greater than $|\mathcal{M}|$, therefore the set of visible epipoles always provides enough constraints for the translation components estimation when the condition (8) is met.

In contrast to the rotational components, for the recovery of the camera center locations a smarter initialization of the LM iteration can also be achieved, as shown in Fig.3. The reference camera is naturally located in the origin, while all the other cameras are located on one epipolar ray corresponding to one visible epipole. For example given that $\tilde{e}_{ij} \in \mathcal{V}$, then C_i is located on the ray identified by \tilde{e}_{ij} at unit distance from the C_j , $C_i = 1 \cdot \tilde{e}_{ij}$, the camera center is then converted in the rf. of the reference camera and the corresponding pose is computed according to (12). As multiple epipoles \tilde{e}_{ij} for a single view i may be available, one is randomly selected for the pose initialization.

3. Epipole Mirroring

In many practical scenarios the geometrical configuration of the multicamera system and the type of lens mounted

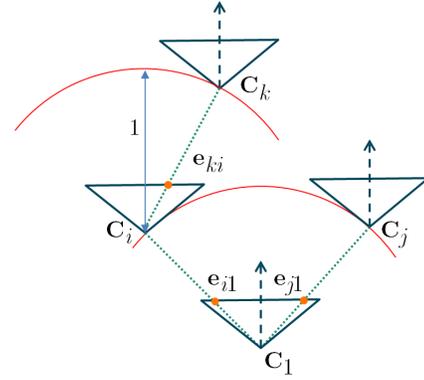


Figure 3: LM Initialization for the translation component recovery.

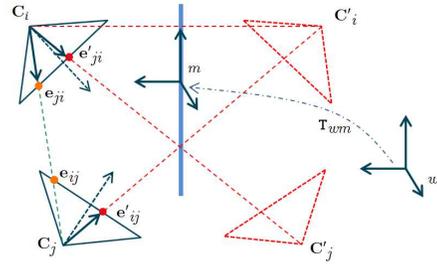


Figure 4: Mirrored Epipoles.

on each camera may result in a calibration image set with a number of visible epipoles, which is not large enough to meet the conditions (8) and (15). An example is given by 360° camera rigs for VR and panorama capture, such as Google JUMP. In order to cope with such configurations we provide a simple solution based on using a planar mirror. It is placed multiple times in front of the camera system, allowing the cameras to observe each other in the reflection and consequently extending the size of the set of visible epipoles. In the next sections the additional epipoles will be denoted as mirrored epipoles (Fig. 4) and we will show how these can be used on their own or in combination with the directly visible epipoles.

3.1. Recovery of the rotational components

In order to simplify the derivation of the geometrical constraints pertaining to the mirrored epipoles, it is convenient to introduce an additional reference system m with the XY plane aligned with the mirror surface, related to the rf. w by the Euclidean transformation $T_{wm} = [R_{wm} | \mathbf{t}_{mw}]$. By applying the basic laws of the reflection geometry one can show that in the rf. m a vector \mathbf{X} and its mirrored version \mathbf{X}' are related by the equation $\mathbf{X} = F_z \mathbf{X}'$, where $F_z = \text{diag}(0, 0, -1)$ is the reflection transformation in-

duced by the plane XY . Therefore, the mirrored epipole \mathbf{e}'_{ij} in the j -th view can be accounted as the image of a virtual cameras located in \mathbf{C}'_i :

$$\mathbf{e}'_{ij} \sim \mathbf{T}_{wj} \circ \mathbf{C}'_i = \mathbf{R}_{wj}^\top (\mathbf{C}'_i - \mathbf{C}_j) , \quad (16)$$

where

$$\mathbf{C}'_i = \mathbf{T}_{wm}^{-1} \circ (\mathbf{F}_z \cdot (\mathbf{T}_{wm} \circ \mathbf{C}_i)) . \quad (17)$$

Following the same steps as in section 2.1, one can easily derive a relation similar to (2) for the mirrored epipoles,

$$\mathbf{e}'_{ij} = -\mathbf{R}_{mj} \mathbf{F}_z \mathbf{R}_{mi}^\top \mathbf{e}'_{ji} = -\mathbf{R}_{wj} \mathbf{R}_{wm}^\top \mathbf{F}_z \mathbf{R}_{wm} \mathbf{R}_{wi}^\top \mathbf{e}'_{ji} . \quad (18)$$

Equation (18) can be expanded by decomposing the rotation matrix \mathbf{R}_{wm} by means of the Euler ZYX parameterization, $\mathbf{R}_{wm} = \mathbf{Z}_{wm} \mathbf{Y}_{wm} \mathbf{X}_{wm}$,

$$\mathbf{e}'_{ij} = -\mathbf{R}_{wj} \mathbf{X}_{wm}^\top \mathbf{Y}_{wm}^\top \mathbf{Z}_{wm}^\top \mathbf{F}_z \mathbf{Z}_{wm} \mathbf{Y}_{wm} \mathbf{X}_{wm} \mathbf{R}_{wi}^\top \mathbf{e}'_{ji} , \quad (19)$$

and then simplified in

$$\mathbf{e}'_{ij} = -\mathbf{R}_{wj} \mathbf{X}_{wm}^\top \mathbf{Y}_{wm}^\top \mathbf{F}_z \mathbf{Y}_{wm} \mathbf{X}_{wm} \mathbf{R}_{wi}^\top \mathbf{e}'_{ji} , \quad (20)$$

using the equivalence $\mathbf{Z}_{wm}^\top \mathbf{F}_z \mathbf{Z}_{wm} = \mathbf{F}_z$.

Each pair of mutually visible mirrored epipoles provides an additional equation in the form of (20), which can be incorporated into the optimization problem (3). One can therefore reformulate (3) by extending the optimization space and including the additional error terms, as:

$$\left\{ \hat{\mathbf{R}}_{wi}, \hat{\mathbf{X}}_{wm}, \hat{\mathbf{Y}}_{wm} \right\}_{\substack{i=1,\dots,N, \\ m=1,\dots,M}} = \arg \min_{\substack{\mathbf{R}_{wi} \in SO(3) \\ \mathbf{X}_{wm} \in SO_X(3) \\ \mathbf{Y}_{wm} \in SO_Y(3)}} E_\rho , \quad (21)$$

where $SO_X(3)$ and $SO_Y(3)$ are the subspace of 3D rotations about the X-axis and the Y-axis and the objective function E_ρ is defined as:

$$\left\{ \begin{array}{l} E_\rho = \sum_{(i,j) \in \mathcal{M}} \|\boldsymbol{\rho}_{i,j}\|^2 + \sum_{(i,j) \in \mathcal{M}'} \|\boldsymbol{\rho}'_{i,j}\|^2 \\ \boldsymbol{\rho}'_{i,j} = \tilde{\mathbf{e}}'_{ij} + \mathbf{R}_{wj} \mathbf{X}_{wm}^\top \mathbf{Y}_{wm}^\top \mathbf{F}_z \mathbf{Y}_{wm} \mathbf{X}_{wm} \mathbf{R}_{wi}^\top \tilde{\mathbf{e}}'_{ji} . \end{array} \right. \quad (22)$$

In (22) the error term $\boldsymbol{\rho}_{i,j}$ is defined according to equation (4), \mathcal{M}' is the set of pairs of mirrored mutually visible epipoles and M is the number of mirror snapshots, that is the number of different mirror poses.

The Jacobian of the objective function can be analytically computed in a (relatively compact) matrix using the same procedure, previously described for the derivation of equation (7), which is omitted for brevity. The solution is then obtained by means of the Levenberg-Marquardt algorithm, where the additional parameters of the mirror poses

are initialized to zero, that is \mathbf{X}_{wm} and \mathbf{Y}_{wm} are initialized as identity matrices.

We remark that the formulation (21) provides the lowest dimensionality of the optimization space as each mirror pose increases the number of parameters by two, namely the additional parameters of the rotations \mathbf{X}_{wm} and \mathbf{Y}_{wm} .

3.2. Recovery of the translational components

Taking advantage of the knowledge of the rotational components of the camera and the mirror poses (21), one can significantly reduce the complexity of the translational component estimation. We first expand the mirror pose components as

$$\left\{ \begin{array}{l} \mathbf{R}_{wm}^\top = [\mathbf{r}_{1wm}, \mathbf{r}_{2wm}, \mathbf{r}_{3wm}] \\ \mathbf{t}_{wm} = [tx_{wm}, ty_{wm}, tz_{wm}] \end{array} \right. .$$

Then by plugging the previous equations in (17) and after a few manipulations, one obtains

$$\left\{ \begin{array}{l} \mathbf{C}'_i = \mathbf{F}_m \mathbf{C}_i + tz_{wm} \mathbf{r}_{3wm} \\ \mathbf{F}_m = \mathbf{X}_{wm}^\top \mathbf{Y}_{wm}^\top \mathbf{F}_z \mathbf{Y}_{wm} \mathbf{X}_{wm} . \end{array} \right. \quad (23)$$

By using equation (23) in (16) one can represent the mirrored epipoles as a function of only the real poses of the cameras and the mirror:

$$\mathbf{e}'_{ij} = \mathbf{R}_{wj} \cdot \mathcal{N} \{ \mathbf{F}_m \mathbf{C}_i + tz_{wm} \mathbf{r}_{3wm} - \mathbf{C}_j \} . \quad (24)$$

Each visible mirrored epipole provides an equation in the form of (24), which can be incorporated into the optimization problem (10). This is reformulated using the estimates of the camera and mirror rotations in place of \mathbf{R}_{wm} , \mathbf{X}_{wm} and \mathbf{Y}_{wm} , extending the optimization space and including the additional error terms:

$$\left\{ \hat{\mathbf{C}}_i, \hat{tz}_{wm} \right\}_{\substack{i=1,\dots,N, \\ m=1,\dots,M}} = \arg \min_{\substack{\mathbf{C}_i \in \mathbb{R}^3 \\ tz_{wm} \in \mathbb{R}}} E_\tau , \quad (25)$$

where the objective function E_τ is defined as:

$$\left\{ \begin{array}{l} E_\tau = \sum_{\mathcal{V}} \|\boldsymbol{\tau}_{i,j}\|^2 + \sum_{\mathcal{V}'} \|\boldsymbol{\tau}'_{i,j}\|^2 \\ \boldsymbol{\tau}'_{i,j} = \tilde{\mathbf{e}}'_{ij} - \hat{\mathbf{R}}_{wj} \cdot \mathcal{N} \{ \hat{\mathbf{F}}_m \mathbf{C}_i + \hat{tz}_{wm} \hat{\mathbf{r}}_{3wm} - \mathbf{C}_j \} . \end{array} \right. \quad (26)$$

In (26), $\boldsymbol{\tau}_{i,j}$ is defined according to equation (11) and \mathcal{V}' is the set of mirrored visible epipoles. Once again, we remark that the formulation (25) provides the lowest dimensionality of the optimization problem as each mirror pose increases the number of parameters only by one, namely the additional parameter tz_{wm} .

The objective function (25) is minimized by means of the LM algorithm, using the same parameterization and the

same initialization for the camera centers as in section 2.2. For the mirror one can also derive a simple initialization following the same model shown in Fig.3.

4. Results

The proposed algorithm has been evaluated on the synthetic and real datasets. In both cases we have tested the algorithm first using only directly visible epipoles and then also introducing the mirrored epipoles in the calibration dataset. The synthetic tests have been organized with the objective of assessing the algorithm stability against increasing noise corrupting the visible epipoles. This is an important point, as in a real scenario the visible epipoles can be localized only with some uncertainty within the image regions corresponding to the visible camera lenses. The algorithm has been further tested using the real multicamera systems, a dome and VR-rig configuration, both comprised of wide angle cameras.

4.1. Synthetic dataset

For the synthetic test a scene has been generated with a cluster of cameras, each one modeled as a super wide angle lens (150° horizontal FoV) and 1024 × 768 pixel sensor.

In the first test we have simulated a system consisting of seven virtual cameras randomly located on a unit sphere and pointing towards its origin. The visible epipoles location has been corrupted by a zero-mean Gaussian noise, with variance ranging in the interval [0; 2] pixels. For each level of noise a new scene has been generated 100 times and the mean and standard deviation of the geometry estimation error has been computed (Fig. 5 (a)). The estimation error of the rotational and translation component of the camera poses has been defined as

$$\begin{cases} \epsilon_{R_i} = \|\Phi(\mathbf{R}_{wi} \cdot \hat{\mathbf{R}}_{wi}^\top)\| \\ \epsilon_{t_i} = \|\mathbf{t}_{wi} - \hat{\mathbf{t}}_{wi}\| \end{cases}, \quad (27)$$

where $\Phi(\mathbf{R})$ is the conversion of a rotation matrix in the axis-angle representation.

In the next test, addressing the case of mirrored epipoles, the virtual cameras have been randomly located on a unit sphere pointing outwards in divergent configuration, so that they cannot see each other directly, but only in the reflection of a virtual mirror, which has been also randomly posed in 3D space. The estimation error has been calculated similar to the previous test and the results are presented in Fig.5 (b).

Both tests on the synthetic data confirm a high robustness against the image projection noise, despite the small size of the calibration dataset.

4.2. Real dataset

The method has been further evaluated on two real multicamera systems, a real dome and a VR-rig, both comprised

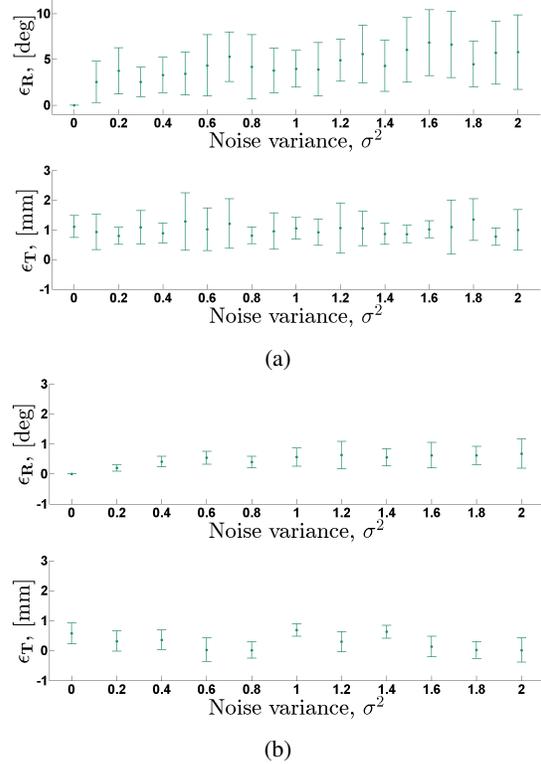


Figure 5: Epipole-based (a) and Mirrored epipole-based approaches, synthetic test. Rotation and translation errors vs. the epipole projection noise variance.

of GoPro HERO Black 3+ cameras. The corresponding system geometry has been estimated using the proposed algorithm (with only visible epipoles or only mirrored epipoles according to the geometry), followed by a BA refinement. For a comparison purpose we have calibrated the system also using the Multi-Camera Self-Calibration tool (MCSC) [16]. In order to ensure a fair comparison the intrinsic parameters of the cameras have been estimated beforehand and used for both calibration pipelines.

4.2.1 Dome configuration

The dome configuration comprises 8 GoPro Hero 3+ cameras arranged on a rigid metal frame, in two circular groups located one on top of the other, each one with 4 cameras (Fig.6 (a)). The system geometry has been estimated first using MCSC with an input of 32000 images (4000 images/points per camera), extracted from a video recording of a green LED moving in the inner volume of the dome. The system geometry has been then recovered using the proposed epipole-based method, after manual selection of the visible epipoles from only 8 input images (1 image per camera). The recovered geometry has been further refined using BA on the set of LED points, (Fig.6 (b)). LED points have

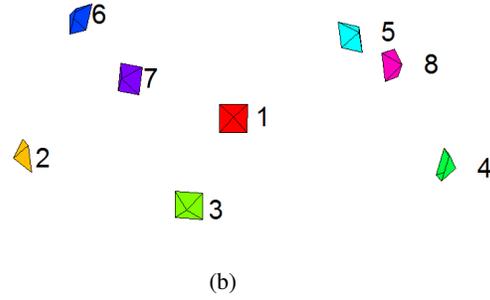
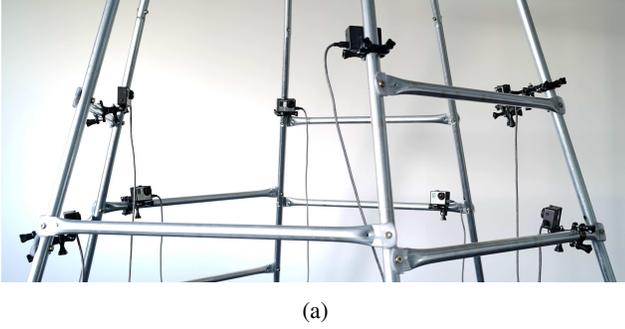


Figure 6: Dome multicamera setup (a). Geometry of the dome multicamera system reconstructed using the proposed Epipole-based approach (b).

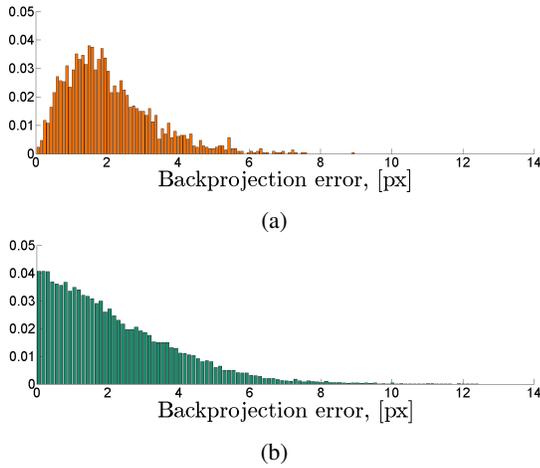


Figure 7: Backprojection error distribution for MCSC (a) and the Epipole-based (b) approach.

been only used in order to allow a direct comparison with MCSC, however in the real scenarios the camera positions can be bundle adjusted using the feature points coming from the actual target scene. To evaluate the estimated geometry the LED point structures have been back-projected onto the camera views using the system geometry, estimated using the two tested approaches. In each view the error has been measured as a Euclidean distance in pixels between the detected LED points and their backprojections and the overall statistical distribution of the backprojection error has been shown in Fig.7 and Table 1.

We observe that this comparison has the shortcoming of computing the evaluation measure from the same data used for calibration. Therefore, we suggest to perform a second test on an independent set of points, provided by a traditional black and white checkerboard pattern, which has been presented to the pairs of camera sharing the same field of view. The grid points have been detected in each view, triangulated in 3D space using the estimated camera geom-

	LED err., px	Pattern err., px
MCSC	$\mu = 2.2, \sigma = 1.9$	$\mu = 3.6, \sigma = 2.1$
Epipole	$\mu = 1.7, \sigma = 1.3$	$\mu = 1.1, \sigma = 0.7$

Table 1: Backprojection error for MCSC and the Epipole-based approach. μ - mean, σ - standard deviation .

etry and projected back to the source views. Similarly to the LED point test, we have computed the mean value and the standard deviation of the backprojection error between the detected grid points and their backprojected counterpart. A visual representation of this test is shown in Fig.9 and the error measures collected in Table 2.

From the inspection of Table 1 we can conclude that our technique is able to successfully recover the system geometry and slightly outcome the performance of MCSC.

4.3. VR rig configuration

The second test has been performed using a VR rig composed of 15 GoPro Hero Black 3+ cameras, 12 pairwise located on a hexagon and 3 positioned on top of the rig at the approximate angle of 45° (Fig.8 (a)).

We have attempted to calibrate the system using MCSC with an input of 180000 images (12000 images per camera), however the system failed due to the peculiarity of this particular configuration, that implies a reduced overlap between the cameras field of views. Our approach instead has successfully recovered the camera system geometry as well as the 3D pose of the mirrors using a small set of views, 12 snapshots, each with a different mirror position. The overall recovered geometry is shown in Fig.8 (b) and Fig.8 (c). We have further refined the system geometry using BA on a set of 5000 feature points extracted from the actual scene. Similar to the dome configuration, the accuracy of the estimated geometry has been evaluated by computing the backprojection error from the reconstructed 3D point cloud and from the pairwise transfer of the BW checkerboard points. The

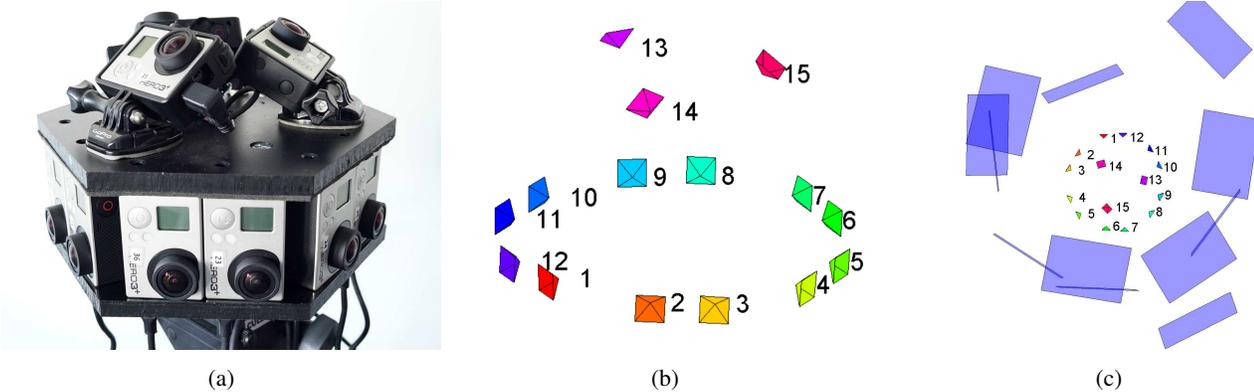


Figure 8: VR rig multicamera system (a). VR rig geometry recovered using the proposed approach, based on mirrored epipoles (b). Top view of the same geometry together with the recovered mirror positions (c).



Figure 9: Checkerboard with the detected (circle) and back-projected (dot) grid corners, using the system geometry, estimated with epipoles.

	Feature err., px	Pattern err., px
Mirrored epipoles	$\mu = 1.4, \sigma = 1.5$	$\mu = 2.4, \sigma = 0.6$

Table 2: Backprojection error for the proposed Mirrored Epipole-based approach.

numerical results are shown in Table 2.

5. Conclusion

We have proposed an algorithm aimed at the estimation of the geometry of a multicamera system, based on the enforcement of simple constraints arising only from the visible epipoles. This approach is particularly suitable for dome-like indoor multicamera systems, typically used for motion capture and 3D reconstruction. In such configurations indeed the number of visible epipoles is usually large enough to provide sufficient constraints, and an acceptable solution can be reached very fast, whereas other techniques based on calibration object are not as flexible and easy to

use. A great advantage of our solution is the capability of a fast recovery of the external calibration of a multicamera system in case of rearrangement, removal or inclusion of cameras, with almost no overhead. Other algorithms instead require each time completion of a complex and time consuming calibration routine. We have also presented an extension of the calibration algorithm aimed at the integration of the epipoles visible in a planar mirror reflection into the calibration dataset. This solution leads to a sensible enlargement of the epipole set and in practice makes the system suitable for any camera geometry, dome-like, fronto-parallel and divergent circular VR-rig configuration, such as Google JUMP rig. We have shown how the mirrored epipoles can be simply integrated within a unified estimation framework by including additional terms into a single calibration objective function. The evaluation on the synthetic and real datasets has shown a high robustness of the proposed approach against image noise, which implies a capability of reaching in a quasi automatic way an accuracy of the camera system geometry high enough for initialization of a successive BA 3D reconstruction pipeline. This is particularly useful, as a typical bundle adjustment pipeline requires a rather precise initial guess in order to converge. The practical advantages such as simplicity and time efficiency of the calibration procedure, reduced dataset collection and high robustness enable the integration of a flexible calibration architecture, capable of handling the rearrangement of the camera geometry without requiring an extra effort.

Acknowledgement

The work has been partially supported by the ERC-Starting Grant (Dynamic MinVIP) and the Cluster of Excellence rebirth. The authors gratefully acknowledge the support.

References

- [1] P. Baker and Y. Aloimonos. Complete calibration of a multi-camera network. pages 134–141, 2000. [1](#)
- [2] P. T. Baker and Y. Aloimonos. Calibration of a multicamera network. *CVPRW 2003. Conference on*, 7(1):72–78, 2003. [1](#)
- [3] A. Chatterjee and V. M. Govindu. Efficient and robust large-scale rotation averaging. In *ICCV*, pages 521–528. IEEE, 2013. [3](#)
- [4] O. Faugeras, Q.-T. Luong, and T. Papadopolou. *The Geometry of Multiple Images: The Laws That Govern The Formation of Images of A Scene and Some of Their Applications*. MIT Press, Cambridge, MA, USA, 2001. [1](#)
- [5] J.-M. Frahm, K. Kser, and R. Koch. Pose estimation for multi-camera systems. *Pattern Recognition, Lecture Notes in Computer Science*, 3175:286–293, 2004. [1](#)
- [6] M. Grossberg and S. Nayar. Modeling the Space of Camera Response Functions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(10):1272–1282, Oct 2004. [1](#)
- [7] R. I. Hartley, J. Trampf, Y. Dai, and H. Li. Rotation averaging. *International Journal of Computer Vision*, 103(3):267–305, 2013. [3](#)
- [8] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521540518, second edition, 2004. [1](#)
- [9] F. Kahl and R. I. Hartley. Multiple-view geometry under the linfinity-norm. *IEEE Trans. Pattern Anal. Mach. Intell.*, 30(9):1603–1617, 2008. [3](#)
- [10] F. Kahlesz, C. Lilge, and R. Klein. Easy-to-use calibration of multiple-camera setups. 2007. [1](#)
- [11] K. Kanatani. Analysis of 3-d rotation fitting. *IEEE Trans. Pattern Anal. Mach. Intell.*, 16(5):543–549, 1994. [3](#)
- [12] R. Kaucic, R. I. Hartley, and N. Y. Dano. Plane-based projective reconstruction. In *ICCV*, pages 420–427, 2001. [3](#)
- [13] R. Kumar, A. Ilie, J.-M. Frahm, and M. Pollefeys. Simple calibration of non-overlapping cameras with a mirror. pages 1–7, June 2008. [2](#)
- [14] D. Martinec and T. Pajdla. Robust rotation and translation estimation in multiview reconstruction. In *CVPR*. IEEE Computer Society, 2007. [3](#)
- [15] C. Rother and S. Carlsson. Linear multi view reconstruction and camera recovery. In *ICCV*, pages 42–51, 2001. [3](#)
- [16] T. Svoboda, D. Martinec, and T. Pajdla. A convenient multicamera self-calibration for virtual environments. *Presence: Teleoper. Virtual Environ.*, 14(4):407–422, 2005. [1](#), [6](#)
- [17] Z. Zhang. A flexible new technique for camera calibration. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(11):1330–1334, 2000. [1](#)
- [18] Z. Zhang. Camera calibration, 2014. in G. Medioni and S.B. Kang, eds., *Emerging Topics in Computer Vision*. [1](#)