Structure from Motion Using Rigidly Coupled Cameras without Overlapping Views

Sandro Esquivel and Reinhard Koch

Christian-Albrechts-University, Kiel, Germany

Abstract. Structure from Motion can be improved by using multi-camera systems without overlapping views to provide a large combined field of view. The extrinsic calibration of such camera systems can be computed from local reconstructions using *hand-eye calibration* techniques. Nevertheless these approaches demand that motion constraints resulting from the rigid coupling of the cameras are satisfied which is in general not the case for decoupled pose estimation. This paper presents an extension to Structure from Motion using multiple rigidly coupled cameras that integrates rigid motion constraints already into the local pose estimation step, based on dual quaternions for pose representation. It is shown in experiments with synthetic and real data that the overall quality of the reconstruction process is improved and pose error accumulation is counteracted, leading to more accurate extrinsic calibration.

1 Introduction

Multi-camera systems play an important role in computer vision and are used in various tasks such as object tracking, 3d reconstruction, and Augmented Reality. Combining cameras with minimal or no shared viewing areas is especially helpful for *Structure from Motion* (SfM) applications, i.e. 3d reconstruction of *a priori* unknown scenes from video sequences, due to their large combined field of view. Existing approaches [18, 8, 7] need an accurate intrinsic and extrinsic calibration of the cameras. While extrinsic calibration is solved in general by matching features between camera images or using calibration objects that are visible in all cameras [12, 19], calibration without overlapping views is far less treated in literature, starting with Caspi & Irani for colocated cameras [1].
It is convenient to reduce this problem to *hand-eye calibration* [4, 5, 11]. These approaches – dubbed *eye-to-eye calibration* in the context of this work – compute the pose transformation between the cameras from the rigidity constraints on relative motions of the cameras. Typically, relative poses are computed first for each camera individually from synchronously captured images using monocular SfM. Afterwards the relative pose of cameras to each other are recovered from the rigidi motion constraints by solving the equation $\mathbf{AX} = \mathbf{XB}$ known from robotic hand-eye calibration, where $\mathbf{A}$, $\mathbf{B}$ are relative pose transformations of the "hand" – a mobile gripper – and the "eye" – a visual sensor mounted on the gripper – at the same instant of time and $\mathbf{X}$ is the Euclidean transformation from the "eye" to the "hand" coordinate frame. However, due to noise resp. estimation errors,

the rigidity constraint between $\mathbf{A}$ and $\mathbf{B}$ is not satisfied in practice, especially when the poses are computed by SfM. This constraint violation induces errors in the calibration step. Lébraly et al. [11] propose to use a bundle adjustment with minimal parametrization for local camera poses – using the local poses of a designated *master* camera for each frame and the time-independent relative poses between the other cameras and the master camera – as a post-processing step while computing the initial extrinsic calibration from deficient local poses.

Our Approach. In this work we will integrate the enforcement of rigidity constraints into the pose estimation step of multi-camera Structure from Motion, in contrast to existing approaches which ignore it or try to remedy it after extrinsic calibration, e.g. using global optimization techniques. This is done by parametrizing motions properly to model the constraints explicitely.
First, we will address the motion constraints on rigidly coupled cameras explicitely and interpret them geometrically. Since the formulation of these constraints depends on the parametrization of the camera poses, we will consider different suitable parametrizations based on *screw motions*.
In the main part we will present a multi-camera SfM framework that considers rigid motion constraints already during intialization and local pose estimation while the eye-to-eye transformation is yet unknown.
We evaluate the initialization and pose estimation steps of rigidly coupled multi-camera SfM with synthetic data, rendered and real video sequences using an implementation of our approach, as well as afterward eye-to-eye calibration.

2 Motion Constraints of Rigidly Coupled Cameras

In the following we will consider N rigidly coupled cameras with local coordinate frames $\mathcal{C}_k$, $k = 1, \dots, N$, captured at M local poses $\mathbf{T}_k^\ell$ with respect to the reference coordinate frames $\mathcal{C}_k$. The transformation $\Delta\mathbf{T}_k$ between the local coordinate frames of the k-th camera and the first camera (*master* camera) – denoted as *eye-to-eye transformation* in the context of this work – satisfies the *rigid motion equation* illustrated in fig. 1:

$$\mathbf{T}_1^\ell \Delta\mathbf{T}_k = \Delta\mathbf{T}_k \mathbf{T}_k^\ell \tag{1}$$

Each $\mathbf{T}$ describes a Euclidean transformation consisting of rotation $\mathbf{R}$ and translation $\boldsymbol{t}$. Hence the rigid motion constraint (1) can be decomposed into a rotation constraint $\mathbf{R}_1^\ell \Delta\mathbf{R}_k = \Delta\mathbf{R}_k \mathbf{R}_k^\ell$ and a translation constraint $\mathbf{R}_1^\ell \Delta\boldsymbol{t}_k + \boldsymbol{t}_1^\ell = \Delta\mathbf{R}_k \boldsymbol{t}_k^\ell + \Delta\boldsymbol{t}_k$ for rigidly coupled motion pairs $(\mathbf{T}_1^\ell, \mathbf{T}_k^\ell)$.
Solving (1) for $\Delta\mathbf{T}_k$ is called *eye-to-eye calibration*. Approaches differ mainly by the motion parametrization. Tsai & Lenz [17] showed that at least two general motions with different rotation axes are needed to solve this problem inambiguously, independent of the chosen pose parametrization.
The rigidity constraints on local motions of rigidly coupled sensors were described within the context of hand-eye calibration by Chen [2] as the *congruence theorem*. The constraints include a *rotation angle constraint* and a constraint on

the rotation axis and the translation parallel to the rotation axis, referred to as the *pitch constraint* in this work. They become obvious when motions are considered as *screw motions*: A general motion of a camera with translation $\boldsymbol{t}$ can be decomposed into a rotation by angle α around a rotation axis $\boldsymbol{r}$, a translation $p\boldsymbol{r}$ with $p = \boldsymbol{t}^\top \boldsymbol{r}$ parallel to the rotation axis $\boldsymbol{r}$, and a translation $\boldsymbol{u} = \boldsymbol{t} - p\boldsymbol{r}$ orthogonal to the rotation axis $\boldsymbol{r}$. This can equally be described by a rotation by α around a line in space with direction $\boldsymbol{r}$ followed by a slide of length p along the line (*Chasles' theorem*), formulated algebraically either as a screw motion [2] or equivalently by means of dual quaternions [3].

Rotation Angle Constraint. Given N rigidly coupled local rotations $\mathbf{R}_k$ with rotation angles $\alpha_k \in [0, \pi]$ around rotation axes $\boldsymbol{r}_k$, all rotation angles α_k must be equal. It can be easily proved from eq. (1) that the following equivalence is valid:

$$\mathbf{R}_1 = \Delta\mathbf{R}_k \mathbf{R}_k \Delta\mathbf{R}_k^\top \quad \Leftrightarrow \quad \boldsymbol{r}_1 = \Delta\mathbf{R}_k \boldsymbol{r}_k \text{ and } \alpha_1 = \alpha_k \quad \text{for all } 2 \leq k \leq N \tag{2}$$

Hence the local rotation estimation problem for N rigidly coupled cameras can be reduced to the estimation of N rotation axes $\boldsymbol{r}_k$ and a single angle $\alpha \in [0, \pi]$.

Pitch Constraint. Given N rigidly coupled local motions $(\mathbf{R}_k, \boldsymbol{t}_k)$, the amount of translation parallel to the rotation axes, the so called *pitch* $p_k = \boldsymbol{t}_k^\top \boldsymbol{r}_k$, must be equal for all motions. This also follows directly from eq. (1):

$$\boldsymbol{t}_k^\top \boldsymbol{r}_k = \underbrace{\Delta\boldsymbol{t}_k^\top (\mathbf{R}_1^\top - \mathbf{I})\boldsymbol{r}_1}_{=0} + \boldsymbol{t}_1^\top \boldsymbol{r}_1 \quad \text{for all } 2 \leq k \leq N \tag{3}$$

Hence the local translation estimation problem for N rigidly coupled cameras can be reduced by $N-1$ parameters when the local rotation axes are known.
The pitch constraint holds when the local translations of all cameras have the same scale. Otherwise, it yields the scalar factor between the coordinate frames of the cameras, given that the motion is not planar. Assume that all translations are only known up to scale, i.e $\hat{\boldsymbol{t}}_k$ with $s_k \hat{\boldsymbol{t}}_k = \boldsymbol{t}_k$ are given. Then the relative scale $\lambda_k = \frac{s_k}{s_1}$ from the k-th to the master coordinate frame is given by:

$$\lambda_k = \frac{\hat{\boldsymbol{t}}_1^\top \boldsymbol{r}_1}{\hat{\boldsymbol{t}}_k^\top \boldsymbol{r}_k} \text{ and } s_k = s_1 \lambda_k \quad \text{for all } 2 \leq k \leq N \tag{4}$$

2.1 Rigidly Coupled Pose Parametrization

To enforce the rotation angle constraint and pitch constraint by using a common parameter for rotation angle and pitch respectively, a pose parametrization is needed that is computationally simple, has minimal redundancy, and explicitely decouples angle and pitch from the remaining pose parameters. Hence, *dual quaternions* provide a natural choice for rigidly coupled pose parametrization.
Quaternions of unit length (see e.g. [10]), represented as 3d-vector/scalar pairs

$\mathbf{q} = (\boldsymbol{q}, q)$, are commonly used for rotation estimation in computer vision and computer graphics since they provide an efficient and only slightly overparameterized representation, do not suffer from singularities, and are still computationally easy to apply. Related to the angle/axis representation, they also explicitely state the rotation angle. *Dual quaternions* $\check{\mathbf{q}} = (\mathbf{q}, \mathbf{q}')$ (see e.g. [3]) provide an equally elegant way to describe rigid motions in 3d space, closely related to the screw axis, angle, and pitch of screw motions. The dual quaternion describing a rotation by angle α around axis $\boldsymbol{r}$ and translation by $\boldsymbol{t}$ is given by the pair:

$$\mathbf{q} = \left(\sin(\frac{\alpha}{2})\boldsymbol{r}, \cos(\frac{\alpha}{2})\right) \quad \mathbf{q}' = \frac{1}{2}\left(\cos(\frac{\alpha}{2})\boldsymbol{t} + \sin(\frac{\alpha}{2})(\boldsymbol{t} \times \boldsymbol{r}), -\sin(\frac{\alpha}{2})\boldsymbol{t}^\top \boldsymbol{r}\right) \quad (5)$$

where $\mathbf{q}$ is referred to as the *real* and $\mathbf{q}'$ as the *dual* part of the dual quaternion. The rigid motion of a 3d point $\boldsymbol{X}$ is described by $\mathbf{R}(\mathbf{q})\boldsymbol{X} + \boldsymbol{t}(\mathbf{q}, \mathbf{q}')$ with:

$$\mathbf{R}(\mathbf{q}) = \mathbf{I} + 2q[\boldsymbol{q}]_\times + 2[\boldsymbol{q}]_\times^2 \quad \boldsymbol{t}(\mathbf{q}, \mathbf{q}') = 2(q\boldsymbol{q}' - q'\boldsymbol{q} + \boldsymbol{q} \times \boldsymbol{q}') \quad (6)$$

where $[\boldsymbol{q}]_\times$ is the skew-symmetric matrix describing the cross product with $\boldsymbol{q}$. The space of such dual quaternions is constrained by $\|\mathbf{q}\| = 1$ and $\mathbf{q}^\top \mathbf{q}' = 0$, i.e. $\|\check{\mathbf{q}}\| = 1$. Since $\mathbf{q}$ and $-\mathbf{q}$ describe the same rotation, we will restrict the real quaternion part to $q \geq 0$. Under this limitation, we conclude from eq. (2) and (3) that all dual quaternions representing rigidly coupled motions have equal scalar parts (q, q'), since these only depend on the rotation angle and motion pitch.

Minimal parametrization. Using unit dual quaternions to represent rigidly coupled motion, the rigid motion constraint can be simply enforced via parameter reduction, i.e. using a single parameter pair (q, q') for the scalar quaternion part. Nonetheless, the parametrization is not minimal, so we will need additional constraints, i.e. unit length constraints $\|\mathbf{q}\| = 1$ and orthogonality constraints $\mathbf{q}^\top \mathbf{q}' = 0$, which is not recommended for nonlinear optimization. We consider a minimal parametrization for nonlinear optimization instead. Assuming w.l.o.g. that r_z is the largest absolute element of the rotation axis $\boldsymbol{r}$ with sign σ_z, we can enforce the unit quaternion constraint by replacing $\boldsymbol{q}$ with $(q_x, q_y, \sigma_z\sqrt{1 - q_x^2 - q_y^2 - q^2})$. This yields a valid parametrization as long as the rotation is not too small. The orthogonality constraint is enforced by replacing $\boldsymbol{q}'$ with $(q_x', q_y', -\frac{q_x q_x' + q_y q_y' + q q'}{\sigma_z \sqrt{1 - q_x^2 - q_y^2 - q^2}})$. A second strategy is to renormalize the non-minimal parameters prior to pose evaluation. This is equivalent to using $\sqrt{1 - q^2}\frac{\boldsymbol{q}}{\|\boldsymbol{q}\|}$ and $\boldsymbol{q}' - \frac{\boldsymbol{q}^\top \boldsymbol{q}' + q q'}{\boldsymbol{q}^\top \boldsymbol{q}}\boldsymbol{q}$ instead of $\boldsymbol{q}$ and $\boldsymbol{q}'$. For an extensive review of minimal rotation parametrization for nonlinear optimization see e.g. [14].

3 Rigidly Coupled Multi-Camera Structure from Motion

In the following we will describe how to enforce rigid motion constraints into the two main stages of SfM for multiple coupled cameras prior to eye-to-eye calibration. We will show that the resulting reconstruction yields more consistent results which will improve the global registration of all cameras.

3.1 Relative Pose Estimation

Relative pose estimation from two views evaluating the epipolar geometry is commonly used as the first step in SfM applications (see [6]). Given n corresponding normalized image points $\boldsymbol{x}_i$, $\boldsymbol{x}'_i$ of 3d points $\boldsymbol{X}_i$ in two camera images $\mathcal{I}$ and $\mathcal{I}'$ with relative pose $(\mathbf{R}, \boldsymbol{t})$, i.e.$\boldsymbol{x} \sim \boldsymbol{X}$, $\boldsymbol{x}' \sim \boldsymbol{X}' = \mathbf{R}\boldsymbol{X} + \boldsymbol{t}$, the essential matrix $\mathbf{E}$ can be written as $\mathbf{E} = \mathbf{R}^\top [\boldsymbol{t}]_\times$. $\mathbf{E}$ is a non-zero 3×3 matrix of rank 2 that satisfies the *epipolar constraint* (a.k.a. *Longuet-Higgins equation* [6]):

$$\boldsymbol{x}_i^\top \mathbf{E} \boldsymbol{x}'_i = 0 \quad \text{for all } i = 1, \ldots, n \tag{7}$$

An illustration for rigidly coupled cameras is shown in fig. 1.

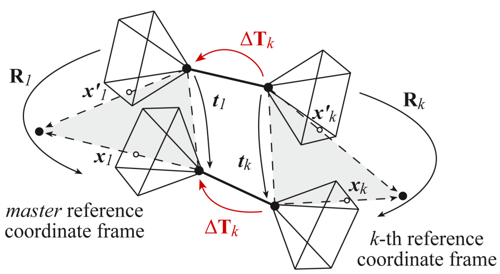

Fig. 1. Local relative pose estimation from 2d-2d correspondences $\boldsymbol{x} \leftrightarrow \boldsymbol{x}'$ for rigidly coupled cameras.

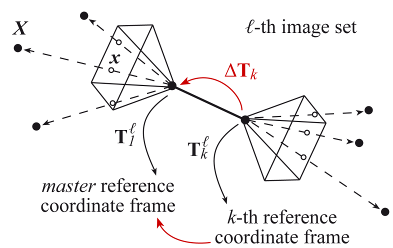

Fig. 2. Local pose estimation from 2d-3d correspondences $\boldsymbol{x} \leftrightarrow \boldsymbol{X}$ for rigidly coupled cameras.

Enforcing the rotation angle constraint. As a first step, the essential matrices $\mathbf{E}_1, \ldots, \mathbf{E}_N$ of all cameras are computed individually from their respective image pairs using traditional algorithms. They are then decomposed into rotations $\mathbf{R}_1, \ldots, \mathbf{R}_N$ resp. unit quaternions $\mathbf{q}_1, \ldots, \mathbf{q}_N$ and translation vectors $\hat{\boldsymbol{t}}_1, \ldots, \hat{\boldsymbol{t}}_N$ up to individual scales (see e.g. [6]). W.l.o.g. we assume that all $\|\hat{\boldsymbol{t}}_i\| = 1$.
The next step is to re-estimate the relative rotations subject to the rotation angle constraint. A straight-forward approach is to average the scalar parts q_k of all unit quaternions $\mathbf{q}_k$ resulting in the constrained rotation parametrization $(q, \boldsymbol{q}_1, \ldots, \boldsymbol{q}_N)$ with $q = \frac{1}{N} \sum_{k=1}^N |q_i|$ and re-normalized vector parts $\boldsymbol{q}_k$. This averaging however does not take the epipolar constraints into account. This can be done by reformulating eq. (7) in terms of unit quaternions and minimizing the joint epipolar error function for all cameras at the same time.
Parametrizing rotations by unit quaternions $\mathbf{q}_k = (\boldsymbol{q}_k, q)$ and translations by unit vectors $\hat{\boldsymbol{t}}_k$, we obtain from eq. (7) for each point correspondence $(\boldsymbol{x}_{k,i}, \boldsymbol{x}'_{k,i})$ one constraint $f_{k,i}^{\text{epi}}(q, \boldsymbol{q}_k, \hat{\boldsymbol{t}}_k) = 0$ which can be formulated as a cubic equation of the parameter vector:

$$f_{k,i}^{\text{epi}}(q, \boldsymbol{q}_k, \hat{\boldsymbol{t}}_k) = \boldsymbol{x}_{k,i}^\top \mathbf{R}((\boldsymbol{q}_k, -q))[\hat{\boldsymbol{t}}_k]_\times \boldsymbol{x}'_{k,i} = 0 \tag{8}$$

Enforcing all $(\boldsymbol{q}_k, q)$ and $\hat{\boldsymbol{t}}_k$ to have unit length in nonlinear optimization of (8) can be achieved by using one of the methods described in 2.1, i.e. minimal parametrization of unit quaternions and unit translation vectors, or renormalizing quaternions and translations. In this work we minimize the joint error function $\sum_{k=1}^{N}\sum_{i=1}^{n_k} f_{k,i}^{\mathrm{epi}}(q, \boldsymbol{q}_k, \hat{\boldsymbol{t}}_k)^2$ using the Levenberg-Marquardt algorithm [13]. The starting point is given by the averaged unit quaternions.

Enforcing the pitch constraint. Given that the motion is non-planar, we can rescale all estimated translations $\hat{\boldsymbol{t}}_k$ with respect to the master coordinate frame as described by eq. (4), yielding absolute translations $\boldsymbol{t}_k$ that satisfy the pitch constraint, i.e. $p = \boldsymbol{t}_1^\top \boldsymbol{r}_1 = \cdots = \boldsymbol{t}_1^\top \boldsymbol{r}_N$ where $\boldsymbol{r}_k = \frac{\boldsymbol{q}_k}{\|\boldsymbol{q}_k\|}$ is the rotation axis of the k-th resulting unit quaternion. The final parameters of the initial camera poses are then given by $(q, q', \boldsymbol{q}_1, \boldsymbol{q}_1', \ldots, \boldsymbol{q}_N, \boldsymbol{q}_N')$ where the dual quaternion parts are $q' = -\frac{1}{2}(\sqrt{1-q^2})p$ and $\boldsymbol{q}_k' = \frac{1}{2}(q\boldsymbol{t} + \boldsymbol{t} \times \boldsymbol{q}_k)$ derived from eq. (5).

3.2 Pose Estimation

After initialization, the local reconstruction part of the SfM algorithm starts, i.e. local 2d-3d correspondences are tracked within subsequent camera images, the local camera poses are re-estimated from 2d-3d correspondences, and new 3d points are computed from 2d-2d correspondences using the current pose. While any approach could be used for local SfM, we recommend to use an approach based on dual quaternions for motion representation such as [15] in order to prevent frequent conversions between different parametrizations. Given n 2d-3d correspondences $(\boldsymbol{x}_i, \boldsymbol{X}_i)$ for an image $\mathcal{I}$ such that $\mathbf{R}^\top(\boldsymbol{X}_i - \boldsymbol{t}) \sim \boldsymbol{x}_i$, the local camera pose $(\mathbf{R}, \boldsymbol{t})$, should minimize the normalized *reprojection error*:

$$f(\mathbf{R}, \boldsymbol{t}) = \sum_{i=1}^{n} \|\mathbf{P}(\mathbf{R}^\top(\boldsymbol{X}_i - \boldsymbol{t})) - \mathbf{P}(\boldsymbol{x}_i)\|^2 \tag{9}$$

where $\mathbf{P} : \mathbb{R}^3 \to \mathbb{R}^2, \boldsymbol{X} \mapsto (X_x/X_z, X_y/X_z)$ denotes the perspective projection. Pose estimation from local 2d-3d correspondences is illustrated in fig. 2.

Enforcing the rigid motion constraint. Given N rigidly coupled cameras and n_k local 2d-3d correspondences $(\boldsymbol{x}_{k,i}, \boldsymbol{X}_{k,i})_{i=1,\ldots,n_k}$ for the k-th camera, the local poses are first computed from individual monocular SfM, and then converted to the dual quaternion representation $(q_k, q_k', \boldsymbol{q}_k, \boldsymbol{q}_k')_{k=1,\ldots,N}$ as described in sec. 2.1. Similar to the relative pose problem, the rigid motion constraint can be enforced by simply averaging the scalar parts in the constrained parametrization $(q, q', \boldsymbol{q}_1, \boldsymbol{q}_1', \ldots, \boldsymbol{q}_N, \boldsymbol{q}_N')$. This parametrization is again not optimal with respect to the reprojection error, but is instead used as an initial solution for minimization of the joint error function $\sum_{k=1}^{N}\sum_{i=1}^{n_k} f_{k,i}^{\mathrm{reproj}}(q, q', \boldsymbol{q}_k, \boldsymbol{q}_k')$ where

$$f_{k,i}^{\mathrm{reproj}}(q, q', \boldsymbol{q}_k, \boldsymbol{q}_k') = \|\mathbf{P}(\mathbf{R}(\boldsymbol{q}_k, -q)[\boldsymbol{X}_{k,i} - \boldsymbol{t}((\boldsymbol{q}_k, q), (\boldsymbol{q}_k', q'))]) - \mathbf{P}(\boldsymbol{x}_{k,i})\|^2 \tag{10}$$

Enforcing the unit length constraint for all dual quaternions in nonlinear optimization of (10) using the Levenberg-Marquardt algorithm is again achieved by the minimal parametrization discussed in 2.1, or renormalizing the parameters.

3.3 Registration and Global Reconstruction

As soon as we have computed local poses that are suitable for eye-to-eye calibration – i.e. the angle between the rotation axes of the first and current motion is above a threshold θ for each camera [17] – the eye-to-eye transformations $\Delta \mathbf{T}_k$ between k-th camera and master camera can be estimated. It is convenient to use a dual quaternion based approach [3] but any approach could be used.
After eye-to-eye calibration the local 3d reconstructions can be merged into a global reconstruction, and optimized by a multi-camera bundle adjustment as described in [11]. This approach refines the reprojection error with respect to 3d points, master camera poses, and eye-to-eye transformations – ensuring that the resulting camera poses satisfy the rigid motion constraint.
Afterwards, SfM can proceed using an approach with known eye-to-eye transformations such as [8] to enhance the stability of the pose tracking. The eye-to-eye transformations can be further refined over time.

4 Tests and Evaluation

4.1 Relative and Absolute Pose Estimation

First we evaluated relative and absolute pose estimation with and without rigid motion constraint enforcement (RMCE) for a large number of random configurations (1000/test) for $N = 2$ cameras with 1 cm distance and $30 - 120°$ rotation difference. Gaussian noise with $\sigma \approx 0.2\%$ image size was added to all points. The *Gold Standard* methods were used for decoupled pose estimation [6]. Relative poses were scaled using the pitch constraint while the master translation length was fixed to 1 m. Fig. 3 shows that the average estimation error improves for absolute rotation estimation, and in general when the number of correspondences is < 15. We also conducted tests with increasing number of cameras and minimal parameters vs. renormalization but the results did not change significantly.

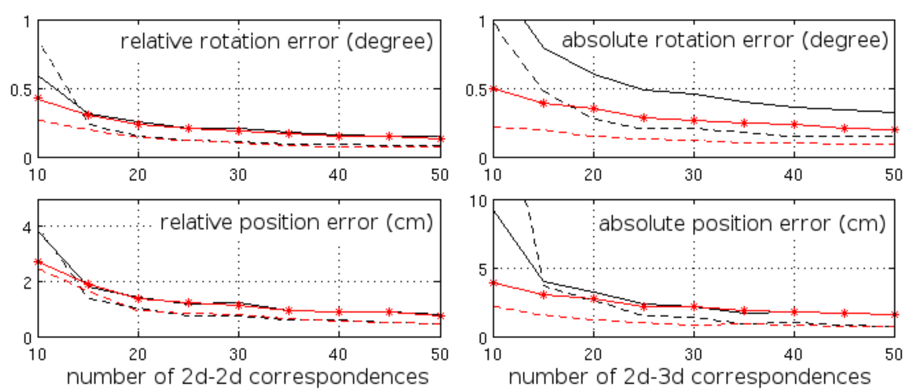

Fig. 3. Average relative/absolute pose error per camera and test, w/o RMCE (upper black line) and with RMCE (lower red line, dashed lines show standard deviation).

To evaluate the impact on eye-to-eye calibration, we further estimated $M = 3, \ldots, 8$ poses with and w/o RMCE and computed the eye-to-eye transformation between $N = 2$ cameras using the method from [3]. The average results for 1000 random datasets per number of poses are improved as shown in fig. 4.

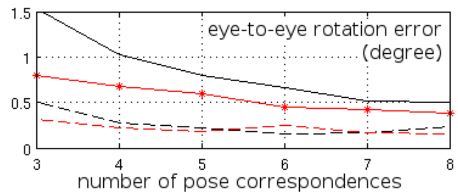

Fig. 4. Average eye-to-eye calibration error from increasing number of random poses estimated w/o RMCE (upper black line) and with RMCE (lower red line).

4.2 Structure from Motion with Rendered Video

The full SfM approach from 4.1 (i.e. without bundle adjustment) with and without RMCE is tested first with a rendered image sequence of a scene consisting of textured boxes. The camera system contains $N = 2$ cameras that are set 25 cm and 15° apart. The motion trajectory covers ca. 1.5 m and 60° rotation. Note that we use a very similar setup as the real scenario in sec. 4.3. Feature points are detected and tracked in subsequent images using a KLT tracker [16]. Relative pose estimation is performed between the 1 and 21st image, followed by absolute pose estimation every 20 images. The relative poses are scaled as in the previous test. Fig. 5 shows that the accuracy of pose estimation is improved by RMCE, especially pose drift over time is counteracted. Eye-to-eye calibration is performed afterwards using every 20-th image. The calibration error is improved from 1.21° and 1.9 cm w/o RMCE to 0.49° and 1.27 cm for this setup.

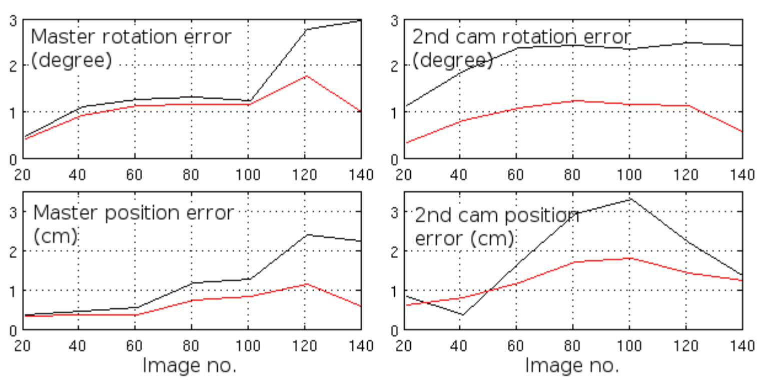

Fig. 5. Pose error during SfM w/o RMCE (upper black line) and with RMCE (lower red line) for a rendered video sequence (total motion covers ca. 1.5m, 60°).

4.3 Structure from Motion with Real Video

A real video sequence was captured with a stereo camera setup consisting of $N = 2$ cameras mounted onto a rig with ca. 24.5 cm $\pm$ 0.8 horizontal offset and $17.2 \pm 0.62^\circ$ rotation difference, viewing a box scene similar to the virtual test case in sec. 4.2 (see example image in fig. 7). Note that we use a setup with partly overlapping views to recover extrinsics for comparison using a default stereo calibration approach. The proposed SfM procedure was applied for every 10-th image and eye-to-eye calibration was computed afterwards from every 20-th estimated local pose pair. The difference between stereo calibration and eye-to-eye calibration is 0.71°, 4.1 cm w/o RMCE, and 0.54°, 2.5 cm with RMCE. Fig. 6 shows how much the rigid motion constraint between master and 2nd camera is violated during unconstrained SfM, especially over time.

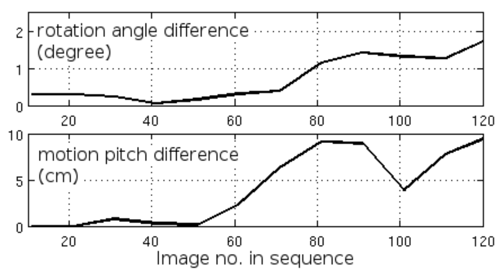

Fig. 6. Rigid motion constraint violation during SfM w/o RMCE for a real video sequence (total motion covers ca. 1m, 50°).

Fig. 7. Example image from real video sequence.

5 Conclusions

In this work we have revisited rigid motion constraints for local pose estimation of rigidly coupled cameras, and described how to incorporate these constraints into existing SfM algorithms. The resulting constrained local poses have been shown to be more robust against drift and improve extrinsic calibration of the multi-camera setup using calibration methods that depend on the rigid motion constraints. Especially SfM resp. extrinsic calibration with few input poses benefits from the constraint enforcement. The results presented here are very useful for systems consisting of few cameras covering the full 360° field of view.

Future Work: The results presented in this work are designed for general motions of the camera system. Further inspection of common degenerate motion classes such as planar motion could be done to modify the multi-camera SfM so that these cases can also be handled by integrating additional knowledge about the scene and motion. Although we have presented a very basic SfM approach here for the sake of clarity, we should also be able to integrate our method easily into more sophisticated and efficient SfM approaches such as e.g. PTAM [9].

References

1. Caspi, Y., Irani, M.: Alignment of non-overlapping sequences. International Journal of Computer Vision 48(1), 39–51 (2002)
2. Chen, H.H.: A screw motion approach to uniqueness analysis of head-eye geometry. In: Computer Vision and Pattern Recognition (CVPR'91). pp. 145–151 (1991)
3. Daniilidis, K.: Hand-eye calibration using dual quaternions. International Journal of Robotics Research 18, 286–298 (1999)
4. Dornaika, F., Chung, C.k.R.: Stereo geometry from 3d ego-motion streams. IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics 33(2), 308–323 (April 2003)
5. Esquivel, S., Woelk, F., Koch, R.: Calibration of a multi-camera rig from non-overlapping views. In: 29th DAGM Symposium on Pattern Recognition (DAGM'07). LNCS, vol. 4713, pp. 82–91. Heidelberg, Germany (September 2007)
6. Hartley, R.I., Zisserman, A.: Multiple View Geometry in Computer Vision (Second Edition). Cambridge University Press, second edn. (2004)
7. Kim, J.H., Li, H., Hartley, R.I.: Motion estimation for nonoverlapping multicamera rigs: Linear algebraic and l_∞ geometric solutions. IEEE Transactions on Pattern Analysis and Machine Intelligence 32(6), 1044–1059 (2010)
8. Kim, J.H., Chung, M.J.: Absolute motion and structure from stereo image sequences without stereo correspondence and analysis of degenerate cases. Pattern Recognition 39(9), 1649–1661 (2006)
9. Klein, G., Murray, D.: Parallel tracking and mapping for small AR workspaces. In: 6th IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR'07) (2007)
10. Kuipers, J.B.: Quaternions and rotation sequences. In: International Conference on Geometry, Integrability and Quantization (GEOM'99). pp. 127–143 (1999)
11. Lébraly, P., Royer, E., Ait-Aider, O., Deymier, C., Dhome, M.: Fast calibration of embedded non-overlapping cameras. In: IEEE International Conference on Robotics and Automation. pp. 221–227. Shanghai, China (May 2011)
12. Luong, Q.T., Faugeras, O.: Self-calibration of a stereo rig from unknown camera motions and point correspondences (1993)
13. Moré, J.J.: The Levenberg-Marquardt algorithm: Implementation and theory. Numerical Analysis 630, 105–116 (1978)
14. Schmidt, J., Niemann, H.: Using quaternions for parametrizing 3-d rotations in unconstrained nonlinear optimization. In: Vision, Modeling, and Visualization (VMV'01). pp. 399–406. Stuttgart, Germany (2001)
15. Shao, L., Walker, M.W.: Estimating 3-d location parameters using dual number quaternions. In: Computer Vision, Graphics, and Image Processing (CVGIP): Image Understanding. vol. 54 (3), pp. 358–367 (1991)
16. Tomasi, C., Kanade, T.: Detection and tracking of point features. Tech. Rep. CMU-CS-91-132, Carnegie Mellon University (1991)
17. Tsai, R.Y., Lenz, R.K.: A new technique for fully autonomous and efficient 3d robotics hand/eye calibration. IEEE Transactions on Robotics and Automation 5(3), 345–358 (1989)
18. Weng, J., Huang, T.S.: Complete structure and motion from two monocular sequences without stereo correspondence. In: International Conference on Pattern Recognition (ICPR'92). pp. 651–654 (1992)
19. Zisserman, A., Beardsley, P.A., Reid, I.D.: Metric calibration of a stereo rig. In: IEEE Workshop on Representations of Visual Scenes. pp. 16–23 (1995)