Solving the Plane-Sphere Ambiguity in Top-Down Structure-from-Motion

Lars Haalck, Benjamin Risse
Institute for Geoinformatics and Institute for Computer Science
University of Münster, Germany
{lars.haalck, b.risse}@uni-muenster.de

Abstract

Drone-based land surveys and tracking applications with a moving camera require three-dimensional reconstructions from videos recorded using a downward facing camera and are usually generated by Structure-from-Motion (SfM) algorithms. Unfortunately, monocular SfM pipelines can fail in the presence of lens distortion due to a critical configuration resulting in a plane-sphere ambiguity which is characterized by severe curvatures of the reconstructions and erroneous relative camera pose estimations. We propose a 4-point minimal solver for the relative pose estimation for two views sharing the same radial distortion parameters (i.e. from the same camera) with a viewing direction perpendicular to the ground plane. To extract 3D reconstructions from continuous videos, the relative pose of pairwise frames is estimated by using the solver with RANSAC and the Sampson error where globally consistent distortion parameters are determined by taking the medial of all values. Moreover, we propose an additional regularizer for the final bundle adjustment to remove any remaining curvature of the reconstruction if necessary. We tested our methods on synthetic and real-world data and our results demonstrate a significant reduction of curvature and more accurate relative pose estimations. Our algorithm can be easily integrated into existing pipelines and is therefore a practical solution to resolve the plane-sphere ambiguity in a variety of top-down SfM applications.

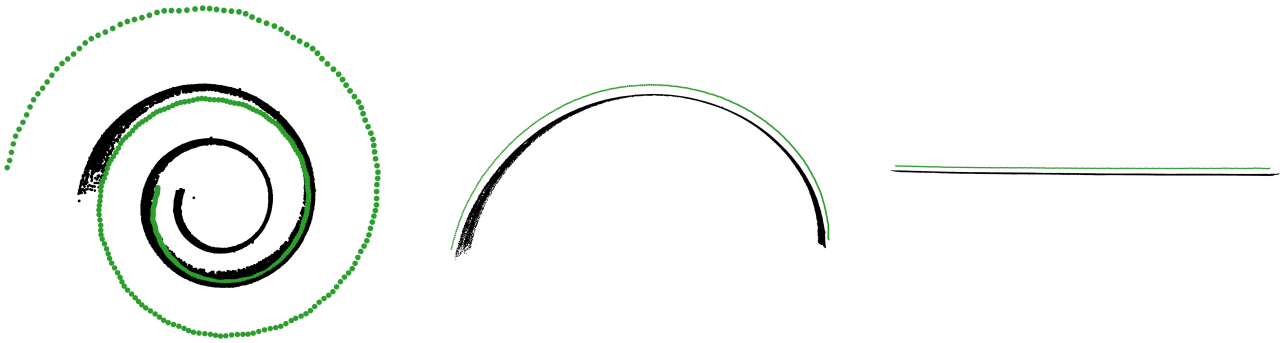
1. Introduction

Computing the intrinsic and extrinsic parameters for a moving camera with an optical axis perpendicular to the ground will inevitably result in a critical configuration for radial self calibration [28], meaning that the radial distortion parameters cannot be determined uniquely. This movement pattern is, however, common in many applications like photogrammetry using UAVs to compute digital elevation maps [2, 21] or top-down animal behavior studies (e.g. for navigation analysis) [7, 24]. For example, top-down video recordings (e.g. from a drone) benefit from image planes parallel to the ground plane to enable direct on-the-ground distance computations without suffering from perspective distortions or occlusions. However, these configurations are known to cause systematic errors in UAV-based reconstructions and DEM computations [16]. The video recordings are often processed using Structure-from-Motion (SfM) pipelines to obtain 3D reconstructions of the recorded environment for subsequent studies. No or insufficient camera calibrations lead to severe doming effects in the resulting reconstruction making accurate measurements in the 3D scene often impossible as illustrated in Fig. 1. The problem arises due to ambiguities between the potential curvature of the ground plane and the lens distortion.

The contributions of this paper are (i) a minimal solver for two views with their viewing direction orthogonal to a shared ground plane and rotations around this viewing direction and (ii) an optional regularizer to mitigate potential remaining curvature. Minimal solvers are an important part for many applications including SfM to find relative rotations and translations between the views. Relative motions are then transformed into globally consistent motions usually followed by bundle adjustment [18, 23]. In an SfM pipeline, our solver can be used instead of the existing minimal solver or as a preprocessing step, to extract distortion parameters that minimize the curvature of the ground plane. In the latter case, SfM pipelines can be used without any modification. The regularizer can be easily integrated into existing SfM pipelines and further penalizes curved surfaces during bundle adjustment.

2. Related Work

Critical configurations in radial self calibration were rigorously analyzed in [28]. In addition to critical surfaces, the author defines some critical motions which are camera motions that always induce ambiguities into the radial distortion estimation. These motions are forward motion and moving on a sphere while pointing to its center (also with infinite radius, i.e. moving on a plane). Forward motion is of interest in robotics and solutions have been proposed



(a) No calibration ($\sigma = 0.91$) (b) Inaccurate calibration ($\sigma = 0.4$) (c) Accurate calibration ($\sigma = 0.02$)

Figure 1. Real-world dataset: Comparison of curvatures in the resulting 3D reconstructions with respect to different calibration scenarios. Scene points and camera centers are illustrated as black and green points respectively. Values for σ are calculated based on the metric defined in 6.2 and quantify curvature (planar reconstruction: low value, curved reconstruction: high value)

like [29], where the authors use a deep-learning approach to predict the intrinsic camera parameters. The latter motion is often discussed in remote sensing applications but only solutions using additional hardware or heuristic solutions have been proposed up to now. For example, the usage of additional sensors like IMUs and GPS have been proposed to prevent curved 3D reconstructions [9], which however prevents the processing of already recorded videos and also imply more complex hardware setups. The author also proposes a heuristic approach to estimate distortion parameters by running bundle adjustment on a grid of different radial distortion parameters and selecting the parameter with the least curvature, which is not feasible for long recordings with many frames due to the size of the bundle adjustment problems. Another heuristic approach using GPS information is proposed in [15]. In [10, 27], the authors suggest using dense ground control points to mitigate the curvature. An iterative approach is proposed in [16] using repeated plane surface approximation with paraboloids and subsequent bundle adjustment.

Minimal solvers for the general relative pose problems requiring five point correspondences have been studied extensively with many improvements made over the years [11, 19]. For more special uses cases, camera motion or scene constraints can be incorporated into the minimal solver usually with the goal of reducing the number of required point correspondences [4, 17, 26]. Besides estimating extrinsic parameters like rotation and translation, minimal solvers have also been proposed for the joint estimation of radial distortion parameters [3, 20]. The goal of these works is to propose a broader class of relative motion minimal solvers. They are thus not applicable for the problem described in this paper, since it requires to assume a certain motion pattern to remove the ambiguity [28].

A calibration free approach for SfM was proposed in [13]. Instead of estimating the parameters through self-calibration, the authors recover a 3D reconstruction using quadrifocal tensors removing the need to estimate focal length, and distortion parameters explicitly. The approach, however, only works for intersecting viewing rays.

3. Problem Formulation

Similarly to [20], we formulate the problem as follows:

$$\begin{aligned} &g\left(x_i', \lambda\right)^T F g\left(x_i, \lambda\right) = 0, \quad \forall i = 1, \ldots, n \\ &F \in \mathcal{P}, \end{aligned} \tag{1}$$

given 2D image correspondences (x_i, x_i') for $i = 1, \ldots, n$, a fundamental matrix F with a set of possible solutions $\mathcal{P}$ for F. The function $g\left(\cdot, \cdot\right)$ calculates the radial undistortion using the Fitzgibbon division model [6]:

$$g\left(x, \lambda\right) = \begin{pmatrix} x^{(1)} \\ x^{(2)} \\ 1 + \lambda \left\|x\right\|^2 \end{pmatrix}, \tag{2}$$

for a single parameter $\lambda \in \mathbb{R}$ and a distorted image point $x = \left(x^{(1)}, x^{(2)}\right) \in \mathbb{R}^2$.

In [25], it has been shown, that the focal length cannot be recovered from two views if the optical axes are parallel (i.e. moving on a plane) or if they are intersecting in a finite point and are equally distant from this point (i.e. moving on a sphere). For this reason we assume the focal length to be estimated with sufficient accuracy. Importantly, a violation of this assumption does not result in non-planar 3D reconstructions when used in an SfM pipeline but will lead to incorrect intrinsic and extrinsic camera parameters. This use-case is especially relevant for inaccurate calibration of the radial distortion parameters.

An essential matrix E can be decomposed as

$$E = R\left[t\right]_\times \tag{3}$$

for a 3D rotation matrix $R \in \mathrm{SO}(3)$ and the cross product matrix notation of a 3D translation vector $t \in \mathbb{R}^3$. Since we assume a rotation around the viewing direction which coincides with the z-axis, the rotation matrix can be simplified to:

$$\tilde{R}_z(\varphi) = \begin{pmatrix} \cos\varphi & -\sin\varphi & 0 \\ \sin\varphi & \cos\varphi & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{4}$$

for an angle $\varphi \in [0, 2\pi)$ which can be parameterized using the half-angle substitution $s = \tan(\varphi/2)$ to

$$R_z(s) = \frac{1}{1+s^2} \begin{pmatrix} 1-s^2 & -2s & 0 \\ 2s & 1-s^2 & 0 \\ 0 & 0 & 1+s^2 \end{pmatrix}, \tag{5}$$

for $s \in \mathbb{R}$ and $|\varphi| \neq \pi$. This constraint is, however, negligible for real-world applications. This parameterization has advantages compared to Eq. (4), since the variable is not bounded and ambiguities at 0 and 2π can be avoided.

The calibration manifold $\mathcal{P}$ from Eq. (1) is consequently defined as:

$$\mathcal{P} = \{R_z(s)\,[t]_\times \in \mathbb{P}^{3\times 3} \mid s \in \mathbb{R}, t \in \mathbb{R}^3\}. \tag{6}$$

Since an essential matrix is only determined up to scale, the scaling $\frac{1}{1+s^2}$ parameter in Eq. (5) can be omitted.

4. Radial Relative Pose on a Common Plane

For each point correspondence (x_i, x_i') for $i = 1, \ldots, 4$, we have one equation:

$$g(x', \lambda)^T \left(R_z(s)\,[t]_\times\right) g(x, \lambda) = 0 \tag{7}$$

Expanding Eq. (7), we get equations that are linear in t. By stacking them row-wise into a matrix, we get the following equation:

$$A \begin{pmatrix} t_1 \\ t_2 \\ t_3 \end{pmatrix} = 0, \tag{8}$$

where $A \in \mathbb{R}[s, \lambda]^{4\times 3}$, whose elements are polynomials in s and λ. The determinants of each of the four 3×3 submatrices of A must be zero since a non-trivial solution for Eq. (8) exists, providing new constraints that are independent of t.

The resulting polynomials can be solved by transforming them into a polynomial eigenvalue problem as commonly done in other minimal solvers [4, 12]. For this, three (of four) of the polynomials from the determinants of the submatrices in Eq. (8) are rewritten as a coefficient matrix multiplied by a vector containing all the monomials:

$$Bm = 0, \tag{9}$$

for coefficients $B \in \mathbb{R}^{3\times 21}$ and the monomial vector

$$m = \big(1, \lambda, \lambda^2, s, s\lambda, s\lambda^2, s^2, s^2\lambda, s^2\lambda^2, s^3, s^3\lambda, s^3\lambda^2, s^4, s^4\lambda, s^4\lambda^2, s^5, s^5\lambda, s^5\lambda^2, s^6, s^6\lambda s^6\lambda^2\big)^T \tag{10}$$

consisting of the 21 monomials. Equation (9) can be rewritten into a polynomial eigenvalue problem:

$$\Psi(s)\,\Lambda = 0 \tag{11}$$

with $\Lambda = \left(1, \lambda, \lambda^2\right)^T$ and $\Psi(s) \in \mathbb{R}[s]^{3\times 3}$ where

$$\Psi(s) \equiv s^6 C_6 + \cdots + s^0 C_0 \tag{12}$$

for square coefficient matrices $C_i \in \mathbb{R}^{3\times 3}$. These matrices can be solved as described in [1] resulting in the 18×18 companion matrix of the polynomial eigenvalue problem:

$$D = \begin{bmatrix} 0 & I & 0 & \ldots & 0 \\ 0 & 0 & I & \ldots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & I \\ -C_6^{-1}C_0 & -C_6^{-1}C_1 & \ldots & \ldots & -C_6^{-1}C_5 \end{bmatrix} \tag{13}$$

The 18 eigenvalues of Eq. (13) are the potential candidates for s. Complex values can be omitted directly. Each real eigenvalue is considered a solution for the relative pose problem, if the corresponding eigenvector fulfills $\Lambda_2/\Lambda_1 = \Lambda_3/\Lambda_2$ given by $\Lambda = \left(1, \lambda, \lambda^2\right)^T$. A solution for λ is then set to this quotient. Given s and λ, the translation t can be determined up to scale as the non-trivial solution of Eq. (8) using Singular Value Decomposition (SVD).

5. Usage in Structure-from-Motion Pipeline

5.1. Sampson Error

The Sampson error [22] is often used in robust estimators like RANSAC [5] as an error metric because it approximates the geometric error and is thus more meaningful than the algebraic error [8].

The algebraic error is defined as:

$$e = g(x', \lambda)^T E g(x, \lambda). \tag{14}$$

The Sampson error is then defined using the Taylor expansion of the geometric error and requires the partial derivatives of the algebraic error:

$$\|\delta_X\|^2 = \frac{e^T e}{JJ^T}. \tag{15}$$

In our case, the algebraic error is a scalar. The partial derivative matrix J is of size 1×4 and is defined with respect to the undistortion function:

$$h' = g(x', \lambda), \quad h = g(x, \lambda) \tag{16}$$

$$J = (Eh_1 + 2\lambda x_1' Eh_3, Eh_2 + 2\lambda x_2' Eh_3, \\ E^T h_1' + 2\lambda x_1 E^T h_3', E^T h_2' + 2\lambda x_2 E^T h_3')^T . \quad (17)$$

5.2. Recovering the Shared Distortion Parameters

Relative poses between pairwise frames are estimated using the minimal solver and RANSAC with the Sampson error as defined in Sec. 5.1. A globally consistent distortion parameter can then be determined by using the median across all pairwise values.

5.3. Axis Regularizer for Bundle Adjustment

Due to numerical inaccuracies, e.g. due to conversions between distortion models or later bundle adjustments, some curvature may still be noticeable in the final reconstruction (see Sec. 6.2). To remove the remaining doming, a regularizer can be added to final the bundle adjustment steps involving radial distortion parameters. In contrast to [16], this does not involve fitting a parametric model into the ground points and is thus applicable in scenarios, where the ground cannot be sufficiently approximated using a paraboloid. The proposed regularizer influences the camera extrinsics which in turn affects the radial distortion parameters and finally the 3D reconstruction.

Given the defined motion constraints (rotation around the viewing direction), the regularizer aims to transform all viewing directions to be parallel to each other. Before bundle adjustment, a dominant viewing direction is determined by gathering all viewing directions and finding the dominant vector $\rho \in \mathbb{R}^3$. The viewing directions are normalized to unit length and stacked row-wise into a matrix A and the dominant vector is determined by solving:

$$\max \|A\rho\|, \quad \text{s.t. } \|\rho\| = 1. \quad (18)$$

A normalized vector maximizing Eq. (18) also maximizes the dot product and therefore minimizes the mean angle between the viewing directions and the solution. The solution of Eq. (18) can be obtained by using SVD and extracting the right-singular vector of A corresponding to the largest singular value.

The regularizer is then defined as:

$$J(\phi_1, \ldots, \phi_N) = \sum_{i=1}^{N} \|\phi_i \times \rho\|^2 \quad (19)$$

for the rotations $\phi_i \in \mathbb{R}^3$ for $i = 1, \ldots, N$ of all N poses in angle-axis representation. The regularizer is minimized if the rotation vector of a pose is parallel (or anti-parallel) to the dominant vector.

6. Experiments

To benchmark the performance of the proposed methods, they are evaluated using a synthetic dataset with a known radial distortion and relative pose, and a real-world dataset consisting of videos captured with different camera modalities[1]. All experiments were done in MATLAB.

6.1. Synthetic Data

The minimal solver is tested on a synthetic dataset by sampling random points. 400 random 3D points are sampled 1000 times in a volume of $[-250, 250]^3$ and transformed using an arbitrary translation and an arbitrary rotation around the viewing direction of an artificial camera to get corresponding points. Subsequently, these points are projected onto images planes and distortion is applied for a random distortion parameter $\lambda \in [-1, 0.2]$. Note that for a severe pincushion distortion of $\lambda > 0.2$ the approximation of the inverse of the radial undistortion functions becomes to inaccurate to generate accurate ground-truth image points.

In Fig. 2, RANSAC is started once per sampling run and the errors in log scale for the distortion, translation and rotation error are illustrated. Median errors are very low being $3.21 \times 10^{-13}, 1.86 \times 10^{-13}, 2.67 \times 10^{-14}$ for distortion, translation and rotation error respectively, which is crucial for real-world scenarios.

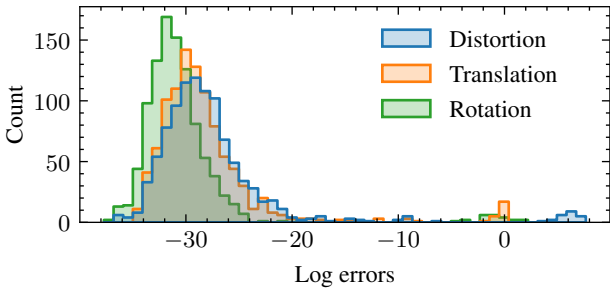
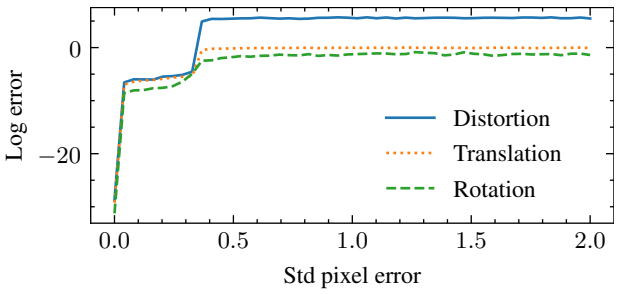


Figure 2. Synthetic dataset: Estimation errors in log scale. For the synthetic dataset described in Sec. 6.1, the histograms of errors in log scale are plotted as the absolute difference of the real distortion and estimated distortion in the first row, the $L2$-norm of the real translation and estimated translation in the second row and absolute angle difference between estimated and real rotation angle in the third row. The median errors are 3.21×10^{-13}, 1.86×10^{-13}, 2.67×10^{-14} for distortion, translation and rotation respectively.

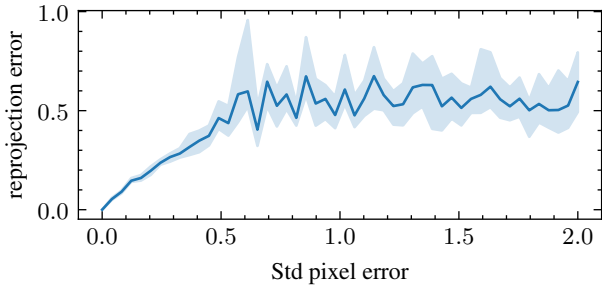
In a second experiment, Gaussian noise is added in 100 steps for standard deviations in $[0, 2]$ given an artificial image of size 1000×1000 with a focal length of 500. Re-

[1]We note that to our knowledge no publicly available dataset for this particular critical configuration is available and existing work on computational aspects focuses on synthetic data to allow precise evaluations; see e.g. [19, 28]

sults are illustrated in Fig. 3a again in log scale. For all errors there is a clear point where accuracy drops around 0.4, while the reprojection error as visualized in Fig. 3b remains small throughout the experiments.



(a) Errors for distortion, translation and rotation in log scale.



(b) Corresponding median reprojection errors with a 95% confidence interval.

Figure 3. Synthetic dataset: Estimation errors and corresponding reprojection errors. For the synthetic dataset described in Sec. 6.1, the median errors for increasing pixel noise are plotted. Pixel noise is quantified using the standard deviation of a Gaussian distribution. Distortion, translation and rotation errors are again measured using the absolute difference, $L2$-norm and absolute angle difference respectively. See Supplement Fig. S1 for different outlier ratios.

To investigate how an inaccurate or unknown focal length impacts the solver, in a final experiment, the known focal length of 500 from a synthetic dataset was shifted by values between -200 and 200. The image size is again 1000×1000. The reprojection error is calculated and shown in Fig. 4, revealing that even for inaccurate focal lengths, the reprojection error stays small, showing that the method is invariant to the focal length but as mentioned in Sec. 3 will just result in a different radial distortion parameter and translation.

6.2. Real-World Data

Real-world experiments are performed with two cameras: a hand-held gimbal stabilized GoPro Hero 5 (H5) with heavy fish-eye distortion operated in the FOV settings medium (m) and wide (w) and a hand-held unstabilized Panasonic HDC TM-900 video camera with minor distortion. In both cases the cameras were attached to a rod to keep a certain distance from the camera operator to

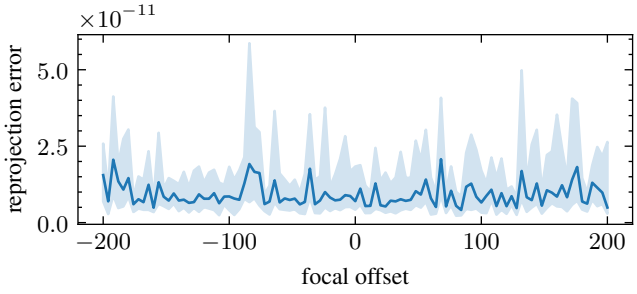


Figure 4. Synthetic dataset: Reprojection errors for inaccurate focal lengths. For the synthetic dataset described in Sec. 6.1, the median reprojection errors for different focal length offsets with 95% confidence intervals are plotted for an artificial image of size 1000×1000.

avoid heavy shadows and the recording of the operators feet. All videos were recorded with a resolution of 1920×1080 pixels. Videos were recorded top-down and subsequently processed in the SfM pipeline OpenMVG [18] on images extract with fives frames per second from the respective video.[2]

To quantify the curvature of the reconstruction the dominant viewing direction is extracted as described in Sec. 5.3. The angle between each viewing direction and this dominant direction is determined and the standard deviation calculated. Consequently small and large deviations are expected in planar and domed scenes respectively.

Results are shown in Tab. 1. In the uncalibrated case, the focal length was set to an arbitrarily chosen value of 1000 for the GoPro videos and a focal length of 1905 for the Panasonic which was extracted from a camera calibration routine. The radial distortion parameters were initially set to zero in all cases. For the minimal solver results the radial distortion parameters were extracted as the median of all image pairs. For the minimal solver and regularized results, the regularizer described in Sec. 5.3 was applied in the final bundle adjustments in addition to the median extracted radial distortion from the previous experiment. The results show that the minimal solver estimates sensible radial distortion for all eight videos even when the camera is not gimbal stabilized (as in v7 and v8). Moreover, the remaining curvature is effectively mitigated by the regularizer.

If the algebraic error is used instead of the Sampson error, the number of usable solutions is an order of magnitude lower in comparison to using the Sampson error as shown in Tab. 2. Across all eight videos, the usable solutions are substantially higher when using the Sampson error. A usable solution is defined as a solution with a realistic radial dis-

[2]The videos of the dataset are publicly available at https://top-down-sfm.cvmls.org

video	v1	v2	v3	v4	v5	v6	v7	v8
camera	H5 (w)	H5 (w)	H5 (w)	H5 (m)	H5 (m)	H5 (m)	TM-900	TM-900
gimbal	✓	✓	✓	✓	✓	✓	✗	✗
uncalibrated	0.847	1.009	0.778	1.047	1.016	0.936	0.293	0.968
minimal solver (ms)	0.034	0.027	0.033	0.027	0.043	0.026	0.069	0.050
ms + regularizer	0.003	0.001	0.019	0.007	0.001	0.018	0.028	0.018

Table 1. Real-world dataset: Curvatures of eight different videos (v1,...,v8) with three different cameras settings: GoPro Hero 5 (H5) with FOV settings wide (w) and medium (m) and a Panasonic TM-900 video recorder.

video	v1	v2	v3	v4	v5	v6	v7	v8
Sampson	78.0	80.6	77.4	79.2	74.9	80.6	83.3	88.6
Algebraic	6.7	11.7	6.2	7.3	6.9	4.8	14.9	24.9

Table 2. Real-world dataset: Algebraic error versus Sampson error for videos v1 to v8. For each video the percentage of usable solutions with $|\lambda| < 1$ are shown.

tortion parameter $|\lambda| < 1$ which includes severe pincushion and barrel distortions.

To analyze the retrieval of radial distortion via median, Fig. 5 illustrates the distribution of radial distortion parameters for all image pairs. As can be seen, the median is a good measure for the radial distortion. Except for a few outliers, the values are all very close to each other. Interestingly this is also true for videos 7 and 8, where no gimbal was used, showing that the proposed method is robust against occasional violations of the motion constraints.

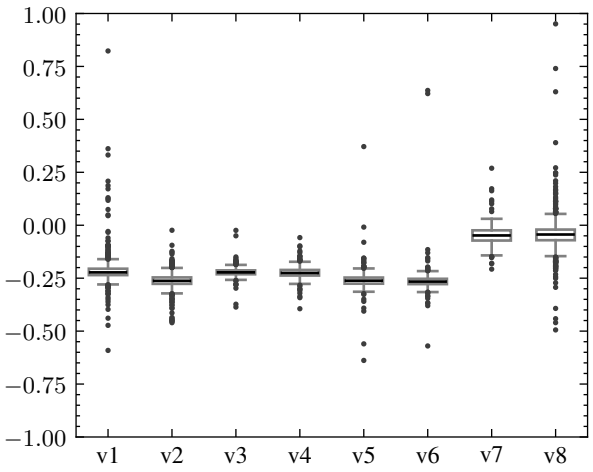


Figure 5. Real-world dataset: Radial distortion parameters for videos v1 to v8 from Tab. 1. For each video the pairwise estimated distortion parameters are plotted in a boxplot. Small variances indicate, that the minimal solver estimated similar values for different image pairs.

Resulting reconstructions for all eight videos are shown in Fig. 6.

To study the impact of the regularizer independently of the minimal solver, we performed an additional experiment. Figure 7 illustrates an example where the regularizer was applied on the fully uncalibrated case without using the minimal solver to find meaningful distortion parameters first. The optimization is unstable and results in an unusable reconstruction. We therefore conclude that the regularizer cannot be used instead of the minimal solver but only as a final adjustment step when the solution is already close the planar scene.

In a final qualitative experiment, the reconstruction of an insufficiently calibrated reconstruction is compared the reconstruction generated using the proposed algorithms. The calibrated reconstruction in Figs. 8a and 8b has significant curvature, making it unusable for many use cases. The reconstruction using the distortion parameter from the minimal solver mitigates this curvature significantly (see Figs. 8c and 8d) and remaining curvature is completely removed using the regularizer from Sec. 5.3 as shown in Figs. 8e and 8f.

7. Discussion & Outlook

In contrast to other critical configurations no formal solution exists to solve the plane-sphere ambiguity in monocular SfM settings during the estimation of the intrinsic and extrinsic camera parameter directly. These ambiguities also exist for movement patterns which deviate slightly from the critical motions like freely-moving non-stabilized cameras. As a consequence, additional hardware such as depth sensors or IMUs, heuristics based post-processing and adjusted flight paths have been utilized to circumvent curved reconstructions and erroneous relative camera pose estimations [9, 10, 15, 16, 27]. While these solutions are applicable for specifically generated and curated data, they often can not be applied to existing recordings and do not solve the intrinsic ambiguity on a formal level.

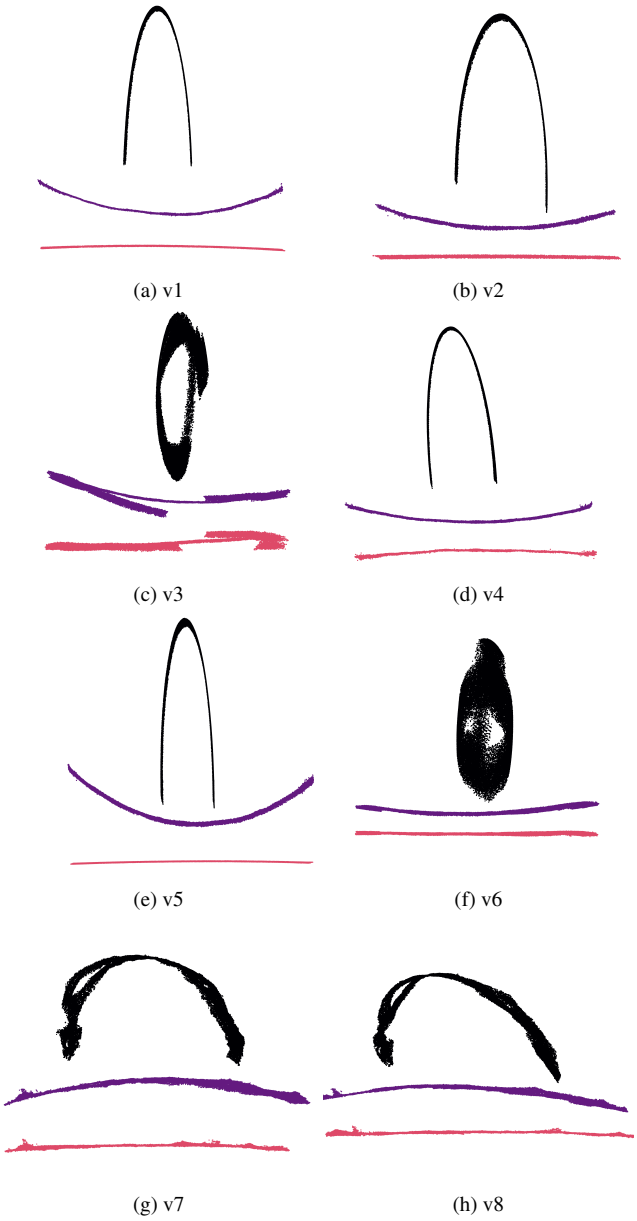


Figure 6. Real-world dataset: Orthographic projections of 3D-reconstruction for videos from Tab. 1. Uncalibrated reconstructions are colored in black [■], reconstructions after setting the radial distortion parameters to value extracted form the minimal solver are colored in purple [■], and after additional regularizer in red [■]. For a a more in-depth view of video v7, the reader is referred to the supplement for an animated rotation around the colored 3D reconstructed point cloud.

In this paper, we presented two effective and practical solutions to remove the radial distortion in top-down SfM settings, namely a minimal solver to estimate relative poses and radial distortion parameters and a regularizer for final bundle adjustment to remove potentially remaining curvature. The novelty of our solver is to exploit a priori motion constraints of top-down recordings yielding a polynomial eigenvalue problem which can be solved using SVD. Combining this solver with RANSAC and the problem specific Sampson error, globally consistent distortion parameters can be extracted which significantly reduce the curvature of the reconstructions. Remaining curvature can be removed by integrating the regularizer into the final bundle adjustment which penalizes non-parallel rotation axes across views. Both methods are easily added to existing SfM pipelines and our results demonstrate that they can remove the curvature of the plane-sphere ambiguity sufficiently even for recordings with occasional violations of the assumed motion constraints.

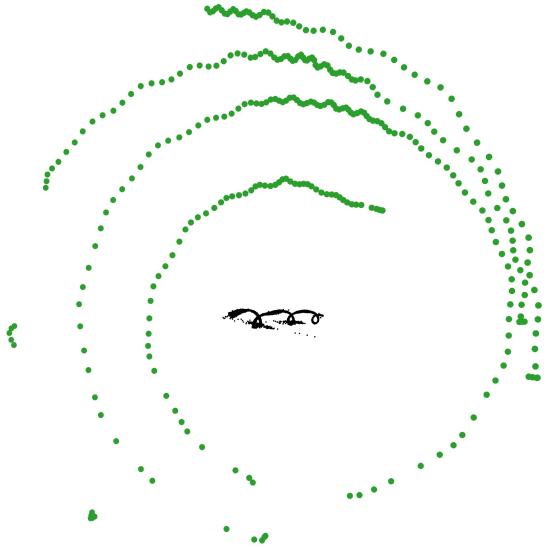
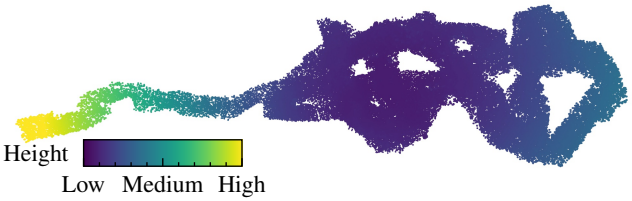


Figure 7. Real-world dataset: Uncalibrated reconstruction using the axis regularizer in the final bundle adjustment without the use of the proposed minimal solver. Scene points and camera centers are illustrated as black and green points respectively. As can be clearly seen regularizing the camera pose alone does not solve the plane-sphere ambiguity.

In the future, we are going to integrate our algorithms into top-down and moving-camera tracking systems. For this, we want to explore different solving techniques for the main equation Eq. (7). The current minimal solver implementation involves many symbolic math operations and large eigenvalue problems. It will be interesting to test automatic solvers [14], which have the potential to improve the efficiency. Moreover, we want to analyze how much deviation from the orthogonal movement pattern still achieves sensible distortion parameters in real-world scenarios.

(a) Calibrated Reconstruction ($\sigma = 0.2$), Side



(b) Calibrated Reconstruction ($\sigma = 0.2$), Top

(c) Minimal solver ($\sigma = 0.07$), Side

(d) Minimal solver ($\sigma = 0.07$), Top

(e) Minimal solver and final axis regularizer ($\sigma = 0.03$), Side

(f) Minimal solver and final axis regularizer ($\sigma = 0.03$), Top

Figure 8. Real-World scenario: Video v7 is taken with a Panasonic TM-900 video recorder with minor distortion. In the left column, the reconstruction (black) and cameras (green) are shown from the side. In the right column, the reconstruction is shown from the top. The colormap in the right column indicates the deviation from the fully planar reconstruction per point (i.e. distance to the ground plane). All point clouds have been scaled, rotated and translated into the same reference coordinate system.

8. Acknowledgments

BR, and LH would like to thank the Human Frontier Science Program [RGP0057/2021]. BR would like to thank the WildDrone MSCA Doctoral Network funded by EU Horizon Europe under grant agreement #101071224. LH also acknowledges support by the Heinrich Böll Foundation.

References

[1] Z. Bai, G. Sleijpen, H. van der Vorst, R. Lippert, and A. Edelman. 9. nonlinear eigenvalue problems. In *Templates for the Solution of Algebraic Eigenvalue Problems*, pages 281–314. Society for Industrial and Applied Mathematics, Jan. 2000. 3

[2] Sean P. Bemis, Steven Micklethwaite, Darren Turner, Mike R. James, Sinan Akciz, Sam T. Thiele, and Hasnain Ali Bangash. Ground-based and UAV-based photogrammetry: A multi-scale, high-resolution mapping tool for structural geology and paleoseismology. *Journal of Structural Geology*, 69:163–178, Dec. 2014. 1

[3] Martin Byrod, Zuzana Kukelova, Klas Josephson, Tomas Pajdla, and Kalle Astrom. Fast and robust numerical solutions to minimal problems for cameras with radial distortion. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, June 2008. 2

[4] Yaqing Ding, Jian Yang, Jean Ponce, and Hui Kong. Minimal solutions to relative pose estimation from two views sharing a common direction with unknown focal length. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, June 2020. 2, 3

[5] Martin A. Fischler and Robert C. Bolles. Random sample consensus. *Communications of the ACM*, 24(6):381–395, June 1981. 3

[6] A.W. Fitzgibbon. Simultaneous linear estimation of multiple view geometry and lens distortion. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*. IEEE Comput. Soc, 2001. 2

[7] Lars Haalck, Michael Mangan, Barbara Webb, and Benjamin Risse. Towards image-based animal tracking in natural environments using a freely moving camera. *Journal of Neuroscience Methods*, 330:108455, Jan. 2020. 1

[8] Richard Hartley and Andrew Zisserman. *Multiple View Geometry*. Cambridge University Press, 3 edition, 2004. 3

[9] Mike R. James and Stuart Robson. Mitigating systematic error in topographic models derived from UAV and ground-based image networks. *Earth Surface Processes and Landforms*, 39(10):1413–1420, June 2014. 2, 6

[10] Marion Jaud, Sophie Passot, Pascal Allemand, Nicolas Le Dantec, Philippe Grandjean, and Christophe Delacourt. Suggestions to limit geometric distortions in the reconstruction of linear coastal landforms by SfM photogrammetry with PhotoScan® and MicMac® for UAV surveys with restricted GCPs pattern. *Drones*, 3(1):2, Dec. 2018. 2, 6

[11] Z. Kukelova, M. Bujnak, and T. Pajdla. Polynomial eigenvalue solutions to the 5-pt and 6-pt relative pose problems. In *Procedings of the British Machine Vision Conference 2008*. British Machine Vision Association, 2008. 2

[12] Zuzana Kukelova, Martin Bujnak, and Tomas Pajdla. Polynomial eigenvalue solutions to minimal problems in computer vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(7):1381–1393, July 2012. 3

[13] Viktor Larsson, Nicolas Zobernig, Kasim Taskin, and Marc Pollefeys. Calibration-free structure-from-motion with calibrated radial trifocal tensors. In *Computer Vision – ECCV 2020*, pages 382–399. Springer International Publishing, 2020. 2

[14] Bo Li and Viktor Larsson. Gaps: Generator for automatic polynomial solvers. *arXiv preprint arXiv:2004.11765*, 2020. 7

[15] Yonglu Li, Yinghao Cai, Dayong Wen, and Yiping Yang. Optimization of radial distortion self-calibration for structure from motion from uncalibrated UAV images. In *2016 23rd International Conference on Pattern Recognition (ICPR)*. IEEE, Dec. 2016. 2, 6

[16] L. Magri and R. Toldo. Bending the doming effect in structure from motion reconstructions through bundle adjustment. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2/W6:235–241, Aug. 2017. 1, 2, 4, 6

[17] Evgeniy Martyushev and Bo Li. Efficient relative pose estimation for cameras and generalized cameras in case of known relative rotation angle. *Journal of Mathematical Imaging and Vision*, 62(8):1076–1086, May 2020. 2

[18] Pierre Moulon, Pascal Monasse, Romuald Perrot, and Renaud Marlet. OpenMVG: Open multiple view geometry. In *International Workshop on Reproducible Research in Pattern Recognition*, pages 60–74. Springer, 2016. 1, 5

[19] D. Nister. An efficient solution to the five-point relative pose problem. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(6):756–770, June 2004. 2, 4

[20] Magnus Oskarsson. Fast solvers for minimal radial distortion relative pose problems. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. IEEE, June 2021. 2

[21] F. Rinaudo, F. Chiabrando, A. Lingua, and A. Spanò. ARCHAEOLOGICAL SITE MONITORING: UAV PHOTOGRAMMETRY CAN BE AN ANSWER. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXIX-B5:583–588, July 2012. 1

[22] Paul D Sampson. Fitting conic sections to "very scattered" data: An iterative refinement of the bookstein algorithm. *Computer Graphics and Image Processing*, 18(1):97–108, Jan. 1982. 3

[23] Johannes Lutz Schönberger and Jan-Michael Frahm. Structure-from-motion revisited. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016. 1

[24] Ariana Strandburg-Peshkin, Damien R Farine, Margaret C Crofoot, and Iain D Couzin. Habitat and social factors shape individual decisions and emergent group structure during baboon collective movement. *eLife*, 6, Jan. 2017. 1

[25] P. Sturm. On focal length calibration from two views. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*. IEEE Comput. Soc, 2001. 2

[26] Chris Sweeney, John Flynn, and Matthew Turk. Solving for relative pose with a partially known rotation is a quadratic eigenvalue problem. In *2014 2nd International Conference on 3D Vision*. IEEE, Dec. 2014. 2

[27] V. Tournadre, M. Pierrot-Deseilligny, and P. H. Faure. UAV LINEAR PHOTOGRAMMETRY. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XL-3/W3:327–333, Aug. 2015. 2, 6

[28] Changchang Wu. Critical configurations for radial distortion self-calibration. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, June 2014. 1, 2, 4

[29] Bingbing Zhuang, Quoc-Huy Tran, Gim Hee Lee, Loong Fah Cheong, and Manmohan Chandraker. Degeneracy in self-calibration revisited and a deep learning solution for uncalibrated SLAM. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, Nov. 2019. 2