Privacy Preserving Localization and Mapping from Uncalibrated Cameras Supplementary Material

Marcel Geppert[1] Viktor Larsson[1] Pablo Speciale[2] Johannes L. Schönberger[2] Marc Pollefeys[1,2]
[1] Department of Computer Science, ETH Zurich [2] Microsoft

1. Overview

In this supplementary material we present;

- The full constraints used in the absolute pose solver with unknown focal length (Section 2).
- Additional details on our implementation (Section 3).
- Additional evaluation of the absolute pose solvers (Section 4).
- An analysis of our cost function for focal length averaging compared to the one used in Sweeney *et al*. [10] (Section 5).
- Reconstruction statistics of the full reconstructions using the mobile phone datasets used for the initialization experiments, and a discussion of a failure case (Section 6).
- Qualitative results of the reconstruction and comparisons with traditional SfM and the calibrated privacy preserving pipeline from [2]. (Section 7)

2. Internal Constraints for Absolute Pose

We use the internal constraints of the projection matrix as presented in Larsson et al. [4],

$$p_{21}p_{31} + p_{22}p_{32} + p_{23}p_{33} = 0 \quad (1)$$
$$p_{11}p_{31} + p_{12}p_{32} + p_{13}p_{33} = 0 \quad (2)$$
$$p_{11}p_{21} + p_{12}p_{22} + p_{13}p_{23} = 0 \quad (3)$$
$$p_{11}^2 + p_{12}^2 + p_{13}^2 - p_{21}^2 - p_{22}^2 - p_{23}^2 = 0 \quad (4)$$
$$p_{13}^2p_{32} - p_{21}^2p_{32} - p_{22}^2p_{32} - p_{12}p_{13}p_{33} - p_{22}p_{23}p_{33} = 0 \quad (5)$$
$$p_{12}p_{13}p_{32} + p_{22}p_{23}p_{32} - p_{12}^2p_{33} + p_{21}^2p_{33} + p_{23}^2p_{33} = 0 \quad (6)$$
$$p_{11}p_{13}p_{32} + p_{21}p_{23}p_{32} - p_{11}p_{12}p_{33} - p_{21}p_{22}p_{33} = 0 \quad (7)$$
$$p_{13}^2p_{31} - p_{22}^2p_{31} + p_{21}p_{22}p_{32} - p_{11}p_{13}p_{33} = 0 \quad (8)$$
$$p_{12}p_{13}p_{31} + p_{22}p_{23}p_{31} - p_{11}p_{12}p_{33} - p_{21}p_{22}p_{33} = 0 \quad (9)$$

Here, Eq. (1)-(3) ensure orthogonality of the rotation matrix rows and columns. Eq. (4) ensures equal norm of the first two rows and columns of the rotation matrix since we estimate a single focal length. Eq. (5)-(9) avoid complex solutions that we would otherwise need to detect and filter in a separate step.

3. Implementation Details

Here we provide some additional details on how the system for our experiments is set up.

Initialization image selection. We use a simple heuristic that is based on the number of pairwise feature matches. At the beginning, we randomly select 10 images with aligned features. For each of these images, we then find three more (aligned) images, so that the set is fully connected in the correspondence graph and has a high number of matches between all pairs. We do not require the maximal number of matches to avoid exhaustive search and limit the runtime. The proposed initialization scheme (Section 3.2 in the main paper) is then run independently on each set of for images, and the set with the highest inlier ratio is used to initialize the full reconstruction. This worked well in our experiments and we did not investigate more complex strategies.

RANSAC variants. In the main paper we refer to RANSAC [1] for robust estimation while in practice we use different variants of RANSAC in different steps of the pipeline. We generally use the implementations provided by the COLMAP library [6]. For absolute pose estimation (Section 3.1 in the paper) we use standard RANSAC with simple inlier counting and without internal local optimization. During initialization we use LO-MSAC [5] for all steps. During the later mapping process we use LO-RANSAC [5] when triangulating new points.

Vanishing point estimation. To estimate vanishing points in the image we directly rely on the implementation provided by COLMAP [6], which is based on LSD [3] line detection followed by computing pairwise line intersections in RANSAC.

4. Additional Evaluation of Pose Estimators

In addition to the results in the main paper we also present the sensitivity of the estimated position to measurement noise. For better readability we show all three plots, including the two that are also shown in the main paper, in Fig. 1. Our line-based solver shows comparable sensitivity

to noise as the keypoint-based counterparts.

In Figure 2 we show the full comparison with the point-based absolute pose solvers that also estimate focal length (c.f. Section 4.2 in the main paper). All of the point-based solvers yield similar results when applied inside RANSAC. For localization, the proposed line-based solver performs slightly worse compared to the keypoint-based counterparts, which is to be expected since it effectively utilizes half the number of geometric constraints (line-to-point vs. point-to-point correspondences).

5. Comparison of our Focal Length Averaging Cost with Sweeney et al. [10]

We estimate globally consistent focal lengths for the four cameras used in initialization based on the estimated pairwise fundamental matrices, similar to Sweeney et al. [10]. However, we found that the cost function used in [10] introduces a bias towards smaller focal lengths, particularly if the initial estimate of the focal length is too far away from the true value. We therefore use a slightly modified cost term to avoid this bias and to obtain a more robust optimization. We visualize the both the original cost function of [10] and our own in Fig. 3.

We also analyze the practical impact of our changes on the datasets that are also used to evaluate the initialization with our method. We follow the same evaluation protocol, initializing the reconstruction from 4 images and then extending the map to up to 50 images (see Section 4.3 in the main paper for more details on the experimental setup). Fig. 4 shows the corresponding pose accuracies. For this dataset we can see that the overall results do not change significantly, except for the *Bedroom* scene.

6. Complete Mobile Phone Reconstructions

As for the reconstruction experiments using the datasets from Strecha *et al*. [9], we report errors and reconstruction statistics for the *Mobile Phone* datasets from Speciale *et al*. [7] used for the initialization evaluation in Table 1. Most notably, the system was not able to register all images in the *Bedroom* scene due to some images with few useful features. Additionally, the reconstruction of the *Lobby* scene leads to large errors in some poses. This is caused by poor focal length estimates for cameras that observe a region with few usable constraints. If the focal length error of a registered camera and consequently the position error of triangulated points becomes too large, subsequently registered cameras observing the same area will be estimated with increasing focal length errors and the system is not able to recover a stable configuration. With wrong focal length estimates, the images' positions will be shifted along the principal axis. This is shown in Fig. 5.

7. Qualitative Results

We present qualitative results for the *Gendarmenmarkt* and *Tower of London* scenes in Fig. 8 in the main paper. As an additional qualitative evaluation we provide a comparison of the reconstructions of all four scenes, namely *Alamo*, *Gendarmenmarkt*, *Madrid Metropolis*, and *Tower of London* using standard COLMAP [6], the privacy preserving SfM pipeline using calibrated cameras from [2], and our pipeline, respectively, in Fig. 6. The results are generally comparable, some differences can be explained with different parameters and thresholds that might not be comparable between the methods.

References

[1] Martin A. Fischler and Robert C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM (CACM)*, 1981. 1

[2] Marcel Geppert, Viktor Larsson, Pablo Speciale, Johannes L. Schönberger, and Marc Pollefeys. Privacy preserving structure-from-motion. In *European Conference on Computer Vision (ECCV)*, 2020. 1, 2, 5

[3] Rafael Grompone von Gioi, Jérémie Jakubowicz, Jean-Michel Morel, and Gregory Randall. LSD: a line segment detector. 2012. 1

[4] Viktor Larsson, Zuzana Kukelova, and Yinqiang Zheng. Making minimal solvers for absolute pose estimation compact and robust. In *International Conference on Computer Vision (ICCV)*, 2017. 1

[5] Karel Lebeda, Jiri Matas, and Ondrej Chum. Fixing the locally optimized RANSAC. In *British Machine Vision Conference (BMVC)*, 2012. 1

[6] Johannes L. Schönberger and Jan-Michael Frahm. Structure-from-motion revisited. In *Computer Vision and Pattern Recognition (CVPR)*, 2016. 1, 2, 5

[7] Pablo Speciale, Johannes L. Schönberger, Sing Bing Kang, Sudipta Sinha, and Marc Pollefeys. Privacy Preserving Image-Based Localization. In *Computer Vision and Pattern Recognition (CVPR)*, 2019. 2

[8] Pablo Speciale, Johannes L. Schönberger, Sudipta N. Sinha, and Marc Pollefeys. Privacy preserving image queries for camera localization. In *International Conference on Computer Vision (ICCV)*, 2019. 3

[9] C. Strecha, W. von Hansen, L. Van Gool, P. Fua, and U. Thoennessen. On benchmarking camera calibration and multi-view stereo for high resolution imagery. In *Computer Vision and Pattern Recognition (CVPR)*, 2008. 2

[10] Chris Sweeney, Torsten Sattler, Tobias Hollerer, Matthew Turk, and Marc Pollefeys. Optimizing the viewing graph for structure-from-motion. In *International Conference on Computer Vision (ICCV)*, 2015. 1, 2, 4

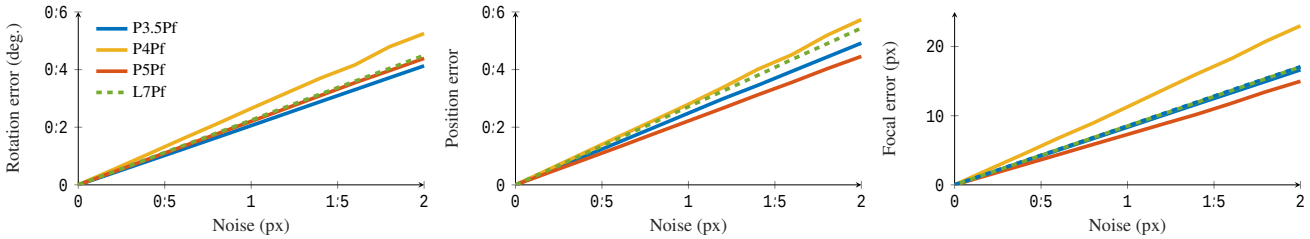

Figure 1. ***Noise Sensitivity.*** The graphs show the median errors in the rotation (*Left*), position (*Middle*) and the focal length (*Right*) for varying noise levels. There is no unit for the position error as we do not enforce a specific scale in our test setup.

Scene	#Images		#Points		Track Length	Rotation (deg)			Position (cm)			Focal Length (%)		
	Total	Reg.	3D	2D		Mean	Std.	Median	Mean	Std.	Median	Mean	Std.	Median
Bedroom	200	179	38.5k	377.0k	9.8	6.0	9.1	3.4	36.3	80.8	10.0	14.2	37.4	2.0
Gatehouse	200	200	85.6k	1026.4k	12.0	2.0	0.4	1.9	27.5	18.9	21.6	3.1	2.7	2.0
Lobby	200	200	61.9k	427.8k	6.9	6.3	2.4	5.9	338.0	551.7	102.3	28.7	43.8	10.0
Sofa	200	199	53.3k	874.0k	16.4	1.7	0.3	1.6	6.5	4.5	5.5	3.0	2.3	2.5
Stable	200	199	64.7k	755.0k	11.7	1.6	0.2	1.5	10.8	7.9	9.3	2.1	1.7	1.9
Whale	200	199	48.1k	327.8k	6.8	1.2	1.7	0.7	110.2	353.4	6.3	7.7	17.2	3.3

Table 1. ***Mobile Phone Datasets***. Full reconstruction statistics for the datasets from Speciale *et al*. [8] to evaluate the initialization step.

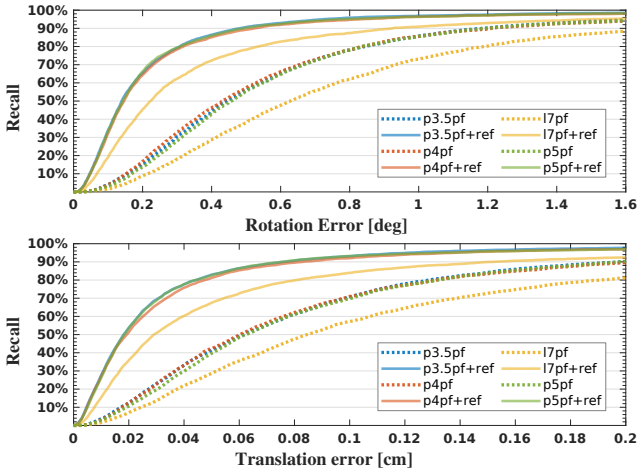

Figure 2. ***Cumulative Rotation and Position Errors***. Our solver compared to keypoint-based solvers that estimate focal length with the image pose. ds

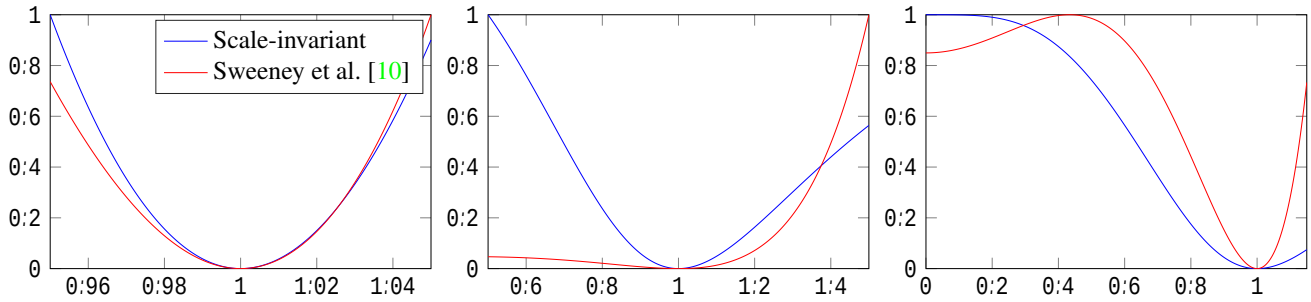

Figure 3. ***Qualitative comparison for the focal length consistency cost.*** The figure shows an example of the cost function used for focal length averaging for a synthetic instance. The x-axis shows the focal length (shared for both cameras) and the ground truth focal length is 1. The three plots show the cost function at different scales. In each plot the costs are normalized to the interval [0,1]. *Left:* Close to the correct focal length, both costs have similar shape. *Middle:* The cost used in [10] rises sharply for focal lengths greater than the ground truth. *Right:* Close to zero, the cost used in [10] has an additional local minimia. Note that the two curves are normalized independently in each plot as we are interested in the shape and not in the absolute values.

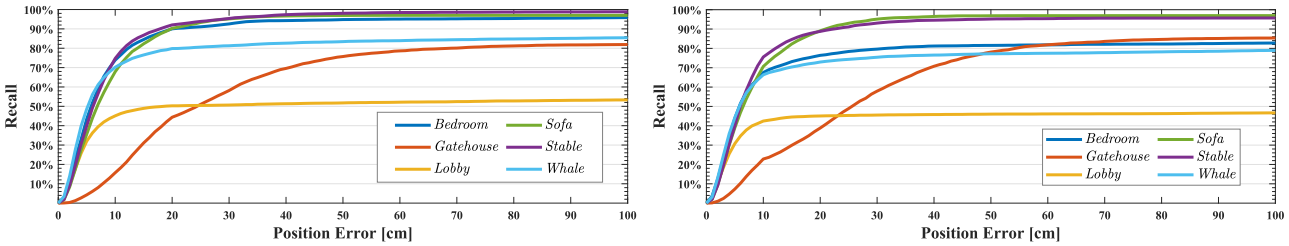

Figure 4. ***Initialization comparison.*** The figure shows a comparison of the reconstruction initialization with our focal length averaging cost function (*Left*, also shown in Fig. 7 in the main paper) or the one from Sweeney *et al*. [10] (*Right*), respectively. While most errors are very similar and the differences can be explained with random factors in the initialization process, the initialization success for the *Bedroom* scene is significantly lower using the cost from [10].

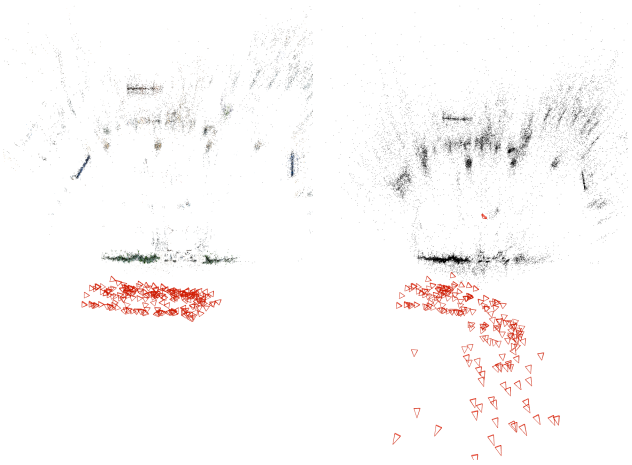

Figure 5. ***Failure Case.*** The figure shows a failure case in the *Lobby* scene where focal lengths are increasingly overestimated. The reference model from keypoints and with known camera parameters (*Left*) vs. the failed reconstruction from lines with unknown camera parameters (*Right*). The wrong focal lengths cause the images to be placed further back.