

# Structure-from-Motion with a Non-Parametric Camera Model

## Supplementary Material

### 1. Detailed Pose Evaluation on ETH3D

This section presents the detailed evaluation of estimated camera poses on original and distorted ETH3D DSLR datasets [5]. Table 1 and Table 2 summarize the AUC score computed at thresholds of  $1^\circ$ ,  $3^\circ$ , and  $5^\circ$ .

The relative performance remains almost unchanged across these thresholds, showing the consistency of evaluated methods.

### 2. Additional Reconstruction Results

Figure 2 presents the corresponding comparative reconstruction results using images captured with fisheye or catadioptric lenses, referring to Figure 5 in the main paper. The consistent reconstruction performance highlights the robustness of our non-parametric method.

### 3. Additional Undistortion Results

We undistort the catadioptric and fisheye images using the calibration map generated during the reconstruction in Figure 2. The undistortion quality demonstrates our non-parametric method’s effectiveness in estimating the underlying calibration, as well as the validity of spline representation.

### 4. Impact of Camera Type on Optimal Control Point Number

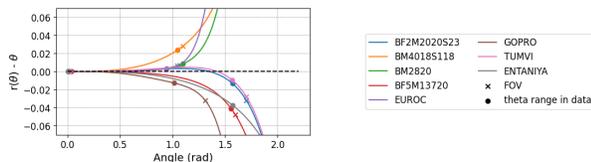


Figure 1. Kalibr [2] calibration patterns

Each dataset of BabelCalib [2] consists of multiple image sets captured with different cameras. We reported in Figure 6 in the main paper as the mean result over all cameras for each dataset. The available camera calibration for BabelCalib Kalibr (covers the widest FOV range) is visualized in Fig 1. Additionally, Fig 4 supplements its individual results. The plot shows that the optimal number of control points is strongly correlated to camera types. However, there are other influencing factors. For instance, BF2M2020S23 and EUROC share similar calibration patterns, however, the observation fractions falling in the near-linear range ( $\theta < 1$  rad) for them are 0.53 and 0.65, respec-

tively, resulting a different trend in Fig 4. For our evaluations, we adopted a fixed 10 control points to demonstrate the general modeling capability of our approach. However, this parameter remains adjustable, allowing flexibility when the camera model is known.

### 5. Generation of Distorted ETH3D dataset

We implement the polynomial radial distortion model in Python with OpenCV, which is defined as:

$$r_{\text{distorted}} = r(1 + k_1 \cdot r + k_2 \cdot r^2), \quad (1)$$

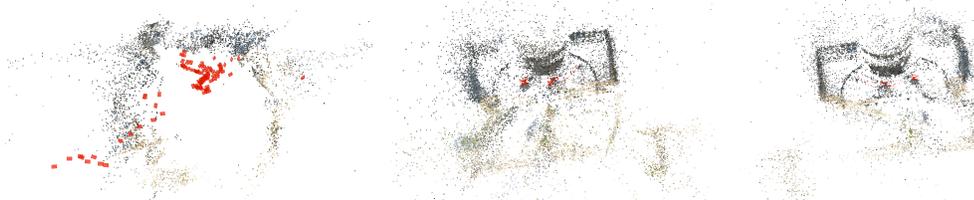
where we set  $k_1 = 0.3, k_2 = 0.3$ . A central crop is applied to remove the black boarder resulting from the fisheye distortion, keeping a scale of 0.7 of the original image dimensions. The cropped image is then resized back to the original resolution using bilinear interpolation.

### References

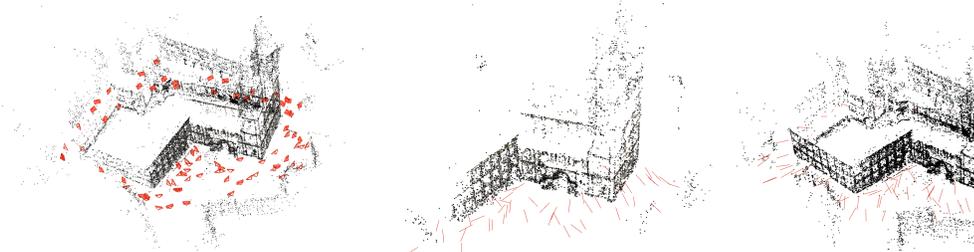
- [1] Viktor Larsson, Nicolas Zobernig, Kasim Taskin, and Marc Pollefeys. Calibration-free structure-from-motion with calibrated radial trifocal tensors. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part V 16*, pages 382–399. Springer, 2020. 2, 4
- [2] Yaroslava Lochman, Kostiantyn Liepeshov, Jianhui Chen, Michal Perdoch, Christopher Zach, and James Pritts. Babelcalib: A universal approach to calibrating central cameras. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 15253–15262, 2021. 1
- [3] Linfei Pan, Marc Pollefeys, and Viktor Larsson. Camera pose estimation using implicit distortion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12819–12828, 2022. 2, 4
- [4] Johannes L Schonberger and Jan-Michael Frahm. Structure-from-motion revisited. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4104–4113, 2016. 2
- [5] Thomas Schops, Johannes L Schonberger, Silvano Galliani, Torsten Sattler, Konrad Schindler, Marc Pollefeys, and Andreas Geiger. A multi-view stereo benchmark with high-resolution images and multi-camera videos. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3260–3269, 2017. 1, 4



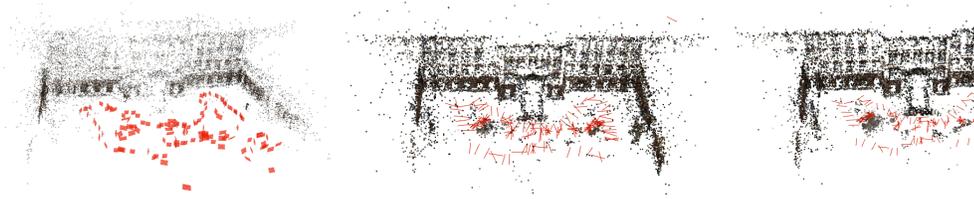
Method	Registered Image	#3D Points
COLMAP	104/108	20.9k
[1] + [3]	108/108	7.1k
Ours	108/108	20.4k



Method	Registered Image	#3D Points
COLMAP	118/120	67.3k
[1] + [3]	120/120	43.4k
Ours	120/120	60.0k



Method	Registered Image	#3D Points
COLMAP	94/94	30.8k
[1] + [3]	48/94	6.9k
Ours	94/94	28.7k



Method	Registered Image	#3D Points
COLMAP	126/157	43.8k
[1] + [3]	157/157	13.6k
Ours	157/157	32.5k

Figure 2. More comparative reconstruction results from catadioptric lens or fisheye lens filmed images. The reconstructions in each row come from COLMAP [4], RadialSfM [1] with pose upgraded and bundle adjusted with [3] and, our pipeline, from left to right.

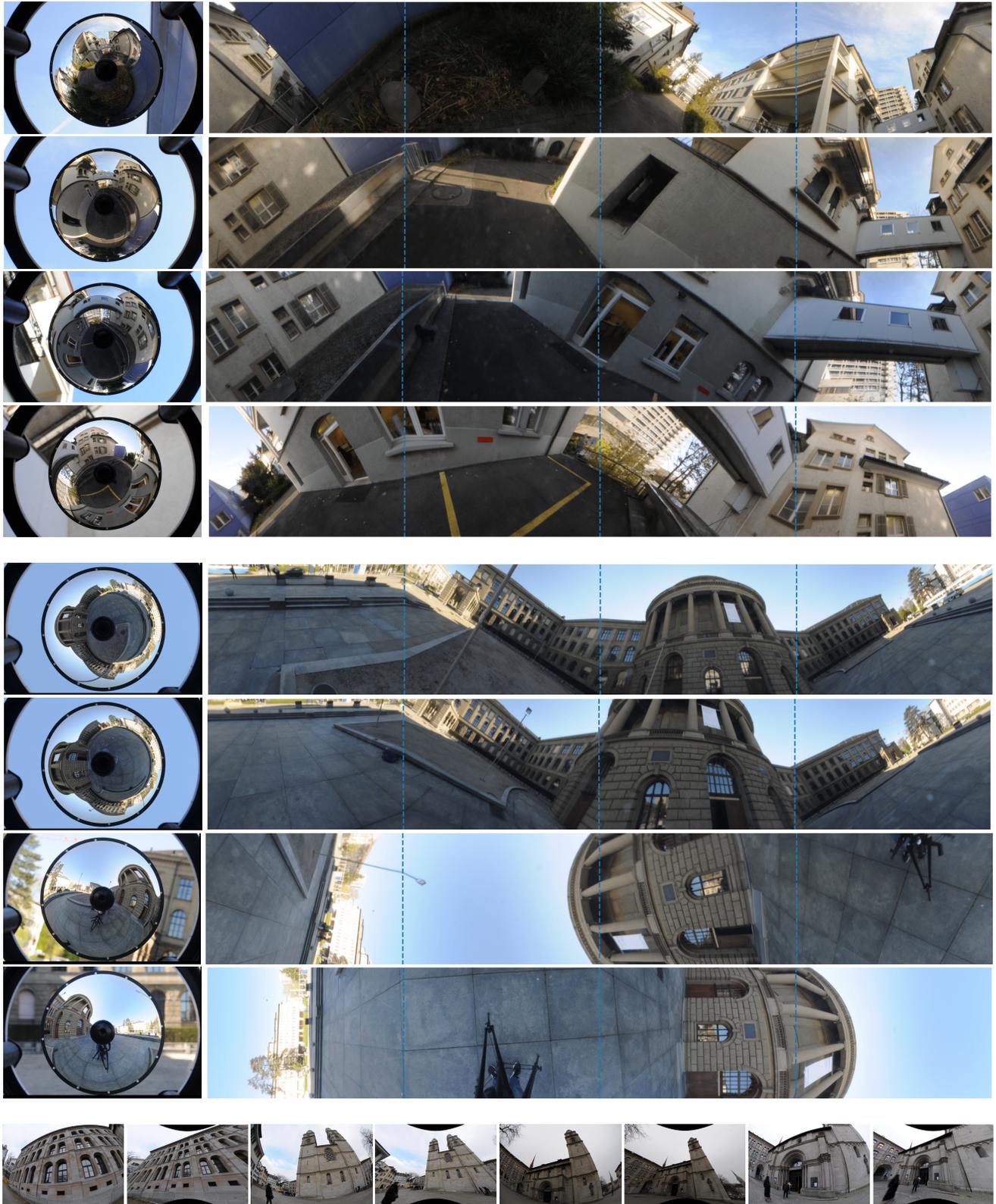


Figure 3. Undistortion results with estimated calibration map on fisheye and catadioptric images.

	AUC @ 1°					AUC @ 3°					AUC @ 5°				
	COLMAP (pinhole)	COLMAP (radial)	COLMAP (thin prism)	[1]+[3]	GenSfM (ours)	COLMAP (pinhole)	COLMAP (radial)	COLMAP (thin prism)	[1]+[3]	GenSfM (ours)	COLMAP (pinhole)	COLMAP (radial)	COLMAP (thin prism)	[1]+[3]	GenSfM (ours)
courtyard	2.9	22.1	56.7	43.1	59.6	5.8	52.6	83.3	59.8	85.2	10.8	67.2	90.0	64.2	91.1
delivery_area	2.5	2.0	70.2	14.7	57.9	5.6	2.7	89.6	23.6	84.0	9.9	3.3	93.7	26.0	90.2
electro	2.0	8.5	56.9	6.1	45.0	4.4	23.7	70.6	10.1	64.2	8.2	30.7	73.9	11.1	69.2
facade	4.2	37.9	72.5	56.6	70.6	18.3	69.7	88.1	68.2	87.0	29.3	79.7	91.7	71.1	91.0
kicker	3.5	21.4	52.4	30.9	55.2	9.2	59.6	76.7	44.8	77.8	18.1	71.8	83.0	48.4	83.6
meadow	1.3	0.9	0.9	-	-	1.3	0.9	0.9	-	-	1.3	0.9	0.9	-	-
office	2.8	15.2	0.3	15.6	23.3	8.0	29.7	0.3	25.6	39.2	15.0	36.4	0.3	28.5	45.7
pipes	6.0	9.3	1.0	-	-	13.3	17.4	1.0	-	-	18.6	20.9	1.0	-	-
playground	2.7	39.2	66.1	15.7	71.6	9.1	65.6	78.4	19.6	89.9	16.2	72.9	80.9	20.6	93.9
relief	17.2	35.8	70.3	36.8	66.2	52.8	68.2	90.1	48.3	87.9	68.6	77.4	94.0	51.0	92.7
relief_2	11.2	27.3	72.6	0.0	59.6	30.6	69.1	90.8	0.0	85.8	50.7	80.7	94.5	0.0	91.4
terrace	5.2	20.5	77.4	17.5	59.2	13.8	55.0	92.5	29.5	85.9	24.3	71.7	95.5	32.5	91.6
terrains	4.7	7.6	0.1	1.0	65.8	16.3	24.3	0.1	1.1	87.8	25.4	37.0	0.1	1.4	92.5
Average	5.1	19.1	45.9	18.3	48.8	14.5	41.4	58.6	25.4	67.3	22.8	50.0	61.5	27.3	71.8

Table 1. Detailed camera pose evaluation on ETH3D [5]. Our method constantly achieves the best or second-best performance across different evaluation thresholds.

	AUC @ 1°					AUC @ 3°					AUC @ 5°				
	COLMAP (pinhole)	COLMAP (radial)	COLMAP (thin prism)	[1]+[3]	GenSfM (ours)	COLMAP (pinhole)	COLMAP (radial)	COLMAP (thin prism)	[1]+[3]	GenSfM (ours)	COLMAP (pinhole)	COLMAP (radial)	COLMAP (thin prism)	[1]+[3]	GenSfM (ours)
courtyard	1.7	1.7	0.1	1.7	55.5	1.7	1.7	0.1	1.8	82.0	1.7	1.8	0.1	2.2	89.1
delivery_area	1.1	1.1	0.1	8.4	55.2	1.2	1.2	0.1	14.4	83.2	1.2	1.2	0.1	16.1	89.7
electro	0.7	0.8	0.1	5.8	41.1	0.7	0.9	0.1	10.0	58.5	0.8	1.0	0.1	11.0	63.0
facade	1.1	1.1	0.1	41.4	70.3	1.3	1.3	0.1	58.3	87.0	1.7	2.1	0.1	62.9	91.1
kicker	2.4	2.2	0.2	2.4	50.3	2.4	2.2	0.2	2.4	75.9	2.4	2.4	0.2	2.5	82.5
meadow	2.7	0.9	0.9	-	-	2.7	0.9	0.9	-	-	2.7	0.9	0.9	-	-
office	0.6	0.4	0.3	-	-	0.6	0.4	0.3	-	-	0.6	0.4	0.3	-	-
pipes	3.1	3.1	1.0	-	-	3.1	3.5	1.0	-	-	3.1	4.3	1.0	-	-
playground	1.1	1.2	0.1	10.7	0.8	1.1	1.2	0.1	12.6	0.8	1.1	1.2	0.1	13.0	0.8
relief	2.6	2.5	0.2	30.2	51.1	5.9	2.5	0.2	39.9	70.9	9.1	2.5	0.2	42.2	75.2
relief_2	1.9	2.4	0.2	22.3	44.2	2.1	2.8	0.3	29.8	61.5	2.9	3.1	0.4	31.4	65.2
terrace	2.1	1.9	0.4	2.3	2.1	2.1	1.9	0.4	3.1	2.1	2.1	1.9	0.4	4.5	2.1
terrains	1.3	1.5	0.1	0.8	1.2	1.4	1.5	0.1	2.0	2.9	1.5	1.7	0.1	3.0	3.6
Average	1.7	1.6	0.3	9.7	28.6	2.0	1.7	0.3	13.4	40.4	2.4	1.9	0.3	14.5	43.2

Table 2. Detailed camera pose evaluation on distorted ETH3D [5]. We remain comparable performance against the original dataset across different evaluation thresholds

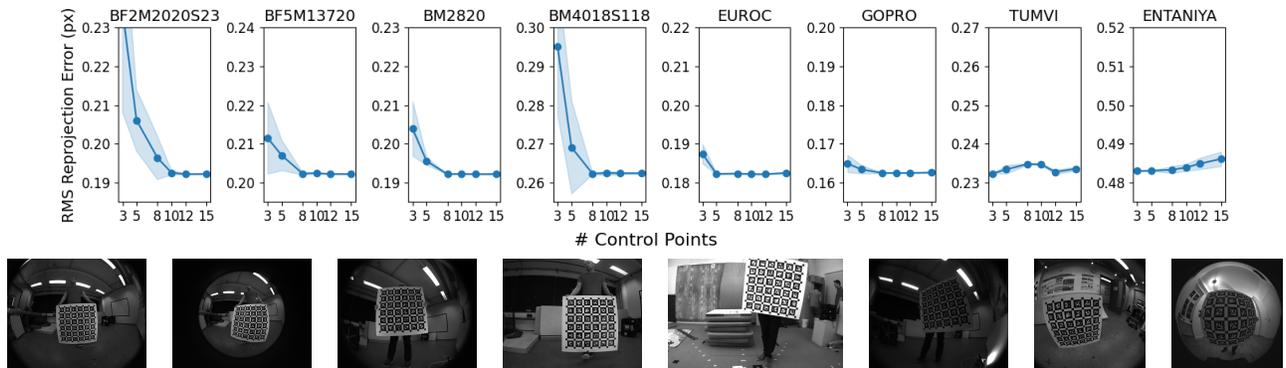


Figure 4. Impact of control points number on calibration and exemplary image (1st image in train set) for each camera in Kalibr.