Fisheye Absolute Pose and Focal Length Estimation Revisited

Supplementary Material

9. More details for dataset processing of `RobotCar-Season(Fisheye)`

As mentioned in Sec. 7.1, `RobotCar-Season` is constructed from the original raw Oxford Robot Car dataset. And we construct `RobotCar-Season(Fisheye)` from `RobotCar-Season`.

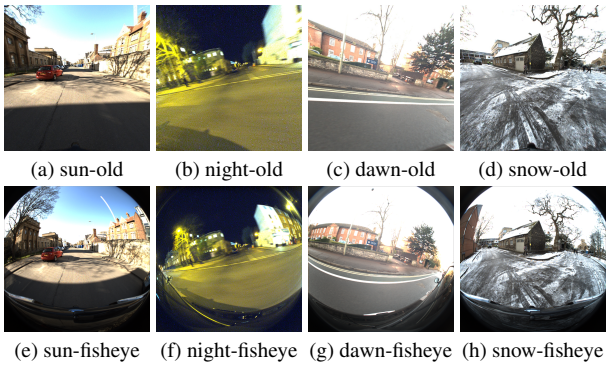
To construct the `RobotCar-Season(Fisheye)`, we replaced as many fisheye raw images as could be found in the original Robot Car dataset. Specifically, $6,079$ images were replaced using the exact same timestamps, $4,359$ were replaced using the closest available timestamps within a $\pm 10\mu s$ interval, and $1,498$ images belonged to subsequences for which no fisheye data were available. In total, `RobotCar-Season(Fisheye)` contains $10,438$ fisheye query images, still a large number of images to demonstrate experiments. Figure 6 presents some examples of replaced fisheye images.

Since the evaluation was performed via submission to the benchmark server that evaluates results on all images, for query images without fisheye data, we used the corresponding pinhole results, *i.e.*, the poses obtained using calibrated PnP-RANSAC. The median runtime reported in Tab. 3 for `RobotCar-Season(Fisheye)` is therefore computed only over the fisheye queries.

10. Experiments on `Calibration` dataset

Given that both `Oxford Day-and-Night` and `RobotCar-Season(Fisheye)` each use only a single fisheye lens, we additionally consider a calibration dataset featuring a diverse set of fisheye cameras and lenses to further evaluate the performance of fisheye solvers. The `Calibration` dataset from BabelCalib [29] was originally collected for camera calibration. It includes four collections from established datasets and one self-recorded collection, and provides the 2D–3D checkerboard corner correspondences. In total, it contains 41 cameras with diagonal FoVs ranging from $88°$ to $268°$.

After removing the catadioptric rigs, 35 cameras with 1526 images remain. We use the calibration by BabelCalib as the ground truth intrinsics and absolute poses. Since the correspondences in this dataset are nearly perfect inliers, we set a small LO-RANSAC threshold of 1 px and a small minimal iteration number of 10. We include these experiments in the supplementary material, as the dataset is relatively easy. Nevertheless, it provides a useful benchmark for sanity checks, so we report the results for completeness.



(a) sun-old (b) night-old (c) dawn-old (d) snow-old

(e) sun-fisheye (f) night-fisheye (g) dawn-fisheye (h) snow-fisheye

Figure 6. Example of replaced fisheye images for `Robot-Season(Fisheye)`.

10.1. Robust Estimation

Table 4 shows the robust estimation results on `Calibration` dataset. As we analyzed in the main paper, RANSAC with good approximate model based solvers generalizes well on this wide range of fisheye lens dataset. **P3.5PF** and **RECALIB** do not work well because of their bad approximate model. LM-optimized fisheye solvers consistently demonstrate better estimation than non-optimized solvers in terms of focal length accuracy and reprojection errors with fewer RANSAC iterations.

10.2. Ablation on focal length priors

P3PSAMP relies on a predefined list of FoV-derived focal lengths. This experiment evaluates how different focal-length priors influence the performance of **P3P** and **P3P-LM**, highlighting the trade-off between prior quality and pose/focal length accuracy. Similarly, the proposed **P5PFL** also requires an initial focal length to enable the linearization described in Sec. 5.4, so we include both **P5PFL** and **P5PFL-LM** in the comparison.

We select five representative fisheye cameras from the `Calibration` dataset, covering a range from large to narrow FoVs, plus one out-of-prior case. Example images are

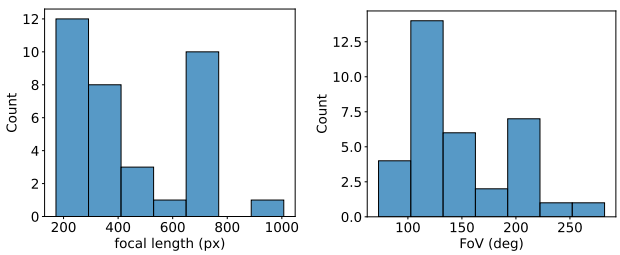
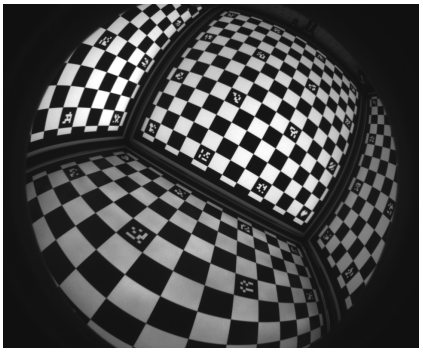


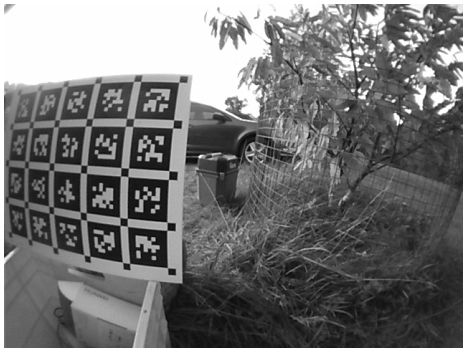
Figure 7. Focal length and FoV distribution of `Calibration`.

Table 4. Results on `Calibration` [29] using LO-RANSAC with threshold 1px. We report the percentage of images correctly calibrated with three thresholds for focal length errors and reprojection errors. Meanwhile, we also report the mean reprojection errors in pixels, the number of RANSAC iterations and the mean runtime in milliseconds. The best results are marked in bold.

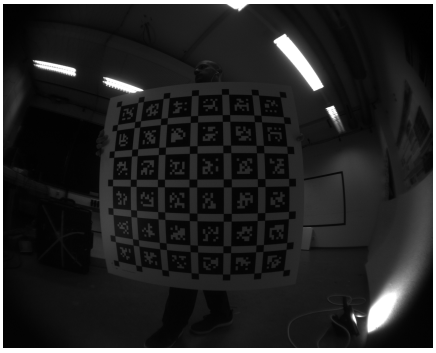
Method	AUC@($f < 1/2/5$ px) (%)	AUC@(reproj $< 0.5/1/2$ px) (%)	mean reproj(px)	#iters	mean runtime(ms)
RECALIB	57.14 / 74.44 / 91.09	70.97 / 93.25 / 97.18	11.3760	162.66	7.26
P3PAC	57.54 / 75.29 / 91.48	72.67 / 93.97 / 97.38	0.8166	40.39	144.65
P3.5PF	38.93 / 52.82 / 66.32	50.59 / 68.48 / 72.21	119.3185	632.39	28.50
P3.5PF-LM	58.78 / **76.34** / 92.53	73.00 / **95.28** / **98.03**	0.4373	30.36	5.50
P3PSAMP	58.65 / 75.88 / 91.48	73.07 / 94.43 / 97.97	0.4451	31.18	**2.59**
P3PSAMP-LM	58.85 / 76.08 / **92.66**	73.07 / 95.22 / **98.03**	0.4373	**29.48**	3.26
P4PFR	58.26 / 75.75 / 91.81	72.80 / 94.89 / 97.90	0.4582	35.78	3.48
P4PFR-LM	**58.91** / 76.21 / 92.53	73.13 / 95.22 / **98.03**	0.4376	30.04	5.29
P5PFR	58.32 / 76.08 / 92.20	72.87 / 95.02 / 97.77	0.6235	51.30	2.94
P5PFR-LM	58.58 / 76.21 / **92.66**	**73.20** / 95.22 / **98.03**	0.4370	38.58	4.29
P5PFL	57.73 / 75.23 / 90.89	72.08 / 93.91 / 96.99	0.5745	72.93	3.55
P5PFL-LM	58.45 / 76.21 / **92.66**	**73.20** / 95.22 / **98.03**	**0.4368**	38.63	4.98



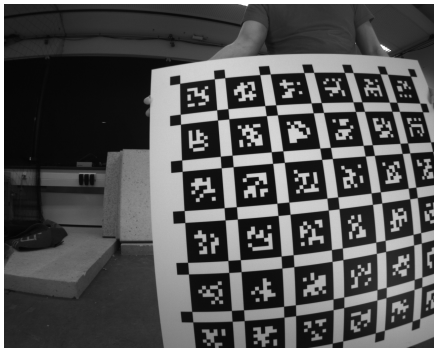
OV-cube-cv00
(large FoV)



UZH-Snapdragon
(median FoV)



Kalibr-GoPro
(median FoV)



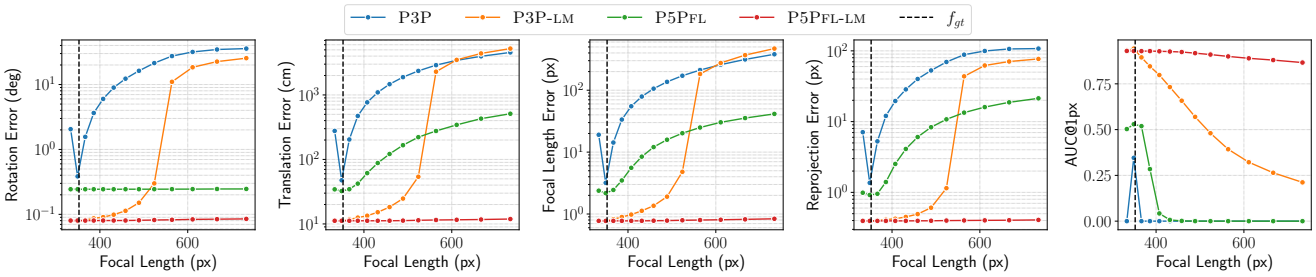
Kalibr-BM4018S118
(narrow FoV)

Kalibr-EUROC
(out-of-prior FoV)

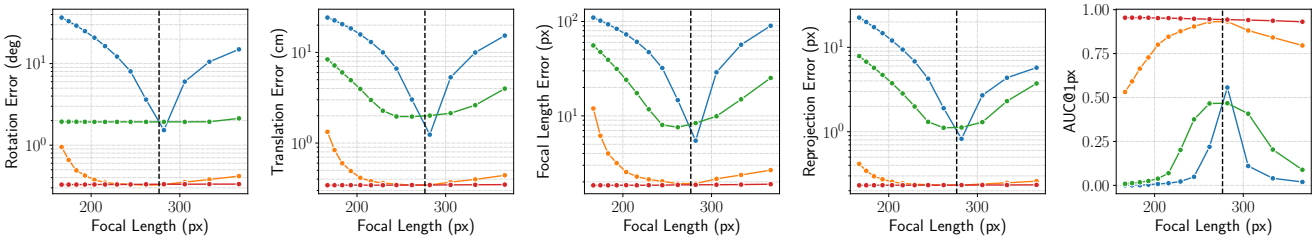
Figure 8. Selected sample frames from the `Calibration` dataset covering a broad range of fisheye fields of view.

shown in Fig. 8. The focal-length priors are obtained from a predefined FoV list ranging from 100° to 220° in 10°, converted to focal length based on the corresponding image size. Note that **P3P-LM** first runs **P3P** on three correspondences and then refines the estimate using four points in the LM step. The same setting as we described in the main paper Sec. 5.2.

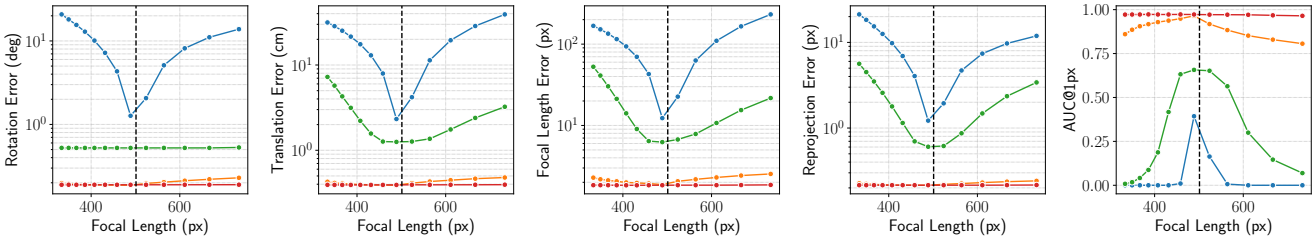
The ablation results in Fig. 9 show that **P5PFL** is consistently more robust to focal-length prior errors than **P3P**. Moreover, **P5PFL-LM** achieves strong performance even under severely inaccurate priors. The reason could be that the accurate rotation estimation provided by **P5PFL** gives LM a reliable initialization.



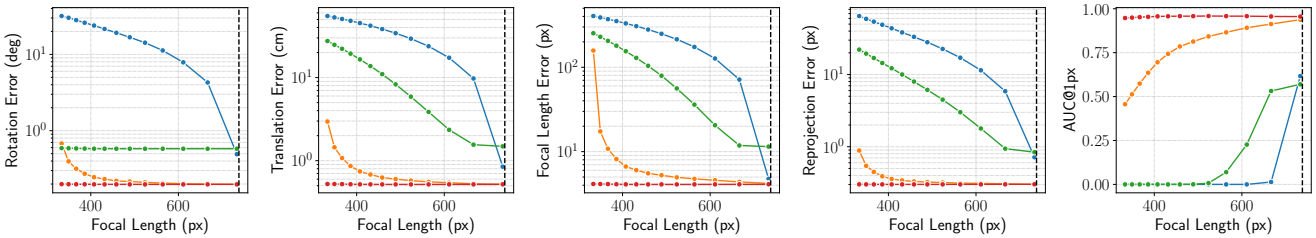
(a) **OV-cube-cv00** (large FoV)



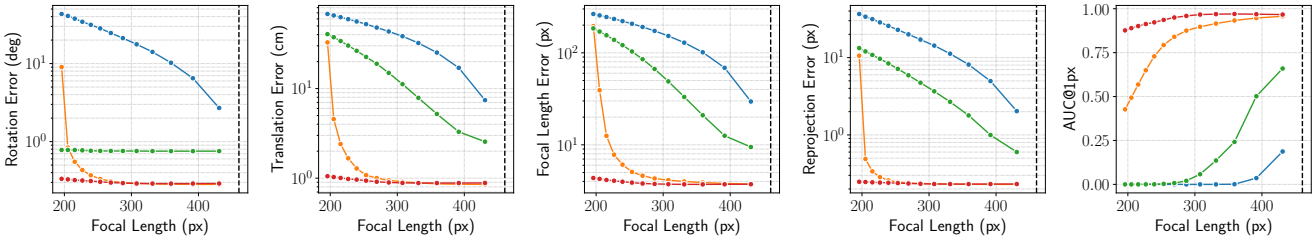
(b) **UZH-Snapdragon-outdoor-forward** (median FoV)



(c) **Kalibr-GOPro** (median FoV)



(d) **Kalibr-BM4018S118** (narrow FoV)



(e) **Kalibr-EUROC** (out-of-prior FoV)

Figure 9. Ablation results on the sensitivity to focal length priors are presented for different solvers across multiple FoV cameras from the `Calibration` dataset. All solvers are evaluated in their minimal configurations. We report rotation error (degrees), translation error (centimeters), focal length error (pixels), reprojection error (pixels), and the percentage of successfully calibrated images with reprojection error below 1 pixel. The results show that **P5PFL** is more robust to the focal length prior than **P3P**, maintaining stable performance even under large prior deviations.

11. RANSAC threshold analysis

We perform a threshold ablation on the day sequences of `Oxford Day-and-Night`, varying the threshold from 4 to 20 px. It can be observed in Fig. 10 that performance is stable within a reasonable range, with 12px providing a good accuracy–robustness trade-off.

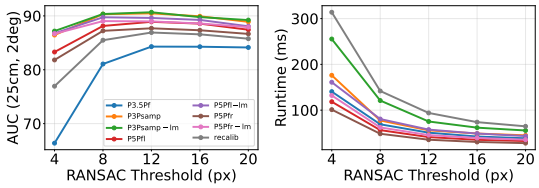


Figure 10. RANSAC threshold ablation on the day sequences of `Oxford Day-and-Night`; 12 px yields a good accuracy–robustness trade-off.

12. Discussion

We note that more complex fisheye models are generally preferable for high-precision offline calibration. Extending the problem studied in this paper to alternative projection models would, however, require substantial re-derivations, together with a careful study of the resulting accuracy–runtime trade-offs. We consider this an interesting direction for future work, but leave it outside the scope of the present paper.

In this work, we focus on the equidistant projection model, $r = f\,\theta$. Despite its simple form, it remains relevant in a fairly broad range of practical settings. Moreover, many real-world fisheye lenses are designed to approximately follow the equidistant mapping.

As shown in [29] and [43] , the Kannala–Brandt (KB) model, which extends the equidistant model, offers strong flexibility and accuracy across a wide range of lenses. A natural extension of the present work would therefore be to use the KB model in the non-linear refinement step.