

Deep Single Fisheye Image Camera Calibration for Over 180-degree Projection of Field of View

Supplementary material

Nobuhiko Wakai
Panasonic Corporation

wakai.nobuhiko@jp.panasonic.com

Takayoshi Yamashita
Chubu University

takayoshi@isc.chubu.ac.jp

1. Comparison of bearing loss and non-grid bearing loss

We describe the effectiveness of our non-grid bearing loss designed for fisheye camera calibration. To compare bearing loss [3] and our non-grid bearing loss, we also train our network using the bearing loss on condition that other settings are corresponded to the experiment in main paper. Although standard image grids in the bearing loss are projected to a unit sphere, there are invalid grid points outer image circles. Moreover, standard grid points are unbalance for incident angles.

Table 1 shows that our non-grid bearing loss has small absolute errors in all parameters compared with the bearing loss. In addition, error distribution in the bearing loss is wider than that in non-grid bearing loss in Fig. 1. In particular, the bearing loss causes large errors in long focal length. Therefore, non-grid bearing loss is suit for fisheye camera calibration with over 180° of field of views (FOV).

2. Experiments using SP360 dataset

For evaluating our method calibrating fisheye cameras, we conduct training and evaluating our network compared with conventional calibration methods. First, we describe dataset and then show experimental results.

2.1. Dataset

We use a large-scale dataset of outdoor panoramas with sun positions called SP360 dataset [1] artificially to make images using arbitrary camera parameters. The difference of SP360 and StreetLearn dataset (Manhattan 2019 subset) [4] is as follows: Image size of SP360 dataset is 3328×1664 pixels, *i.e.*, half width and height of StreetLearn dataset. Additionally, SP360 dataset is divided into train (including validation) and test sets of 19,038 and 55 images, respectively. We render 27 and 300 image patches for train and test sets, respectively. Train, validation, and

test sets have 502, 204, 11, 822, and 16, 500 images, respectively because we use validation rate 0.023 for train and validation division. Note that the number of images in SP360 dataset is approximately one third of that in StreetLearn dataset. Thus, we render three times image patches in SP360 dataset compared with that in StreetLearn dataset. In experiment using SP360 dataset, we follows the main paper settings except for the dataset.

2.2. Experimental results

We show experimental results using SP360 dataset in the same manner of main paper evaluation.

2.2.1 Error distribution of our network

We show the error distribution of our network using the test set of SP360. We show the absolute errors between ground-truth and predicted values among these parameters in Fig. 2.

In our method, the absolute errors between ground-truth and predicted parameters in tilt angle θ , roll angle ψ , and focal length f are 6.62 ± 13.21 [deg], 9.34 ± 19.89 [deg], and 0.276 ± 0.257 [mm], respectively in Tab. 1. In López-Antequera’s method [3], the absolute errors in tilt angle θ and roll angle ϕ are 32.25 ± 33.79 [deg] and 44.96 ± 25.94 [deg], respectively. Therefore, our method precisely calibrates rotation angles compared with López-Antequera’s method.

2.2.2 Reprojection error

In the test set of SP360, the reprojection errors of our method and López-Antequera’s method are 17.46 ± 15.43 and 41.80 ± 47.81 pixels, respectively. In addition to rotation errors, our method has small reprojection errors compared with López-Antequera’s method that has large reprojection errors due to large angle errors. This trend is the same using StreetLearn dataset in evaluation.

Loss	θ error [deg]	ψ error [deg]	f error [mm]
Bearing loss [3]	7.29 ± 14.03	10.12 ± 20.78	0.318 ± 0.301
Non-grid bearing loss	6.62 ± 13.21	9.34 ± 19.89	0.276 ± 0.257

Table 1: Comparison of bearing loss and non-grid bearing loss using our network.

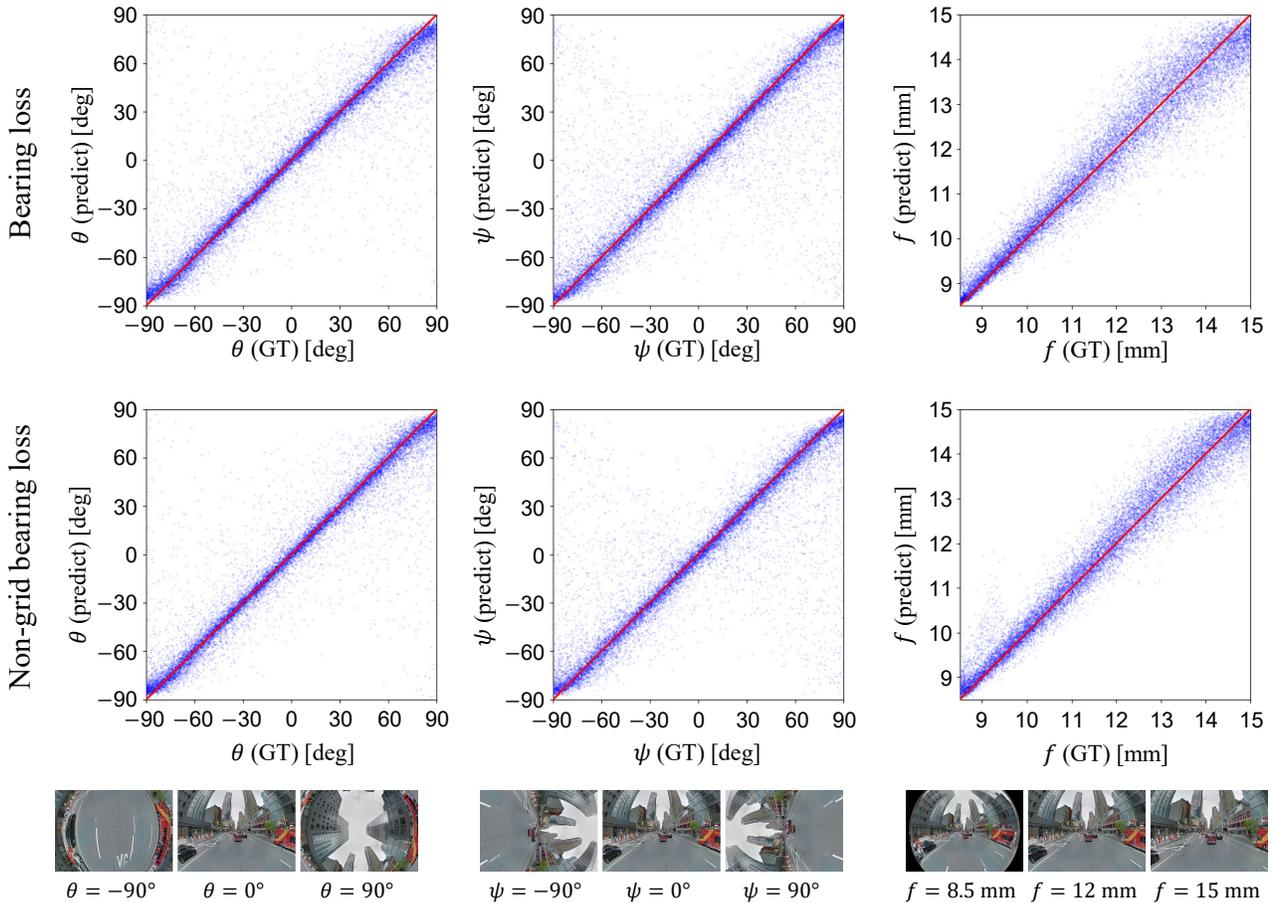


Figure 1: Error distribution on the test set of 16, 100 images in StreetLearn dataset. The horizontal axis indicates ground truth values of parameters. The vertical axis indicates predicted parameters. The diagonal red lines indicate perfect prediction. The bottom images are examples of rendered images using notated camera parameters.

2.2.3 Comparison using PSNR and SSIM

We evaluate the peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) [6] for intrinsic parameters. In image rectification task, extrinsic parameters are ignored because the image rectification is carried out using only intrinsic parameters.

Table 2 shows comparison of PSNR and SSIM in our test set of SP360. Our method outperforms conventional methods in both PSNR and SSIM the same as StreetLearn dataset evaluation. Note that we exclude Santana-Cedr s’s method for quantitative evaluation because it does not work in many images where the line detector fails to extract lines.

Figure 3 shows that the qualitative rectification results on our test dataset generated by our method and the others. In various street images, our method obtains overall the most similar to the ground truth images even if cameras are rotated with large angles.

As described above, our method outperforms conventional methods using SP360 as well as using StreetLearn dataset.

References

[1] S. Chang, C. Chiu, C. Chang, K. Chen, C. Yao, R. Lee, and H. Chu. Generating 360 outdoor panorama dataset with reliable

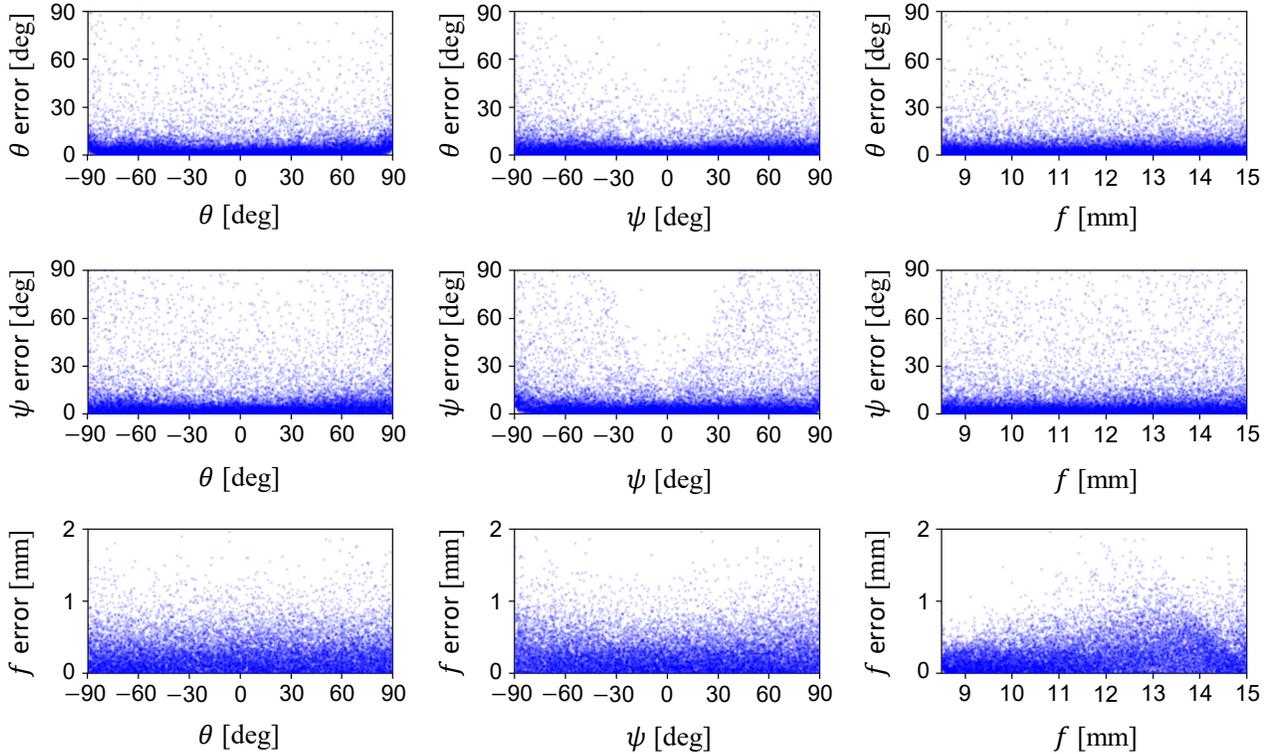


Figure 2: Errors on the test set of 16,500 images in SP360 dataset. The horizontal axis denotes ground truth values while the vertical axis denotes the absolute error between ground-truth and predicted values.

Method	Learning	Extrinsics	Intrinsics	Projection	Over 180° FOV	PSNR \uparrow	SSIM \uparrow
Santana-Cedr�s [5] ¹			✓	Perspective		-	-
Yin [7]	✓		✓	Fisheye ²	✓	15.46 \pm 1.59	0.3914 \pm 0.1024
Liao [2]	✓		✓	Perspective		15.75 \pm 1.90	0.4413 \pm 0.0100
L�pez-Antequera [3]	✓	✓	✓	Perspective		16.67 \pm 3.02	0.5034 \pm 0.0159
Ours	✓	✓	✓	Fisheye	✓	21.60 \pm 5.06	0.6434 \pm 0.1818

¹ Exclusion for evaluation due to failure of line detection in many images.

² Using generalized fisheye camera models.

Table 2: Comparison of conventional methods and our method using the test set of SP360.

- sun position estimation. In *Proceedings of SIGGRAPH Asia*, 2018. 1
- [2] K. Liao, C. Lin, and Y. Zhao. A deep ordinal distortion estimation approach for distortion rectification. *arXiv preprint arXiv:2007.10689*, 2020. 3, 4
- [3] M. L pez-Antequera, R. Mar , P. Gargallo, Y. Kuang, J. Gonzalez-Jimenez, and G. Haro. Deep single image camera calibration with radial distortion. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11809–11817, 2019. 1, 2, 3, 4
- [4] P. Mirowski, A. Banki-Horvath, K. Anderson, D. Teplyashin, K. M. Hermann, M. Malinowski, M. K. Grimes, K. Simonyan, K. Kavukcuoglu, A. Zisserman, and R. Hadsell. The StreetLearn environment and dataset. *arXiv preprint arXiv:1903.01292*, 2019. 1
- [5] D. Santana-Cedr s, L. Gomez, M. Alem n-Flores, A. Salgado, J. Esclar n, L. Mazorra, and L. Alvarez. An iterative optimization algorithm for lens distortion correction using two-parameter models. *Image Processing On Line*, 6:326–364, 2016. 3, 4
- [6] Z. Wang and A. C. Bovik. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing (TIP)*, 13(4):600–612, 2004. 2
- [7] X. Yin, X. Wang, J. Yu, M. Zhang, P. Fua, and D. Tao. FishEyeRecNet: A multi-context collaborative deep network for fisheye image rectification. In *Proceedings of European Conference on Computer Vision (ECCV)*, 2018. 3, 4

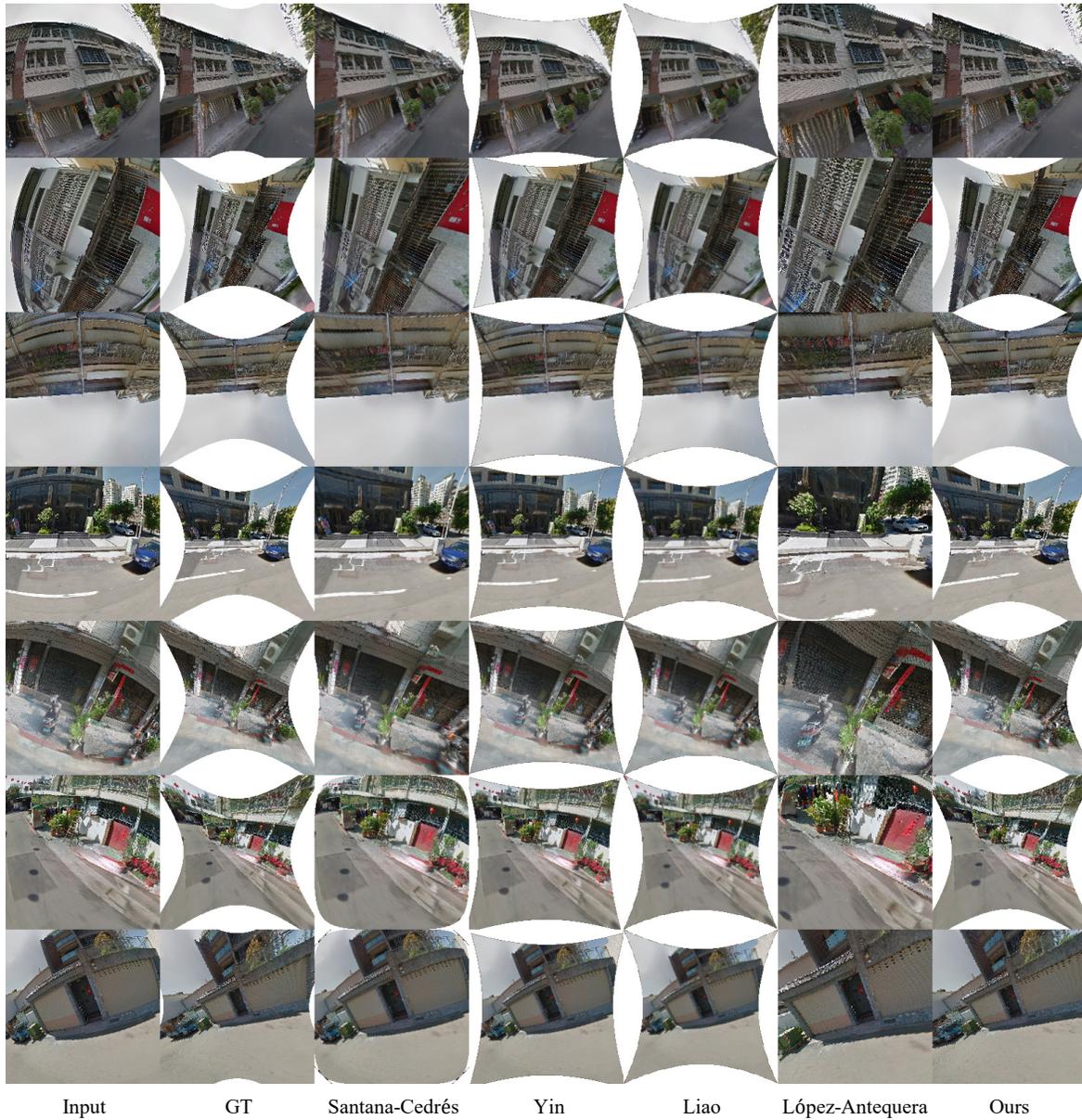


Figure 3: Qualitative results on our test images of SP360. We show the input image, the ground truth image, and results of the compared methods: Santana-Cedrés [5], Yin [7], Liao [2], López-Antequera [3], and our method from left to right.