Revisiting Calibration of Wide-Angle Radially Symmetric Cameras – Supplementary Material –

Andrea Porfiri Dal Cin[1], Francesco Azzoni[1], Giacomo Boracchi[1], and Luca Magri[1]

DEIB, Politecnico di Milano, Italy
firstname.lastname@polimi.it

In this document, we provide additional details concerning the main paper that could not be included in the manuscript due to space limitations.

7 Additional Details on Camera Models

7.1 Camera Model Ambiguity

In Sec. 1 of the main paper, we refer to the problem of camera *ambiguity* that makes end-to-end calibration ill-posed and challenging for camera models with inherent ambiguities. This section shows that the EUCM and DSCM camera models are ambiguous, while the UCM camera is not. We formally define an ambiguous camera model to better elucidate the issues of ambiguity.

Definition (Ambiguous Camera Model). A camera model $\mathcal{M}$ is considered *ambiguous* if there exists at least one pair of intrinsic parameter configurations $\mathbf{i}_{\mathcal{M}} = (f, d_1, \dots, d_n)$, $\mathbf{i}'_{\mathcal{M}} = (f', d'_1, \dots, d'_n)$ that result in the same projection equations.

It is important to emphasize that, in practical scenarios, identifying whether a camera model is ambiguous is impractical due to the complexity inherent in the projection equations, making analytical approaches largely ineffective. Consequently, establishing whether a camera model $\mathcal{M}$ is ambiguous often involves an exhaustive search across potential parameter configurations to identify at least one pair resulting in identical camera projection equations.

Determining whether two configurations (sets of camera parameters) are ambiguous, *i.e.*, if they represent the same physical camera, hinges on the reprojection error calculated by back-projecting image points onto the unit sphere with one set of parameters and then reprojecting these points onto the image plane with the other set of parameters. When the reprojection error is zero or within numerical precision for two different sets of parameters, then projection equations are equal, and thus, the camera model is ambiguous.

We investigate camera model ambiguities by proceeding as follows:

(i) We confine our search for ambiguities to baseline configurations by setting the sensor size at $400 \times 400px$ and the Angular Field of View (AFOV) to 150°.

This limits the search scope while retaining comprehensiveness. We experienced that, in most cases, ambiguities are detected within this range of parameters. Nevertheless, if we do not find ambiguities in those ranges, further exploration is performed for other AFOV values.

(ii) We systematically sampled parameters from model $\mathcal{M}$, excluding focal distance f (designated as p_1 or q_1). Parameters $(d_2,\ldots,d_n)$ were uniformly selected, allowing us to construct all potential configurations. Specifically, for the UCM [2], we varied parameter a across 101 points within $[0,1]$. For the EUCM [3], a was tested at 11 values within $[0,1]$, and b at 11 values within $[0,2]$, generating 121 unique configurations. A similar approach was applied to the DSCM [7], where ξ spanned $[-1,1]$.

(iii) Each configuration's reprojection error Re was compared against others and illustrated as heatmaps. The lower the value of the reprojection error, the more similar the configurations. Low values are only located near the main diagonal for a model without ambiguous configurations, where a configuration is compared with itself. On the other hand, ambiguous models also present low values far from the diagonal that correspond to different parameter configurations.

We studied the ambiguity of the UCM [2] and proved that it is free from ambiguities, as detailed in Fig. 1. Then, we studied the EUCM and DSCM camera models since they are the most interesting ones for our synthetic dataset generation procedure. Both models present some ambiguous configuration pairs (Fig. 2 and Fig. 3).

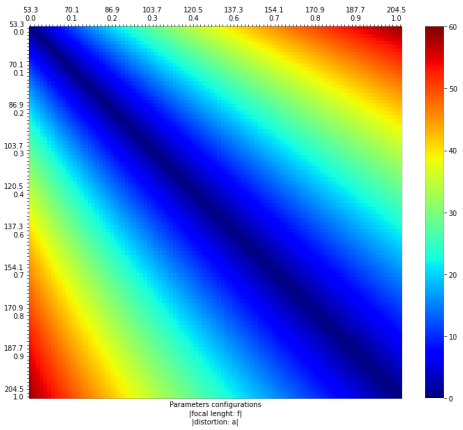

Fig. 1: Ambiguity heatmap of the UCM [2]. The value of the heatmap is the reprojection error; the lower, the more similar the cameras represented by the configurations on the X and Y axis. The configurations on the X and the Y are the same and in the same order thus the error on the main diagonal is zero, and the heatmap is symmetric. Due to the absence of low errors far from the diagonal, we can deduce that this model is not ambiguous.

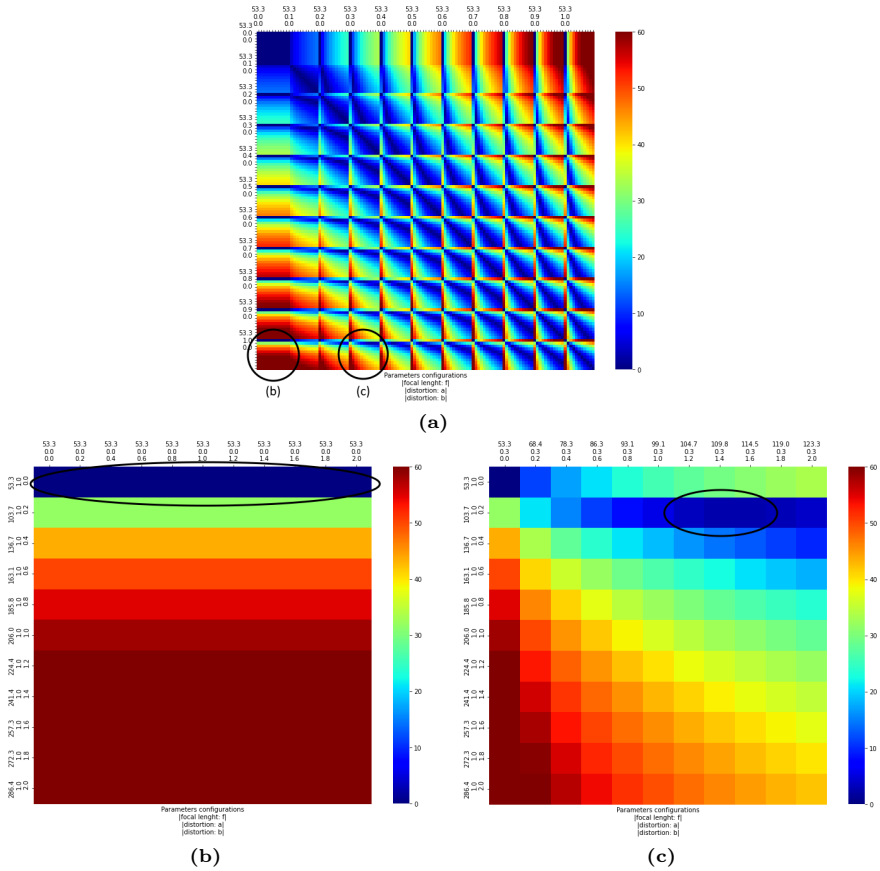

Fig. 2: Ambiguity heatmap of the EUCM [3]. In (a), the entire heatmap differs from the UCM one. Low error values are also present far from the main diagonal, meaning that different configurations lead to similar cameras. In (b) and (c), we report two subsets of the heatmap that contain low errors (identified by a black ellipse). An example of an ambiguous configuration pair in (c) is composed of configuration $(103.7, 1, 0.2)$ on the rows and $(109.8, 0.3, 1.4)$ on the columns.

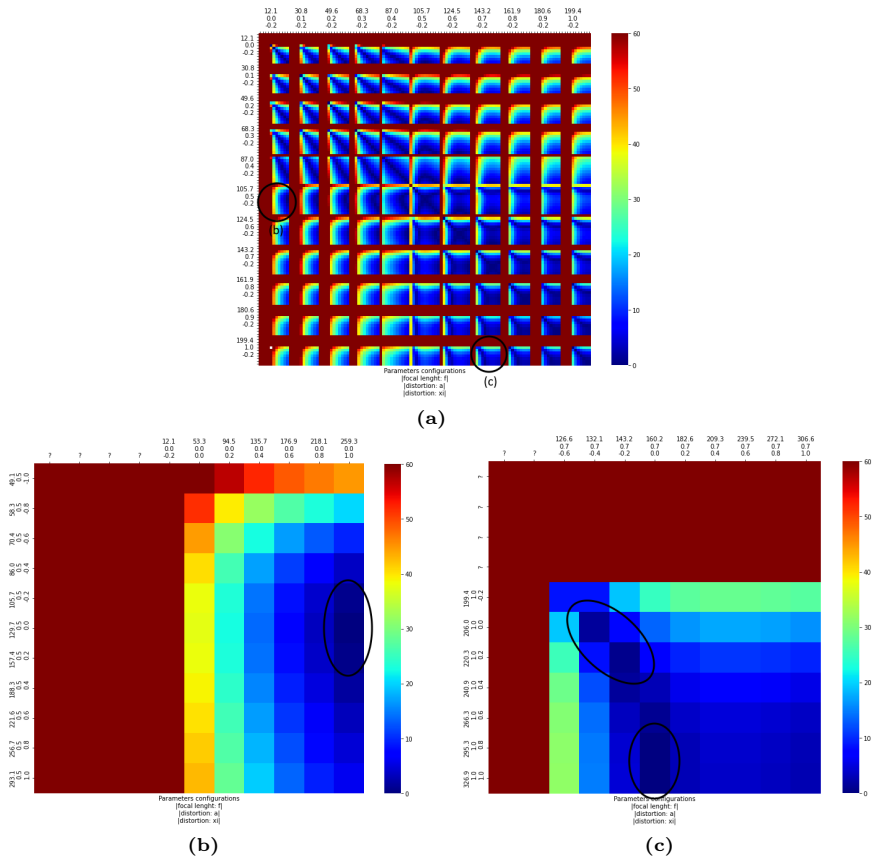

Fig. 3: Ambiguity heatmap of the DSCM [7]. In (a) the entire heatmap, with the same scale as the other ones (Fig. 2). Also, low error values are present far from the main diagonal in this case. Thus, this model is ambiguous. In (b) and (c), we report two subsets of the heatmap that contain low errors (identified by a black ellipse). An example of an ambiguous configuration pair in (c) is composed of configuration $(206, 1, 0)$ on the rows and $(132.1, 0.7, -0.4)$ on the columns.

7.2 Derivation of the focal length from the AFOV

This section elaborates on the content of Sec. 4.3 in the main paper, where we discuss how we generate synthetic datasets by computing the focal length from the Angular Field of View (AFOV) and other intrinsic parameters of the camera.

We report the steps performed to obtain the equation of the focal length f given the AFOV, the sensor size S, and the other intrinsic parameters a and ξ for the Double Sphere Camera Model (DSCM) [7]. Starting from the following AFOV definition:

$$\text{AFOV} = 2\,\mathcal{P}^{-1}_{r,\text{DSCM}}\left(\frac{S}{2}, f, a, \xi\right) , \tag{1}$$

our goal is to recover f. Starting from the previous equation, we follow these steps:

1. Divide by 2 and apply the radial projection function $\mathcal{P}_{r,\text{DSCM}}$ on both terms:

$$\mathcal{P}_{r,\text{DSCM}}\left(\frac{\text{AFOV}}{2}, f, a, \xi\right) = \mathcal{P}_{r,\text{DSCM}}\left(\mathcal{P}^{-1}_{r,\text{DSCM}}\left(\frac{S}{2}, f, a, \xi\right), f, a, \xi\right)$$
$$\mathcal{P}_{r,\text{DSCM}}\left(\frac{\text{AFOV}}{2}, f, a, \xi\right) = \frac{S}{2}$$

2. Isolate the focal length f from the the radial projection function $\mathcal{P}_{r,\text{DSCM}}$, obtained from Eq. 2, as:

$$\mathcal{P}_r(\theta) = f\,\mathcal{D}(\theta, a, \xi) = f\,\frac{\sin\theta}{\alpha e_1 + (1-\alpha)(\xi + \cos\theta)}$$
$$e_1 = \sqrt{1 + \xi^2 + 2\xi\cos\theta}$$

3. Substitute the new projection function in Step 1. and obtain f:

$$f\,\mathcal{D}\left(\frac{\text{AFOV}}{2}, a, \xi\right) = \mathcal{P}_{r,\text{DSCM}}\left(\frac{\text{AFOV}}{2}, f, a, \xi\right) = \frac{S}{2}$$
$$f = \frac{S}{2\,\mathcal{D}\left(\frac{\text{AFOV}}{2}, a, \xi\right)}$$

7.3 Camera Model Representative Power

Choosing between camera models for the synthetic data generation process described in Sec. 4.3 requires evaluating the representative power of these models, *i.e.*, how broad the range of cameras they can represent. We follow the approach used in [8] for this.

Given a camera model, we uniformly sample its camera parameters, generating 10^3 ground truth configurations. Then, we compute the VaCR for the sampled camera configurations and use the robust fitting procedure in Sec. 4.4 to find the set of parameters for all the other considered camera models in our evaluation. We compare the reprojection errors attained by fitting the VaCR of the first model to the others and report results in Tab. 1, where each row corresponds to the model used as ground truth, and the columns report the reprojection errors of the other models. A low reprojection error means that a model (on the columns) can represent the cameras of the model used as ground truth (on the rows) well. According to the results, the EUCM can approximate the UCM well (as expected, since the former is a generalization of this model) and the KB3 well but struggles when considering the DSCM since the reprojection errors are significantly higher. Nevertheless, the errors of the DSCM when modeling the EUCM are 0.449429, which is significantly smaller than the vice-versa: 0.521680. In other words, the DSCM can reasonably represent the cameras

	UCM	EUCM	DS	KB3
UCM	-	**0.019747**	0.030701	3.337970
EUCM	1.296446	-	**0.449429**	3.923736
DS	1.103596	**0.521680**	-	2.690434
KB3	0.081717	**0.020336**	0.068577	-

Table 1: Comparison between the camera models using the mean reprojection error Re. The error reported in a cell is the mean Re of the model on the column calibrated on 10^3 configurations of the model on the row. For each reference model, *i.e.*, each row, we report in **bold** the model that best approximates it. The EUCM is the best approximation for the UCM and KB3. At the same time, the DSCM is the model with the best overall representative power (note the significant difference between the underlined results).

modeled by the EUCM, while the EUCM struggles to represent the cameras of the DSCM. Moreover, the DSCM also suits the cameras of UCM and KB3, as we can see from the small gap with the best scores.

In conclusion, the UCM and the KB3 have the lowest representative power, as we expected, since they are less complex. They both have only two parameters: focal length f and one parameter for distortion. On the other hand, both DSCM and EUCM have three parameters (focal length f and two distortion parameters) and achieve better results. A similar study that compared those models' time complexity and calibration accuracy (except KB3) is reported in [7]. The conclusion is the same as ours: the DSCM achieves higher accuracy than the EUCM when modeling real lenses at the cost of a slightly increased computation time.

8 Radial Symmetry Proof of DSCM Camera Model

In this section, we report proof that the Double Sphere Camera Model (DSCM) [7] is radially symmetric, given the constraints specific to our analysis, as mentioned in Sec. 4.1 of the main paper. Beginning with the Cartesian projection function $\mathcal{P}$, we derive the radial projection function $\mathcal{P}_r$. This transformation results in a model where the projection relies solely on the incident angle θ, rendering the azimuthal angle ϕ invariant. This simplification underscores the radial symmetry inherent in the DSCM under our problem's assumptions, confirming its suitability for representing wide-angle lenses without azimuthal distortion.

1. Cartesian formulation:

$$\mathbf{m} = \mathcal{P}(\mathbf{x_c}) = f \frac{1}{\alpha d_2 + (1-\alpha)(\xi d_1 + z)} \begin{bmatrix} x \\ y \end{bmatrix}$$
$$d_1 = \mathbf{x}_c = \sqrt{x^2 + y^2 + z^2}$$
$$d_2 = \sqrt{x^2 + y^2 + (\xi d_1 + z)^2}$$

2. Variable substitution:

$$\begin{aligned} d_1 &= \rho \\ d_2 &= \sqrt{x^2 + y^2 + z^2 + \xi^2 d_1^2 + 2\xi d_1 z} \\ &= \sqrt{\rho^2 + \xi^2\rho^2 + 2\xi\rho(\rho\cos\theta)} \\ &= \sqrt{\rho^2(1 + \xi^2 + 2\xi\cos\theta)} \\ &= \rho\sqrt{1 + \xi^2 + 2\xi\cos\theta} = \rho\, e_1 \end{aligned}$$

3. During the substitution in the main projection the variable ρ is simplified:

$$\begin{aligned} \mathbf{m} = \mathcal{P}(\mathbf{x_r}) &= f\,\frac{1}{\alpha\rho\, e_1 + (1-\alpha)(\xi\rho + \rho\cos\theta)} \begin{bmatrix} \rho\sin\theta\cos\phi \\ \rho\sin\theta\sin\phi \end{bmatrix} \\ &= f\,\frac{\rho\sin\theta}{\rho(\alpha e_1 + (1-\alpha)(\xi + \cos\theta))} \begin{bmatrix} \cos\phi \\ \sin\phi \end{bmatrix} \\ &= f\,\frac{\sin\theta}{\alpha e_1 + (1-\alpha)(\xi + \cos\theta)} \begin{bmatrix} \cos\phi \\ \sin\phi \end{bmatrix} \end{aligned}$$

4. Express the image point $\mathbf{m}$ in polar coordinates:

$$\mathbf{m} = \begin{bmatrix} m_\rho\cos(m_\phi) \\ m_\rho\sin(m_\phi) \end{bmatrix} = m_\rho \begin{bmatrix} \cos(m_\phi) \\ \sin(m_\phi) \end{bmatrix} = f\,\frac{\sin\theta}{\alpha e_1 + (1-\alpha)(\xi + \cos\theta)} \begin{bmatrix} \cos\phi \\ \sin\phi \end{bmatrix}$$

5. We can conclude that the radial projection function does not alter the value of ϕ:

$$\begin{aligned} \begin{bmatrix} \cos(m_\phi) \\ \sin(m_\phi) \end{bmatrix} &= \begin{bmatrix} \cos\phi \\ \sin\phi \end{bmatrix} \\ m_\phi &= \phi \end{aligned}$$

Moreover, the radius of the projected point on the image plane depends only on θ:

$$\begin{aligned} m_\rho = \mathcal{P}_r(\theta) &= f\,\frac{\sin\theta}{\alpha e_1 + (1-\alpha)(\xi + \cos\theta)} \\ e_1 &= \sqrt{1 + \xi^2 + 2\xi\cos\theta} \end{aligned} \tag{2}$$

9 Extended Ablation Study

9.1 Choice of Angular Resolution W_p

In this section, we describe the reason behind our choice for the angular resolution, or equivalently the polar width W_p, when transforming a Cartesian image

$\mathcal{I}$ in the Polar domain $\mathcal{I}_p$. First, we denote the number of pixels in the image $\mathcal{I}$ that have a radial distance equal to r as $p(r)$. A good approximation for this value is the length of the circumference with radius r:

$$p(r) \approx 2\pi r \tag{3}$$

Then, we denote as W_p the width chosen for the polar image. Intuitively, W_p is the number of pixels for each value of r in $\mathcal{I}_p$, and differently from the Cartesian case, it is constant:

$$p_p(r) = W_p \tag{4}$$

To help understanding the implication of the choice of W_p, we introduce the two measures that depend on the radius r. The first is the number of Cartesian pixels that project to the same polar pixel $\beta_{c2p}(r)$, and the second is its inverse, *i.e.*, the number of polar pixels that project onto the same Cartesian pixel $\beta_{p2c}(r)$. Their definition is:

$$\begin{aligned} \beta_{c2p}(r) &= \frac{p(r)}{p_p(r)} = \frac{2\pi r}{W_p} \\ \beta_{p2c}(r) &= \frac{p_p(r)}{p(r)} = \frac{W_p}{2\pi r} \end{aligned} \tag{5}$$

Intuitively, for low values of the radius r (close to the center of the image), the number of Cartesian pixels is low, and $\beta_{p2c}(r)$ reach its maximum. In fact, the first rows of a polar image are the ones with the highest deformation, and contain few information. Going further away from the center, thus increasing the value of r, the values of $\beta_{p2c}(r)$ decreases, while $\beta_{c2p}(r)$ increases. Note that, if for some values of r, we have $\beta_{c2p}(r) > 1$, we are loosing a certain degree of information contained in the Cartesian image. As an example, if $\beta_{c2p}(100) = 3$, at radius $r = 100$ each polar pixel contains the combined information of 3 Cartesian pixels, and it is worse for values of $r > 100$. Ideally, to avoid any loss of information, we could set $W_{\mathbf{polar}}$ such that $\beta_{c2p}(r_{max}) = 1$, where $r_{max} = \frac{S}{2}$. From Eq. 5, we derive:

$$W_p = 2\pi \frac{S}{2} = \pi S \tag{6}$$

Nevertheless, for all the other values of the radius $\beta_{p2c}(r) > 1$, thus, we are introducing a lot of redundant information. Furthermore, the increase in width introduce the need for a larger receptive field, other than an excessive deformation of the image features. We found out that a good compromise is achieved by setting $W_p = S$, which corresponds to $\beta_{c2p}\left(\frac{r_{max}}{\pi}\right) = 1$, and $\beta_{p2c}(r_{max}) = \pi$.

9.2 Choice of downsampling rate for the VaCR

Referring to Sec. 4.2 of the main paper, we explore the impact of different downsampling rates k on calibration accuracy.

The downsampling process is responsible for producing a smaller feature volume $\mathcal{F}_{\text{gap}}$, sized $C \times \frac{S}{2k} \times 1$, compared to the original image size S, which consequently reduces the number of regression paths in our network for the

k / σ	$\sigma = 0.0$	$\sigma = 0.5$	$\sigma = 1.0$	$\sigma = 2.0$
$k = 1$	0.0	2.058	2.359	2.834
$k = 2$	0.0	2.058	2.360	2.836
$k = 4$	0.0	2.059	2.361	2.839
$k = 8$	0.0	2.061	2.363	2.843
$k = 10$	0.0	2.098	2.390	2.879

Table 2: Comparison of reprojection error Re for varying downsampling factor k and noise standard deviation σ.

estimation of VACR entries. This increase in sparsity in the estimated VACR unavoidably provides less data for the robust fitting procedure for estimating camera parameters specific to the input camera model. A denser VACR provides more information for parameter estimation, which could theoretically improve calibration accuracy.

Our evaluation focuses on the calibration accuracy concerning the downsampling rate k, as defined in the main paper in Sec. 4.2, which, in turn, determines the number of entries in the estimated VACR. To assess this, we perform synthetic tests by sampling a set of ground truth camera parameters for the DSCM camera model and computing the corresponding VACR. We then add Gaussian noise to the VACR to emulate real-world inaccuracies in its estimation by our CNN. Following this, we use the robust fitting procedure of Sec. 4.4 to estimate camera parameters for DSCM given the noisy VACR. We then measure the resulting reprojection error Re under varying: **(i)** downsampling factor k, **(ii)** standard deviation σ of the Gaussian noise added to the VACR. For each (k, σ) pair, we perform 100 independent tests, each with a different set of camera parameters, and report the average Re in Tab. 2.

Results reveal that reprojection errors tend to escalate with higher noise standard deviation σ, yet are mitigated when lower downsampling factors k are applied, counteracting the noise in VACR estimates. Conversely, when using a high downsampling factor of $k \geq 8$, there is a marked degradation in performance, indicating that employing larger images or enhancing feature volumes via transposed convolutions in network architecture has limited benefit in improving overall calibration accuracy. As mentioned in the main paper, we set $k = 8$ for a good trade-off between calibration accuracy and efficiency.

9.3 Extrinsic Parameter Estimation

In our framework, we have integrated additional regression heads to predict camera tilt and roll, employing the architecture outlined by Wakai [8]. Tab. 3 presents the performance comparison of our method against the benchmarks set by López-Antequera [5] and Wakai [8]. As expected, our approach yields error rates on the regressed parameters that are comparable to those of Wakai, with variations leading to both superior and inferior outcomes. In contrast, the

	KITTI-360 [4]		Streetlearn [6]		SILDa [1]		Woodscape [9]	
Method	Tilt θ [deg]	Roll ψ [deg]	Tilt θ [deg]	Roll ψ [deg]	Tilt θ [deg]	Roll ψ [deg]	Tilt θ [deg]	Roll ψ [deg]
López-Antequera [5]	29.15	33.67	24.70	29.01	35.81	29.20	31.47	28.03
Wakai [8]	**5.20**	6.87	6.02	**4.98**	**5.78**	**4.60**	7.13	8.46
Ours	5.38	**6.45**	**5.87**	5.39	5.80	4.78	**7.01**	**8.39**

Table 3: Extrinsic Parameter Calibration Evaluation. Comparison of the absolute extrinsic parameter errors, in degrees. For all metrics, lower is better.

method by López-Antequera [5] consistently underperforms across all evaluated metrics.

10 Metrics

In this section, we define the reprojection error Re, as utilized in Sec. 5 of the main document. The reprojection error is expressed as follows:

$$Re = \sum_{\mathbf{u} \in \mathcal{I}} ||\mathbf{u} - P_{\mathcal{M}}(P_{\mathcal{M}'}^{-1}(\mathbf{u}, \mathbf{i}_{\mathcal{M}',\text{gt}}), \mathbf{i}_{\mathcal{M}})||^2 , \tag{7}$$

where **(i)** $\mathcal{M}'$ represents the camera model used to generate the image; **(ii)** $\mathbf{i}_{\mathcal{M}',\text{gt}}$ denotes the ground truth parameters for the camera model $\mathcal{M}'$; **(iii)** $P_{\mathcal{M}}$ is the projection function of the input camera model; and **(iv)** $\mathbf{i}_{\mathcal{M}}$ are the parameters estimated according to the input camera model. The function $P_{\mathcal{M}}$ projects a 3D point, specifically a bearing vector, onto the 2D image plane, given a set of model parameters for $\mathcal{M}$. Conversely, $P_{\mathcal{M}}^{-1}$ back-projects a 2D image point into its corresponding 3D bearing vector, considering the parameters of model $\mathcal{M}$.

References

1. Balntas, V.: SILDa: A Multi-Task Dataset for Evaluating Visual Localization. Medium (Apr 2019), https://medium.com/scape-technologies/silda-a-multi-task-dataset-for-evaluating-visual-localization-7fc6c2c56c74 10
2. Geyer, C., Daniilidis, K.: A unifying theory for central panoramic systems and practical implications. In: Computer Vision—ECCV 2000: 6th European Conference on Computer Vision Dublin, Ireland, June 26–July 1, 2000 Proceedings, Part II 6. pp. 445–461. Springer (2000) 2
3. Khomutenko, B., Garcia, G., Martinet, P.: An enhanced unified camera model. IEEE Robotics and Automation Letters **1**(1), 137–144 (2015) 2, 3
4. Liao, Y., Xie, J., Geiger, A.: Kitti-360: A novel dataset and benchmarks for urban scene understanding in 2d and 3d. IEEE Transactions on Pattern Analysis and Machine Intelligence **45**(3), 3292–3310 (2022) 10
5. Lopez, M., Mari, R., Gargallo, P., Kuang, Y., Gonzalez-Jimenez, J., Haro, G.: Deep single image camera calibration with radial distortion. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11817–11825 (2019) 9, 10
6. Mirowski, P., Banki-Horvath, A., Anderson, K., Teplyashin, D., Hermann, K.M., Malinowski, M., Grimes, M.K., Simonyan, K., Kavukcuoglu, K., Zisserman, A., et al.: The streetlearn environment and dataset. arXiv preprint arXiv:1903.01292 (2019) 10
7. Usenko, V., Demmel, N., Cremers, D.: The double sphere camera model. In: 2018 International Conference on 3D Vision (3DV). pp. 552–560. IEEE (2018) 2, 4, 6
8. Wakai, N., Sato, S., Ishii, Y., Yamashita, T.: Rethinking generic camera models for deep single image camera calibration to recover rotation and fisheye distortion. In: European Conference on Computer Vision. pp. 679–698. Springer (2022) 5, 9, 10
9. Yogamani, S., Hughes, C., Horgan, J., Sistu, G., Varley, P., O'Dea, D., Uricár, M., Milz, S., Simon, M., Amende, K., et al.: Woodscape: A multi-task, multi-camera fisheye dataset for autonomous driving. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 9308–9318 (2019) 10