360Loc: A Dataset and Benchmark for Omnidirectional Visual Localization with Various Cameras
- Supplementary -

Huajian Huang[1*] Changkun Liu[1*] Yipeng Zhu[1] Hui Cheng[2] Tristan Braud[1] Sai-Kit Yeung[1]
[1]The Hong Kong University of Science and Technology [2]Sun Yat-sen University
* equal contribution
{hhuangbg, cliudg, yzhudg}@connect.ust.hk, chengh9@mail.sysu.edu.cn, {braudt, saikit}@ust.hk

1. Appendix

1.1. Pinhole SfM and 360° Mapping

Owing to the incompatibility of current SfM methods, 360° images are usually cropped into pinhole images for reconstruction. However, SfM based on pinhole images is vulnerable in particular for scenes with symmetric or repetitive features. To clarify the weakness of Pinhole SfM, we obtain 632×2 random cropped images from c_4 (see Table 2 in the main paper) and 632×5 cube map images from 632 360° images. The bottom faces of the cube map are discarded because of capturing the ground and platform itself. After that we use HLoc [9] (NetVLAD [1] top$k$=20, SuperPoint [3] + SuperGlue [10]) and COLMAP [12] for constructing sparse point cloud models. The reconstruction results are demonstrated in Figure 1. Compared to ground truth Figure 1c, using cube map pinhole images Figure 1b encounters disaster in trajectory recovery, while the pose estimations in Figure 1a have large drift. This example demonstrates that making use of omnidirectional FoV of 360° images is important for accurate and efficient mapping.

1.2. The Graph-based SLAM for Reference Map Collection

In the process of reference map collection, we utilized a graph-based SLAM technique [6] with a loop detection algorithm. The core of the graph-based method is to minimize the errors between the parameters and the constraints through optimization. We choose the ground plane feature coefficients $\boldsymbol{\zeta}_0 = [n_x, n_y, n_z, \delta]^T = [0, 0, 0, 1]$ as a constraint to optimize the lidar pose $\boldsymbol{\eta}_t = [\mathbf{R}_t|\mathbf{t}_t] \in SO(3) \times \mathbb{R}^3$ at frame t, ensuring that the ground plane detected in each observation becomes consistent. $\boldsymbol{n} = \left[n_x,\ n_y,\ n_z\right]$is the normal vector of the constrained plane, which points to the Z direction of the $\mathbf{O_w}$-$\mathbf{XYZ}$, and δ is the length of the intercept. So, the transformed ground plane constraints in $\mathbf{O_l}$-$\mathbf{XYZ}$ is:

$$\left[n'_x,\ n'_y,\ n'_z\right] = \mathbf{R}_t \cdot \left[n_x,\ n_y,\ n_z\right], \tag{1}$$

$$\delta' = \delta - \mathbf{t}_t \cdot \left[n'_x,\ n'_y,\ n'_z\right]^T, \tag{2}$$

where, $\boldsymbol{\zeta}'_0 = \left[n'_x,\ n'_y,\ n'_z,\ \delta'\right]$. Thus, the error ϵ_t between a lidar pose and the ground plane feature is defined as follows:

$$\boldsymbol{\tau}(\boldsymbol{\zeta}) = \left[\arctan\left(\frac{n_y}{n_x}\right),\ \arctan\left(\frac{n_z}{|\mathbf{n}|}\right),\ \delta\right], \tag{3}$$

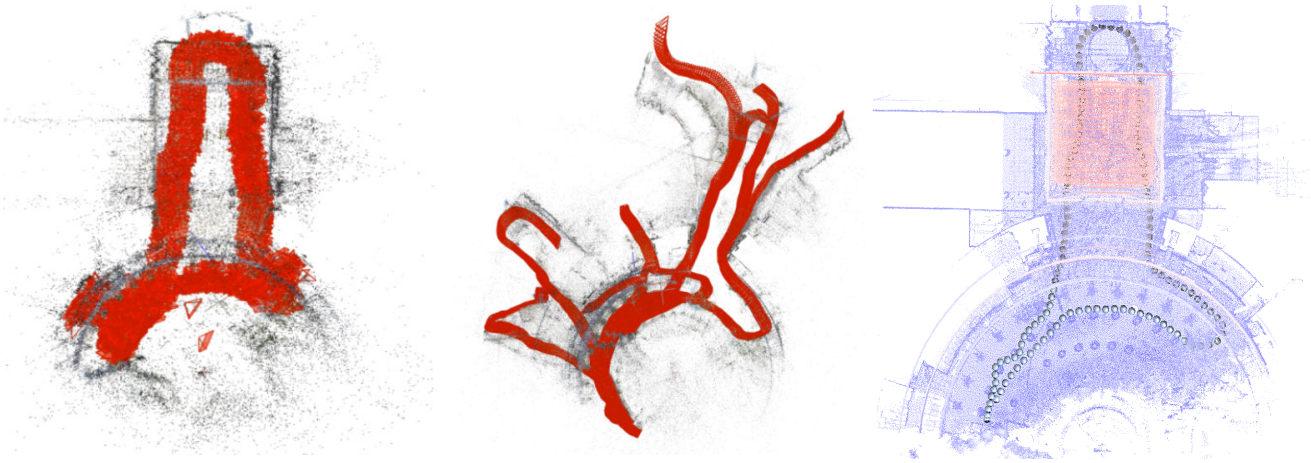
$$\boldsymbol{\epsilon}_t = \boldsymbol{\tau}\left(\boldsymbol{\zeta}'_0\right) - \boldsymbol{\tau}\left(\boldsymbol{\zeta}_t\right), \tag{4}$$

where $\boldsymbol{\zeta}_t$ is the detected ground plane at frame t. With this constraint, we can now get the optimized frame-by-frame values of $\boldsymbol{\eta}_t$ in real time while collecting.

1.3. Details about Dataset Collection

Firstly, it should be noted that during the collection of the reference map, due to the diverse data types and large data volume, we employed a multi-threaded approach in the programming. However, when saving the data, we recorded the timestamps corresponding to their collection triggers and performed offline synchronization. Here, we provide detailed explanations of several methods used for data collection, as introduced in Section 3 of the main paper.

For the image collection process, we set the camera's frame rate to 25 frames per second (fps) and traversed the entire current scene using different routes while manually holding the collection device. This process extended over multiple days and encompassed diverse time periods. Thus, each scene in the dataset includes images in daytime and nighttime. Besides, the images have appearance changes due to the differences in dates as shown in Figure 2. The data is collected using a handheld device and three participants who recorded sequences were asked to freely walk through each scene with different routines as shown in Figure 5. Collected images feature diverse capture angles, and



(a) Random pinhole mapping from c_4. (b) Cube map pinhole mapping. (c) 360° mapping.

Figure 1. Subfigure (a)(b) show that the pinhole cameras are incapable of mapping in ambiguous scenes. Ground truth traces are white spheres in Subfigure (c).

some may exhibit motion blur. These characteristics make the 360Loc dataset rich, comprehensive, and challenging in nature.

1.4. Image Retrieval Evaluation

We evaluate global descriptors computed by NetVLAD [1], CosPlace [2], OpenIBL [4] and AP-GeM [5]. The query image is deemed correctly localized if at least one of the top k retrieved database images is within $d = 5m$ from the ground truth position of the query for Concourse and $d = 10m$ for the other three scenes. The average IR results over 4 scenes are shown in Table 4 of the main paper. We provide the IR results of each scene in this supplementary material in Table 1, Table 3, Table 2, and Table 4. The trend is similar to the Table 4 shown in the main paper.

Among all global feature descriptor methods, the 360° query exhibits the best precision and recall in most cases, while the pinhole query performs the worst. The performance of recall and precision is correlated with the FoV, where a larger FoV results in higher performance. The remap method (VC1) provides limited improvement for pinhole queries but yields higher improvement for fisheye1, fisheye2, and fisheye3 queries. The reason is that the FoV of pinhole cameras is only 85°. Consequently, VC1 results in significant black borders when converting to a 360° image due to the limited coverage of the pinhole camera as shown in Figure 3.

The rectify method (VC2) significantly improves pinhole, fisheye1, fisheye2, and fisheye3 queries by eliminating the domain gap in IR. However, the pinhole, fisheye1, and fisheye2 queries' recall and precision are still much lower than those of the 360° query. Only the query from fisheye3 (widest FoV) approaches the performance of 360° query. The domain gap mainly affects the precision and recall of fisheye3. Both remap (VC1) and crop (VC2) significantly improve IR performance for fisheye3. On the other hand, pinhole queries are more prone to being mistaken as erroneous locations with similar structures due to their narrower FoV even there is no cross-device domain gap during IR by applying VC2 as shown in Figure 3.

1.5. Local Feature Matching Evaluation

The average local feature matching results over 4 scenes are shown in Table 5 of the main paper. We provide the results of each scene in this supplementary material in Table 5, Table 7, Table 6, and Table 8 with extra SIFT [8] + Nearest Neighbor (NN) setting. The trend is similar to the Table 5 shown in the main paper.

We present two groups of examples of 2D-2D matching. In each group in Figure 4. Query frames from c_1, c_2, c_3, c_4 are cropped from the same 360° query, and all corresponding retrieved images overlap with query frames. However, the pinhole query from c_4 suffers from erroneous matches due to interference from symmetrical and repetitive structures, while fisheye query frames from c_1, c_2, c_3, and 360° query frames have better matches as the wide FoV allows to capture more unique features. This finding suggests that cross-device visual localization on a 360-camera database in challenging ambiguous scenarios requires more robust local matching approaches.

1.6. Absolute Pose Regression

The average local median translation and rotation errors over 4 scenes are shown in Figure 5 of the main paper. We provide the results of each scene in this supplementary material in Table 9, Table 10, Table 11 and Table 12. The trend is similar to the Figure 5 shown in the main paper.

Figure 2. Our datasets feature with appearance changes.

When we introduce images from virtual cameras for data augmentation, PN^{vc2} and MS-T^{vc2} exhibit significantly reduced translation and rotation errors across all cameras in most cases, particularly during daytime.

	Query	R@1	R@5	P@5	R@10	P@10	R@20	P@20
NetVLAD [1]	pinhole	0.19	0.43	0.20	0.56	0.19	0.71	0.17
	+VC1	0.28	0.54	0.28	0.68	0.27	0.79	0.24
	+VC2	0.55	0.74	0.52	0.82	0.49	0.88	0.41
	fisheye1	0.42	0.68	0.40	0.79	0.38	0.89	0.34
	+VC1	0.53	0.77	0.50	0.86	0.47	0.92	0.41
	+VC2	0.79	0.95	0.66	0.97	0.56	0.98	0.46
	fisheye2	0.46	0.73	0.44	0.83	0.42	0.91	0.37
	+VC1	0.54	0.78	0.52	0.87	0.49	0.93	0.43
	+VC2	0.81	0.95	0.68	0.97	0.58	0.98	0.48
	fisheye3	0.59	0.82	0.56	0.89	0.52	0.95	0.46
	+VC1	0.65	0.86	0.62	0.92	0.58	0.96	0.50
	+VC2	**0.82**	**0.96**	**0.70**	**0.98**	**0.61**	**0.99**	0.50
	360	0.68	0.71	0.65	0.71	**0.61**	0.72	**0.52**
CosPlace [2]	pinhole	0.16	0.29	0.15	0.38	0.15	0.50	0.15
	+VC1	0.21	0.35	0.21	0.45	0.21	0.58	0.20
	+VC2	0.31	0.44	0.31	0.51	0.30	0.60	0.27
	fisheye1	0.32	0.52	0.32	0.63	0.31	0.74	0.29
	+VC1	0.39	0.54	0.38	0.64	0.36	0.76	0.33
	+VC2	0.62	0.86	0.47	0.94	0.39	0.97	0.31
	fisheye2	0.35	0.56	0.36	0.67	0.35	0.77	0.32
	+VC1	0.39	0.54	0.38	0.64	0.37	0.76	0.34
	+VC2	0.63	0.88	0.49	0.94	0.41	0.97	0.33
	fisheye3	0.45	0.67	0.45	0.77	0.44	0.86	0.41
	+VC1	0.52	0.67	0.50	0.76	0.49	0.86	0.44
	+VC2	0.70	0.91	0.56	0.95	0.48	0.97	0.39
	360	**0.94**	**0.97**	**0.93**	**0.98**	**0.90**	**0.98**	**0.78**
OpenIBL [4]	pinhole	0.15	0.32	0.15	0.44	0.14	0.60	0.14
	+VC1	0.21	0.42	0.21	0.55	0.20	0.69	0.19
	+VC2	0.55	0.72	0.53	0.78	0.49	0.86	0.42
	fisheye1	0.37	0.59	0.35	0.72	0.34	0.83	0.31
	+VC1	0.55	0.75	0.52	0.84	0.48	0.89	0.41
	+VC2	0.75	0.92	0.62	0.96	0.53	0.98	0.43
	fisheye2	0.41	0.66	0.39	0.77	0.37	0.86	0.34
	+VC1	0.58	0.78	0.55	0.85	0.51	0.90	0.44
	+VC2	0.77	0.93	0.63	0.97	0.55	0.99	0.45
	fisheye3	0.54	0.78	0.52	0.87	0.49	0.93	0.44
	+VC1	0.69	0.87	0.67	0.92	0.62	0.96	0.53
	+VC2	0.80	0.94	0.66	0.97	0.57	**0.99**	0.47
	360	**0.91**	**0.96**	**0.90**	**0.97**	**0.84**	0.98	**0.72**
AP-GeM [5]	pinhole	0.17	0.36	0.18	0.47	0.18	0.61	0.17
	+VC1	0.26	0.46	0.26	0.58	0.25	0.72	0.23
	+VC2	0.50	0.71	0.48	0.79	0.45	0.88	0.39
	fisheye1	0.37	0.60	0.35	0.71	0.33	0.80	0.30
	+VC1	0.46	0.68	0.45	0.78	0.42	0.86	0.38
	+VC2	0.66	0.90	0.54	0.95	0.47	0.98	0.38
	fisheye2	0.40	0.64	0.38	0.73	0.36	0.82	0.33
	+VC1	0.49	0.71	0.48	0.80	0.45	0.88	0.40
	+VC2	0.68	0.91	0.56	0.96	0.49	0.98	0.40
	fisheye3	0.47	0.69	0.45	0.77	0.42	0.85	0.37
	+VC1	0.59	0.78	0.56	0.85	0.53	0.92	0.46
	+VC2	0.69	0.92	0.58	0.96	0.52	0.97	0.43
	360	**0.84**	**0.94**	**0.80**	**0.97**	**0.75**	**0.98**	**0.64**

Table 1. *Concourse*. Image retrieval results based on 360° reference database for the top k retrieved images, k = 1, 5, 10, 20. # indicates the highest value of R@k and P@k for each device w and w/o virtual cameras (VC1, VC2). Best results for all devices of R@k and P@k are in bold with #.

References

[1] Relja Arandjelovic, Petr Gronat, Akihiko Torii, Tomas Pajdla, and Josef Sivic. Netvlad: Cnn architecture for weakly supervised place recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5297–5307, 2016.

[2] Gabriele Berton, Carlo Masone, and Barbara Caputo. Rethinking visual geo-localization for large-scale applications. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4878–4888, 2022.

[3] Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. Superpoint: Self-supervised interest point detection and description. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 224–236, 2018.

[4] Yixiao Ge, Haibo Wang, Feng Zhu, Rui Zhao, and Hongsheng Li. Self-supervising fine-grained region similarities for large-scale image localization. In *European Conference on Computer Vision*, 2020.

[5] A. Gordo, J. Almazan, J. Revaud, and D. Larlus. End-to-end learning of deep visual representations for image retrieval. *IJCV*, 2017.

[6] Kenji Koide, Jun Miura, and Emanuele Menegatti. A portable three-dimensional lidar-based system for long-term and wide-area people behavior measurement. *International Journal of Advanced Robotic Systems*, 16(2): 1729881419841532, 2019.

[7] Philipp Lindenberger, Paul-Edouard Sarlin, and Marc Pollefeys. LightGlue: Local Feature Matching at Light Speed. In *ICCV*, 2023.

[8] David G Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60:91–110, 2004.

[9] Paul-Edouard Sarlin, Cesar Cadena, Roland Siegwart, and Marcin Dymczyk. From coarse to fine: Robust hierarchical localization at large scale. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12716–12725, 2019.

[10] Paul-Edouard Sarlin, Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. Superglue: Learning feature

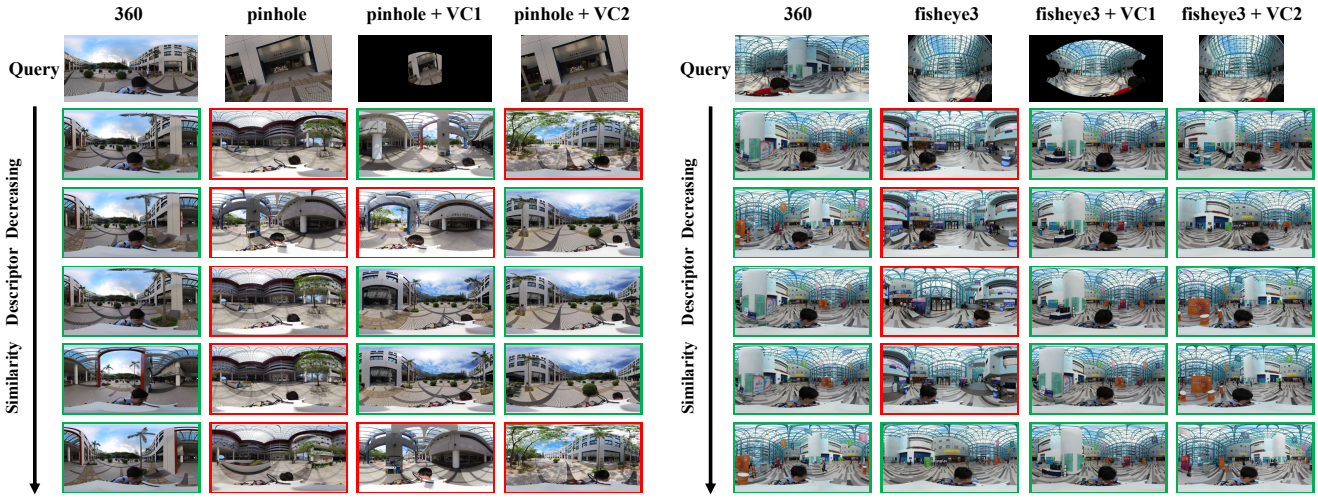


Figure 3. Example queries and top-5 retrieved images. The green boxes represent images that have overlap regions with the query image, and images with red boxes are wrong images.

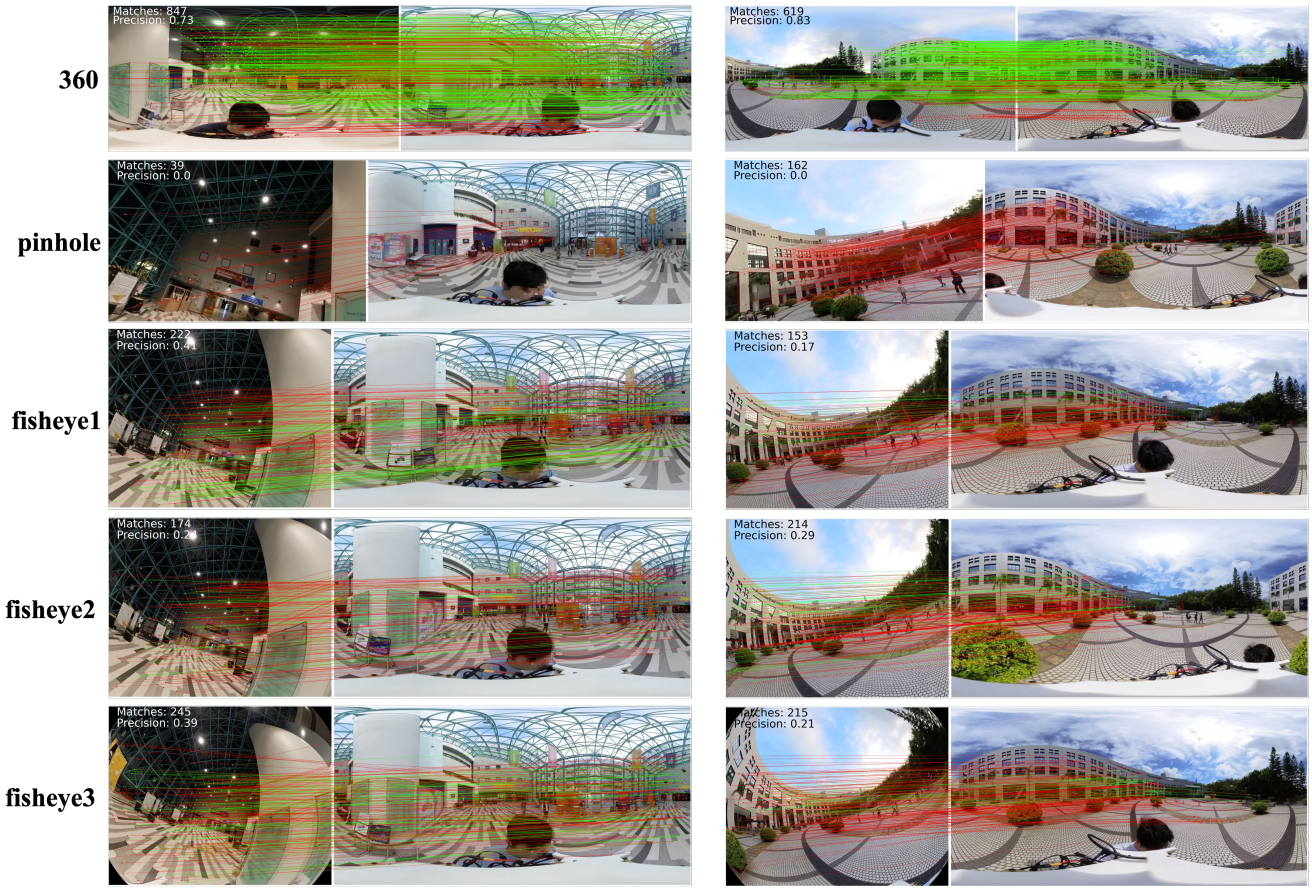
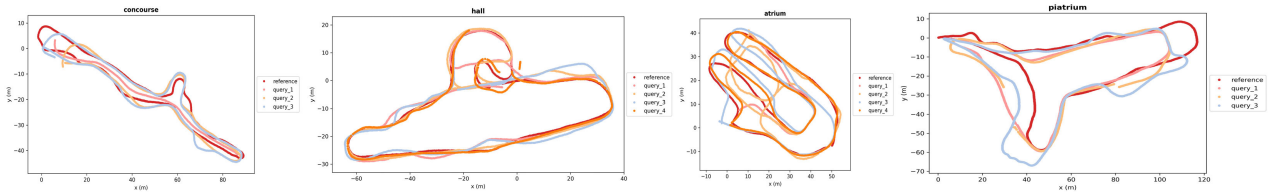


Figure 4. Two groups of local 2D-2D matching pairs (Superpoint + LightGlue [7]). For each pair, left side is the query frame and right side is the top-1 retrieved reference 360° image. In this figure, all retrieved images are correct. The reprojection error threshold is 5px (green means correct matches, red means wrong matches).



(a) Concourse (b) Hall (c) Atrium (d) Piatrium

Figure 5. Trajectories of reference and query.

	Query	R@1	R@5	P@5	R@10	P@10	R@20	P@20
NetVLAD [1]	pinhole	0.30	0.55	0.29	0.69	0.28	0.81	0.25
	+VC1	0.24	0.43	0.24	0.53	0.23	0.64	0.21
	+VC2	0.55	0.69	0.53	0.76	0.52	0.84	0.48
	fisheye1	0.56	0.79	0.54	0.87	0.51	0.93	0.46
	+VC1	0.57	0.75	0.55	0.82	0.53	0.89	0.48
	+VC2	0.76	0.93	0.70	0.98	0.65	0.99	0.57
	fisheye2	0.60	0.82	0.57	0.89	0.55	0.94	0.50
	+VC1	0.61	0.78	0.59	0.85	0.57	0.90	0.52
	+VC2	0.77	0.94	0.72	0.98	0.67	0.99	0.59
	fisheye3	0.69	0.87	0.66	0.92	0.64	0.96	0.58
	+VC1	0.70	0.85	0.69	0.91	0.66	0.95	0.61
	+VC2	0.81	**0.96**	0.75	**0.99**	0.70	**1.00**	0.63
	360	**0.90**	**0.96**	**0.89**	0.98	**0.86**	**1.00**	**0.81**
CosPlace [2]	pinhole	0.19	0.32	0.19	0.42	0.19	0.55	0.19
	+VC1	0.25	0.37	0.24	0.46	0.24	0.56	0.23
	+VC2	0.35	0.43	0.34	0.49	0.33	0.57	0.31
	fisheye1	0.39	0.54	0.38	0.65	0.38	0.76	0.36
	+VC1	0.38	0.52	0.37	0.60	0.36	0.68	0.33
	+VC2	0.66	0.88	0.56	0.94	0.49	0.98	0.41
	fisheye2	0.41	0.59	0.42	0.68	0.42	0.79	0.40
	+VC1	0.40	0.54	0.39	0.61	0.37	0.70	0.34
	+VC2	0.67	0.89	0.59	0.95	0.52	0.98	0.43
	fisheye3	0.54	0.70	0.54	0.78	0.53	0.87	0.50
	+VC1	0.52	0.67	0.53	0.74	0.51	0.81	0.48
	+VC2	0.73	0.91	0.64	0.96	0.58	0.99	0.49
	360	**0.96**	**0.98**	**0.96**	**0.99**	**0.94**	**1.00**	**0.89**
OpenIBL [4]	pinhole	0.23	0.45	0.23	0.58	0.23	0.71	0.21
	+VC1	0.24	0.40	0.22	0.50	0.21	0.61	0.20
	+VC2	0.53	0.69	0.52	0.76	0.50	0.84	0.45
	fisheye1	0.48	0.71	0.47	0.81	0.45	0.89	0.41
	+VC1	0.55	0.73	0.53	0.81	0.50	0.88	0.45
	+VC2	0.77	0.94	0.69	0.98	0.62	0.99	0.52
	fisheye2	0.53	0.75	0.52	0.84	0.49	0.91	0.44
	+VC1	0.59	0.76	0.57	0.84	0.54	0.90	0.48
	+VC2	0.79	0.95	0.71	0.98	0.64	**1.00**	0.55
	fisheye3	0.65	0.84	0.63	0.91	0.60	0.95	0.54
	+VC1	0.72	0.86	0.70	0.92	0.67	0.96	0.60
	+VC2	0.82	0.96	0.74	**0.99**	0.68	**1.00**	0.59
	360	**0.95**	**0.98**	**0.94**	**0.99**	**0.91**	**1.00**	**0.83**
AP-GeM [5]	pinhole	0.31	0.50	0.30	0.61	0.29	0.72	0.27
	+VC1	0.34	0.54	0.33	0.64	0.32	0.74	0.29
	+VC2	0.63	0.79	0.62	0.87	0.59	0.93	0.54
	fisheye1	0.49	0.69	0.47	0.78	0.45	0.86	0.41
	+VC1	0.52	0.70	0.50	0.78	0.47	0.85	0.43
	+VC2	0.75	0.94	0.69	**0.98**	0.63	**0.99**	0.55
	fisheye2	0.52	0.71	0.50	0.80	0.48	0.87	0.44
	+VC1	0.55	0.73	0.53	0.81	0.50	0.87	0.46
	+VC2	0.77	0.94	0.70	**0.98**	0.64	**0.99**	0.56
	fisheye3	0.58	0.78	0.56	0.86	0.54	0.92	0.49
	+VC1	0.66	0.83	0.64	0.89	0.61	0.93	0.56
	+VC2	0.78	0.95	0.72	**0.98**	0.67	**0.99**	0.58
	360	**0.90**	**0.96**	**0.87**	**0.98**	**0.83**	**0.99**	**0.76**

Table 2. *Hall*. Image retrieval results based on 360° reference database for the top k retrieved images, k = 1, 5, 10, 20. # indicates the highest value of R@k and P@k for each device w and w/o virtual cameras (VC1, VC2). Best results for all devices of R@k and P@k are in bold with # .

matching with graph neural networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4938–4947, 2020.

[11] Torsten Sattler, Will Maddern, Carl Toft, Akihiko Torii, Lars Hammarstrand, Erik Stenborg, Daniel Safari, Masatoshi Okutomi, Marc Pollefeys, Josef Sivic, et al. Benchmarking 6dof outdoor visual localization in changing conditions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8601–8610, 2018.

[12] Johannes L Schonberger and Jan-Michael Frahm. Structure-from-motion revisited. In *IEEE conference on computer vision and pattern recognition*, pages 4104–4113, 2016.

	Query	R@1	R@5	P@5	R@10	P@10	R@20	P@20
NetVLAD [1]	pinhole	0.22	0.44	0.21	0.57	0.21	0.69	0.20
	+VC1	0.24	0.44	0.24	0.57	0.23	0.72	0.22
	+VC2	0.48	0.67	0.47	0.76	0.46	0.85	0.42
	fisheye1	0.39	0.63	0.37	0.74	0.35	0.85	0.33
	+VC1	0.48	0.71	0.46	0.82	0.45	0.90	0.41
	+VC2	0.75	0.94	0.63	0.98	0.55	0.99	0.47
	fisheye2	0.42	0.66	0.40	0.78	0.39	0.88	0.36
	+VC1	0.51	0.74	0.50	0.84	0.48	0.91	0.43
	+VC2	0.75	0.95	0.65	0.98	0.58	**1.00**	0.49
	fisheye3	0.52	0.77	0.51	0.86	0.48	0.93	0.45
	+VC1	0.60	0.81	0.59	0.89	0.56	0.95	0.51
	+VC2	0.79	0.96	0.68	**0.99**	0.60	**1.00**	0.52
	360	**0.85**	**0.95**	**0.82**	0.98	**0.79**	**1.00**	**0.70**
CosPlace [2]	pinhole	0.15	0.23	0.15	0.29	0.15	0.37	0.14
	+VC1	0.21	0.33	0.21	0.42	0.21	0.52	0.21
	+VC2	0.34	0.43	0.34	0.49	0.33	0.59	0.32
	fisheye1	0.22	0.32	0.22	0.39	0.22	0.49	0.21
	+VC1	0.36	0.48	0.36	0.55	0.36	0.63	0.34
	+VC2	0.65	0.88	0.53	0.94	0.45	0.98	0.36
	fisheye2	0.23	0.35	0.24	0.41	0.23	0.52	0.22
	+VC1	0.38	0.50	0.37	0.57	0.37	0.64	0.36
	+VC2	0.67	0.89	0.55	0.95	0.46	0.98	0.37
	fisheye3	0.30	0.44	0.31	0.52	0.30	0.63	0.30
	+VC1	0.49	0.60	0.48	0.66	0.47	0.73	0.45
	+VC2	0.72	0.91	0.59	**0.96**	0.50	**0.99**	0.41
	360	**0.92**	**0.94**	**0.91**	**0.96**	**0.90**	0.97	**0.85**
OpenIBL [4]	pinhole	0.17	0.35	0.17	0.46	0.17	0.60	0.16
	+VC1	0.23	0.42	0.23	0.54	0.23	0.68	0.22
	+VC2	0.51	0.69	0.50	0.78	0.48	0.86	0.44
	fisheye1	0.33	0.53	0.31	0.64	0.30	0.76	0.28
	+VC1	0.52	0.72	0.51	0.82	0.49	0.90	0.45
	+VC2	0.77	0.95	0.64	0.98	0.55	**1.00**	0.46
	fisheye2	0.35	0.57	0.34	0.69	0.33	0.80	0.31
	+VC1	0.56	0.74	0.54	0.83	0.52	0.90	0.48
	+VC2	0.80	0.96	0.66	0.98	0.57	**1.00**	0.47
	fisheye3	0.48	0.70	0.46	0.81	0.45	0.91	0.41
	+VC1	0.65	0.81	0.64	0.89	0.61	0.94	0.56
	+VC2	0.83	**0.97**	0.71	**0.99**	0.61	**1.00**	0.51
	360	**0.90**	**0.97**	**0.89**	0.98	**0.85**	**1.00**	**0.75**
AP-GeM [5]	pinhole	0.15	0.29	0.15	0.38	0.15	0.49	0.15
	+VC1	0.19	0.34	0.19	0.44	0.19	0.56	0.19
	+VC2	0.46	0.64	0.45	0.74	0.44	0.83	0.40
	fisheye1	0.22	0.41	0.22	0.52	0.22	0.63	0.22
	+VC1	0.36	0.57	0.36	0.67	0.35	0.79	0.34
	+VC2	0.63	0.89	0.55	0.95	0.49	0.98	0.43
	fisheye2	0.25	0.44	0.24	0.55	0.24	0.67	0.24
	+VC1	0.40	0.60	0.39	0.70	0.38	0.81	0.36
	+VC2	0.64	0.90	0.56	0.96	0.50	0.99	0.44
	fisheye3	0.33	0.55	0.33	0.66	0.32	0.78	0.31
	+VC1	0.48	0.68	0.48	0.78	0.46	0.88	0.44
	+VC2	0.65	**0.92**	0.57	**0.97**	0.51	**0.99**	0.45
	360	**0.73**	0.88	**0.72**	0.94	**0.68**	0.98	**0.62**

Table 3. *Atrium*. Image retrieval results based on 360° reference database for the top k retrieved images, k = 1, 5, 10, 20. # indicates the highest value of R@k and P@k for each device w and w/o virtual cameras (VC1, VC2). Best results for all devices of R@k and P@k are in bold with # .

	Query	R@1	R@5	P@5	R@10	P@10	R@20	P@20
NetVLAD [1]	pinhole	0.20	0.39	0.19	0.51	0.18	0.64	0.17
	+VC1	0.21	0.38	0.20	0.50	0.20	0.61	0.19
	+VC2	0.43	0.59	0.42	0.67	0.41	0.76	0.38
	fisheye1	0.32	0.56	0.31	0.67	0.30	0.78	0.28
	+VC1	0.45	0.64	0.44	0.72	0.42	0.80	0.39
	+VC2	0.62	0.82	0.54	0.87	0.49	0.90	0.42
	fisheye2	0.34	0.59	0.34	0.71	0.33	0.81	0.31
	+VC1	0.49	0.66	0.47	0.74	0.44	0.81	0.41
	+VC2	0.64	0.84	0.56	0.88	0.50	0.91	0.43
	fisheye3	0.47	0.70	0.45	0.78	0.43	0.85	0.40
	+VC1	0.57	0.72	0.55	0.78	0.52	0.83	0.47
	+VC2	0.67	**0.85**	0.58	**0.89**	0.52	**0.92**	0.46
	360	**0.71**	0.82	**0.70**	0.86	**0.67**	0.89	**0.60**
CosPlace [2]	pinhole	0.12	0.19	0.11	0.25	0.11	0.34	0.11
	+VC1	0.17	0.27	0.18	0.32	0.18	0.39	0.17
	+VC2	0.29	0.36	0.29	0.42	0.28	0.51	0.28
	fisheye1	0.20	0.32	0.21	0.40	0.20	0.49	0.20
	+VC1	0.30	0.39	0.30	0.43	0.29	0.49	0.28
	+VC2	0.58	0.80	0.47	0.87	0.39	0.91	0.32
	fisheye2	0.23	0.35	0.23	0.42	0.23	0.51	0.22
	+VC1	0.32	0.40	0.32	0.45	0.31	0.51	0.29
	+VC2	0.60	0.81	0.48	0.87	0.41	0.91	0.34
	fisheye3	0.32	0.45	0.32	0.53	0.31	0.61	0.30
	+VC1	0.41	0.50	0.41	0.55	0.41	0.60	0.38
	+VC2	0.62	0.82	0.51	0.88	0.44	0.91	0.37
	360	**0.85**	**0.89**	**0.84**	**0.90**	**0.83**	**0.93**	**0.78**
OpenIBL [4]	pinhole	0.17	0.33	0.17	0.43	0.16	0.55	0.16
	+VC1	0.17	0.31	0.17	0.40	0.17	0.52	0.16
	+VC2	0.44	0.59	0.43	0.66	0.42	0.75	0.39
	fisheye1	0.30	0.50	0.30	0.59	0.28	0.69	0.26
	+VC1	0.47	0.61	0.45	0.70	0.44	0.78	0.40
	+VC2	0.66	0.83	0.55	0.88	0.48	0.91	0.40
	fisheye2	0.34	0.52	0.33	0.62	0.31	0.71	0.29
	+VC1	0.50	0.65	0.49	0.72	0.47	0.79	0.43
	+VC2	0.67	0.84	0.57	0.89	0.50	**0.92**	0.42
	fisheye3	0.46	0.65	0.44	0.73	0.42	0.80	0.39
	+VC1	0.61	0.74	0.59	0.79	0.56	0.84	0.51
	+VC2	0.70	**0.86**	0.59	**0.90**	0.52	**0.92**	0.44
	360	**0.80**	0.85	**0.77**	0.87	**0.74**	0.89	**0.65**
AP-GeM [5]	pinhole	0.18	0.34	0.18	0.42	0.17	0.52	0.17
	+VC1	0.21	0.36	0.21	0.45	0.21	0.56	0.20
	+VC2	0.42	0.59	0.41	0.67	0.40	0.76	0.37
	fisheye1	0.33	0.52	0.32	0.61	0.31	0.71	0.29
	+VC1	0.38	0.54	0.37	0.63	0.36	0.72	0.34
	+VC2	0.57	0.80	0.49	0.86	0.45	**0.90**	0.39
	fisheye2	0.35	0.56	0.34	0.65	0.33	0.73	0.32
	+VC1	0.40	0.57	0.39	0.65	0.38	0.73	0.36
	+VC2	0.58	**0.81**	0.51	**0.87**	0.46	**0.90**	0.40
	fisheye3	0.42	0.62	0.39	0.70	0.38	0.77	0.36
	+VC1	0.47	0.62	0.45	0.70	0.43	0.78	0.41
	+VC2	0.58	**0.81**	0.51	**0.87**	0.46	**0.90**	0.41
	360	**0.71**	0.80	**0.68**	0.85	**0.64**	0.88	**0.58**

Table 4. *Piatrium*. Image retrieval results based on 360° reference database for the top k retrieved images, k = 1, 5, 10, 20. # indicates the highest value of R@k and P@k for each device w and w/o virtual cameras (VC1, VC2). Best results for all devices of R@k and P@k are in bold with # .

Scene		Day												
Concourse		Query camera												
Retrieval	local Matching	pinhole	+ VC1	+VC2	fisheye1	+ VC1	+VC2	fisheye2	+ VC1	+VC2	fisheye3	+ VC1	+VC2	360
NetVLAD	SIFT + NN	4.8/9.0/13.4	9.0/12.4/16.3	15.9/20.2/24.8	2.0/4.5/10.5	4.7/ 9.1 /15.3	3.9 /9.0/ 17.8	1.9/4.7/10.2	4.2/8.7/15.9	5.3/10.9/19.8	3.8/8.3/15.6	6.3/11.3/18.3	6.7/12.6/21.7	**19.2/29.0/48.9**
	DISK + LG	7.7/14.5/27.6	12.6/19.2/27.5	**16.4** / 23.2/33.6	1.6/5.0/21.4	3.0/8.3/25.2	4.4/9.8/32.8	1.3/4.9/23.8	3.9/9.2/26.5	4.6/11.3/33.1	4.8/10.5/32.0	4.6/13.0 /34.3	4.5/12.7/ 40.1	**14.3/25.8/58.9**
	SP + LG	11.0/19.6/35.0	14.9/22.0/33.6	**22.3/31.7** /47.4	2.7/7.1/22.8	4.3 /10.5/29.6	4.2/ 12.6/38.1	2.3/8.4/26.3	4.4/11.9/29.5	6.1/15.9/40.6	3.6/9.9/30.9	6.2 /15.5/37.1	5.4/ 16.9/44.8	**17.7/31.7/56.8**
	SP + SG	11.2/19.3/33.1	14.7/23.1/32.6	**22.4/31.5** /46.7	2.0/5.5/20.3	2.7/9.4/27.3	4.5/11.8/31.3	2.2/6.4/21.4	3.9 /9.9/25.9	3.6/ 11.8/35.2	4.2/10.5/26.5	6.0 /13.7/33.1	5.9/ 14.8/37.5	14.8/26.5/ **53.0**
CosPlace	SIFT + NN	6.2/8.8/12.4	6.0/8.9/12.0	10.4/13.1/15.4	1.4/3.0/6.8	2.9/5.5/10.3	3.6/7.2/13.2	1.9/4.0/8.3	3.5/7.0/11.1	4.0/7.8/13.7	2.9/6.0/12.1	5.2 /8.9/15.3	5.1/ 9.2/17.8	**22.8/34.1/55.6**
	DISK + LG	7.7/13.5/24.4	7.8/12.8/21.7	9.6/14.2/21.8	1.2/3.4/16.0	2.7 /6.3/20.7	2.5/ 7.4/23.3	1.6/4.3/18.8	2.1/7.0/19.9	4.0/8.0/25.4	2.1/6.2/25.5	3.9/8.9/28.9	4.6/9.9/32.0	**14.3/27.5/59.2**
	SP + LG	9.5/17.9/28.4	11.0/18.0/25.9	13.4/19.3/29.8	1.5/4.8/16.3	2.0/6.9/19.0	3.5/9.4/28.4	1.4/4.7/18.9	3.0 /7.7/20.7	3.0/9.4/30.7	3.8/8.7/24.5	4.7/11.8/28.5	6.1/14.1/35.4	**18.4/33.1/58.7**
	SP + SG	10.5/17.4/26.7	10.5/16.3/23.9	13.4/20.0/29.7	0.8/4.5/15.1	1.8 /5.1/18.5	1.6/ 7.5/23.4	1.6/4.9/14.1	2.5/7.7/18.1	2.6/8.7/24.6	2.5/6.7/20.4	4.6/11.0 /26.9	3.8/9.9/ 29.1	**16.4/30.5/58.9**
		Night												
NetVLAD	SIFT + NN	1.4/3.2/6.6	2.8/4.4/5.5	7.1/9.5/10.9	1.2/2.2/5.1	2.5/4.9/7.3	3.3/6.3/9.2	1.0/2.1/5.3	2.4/4.8/8.0	3.9/6.7/10.8	1.7/4.7/7.2	5.5 /9.4/13.3	5.3/ 9.8/15.3	**21.6/29.0/36.4**
	DISK + LG	5.0/11.5/19.7	6.5/10.7/16.1	10.4/17.2/25.7	1.1/4.4/13.4	2.4/6.6/17.6	2.8/8.9/21.7	1.3/4.6/14.5	2.2 /5.4/17.7	2.2/9.0/27.1	2.7/7.2/22.3	3.5/10.8/27.2	5.4/12.5/31.3	**17.7/32.1/55.1**
	SP + LG	6.2/14.2/27.4	8.7/14.4/21.8	16.0/24.3/34.4	1.7/4.9/15.7	3.6/10.1/24.0	4.3/12.5/29.8	2.4/7.1/18.7	4.6/11.8/22.1	4.7/14.9/33.9	3.4/8.8/25.3	4.4/12.8/30.0	5.5/15.6/38.2	**13.2/26.8/49.6**
	SP + SG	6.3/13.9/24.2	7.8/14.6/20.5	**14.3** / 23.2/33.1	0.9/3.6/10.8	2.5 /7.3/17.4	2.3/ 10.3/24.6	1.6/5.2/15.1	2.5/6.9/19.6	3.1/10.7/25.2	1.8/6.8/18.4	3.9/9.5/23.5	4.9/13.9/32.1	14.2/ **27.8/50.8**
CosPlace	SIFT + NN	1.7/2.9/5.8	1.7/2.9/4.7	2.8/3.4/4.9	0.4/1.1/2.5	1.3/2.5/4.4	2.9/4.4/6.4	0.5/1.8/3.7	1.2/2.5/4.6	2.5/5.6/8.2	1.9/4.0/5.8	1.8/4.3/6.8	4.3/7.3/11.1	**23.9/31.1/39.9**
	DISK + LG	4.4/8.5/15.7	4.4/8.5/14.3	6.1/8.3/12.9	1.0/3.3/11.1	1.6/4.0/12.5	1.9/5.6/18.0	1.2/3.7/12.9	1.2/3.8/12.1	1.8/6.4/19.3	2.1/5.4/15.1	2.6/7.2/20.5	4.7/10.7/28.7	**17.5/34.0/59.9**
	SP + LG	5.2/11.8/20.2	4.8/9.8/16.4	7.2/10.8/19.6	1.2/2.8/12.7	2.0/5.4/14.5	3.3/9.0/24.6	1.6/3.7/15.7	1.9/4.7/13.4	4.0/11.6/27.1	2.4/5.8/17.5	3.4/7.5/19.4	5.4/14.1/30.5	**14.2/29.8/55.3**
	SP + SG	5.3/11.4/19.4	5.0/9.7/16.1	7.0/12.0/18.6	0.8/2.9/11.4	1.3/4.1/10.8	2.1/7.2/19.7	1.4/3.5/11.7	1.3/4.5/11.2	2.4/7.6/21.0	1.5/4.7/13.5	2.7/7.4/16.4	3.6/8.3/24.2	**14.0/28.2/52.7**

Table 5. *Concourse*. Local matching localization results. Percentage of predictions with high (0.25m, 2°), medium (0.5m, 5°), and low (5m, 10°) accuracy [11] (higher is better). # indicates the highest value for each device w and w/o virtual cameras (VC1, VC2) of each accuracy level. The best results for all devices of each accruacy level are in bold with # .

Scene		Day												
Hall		Query camera												
Retrieval	local Matching	pinhole	+ VC1	+VC2	fisheye1	+ VC1	+VC2	fisheye2	+ VC1	+VC2	fisheye3	+ VC1	+VC2	360
NetVLAD	SIFT + NN	8.9/13.8/20.7	11.4/16.3/20.6	35.8/45.1/52.2	5.3/11.6/25.7	11.1/21.8/37.3	12.6/24.3/42.7	7.6/14.0/27.5	12.5/24.8/41.3	14.1/25.6/44.4	10.6/19.6/37.5	16.4/32.3/50.2	15.1/28.4/47.7	**36.0/54.7/76.0**
	DISK + LG	11.4/17.8/27.3	13.4/18.0/22.4	27.2/36.5/44.3	3.8/9.9/26.7	7.3/19.4/40.4	8.4/21.6/43.6	3.9/11.0/30.5	10.0/23.2/44.3	9.4/22.8/46.7	8.0/19.3/41.1	14.7/31.8/56.0	11.7/27.6/53.7	**36.0/57.2/80.6**
	SP + LG	14.4/22.4/33.7	14.5/20.6/26.3	**34.0** / 44.7/53.6	4.1/9.8/27.6	8.0/19.7/41.9	8.5/22.3/47.6	3.9/11.8/32.0	9.7/23.9/44.9	9.7/24.5/50.4	7.7/19.7/42.7	11.1/ 28.9/54.9	11.4 /27.4/53.3	32.9/ **55.4/75.3**
	SP + SG	14.1/22.4/33.7	15.1/21.0/27.2	**36.7** / 50.4/61.4	3.2/8.7/26.0	6.3/17.3/37.7	6.2/18.4/42.9	3.7/10.4/28.5	8.3/20.7/43.5	7.2/20.4/44.9	6.3/15.9/39.5	11.2/27.3/53.3	10.6/24.2/47.7	27.5/ **49.8/73.6**
CosPlace	SIFT + NN	3.9/6.3/9.8	9.0/12.8/16.2	15.1/20.4/25.2	3.0/7.3/14.5	7.0/12.7/22.5	8.4/18.3/32.1	3.7/7.3/16.7	8.0/14.6/23.4	8.2/17.4/32.8	7.7/15.0/28.3	11.3 /20.7/34.8	11.0/ 22.0/39.3	**37.1/54.7/77.1**
	DISK + LG	6.2/9.0/15.9	10.4/15.3/20.3	14.0/19.4/24.6	1.7/5.0/16.6	4.9/11.4/23.7	6.1/14.4/33.3	2.8/7.3/19.7	6.3 /14.3/28.1	5.7/ 14.9/36.3	6.2/14.1/31.8	10.6/21.9 /41.3	9.6/20.4/ 44.6	**36.7/57.3/79.8**
	SP + LG	5.8/10.4/18.8	10.4/15.5/20.7	16.1/22.0/27.0	1.7/6.5/17.4	4.1/10.8/23.4	5.1/14.6/34.8	2.8/7.3/19.2	5.1/13.3/27.3	5.5/15.5/38.0	5.7/13.9/30.9	8.3/20.9/39.5	9.0/22.8/44.4	**31.6/52.7/77.0**
	SP + SG	6.0/10.4/19.1	11.6/16.5/22.2	17.9/24.2/31.1	2.0/5.8/15.9	3.7/10.4/23.5	4.4/12.6/32.3	2.3/6.8/18.8	4.4/12.4/25.4	5.3/13.9/34.2	4.1/12.2/29.7	8.1 /18.6/38.2	7.1/ 18.7/40.4	**29.4/49.9/73.6**
		Night												
NetVLAD	SIFT + NN	0.0 /0.0/0.1	0.0 /0.0/0.0	0.0/0.4/0.7	0.0/0.2/0.8	0.1/0.5/1.3	0.2/0.6/1.5	0.0/0.1/0.8	0.0/0.4/1.2	0.2/0.7/1.9	0.3/0.7/1.6	0.1/0.8/2.1	0.3/1.0/2.0	**2.1/5.3/13.5**
	DISK + LG	0.4/1.6/4.5	0.5/1.3/2.5	1.9/4.7/8.2	0.4/1.5/6.3	0.2/1.6/6.4	0.3/3.0/12.8	0.3/1.8/8.2	0.4/2.8/9.2	0.7/3.9/14.8	0.6/4.1/14.0	1.1/4.4/16.3	1.2/6.3/22.1	**7.2/22.1/53.5**
	SP + LG	0.6/2.1/6.7	0.6/1.7/3.9	2.1/5.4/10.5	0.5/2.2/10.5	0.9/3.4/11.2	1.1/5.1/20.4	0.4/2.6/12.1	0.9/4.1/14.0	1.2/5.6/23.9	0.8/4.3/16.5	1.7 /6.6/21.3	1.7/7.9/29.2	**6.1/21.6/54.6**
	SP + SG	0.6/2.4/6.2	0.7/1.5/4.2	2.5/6.9/13.1	0.4/1.8/8.5	0.4/2.2/10.0	0.6/4.1/18.9	0.4/2.2/9.8	0.5/3.0/11.9	1.0/5.4/22.2	0.8/4.3/15.0	1.2 /5.6/18.5	1.0/ 6.2/26.4	**5.7/17.6/50.1**
CosPlace	SIFT + NN	0.0/0.1/0.2	0.0/0.0/0.0	0.1/0.1/0.4	0.0/0.1/0.2	0.0/0.0/0.1	0.1/0.3/0.9	0.0/0.3/0.5	0.0/0.0/0.0	0.0/0.2/1.1	0.0 /0.2/0.8	0.0 /0.2/0.6	0.0/0.6/1.7	**2.1/6.7/14.9**
	DISK + LG	0.8/1.8/4.1	0.6/1.4/3.7	1.2/3.3/6.8	0.3/1.3/4.9	0.0/0.7/3.5	0.7/2.6/10.0	0.3/1.6/6.6	0.1/0.8/4.2	0.3/2.3/10.5	0.2/2.2/11.0	0.7/2.7/9.2	0.6/4.5/15.4	**7.0/23.6/62.3**
	SP + LG	0.5/1.8/5.2	0.7/1.6/4.0	1.5/4.5/8.4	0.3/1.6/9.0	0.3/0.9/4.5	0.8/3.5/16.0	0.1/1.0/8.8	0.1/1.2/5.3	1.0/4.2/17.0	0.6/2.9/13.5	0.9/3.7/12.2	1.4/6.1/21.4	**5.2/22.1/64.7**
	SP + SG	0.6/1.5/5.1	0.6/2.1/4.4	2.5/5.2/10.8	0.2/1.5/6.7	0.1/1.1/3.7	0.7/3.6/14.5	0.5/1.8/7.4	0.1/1.0/4.7	0.8/3.7/16.1	0.4/3.0/11.6	0.8/3.5/10.8	1.0/4.3/19.6	**5.8/20.1/60.9**

Table 6. *Hall*. Local matching localization results. Percentage of predictions with high (0.25m, 2°), medium (0.5m, 5°), and low (5m, 10°) accuracy [11] (higher is better). # indicates the highest value for each device w and w/o virtual cameras (VC1, VC2) of each accuracy level. The best results for all devices of each accruacy level are in bold with # .

Scene		Day												
Atrium		Query camera												
Retrieval	local Matching	pinhole	+ VC1	+VC2	fisheye1	+ VC1	+VC2	fisheye2	+ VC1	+VC2	fisheye3	+ VC1	+VC2	360
NetVLAD	SIFT + NN	0.9/2.1/6.9	2.2/4.6/10.1	6.1/10.5/20.5	0.4/1.0/6.5	1.3/3.1/14.5	2.2/6.0/20.2	1.3/2.2/8.7	2.2/5.4/16.8	2.4/5.4/21.6	1.4/3.7/13.5	2.8/6.5 /21.5	2.5/6.3/ 26.5	**12.8/24.5/56.7**
	DISK + LG	2.5/6.9/24.7	3.5/9.6/23.5	6.5/14.8/35.7	0.2/1.3/11.1	0.8/3.1/22.4	1.0/4.8/30.7	0.5/1.5/14.6	1.0/ 4.0 /27.6	1.1 /3.9/ 30.2	0.9/3.0/23.5	1.3/5.0/33.5	2.5/7.1/38.2	**7.9/17.4/64.1**
	SP + LG	3.5/8.9/29.3	4.7/11.4/24.2	**10.7** / 21.3/37.2	0.2/1.5/14.3	1.6/5.1/28.2	2.0/6.9/35.1	0.8/3.1/17.5	1.5/5.7/30.9	2.1/7.3/38.6	2.2/5.6/26.2	2.6/8.3/41.0	3.1/9.9/46.6	10.5/ **24.3/65.4**
	SP + SG	4.3/9.5/29.7	5.8/12.5/26.6	**12.7** / 24.7/46.6	0.7/2.1/14.1	1.0/4.6/28.1	1.5/6.2/34.6	0.4/1.5/13.6	2.1 /5.3/30.2	1.8/ 6.7/36.9	1.7/4.3/22.8	2.5 /7.8/33.9	2.5/8.3/40.7	10.2/ **23.3/59.5**
CosPlace	SIFT + NN	0.7/2.2/6.5	1.3/2.9/8.8	3.0/5.7/13.7	0.3/0.8/5.4	1.2/2.9/11.4	1.6/3.7/15.7	0.2/1.3/5.2	1.0/ 3.0 /11.3	1.2 /2.7/ 15.4	0.9/2.3/8.8	2.0 /4.6/18.2	1.9/ 5.1/19.4	**12.8/23.9/60.7**
	DISK + LG	1.8/4.9/19.0	3.2/8.1/24.5	4.6/10.6/28.0	0.0/0.5/8.6	0.5 /1.8/17.5	0.3/ 2.6/23.0	0.4/1.3/10.2	0.6 /1.9/18.1	0.5/ 2.3/23.1	0.5/1.7/15.5	1.0/4.6 /26.3	1.0 /3.3/ 31.7	**9.8/19.8/71.5**
	SP + LG	2.3/7.6/23.9	3.5/9.0/23.7	6.7/13.1/27.3	0.4/1.2/9.6	0.9/2.9/20.1	1.4/4.2/27.0	0.2/1.7/12.3	1.0/2.9/21.8	1.5/4.5/29.9	0.5/2.7/16.4	2.7 /6.1/32.5	1.8/ 6.9/34.9	**12.2/26.7/72.2**
	SP + SG	3.0/8.2/22.2	3.3/10.2/24.8	7.7/15.6/33.4	0.5/1.5/9.4	1.1/2.7/18.5	1.3/4.0/27.7	0.5/1.9/9.2	1.1/3.7/21.3	1.0/4.5/28.5	0.7/2.1/14.4	1.7/5.8/28.1	2.0/6.0/33.2	**10.5/24.8/69.6**
		Night												
NetVLAD	SIFT + NN	0.0/0.2/0.7	0.1/0.4/0.6	0.9/1.6/2.5	0.0/0.0/0.7	0.1/0.6/1.9	0.5/1.0/3.1	0.2/0.4/1.1	0.2/0.8/2.5	0.5/1.2/3.4	0.4/0.9/2.3	0.3/ 1.4 /3.1	0.5 /1.3/ 4.3	**3.9/6.5/15.5**
	DISK + LG	1.1/4.0/13.5	1.6/3.6/10.6	3.9/8.3/17.0	0.4/1.2/8.4	0.5/2.2/11.2	1.0/3.2/18.1	0.2/1.4/9.7	0.6/2.7/13.3	1.1/4.8/21.7	0.8/2.6/14.2	1.4/4.6/20.2	1.6/5.7/25.8	**7.1/19.4/56.4**
	SP + LG	1.5/4.2/14.5	1.9/3.9/11.8	5.5/10.3/18.1	0.4/1.6/11.1	0.7/3.2/17.6	1.6/5.4/26.5	0.4/1.8/11.9	1.1/4.0/18.5	1.6/6.1/29.3	0.7/3.8/16.0	1.6/5.5/25.0	1.9/7.6/32.2	**7.0/19.7/51.7**
	SP + SG	1.6/4.1/14.4	2.1/4.9/12.3	5.9/11.7/21.7	0.3/1.6/9.8	0.7/2.4/13.7	1.4/5.5/24.8	0.6/2.1/11.3	0.9/3.2/16.8	1.6/5.7/26.3	0.5/2.7/15.3	1.6/4.8/21.8	2.1/6.8/30.4	**6.3/18.7/50.5**
CosPlace	SIFT + NN	0.0/0.1/0.2	0.2/0.3/0.7	0.6/0.9/1.4	0.0/0.1/0.5	0.0/0.0/0.8	0.3/0.5/2.0	0.0/0.1/0.7	0.1/0.3/1.0	0.1/0.7/2.5	0.1/0.3/0.8	0.2/0.7/2.2	0.5/1.1/3.0	**3.5/7.9/19.0**
	DISK + LG	0.9/3.0/10.6	1.6/4.3/11.8	2.7/5.9/13.0	0.2/0.8/4.9	0.3/1.3/10.7	0.7/2.5/15.0	0.2/0.9/5.8	0.4/1.6/10.7	0.8/2.7/17.9	0.2/1.4/9.9	0.9/3.0/16.7	1.4/4.8/24.6	**7.8/22.5/62.3**
	SP + LG	1.4/3.8/13.1	2.3/5.0/13.5	3.0/6.5/14.2	0.3/1.1/7.7	0.3/1.6/12.9	0.8/4.0/21.7	0.2/0.9/8.7	0.9/2.8/14.9	1.2/4.3/24.7	0.6/2.3/11.6	0.9/3.5/19.2	1.2/5.6/29.2	**6.4/18.0/55.7**
	SP + SG	1.5/4.1/12.4	2.4/5.5/13.5	3.8/8.2/16.9	0.1/0.8/6.4	0.4/1.8/12.0	0.6/3.9/20.8	0.3/1.0/6.8	0.7/2.3/11.6	1.0/3.9/22.2	0.3/1.7/9.7	0.9/2.9/15.8	1.4/6.0/26.9	**6.4/17.3/53.1**

Table 7. *Atrium*. Local matching localization results. Percentage of predictions with high (0.25m, 2°), medium (0.5m, 5°), and low (5m, 10°) accuracy [11] (higher is better). # indicates the highest value for each device w and w/o virtual cameras (VC1, VC2) of each accuracy level. The best results for all devices of each accruacy level are in bold with # .

Scene		Day												
Piatrium		Query camera												
Retrieval	local Matching	pinhole	+ VC1	+VC2	fisheye1	+ VC1	+VC2	fisheye2	+ VC1	+VC2	fisheye3	+ VC1	+VC2	360
NetVLAD	SIFT + NN	1.2/2.5/7.9	4.1/7.6/13.1	10.1/16.2/27.5	0.9/2.1/8.4	1.9/5.5/18.0	3.0/6.9/21.0	1.2/2.3/8.1	3.2/7.0/18.9	2.2/6.0/21.6	2.5/5.7/15.3	4.1/9.0/24.7	3.7/8.0/23.2	**13.4/24.8/51.3**
	DISK + LG	2.3/6.1/18.8	4.4/9.4/20.5	6.8/14.2/28.5	0.7/1.4/11.6	2.1/6.0/22.2	1.9/5.6/25.1	0.8/2.3/14.5	2.2/6.6 /25.3	2.2/6.1/ 27.7	1.3/5.1/22.4	3.1/8.9 /34.0	2.3/8.2/ 35.1	**10.0/22.9/60.9**
	SP + LG	3.2/8.7/25.7	7.5/14.1/26.1	**12.3** / 21.0/33.4	0.8/3.1/15.6	2.5/7.2/29.0	2.5/7.8/31.8	0.7/3.6/17.0	3.3/8.3/31.0	2.6/8.4/34.7	2.4/6.7/26.6	4.3/12.3/40.9	3.7/11.7/40.5	11.9/ **26.9/59.2**
	SP + SG	3.9/9.6/26.2	8.1/14.7/27.7	**14.4/26.4** / 43.9	0.4/2.5/13.2	2.0/6.8/25.4	2.2/7.6/29.1	0.4/2.4/14.5	2.1/6.6/26.5	2.3/7.0/30.0	1.6/5.5/24.8	3.6/10.1 /36.1	2.7/9.6/ 36.2	10.7/25.1/ **55.5**
CosPlace	SIFT + NN	0.7/1.1/3.8	2.6/4.5/10.0	4.2/7.9/15.0	0.5/1.4/4.6	1.3/3.0/11.9	2.1/4.3/14.6	0.6/1.5/5.6	1.5/3.7/12.8	2.1/4.7/15.6	1.3/3.2/9.8	2.8/5.9 /16.3	2.3/5.3/ 19.0	**10.3/21.3/50.9**
	DISK + LG	1.2/3.8/12.7	3.1/6.9/17.7	3.8/8.2/19.7	0.4/1.2/6.1	0.9/2.7/15.6	1.0/3.4/21.6	0.5/1.2/8.3	0.9/2.9/16.5	1.2/4.1/23.5	1.0/3.2/14.7	2.0 /5.3/24.0	2.0/5.7/30.2	**9.7/22.8/61.8**
	SP + LG	1.6/5.1/17.3	5.2/10.2/19.7	6.4/11.3/22.5	0.3/1.5/8.8	1.6/3.9/16.9	1.0/4.4/26.0	0.6/1.7/10.5	1.0/4.2/18.8	1.6/6.4/29.3	1.1/3.4/17.4	2.7/7.5 /27.7	1.8/6.8/ 32.3	**12.4/27.0/64.5**
	SP + SG	2.2/5.5/16.5	5.2/10.9/20.3	7.4/14.2/27.8	0.5/1.6/8.2	1.1/3.7/15.9	1.0/4.3/23.4	0.6/1.9/9.4	0.9/3.4/16.1	1.2/4.7/24.4	0.9/2.7/15.5	2.8 /6.4/25.7	2.5/ 7.3/27.5	**10.1/25.4/60.8**
		Night												
NetVLAD	SIFT + NN	0.1/0.1/0.1	0.0/0.0/0.1	0.0/ 0.1/0.1	0.0/0.1/0.3	0.0/0.1/0.4	0.1/0.4/0.9	0.0/0.0/0.0	0.0/0.1/0.4	0.0/0.1/0.4	0.0/0.0/0.3	0.0/0.1/0.4	0.1/0.2/1.1	**0.4/1.0/2.9**
	DISK + LG	0.2/0.6/3.6	0.4 /0.7/2.4	0.1/ 0.9/3.6	0.0/0.1/1.6	0.0/0.4/3.1	0.1/0.8/5.7	0.0/0.1/2.2	0.1/1.1/4.5	0.2/1.1/5.7	0.0/0.6/4.6	0.1/1.2/7.0	0.2/1.6/10.9	**1.9/6.9/24.8**
	SP + LG	0.6/1.4/5.5	0.4/1.1/2.9	0.8/1.6/4.4	0.2/0.4/4.7	0.3/0.8/7.0	0.5/2.5/10.5	0.1/0.4/4.4	0.4 /1.9/8.5	0.4/2.0/11.9	0.2/1.4/7.8	0.5/2.5/11.3	0.9/3.3/16.6	**1.6/6.6/25.3**
	SP + SG	0.6/1.9/4.3	0.5/1.4/2.5	1.0/2.4/5.9	0.3/0.4/3.8	0.1/0.5/5.9	0.2/1.5/9.1	0.1/0.4/3.4	0.3/1.4/6.5	0.4/1.7/11.3	0.2/1.2/6.6	0.5/2.2/9.4	0.4/2.4/14.8	**2.0/6.9/19.9**
CosPlace	SIFT + NN	0.1/0.1/0.2	0.0/0.0/0.1	0.1/0.1/0.2	0.0/0.0/0.3	0.0/0.0/0.3	0.0/0.0/0.3	0.0/0.0/0.1	0.1/0.1/0.3	0.0/0.0/0.6	0.0/ 0.2 /0.3	0.0/0.0/0.1	0.1 /0.1/ 0.9	**0.4/2.0/6.2**
	DISK + LG	0.1/ 0.8/3.9	0.1/0.3/3.0	0.2/0.8/3.9	0.0/0.1/2.5	0.1/0.4 /2.7	0.1/0.4/5.7	0.0/0.1/2.4	0.1/ 0.5 /2.8	0.1 /0.3/ 5.8	0.0/0.4/4.3	0.1 /0.4/4.6	0.1/0.8/9.3	**2.6/8.0/39.3**
	SP + LG	0.4/1.4/5.8	0.3/1.5/4.4	0.4 /1.0/3.6	0.1/0.3/3.5	0.1/0.9/4.0	0.4/1.9/9.7	0.1/0.5/3.3	0.3 /0.8/3.9	0.2/ 1.3/11.7	0.1/0.6/7.1	0.4/1.7/7.0	0.7/2.8/14.3	**3.3/10.2/38.0**
	SP + SG	0.9/1.7/4.9	0.5/1.3/4.2	0.6/1.6/4.8	0.1/0.3/3.6	0.1/0.8/2.9	0.4/1.4/9.8	0.1/0.3/2.9	0.0/0.5/4.4	0.3/1.8/11.3	0.2/0.9/5.5	0.3/1.4/6.2	0.4/1.9/12.1	**2.3/9.2/35.0**

Table 8. *Piatrium*. local matching localization results. Percentage of predictions with high (0.25m, 2°), medium (0.5m, 5°), and low (5m, 10°) accuracy [11] (higher is better). # indicates the highest value for each device w and w/o virtual cameras (VC1, VC2) of each accuracy level. The best results for all devices of each accruacy level are in bold with # .

scene	Concourse									
query	pinhole		fisheye1		fisheye2		fisheye2		360	
APR	PN	PN^{vc2}	PN	PN^{vc2}	PN	PN^{vc2}	PN	PN^{vc2}	PN	PN^{vc2}
Day	18.6/95.0	7.7/26.6	10.7/88.7	3.5/14.0	10.1/87.7	3.3 / **13.8**	7.7/85.9	3.0/15.2	2.6/39.0	**2.1** / 38.5
Night	19.8/96.2	11.6/46.5	13.3/85.5	8.4/22.0	13.0/88.4	8.3 / **21.3**	11.1/87.5	8.9/20.4	**5.5** / 45.6	8.1/46.6
APR	MS-T	MS-T^{vc2}	MS-T	MS-T^{vc2}	MS-T	MS-T^{vc2}	MS-T	MS-T^{vc2}	MS-T	MS-T^{vc2}
Day	17.0/69.1	4.6/20.4	11/60.8	2.0/10.5	10.5/59.6	1.9/10.3	9.9/63.4	**1.7** / 10.6	4.9/34.6	1.9 / **10.2**
Night	22.1/74.7	8.9/36.6	17.7/69.5	5.9/18.6	17.0/68.3	**5.7** / **18.1**	17.2/72.7	5.8/19.2	8.4/57.3	7.0 / **18.1**

Table 9. *Concourse*. The median translation and rotation errors (m/°) of different APRs during daytime (lower is better). # indicates the lowest value of error for each device of APR and APR^{vc2}. Best results for all devices of APR and APR ^{vc2} are in bold with # .

scene	Hall									
query	pinhole		fisheye1		fisheye2		fisheye2		360	
APR	PN	PN^{vc2}	PN	PN^{vc2}	PN	PN^{vc2}	PN	PN^{vc2}	PN	PN^{vc2}
Day	18.6/99.4	8.5/35.5	11.1/95.7	3.2 / **19.7**	10.4/95.9	3.0/20.0	7.5/92.3	2.8/20.3	2.7/97.3	**1.9** / 76.5
Night	22.8/95.6	20.5/91.3	18.3/93.8	15.1/65.5	17.9/93.6	14.8 / **63.2**	16.8/91.3	**14.4** / 63.8	10.9 /100.1	15.2/ 77.1
APR	MS-T	MS-T^{vc2}	MS-T	MS-T^{vc2}	MS-T	MS-T^{vc2}	MS-T	MS-T^{vc2}	MS-T	MS-T^{vc2}
Day	21.8/96.0	5.0/37.9	14.1/94.0	2.0/22.7	13.5/92.7	1.8/21.2	11.2/91.7	**1.6** / 21.9	11.4/88.5	1.7 / **19.6**
Night	32.1/96.1	29.4/88.6	29.9/94.2	18.6 / **28.9**	29.7/93.6	18.4/76.5	31.8/93.1	18.5/76.0	22.3/94.4	**14.7** / 75.0

Table 10. *Hall*. The median translation and rotation errors (m/°) of different APRs during daytime (lower is better). # indicates the lowest value of error for each device of APR and APR^{vc2}. Best results for all devices of APR and APR ^{vc2} are in bold with # .

scene	Atrium									
query	pinhole		fisheye1		fisheye2		fisheye2		360	
APR	PN	PN^{vc2}	PN	PN^{vc2}	PN	PN^{vc2}	PN	PN^{vc2}	PN	PN^{vc2}
Day	15.5/85.5	10.8/37.0	11.8/79.9	5.0 / **17.3**	11.3/80.4	4.7 / **17.3**	10.1/79.8	4.2/17.7	4.5/82.6	**3.4** / 78.4
Night	19.8/90.8	16.1/83.1	16.4/87.3	10.0/46.5	16.1/86.7	9.9/46.6	15.7/83.5	10.3/47.1	**8.7/22.8**	10.8/30.7
APR	MS-T	MS-T^{vc2}	MS-T	MS-T^{vc2}	MS-T	MS-T^{vc2}	MS-T	MS-T^{vc2}	MS-T	MS-T^{vc2}
Day	20.1/77.7	10.5/46.9	14.5/73.7	5.4/28.0	14.0/71.7	5.1 / **27.2**	12.5/65.6	**4.6** / 29.1	13.2/54.7	5.2/41.1
Night	24.8/84.7	17.7/79.7	17.8/80.5	10.9/72.0	17.0/80.1	10.4/72.3	15.1/80.0	10.3/73.5	**5.0/46.3**	8.5/68.6

Table 11. *Atrium*. The median translation and rotation errors (m/°) of different APRs during daytime (lower is better). # indicates the lowest value of error for each device of APR and APR^{vc2}. Best results for all devices of APR and APR ^{vc2} are in bold with # .

scene	Piatrium									
query	pinhole		fisheye1		fisheye2		fisheye2		360	
APR	PN	PN^{vc2}	PN	PN^{vc2}	PN	PN^{vc2}	PN	PN^{vc2}	PN	PN^{vc2}
Day	21.3/88.0	12.2/38.4	13.8/86.5	4.9 / **18.7**	13.0/87.1	4.6/18.9	10.9/84.4	4.4/19.6	5.7/69.8	**3.8** / 59.5
Night	33.2/89.7	28.8/90.4	29.3/89.6	**23.2** / 64.6	28.8/90.4	23.3 / **64.1**	29.2/89.8	23.5/67.0	30.6/98.5	23.5/80.8
APR	MS-T	MS-T^{vc2}	MS-T	MS-T^{vc2}	MS-T	MS-T^{vc2}	MS-T	MS-T^{vc2}	MS-T	MS-T^{vc2}
Day	30.0/78.5	10.8/35.8	23.4/71.3	4.0/21.1	22.7/72.7	3.8 / 20.7	19.3/72.6	**3.7** / 20.4	11.3/52.0	4.1 / **17.6**
Night	37.7/91.3	30.0/87.4	37.3/89.1	24.3/75.1	36.2/89.0	23.8/74.5	35.7/85.6	23.9/75.1	37.2/83.0	**22.1** / **53.7**

Table 12. *Piatrium*. The median translation and rotation errors (m/°) of different APRs during daytime (lower is better). # indicates the lowest value of error for each device of APR and APR^{vc2}. Best results for all devices of APR and APR ^{vc2} are in bold with # .