CrowdDriven: A New Challenging Dataset for Outdoor Visual Localization - Supplementary Material

Ara Jafarzadeh[1] Manuel López Antequera[2] Pau Gargallo[2] Yubin Kuang[2] Carl Toft[1] Fredrik Kahl[1] Torsten Sattler[3]

[1]Chalmers University of Technology [2]Facebook [3]Czech Technical University in Prague

This supplementary material provides the following information: Sec. 1 provides details about the PixLoc baseline used in the main paper. Sec. 2 provides a more detailed description of the sequence-based localization approaches used in this work and presents experimental results. Sec. 3 shows images from the scenes included in our proposed benchmark.

1. PixLoc Details

Given an initial pose estimate, a set of database images, and a set of 3D points potentially visible in it, PixLoc [12] refines the initial estimate by minimizing a feature-metric error: PixLoc computes feature maps for the images, projects the 3D points into the test images and the database images, and minimizes the difference in the features belonging to the projections. For a test image, we consider each database image from the same scene. For each database image $\mathcal{I}_D$, we identify the four other database images taken from the most similar positions and use the 3D points visible in these five database images. The pose of $\mathcal{I}_D$ is used to initialize the pose of the test image. This results in N poses for the test image, one for each of the N database images. We tried selecting the pose with the smallest feature-metric error but found that this approach did not work well. Likely, this is due to comparing poses that observe different 3D points. Rather, we use a simple scoring function for each pose: let $\mathbf{c}_i$ be the test image position estimated for the i-th database image. The score s_i for the i-th pose (computed from the i-th database image) is given as

$$s_i = \sum_{j \neq j} \frac{1}{||\mathbf{c}_i - \mathbf{c}_j||_2 + \varepsilon} \ , \tag{1}$$

where ε is a small constant to avoid division by zero. Intuitively, the score for a pose is large if there are multiple other pose estimates nearby. As such, the scoring is based on the assumption that we will have multiple pose estimates close to the true test pose while incorrect poses are rather far from each other.

For PixLoc, we use the 3D models created for the HLoc method [10, 11] as we got better results with these models compared to those build using SIFT [8] and D2-Net [3] features.

2. Multi-image localization

As discussed in the main paper, in addition to single-image localization, we have evaluated a multi-image (sequence-based) localization approach on our dataset. Using known relative poses, this approach models a sequence of images as a generalized camera [9], *i.e.*, a camera with multiple centers of projections. This enables us to estimate the poses of all images in the sequence at the same time [2, 5, 7, 14, 16]. The advantage of this approach is that it enables localizing multiple images even if none of them individually has enough correct matches to facilitate successful pose estimation.

We use the multi-image approach from [17], using the code publicly released by the authors.[1] This method use a minimal solver [5] (inside a LO-RANSAC [6] loop) that estimates both the pose of the generalized camera and its intrinsic scale, *i.e.*, the scale of the distances between the individual images in a sequence. This models the fact that some approaches, *e.g.*, monocular SLAM, might not be able to estimate the scale. Each new best model found inside RANSAC is optimized using local optimization [6]. This includes a non-linear optimization of the sum of squared reprojection errors over the inlier matches of the estimated generalized poses. Finally, the best pose found by RANSAC is optimized using the same optimization method. Non-linear optimization is implemented using the Ceres library [1]. The 2D-3D matches required for estimating the pose of a generalized camera are provided by the baselines used in the main paper. *I.e.*, we use the 2D-3D matches found for each individual image in a generalized camera.

[1]https://github.com/tsattler/MultiCameraPose

We use sequences of length k to define the generalized cameras. More precisely, for the i-th image in the sequence and a given sequence length k, we create a generalized camera containing images i to $i+k$.[2] As a result, each image is contained in multiple generalized cameras. Thus, there are multiple pose candidates for each test image, corresponding to the poses of the generalized cameras it is part of. We select the pose corresponding to the generalized camera pose with the largest number of inliers.

As in [13], we obtain the *relative* poses required to form generalized cameras using the ground truth poses. This allows us to obtain an upper bound on the pose accuracy that can be obtained via multi-image queries. As can be seen from the results shown in Tab. 3 of the main paper, HLoc [10, 11] and D2-Net [3] clearly outperform the Rectified SIFT [15] and S2DHM [4] baselines. In our experiments, we thus focus on HLoc and D2-Net. Note that we are not using PixLoc as the multi-image localization approach described above requires 2D-3D correspondences while PixLoc is not providing 2D-3D matches.

Table 1 provides detailed results on the easy (light gray), medium (gray), and hard (dark gray) parts of our benchmark. We use all images in a sequence to define the sequence length k. The results on the easy scenes are included as a form of sanity check to show that sequence-based localization works as intended. For reference, we also include Tab. 3 from the main paper, which provides single-image localization results, as Tab. 2 here.

As can be seen by comparing the two tables, using sequences typically improves the performance for the easy datasets, especially in terms of median position and orientation errors. For the medium datasets (standard gray), we observe that using sequences instead of single images for localization again improves results for many datasets. These improvements can be substantial, *e.g.*, for Boston2, Boston4, Cambridge, Massachusetts3, and Thuringia. However, there are some exceptions, *e.g.*, Boston1 and Burgundy2, where using sequences can actually lead to less accurate results. We attribute this to instabilities in the scale estimate refined by the local optimization, *e.g.*, due to the distribution of 2D-3D matches over the images in the sequences. Looking at the hard datasets (dark gray), we observe that using sequence-based localization does not enable us to achieve significantly better results, even though we are using all images in a sequence and ground truth poses to define the generalized cameras. This result shows that our benchmark contains challenges that cannot be easily solved by simply using image sequences for localization.

[2]For images in the end of the sequence, the generalized camera might contain less than k images.

Figure 1: Bayern, Category : Easy

Figure 2: Clermont-Ferrand, Category : Easy

3. Images from CrowdDriven

In order to provide an overview over the type of scenes and conditions contained in our benchmark, Figures 1 to 39 show example images from all of our datasets and include the category (Easy, Medium, Hard) of the dataset. In each pair, the left image comes from the set of reference / training images, while the right comes from the set of query / test images. Note that for illustration purposes, the images have been resized to the same size. The aspect ratio in the figures thus differs from the aspect ratios of the images contained in the benchmark.

References

[1] Sameer Agarwal, Keir Mierle, and Others. Ceres solver. http://ceres-solver.org. 1

[2] Federico Camposeco, Torsten Sattler, and Marc Pollefeys. Minimal Solvers for Generalized Pose and Scale Estimation from Two Rays and One Point. In *ECCV*, 2016. 1

[3] Mihai Dusmanu, Ignacio Rocco, Tomas Pajdla, Marc Pollefeys, Josef Sivic, Akihiko Torii, and Torsten Sattler. D2-net: A trainable cnn for joint detection and description of local features. *arXiv preprint arXiv:1905.03561*, 2019. 1, 2

[4] Hugo Germain, Guillaume Bourmaud, and Vincent Lepetit. Sparse-to-dense hypercolumn matching for long-term visual

	D2-Net			HLoc			
name	pos. err	rot. err	% of localized 0.5/1.0/5.0/10.0 (m) 2/5/10/20 (°)	pos. err	rot. err	% of localized 0.5/1.0/5.0/10.0 (m) 2/5/10/20 (°)	Changes
Angers1	36.62	179.54	F	31.98	178.82	F	📏
Angers2	70.23	162.00	F	11.20	5.10	0/ 0/6.52/ 45.65	📏
Bayern	0.01	0.06	100/ 100/ 100/ 100	0.01	0.05	100/ 100/ 100/ 100	💡☁️👷
Besançon2	20.67	178.70	F	124.43	147.95	F	🌳📏
Besançon3	43.82	173.62	F	69.34	162.69	F	🌳📏
Besançon4	104.50	154.95	F	67.19	143.19	F	☁️🌳📏
Boston1	32.67	3.13	F	23.96	2.38	0/ 0/ 0/ 2.13	🌒👷
Boston2	5.27	0.28	0/ 0/ 48.98/ 100	4.15	0.22	0/ 0/ 83.67/ 100	🌒
Boston3	8.90	2.76	0/ 0/ 22.58/ 61.29	3.60	3.56	0/ 0/ 64.52/ 74.19	💡📐👷
Boston4	4.66	1.40	2.94/ 2.94/ 55.88/ 91.18	4.61	1.99	5.88/ 14.71/ 64.71/ 88.24	🌒👷
Boston5	10.73	2.28	0/ 0/ 26.47/ 47.06	2.34	1.80	0/ 0/ 67.65/ 67.65	🌒👷
Brittany	59.55	142.10	F	20.51	144.92	0/ 0/ 9.68/ 9.68	🌳📏👷
Brourges	22.36	175.96	F	24.28	172.75	F	🌳📏👷
Burgundy2	4.15	4.18	0/ 0/ 58.00/ 96.00	8.05	3.39	0/ 0/ 30.00/ 62.00	💡☁️💦
Cambridge	0.30	0.54	84.85/ 100/ 100/ 100	0.28	0.38	100/ 100/ 100/ 100	🌒👷
Clermont-Ferrand	0.17	0.20	100/ 100/ 100/ 100	0.15	0.10	100/ 100/ 100/ 100	💡☁️👷
Curitiba	0.01	0.03	100/ 100/ 100/ 100	0.02	0.02	100/ 100/ 100/ 100	💡☁️👷
Ile-de-France	32.65	178.20	F	100.58	165.77	F	📏👷
Le-Mans	54.45	175.24	F	66.17	172.88	F	☁️📏👷
Leuven	14.28	175.78	F	12.03	175.96	F	☁️📏👷
Massachusetts1	0.03	0.02	100/ 100/ 100/ 100	0.10	0.04	100/ 100/ 100/ 100	💡☁️🌳👷
Massachusetts2	8.13	2.85	0/ 4.17/ 12.50/ 58.33	33.68	86.38	0/ 0/ 4.17/ 12.50	🌒👷
Massachusetts3	2.34	2.72	0/ 18.18/ 86.36/ 95.45	2.03	3.41	0/ 34.09/ 95.45/ 97.73	🌳🌒👷
Massachusetts4	1.30	0.86	0/ 2.56/ 100/ 100	1.84	1.01	0/ 7.69/ 97.44/ 100	🌒👷
Melbourne	0.05	0.03	100/ 100/ 100/ 100	0.06	0.07	100/ 100/ 100/ 100	💡👷
Muehlhausen	0.02	0.07	100/ 100/ 100/ 100	0.03	0.09	100/ 100/ 100/ 100	☁️👷
Nouvelle-Aquitaine1	52.11	173.29	F	55.75	176.71	F	❄️🍂📏
Nouvelle-Aquitaine2	47.36	173.73	F	67.90	166.50	F	📏
Orleans1	30.21	179.61	F	38.52	179.27	F	📏
Orleans2	49.18	173.41	F	22.35	174.73	F	📏👷
Pays de la Loire	27.04	166.66	F	26.04	177.43	F	🌳📏👷
Poing	0.02	0.01	100/ 100/ 100/ 100	0.02	0.01	100/ 100/ 100/ 100	💡🌳👷
Portland	0.04	0.04	100/ 100/ 100/ 100	0.06	0.03	100/ 100/ 100/ 100	💡☁️👷
Savannah	0.03	0.01	100/100/100/100	0.02	0.02	100/100/100/100	💡👷
Skåne	12.34	2.90	0/ 0/ 30.00/ 45.00	5.98	3.25	0/ 0/ 40.00/ 80.00	💡📐👷
Subcarpathia	0.30	0.09	82.35/ 100/ 100/ 100	0.32	0.08	82.35/ 100/ 100/ 100	❄️🍂👷
Sydney	0.04	0.02	100/ 100/ 100/ 100	0.04	0.03	100/ 100/ 100/ 100	💡☁️👷
Thuringia	0.42	0.11	81.82/ 90.91/ 100/ 100	0.26	0.09	100/ 100/ 100/ 100	💡👷
Tsuru	0.01	0.04	100/ 100/ 100/ 100	0.02	0.07	100/ 100/ 100/ 100	💡☁️👷
Washington	0.69	0.36	10.00/ 100/ 100/ 100	1.00	0.87	0/ 50.00/ 100/ 100	💡

Table 1: Performance of the baseline methods on our CrowdDriven benchmark when using sequence-based localization. We report the median position (in meters) and orientation (in degrees) errors, as well as the percentage of test images localized within certain error bounds on the position and orientation errors. We report results for using all test images in a scene to define the generalized camera used for sequence-based localization. Easy, medium, and hard datasets are color-coded in light, standard, and dark gray, respectively. The right side of the table provides information about the type of change between the training and test sequences: illumination: 💡, overcast: ☁️, foliage: 🌳, snow: ❄️, seasonal: 🍂, day-night: 🌒, small viewpoint: 📐, rain: 💦, strong viewpoint: 📏, man-made changes: 👷 . 'F' stands for failure to localize any image within the coarsest precision regime.

name	D2-Net pos. err	D2-Net rot. err	D2-Net % of localized 0.5/1.0/5.0/10.0 (m) 2/5/10/20 (°)	S2DHM pos. err	S2DHM rot. err	S2DHM % of localized 0.5/1.0/5.0/10.0 (m) 2/5/10/20 (°)	HLoc pos. err	HLoc rot. err	HLoc % of localized 0.5/1.0/5.0/10.0 (m) 2/5/10/20 (°)	Rectified SIFT pos. err	Rectified SIFT rot. err	Rectified SIFT % of localized 0.5/1.0/5.0/10.0 (m) 2/5/10/20 (°)	PixLoc pos. err	PixLoc rot. err	PixLoc % of localized 0.5/1.0/5.0/10.0 (m) 2/5/10/20 (°)	Changes
Angers1	28.02	177.43	F	97.81	171.26	F	46.39	161.65	F	191.62	148.27	F	21.03	175.11	F	📏
Angers2	35.51	165.56	F	174.61	153.78	F	68.34	122.82	0/ 0/ 0/ 6.52	437.63	132.38	F	45.26	173.18	F	📏
Bayern	0.03	0.06	96.15/ 96.15/ 96.15/ 96.15	0.09	0.30	80.77/ 80.77/ 80.77/ 80.77	0.02	0.07	80.77/ 80.77/ 80.77/ 80.77	0.04	0.11	80.77/ 84.62/ 84.62/ 84.62	0.09	0.12	61.54/ 61.54/ 65.38/ 73.08	💡☁👷
Besançon2	81.45	160.22	F	-	-	-	59.02	152.43	F	128.70	121.79	F	34.30	169.03	F	🌳📏
Besançon3	48.16	162.30	F	258.71	148.52	F	71.61	162.18	F	107.51	148.63	F	36.86	168.47	F	🌳📏
Besançon4	117.25	151.57	F	-	-	-	108.59	141.24	F	287.57	134.02	F	69.80	172.84	F	☁🌳📏
Boston1	28.78	4.99	0/ 0/ 4.26/ 23.40	239.57	140.05	F	31.75	8.47	0/ 0/ 0/ 10.64	125.13	129.79	F	27.09	15.55	0/ 0/ 0/ 2.13	🌒👷
Boston2	6.46	0.96	0/ 0/ 24.49/ 97.96	496.16	86.71	0/ 0/ 8.16/ 16.33	4.68	0.82	0/ 0/ 63.27/ 95.92	87.21	150.42	F	13.26	7.22	0/ 0/ 6.12/ 38.78	🌒👷
Boston3	6.66	4.20	0/ 0/ 29.03/ 51.61	196.98	114.24	F	27.57	32.83	0/ 0/ 12.90/ 19.35	111.70	155.53	F	20.72	16.12	0/ 0/ 0/ 19.35	💡👷📐
Boston4	12.94	2.51	0/ 0/ 20.83/ 41.67	-	-	-	15.57	6.19	0/ 0/ 26.47/ 38.24	-	-	-	18.90	5.81	0/ 0/ 11.76/ 26.47	🌒👷
Boston5	18.08	2.26	0/ 0/ 11.76/ 17.65	97.70	74.40	0/ 0/ 0/ 5.88	16.59	4.35	0/ 0/ 26.47/ 26.47	91.74	157.72	F	13.36	12.36	0/ 0/ 0/ 26.47	🌒👷
Brittany	14.74	147.24	F	177.38	143.03	F	36.44	137.46	0/ 3.23/ 3.23/ 3.23	305.68	121.02	F	14.96	162.08	F	🌳👷📏
Brourges	31.84	153.74	F	-	-	-	22.24	153.54	F	57.10	97.23	F	14.64	177.38	F	🌳📏👷
Burgundy2	4.41	3.72	0/ 4.00/ 60/ 76.00	57.78	34.59	F	7.32	5.36	0/ 4.00/ 40/ 58.00	554.19	166.83	F	20.97	14.82	0/ 0/ 0/ 26.00	💡☁💦
Cambridge	0.50	0.87	51.52/ 90.91/ 93.94/ 96.97	94.58	82.57	9.09/ 12.12/ 12.12/ 12.12	0.37	0.43	69.70/ 100/ 100/ 100	58.58	135.56	3.03/ 6.06/ 18.18/ 21.21	30.06	15.24	0/ 0/ 3.03/ 6.06	🌒👷
Clermont-Ferrand	0.22	0.24	100/ 100/ 100/ 100	0.25	0.47	100/ 100/ 100/ 100	0.15	0.27	100/ 100/ 100/ 100	0.21	0.33	80/ 93.33/ 93.33/ 93.33	0.19	0.20	93.33/ 93.33/ 93.33/ 93.33	💡☁👷
Curitiba	0.03	0.07	100/ 100/ 100/ 100	0.21	0.36	84.21/ 89.47/ 100/ 100	0.04	0.06	100/ 100/ 100/ 100	0.06	0.08	89.47/ 89.47/ 89.47/ 89.47	0.06	0.09	84.21/ 84.21/ 84.21/ 84.21	💡☁👷
Ile-de-France	58.17	159.89	F	-	-	-	121.89	120.76	F	325.41	159.15	F	24.99	175.11	F	📏👷
Le-Mans	53.53	160.64	F	63.90	165.43	F	52.82	163.53	F	249.61	138.46	F	40.68	176.24	F	☁📏👷
Leuven	10.85	164.76	F	48.83	154.03	F	12.54	173.48	F	69.29	132.33	F	8.42	140.80	F	☁📏👷
Massachusetts1	0.08	0.06	100/ 100/ 100/ 100	0.59	0.36	40/ 72.00/ 96.00/ 96.00	0.12	0.07	100/ 100/ 100/ 100	0.16	0.06	96.00/ 96.00/ 100/ 100	0.20	0.10	76.00/ 84.00/ 84.00/ 84.00	💡☁🌳👷
Massachusetts2	5.80	0.24	0/ 0/ 0/ 100	-	-	-	18.34	9.47	0/ 0/ 0/ 8.33	111.63	168.17	F	6.70	10.76	0/ 0/ 0/ 66.67	🌒👷
Massachusetts3	23.95	25.32	0/ 10.53/ 36.84/ 42.11	498.52	123.61	F	3.91	5.85	0/ 25.00/ 52.27/ 59.09	680.80	104.46	F	18.02	10.24	0/ 0/ 4.55/ 22.73	🌳🌒👷
Massachusetts4	1.57	1.01	0/ 0/ 97.44/ 100	69.55	24.08	0/ 8.11/ 32.43/ 32.43	2.23	1.14	0/ 0/ 100/ 100	40.25	115.17	0/ 0/ 0/ 4.00	10.38	5.70	0/ 0/ 30.77/ 46.15	🌒👷
Melbourne	0.07	0.07	100/ 100/ 100/ 100	0.21	0.22	83.33/ 100/ 100/ 100	0.09	0.16	100/ 100/ 100/ 100	0.16	0.16	100/ 100/ 100/ 100	14.07	0.93	16.67/ 16.67/ 16.67/ 41.67	💡👷
Muehlhausen	0.03	0.07	100/ 100/ 100/ 100	0.32	0.53	80/ 100/ 100/ 100	0.04	0.07	100/ 100/ 100/ 100	0.04	0.11	100/ 100/ 100/ 100	0.04	0.10	100/ 100/ 100/ 100	☁👷
Nouvelle-Aquitaine1	40.23	172.90	F	328.30	150.36	F	34.02	160.82	F	418.74	137.36	F	44.17	170.42	F	❄🍂📏
Nouvelle-Aquitaine2	81.48	126.82	F	35.89	161.41	F	67.65	147.96	F	90.10	150.32	F	28.51	169.94	F	📏
Orleans1	17.42	178.86	F	256.48	149.29	F	33.42	175.46	F	87.40	157.44	0/ 0/ 3.03/ 3.03	25.87	178.74	F	📏
Orleans2	175.99	127.58	F	-	-	-	32.95	165.35	F	359.37	154.26	F	15.30	177.55	F	📏👷
Pays de la Loire	25.32	159.23	F	52.74	156.08	F	34.11	166.72	F	131.31	135.25	0/ 0/ 0/ 4.76	17.56	175.95	F	🌳📏👷
Poing	0.05	0.07	100/ 100/ 100/ 100	0.45	0.61	60/ 85.00/ 100/ 100	0.06	0.07	100/ 100/ 100/ 100	183.46	85.13	20/ 20/ 20/ 20.00	0.08	0.04	85.00/ 85.00/ 85.00/ 85.00	💡🌳👷
Portland	0.13	0.16	100/ 100/ 100/ 100	0.40	0.46	66.67/ 95.24/ 100/ 100	0.11	0.14	100/ 100/ 100/ 100	0.16	0.12	95.24/ 100/ 100/ 100	0.13	0.16	85.71/ 85.71/ 85.71/ 85.71	💡☁👷
Savannah	0.08	0.05	100/ 100/ 100/ 100	0.25	0.28	83.33/ 94.44/ 100/ 100	0.08	0.05	100/ 100/ 100/ 100	0.07	0.05	100/ 100/ 100/ 100	0.09	0.08	94.44/ 94.44/ 94.44/ 94.44	💡👷
Skåne	5.14	2.97	0/ 0/ 50/ 85.00	392.36	120.41	0/ 0/ 0/ 5.00	4.26	3.70	0/ 5.00/ 55.00/ 90.00	609.69	154.13	F	35.93	25.73	F	💡📐
Subcarpathia	0.54	0.46	47.06/ 70.59/ 88.24/ 94.12	6.39	4.09	0/ 0/ 41.18/ 58.82	0.35	0.26	70.59/ 76.47/ 100/ 100	116.17	80.70	11.76/ 11.76/ 29.41/ 35.29	66.44	9.84	5.88/ 5.88/ 11.76/ 11.76	❄🍂👷
Sydney	0.18	0.11	100/ 100/ 100/ 100	1.39	0.82	35.71/ 35.71/ 85.71/ 92.86	0.18	0.15	85.71/ 100/ 100/ 100	2.81	1.86	0/ 25.00/ 75.00/ 75.00	0.18	0.20	64.29/ 71.43/ 71.43/ 71.43	💡☁👷
Thuringia	0.57	0.26	45.45/ 90.91/ 100/ 100	0.95	0.60	18.18/ 54.55/ 100/ 100	0.37	0.25	81.82/ 100/ 100/ 100	431.41	121.14	F	7.03	6.51	0/ 0/ 27.27/ 63.64	💡👷
Tsuru	0.01	0.04	100/ 100/ 100/ 100	0.06	0.30	100/ 100/ 100/ 100	0.03	0.03	100/ 100/ 100/ 100	0.02	0.04	100/ 100/ 100/ 100	0.03	0.04	100/ 100/ 100/ 100	💡☁👷
Washington	1.00	0.42	0/ 50/ 100/ 100	3.96	0.66	0/ 0/ 100/ 100	1.07	1.47	10/ 30/ 100/ 100	0.90	0.68	40/ 50/ 90/ 100	2.46	1.20	0/ 0/ 70.00/ 70.00	💡

Table 2: Tab. 3 from the main paper, provided for reference: Localization performance of the baseline methods on our CrowdDriven benchmark. We report the median position (in meters) and orientation (in degrees) errors, as well as the percentage of test images localized within certain error bounds on the position and orientation errors. Easy, medium, and hard datasets are color-coded in light, standard, and dark gray, respectively. The right side of the table provides information about the type of change between the training and test sequences: illumination: 💡, overcast: ☁, foliage: 🌳, snow: ❄, seasonal: 🍂, day-night: 🌒, small viewpoint: 📐, rain: 💦, strong viewpoint: 📏, man-made changes: 👷. 'F' stands for failure to localize any image within the coarsest precision regime.

Figure 3: Curitiba, Category : Easy

Figure 4: Massachusetts1, Category : Easy

localization. In *2019 International Conference on 3D Vision (3DV)*, pages 513–523. IEEE, 2019. 2

[5] Zuzana Kukelova, Jan Heller, and Andrew Fitzgibbon. Efficient Intersection of Three Quadrics and Applications in

Figure 5: Melbourne, Category : Easy

Figure 6: Muehlhausen, Category : Easy

Figure 7: Poing, Category : Easy

Computer Vision. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016. 1

[6] Karel Lebeda, Juan E. Sala Matas, and Ondřej Chum. Fixing the Locally Optimized RANSAC. In *British Machine Vision Conference (BMVC)*, 2012. 1

[7] Gim Hee Lee, Bo Li, Marc Pollefeys, and Friedrich Fraundorfer. Minimal solutions for the multi-camera pose estimation problem. *IJRR*, 34(7):837–848, 2015. 1

[8] David G. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. *IJCV*, 60(2), 2004. 1

[9] R. Pless. Using Many Cameras as One. In *CVPR*, 2003. 1

Figure 8: Portland, Category : Easy

Figure 9: Savannah, Category : Easy

Figure 10: Subcarpathia, Category : Easy

[10] 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 Conference on Computer Vision and Pattern Recognition*, pages 12716–12725, 2019. 1, 2

[11] Paul-Edouard Sarlin, Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. SuperGlue: Learning feature matching with graph neural networks. In *CVPR*, 2020. 1, 2

[12] Paul-Edouard Sarlin, Ajaykumar Unagar, Mans Larsson, Hugo Germain, Carl Toft, Viktor Larsson, Marc Pollefeys, Vincent Lepetit, Lars Hammarstrand, Fredrik Kahl, and

Figure 11: Sydney, Category : Easy

Figure 12: Tsuru, Category : Easy

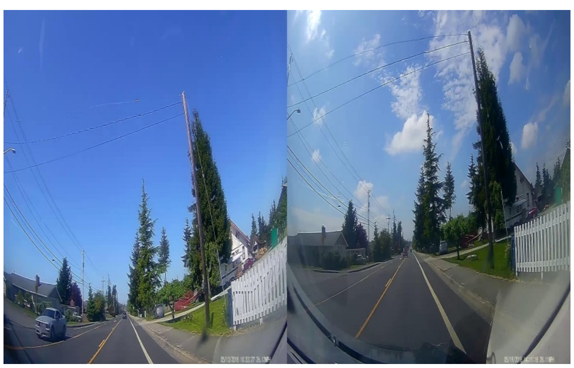

Figure 13: Washington, Category : Medium

Figure 14: Boston1, Category : Hard

Figure 15: Boston2, Category : Medium

Figure 16: Boston3, Category : Medium

Torsten Sattler. Back to the feature: Learning robust camera localization from pixels to pose. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021. 1

[13] Torsten Sattler, Will Maddern, Carl Toft, Akihiko Torii, Lars Hammarstrand, Erik Stenborg, Daniel Safari, Masatoshi Okutomi, Marc Pollefeys, Josef Sivic, Fredrik Kahl, and Tomas Pajdla. Benchmarking 6DOF Urban Visual Localization in Changing Conditions. In *CVPR*, 2017. 2

[14] Chris Sweeney, Victor Fragoso, Tobias Höllerer, and Matthew Turk. gDLS: A Scalable Solution to the Generalized Pose and Scale Problem. In *ECCV*. Springer, 2014. 1

[15] Carl Toft, Daniyar Turmukhambetov, Torsten Sattler, Fredrik Kahl, and Gabriel J Brostow. Single-image depth prediction makes feature matching easier. In *European Conference on Computer Vision*, pages 473–492. Springer, 2020. 2

[16] Jonathan Ventura, Clemens Arth, Gerhard Reitmayr, and Dieter Schmalstieg. A Minimal Solution to the Generalized Pose-and-Scale Problem. In *CVPR*, 2014. 1

[17] Johanna Wald, Torsten Sattler, Stuart Golodetz, Tommaso Cavallari, and Federico Tombari. Beyond controlled environments: 3D camera re-localization in changing indoor scenes.

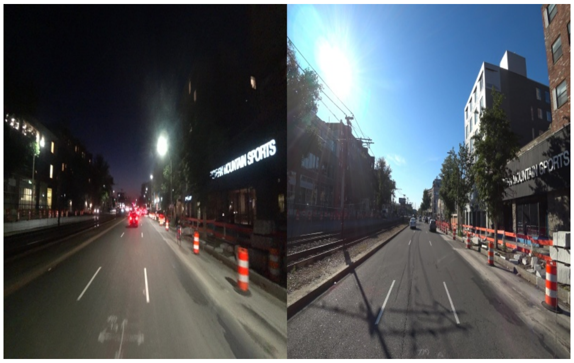

Figure 17: Boston4, Category : Medium

Figure 18: Boston5, Category : Medium

Figure 19: Burgundy2, Category : Medium

In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2020. 1

Figure 20: Massachusetts2, Category : Hard

Figure 21: Massachusetts3, Category : Medium

Figure 22: Massachusetts4, Category : Medium

Figure 23: Skåne, Category : Medium

Figure 24: Thuringia, Category : Easy

Figure 25: Angers1, Category : Difficult

Figure 26: Angers2, Category : Difficult

Figure 27: Besançon2, Category : Difficult

Figure 28: Besançon3, Category : Difficult

Figure 29: Besançon4, Category : Difficult

Figure 30: Brittany, Category : Difficult

Figure 31: Brourges, Category : Difficult

Figure 32: Ile-de-France, Category : Difficult

Figure 33: Le-Mans, Category : Difficult

Figure 34: Leuven, Category : Difficult

Figure 35: Nouvelle-Aquitaine1, Category : Difficult

Figure 36: Nouvelle-Aquitaine2, Category : Difficult

Figure 37: Orleans1, Category : Difficult

Figure 38: Orleans2, Category : Difficult

Figure 39: Pays de la Loire, Category : Difficult