4Seasons: A Cross-Season Dataset for Multi-Weather SLAM in Autonomous Driving

Patrick Wenzel[1,2], Rui Wang[1,2], Nan Yang[1,2], Qing Cheng[2], Qadeer Khan[1,2],
Lukas von Stumberg[1,2], Niclas Zeller[2], and Daniel Cremers[1,2]

[1] Technical University of Munich
[2] Artisense
wenzel@cs.tum.edu

Abstract. We present a novel dataset covering seasonal and challenging perceptual conditions for autonomous driving. Among others, it enables research on visual odometry, global place recognition, and map-based re-localization tracking. The data was collected in different scenarios and under a wide variety of weather conditions and illuminations, including day and night. This resulted in more than 350 km of recordings in nine different environments ranging from multi-level parking garage over urban (including tunnels) to countryside and highway. We provide globally consistent reference poses with up-to centimeter accuracy obtained from the fusion of direct stereo visual-inertial odometry with RTK-GNSS. The full dataset is available at www.4seasons-dataset.com.

Keywords: autonomous driving, long-term localization, SLAM, visual learning, visual odometry

1 Introduction

During the last decade, research on visual odometry (VO) and simultaneous localization and mapping (SLAM) has made tremendous strides [30,12,29,11] particularly in the context of autonomous driving (AD) [9,44,46,28]. One reason for this progress has been the publication of large-scale datasets [7,14,6] tailored for benchmarking these methods. Naturally, the next logical step towards progressing research in the direction of visual SLAM has been to make it robust under dynamically changing and challenging conditions. This includes VO, *e.g.* at night or rain, as well as long-term place recognition and re-localization against a pre-built map. In this regard, the advent of deep learning has exhibited itself to be a promising potential in complementing the performance of visual SLAM [8,22,39,20]. Therefore, it has become all the more important to have datasets that are commensurate with handling the challenges of any real-world environment while also being capable of discerning the performance of state-of-the-art approaches.

To accommodate this demand, we present in this paper a versatile cross-season and multi-weather dataset on a large-scale focusing on long-term localization for autonomous driving. By traversing the same stretch under different
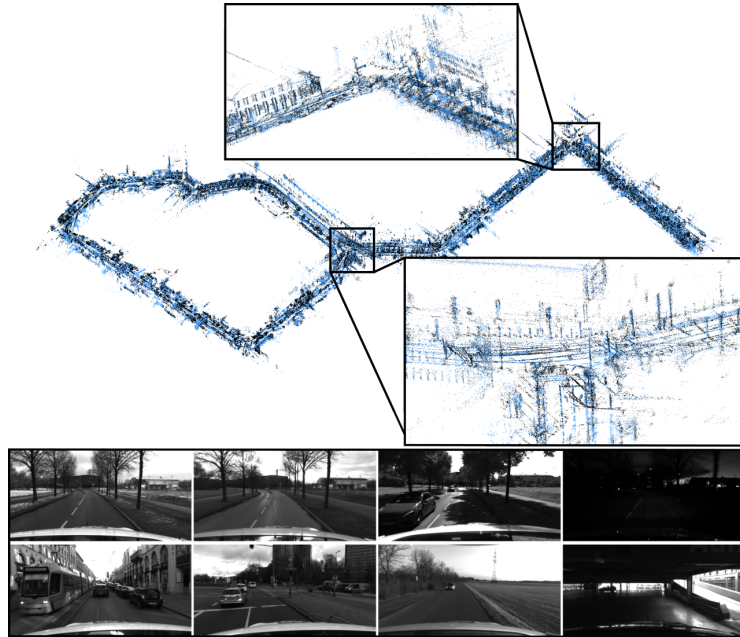
Fig. 1: **Dataset overview.** Top: overlaid maps recorded at different times and environmental conditions. The points from the reference map (black) align well with the points from the query map (blue), indicating that the reference poses are indeed accurate. Bottom: sample images demonstrating the diversity of our dataset. The first row shows a collection from the same scene across different weather and lighting conditions: snowy, overcast, sunny, and night. The second row depicts the variety of scenarios within the dataset: inner city, suburban, countryside, and a parking garage.

conditions and over a long-term time horizon, we capture variety in illumination and weather as well as in the appearance of the scenes. Figure 1 visualizes two overlaid 3D maps recorded at different times as well as sample images of the dataset.

In detail this work adds the following contributions to the state-of-the-art:

– A cross-season/multi-weather dataset for long-term visual SLAM in automotive applications containing more than 350 km of recordings.
– Sequences covering nine different kinds of environments ranging from multi-level parking garage over urban (including tunnels) to countryside and highway.
– Global six degrees of freedom (6DoF) reference poses with up-to centimeter accuracy obtained from the fusion of direct stereo visual-inertial odometry (VIO) with RTK-GNSS.
– Accurate cross-seasonal pixel-wise correspondences to train dense feature representations.

2   Related Work

There exists a variety of benchmarks and datasets focusing on VO and SLAM for AD. Here, we divide these datasets into the ones which focus only on the task of VO as well as those covering different weather conditions and therefore aiming towards long-term SLAM.

2.1   Visual Odometry

The most popular benchmark for AD certainly is KITTI [14]. This multi-sensor dataset covers a wide range of tasks including not only VO, but also 3D object detection, and tracking, scene flow estimation as well as semantic scene understanding. The dataset contains diverse scenarios ranging from urban over countryside to highway. Nevertheless, all scenarios are only recorded once and under similar weather conditions. Ground truth is obtained based on a high-end inertial navigation system (INS).

Another dataset containing LiDAR, inertial measurement unit (IMU), and image data at a large-scale is the Málaga Urban dataset [4]. However, in contrast to KITTI, no accurate 6DoF ground truth is provided and therefore it does not allow for a quantitative evaluation based on this dataset.

Other popular datasets for the evaluation of VO and VIO algorithms not related to AD include [40] (handheld RGB-D), [5] (UAV stereo-inertial), [10] (handheld mono), and [36] (handheld stereo-inertial).

2.2   Long-term SLAM

More related to our work are datasets containing multiple traversals of the same environment over a long period of time. With respect to SLAM for AD the Oxford RobotCar Dataset [27] represents a kind of pioneer work. This dataset consists of large-scale sequences recorded multiple times for the same environment over a period of one year. Hence, it covers large variations in the appearance and structure of the scene. However, the diversity of the scenarios is only limited to an urban environment. Also, the ground truth provided for the dataset is not accurate up-to centimeter-level and therefore, requires additional manual effort to establish accurate cross-sequence correspondences.

The work by [34] represents a kind of extension to [27]. This benchmark is based on subsequences from [27] as well as other datasets. The ground truth of the RobotCar Seasons [34] dataset is obtained based on structure from motion (SfM) and LiDAR point cloud alignment. However, due to inaccurate GNSS measurements [27], a globally consistent ground truth up-to centimeter-level can not be guaranteed. Furthermore, this dataset only provides one reference traversal in the overcast condition. In contrast, we provide globally consistent reference models for all traversals covering a wide variety of conditions. Hence, every traversal can be used as a reference model that allows further research, *e.g.* on analyzing suitable reference-query pairs for long-term localization and mapping.

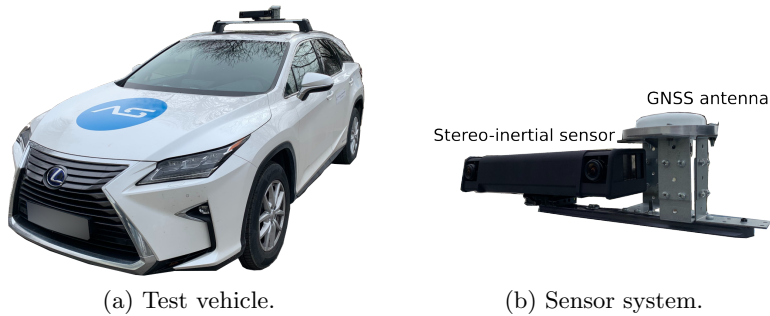(a) Test vehicle.   (b) Sensor system.

Fig. 2: **Recording setup.** Test vehicle and sensor system used for dataset recording. The sensor system consists of a custom stereo-inertial sensor with a stereo baseline of 30 cm and a high-end RTK-GNSS receiver from Septentrio.

2.3   Other Datasets

Examples of further multi-purpose AD datasets which also can be used for VO are [7,45,19,6].

As stated in Section 1, our proposed dataset differentiates from previous related work in terms of being both large-scale (similar to [14]) as well as having high variations in appearance and conditions (similar to [27]). Furthermore, we are providing accurate reference poses based on the fusion of direct stereo VIO and RTK-GNSS.

3   System Overview

This section presents the sensor setup which is used for data recording (Section 3.1). Furthermore, we describe the calibration of the entire sensor suite (Section 3.2) as well as our approach to obtain up-to centimeter-accurate global 6DoF reference poses (Section 3.3).

3.1   Sensor Setup

The hardware setup consists of a custom stereo-inertial sensor for 6DoF pose estimation as well as a high-end RTK-GNSS receiver for global positioning and global pose refinement. Figure 2 shows our test vehicle equipped with the sensor system used for data recording.

Stereo-Inertial Sensor. The core of the sensor system is our custom stereo-inertial sensor. This sensor consists of a pair of monochrome industrial-grade global shutter cameras (Basler acA2040-35gm) and lenses with a fixed focal length of $f = 3.5$ mm (Stemmer Imaging CVO GMTHR23514MCN). The cameras are mounted on a highly-rigid aluminum rail with a stereo baseline of 30 cm.

On the same rail, an IMU (Analog Devices ADIS16465) is mounted. All sensors, cameras, and IMU are triggered over an external clock generated by an field-programmable gate array (FPGA). Here, the trigger accounts for exposure compensations, meaning that the time between the centers of the exposure interval for two consecutive images is always kept constant (1/[frame rate]) independent of the exposure time itself.

Furthermore, based on the FPGA, the IMU is properly synchronized with the cameras. In the dataset, we record stereo sequences with a frame rate of 30 fps. We perform pixel binning with a factor of two and crop the image to a resolution of 800 × 400. This results in a field of view of approximately 77° horizontally and 43° vertically. The IMU is recorded at a frequency of 2000 Hz. During recording, we run our custom auto-exposure algorithm, which guarantees equal exposure times for all stereo image pairs as well as a smooth exposure transition in highly dynamic lighting conditions, as it is required for visual SLAM. We provide those exposure times for each frame.

GNSS Receiver. For global positioning and to compensate drift in the VIO system we utilize an RTK-GNSS receiver (mosaic-X5) from Septentrio in combination with an Antcom Active G8 GNSS antenna. The GNSS receiver provides a horizontal position accuracy of up-to 6 mm by utilizing RTK corrections. While the high-end GNSS receiver is used for accurate positioning, we use a second receiver connected to the time-synchronization FPGA to achieve synchronization between the GNSS receiver and the stereo-inertial sensor.

3.2 Calibration

Aperture and Focus Adjustment. The lenses used in the stereo-system have both adjustable aperture and focus. Therefore, before performing the geometric calibration of all sensors, we manually adjust both cameras for a matching average brightness and a minimum focus blur [18], across a structured planar target in 10 m distance.

Stereo Camera and IMU. For the intrinsic and extrinsic calibration of the stereo cameras as well as the extrinsic calibration and time-synchronization of the IMU, we use a slightly customized version of *Kalibr*[1] [32]. The stereo cameras are modeled using the Kannala-Brandt model [23], which is a generic camera model consisting of in total eight parameters. To guarantee an accurate calibration over a long-term period, we perform a feature-based epipolar-line consistency check for each sequence recorded in the dataset and re-calibrate before a recording session if necessary.

GNSS Antenna. Since the GNSS antenna does not have any orientation but has an isotropic reception pattern, only the 3D translation vector between one

---

[1] github.com/ethz-asl/kalibr

of the cameras and the antenna within the camera frame has to be known. This vector was measured manually for our sensor setup.

3.3   Ground Truth Generation

Reference poses (*i.e.* ground truth) for VO and SLAM should provide high accuracy in both local relative 6DoF transformations and global positioning. To fulfill the first requirement, we extend the state-of-the-art stereo direct sparse VO [44] by integrating IMU measurements [43], achieving a stereo-inertial SLAM system offering average tracking drift around 0.6 % of the traveled distance. To fulfill the second requirement, the poses estimated by our stereo-inertial system are integrated into a global pose graph, each with an additional constraint from the corresponding RTK-GNSS measurement. Our adopted RTK-GNSS system can provide global positioning with up-to centimeter accuracy. The pose graph is optimized globally using the Gauss-Newton method, ending up with 6DoF camera poses with superior accuracy both locally and globally. For the optimization, we make use of the g2o library [25].

One crucial aspect for the dataset is that the reference poses which we provide are actually accurate enough, even though some of the recorded sequences partially contain challenging conditions in GNSS-denied environments. Despite the fact that the stereo-inertial sensor system has an average drift around 0.6 %, this cannot be guaranteed for all cases. Hence, for the reference poses in our dataset, we report whether a pose can be considered to be reliable by measuring the distance to the corresponding RTK-GNSS measurement. Only RTK-GNSS measurements with a reported standard deviation of less than 0.01 m are considered as accurate. For all poses, without corresponding RTK-GNSS measurement we do not guarantee a certain accuracy. Nevertheless, due to the highly accurate stereo-inertial odometry system, these poses still can be considered to be accurate in most cases even in GNSS-denied environments, *e.g.* tunnels or areas with tall buildings.

4   Scenarios

This section describes the different scenarios we have collected for the dataset. The scenarios involve different sequences – ranging from urban driving to parking garage and rural areas. We provide complex trajectories, which include partially overlapping routes, and multiple loops within a sequence. For each scenario, we have collected multiple traversals covering a large range of variation in environmental appearance and structure due to weather, illumination, dynamic objects, and seasonal effects. In total, our dataset consists of nine different scenarios, *i.e.* industrial area, highway, local neighborhood, ring road, countryside, suburban, inner city, monumental site, and multi-level parking garage.

We provide reference poses and 3D models generated by our ground truth generation pipeline (*c.f.* Figure 3) along with the corresponding raw image
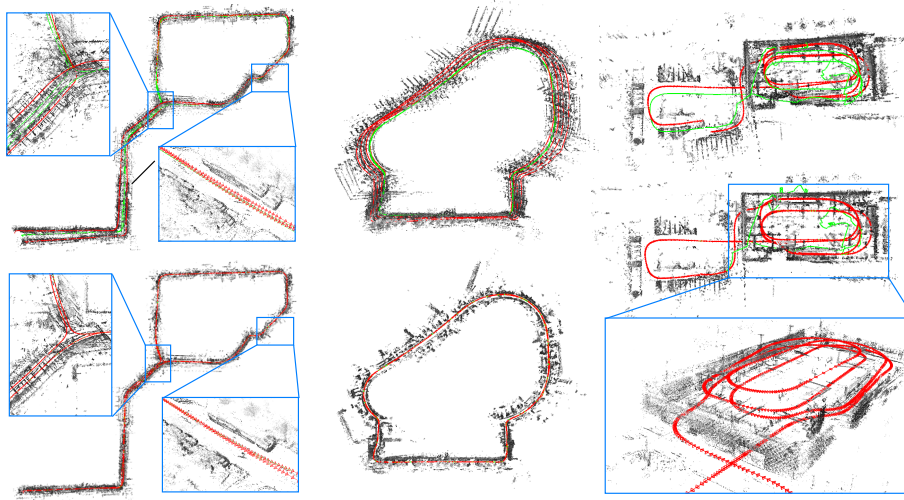
Fig. 3: **3D models of different scenarios contained in the dataset**. The figure shows a loop around an industrial area (left), multiple loops around an area with high buildings (middle), and a stretch recorded in a multi-level parking garage (right). The green lines encode the GNSS trajectories, and the red lines encode the VIO trajectories. Top: shows the trajectories before the fusion using pose graph optimization. Bottom: shows the result after the pose graph optimization. Note that after the pose graph optimization the reference trajectory is well aligned.

frames and raw IMU measurements. Figure 4 shows another example of the optimized trajectory, which depicts the accuracy of the provided reference poses.

The dataset will challenge current approaches on long-term localization and mapping since it contains data from various seasons and weather conditions as well as from different times of the day as shown in the bottom part of Figure 1.

4.1   Ground Truth Validation

The top part of Figure 1 shows two overlaid point clouds from different runs across the same scene. Note that despite the weather and seasonal differences the point clouds align very well. This shows that our reference poses are indeed very accurate. Furthermore, a qualitative assessment of the point-to-point correspondences is shown in Figure 5. The figure shows a subset of very accurate pixel-wise correspondences across different seasons (*autumn/winter*) in the top and different illumination conditions (*sunny/night*) in the bottom. These point-to-point correspondences are a result of our up-to centimeter-accurate global reference poses and are obtained in a completely self-supervised manner. This makes them suitable as training pairs for learning-based algorithms. Recently, there has been an increasing demand for pixel-wise cross-season correspondences
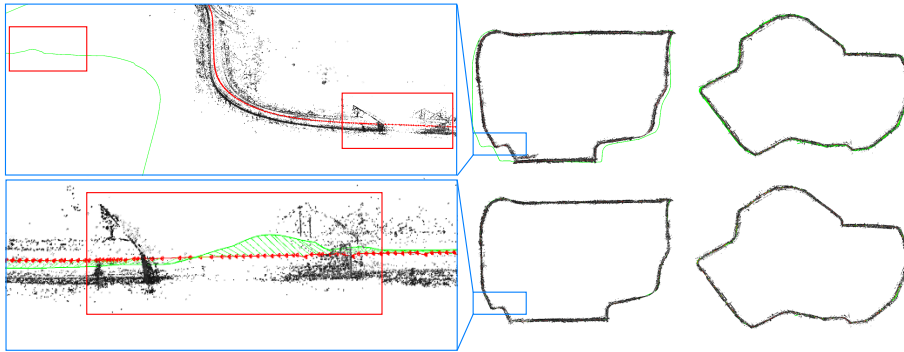
Fig. 4: **Reference poses validation.** This figure shows two additional 3D models of the scenarios collected. Note that these two sequences are quite large (more than 10 km and 6 km, respectively). Top: before the fusion using pose graph optimization. Bottom: results after optimization. The green lines encode the GNSS trajectories, the red lines show the VIO trajectories (before fusion) and the fused trajectories (after fusion). The left part of the figure shows a zoomed-in view of a tunnel, where the GNSS signal becomes very noisy as highlighted in the red boxes. Besides, due to the large size of the sequence, the accumulated tracking error leads to a significant deviation of the VIO trajectory from the GNSS recordings. Our pose graph optimization, by depending globally on GNSS positions and locally on VIO relative poses, successfully eliminates global VIO drifts and local GNSS positioning flaws.

which are needed to learn dense feature descriptors [38,8,33]. However, there is still a lack of datasets to satisfy this demand. The KITTI [14] dataset does not provide cross-seasons data. The Oxford RobotCar Dataset [27] provides cross-seasons data, however, since the ground truth is not accurate enough, the paper does not recommend benchmarking localization and mapping approaches.

Recently, RobotCar Seasons [34] was proposed to overcome the inaccuracy of the provided ground truth. However, similar to the authors of [38], we found that it is still challenging to obtain accurate cross-seasonal pixel-wise matches due to pose inconsistencies. Furthermore, this dataset only provides images captured from three synchronized cameras mounted on a car, pointing to the rear-left, rear, and rear-right, respectively. Moreover, the size of the dataset is quite small and a significant portion of it suffers from strong motion blur and low image quality.

To the best of our knowledge, our dataset is the first that exhibits accurate cross-season reference poses for the AD domain.

5   Tasks

This section describes the different tasks of the dataset. The provided globally consistent 6DoF reference poses for diverse conditions will be valuable to develop
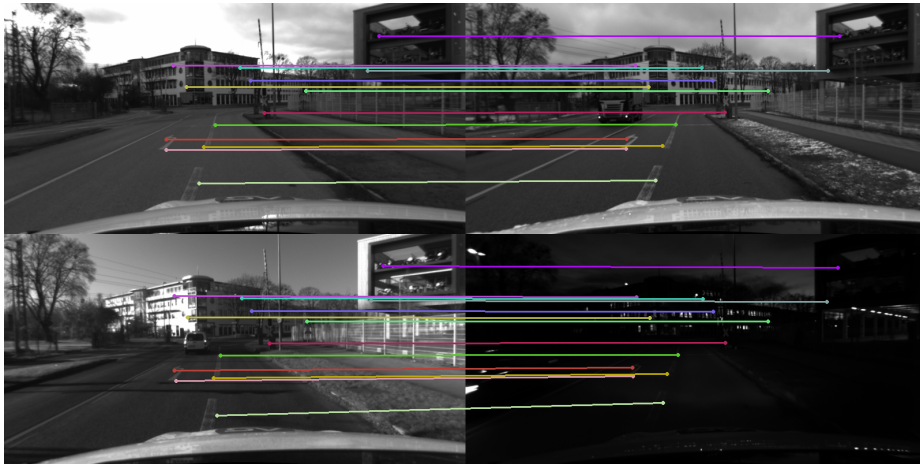
Fig. 5: **Accurate pixel-wise correspondences, making cross-seasonal training possible.** Qualitative assessment of the accuracy of our data collection and geometric reconstruction method for a sample of four different conditions (from top left in clockwise order: *overcast, snowy, night, sunny*) across the same scene. Each same colored point in the four images corresponds to the same geometric point in the world. The cameras corresponding to these images have different poses in the global frame of reference. Please note that the points are not matched but rather a result of our accurate reference poses and geometric reconstruction. This way we are capable of obtaining sub-pixel level accuracy. On average we get more than 1000 of those correspondences per image pair.

and improve the state-of-the-art for different SLAM related tasks. Here the major tasks are robust VO, global place recognition, and map-based re-localization tracking.

In the following, we will present the different subtasks for our dataset.

5.1 Visual Odometry in Different Weather Conditions

VO aims to accurately estimate the 6DoF pose for every frame relative to a starting position. To benchmark the task of VO there already exist various datasets [15,40,10]. All of these existing datasets consist of sequences recorded at rather homogeneous conditions (indoors, or sunny/overcast outdoor conditions). However, especially methods developed for AD use cases must perform robustly under almost any condition. We believe that the proposed dataset will contribute to improving the performance of VO under diverse weather and lighting conditions in an automotive environment. Therefore, instead of replacing existing benchmarks and datasets, we aim to provide an extension that is more focusing on challenging conditions in AD. As we provide frame-wise accurate poses for large portions of the sequences, metrics well known from other benchmarks like

Fig. 6: **Challenging scenes for global place recognition.** Top: two pictures share the same location with different appearances. Bottom: two pictures have similar appearance but are taken at different locations.

absolute trajectory error (ATE) or relative pose error (RPE) [15,40] are also applicable to our data.

5.2   Global Place Recognition

Global place recognition refers to the task of retrieving the most similar database image given a query image [26]. In order to improve the searching efficiency and the robustness against different weather conditions, tremendous progress on global descriptors [21,3,1,13] has been seen. For the re-localization pipeline, visual place recognition serves as the initialization step to the downstream local pose refinement by providing the most similar database images as well as the corresponding global poses. Due to the advent of deep neural networks [37,24,17,41], methods aggregating deep image features are proposed and have shown advantages over classical methods [2,16,31,42].

The proposed dataset is challenging for global place recognition since it contains not only cross-season images that have different appearances but share a similar geographical location but also the intra-season images which share similar appearances but with different locations. Figure 6 depicts example pairs of these scenarios. We suggest to follow the standard metric widely used for global place recognition [2,3,35,16].

5.3   Map-based Re-Localization Tracking

Map-based re-localization tracking [39] refers to the task of locally refining the 6DoF pose between reference images from a pre-built reference map and images

from a query sequence. In contrast to wide-baseline stereo matching, for re-localization tracking, it is also possible to utilize the sequential information of the sequence. This allows us to estimate depth values by running a standard VO method. Those depth estimates can then be used to improve the tracking of the individual re-localization candidates.

In this task we assume to know the mapping between reference and query samples. This allows us to evaluate the performance of local feature descriptor methods in isolation. In practice, this mapping can be found using image retrieval techniques like NetVLAD [2] as described in Section 5.2 or by aligning the point clouds from the reference and query sequences [34], respectively.

Accurately re-localizing in a pre-built map is a challenging problem, especially if the visual appearance of the query sequence significantly differs from the base map. This makes it extremely difficult especially for vision-based systems since the localization accuracy is often limited by the discriminative power of feature descriptors. Our proposed dataset allows us to evaluate re-localization tracking across multiple types of weather conditions and diverse scenes, ranging from urban to countryside driving. Furthermore, our up to centimeter-accurate ground truth allows us to create diverse and challenging re-localization tracking candidates with an increased level of difficulty. By being able to precisely changing the re-localization distances and the camera orientation between the reference and query samples, we can generate more challenging scenarios. This allows us to determine the limitations and robustness of current state-of-the-art methods.

6   Conclusion

We have presented a cross-season dataset for the purpose of multi-weather SLAM, global visual localization, and local map-based re-localization tracking for AD applications. Compared to other datasets, like KITTI [14] or Oxford RobotCar [27], the presented dataset provides diversity in both multiplicities of scenarios and environmental conditions. Furthermore, based on the fusion of direct stereo VIO and RTK-GNSS we are able to provide up-to centimeter-accurate reference poses as well as highly accurate cross-sequence correspondences. One drawback of the dataset is that the accuracy of the reference poses can only be guaranteed in environments with good GNSS receptions. However, due to the low drift of the stereo VIO system, the obtained reference poses are also very accurate in GNSS-denied environments, *e.g.* tunnels, garages, or urban canyons.

We believe that this dataset will help the research community to further understand the limitations and challenges of long-term visual SLAM in changing conditions and environments and will contribute to advance the state-of-the-art. To the best of our knowledge, ours is the first large-scale dataset for AD providing cross-seasonal accurate pixel-wise correspondences for diverse scenarios. This will help to vastly increase robustness against environmental changes for deep learning methods. The dataset is made publicly available to facilitate further research.

References

1. Angeli, A., Filliat, D., Doncieux, S., Meyer, J.A.: Fast and incremental method for loop-closure detection using bags of visual words. IEEE Transactions on Robotics (T-RO) 24(5), 1027–1037 (2008)
2. Arandjelovic, R., Gronat, P., Torii, A., Pajdla, T., Sivic, J.: NetVLAD: CNN architecture for weakly supervised place recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 5297–5307 (2016)
3. Arandjelovic, R., Zisserman, A.: All about VLAD. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1578–1585 (2013)
4. Blanco-Claraco, J.L., Ángel Moreno-Dueñas, F., González-Jiménez, J.: The Málaga urban dataset: High-rate stereo and LiDAR in a realistic urban scenario. International Journal of Robotics Research (IJRR) 33(2), 207–214 (2014)
5. Burri, M., Nikolic, J., Gohl, P., Schneider, T., Rehder, J., Omari, S., Achtelik, M.W., Siegwart, R.: The EuRoC micro aerial vehicle datasets. International Journal of Robotics Research (IJRR) 35(10), 1157–1163 (2016)
6. Caesar, H., Bankiti, V., Lang, A.H., Vora, S., Liong, V.E., Xu, Q., Krishnan, A., Pan, Y., Baldan, G., Beijbom, O.: nuScenes: A multimodal dataset for autonomous driving. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 11621–11631 (2020)
7. Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S., Schiele, B.: The cityscapes dataset for semantic urban scene understanding. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 3213–3223 (2016)
8. Dusmanu, M., Rocco, I., Pajdla, T., Pollefeys, M., Sivic, J., Torii, A., Sattler, T.: D2-Net: A trainable CNN for joint detection and description of local features. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 8092–8101 (2019)
9. Engel, J., Stückler, J., Cremers, D.: Large-scale direct SLAM with stereo cameras. In: Proceedings of the IEEE/RSJ Conference on Intelligent Robots and Systems (IROS). pp. 1935–1942 (2015)
10. Engel, J., Usenko, V., Cremers, D.: A photometrically calibrated benchmark for monocular visual odometry. In: arXiv preprint arXiv:1607.02555 (2016)
11. Engel, J., Koltun, V., Cremers, D.: Direct sparse odometry. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI) 40(3), 611–625 (2017)
12. Engel, J., Schöps, T., Cremers, D.: LSD-SLAM: Large-scale direct monocular SLAM. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 834–849 (2014)
13. Gálvez-López, D., Tardos, J.D.: Bags of binary words for fast place recognition in image sequences. IEEE Transactions on Robotics (T-RO) 28(5), 1188–1197 (2012)
14. Geiger, A., Lenz, P., Stiller, C., Urtasun, R.: Vision meets robotics: The KITTI dataset. International Journal of Robotics Research (IJRR) 32(11), 1231–1237 (2013)
15. Geiger, A., Lenz, P., Urtasun, R.: Are we ready for autonomous driving? the KITTI vision benchmark suite. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 3354–3361 (2012)
16. Gordo, A., Almazán, J., Revaud, J., Larlus, D.: Deep image retrieval: Learning global representations for image search. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 241–257 (2016)

17. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 770–778 (2016)
18. Hu, H., de Haan, G.: Low cost robust blur estimator. In: Proceedings of the IEEE International Conference on Image Processing (ICIP). pp. 617–620 (2006)
19. Huang, X., Cheng, X., Geng, Q., Cao, B., Zhou, D., Wang, P., Lin, Y., Yang, R.: The ApolloScape dataset for autonomous driving. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). pp. 954–960 (2018)
20. Jaramillo, C.: Direct multichannel tracking. In: Proceedings of the International Conference on 3D Vision (3DV). pp. 347–355 (2017)
21. Jégou, H., Douze, M., Schmid, C., Pérez, P.: Aggregating local descriptors into a compact image representation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 3304–3311 (2010)
22. Jung, E., Yang, N., Cremers, D.: Multi-frame GAN: Image enhancement for stereo visual odometry in low light. In: Conference on Robot Learning (CoRL). pp. 651–660 (2019)
23. Kannala, J., Brandt, S.S.: A generic camera model and calibration method for conventional, wide-angle, and fish-eye lenses. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI) 28(8), 1335–1340 (2006)
24. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Neural Information Processing Systems (NeurIPS). pp. 1097–1105 (2012)
25. Kümmerle, R., Grisetti, G., Strasdat, H., Konolige, K., Burgard, W.: g2o: A general framework for graph optimization. In: Proceedings of the IEEE International Conference on Robotics and Automation (ICRA). pp. 3607–3613 (2011)
26. Lowry, S., Sünderhauf, N., Newman, P., Leonard, J.J., Cox, D., Corke, P., Milford, M.J.: Visual place recognition: A survey. IEEE Transactions on Robotics (T-RO) 32(1), 1–19 (2015)
27. Maddern, W., Pascoe, G., Linegar, C., Newman, P.: 1 year, 1000 km: The oxford robotcar dataset. International Journal of Robotics Research (IJRR) 36(1), 3–15 (2017)
28. Mur-Artal, R., Tardós, J.D.: ORB-SLAM2: An open-source SLAM system for monocular, stereo, and RGB-D cameras. IEEE Transactions on Robotics (T-RO) 33(5), 1255–1262 (2017)
29. Mur-Artal, R., Montiel, J.M.M., Tardos, J.D.: ORB-SLAM: a versatile and accurate monocular SLAM system. IEEE Transactions on Robotics (T-RO) 31(5), 1147–1163 (2015)
30. Newcombe, R.A., Lovegrove, S.J., Davison, A.J.: DTAM: dense tracking and mapping in real-time. In: Proceedings of the International Conference on Computer Vision (ICCV). pp. 2320–2327 (2011)
31. Radenović, F., Tolias, G., Chum, O.: Fine-tuning CNN image retrieval with no human annotation. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI) 41(7), 1655–1668 (2018)
32. Rehder, J., Nikolic, J., Schneider, T., Hinzmann, T., Siegwart, R.: Extending kalibr: Calibrating the extrinsics of multiple IMUs and of individual axes. In: Proceedings of the IEEE International Conference on Robotics and Automation (ICRA). pp. 4304–4311 (2016)
33. Revaud, J., Weinzaepfel, P., de Souza, C.R., Humenberger, M.: R2D2: repeatable and reliable detector and descriptor. In: Neural Information Processing Systems (NeurIPS). pp. 12405–12415 (2019)

34. Sattler, T., Maddern, W., Toft, C., Torii, A., Hammarstrand, L., Stenborg, E., Safari, D., Okutomi, M., Pollefeys, M., Sivic, J., Kahl, F., Pajdla, T.: Benchmarking 6DOF outdoor visual localization in changing conditions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 8601–8610 (2018)
35. Sattler, T., Weyand, T., Leibe, B., Kobbelt, L.: Image retrieval for image-based localization revisited. In: Proceedings of the British Machine Vision Conference (BMVC) (2012)
36. Schubert, D., Goll, T., Demmel, N., Usenko, V., Stückler, J., Cremers, D.: The TUM VI benchmark for evaluating visual-inertial odometry. In: Proceedings of the IEEE/RSJ Conference on Intelligent Robots and Systems (IROS). pp. 1680–1687 (2018)
37. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. In: Proceedings of the International Conference on Learning Representations (ICLR) (2015)
38. Spencer, J., Bowden, R., Hadfield, S.: Same features, different day: Weakly supervised feature learning for seasonal invariance. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 6459–6468 (2020)
39. von Stumberg, L., Wenzel, P., Khan, Q., Cremers, D.: GN-Net: The gauss-newton loss for multi-weather relocalization. IEEE Robotics and Automation Letters (RA-L) 5(2), 890–897 (2020)
40. Sturm, J., Engelhard, N., Endres, F., Burgard, W., Cremers, D.: A benchmark for the evaluation of RGB-D SLAM systems. In: Proceedings of the IEEE/RSJ Conference on Intelligent Robots and Systems (IROS). pp. 573–580 (2012)
41. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1–9 (2015)
42. Tolias, G., Sicre, R., Jégou, H.: Particular object retrieval with integral max-pooling of CNN activations. arXiv preprint arXiv:1511.05879 (2015)
43. Von Stumberg, L., Usenko, V., Cremers, D.: Direct sparse visual-inertial odometry using dynamic marginalization. In: Proceedings of the IEEE International Conference on Robotics and Automation (ICRA). pp. 2510–2517 (2018)
44. Wang, R., Schwörer, M., Cremers, D.: Stereo DSO: Large-scale direct sparse visual odometry with stereo cameras. In: Proceedings of the International Conference on Computer Vision (ICCV). pp. 3903–3911 (2017)
45. Wang, S., Bai, M., Mattyus, G., Chu, H., Luo, W., Yang, B., Liang, J., Cheverie, J., Fidler, S., Urtasun, R.: TorontoCity: Seeing the world with a million eyes. In: Proceedings of the International Conference on Computer Vision (ICCV) (2017)
46. Yang, N., Wang, R., Stückler, J., Cremers, D.: Deep virtual stereo odometry: Leveraging deep depth prediction for monocular direct sparse odometry. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 817–833 (2018)