2023 IEEE International Conference on Robotics and Automation (ICRA)
Workshop on Active Methods in Autonomous Navigation
29 May – 2 June, 2023, London, UK

Low Visibility SLAM: A Dataset for Evaluating Simultaneous Localisation and Mapping under Varying Visibility Conditions

Christopher J. Holder, and Muhammad Shafique, *Senior Member, IEEE*

***Abstract*— Simultaneous localisation and mapping of unexplored environments is a challenging problem in robotics, and while several recent works have demonstrated highly accurate and robust approaches, the additional challenge of performing this task under severely reduced visibility is relatively unresolved. With a view towards working to address this, we introduce the Low Visibility SLAM (LVS) dataset, comprising a total of 36 image sequences paired with ground truth pose data captured in simulated environments under varying levels of aerosol density (i.e. fog). In this work we describe the LVS dataset and use it to evaluate a state-of-the-art deep learning-based SLAM algorithm, demonstrating that this is a challenging problem for future works in the area to address. The LVS dataset is available for download at http://kaggle.com/datasets/ebrainlab/lowvisslam.**

I. Introduction

Simultaneous Localisation and Mapping (SLAM) involves an agent simultaneously building a map as it explores an unknown environment while tracking its pose within that environment, and is a significant challenge in the field of robotics, particularly in applications such as autonomous navigation, augmented reality, and 3D reconstruction. Over the past few decades, there have been numerous efforts to develop efficient SLAM algorithms [1][2][3] and datasets that can be used for benchmarking and validation. In this context, the availability of high-quality SLAM datasets plays a crucial role in advancing the state-of-the-art.

The KITTI dataset [4] has been widely used to evaluate SLAM methods in realistic urban driving scenarios and includes several sensor modalities, such as stereo cameras, LiDAR, and GPS, and provides ground truth poses for evaluation. The EuRoC dataset [5] includes various indoor and outdoor scenes, and provides both stereo cameras and IMU data for evaluating SLAM algorithms. The TUM-VI dataset [6] includes a wide range of indoor and outdoor scenes, captured using a synchronized camera and IMU setup, and provides ground truth poses for evaluation.

More recently, photorealistic simulated environments have enabled datasets encompassing a broader variety of scenarios and more accurate ground truth data than can be collected in real environments. Tartanair [7] comprises 30 environments built in the Unreal Engine, encompassing both indoor and outdoor scenarios and difficult lighting and weather conditions.

SLAM algorithms have been demonstrated utilizing a variety of sensor modalities, including monocular imagery, depth sensors, and LIDAR, however in almost all cases poor visibility conditions, such as fog or smoke, can severely impact the capability to extract and track environmental features, drastically reducing accuracy of both mapping and localisation. In this work, we aim to take a step towards addressing this limitation with the Low Visibility SLAM (LVS) dataset, intended to assess the impact of varying densities of visible aerosols, such as fog or smoke, on the performance of visual SLAM algorithms utilising monocular imagery. The dataset comprises 36 sequences of colour images along with corresponding timestamps and camera poses captured in simulated environments rendered in the Unity 3D engine.

Figure 1. Example image from our Forest environment in conditions of good visibility (top) and dense fog (bottom).

The 36 sequences encompass 3 paths of varying lengths through each of 3 environments, with each path repeated with 4 levels of visible aerosol density, enabling direct evaluation of the impact of visibility conditions on SLAM algorithms. Figure 1 shows a sample from the Forest environment of our dataset with good visibility and with dense fog.

II. Dataset

A. Overview

Three simulated environments are used in the dataset: a large indoor environment named "Temple", a small indoor

C.J. Holder and M. Shafique are with the Division of Engineering, New York University Abu Dhabi, P.O. Box 129188, Saadiyat Island, Abu Dhabi, United Arab Emirates. E-mail: {chris.holder, ms12713}@nyu.edu.

Figure 2. Test image used to measure visibility in foggy environments. The visibility distance in fog (bottom) is computed as the distance at which the contrast between the white text and black background of the test image drops below 5% of that measured with no fog (top)

environment named "Cabin" and a large outdoor environment named "Forest". Sequences are generated from three distinct paths through each of these environments, labelled "Short", "Medium" and "long". Each path is created via the manual movement of a virtual camera through the environment, with free movement in the x, y and z axes, and free rotation around the x and y axes (pitch and yaw), while roll remains fixed at zero.

For each path, four sequences are generated with visible aerosol levels of "none", "light", "mid" and "dense". In the two indoor environments an exponential depth-based screen-space fog is used, while in the outdoor environment volumetric fog is used.

Table I lists the three environments used, their dimensions and the lengths of the three paths taken through each in time taken in seconds and the number of images each comprises. Table II lists details of the three low visibility sequences generated for each path (each path also includes a fourth sequence without reduced visibility). In environments where depth-based fog is used, the density value refers to the variable v in the exponential fog equation: $f = 1/e^{dv}$, where f refers to the visibility of an object at a distance of d from the camera. Volumetric fog is computed as described in [8], with the density values listed in Table II referring to the fog attenuation distance in metres.

We measure the visibility in each environment as the distance in metres from the test image shown Figure 2 at which the contrast between the light and dark regions reaches 5% of its value under perfect visibility. Figure 3 shows a sample image from each environment at each visibility level.

TABLE I. DETAILS OF THE ENVIRONMENTS & PATHS IN THE LVS DATASET

Environment	Temple			Cabin			Forest		
Type	Large indoor			Small indoor			Outdoor		
Dimensions	40m × 40m			10m × 10m			3000m × 3000m		
Path	Short	Medium	Long	Short	Medium	Long	Short	Medium	Long
Time (s)	25	49	94	20	52	82	32	60	126
Images	107	204	394	86	219	341	133	252	528

TABLE II. DETAILS OF THE VISIBILITY CONDITIONS OF EACH SEQUENCE IN THE LVS DATASET

Environment	Temple			Cabin			Forest		
Simulated Aerosol Type	Depth based screen-space fog			Depth based screen-space fog			Volumetric fog		
Density	Light	Mid	Dense	Light	Mid	Dense	Light	Mid	Dense
	0.05	0.1	0.15	0.1	0.2	0.3	160	80	40
Visibility (m)	34.3	17.5	11.7	17.5	8.9	6	198	126	87

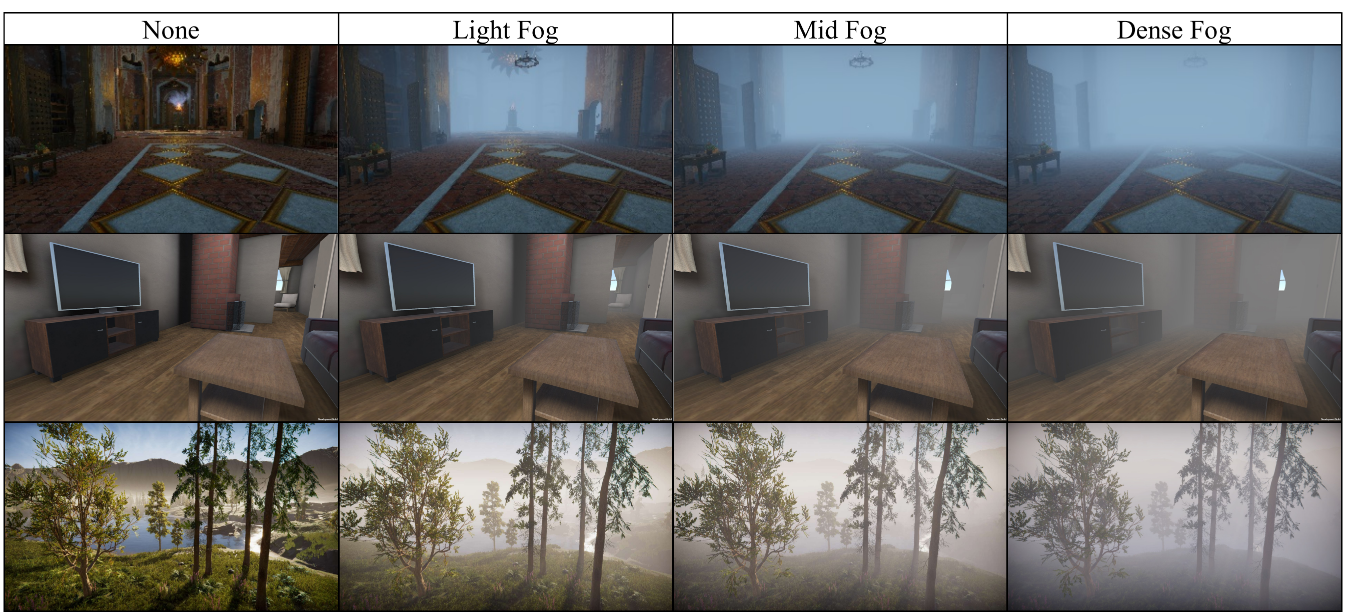

Figure 3. Example images from each environment of the LVS dataset illustrating the four aerosol density levels simulated. Top to bottom: Temple, Cabin, and Forest environments.

TABLE III. EVALUATION OF DROID SLAM ON ALL 36 SEQUENCES OF OUR LVS DATASET

	Temple						Cabin						Forest					
	Short		Medium		Long		Short		Medium		Long		Short		Medium		Long	
	RMSE	Max	RMSE	Max	RMSE	Max	RMSE	Max	RMSE	Max	RMSE	Max	RMSE	Max	RMSE	Max	RMSE	Max
No Fog	5.35	11.59	4.42	10.53	12.44	30.66	0.85	2.69	1.05	1.61	1.00	2.05	9.38	20.53	254.04	778.52	26.89	70.80
Light Fog	5.47	12.39	4.42	10.54	12.30	30.48	0.86	2.69	1.05	1.61	1.00	2.05	9.37	20.50	224.81	606.39	27.82	70.35
Mid Fog	5.05	11.39	4.46	10.65	12.62	30.19	0.86	2.69	1.05	1.60	1.00	2.05	8.99	20.29	219.63	595.56	30.21	71.13
Dense Fog	4.45	11.42	10.54	30.91	12.72	32.87	0.85	2.69	1.05	1.62	1.00	2.05	26.13	97.03	218.09	592.86	134.80	300.66

B. Format

The LVS dataset is available for public download at http://kaggle.com/datasets/ebrainlab/lowvisslam. This section describes the structure and format of the downloadable dataset. For each of the 9 paths that make up the dataset, there is a directory named in the format "[Environment][Path Length]" e.g. "TempleMedium". Inside this directory is a text file named "frames.txt" and 4 subdirectories, each containing one sequence of images, with the subdirectory name referring to the aerosol density value of that sequence e.g. "0.1". These 4 sequences are identical in every way apart from the aerosol density level, therefore the timestamps and poses listed in the single "frames.txt" file applies to all sequences.

The file "frames.txt" contains one line for each frame captured along the path in the format "[Timestamp] [Position] [Rotation]", where Timestamp is the time in milliseconds since the simulation was started, Position is the global camera position within the environment in metres in the format "[px] [py] [pz]", and Rotation is a quaternion that describes global camera rotation in the format "[qx] [qy] [qz] [qw]".

Images are in stored in lossless png format in RGB colour with a resolution 1920×1080 pixels, and were captured at a rate of approximately 5 frames per second. Image filenames correspond to the timestamps listed in the "frames.txt" file. The dataset root folder also contains a camera calibration file in yaml format which is applicable to all sequences.

III. EXPERIMENTS

We perform an evaluation on our dataset using the state-of-the-art SLAM method DROID SLAM [3], which is a deep learning-based approach capable of monocular, stereoscopic and RGBD SLAM. The results of this evaluation are listed in Table III. In each case, a baseline is set for each path using the sequence with no fog, which performance on light, mid and dense -fog sequences are subsequently compared to. In each case, we record the root mean squared error (RMSE) over the sequence as well as the maximum error, in metres.

IV. CONCLUSION

This work describes our novel Low Visibility SLAM (LVS) dataset, which we introduce for the evaluation of the performance of SLAM algorithms under poor visibility conditions. The dataset comprises 36 image sequences with ground truth localisation data, encompassing 9 routes through 3 simulated environments, each repeated under 4 levels of visibility. We have evaluated a state-of-the-art SLAM method on our dataset, demonstrating that localisation and mapping under conditions of poor visibility is a challenging problem for future work to address.

ACKNOWLEDGEMENTS

This work was partially supported by the NYUAD Center for Artificial Intelligence and Robotics (CAIR), funded by Tamkeen under the NYUAD Research Institute Award CG010. This work was also partially supported by the project "eDLAuto: An Automated Framework for Energy-Efficient Embedded Deep Learning in Autonomous Systems", funded by the NYUAD Research Enhancement Fund (REF).

REFERENCES

[1] Czarnowski, J., Laidlow, T., Clark, R. and Davison, A. J. (2020), 'Deepfactors: Real-time probabilistic dense monocular slam', IEEE Robotics and Automation Letters 5(2), 721--728.

[2] Campos, C., Elvira, R., Rodriguez, J. J. G., Montiel, J. M. and Tardys, J. D. (2021), 'Orb-slam3: An accurate open-source library for visual, visual–inertial, and multimap slam', IEEE Transactions on Robotics 37(6), 1874--1890.

[3] Teed, Z. and Deng, J. (2021), 'Droid-slam: Deep visual slam for monocular, stereo, and rgb-d cameras', Advances in neural information processing systems 34, 16558--16569.

[4] Geiger, A., Lenz, P., Stiller, C. and Urtasun, R. (2013), 'Vision meets robotics: The kitti dataset', The International Journal of Robotics Research 32(11), 1231--1237.

[5] Burri, M., Nikolic, J., Gohl, P., Schneider, T., Rehder, J., Omari, S., Achtelik, M. W. and Siegwart, R. (2016), 'The EuRoC micro aerial vehicle datasets', The International Journal of Robotics Research 35(10), 1157--1163.

[6] Schubert, D., Goll, T., Demmel, N., Usenko, V., Stückler, J. and Cremers, D. (2018), The TUM VI benchmark for evaluating visual-inertial odometry, in '2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)', pp. 1680--1687.

[7] Wang, W., Zhu, D., Wang, X., Hu, Y., Qiu, Y., Wang, C., Hu, Y., Kapoor, A. and Scherer, S. (2020), Tartanair: A dataset to push the limits of visual slam, in '2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)', pp. 4909--4916.

[8] Wronski, B. (2018), 'Volumetric fog and lighting', GPU Pro 360, 321.