MESA: A Multi-Environment Synthetic Adaptation Dataset for visual SLAM Evaluation and Feature Learning

Anastasios Agakidis*
Production and Management Engineering
Democritus University of Thrace
Xanthi, Greece
aagakidi@pme.duth.gr

Panagiotis Bakirtzis*
Electrical and Computer Engineering
Democritus University of Thrace
Xanthi, Greece
pbakirtz@ee.duth.gr

Loukas Bampis
Electrical and Computer Engineering
Democritus University of Thrace
Xanthi, Greece
lbampis@ee.duth.gr

***Abstract*—Accurate localization and mapping are fundamental properties for any autonomous system, commonly addressed concurrently by means of Simultaneous Localization and Mapping (SLAM) techniques. However, many SLAM algorithms struggle under varying environmental conditions, such as lighting changes and atmospheric disturbances. In this work, we introduce a novel synthetic dataset designed to facilitate both visual SLAM benchmarking and feature learning in diverse environmental conditions. Our dataset consists of over *140,000* high-resolution images recorded in 5 different worlds developed in NVIDIA Omniverse and Unreal Engine 5. Multiple environmental conditions, including extreme variations in lighting, fog, and dynamic objects, were simulated, with a mobile platform (camera) performing identical trajectories in order to ensure perfect alignment of the corresponding frames at pixel-level. To demonstrate the dataset's usability, we perform a benchmarking case-study evaluating ORB-SLAM3 and HFNET-enhanced ORB-SLAM3. Experimental results validate the dataset's potential for SLAM evaluation, robustness testing, and learning-based adaptation.**

***Index Terms*—synthetic dataset, SLAM benchmarking, autonomous navigation, photorealistic environments, feature extraction**

I. Introduction

In order for an autonomous robot to operate in complex real-world environments accurate perception is required. Tasks such as Simultaneous Localization and Mapping (SLAM) and Visual Odometry (VO) usually rely on the ability to detect, track, and match visual features under a variety of environmental conditions.

Datasets captured in the real-world (e.g., KITTI [1], TUM RGB-D [2]) provide valuable benchmarks. However, they do not always offer the controlled and repeatable testing conditions needed for SLAM and VO performance evaluation across different conditions, especially dynamic ones. Thus, records of synthetic environments have been developed over the last decade to fill this gap [3]–[6]. Such environments allow for a wide variety of robotics applications to be validated and tested since they offer controlled and specifically defined characteristics for multiple environments and conditions, ease and reduced cost for collecting excessive amounts of data, and the ability of a simulated agent to interact with its environment.

In this paper we present MESA[1], a novel large-scale synthetic dataset developed using the state-of-the-art rendering engines NVIDIA Omniverse[2] and Unreal Engine 5 (UE5)[3]. Our proposal follows the principles of *Illumination Conditions Adaptation* (ICA) [5], which showed that learned-based feature detectors and descriptors, can be effectively trained for challenging scenarios using synthetic datasets, which can generalize in real-world benchmarks. MESA is tailored for SLAM benchmark and feature extraction research and provides multi-condition image sequences, offering both benchmarking and training capabilities under controlled lighting, weather, and atmospheric variations. In addition, the recorded sequences were captured under identical camera poses across the different conditions, providing pixel-level correspondences, which allows them to be used as condition augmentations of the same scene within the scope of Deep Learning architectures.

The remainder of this paper is organized as follows: Section II reviews existing datasets and their applications in SLAM and feature learning. Section III details the dataset generation process, including the rendering pipeline and environmental variations. Section IV presents our experimental setup and results, benchmarking SLAM and feature extraction techniques across different environmental conditions. Finally, Section V summarizes our findings and discusses future directions for synthetic dataset development and evaluation in robotics and computer vision.

II. Related Work

The available literature for real-world and synthetic datasets dedicated to SLAM applications can adhere to the following three principles: **dynamic characteristics**, referring to the presence of moving elements and human interactions that increase environmental complexity and test the responsiveness

*These authors contributed equally to this work.

[1] https://learner.ee.duth.gr/datasets.html

[2] https://developer.nvidia.com/nvidia-omniverse-platform

[3] https://www.unrealengine.com

of SLAM and path planning systems; **illumination conditions**, which assess a system's adaptability to diverse visibility scenarios and highlight the algorithm's performance under challenging lighting; and **data quality**, emphasizing the need for large-scale, diverse datasets that support robust validation, evaluation, and debugging of SLAM systems, particularly the more sensitive Visual SLAM methods.

Based on the above distinction the **TUM RGB-D** [2] refers to a real world dataset that covers scenarios with sequences of static and dynamic environments, addressing challenges like dynamic object tracking, occlusion handling, and motion blur. Most sequences are captured under controlled lighting conditions with minimal natural light interference and few shadows and reflections. The ground truth trajectories are recorded with motion capture systems, providing almost accurate pose estimations. **EuRoC MAV** [7] is also a real world dataset which features high-quality challenging indoor sequences ideal for Visual-Inertial SLAMs. The use of a drone in closed spaces produces frames with fast, unpredictable, and jerky movements, creating motion blur, rolling shutter distortions, and sometimes even ground truth drift. The dataset consists mostly of static environments constructed with well-lit lighting, without offering enough dark indoor areas.

In contrast to the above, the **ICL-NUIM** [8] is a virtual dataset created using a physics-based rendering engine. It consists of predefined, smooth, and continuous camera trajectories through two different indoor scenes. The scenes are completely static with uniformly distributed synthetic lighting. The ground truth accuracy is 100% due to the synthetic nature of the dataset. Similarly, the **Replica** [9] is a high-fidelity synthetic dataset featuring predefined camera trajectories that mimic robotic navigation with smooth, realistic movements through a fully static environment. It provides a realistic lighting simulation with soft shadows, reflections, and indirect illumination, although without time-of-day variations. The ground truth data is noise-free. Finally, the **Virtual KITTI Dataset** [3] is a synthetic counterpart of the real KITTI dataset. It follows a predefined path with smooth forward motion, turns, and stops. Dynamic obstacles like cars, pedestrians, and cyclists are added to simulate realistic traffic flow. It includes varying weather and lighting conditions, introducing visibility challenges like fog and rain.

III. Methodology

To support robust evaluation and training of SLAM and learned feature extraction methods under varying environmental conditions, a synthetic dataset was constructed featuring photorealistic indoor sequences of four different virtual worlds, formulating five distinct subsets as listed in Table I. The sequences were recorded from each world and consist of frames captured along identical camera trajectories rendered under different conditions, such as changes in lighting and atmospheric effects. To achieve this, we designed fixed camera motion scripts that maintained consistent pose and transformation across multiple environmental configurations. This design allows the evaluation of SLAM approaches by removing complexity of the camera's motion or the environment's structure and isolating only effect of different lighting or attenuation conditions. However, our approach also enables the use of domain adaptation techniques such as ICA [5], where feature detectors can be trained by leveraging features extracted from well-lit or clear-condition frames as pseudo-ground-truth for their corresponding degraded-condition counterparts (e.g., nighttime, foggy, or smoke-filled environments). By aligning identical camera trajectories across diverse conditions, MESA facilitates consistent supervision between easy and challenging scenes, supporting the development of feature extraction models that maintain reliability even under extreme visual degradation.

TABLE I: Dataset subsets and corresponding number of sequences recorded for different conditions, together with their respective frames count.

Subset	#Seq.	#Frames/Seq.
Omniverse Warehouse	4	15,000
Omniverse Warehouse (Dynamic)	1	30,000
Laboratory Office	4	7,500
Interior Apartment	2	3,580
UE5 Warehouse	2	8,000

We focused on cluttered indoor scenes containing a wide range of geometric features and high quality textures. The camera intrinsic parameters were also exported to support geometric benchmarking or training scenarios, while the output resolution and frame rate were kept the same, for each individual sequence.

The sequences generated in NVIDIA Omniverse were rendered using a GeForce RTX 4090 GPU while those in UE5 were rendered with a GeForce RTX 3080 GPU. All sequences share the following rendering settings:

- **Output Resolution:** 1920 × 1080
- **Frame Rate:** 30 FPS
- **Output Format:** PNG sequence, MP4
- **Camera:** RGB camera with a predefined motion path

All sequences are provided with ground-truth trajectories, enabling quantitative evaluation through established SLAM metrics such as Absolute Trajectory Error (ATE RMSE) and Relative Pose Error (RPE Mean).

A. NVIDIA Omniverse Sequences

NVIDIA Omniverse provides a high-fidelity simulation environment leveraging real-time physics-based rendering through its built-in RTX renderer by utilizing ray tracing directly on NVIDIA RTX GPUs. The platform is capable of providing fully dynamic illumination without any light baking, by computing separate passes such as direct lighting with ray-traced shadows, ambient occlusion, and reflections.

A single large-scale indoor industrial warehouse environment was created using prebuilt assets, expanded within the NVIDIA Omniverse platform. The warehouse environment was selected for its relevance to real-world robotics applications and its potential to simulate complex indoor settings. It

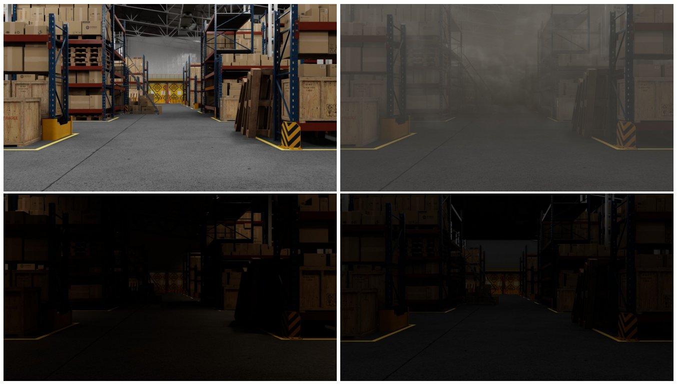


Fig. 1: Representative examples of the *Omniverse Warehouse* environment rendered under four different conditions. From top-left to bottom-right: (i) bright, (ii) smoke/fog, (iii) low-light, and (iv) dark. All sequences use a consistent camera path to ensure lighting remains the only altered variable.

features wide and narrow aisles, pallets, shelving units, loading docks, and storage zones.

Variations of the warehouse environment include the (i) bright, (ii) smoke/fog, (iii) low-light, and (iv) dark sequences, which consist of static elements. In addition, a dynamic elements sequence was developed that features moving obstacles, occlusions, varying illumination, and dynamic actors such as forklifts, pallet jacks, falling objects, liquid drops, and real-time lighting changes.

Figure 1 contains representative examples of all four sequences recorded from the static *Omniverse Warehouse*, while Fig. 2 illustrates a top-down view of the environment, overlaid with the corresponding camera trajectory. Furthermore, Fig. 3 depicts representative examples of the dynamic elements *Omniverse Warehouse* case from different instances of the dynamic events that contains. Finally, Fig. 4 presents the respective camera trajectory followed for the formulation of this sequence.

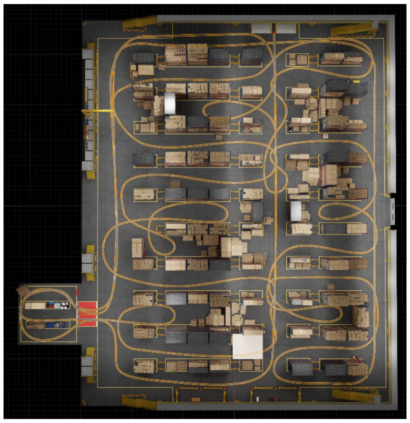


Fig. 2: Top-down visualization with overlaid ground-truth camera trajectory of the *Omniverse Warehouse* environment, in the static elements sequences.

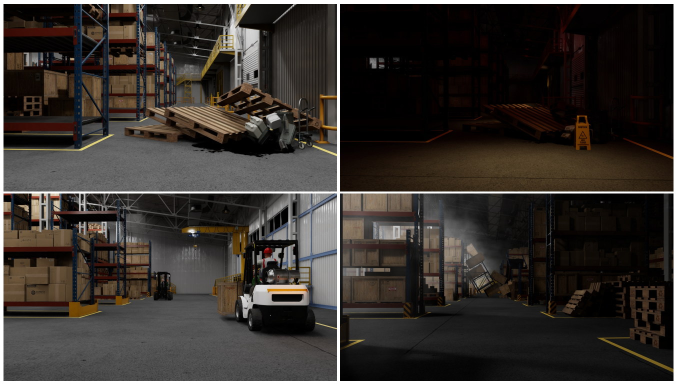


Fig. 3: Visualization of the *dynamic elements* sequence of the *Omniverse Warehouse* environment.

B. Unreal Engine 5 Sequences

UE5 was used to generate additional high-quality indoor sequences, leveraging its advanced rendering technologies to simulate realistic yet controlled lighting conditions. UE5 uses *Lumen*, a fully dynamic global illumination system that provides real-time indirect lighting and reflections. Unlike traditional baked lighting solutions, *Lumen* adapts to dynamic scene changes and supports soft, natural light bounces and accurate shadowing, making it ideal for simulating scenarios such as day-to-night transitions and varying visibility conditions.

Three different indoor environments were developed, each with a unique camera trajectory capturing continuous motion with fixed transformation and rotation parameters under multiple environmental conditions:

- **Laboratory Office:** A larger, research-themed workspace captured under five different conditions: (i) bright, (ii) daylight with smoke/fog, (iii) low-light, and (iv) dark.
- **Interior Apartment:** A compact, office-style living space rendered under two lighting conditions: bright and low-light.

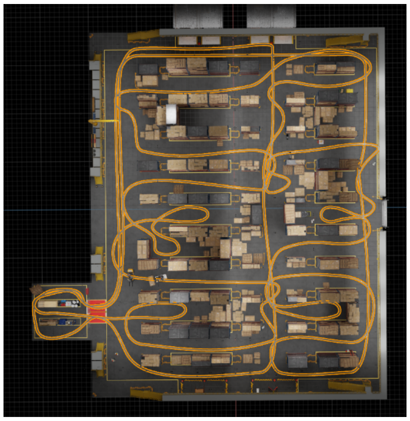


Fig. 4: Top-down visualization with overlaid ground-truth camera trajectory of the *Omniverse Warehouse* environment's dynamic elements sequence.

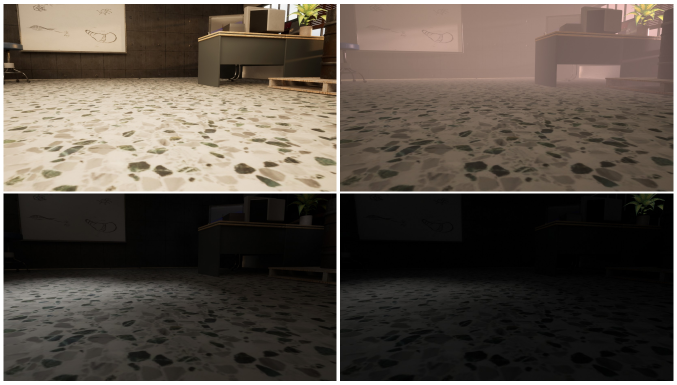


Fig. 5: Representative examples of the *Laboratory Office* environment rendered under four different lighting conditions in UE5. From top-left to bottom-right: (i) bright, (ii) daylight with smoke/fog, (iii) low-light, and (iv) dark.

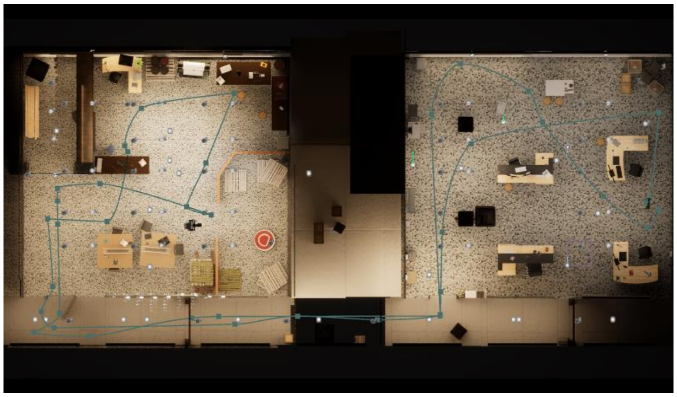


Fig. 6: Top-down visualization with overlaid ground-truth camera trajectory of the *Laboratory Office* environment.

- **UE5 Warehouse:** An industrial warehouse environment, densely populated with storage items and obstacles, rendered in bright and and low-light conditions.

IV. Experiments

To assess the applicability of the MESA dataset for evaluating SLAM performance under varying illumination and environmental conditions, we conducted a series of benchmark experiments using two different SLAM systems:

- **ORB-SLAM3**: a state-of-the-art feature-based visual SLAM system [10] and
- **HF-Net + ORB-SLAM3**: a modified version of ORB-SLAM3, where traditional ORB features are replaced with HF-Net [11] keypoints and descriptors, integrated through the HFNet-SLAM framework[4].

Representative frames of the *Laboratory office* and the top-down visualization along with the camera path are shown in Fig. 5 and Fig. 6, respectively, while the same is done with *UE5 Warehouse* and *Interior Apartment* in Fig. 7 and Fig. 8.

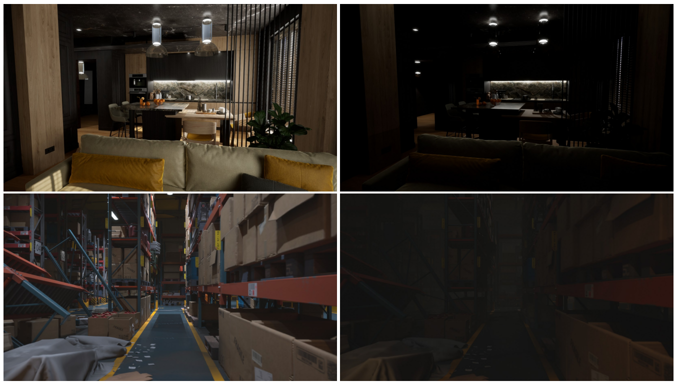


Fig. 7: The left column shows scenes under bright lighting, while the right column presents their darkened variants. The top row corresponds to the *Interior Apartment* environment, and the bottom row to the *UE5 Warehouse* environment.

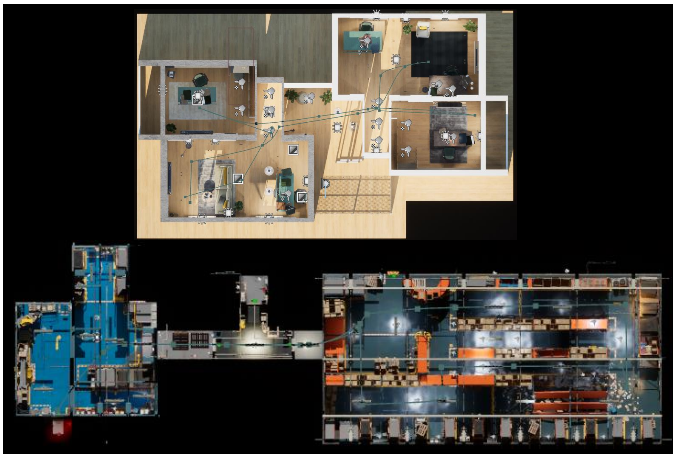


Fig. 8: Top-down visualization with overlaid ground-truth camera trajectory of the (top) *Interior Apartment* and the (bottom) *UE5 Warehouse* environments.

Each image sequence of the MESA dataset was tested using both systems. These sequences span a range of lighting and environmental variations, as detailed in Section III, while maintaining fixed camera trajectories within each scene with the exception of the *Omniverse Warehouse Dynamic Elements* which uses a separate path. Ground-truth trajectories were used to evaluate SLAM performance, both qualitatively, by overlaying the estimated trajectories with ground truth one, and quantitatively by computing standard SLAM evaluation metrics, including ATE RMSE and RPE Mean. Note that in cases where a SLAM approach lost track of the visual features and failed in producing a complete estimated trajectory, it is considered unsuccessful and excluded from the presented benchmarking results, highlighting the challenging cases of the proposed dataset.

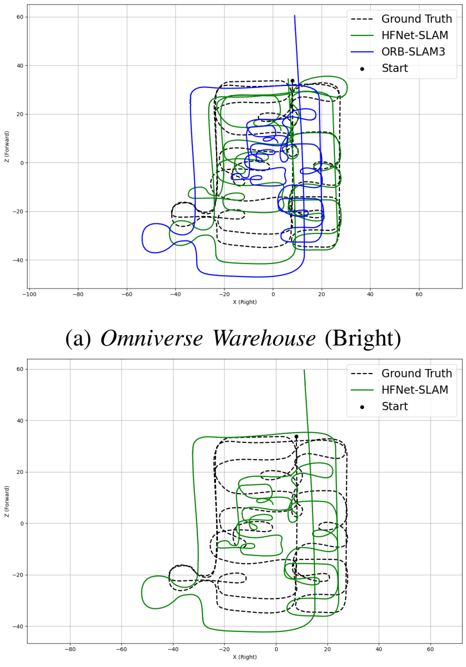
A. Results on the Omniverse Sequences

The *Omniverse Warehouse* static sequences, are longer and more complex spanning to 15,000 frames. Figure 9 shows

[4] https://github.com/LiuLimingCode/HFNet_SLAM

TABLE II: ATE and RPE mean for HFNet-SLAM and ORB-SLAM3 across environments. Failed sequences are excluded.

Method	Omniverse Warehouse				Lab Office				Interior Aptartment		UE5 Warehouse	
	Well-Lit		Smoke/Fog		Well-Lit		Low-Light		Bright		Bright	
	ATE	RPE	ATE	RPE	ATE	RPE	ATE	RPE	ATE	RPE	ATE	RPE
HFNet-SLAM	4.979	0.0121	10.353	0.0247	0.593	0.0039	0.651	0.0044	0.357	0.0035	1.008	0.0067
ORB-SLAM3	12.561	0.0289	–	–	3.997	0.0083	10.566	0.0403	0.553	0.0057	1.903	0.0113



(a) *Omniverse Warehouse* (Bright)

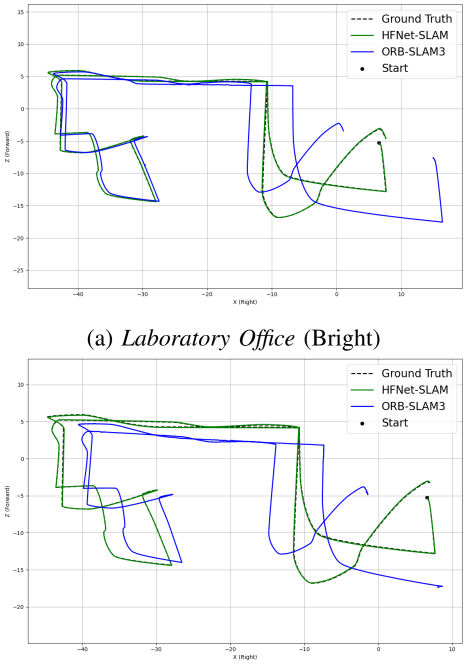
(b) *Omniverse Warehouse* (Smoke/Fog)

Fig. 9: Trajectory comparison between HFNet-SLAM, ORB-SLAM3, and ground-truth on *Omniverse Warehouse* under different visibility conditions.

qualitative comparisons for the bright and smoke/fog variants. HFNet-SLAM remains functional across both sequences, with expected performance degradation due to the difficultness of the dataset, especially under the influence of smoke/fog. ORB-SLAM3 performs acceptably in the bright light variant but fails in the smoke/fog scenario, unable to maintain consistent tracking. In the dynamic elements sequence of the *Omniverse Warehouse* environment, spanning to 30,000 frames, both systems failed to generate a complete trajectory.

B. Results on the UE5 Sequences

Figures 10 and 11 present qualitative trajectory comparisons between ORB-SLAM3, HFNet-SLAM, and the ground-truth across different scenes and environmental conditions. Results from the *Laboratory Office* environment under bright and low-light settings can be seen in Fig. 10. Both systems complete the sequences successfully. However, HFNet-SLAM shows notably improved trajectory alignment, particularly under the more challenging low-light setting. Runs on the dark and daylight with smoke/fog sequences of the *Laboratory Office* environment, failed to generate a complete SLAM trajectory.



(a) *Laboratory Office* (Bright)

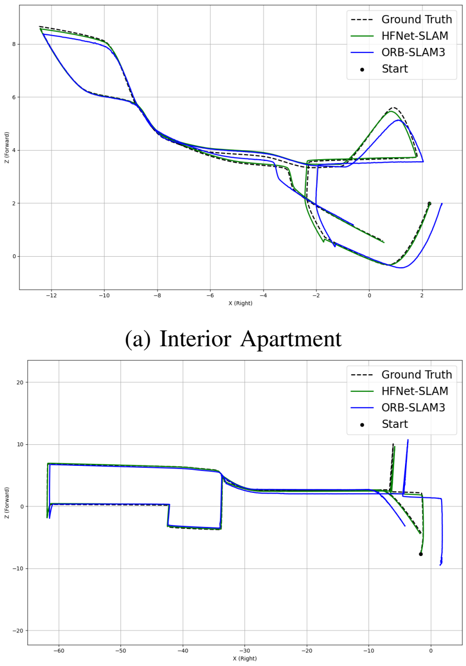
(b) *Laboratory Office* (Low-Light)

Fig. 10: Trajectory comparison between HFNet-SLAM, ORB-SLAM3, and ground-truth on the *Laboratory Office* dataset under different lighting conditions.

In Fig. 11, both pipelines succeed in the brightly lit versions of the *Interior Apartment* and *UE5 Warehouse* environments. However, all attempts to run these sequences under their corresponding dark variants failed for both systems. These failure cases underscore the importance of improved robustness to low-light conditions in SLAM research.

C. Quantitative Benchmarking

To provide a complete overview of system performance across the proposed environments, Table II presents the ATE RMSE and RPE Mean for both SLAM systems. HFNet-SLAM consistently outperforms ORB-SLAM3 in all scenarios showing that Deep Learning feature extraction can handle challenging environmental conditions, such as the ones included in the proposed dataset, more effectively than hand-crafted approaches. This is more evident in the case of low-light and smoke/fog conditions, which show significant performance gaps, with HFNet-SLAM maintaining better consistency in difficult visual conditions. The sequence where ORB-SLAM3 failed to track is marked as "–". In addition to quantitative results, Table III presents a summary of success/failure cases for



(a) Interior Apartment

(b) UE5 Warehouse

Fig. 11: Trajectory comparison between HFNet-SLAM, ORB-SLAM3, and ground-truth on the (a) *Interior Apartment* and (b) *UE5 Warehouse* datasets in bright conditions.

each SLAM system across all tested conditions. A sequence is marked as "Failure" when tracking was lost before concluding the camera's complete trajectory, or when the estimated trajectory significantly deviated without convergence.

TABLE III: Successes (✓) and Failures (✗) of the benchmarked SLAM techniques across all evaluated sequences.

Scene	**HFNet-SLAM**	**ORB-SLAM3**
Omni. Warehouse (Well-Lit)	✓	✓
Omni. Warehouse (Smoke/Fog)	✓	✗
Omni. Warehouse (Low-Light)	✗	✗
Omni. Warehouse (Dark)	✗	✗
Omni. Warehouse (Dynamic)	✗	✗
Lab Office (Well-Lit)	✓	✓
Lab Office (Daylight w/ Smoke/Fog)	✓	✓
Lab Office (Low-Light)	✗	✗
Lab Office (Dark)	✗	✗
Interior Apartment (Bright)	✓	✓
Interior Apartment (Dark)	✗	✗
UE5 Warehouse (Bright)	✓	✓
UE5 Warehouse (Dark)	✗	✗

V. Conclusion

In this paper we presented MESA, a large-scale synthetic dataset built using UE5 and NVIDIA Omniverse, designed to benchmark and train SLAM and learned feature extraction systems under systematically varied environmental conditions. MESA offers photorealistic indoor sequences rendered under diverse lighting and visibility settings including low-light, fog, and dynamic scenes, while preserving consistent camera trajectories across each variation. Experimental results using ORB-SLAM3 and HFNet-SLAM demonstrate that MESA comprises a set of challenging scenarios, highlighting the need for more robust monocular SLAM and feature extraction techniques. The dataset is made publicly available with the view to serve both as an evaluation tool, but also as a foundation for training models that generalize to real-world scenarios. As shown in ICA [5], synthetic datasets like MESA can support the development of robust learned feature detectors that transfer effectively to real environments. Future extensions of the dataset will include outdoor scenes, LiDAR, and depth sensor integration, as well as human–robot interaction environments, further expanding its applicability in the field of autonomous systems.

Acknowledgments

This research is implemented in the framework of H.F.R.I call "Basic research Financing (Horizontal support of all Sciences)" under the National Recovery and Resilience Plan "Greece 2.0" funded by the European Union – NextGenerationEU (H.F.R.I. Project Number: 15339).

References

[1] A. Geiger and P. Lenz and C. Stiller and R. Urtasun, "Vision meets robotics: The KITTI dataset," *The International Journal of Robotics Research*, vol. 32, no. 11, pp. 1231–1237, 2013.

[2] Sturm, Jürgen and Engelhard, Nikolas and Endres, Felix and Burgard, Wolfram and Cremers, Daniel, "A benchmark for the evaluation of RGB-D SLAM systems," in *In Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2012, pp. 573–580.

[3] Gaidon, Adrien and Wang, Qiao and Cabon, Yohann and Vig, Eleonora, "Virtual Worlds as Proxy for Multi-Object Tracking Analysis," in *In Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[4] Tremblay, Jonathan and To, Thang and Birchfield, Stan, "Falling Things: A Synthetic Dataset for 3D Object Detection and Pose Estimation," in *In Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2018.

[5] Agakidis, Anastasios and Bampis, Loukas and Gasteratos, Antonios, "Illumination Conditions Adaptation for Data-Driven Keypoint Detection under Extreme Lighting Variations," in *In Proc. of the IEEE International Conference on Imaging Systems and Techniques (IST)*, 2023, pp. 1–6.

[6] Zhang, Yi and Qiu, Weichao and Chen, Qi and Hu, Xiaolin and Yuille, Alan, "UnrealStereo: Controlling Hazardous Factors to Analyze Stereo Vision," in *In Proc. of the International Conference on 3D Vision (3DV)*, 2018, pp. 228–237.

[7] Michael Burri and Janosch Nikolic and Pascal Gohl and Thomas Schneider and Joern Rehder and Sammy Omari and Markus W Achtelik and Roland Siegwart, "The EuRoC micro aerial vehicle datasets," *The International Journal of Robotics Research*, vol. 35, no. 10, pp. 1157–1163, 2016.

[8] Handa, Ankur and Whelan, Thomas and McDonald, John and Davison, Andrew J., "A Benchmark for RGB-D Visual Odometry, 3D Reconstruction and SLAM," in *In Proc. of the IEEE International Conference on Robotics and Automation (ICRA)*, 2014, pp. 1524–1531.

[9] J. Straub, T. Whelan, L. Ma, Y. Chen, E. Wijmans, S. Green, J. J. Engel, R. Mur-Artal, C. Ren, S. Verma *et al.*, "The replica dataset: A digital replica of indoor spaces," *arXiv preprint arXiv:1906.05797*, 2019.

[10] Campos, Carlos and Elvira, Richard and Rodríguez, Juan J. Gómez and M. Montiel, José M. and D. Tardós, Juan, "ORB-SLAM3: An Accurate Open-Source Library for Visual, Visual–Inertial, and Multimap SLAM," *IEEE Transactions on Robotics*, vol. 37, no. 6, pp. 1874–1890, 2021.

[11] Sarlin, Paul-Edouard and Cadena, Cesar and Siegwart, Roland and Dymczyk, Marcin, "From Coarse to Fine: Robust Hierarchical Localization at Large Scale," in *In Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.