Enhancing Vision based SLAM through Shadow Removal Preprocessing

Hanxi Wan
Robotics
University of Michigan
Ann Arbor, USA
wanhanxi@umich.edu

Kanisius Kusumadjaja
ECE
University of Michigan
Ann Arbor, USA
jehiank@umich.edu

Seung Hun Lee
ECE
University of Michigan
Ann Arbor, USA
armyhuni@umich.edu

Tung Do
ECE
University of Michigan
Ann Arbor, USA
tungsdo@umich.edu

***Abstract*—In general, vision-based SLAM struggles to detect dynamic objects, which complicates tracking for Unmanned Ground Vehicles (UGVs). This issue arises because vision-based SLAM is susceptible to environmental factors such as shadows or significant changes in illumination, which can affect object detection. Particularly, raw data that excludes dynamic objects but includes shadows does not accurately represent the real environment. Our objective is to implement a shadow removal algorithm that addresses both static and dynamic objects at the front end of the SLAM pipeline to see whether it improves our SLAM accuracy results.**

***Index Terms*—Ground Vehicle, Vision-based SLAM, Shadow Removal**

I. Introduction

Vision-based Simultaneous Localization and Mapping (SLAM) inherently struggles with detecting dynamic objects, presenting significant challenges in tracking for Unmanned Ground Vehicles (UGVs). Research indicates that the removal of dynamic objects and their associated shadows can substantially improve SLAM outcomes [1], [2]. The susceptibility of vision-based SLAM to environmental influences, such as shadows or significant illumination changes, often disrupts object detection. Earlier studies, such as those by Bescos et al. [1] and Wimbauer et al. [3], have tackled this issue by segmenting and masking or completely removing dynamic objects. However, these methods often left residual shadows in the raw data, compromising the representation of the actual environment. Our research aims to refine this approach by developing and implementing a shadow removal algorithm that effectively processes both static and dynamic objects at the front end of the SLAM pipeline. This study explores whether such preprocessing enhances the accuracy of SLAM results, aligning with our goal to enhance vision-based SLAM through shadow removal.

II. Related Work

This section reviews pertinent literature in the areas of vision-based SLAM, focusing on adaptations for dynamic environments and illumination variations, as well as shadow removal techniques.

A. Vision-based SLAM

Vision-based SLAM systems create a map of an unknown environment and simultaneously keep track of their location within it. These systems predominantly use visual inputs from cameras, which makes them highly sensitive to changes in the environment.

1) Dynamic SLAM: Dynamic SLAM focuses on adapting standard SLAM approaches to handle environments where objects are in motion, which is a common scenario in urban and indoor settings. Traditional SLAM systems struggle in such conditions due to the static nature of the map they generate. Previous studies have introduced approaches to identify and exclude dynamic objects from the SLAM process, enhancing tracking accuracy and map fidelity [4].

2) Twilight SLAM: Twilight SLAM deals with the challenges posed by low-light conditions or significant changes in illumination, which are critical issues in vision-based systems. Researchers have explored various enhancement techniques such as integrating depth sensors or employing advanced machine learning algorithms to improve feature detection under varying lighting conditions. Previous work has demonstrated improvements in SLAM performance by adapting the feature extraction and matching processes to be more robust to these changes [5].

B. Shadow Removal

The presence of shadows can significantly impact the performance of vision-based SLAM systems, as shadows can alter the appearance of objects and surfaces, leading to incorrect assumptions about the environment. Recent advancements in shadow removal have been aimed at enhancing image quality for better object recognition and localization. Techniques range from simple color space transformations to complex neural networks that learn to differentiate between shadows and objects. Such techniques are pivotal for ensuring that SLAM systems are not misled by the transient changes in visual appearance caused by moving shadows.

Each of these areas contributes to the broader goal of improving the robustness and accuracy of vision-based SLAM systems in dynamic and poorly lit environments. Our work builds upon these studies by integrating a shadow removal algorithm directly into the SLAM pipeline, aiming to refine both the mapping and tracking components of the system.

III. Proposed Methodology

Our proposed methodology consists of two processes: preprocessing for shadow removal and ORB-SLAM2 for visual SLAM representation. The figure below shows the overall pipeline we have considered (see Fig. 1).

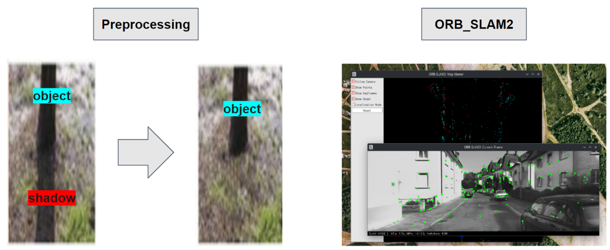

Fig. 1: Overview of our proposed methodology.

A. Shadow Removal

Lots of shadow removal algorithms are available, such as SpA-Former-shadow-removal [6], SID [7], and TBRNet [8]. We have implemented and tested several shadow removal algorithms from current research and have integrated it with the ORB-SLAM2 framework.

1) Not Learning Method:

Algorithm 1 Shadow Detection Algorithm in LAB color space

Convert the RGB image to a LAB image
Compute the mean values of the pixels in L, A, and B planes:
$\mu_L \leftarrow \text{mean}(L)$
$\mu_A \leftarrow \text{mean}(A)$
$\mu_B \leftarrow \text{mean}(B)$
if $\mu_A + \mu_B \leq 256$ **then**
Classify pixels with $L \leq (\mu_L - \sigma_L/3)$ as shadow pixels.
else
Classify pixels with lower values in both L and B planes as shadow pixels.
end if

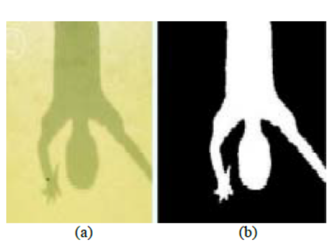

Fig. 2: Shadow Area detected using Algorithm 1. (a) original image; (b) the detected shadow region in white

- **LAB Color Space.** This method of detecting and removing shadows is based on the paper by Murali and Govindan [9] utilizing the LAB color space which represents color in terms of lightness (L) and the A and B channel which represent Green to Red ratio (A) and the Yellow and Blue ratio (B). The approach to detect shadows utilizes the L and B channels. The L channel provides lightness information and the B channel are generally smaller in most outdoor images. Combining the values from L and B, we can generate a threshold to determine whether pixels are a shadow or not. The general algorithm is shown in Algorithm 1. To remove the shadow we multiply the RGB channels of the shadow pixels by a constant as determined by the ratio of the average of the surrounding non-shadow area to that in the shadow area.
- **YCbCr Color Space.** This method introduces a straightforward framework utilizing the YCbCr color space, specifically the luminance, chroma: blue, and chroma: red components, to detect and remove shadows from images [10]. Initially, it employs statistical analysis of intensity within the YCbCr color space to identify shadows. Once detected, a shadow density model segments the image into regions of uniform density. The removal process then involves relighting each pixel within the YCbCr color space and subsequently adjusting the colors in the shadowed areas within the RGB color space to ensure seamless transitions. A key advantage of this approach is the preservation of detail and the avoidance of harsh transitions between shadowed and non-shadowed areas.

2) Learning Method:

- **SpA.** The SpA-Former method [6] innovatively integrates shadow detection and removal into a single efficient Transformer-based stage, thereby streamlining the process and enhancing computational efficiency. It utilizes a Gated Feed-Forward Network (GFFN) within the Transformer encoder to capture global dependencies effectively. This is complemented by a CNN decoder that incorporates elements of Generative Adversarial Networks (GANs), including Two-Wheel RNN joint spatial attention (TWRNN) and Fourier transform residual blocks (FTR). These elements are critical for accurately identifying and removing shadows, ensuring detailed reconstruction of shadow-free images. Notably, SpA-Former is designed to be lightweight, with a total size of only 0.47MB and requiring 15G FLOPS, making it suitable for real-time applications. This method not only reduces the complexity of traditional shadow removal processes but also maintains high performance on standard datasets like ISTD and SRD, making it an excellent enhancement for SLAM systems that require high-quality, shadow-free visual input for accurate mapping and localization.
- **TBRNet.** TBRNet (Three-branch residual network) [8] is composed of three branches dedicated to three distinct tasks (shadow image reconstruction, shadow matte esti-

mation, and shadow removal). The shadow image reconstruction branch maximally preserves detailed information of an input image (i.e., shadow image). The shadow matte estimation branch identifies shadow positions and computes pixel-level illumination adjustments. Lastly, the shadow removal branch restores the light of shadow areas according to the light of nonshadow areas, and finally, outputs a shadow-free image without artifacts.

B. ORB-SLAM2

ORB-SLAM2 is an advanced SLAM system designed for monocular, stereo, and RGB-D cameras. It is capable of real-time operation on standard CPUs across varied environments and is equipped with features like map reuse, loop closing, and relocalization. It excels in accuracy due to its backend based on bundle adjustment and includes a lightweight localization mode for zero-drift localization in known maps. The system is open-source, making it a valuable tool for research and practical applications alike [11].

IV. Datasets

1) KITTI (Karlsruhe Institute of Technology and Toyota Technological Institute): is a popular dataset in the field of SLAM and autonomous navigation consiting of hours worth of recorded city and highway traffic scenarios in several camera set ups such as RGB, grayscale stereo, and 3D laser scanning. The KITTI dataset provides the entire sequence of images from a vehicle navigating through different city and highway streets which allow us to apply our shadow removal algorithms to each image. In our study of applying shadow removal algorithms we utilized sequence 00 in both RGB and grayscale stereo. While our primary emphasis is on removing shadows in the scenes and seeing the results in ORB-SLAM2, the image sequences also provide insight into the effectiveness of shadow removal methods in various environmental lighting scenarios.

2) FinnForest: The FinnForest dataset [12] is a novel and challenging dataset of SLAM testing material in a forest landscape, particularly for mobile robotics working in semi-natural environments. In contrast to common urban structures, such as the dataset provided by KITTI, the FinnForest dataset provides an unregulated natural environment to exemplify sub-urban and forest environment. In the forest environment, shadows can be more present than in scenarios such as highway driving and the shadows are much less structured and defined such as those found in urban environments. The dataset is collected from a vehicle equipped with a sensor rig that constitutes four RGB cameras, an Inertial Measurement Unit, and a Global Navigation Satellite System receiver. The sensors are synchronized based on non-drifting timestamps. The dataset provides trajectories of varying complexity both for the state of the art visual odometry approaches and visual simultaneous localization and mapping algorithms. We select the summer sequence S01_8Hz_summer_seq1_shortLoop for our testing as it provides the opportunity for loop closure and has daylight which allows us to apply our shadow-removal algorithms.

V. Results

Our results were driven by different shadow removal algorithms. We compared the results from two non-learning methods on the KITTI dataset and also evaluated the results from two learning methods on the same dataset. Then, we chose the best algorithm from each method and applied it to the Finforest dataset, where we observed more shadows.

A. Non-Learning Method on KITTI Dataset

1) LAB Shadow Removal on KITTI Dataset: Figure 9a compares the shadow detection and removal using the LAB method on the KITTI 00 sequence. We see on the right hand side that the LAB algorithm is able to highlight parts of the scene that was previously occluded by a shadow which should presumably increase the performance of Visual SLAM algorithms. However, for the large shadow in the center of the scene, other than highlighting the shadow and showing the road more clearly, there is still a persistent outline of the shadow on the road, this is known as the ghosting effect and has been an issue for many shadow removal algorithms. Today's shadow removal algorithms are better able to remove the ghosting effect through the use of deep learning methods.

(a) Before Removal

(b) After Removal

Fig. 3: LAB Result.

2) YCbCr Shadow Removal on KITTI Dataset: Figure 4 shows a comparison between the original and shadow-removed KITTI data using the YCbCr algorithm. This method can also detect shadows; however, it mistakenly recognizes some objects as shadows. Since YCbCr separates the image into luminance (Y) and chrominance (Cb and Cr) component which exploits the human visual system's lesser sensitivity to fine color details, it allows for efficient compression compared to other color space method. Consequently, some parts of objects are incorrectly identified as shadows, which is far from our expectations. Therefore, in the non-learning methodology, the LAB color space shows better results compared to the YCbCr algorithm for shadow removal.

(a) Before Removal

(b) After Removal

Fig. 4: YCbCr Result.

(a) Before Removal

(b) After Removal

Fig. 5: SpA Result.

B. Learning Method on KITTI Dataset

1) SpA on KITTI Dataset: Figure 5 presents a side-by-side evaluation of the SpA method's shadow detection and removal capabilities on the KITTI dataset, with ellipses marking the shadow regions in the top image. The SpA method not only adeptly identifies but also effectively eradicates shadows, yielding a scene where areas once under shadow now exhibit enhanced brightness and uniformity akin to their surroundings. This meticulous preservation of details, apparent in the undisturbed textures of the road and building facade, is crucial for SLAM applications where feature extraction demands high accuracy. Moreover, the SpA approach manages to maintain color fidelity, avoiding the unnatural color shifts that often accompany shadow removal. There are no discernible edge artifacts post-removal, indicating a smooth transition between shadowed and lit regions. Given the real-world conditions mirrored by the KITTI dataset, the success of the SpA method points to its potential applicability in autonomous driving systems that necessitate precise environmental interpretation. The SpA method emerges as a potent preprocessing tool for vision-based SLAM, promising to significantly refine the system's mapping and navigational precision by providing clean, consistent visual data.

2) TBRNet on KITTI Dataset: Figure 6 presents a comparison between the original frame and its corresponding shadow-removed counterpart, both extracted from the KITTI dataset. The efficacy of this method lies in its adeptness at discerning and rectifying substantial shadow coverage, owing to its innovative approach featuring three distinct branches dedicated to various aspects of shadow removal. By strategically addressing different facets of this challenge, the method achieves commendable results in reclaiming obscured regions. Consequently, the rejuvenated areas not only enhance visual clarity but also furnish a richer array of feature points for algorithms like ORB-SLAM to leverage, thereby amplifying overall performance. Nevertheless, the efficacy of TBRNet diminishes when confronted with diminutive or irregular shadows, posing a notable challenge for complete removal and recovery in such scenarios. Further refinement may be necessary to address these nuanced instances comprehensively.

(a) Before Removal

(b) After Removal

Fig. 6: TBRNet Result.

C. ORB-SLAM2 on KITTI Dataset

In this section, ORB-SLAM2 is run on preprocessed data and the results are evaluated in terms of APE (absolute pose error), as shown in Table I. The APE value along timesteps and the box plot are shown in Figure 7 and 8.

Generally, learning methods tend to perform better than non-learning methods due to their superior ability to remove shadows and reduce noise, which leads to more robust feature extraction. By precisely removing shadows, it allows for more accurate feature detection and consistent feature matching, which is critical for the SLAM algorithm. The method also promotes steadier tracking by eliminating impermanent shadow features and enhances the quality of the SLAM-generated map by supplying shadow-free images. Furthermore,

	rmse	mean	median
Color_original	3.488324	3.180059	3.246839
Color_SpA	3.563915	3.264284	3.328308
Color_TRB	**3.350159**	**3.06422**	**3.137969**
Gray_original	1.34489	1.226514	1.144547
Gray_SpA	**1.303871**	**1.174259**	**1.131656**
Gray_TRB	1.37854	1.237017	1.223835
Gray_YcbCr	1.576639	1.388436	1.201032
Garay_LAB	1.538319	1.397639	1.222206

TABLE I: KITTI ORB-SLAM2 APE Results.

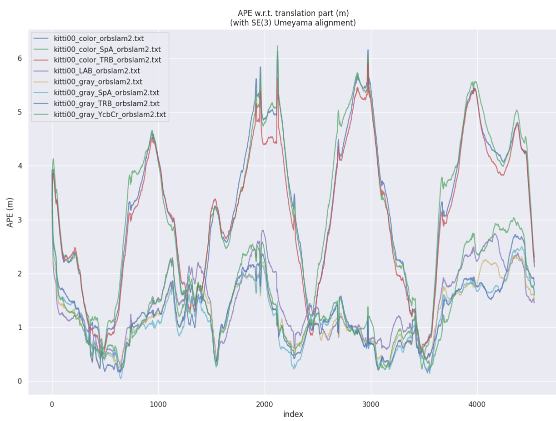

Fig. 7: ORB-SLAM2 APE Result.

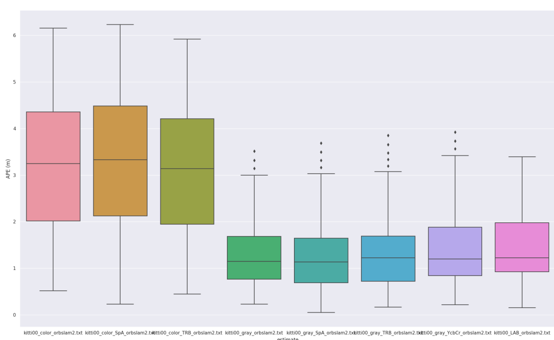

Fig. 8: ORB-SLAM2 APE Boxplot.

its learning-based design grants it the flexibility to handle various lighting conditions more adeptly compared to non-learning methods, which may not perform as uniformly across diverse environments.

Grayscale images hold an advantage over colored ones due to their inherent stability across diverse lighting conditions and robustness to illumination fluctuations. This stability ensures that features extracted from grayscale imagery maintain consistency and reliability. In the domain of colored imagery, the efficacy of TRBNet shines through. By adeptly recovering dark areas, TRBNet enriches the visual data with additional information, particularly beneficial for feature extraction processes. This augmentation of the image data enhances the discriminative power of the features extracted from colored images. Within the realm of grayscale imagery, SpA stands out as a superior method. By effectively eliminating shadows and accurately representing features on the road, SpA enhances the perceptual clarity of the scene. This precision in shadow removal not only improves visual fidelity but also aids in the extraction of salient features critical for SLAM algorithms.

D. Shadow Removal on Finforest Dataset

Now we investigate how our Shadow Removal algorithms perform on the FinnForest dataset.

1) LAB Shadow removal on FinnForest Dataset: We see that the LAB Shadow removal algorithm on the FinnForest dataset 9 is also effective to lighten up shadows and to remove much of the occlusion caused by the original shadow, particularly in the bottom left and middle right side of the image. However, it is not without flaws, as we see that the thin shadows in the center are not removed, this is likely because the shadows are not dark enough and therefore not past the threshold to be classified as a shadow in the LAB shadow detection algorithm.

(a) Before Removal

(b) After Removal

Fig. 9: LAB Result on FinnForest

2) SpA Shadow removal on FinnForest Dataset: Figures 10 and 11 reveal challenges encountered by the SpA method when applied to the FinnForest dataset, an environment that proves more demanding than the urban landscapes of the KITTI dataset. In the top image of 10, while shadows are recognized, they are not entirely removed, potentially due to the intricate textures and diverse lighting within forest settings that confuse the algorithm's ability to differentiate between actual shadows and naturally dark areas. Moreover, the top image of Figure 11 points out erroneous shadow detection, likely caused by the intricate patterns of light filtering through the trees and the complex terrain. The forest's nuanced texture and lighting, markedly distinct from urban environs, contribute to the SpA method's decreased efficacy, necessitating further model training on datasets akin to FinnForest. Additionally, the method's adaptation to new domains may benefit from domain adaptation techniques and multimodal data integration, like combining visual data with depth sensing, to refine shadow detection and removal for such multifaceted natural scenes.

(a) Before Removal

(b) After Removal

Fig. 10: Spa on FinnForest Result.

(a) Before Removal

(b) After Removal

Fig. 11: SpA on FinnForest Result.

VI. Discussion

A. Contributions

The goal of this project is to propose a novel methodology for minimizing one of the major environmental factors—shadow removal—to enhance visual-based SLAM. We attempted to find a more adequate algorithm for removing shadows by comparing non-learning and learning methods. We found that while the non-learning method can be efficient and applied in real-time SLAM, its results are not as good as those of the learning method. Consequently, the learning method has shown better performance not only in detecting shadows but also in removing them from objects. We evaluate a novel pipeline that we proposed in this project and show that our preprocessing improves Visual-based SLAM.

B. Limitations and Future Works

We have addressed several challenges associated with deterministic shadow removal using both mathematical algorithms and learning methods. However, there is still room for improvement.

First, removing shadows deterministically through mathematical algorithms presents substantial difficulties. One of the major issues encountered is ghosting, where remnants of shadows persist in the output image, reducing both clarity and quality. Furthermore, attempts to mitigate shadows can inadvertently introduce noise into the image. Particularly in the LAB color space, efforts to lighten shadows can lead to over-illumination, which distorts the true colors and brightness of the scene. Another significant challenge is the difficulty in distinguishing between shaded objects and actual shadows. This differentiation is crucial for accurate shadow removal but remains a complex problem due to subtle variations in light and color that need to be interpreted. Given these challenges, the future of shadow removal likely lies with machine learning methods. These approaches still face particular issues in the context of real-time applications, which struggle with complexity and running time. However, as shown in our results, machine learning can potentially provide more adaptive, efficient, and accurate solutions for shadow removal. One innovative approach could involve pre-masking shadows subsequent to object detection. By matching the detected object with its shadow, it might be possible to more precisely identify and remove the shadow, thereby enhancing the overall image quality. This method integrates object recognition with shadow detection, leveraging the strengths of machine learning to address the inherent complexities of shadow removal in dynamic environments. While current deterministic methods face significant hurdles, the integration of machine learning techniques presents a viable pathway toward more effective shadow removal solutions, promising substantial improvements in visual-based SLAM.

VII. Conclusion

In our exploration of enhancing vision-based SLAM with shadow removal preprocessing, we have demonstrated the potential to improve the robustness and accuracy of SLAM systems in varied environments. Our methodology integrates advanced shadow detection and removal algorithms—both learning-based and non-learning—into the front end of the SLAM pipeline. The successful application of the SpA method on the KITTI dataset, and the investigation of its limitations in the FinnForest dataset, paves the way for future research. Despite challenges such as distinguishing subtle shadow nuances and real-time processing demands, the advancements in learning-based shadow removal techniques are promising. Future work will focus on refining these methods and tailoring them to the unique challenges presented by different environments. This research has the potential to significantly reduce the impact of environmental factors on SLAM systems, a stride forward for autonomous navigation technologies.

VIII. Resources

GitHub: https://github.com/dyingplant/mobrob11

Presentation: https://www.youtube.com/watch?v=r6X7FIWKkBw

References

[1] B. Bescos, J. M. Facil, J. Civera, and J. Neira, "Dynaslam: Tracking, mapping, and inpainting in dynamic scenes," *IEEE Robotics and Automation Letters*, vol. 3, no. 4, p. 4076–4083, Oct. 2018. [Online]. Available: http://dx.doi.org/10.1109/LRA.2018.2860039

[2] B. Bescos, J. Neira, R. Siegwart, and C. Cadena, "Empty cities: Image inpainting for a dynamic-object-invariant space," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 5460–5466.

[3] F. Wimbauer, N. Yang, L. von Stumberg, N. Zeller, and D. Cremers, "Monorec: Semi-supervised dense reconstruction in dynamic environments from a single moving camera," 2021.

[4] H. Yin, S. Li, Y. Tao, J. Guo, and B. Huang, "Dynam-slam: An accurate, robust stereo visual-inertial slam method in dynamic environments," *IEEE Transactions on Robotics*, vol. 39, no. 1, pp. 289–308, 2022.

[5] S. P. Singh, B. Mazotti, S. Mayilvahanan, G. Li, D. M. Rajani, and M. Ghaffari, "Twilight slam: A comparative study of low-light visual slam pipelines," *arXiv preprint arXiv:2304.11310*, 2023.

[6] X. Zhang, Y. Zhao, C. Gu, C. Lu, and S. Zhu, "Spa-former:an effective and lightweight transformer for image shadow removal," in *2023 International Joint Conference on Neural Networks (IJCNN)*, 2023, pp. 1–8.

[7] H. Le and D. Samaras, "Physics-based shadow image decomposition for shadow removal," 2020.

[8] J. Liu, Q. Wang, H. Fan, J. Tian, and Y. Tang, "A shadow imaging bilinear model and three-branch residual network for shadow removal," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–15, 2023.

[9] S. Murali and G. V K, "Shadow detection and removal from a single image using lab color space," *Cybernetics and Information Technologies*, vol. 13, 03 2013.

[10] K. Deb and A. H. Suny, "Shadow detection and removal based on ycbcr color space," *SmartCR*, vol. 4, no. 1, pp. 23–33, 2014.

[11] R. Mur-Artal and J. D. Tardós, "ORB-SLAM2: an open-source SLAM system for monocular, stereo and RGB-D cameras," *IEEE Transactions on Robotics*, vol. 33, no. 5, pp. 1255–1262, 2017.

[12] I. Ali, A. Durmush, O. Suominen, J. Yli-Hietanen, S. Peltonen, J. Collin, and A. Gotchev, "Finnforest dataset: A forest landscape for visual slam," *Robotics and Autonomous Systems*, vol. 132, p. 103610, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0921889020304504