SCOPE: Stochastic Cartographic Occupancy Prediction Engine for Uncertainty-Aware Dynamic Navigation

Zhanteng Xie, *Member, IEEE*, and Philip Dames, *Member, IEEE*

***Abstract*—This article presents a family of Stochastic Cartographic Occupancy Prediction Engines (SCOPEs) that enable mobile robots to predict the future states of complex dynamic environments. They do this by accounting for the motion of the robot itself, the motion of dynamic objects, and the geometry of static objects in the scene, and they generate a range of possible future states of the environment. These prediction engines are software-optimized for real-time performance for navigation in crowded dynamic scenes, achieving up to 89 times faster inference speed and 8 times less memory usage than other state-of-the-art engines. Three simulated and real-world datasets collected by different robot models are used to demonstrate that these proposed prediction algorithms are able to achieve more accurate and robust stochastic prediction performance than other algorithms. Furthermore, a series of simulation and hardware navigation experiments demonstrate that the proposed predictive uncertainty-aware navigation framework with these stochastic prediction engines is able to improve the safe navigation performance of current state-of-the-art model- and learning-based control policies.**

***Index Terms*—Deep Learning in Robotics and Automation, Reactive and Sensor-Based Planning, Learning and Adaptive Systems, Environment Prediction.**

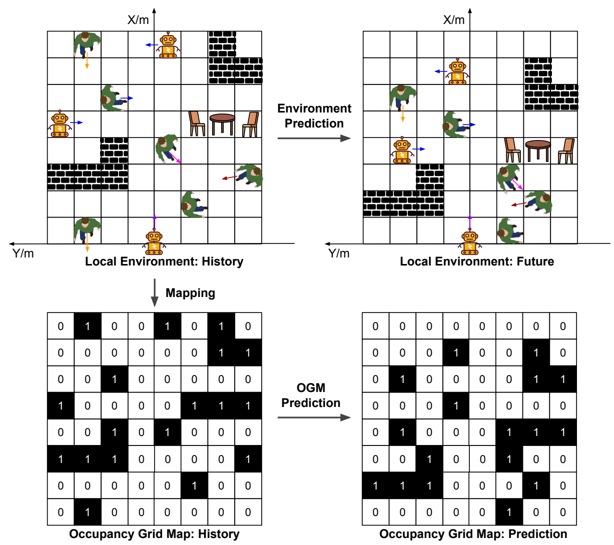


Fig. 1. A simple illustration of the occupancy grid map prediction problem. In a complex dynamic environment with many pedestrians, robots, tables, chairs and walls, colored arrows indicate the velocity of each agent.

I. Introduction

Autonomous mobile robots are beginning to enter people's lives and are trying to help us provide different last mile delivery services, such as moving goods in warehouses or hospitals and assisting grocery shoppers [1]–[3]. To realize this vision, mobile robots are required to safely and efficiently navigate through complex and dynamic environments filled not only with static obstacles (*e.g.,* tables, chairs, and walls) but also with many moving people and/or other mobile robots. The first prerequisite for robots to navigate and perform tasks is to use their sensors to perceive the surrounding environment. This work focuses on the next step, which is to accurately and reliably predict how the surrounding environment will change based on these sensor data, as shown in Fig. 1. This will allow a robot to proactively act based on its predictions and the associated uncertainty to avoid potential future collisions, a key part of improving autonomous robot navigation. Note that since this general perception-prediction-control navigation framework is a complex and resource-intensive system, it is very important to make these algorithms hardware-friendly (*e.g.,* using smaller computational power, memory usage, and storage usage) and run in real-time, especially for mobile robots with limited resources. A well-performing predictor is useless for practical robotics applications if it consumes a lot of memory and/or cannot run in real-time on a resource-limited robot.

In this article, we propose a family of deep neural network (DNN)-based Stochastic Cartographic Occupancy Prediction Engines (*i.e.,* SCOPE[1], SCOPE++, and SO-SCOPE) for resource-constrained mobile robots to provide stochastic future state predictions, as shown in Fig. 2, and enable uncertainty-aware navigation in crowded dynamic scenes, as shown in Fig. 3. Specifically, this article presents six contributions:

1) We design an algorithmic pipeline called SCOPE++ that can use a short history of robot odometry and lidar measurements to predict a distribution of potential future robot/environment states. SCOPE++ includes modules to compensate for the ego-motion of the robot, to segment static/dynamic objects in the scene, to predict future scenes using a ConvLSTM network, and to sample other future scenes using a variational autoencoder (VAE).
2) We analyze the running time and memory usage of

*This work was funded by Temple University.

Zhanteng Xie and Philip Dames are with the Department of Mechanical Engineering, Temple University, Philadelphia, PA, USA {zhanteng.xie, pdames}@temple.edu

Multimedia are available: https://youtu.be/8TtHTtJzuc8

[1]In our previous work [4] we used the acronym SOGMP (Stochastic Occupancy Grid Map Predictor) instead of SCOPE.

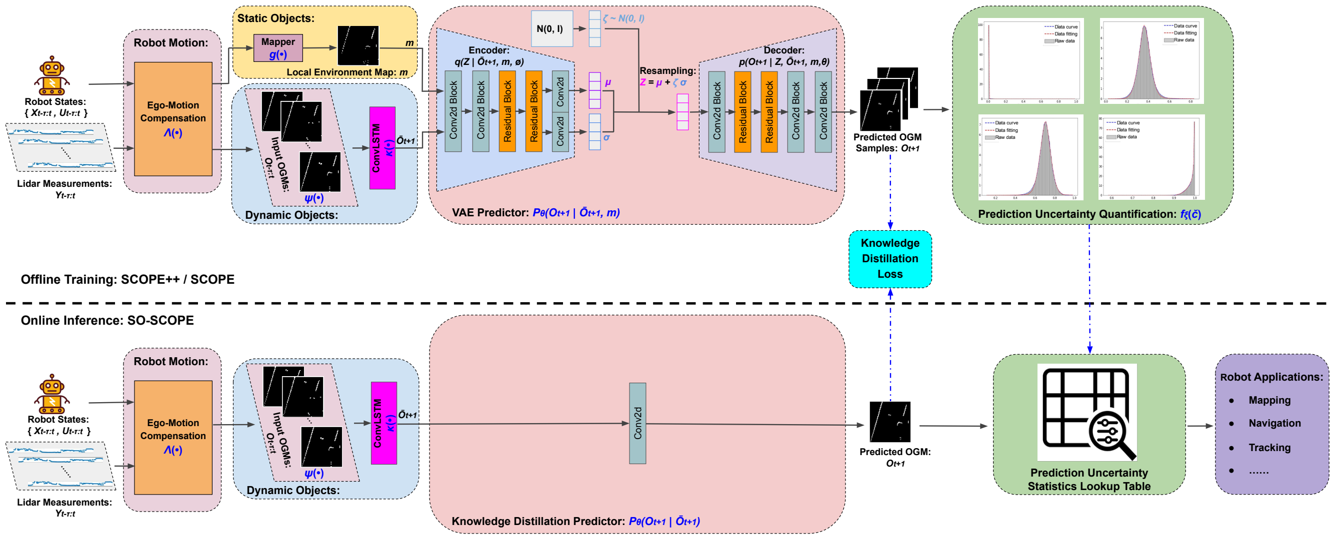


Fig. 2. System architectures of the SCOPE++ predictor, SCOPE predictor, and its software-accelerated SO-SCOPE predictor (note that SCOPE omits the Static Objects block compared to SCOPE++). The basic process of the SCOPE++ predictor is: 1) based on a history of robot states, the robot transfers the lidar measurement history to the predicted coordinate frame of the robot to compensate for the ego-motion, 2) these compensated lidar measurements are used to generate a local environment map to account for static objects, and a set of OGMs to account for dynamic objects, and 3) the local map of static objects and the predicted OGM of dynamic objects are fed into an variational autoencoder to predict the future OGM. To accelerate the SCOPE++ predictor, we first follow the SCOPE network architecture and replace the VAE network with a single convolutional layer, then use knowledge distillation technology to train the SO-SCOPE network to obtain the prediction information, and finally, we model and quantify the prediction uncertainty of the SCOPE++ to obtain uncertainty statistics and use them to generate uncertainty estimates of SO-SCOPE.

each module of SCOPE++ to identify computational bottlenecks. Based on this, we compress the VAE by performing an in-depth statistical analysis of its output and by using knowledge distillation techniques. The resulting software-optimized SCOPE (SO-SCOPE) achieves slightly better performance while consuming less memory, performing faster inference, and running in real-time with other resource-intensive algorithms on resource-constrained mobile robot hardware.

3) We validate the ability of our SCOPE predictors (*i.e.,* SCOPE++, SCOPE, and SO-SCOPE) to predict OGMs using three OGM datasets (each of which comes from a different robot model) and provide a comprehensive benchmark of prediction performance and resource usage using six state-of-the-art algorithms. We find that the SCOPE family achieves smaller absolute errors, higher structural similarity, higher tracking accuracy, and lower computational resource requirements than other state-of-the-art methods (*i.e.,* ConvLSTM [5], DeepTracking [6], PhyDNet [7], SAAConvLSTM [8], TAAConvLSTM [8], and LOPR [9]). We also perform a detailed analysis of the correctness, diversity, and consistency of the uncertainty estimates from the SCOPE family.
4) We propose a costmap-based predictive uncertainty-aware navigation framework to incorporate OGM prediction and its uncertainty information into current existing navigation control policies to improve their safe navigation performance in crowded dynamic scenes.
5) We validate the navigation performance in simulated 3D environments with varying crowd densities and real-world experiments. We find that the predictive uncertainty-aware navigation framework combined with our proposed SCOPE family can improve the navigation performance and safety of extant control policies relative to state-of-the-art solutions, including a model-based controller [10], a supervised learning-based approach [11], and two deep reinforcement learning (DRL)-based approaches [12], [13].
6) We open source the OGM prediction code with the OGM dataset [14] (https://github.com/TempleRAIL/scope) and its predictive uncertainty-aware navigation framework (https://github.com/TempleRAIL/scope_nav).

Note that the full version of the paper and more details can be found in [15].

II. Qualitative Results

A. OGM Prediction Results

Figure 4 and the attached Multimedia illustrate the future OGM predictions generated by our proposed predictors and the baselines. We observe three interesting phenomena. First, the image-based baselines, especially the PhyDNet, generate blurry future predictions after 5-time steps, with only blurred shapes of static objects (*i.e.,* walls) and missing dynamic objects (*i.e.,* pedestrians). We believe this is because these six baselines are deterministic models that use less expressive network architectures, only treat time series OGMs as images/video, and cannot capture and utilize the kinematics and dynamics of the robot itself, dynamic objects, and static objects. Second, the SCOPE++ with a local environment map has a sharper and more accurate surrounding scene geometry (*i.e.,* right walls) than the SCOPE without it. This difference indicates that the local environment map for static objects is beneficial and plays a key role in predicting surrounding scene geometry. Third, our proposed software-optimized SO-SCOPE can achieve clear

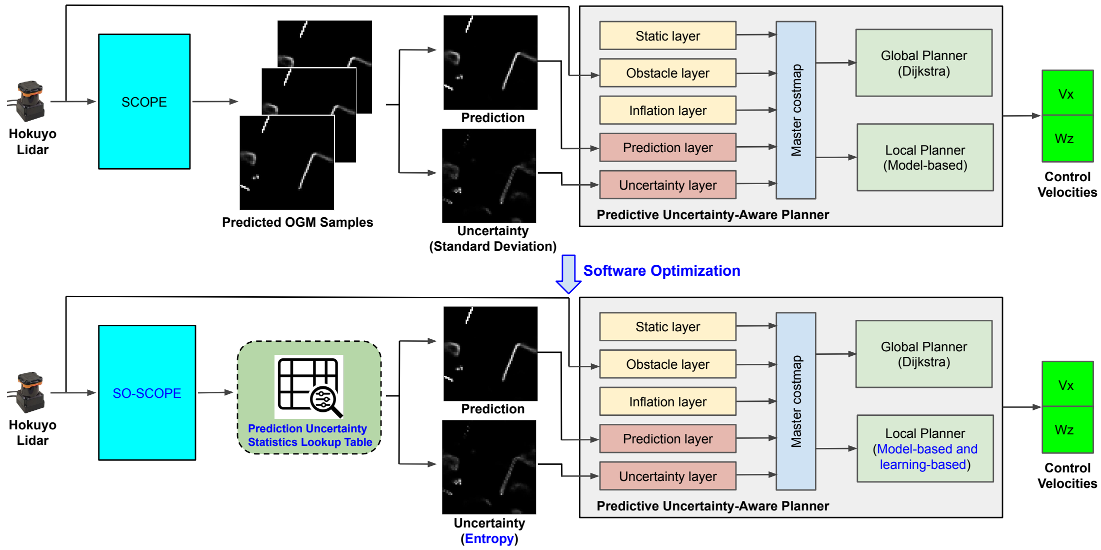


Fig. 3. System architectures of the SCOPE-based and SO-SCOPE-based predictive uncertainty-aware navigation planners. The blue font emphasizes the difference between the SCOPE-based navigation framework and the SO-SCOPE-based navigation framework. The basic process of our proposed navigation framework is as follows: first, the lidar data is also fed into our SCOPE or SO-SCOPE predictor to generate predicted OGM samples or lookup statistics from the prediction uncertainty statistics lookup table. Then, we can easily generate the prediction mean map and uncertainty map from these samples or the statistics. Finally, we create the prediction costmap layer and the uncertainty costmap layer, combine them into the master costmap, and obtain our SCOPE-based or SO-SCOPE-based predictive uncertainty-aware planners.

and sharp OGM predictions similar to its "teacher" SCOPE, which demonstrates the effectiveness of applying knowledge distillation techniques to optimize our SCOPE/SCOPE++.

B. Uncertainty-Aware Navigation Results

From the attached Multimedia and Figs. 5a and 5b, we can see how our robot deployed with DWA/SCOPE/PU can actively avoid collisions with walking students crossing the hallway, safely avoid static students standing, and reach predefined goals by following predictive uncertainty-aware nominal paths, traveling a total length of 76.10 m and an average speed of 0.42 m/s. It demonstrates the real-world effectiveness of our proposed SCOPE predictor and DWA/SCOPE/PU planner. In addition, from the attached Multimedia and Figs. 5c and 5d, we can see that even with three learning-based blocks, our robot deployed with DRL-VO/SO-SCOPE/PU is still able to quickly and actively avoid collisions with walking students crossing the hallway and reach predefined goals by following predictive uncertainty-ware nominal paths, traveling a total length of 86.41 m and an average speed of 0.47 m/s. It demonstrates our software-optimized SO-SCOPE predictor is hardware friendly and our software-optimized predictive uncertainty-aware navigation framework can be combined with different high computational load learning-based algorithms.

REFERENCES

[1] R. ED, "Types and applications of autonomous mobile robots," https://www.conveyco.com/blog/types-and-applications-of-amrs, July 2022, (Accessed on 08/20/2022).

[2] J.-u. Kim, "Keimyung hospital demonstrates smart autonomous mobile robot," https://www.koreabiomed.com/news/articleView.html?idxno=10585, Mar 2021, (Accessed on 08/20/2022).

[3] SICK, "Revolutionizing grocery shopping with mobile robots," https://sickusablog.com/revolutionizing-grocery-shopping-mobile-robots, Mar 2021, (Accessed on 08/20/2022).

[4] Z. Xie and P. Dames, "Stochastic occupancy grid map prediction in dynamic scenes," in *Proceedings of The 7th Conference on Robot Learning*, ser. Proceedings of Machine Learning Research, vol. 229. PMLR, 06–09 Nov 2023, pp. 1686–1705. [Online]. Available: https://proceedings.mlr.press/v229/xie23a.html

[5] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-c. Woo, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," in *Advances in Neural Information Processing Systems*, vol. 28, 2015.

[6] P. Ondruska and I. Posner, "Deep tracking: Seeing beyond seeing using recurrent neural networks," in *Thirtieth AAAI conference on artificial intelligence*, 2016.

[7] V. L. Guen and N. Thome, "Disentangling physical dynamics from unknown factors for unsupervised video prediction," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 11 474–11 484.

[8] B. Lange, M. Itkina, and M. J. Kochenderfer, "Attention augmented convlstm for environment prediction," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2021, pp. 1346–1353.

[9] ——, "Lopr: Latent occupancy prediction using generative models," *arXiv preprint arXiv:2210.01249*, 2022.

[10] D. Fox, W. Burgard, and S. Thrun, "The dynamic window approach to collision avoidance," *IEEE Robotics & Automation Magazine*, vol. 4, no. 1, pp. 23–33, 1997.

[11] Z. Xie, P. Xin, and P. Dames, "Towards safe navigation through crowded dynamic environments," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, Sep. 2021. [Online]. Available: https://doi.org/10.1109/IROS51168.2021.9636102

[12] R. Guldenring, M. Görner, N. Hendrich, N. J. Jacobsen, and J. Zhang, "Learning local planners for human-aware navigation in indoor environments," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 6053–6060.

[13] Z. Xie and P. Dames, "DRL-VO: Learning to navigate through crowded dynamic scenes using velocity obstacles," *IEEE Transactions on Robotics*, vol. 39, no. 4, pp. 2700–2719, Aug. 2023. [Online]. Available: https://doi.org/10.1109/TRO.2023.3257549

[14] ——, "Stochastic Occupancy Grid Map Prediction in Dynamic Scenes: Dataset," https://doi.org/10.5281/zenodo.7051560.

[15] ——, "Scope: Stochastic cartographic occupancy prediction engine for uncertainty-aware dynamic navigation," *IEEE Transactions on Robotics*, vol. 41, pp. 4139–4158, 2025.

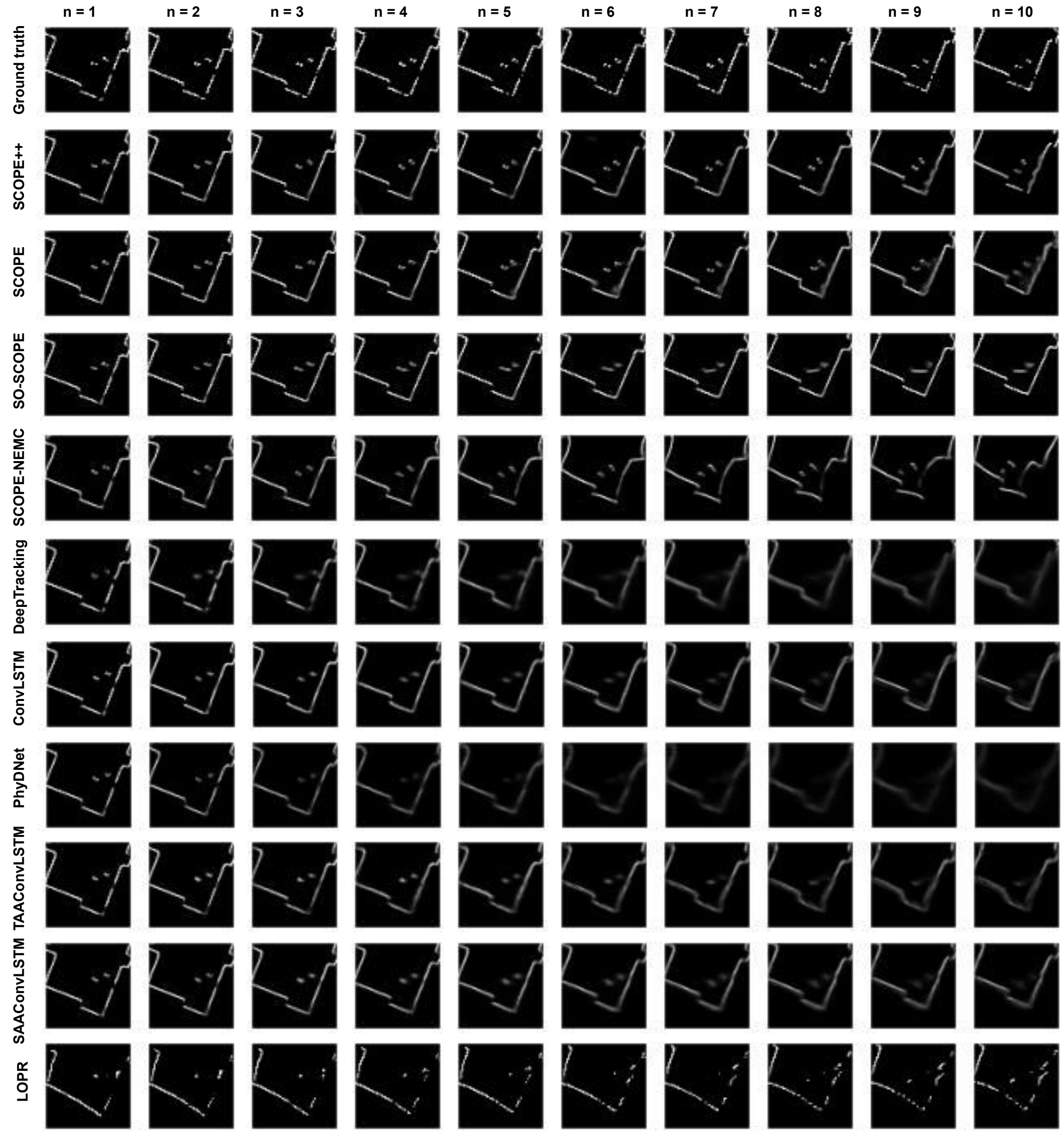
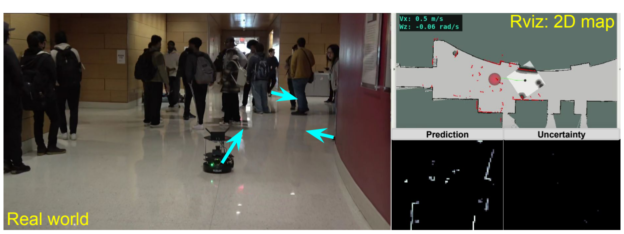
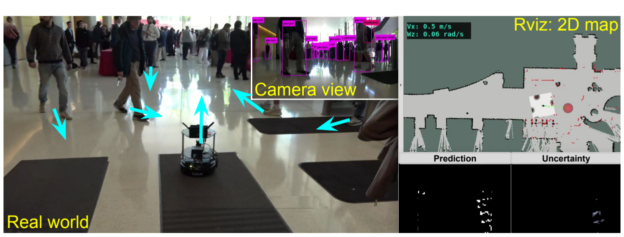


Fig. 4. Prediction showcase of ten OGM predictors tested on the OGM-Turtlebot2 dataset. The black and white areas are free and occupied space, respectively.

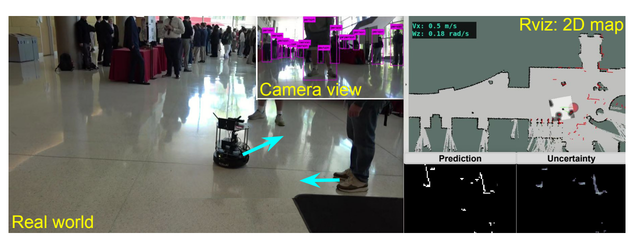


(a) Robot deployed with DWA/SCOPE/PU reactions (time t).

(b) Robot deployed with DWA/SCOPE/PU reactions (time $t + 3$).



(c) Robot deployed with DRL-VO/SO-SCOPE/PU reactions (time t).



(d) Robot deployed with DRL-VO/SO-SCOPE/PU reactions (time $t + 3$).

Fig. 5. Robot deployed with different uncertainty-aware control policies reactions to moving pedestrians in the indoor hallway with high crowd density at different times. (a-b) DWA/SCOPE/PU. (c-d) DRL-VO/SCOPE/PU