



**HAL**  
open science

# A Navigation-Based Evaluation Metric for Probabilistic Occupancy Grids: Pathfinding Cost Mean Squared Error

Jean-Baptiste Horel, Robin Baruffa, Lukas Rummelhard, Alessandro Renzaglia, Christian Laugier

## ► To cite this version:

Jean-Baptiste Horel, Robin Baruffa, Lukas Rummelhard, Alessandro Renzaglia, Christian Laugier. A Navigation-Based Evaluation Metric for Probabilistic Occupancy Grids: Pathfinding Cost Mean Squared Error. ITSC 2023 - 26th IEEE International Conference on Intelligent Transportation Systems, IEEE, Sep 2023, Bilbao, Spain, Spain. pp.1-6. hal-04211125

**HAL Id: hal-04211125**

**<https://hal.science/hal-04211125v1>**

Submitted on 19 Sep 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# A Navigation-Based Evaluation Metric for Probabilistic Occupancy Grids: Pathfinding Cost Mean Squared Error

Jean-Baptiste Horel<sup>1</sup>, Robin Baruffa<sup>1</sup>, Lukas Rummelhard<sup>1</sup>, Alessandro Renzaglia<sup>2</sup> and Christian Laugier<sup>1</sup>

**Abstract**—While robotics increasingly relies on occupancy grids for environment perception, the lack of specifically-designed metrics leads existing research to employ Image Quality Assessment (IQA) metrics and topological evaluations, which were primarily designed for binary occupancy grids. While appropriate as a first approximation, not taking into account the particular nature and usage of probabilistic occupancy grids limits the accuracy of their evaluation. In this paper, we propose the PathFinding Cost Mean Squared Error (PFC-MSE), a new probabilistic occupancy grid comparison metric designed to incorporate their main usage and attributes. Emulating grid-based navigation methods, the metric defines the difference between two grids as the measured spread between the navigation behavior their use induces, which emphasizes variations in general topology over local cell-value fluctuations. Experimental results on 10,000 driving scenes exhibit the relevance of the approach in quantifying grid disparities compared to existing approaches.

## I. INTRODUCTION

First introduced in [1], Occupancy Grids (OG) provide a probabilistic representation of the environment as perceived by robotic systems. This representation plays a crucial role in constructing an accurate model of the environment, effectively capturing the uncertainty inherent to the sensors and the perception systems. They allow the system to create a discrete and rasterized internal representation of the environment in order to detect and localize elements such as pedestrians, vehicles, drivable space, and static obstacles. In the robotic field, OGs are used for a wide range of navigation tasks such as Automated Driving Systems (ADS), Advanced Driver-Assistance System (ADAS), collision avoidance, localization and mapping. To optimize, test and validate OG inference models for a specific application, a metric capable of assessing the adequacy of the environment perception must be employed, quantitatively comparing the inferred OG to an aimed Ground Truth (GT) OG of the environment.

In the literature, most publications rely on general-purpose metrics derived from the computer vision literature such as Intersection Over Union (IoU), Mean Squared Error (MSE) and Cross-Entropy [2], [3], [4]. These IQA metrics, measuring visual resemblance between two OGs, may not be the best indicators that the inferred OG will lead to the desired navigation behavior. The selection of an appropriate perception assessment metric should rely on the aimed use case of the robotic platform.

In this paper, we introduce the PathFinding Cost MSE (PFC-MSE) metric, an ego-centric similarity metric capable

of evaluating the similarity between two probabilistic OGs by simulating the behavior of a navigation algorithm. The main postulate being that if a navigation algorithm generates similar trajectories using both the GT and the inferred OG, the two are alike for navigation purposes. Unlike existing metrics that often perform pixelwise comparison, PFC-MSE is sensitive to inference errors perturbing the behavior of the pathfinding algorithm, while being insensitive to errors that do not. Using a 10,000 scenes driving dataset with inferences and GT OGs, we analyze its behavior and compare it to existing metrics.

## II. STATE OF THE ART

Existing works using probabilistic OGs usually focus on quantitative measurements such as computation time, algorithmic complexity and theoretical advantages but lack a quantitative similarity measurement between two OGs [5]. Due to the scarcity of established probabilistic OG metrics in the literature, we were prompted to explore related research fields in search of relevant existing works. Existing OG metrics can be categorized as Image Quality Assessment-based, localization-based, geometry-based, and topology-based.

### A. Image Quality Assessment (IQA)

Many state-of-the-art perception algorithms are compared on benchmark datasets using metrics derived from computer vision such as IoU, MSE and cross-entropy [2], [5]. These metrics are often pixel-to-pixel comparison between the GT and the inference OG. If the inference is slightly offset, rotated or deformed, as it is often the case on real robotic systems, these metrics may yield a totally different result. Multi-scale approaches such as Gaussian Pyramid and IW-SSIM methods tackle this problem by calculating similarity at different resolutions [6], [7]. Instead of comparing images in the spatial domain, some IQA metrics compare the images in the frequency or wavelet domain, which capture the frequency local distributions of the image [6]. IQA metrics are often purpose-built to emulate human vision quality assessment. These vision-based metrics are not necessarily the best for OG assessment as two OGs can visually look similar but lead to a drastically different behavior in the decision making algorithms [2]. Human vision is more sensitive to localized narrow-band distortion [8], while for navigation use-cases, low-frequency distributions are often more important to capture the overall environment topology. Metric such as SSIM and CW-SSIM are purposely insensitive to luminance and low contrast changes [8]. Using these metrics on probabilistic OGs would ignore the mean probability of

<sup>1</sup>Univ. Grenoble Alpes, INRIA, 38000 Grenoble, France; e-mail: `firstname.lastname@inria.fr`

<sup>2</sup>Univ Lyon, Inria, INSA Lyon, CITI, F-69621 Villeurbanne, France.

the OG and only evaluate the relative variation of probability. Furthermore, while manipulated data structures are similar (i.e. matrices) in images and OGs, the information stored in each cell/pixel are either occupancy probability or luminance, representing distinct concepts. Applying IQA metrics on probabilistic OGs does not allow taking into account the specific meaning of the representation.

### B. Localization-based metrics

Mapping algorithms generate a binary 2D map of the environment for path planning and localization purposes. In the context of mapping algorithms, most authors postulate that a correct inference is one that leads to a precise localization of the robot [9], [10], [11]. A localization algorithm, such as Monte-Carlo Localization (MCL), locates the robot in the 2D map using sensor readings. The positional error is calculated with respect to the real world measurement of the robot position and used for performance assessment. This approach is widely used in the Simultaneous Localization and Mapping (SLAM) literature and the standard localization metric is the Root Mean Squared Error (RMSE) between the GT trajectory and the inferred trajectory [9]. This approach has the advantage of working without the need for a GT map, which is often hard to obtain [10]. On the other hand, it is limited to a single trajectory and therefore cannot capture the entirety of the environment topology.

### C. Geometry-based metrics

Geometry-based metrics have been proposed to evaluate the accuracy of maps for localization and planning purposes. Some approaches compare the positions of corner features in both the GT and the inferred map. [12], [13], [14] also present such a geometrical feature-based metric, the latter proposing to use the SIFT algorithm to generate geometrical features. These approaches work best for structured environment where a purpose-built feature extractor can be used. Performing the error calculation using a limited number of geometrical features inherently leads to a loss of information. Parts of the OG that might be relevant for navigation may be ignored altogether.

### D. Topology metrics

A different approach consists in evaluating the similarity of two maps using topology graphs. In [15], a topological skeleton generated from the inferred OG is overlaid on top of the GT to count the number of vertices that are false positives and false negatives. However the authors did not present a generic method to generate the skeleton graphs. Similarly, [16] introduces the use of Voronoi diagrams to generate topological graphs for both the OG and inference. Both graphs are compared using the number of matching vertices between both graphs, the global and relative error in the location of graph vertices. As the final comparison between two OGs is performed using a finite number of vertices, this method suffers from the same information-loss shortcomings as the geometry-based metrics. Only points

far away from obstacles are evaluated, and the close-to-obstacle ones are purposely ignored, even though they could be critical for navigation. Not relying on a unique topological graph, [17] compared the length of trajectories generated using the A\* pathfinding algorithm in order to compare the navigation behavior between hexagonal tiled OGs and squared tiled OGs. While topological approaches are relevant for navigation-based metrics, they are designed for indoor binary OGs, which are different from the probabilistic ego-centered use-case.

## III. METHODOLOGY

The objective of our new navigation-oriented metric is to address the limitations identified in existing methods. PFC-MSE is specifically designed to assess the similarity between OGs by considering the behavior of an ego vehicle navigating through these grids. Our metric evaluates the similarity between the GT OG, an aimed representation of the environment surrounding the ego vehicle (Figure 1a), and the inference of this environment made by the vehicle perception system (Figure 1c). Both grids are represented using a single channel probabilistic OG where each cell represents the likelihood of it being occupied by an obstacle. Figure 1 illustrates intermediate steps of the metric evaluation process. PFC-MSE is computed using the following steps:

- 1) For each OG we find its shortest-path tree, composed of all the shortest paths from the ego position to every cell of the OG (some example of paths are drawn in red in Figure 1b and 1d).
- 2) Both OGs are transformed into cost grids using the cumulative costs of the paths from their respective shortest-path trees (e.g. Figure 1b and Figure 1d).
- 3) An intermediate distortion grid is computed by performing the cell-wise absolute error between both cost grids (e.g. Figure 1e).
- 4) The metric is evaluated by computing the MSE of the distortion grid weighed by the disjunctive probability of free occupancy on both grids.

### A. Shortest-path tree and cost grid

Using Dijkstra's pathfinding algorithm [18], the shortest-path tree of an OG is computed starting from the ego vehicle position. For the pathfinding algorithm to generate meaningful paths, the free cells should have a positive minimal cost. The cell values are interpolated between 1 and a scaling parameter *ratio*.

$$Grid_{scaled} = (ratio - 1) \times Grid + 1 \quad (1)$$

The pathfinding is allowed to cross any cell, even occupied ones. The ratio parameter impacts how detours are made to avoid occupied cells. With a *ratio* of 100, crossing 100 free cells is more cost-effective than crossing one occupied cell. Its impact and the value we have chosen for our experiments will be discussed later in section IV-A.

The OG is then interpreted as an oriented graph where each cell is linked to its 4 adjacent cells and 4 diagonal cells. The edge costs are the euclidean distance between the

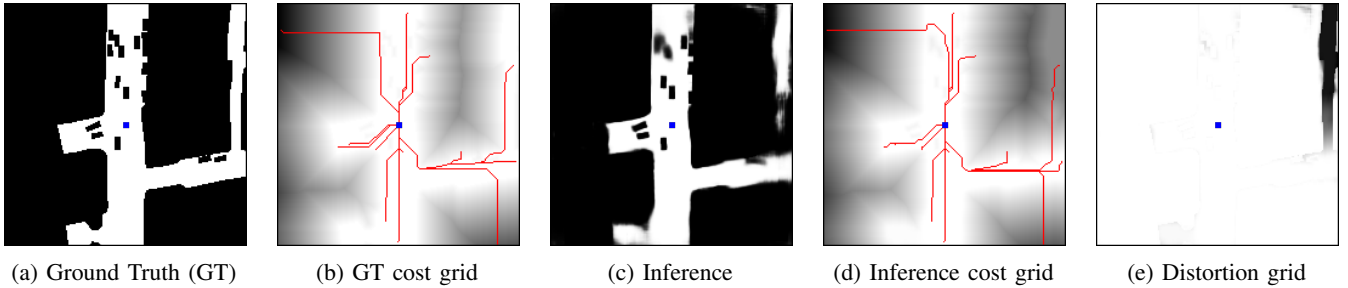


Fig. 1: Illustration of the metric evaluation process. 1a and 1b are the Ground Truth (GT) and its cost grid, 1c and 1d are the inference and its cost grid. Examples of paths are drawn in red on both cost grid. The resulting distortion grid 1e is the pixel-wise absolute error between both cost grids, it is also weighed by the disjunctive probability of free occupancy on the GT or the inference. PFC-MSE, IoU and MSE scores can be found in Table I scene 2.

cells, it allows diagonal movements on the OG. Given two cells  $c_1$  and  $c_2$  of the OG, the cost of the oriented edge from  $c_1$  to  $c_2$  is the occupancy of  $c_2$  multiplied by the distance between the cells:

$$\text{cost}(c_1 \rightarrow c_2) = \text{Grid}_{\text{scaled}}(c_2) \|c_2 - c_1\|_2 \quad (2)$$

Dijkstra’s algorithm is applied on the obtained graph using the cell representing the ego vehicle  $c_{ego}$  as the starting node. For each cell  $c$  of the OG we obtain the shortest path from  $c_{ego}$  to  $c$ :

$$\begin{aligned} \text{Path}(c) &= (c_0, c_1 \dots c_L) \\ \text{with } c_0 &= c_{ego} \text{ and } c_L = c \end{aligned} \quad (3)$$

Examples of paths on a GT and an inference are drawn in red in Figure 1.

The value of each cell from the cost grid is the total cost of the path leading to it, taken from the shortest-path tree. These costs are computed from scaled grid values that were scaled for the pathfinding, then they are scaled back so that they correspond to occupancy values. Therefore, the total cost of a path is the sum of the occupation probability of each cell along it, in other words the quantity of occupation crossed by the path.

$$\text{Grid}_{\text{cost}}(c) = \frac{\sum_{i=0}^{L-1} \text{cost}(c_i \rightarrow c_{i+1}) - L}{\text{ratio} - 1} \quad (4)$$

The transformation from a probabilistic OG to a cost grid is bijective, meaning no information is lost when using cost grids as proxies to compare two OGs.

### B. Distortion measurement

To measure the distortion between two OGs, we compute the pixel-wise absolute error between their respective cost grids. Each cost grid cell value is then weighed by the disjunctive probability of the cell being free either on the GT OG or on the inference OG. This distortion grid is an intermediate representation of the differences between both OGs and is later pooled using MSE and normalized by the probabilistic weighing, to evaluate the score of the metric.

The cost grids of two identical OGs are the same as all the shortest paths and their costs would be the same. Thus, no differences can be observed from their cost grids and the

norm of their distortion grid is zero. For two similar grids with small dissimilarity like in Figure 1, the differences in the grids will be passed on the cost grids and the distortion grid in two ways:

- 1) The final cost of a path is that of the destination cell. A dissimilarity on this cell will directly create a difference of cost between the path from the GT and the path from the inference. On the distortion grid, this difference will be localized at the cell where the dissimilarity is.
- 2) A dissimilarity on a cell impacts the shortest path tree and the cost of the paths. New shortest paths can emerge on the inference that are different from the GT paths. Depending on the dissimilarity, the paths can make detours to avoid a cost increase or to benefit from a cost decrease. The shortest paths can also remain the same but with changed costs. Overall, this difference will be localized on the distortion grid at the destination cell of the paths impacted by the dissimilarity.

This gives several properties to our PFC-MSE metric. Firstly, dissimilarities are not only measured on cells having different values but also on the costs of paths impacted by the dissimilarity. It allows us to not only measure local distortions, but also evaluate the topological differences between the GT and the inference from a pathfinding point of view. Secondly, cells crossed by more paths are given more importance. A cell further away from the ego-vehicle will be crossed by a smaller number of paths overall and will therefore have less impact on the final score. In specific cases such as Figure 1e, the cost of the paths downstream of the bottleneck depends on the cell values crossed in the bottleneck. Therefore, using pathfinding allows us to not only measure the localized differences between two OGs, but also measure the impact of this differences on the topology of the OGs and on the navigation behavior.

The final evaluation of PFC-MSE is performed by computing the squared norm of the distortion grid weighed by the probability of free occupancy of the cell. The probability weighting gives more importance to costs of cells that are free either on the GT or the inference.

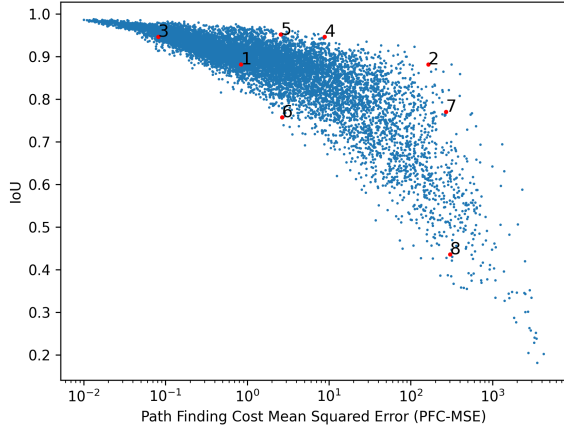


Fig. 2: Correlation graph between PFC-MSE and IoU. Each point corresponds to an evaluation from the 10,000 scenes dataset. Being based on MSE, best scores of PFC-MSE are close to zero. The highlighted points are the 8 scenes shown in Table I. Scenes 1, 2 and scene 3, 4 have similar IoU and MSE. Scenes 5, 6 and scenes 7, 8 have similar PFC-MSE.

$$\text{PFC-MSE}(GT, I) = \frac{\sum_c w_c \|GT_{cost}(c) - I_{cost}(c)\|_2}{\sum_c w_c} \quad (5)$$

with  $w_c = p(c = free) = 1 - GT(c)I(c)$

When evaluating metrics for classification tasks, the confusion matrix is often used as a standard analysis tool. OG inference is a classification problem where occupied space is the negative condition and free space is the positive condition. PFC-MSE evaluation is based on true positives, false positives and false negatives which are the cells on which the navigation task is performed. The probabilistic weighting is close to zero when a cell is classified as occupied on both OGs, therefore PFC-MSE gives little to no importance to true negatives (cells correctly classified). On the other hand, the emphasis is on the cells classified as free on the GT or on the inference, which are true positives, false positives and false negatives.

#### IV. EXPERIMENTS

To evaluate the effectiveness of our metric and compare it against others, we conducted experiments using the Nuscene dataset [3]. This dataset includes a large number of scenes from diverse driving scenarios, such as crowded urban streets, highways, and rural roads. Different sensor readings are included, such as 6 cameras, LIDAR and radar point-clouds. We generated inference OGs using LIDAR-Aided Perspective Transform Network (LAPNet-FPN) [4]. It is a machine learning model that combines LIDAR and camera data to generate semantic grids for different classes (vehicles, pedestrians, drivable area, etc.) in the Bird-Eye-View (BEV) frame. BEV networks usually apply a threshold to produce binary inferences used for IoU evaluation. Since we are

interested in probabilistic OGs, we removed this threshold so that the inference has continuous values. We combine the vehicles, pedestrians and drivable area segmentation masks generated by LAPNet-FPN to reconstruct a free or occupied space representation of the environment. Figure 1a shows an example of a GT OG from Nuscene and Figure 1c the corresponding OG inference generated using LAPNet-FPN. We have generated 10,000 combinations of GT and inferences with a size of 200x200 cells and a resolution of 0.5m per cell, equating to a total perception area of 100x100m around the ego vehicle [4]. For each scenes, we evaluated PFC-MSE, IoU and MSE. Since IoU only works on binary OGs, we used the same threshold of 0.5 as done by LAPNet-FPN [4].

We experimented with different values of *ratio* (10, 50, 100, 200) and found no meaningful differences. We used a *ratio* of 100, meaning a path will cross an occupied cell if it does not have a possible detour crossing less than 100 free cells. We find this choice relevant as our OGs have a size of 200x200 and the ego vehicle is at the center.

##### A. PFC-MSE behavior

In this section we illustrate and analyse the behavior of PFC-MSE. Table I showcases a subset of eight scenes chosen from our dataset illustrating the GT alongside its respective inference generated by LAPNet-FPN and their IoU, MSE and PFC-MSE scores.

1) *Topological errors*: Scene 2 from Table I illustrates how PFC-MSE evaluates topological errors. An incorrect inference classification leads to an important modification of the environment topology by incorrectly blocking access to a section of the road. On the GT, paths going to this road section only cross free space, thus their cost are low, (cf. Figure 1b), while on the inference, the paths cross occupied cells, thus increasing their costs (cf. Figure 1b). Even if cells are well classified on the inference, their respective paths accumulate errors from upstream cells as displayed by the distortion grid Figure 1e and Table I scene 2. PFC-MSE penalizes topological errors by evaluating how the navigation leading to cells downstream of the errors is disturbed.

2) *Distance from the ego position*: Scenes 2, 7 and 8 in Table I illustrate how errors close to the ego position are accentuated in the PFC-MSE score. For each distortion grid, gradients of error start at the closest misclassification from the ego position and increase while moving away from the ego position. Paths with destinations closer to the ego position cross less misclassified cells and thus show few distortion. Paths going further accumulate more error as they cross more misclassified cells. In other words, the outermost cells will influence only few paths, while cells closer to the ego position will affect all paths going downstream.

3) *Uncertainty*: On the binary GT, paths avoid occupied cells whenever possible by taking detours through free space. However, paths on the inference OG may need to cross cells with uncertain classification, leading to an increased cost. Scenes 5 and 6 from Table I show how uncertainty causes slight distortions to paths crossing the areas of uncertainty.

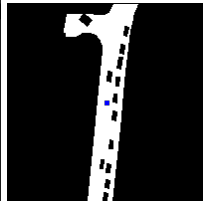
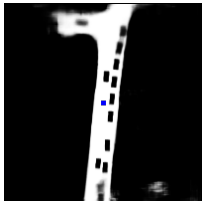



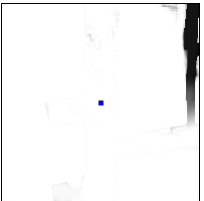

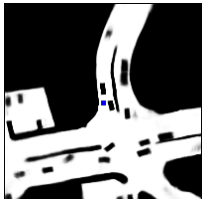



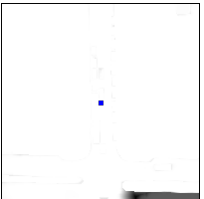

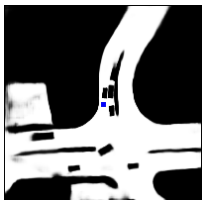
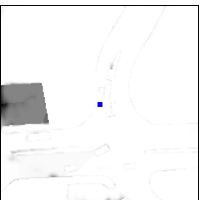
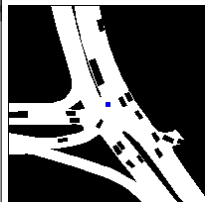

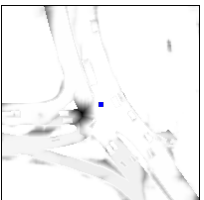
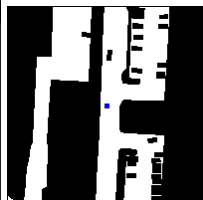

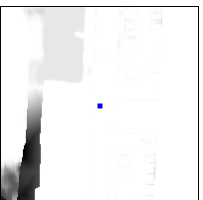

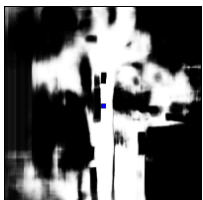
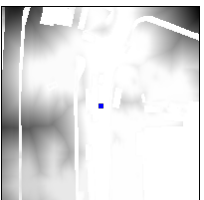
Sc	Ground Truth	Inference	Distortion	Sc	Ground Truth	Inference	Distortion
1				2			
	IoU: <b>0.882</b> MSE: 0.012 PFC-MSE: 0.838 Max distortion: 5.021				IoU: <b>0.882</b> MSE: 0.023 PFC-MSE: 163.4 Max distortion: 41.47		
3				4			
	IoU: <b>0.948</b> MSE: 0.017 PFC-MSE: 0.081 Max distortion: 4.191				IoU: <b>0.947</b> MSE: 0.012 PFC-MSE: 8.863 Max distortion: 16.90		
5				6			
	IoU: 0.952 MSE: 0.015 PFC-MSE: <b>2.565</b> Max distortion: 9.234				IoU: 0.758 MSE: 0.078 PFC-MSE: <b>2.658</b> Max distortion: 12.61		
7				8			
	IoU: 0.771 MSE: 0.073 PFC-MSE: <b>268.8</b> Max distortion: 56.81				IoU: 0.436 MSE: 0.387 PFC-MSE: <b>301.5</b> Max distortion: 65.86		

TABLE I: Collection of eight scenes extracted from our dataset, showing the Ground Truth (GT) alongside the corresponding inference and distortion grid for each scene. The two sets of scenes 1-2 and 3-4 have similar IoU and MSE scores while differing in their PFC-MSE score. The two sets of scenes 5-6 and 7-8 have similar PFC-MSE but different IoU and MSE scores. For IoU higher is better, while for PFC-MSE and MSE lower is better. Distortions grids are weighed by the occupancy probability. The color scales is specific to each grid, the maximum value of the distortion grid (the darkest cell value) is shown to account for the color scale.

When a possible detour around an uncertain area is possible, the distortion caused is low. However if no detour exists, crossing the uncertain area leads to a greater distortion.

### B. Correlation between PFC-MSE and other metrics

We have decided to compare PFC-MSE against IoU and MSE as they are the most widely used metrics in the literature. Figure 2 shows the correlation between PFC-MSE and the IoU. The correlation between PFC-MSE and MSE is not shown as we observed a strong correlation between IoU and MSE in our dataset, therefore making it redundant. PFC-MSE increases exponentially therefore we represented it using a logarithmic scale. We have observed no correlation between PFC-MSE and metrics such as IoU, MSE, SSIM, Gaussian Pyramid MSE, Gaussian Pyramid SSIM, CW-SSIM [8] and IW-SSIM [19]. This lack of correlation shows that PFC-MSE is not a surrogate of existing metrics and

exhibits an original behavior.

### C. Comparison with existing metrics

Table I shows 8 scenes from our dataset illustrating the comparison with IoU and MSE and highlighting the benefits of PFC-MSE.

1) *Similar IoU - Different PFC-MSE*: All three metrics give good scores to scenes 1 and 3 as the cell-wise error between the inference and the GT is low and the topology accurate. However, PFC-MSE behaves very differently in scenes 2 and 4 due to the changes in topology, impacting navigation on the OGs. PFC-MSE penalizes the incorrectly blocked road on the left of the OGs in scene 2 and the blocked road at the bottom of the OGs in scene 4. Compared to IoU and MSE, PFC-MSE is able to measure how the error changes the topology and how it disturbs navigation on the grid, independently of the cell-wise errors. Thus, PFC-MSE

penalizes more scenes 2 and 4.

2) *Similar PFC-MSE - Different IoU*: PFC-MSE evaluates similarly scenes 5 and 6 from Table I but their IoU scores are different. In scene 5, the square area on the left of the OG, despite being well classified as unoccupied, shows significant distortions. At first glance, Scene 6 shows more uncertainty and misclassifications. However, these errors have little impact on the topology of the OG. PFC-MSE therefore gives a better score than IoU and MSE. While uncertain inferences disrupt navigation and lead to increased costs, they are not as detrimental as classification errors that impact the access to free space areas.

PFC-MSE evaluates scenes 7 and 8 as equally inaccurate, while IoU and MSE evaluation give high scores to scene 7. Scene 7 contains an inference error resulting in the isolation of an area of free space that, although disconnected from the rest, is drivable space and should have been classified as such. The erroneously classified areas in scene 8, despite the scene having a significantly larger free space, is comparable in proportion to the errors observed in scene 7. This indicates that the extent of errors in scene 8, relative to the free space area, is similar to that of scene 7.

Through these examples, we show how PFC-MSE differs from other metrics and is more suited to compare probabilistic OGs with navigation in mind. By design, it is able to assert how differences between two OGs will change the navigation behavior of a vehicle and penalizes it.

## V. CONCLUSION

This paper introduces PathFinding Cost MSE (PFC-MSE), a new navigation-based metric for occupancy grid similarity evaluation. This metric addresses the limitations of existing metrics by evaluating the differences of pathfinding behaviour between the ground truth and the inference occupancy grids. Unlike other techniques, it takes into account the occupancy probability of each cells. Using a 10,000 scenes dataset, we confirmed that our metric is sensitive to changes in topology while being insensitive to other perturbations that do not significantly alter the navigation algorithm. Inference errors closer to the ego vehicle get more penalized than far away ones since they influence fewer trajectories. We showed that PFC-MSE is not correlated to other metrics, serving as compelling evidence that PFC-MSE exhibits a distinct behavior while taking into account the particular nature and usage of probabilistic OGs.

Two limitations arise from our work that could be addressed in future research. Firstly, Dijkstra pathfinding algorithm is insufficient for emulating navigation behaviors constrained by a kinematic or a geometrical model. Using cost from the graph-based shortest path as the only criterion for choosing the optimal path is a simplified approach. Incorporating the smoothness of the generated trajectory could be a valuable improvement to further shorten the gap between the emulated navigation and the real-world system.

Secondly, PFC-MSE is limited to single channel probabilistic OGs. It currently does not account for multiple

classes of occupancy (e.g. static, dynamic and unknown). The addition of these new classes would require managing classification errors with weights depending on the type of misclassification (e.g. a pedestrian misclassified as a crosswalk is worse than a car misclassified as a bus).

## ACKNOWLEDGMENT

This work has been supported by the PRISMA project, co-financed by the French "Grand Défi" on Trustworthy AI for Industry, and by the "Institut de Recherche Technologique NanoElec" project, founded by the French program "Investissement d'Avenir" ANR-10-AIRT-05.

## REFERENCES

- [1] A. Elfes, "Using occupancy grids for mobile robot perception and navigation," *Computer*, vol. 22, no. 6, pp. 46–57, 1989.
- [2] A. Pandharipande, C.-H. Cheng, J. Dauwels, S. Z. Gurbuz, J. Ibanex-Guzman, G. Li, A. Piazzoni, P. Wang, and A. Santra, "Sensing and machine learning for automotive perception: A review," *IEEE Sensors Journal*, pp. 1–1, 2023.
- [3] H. Caesar, V. Bankiti, A. H. Lang, S. Vora, V. E. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan, and O. Beijbom, "nusScenes: A multimodal dataset for autonomous driving," in *CVPR*, 2020.
- [4] M. A. Diaz-Zapata, D. S. González, Ö. Ercent, J. Dibangoye, and C. Laugier, "Laptnet-fpn: Multi-scale lidar-aided projective transform network for real time semantic grid prediction," *arXiv preprint arXiv:2302.06414*, 2023.
- [5] I. Varsadan, A. Birk, and M. Pfingsthorn, "Determining map quality through an image similarity metric," in *RoboCup 2008: Robot Soccer World Cup XII*, L. Iocchi, H. Matsubara, A. Weitzenfeld, and C. Zhou, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 355–365.
- [6] G. Zhai and X. Min, "Perceptual image quality assessment: a survey," *Science China Information Sciences*, vol. 63, 11 2020.
- [7] Z. Wang and Q. Li, "Information content weighting for perceptual image quality assessment," *IEEE Transactions on Image Processing*, vol. 20, no. 5, pp. 1185–1198, 2011.
- [8] M. P. Sampat, Z. Wang, S. Gupta, A. C. Bovik, and M. K. Markey, "Complex wavelet structural similarity: A new image similarity index," *IEEE Transactions on Image Processing*, vol. 18, no. 11, pp. 2385–2401, 2009.
- [9] A. Huletski, D. Kartashov, and K. Krinkin, "Evaluation of the modern visual slam methods," 11 2015, pp. 19–25.
- [10] D. Han, Y. Li, S. Tao, and L. Zhenyang, "Multi-objective optimization of loop closure detection parameters for indoor 2d simultaneous localization and mapping," *Sensors*, vol. 20, p. 1906, 03 2020.
- [11] O. Wulf, A. Nuchter, J. Hertzberg, and B. Wagner, "Ground truth evaluation of large urban 6d slam," 10 2007, pp. 650 – 657.
- [12] J. Pellenz, "Mapping and map scoring at the robocuprescue competition," 01 2008.
- [13] A. Wagan, A. Godil, and X. Li, "Map quality assessment," pp. 278–282, 08 2008.
- [14] G. Fontana, M. Matteucci, and D. G. Sorrenti, *Rawseeds: Building a Benchmarking Toolkit for Autonomous Robotics*. Cham: Springer International Publishing, 2014, pp. 55–68. [Online]. Available: [https://doi.org/10.1007/978-3-319-00272-9\\_4](https://doi.org/10.1007/978-3-319-00272-9_4)
- [15] B. Balaguer, S. Balakirsky, S. Carpin, and A. Visser, "Evaluating maps produced by urban search and rescue robots: Lessons learned from robocup," *Autonomous Robots*, vol. 27, pp. 449–464, 11 2009.
- [16] S. Schwertfeger and A. Birk, "Map evaluation using matched topology graphs," *Autonomous Robots*, vol. 40, 09 2015.
- [17] J. Zhang, X. Wang, L. Xu, and X. Zhang, "An occupancy information grid model for path planning of intelligent robots," *ISPRS International Journal of Geo-Information*, vol. 11, no. 4, 2022. [Online]. Available: <https://www.mdpi.com/2220-9964/11/4/231>
- [18] E. W. Dijkstra, "A note on two problems in connexion with graphs," *Numerische Mathematik*, vol. 1, pp. 269–271, 1959.
- [19] Z. Wang and Q. Li, "Information content weighting for perceptual image quality assessment," *IEEE Transactions on Image Processing*, vol. 20, no. 5, pp. 1185–1198, 2011.