



HAL
open science

Next-Best-View selection from observation viewpoint statistics

Stéphanie Aravecchia, Antoine Richard, Marianne Clausel, Cédric Pradalier

► **To cite this version:**

Stéphanie Aravecchia, Antoine Richard, Marianne Clausel, Cédric Pradalier. Next-Best-View selection from observation viewpoint statistics. International Conference on Intelligent Robots and Systems (IROS), IEEE, Oct 2023, Detroit, Michigan/USA, United States. pp.10505-10510, 10.1109/IROS55552.2023.10341982 . hal-04128252

HAL Id: hal-04128252

<https://hal.science/hal-04128252>

Submitted on 14 Jun 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.

Next-Best-View selection from observation viewpoint statistics

Stéphanie Aravecchia¹, Antoine Richard², Marianne Clausel³, Cédric Pradalier¹

Abstract—This paper discusses the problem of autonomously constructing a qualitative map of an unknown 3D environment using a 3D-Lidar. In this case, how can we effectively integrate the quality of the 3D-reconstruction into the selection of the Next-Best-View? Here, we address the challenge of estimating the quality of the currently reconstructed map in order to guide the exploration policy, in the absence of ground truth, which is typically the case in exploration scenarios. Our key contribution is a method to build a prior on the quality of the reconstruction from the data itself. Indeed, we not only prove that this quality depends on statistics from the observation viewpoints, but we also demonstrate that we can enhance the quality of the reconstruction by leveraging these statistics during the exploration. To do so, we propose to integrate them into Next-Best-View selection policies, in which the information gain is directly computed based on these statistics. Finally, we demonstrate the robustness of our approach, even in challenging environments, with noise in the robot localization, and we further validate it through a real-world experiment.

I. INTRODUCTION

The primary objective of autonomous exploration is to answer the question "where to go next?", but this question is intrinsically linked to what we are exploring and our purpose or goals. When a robot explores an unknown space, the complete volume is unknown at the beginning of the exploration. When the robot moves in the space, it gathers information and updates a map. In robotics, autonomously exploring environments consists in finding the next most interesting area to visit, i.e. determining a goal destination. This goal is usually selected by optimizing a policy, usually balancing information gain with some cost. In this work, we are interested in autonomous exploration tasks in large scale sparse environments, that is, where the proportion of the volume containing objects is small with respect to the scale of the scene. We are particularly interested in large scale natural environments, such as parks or forests, sometimes called "unstructured" [1], [2], in opposition to environments containing large geometrical shapes, such as buildings, called "structured". In a large scale sparse unstructured environment, computing an accurate 3D map with a 3D-Lidar is a challenging task. Among the main difficulties, we can point out the sparsity of the data and the errors in the laser measurements. This is amplified by the very complex nature of the trees (the small surfaces they offer to the laser, the leaves that act as semi-transparent reflectors, etc. [1]–[3]). The localization of the robot in the map is an additional

source of errors because the 3D-reconstruction is a probabilistic accumulation of all the point-clouds transformed in the localization frame.

When the objective of the exploration is constructing an accurate map, a central question is "how accurate is the map currently, in each section that is currently mapped?". Additionally, when working in natural environments, there is generally a lack of any ground-truth. This raises the second question: "how can we derive an exploration policy to improve the quality of a map, without access to any ground-truth?".

In this study, we propose a methodology to compare the reconstruction to the ground-truth at a local level. Then, we propose several computationally light viewpoint statistics, and a methodology to prove that they are indicators of the local reconstruction quality. Those indicators point out which areas are worth re-observing and how to re-observe them. Conversely, they allow discarding areas that are not worth re-observing, i.e. areas that do not significantly improve the reconstruction quality. We demonstrate statistically that the quality of the map can be estimated from statistics collected from the observations themselves. Finally, we integrate those statistics in the choice of the Next-Best-View in an exploration task, demonstrating that those statistics can indeed be used in an exploration policy to improve the overall map quality.

To summarize, our contributions are:

- a methodology to compare reconstruction to ground-truth at a local level;
- the proposition of four indicators of the map quality based on viewpoint statistics only;
- a methodology to validate that those indicators are valid indicators of the map quality;
- the integration of those indicators in Next-Best-View policies.

II. RELATED WORK

A. Map building

In this paper, we focus on the most common map representation for outdoor robotics: volumetric maps, i.e. 3D-grids. Such a map is a discretization of the space in voxels, each voxel containing some information, such as its occupancy likelihood. [4] presented Octomap, a method to build and store an octree instead of a 3D-grid, saving memory and computation. The open-source Octomap library implements the complete probabilistic map building process. The map is constructed in a given map frame, and every point-cloud in the Lidar frame updates the map. For each point in the

¹ are with IRL2958 GT-CNRS, 2 rue Marconi, 57070 Metz, France
first.last@georgiatech-metz.fr

² is with University of Luxembourg, Luxembourg

³ is with Université de Lorraine, Nancy, France

point-cloud, a ray-casting operation is performed. Between the robot and the returned point, the probability of occupancy of the leaves in the octree is decreased. The returned points are updated as occupied, their probability of occupancy is increased. We use Octomap to build the probabilistic volumetric map denoted as “reconstruction” in this paper.

It is also important to note that the quality of the map is intrinsically linked to the precision of the robot localization. Since the point-clouds are expressed in the Lidar frame (which is based on the robot’s localization), any error in the transformation from the Lidar frame to the map frame will lead to errors in the map [1]. For those reasons, we evaluate the robustness of our method with respect to the localization precision, with different level of noise in simulation, and with a real experiment.

B. Next-Best-View Policies

In autonomous exploration, traditional policies consist in setting a goal destination within the known volume, but with at least one unknown neighbor. This frontier-based exploration was first introduced by [5], where the goal is the nearest accessible frontier point (called *closest-frontier* in this work). Another traditional policy is to sample randomly the goal among the frontier points (*random-frontier*). Similarly, [6] set goals by filtering and clustering frontier points at different levels in Octomap’s octree. In that case, however, the overall objective is to rapidly decrease the percentage of unknown voxels in the map. By contrast, in this work, the quality of the map is a central feature of the overall objective. The policy is therefore intrinsically linked to the robot perception, making this a *Next-Best-View* (NBV) problem: the objective of the next goal selection is a balance between the need to maximize the efficiency of the perception (the information gain), and some cost. Often, in that case, exploration policies are linked to simultaneous localization and mapping, and the gain is formulated to reduce both the uncertainty in the map and in the localization, as in [7]. In this work, we consider that the localization is an input to the mapping problem.

Commonly, improving the 3D-reconstruction quality can be seen as an inspection task: the focus is on the reconstruction of the surface, and not on the exploration of the 3D volume. In that context, the NBV is generally the candidate view that maximizes the expected gain on the reconstruction quality. [8] and [9] select the NBV to create a high quality surface of a single small scale object. Both methods require heavy calculations, such as estimating at runtime the surfaces and their normals. This prevents them from scaling to large environments. The scaling is tackled by [10] and [11]. They focus on selecting the NBV to improve the quality of the 3D-reconstruction of large and complex structures. Whereas [10] selects the NBV after performing an initial scan to obtain a rough model of the structure to reconstruct, [11] estimates at runtime the quality of the surfaces through the computation of surface points, where each point is assigned a confidence level via the integration of Truncated Signed Distance Fields (TSDFs).

However, some methods focus on exploring a 3D-volume and reconstructing a 3D-surface at the same time. For example, [12] explores an unknown volume, and proposes several information gain formulations to select the NBV, all aiming to be informative to the reconstruction by using the entropy of the voxels. [13] explores a large scale unknown volume with multiple UAVs and reconstructs a 3D-model of a large scale object of interest in this volume. The NBV is guided by the expected information gain on the 3D-reconstruction with an estimation of the current surface with TSDFs.

All those methods assume in their information gain formulations that the object of interest is large with respect to the scale of the scene. That is the main limitation of those methods: they make assumptions on the scene or object to reconstruct, and are not intended to work in large scale sparse unstructured environments, with their specific challenges. Indeed, in that type of environment, not only are there few objects to reconstruct in a large scene, but also the noise in the Lidar and in the localization is important, the sampling of the objects in the 3D-grid is not uniform, we face unpredictable occlusions, etc... On the contrary, our method makes no such restrictive assumptions, and our NBV policies rely on observation viewpoint statistics only that discard the need for any assumption on the scene to reconstruct.

III. METHOD

The objective of our work is to prove that some indicators, based on viewpoint statistics only, are presumptive of the map quality, and to show their interest when integrated in the choice of the NBV. First, we explain our methodology to compare 3D-reconstruction to ground-truth, and to measure the quality of the reconstruction at a local level. To do so, we discretize the space into cuboid regions and calculate a reconstruction metric for each cuboid. Second, we define the indicators, i.e the statistics from the observation viewpoints, in the same cuboid regions. Third, we evaluate those indicators with respect to their ability to be presumptive of the map quality. Finally, we integrate those indicators in the NBV selection. All the code is available open-source¹.

A. Cuboid region comparison

This section explains how we compare the reconstruction to the ground-truth locally, on cuboid regions. To enable this local comparison, we discretize the volume the reconstruction and the ground-truth represent into two 3D-grids, of the same resolution RES and in the same reference frame R_f . A local region is then a cuboid C . To compare the two local regions, we compute metrics between intersecting cuboids. Alg. 1 summarizes this section.

Cuboid extraction: The 3D-reconstruction is obtained from the full probability map constructed by Octomap, the ground-truth is either the mesh of the scene used in simulation, or the point-cloud obtained when scanning the area with a Total Station (Sec. IV-A). For the ground-truth, in the simulated environment, we first slice the ground-truth

¹<https://github.com/stephanie-aravecchia/obs-stats-NBV>

mesh with horizontal planes, with a vertical space of res , our spatial resolution (with $res < RES$). In each plane, we then calculate the intersection of the mesh and the plane, and we store it in a 3D matrix of voxel size res . Each voxel on the intersection is occupied and has a value of 1.0, all the remaining voxels are empty and a value of 0.0. In the real experiment, we construct a 3D matrix of voxel size res containing zeros. We iterate on the point-cloud from the Total Station, and for each point, we set the value of the corresponding voxel to 1.0. Our ground-truth dataset, \mathcal{D}_{gt} , is then a 3D matrix of size (h_1, w_1, n_1) . Its associated volume in space, in the reference frame R_f , is the bounding-box B_{box_1} . For the reconstruction, we first create a full probability map with Octomap in the same reference frame R_f , and with the same resolution res . From this octree, and its bounding-box in R_f , we initialize a 3D matrix as unknown space, i.e. values of 0.5, corresponding to the equal probability of the voxel to be occupied or empty. We then iterate on all the leaves in the octree. For each leaf, we set the probability of its associated voxels in the 3D matrix to the probability of the leaf (the unknown space is implicitly described in Octomap with absent leaves). We finally obtain the reconstruction dataset, \mathcal{D}_{rec} , a 3D matrix of size (h_2, w_2, n_2) , with a voxel resolution res , with its associated bounding-box, B_{box_2} , in the same reference frame, R_f .

Finally, we do the comparison in $B_{\text{box}_1} \cap B_{\text{box}_2}$. First, we load the intersection of both datasets in two 3D matrices, M_{gt} and M_{rec} : a voxel v_{gt}^{ijk} from M_{gt} corresponds to the same volume in R_f than v_{rec}^{ijk} from M_{rec} . Second, we go through both 3D matrices, and compare cuboid region by cuboid region, the reconstruction and the ground truth. A cuboid region is a group of $n \times n \times n$ voxels in each 3D matrix, and is noted C_{gt} or C_{rec} . Those cuboid regions are in fact large voxels of size $RES = n \times res$, in the 3D-grid $B_{\text{box}_1} \cap B_{\text{box}_2}$, in R_f , and each element in C_{rec} and C_{gt} stores the occupancy likelihood of its corresponding voxel of size n in R_f .

Cuboid metrics: Since this method is developed with the objective to work also on large scale environments with sparse objects, the reconstructed volume may contain more empty space than actual objects to reconstruct. A measure that would give information on occupied space only may not be representative of the complete volume. For this reason, we found it interesting to measure not only how well objects have been reconstructed, but also how well the empty space has been reconstructed. To do so, we first define two sets: \mathcal{U}_{occ} , the set of the cuboids regions containing at least one occupied voxel in C_{gt} and \mathcal{U}_{empty} , its complement. Then, we measure the quality of reconstruction of the cuboids with a different metric in each set. First, to measure the quality of reconstruction of a cuboid region in \mathcal{U}_{occ} we use the surface coverage cov . We apply the classical methodology in a 3D-grid, as [12]: all the reconstructed points in C_{rec} are considered belonging to the surface if their occupancy likelihood is above a threshold. We calculate the Euclidean distance between a ground-truth point S and the closest reconstructed point P . If the distance SP is below a registration distance,

the point is considered reconstructed. The total number of ground-truth points is n_S , the total number of considered reconstructed points is n_{rec} . The surface coverage is then $cov = n_{rec}/n_S$. Conversely, when we measure the quality of reconstruction of a cuboid region in \mathcal{U}_{empty} we actually want to measure the distance between C_{rec} and a ground-truth cuboid containing only zeros. The L_1 norm provides that distance: $L_1 = \sum_{\substack{(i,j,k)=(n,n,n) \\ (i,j,k)=(0,0,0)}} C_{rec,i,j,k}$.

Algorithm 1 Reconstruction and ground-truth comparison

```

//  $\mathcal{D}_{gt}$  and  $\mathcal{D}_{rec}$  are the datasets
 $\mathcal{D}_{gt}(B_{\text{box}_1}, (h_1, w_1, n_1), R_f)$ ,  $\mathcal{D}_{rec}(B_{\text{box}_2}, (h_2, w_2, n_2), R_f)$ 
 $B_{\text{box}} \leftarrow B_{\text{box}_1} \cap B_{\text{box}_2}$ 
 $M_{gt} \leftarrow \mathcal{D}_{gt}(B_{\text{box}})$ ,  $M_{rec} \leftarrow \mathcal{D}_{rec}(B_{\text{box}})$ 
for all cuboid  $\in B_{\text{box}}$  do:
     $C_{gt} \leftarrow M_{gt}(\text{cuboid})$ ,  $C_{rec} \leftarrow M_{rec}(\text{cuboid})$ 
    if cuboid  $\in \mathcal{U}_{occ}$  then:
        cuboid.metrics  $\leftarrow$  computeCov( $C_{rec}$ ,  $C_{gt}$ )
    else:
        cuboid.metrics  $\leftarrow$  computeL1( $C_{rec}$ )
    end if
end for

```

B. Statistics from the observation viewpoints

This section describes the statistics from the observation viewpoints we consider, it explains how these statistics are computed, and introduces the new indicators we propose: the spherical variance, and the covered angular sectors, to express the diversity of the viewpoints. The underlying idea is that in sparse and unstructured environments particularly, the quality of the reconstruction should be influenced by the viewpoint diversity, because of the specific problems indicated in Sec. II-B.

We consider the following statistics: 1) n_{obs} , the number of times the cuboid region is observed; 2) r_{min} , the minimum range of the viewpoints; 3) n_{Ω} , the number of angular sectors covered by the viewpoints; 4) σ_{θ} , the spherical variance of the viewpoints. The two last are detailed later.

Viewpoint statistics update: To begin with, we describe how the statistics are stored and updated. We first construct and initialize \mathcal{G} , a 3D-grid containing observation viewpoint statistics. \mathcal{G} is in the same reference frame R_f than the reconstructed map and the ground-truth, and each element \mathcal{C}_s in this grid spatially corresponds to a cuboid region described previously (same resolution RES , same position in R_f). \mathcal{C}_s stores the viewpoint statistics of its corresponding cuboid region. Each time we receive a new point-cloud from the Lidar, we update at most once \mathcal{C}_s , the elements of \mathcal{G} , to keep track of the viewpoint statistics. We know that each point P of the point cloud is observed from the center of the Lidar, O . The laser ray is $[PO]$. We also know that each element along $[PO]$ has been observed from the same point O . We calculate the intersection of the segment $[PO]$ and the grid \mathcal{G} . For each element in this intersection, we update the statistics from the observations, considering the center of \mathcal{C}_s is observed from O . We perform the ray-casting operation to compute the viewpoint statistics regardless that the cuboid region contains an object or empty space. The update of \mathcal{G} is fast to compute and easily runs in real-time.

Angular Sectors: The objective here is to count the number of angular sectors each cuboid region has been observed from. To do so, we divide the horizontal plane (x, y) going through the center of the considered cuboid region $C = (x_c, y_c, z_c)$ into n angular sectors, represented with Ω , a boolean vector of size n . For each observation from a point $O = (x_o, y_o, z_o)$, we compute the azimuthal angle in the (x, y) plane: $\theta = \text{atan2}(y_o - y_c, x_o - x_c)$. From θ , we compute the angular sector index i in Ω : $i = \theta \times n / 2\pi$, and set Ω_i to 1. The statistic we propose is simply: $n_\Omega = \sum_{i=0}^{i=n-1} \Omega_i$.

Spherical Variance: Calculating the spherical variance, defined in [14], consists in encoding each viewpoint through its spherical angle with the coordinate axes, $\mathbf{U} = [U_x, U_y, U_z]$ defined as the unit vector. Assume now that we are given n observation viewpoints $\mathbf{U}(1), \dots, \mathbf{U}(n)$. The spherical variance is defined as follows :

$$\sigma_\theta = 1 - R/n \text{ with } R = \sqrt{X_{sum}^2 + Y_{sum}^2 + Z_{sum}^2} \quad (1)$$

where we denote

$$X_{sum}, Y_{sum}, Z_{sum} = \sum_{k=0}^{k=n} U_x(k), \sum_{k=0}^{k=n} U_y(k), \sum_{k=0}^{k=n} U_z(k) \quad (2)$$

We compute this spherical variance for each update, where we derive \mathbf{U} from $[CO]$, and then calculate the resultant length R from the history of $\mathbf{U}(n)$, from observation 0 to the current one n , in the considered cuboid region \mathcal{C}_S .

C. Next-Best-View Selection

This paragraph presents our method where the previously described grid of viewpoint statistics \mathcal{G} is integrated into the NBV selection. We implement four different policies, each of them considering the expected information gain on a single statistic: either on n_{obs} , r_{min} , n_Ω or σ_θ .

The candidates NBV are all the reachable candidates, in the known free space, in a selected range around the current robot position, R . For each target candidate T , let us consider a_T the action of moving the robot from R to T . Following [7], our NBV selection policy is based on two quantities: $E[I(a_T)]$, the expected information gain of taking a_T , and its associated cost $cost(a_T)$.

Expected Information Gain: For each candidate NBV, T , we calculate $E[I(a_T)]$.

First, we consider the current state, \mathcal{G}_0 , and the predicted state $\hat{\mathcal{G}}$. We compute $\hat{\mathcal{G}}$ by updating every visible cuboid \mathcal{C}_s of \mathcal{G}_0 with a new expected observation after simulating taking a_T (i.e. a new observation from T). For n_Ω , we update all the angular sectors in Ω_i covered by the displacement of the robot from R to T . For the three other statistics, we simply consider the new observation from T .

Second, given the two states \mathcal{G}_0 and $\hat{\mathcal{G}}$, $E[I(a_T)]$ is the accumulation of the considered statistic's updates over all visible cuboids. Given a cuboid \mathcal{C}_s , one indicator s_0^i in $[n_{obs}, r_{min}, n_\Omega, \sigma_\theta]$, in state \mathcal{G}_0 and the updated indicator \hat{s}^i in state $\hat{\mathcal{G}}$, the indicator update follows the central limit theorem $g^i = \log(\hat{s}^i) - \log(s_0^i)$ if $\hat{s}^i > s_0^i$, else $g^i = 0$ (symmetric for r_{min}).

metrics M	indicator s	threshold s^*	alternative hypothesis H_1
cov	$n_{obs}, n_\Omega, \sigma_\theta$	$n_{obs}^*, n_\Omega^*, \sigma_\theta^*$	$M[s \geq s^*] > M[s < s^*]$
cov	r_{min}	r_{min}^*	$M[s < s^*] > M[s \geq s^*]$
L_1	$n_{obs}, n_\Omega, \sigma_\theta$	$n_{obs}^*, n_\Omega^*, \sigma_\theta^*$	$M[s \geq s^*] < M[s < s^*]$
L_1	r_{min}	r_{min}^*	$M[s < s^*] < M[s \geq s^*]$

TABLE I: Details of the Mann-Whitney U tests populations and hypothesis

Finally, for one statistic: $E[I(a_T)] = 1/n \sum_{i=1}^{i=n} g^i$, where g^i depends on the considered statistic, and n is the number of visible cuboids from T .

NBV selection policy: The cost $cost(a_T)$ is a function of the distance the robot has to drive from R to T , avoiding obstacles, computed with a Breadth-first algorithm.

The NBV is then selected as in [7]. It is the candidate T , with a_T^* the action with the highest expected utility: $a_T^* = \text{argmax}(E[U(a_T)])$, where the expected utility of a_T is $E[U(a_T)] = E[I(a_T)] - \alpha \times cost(a_T)$, and α is a weighting factor.

D. Evaluation Methodology

1) *Validation of the indicators with statistical tests:* The objective of this section is to prove that the observation viewpoint statistics we consider are closely related to the map quality, and to validate that they are indeed relevant indicators of the quality of reconstruction. Each cuboid region yield a measure of the quality of its reconstruction (cov or L_1 , sec. III-A), and statistics from its observation viewpoints (n_{obs} , r_{min} , n_Ω , σ_θ , sec. III-B). Let M be the value of its measured quality and s the value of the statistic. We split the population of cuboids regions in two populations, using a threshold for the considered indicator, and we perform a statistical test to prove that the quality of reconstruction M of the population of cuboids for which the indicator s is above (or below) the threshold is significantly better than the other population.

To perform the tests, we set the threshold s^* to the median of s . Then, following [15], we compare the two populations with the one-sided non-parametric Mann-Whitney U tests indicated in Table I, available in the scipy-stats library². If the p-value is small, then we reject the null hypothesis H_0 , that the two populations of voxels are equally well reconstructed, in favor of the alternative H_1 , corresponding to a significant better reconstruction when the value of the indicator of interest is above s^* (or below, for r_{min}). In other words, a small p-value proves the considered viewpoints statistic is undeniably an indicator of the reconstruction quality.

2) *Evaluation of the NBV selection:* We evaluate several exploration tasks w.r.t their produced map quality. The map quality is evaluated globally. For each experimental exploration, we consider \mathcal{Q} , the quality of reconstruction of the complete volume, as the weighted mean of the normalized metrics calculated on the cuboids regions of \mathcal{U}_{occ} and \mathcal{U}_{empty} . If \mathcal{U}_{occ} and \mathcal{U}_{empty} contains respectively n and m elements:

$$\mathcal{Q} = \frac{1}{n+m} \left(\sum_{k=0}^{k=n} \frac{cov_k - \min(cov)}{\max(cov) - \min(cov)} + \sum_{k=0}^{k=m} \left(1 - \frac{L_{1_k} - \min(L_1)}{\max(L_1) - \min(L_1)} \right) \right) \quad (3)$$

²<https://docs.scipy.org/doc/scipy/reference/stats.html>

Finally, we want to compare different policies based on how the map quality evolves during the exploration process, that is, how \mathcal{Q} changes when the amount of unknown space in the map decreases. We compare our four NBV selection policies, where the information gain is based on n_{obs} , r_{min} , n_{Ω} or σ_{θ} (III-C), with three other goal selections: *random-frontier* and *closest-frontier* presented in II-B, and another goal selection called *random-free* based on random sampling of goals in the known space. We add the latter, because by construction our policies may sample goals in the known space, and it may not be fair to compare it only against frontier algorithms that exclusively sample goals in the frontier area close to the unknown space. Comparing to a policy that samples goals randomly inside the known space remove this potential bias in our opinion.

IV. EXPERIMENTS AND RESULTS

Our experiments are done in simulation and in the field. In both cases, we use the ROS framework and the robot is a Clearpath Husky, equipped with a 3D Lidar Ouster OS1-16.

A. Experimental Setup in Simulation

The simulator is Gazebo. A mixture of noise is applied to the Lidar to better simulate the behaviour of a Lidar in outdoor natural environments.

We generate randomly several environments. Each environment is a plane of dimension $60m \times 60m$ on which we place assets with a Poisson Cluster Point Process, to reproduce the natural spatial distribution of trees. The assets can be, with an increasing level of difficulty w.r.t the reconstruction, either rectangular cuboids, cross-extruded shapes, helicoidal cones, or randomly selected in our 15-trees library mimicking winter bare trees. Random factors are applied on the assets’ dimensions and orientation. With the same point process generation, we create four different synthetic environments, one per type of asset.

Since the quality of the reconstruction is also directly linked to the localization of the robot, we incrementally add a Gaussian noise to the perfect localization. The noise levels considered in this study are (with the standard deviations σ_p , on position, in meter, σ_q , on orientation, in radian): 0) perfect localization; 1) $\sigma_p = 0.005$, $\sigma_q = 0.005$; 2) $\sigma_p = 0.05$, $\sigma_q = 0.01$.

In this work, the sensor range is set to 20m, the map resolution to $res = 0.1m$, and the cuboid resolution to $RES = 1m$. For the NBV selection, we select candidates in a range of 20m around the robot.

We generate 12 environments (3 different point processes with 4 different assets). For each environment, for each of the 3 noise levels, we run 3 simulations with each policy (*random-frontier*, *closest-frontier*, *random-free*, ours[n_{obs}], ours[r_{min}], ours[n_{Ω}] and ours[σ_{θ}]), for a total of 756 experiments in simulation. Each experiment consists in teleporting the robot to the goal, accumulating the point-cloud, and teleporting it to the next goal. We repeat until the ratio of discovered space in the ground-plane reaches 70%.

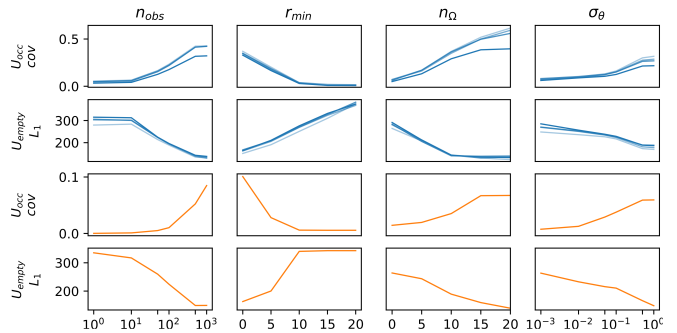


Fig. 1: Mean quality of reconstruction expressed as a function of viewpoint statistics. Blue: in simulation, orange: real experiment.

B. Experimental Setup in the field

The experimental field is a small park containing trees, of an approximate area of $1500m^2$. Although it is not possible to acquire the ground-truth of such an environment, we consider that the 3D point-cloud obtained from the scan of the environment with a Total Station Leica TS60 is precise and dense enough to be considered as ground truth. The horizontal and vertical angular resolution of the scanning is set to 0.05 degrees. To scan the area, the Total Station is placed on three different locations, to have different viewpoints on the trees. Each scan takes 40 to 60 minutes. At the end, we obtain a single consistent point-cloud from the area. To localize the robot in this area, it is equipped with a prism tracked by the Total Station. The pose is recovered in a post-processing phase. With this setup, we can consider we have a “good” localization for an outdoor robotics application. For this experiment, we drive the robot manually, and we obtain one reconstruction with Octomap (max-range 10m), with $res = 0.1m$, that we compare to the ground-truth. This experiment is designed to validate that the statistics from the observation viewpoints remain indicators of the map quality even in a real environment.

Ultimately, the maps from the simulation and the real experiment are compared to the ground-truth as described in sec. III, with a cuboid resolution of $RES = 1m$.

C. Results

Fig. 1 shows how the statistics from the observation viewpoints are linked to the quality of reconstruction. Following what could seem intuitive, the quality of reconstruction increases (cov increases and L_1 decreases) when n_{obs} , n_{Ω} or σ_{θ} increase (the viewpoints are more numerous / more diverse), and when r_{min} decreases (seen from closer). For the experiments in simulation, the figure displays the results only with the first level of noise, and the experiments are grouped by type of asset. Each line corresponds to the mean of the statistic for the 252 experiments selected. The darker the line, the more challenging the type of environment. All the curves follow the same trend, even though some statistics seem more robust when the environment is more difficult (r_{min} in particular). The figure also displays the information on the real experiment. The trend of the curves remains the same, despite the different vertical scale on cov , due to the

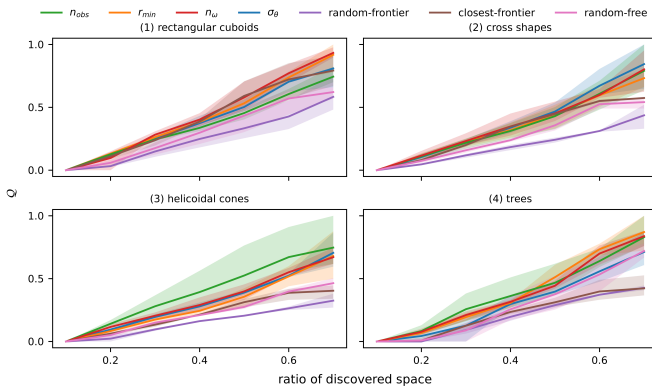


Fig. 2: Quality of reconstruction expressed as a function of the proportion of discovered space in the ground plane. The plots are arranged with an increasing difficulty in the scene. In each figure, the line is the mean of the experiments, the area is between the min and the max. Our proposed methods (n_{obs} , r_{min} , n_{Ω} , σ_{θ}) are compared to the baselines (*random-frontier*, *closest-frontier*, *random-free*).

fact that the real environment is more difficult to reconstruct than the simulation. Although we do not show it here due to lack of space, for the experiments in simulation, the curves are nearly superimposed if we fix the type of asset, and vary the noise in the localization, indicating that the statistics are robust to noise.

To prove that the trend we see is statistically significant, we perform the statistical tests described in III-D for each considered indicator. To avoid any bias in the viewpoint statistics, we perform the tests only on the *random-frontier* and *random-free* experiments. The results of the 162 experiments are presented in Tab. II. We use a common threshold for the p-value: 0.05. In simulation, generally, more than 87% of the tests leads to $p < 0.05$, rejecting H_0 in favor of H_1 : this validates statistically that our hypothesis that the viewpoint statistics we consider are indicators of the map quality, even in challenging environment, and with noise in the localization. In the real experiment, $p < 0.05$ for all the tests, validating that the statistics’ ability to indicate the quality of the reconstruction remain true in a real environment. We cannot see any trend in the tests where $p > 0.05$, they are spread on the different type of assets and on the different level of noise. The lower reliability of n_{obs} and n_{Ω} at indicating the quality of reconstruction of occupied space suggest that a reliable policy should optimize on all the statistics and not only one at the time, as we do currently. This is left for future work.

	<i>cov</i>	L_1		<i>cov</i>	L_1
n_{obs}	0.74	0.90	r_{min}	0.89	1.00
n_{Ω}	0.74	0.87	σ_{θ}	0.88	0.93

TABLE II: Ratio of statistical tests with pvalue $p < 0.05$

Fig. 2 displays how the quality of the map Q evolves during the exploration phase, when the space is discovered. The results are shown for the experiments in four environments, where the asset distribution is the same, but the assets vary, and the perfect localization (84 experiments). Generally, the curves representing our methods are higher than the others. This indicates that basing NBV policies on viewpoint statistics significantly improves the map quality

during the exploration process. This is particularly true when the difficulty in the scene increases.

V. CONCLUSION

We have demonstrated that the four observation viewpoint statistics we consider in this work are undeniably indicators of the map quality, while being extremely simple to compute in real-time for any environment. We also have demonstrated that basing the choice of the Next-Best-View on one of those indicators allows exploring while optimizing implicitly, but significantly, the quality of the 3D-reconstruction. So far, we have used only one observation viewpoint statistic at a time. Our intuition, consistent with statistical testing, is that building upon the four statistics, we will be able to improve the prediction of the quality of reconstruction, and further integrate that prediction in the Next-Best-View selection. In doing so, the definition of the information gain may become more challenging. Future work will focus on solving this problem with Deep Reinforcement Learning.

REFERENCES

- [1] G. Chahine *et al.*, “Mapping in unstructured natural environment: a sensor fusion framework for wearable sensor suites,” *SN Applied Sciences*, no. 5, pp. 1–14.
- [2] P. Babin *et al.*, “Large-Scale 3D Mapping of Subarctic Forests,” *Springer Proceedings in Advanced Robotics*, vol. 16, pp. 261–275, 2021.
- [3] J. Laconte *et al.*, “Lidar measurement bias estimation via return waveform modelling in a context of 3D mapping,” *Proceedings - IEEE International Conference on Robotics and Automation*, vol. 2019-May, pp. 8100–8106, 2019.
- [4] A. Hornung *et al.*, “OctoMap: An efficient probabilistic 3D mapping framework based on octrees,” *Autonomous Robots*, vol. 34, no. 3, pp. 189–206, 2013.
- [5] B. Yamauchi, “Frontier-based approach for autonomous exploration,” *Proceedings of IEEE International Symposium on Computational Intelligence in Robotics and Automation, CIRA*, pp. 146–151, 1997.
- [6] A. Batinovic *et al.*, “A Multi-Resolution Frontier-Based Planner for Autonomous 3D Exploration,” *IEEE Robotics and Automation Letters*, vol. 6, no. 3, pp. 4528–4535, 2021.
- [7] C. Stachniss *et al.*, “Information gain-based exploration using rao-blackwellized particle filters,” *Robotics: Science and Systems*, vol. 1, pp. 65–72, 2005.
- [8] N. A. Massios and R. B. Fisher, “A Best Next View Selection Algorithm Incorporating a Quality Criterion,” *Department of Artificial Intelligence, University of Edinburgh*, 1998.
- [9] S. Kriegel *et al.*, “Efficient next-best-scan planning for autonomous 3D surface reconstruction of unknown objects,” *Journal of Real-Time Image Processing*, vol. 10, no. 4, 2015.
- [10] R. Almadhoun *et al.*, “Guided next best view for 3D reconstruction of large complex structures,” *Remote Sensing*, vol. 11, no. 20, pp. 1–20, 2019.
- [11] S. Song and S. Jo, “Surface-Based Exploration for Autonomous 3D Modeling,” *Proceedings - IEEE International Conference on Robotics and Automation*, pp. 4319–4326, 2018.
- [12] S. Isler *et al.*, “An information gain formulation for active volumetric 3D reconstruction,” *Proceedings - IEEE International Conference on Robotics and Automation*, pp. 3477–3484, 2016.
- [13] G. Hardouin *et al.*, “Next-Best-View planning for surface reconstruction of large-scale 3D environments with multiple UAVs,” *IEEE International Conference on Intelligent Robots and Systems*, pp. 1567–1574, 2020.
- [14] D. E. Tyler, “Statistical Analysis for the Angular Central Gaussian Distribution on the Sphere,” *Biometrika*, vol. 74, no. 3, pp. 579–589, jan 1987.
- [15] M. P. Fay and M. A. Proschan, “Wilcoxon-Mann-Whitney or T-test? on assumptions for hypothesis tests and multiple interpretations of decision rules,” *Statistics Surveys*, vol. 4, pp. 1–39, 2010.