2018 IEEE International Conference on Robotics and Automation (ICRA)
May 21-25, 2018, Brisbane, Australia

Visual Saliency–aware Receding Horizon Autonomous Exploration with Application to Aerial Robotics

Tung Dang, Christos Papachristos, and Kostas Alexis

Abstract— **This paper presents a novel strategy for autonomous visual saliency–aware receding horizon exploration of unknown environments using aerial robots. Through a model of visual attention, incrementally built maps are annotated regarding the visual importance and saliency of different objects and entities in the environment. Provided this information, a path planner that simultaneously optimizes for exploration of unknown space, and also directs the robot's attention to focus on the most salient objects, is developed. Following a two–step optimization paradigm, the algorithm first samples a random tree and identifies the branch maximizing for new volume to be explored. The first viewpoint of this path is then provided as a reference to the second planning step. Within that, a new tree is spanned, admissible branches arriving at the reference viewpoint while respecting a time budget dependent on the robot endurance and its environment exploration rate are found and evaluated in terms of reobserving salient regions at sufficient resolution. The best branch is then selected and executed by the robot, and the whole process is iteratively repeated. The proposed method is evaluated regarding its ability to provide increased attention toward salient objects, is verified to run onboard a small aerial robot, and is demonstrated in a set of challenging experimental studies.**

I. INTRODUCTION

Recent breakthroughs in aerial robotics research have enabled their wide utilization in a variety of critical tasks including those of infrastructure inspection, precision agriculture, search and rescue, and surveillance. In most such applications, autonomous operation typically refers to the task of information gathering and mapping, either based on a prescribed inspection path or based on exploration path planning functionality. However, mapping in that sense treats the problem in a manner that is agnostic to the specific visual importance of different objects and entities in the environment. Unless explicitly prescribed (e.g. through predefined semantics), salient objects such as a painting, a warning sign, a crack on a surface, or a person, are treated as equally important as flat plain walls. Significant progress can be achieved if the visual saliency of different entities is accounted for at the planning and viewpoint selection stage. Salient objects can then get the attention that is required to perceive and map them in detail, while self–similar parts of the environment can be rapidly explored.

Especially for the case of an aerial robot aiming to operate autonomously in order to efficiently explore and map unknown, large and complex environments, the ability to know where to focus its perceptual attention is critical in order to achieve optimized information gathering. This is *especially* the case when the endurance of the system does not permit a fully and uniformly detailed mapping, and the object classes to be detected are not known a priori.

Fig. 1. Instance of visual saliency–aware exploration.

To enable autonomous exploration that is aware of the visual importance of different objects in the environment, in this research we first looked at the field of visual attention modeling (*"finding a function that minimizes the error on eye fixation prediction"* [1]) and saliency [2–10]. To incorporate the relevant information in a representation that enables efficient planning, saliency–annotated volumetric mapping was developed and further incorporated the concept of Inhibition Of Return (IOR) [6, 11]. IOR is motivated by research in neuroscience and refers to the relative suppression of orienting toward objects that had recently been the focus of attention, therefore encouraging orienting towards novelty.

Provided this mapping functionality, the proposed *visual saliency–aware receding horizon exploration* path planning strategy simultaneously optimizes for efficient online exploration of unknown environments, and directing the robot's perceptual attention (and online mapping operation) towards the most salient objects and entities of the environment. For every iteration of the planner, an optimized sequence of viewpoints for exploration of unknown volume is derived based on the principles of sampling–based planning. The first viewpoint of this sequence is selected to be visited, but the path to reach it is computed through a second planning step that optimizes for directing the robot's sensors towards the visually salient parts of the map. Subsequently, the whole process is repeated in a receding horizon fashion. During the entire robot operation, the saliency–annotated occupancy map gets updated based on the new sensor stimuli.

This material is based upon work supported by the Department of Energy under Award Number [DE-EM0004478].

The authors are with the Autonomous Robots Lab, University of Nevada, Reno, 1664 N. Virginia, 89557, Reno, NV, USA tung.dang@nevada.unr.edu

The proposed saliency–aware exploration strategy is evaluated both in simulation, as well as in challenging experimental studies. Computational complexity analysis is further provided. All experiments were conducted indoors with the robot relying only on onboard visual–inertial localization. The planned paths, the derived saliency–based volumetric maps, and metrics demonstrating the increased focus of the robot's attention towards salient areas are presented. The algorithm implementation is released as an open source package accompanied with a relevant dataset [12].

The remainder of the paper is structured as follows: Section II provides an overview of related work, while the considered problem is defined in III. The proposed algorithm is detailed in Section IV. Simulation studies are presented in Section V, followed by experimental evaluation in Section VI. Finally, conclusions are drawn in Section VII.

II. RELATED WORK

A rich body of work exists for the problems of robotic exploration and informative path planning [13–22]. Particularly in the field of exploration, early work includes the sampling of "next–best–views" [21], frontiers–based exploration [22] and more recent efforts [18–20]. However, exploration in that manner operates in ways that are agnostic towards the varying visual importance of different objects. This limitation can be alleviated if a model of visual attention is exploited by the robot planner. Indeed, the computer vision, neuroscience, psychology and developmental robotics communities have put significant effort in developing insightful models of saliency and pre–attentive vision through cognitive, Bayesian, information theoretic, graphical, spectral analysis, and pattern classification approaches [2–10], but only few efforts focus on mobile robotic information gathering applications [23]. The contribution proposed in this paper aims to bridge this gap and develop a path planning strategy to simultaneously optimize for the extrinsic objective of unknown environment exploration, as well as intrinsic rewards related to the visual saliency of different objects and entities of the environment. Through this process, a robotic behavior focused on autonomous exploration, but also actively pursuing to gather information which corresponds to salient subsets of the environment, is achieved.

III. PROBLEM DESCRIPTION

The problem considered in this work is that of exploring a bounded volume $V^E \subset \mathbb{R}^3$, while *simultaneously* aiming to optimize the information sampling over the areas presenting visual saliency given a model of an onboard camera. The exploration problem is casted globally and refers to determining which parts of the initially unmapped space $V_{unm} \overset{init}{=} V^E$ are free $V_{free} \subset V^E$ or occupied $V_{occ} \subset V^E$. In this work, the volume is discretized in an occupancy map $\mathcal{M}$ that is comprised of cubical voxels $m \in \mathcal{M}$ with edge length r. Since for most sensors –including cameras– perception stops at surfaces, sometimes hollow spaces or narrow pockets cannot be explored, thus leading to a residual volume:

Definition 1 (Residual Volume) Let Ξ be the simply connected set of collision free configurations and $\bar{\mathcal{V}}_m^E \subseteq \Xi$ the set of configurations from which the voxel m can be perceived. The residual volume is $V_{res}^E = \bigcup_{m \in \mathcal{M}} (m |\ \bar{\mathcal{V}}_m^E = \emptyset)$.

The exploration problem is then defined as:

Problem 1 (Volumetric Exploration) Given a bounded volume V^E, find a collision free path σ starting at an initial configuration $\xi_{init} \in \Xi$ that leads to identifying the free and occupied parts V_{free}^E and V_{occ}^E, such that there does not exist any collision free configuration from which any piece of $V^E \setminus \{V_{free}^E, V_{occ}^E\}$ could be perceived. Thus, $V_{free}^E \cup V_{occ}^E = V^E \setminus V_{res}^E$. Feasible paths σ of this problem are subject to the limited Field of View (FoV) of the sensor, its effective sensing distance, and robot dynamics constraints.

Volumetric exploration is considered as an extrinsic objective for the robot to be optimized with priority. Maximization of information sampling over visually salient areas can then be casted as a local nested path optimization given the initial and final robot configurations ξ_{init} and ξ_{final} corresponding to sequential viewpoints of the exploration planning step.

Problem 2 (Saliency-aware Planning subject to Time Budget Constraints) Given a bounded volume $V^S \subseteq V^E$ enclosing ξ_{init} and ξ_{final} and expected exploration rate e_r^k, find a collision free path σ^S connecting $\xi_{init} \in \Xi$ and ξ_{final}, and maximizing an information sampling objective over areas of the environment that present visual saliency, given a model of visual saliency on occupancy maps $\mathcal{S}(\mathcal{M})$. Feasible paths σ_f^S of this problem are subject to a) the limited sensor FoV and effective sensing distance, b) an allowed time budget $t_S^{\max}$, as well as c) constraints imposed by robot dynamics.

IV. PROPOSED APPROACH

The proposed method enables robots to autonomously explore unknown environments, while simultaneously directing their attention towards the most salient subsets of their environment to achieve an information–gathering optimized behavior. This saliency–aware planning is achieved through a model of visual saliency, the corresponding updates to the online reconstructed occupancy map, and the respective path planning methods. In particular, a nested optimization, two–step planning approach is considered. In the first step, a finite–depth random tree of robot configurations is sampled, and the branch that optimizes for exploration of unknown volume (*extrinsic reward*) is identified. The first vertex of this branch is selected and is used as the reference to a second planning stage. During this second step, the method samples a new random tree and identifies the branch that locally optimizes for directing the robot's attention (sensors) towards salient objects and entities (*intrinsic reward*), while ensuring arrival to the first planning layer reference, and maintaining a bounded travel cost compared to the optimal one. This branch is then executed by the robot and the whole process is repeated in a receding horizon fashion.

A camera, more specifically for this work a stereo visual–inertial unit, is used to enable robotic visual saliency calculation and map building. The employed volumetric rep-

resentation is an efficient occupancy map $\mathcal{M}$ based on octrees [24] dividing the space V^E in cubical volumes $m \in \mathcal{M}$, that can either be marked as unexplored, free, occupied-"salient", -"inhibited" and -"normal". Paths are only planned inside the iteratively explored free space V^E_{free}, relevant vehicle configurations are denoted as $\xi \in \Xi$, while paths are given by $\sigma : \mathbb{R}^n \rightarrow \xi$ and specifically from ξ_{k-1} to ξ_k by $\sigma_{k-1,k}(s),\ s \in [0,1]$ with $\sigma_{k-1,k}(0) = \xi_{k-1}$ and $\sigma_{k-1,k}(1) = \xi_k$. The sampled paths have to be collision–free and feasible for the vehicle to track given its dynamic constraints. Within the algorithm implementation, a robot configuration is defined by the state $\xi = [x, y, z, \psi]^T$ with roll (ϕ) and pitch (θ) considered near–zero. For slow maneuvering it is considered that the robot flies in straight paths $\sigma_{k-1,k} = s\xi_k + (s-1)\xi_{k-1},\ s \in [0,1]$ with constant velocity v_t. The subsections below detail the proposed approach.

A. Visual Saliency Model

Visual saliency is the distinct subjective perceptual quality of certain objects and entities that makes them stand out from their neighbors and immediately attract the attention of humans, or broadly primates and many other mammals [4, 6]. This attraction of our attention to visually salient stimuli is believed to have evolutionary roots and relates to the historical need to rapidly detect potential prey, predators, or mates in cluttered environments. It allows attention to focus on the sub–regions of the environment that mostly matter, enables respective active vision and is considered to enable serialization of advanced processing of the salient areas [6].

1) Visual Saliency on Images: Aiming for a high performance and computationally efficient method, this work employs and augments the use of the cognitive model of visual saliency detailed in [8] which revisits the seminal work in [4]. According to this model, saliency is independent of the nature of a particular task and is driven in a bottom-up manner. It can therefore be considered as an *intrinsic motivation* for the robot. Given a visual input to the robot (a camera frame $\mathcal{C}_\ell$), the saliency model $\mathcal{S}_\ell = \mathcal{S}(\mathcal{C}_\ell)$ aims to calculate a centralized two–dimensional "Saliency Map" that predicts pixel–wise the expected human eye–fixation. To achieve this task, the input image $\mathcal{C}_\ell$ is decomposed to a set of pre-attentive feature maps $\mathcal{F}^\kappa_\ell$ related to color and intensity that operate in parallel over the entire frame as shown in Figure 2. Image pyramids are employed for multi–scale computation, while spatial contrasts are computed using Gaussian smoothing. A scale–space structure and a center–surround ratio are implemented towards reliable salient object segmentation. The individual "conspicuity maps" per feature map are then calculated and fused into a unique map that encodes for saliency, irrespective of the feature channel in which stimuli appeared salient. A brief summary of the method follows.

First, each camera frame is converted into the opponent color space described in [25]. Intensity (I_ℓ), Red-Green (RG_ℓ) and Blue-Yellow (BY_ℓ) channels are then computed:

$$I_\ell = \frac{R_\ell + G_\ell + B_\ell}{3},\ RG_\ell = R_\ell - G_\ell,\ BY_\ell = B_\ell - \frac{R_\ell + G_\ell}{2} \quad (1)$$

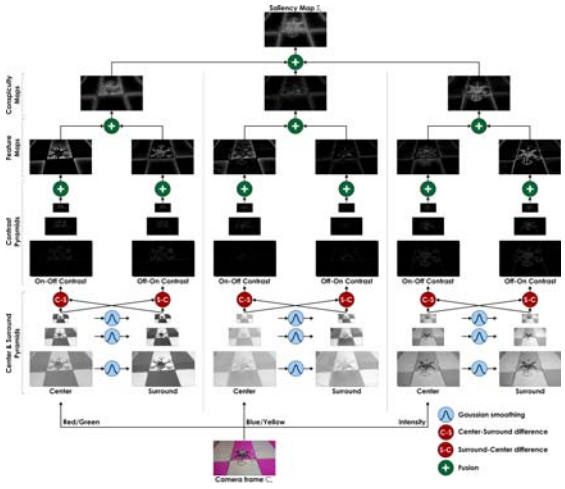

Fig. 2. Illustration of the bottom–up visual saliency model.

where R_ℓ, G_ℓ, B_ℓ are the red, green and blue channels of the image $\mathcal{C}_\ell$ respectively. For each of the I_ℓ, RG_ℓ and BY_ℓ channels, a scale–space using scales and octaves is computed, while twin pyramids consisting of one center $\mathbf{C}_\ell = [\mathbf{c}_{\ell,0}, ..., \mathbf{c}_{\ell,L}]$ and one surround pyramid $\mathbf{S}_\ell = [\mathbf{s}_{\ell,0}, ..., \mathbf{s}_{\ell,L}]$ are computed using Gaussian smoothing:

$$G(u,v) = \frac{1}{2\pi\sigma_x^2} e^{-\frac{u^2+v^2}{2\sigma_x^2}},\ [u,v] \rightarrow \text{pixel coordinates} \quad (2)$$

with σ_x corresponding to a center–surround ratio ($\sigma_x = \sqrt{\sigma_s^2 - \sigma_c^2}$, where σ_c is the value to obtain the center image $\mathbf{c}^i_k$ and σ_s is the smoothing factor for the surround image $\mathbf{s}^i_k$). One twin pyramid $[\mathbf{C}^\kappa_\ell, \mathbf{S}^\kappa_\ell],\ \kappa \rightarrow \{I_\ell, RG_\ell, BY_\ell\}$ is computed for each of the I_ℓ, RG_ℓ and BY_ℓ channels. Provided $[\mathbf{C}^\kappa_\ell, \mathbf{S}^\kappa_\ell]$, on–off and off–on center-surround contrast maps for every layer i of the pyramid are derived:

$$\begin{aligned} \mathbf{X}^\kappa_{\ell,\mathbf{i}} &= \mathbf{c}^\kappa_{\ell,\mathbf{i}} - \mathbf{s}^\kappa_{\ell,\mathbf{i}} \quad \text{(on–off contrasts)} \\ \mathbf{Y}^\kappa_{\ell,\mathbf{i}} &= \mathbf{s}^\kappa_{\ell,\mathbf{i}} - \mathbf{c}^\kappa_{\ell,\mathbf{i}} \quad \text{(off–on contrasts)} \end{aligned} \quad (3)$$

Subsequently, the images from each contrast pyramid are combined with across–scale addition $\oplus$ in order to derive the feature maps $\mathcal{F}^\kappa_\ell,\ \kappa \rightarrow \{I_\ell, RG_\ell, BY_\ell\}$:

$$\begin{aligned} \mathcal{F}^\kappa_{\ell,on-off} &= \oplus_i X^\kappa_{\ell,i},\ i \in \{1, ..., L\} \\ \mathcal{F}^\kappa_{\ell,off-on} &= \oplus_i Y^\kappa_{\ell,i},\ i \in \{1, ..., L\} \end{aligned} \quad (4)$$

Finally, the feature maps of each channel are fused into conspicuity maps $\epsilon^\kappa_\ell,\ \kappa \rightarrow \{I_\ell, RG_\ell, BY_\ell\}$ which are then fused into a single saliency map $\mathcal{S}_\ell$:

$$\begin{aligned} \epsilon^\kappa_\ell &= \mathbf{f}(\mathcal{F}^\kappa_{\ell,on-off}, \mathcal{F}^\kappa_{\ell,off-on}),\ \kappa \rightarrow \{I_\ell, RG_\ell, BY_\ell\} \\ \mathcal{S}_\ell &= \mathbf{g}(\epsilon^I_\ell, \epsilon^{RG}_\ell, \epsilon^{BY}_\ell) \end{aligned} \quad (5)$$

where $\mathbf{f}, \mathbf{g}$ denote fusion operations by arithmetic mean per pixel value among the different conspicuity map images. Once the saliency map $\mathcal{S}_\ell$ is derived, histogram equalization is performed, and low saliency values are eliminated.

2) Saliency–based Occupancy Map: To enable saliency–aware path planning in the robot configuration space, the saliency maps per image $\mathcal{C}_\ell$ are utilized to derive a Saliency–based 3D occupancy map by appending camera saliency values to the map voxels $m \in \mathcal{M}$. At every associated ℓ–th update of the occupancy map, five type of voxels are considered, namely unexplored, free, occupied-"salient", -"inhibited" and -"normal" (not salient). Salient voxels have a saliency value $S_m^\ell > 0$, while for implementation robustness a threshold S_m^{thres} is employed and consequently, salient voxels are such that $S_m^\ell \geq S_m^{\text{thres}}$. Similarly for normal voxels $S_m^\ell < S_m^{\text{thres}}$. Provided the saliency maps $\mathcal{S}_\ell$ derived per image, the robot pose $\mathbf{p}_\ell$ at the corresponding viewpoint configuration, the transformation between the camera coordinate frame $\mathcal{K}$ and the body–fixed frame $\mathcal{B}$, the camera model and the occupancy map $\mathcal{M}$, geometric projection allows finding the subset of pixels of the saliency map $\mathcal{S}_\ell$ that refers to each of the visible voxels from $\mathbf{p}_\ell$. For such a voxel m, the projected volumetric saliency value S_m^{proj} is calculated:

$$S_m^{\text{proj}} \quad = \quad \frac{1}{N} \sum_{[u,v] \in \text{proj}_{\mathcal{S}_\ell}^m} \mathcal{S}_\ell(u,v) \tag{6}$$

where $\text{proj}_{\mathcal{S}_\ell}^m$ denotes the saliency map pixels the projection of which lies on voxel m. For a voxel first perceived to be salient, the algorithm sets $S_m^\ell \leftarrow S_m^{\text{proj}}$. For salient voxels that are reobserved from different perspectives, which in turn affects the saliency map calculation for the same volumetric subsets, update of the saliency value takes place:

$$S_m^\ell \quad = \quad S_m^{\ell-1} + \gamma(S_m^{\text{proj}} - S_m^{\ell-1}) \tag{7}$$

where $\gamma \in [0,1]$ is a weighting factor allowing the voxel saliency to emphasize on recent observations ($\gamma > 0.5$), first impression ($\gamma < 0.5$) or average uniformly ($\gamma = 0.5$).

3) Inhibition of Return in Occupancy Maps: The model of visual saliency approximates pre–attentive vision mechanisms observed in humans and enables to develop intrinsic rewards to adaptively direct the attention of the robot. Saliency maps however raise a side–problem related to how we can prevent attention from permanently focusing into the most salient location [6]. To model the respective behavior in primates, the concept of Inhibition of Return (IOR) has been proposed [11] and refers to the relative suppression of orienting toward stimuli which have recently been the focus of attention, therefore encouraging orienting towards novelty.

In saliency research considering static cameras, IOR has been approximated by modeling its short–term memory effect as a "discharging" function applied per pixel value. However, such a method is not directly transferable to a moving robot. Instead, we implement a 3D IOR system applied on the saliency value S_m^k of each voxel. The following update takes place for each salient voxel:

$$S_m^{\ell-1} \quad \leftarrow \quad S_m^{\ell-1} e^{-\beta \Delta T}, \; \Delta T = t_\ell - t_{\ell-1} \tag{8}$$

where β is a tuning factor and $t_\ell, t_{\ell-1}$ are the timestamps of camera frames $\mathcal{C}_\ell, \mathcal{C}_{\ell-1}$. This step takes place before applying (7). Once S_m^ℓ of a salient voxel decreases below a threshold $S_m^\ell < S_m^{\text{IOR}}$, then it is considered "inhibited".

B. Visual Saliency–aware Exploration Path Planning

The Visual Saliency–aware Exploration Planner (VSEP), overviewed in Figure 3, is detailed in this sub-section.

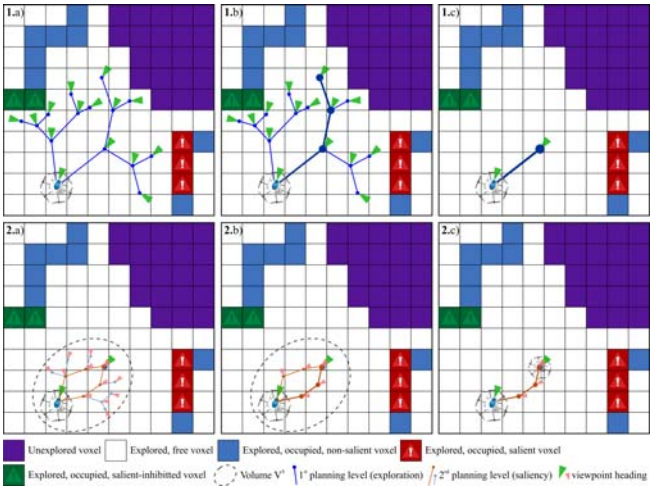

Fig. 3. Illustration of the visual saliency–aware 2–step exploration planner.

1) Exploration Planning: The proposed method first optimizes for the extrinsic reward of unknown area exploration. At every iteration and for the current occupancy map $\mathcal{M}$ of the world, the set of visible but unmapped voxels from a robot configuration ξ is denoted as $\mathbf{Visible}(\mathcal{M}, \xi)$. Every voxel $m \in V_\xi^V$ lies in the unexplored area V_{unm}^E ($V_\xi^V \subset V_{unm}^E$), the direct line of sight does not cross occupied voxels and respects the sensor model. A camera sensor with fixed mounting angle η^m, limited vertical and horizontal FoV angles $[a_v, a_h]$, and a maximum effective sensing distance $d_{\max}^{\text{planner}}$ is considered. For robustness, the value of $d_{\max}^{\text{planner}}$ used is smaller than the actual effective distance of the sensor $d_{\max}^{\text{sensor}}$. Starting from the current configuration ξ_0, a geometric random tree $\mathbb{T}^E$ with maximum edge length ℓ_E is incrementally built in the configuration space using the RRT algorithm [26]. The resulting tree is comprised of $N_{\mathbb{T}}^E$ nodes and a set of collision free paths σ^E. In order to sufficiently explore the configuration space but also limit the computational cost, a minimum of $N_{\min}^E$ and maximum of N_{TOT}^E nodes are defined. The information gain of a node $\mathbf{ExplorationGain}(n^E)$ is the cumulative unmapped volume that can be explored from the nodes along the branch. For node k:

$$\begin{aligned}\mathbf{ExplorationGain}(n_k^E) = \mathbf{ExplorationGain}(n_{k-1}^E) + \\ \mathbf{Visible}(\mathcal{M}, \xi_k) e^{-\lambda c(\sigma_{k-1,k}^E)}\end{aligned} \tag{9}$$

where $c(\sigma_{k-1,k}^E)$ is the length of the path, and λ is a tunable factor to penalize long paths. Admissible paths are considered only among those that their travel time $c(\sigma_{k-1,k}^E)/v_t$ is smaller than the remaining robot endurance t_r. Once this level of the path planning iteration

is completed, the first segment σ^E_{RH} of the best branch $\mathbf{ExtractBestPathSegment}(n^E_{best})$, the vertex at the end of this segment n^E_{RH}, and the associated pose configuration ξ_{RH} are extracted. If no positive gain can be found, the exploration process is terminated. This step employs and extends the concept of previous own work [18, 19].

Algorithm 1 Proposed Planner - Iterative Step

$\xi_0 \leftarrow$ current vehicle configuration
Initialize $\mathbb{T}^E$ with ξ_0
$g^E_{best} \leftarrow 0$ ▷ Set best exploration gain to zero
$n^E_{best} \leftarrow n_0(\xi_0)$ ▷ Set best exploration node to root
$N^E_{\mathbb{T}} \leftarrow$ Number of initial nodes in $\mathbb{T}^E$
while $N^E_{\mathbb{T}} < N^E_{\max}$ **or** $g^E_{best} = 0$ **do**
  Incrementally build $\mathbb{T}^E$ by adding $n^E_{new}(\xi_{new})$
  $N^E_{\mathbb{T}} \leftarrow N^E_{\mathbb{T}} + 1$
  if $\mathbf{ExplorationGain}(n^E_{new}) > g^E_{best}$ **then**
    $n^E_{best} \leftarrow n^E_{new}$
    $g^E_{best} \leftarrow \mathbf{ExplorationGain}(n^E_{new})$
  end if
  if $N^E_{\mathbb{T}} > N^E_{TOL}$ **then**
    Terminate planning
  end if
end while
$\sigma^E_{RH}, n^E_{RH}, \xi_{RH} \leftarrow \mathbf{ExtractBestPathSegment}(n^E_{best})$
$\mathbb{S}_{\xi_{RH}} \leftarrow \mathbf{LocalSet}(\xi_{RH})$
$\alpha \leftarrow 1$ ▷ number of admissible paths
$g^S_\alpha \leftarrow \mathbf{SaliencyGain}(\sigma^E_{RH})$
$g^S_{best} \leftarrow g^S_\alpha$ ▷ straight path saliency gain
$\sigma^S_{best} \leftarrow \sigma^S_{RH}$ ▷ Set best saliency–gain path
while $N^S_{\mathbb{T}} < N^S_{\max}$ **or** $\mathbb{V}(\mathbb{T}^S) \cap \mathbb{S}_{\xi_{RH}} = \emptyset$ **do**
  Incrementally build $\mathbb{T}^S$ by adding $n^S_{new}(\xi_{new})$
  if $\xi_{new} \in \mathbb{S}_{\xi_{RH}}$ **then**
    Add new vertex n^S_{new} at ξ_{RH} and connect
    $\alpha \leftarrow \alpha + 1$
    $\sigma^S_\alpha \leftarrow \mathbf{ExtractBranch}(n^S_{new})$
    $g^S_\alpha \leftarrow \mathbf{SaliencyGain}(\sigma^S_\alpha)$
    if $g^S_\alpha > g^S_{best}$ **then**
      $\sigma^S \leftarrow \sigma^S_\alpha$
      $g^S_{best} \leftarrow g^S_\alpha$
    end if
  end if
end while
return σ^S

2) Saliency–aware Planning: Provided the current and goal robot configurations ξ_0, ξ_{RH} for the current iteration of the first layer of the planner, the second planning step aims to direct the robot to effectively shift its attention toward salient regions by identifying a new path σ^S that connects ξ_0 to ξ_{RH}, maximizes an associated information gain and has a travel cost bounded by an adaptively computed time–budget. Within a volume V^S that includes ξ_0 and ξ_{RH}, a new random tree $\mathbb{T}^S$ is spanned, has a maximum edge length ℓ_S and satisfies constraints related to maximum yaw rate $\dot{\psi}_{\max}$. A total set of $N^S_{\mathbb{T}}$ vertices within V^S are sampled with $N^S_{\mathbb{T}} < N^S_{\max}$ and admissible paths (σ^E_{RH} treated as one of them) that a) start from ξ_0 and arrive in a local connected set $\mathbb{S}_{\xi_{RH}}$ around the reference ξ_{RH} provided by the first planning layer, and b) have a maximum allowed travel cost $t^{\max}_S$, are derived. Since admissible paths are those arriving into the local set $\mathbb{S}_{\xi_{RH}}$ ($\mathbb{V}(\mathbb{T}^S) \cap \mathbb{S}_{\xi_{RH}} \neq \emptyset$) and not necessarily exactly on ξ_{RH}, an additional connection is introduced to ensure that the robot reaches precisely the configuration sampled by the first planning step. Subsequently, for all the $N^S_{\mathbb{T}} - 1$ edges of the $\mathbb{T}^S$ tree (including the additional edges to n_{RH}), calculation of *saliency gain* takes place and allows the identification of the admissible branch that optimizes for directing the robot's attention to the salient subsets of the map. The main steps of this process are detailed below.

Calculation of the maximum available time budget $t^{\max}_S$ is considered in relation to the expected exploration rate e_{rate} assuming only first–layer planning steps (extrapolated from the last ν planning iterations), the percentage $\epsilon = (V^E_{occ} + V^E_{free})/V^E$ of currently explored space, and the expected remaining time for the exploration mission, the remaining robot endurance t_r, as well as the travel cost t_E of the straight edge connection between ξ_0 and ξ_{RH}:

$$\begin{aligned} t^{\max}_S &= (1+\zeta)t_E \\ \zeta &= (t_r - t_{\text{req}})/t_{\text{req}}\big|^{\zeta_{\max}}_0 \,,\; t_{req} = (1-\epsilon)/e_{\text{rate}} \end{aligned} \tag{10}$$

where $\cdot\big|^{\zeta_{\max}}_0$ denotes that the value of ζ is constrained between 0 and a tunable $\zeta_{\max}$ that enhances robustness against environments of unpredictable complexity and avoids "opportunistic" behaviors that spend a very major part of the robot's time to observe salient regions. Considering a constant velocity v_t, $t^{\max}_S$ translates to a limit $d^{\max}_S$ for the length of admissible branches at the second planning step.

Given the identified admissible branches σ^S_α, each associated with a set of μ_α vertices $\{n^S_{\alpha,i}\},\ i = (1, ..., \mu_\alpha)$, as well as the current occupancy map $\mathcal{M}$ with salient voxels of value S^k_m, calculation of saliency gain can take place. The proposed gain formulation aims to simultaneously optimize for the amount of salient voxels observed, as well as the resolution of the associated observations. Considering a pinhole camera model and a certain resolution, it employs two terms, namely a) the "Information per Area" (IA) referring to the salient voxels perceived by the camera pixels at a certain viewpoint, and b) the "Area per Pixel" (AP) that relates to the resolution of the observations [27]. More specifically, the saliency gain across a path σ^S_α takes the form:

$$\begin{aligned} \mathbf{SaliencyGain}(\sigma^S_\alpha) &= \sum_{i=1}^{\mu_\alpha} \sum_{m \in \mathbf{C}(n^S_{\alpha,i})} \overbrace{S^k_m}^{\text{IA}}\, e^{-\tau \overbrace{f_m}^{\text{AP}}} \\ f_m &= \frac{z^2_{n_{\alpha,i},m}}{f_x f_y} \end{aligned} \tag{11}$$

where $m \in \mathbf{C}(n^S_{\alpha,i})$ denotes the voxels m that are perceived by the camera frustum attached to vertex $n^S_{\alpha,j}$ of path σ^S_α, $z_{n_{\alpha,i},m}$ is the distance from the robot configuration of vertex $n^S_{\alpha,i}$ to the m voxel, f_x, f_y denote the focal length of the camera based on the pinhole model, and τ is a tuning factor.

This saliency gain formulation allows the identification of a branch that optimally balances the need to direct the robot attention towards as many salient regions as possible, but also factor in the importance of high resolution observations. It therefore aims to optimize the aggregate information per camera pixel across the robot path. Note that only occupied "salient" voxels add to this gain, and not "inhibited" ones.

3) Post Exploration Saliency–driven Planning: When the proposed saliency–aware exploration planner achieves the mapping of the complete volume (or a tuned large percentage of it), a final stage of planning is engaged for the remaining endurance of the system. During this *saliency–driven* planning phase, the sole goal is to re-observe salient voxels that are not inhibited yet. In a single planning level, a random tree $\mathbb{T}^{SD}$ is spanned in V^E. Saliency gain is computed for all admissible branches $\{\sigma^{SD}\}$ based on (11) and the best one is identified. The first step of this branch is conducted and the procedure is repeated in a receding horizon fashion until all salient voxels become "inhibited".

C. Computational Complexity

Computational complexity analysis is provided for the first and second planning step (the saliency–driven step is similar to the first) of the algorithm to illustrate its scalability. Checking the status of a voxel, and other queries for the full occupancy map, have logarithmic complexity $\mathcal{O}(\log(V^{\mathcal{J}}/r^3))$, $\mathcal{J} \to E, S$ in the number of voxels [19, 24]. The construction of RRTs in fixed environments scales with $\mathcal{O}(N_{\mathbb{T}}^{\mathcal{J}} \log N_{\mathbb{T}}^{\mathcal{J}})$, $\mathcal{J} \to E, S$ and the query for the best node scales with $\mathcal{O}(N_{\mathbb{T}}^{\mathcal{J}})$, $\mathcal{J} \to E, S$. The number of voxels in the volume around an edge to check for collision scales with $1/r^3$ and the complexity to check the $N_{\mathbb{T}}^{\mathcal{J}} - 1$, $\mathcal{J} \to E, S$ edges is $\mathcal{O}(N_{\mathbb{T}}^{\mathcal{J}}/r^3 \log(V^E/r^3))$, $S \to E, S$. For both planning steps, the relevant gain has to be computed. As the camera frustum volume is $V_{sensor} \propto (d_{\max}^{\text{planner}})^3$, the number of voxels to be tested is $N_{vox} \approx V_{sensor}/r^3$. Visibility check complexity is $\mathcal{O}(d_{\max}^{\text{planner}}/r \log(V^{\mathcal{J}}/r^3))$, $\mathcal{J} \to E, S$ as it requires ray casting for every voxel and the number of voxels on the ray is $\mathcal{O}(d_{\max}^{\text{planner}}/r)$. Consequently, the complexity terms for the first– and second–level planning gain calculations take the form $\mathcal{O}((d_{\max}^{\text{planner}}/r)^4 \log(V^{\mathcal{J}}/r^3))$, $\mathcal{J} \to E, S$. Therefore, the overall order depends logarithmically on the planning volume and tree sampling iterations. Scalability can be maintained given the correlation of V^E, V^S with r.

V. SIMULATION BASED EVALUATION

In order to evaluate the proposed Visual Saliency–aware Exploration Planner (VSEP) prior to its experimental verification, multiple simulation studies were conducted. The open–source RotorS Simulator [28] was employed, as it can provide both state feedback as well as the necessary onboard localization and mapping information. A hexacopter aerial robot equipped with a stereo camera providing raw and depth images ($[a_v, a_h] = [60, 90]^\circ$, $\eta^m = 15^\circ$ downward) was simulated. Environments of varying density of salient entities (paintings, furniture and a human) were considered. All environments are initially unknown to the robot.

Figure 4 presents the environments used in simulation, indicative results of the path planning process and statistical metrics that demonstrate the ability of the proposed approach to direct the robot's attention to the most salient regions. The proposed planner (VSEP) is compared with the case of running only the exploration step which corresponds to our own previous work in [19] and is denoted as NBVP.

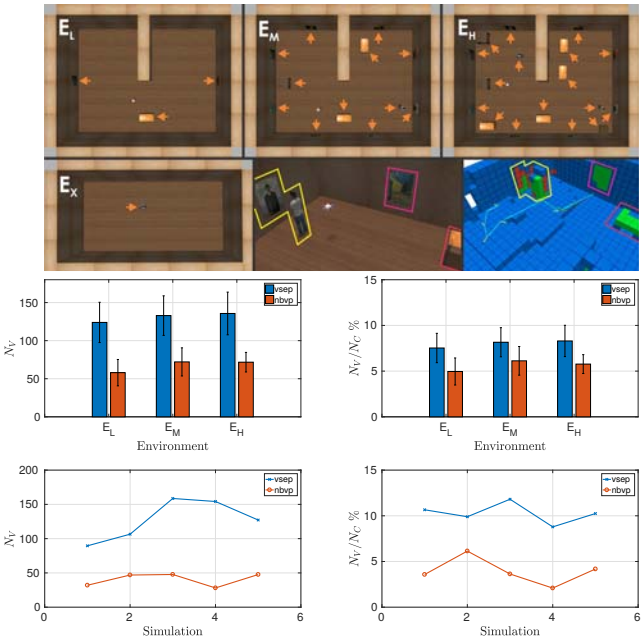

Fig. 4. Upper left plot: Average number of viewpoints N_V toward each salient voxel when running the VSEP and when running NBVP for three environments (E_L, E_M, E_H) of increasing density of salient objects (pointed using an arrow) and for multiple simulations (5 per environment and planner) to derive statistical averages and variance. As shown, the proposed new method ensures more than double numbers of viewpoints toward salient voxels on avarage with 33% increase in mission time (400s for VSEP, and 300s for NBVP). Upper right plot: Percentage of average number of viewpoints toward each salient voxel N_V versus the total number of camera frames N_C used for updating the saliency–based occupancy map for each test case (which linearly depends on mission time). Lower plot: Comparison of the same metrics in multiple simulation studies for a smaller environment E_X with a unique salient object.

VI. EXPERIMENTAL EVALUATION

A challenging set of experiments were conducted to evaluate and demonstrate the performance of the visual saliency–aware exploration algorithm. All estimation, control and planning functionalities were conducted onboard and without any support from motion capture systems.

A. Platform Overview

A custom–built hexarotor platform is employed and has a weight of 2.5kg. The system relies on a Pixhawk–autopilot for its attitude control and further integrates an Intel NUC5i7RYH, and a stereo visual–inertial sensor providing synchronized stereo frames (using a StereoLabs ZED) and IMU data (UM7). All the planning, localization and mapping, and position control loops are running on the NUC5i7RYH with the support of the Robot Operating System (ROS). The system performs visual–inertial odometry

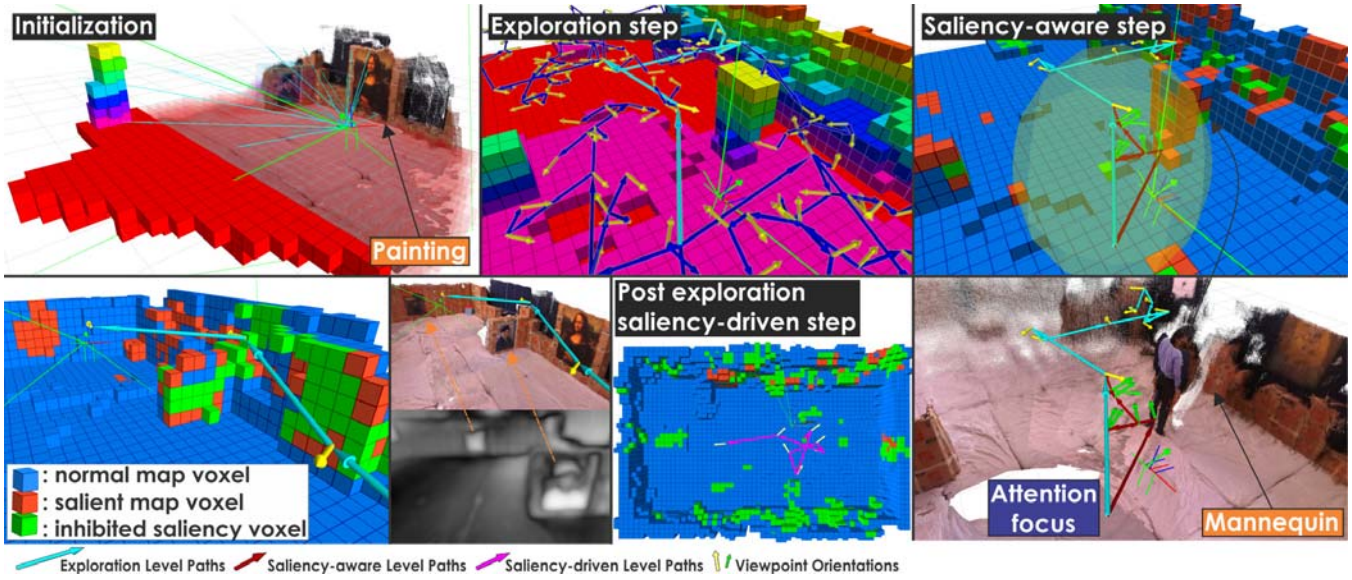

Fig. 5. a) Initialization of VSEP in a saliency-enhanced mockup environment, 3D pointcloud and volumetric height-map, b) 1^{st} layer: next-best-view random sampling for pure volumetric exploration, c) 2^{nd} layer: saliency-aware resampling in local configuration space, robot deviates and directs its attention towards visually salient regions (voxels on the right) while respecting mission time-constraints, d) Saliency-encoded octomap with Inhibition of Return: at a later mission step, saliency regions that were observed first (paintings on the right) have become inhibited, while the latest discovered ones (painting and colored box at the far end) are still salient, e) Pointcloud representation of previous phase and sample saliency image frame, f) Mission completion: post-exploration phase focuses on the last upper side regions that have not become inhibited yet, g) Pointcloud at 2^{nd} layer snapshot: part of the salient regions are in fact a human (mannequin) standing in front of a wall painting.

using ROVIO [29] and mapping through the point cloud from the stereo camera and the robot pose estimates. For position tracking, the model predictive controller in [30] is utilized.

B. Experimental Studies

The first experimental scenario refers to an indoor environment sized $11 \times 6 \times 2$m that includes vertical and T–shaped walls adding geometric complexity, while 5 posters of paintings (including portraits), 3 colorful boxes, and a dressed mannequin, were distributed to create objects of varying visual saliency. A summary of the most important algorithm parameters is provided in Table I, where ℓ'_E refers to the per iteration length of the selected edge of the first planning level. Figure 5 presents multiple instances of the experiment and demonstrates how the visual saliency–aware second planning layer adjusts the robot's perceptual attention. The result is an overall exploration path appropriately adjusted to focus on the salient regions, while the IOR functionality encourages the robot to shift its attention. As shown in Figure 7 this leads to significantly more viewpoints toward the salient voxels compared to the case of the robot simply following the exploration steps of the first planning layer, which leads to enhanced mapping. Once exploration of the unknown volume is completed, the robot spends the remaining of its 300s considered mission–time for saliency–driven reobservation of salient voxels that are not inhibited.

The second experimental scenario refers to the exploration of a machine–shop like environment sized $15 \times 4 \times 1.5$m that includes a set of machinery, warning signs, fire extinguishers and other components that overall lead to a complex environment in terms of visual saliency. Figure 6 presents instances of this experiment and demonstrates how the second visual saliency–based planning layer adjusts the robot's sensing attention during exploration. Aligned with intuition, the warning signs and the fire extinguisher attract the robot's attention and the planner appropriately defines viewpoints to observe and map them in detail.

TABLE I
EXPERIMENT PARAMETERS

Parameter	Value	Parameter	Value
v_t	0.125m/s	$\dot{\psi}_{\max}$	$\pi/12$rad/s
FoV $[a_v, a_h]$	$[55, 85]^\circ$	η^m	17.5°
$[d_{\max}^{\text{planner}}, d_{\max}^{\text{sensor}}]$	[4.5, 5]m	r	0.2m
λ	0.5	$[\ell_E, \ell_S]$	$[2.0, \ell'_E]$m
$[N_{\max}^E, N_{\max}^S]$	$[150, 500]$	$S_m^{\text{IOR}} = S_m^{\text{thres}}$	125
S_m^{thres}	125	Collision box	0.85x0.85x0.25m
τ	10^5	$[f_x, f_y]$	$[704.58, 705.71]$
γ	0.7	β	0.0008

VII. CONCLUSIONS

In this work, a visual saliency–aware exploration path planning strategy is proposed and demonstrated to efficiently adjust the robot's perceptual attention in order to focus on the most salient subsets of the environment. The algorithm employs an iterative two–step planning paradigm within which it first identifies the next–best–view for exploration and subsequently derives a path that optimizes for saliency observation to reach that viewpoint. As a third step, purely saliency–driven reobservation takes place once the unknown area is explored and given that remaining flight time is available. Through a sequence of simulation studies and experimental results, it is demonstrated both by example and

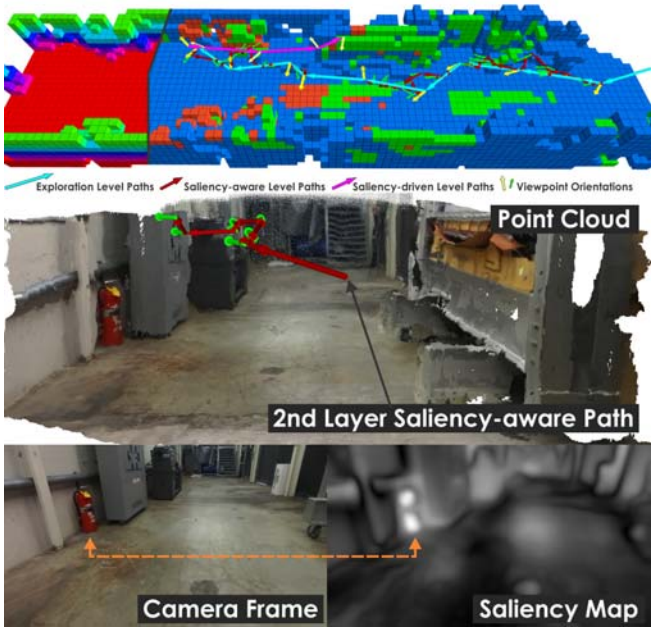

Fig. 6. Visual saliency–aware exploration for a machine shop environment involving several types of salient objects such as warning signs and a fire extinguisher. The robot adapts its exploration path to direct its attention to the corresponding voxels of the map. Orange voxels represent occupied–"salient" voxels, green voxels represent occupied–"inhibited" voxels, while blue represents occupied–"normal" voxels. The second row presents the online reconstructed point cloud and the case of one saliency–based path focusing primarily on the fire extinguisher, while relevant camera frames and corresponding saliency maps are shown at the bottom of the figure.

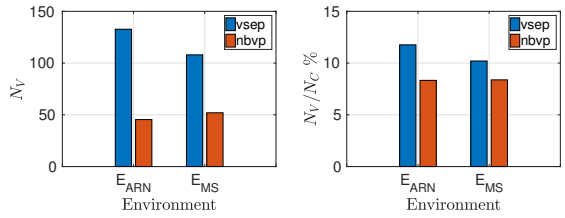

Fig. 7. Left: Average number of viewpoints N_V toward each salient voxel for the full solution of the planner (VSEP) and when only the exploration viewpoints are provided as commands to the robot (NBVP) for both the ARENA (E_{ARN}) and Machine Shop (E_{MS}) experiments. Right: Percentage of average number viewpoints attending each salient voxel versus total number of camera frames N_C used to update the saliency–based occupancy map for the two planning cases.

statistically that the method identifies paths that effectively adjust the robot's attention to salient parts of its map.

References

[1] A. Borji, and L. Itti, "State-of-the-art in visual attention modeling," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 1, pp. 185–207, 2013.

[2] T. Judd, F. Durand, and A. Torralba, "A benchmark of computational models of saliency to predict human fixations," 2012.

[3] J. K. Tsotsos, *A computational perspective on visual attention*. MIT Press, 2011.

[4] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 20, no. 11, 1998.

[5] L. Itti and P. Baldi, "Bayesian surprise attracts human attention," *Vision research*, vol. 49, no. 10, pp. 1295–1306, 2009.

[6] L. Itti and C. Koch, "Computational modelling of visual attention," *Nature reviews neuroscience*, vol. 2, no. 3, pp. 194–203, 2001.

[7] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," in *Advances in neural information processing systems*, 2007.

[8] S. Frintrop, T. Werner, and G. Martin Garcia, "Traditional saliency reloaded: A good old model in new shape," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.

[9] J. Gottlieb, P.-Y. Oudeyer, M. Lopes, and A. Baranes, "Information-seeking, curiosity, and attention: computational and neural mechanisms," *Trends in cognitive sciences*, vol. 17, no. 11, 2013.

[10] P.-Y. Oudeyer, F. Kaplan, V. V. Hafner, and A. Whyte, "The playground experiment: Task-independent development of a curious robot," in *AAAI Spring Symposium on Developmental Robotics*, 2005.

[11] R. M. Klein, "Inhibition of return," *Trends in cognitive sciences*, vol. 4, no. 4, pp. 138–147, 2000.

[12] T. Dang, C. Papachristos, and K. Alexis, "Visual Saliency–aware Receding Horizon Autonomous Exploration with Application to Aerial Robotics." [Online]. Available: https://github.com/unr-arl/vseplanner

[13] R. Marchant and F. Ramos, "Bayesian optimisation for informative continuous path planning," in *Robotics and Automation (ICRA), 2014 IEEE International Conference on*. IEEE, 2014, pp. 6136–6143.

[14] A. Jones, M. Schwager, and C. Belta, "A receding horizon algorithm for informative path planning with temporal logic constraints," in *IEEE International Conference on Robotics and Automation*. IEEE, 2013.

[15] D. Berenson, P. Abbeel, and K. Goldberg, "A robot path planning framework that learns from experience," in *Robotics and Automation (ICRA), 2012 IEEE International Conference on*.

[16] T. Stoyanov, M. Magnusson, H. Andreasson, and A. J. Lilienthal, "Path planning in 3d environments using the normal distributions transform," in *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*. IEEE, 2010, pp. 3263–3268.

[17] M. Popovic, G. Hitz, J. Nieto, I. Sa, R. Siegwart, and E. Galceran, "Online informative path planning for active classification using uavs," *arXiv preprint arXiv:1609.08446*, 2016.

[18] C. Papachristos, S. Khattak, and K. Alexis, "Uncertainty–aware receding horizon exploration and mapping using aerial robots," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2017.

[19] A. Bircher, M. Kamel, K. Alexis, H. Oleynikova and R. Siegwart, "Receding horizon "next-best-view" planner for 3d exploration," in *IEEE International Conference on Robotics and Automation (ICRA)*, May 2016. [Online]. Available: https://github.com/ethz-asl/nbvplanner

[20] L. Yoder and S. Scherer, "Autonomous exploration for infrastructure modeling with a micro aerial vehicle," in *Field and Service Robotics*. Springer, 2016, pp. 427–440.

[21] C. Connolly *et al.*, "The determination of next best views," in *IEEE International Conference on Robotics and Automation 1985*.

[22] B. Yamauchi, "A frontier-based approach for autonomous exploration," in *CIRA'97*. IEEE, 1997, pp. 146–151.

[23] C. Craye, D. Filliat, and J.-F. Goudou, "Rl-iac: An exploration policy for online saliency learning on an autonomous mobile robot," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2016, pp. 4877–4884.

[24] A. Hornung, K. M. Wurm, M. Bennewitz, C. Stachniss, and W. Burgard, "OctoMap: An efficient probabilistic 3D mapping framework based on octrees," *Autonomous Robots*, 2013.

[25] D. A. Klein and S. Frintrop, "Salient pattern detection using w2 on multivariate normal distributions," in *Joint DAGM (German Association for Pattern Recognition) and OAGM Symposium*. Springer, 2012.

[26] S. LaValle and J. Kuffner, J.J., "Randomized kinodynamic planning," in *IEEE ICRA*, 1999, pp. 473–479 vol.1.

[27] M. Schwager, B. J. Julian, M. Angermann, and D. Rus, "Eyes in the sky: Decentralized control for the deployment of robotic camera networks," *Proceedings of the IEEE*, vol. 99, no. 9.

[28] F. Furrer, M. Burri, M. Achtelik, and R. Siegwart, "Rotors-a modular gazebo mav simulator framework," in *Robot Operating System (ROS)*. Springer, 2016, pp. 595–625.

[29] M. Bloesch, S. Omari, M. Hutter, and R. Siegwart, "Robust visual inertial odometry using a direct ekf-based approach," in *Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on*. IEEE, 2015, pp. 298–304.

[30] M. Kamel, T. Stastny, K. Alexis, and R. Siegwart, "Model predictive control for trajectory tracking of unmanned aerial vehicles using ros," *Springer Book on Robot Operating System (ROS)*.