

# APF-PF: Probabilistic Depth Perception for 3D Reactive Obstacle Avoidance

Shakeeb Ahmad<sup>1</sup>, Zachary N. Sunberg<sup>1</sup> and J. Sean Humbert<sup>2</sup>

**Abstract**—This paper proposes a framework for 3D obstacle avoidance in the presence of partial observability of environment obstacles. The method focuses on the utility of the Artificial Potential Function (APF) controller in a practical setting where noisy and incomplete information about the proximity is inevitable. We propose a Particle Filter (PF) approach to estimate potential obstacle locations in an input depth image stream. The probable candidates are then used to generate an action that maneuvers the robot towards the negative gradient of potential at each time instant. Rigorous experimental validation on a quadrotor UAV highlights the robustness and reliability of the method when robot’s sensitivity to incorrect perception information can be concerning. The proposed perception and control stack is run onboard the UAV, demonstrating the computational feasibility for real-time applications and agile robots.

## I. INTRODUCTION

There has been an increasing interest in the area of active perception over the last decade or two. A recent DARPA challenge [1] poses a problem of exploration of unknown, unstructured, large-scale and cluttered underground environments. Like many other sub-problems from the challenge, the path planning problem suffers from uncertain and incomplete sensor information due to noise and limited field of view respectively. Moreover, the computation time and the effectiveness are the essential attributes of a planning method. Mapping-based techniques perform well to generate a path according to a global mission strategy. Commonly used mapping tools like OctoMap [2] and Voxelblox [3] have been popular recently, in generating a map representation of an environment using 3D pointclouds. However, mapping is a slow processes. For large environments, planning over such global information contributes to additional computation time overhead. Moreover, the coarseness of the map being used for planning over large scale environments may not be suitable to detect intricacies in the environment. A fundamental requirement for a mapping-based planner to be effective, is a precise position feedback of the robot. In practice, such estimates are expensive to obtain, especially locally, in terms of both computation and payload. Therefore, there is a high chance that an agile robot, such as an MAV, does not map small and thin obstacles. Such obstacles, like hanging cables and pipes, are inherent to many complex indoor environments like caves, mines and warehouses. Hence, there is a need to

This work was supported through the DARPA Subterranean Challenge, cooperative agreement number HR0011-18-2-0043

<sup>1</sup> Department of Aerospace Engineering Sciences, University of Colorado Boulder, CO, {shakeeb.ahmad, zachary.sunberg}@colorado.edu

<sup>2</sup> Department of Mechanical Engineering, University of Colorado Boulder, CO, sean.humbert@colorado.edu

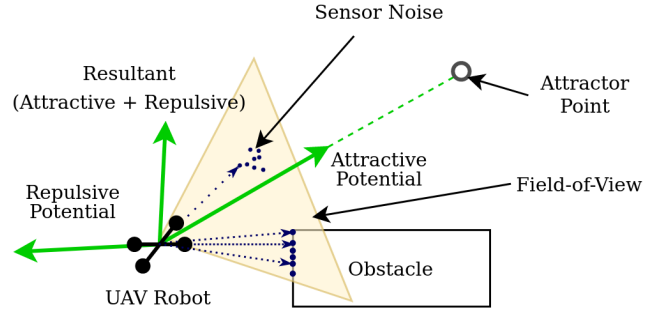


Fig. 1. An illustration of the APF-PF problem. The blue dots represent the depth image points as projected in 3D. The blue arrows show the candidates for the obstacle position vector,  $x_{obs}$  (densest dotted line represents the most probable candidate).

use faster and higher resolution raw sensor measurements to guide a robot. Moreover, a full 3D solution is required to navigate through uncertain and complicated structures. The method proposed in this paper, utilizes raw sensor measurements to generate robot actions without relying on a map representation of an environment.

Researches have shown in past that living organisms [4] like insects [5], [6] rely on instantaneous optic flow signals to provide visual cues for reactive obstacle avoidance. One component of an optic flow signal is the sense of how far objects are located in an environment. Depth cameras provide this information in the form of depth images relative to their sensors. A depth image stream can have varying noise levels depending on the type of sensors, the calculation methods involved and abrupt reflections from the objects in the environments. This makes it challenging to obtain useful information to feed into most motion planners that demonstrate well in the perfect observability settings.

Fragoso *et al.* [7] proposed a way of keeping only a local egocentric map of the environment in the form of a cylinder that updates in real-time through temporal fusion of limited field-of-view visual information. Some other works [8], [9] and [10] propose inflation of depth images using C-Space expansion to facilitate path generation inside a disparity space. Later work by Dubey *et al.* [11] highlights the importance of considering small obstacles like wires, in autonomous robot navigation, by explicitly detecting them using convolutional neural networks. Authors in [8] and [12] demonstrated closed-loop RRT and LQR based techniques,

respectively, for path planning inside a depth image space. Reactive controllers, especially the ones based on Artificial Potential Function (APF) [13], and bio-inspired [14]–[16] methods have demonstrated speed and light computational burden.

This paper proposes a solution for the obstacle avoidance problem that 1) generates robot actions directly from an imperfect depth image stream hence providing an end-to-end solution, 2) has a fast update rate to cater for the needs of agile robots, 3) is computationally feasible for robots with limited on-board resources such as MAVs, 4) adapted for robot maneuvers in 3D to make it useful for arbitrarily structured environments. The perception solution to make such control feasible is the foremost contribution of the paper. A particle filter technique is proposed to probabilistically identify and track regions corresponding to potential obstacles, inside a sequence of depth images. At each time instant, the repulsive potential applied by each candidate for such a region, is weighted according to its probability to compute a total repulsive potential. To close the loop, a user-defined goal point generates an attractive potential and finally the robot is commanded to follow the negative gradient of the resultant potential. Although not uniquely defined in the literature, the term *reactive obstacle avoidance* typically refers to the methods that require only local environment information to generate steering commands away from obstacles and towards the goal, instead of planning paths inside a global map. Our method slightly deviates from the traditional context of reactive obstacle avoidance in that it uses particle filter on a *history* of depth images in order to estimate the local environment information. However, the APF controller in-loop only requires the local information and no map information is available or saved during flight, hence, the method proposed in this paper falls under the umbrella of reactive methods for obstacle avoidance. Fig. 2 shows the simplified process flow, various components of which are explained in the following sections.

The rest of the paper is organized as follows. Section II highlights the problem formulation followed by the description of the filter design, in Section III. Section IV provides a description of how the control loop is closed with the perception information. Section V presents the results and analysis from physical experiments. Finally, Section VI concludes the discussion.

## II. PROBLEM SETUP

**Problem 1:** Let  $\mathcal{X}_{free}$  be a set of all collision free robot configurations,  $\mathcal{X}_{goal} \subset \mathcal{X}_{free}$  be the goal region and  $\mathcal{A}$  be a set of available robot actions. Given a history of depth images  $\mathbf{I}_{[0:t]}$ , at any time instant  $t$ , choose an action  $\mathbf{a}_t \in \mathcal{A}$  such that  $\mathbf{x}_{t+1} \in \mathcal{X}_{free}$  and  $\mathbf{x}_{t_f} \in \mathcal{X}_{goal}$ , at some finite time  $t_f$ .

We aim to solve a well-defined problem of moving a robot from an initial to a final point while avoiding collisions. In the APF approach, the robot behaves as a point charge. The goal point generates an attractive potential for the robot which vanishes as the robot gets closer to the goal. The

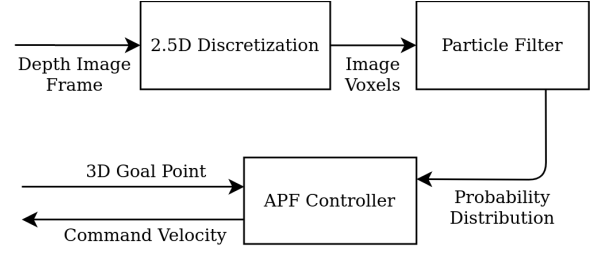


Fig. 2. Simplified process flow.

proximity obstacles are treated as a collection of point charges to generate a repulsive potential to the robot. The robot maneuvers towards the goal, following the negative gradient of the net potential. If  $\mathbf{x}_{goal} \in \mathbb{R}^3$  is the vector from the robot's body frame origin to the goal and  $\mathbf{x}_{obs} \in \mathbb{R}^3$  is the obstacle's position in the robot's body frame, the attractive and repulsive potential functions,  $U_{att} : \mathbb{R}^3 \rightarrow [0, \infty)$  and  $U_{rep} : \mathbb{R}^3 \rightarrow [0, \infty)$ , are given by,

$$U_{att}(\mathbf{x}_{goal}) = \begin{cases} \rho_r \xi \|\mathbf{x}_{goal}\|, & \|\mathbf{x}_{goal}\| > \rho_r. \\ \frac{1}{2} \xi \|\mathbf{x}_{goal}\|^2, & \|\mathbf{x}_{goal}\| \leq \rho_r. \end{cases} \quad (1)$$

$$U_{rep}(\mathbf{x}_{obs}) = \begin{cases} \frac{1}{2} \eta \left( \frac{1}{\|\mathbf{x}_{obs}\|} - \frac{1}{\rho_0} \right)^2, & \|\mathbf{x}_{obs}\| < \rho_0. \\ 0, & \text{otherwise.} \end{cases} \quad (2)$$

Here,  $\rho_0$  defines the sensing horizon over which the repulsive potential is effective. A quadratic potential well is a typical choice for the attractive potential function since it leads to a control law that is linear and all other potentials approximate to quadratic for small perturbations in  $\mathbf{x}_{goal}$  [17]. However, the conic potential [18] provides a constant control law except at the goal [19]. This behavior is more suitable around the obstacles. The conic attractive function is, therefore, often used close to obstacles and is switched to the parabolic function in close proximity of the goal. In (1),  $\rho_r$  refers to the radius around the goal where the conic potential function switches to the parabolic potential function. In order to ensure continuity, the conic and quadratic potential gains are set to  $\rho_r \xi$  and  $\xi$  respectively.

An accurate knowledge of an obstacle's location relative to the robot,  $\mathbf{x}_{obs}$ , is an important variable for the effectiveness of the APF method. Unlike an RGB image, a single channel depth image encodes this information up to an allowable camera resolution. However, depth cameras may have various types of imperfections, making it difficult to rely on the raw output, especially when a system is very sensitive to an incorrect feedback. The following section explains the proposed observer design to estimate such information from a sequence of depth images. Fig. 1 shows a depiction of the APF-PF problem.

## III. PERCEPTION

In a recent effort to simulate a No-Depth-Return (NDP) type of noise, authors in [20] referred to depth imaging as an area in robotic perception that suffers from a large simulation-reality gap due to scene-dependent noise. This

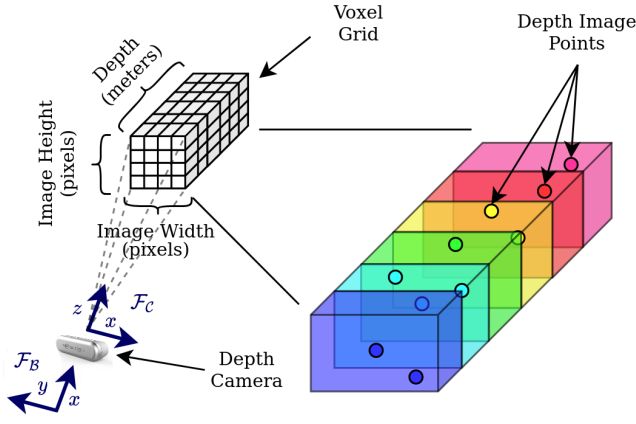


Fig. 3. Discretization of a 2.5D depth image. The depth camera's optical frame,  $\mathcal{F}_C$ , and the robot's body frame,  $\mathcal{F}_B$ , are shown in the right-handed coordinate system.

sort of noise is exceptionally difficult to model hence the literature provides little insight into this topic. We base our model on the noise characteristics that adversely affect the detection of potential obstacles in a depth image. These considerations include the density of the clusters, their distances from the camera and their ability to appear, disappear and move during the consecutive image frames. The goal of this problem is to detect candidates for an occupied region inside a depth image. Particle filtering is an appropriate sampling-based approach to the state estimation that can handle large state spaces.

The particle filter problem parameters can be defined as a tuple  $(\mathcal{S}, \mathcal{O}, \mathcal{T}, \mathcal{Z})$ , where  $\mathcal{S}$  and  $\mathcal{O}$  define the set of all possible states and observations respectively. In our case,  $\mathcal{S}$  is a set of *perception* states which refers to the occupied regions inside a depth image. This encodes the information about  $\mathbf{x}_{\text{obs}}$  without requiring the robot and the obstacles to be mapped in a global frame to generate the repulsive potentials. The probability of a state  $\mathbf{s}_t \in \mathcal{S}$ , at any time instant  $t$ , to transition to a state  $\mathbf{s}_{t+1} \in \mathcal{S}$ , at the next time instant, is defined as  $\mathcal{T}(\mathbf{s}_t, \mathbf{s}_{t+1})$ . Similarly,  $\mathcal{Z}(\mathbf{s}_{t+1}, \mathbf{o}_{t+1})$  is the probability of an observation  $\mathbf{o}_{t+1} \in \mathcal{O}$ , given a transition to a state  $\mathbf{s}_{t+1} \in \mathcal{S}$ . A particle filter outputs a *belief* at every iteration which refers to the probability distribution over all possible states given an observation history. This encodes the relevant information from the entire sequence of observations without needing to explicitly store it.

The following subsections explain various components of the perception problem.

#### A. State Space Transition Models

Any point in an image can correspond to an obstacle in a completely unknown environment and, hence, can potentially be a state. However, at this resolution, the state space size can be large. For a  $640 \times 480$  image, it is 307200 times the number of possible depths at which an obstacle can be located. This requires the flexibility to discretize such information. We discretize the input image in a 2.5D fashion. This

translates to discretization over pixels in the plane parallel to the camera and over depth in the plane perpendicular to the camera. This allows us to make detections directly inside a depth image space in the camera frame,  $\mathcal{F}_C$ , without projecting the whole image to a 3D space for processing. Moreover, the discretization reduces the state space size up to a user-defined refinement depending on the available computational power. This results in a *voxel* grid where the 2.5D coordinate of each voxel belongs to the state space. When projected to 3D, it forms a pyramid in case of a pinhole camera model, with apex corresponding to the optical center of the camera. This, hence, does not require explicit knowledge of the camera model during the perception process. The 2.5D voxel grid covers the entire sensing horizon (field-of-view). Fig. 3 shows such a discretization process, along with the relevant frames of reference. The full state space can be written as,

$$\mathcal{S} = \{(i, j, k) : i \in \{0, k_w, 2k_w, \dots, (N_w - 1)k_w\}, \\ j \in \{0, k_h, 2k_h, \dots, (N_h - 1)k_h\}, \\ k \in \{0, k_d, 2k_d, \dots, (N_d - 1)k_d\}\} \cup \{\mathbf{s}_b\},$$

where  $k_w, k_h \in \mathbb{Z}^+$  are the pixel discretization steps along the width and height of an image and  $k_d \in \mathbb{R}^+$  refers to the discretization step along the depth. Similarly,  $N_w, N_h$  and  $N_d$  are the number of discretization steps along the respective axes. In order for the belief to sum to 1 at each time instant, the state space model also includes a state,  $\mathbf{s}_b$ , referring to obstacle-free sensing horizon. The probability of this state, in the belief vector, informs about the likelihood of there being no obstacle in the field-of-view. In the state space model, this state refers to the voxel grid boundary. The idea is that when there is no obstacle inside the sensing horizon, there is always an obstacle surrounding the voxel grid almost surely, and with non-zero probability of transitioning to the voxel grid.

The state transition probability is defined as the product of two factors:

$$\mathcal{T}(\mathbf{s}_t, \mathbf{s}_{t+1}) = P(\mathbf{s}_{t+1} | \mathbf{s}_t) \propto P_{\text{move}}(\mathbf{s}_{t+1} | \mathbf{s}_t) P_{\text{dist}}(\mathbf{s}_{t+1}). \quad (3)$$

The term,  $P_{\text{move}}(\mathbf{s}_{t+1} | \mathbf{s}_t)$ , models the relative motion of a scene obstacle by constraining the Manhattan distance  $|\text{diag}(k_w, k_h, k_d)^{-1}(\mathbf{s}_{t+1} - \mathbf{s}_t)|$  to follow the normal distribution  $2\mathcal{N}(0, \sigma_s)$ , with  $\sigma_s$  depending on the expected speed of the environment obstacles relative to the robot. This means that an occupied voxel is more likely to transition to a nearby voxel than to a farther one. The states that are close to the voxel grid boundary are highly likely to transition in and out of the field-of-view. The second term,  $P_{\text{dist}}(\mathbf{s}_{t+1})$ , biases the transitions towards the vehicle. Specifically, this bias is defined with a normal distribution centered at zero,  $P_{\text{dist}}(\mathbf{s}_{t+1}) = P_{\text{dist}}(s_{t+1}^z) \propto 2\mathcal{N}(0, \sigma_z)$ , where  $s_{t+1}^z$  is the depth associated with the state  $\mathbf{s}_{t+1}$ . The bias helps the perception module to make a more conservative prediction

of an obstacle's behavior *i.e.*, coming towards the robot with high probability. Moreover, the closer proximity obstacles have a significantly higher contribution to the repulsive potential than the farther ones. The bias towards approaching obstacles plays an important role in ensuring a safe maneuver for a fast moving robot.

### B. Observation Model

An observation  $\mathbf{o} \in \mathcal{O}$  is given by,

$$\mathbf{o} = \{n_i \in \mathbb{Z}^+ : i \in \{0, 1, 2, \dots, (N_w N_h N_d - 1)\}\}.$$

In the expression above,  $n_i$  represents the number of points contained in the voxel  $i$ . The observation model refers to the probability of an observation,  $\mathbf{o}_{t+1}$ , after transitioning to a state  $\mathbf{s}_{t+1}$  *i.e.*,  $\mathcal{Z}(\mathbf{s}_{t+1}, \mathbf{o}_{t+1}) = P(\mathbf{o}_{t+1} | \mathbf{s}_{t+1})$ . For any state inside the voxel grid,  $\mathbf{s}_{t+1} \neq \mathbf{s}_b$ , this can be written as,

$$P(\mathbf{o}_{t+1} | \mathbf{s}_{t+1}) = P(n_0, n_1, \dots, n_{N_w N_h N_d - 1} | \mathbf{s}_{t+1}) = \quad (4)$$

$$\frac{P(\mathbf{s}_{t+1} | n_0, n_1, \dots, n_{N_w N_h N_d - 1}) P(n_0, n_1, \dots, n_{N_w N_h N_d - 1})}{P(\mathbf{s}_{t+1})}.$$

Dense cluster of points inside a depth image serves as an important feature to determine that a region is a valid projection of a physical environment object. Therefore, in order to gain confidence on a voxel, the number of points contained within it is observed. Hence, we can conveniently exploit independence of variables. Equation 4 can then be reduced to,

$$P(\mathbf{o}_{t+1} | \mathbf{s}_{t+1}) = \frac{P(\mathbf{s}_{t+1} | n_{\mathbf{s}_{t+1}}) P(n_0, n_1, \dots, n_{N_w N_h N_d - 1})}{P(\mathbf{s}_{t+1})}, \quad (5)$$

where  $n_{\mathbf{s}_{t+1}}$  is the number of points contained in the state  $\mathbf{s}_{t+1}$ .

Using Bayes's rule on the term  $P(\mathbf{s}_{t+1} | n_{\mathbf{s}_{t+1}})$ , and assuming no prior information about  $n_i$  for any voxel  $i$ ,

$$P(\mathbf{o}_{t+1} | \mathbf{s}_{t+1}) \propto P(n_{\mathbf{s}_{t+1}} | \mathbf{s}_{t+1}) = 2\mathcal{N}(k_w k_h, \sigma_o). \quad (6)$$

Here  $k_w k_h$  is the maximum number of points that can be contained in a voxel and depends on the discretization intervals. The tuning parameter,  $\sigma_o$ , determines the number of points required to gain confidence on any state inside the voxel grid. Low values result in conservative detections where large number of points are required to be certain about a state.

For  $\mathbf{s}_{t+1} = \mathbf{s}_b$ ,

$$P(\mathbf{o}_{t+1} | \mathbf{s}_{t+1}) \quad (7)$$

$$= P(n_0 | \mathbf{s}_{t+1}) P(n_1 | \mathbf{s}_{t+1}) \dots P(n_{N_w N_h N_d - 1} | \mathbf{s}_{t+1})$$

$$\propto P(\max_i n_i | \mathbf{s}_{t+1}) = 2\mathcal{N}(0, \sigma_n).$$

The probability that there is no obstacle inside the voxel grid depends on the number of points in the most populated voxel. This is designed so that, for the system to believe that there is an obstacle in the field-of-view, a significantly large

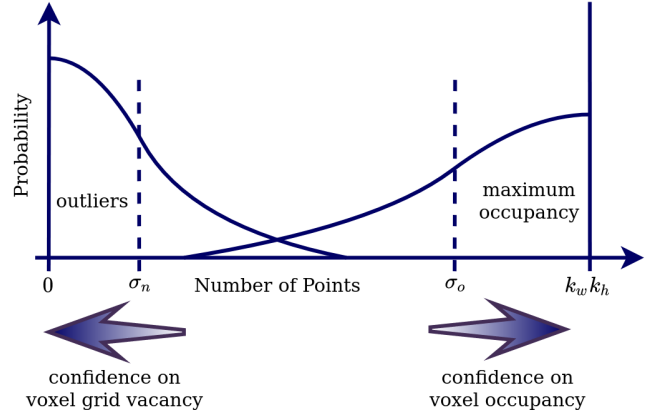


Fig. 4. A depiction of the observation model. The horizontal scale represents the number of points contained in a voxel. The number of points towards the left of the scale is more likely to represent outliers. The number of points towards the right of the scale is more likely to represent that the voxel is occupied with an obstacle.

cluster of points needs to exist in atleast one of the voxels. This models the behavior of the outliers in a depth image and helps filter out sparse noise very efficiently. A larger value of  $\sigma_n$  helps reject greater number of outliers per voxel. The coexistence of the models defined by (6) and (7) is shown in Fig. 4. By appropriately varying  $\sigma_n$  and  $\sigma_o$ , a variety of different filtering behaviors can be achieved.

### C. Particle Filter Algorithm

The particle filter algorithm is executed on the state and observation models keeping in view the applicability of the approach to systems that require fast and real-time feedback. The transition probabilities,  $\mathcal{T}$ , depend on the image and discretization parameters which are kept constant throughout the process. Therefore, the first step is to calculate a transition probability lookup table for all possible states. At the start of the run, the belief is initialized with equal number of particles for every state. Every time an image is received, it is first discretized and each voxel is assigned the number of points contained within it. All the particles in the belief vector are then propagated forward in time by randomly drawing, for each particle, a next state from the distribution defined by  $\mathcal{T}$ . Each propagated particle is then assigned a weight based on the observation model (6), (7). Finally, the particles are drawn at random according to the distribution of weights [21]. The probability distribution over the entire state space given the history of depth images, can be computed from the belief vector at any time step. Algorithm 1 summarizes the estimation process. In the particle filter formulation, the belief about a state is represented by the number of particles associated with that state. For the rest of the paper, the posterior,  $P(\mathbf{s}_t | \mathbf{o}_0, \mathbf{o}_1, \mathbf{o}_2, \dots, \mathbf{o}_t \in \mathcal{O})$ , is represented by the normalized belief vector  $\mathbf{b}_t$  of length equal to the total number of states.

**Algorithm 1** Particle Filter Estimation

---

```

 $b \leftarrow \mathcal{U}(0, N)$ 
 $\mathcal{T} \leftarrow$  generate transition probability table given image
dimensions and discretization intervals
repeat
   $b \leftarrow \text{update\_belief}(\text{depth image}, b)$ 
function  $\text{update\_belief}(\text{depth image}, b)$ 
   $b' \leftarrow \emptyset$ 
  for  $i \leftarrow 1$  to  $N_w N_h N_d - 1$ 
     $n_i \leftarrow$  number of points inside voxel  $i$ 
    for  $i \leftarrow 1$  to  $|b|$ 
       $s_i \leftarrow$  random state in  $b$ 
      draw  $k$  with probability proportional to  $\mathcal{T}(s_i, \cdot)$ 
       $w(i) \leftarrow \text{observation\_weight}(n_k)$ 
       $s'_i \leftarrow s_k$ 
    for  $i \leftarrow 1$  to  $|b|$ 
      draw  $k$  with probability proportional to  $w$ 
      Add  $s'_k$  to  $b'$ 
  return  $b'$ 

```

---

## IV. APF CONTROL

The mapping between a pixel inside a depth image to its 3D coordinates in the camera's optical frame,  $\mathcal{F}_C$ , is defined by the camera's projection model. For a pinhole camera model, it is governed by the following set of equations,

$$\begin{aligned} s_t^x &= p_t^x(f_x/p_t^z) + c_x \\ s_t^y &= p_t^y(f_y/p_t^z) + c_y \\ s_t^z &= p_t^z, \end{aligned} \quad (8)$$

where the intrinsics are defined by the camera's focal lengths,  $f_x, f_y$ , and the location of its optical axis,  $(c_x, c_y)$ , in pixels. The coordinates of a voxel  $s_t \in \mathcal{S}$  at time instant  $t$  are referred to as  $(s_t^x, s_t^y, s_t^z)$ , with the projected 3D point being  $\mathbf{p}_t = (p_t^x, p_t^y, p_t^z) \in \mathbb{R}^3$ . Using (1), (2), (8), and a belief,  $\mathbf{b}_t$ , the net potential,  $U_{\text{net}} : \mathbb{R}^3 \times \mathbb{R}^3 \rightarrow [0, \infty)$ , on the robot is found as,

$$U_{\text{net}}(\mathbf{x}_{\text{goal}}, \mathbf{x}_{\text{obs}}) = \sum_{s \in \mathcal{S}} (U_{\text{att}}(\mathbf{x}_{\text{goal}}) + U_{\text{rep}}(\mathbf{R}_C^B \mathbf{p}_t)) b_t(s). \quad (9)$$

where,  $\mathbf{R}_C^B$  is the transformation from the camera frame,  $\mathcal{F}_C$ , to the robot's body frame,  $\mathcal{F}_B$ , and  $b_t(s)$  is the belief associated with a state  $s$  at time instant  $t$  when the potential is computed.

## V. IMPLEMENTATION AND RESULTS

For the ease of integration on an actual robot, we program the perception and control stack in Robot Operating System (ROS) C++ environment. The first subsection examines filter's performance to detect a thin cable in the presence of varying levels of added noise. The second subsection shows results from the flight testing on a real quadrotor UAV robot. For both of the tests, an Intel Realsense D435 [22] camera is

used to obtain the depth information. The image resolution is set to  $640 \times 480$  pixels.

## A. Thin Cable Test

Thin and small obstacles pose challenge to flying robots since they are extremely difficult to be mapped. Hanging cables and wires serve as one of the worst case scenarios for this category of obstacles. An extension cord, 8 mm in diameter, is chosen to put filter's performance to test while keeping the robot static. Fig. 5(a) shows the RGB image of the hanging cord as seen by the forward facing camera on the stationary robot. Figures 5(b),(c) show two snapshots of the depth image when projected to the robot's body frame coordinates. Each 3D point in the figures corresponds to a pixel in the depth image. The maximum number of points an object can occupy in a  $640 \times 480$  image is 307,200. However, the camera only manages to see less than 300 points on the cord, diminishing to far less than that in many frames.

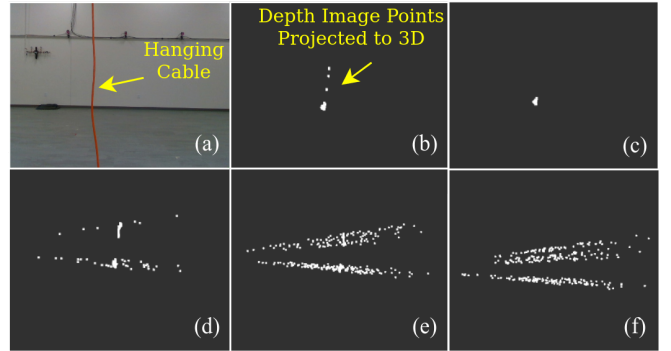


Fig. 5. A thin cable hanging in front of the robot. (a) First person view of the cable from the robot. (b),(c) Pointcloud from the side view of the cable, two instances shown. (d) Pointcloud from the side view of the cable with gaussian noise of 0.2 m added to every  $10^{th}$  pixel. (e) Pointcloud from the side view of the cable with gaussian noise of 0.2 m added to every other pixel. (f) Pointcloud from the side view of the cable with gaussian noise of 0.2 m added to every pixel.

In three different tests, we added gaussian noise of standard deviation 0.2 m to the depth associated with every pixel, every other pixel and every  $10^{th}$  pixel in every input image. A fourth test is performed as a control without any noise. The discretization intervals are chosen to be  $k_w = 50, k_h = 50, k_d = 0.1$ . The state transition and observation models' covariance parameters are set to  $\sigma_s = 8, \sigma_z = 0.4, \sigma_o = 300, \sigma_n = 1$ . Additionally, 20,000 particles are used for particle filter estimation. In the APF-PF framework, the repulsive potential applied by each depth image voxel is weighted according to its probability. The probabilities, and hence the influence of the repulsive potentials generated by different voxels, are relative. Therefore, observing the most probable voxel provides a useful insight into the perception performance while keeping the illustration legible. Fig. 6 shows the depth of the most probable voxel transitioning over time for all four noise levels. The size of each vertex in the plot is set proportional to the belief of the corresponding



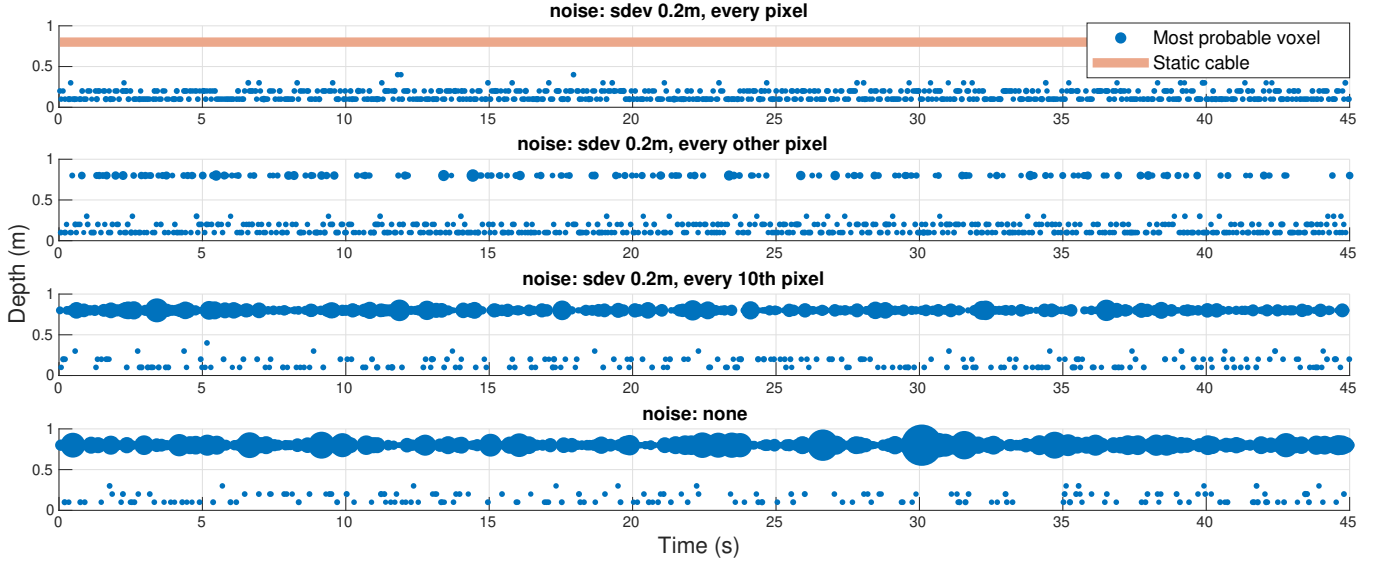


Fig. 6. The most probable occupied voxel location over time for a thin static cable test. The thickness of points is proportional to their respective probabilities.

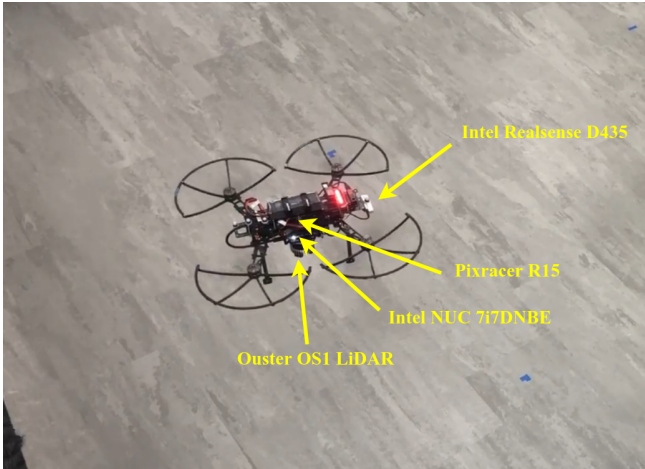


Fig. 7. An aerial view of the flying quadrotor UAV.

voxel. When the noise is added to every pixel on the cable, there is no information in the image and hence the detections do not necessarily correspond to any physical object in the scene. A bias can be seen towards closer distances due to  $P_{\text{dist}}(s_{t+1})$  term in (3). As the noise gets less dense, the detections appear to happen where the cord is located, with increasing confidence as the density of noise decreases. Like any other observer, there is a performance trade-off if an object of interest resembles the expected sensor noise, as is the case with the cord. Therefore, even with no added noise, the valid detections get less certain during some time instances. This is also because in some frames, for instance Fig. 5(b), the structure of the cord disappears to very less

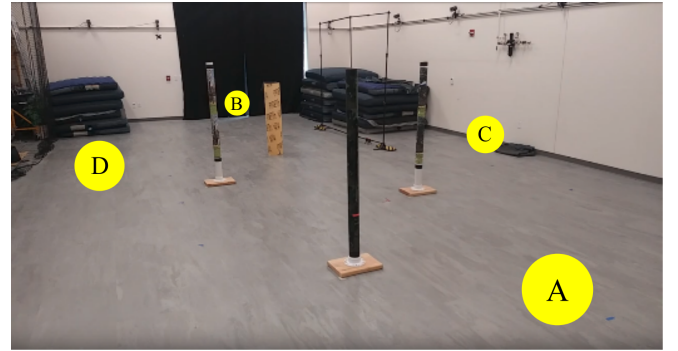


Fig. 8. The high bay flight area.

number of points before reappearing again in the depth image. The filter does reasonably well in keeping track of the object during those instances. Nevertheless, if the objects of interest in the depth images are more observable or if the noise is more sparse,  $\sigma_o$  and  $\sigma_n$  can be varied to get more accurate detections.

### B. Physical UAV Experiments

We performed several flight tests in a high bay area approximately 17 m long and 8 m wide. Results from all of the flight tests are shown in this section. The quadrotor UAV (Fig. 7) is built out of a 540 mm airframe equipped with a forward facing Intel RealSense D435 depth camera. The localization estimates on the UAV are provided by a 3D Ouster LiDAR [23], [24] along with a PX4-based flight controller [25]. The Ouster LiDAR being used has a maximum range of approximately 100 m, a resolution of

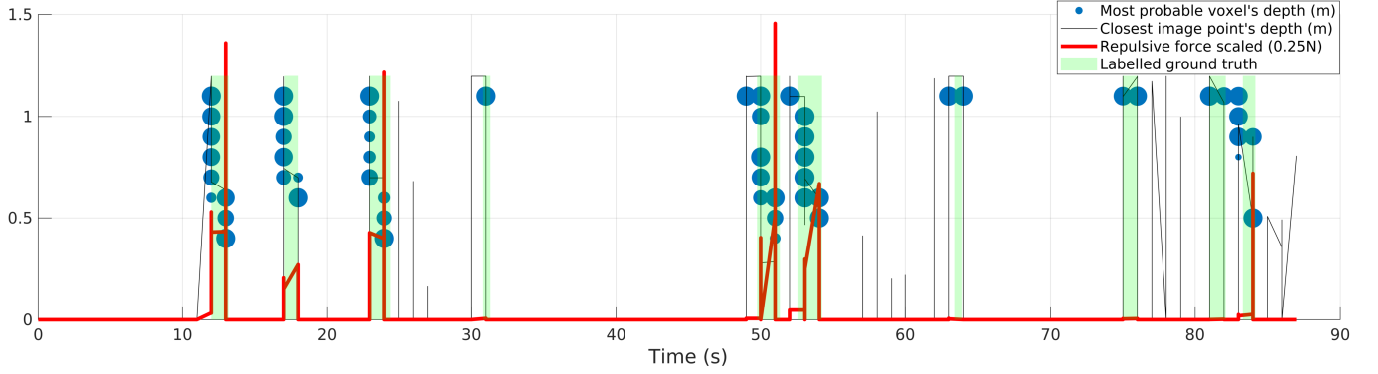


Fig. 9. A closeup of one A-B, B-C and C-D run. The green shaded regions represent carefully hand-labelled intervals during which a physical object was inside the camera's field-of-view.

1024×64 with horizontal and vertical field-of-views of 360 and 45 degrees, respectively. This makes it very well-suited for localization, but the sparseness of its depth image and the performance degradation at close ranges motivates us to use the Intel Realsense for small obstacle avoidance. Since, we are relying on one camera for these experiments, its utility is increased by adding the non-holonomic constraints on the robot in the  $x-y$  plane of its body frame,  $\mathcal{F}_B$ . The steering command is generated so that the UAV faces the negative gradient of potential,  $\nu = -\nabla U_{\text{net}}/|\nabla U_{\text{net}}|$ , defined in  $\mathcal{F}_B$ .

$$\begin{aligned} v_t^\psi &= \frac{1}{\pi} v_{\max}^\psi (\arctan 2(\nu^y, \nu^x)), \\ v_t^x &= v_{\max}^x \cos(\arctan 2(\nu^y, \nu^x)), \\ v_t^z &= v_{\max}^z \nu^z, \end{aligned} \quad (10)$$

where  $v_t^x$ ,  $v_t^z$ , and  $v_t^\psi$  are the forward and vertical velocities and the yaw rate in  $\mathcal{F}_B$ , respectively. The maximum velocities are regulated by  $v_{\max}^x$ ,  $v_{\max}^z$ , and  $v_{\max}^\psi$ . Additionally,  $\nu^x$ ,  $\nu^y$ , and  $\nu^z$  are the  $x$ ,  $y$  and  $z$  components of  $\nu$ . The APF method inherently assumes a holonomic vehicle model. However, by setting  $v_{\max}^\psi$  sufficiently larger than  $v_{\max}^x$ , the vehicle's translation closely approximates the holonomic behavior by making yaw to converge much faster than any translation component. In order to avoid the forward velocity to be negative,  $v_t^x$  is lower bounded by zero.

The covariance parameters,  $\sigma_o$  and  $\sigma_n$ , for the particle filter, are set to 170 and 60 respectively, while keeping all other parameters the same. The control parameters are used as,  $\rho_r = 0.5$ ,  $\xi = 0.4$ ,  $\eta = 1.1$ ,  $v_{\max}^x = 0.6$ ,  $v_{\max}^z = 0.6$ ,  $v_{\max}^\psi = 1.0$ . Fig. 8 shows the setup with randomly placed obstacles of different reflectivity and of diameters, 270 mm, 90 mm and 22 mm in the course. Each UAV flight path starts from one of the four regions A, B, C or D and ends in one of the three regions B, C or D. The attractor point corresponding to each region is located at its center. The combined duration of the flight tests is approximately 8 min with more than 150 m of distance traveled by the UAV safely. The UAV performs 3 A to B (A-B), 9 B to C (B-C), 8 C to D (C-D), 1 D to C (D-C) and 6 D to B (D-B) flights. The paths followed

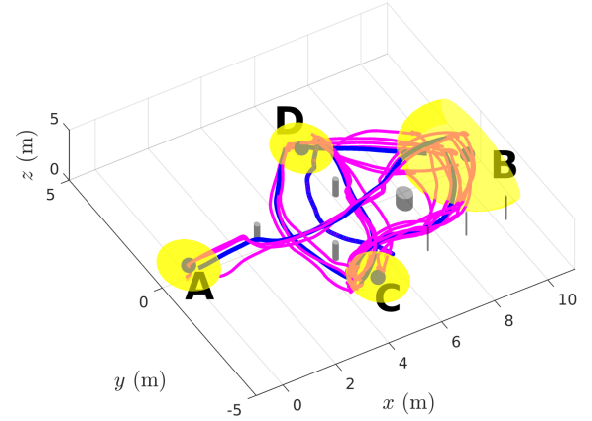


Fig. 10. Paths followed by the UAV for multiple A-B, B-C, C-D, D-C and D-B flights. Yellow spheres show the regions A, B, C and D. The blue dots represent the attractor points. The paths followed by the UAV are shown in magenta. For each flight, the UAV takes a different route depending on the initial conditions. The average path over each route is shown by a blue line.

by the UAV for all the flights, as reported by the LiDAR odometry, with the locations of the obstacles and attractors are shown in Fig. 10. For the purpose of further analysis, Fig. 9 shows the performance of the proposed method over one randomly chosen A-B-C-D run. If the APF method was applied directly on the raw sensor measurements, the area of lowest depth would make the most contribution to the repulsive potential. In that case, even low levels of noise can prove highly compromising to system's safety especially if such an area corresponds to noise. The plot shows the depth of the most probable voxel compared against the depth of the closest point inside the raw depth image, at each time instant. The magnitude of the repulsive force is overlaid on the plot. In order to highlight the frequency of raw sensor noise, we hand-labelled the sections in the plot where the UAV had a physical object inside its field-of-view. These sections are shaded green in the figure. At various instances throughout

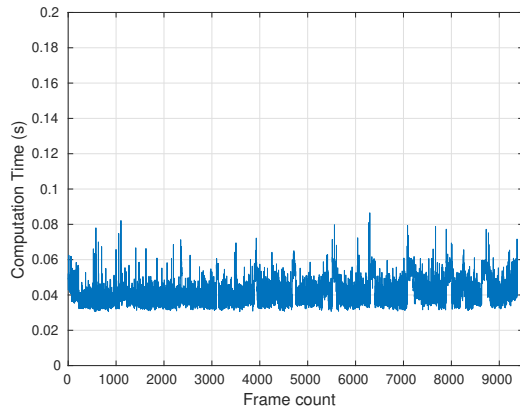


Fig. 11. Compute time over several minutes long flight data.

the flight path, the UAV perceived noise that did not belong to any physical object. The repulsive force magnitude is a direct measure of the vehicle's response to those observations. Lastly, the computation time for the complete software stack, over several thousand frames, is shown in Fig. 11.

## VI. CONCLUSION

The paper proposed a perception technique to effectively use an APF-based motion planner on a raw depth image stream. Unlike popular map-based planners, the method is well suited for real-time maneuvers in agile robots especially when precisely building a map is not feasible. The technique is shown to work effectively to perceive and avoid small and thin obstacles from a sequence of noisy depth inputs. The noise level, density and frequency varies with lighting conditions, reflectivity of the surrounding objects, and other environment artifacts. Moreover, depth from stereo is governed by different optics and computation methods than that of a Time-of-Flight (TOF) technology. The method proposed in this paper provides a flexible probabilistic framework to filter out various noise types by carefully tuning the relevant parameters. Finally, the robustness and efficiency of the perception and control stack is demonstrated through physical UAV experiments.

A video of all the flight tests is posted at <https://youtu.be/oB-WNkrNwR8>.

## ACKNOWLEDGMENT

This work was supported through the DARPA Subterranean Challenge, cooperative agreement number HR0011-18-2-0043.

## REFERENCES

- [1] DARPA. (2020) Unearthing the subterranean environment. [Online]. Available: <https://www.subtchallenge.com/>
- [2] 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, software available at <http://octomap.github.com>. [Online]. Available: <http://octomap.github.com>
- [3] H. Oleynikova, Z. Taylor, M. Fehr, R. Siegwart, and J. Nieto, "Voxblox: Incremental 3d euclidean signed distance fields for on-board mav planning," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2017.
- [4] W. H. Warren, B. A. Kay, W. D. Zosh, A. P. Duchon, and S. Sahuc, "Optic flow is used to control human walking," *Nature neuroscience*, vol. 4, no. 2, pp. 213–216, 2001.
- [5] M. A. Frye and M. H. Dickinson, "Fly flight: a model for the neural control of complex behavior," *Neuron*, vol. 32, no. 3, pp. 385–388, 2001.
- [6] A. Borst and J. Haag, "Neural networks in the cockpit of the fly," *Journal of Comparative Physiology A*, vol. 188, no. 6, pp. 419–437, 2002.
- [7] A. T. Frago, C. Cigla, R. Brockers, and L. H. Matthies, "Dynamically feasible motion planning for micro air vehicles using an egocylinder," in *Field and Service Robotics*. Springer, 2018, pp. 433–447.
- [8] L. Matthies, R. Brockers, Y. Kuwata, and S. Weiss, "Stereo vision-based obstacle avoidance for micro air vehicles using disparity space," in *2014 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2014, pp. 3242–3249.
- [9] S. Scherer, D. Ferguson, and S. Singh, "Efficient c-space and cost function updates in 3d for unmanned aerial vehicles," in *2009 IEEE International Conference on Robotics and Automation*. IEEE, 2009, pp. 2049–2054.
- [10] G. Dubey, S. Arora, and S. Scherer, "Droan—disparity-space representation for obstacle avoidance," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2017, pp. 1324–1330.
- [11] G. Dubey, R. Madaan, and S. Scherer, "Droan-disparity-space representation for obstacle avoidance: Enabling wire mapping & avoidance," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018, pp. 6311–6318.
- [12] S. Ahmad and R. Fierro, "Real-time quadrotor navigation through planning in depth space in unstructured environments," in *2019 International Conference on Unmanned Aircraft Systems (ICUAS)*. IEEE, 2019, pp. 1467–1476.
- [13] O. Khatib, "Real-time obstacle avoidance for manipulators and mobile robots," in *Autonomous robot vehicles*. Springer, 1986, pp. 396–404.
- [14] J. Conroy, G. Gremillion, B. Ranganathan, and J. S. Humbert, "Implementation of wide-field integration of optic flow for autonomous quadrotor navigation," *Autonomous robots*, vol. 27, no. 3, p. 189, 2009.
- [15] J. S. Humbert and A. M. Hyslop, "Bioinspired visuomotor convergence," *IEEE Transactions on Robotics*, vol. 26, no. 1, pp. 121–130, 2009.
- [16] M. T. O'Hradzansky, A. B. Mills, E. R. Rush, D. G. Riley, E. W. Frew, and J. S. Humbert, "Reactive control and metric-topological planning for exploration," in *2020 IEEE International Conference on Robotics and Automation*. IEEE, 2020.
- [17] R. Volpe and P. Khosla, "Manipulator control with superquadric artificial potential functions: Theory and experiments," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 20, no. 6, pp. 1423–1436, 1990.
- [18] J. Andrews and N. Hogan, "Control of manufacturing processes and robotic systems, chapter impedance control as a framework for implementing obstacle avoidance in a manipulator," 1983.
- [19] S. S. Ge and Y. J. Cui, "New potential functions for mobile robot path planning," *IEEE Transactions on robotics and automation*, vol. 16, no. 5, pp. 615–620, 2000.
- [20] C. Sweeney, G. Izatt, and R. Tedrake, "A supervised approach to predicting noise in depth images," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 796–802.
- [21] M. J. Kochenderfer, *Decision making under uncertainty: theory and application*. MIT press, 2015.
- [22] Intel. (2020) Depth camera d435 - intel® realSense™ depth and tracking cameras. [Online]. Available: <https://www.intelrealsense.com/depth-camera-d435/>
- [23] Ouster. (2020) High-resolution os1 lidar sensor: robotics, trucking, mapping. [Online]. Available: <https://ouster.com/products/os1-lidar-sensor/>
- [24] W. Hess, D. Kohler, H. Rapp, and D. Andor, "Real-time loop closure in 2d lidar slam," in *2016 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2016, pp. 1271–1278.
- [25] Dronecode. (2020) mro pixracer - px4 v1.9.0 user guide. [Online]. Available: <https://docs.px4.io/v1.9.0/en/flight-controller/pixracer.html>