2019 International Conference on Robotics and Automation (ICRA)
Palais des congres de Montreal, Montreal, Canada, May 20-24, 2019

Evaluating the Effectiveness of Perspective Aware Planning with Panoramas

Daniel Mox, Anthony Cowley, M. Ani Hsieh, C. J. Taylor

***Abstract*— In this work, we present an information based exploration strategy tailored for the generation of high resolution 3D maps. We employ RGBD panoramas because they have been shown to provide memory efficient high quality representations of space. Robots explore the environment by selecting locations with maximal Cauchy-Schwarz Quadratic Mutual Information (CSQMI) computed on an angle enhanced occupancy grid to collect these RGBD panoramas. By employing the angle enhanced occupancy grid, the resulting exploration strategy emphasizes perspective in addition to binary coverage. Furthermore, the goal selection strategy is improved by using image morphology to reduce the search space over which CSQMI is computed. We present experimental results demonstrating the improved performance in perception related tasks by capturing panoramas using this approach, near frontier exploration, and a control of logging images at regular intervals while teleoperating the robot through the workspace. Collect imagery was passed through an object detection library with our perspective aware approach yielding a greater number of successful detections compared to near frontier exploration.**

I. INTRODUCTION

A central pillar of robotics research is the development of efficient autonomous mapping and exploration strategies with the attendant development of suitable representations of the environment. Suitable representations serve as a prerequisite for fundamental tasks such as exploration and localization, with a plethora of forms emerging to suit each case. Metric maps that render spatial attributes in terms of their location in a shared coordinate frame have enjoyed immense popularity in the literature. Examples of *de facto* standards for metric maps include occupancy grids for exploration [6] and landmark maps for simultaneous localization and mapping (SLAM) [18]. On the other hand, topological maps seek to relate semantic information about the environment gleaned from sensor observations, and have been used to aid in rapid exploration [4] and efficient task allocation [16].

In typical SLAM scenarios, map representations are tailored to suite the localization and mapping system. However, choices made in support of those goals may not be well-suited for other uses of map data, such as object recognition. An example can be found in landmark based maps where only a few feature patches are extracted from an entire source image, while in voxel based grids the detailed geometry of point clouds are lumped into a small number of cells. This compression of sensor data has only been magnified with the emergence of RGBD cameras and LiDAR. These sensors combine depth and visual information that can be used to create dense, high fidelity three dimensional maps [19], but the torrent of available data must must be carefully navigated.

New approaches using RGBD cameras include the influential work of Newcombe in which dense mapping and tracking is achieved by fusing depth frames for surface reconstruction [14]. Henry employed surfels, *e.g.*, surface orientation, patch size, and color, gleaned from RGBD frames to build 3D maps of indoor environments [8]. Recent work by Taylor [17] focuses on panoramic views comprising multiple depth frames captured at a location. Panoramic depth images are advantageous for mapping because they provide both spatial information for motion planning and fine grained detail for object recognition. Furthermore, when compared to traditional grid based approaches, panoramic images have the potential to provide significantly more detail while consuming only a modest amount of memory resulting in better scaling characteristics.

Recent advances in our understanding of information theory and how to apply it to the robot exploration problem have lead to a slew of new algorithms based on a powerful principle: the map provides hints about what observations to expect at different locations and vice versa. This notion is at the heart of mutual information, which seeks to quantify the amount of information one random variable contains about another [3]. Julian demonstrated that mutual information eventually drives the robot towards unknown space and used it as an objective function for autonomous exploration using a range sensor [10]. A similar approach was shown by Charrow who used Cauchy-Schwarz Quadratic Mutual Information (CSQMI), a metric closely related to mutual information, to guide a robot equipped with an RGBD camera to map an unknown environment [2].

Other works combine object detection with robot mobility. As a part of the Multi-Autonomous Ground-robotic International Challenge (MAGIC), Reid demonstrated a multi-robot exploration approach which detected faces and objects of interest by periodically sampling a camera attached to the robot [15]. Ekvall developed a system where a service robot augmented a metric map with semantic information and detected object locations while being lead through a space [5]. And Menger et. al. won the Semantic Robot Vision Challenge with a robot that explores an unknown environment while seeking to recognize objects by selecting views for a camera weighted towards unobserved space [12].

We build upon our previous work in [13] which developed an information maximizing exploration strategy for teams of

this research was supported by ARL grant W911NF-08-2-0004 and ARO grant W911NF-13-1-0350

the authors are with the General Robotics Actuation Sensing and Perception (GRASP) Lab at the University of Pennsylvania {mox,acowley,mya,cjtaylor}@seas.upenn.edu

robots to map an unknown space. In that work we introduced a novel spatial grid representation that emphasizes *perspective* in addition to coverage to guide robots in exploring unknown spaces while collecting panoramas used to create a detailed map of the environment. Building on the same foundation, this work introduces an improved goal selection algorithm and details a set of experiments comparing the performance of our approach. In particular we pass imagery from panoramas collected using perspective-aware planning through an object detection library and compare the results against (1) panoramas collected using near frontier exploration and (2) images logged at regular intervals by teleoperating the robot through the space with the camera pointed at scene geometry.

II. METHODOLOGY

To provide a complete overview of our approach we start with a summary of our previous work in Sections II-A and II-B and detail the improved goal selection strategy in Section II-C. The objective of our work is to effectively explore an unknown environment and to collect high resolution RGBD panoramic images that can be used to produce a detailed, memory efficient representation of space.

There are a number of advantages to using panoramic imagery to represent the environment over more common grid based methods [17]. Unlike grid or voxel methods, which devote significant resources to represent regions of empty space, panoramas capture interesting scene geometry and nothing more as free space is implicit. Thus, they afford a high level of detail for the memory consumed, a desirable property when exploring realistic, expansive environments. Another advantage is that panorama synthesis can be performed on-board the robot alleviating the need for costly network transfers. Further information regarding RGBD panoramas and their construction can be found in [17].

In this work we focus on the development of an exploration strategy that selects the locations in an unknown environment at which the robot captures panoramas. This goal is comprised of two main tasks: 1) capturing panoramas an constructing local maps used for planning and 2) selecting new locations at which to collect additional panoramas. The following subsections detail each part of our approach.

A. Angle Occupancy Grid

In this work, we employ the 2D angle occupancy grid [13] to represent the workspace. The angle occupancy grid takes the familiar grid map and adds perspective by dividing each cell into a fixed number of bins, each an evenly proportioned slice of the 360° horizontal field of view. For example, an angle occupancy grid with four bins would account for the cardinal directions. In general, typical mapping strategies emphasize coverage of space with no consideration for perspective. While this is suitable in some situations, e.g., navigation, a map that can offer clear and diverse views of surfaces within the space would be useful and advantageous to a wide variety of applications. The 2D angle occupancy grid accounts for both *perspective* as well as coverage. To integrate new observations into a map represented using a 2D angle occupancy grid, only the bin corresponding to the angle at which the cell was observed needs to be updated, with the exception of free cells which appear the same regardless of the angle of incidence. Thus, a previously updated cell appears as unobserved when viewed from a new perspective and can significantly impact the selection of frontier locations for exploration.

B. CSQMI

The goal planner is tasked with identifying panorama capture locations that both expand the current understanding of the environment and afford multiple and diverse views of space. We build on our previous work in exploration [13] based on the algorithm presented in [2], which uses Cauchy-Schwarz Quadratic Mutual Information to select poses at which a predicted sensor measurement maximally decreases the uncertainty of the underlying occupancy grid. A more familiar objective function often used in exploration scenarios is mutual information (MI), which can be expressed in terms of Shannon's entropy and conditional entropy. While MI and CSQMI are not directly related, they have been shown to produce consistent results in an exploration setting [1]. For our problem, the key difference between MI and CSQMI is computational efficiency; while MI requires costly numerical integration to solve [10], CSQMI admits a closed form solution which can easily be computed online.

We represent panoramas as collections of range measurements, modeled as beams with Gaussian noise, evenly distributed 360° about the capture pose and use the aforementioned angle occupancy grid to infer a distribution over possible future measurements. Then, the expected information gain of a single beam k in a panorama, z_k, at a candidate location, x, can be computed by considering the cells in the angle occupancy grid intercepted by the beam:

$$\begin{aligned} I_{CS}[c; z_k \mid x] &= \log \sum_{l=0}^{C} w_l \mathcal{N}(0, 2\sigma^2) \\ &+ \log \prod_{i=1}^{C} (o_i^2 + (1-o_i)^2) \sum_{j=0}^{C} \sum_{l=0}^{C} p(e_j) p(e_l) \mathcal{N}(\mu_l - \mu_j, 2\sigma^2) \\ &- 2 \log \sum_{j=0}^{C} \sum_{l=0}^{C} p(e_j) w_l \mathcal{N}(\mu_l - \mu_j, 2\sigma^2), \end{aligned} \tag{1}$$

where $C = |c|$, μ_i is the distance along the beam of the center of cell c_i, $o_i = p(c_i = 1)$ is the probability that the i^{th} cell in c is occupied, $p(e_j)$ is the probability that the first occupied cell in c is c_j and $\mathcal{N}(x, \sigma) = 1/\sqrt{2\pi\sigma^2} \cdot \exp\{-x^2/2\sigma^2\}$. Additionally, each w_l is calculated from:

$$w_l = p^2(e_l) \prod_{j=l+1}^{C} (o_j^2 + (1-o_j)^2), \tag{2}$$

with $0 < l < C$, $w_0 = p(e_0)$, and $w_C = p^2(e_C)$.

Following the beam independence assumption, the expected information gained from capturing a panorama at x can be found by combining the contribution from each beam:

$$I_{CS}[m; z \mid x] = \sum_{i=0}^{K} I_{CS}[c; z_i \mid x]. \tag{3}$$

We refer the interested reader to [2] for a complete derivation of Equations 1 - 3 along with a thorough treatment of predicting sensor measurements and computational complexity. Note that beams often intercept the same cells along their length violating the CSQMI independence assumption. However, taking beam dependence into account requires computing the joint distribution over all overlapping beams in the panorama which is intractable. This shortcoming can be assuaged by computing CSQMI over a subset of beams in the panorama that can reasonably be considered independent [2].

From Equation 3, an information maximizing exploration policy emerges of the form:

$$x^* = \arg\max_{x \in \mathcal{X}} I_{CS}[m; z|x], \tag{4}$$

where $\mathcal{X}$ represents the set of all candidate capture locations.

C. Goal Selection

While approximations can be made to Equation 3 in order to boost efficiency, solving Equation 4 by searching over all free space is computationally wasteful and becomes costly as the map grows. In our previous work, a fixed number of candidate poses at which to compute CSQMI were selected by randomly sampling free space. However, as free space grows, the sampling becomes more and more sparse ultimately inhibiting exploration due to a lack of samples near unexplored regions.

In this paper we propose a new candidate goal selection approach based on image morphology. Beginning with the 2D occupancy grid representation of the environment, a binary image is generated by applying a threshold to the occupancy values. Noise inherent in the map is reduced by performing dilation on the occupied space, then the remaining free space is reduced to a pixel wide connected skeleton using a thinning algorithm. This skeleton represents a compressed version of free space and also acts as a Voronoi diagram, with skeleton pixels residing at roughly equal distances from nearby obstacles. Finally, CSQMI can be computed for the cell locations belonging to the skeleton and a suitable goal chosen.

While a search over the skeleton will yield the point with the greatest CSQMI, there are often local maxima in other regions of space also yielding large information gains. Identifying these alternative goals becomes crucial if a team of robots is tasked with exploration, wherein each agent needs a distinct goal to visit, or if other factors are considered such as travel cost, where a closer goal with slightly lower CSQMI might be chosen over the maximum but distant solution. One could apply a distance threshold or clustering algorithm to seek out these local minimum, however, an even simpler solution exists. At the location of a captured panorama, CSQMI is at a minimum as all cells within the sensors range have been observed, with additional observations from the same perspective yielding diminishing information gains. Additionally, since panorama capture locations lie on or nearby the skeleton, we can look to past capture points as a natural means of partitioning the skeleton into regions. Then, local maxima can be found by searching for the maximum along each segment of the partitioned skeleton.

Another consideration is the relative proximity of the skeleton max to the true max, found by exhaustive computation over all free space. While we make no formal claims in this regard, we note that in practice the skeleton max rarely strays more than a fraction of a meter away from the true max with the two often coincident. This should come as no surprise since skeleton points lie at points equally spaced between obstacles and thus are excellent vantage points of surrounding surfaces. The CSQMI reward surface computed over the entirety of free space along with the skeleton are shown in Fig. 1.

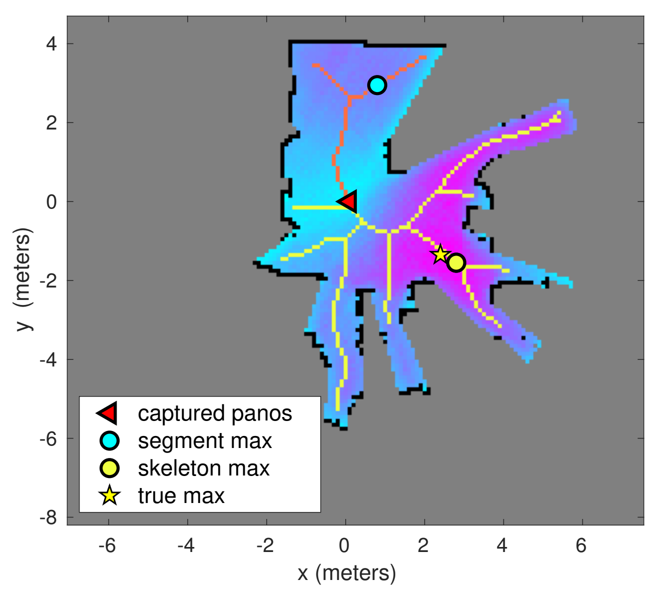


Fig. 1: CSQMI reward surface for an indoor office environment displayed as a cool heat map (turquoise is low and magenta is high) with the skeleton segments overlaid. The maximum computed over all free space lies close by the skeleton max. An additional goal is found for the upper skeleton segment.

III. Experiments

A challenge facing any evaluation of exploration strategies is defining what makes one exploration of an unknown environment better than another. While a map itself is generally intended for navigation, we are producing a map annotated with panoramic images, so we can consider an axis of measurement beyond accurate representation of navigation hazards and speed of mapping: are the collected images generally useful?

To focus on one example of using image data garnered from an exploration, we evaluate how useful the collected imagery is to a subsequent object recognition task. The overall claim is that a perspective aware exploration strategy yields more thorough, better quality image coverage of the environment. We evaluate this claim by considering an exploration task carried out by a robot building a map sparsely annotated with panoramic images. The robot explores an environment, and its collected data is then fed into an object recognition task that can identify what was found in the environment. We then compare the performance of the data garnered from our approach, using CSQMI on an angle occupancy grid, with (1) exploration driven by teleoperation, in which a human continuously aims the camera at objects of interest, and (2) an autonomous frontier exploration strategy operating on a binary occupancy grid that fills out a map without considering image quality. During frontier and CSQMI exploration panoramas are captured at the goal locations prescribed by each algorithm while during teleoperation images are logged at regular intervals while the robot is driven through the space. The frontier method is a common baseline for exploration strategies being both simple and effective at building a map of unknown space. Teleoperating the robot with the camera aimed at scene geometry represents a best case scenario where objects are clearly captured in the logged images and easily decipherable by a detection library. Each approach to exploring the environment builds a useful navigation map, but we can quantify the utility of the image data collected during the exploration.

The presented perspective aware mapping and exploration system was implemented in C++ with ROS and run on the scarab platform developed at the University of Pennsylvania (Fig. 2a). The Scarab is a differential drive ground platform with an ASUS Xtion Pro Live camera providing RGBD imagery and a Hokuyo UTM-30LX for laser based localization using the popular gmapping ROS packaged which implements the SLAM approach outlined in [7]. Live experiments were conducted at the Penn Engineering Research and Collaboration Hub (PERCH), a visually dense environment providing a diverse set of geometry and objects interesting for panorama capture (see Fig. 2b).

IV. RESULTS

We evaluate the presented perspective aware planning approach based on its ability to collect imagery useful for object detection. Five objects from the COCO dataset [11] where distributed throughout the exploration space and the collected imagery was processed using the TensorFlow object detection library using one of the available models trained on the COCO dataset, in particular `faster_rcnn_resnet50_coco` [9]. The performance (i.e. number of objects detected) of our approach was pitted against the near frontier exploration algorithm, where the nearest border between free and unknown space is selected as the next goal, and a "ground truth" run where the robot was teleoperated through the space while logging images with the camera kept facing surface geometry (a best case scenario for object detection). Four sets of trials were conducted and between each the target objects were redistributed throughout the space in locations specifically chosen to provide a challenge to the autonomous exploration methods, namely in orientations that only visible from certain viewpoints.

(a) (b)

Fig. 2: (a) the scarab differential drive ground platform and (b) a portion of the space at PERCH used in the experiments.

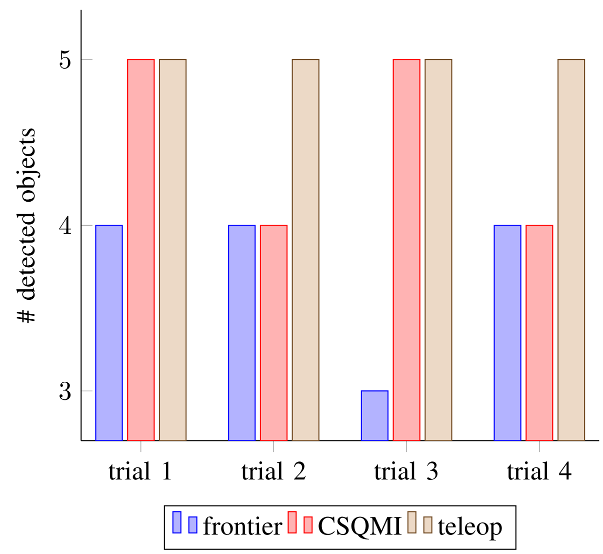


Fig. 3: number of objects detected for each planning method for each trial.

	trial 1	trial 2	trial 3	trial 4
CSQMI	8	9	8	8
frontier	5	6	6	6

TABLE I: the number of panoramas captured by each method during each trial.

Shown in Fig. 3 are the results of the four sets of trials with the corresponding number of panoramas captured by each method shown in TABLE I. Fig. 4 shows the paths and panorama capture locations of the robot during a set of trials. As expected, the ground truth run performed the best followed by our perspective aware planning approach and the near frontier method.

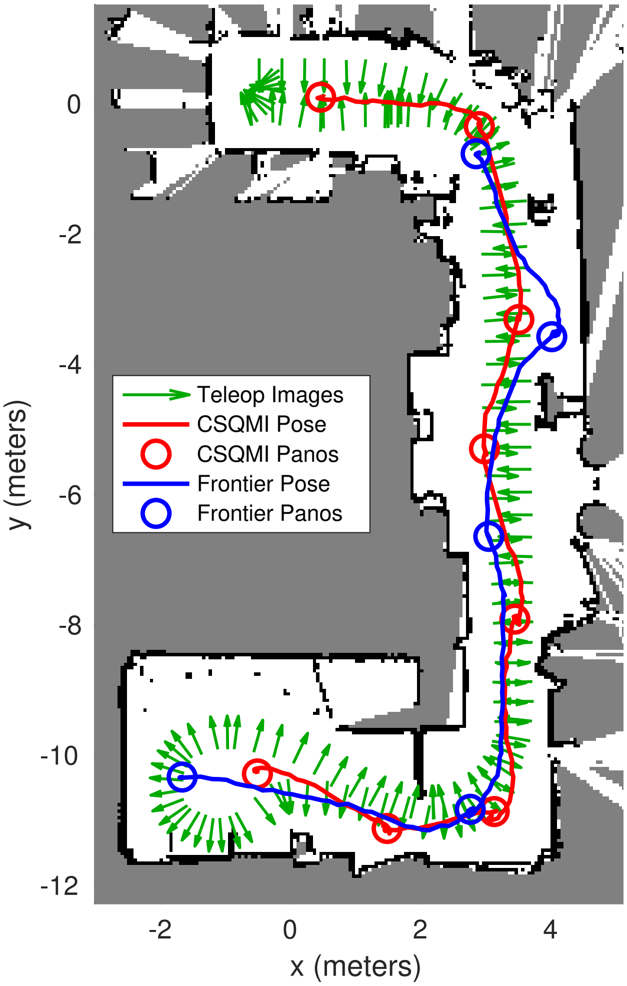
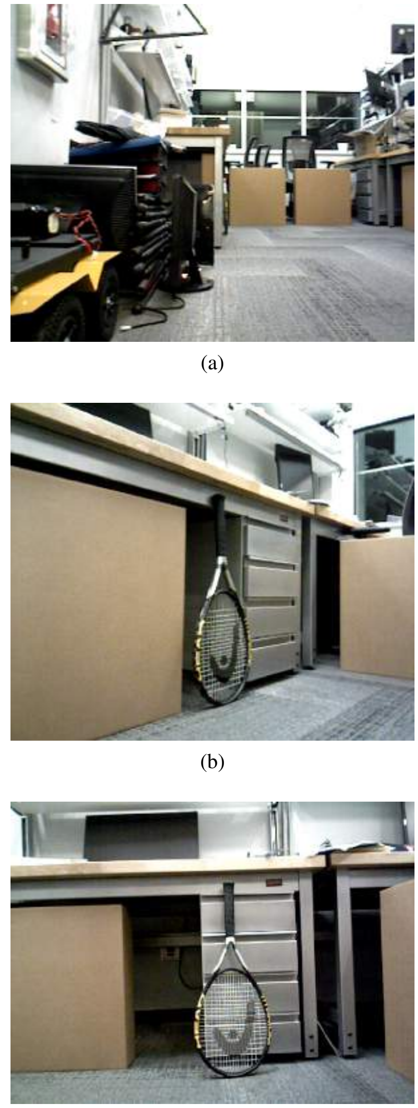


Fig. 4: robot pose and panorama capture locations during trial 1 for perspective aware exploration and frontier exploration as well as locations and camera orientation of images logged while teleoperation



(a)

(b)

(c)

Fig. 5: panorama frames captured during trial 1 showing the best view of the tennis racket in the corridor at the top of the map from (a) near frontier and (b) CSQMI and (c) the clearest image collected during teleoperation.

The impact of perspective aware planning is illustrated by Fig. 4. Frontier exploration captures fewer, sparsely distributed panoramas and concludes once the space has been completely covered, neglecting to travel to the end of the corridor in the top of the space. The perspective aware approach chooses to collect an additional panorama in this region as it yields sufficiently new views of surfaces, a consequence of the angle occupancy grid, ultimately leading to the detection of an object located there. Fig. 5 shows a stark difference. The tennis racket is clearly captured by our perspective aware approach (Fig. 5b) while it is hardly visible but for a side profile in the frontier panorama and goes unrecognized by the detection library (see center of Fig. 5a). In fact, poor views of objects of interest was the leading cause of detection failures. Note that no knowledge of object locations was given to the autonomous exploration strategies; instead we rely on their implicit determination of map completeness.

It may seem trivial that the greater number of panoramas captured by our angle-aware approach would lead to more object detections than the frontier method. However, simply capturing more imagery does not solve the problem. Consider a simple heuristic directing the robot to collect an additional panorama at the midpoint of the path to the nearest frontier. In this case the frontier approach would collect 9 panoramas as opposed to 8 by our method. But after collecting the 9th panorama at the blue circle (approximately $(x, y) = (3, -1)$ in the map shown in Fig. 4) no more frontiers exist in the map and exploration would conclude without additional views of the upper segment of the environment. This limitation was also noted in our previous simulations in [13] and highlights the limitation of binary coverage as a proxy for map quality. The merit of our method lies not the number of panoramas collected but where they are captured, offering diverse views of space that result in more successful obstacle detection.

V. CONCLUSIONS

In this paper we have presented an information based exploration algorithm using panoramas that accounts for perspective in addition to coverage. We introduced a goal selection strategy that efficiently identifies useful panorama capture locations by computing mutual information over a skeleton of free space. The effectiveness of our approach was demonstrated through a set of experiments where the robot was tasked with exploring a space, and the collected panoramas where then used to detect objects. Our method resulted in more objects detected as compared to near frontier exploration.

REFERENCES

[1] Benjamin Charrow. Information-theoretic active perception for multi-robot teams. 2015.

[2] Benjamin Charrow, Sikang Liu, Vijay Kumar, and Nathan Michael. Information-theoretic mapping using Cauchy-Schwarz Quadratic Mutual Information. *2015 IEEE International Conference on Robotics and Automation (ICRA)*, pages 4791–4798, 2015.

[3] Thomas M Cover and Joy A Thomas. *Elements of information theory*. John Wiley & Sons, 2012.

[4] Anthony Cowley, Camillo J. Taylor, and Ben Southall. Rapid multi-robot exploration with topometric maps. *Proceedings - IEEE International Conference on Robotics and Automation*, pages 1044–1049, 2011.

[5] Staffan Ekvall, Patric Jensfelt, and Danica Kragic. Integrating active mobile robot object recognition and slam in natural environments. In *IROS*, pages 5792–5797, 2006.

[6] Alberto Elfes. Using occupancy grids for mobile robot perception and navigation. *Computer*, 22(6):46–57, 1989.

[7] Giorgio Grisetti, Cyrill Stachniss, Wolfram Burgard, et al. Improved techniques for grid mapping with rao-blackwellized particle filters. *IEEE transactions on Robotics*, 23(1):34, 2007.

[8] Peter Henry, Michael Krainin, Evan Herbst, Xiaofeng Ren, and Dieter Fox. RGB-D mapping: Using depth cameras for dense 3D modeling of indoor environments. *Springer Tracts in Advanced Robotics*, 79:477–491, 2014.

[9] Jonathan Huang, Vivek Rathod, Chen Sun, Menglong Zhu, Anoop Korattikara, Alireza Fathi, Ian Fischer, Zbigniew Wojna, Yang Song, Sergio Guadarrama, and Kevin Murphy. Speed/accuracy trade-offs for modern convolutional object detectors. *CoRR*, abs/1611.10012, 2016.

[10] B. J. Julian, S. Karaman, and D. Rus. On mutual information-based control of range sensing robots for mapping applications. *The International Journal of Robotics Research*, 33(10):1375–1392, 2014.

[11] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755. Springer, 2014.

[12] David Meger, Per-Erik Forssén, Kevin Lai, Scott Helmer, Sancho McCann, Tristram Southey, Matthew Baumann, James J Little, and David G Lowe. Curious george: An attentive semantic robot. *Robotics and Autonomous Systems*, 56(6):503–511, 2008.

[13] Daniel Mox, Anthony Cowley, M Ani Hsieh, and CJ Taylor. Information based exploration with panoramas and angle occupancy grids. In *Distributed Autonomous Robotic Systems: The 13th International Symposium*, volume 6, page 45. Springer, 2018.

[14] Richard A Newcombe, Shahram Izadi, Otmar Hilliges, David Molyneaux, David Kim, Andrew J Davison, Pushmeet Kohi, Jamie Shotton, Steve Hodges, and Andrew Fitzgibbon. Kinectfusion: Real-time dense surface mapping and tracking. In *Mixed and augmented reality (ISMAR), 2011 10th IEEE international symposium on*, pages 127–136. IEEE, 2011.

[15] Robert Reid, Andrew Cann, Calum Meiklejohn, Liam Poli, Adrian Boeing, and Thomas Braunl. Cooperative multi-robot navigation, exploration, mapping and object detection with ros. In *Intelligent Vehicles Symposium (IV), 2013 IEEE*, pages 1083–1088. IEEE, 2013.

[16] Cyrill Stachniss, Óscar Martínez Mozos, and Wolfram Burgard. Efficient exploration of unknown indoor environments using a team of mobile robots. *Annals of Mathematics and Artificial Intelligence*, 52(2):205–227, 2008.

[17] Camillo J. Taylor, Anthony Cowley, Rafe Kettler, Kai Ninomiya, Mayank Gupta, and Boyang Niu. Mapping with Depth Panoramas. *Intelligent Robots and Systems (IROS 2015), 2015 IEEE/RSJ International Conference on*, pages 6265–6272, 2015.

[18] Sebastian Thrun, Wolfram Burgard, and Dieter Fox. *Probabilistic robotics*. MIT press, 2005.

[19] Thomas Whelan, Michael Kaess, Maurice Fallon, Hordur Johannsson, John Leonard, and John McDonald. Kintinuous: Spatially extended kinectfusion. *MIT-CSAIL-TR-2012-020*, 2012.