HAL
archives-ouvertes.fr

Combined Vision and Frontier-Based Exploration Strategies for Semantic Mapping

Islem Jebari, Stéphane Bazeille, David Filliat

▶ To cite this version:

Combined Vision and Frontier-Based Exploration Strategies for Semantic Mapping

Islem Jebari[1], Stéphane Bazeille[1], and David Filliat[1]

[1] Electronics & Computer Science Laboratory, ENSTA ParisTech, Paris, France

Abstract. We present an approach to multi-objective exploration whose goal is to autonomously explore an unknown indoor environment. Our objective is to build a semantic map containing high-level information, namely rooms and the objects laid in these rooms. This approach was developed for the Panoramic and Active Camera for Object Mapping (PACOM)[1] project in order to participate in a French exploration and mapping contest called CAROTTE[2]. To achieve efficient exploration, we combine two classical approaches: frontier-based exploration for 2D laser metric mapping and next-best view computation for visual object search. Based on a stochastic sampling strategy, this approach looks for a position that maximizes a multi-objective cost function. We show the advantage of using this combined approach compared to each particular approach in isolation. Additionally, we show how an uncertainty reduction strategy makes it possible to reduce object localization error after exploration.

Keywords: SLAM, multi-objective exploration, semantic mapping.

1 Introduction

The Panoramic and Active Camera for Object Mapping (PACOM)[1] project aims at developing a robotic system dedicated to indoor autonomous exploration and mapping. In particular, we seek to extract high-level semantic information that is easy to understand and important to the robot users such as surrounding objects and the environment structure.

The semantic map we build is based on a 2D occupancy grid to which the location of objects detected using vision is added. The exploration strategy should therefore both ensure that the coverage of the occupancy grid is complete and that all the objects of the environment have been perceived. We initially considered a purely vision-based exploration strategy that gave overall correct results and covered almost all the area using vision in most cases but had disadvantages such as incoherent and unnecessary long trajectory of the robot. In order to solve these problems, we chose to adopt a multi-objective exploration strategy in which we fuse a frontier-based exploration criteria dedicated to occupancy grid construction, a vision-based criteria dedicated to object search and a distance criteria in order to shorten exploration trajectories.

This paper is organized as follow: next section overviews the related work. Section 3 describes the semantic map construction process. The multi-objective exploration is described in Section 4 and Section 5 gives experimental results.

2 Related Work

Autonomous exploration techniques can be classified according to two distinct approaches. In the first one, exploration strategies make use of predefined trajectories [1] which are static and defined off-line, exploiting some a priori information about the environment. In the second approach, exploration is

viewed as a sequence of steps, each one composed of a movement towards a location and of an observation with which the robot acquires data about the environment. The exploration of an unknown environment using a 2D scanning laser sensor is often performed using the frontier-based exploration algorithm [2] or an exploration strategy choosing the next best position for the robot given the utility of this position for the mapping problem [3]. In the context of our project, the exploration should take into account the need to search for objects. This general problem of optimally covering the whole environment with several viewpoints is referred to as the "art gallery problem" [4] and is known to be NP-hard. However, several practical solutions exist, for example for optimizing the probability of finding the target given a limited number of actions [5].

3 Semantic Mapping

The semantic map build by our robot integrates the 2D metric map of the environment, the location of objects detected using vision and the segmentation of the free space into distinct rooms. The object recognition algorithm we developed is able to detect initially known objects in the robot's visual environment and to estimate their location in the camera reference frame (see [6] for details). Object detection is performed using a two-step method: first, a fast extraction of the salient regions [7] is performed in order to segment multiple possible objects inside an image. Second, a local features extraction [8] and a bag of visual words algorithm [9] is used to recognize the segmented objects. Once an object is detected, its distance from the camera is estimated using the scale difference between current and reference image, using SIFT keypoint matching between the two images.

   The position, name and images of the detected objects along with the segmented rooms are added to the 2D occupancy grid to create the semantic map. Rooms are detected in the occupancy grid through an algorithm that detects doors and analyses the resulting connected components of open space, recognizing as rooms the areas with a sufficient size. The positions in the absolute reference frame of the detected objects are estimated using a Kalman filter for each object.

4 Multi-Objective Exploration for Semantic Mapping

   The main goal of our project is to build such a semantic map to describe as much of the environment as possible, as quickly as possible. Exploration is therefore required to completely map the environment using the laser scanner and to detect by image processing using a pan-tilt camera the objects laid on the floor. We took advantage of the fact that the two sensors have similar field of view (a semi-circle in front of the robot) to integrate these two objectives into a single function. Additionally, we used a map improvement strategy in order to improve the precision of object localization after exploration is completed.

   Computing an optimal set of positions to cover the environment is a NP-hard problem. We therefore use a stochastic sampling strategy inspired by the Randomized Art-Gallery Algorithm [10] to search for the next robot position that discovers the more unmapped area. For each sampled position, a score is computed integrating several criteria:

$$S = \lambda_{obj} S_{obj} + \lambda_{map} S_{map} + \lambda_{dist} S_{dist} \qquad (1)$$

where S_{obj} is a score giving the area potentially discovered by vision, S_{map} is a score related to unknown environment discovery using laser and S_{dist} is a score giving the cost of going to this position from the current position. Each score is normalized between 0 and 1 in order to remove the influence of their different orders of magnitude. The sampled position with the highest score is taken as the next position the robot should reach using the Navigation subsystem while performing SLAM (Simultaneous Localization and Mapping) and object detection. When the position is reached, the camera is swept from left to right to perform object detection and a new exploration position is computed. The exploration is performed until a given percentage (95% in this paper) of the free area has been observed by the camera.

   For object search, the score S_{obj} is computed as the number of cells of the open area visible through the camera that has not been observed yet. The visible area takes the pan-tilt capability of the camera into

account to compute the left and right limit of the field of view. The minimal resolution needed for object recognition and the minimal object size are taken into account to compute the minimum and maximum perception distance. The unobserved area is computed using a 2D view-map that records the area previously perceived by the robot camera. The view-map is updated during robot movements using the camera angle of view, the camera being oriented in front of the robot. When the robot reaches an exploration goal, the area perceived by moving the camera left to right is recorded. Once the view map is updated, the connected components of the unobserved open areas are computed (see the colored areas in the right part of Fig. 1 for example).
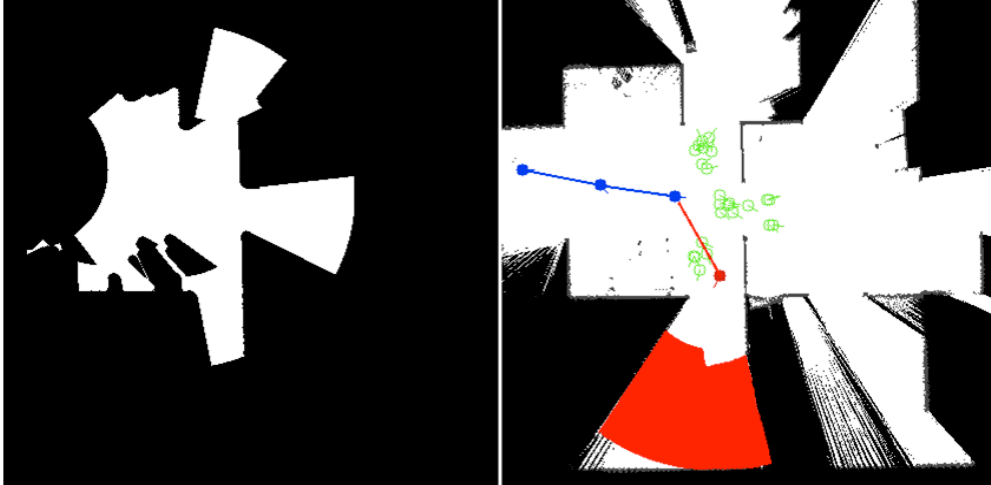
Fig. 1. Left: View-map recording already seen areas. Right: Exploration trajectory with the sampled positions *(in green)*, the selected next position and the next observed area *(in red)*.

The score S_{map} is linked to frontiers discovery. The idea behind frontier-based exploration [11] is to gain new information about the world by moving to the boundary between open space and unexplored space. Then, the robot can see into unexplored space and add the new information to its map. Any open cell adjacent to an unknown cell is labeled as frontier cell. Connected frontier cells are grouped into frontier regions that will be used in the sampling strategy. The score S_{map} is computed as the number of frontier cells between free and unknown area that are visible through the laser sensor from the sampled position.

The score S_{dist} is computed as the opposite of the travel distance from the current robot position. The distance of all grid cells to the current robot position is pre-computed before sampling the positions using the Dijkstra algorithm in order to rapidly evaluate this criterion for any sampled position.

As our approach is a greedy method that selects the next best position, but does not take the whole trajectory into account, the sampling strategy used to select the positions to be evaluated (50 in this paper) may have an influence on the resulting exploration trajectory. In order to favor more continuous trajectories, the positions are first sampled inside the free space in a semi-circular area in front of the robot, with random orientations. If no reachable position can be found in this area or if all the positions have an exploration score below a threshold, positions are sampled from the whole map in order to ensure global covering. These positions are sampled from the connected components of unseen areas and frontier cells by computing robot positions inside the open space that will observe these areas.

A complementary strategy has been implemented to reduce the error position of the detected objects at the end of the multi-objective exploration. We select in our map the objects whose uncertainty is above a threshold and we sample uncertainty reduction positions along the small axis of the object uncertainty ellipses. We then select the position that is the closest to the current robot position. When the position is reached, the Kalman filters are updated to reduce this error if the object is correctly detected. If the object is not detected, it is removed from the semantic map.

5 Experimental Results

Experiments have been performed in a 63 m^2 indoor environment composed of 2 rooms and 2 corridors. We use a Pioneer 3 dx mobile base, with a SICK Laser Range Finder and a Pan-Tilt-Zoom camera. We performed several experiments using vision criteria, frontier criteria and both in order to demonstrate the interest of our approach.

5.1 Vision-Based Exploration

The vision-based exploration strategy using parameters $(\lambda_{obj}, \lambda_{map}, \lambda_{dist}) = (1, 0, 0.6)$ is the first approach we explored because it should be sufficient for correct mapping. Indeed, the laser range is larger than the camera range. So, by trying to cover visually the environment to detect objects we should also discover the whole map using laser.
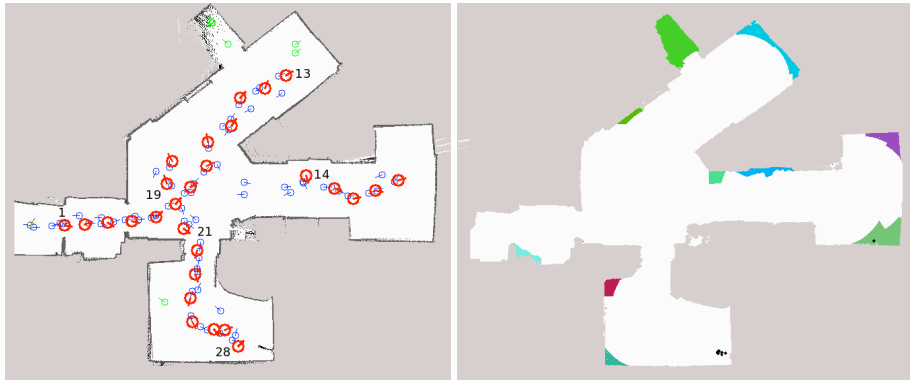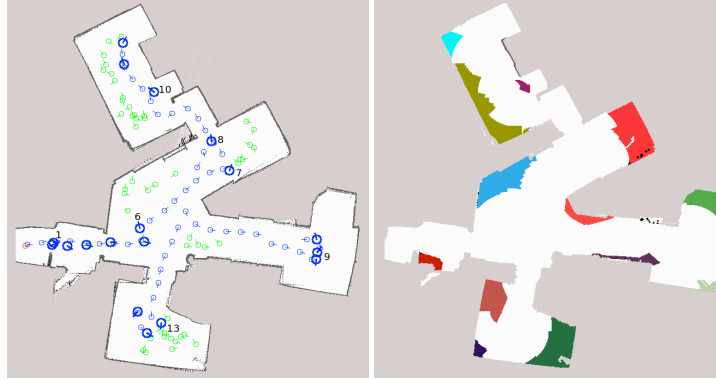

Fig. 2. Left: Vision-based exploration map: red circles correspond to exploration points with the robot orientation. Right: view-map showing areas perceived by the camera *(white areas)*.

Moreover, using this strategy, images are taken mostly when the robot is stopped, resulting in good image quality. However, as seen on the left of Fig. 2, the trajectory generated is disordered and not optimal. Indeed, as we are looking for the larger unseen area, the robot travels more than it needs. As a result, we obtain a complicated trajectory and a slow exploration. A second limitation is that the map doesn't completely cover the area to be explored (the room on the top left wasn't explored) because the robot did not get close enough to discover the room using the laser.

5.2 Frontier-Based Exploration

The frontier-based exploration strategy using parameters $(\lambda_{obj}, \lambda_{map}, \lambda_{dist}) = (0, 1, 0.6)$ is the most suitable to give a complete laser map. It is the fastest way to discover the environment and the resulting trajectory is consistent because exploration points are limited. As an example, since we are looking for edges in the laser map it is not possible to miss the top left room (see Fig. 3). However, this method presents some limitations. Using only this strategy many parts are not explored visually and as images are all recorded during motion between exploration points, the object detection is sometimes missed or less accurate because of blurry images.
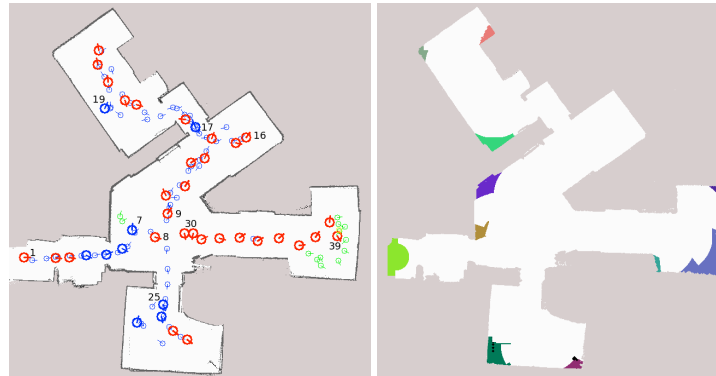
Fig. 3. Left: Frontier-based exploration map: dark blue circles correspond to exploration points with the robot orientation. Right: view-map showing areas perceived by the camera *(white areas)*.

5.3 Multi-Objective Exploration Strategy

Fig. 4 shows a map generated with parameters $(\lambda_{obj}, \lambda_{map}, \lambda_{dist}) = (1, 1, 0.6)$. This multi-objective exploration strategy combines the advantages of the two previous methods. As they are complementary the resulting semantic map is complete from a laser point of view, the vision covers nearly the entire laser map and the exploration trajectory is rather consistent. In addition, thanks to the distance parameter, the robot avoids coming back to a previously visited area (a room is completely visited the first time the robot comes inside).

As it is shown on Fig. 4, the multi-objective strategy can switch automatically between visual exploration and laser exploration. During the exit of the small starting room, the score of the first exploration points are mostly due to the frontier criteria because frontier edges are very large. After that, the laser exploration score gets smaller and vision is privileged in order to complete the map. When every near area is seen, frontier exploration becomes stronger again and lead to discover a new room.

Fig. 4. Left: Frontier & Vision-based exploration map: dark blue circles correspond to exploration points where frontier score is higher than vision score; red points correspond to positions where vision score is higher. Right: view-map showing areas perceived by the camera *(white areas)*.

5.4 Semantic Map Improvement

The final result corresponding to the semantic map obtained at the end of the exploration run is shown on Fig. 5 (left). Green color corresponds to a correct detection with a correct label. Red color refers to a false detection and orange color refers to missed objects. Some images of the detected objects are shown with frames using the same code color as the ellipse of uncertainty. Room detection algorithm successfully detected the rooms and the main corridor. The map improvement module was developed to verify the object detection. That is to say, to delete false detection and to reduce the ellipses of uncertainty. Purple points in Fig. 5 (left) show the positions the robot reached in order to reduce the

ellipse of uncertainty, along with some images taken by the robot from these positions. Fig. 5 (right) shows the semantic map after the map improvement module. The ellipses of uncertainty are smaller and one false detection is deleted. Room detection algorithm successfully detects the rooms and the main corridor.

Fig. 5. Left: Semantic map showing the environment structure and the objects position with their ellipses of uncertainty. Right: Semantic map after the map improvement module.

6 Conclusion & Perspectives

In this paper, we present our novel multi-objective autonomous exploration strategy in order to build a complete semantic map of an unknown indoor environment. Our multi-objective strategy provides better performances than single-objective explorations since it successfully integrates both frontier-based and vision-based strategies to benefit from their advantages. Our solution leads to complete laser-based exploration, complete vision-based area coverage, a consistent and nearly optimal trajectory and a better quality of detection as we added a map improvement module. It has to be noticed that our strategy is based on random sampling and may produce results that are intuitively non-optimal, depending on the choice of the multi-objective function parameters. Experimental tuning of the parameters was done for this paper to make it possible to reach good behaviors, but we are developing quantitative experiments using a simulator to have a better estimate of the pros and cons of the proposed approach. However, a difficulty is to find objective performance metrics for the semantic mapping problem.

Currently, our final map improvement strategy is a separate behavior executed after exploration is complete. In future work, we plan to integrate this objective directly in the multi-objective function in order to give a more accurate position of the object after exploration. This should make the robot to move and to take more images of an object after the first detection and should avoid coming back later.

References

1. Bourque, E., Dudek, G.: Viewpoint selection-an autonomous robotic system for virtual environment creation. In: IROS, vol. 1, 1998, pp. 526–532.
2. Yamauchi, B.: Frontier-based exploration using multiple robots. In: Proceedings of the Second international Conference on Autonomous Agents, Minneapolis, Minnesota, United States (1998)
3. Tovar, B., Muñoz-Gomez, L., Murrieta-Cid, R., Alencastre-Miranda, M., Monroy, R., Hutchinson, S.: Planning Exploration Strategies for Simultaneous Localization and Mapping. In: Journal on Robotics and Autonomous Systems, vol. 54(4), pp. 314-331 (2006)
4. Shermer, T.C.: Recent results in art galleries. In: Proceedings of the IEEE, pp. 1384-1399 (1992)
5. Tsotsos, J.K., Shubina, K.: Attention and Visual Search: Active Robotic Vision Systems that Search. In: The 5th International Conference on Computer Vision Systems (2007)
6. Jebari , I. & al.: Multi-sensor semantic mapping and exploration of indoor environments. In: Proceedings of the IEEE International Conference on Technologies for Practical Robot Applications (TePRA2011), Massachusetts, USA (2011)

7. Butko, N.J., Zhang, L., Cottrel, G.W., Movellan, J.R.: Visual Saliency Model for Robot Cameras. In: International Conference on Robotics and Automation (2008)
8. Filliat, D.: A visual bag of words method for interactive qualitative localization and mapping. In: Proceedings of the International Conference on Robotics and Automation (ICRA) (2007)
9. Sivic, J., Zisserman, A.: Video Google: A Text Retrieval Approach to Object Matching in Videos. In: Proceedings of the International Conference on Computer Vision (2003)
10. Gonzalez-Banos, Latombe, J.: A randomized Art-Gallery Algorithm for Sensor Placement. In: Annual Symposium on Computational Geometry (2001)
11. Yamauchi, B.: A frontier-based approach for autonomous exploration. In: Pproceedings of the 1997 IEEE International Symposium on Computational Intelligence in Robotics and Automation (1997)