Active Perception in SLAM Exploration: Problem Formulation and Methods Review

Wenlong Du

Sino-British College, University of Shanghai for Science and Technology, Shanghai, China

sbc-20-1045@sbc.usst.edu.cn

Abstract. Active SLAM (A-SLAM) aims to enable an agent's autonomous exploration and mapping task, provide robustness and accuracy in both localization and mapping and enhance environmental adaptivity in applications. Its working principle, as a holistic probabilistic-based problem in model formulation, can be divided into three decoupled working stages and accordingly solved, these stages are brief action identification, cost and utility computation, and action execution. The identification stage needs reasoning on what and where to look at in order to diminish uncertainty in an efficient fashion, by employing sensory methods and performing searching or optimization tasks with appropriate environmental representations, to feed the following stages with a discretized set of planning goals. This is the active perception problem by definition. The activeness of active SLAM is mostly decided at the perception stage with appropriate quantifying measures. This paper revisits the formulation of the A-SLAM problem especially by the stage of active perception, reviews its environmental representations (maps) and related modeling measures for information and uncertainty, underlines a sample-based aim of maximizing the information gain from sensors with respect to corresponding maps, and make comments on some novel approaches.

Keywords: Active perception, Simultaneous localization and mapping, Robotic exploration.

1. Introduction

Fully cognition to environment of an autonomous robot has always been a dream for robotics researchers. On the way towards this goal, people had formulated a framework to achieve simultaneous localization and mapping (SLAM), which can effectively handle sensory data from cameras, IMUs, etc., and apply those data to make an estimate of an agent's current state. Cumulative error in state estimation will be significant throughout time, irrespective of which kind of estimate method is applied if performed without loop closure detection [1].

Loop closure detection, by re-visiting previous landmarks, can take trim on state error and generate a globally consistent map. This approach works inherently deliberately toward the objective of error correction and can be assigned further automation. Fig.1 briefly illustrates the working principle for visual SLAM, with a stress on the significance of loop closure detection.

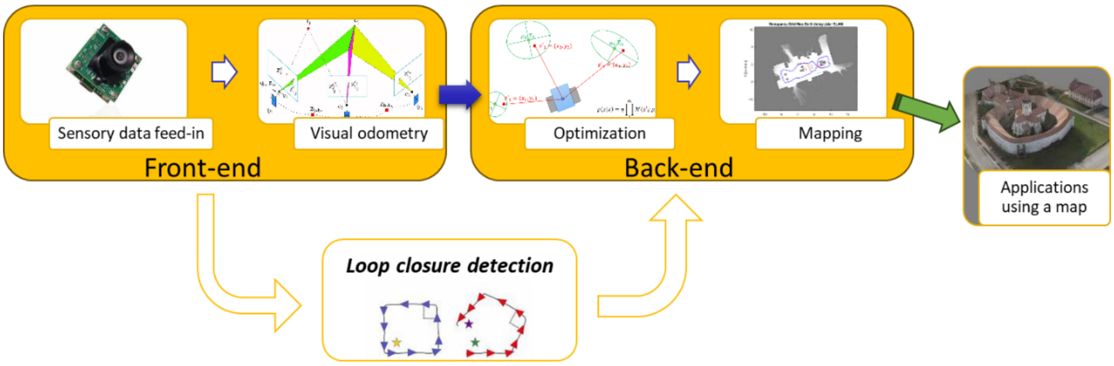


Figure 1. An illustration of general working principle for visual SLAM

This classical framework works well when passive guidance is presented in creating a map. Applications for SLAM, however, address their expectation for robots on a higher autonomous and robustness level. An agent is expected to explore actively by itself to generate a map of unfamiliar environment. In obstacle-free outdoor open areas, exploration tasks can be guided by GNSS for it continuously reports the absolute coordinate of an agent.

However, in some other exploration scenarios losing tracking with ground-truth, for example, underwater, indoor, where satellite signal is denied, traditional point-to-point GNSS-based navigation cannot be applied. Fig.2 illustrates these situations. To bridge the gap between SLAM technology and exploration tasks in a complex environment, the concept of active SLAM (A-SLAM) has been formulated and continuously attracting research attention. It aims to conduct an autonomous exploration and mapping process, building a high-fidelity representation of surroundings with minimum uncertainty and sufficiently low cost in energy consumption, trajectory length, etc.

Figure 2. Intended applications for A-SLAM

The problem of A-SLAM has different probabilistic-based mathematical representations, there also exists applications with control theoretical formulation like [2], however, most approaches decouple the entire problem into 3 sub-stages [3, 4]. The first stage identifies potential points of interest as action goals from a map, which is highly dependent on sensory apparatus as well as environmental representation, i.e. active perception methods that this paper major concerns. The second and third stages are problem-solving stages by information acquired in the first stage, leveraging current optimizers and path planners. This paper will explain the necessity of applying sensors and maps to achieve active perception in the first stage, by examining its motivation (why) and reviewing state-of-the-art approaches, formulating the methodology (how), and proposing some novel perspectives based on research trends.

2. Motivation for active perception in SLAM exploration

The concept of active perception, first proposed by [5], has motivated various ideas in robotics, including active SLAM. Revised by [6] with a modern view, it dedicates to reasoning an agent's action with respect to expectations in a particular task, by which mostly concerns area coverage [2], [7, 8]. In the early practices of active robot exploration, due to the absence of reasoning, a robot just randomly chose its observation goal or required human interaction until the concept of frontiers had been proposed [4].

Frontiers are the boundaries between known and unknown regions, it needs an appropriate environmental representation, e.g. an occupancy grid (OG) map as a prior, to be sufficiently discovered. In early practices with the OG map, those discretized grids with more connections to their unoccupied neighbor within an appropriate view were recognized as having more discovery rewards when reached [9]. This primitive approach is considered information theoretic driven, for it intuitively recognizes frontier pixels/voxels as those with higher information gain because a robot at the frontier can observe more unknown zones towards the proposal of area coverage. It has also motivated the topological representation in more recent approaches, which uses the conception of graph connectivity developed from pixel connectivity. Due to the increase of available computational power and advances of digital cameras, visual-based maps were given their rise but the idea of OG map still holds, which will be discussed in the later section.

Such approaches soon got their further utilization with respect to quantifying measures driven by maximizing mutual information or other metrics with regard to information theory, which then became an optimization formulation considering information gain [10]. Leveraged by frontier information and map mutually, especially by knowing the optimization objective in the form of waypoints illustrated in such a map whose exploration rewards are decided by quantification of uncertainty, people can develop path planners within a path searching framework, and take advantages of existing algorithms [11].

Besides an environmental representation, the way that enables access to the map, i.e. the sensor, is also important. Early approaches used sonars to construct a distance-based environmental representation like [9]. However, depth sensors of sonar and lidar can only obtain point cloud representation and lose texture and semantic information about the environment, where cameras can compensate for these disadvantages with adequate information acquisition. From recent approaches [12-14] the trend of using combined panoramic visual methods with an even larger visual field can be sought. With increased visual range enabled, an agent can get a vaster knowledge of the environment, this will benefit A-SLAM in multiple ways. Sensor fusion approaches also allow robots to gain much more information with different metrics, among those the most popular are fused visual and distance detectors. To generalize the idea, a proposal for maximizing information gain can be addressed.

As aforementioned, loop closure detection can contribute substantially to trimming localization and mapping error, thus loop closing and landmark revisiting can be induced actively to enhance SLAM accuracy. In contrast to the frontier detection method, active loop closing deliberately revisits known landmarks, also known as exploiting. There is a dilemma between exploration and exploitation, and this still remains an open problem in the research field of SLAM [1].

Utility and cost are calculated by the last two stages as optimization problems and iterations are performed among the three cascaded stages until some stopping criteria [15] is met. It is hoped that there is sufficient information gained from the perception stage, to avoid action policy falling into suboptimal due to perception defect, simply speaking, not able to look at things out of vision. Fig.3 shows a working principle of how these stages cascaded using a block diagram.

There are two aspects of the approach of maximizing the information gain, firstly a representation of the environment is employed as a quantitative measure to conduct the optimization task, give a metric for numerating information gain, secondly sensory methods are considered important for providing such a map, and thus maximizing the measure.

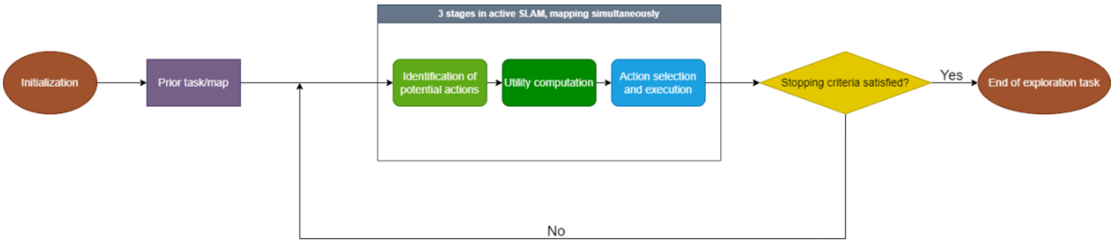


Figure 3. Working principle of decoupled stages in A-SLAM

Through these approaches a general aim for active perception can be addressed as, by all means of utilizing camera positions, different implementations, or employing novel sensory technology, to maximize information gain. The benefit of maximizing information gain is threefold: firstly, for a cascaded Bayesian inference problem it reduces the possibility of resulting in local minima; secondly, if the vision range is large enough, an agent can be more aware of its vicinity, especially useful when in landmark re-seeking; thirdly, more information gained per gaze can significantly reduce relocating the orientation of a robot and enhance positioning accuracy & efficiency thereby. In the next section by reviewing state of the art, novel implementations and technologies are demonstrated.

3. Environmental representations

As reviewed in previous motivation-stating section, a choice for representation of the environment is not solely related to which sensor is in use, however, it can also be highly determined by what tasks are assigned to an agent that performs active perception. Some application tasks like area coverage in feature-sparse scenarios need non-myopic perception and planning [16, 17], while others, for example, indoor navigation, deal with frequent short-term redirecting [13], where a weight is put more on its vicinity. The abundance of visual features can be a key factor in determining which representation to apply with, the importance of uncertainty representation is called [18, 19].

Continued with previous discussion of motivation, this review section starts with conventional OG-based approaches employed in recent work with adequate visual features, to difficult situations with low quantities of available features. Conventional approaches provide successors with their highly understandable means of reasoning for frontiers as previously stated, and their motivations always provide researchers with intuition in fusing state-of-the-art technologies like [20]. On the rise of semantic representations of SLAM [21], the compactness of the map helps a robot to gain a high-level understanding of the environment, enhancing robustness during exploration missions. [20] reports a metric-, 3D-voxel based occupancy map with an additional semantic layer, where the exploration task is conducted by incremental frontier-based searching, leveraging information gain optimization. The accuracy and robustness proposal are handled by loop closure detection using the semantic layer of the map. Fig.4 gives an illustration of the visualization of various environmental representations and the concept of a frontier.

Figure 4. Left subfigure: above [22] and middle [20] both depict metric-semantic map, below [23] depicts the belief space representation; right figure [24] visualizes frontier in a 2D map

While often subjected to a collision-free environment, applications for area coverage face challenges of inadequate visual features as well as lost track of ground truth, rendering navigation a difficult task. Grid/cell-based maps failed to measure localization and mapping uncertainty, even enhanced with semantic layer, the author of this paper argues, because the diversity of meaningful semantic objects is also of lacking as applications examples [14, 25] revealed. Authors of [25] incorporated their early work and proposed an approach of evaluating visual saliency from a relevant map. They utilized a threshold to filter rewardable waypoint candidates.

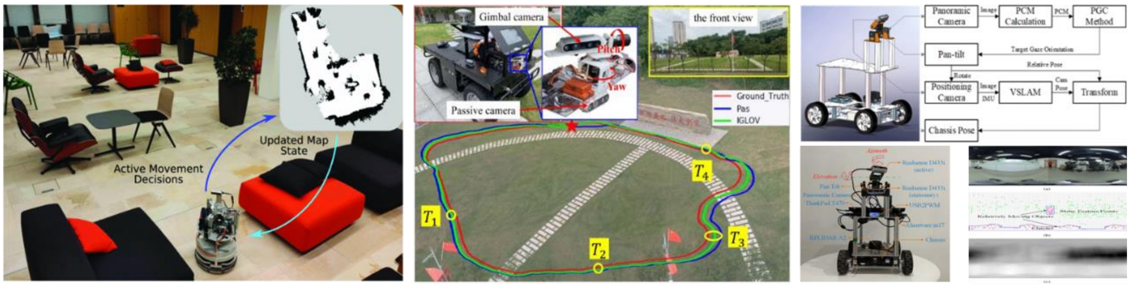
4. Sensory methods review

The aim of maximizing information gain can be directly fulfilled by increasing the field of view (FOV), using a panoramic or fisheye camera. Increased FOV can also benefit visual odometry, providing more accurate localization [26]. However, a larger view, along with post-processing for image distortion means more computational resources are required. [12] has combined a panoramic camera whose sensory data is used to generate a cost map, in order to guide the pan-tilt camera gazing at the best feature it should look at.

[14] proposed a combined camera set, with a fixed camera passively observing the vicinity of the ego and a gimbal camera gazing at the best view, inducing minimal estimation uncertainty. This approach fits the perception problem into a next-best-view problem. Compared to state-of-the-art, this work has shown its strong adaptability to featureless environments. There is also a work that developed a gaze selection system aimed at enhancing the robustness of visual odometry, by choosing high-texture quality areas as interests [27], but it does not respond to the A-SLAM problem.

literature[13] had developed a rotational camera robotics platform to add an extra degree of freedom in visual sensing. Compared to normal approaches (under the same exploration strategy), their work demonstrated better performance in terms of entropy, trajectory error, and energy consumption. [20] is a systematic approach leveraging an UAV's mobility to achieve active loop closure, based on a metric-semantic map.

Although the camera can provide adequate information, the data must undergo a reconstruction process to generate representations for frontier detection. A Lidar-based application for indoor greenhouse navigation [28], using a dynamic window approach, has shown its advance in accuracy with a GPS-denied environment. Early research [29] discovered fusion of monocular vision and laser range data.



(Bonetto et al., 2021 ECMR) (Wang et al., 2024) (Yuwen et al., 2023)

Figure 5. Recent work on sampling and visual perception employing orientation control of camera, left [13] and right [12] work are based on indoor environment, while middle [14] concerns featureless outdoor usage

A summary as Fig.5 indicates: these works are sample-based approaches towards better information gain, but some also provide novel mechanisms which put reasoning on map [12, 14, 20].

For a further glance, a survey article [30] reports the up-to-date research progress for extreme underground SLAM tasks, incorporating the perception advances using multiple types of sensors including IMU, rotary encoder, LIDAR, vision, and thermal.

5. Conclusion

In this paper, the intricacies of active perception in SLAM exploration task, particularly under the light of active SLAM, are thoroughly reviewed. The perspectives are selected as sensory advances and environmental representations, covering both the sensing and reasoning phase of a perception action. In reviewing these representations, the historical view of an occupancy grid map is included because the essence of it, especially the concept of frontier, has enlightened further effective and robust research. Also, sensory methods are carefully reviewed, with a selection on visual approaches because they provide rich information at per game.

From aforementioned approaches, it is argued that, by exploiting sensor capability, involving multi-modular sensing technology, assigning redundant degrees of freedom, efficiency, and robustness of exploration may be enhanced. Techniques can be combined but association problems resulting in precision loss may occur. However, an event camera-based active perception approach, using an up-to-date vision modality, is not proposed yet as far as the author known.

This paper concludes that, with respect to current A-SLAM framework, active perception problem at the front stage of automatic robot exploration can be partly interpreted as maximizing information gain, solved by the sensor-based approaches aforementioned. These solutions, of course, should be aligned with appropriate representations of the environment, with considerations from visual feature abundance to computational resources available.

References

[1] Cadena C, Carlone L, Carrillo H, Latif Y, Scaramuzza D, Neira J, Reid I, Leonard J J. Past, Present, and Future of Simultaneous Localization And Mapping: Towards the Robust-Perception Age. IEEE Trans. Robot. 2016, 32 (6), 1309–1332.

[2] Chen Y, Huang S, Fitch R. Active SLAM for Mobile Robots With Area Coverage and Obstacle Avoidance. IEEE/ASME Transactions on Mechatronics 2020, 25 (3), 1182–1192.

[3] Ahmed M F, Masood K, Fremont V, Fantoni I. Active SLAM: A Review on Last Decade. Sensors 2023, 23 (19), 8097.

[4] Placed J A, Strader J, Carrillo H, Atanasov N, Indelman V, Carlone L, Castellanos J A. A Survey on Active Simultaneous Localization and Mapping: State of the Art and New Frontiers. IEEE Transactions on Robotics 2023, 39 (3), 1686–1705.

[5] Bajcsy R. Active Perception. Proceedings of the IEEE 1988, 76 (8), 966–1005.

[6] Bajcsy R, Aloimonos Y, Tsotsos J K. Revisiting Active Perception. Auton Robot 2018, 42 (2), 177–196.

[7] Kim A, Eustice R M. Active Visual SLAM for Robotic Area Coverage: Theory and Experiment. The International Journal of Robotics Research 2015, 34 (4–5), 457–475.

[8] Paull L, Seto M, Leonard J J, Li H. Probabilistic Cooperative Mobile Robot Area Coverage and Its Application to Autonomous Seabed Mapping. The International Journal of Robotics Research 2018, 37 (1), 21–45.

[9] Elfes A. Using Occupancy Grids for Mobile Robot Perception and Navigation. Computer 1989, 22 (6), 46–57.

[10] Deng D, Xu Z, Zhao W, Shimada K. Frontier-Based Automatic-Differentiable Information Gain Measure for Robotic Exploration of Unknown 3D Environments. arXiv November 10, 2020.

[11] Zhao Y, Xiong Z, Zhou S, Wang J, Zhang L, Campoy P. Perception-Aware Planning for Active SLAM in Dynamic Environments. Remote Sensing 2022, 14 (11), 2584.

[12] Yuwen X, Zhang H, Yan F, Chen L. Gaze Control for Active Visual SLAM via Panoramic Cost Map. IEEE Transactions on Intelligent Vehicles 2023, 8 (2), 1813–1825.

[13] Bonetto E, Goldschmid P, Black M J, Ahmad A. Active Visual SLAM with Independently Rotating Camera. In 2021 European Conference on Mobile Robots (ECMR); 2021; pp 1–8.

[14] Wang Z, Chen H, Zhang S, Lou Y. Active View Planning for Visual SLAM in Outdoor Environments Based on Continuous Information Modeling. IEEE/ASME Transactions on Mechatronics 2024, 29 (1), 237–248.

[15] Placed J A, Castellanos J A. Enough Is Enough: Towards Autonomous Uncertainty-Driven Stopping Criteria. IFAC-PapersOnLine 2022, 55 (14), 126–132.

[16] Atanasov N, LeNy J, Daniilidis K, Pappas G J. Decentralized Active Information Acquisition: Theory and Application to Multi-Robot SLAM. In 2015 IEEE International Conference on Robotics and Automation (ICRA); 2015; pp 4775–4782.

[17] Kitanov A, Indelman V. Topological Belief Space Planning for Active SLAM with Pairwise Gaussian Potentials and Performance Guarantees. The International Journal of Robotics Research 2024, 43 (1), 69–97.

[18] Rodríguez-Arévalo M L, Neira J, Castellanos J A. On the Importance of Uncertainty Representation in Active SLAM. IEEE Transactions on Robotics 2018, 34 (3), 829–834.

[19] Nordlöf J, Hendeby G, Axehill D. Belief Space Planning Using Landmark Density Information. In 2020 IEEE 23rd International Conference on Information Fusion (FUSION); 2020; pp 1–8.

[20] Tao Y, Liu X, Spasojevic I, Agarwal S, Kumar V. 3D Active Metric-Semantic SLAM. IEEE Robot. Autom. Lett. 2024, 9 (3), 2989–2996.

[21] Chen W, Shang G, Ji A, Zhou C, Wang X, Xu C, Li Z, Hu K. An Overview on Visual SLAM: From Tradition to Semantic. Remote Sensing 2022, 14 (13), 3010.

[22] Yang Z, Liu C. TUPPer-Map: Temporal and Unified Panoptic Perception for 3D Metric-Semantic Mapping. In 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS); 2021; pp 1094–1101.

[23] Patil S, Kahn G, Laskey M, Schulman J, Goldberg K, Abbeel P. Scaling up Gaussian Belief Space Planning Through Covariance-Free Trajectory Optimization and Automatic Differentiation. In Algorithmic Foundations of Robotics XI: Selected Contributions of the Eleventh International Workshop on the Algorithmic Foundations of Robotics; Akin, H. L., Amato, N. M., Isler, V., van der Stappen, A. F., Eds.; Springer International Publishing: Cham, 2015; pp 515–533.

[24] Topiwala A, Inani P, Kathpal A. Frontier Based Exploration for Autonomous Robot; 2018.

[25] Kim A, Eustice R M. Perception-Driven Navigation: Active Visual SLAM for Robotic Area Coverage. In 2013 IEEE International Conference on Robotics and Automation; 2013; pp 3196–3203.

[26] Zhang Z, Rebecq H, Forster C, Scaramuzza D. Benefit of Large Field-of-View Cameras for Visual Odometry. In 2016 IEEE International Conference on Robotics and Automation (ICRA); 2016; pp 801–808.

[27] Manderson T, Holliday A, Dudek G. Gaze Selection for Enhanced Visual Odometry During Navigation. In 2018 15th Conference on Computer and Robot Vision (CRV); 2018; pp 110–117.

[28] Jiang S, Wang S, Yi Z, Zhang M, Lv X. Autonomous Navigation System of Greenhouse Mobile Robot Based on 3D Lidar and 2D Lidar SLAM. Front. Plant Sci. 2022, 13.

[29] Sun F, Zhou Y, Li C, Huang Y. Research on Active SLAM with Fusion of Monocular Vision and Laser Range Data. In 2010 8th World Congress on Intelligent Control and Automation; 2010; pp 6550–6554.

[30] Ebadi K, Bernreiter L, Biggie H, Catt G, Chang Y, Chatterjee A, Denniston C E, Deschênes S P, Harlow K, Khattak S, Nogueira L, Palieri M, Petráček P, Petrlík M, Reinke A, Krátký V, Zhao S, Agha-mohammadi A, Alexis K, Heckman C, Khosoussi K, Kottege N, Morrell B, Hutter M, Pauling F, Pomerleau F, Saska M, Scherer S, Siegwart R, Williams J L, Carlone L. Present and Future of SLAM in Extreme Environments: The DARPA SubT Challenge. IEEE Transactions on Robotics 2024, 40, 936–959.