Appeared at ICRA@40 2024 Sept 23-26 2024 Rotterdam, Netherlands

Perception-aware Planning for Robotics: Challenges and Opportunities

Qingxi Meng, Carlos Quintero-Peña, Zachary Kingston, Vaibhav Unhelkar, and Lydia E. Kavraki

***Abstract*— In this work, we argue that new methods are needed to generate robot motion for navigation or manipulation while effectively achieving perception goals. We support our argument by conducting experiments with a simulated robot that must accomplish a primary task, such as manipulation or navigation, while concurrently monitoring an object in the environment. Our preliminary study demonstrates that a decoupled approach fails to achieve high success in either action-focused motion generation or perception goals, motivating further developments of approaches that holistically consider both goals.**

I. Introduction

Recent progress in machine learning and computer vision has significantly enhanced robots' perception capabilities [1], opening possibilities for new robotic applications. However, designing methods that efficiently integrate perception and action objectives remains a non-trivial challenge. This is a requirement for many promising robotic applications such as collaborative robots [2], agile quadrotor flying [3] and autonomous security robots [4].

While existing mobile robotics tasks such as inspection planning [5] and surveillance [6] often require achieving visibility of landmarks, there is a gap in understanding how to integrate additional degrees of freedom (DoF) when addressing field-of-view constraints. Recent hierarchical tracking methods for manipulators using nullspace projections and impedance control [7], while related, have yet to fully address this challenge.

In this paper, we argue in favor of new methods capable of generating robot motion for navigation or manipulation while effectively accomplishing perception goals. Existing methods focus on how to plan robot motion in the presence of unseen [8] or dynamic [9] obstacles, or how to improve robot localization [10, 11]. Furthermore, methods that consider point of interest constraints simply rely on keeping the centroid of the tracked features at the center of the image plane [3, 12]. Finally, the majority of existing methods are designed for unmanned aerial vehicles [3, 9–12] and do not readily generalize to high-DoF robots such as mobile manipulators or robots with motion constraints. *We posit that approaches that holistically consider perception and motion goals are needed to achieve effective multi-task capable robots*—i.e.*, with simultaneous perception and action goals.*

We support our argument by conducting experiments of a simulated robot that must accomplish primary tasks, such as manipulation or navigation, while concurrently maintaining continuous monitoring of an object in the environment. To attain perception goals we use off-the-shelf computer vision models for object detection and tracking [13, 14], while for motion generation we use state-of-the-art motion planning algorithms [15]. Our preliminary study demonstrates that the prevalent decoupled approach falls short in achieving high success rates for both action-centric motion and perception.

This work was supported in part by NSF award #2222876 and #2326390, the Houston Methodist Hospital, and Rice University. Rice University, Department of Computer Science, Houston TX, USA. {qm15, carlosq, zak, vu2, kavraki}@rice.edu.

II. Experiments and Results

In our experiments, we employ sampling-based motion planning (SBMP) [16] for motion generation due to its efficacy in navigating environments with numerous obstacles and its algorithmic guarantee of probabilistic completeness. We use object tracking as the perception goal because it is popular in many applications such as human-robot interaction and navigation [1].

To integrate monitoring into SBMP for high-DoF robots, we introduce three methods that ensure objects remain within the robot's line of sight or camera frustum:

*1) **Path Post-Processing**:* This method involves motion planning followed by post-processing to enforce monitoring. Monitoring is not guaranteed due to joint limits and the original path's ignorance of monitoring needs.

*2) **Rejection Sampling**:* Here, SBMP rejects samples that do not satisfy the monitoring constraint, i.e., if the object is in the robot's camera frustum.

*3) **Planning with Manifold Constraints**:* Here, we used manifold-constrained SBMP [17] to satisfy monitoring constraints. We use two models: the distance between the object's center and robot's line of sight and the signed distance between the object and the camera's frustum.

We evaluate these methods along with a perception-unaware baseline using a simulated Stretch RE1 mobile manipulator [18] that must navigate to a goal configuration while continuously monitoring an object (cup, suitcase or monitor). We utilize YOLOv9 [13] and DeepSORT [14] for real-time perception performance assessment and OMPL [19] for motion planning.

Table I presents results across 1,500 planning problems (500 per object), with metrics such as planning time, path length, success rate, and constraint satisfaction rate. Perception performance is gauged by detection rate, average confidence score, tracking rate, and normalized tracking box variance.

We observe the **Unaware** method's monitoring rate is less than 50%, which is insufficient for practical applications. The **Post** method increases monitoring but is still suboptimal due to kinematic constraints. Methods enforcing monitoring during planning (**Rejection**, **Manifold-line**, **Manifold-frustum**) satisfy the constraints near-perfectly but at the expense of

Method	Motion Planning				Object Tracking			Norm. Track Box Var.	
	Plan. Time (s)	Path Len. (rad)	Cntr. Sat.	Succ. Rate	Det. Rate	Conf. Score	Track Rate	Width	Height
Unaware	**0.02**	6.53	48.95%	99.87%	42.41%	0.674	69.10%	9.45%	11.93%
Post	0.04	**5.84**	59.03%	**99.93%**	49.86%	0.669	78.68%	8.90%	11.55%
Rejection	0.66	7.58	**100.00%**	79.80%	72.11%	0.681	92.18%	8.96%	10.75%
Manifold-line	2.14	7.75	99.97%	95.60%	73.54%	0.684	**92.75%**	**8.78%**	**10.37%**
Manifold-frustum	8.17	7.64	99.99%	90.73%	**73.93%**	**0.685**	91.56%	8.89%	10.57%

Table I. Performance metrics of the evaluated methods. "Plan. Time" is the average planning time. "Path Len." is the average length of the path. "Cntr. Sat. Rate" is the percentage of the path where the monitoring constraints are satisfied. "Succ. Rate" is motion planning's success rate. "Det. Rate" is the object detection rate. "Conf. Score" is the average confidence score of detections. "Track Rate" is the percentage of the frames that the object is tracked. "Norm. Track Box Var." denotes the variances of the bounding box, normalized as percentages of the image's width and height.

longer planning times, increased path lengths, and lower success rates.

The object tracking performance of our methods reveals a perception-specific challenge, with detection rates below 75% and average confidence scores below 0.7. While the **Rejection**, **Manifold-line**, and **Manifold-frustum** methods achieve tracking rates slightly above 90%, the normalized variance in tracking box dimensions suggests that the tracking quality may not be consistently high. The results motivate the need of novel perception-aware algorithms. Furthermore, the off-the-shelf object tracking methods may need enhancement to better leverage the motion characteristic of the problem.

III. Conclusion and Future Work

Our experiments reveals the limitations of directly integrating object monitoring into motion planning through the line of sight or camera frustum. One key limitation that is present in current decoupled approaches is the lack of consideration for occlusion, which is a common issue in dynamic environments. To address these shortcomings, there is a clear need for holistic approaches to motion planning that take perception goals into account. For example, more sophisticated models for mapping perception goals to the cost function or constraints of the motion planning should be explored.

Building on these findings, our research shows that the approach of separately handling robot motion generation for navigation or manipulation—typically through motion planning, control, or policy learning—and common perception objectives like object tracking and scene understanding may not be adequate. Consequently, there is a need for novel research that integrates both perception and motion objectives. The advancement of such research initiatives is crucial for more versatile and robust robotic systems capable of handling a broader range of tasks with simultaneous perception and action goals.

References

[1] F. Chen, X. Wang, Y. Zhao, S. Lv, and X. Niu. "Visual object tracking: A survey". In: *Computer Vision and Image Understanding* 222 (2022), p. 103508.

[2] A. Borboni, K. V. V. Reddy, I. Elamvazuthi, M. S. AL-Quraishi, E. Natarajan, and S. S. Azhar Ali. "The expanding role of artificial intelligence in collaborative robots for industrial applications: a systematic review of recent works". In: *Machines* 11.1 (2023), p. 111.

[3] D. Falanga, P. Foehn, P. Lu, and D. Scaramuzza. "PAMPC: Perception-Aware Model Predictive Control for Quadrotors". In: *IEEE/RSJ Int. Conf. on Intell. Robots and Syst.* 2018.

[4] N. Hochgeschwender, G. Cornelius, and H. Voos. "Arguing security of autonomous robots". In: *IEEE/RSJ Int. Conf. on Intell. Robots and Syst.* IEEE. 2019, pp. 7791–7797.

[5] R. Almadhoun, T. Taha, L. Seneviratne, J. Dias, and G. Cai. "A survey on inspecting structures using robotic systems". In: *International Journal of Advanced Robotic Systems* 13.6 (2016), p. 1729881416663664.

[6] A. Bircher, K. Alexis, M. Burri, P. Oettershagen, S. Omari, T. Mantel, and R. Siegwart. "Structural inspection path planning via iterative viewpoint resampling with application to aerial robotics". In: *IEEE Int. Conf. Robot. Autom.* 2015, pp. 6423–6430.

[7] G. Garofalo and C. Ott. "Hierarchical Tracking Control With Arbitrary Task Dimensions: Application to Trajectory Tracking on Submanifolds". In: *IEEE Robotics and Automation Letters* 5.4 (2020), pp. 6153–6160.

[8] G. Goretkin, L. P. Kaelbling, and T. Lozano-Pérez. "Look Before You Sweep: Visibility-Aware Motion Planning". In: *Algorithmic Foundations of Robotics XIII*. Ed. by M. Morales, L. Tapia, G. Sánchez-Ante, and S. Hutchinson. Cham: Springer International Publishing, 2020, pp. 373–388.

[9] J. Tordesillas and J. P. How. "PANTHER: Perception-Aware Trajectory Planner in Dynamic Environments". In: *IEEE Access* 10 (2022), pp. 22662–22677.

[10] L. Bartolomei, L. Teixeira, and M. Chli. "Perception-aware Path Planning for UAVs using Semantic Segmentation". In: *IEEE/RSJ Int. Conf. on Intell. Robots and Syst.* 2020, pp. 5808–5815.

[11] B. Ichter, B. Landry, E. Schmerling, and M. Pavone. "Perception-Aware Motion Planning via Multiobjective Search on GPUs". In: *Robotics Research*. Ed. by N. M. Amato, G. Hager, S. Thomas, and M. Torres-Torriti. Cham: Springer International Publishing, 2020, pp. 895–912.

[12] H. Lu, Q. Zong, S. Lai, B. Tian, and L. Xie. "Real-Time Perception-Limited Motion Planning Using Sampling-Based MPC". In: *IEEE Transactions on Industrial Electronics* 69.12 (2022), pp. 13182–13191.

[13] C.-Y. Wang and H.-Y. M. Liao. "YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information". In: *arXiv preprint arXiv:2402.13616* (2024).

[14] N. Wojke and A. Bewley. "Deep Cosine Metric Learning for Person Re-identification". In: *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE. 2018, pp. 748–756.

[15] J. J. Kuffner and S. M. LaValle. "RRT-connect: An efficient approach to single-query path planning". In: *Proceedings 2000 ICRA. Millennium Conference. IEEE International Conference on Robotics and Automation. Symposia Proceedings (Cat. No.00CH37065)*. Vol. 2. 2000, 995–1001 vol.2.

[16] A. Orthey, C. Chamzas, and L. E. Kavraki. "Sampling-based motion planning: A comparative review". In: *Annual Review of Control, Robotics, and Autonomous Systems* 7 (2023).

[17] Z. Kingston, M. Moll, and L. E. Kavraki. "Sampling-based methods for motion planning with constraints". In: *Annual review of control, robotics, and autonomous systems* 1 (2018), pp. 159–185.

[18] C. C. Kemp, A. Edsinger, H. M. Clever, and B. Matulevich. "The design of stretch: A compact, lightweight mobile manipulator for indoor human environments". In: *IEEE Int. Conf. Robot. Autom.* IEEE. 2022, pp. 3150–3157.

[19] I. A. Şucan, M. Moll, and L. E. Kavraki. "The Open Motion Planning Library". In: *IEEE Robotics & Automation Magazine* 19.4 (Dec. 2012), pp. 72–82.