Teaching Robots To "Get It Right"

Joydeep Biswas
Department of Computer Science
The University of Texas at Austin
2317 Speedway, Stop D9500
Austin, Texas 78702

Abstract

We are interested in building and deploying service mobile robots to assist with arbitrary end-user tasks in everyday environments. In such open-world settings, how can we ensure that robots 1) are robust to environmental changes; 2) navigate the world in ways consistent with social and other unwritten norms; and 3) correctly complete the tasks expected of them? In this work, we survey these technical challenges and present several promising directions to address them. To "get it right", robots will have to reason about unexpected sources of failures in the real world and learn to overcome them; glean appropriate contextual information from perception to understand how to navigate in the world; and infer what correct task execution actually entails.

Motivation

We are approaching the golden era of robotics — we are finally starting to see robots in homes, hospitals, and on sidewalks. It may be tempting to declare victory, but in fact, we are still far from being able to truly rely on our robot assistants to consistently and reliably complete their tasks in the real world. A home robot may exhibit precise navigation under nominal circumstances, but moving the couch around, or even dropping a few unfortunately placed backpacks is sufficient to confuse the robot about its location in the world, and waylay its ability to navigate to the kitchen. A delivery robot may be fully capable of robustly traversing a clearly marked sidewalk on a well-lit sunny day, but ask it to deliver a package in the rain, and it may slide off the sidewalk — or perhaps even trample the customer's precious petunias in its zeal to get straight to the door. Beyond navigation, a service robot may also fail to correctly complete tasks requested of it — when asked to find a vacant conference room with a white board, it may come back with no options, or options that are not vacant, or one with no whiteboard.

There are many ways in which a service mobile robot may fail to understand how to "get it right" in various environments (Figure 1). The underlying causes for such limitations, and the open research questions that arise from them, are myriad, and in this work we survey three key aspects of the problem: 1) competence-aware autonomy; 2) context-aware navigation; and 3) learning to perform end-user tasks.

Figure 1: Teaching robots to "get it right" in environments ranging from indoor offices to offroad settings, and outdoor sidewalks.

Competence-Aware Autonomy is the ability of a robot to reason about its limitations and to adapt its behavior to overcome them. This includes reasoning about novel limitations discovered at runtime, going beyond what it may have seen or been trained with during development. It also requires the robot to reason about what actions it can take to overcome these novel limitations, and how to plan to complete tasks to minimize the impact of these limitations. To exhibit competence-aware autonomy, a robot must be able to autonomously discover and reason about limitations in deployment environments and adapt its behavior to overcome them. A competence-aware robot should thus, for example, be capable of autonomously discovering that a wet marble sidewalk leads to reflective and slippery surfaces that pose challenges to visual navigation, and adapt its behavior to pick alternate routes when possible, or slow down on the wet marble sidewalk if no other route is available.

Context-Aware Navigation is the ability of a robot to identify key relevant factors in its environment for the situation at hand, and to use this information to modify its navigation behavior. This is an essential skill for navigating in dense urban environments — if a sidewalk has a temporary hazard such as a fallen tree, a last-mile delivery robot should be able to reason about when it is safe to step onto the road to bypass it. When navigating on a university campus, the robot should step aside to let delivery vehicles pass, move with the flow of the crowd of pedestrians in accordance with social norms, and avoid driving on the grass when a paved path is available. Context-aware navigation thus requires the robot to *perceive* the relevant context, and to *reason* about how to modify its navigation behavior based on this context.

While it may be tempting to pre-specify all potential objects and entities of relevance for context-aware navigation, this is infeasible in practice since the real world will inevitably present novel objects and entities. An open research question is thus how to both perceive and reason about context in a way that generalizes to novel scenarios — new objects, entities, and environments.

Learning to Perform End-User Tasks is the ability of a robot to learn to perform tasks that are not explicitly specified during development. This may include learning from demonstrations, from natural language commands, from teleoperation, or from combinations of these. The challenges here are to learn from limited demonstrations, to be robust to noisy or partial demonstrations, to learn how to generalize the learned task to novel situations, and to ensure that the learned task meets the users' expectations.

In this paper, we review recent work in these three areas, present several promising directions to address them, and discuss open challenges and avenues for future work. Collectively, these three aspects of "getting it right" are essential for building and deploying service mobile robots to assist with arbitrary end-user tasks in everyday environments.

Competence-Aware Autonomy

Robots that are deployed in the real world over extended durations will inevitably encounter novel failure modes that were not anticipated during development. The goal of competence-aware autonomy is to enable robots to autonomously discover and reason about such failures, and to adapt their behavior to overcome them. Failures of perception may lead to erroneous state estimates (*e.g.,* obstacle detection missing a thin reflective chair), while failures of planning may lead to suboptimal or dangerous actions (*e.g.,* driving too fast on a slippery surface). In both cases, the source of uncertainty leading to failures may be due to *aleatoric* or *epistemic* uncertainty. Aleatoric uncertainty includes inherent stochasticity in the sensing and actuation, and the partial observability of the world. Epistemic uncertainty includes limitations of the computational models, algorithmic limitations and approximations, and (in the case of machine-learned models) the limitations of the training data.

While there has been significant progress in *uncertainty quantification*, most such approaches are either unable to reason about epistemic uncertainty, or fail to accurately estimate epistemic uncertainty in out-of-distribution settings. We seek to address this gap by introducing *introspective perception* [22], a novel approach to uncertainty estimation that enables robots to autonomously discover perceptual failures during deployment, accounting for both aleatoric and epistemic uncertainty.

Introspective Perception

We formulate introspective perception as a higher-order function that takes as input a perception function and learns an introspection function that predicts what parts of the sensed input to the perception function are likely to lead to perceptual failures. Introspective perception exploits *consistency* at several levels to identify failures — by leveraging multi-modal sensing consistency, spatio-temporal consistency, and algorithmic consistency naturally present in data collected by a mobile robot, it can learn an empirical model of the error distribution of perception algorithms in the deployment environment, in an autonomously supervised manner.

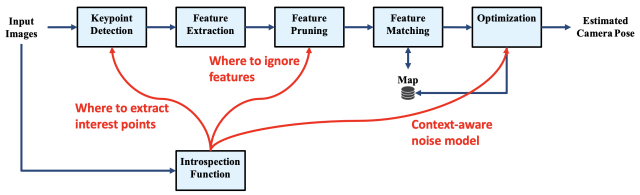

Figure 2: Introspective Vision for Simultaneous Localization and Mapping (IV-SLAM) autonomously learns where to extract interest points, which features to ignore, and context-aware noise models for competence-aware SLAM.

With Introspective Vision for Obstacle Avoidance (IVOA) [24], we leveraged occasionally available supervisory sensing to autonomously detect failures in depth estimation. Given a black-box stereo vision algorithm, IVOA is able to predict which parts of sensed images are likely to result in failures of obstacle sensing, and the types of distinct failure modes. We further investigated competence-aware planning using the results of IVOA to identify locations in the world where a robot is likely to experience perceptual failures. We introduced competence-aware path planning via introspective perception (CPIP) [21], a Bayesian framework to iteratively learn and exploit task-level competence in novel deployment environments.

We further introduced Introspective Vision for Simultaneous Localization and Mapping (IV-SLAM) [23] to autonomously learn context-aware noise models for features extracted for visual SLAM. IV-SLAM leverages spatio-temporal consistency of 3D landmarks in visual SLAM to autonomously identify when regions in captured images were likely to lead to tracking failures. Using this autonomously supervised data collection, IV-SLAM learns a context-aware noise model to predict feature re-projection errors (Figure 2). We empirically demonstrated on standard datasets and real-world data with the UT Jackal, that IV-SLAM 1) is accurately able to predict sources of tracking error, 2) reduces tracking error compared to visual SLAM, and 3) increases the mean distance between tracking failures by more than 70% compared to V-SLAM in challenging real-world settings.

Open Research Questions

While we have made significant progress in introspective perception, several open research questions remain. First, how can we learn to reason about the impact of conditions and context on the performance of perception algorithms? For example, a wet marble sidewalk may be slippery and hence dangerous to drive on, but when it is dry, it may be safe — hence competence is dependent on both the context (marble sidewalk) and the conditions (wet or dry). Second,

how can we learn to reason about the impact of perception failures on high-level task execution? We believe addressing these questions will help significantly improve the robustness and reliability of robots in real-world settings.

Context-Aware Navigation

Context-aware navigation can be broken down into its perception and planning components, where perception needs to identify entities and factors that will affect navigation, and planning needs to reason about how different actions will be affected by these entities and factors. We specifically focus on perceiving entities of relevance, terrain-aware navigation, and social navigation.

Perceiving Entities of Relevance

A common approach to context-aware navigation is to prespecify a set of entities that are relevant for navigation, and to use supervised learning to detect these entities, their poses, and track them over time. This approach has most notably been effective in the autonomous vehicles (AV) domain, where large-scale labeled training datasets such as the KITTI [7], Waymo [29], and NuScenes [3] datasets have enabled the development of highly accurate perception algorithms for detecting entities of relevance to autonomous driving. However, such datasets rely on higher fidelity sensor suites, encounter different geometric and semantic entities, and have different sensor viewpoints compared to urban robots. This causes perception models trained on AV datasets to perform poorly on robots in urban settings.

To address this gap, we contributed the UT Campus Object Dataset (CODa) [34], a large-scale annotated multi-modal dataset for training and benchmarking egocentric 3D perception for robots in urban environments. Our dataset is comprised of 23 sequences in indoor and outdoor settings on a university campus and contains repeated traversals from different viewpoints, weather conditions, and scene densities. CODa contains 1.3 million ground truth 3D bounding box annotations, instance IDs, and occlusion values for objects in the 3D point cloud. Furthermore, it includes 5000 frames of 3D terrain segmentation annotations for 3D point clouds. All annotations are provided by human annotators, and labeled at 10Hz for 3D bounding boxes, and 2-10Hz for terrain semantic segmentation. Compared to similar 3D perception datasets, CODa has far more class diversity, containing 53 object classes and 23 urban terrain types. This includes classes that are useful to urban navigation, such as doors, railings, stairs, emergency phones, and signs.

An open research question is how to go beyond supervised learning in perceiving entities of relevance — unfortunately the state of the art in unsupervised perception significantly lag behind supervised approaches. We hope that CODa will be a valuable resource to boot-strap research in unsupervised 3D perception for urban robots, and eventually enable robots to perceive entities of relevance in urban environments without requiring expensive labeled data.

Terrain-Aware Navigation

Robots deployed in outdoor environments must reason about different types of terrain for both safety (*e.g.,* prefer dirt over mud) and deployer preferences (*e.g.,* prefer dirt path over flower beds). Most existing solutions to this preference-aware path-planning problem use semantic segmentation [14] to classify terrain types from camera images, and then ascribe costs to each type [31]. Unfortunately, there are three key limitations of such approaches – they 1) require pre-enumeration of the discrete terrain types, 2) are unable to handle hybrid terrain types (*e.g.,* grassy dirt), and 3) require expensive labeled data to train semantic segmentation.

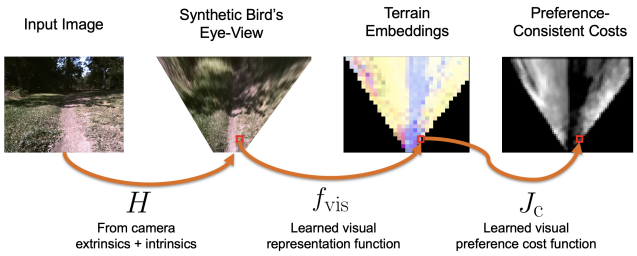

Figure 3: Leveraging learned visual representations for preference-aware path planning.

To overcome these limitations of discrete segmentation-based approaches, we have been pursuing several variations of *representation learning* to teach robots to distinguish between different terrain types without requiring explicit labels (Figure 3). In general, these approaches first convert RGB images to learned embeddings that encode terrain types, and then use these embeddings to predict costs for different terrain types. There are several variations of this approach, including learning embeddings via supervised contrastive learning [18], self-supervised representation learning [2], and autonomous domain adaptation [17].

Visual Representation Learning for Preference-Aware Path Planning (VRL-PAP) [26] leverages unlabeled human demonstrations of navigation to autonomously learn visual embeddings of terrains that distinguish terrains treated by the human demonstrator as having different preferences. VRL-PAP relies on demonstrations that show the human avoiding certain terrains and preferring others, and uses these demonstrations to learn a mapping from images to representations that distinguishes between the different terrains. In some scenarios, it may be infeasible to provide demonstrations where the human clearly avoids certain terrains — for such scenarios, we introduce Self-supervised Terrain Representation Learning (STERLING) [16], which employs a novel multi-modal self-supervision objective through non-contrastive representation learning to learn relevant terrain representations for terrain-aware navigation. To adapt to novel terrains, we introduce Preference Extrapolation for Terrain-aware Robot Navigation (PATERN) [17], which extrapolates operator preferences for visually novel terrains by identifying similarity in the inertial-proprioceptive-tactile space. The key insight of PATERN is that the operator's preferences for novel terrains can often be inferred by comparing to other terrains that though visually different, have similar inertial-proprioceptive-tactile responses when traversed.

Social Navigation

A key challenge with social navigation is that "socially acceptable" navigation behaviors are both ill-specified and vary significantly between scenarios, individuals, and cultures. A summary of the relevant challenges, scenarios, metrics, and benchmarks is included in our recent survey [6] on social navigation. Given the lack of precise specifications for socially acceptable navigation, one appealing approach is to learn social costs and navigation policies directly from human demonstrations. In support of this, we introduced Socially Compliant Navigation Dataset (SCAND) [15], a large-scale dataset of manually driven robots in challenging social settings to demonstrate how to navigate in a socially compliant manner. SCAND contains 8.7 hours, 138 trajectories, 25 miles of socially compliant, human tele-operated driving demonstrations, with logs of 3D LIDAR, joystick commands, odometry, visual, and inertial sensing collected on a Boston Dynamics Spot and a Clearpath Jackal.

Using SCAND, we recently showed that, surprisingly, even in socially challenging settings, pure geometric navigation can be sufficient for more than 80% of the time, and a hybrid approach that learns to switch between geometric and learning-based navigation on a context-specific basis can thus be effective in practice [25].

Open Research Questions

While we have made significant progress in learning to perceive entities of relevance, terrain-aware navigation, and social navigation, several open research questions still remain. First, how can we learn to identify and reliably detect entities of relevance in urban environments without requiring expensive labeled data? Second, how can we learn to account for multiple factors simultaneously, including terrain types, social norms, and other contextual factors, in a unified framework? Third, how can we teach robots how to gracefully handle unexpected or unusual combinations of circumstances, such as emergency vehicles at a construction site, or a parade on a rainy day? We expect that addressing these questions will assist in making robots more robust and capable of navigating complex urban environments.

Performing End-User Tasks

While the prevalence of robots in the real world is increasing, they are still programmed to perform a limited set of pre-defined tasks. We contend that to truly be useful in the real world, robots must be able to learn to perform tasks not explicitly specified during development. Such end-user tasks can be specified in natural language, demonstrated by a human, or inferred from context. We survey several recent approaches to learning to perform end-user tasks, including learning from limited demonstrations and natural language.

Learning From Limited Demonstration

Learning from demonstration is a promising paradigm for teaching robots to perform tasks that are not explicitly specified during development. The key challenge here is to learn from a small number of demonstrations and to generalize the learned task to novel situations. To address this challenge, we introduce several approaches that rely on *programmatic imitation learning* (PIL) to represent tasks as programs in a domain-specific language (DSL). The DSL provides a structured representation of the task that can be used to generalize to novel situations and further provides strong inductive biases that promote data-efficient learning.

Despite its merits, PIL has several limitations, including the need for computationally expensive search in program space, the inability to handle noisy demonstrations, and the necessity of providing labeled demonstrations. We introduce layered dimension-informed program synthesis (LDIPS) [10] to address the first limitation, by using physics-informed dimension constraints to restrict the search space of programs to physically meaningful expressions. LDIPS prohibits synthesizing programs that violate dimension constraints (*e.g.,* comparing a speed to a distance), which simultaneously reduces the search space and improves the generalizability of the learned programs. Iterative dimension-informed program synthesis (IDIPS) [8] further extends LDIPS by iteratively refining the programs in a lifelong learning setting, where the robot continually improves its performance from additional demonstrations over time. We introduce SAT modulo theory (SMT) -based robot transition repair (SRTR) [9] to correct learned programs during demonstrations by using an SMT solver to modify program parameters such that the result of running the program matches human-provided corrections. To learn programs in environments with an open set of objects, we introduce PROLEX [20] to learn programmatic structures from demonstration traces, and to prune the search space by reasoning about semantic relations between potential objects and actions. We have also recently demonstrated that PIL can be formulated using probabilistic programming to learn programs from noisy demonstrations and in the absence of action labels: PLUNDER [33] uses an expectation-maximization framework to simultaneously infer action labels from unlabeled motor demonstrations and synthesize a program that is consistent with such labels.

Natural Language To Task Programs

Recent advancements in large language models (LLMs) have spurred interest in using them for generating robot programs from natural language, with promising initial results [5, 30, 13, 12, 1, 19, 27, 32, 4, 28]. We investigate the use of LLMs to generate programs to perform long-horizon tasks where *accurate sequencing and ordering* of actions is crucial for success. We contribute CodeBotler and RoboEval [11] — CodeBotler is an open-source robot-agnostic tool to program service mobile robots from natural language, and RoboEval a benchmark for evaluating LLMs' capabilities of generating programs to complete service robot tasks. RoboEval evaluates the correctness of generated programs by checking execution traces starting with multiple initial states, and checking whether the traces satisfy temporal logic properties that encode correctness for each task. Our findings from RoboEval show that even the largest of LLMs struggle to generate programs that accurately perform end-user tasks. We believe RoboEval will be a valuable resource for the community to benchmark progress in gener-

ating robot programs from natural language.

Open Research Questions

While we have made significant progress in learning to perform end-user tasks, several key challenges remain. First, how can we learn to infer the implicit intent behind user tasks, beyond the explicit demonstration or natural language description? Second, there is a trade-off between expressiveness and robustness — approaches that are capable of learning a wider and richer variety of novel behaviors are often less robust to domain shifts and noisy demonstrations. How can we design algorithms that are both expressive and robust? Third, how can we provide feedback or some form of probabilistic guarantees to users about the correctness of the learned task? We believe that addressing these questions will be essential for building and deploying general-purpose service mobile robots.

Conclusion

In this paper, we have surveyed three key aspects of "getting it right" for service mobile robots: competence-aware autonomy, context-aware navigation, and learning to perform end-user tasks. We have presented several promising directions to address these challenges, and open questions and avenues for future work.

Acknowledgements

The insights, results, and conclusions presented in this paper are the result of research in the Autonomous Mobile Robotics Lab at the University of Texas at Austin. I would like to thank my students, collaborators, and funding agencies for their support and contributions to this work.

References

[1] Michael Ahn, Anthony Brohan, et al. "Do As I Can, Not As I Say: Grounding Language in Robotic Affordances". In: *arXiv:2204.01691*. 2022.

[2] Adrien Bardes, Jean Ponce, and Yann Lecun. "VICReg: Variance-Invariance-Covariance Regularization For Self-Supervised Learning". In: *ICLR 2022-International Conference on Learning Representations*. 2022.

[3] Holger Caesar, Varun Bankiti, Alex H Lang, Sourabh Vora, Venice Erin Liong, Qiang Xu, Anush Krishnan, Yu Pan, Giancarlo Baldan, and Oscar Beijbom. "nuscenes: A multimodal dataset for autonomous driving". In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2020, pp. 11621–11631.

[4] Boyuan Chen, Fei Xia, et al. "Open-vocabulary Queryable Scene Representations for Real World Planning". In: *IEEE International Conference on Robotics and Automation (ICRA), pp. 11509-11522* (2022).

[5] Yan Ding, Xiaohan Zhang, Chris Paxton, and Shiqi Zhang. "Task and Motion Planning with Large Language Models for Object Rearrangement". In: *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 2086-2092* (2023).

[6] Anthony Francis, Claudia Perez-D'Arpino, Chengshu Li, Fei Xia, Alexandre Alahi, Rachid Alami, Aniket Bera, Abhijat Biswas, Joydeep Biswas, Rohan Chandra, Hao-Tien Lewis Chiang, Michael Everett, Sehoon Ha, Justin Hart, Jonathan P. How, Haresh Karnan, Tsang-Wei Edward Lee, Luis J. Manso, Reuth Mirksy, Soeren Pirk, Phani Teja Singamaneni, Peter Stone, Ada V. Taylor, Peter Trautman, Nathan Tsoi, Marynel Vazquez, Xuesu Xiao, Peng Xu, Naoki Yokoyama, Alexander Toshev, and Roberto Martin-Martin. *Principles and Guidelines for Evaluating Social Robot Navigation Algorithms*. 2023. arXiv: 2306.16740 [cs.RO].

[7] Andreas Geiger, Philip Lenz, Christoph Stiller, and Raquel Urtasun. "Vision meets robotics: The kitti dataset". In: *The International Journal of Robotics Research* 32.11 (2013), pp. 1231–1237.

[8] Jarrett Holtz, Simon Andrews, Arjun Guha, and Joydeep Biswas. "Iterative Program Synthesis for Adaptable Social Navigation". In: *Intelligent Robots and Systems (IROS), IEEE/RSJ International Conference on*. 2021, pp. 6256–6261. DOI: 10.1109/IROS51168.2021.9636540.

[9] Jarrett Holtz, Arjun Guha, and Joydeep Biswas. "Interactive Robot Transition Repair With SMT". In: *International Joint Conference on Artificial Intelligence (IJCAI)*. 2018, pp. 4905–4911. DOI: 10.24963/ijcai.2018/681.

[10] Jarrett Holtz, Arjun Guha, and Joydeep Biswas. "Robot Action Selection Learning via Layered Dimension Informed Program Synthesis". In: *Conference on Robot Learning*. 2020, pp. 1471–1480.

[11] Zichao Hu, Francesca Lucchetti, Claire Schlesinger, Yash Saxena, Anders Freeman, Sadanand Modak, Arjun Guha, and Joydeep Biswas. "Deploying and Evaluating LLMs to Program Service Mobile Robots". In: *IEEE Robotics and Automation Letters* 9.3 (2024), pp. 2853–2860. DOI: 10.1109/LRA.2024.3360020.

[12] Wenlong Huang, Pieter Abbeel, Deepak Pathak, and Igor Mordatch. "Language Models as Zero-Shot Planners: Extracting Actionable Knowledge for Embodied Agents". In: *International Conference on Learning Representations* (2022).

[13] Wenlong Huang, Chen Wang, et al. "VoxPoser: Composable 3D Value Maps for Robotic Manipulation with Language Models". In: *Proceedings of The 7th Conference on Robot Learning, PMLR vol. 229, pp. 540-562* (2023).

[14] Peng Jiang, Philip Osteen, Maggie Wigness, and Srikanth Saripalli. "Rellis-3d dataset: Data, benchmarks and analysis". In: *2021 IEEE international*

conference on robotics and automation (ICRA). IEEE. 2021, pp. 1110–1116.

[15] Haresh Karnan, Anirudh Nair, Xuesu Xiao, Garrett Warnell, Soeren Pirk, Alexander Toshev, Justin Hart, Joydeep Biswas, and Peter Stone. "Socially CompliAnt Navigation Dataset (SCAND): A Large-Scale Dataset Of Demonstrations For Social Navigation". In: *IEEE Robotics and Automation Letters* 7.4 (2022), pp. 11807–11814. DOI: 10.1109/LRA.2022.3184025.

[16] Haresh Karnan, Elvin Yang, Daniel Farkash, Garrett Warnell, Joydeep Biswas, and Peter Stone. "STERLING: Self-Supervised Terrain Representation Learning from Unconstrained Robot Experience". In: *7th Annual Conference on Robot Learning*. 2023.

[17] Haresh Karnan, Elvin Yang, Garrett Warnell, Joydeep Biswas, and Peter Stone. "Wait, That Feels Familiar: Learning to Extrapolate Human Preferences for Preference Aligned Path Planning". In: *2024 IEEE International Conference on Robotics and Automation (ICRA)*. 2024.

[18] Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, Ce Liu, and Dilip Krishnan. "Supervised contrastive learning". In: *Advances in neural information processing systems* 33 (2020), pp. 18661–18673.

[19] Jacky Liang, Wenlong Huang, Fei Xia, Peng Xu, Karol Hausman, Brian Ichter, Pete Florence, and Andy Zeng. "Code as Policies: Language Model Programs for Embodied Control". In: *arXiv:2209.07753*. 2022.

[20] Noah Patton, Kia Rahmani, Meghana Missula, Joydeep Biswas, and Isil Dillig. "Programming-by-Demonstration for Long-Horizon Robot Tasks". In: *Proceedings of the ACM on Programming Languages* 8.POPL (2024), pp. 512–545.

[21] Sadegh Rabiee, Connor Basich, Kyle Hollins Wray, Shlomo Zilberstein, and Joydeep Biswas. "Competence-Aware Path Planning via Introspective Perception". In: *IEEE Robotics and Automation Letters* (2022), pp. 3218–3225. DOI: 10.1109/LRA.2022.3145517.

[22] Sadegh Rabiee and Joydeep Biswas. "Introspective Perception for Mobile Robots". In: *Artificial Intelligence* (2023), p. 103999. DOI: 10.1016/j.artint.2023.103999.

[23] Sadegh Rabiee and Joydeep Biswas. "IV-SLAM: Introspective Vision for Simultaneous Localization and Mapping". In: *Conference on Robot Learning*. 2020, pp. 1100–1109.

[24] Sadegh Rabiee and Joydeep Biswas. "IVOA: Introspective Vision for Obstacle Avoidance". In: *Intelligent Robots and Systems (IROS), IEEE/RSJ International Conference on*. IEEE. 2019, pp. 1230–1235. DOI: 10.1109/IROS40897.2019.8968176.

[25] Amir Hossain Raj, Zichao Hu, Haresh Karnan, Rohan Chandra, Amirreza Payandeh, Luisa Mao, Peter Stone, Joydeep Biswas, and Xuesu Xiao. "Rethinking social robot navigation: Leveraging the best of two worlds". In: *2024 IEEE International Conference on Robotics and Automation (ICRA). IEEE*. 2024.

[26] Kavan Singh Sikand, Sadegh Rabiee, Adam Uccello, Xuesu Xiao, Garrett Warnell, and Joydeep Biswas. "Visual Representation Learning for Preference-Aware Path Planning". In: *Robotics and Automation (ICRA), IEEE International Conference on*. 2022, pp. 11303–11309. DOI: 10.1109/ICRA46639.2022.9811828.

[27] Ishika Singh, Valts Blukis, et al. "ProgPrompt: Generating Situated Robot Task Plans using Large Language Models". In: *ICRA 2023*. 2023.

[28] Chan Hee Song, Jiaman Wu, et al. "LLM-Planner: Few-Shot Grounded Planning for Embodied Agents with Large Language Models". In: *Proceedings of IEEE/CVF International Conference on Computer Vision (ICCV)* (2023).

[29] Pei Sun, Henrik Kretzschmar, Xerxes Dotiwalla, Aurelien Chouard, Vijaysai Patnaik, Paul Tsui, James Guo, Yin Zhou, Yuning Chai, Benjamin Caine, et al. "Scalability in perception for autonomous driving: Waymo open dataset". In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2020, pp. 2446–2454.

[30] Huaxiaoyue Wang, Gonzalo Gonzalez-Pumariega, Yash Sharma, and Sanjiban Choudhury. "Demo2Code: From Summarizing Demonstrations to Synthesizing Code via Extended Chain-of-Thought". In: *37th Conference on Neural Information Processing Systems* (2023).

[31] Maggie Wigness, John G. Rogers, and Luis E. Navarro-Serment. "Robot Navigation from Human Demonstration: Learning Control Behaviors". In: *2018 IEEE International Conference on Robotics and Automation (ICRA)*. 2018, pp. 1150–1157. DOI: 10.1109/ICRA.2018.8462900.

[32] Jimmy Wu, Rika Antonova, et al. "TidyBot: Personalized Robot Assistance with Large Language Models". In: *Autonomous Robots, vol. 47, no. 8, pp. 1087–1102* (2023).

[33] Jimmy Xin, Linus Zheng, Jiayi Wei, Kia Rahmani, Jarrett Holtz, Isil Dillig, and Joydeep Biswas. *PLUNDER: Probabilistic Program Synthesis for Learning from Unlabeled and Noisy Demonstrations*. 2023. arXiv: 2303.01440 [cs.RO].

[34] Arthur Zhang, Chaitanya Eranki, Christina Zhang, Ji-Hwan Park, Raymond Hong, Pranav Kalyani, Lochana Kalyanaraman, Arsh Gamare, Arnav Bagad, Maria Esteva, and Joydeep Biswas. *Towards Robust Robot 3D Perception in Urban Environments: The UT Campus Object Dataset*. 2023. arXiv: 2309.13549 [cs.RO].