Empowering Mobile Manipulators with Internet-of-Things Sensing for Real-world Unstructured Environments

Sørensen, Sune Lundø; Kjærgaard, Mikkel Baun

Publication date:
2022

Document version:
Final published version

Citation for pulished version (APA):
Sørensen, S. L., & Kjærgaard, M. B. (2022). *Empowering Mobile Manipulators with Internet-of-Things Sensing for Real-world Unstructured Environments*. 1-5. Paper presented at ROS 2022 Workshop on Mobile Manipulation and Embodied Intelligence (MOMA), Kyoto, Japan.

Empowering Mobile Manipulators with Internet-of-Things Sensing for Real-world Unstructured Environments

Sune Lundø Sørensen[1†] and Mikkel Baun Kjærgaard[1]

Abstract— **Maintaining a world model is important for planning mobile manipulation tasks. The process of maintaining the correspondence between objects in the world model (symbols) and perceptual data relating to the same physical objects is called *perceptual anchoring*. Many approaches perform anchoring using robot-mounted sensors, while others instrument the environment with stationary sensors to exploit the mobility of the robot and the high visual coverage of the stationary sensors. However, many of the latter works use custom sensors or make assumptions which are not applicable in real-world scenarios. We propose that the way towards world modeling in real-world scenarios is by using off-the-shelf commercially available sensors. In this paper, we compare this approach to existing world modeling works and discuss the challenges that lie ahead.**

I. INTRODUCTION

For planning its actions an autonomous mobile manipulator needs a world model. The model can have various types of information, depending on the robot's tasks. For a warehouse robot, this model could be as simple as a 2D occupancy grid [1], but a robot performing complicated manipulation tasks in a dynamic and unstructured environment may need a more complex model, including rich symbols with object semantics and information about persons blocking its path. The process of maintaining the correspondence between the symbols in the world model and perceptions of the physical objects is called *perceptual anchoring* [2]. Many works perform anchoring using only data from the sensors mounted on a robot [3], [4], [5]. This presents a challenge when the robot is deployed in a large and dynamic environment. Due to its limited field of view, it can only observe a part of the environment and therefore only update a part of its world model at a time. Ideally, it should have a complete and recent view of the world for planning actions most efficiently and avoid re-planning. Other works approach anchoring by instrumenting the environment with stationary sensors which observe the environment, even when the robot is not present [6], [7]. To illustrate how the two approaches compare, consider a robot in a canteen at a large university. Many people eat their lunch here at different times between 11.00 and 14.00. To ensure that there are always some clean tables to sit at, the robot is tasked with clearing and cleaning tables, when people have finished eating. Using a single robot, it must systematically survey the canteen for tables to clean, while avoiding people. By instrumenting the canteen with stationary sensors, all tables can be monitored simultaneously and the observations can be combined in a complete and updated world model. The robot can then go to a table as soon as people have left it, without having observed it in advance itself. Similar challenges are present in many other settings, such as airport waiting areas, restaurants and in other robot applications. These tasks could also be solved with multi-robot systems, but adding more robots to a system is usually more expensive than adding stationary sensors.

Existing solutions using stationary sensors depend on (sometimes expensive), custom-built sensors, which are limiting them to laboratory setups. We propose that by using commercially available, low-cost IoT sensors, mobile manipulators can be deployed in many different real-world environments. However, these sensors are not without limitations. To keep them low-cost, they are often 2D or 3D with a lower resolution than the often more expensive robot-mounted sensors. IoT cameras are often mounted high up on walls, so they cover a greater area. In the example above, this would mean that they are further away from the objects to be cleared from the tables, than the robot-mounted sensors, which could be driven relatively close. Consequently, the robot-sensors would be able to acquire much more detail than an IoT sensor with the same resolution. In Figure 1 this is illustrated for different 3D cameras observing an object of 10-by-10 cm at different distances. Another limitation is that not all sensors allow access to raw observations. For privacy reasons some sensors only provide the result of processing their observations, which may not contain all the information required by a world model. Lastly, to keep the sensors low-cost and energy efficient their processing power is often limited. Consequently, the rate at which they provide data may not very high.

In this paper we will compare the use of low-cost IoT sensors in combination with a robot, to existing solutions and highlight the challenges that need to be addressed, including representation, data association and fusion, maintaining a world model over time and privacy. Figure 2 illustrates the context of and relations between the challenges. The remainder of this paper is structured as follows: Section II will present existing single-robot and environment-instrumenting approaches. In Section III we will discuss the use of low-cost IoT devices for mobile manipulating in real-world environments, including the

*This work was supported by the Innovation Fund Denmark

[1]The authors are with the University of Southern Denmark, Faculty of Engineering, Mærsk Mc-kinney Møller institute, Software engineering section, Campusvej 55, 5230 Odense M, Denmark. {slso,mbkj}@mmmi.sdu.dk

† Corresponding author.

challenges and benefits of these approaches. Section IV will describe the FacilityCobot project and how we work with some of the challenges. Finally, we will conclude in Section V.

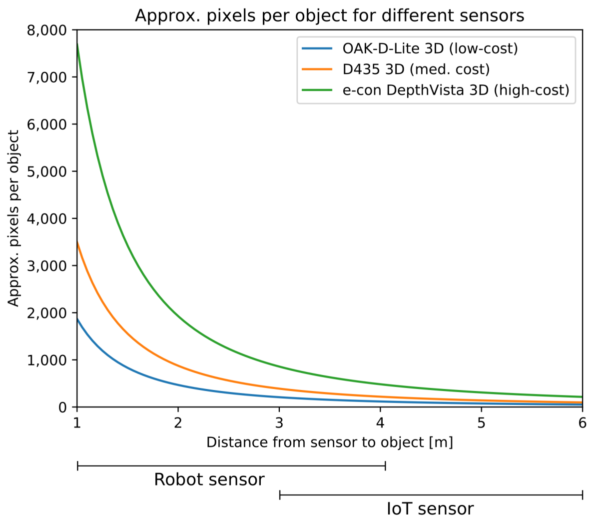

Fig. 1. Pixels per object for different sensors observing a table-top object of size 10-by-10 cm at different distances.

II. BACKGROUND

In this section, we will present existing single-robot and instrumented-environment approaches. All of the works focus on modeling objects and their various properties, which are most often the interest for manipulation. Consequently, many of the works also address the anchoring problem [2] as mentioned in Section I.

A. Single-robot approaches

There exists many works on world modeling with a single robot. Here we will present a few. Günther et al. [3] perform modeling based on point clouds from an RGB-D sensor. They exploit the context of objects to improve classification results using a conditional random field (CRF) and a probabilistic graphical model. They perform anchoring by computing similarities between anchors and objects using a support vector machine and perform data association with the Hungarian algorithm [8]. Ruiz-Sarmiento et al. [4] also employ a CRF to exploit not only the relation between objects but also between objects and rooms. Instead of keeping only the most probable results of inference on the CRF, they store multiple hypotheses on the state of the world. They perform anchoring by comparing the overlap of the bounding boxes of anchors and percepts. Elfring et al. [5] also maintains multiple hypotheses on the world state. They use motion models to predict future object positions using a Kalman filter, but their framework allows for other kinds of prior knowledge to be used for predicting future objects states. Hiller et al. [9] models objects by their pose, dynamics (velocity), shape and class. They also provide methods for producing standard representations, such as occupancy grids [1], from the world model.

B. Environment-sensor approaches

One of the main limitations of performing world modeling with a single robot is the fact that it can only observe a limited part of its environment at a time. This is one of the motivations for using environmental sensors in combination with robots [10]. Furthermore, many homes, public and work spaces are already equipped with smart sensors, making deployment more cost-effective.
LeBlanc and Saffiotti [7] present a general anchoring framework for cooperative anchoring using heterogeneous information. A local anchor space is defined for each robot and stationary sensor, along with functions for mapping them into the global anchor space. They implement the framework using fuzzy logic and the PEIS framework [11]. Daoutis et al. [6] also present a framework for cooperative anchoring. Similar to LeBlanc and Saffiotti [7] they use local anchor spaces for each agent, which are then combined in a global anchor space. Furthermore, they connect the anchors to semantic knowledge for reasoning and human-robot interaction. They also implement and evaluate their approach in the PEIS framework [11].

Table I summarizes the advantages and limitations of the two approaches presented here and using commercially available IoT sensors together with a robot.

III. CHALLENGES

The world modeling approaches in Section II-B which use environmental cameras together with mobile robots, all have one thing in common: They use specialized sensors, which are not commercially available or are not designed to work under the conditions of real-world applications. We propose that the way to perform world modeling for mobile manipulation in real-world environments is to use commercially available "off-the-shelf" IoT sensors. Due to the wide variety of these, and the limitations described in Section I, several challenges need to be addressed. In this section we will discuss some of the most important.

A. Representing sensing data from multiple sources

Different environments and tasks require different sensors. A world modeling approach should be able to accommodate these with a minimal compromise in accuracy and reliability. This requires the ability to represent diverse data, due to the heterogeneity of the IoT and robot sensors and the requirements of the application. A suitable candidate for this challenge is *Conceptual spaces*, a framework developed by Gardenförs [12]. Simply put, it is a metric space, where each axis represents a property such as x-coordinate, category, color, etc. This allows efficient computations of e.g., similarity, also of heterogeneous data. It has already been shown to work well in many anchoring systems [5], [7], [13] to represent objects, their properties and related

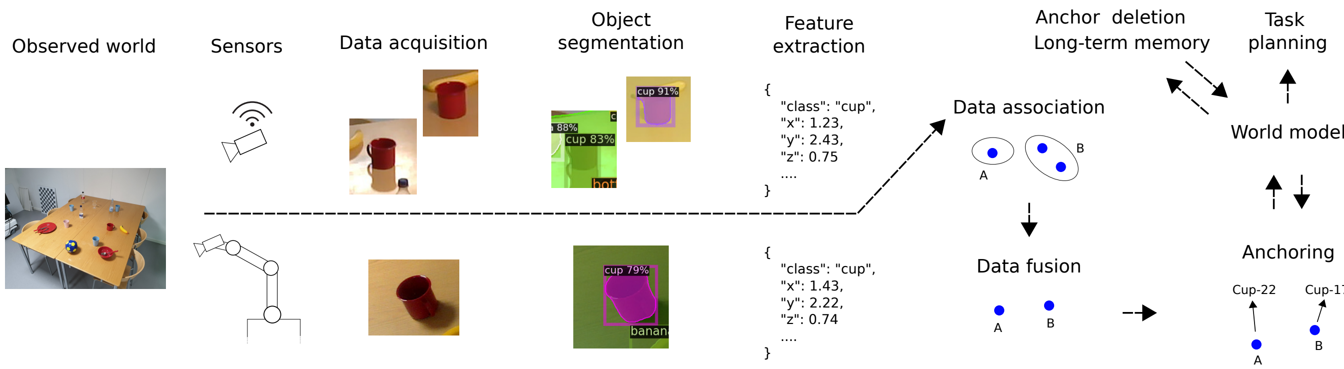

Fig. 2. Conceptual schematic of a world modeling pipeline using robot and IoT sensing data.

TABLE I

COMPARING THE ADVANTAGES AND LIMITATOINS OF THE DIFFERENT STRATEGIES.

Approach	**Advantages**	**Limitations**
Self-contained robot	No need to install other sensors No extrinsic calibration between systems	Limited immediate coverage Scales poorly to large environments
Robot and custom sensors	Greater combined immediate coverage Raw data contains high amount of information Customizable interface Scales well to larger environments	High cost Not commercially available Potentially not suitable for real-world conditions Privacy may not be considered Need extrinsic calibration
Robot and IoT sensors	Commercially available sensors Sensors are often lower cost Many buildings already have sensors installed Closed system (no raw data) is more privacy-friendly Scales well to larger areas	Need extrinsic calibration Output may have limited details

uncertainties. Beside representing objects and their properties is the challenge of representing conceptual knowledge, such as ”cups can contain liquids”. One way to represent this kind of information is to use ontologies as done by Melchert et al. [14]. Finally, the relations between objects also contain valuable information, both in the conceptual relations (e.g. between a plate and a fork) and the spatial relations of nearby objects.

B. Combining sensing data from multiple sources

The advantage of using stationary sensors and robots together depend on the ability of the system to correctly combine the data from both sides. This requires proper data association and fusion. Much work on data association has been done in multi-target multi-sensor tracking, using spatial data, classification results and visual appearance [15]. In the best case scenario, one would be able to obtain the transformation between the IoT sensors and the robot through calibration. But in cases where this is not possible (if for example a system contains a sensor which does not support detection of a calibration board), the problem of data association becomes more difficult. The problem is somewhat similar to registration of point clouds, but with more complex data points.

Data association and fusion algorithms should also incorporate the reliability and accuracy of the different sensors. For example, IoT cameras are often mounted further away from objects of interest and therefore might not provide accurate pose estimates, while the robot could be driven close to the objects and therefore provide more accurate data. In this example the association and fusion algorithms should put more weight on the robot than the IoT data. A challenge is that commercially available sensors might not provide e.g., the covariance for a position or the confidence of a detection result. In this case one approach could be to learn the sensor's accuracy over time.

Lastly, the challenge of heterogeneous data should also be addressed. Many different IoT and robot sensors exist, and they do not all provide the same data. Data association and fusion algorithms should be able to handle heterogeneous data points. Some work has already been done by LeBlanc and Saffiotti [7] using fuzzy logic. However, their approach is not quantitatively evaluated.

C. Managing perceptual data over time

A very important aspect of world modeling is to update its contents according to changes in the physical environment. As mentioned in Section I, perceptual anchoring is the process of maintaining the correspondence between symbols and percepts relating to the same physical object [2]. In other words, it is the process of maintaining the objects in the world model over time. Many approaches to anchoring have been proposed [2], [3], [4], [5], [6], [7], [13], but at least two important challenges have potential for improved solutions: Deletion of anchors and long-term memory. Just as it is important for manipulation planning to know which objects are present in the environment, it is also important to know when an object has disappeared. This involves determining if an object is occluded, not detected to due e.g., lighting, moved to another location, changed appearance or has been removed. Loutfi et al. [16] employ a simple strategy for determining when an object is no longer present and its anchor should be deleted, but note that a more sophisticated approach is needed. Along with Chella et al. [13] they also mention the possibility of a long-term memory to store anchors which are not currently active, but may be useful in the future. Both aspects of the anchor life cycle are very important.

D. Privacy

For any environment where humans are present, privacy is important to consider. For IoT devices and robots a part of the solution could be to only provide the processed data and never the raw data, which might contain privacy-sensitive information. However, while the data from the individual devices may not pose a privacy issue, the world modeling approach should be aware of risks that appear when the data is combined. Additionally, some IoT devices process data in their own cloud services, complicating the data handling. There exists tools such as the privacy risk assessment tool-chain by Schwee et al. [17], which can be used to identify privacy issues to address, before the world model is deployed.

IV. Addressing the Challenges in the FacilityCobot Project

The challenges in mentioned in Section III are all addressed in existing works, some closing more gaps than others. In the FacilityCobot project[1] we work with them in the context of a facility management application. In particular, we address challenges III-B and III-C, on combining data and maintaining the world model over time. In this section we will outline the project and how we address the two challenges.

FacilityCobot is a cross-disciplinary project combining mobile manipulation, computer vision and software engineering in an effort to automate facility management tasks, such as clearing and cleaning in canteens, airport waiting areas and many other domains. The hardware used in this project consists of several 2D IoT cameras and a mobile manipulator. The IoT cameras are able to recognize humans, tables and other objects and are mounted on the ceiling, so they cover the operation area. The mobile manipulator is equipped with an RGB-D camera, a gripper and a cleaning tool, and software for navigation, motion planning and computer vision algorithms. To provide a foundation for planning efficient execution of the robot's tasks, we work with combining data from both the robot's sensors and the IoT cameras into a world model and maintain said model over time.

As an initial effort to address the challenge of combining data from the IoT sensors and the robot, we are currently working on quantifying the performance of clustering methods and a global nearest neighbor method for data association of IoT and robot sensing data. The next step for this challenge is to integrate more information into the data association process. Firstly, the uncertainties of the different sensors and object properties (e.g., position covariance classification confidence, etc.) and secondly the field of view of the robot for filtering which concurrent IoT percepts should be considered for association as in [18]. Lastly, we would like to integrate domain-specific knowledge for data association.

In our current work, we evaluate the data association methods on a data set with static scenes. To address the challenge of maintaining a world model over time, we will extend the evaluation to dynamic scenes. To achieve this we will first collect a data set of IoT and robot observations of dynamic scenes. The data set should include scenes where new objects are added, objects are removed, occluded and moved. This ensures that the data set also can be used for evaluating methods for determining when objects has been removed, as mentioned in Section III-C. To design a method for this problem, we will leverage existing work on occlusions in perceptual anchoring [19]. The data publication frequency of the IoT sensors should also be considered in this method, since it is lower for these than the robot, as mentioned in Section I. Finally, we will also explore designs of a long-term anchor memory (LTAM) for storing anchors that are not currently active, but may become useful in the future. The design of the LTAM should among other things consider the semantic data of an object. E.g., a used napkin is unlikely to return to the environment, but the cart with dirty dishes is brought back empty every morning.

V. Conclusions

In this paper we have highlighted the challenges of performing world modeling, in particular perceptual anchoring, for mobile manipulation tasks using robot-mounted and low-cost IoT sensors and described how some of them are addressed in systems of custom IoT sensors and robots. We have also briefly presented the FacilityCobot project, which is the context for our current and future work on particularly the challenges of combining data from IoT and robot sensors and maintaining a world model over time.

References

[1] H. Moravec and A. Elfes, "High resolution maps from wide angle sonar," in *Proceedings. 1985 IEEE International Conference on Robotics and Automation*, vol. 2. St. Louis, MO, USA: Institute of Electrical and Electronics Engineers, 1985, pp. 116–121. [Online]. Available: http://ieeexplore.ieee.org/document/1087316/

[2] S. Coradeschi and A. Saffiotti, "An introduction to the anchoring problem," *Robotics and Autonomous Systems*, vol. 43, no. 2-3, pp. 85–96, May 2003. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0921889003000216

[3] M. Günther, J. Ruiz-Sarmiento, C. Galindo, J. González-Jiménez, and J. Hertzberg, "Context-aware 3D object anchoring for mobile robots," *Robotics and Autonomous Systems*, vol. 110, pp. 12–32, Dec. 2018. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0921889017307856

[1] `https://www.sdu.dk/en/facilitycobot`

[4] J.-R. Ruiz-Sarmiento, C. Galindo, and J. Gonzalez-Jimenez, "Building Multiversal Semantic Maps for Mobile Robot Operation," *Knowledge-Based Systems*, vol. 119, pp. 257–272, Mar. 2017. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0950705116305184

[5] J. Elfring, S. van den Dries, M. van de Molengraft, and M. Steinbuch, "Semantic world modeling using probabilistic multiple hypothesis anchoring," *Robotics and Autonomous Systems*, vol. 61, no. 2, pp. 95–105, Feb. 2013. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0921889012002163

[6] M. Daoutis, S. Coradeschi, and A. Loutfi, "Cooperative Knowledge Based Perceptual Anchoring," *International Journal on Artificial Intelligence Tools*, vol. 21, no. 03, p. 43, June 2012. [Online]. Available: https://www.worldscientific.com/doi/abs/10.1142/S0218213012500121

[7] K. LeBlanc and A. Saffiotti, "Cooperative anchoring in heterogeneous multi-robot systems," in *2008 IEEE International Conference on Robotics and Automation*. Pasadena, CA, USA: IEEE, May 2008, pp. 3308–3314. [Online]. Available: http://ieeexplore.ieee.org/document/4543715/

[8] H. W. Kuhn, "The hungarian method for the assignment problem," *Naval Research Logistics Quarterly*, vol. 2, no. 1-2, pp. 83–97, 1955. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/nav.3800020109

[9] M. Hiller, F. Particke, C. Hofmann, H. Bey, and J. Thielecke, "World Modeling for Mobile Platforms Using a Contextual Object-Based Representation of the Environment," in *2018 IEEE 8th Annual International Conference on CYBER Technology in Automation, Control, and Intelligent Systems (CYBER)*. Tianjin, China: IEEE, July 2018, pp. 187–191. [Online]. Available: https://ieeexplore.ieee.org/document/8688197/

[10] H. Nakashima, H. Aghajan, and J. C. Augusto, Eds., *Handbook of Ambient Intelligence and Smart Environments*. Boston, MA: Springer US, 2010. [Online]. Available: http://link.springer.com/10.1007/978-0-387-93808-0

[11] A. Saffiotti and M. Broxvall, "PEIS ecologies: ambient intelligence meets autonomous robotics," in *Proceedings of the 2005 joint conference on Smart objects and ambient intelligence innovative context-aware services: usages and technologies - sOc-EUSAI '05*. Grenoble, France: ACM Press, 2005, p. 277. [Online]. Available: http://portal.acm.org/citation.cfm?doid=1107548.1107615

[12] P. Gärdenfors, *Conceptual spaces: the geometry of thought*. Cambridge London: MIT Press, 2004.

[13] A. Chella, S. Coradeschi, M. Frixione, and A. Saf, "Perceptual Anchoring via Conceptual Spaces," p. 6, 2004.

[14] J. Melchert, S. Coradeschi, and A. Loutfi, "Knowledge Representation and Reasoning for Perceptual Anchoring," in *19th IEEE International Conference on Tools with Artificial Intelligence(ICTAI 2007)*. Patras, Greece: IEEE, Oct. 2007, pp. 129–136. [Online]. Available: http://ieeexplore.ieee.org/document/4410274/

[15] K. Yoon, Y.-m. Song, and M. Jeon, "Multiple hypothesis tracking algorithm for multi-target multi-camera tracking with disjoint views," *IET Image Processing*, vol. 12, no. 7, pp. 1175–1184, 2018. [Online]. Available: https://ietresearch.onlinelibrary.wiley.com/doi/abs/10.1049/iet-ipr.2017.1244

[16] A. Loutfi, S. Coradeschi, and A. Saffiotti, "Maintaining Coherent Perceptual Information Using Anchoring," p. 6, 2005.

[17] J. H. Schwee, F. C. Sangogboye, F. D. Salim, and M. B. Kjærgaard, "Tool-chain for supporting privacy risk assessments," in *BuildSys '20: The 7th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation, Virtual Event, Japan, November 18-20, 2020*. ACM, 2020, pp. 140–149. [Online]. Available: https://doi.org/10.1145/3408308.3427605

[18] L. L. Wong, L. P. Kaelbling, and T. Lozano-Pérez, "Data association for semantic world modeling from partial views," *The International Journal of Robotics Research*, vol. 34, no. 7, pp. 1064–1082, June 2015. [Online]. Available: http://journals.sagepub.com/doi/10.1177/0278364914559754

[19] P. Zuidberg Dos Martires, N. Kumar, A. Persson, A. Loutfi, and L. De Raedt, "Symbolic Learning and Reasoning With Noisy Data for Probabilistic Anchoring," *Frontiers in Robotics and AI*, vol. 7, p. 100, July 2020. [Online]. Available: https://www.frontiersin.org/article/10.3389/frobt.2020.00100/full