Towards an Artificial Perception Framework for Autonomous Robots in Logistics

Christopher Mayershofer and Johannes Fottner

Technical University of Munich (TUM), Garching, Germany,
{christopher.mayershofer, j.fottner}@tum.de *

Abstract. Autonomous robots in logistics are a promising approach towards a fully automated material flow. In order to use their full potential however, they must be able to extract semantic information from logistics environments. In contrast to other application areas of autonomous robots (e.g. autonomous driving, service robotics) the logistics domain lacks a common dataset and benchmark suite covering multiple sensor modalities and perception tasks. This paper conceptualizes a framework for artificial perception research in logistics that aims to close this gap in a sustainable, data-driven way. Our framework consists of three components: (1) A foundation, based on logistics-specific standards, concepts and requirements. (2) An open dataset, covering multiple sensor modalities and perception tasks and (3) a standardized benchmark suite. As shown in other research areas, a common and open platform for data-driven research facilitates novel developments and makes results comparable and traceable over time.

Keywords: Autonomous systems, Artificial perception, Logistics

1 Introduction

Increasing demand, diversification of product variants and the lack of skilled workers in industrialized nations are reasons for a cross-sectoral change in logistics [1]. New technologies and innovations not only address these challenges, but can also generate a sustainable competitive advantage. Especially development in the field of robotics is increasingly applied in logistics and is thus considered a key technology towards a fully automated and flexible material flow [2]. These robots are gradually transforming from rigid automatons to intelligent machines. When robots found their way into production facilities of automotive manufacturers in the early 1960s, they were only capable of performing repetitive tasks by previously programming a fixed sequence of process steps. However, dynamic and less structured applications, as often found in logistics, require more sophisticated approaches in order to safely, robustly and autonomously execute complex process chains [1]. Hence, a shift towards partly intelligent machines is observable that mimic the human behavior with respect to adaptivity, flexibility and robustness. Perception is one enabling technology for these systems and

* This work was supported by BMW Group.

describes the process of semantically understanding the outside world similar to how we humans do. In contrast to other application areas of autonomous systems (e.g. autonomous driving, service robotics) the logistics domain lacks a common dataset and benchmark suite covering multiple sensor modalities and perception tasks. W e therefore propose a holistic, open, data-driven framework to support artificial perception research for autonomous robots in logistics. This paper describes our framework concept aiming to enable future research in the area of artificial perception within the application field of logistics. It focuses on delivering a solid foundation based on standards and concepts, multi-modal and multi-task datasets as well as standardized benchmarks.

2 Motivation

2.1 Robotics in Logistics

Logistics is a multi-trillion a year worldwide industry. As mentioned above, increasing demand, diversification of product variants and the lack of skilled workers are examples for cross-sectoral challenges which future logistics attempts to solve [1]. In e-commerce, for example, a worldwide increase in turnover of almost 20 percent was forecast for 2020.[1] In the manufacturing industry, on the other hand, the steady increase in product variants is also a contributing factor that poses challenges for logistics operations. A vivid example of this is the German automotive industry, where a model increase of almost 50% from 2007 to 2019 was predicted.[2] This contrasts with the shortage of skilled workers in logistics. In addition to demographic change (especially in industrialized countries), this is further supported by forecasts of an increase in the level of education [3]. In its labor market forecast, the German federal government reports an expected decline in the total number of employees by 1.4 million in 2030 compared to 2010.[3]

These challenges are countered by new technologies and innovations that not only solve them, but also generate a competitive advantage. Especially developments in the field of robotics are increasingly applied in logistics and are regarded a key technology for achieving the vision of a fully automated material flow. These automated systems stand out among other things due to scalability [1], independence from the labour market [4] and cost-efficiency [1]. However, today's systems are still far from the vision of flexible, fully automated material flow. The reasons for this include the lack of robustness and flexibility of the systems in the complex industrial environment as well as the high variance of logistics processes [2]. In practice, these problems are currently being countered with humans in key positions who have a general, non-task-specific flexibility [2].

Meanwhile, the logistics industry is perceiving mentioned problems as novel requirements for logistics systems of the future. In particular, there is a need

[1] https://www.statista.com/statistics/379046/worldwide-retail-e-commerce-sales/
[2] https://www.pwc-wissen.de/pwc/de/shop/publikationen/How+to+stay+No+1!/?card=13052
[3] https://www.bmas.de/SharedDocs/Downloads/DE/PDF-Publikationen/a756-arbeitsmarktprognose-2030.html

to automate complex, difficult-to-automate processes in a sustainable manner using flexible, adaptive, cost-efficient and yet robust systems.

2.2 Towards Intelligent Machines

In the future, these systems are expected to achieve more and more human-like capabilities, including flexibility, adaptivity and rationality, through techniques of artificial intelligence and advanced robotics. Broadly speaking, the field of artificial intelligence is concerned with understanding, formulating, testing and ultimately artificially replicating the foundations of human intelligence. The term Intelligent Agent (IA) accumulates the ideas of a human-like, artificial intelligence into an abstract capsule, and hence formulates the ultimate goal of intelligent robotics. Similar to humans, an IA perceives the environment, draws conclusions and makes decisions before finally acting in order to optimally achieve a certain goal [5]. The underlying infrastructure for information acquisition, processing, reasoning and (re)action is called cognitive architecture [6]. In order to be able to reason about what action to take next, information about the outside world needs to be acquired first. Therefore, the ability to observe and understand the environment is a key technology for IAs [5].

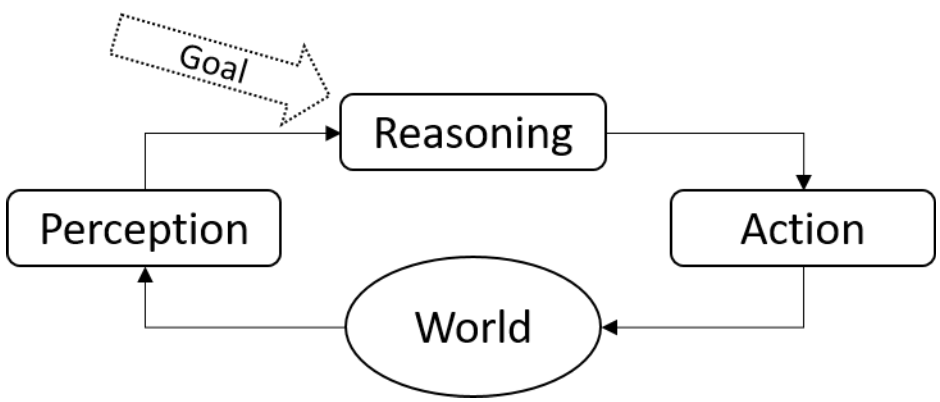

Fig. 1. Model of an intelligent agent from the perspective of *Grounded Cognition* [7]

Based on the same human capability to autonomously acquire information from the environment, this process is referred to as *perception*. Psychology describes human perception as a means of forming an image of the physical outside world which is appropriate for our actions. This image is built upon our everyday intuitions and is acquired by our senses [8]. Perception thus enables a semantic understanding of the outside world and arranges information gained into an internal representation. Next, we briefly summarize state-of-the-art perception research related to this work.

3 Related Work

3.1 Perception for Autonomous Driving

Especially in the field of autonomous driving the necessity of artificial perception is confirmed in literature [9]. Several different sensor modalities are applied here. Besides low to high range radars, cameras, sonars as well as costly multi-line lidars are used. From the acquired multi-modal raw data semantic information is extracted next. Starting with classical object detection (e.g. vehicles and people [10]) over semantic segmentation (e.g. path or free-space[4]) up to the analysis of temporal relationships (e.g. tracking [11]) various methods are applied. The extracted information is subsequently used to derive situation-specific decisions [9]. These often data-driven methods are developed using publicly available datasets and evaluated against their benchmarks. Geiger et al. presented the KITTI Vision Benchmark Suite [12] in 2012, which has been continuously expanded over the years [13, 14]. In addition, further datasets and benchmarks were published covering novel challenges, such as [15, 16].

3.2 Perception for Service Robotics

Service robots are to provide services for humans in the future, such as care taking or the preparation of meals [17]. Correspondingly, service robots also need information from the environment in order to carry out autonomous actions to achieve their goal. Here, too, various exteroceptive sensor modalities are installed, including cameras, microphones and single-line laser scanners. Again, application-specific information is extracted from raw data using methods for object detection (e.g. kitchen utensils [18] or household items [19]) and recognition of object properties and states (e.g. cracked egg [20]). Again, data sets and benchmarks are available to track progress of the research community and make results comparable. Similar to KITTI in the field of autonomous driving, OpenLORIS comprises two tasks and various sensor modalities. In contrast to the comparatively limited object classes to be expected in the field of autonomous driving, service robots have to understand and handle a much higher object varience. Correspondingly, data sets were presented that cover a large number of the objects to be recognized, for example [18, 19].

3.3 Perception for Industrial and Logistics Robotics

Industry is also striving for robots that perceive the environment in a human-like manner. Again, different sensor modalities are used to acquire raw data from the environment. Among others, (depth) cameras, lidar and tactile bumpers as well as sonars are used. Again, methods are used to locate and classify objects (e.g. people [21], load carriers [22]) or to predict information about an object (e.g. 6DoF gripping point [23], pallet pickup pose [24]). Building on this high-level information, robot systems are supposed to make smart decisions and act human-alike. In contrast to the application areas mentioned above, only a few,

[4] https://developer.nvidia.com/drive/drive-perception

very specialized data sets are available in the logistics sector. So far, no dataset and benchmark suite is known in the literature. Reference [25] provides a lidar dataset for pallet recognition. Reference [23] provides data(on request) on 71 supermarket objects. Given the absence of necessary datasets, [22,26,27] describe the evaluation on custom recorded data. Thus a valid comparison is not possible.

4 Problem Formulation and Objective

We aim to enable high level robotic perception within logistics environments to overcome the above detailed challenges. To do so, we first formulate the key challenges, before introducing our framework concept.

Artificial perception is widely researched in the context of autonomous systems. These methods are applied in different types of systems such as autonomous vehicles, as well as service and logistics robots. In particular within the field of logistics, however, only isolated functions (e.g. object detection) on individual sensor modalities (e.g. camera, lidar) have been investigated so far. To the best of our knowledge, a multi-modal, multi-task perception system that mimics human perception has not yet been proposed.

Such a multi-modal, multi-task perception must be learned by the system from experience. While "hard-coded" perceptual skills are feasible for standardized and restricted scenarios under given developmental efforts, it is to be assumed that generic perception has to be (primarily) learned from experience. These experiences must be made available to the system in the form of data. In contrast to other fields of application, multi-modal, multi-task datasets dealing with object recognition and scene understanding in logistics are not available.

Finally, research results in the context of artificial perception in logistics must be comparable and make progress traceable. In order to achieve the goal of artificial perception, standardized and thus comparable tests are required. These tests enable the systematic analysis of research results and allow conclusions to be drawn about research progress over time. Although methods developed in logistics research have been evaluated, they rarely allow direct comparison due to different test setups (e.g. test design, sensor technology, environmental conditions).

5 Artificial Perception Framework for Autonomous Robots in Logistics

After formulating the problem, this section presents our concept towards a holistic artificial perception framework for autonomous robots in logistics. Our concept is a framework aiming to enable future research focusing on various artificial perception capabilities within the application field of logistics. It aims as a ground work, delivering three essential building blocks:

1) Foundation. Firstly, requirements for an Artificial Perception System (APS) need to be analyzed, defined and documented. The framework is based on a thorough consideration of all requirements, analyzed using information sources ranging from theoretical to practical ones. We differentiate between hardware, software and process requirements. Available information such as literature and international standards, logistics processes as well as market research data get processed in order to form a baseline of requirements which need to be considered when developing APS for logistics. This building block tries to find answers to questions such as, “Which semantic information is useful?”, “Which sensors can be used to acquire it?”, “How can this information be modeled?” and derive perception features suitable for autonomous systems in logistics.

2) Open dataset. Secondly, an open dataset is conceptualized, developed and compiled. The dataset allows the logistics community to get started with and develop novel, application-specific solutions towards artificial perception within logistics. The dataset will consist of data captured in different logistics environments (e.g. automotive industry, distribution centers, wholesale) using multiple sensor modalities (e.g. cameras, lidar) covering various perception capabilities. In addition, we encourage third parties to contribute high quality annotations and datasets. The dataset is composed of several subsets. Subsets offer the possibility to model various perception capabilities required for different use cases and to extend the framework with additional annotations as well as data in the future. We differentiate the subsets by sensor modality used, origin of the data and the tasks annotated. Examples for different sensor modalities used in logistics are color and depth cameras, 2d laser rangefinders and sonars. The data can originate from real logistics environments (natural), simulations and virtual environments (artificial) or from the Internet (online). Tasks to be annotated include object detection, depth estimation, semantic and panoptic segmentation or gripping point estimation. Table 1 lists possible subset candidates.

3) Benchmark suite. The third and final component of our framework is a standardized benchmark suite. This allows drawing comparable conclusions from defined benchmarks in order to make research progress traceable over time. The benchmark suite is envisioned to contain both, a decisive definition of performance indicators on the one hand and standardized test scenarios on the other hand. The benchmarks are based on the previously introduced dataset. They define the development process (e.g. definition of the training data to be used) as well as the evaluation process (e.g. definition of the test set) and thus enable a comparison of different models and methods. This allows, for instance, to test and analyze object detection models for the logistics domain in a standardized manner. Moreover, splitting the dataset into subsets with different properties (i.e. modality, data source and annotated tasks) allows designing benchmarks with increasing levels of complexity. This, for example, allows benchmarking the transfer of synthetically trained models into real applications. Evaluation metrics to be used are derived from relevant literature and extended if necessary. The flexible design of individual benchmarks enables the research community to jointly and transparently push the state-of-the-art.

Table 1. Subset candidates for the open dataset. The proposed framework provides an open dataset consisting of different subsets which we categorize by sensor modality, data source and annotated task

Sensor modality	Data source	Annotated task
Color camera	Natural	Object detection
Color camera	Online	Depth estimation
Color camera	Artificial	Panoptic segmentation
Depth camera	Natural	Object detection
Depth camera	Artificial	Panoptic segmentation
2D laser rangefinder	Natural	Object detection
2D laser rangefinder	Artificial	Panoptic segmentation

6 Conclusion and Future Work

Autonomous robots in Logistics are a promising approach towards a fully automated material flow. In order to use their full potential, they must be able to extract semantic information from logistics environments. We motivate the necessity of artificial perception for logistics and formulate a key challenge. In contrast to other application areas, logistics research has so far not dealt with creating a data-driven foundation that allows developing and comparing novel perception capabilities. In this paper, we conceptualize an artificial perception framework for autonomous robot research in logistics. The framework lays a foundation by distilling logistics-specific standards, concepts and requirements, provides an open, multi-modal, multi-task dataset and defines standardized benchmarks.

Future work will focus on creating and releasing the frameworks initial version. Accordingly, we are working on defining the requirements as well as annotating and providing initial data. As a first version, we plan on releasing lidar, camera and synthetic data data for logistics-specific 2D object detection.

References

1. F. Molzow-Voit, M. Quandt, M. Freitag, and G. Spöttl, *Robotik in der Logistik.* Wiesbaden: Springer Fachmedien Wiesbaden, 2016.
2. A.-K. Rohde, "Robotik in der Logistik – Einsatzpotenziale, Herausforderungen und Trends," in *Robotik in der Logistik*, 2016, pp. 23–40.
3. J. Allmendinger and C. Ebner, "Arbeitsmarkt und demografischer wandel," *Zeitschrift für Arbeits- und Organisationspsychologie*, vol. 50, no. 4, pp. 227–239.
4. C. Hahn-Woernle, "Neue Anforderungen für die Logistik des 21. Jahrhunderts," in *Internet der Dinge in der Intralogistik*, 2010, pp. 9–13.
5. S. J. Russell and P. Norvig, *Artificial intelligence: A modern approach*, 2016.
6. P. Langley, J. E. Laird, and S. Rogers, "Cognitive architectures: Research issues and challenges," *Cognitive Systems Research*, vol. 10, no. 2, pp. 141–160, 2009.
7. L. W. Barsalou, "Grounded cognition," *Annual review of psychology*, vol. 59, pp. 617–645, 2008.
8. R. Mausfeld, "Wahrnehmungspsychologie," in *Psychologie. Eine Einführung in ihre Grundlagen und Anwendungsfelder*, 2010.

9. J. van Brummelen, M. O'Brien, D. Gruyer, and H. Najjaran, "Autonomous vehicle perception: The technology of today and tomorrow," *Transportation Research Part C: Emerging Technologies*, vol. 89, pp. 384–406, 2018.
10. X. Du, M. H. Ang, S. Karaman, and D. Rus, "A general pipeline for 3d detection of vehicles," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*.
11. P. Voigtlaender, M. Krause, A. Osep, J. Luiten, B. B. G. Sekar, A. Geiger, and B. Leibe, "MOTS: Multi-object tracking and segmentation," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.
12. A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the kitti vision benchmark suite," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012.
13. J. Fritsch, T. Kuehnl, and A. Geiger, "A new performance measure and evaluation benchmark for road detection algorithms," in *International Conference on Intelligent Transportation Systems (ITSC)*, 2013.
14. M. Menze and A. Geiger, "Object scene flow for autonomous vehicles," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.
15. P. Sun, H. Kretzschmar et al., "Scalability in perception for autonomous driving: Waymo open dataset," in *Proc. of the IEEE/CVF International Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2019.
16. A. Patil, S. Malla, H. Gang, and Y. Chen, "The h3d dataset for full-surround 3d multi-object detection and tracking in crowded urban scenes," in *2019 International Conference on Robotics and Automation (ICRA)*, 2019, pp. 9552–9557.
17. R. D. Schraft and G. Schmierer, *Service Robots*. A K Peters/CRC Press, 2000.
18. T.-Y. Lin, M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, P. Perona, D. Ramanan, C. L. Zitnick, and P. Dollr, "Microsoft coco: Common objects in context," in *European Conference on Computer Vision (ECCV) Proceedings*, 2014.
19. A. Kasper, Z. Xue, and R. Dillmann, "The kit object models database: An object model database for object recognition, localization and manipulation in service robotics," *The International Journal of Robotics Research*, vol. 31, no. 8, pp. 927–934, 2012.
20. A. B. Jelodar and Y. Sun, "Joint object and state recognition using language knowledge," in *2019 IEEE International Conference on Image Processing (ICIP)*.
21. A. Lang and J. Fottner, "Konzeption eines kamerabasierten kollisionswarnsystems zur prävention von arbeitsunfällen an gabelstaplern."
22. C. Poss, T. Irrenhauser, M. Prueglmeier, D. Goehring, F. Zoghlami, and V. Salehi, "Perceptionbased intelligent materialhandling in industrial logistics environments," in *11th International Conference on Computer and Automation Engineering 2019*.
23. R. Bormann, B. F. de Brito, J. Lindermayr, M. Omainska, and M. Patel, "Towards automated order picking robots for warehouses and retail," in *ICVS*, 2019.
24. B. Molter and J. Fottner, "Real-time pallet localization with 3d camera technology for forklifts in logistic environments," in *2018 IEEE International Conference on Service Operations and Logistics, and Informatics*, pp. 297–302.
25. I. S. Mohamed, A. Capitanelli, F. Mastrogiovanni, S. Rovetta, and R. Zaccaria, "A 2d laser rangefinder scans dataset of standard eur pallets," *Data in Brief*, 2019.
26. D. Hoang, T. Stoyanov, and A. J. Lilienthal, "Object-rpe: Dense 3d reconstruction and pose estimation with convolutional neural networks for warehouse robots," in *2019 European Conference on Mobile Robots (ECMR)*, 2019, pp. 1–6.
27. C. Mayershofer, T. Ge, and J. Fottner, "Towards Fully-Synthetic Training for Industrial Applications," in *IEEE International Conference on Logistics, Informatics and Service Sciences (LISS2020)*.