Semantics for Robotic Mapping, Perception and Interaction: A Survey

Other titles in Foundations and Trends® in Robotics

Embodiment in Socially Interactive Robots
Eric Deng, Bilge Mutlu and Maja J Mataric
ISBN: 978-1-68083-546-5

Modeling, Control, State Estimation and Path Planning Methods for Autonomous Multirotor Aerial Robots
Christos Papachristos, Tung Dang, Shehryar Khattak, Frank Mascarich, Nikhil Khedekar and Kostas Alexis
ISBN: 978-1-68083-548-9

An Algorithmic Perspective on Imitation Learning
Takayuki Osa, Joni Pajarinen, Gerhard Neumann, J. Andrew Bagnell, Pieter Abbeel and Jan Peters
ISBN:978-1-68083-410-9

Semantics for Robotic Mapping, Perception and Interaction: A Survey

Sourav Garg
Niko Sünderhauf
Feras Dayoub
Douglas Morrison
Akansel Cosgun
Gustavo Carneiro
Qi Wu
Tat-Jun Chin
Ian Reid
Stephen Gould
Peter Corke
Michael Milford

now
the essence of knowledge
Boston — Delft

Foundations and Trends® in Robotics

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
United States
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is

S. Garg, N. Sünderhauf, F. Dayoub, D. Morrison, A. Cosgun, G. Carneiro, Q. Wu, T.-J. Chin, I. Reid, S. Gould, P. Corke and M. Milford. *Semantics for Robotic Mapping, Perception and Interaction: A Survey*. Foundations and Trends® in Robotics, vol. 8, no. 1–2, pp. 1–224, 2020.

ISBN: 978-1-68083-769-8

Foundations and Trends® in Robotics
Volume 8, Issue 1-2, 2020

Editorial Scope

Topics

Foundations and Trends® in Robotics publishes survey and tutorial articles in the following topics:

- Mathematical modelling
- Kinematics
- Dynamics
- Estimation Methods
- Robot Control
- Planning
- Artificial Intelligence in Robotics
- Software Systems and Architectures
- Mechanisms and Actuators
- Sensors and Estimation
- Planning and Control
- Human-Robot Interaction
- Industrial Robotics
- Service Robotics

Information for Librarians

Foundations and Trends® in Robotics, 2020, Volume 8, 4 issues. ISSN paper version 1935-8253. ISSN online version 1935-8261. Also available as a combined paper and online subscription.

Contents

Semantics for Robotic Mapping, Perception and Interaction: A Survey

Sourav Garg[1], Niko Sünderhauf[1], Feras Dayoub[1], Douglas Morrison[1], Akansel Cosgun[2], Gustavo Carneiro[3], Qi Wu[3], Tat-Jun Chin[3], Ian Reid[3], Stephen Gould[4], Peter Corke[1] and Michael Milford[1]

[1]*QUT Centre for Robotics and School of Electrical Engineering and Robotics, Queensland University of Technology, Australia*
[2]*Department of Electrical and Computer Systems Engineering, Monash University, Australia*
[3]*School of Computer Science, University of Adelaide, Australia*
[4]*College of Engineering and Computer Science, Australian National University, Australia*

ABSTRACT

For robots to navigate and interact more richly with the world around them, they will likely require a deeper *understanding* of the world in which they operate. In robotics and related research fields, the study of understanding is often referred to as semantics, which dictates *what does the world "mean" to a robot*, and is strongly tied to the question of *how to represent that meaning*. With humans and robots increasingly operating in the same world, the prospects of human–robot interaction also bring *semantics and ontology of natural language* into the picture. Driven by need, as well as by enablers like increasing availability of training data and computational resources, semantics is

---

Sourav Garg, Niko Sünderhauf, Feras Dayoub, Douglas Morrison, Akansel Cosgun, Gustavo Carneiro, Qi Wu, Tat-Jun Chin, Ian Reid, Stephen Gould, Peter Corke and Michael Milford (2020), "Semantics for Robotic Mapping, Perception and Interaction: A Survey", Foundations and Trends® in Robotics: Vol. 8, No. 1–2, pp 1–224. DOI: 10.1561/2300000059.

a rapidly growing research area in robotics. The field has received significant attention in the research literature to date, but most reviews and surveys have focused on particular aspects of the topic: the technical research issues regarding its use in specific robotic topics like mapping or segmentation, or its relevance to one particular application domain like autonomous driving. A new treatment is therefore required, and is also timely because so much relevant research has occurred since many of the key surveys were published. This survey therefore provides an overarching snapshot of where semantics in robotics stands today. We establish a taxonomy for semantics research in or relevant to robotics, split into four broad categories of activity, in which semantics are extracted, used, or both. Within these broad categories we survey dozens of major topics including fundamentals from the computer vision field and key robotics research areas utilizing semantics, including mapping, navigation and interaction with the world. The survey also covers key practical considerations, including enablers like increased data availability and improved computational hardware, and major application areas where semantics is or is likely to play a key role. In creating this survey, we hope to provide researchers across academia and industry with a comprehensive reference that helps facilitate future research in this exciting field.

1

Introduction

For robots to move beyond the niche environments of fulfilment warehouses, underground mines and manufacturing plants into widespread deployment in industry and society, they will need to understand the world around them. Most mobile robot and drone systems deployed today make relatively little use of explicit higher level "meaning" and typically only consider geometric maps of environments, or three dimensional models of objects in manufacturing and logistics contexts. Despite considerable success and uptake to date, there is a large range of domains with few, if any, commercial robotic deployments; for example: aged care and assisted living facilities, autonomous on-road vehicles, and drones operating in close proximity to cluttered, human-filled environments. Many challenges remain to be solved, but we argue one of the most significant is simply that robots will need to better *understand* the world in which they operate, in order for them to move into useful and safe deployments in more diverse environments. This need for *understanding* is where *semantics* meet robotics.

Semantics is a widely used term, not just in robotics but across fields ranging from linguistics to philosophy. In the robotics domain, despite widespread *usage* of semantics, there is relatively little formal definition of what the term means. In this survey, we aim to provide a taxonomy rather than specific definition of semantics, and note that the surveyed research exists along a spectrum from traditional, non-semantic approaches to those which are primarily semantically-based. Broadly speaking, we can consider semantics in a robotics context to be about the *meaning* of things: the meaning of places, objects, other entities occupying the environment, or even the language used in communicating between robots and humans or between robots themselves. There are several areas of relevance to robotics where semantics have been a strong focus in recent years, including SLAM (Simultaneous Localization And Mapping), segmentation and object recognition.

Given the importance of semantics for robotics, how can they be equipped, or learn about meaning in the world? There are multiple methods, but they can be split into two categories. Firstly, *provided semantics* describes situations where the robot is given the knowledge beforehand. *Learnt semantics* describes situations where the robot, either beforehand or during deployment, learns this information. The learning mechanism leads to further sub categorization: learning from observation of the entities of interest in the environment, and actual interaction with the entities of interest, such as through manipulation. Learning can occur in a supervised, unsupervised or semi-supervised manner.

Semantics as a research field in robotics has grown rapidly in the past decade. This growth has been driven in part by the opportunity to use semantics to improve the capabilities of robotic systems in general, but several other factors have also contributed. The popularization of deep learning over the past decade has facilitated much of the research in semantics for robotics, enabling capabilities like high performance general object recognition – *object class* is an important and useful "meaning" associated with the world. Increases in dataset availability, access to the cloud and compute resources have also been critical, providing a much richer source of information from which robots can learn, and the computational power with which to do so, rapidly and at scale. Given the popularity of the field, it has been the focus of a number of key review and survey papers, which we cover here.

1.1 Past Coverage Including Survey and Review Papers

As one of the core fields of research in robotics, SLAM has been the subject of a number of reviews, surveys and tutorials over the past few decades, including coverage of semantic concepts. A recent review of SLAM by Cadena *et al.* [1] positioned itself and other existing surveys as belonging to either classical [2]–[5], algorithmic-analysis [6] or the "robust perception" age [1]. This review highlighted the evolution of SLAM, from its earlier focus on probabilistic formulations and analyses of fundamental properties to the current and growing focus on robustness, scalability and high-level understanding of the environment – where semantics comes to the fore. Also discussed were robustness and scalability in the context of long-term autonomy and how the underlying representation models, metric *and* semantic, shape the SLAM problem formulation. For the high-level semantic representation and understanding of the environment, [1] discussed literature where SLAM is used for inferring semantics, and semantics are used to improve SLAM, paving the way for a joint optimization-based semantic SLAM system. They further highlight the key challenges of such a system, including consistent fusion of semantic and metric information, task-driven semantic representations, dynamic adaptation to changes in the environment, and effective exploitation of semantic reasoning.

Kostavelis and Gasteratos [7] reviewed semantic mapping for mobile robotics. They categorized algorithms according to scalability, inference model, temporal coherence and topological map usage, while outlining their potential applications and emphasizing the role of human interaction, knowledge representation and planning. More recently, [8] reviewed current progress in SLAM research, proposing that SLAM including semantic components could evolve into "Spatial AI", either in the form of autonomous AI, such as a robot, or as Intelligent Augmentation (IA), such as in the form of an AR headset. A Spatial AI system would not only capture and represent the scene intelligently but also take into account the embodied device's constraints, thus requiring joint innovation and co-design of algorithms, processors and sensors. More recently, [9] presented Gaussian Belief Propagation (GBP) as an algorithmic framework suited to the needs of a Spatial AI system, capable of delivering high performance despite resource constraints. The proposals in both [8] and [9] are significantly motivated by rapid developments in processor hardware, and touch on the opportunities for closing the gap between intelligent perception and resource-constrained deployment devices. More recently, [10] surveyed the use of semantics for visual SLAM, particularly reviewing the integration of "semantic extractors" (object recognition and semantic segmentation) within modern visual SLAM pipelines.

Much of the growth in semantics-based approaches has coincided with the increase in capabilities brought about by the modern deep learning revolution. Schmidhuber [11] presented a detailed review of deep learning in neural networks as applied through supervised, unsupervised and reinforcement learning regimes. They introduced the concept of "Credit Assignment Paths" (CAPs) representing chains of causal links between events, which help understand the level of depth required for a given Neural Network (NN) application. With a vision of creating general purpose learning algorithms, they highlighted the need for a brain-like learning system following the rules of fractional neural activation and sparse neural connectivity. Liu *et al.* [12] presented a more focused review of deep learning, surveying generic object detection. They highlighted the key elements involved in the task such as the accuracy–efficiency trade-off of detection frameworks, the choice and evolution of backbone networks, the robustness of object representation and reasoning based on additionally available context.

A significant body of work has focused on extracting more meaningful abstractions of the raw data typically obtained in robotics such as 3D point clouds. Towards this end, a number of surveys have been conducted in recent years for point cloud filtering [13] and description [14], 3D shape/object classification [15], [16], 3D object detection [15]–[19], 3D object tracking [16] and 3D semantic segmentation [15]–[17], [20]–[24]. With only a couple of exceptions, all of these surveys have particularly reviewed the use of *deep learning* on 3D point clouds for respective tasks. Segmentation has also long been a fundamental component of many robotic and autonomous vehicle systems, with *semantic* segmentation focusing on labeling areas or pixels in an image by class type. In particular the overall goal is to label by class, not by instance. For example, in an autonomous vehicle context this goal constitutes labeling pixels as belonging to *a* vehicle, rather than as a specific instance of a vehicle (although that is also an important capability). The topic has been the focus of a large quantity of research with resulting survey papers that focus primarily on semantic segmentation, such as [22], [25]–[30].

Beyond these flagship domains, semantics have also been investigated in a range of other subdomains. Ramirez-Amaro *et al.* [31] reviewed the use of semantics in the context of understanding human actions and activities, to enable a robot to execute a task. They classified semantics-based methods for *recognition* into four categories: syntactic methods based on symbols and rules, affordance-based understanding of objects in the environment, graph-based encoding of complex variable relations, and knowledge-based methods. In conjunction with recognition, different methods to *learn and execute* various tasks were also reviewed including learning by demonstration, learning by observation and execution based on structured plans. Likewise for the service robotics field, [32] presented a survey of vision-based semantic mapping, particularly focusing on its need for an effective human–robot interface for service robots, beyond pure navigation capabilities. Paulius and Sun [33] also surveyed knowledge representations in service robotics.

Many robots are likely to require image retrieval capabilities where semantics may play a key role, including in scenarios where humans are interacting with the robotic systems. Enser and Sandom [34] and Liu *et al.* [35] surveyed the "semantic gap" in current content-based image retrieval systems, highlighting the discrepancy between the limited descriptive power of low-level image features and the typical richness of (human) user semantics. Bridging this gap is likely to be important for both improved robot capabilities *and* better interfaces with humans. Some of the reviewed approaches to reducing the semantic gap, as discussed in [35], include the use of object ontology, learning meaningful associations between image features and query concepts and learning the user's intention by relevance feedback. This semantic gap concept has gained significant attention and is reviewed in a range of other papers including [36]–[42]. Acting upon the enriched understanding of the scene, robots are also likely to require sophisticated grasping capabilities, as reviewed in [43], covering vision-based robotic grasping in the context of object localization, pose estimation, grasp detection and motion planning. Enriched interaction with the environment based on an understanding of what can be done with an object – its "affordances" – is also important, as reviewed in [44]. Enriched interaction with humans is also likely to require an understanding of language, as reviewed recently by [45]. This review covers some of the key elements of language usage by robots: collaboration via dialogue with a person, language as a means to drive learning and understanding natural language requests, and deployment, as shown in application examples.

1.2 Summary and Rationale for This Survey

The majority of semantics coverage in the literature to date has occurred with respect to a specific research topic, such as SLAM or segmentation, or targeted to specific application areas, such as autonomous vehicles. As can be seen in the previous subsection, there has been both extensive research across these fields as well as a number of key survey and review papers summarizing progress to date. These deep dives into specific sub-areas in robotics can provide readers with a deep understanding of technical considerations regarding semantics in that context. As the field continues to grow however there is increasing need for an overview that more broadly covers semantics across all of robotics, whilst still providing sufficient technical coverage to be of use to practitioners working in these fields. For example, while [1] extensively considers the use of semantics primarily within SLAM research, there is a need to cover the role of semantics more broadly in various robotics tasks and competencies which are closely related to each other. The task, "bring a cup of coffee", likely requires semantic understanding borne out of both the underlying SLAM system *and* the affordance-grasping pipeline. This survey therefore goes beyond specific application domains or methodologies to provide an overarching survey of semantics across all of robotics, as well as the semantics-enabling research that occurs in related fields like computer vision and machine learning. To encompass such a broad range of topics in this survey, we have divided our coverage of research relating to semantics into (a) the fundamentals underlying the current and potential use of semantics in robotics, (b) the widespread use of semantics in robotic mapping and navigation systems, and (c) the use of semantics to enhance the range of interactions robots have with the world, with humans, and with other robots.

This survey is also motivated by timeliness: the use of semantics is a rapidly evolving area, due to both significant current interest in this field, as well as technological advances in local and cloud compute, and the increasing availability of data that is critical to developing or training these semantic systems. Consequently, with many of the key papers now half a decade old or more, it is useful to capture a snapshot of the field as it stands now, and to update the treatment of various topic areas based on recently proposed paradigms. For example, this survey discusses recent semantic mapping paradigms that mostly post-date key papers by [1], [7], such as combining single- and multi-view point clouds with semantic segmentation to directly obtain a local semantic map [46]–[50]. Whilst contributing a new overview of the use of semantics across robotics in general, we are also careful to adhere where possible to recent proposed taxonomies in specific research areas. For example, in the area of 3D point clouds and their usage for semantics, within Subsection 3.3, with the help of key representative papers, we *briefly* describe the recent research evolution of using 3D point cloud representations for learning object- or pixel-level semantic labeling, in line with the taxonomy proposed by existing comprehensive surveys [15]–[17], [22], [23]. Finally, beyond covering new high level conceptual developments, there is also the need to simply update the paper-level coverage of what has been an incredibly large volume of research in these fields even over the past five years. The survey refers to well over 100 research works from the past year alone, representative of a much larger total number of research works. This breadth of coverage would normally come at the cost of some depth of coverage: here we have attempted to cover the individual topics in as much detail as possible, with over 900 referenced works covered in total. Where appropriate we also make reference to prior survey and review papers where further detailed coverage may be of interest, such as for the topic of 3D point clouds and their usage for semantics.

Moving beyond single application domains, we also provide an overview of how the use of semantics is becoming an increasingly integral part of many trial (and in some cases full scale commercial) deployments including in autonomous vehicles, service robotics and drones. A richer understanding of the world will open up opportunities for robotic deployments in contexts traditionally *too difficult* for safe robotic deployment: nowhere is this more apparent perhaps than for on-road autonomous vehicles, where a subtle, nuanced understanding of all aspects of the driving task is likely required before robot cars become comparable to, or ideally superior to, human drivers. Compute and data availability has also enabled many of the advancements in semantics-based robotics research; likewise these technological advances have also facilitated investigation of their deployment in robotic applications that previously would have been unthinkable – such as enabling sufficient on-board computation for deploying semantic techniques on power- and weight-limited drones. We cover the current and likely future advancements in computational technology relevant to semantics, both local and online versions, as well as the burgeoning availability of rich, informative datasets that can be used for training semantically-informed systems.

In summary, this survey aims to provide a unifying overview of the development and use of semantics across the entire robotics field, covering as much detailed work as feasible whilst referencing the reader to further details where appropriate. Beyond its breadth, the survey represents a substantial update to the semantics topics covered in survey and review papers published even only a few years ago. By surveying the technical research, the application domains and the technology enablers in a single treatment of the field, we can provide a unified snapshot of what is possible now and what is likely to be possible in the near future.

1.3 Taxonomy and Survey Structure

Existing literature covering the role of *semantics* in robotics is fragmented and is usually discussed in a variety of task- and application-specific contexts. In this survey, we consolidate the disconnected semantics research in robotics; draw links with the fundamental computer vision capabilities of extracting semantic information; cover a range of potential applications that typically require high-level decision making; and discuss critical upcoming enhancers for improving the scope and use of semantics. To aid in navigating this rapidly growing and already sizable field, here we propose a taxonomy of semantics as it pertains to robotics (see Figure 1.1). We find the relevant literature can be divided into four broad categories:

1. *Static or Un-embodied Scene Understanding*, where the focus of research is typically on developing intrinsic capability to extract semantic information from images, for example, object recognition and image classification. The majority of research in this direction uses single image-based 2D input to infer the underlying semantic or 3D content of that image. However, image acquisition and processing in this case is primarily static in nature (including videos shot by a static camera), separating it conceptually from a mobile embodied agent's dynamic perception of the environment due to motion of the agent. Because RGB cameras are widely used in robotics, and the tasks being performed, such as object recognition, are also performed by robots, advances in this area are relevant to robotics research. In Section 2, we introduce the fundamental components of semantics that relate to or enable robotics, focusing on topics that have been primarily or initially investigated in non-robotics but related research fields, such as computer vision. We cover the key components of semantics as regards object detection, segmentation, scene representations and image retrieval, all highly relevant capabilities for robotics, even if not all the work has yet been demonstrated on robotic platforms.

2. *Mobile Environment Understanding and Mapping*, where the research is typically motivated by the mobile or dynamic nature of robots and their surroundings. The research literature in this category includes the task of semantic mapping, which could be topological, or a dense and precise 3D reconstruction. These mapping tasks can often leverage advances in static scene understanding research, for example, place categorization (image classification)

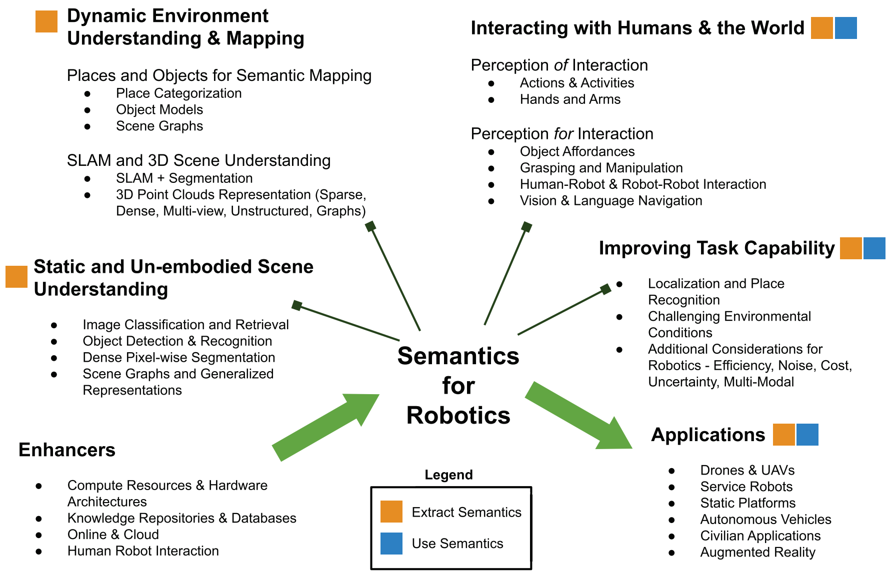

Figure 1.1: A taxonomy for semantics in robotics. Four broad categories of semantics research are complemented by technological, knowledge and training-related enhancers and lead to a range of robotic applications. Areas can be primarily involved in extracting semantics, using semantics or a combination of both.

forming the basis of semantic topological mapping, or pixel-wise semantic segmentation being used as part of a semantic 3D reconstruction pipeline. Semantic maps provide a representation of information and understanding at an environment or spatial level. With the increasing use of 3D sensing devices, along with the maturity of visual SLAM, research on semantic understanding of 3D point clouds is also growing, aimed at enabling a richer semantic representation of the 3D world. In Section 3, we cover the use of semantics for developing representations and understanding at an environment level. This includes the use of places, objects and scene graphs for semantic mapping, and 3D scene understanding through Simultaneous Localization And Mapping (SLAM) and point clouds processing.

3. *Interaction*, where the existing research "connects the dots" between the ability to *perceive* and the ability to *act*. The literature

in this space can be further divided into the "perception *of* interaction" and "perception *for* interaction". The former includes the basic abilities of understanding actions and activities of humans and other dynamic agents, and enabling robots to learn from demonstration. The latter encompasses research related to the use of the perceived information to act or perform a task, for example, developing a manipulation strategy for a detected object. In the context of robotics, detecting an object's affordances can be as important as recognizing that object, enabling semantic reasoning relevant to the task and affordances (e.g., "cut" and "contain") rather than to the specific object category (e.g., "knife" and "jar"). While object grasping and manipulation relate to a robot's interaction with the environment, research on interaction with other *humans* and *robots* includes the use of natural language to generate inverse semantics, or to follow navigation instructions. Section 4 addresses the use of semantics to facilitate robot interaction with the world, as well as with the humans and robots that inhabit that world. It looks at key issues around affordances, grasping, manipulation, higher-level goals and decision making, human–robot interaction and vision-and-language navigation.

4. *Improving Task Capability*, where researchers have focused on utilizing semantic representations to improve the capability of other tasks. This includes for example the use of semantics for high-level reasoning to improve localization and visual place recognition techniques. Furthermore, semantic information can be used to solve more challenging problems such as dealing with challenging environmental conditions. Robotics researchers have also focused on techniques that unlock the full potential of semantics in robotics, since existing research has not always been motivated by or had to deal with the challenges of real world robotic applications, by addressing challenges like noise, clutter, cost, uncertainty and efficiency. In Section 5, we discuss various ways in which researchers extract or employ semantic representations for localization and visual place recognition, dealing with challenging environmental

conditions, and generally enabling semantics in a robotics context through addressing additional challenges.

The four broad categories presented above encompass the relevant literature on how semantics are defined or used in various contexts in robotics and related fields. This is also reflected in Figure 1.1 through "extract semantics" and "use semantics" labels associated with different sections of the taxonomy. Extracting semantics from images, videos, 3D point clouds, or by actively traversing an environment are all methods of creating semantic representations. Such semantic representations can be input into high-level reasoning and decision-making processes, enabling execution of complex tasks such as path planning in a crowded environment, pedestrian intention prediction, and vehicle trajectory prediction. Moreover, the use of semantics is often fine-tuned to particular applications like agricultural robotics, autonomous driving, augmented reality and UAVs. Rather than simply being exploited, the semantic representations themselves can be jointly developed and defined in consideration of how they are then used. Hence, in Figure 1.1, the sections associated with "use semantics" are also associated with "extract semantics". These high-level tasks can benefit from advances in fundamental and applied research related to semantics. But this research alone is not enough: advances in other areas are critical, such as better cloud infrastructure, advanced hardware architectures and compute capability, and the availability of large datasets and knowledge repositories. Section 6 reviews the influx of semantics-based approaches for robotic deployments across a wide range of domains, as well as the critical technology enablers underpinning much of this current and future progress. Finally, Section 7 discusses some of the key remaining challenges in the field and opportunities for addressing them through future research, concluding coverage of what is likely to remain an exciting and highly active research area into the future.

References

[1] C. Cadena, L. Carlone, H. Carrillo, Y. Latif, D. Scaramuzza, J. Neira, I. Reid, and J. J. Leonard, "Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age," *IEEE Transactions on Robotics*, vol. 32, no. 6, pp. 1309–1332, 2016.

[2] H. Durrant-Whyte and T. Bailey, "Simultaneous localization and mapping: Part i," *IEEE Robotics and Automation Magazine*, vol. 13, no. 2, pp. 99–110, 2006.

[3] T. Bailey and H. Durrant-Whyte, "Simultaneous localization and mapping (SLAM): Part ii," *IEEE Robotics and Automation Magazine*, vol. 13, no. 3, pp. 108–117, 2006.

[4] S. Thrun, "Probabilistic robotics," *Communications of the ACM*, vol. 45, no. 3, pp. 52–57, 2002.

[5] C. Stachniss, J. J. Leonard, and S. Thrun, "Simultaneous localization and mapping," in *Springer Handbook of Robotics*, Cham, Switzerland: Springer, 2016, pp. 1153–1176.

[6] G. Dissanayake, S. Huang, Z. Wang, and R. Ranasinghe, "A review of recent developments in simultaneous localization and mapping," in *2011 6th International Conference on Industrial and Information Systems*, IEEE, pp. 477–482, 2011.

[7] I. Kostavelis and A. Gasteratos, "Semantic mapping for mobile robotics tasks: A survey," *Robotics and Autonomous Systems*, vol. 66, pp. 86–103, 2015.

[8] A. J. Davison, "FutureMapping: The computational structure of spatial AI systems," *arXiv preprint arXiv:1803.11288*, 2018.

[9] A. J. Davison and J. Ortiz, "Futuremapping 2: Gaussian belief propagation for spatial ai," *arXiv preprint arXiv:1910.14139*, 2019.

[10] L. Xia, J. Cui, R. Shen, X. Xu, Y. Gao, and X. Li, "A survey of image semantics-based visual simultaneous localization and mapping: Application-oriented solutions to autonomous navigation of mobile robots," *International Journal of Advanced Robotic Systems*, vol. 17, no. 3, Art. no. 1729881420919185, 2020.

[11] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Networks*, vol. 61, pp. 85–117, 2015.

[12] L. Liu, W. Ouyang, X. Wang, P. Fieguth, J. Chen, X. Liu, and M. Pietikäinen, "Deep learning for generic object detection: A survey," *International Journal of Computer Vision*, vol. 128, no. 2, pp. 261–318, 2020.

[13] X.-F. Han, J. S. Jin, M.-J. Wang, W. Jiang, L. Gao, and L. Xiao, "A review of algorithms for filtering the 3D point cloud," *Signal Processing: Image Communication*, vol. 57, pp. 103–112, 2017.

[14] X.-F. Hana, J. S. Jin, J. Xie, M.-J. Wang, and W. Jiang, "A comprehensive review of 3d point cloud descriptors," *arXiv preprint arXiv:1802.02297*, 2018.

[15] W. Liu, J. Sun, W. Li, T. Hu, and P. Wang, "Deep learning on point clouds and its application: A survey," *Sensors*, vol. 19, no. 19, p. 4188, 2019.

[16] Y. Guo, H. Wang, Q. Hu, H. Liu, L. Liu, and M. Bennamoun, "Deep learning for 3d point clouds: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020.

[17] A. Ioannidou, E. Chatzilari, S. Nikolopoulos, and I. Kompatsiaris, "Deep learning advances in computer vision with 3d data: A survey," *ACM Computing Surveys (CSUR)*, vol. 50, no. 2, pp. 1–38, 2017.

[18] E. Arnold, O. Y. Al-Jarrah, M. Dianati, S. Fallah, D. Oxtoby, and A. Mouzakitis, "A survey on 3d object detection methods for autonomous driving applications," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 10, pp. 3782–3795, 2019.

[19] M. M. Rahman, Y. Tan, J. Xue, and K. Lu, "Recent advances in 3D object detection in the era of deep neural networks: A survey," *IEEE Transactions on Image Processing*, vol. 29, pp. 2947–2962, 2019.

[20] E. Grilli, F. Menna, and F. Remondino, "A review of point clouds segmentation and classification algorithms," *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 42, p. 339, 2017.

[21] J. Zhang, X. Zhao, Z. Chen, and Z. Lu, "A review of deep learning-based semantic segmentation for point cloud (november 2019)," *IEEE Access*, vol. 7, pp. 179 118–179 133, 2019.

[22] F. Lateef and Y. Ruichek, "Survey on semantic segmentation using deep learning techniques," *Neurocomputing*, vol. 338, pp. 321–348, 2019.

[23] Y. Xie, J. Tian, and X. Zhu, "A review of point cloud semantic segmentation," *IEEE Geoscience and Remote Sensing Magazine (GRSM)*, 2020.

[24] E. Malinverni, R. Pierdicca, M. Paolanti, M. Martini, C. Morbidoni, F. Matrone, and A. Lingua, "Deep learning for semantic segmentation of 3D point cloud," *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 42, no. 2, pp. 735–742, 2019.

[25] A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, P. Martinez-Gonzalez, and J. Garcia-Rodriguez, "A survey on deep learning techniques for image and video semantic segmentation," *Applied Soft Computing*, vol. 70, pp. 41–65, 2018.

[26] Y. Guo, Y. Liu, T. Georgiou, and M. S. Lew, "A review of semantic segmentation using deep neural networks," *International Journal of Multimedia Information Retrieval*, vol. 7, no. 2, pp. 87–93, 2018.

[27] B. Zhao, J. Feng, X. Wu, and S. Yan, “A survey on deep learning-based fine-grained object classification and semantic segmentation,” *International Journal of Automation and Computing*, vol. 14, no. 2, pp. 119–135, 2017.
[28] M. Siam, S. Elkerdawy, M. Jagersand, and S. Yogamani, “Deep semantic segmentation for automated driving: Taxonomy, roadmap and challenges,” in *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*, IEEE, pp. 1–8, 2017.
[29] M. Thoma, “A survey of semantic segmentation,” *arXiv preprint arXiv:1602.06541*, 2016.
[30] H. Zhu, F. Meng, J. Cai, and S. Lu, “Beyond pixels: A comprehensive survey from bottom-up to semantic image segmentation and cosegmentation,” *Journal of Visual Communication and Image Representation*, vol. 34, pp. 12–27, 2016.
[31] K. Ramirez-Amaro, Y. Yang, and G. Cheng, “A survey on semantic-based methods for the understanding of human movements,” *Robotics and Autonomous Systems*, vol. 119, pp. 31–50, 2019.
[32] Q. Liu, R. Li, H. Hu, and D. Gu, “Extracting semantic information from visual data: A survey,” *Robotics*, vol. 5, no. 1, 2016.
[33] D. Paulius and Y. Sun, “A survey of knowledge representation in service robotics,” *Robotics and Autonomous Systems*, vol. 118, pp. 13–30, 2019.
[34] P. Enser and C. Sandom, “Towards a comprehensive survey of the semantic gap in visual image retrieval,” in *International Conference on Image and Video Retrieval*, Berlin and Heidelberg, Germany: Springer, pp. 291–299, 2003.
[35] Y. Liu, D. Zhang, G. Lu, and W. Y. Ma, “A survey of content-based image retrieval with high-level semantics,” *Pattern Recognition*, vol. 40, no. 1, pp. 262–282, 2007.
[36] X. Li, T. Uricchio, L. Ballan, M. Bertini, C. G. Snoek, and A. D. Bimbo, “Socializing the semantic gap: A comparative survey on image tag assignment, refinement, and retrieval,” *ACM Computing Surveys (CSUR)*, vol. 49, no. 1, pp. 1–39, 2016.

[37] D. Sudha and J. Priyadarshini, "Reducing semantic gap in video retrieval with fusion: A survey," *Procedia Computer Science*, vol. 50, pp. 496–502, 2015.

[38] A. Ramkumar and B. Poorna, "Ontology based semantic search: An introduction and a survey of current approaches," in *2014 International Conference on Intelligent Computing Applications*, IEEE, pp. 372–376, 2014.

[39] D. Zhang, M. M. Islam, and G. Lu, "A review on automatic image annotation techniques," *Pattern Recognition*, vol. 45, no. 1, pp. 346–362, 2012.

[40] M. Ehrig, *Ontology Alignment: Bridging the Semantic Gap*, vol. 4. Boston, MA: Springer Science and Business Media, 2006.

[41] J. S. Hare, P. H. Lewis, P. G. Enser, and C. J. Sandom, "Mind the gap: Another look at the problem of the semantic gap in image retrieval," in *Multimedia Content Analysis, Management, and Retrieval 2006*, International Society for Optics and Photonics, vol. 6073, Art. no. 607309, 2006.

[42] C. Wen and G. Geng, "Review and research on 'semantic gap' problem in the content based image retrieval," *Journal of Northwest University (Natural Science Edition)*, vol. 35, no. 5, pp. 536–540, 2005.

[43] G. Du, K. Wang, S. Lian, and K. Zhao, "Vision-based robotic grasping from object localization, object pose estimation to grasp estimation for parallel grippers: A review," *Artificial Intelligence Review*, pp. 1–58, 2020.

[44] P. Ardón, È. Pairet, K. S. Lohan, S. Ramamoorthy, and R. Petrick, "Affordances in robotic tasks—A survey," *arXiv preprint arXiv:2004.07400*, 2020.

[45] S. T. Tellex, N. Gopalan, H. Kress-Gazit, and C. Matuszek, "Robots that use language," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 3, pp. 22–55, 2020.

[46] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, "PointNet: Deep learning on point sets for 3D classification and segmentation," in *Proceedings—30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, pp. 601–610, 2017.

[47] F. J. Lawin, M. Danelljan, P. Tosteberg, G. Bhat, F. S. Shahbaz Khan, and M. Felsberg, "Deep projective 3D semantic segmentation," in *International Conference on Computer Analysis of Images and Patterns*, pp. 95–107, 2017.

[48] J. Huang and S. You, "Point cloud labeling using 3d convolutional neural network," in *2016 23rd International Conference on Pattern Recognition (ICPR)*, IEEE, pp. 2670–2675, 2016.

[49] A. Boulch, J. Guerry, B. Le Saux, and N. Audebert, "SnapNet: 3D point cloud semantic labeling with 2D deep segmentation networks," *Computers and Graphics (Pergamon)*, vol. 71, pp. 189–198, 2018. DOI: 10.1016/j.cag.2017.11.010.

[50] A. Dai and M. Nießner, "3DMV: Joint 3D-multi-view prediction for 3D semantic scene segmentation," in *Proceedings of the European Conference on Computer Vision (ECCV)*, vol. 11214 LNCS, pp. 458–474, 2018.

[51] L.-J. Li, R. Socher, and L. Fei-Fei, "Towards total scene understanding: Classification, annotation and segmentation in an automatic framework," in *2009 IEEE Conference on Computer Vision and Pattern Recognition,* IEEE, pp. 2036–2043, 2009.

[52] L. J. Li, H. Su, E. P. Xing, and L. Fei-Fei, "Object Bank: A high-level image representation for scene classification and semantic feature sparsification," in *Advances in Neural Information Processing Systems 23: 24th Annual Conference on Neural Information Processing Systems 2010, NIPS 2010*, pp. 1–9, 2010.

[53] J. Yao, S. Fidler, and R. Urtasun, "Describing the scene as a whole: Joint object detection, scene classification and semantic segmentation," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 702–709, 2012.

[54] J. Sánchez, F. Perronnin, T. Mensink, and J. Verbeek, "Image classification with the fisher vector: Theory and practice," *International Journal of Computer Vision*, vol. 105, no. 3, pp. 222–245, 2013.

[55] Y. Qi, G. Zhang, and Y. Li, "Image classification model using visual bag of semantic words," *Pattern Recognition and Image Analysis*, vol. 29, no. 3, pp. 404–414, 2019.

[56] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, pp. 248–255, 2009.

[57] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, pp. 1097–1105, 2012.

[58] B. Zhou, A. Lapedriza Garcia, J. Xiao, A. Torralba, and A. Oliva, "Learning deep features for scene recognition using places database," in *MIT Web Domain*, pp. 487–495, 2014.

[59] B. Zhou, A. Lapedriza, A. Khosla, A. Oliva, and A. Torralba, "Places: A 10 million image database for scene recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 6, pp. 1452–1464, 2018.

[60] B. Zhou, A. Khosla, À. Lapedriza, A. Oliva, and A. Torralba, "Object detectors emerge in deep scene CNNs," in *ICLR*, 2015.

[61] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning deep features for discriminative localization," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016.

[62] K. Meng, X. Fan, H. Li, F. Sun, Z. Wang, and Z. Luo, "A deep multi-modal fusion approach for semantic place prediction in social media," in *MUSA2 2017—Proceedings of the Workshop on Multimodal Understanding of Social, Affective and Subjective Attributes, Co-Located with MM 2017*, pp. 31–37, 2017.

[63] T. Wei, W. Hu, X. Wu, Y. Zheng, H. Ye, J. Yang, and L. He, "Aggregating rich deep semantic features for fine-grained place classification," in *International Conference on Artificial Neural Networks*, vol. 11729 LNCS, pp. 55–67, Springer International Publishing, 2019. DOI: 10.1007/978-3-030-30508-6_5.

[64] A. Wang, J. Cai, J. Lu, and T.-J. Cham, "Modality and component aware feature fusion for RGB-D scene classification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5995–6004, 2016.

[65] S. Gupta, J. Hoffman, and J. Malik, "Cross modal distillation for supervision transfer," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2827–2836, 2016.

[66] D. Du, L. Wang, H. Wang, K. Zhao, and G. Wu, "Translate-to-recognize networks for rgb-d scene recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 11 836–11 845, 2019.

[67] R. Hong, Y. Yang, M. Wang, and X.-S. Hua, "Learning visual semantic relationships for efficient visual retrieval," *IEEE Transactions on Big Data*, vol. 1, no. 4, pp. 152–161, 2016.

[68] F. Huang, C. Jin, Y. Zhang, K. Weng, T. Zhang, and W. Fan, "Sketch-based image retrieval with deep visual semantic descriptor," *Pattern Recognition*, vol. 76, pp. 537–548, 2018.

[69] N. V. Hoang, M. Rukoz, and M. Manouvrier, "Latent semantic minimal hashing for image retrieval," *IEEE Transactions on Image Processing on Image Processing*, vol. 26, no. 1, pp. 1–20, 2009.

[70] A. Gordo and D. Larlus, "Beyond instance-level image retrieval: Leveraging captions to learn a global visual representation for semantic retrieval," in *Proceedings—30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, vol. 2017-Jan., pp. 5272–5281, 2017.

[71] C. Bai, J. Nan Chen, L. Huang, K. Kpalma, and S. Chen, "Saliency-based multi-feature modeling for semantic image retrieval," *Journal of Visual Communication and Image Representation*, vol. 50, no. Dec. 2017, pp. 199–204, 2018. DOI: 10.1016/j.jvcir.2017.11.021.

[72] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[73] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987, 2002.

[74] M. M. Cheng, N. J. Mitra, X. Huang, P. H. Torr, and S. M. Hu, "Global contrast based salient region detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 3, pp. 569–582, 2015.

[75] M. H. Quinn, E. Conser, J. M. Witte, and M. Mitchell, "Semantic image retrieval via active grounding of visual situations," *Proceedings—12th IEEE International Conference on Semantic Computing, ICSC 2018*, vol. 2018-Jan., pp. 172–179, 2018.

[76] B. Barz and J. Denzler, "Hierarchy-based image embeddings for semantic image retrieval," in *Proceedings—2019 IEEE Winter Conference on Applications of Computer Vision, WACV 2019*, pp. 638–647, 2019.

[77] G. A. Miller, *WordNet: An Electronic Lexical Database*. Cambridge, MA: MIT Press, 1998.

[78] J. Lee, D. Kim, J. Ponce, and B. Ham, "Sfnet: Learning object-aware semantic correspondence," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2278–2287, 2019.

[79] K. Li, Y. Zhang, K. Li, Y. Li, and Y. Fu, "Visual semantic reasoning for image-text matching," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 4654–4662, 2019.

[80] S. Ren, K. He, and R. Girshick, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Advances in Neural Information Processing Systems*, pp. 1–9, 2015.

[81] M. Engilberge, L. Chevallier, P. Perez, and M. Cord, "Finding beans in burgers: Deep semantic-visual embedding with localization," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 3984–3993, 2018.

[82] X. Chang, Z. Ma, Y. Yang, Z. Zeng, and A. G. Hauptmann, "Bi-level semantic representation analysis for multimedia event detection," *IEEE Transactions on Cybernetics*, vol. 47, no. 5, pp. 1180–1197, 2017.

[83] D. G. Lowe, “Object recognition from local scale-invariant features,” in *Proceedings of the Seventh IEEE International Conference on Computer Vision*, IEEE, vol. 2, pp. 1150–1157, 1999.
[84] H. Bay, T. Tuytelaars, and L. Van Gool, “Surf: Speeded up robust features,” in *European Conference on Computer Vision*, Springer, pp. 404–417, 2006.
[85] R. Girshick, “Fast R-CNN,” in *Proceedings of the IEEE International Conference on Computer Vision*, vol. 2015, pp. 1440–1448, 2015.
[86] K. He, X. Zhang, S. Ren, and J. Sun, “Spatial pyramid pooling in deep convolutional networks for visual recognition,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 9, pp. 1904–1916, 2015.
[87] C. Cadena, A. Dick, and I. Reid, “A fast, modular scene understanding system using context-aware object detection,” in *Proc. International Conf. on Robotics and Automation*, 2015.
[88] M. Simon, K. Amende, A. Kraus, J. Honer, T. Samann, H. Kaulbersch, S. Milz, and H. Michael Gross, “Complexer-YOLO: Real-time 3D object detection and tracking on semantic point clouds,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2019.
[89] R. Yao, G. Lin, C. Shen, Y. Zhang, and Q. Shi, “Semantics-aware visual object tracking,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 29, no. 6, pp. 1687–1700, 2019.
[90] Z. Zhang, S. Qiao, C. Xie, W. Shen, B. Wang, and A. L. Yuille, “Single-shot object detection with enriched semantics,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 5813–5821, 2018.
[91] Q. Li, X. Zhao, R. He, and K. Huang, “Visual-semantic graph reasoning for pedestrian attribute recognition,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, pp. 8634–8641, 2019.

[92] S. Zakharov, W. Kehl, A. Bhargava, and A. Gaidon, "Autolabeling 3D objects with differentiable rendering of sdf shape priors," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 12 224–12 233, 2020.

[93] L. Lucchese and S. K. Mitra, "Colour image segmentation: A state-of-the-art survey," in *Proceedings-Indian National Science Academy Part A*, vol. 67, no. 2, pp. 207–222, 2001.

[94] L. Vincent and P. Soille, "Watersheds in digital spaces: An efficient algorithm based on immersion simulations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, no. 6, pp. 583–598, 1991.

[95] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 603–619, 2002.

[96] Y. Schnitman, Y. Caspi, D. Cohen-Or, and D. Lischinski, "Inducing semantic segmentation from an example," in *Asian Conference on Computer Vision*, Springer, pp. 373–384, 2006.

[97] T. Athanasiadis, P. Mylonas, Y. Avrithis, and S. Kollias, "Semantic image segmentation and object labeling," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 3, pp. 298–312, 2007.

[98] V. Lempitsky, A. Vedaldi, and A. Zisserman, "Pylon model for semantic segmentation," in *Advances in Neural Information Processing Systems*, pp. 1485–1493, 2011.

[99] B. Fulkerson, A. Vedaldi, and S. Soatto, "Class segmentation and object localization with superpixel neighborhoods," in *2009 IEEE 12th International Conference on Computer Vision*, IEEE, pp. 670–677, 2009.

[100] J. Shotton, J. Winn, C. Rother, and A. Criminisi, "Textonboost: Joint appearance, shape and context modeling for multi-class object recognition and segmentation," in *European Conference on Computer Vision*, Springer, pp. 1–15, 2006.

[101] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 11, pp. 1222–1239, 2001.

[102] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, pp. 640–651, 2015.

[103] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," in *ICLR*, 2016.

[104] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 12, pp. 2481–2495, 2017.

[105] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Cham, Switzerland: Springer, pp. 234–241, 2015.

[106] A. Paszke, A. Chaurasia, S. Kim, and E. Culurciello, "Enet: A deep neural network architecture for real-time semantic segmentation," *arXiv preprint arXiv:1606.02147*, 2016.

[107] L. C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 834–848, 2018.

[108] T. Takikawa, D. Acuna, V. Jampani, and S. Fidler, "Gated-SCNN: Gated shape CNNs for semantic segmentation," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 5229–5238, 2019.

[109] C. Sakaridis, D. Dai, and L. Van Gool, "Semantic foggy scene understanding with synthetic data," *International Journal of Computer Vision*, vol. 126, no. 9, pp. 973–992, 2018. DOI: 10.1007/s11263-018-1072-8.

[110] Ö. Erkent and C. Laugier, "Semantic segmentation with unsupervised domain adaptation under varying weather conditions for autonomous vehicles," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 3580–3587, 2020.

[111] A. Kundu, V. Vineet, and V. Koltun, "Feature space optimization for semantic video segmentation," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2016-Dec., pp. 3168–3175, 2016.

[112] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, "The cityscapes dataset for semantic urban scene understanding," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 29, pp. 270–275, 2016.

[113] G. Guarino, T. Chateau, C. Teuliere, and V. Antoine, "Temporal information integration for video semantic segmentation," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 2020.

[114] Y. He, W. C. Chiu, M. Keuper, and M. Fritz, "STD2P: RGBD semantic segmentation using spatio-temporal data-driven pooling," in *Proceedings—30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, vol. 2017-Jan., pp. 7158–7167, 2017.

[115] R. S. A. M. P. Lottes and C. S. M. B. T. Schultz, "Gradient and log-based active learning for semantic segmentation of crop and weed for agricultural robots," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 2020.

[116] F. Perazzi, J. Pont-Tuset, B. McWilliams, L. Van Gool, M. Gross, and A. Sorkine-Hornung, "A benchmark dataset and evaluation methodology for video object segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 724–732, 2016.

[117] S. Caelles, K.-K. Maninis, J. Pont-Tuset, L. Leal-Taixé, D. Cremers, and L. Van Gool, "One-shot video object segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 221–230, 2017.

[118] W. Wang, J. Shen, and L. Shao, "Video salient object detection via fully convolutional networks," *IEEE Transactions on Image Processing*, vol. 27, no. 1, pp. 38–49, 2017.

[119] W. Wang, J. Shen, R. Yang, and F. Porikli, "Saliency-aware video object segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 1, pp. 20–33, 2017.

[120] F. Perazzi, A. Khoreva, R. Benenson, B. Schiele, and A. Sorkine-Hornung, "Learning video object segmentation from static images," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2663–2672, 2017.

[121] D. Pathak, R. Girshick, P. Dollár, T. Darrell, and B. Hariharan, "Learning features by watching objects move," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2701–2710, 2017.

[122] D. Held, D. Guillory, B. Rebsamen, S. Thrun, and S. Savarese, "A probabilistic framework for real-time 3D segmentation using spatial, temporal, and semantic cues," in *Robotics: Science and Systems*, vol. 12, 2016.

[123] K. He, G. Gkioxari, P. Dollar, and R. Girshick, "Mask R-CNN," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2961–2969, 2017.

[124] M. Bai and R. Urtasun, "Deep watershed transform for instance segmentation," in *Proceedings—30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, pp. 5221–5229, 2017.

[125] A. W. Harley, K. G. Derpanis, and I. Kokkinos, "Segmentation-aware convolutional networks using local attention masks," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 5038–5047, 2017.

[126] A. Newell, Z. Huang, and J. Deng, "Associative embedding: End-to-end learning for joint detection and grouping," in *Advances in Neural Information Processing Systems*, pp. 2277–2287, 2017.

[127] Z. Huang, L. Huang, Y. Gong, C. Huang, and X. Wang, "Mask scoring R-CNN," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 6409–6418, 2019.

[128] A. Kirillov, K. He, R. Girshick, C. Rother, and P. Dollár, "Panoptic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 9404–9413, 2019.

[129] A. Kirillov, R. Girshick, K. He, and P. Dollár, "Panoptic feature pyramid networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 6399–6408, 2019.

[130] Y. Xiong, R. Liao, H. Zhao, R. Hu, M. Bai, E. Yumer, and R. Urtasun, "Upsnet: A unified panoptic segmentation network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 8818–8826, 2019.

[131] J. Li, A. Bhargava, A. R. R. Knohr, and A. D. Gaidon, "Fusing predictions for end-to-end panoptic segmentation," Mar. 12, US Patent App. 16/125,529, 2020.

[132] R. Hou, J. Li, A. Bhargava, A. Raventos, V. Guizilini, C. Fang, J. Lynch, and A. Gaidon, "Real-time panoptic segmentation from dense detections," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8523–8532, 2020.

[133] B. Cheng, M. D. Collins, Y. Zhu, T. Liu, T. S. Huang, H. Adam, and L.-C. Chen, "Panoptic-deeplab: A simple, strong, and fast baseline for bottom-up panoptic segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 12 475–12 485, 2020.

[134] H. Wang, Y. Zhu, B. Green, H. Adam, A. Yuille, and L.-C. Chen, "Axial-deeplab: Stand-alone axial-attention for panoptic segmentation," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2020.

[135] N. Silberman, D. Hoiem, P. Kohli, and R. Fergus, "Indoor segmentation and support inference from RGBD images," in *European Conference on Computer Vision*, pp. 746–760, Berlin and Heidelberg, Germany: Springer, 2012.

[136] S. Gupta, P. Arbeláez, R. Girshick, and J. Malik, "Indoor scene understanding with RGB-D images: Bottom-up segmentation, object detection and semantic segmentation," *International Journal of Computer Vision*, vol. 112, no. 2, pp. 133–149, 2015.

[137] C. Cadena and I. Reid, "Multi-modal auto-encoders as joint estimators for robotics scene understanding," in *Robotics: Science and Systems*, 2016.

[138] J. Ngiam, A. Khosla, M. Kim, J. Nam, H. Lee, and A. Y. Ng, "Multimodal deep learning," in *ICML*, 2011.

[139] A. Valada, G. L. Oliveira, T. Brox, and W. Burgard, "Deep multispectral semantic scene understanding of forested environments using multimodal fusion," in *International Symposium on Experimental Robotics*, Cham, Switzerland: Springer, pp. 465–477, 2016.

[140] Y. Chen, W. Li, X. Chen, and L. V. Gool, "Learning semantic segmentation from synthetic data: A geometrically guided input-output adaptation approach," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1841–1850, 2019.

[141] K. Wong, S. Wang, M. Ren, M. Liang, and R. Urtasun, "Identifying unknown instances for autonomous driving," in *Conference on Robot Learning*, pp. 384–393, 2020.

[142] M. Bai, W. Luo, K. Kundu, and R. Urtasun, "Exploiting semantic information and deep matching for optical flow," in *European Conference on Computer Vision*, vol. 9910 LNCS, pp. 154–170, 2016.

[143] G. Yang, H. Zhao, J. Shi, Z. Deng, and J. Jia, "SegStereo: Exploiting semantic information for disparity estimation," *Eccv*, pp. 1–16, 2018.

[144] P. L. Dovesi, M. Poggi, L. Andraghetti, M. Martí, H. Kjellström, A. Pieropan, and S. Mattoccia, "Real-time semantic stereo matching," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 2020.

[145] Z. Wu, X. Wu, X. Zhang, S. Wang, and L. Ju, "Semantic stereo matching with pyramid cost volumes," *Iccv*, pp. 7484–7493, 2019.

[146] V. Guizilini, R. Hou, J. Li, R. Ambrus, and A. Gaidon, "Semantically-guided representation learning for self-supervised monocular depth," in *2020 International Conference on Learning Representations (ICLR)*, 2020.

[147] D. Neven, B. De Brabandere, S. Georgoulis, M. Proesmans, and L. Van Gool, "Fast scene understanding for autonomous driving," *Proceedings DLVP2017*, pp. 1–5, 2017.

[148] I. Kokkinos, "UberNet: Training a universal convolutional neural network for low-, mid-, and high-level vision using diverse datasets and limited memory," in *Proceedings—30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, vol. 99, pp. 290–303, 2017.

[149] V. Nekrasov, T. Dharmasiri, A. Spek, T. Drummond, C. Shen, and I. Reid, "Real-time joint semantic segmentation and depth estimation using asymmetric annotations," in *Proceedings—IEEE International Conference on Robotics and Automation*, vol. 2019-May, pp. 7101–7107, 2019.

[150] J. McCormac, A. Handa, A. Davison, and S. Leutenegger, "SemanticFusion: Dense 3D semantic mapping with convolutional neural networks," in *Proceedings—IEEE International Conference on Robotics and Automation,* IEEE, pp. 4628–4635, Jul. 2017.

[151] K. Yang, L. M. Bergasa, E. Romera, X. Huang, and K. Wang, "Predicting polarization beyond semantics for wearable robotics," vol. 2018 Nov., in *IEEE-RAS International Conference on Humanoid Robots*, pp. 96–103, 2019.

[152] H. Badino, U. Franke, and D. Pfeiffer, "The stixel world—A compact medium level representation of the 3d-world," *Joint Pattern Recognition Symposium*, pp. 51–60, 2009.

[153] L. Schneider, M. Cordts, T. Rehfeld, D. Pfeiffer, M. Enzweiler, U. Franke, M. Pollefeys, and S. Roth, "Semantic stixels: Depth is not enough," in *IEEE Intelligent Vehicles Symposium, Proceedings*, vol. 2016-Aug., no. Iv, pp. 110–117, 2016.

[154] A. W. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 12, pp. 1349–1380, 2000.

[155] S. Santini, A. Gupta, and R. Jain, "A user interface for emergent semantics in image databases," in *Database Semantics*, Boston, MA: Springer, 1999, pp. 123–143.

[156] W. T. Freeman and E. H. Adelson, "Steerable filters for early vision, image analysis, and wavelet decomposition," in *[1990] Proceedings Third International Conference on Computer Vision*, IEEE, pp. 406–415, 1990.

[157] G. Carneiro, A. B. Chan, P. J. Moreno, and N. Vasconcelos, "Supervised learning of semantic classes for image annotation and retrieval," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 3, pp. 394–410, 2007.

[158] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "Slic superpixels compared to state-of-the-art superpixel methods," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 11, pp. 2274–2282, 2012.

[159] C. Schmid and R. Mohr, "Local grayvalue invariants for image retrieval," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 5, pp. 530–535, 1997.

[160] M. J. Swain and D. H. Ballard, "Indexing via color histograms," in *Active Perception and Robot Vision*, Springer, 1992, pp. 261–273.

[161] J. Sivic and A. Zisserman, "Video google: A text retrieval approach to object matching in videos," in *Proceedings 9th IEEE International Conference on Computer Vision*, vol. 2, pp. 1470–1477, 2003. DOI: 10.1109/ICCV.2003.1238663.

[162] B. Williams, M. Cummins, J. Neira, P. Newman, I. Reid, and J. Tardós, "A comparison of loop closing techniques in monocular SLAM," *Robotics and Autonomous Systems*, vol. 57, no. 12, pp. 1188–1197, 2009.

[163] D. Lin, S. Fidler, and R. Urtasun, "Holistic scene understanding for 3D object detection with RGBD cameras," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1417–1424, 2013.

[164] J. Johnson, R. Krishna, M. Stark, L. J. Li, D. A. Shamma, M. S. Bernstein, and F. F. Li, "Image retrieval using scene graphs," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 07-12-Jun., pp. 3668–3678, 2015.

[165] S. Schuster, R. Krishna, A. Chang, L. Fei-Fei, and C. D. Manning, "Generating semantically precise scene graphs from textual descriptions for improved image retrieval," in *Proceedings of the Fourth Workshop on Vision and Language*, pp. 70–80, 2015.

[166] D. Xu, Y. Zhu, C. B. Choy, and L. Fei-Fei, "Scene graph generation by iterative message passing," in *Proceedings—30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, vol. 2017-Jan., pp. 3097–3106, 2017.

[167] Y. Li, W. Ouyang, B. Zhou, K. Wang, and X. Wang, "Scene graph generation from objects, phrases and region captions," in *Proceedings of the IEEE International Conference on Computer Vision*, vol. 2017-Oct., pp. 1270–1279, 2017.

[168] R. Zellers, M. Yatskar, S. Thomson, and Y. Choi, "Neural motifs: Scene graph parsing with global context," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 5831–5840, 2018.

[169] R. Herzig, M. Raboh, G. Chechik, J. Berant, and A. Globerson, "Mapping images to scene graphs with permutation-invariant structured prediction," in *Advances in Neural Information Processing Systems*, pp. 7211–7221, 2018.

[170] J. Johnson, A. Gupta, and L. Fei-Fei, "Image generation from scene graphs," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1912–1228, 2018.

[171] O. Ashual and L. Wolf, "Specifying object attributes and relations in interactive scene generation," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 4561–4569, 2019.

[172] V. S. Chen, P. Varma, R. Krishna, M. Bernstein, C. Re, and L. Fei-Fei, "Scene graph prediction with limited labels," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2580–2590, 2019.

[173] J. Gu, H. Zhao, Z. Lin, S. Li, J. Cai, and M. Ling, "Scene graph generation with external knowledge and image reconstruction," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1969–1978, 2019.

[174] Y. Xian, C. H. Lampert, B. Schiele, and Z. Akata, "Zero-shot learning—A comprehensive evaluation of the good, the bad and the ugly," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 9, pp. 2251–2265, 2018.

[175] C. H. Lampert, H. Nickisch, and S. Harmeling, "Attribute-based classification for zero-shot visual object categorization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 3, pp. 453–465, 2013.

[176] M. Palatucci, D. Pomerleau, G. Hinton, and T. M. Mitchell, "Zero-shot learning with semantic output codes," in *Advances in Neural Information Processing Systems 22—Proceedings of the 2009 Conference*, pp. 1410–1418, 2009.

[177] Z. Zhang and V. Saligrama, "Zero-shot learning via semantic similarity embedding," in *Proceedings of the IEEE International Conference on Computer Vision*, vol. 2015 Inter, pp. 4166–4174, 2015.

[178] L. Chen, H. Zhang, J. Xiao, W. Liu, and S. F. Chang, "Zero-shot visual recognition using semantics-preserving adversarial embedding networks," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 1043–1052, 2018.

[179] V. Kumar Verma, G. Arora, A. Mishra, and P. Rai, "Generalized zero-shot learning via synthesized examples," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4281–4289, 2018.

[180] R. Felix, V. B. Kumar, I. Reid, and G. Carneiro, "Multi-modal cycle-consistent generalized zero-shot learning," in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 21–37, 2018.

[181] Y. Atzmon and G. Chechik, "Adaptive confidence smoothing for generalized zero-shot learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 11 671–11 680, 2019.

[182] R. Felix, B. Harwood, M. Sasdelli, and G. Carneiro, "Generalised zero-shot learning with domain classification in a joint semantic and visual space," in *2019 Digital Image Computing: Techniques and Applications (DICTA)*, IEEE, pp. 1–8, 2019.

[183] L. Niu, J. Cai, A. Veeraraghavan, and L. Zhang, "Zero-shot learning via category-specific visual-semantic mapping and label refinement," *IEEE Transactions on Image Processing*, vol. 28, no. 2, pp. 965–979, 2019.

[184] M. R. Vyas, H. Venkateswara, and S. Panchanathan, "Leveraging seen and unseen semantic relationships for generative zero-shot learning," in *European Conference on Computer Vision*, Cham, Switzerland: Springer, pp. 70–86, 2020.

[185] R. A. Epstein, E. Z. Patai, J. B. Julian, and H. J. Spiers, "The cognitive map in humans: Spatial navigation and beyond," *Nature Neuroscience*, vol. 20, no. 11, p. 1504, 2017.

[186] R. A. Brooks, "A robot that walks; emergent behaviors from a carefully evolved network," *Neural Computation*, vol. 1, no. 2, pp. 253–262, 1989.

[187] V. Braitenberg, *Vehicles: Experiments in Synthetic Psychology*. Cambridge, MA: MIT Press, 1986.

[188] P. Maes and R. A. Brooks, "Learning to coordinate behaviors," in *AAAI*, vol. 90, pp. 796–802, 1990.

[189] R. Zlot and M. Bosse, "Efficient large-scale 3D mobile mapping and surface reconstruction of an underground mine," in *Field and Service Robotics*, Berlin and Heidelberg, Germany: Springer, pp. 479–493, 2014.

[190] R. W. Wolcott and R. M. Eustice, "Visual localization within lidar maps for automated urban driving," in *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, IEEE, pp. 176–183, 2014.

[191] J. Guivant, E. Nebot, J. Nieto, and F. Masson, "Navigation and mapping in large unstructured environments," *The International Journal of Robotics Research*, vol. 23, no. 4–5, pp. 449–472, 2004.

[192] M. Schreiber, C. Knoppel, and U. Franke, "LaneLoc: Lane marking based localization using highly accurate maps," in *IEEE Intelligent Vehicles Symposium, Proceedings*, IEEE, pp. 449–454, 2013.

[193] A. Jacobson, F. Zeng, D. Smith, N. Boswell, T. Peynot, and M. Milford, "What localizes beneath: A metric multisensor localization and mapping system for autonomous underground mining vehicles," *Journal of Field Robotics*, 2020.

[194] A. Jain, S. Casas, R. Liao, Y. Xiong, S. Feng, S. Segal, and R. Urtasun, "Discrete residual flow for probabilistic pedestrian behavior prediction," in *Conference on Robot Learning*, PMLR, pp. 407–419, 2020.

[195] D. Frossard, E. Kee, and R. Urtasun, "Deepsignals: Predicting intent of drivers through visual signals," in *2019 International Conference on Robotics and Automation (ICRA)*, IEEE, pp. 9697–9703, 2019.

[196] N. Homayounfar, W.-C. Ma, J. Liang, X. Wu, J. Fan, and R. Urtasun, "Dagmapper: Learning to map by discovering lane topology," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2911–2920, 2019.

[197] A. Rosinol, M. Abate, Y. Chang, and L. Carlone, "Kimera: An open-source library for real-time metric-semantic localization and mapping," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, pp. 1689–1696, 2020.

[198] R. F. Salas-Moreno, R. A. Newcombe, H. Strasdat, P. H. Kelly, and A. J. Davison, "SLAM++: Simultaneous localisation and mapping at the level of objects," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 1352–1359, 2013.

[199] N. Suenderhauf, T. T. Pham, Y. Latif, M. Milford, and I. Reid, "Meaningful maps with object-oriented semantic mapping," in *IEEE International Conference on Intelligent Robots and Systems,* IEEE, vol. 2017-Sep., pp. 5079–5085, Dec. 2017.

[200] A. Gawel, C. Del Don, R. Siegwart, J. Nieto, and C. Cadena, “X-view: Graph-based semantic multiview localization,” *IEEE Robotics and Automation Letters*, vol. 3, no. 3, pp. 1687–1694, 2018.

[201] S. Garg, N. Suenderhauf, and M. Milford, “LoST? Appearance-invariant place recognition for opposite viewpoints using visual semantics,” in *Proceedings of Robotics: Science and Systems XIV*, 2018.

[202] A. Pronobis and P. Jensfelt, “Large-scale semantic mapping and reasoning with heterogeneous modalities,” in *Proceedings—IEEE International Conference on Robotics and Automation*, pp. 3515–3522, 2012.

[203] A. Rosinol, A. Gupta, M. Abate, J. Shi, and L. Carlone, “3D dynamic scene graphs: Actionable spatial perception with places, objects, and humans,” in *Robotics: Science and Systems*, 2020.

[204] A. J. Davison, “Real-time simultaneous localisation and mapping with a single camera,” in *International Conference on Computer Vision*, 2003.

[205] A. J. Davison, I. D. Reid, N. D. Molton, and O. Stasse, “Monoslam: Real-time single camera SLAM,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 6, pp. 1052–1067, 2007.

[206] G. Klein and D. Murray, “Parallel tracking and mapping on a camera phone,” in *2009 8th IEEE International Symposium on Mixed and Augmented Reality*, IEEE, pp. 83–86, 2009.

[207] C. Mei, G. Sibley, M. Cummins, P. Newman, and I. D. Reid, “A constant-time efficient stereo SLAM system,” in *British Machine Vision Conference*, 2009.

[208] H. Strasdat, J. M. M. Montiel, and A. J. Davison, “Visual SLAM: Why filter?” *Image and Vision Computing*, vol. 30, no. 2, pp. 65–77, 2012.

[209] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardos, “ORB-SLAM: A versatile and accurate monocular SLAM system,” *IEEE Transactions on Robotics*, vol. 31, no. 5, pp. 1147–1163, 2015.

[210] R. A. Newcombe, S. J. Lovegrove, and A. J. Davison, "DTAM: Dense tracking and mapping in real-time," in *2011 International Conference on Computer Vision*, IEEE, pp. 2320–2327, 2011.

[211] J. Engel, J. Sturm, and D. Cremers, "Semi-dense visual odometry for a monocular camera," in *International Conference on Computer Vision*, pp. 1449–1456, 2013.

[212] C. Forster, M. Pizzoli, and D. Scaramuzza, "SVO: Fast semi-direct monocular visual odometry," in *International Conference on Robotics and Automation*, pp. 15–22, 2014.

[213] J. Engel, V. Koltun, and D. Cremers, "Direct sparse odometry," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 3, pp. 611–625, 2018.

[214] J. Geng, "Structured-light 3d surface imaging: A tutorial," *Adv. Opt. Photon.*, vol. 3, no. 2, pp. 128–160, 2011.

[215] P. Henry, M. Krainin, E. Herbst, X. Ren, and D. Fox, "RGB-D mapping: Using depth cameras for dense 3D modeling of indoor environments," in *Experimental Robotics—The 12th International Symposium on Experimental Robotics, ISER 2010,* Dec. 18–21, 2010, New Delhi and Agra, India, O. Khatib, V. Kumar, and G. S. Sukhatme, Eds., ser. Springer Tracts in Advanced Robotics, vol. 79, pp. 477–491, Berlin and Heidelberg, Germany: Springer, 2010. DOI: 10.1007/978-3-642-28572-1_33.

[216] R. A. Newcombe, S. Izadi, O. Hilliges, D. Molyneaux, D. Kim, A. J. Davison, K. P, J. Shotton, S. Hodges, and A. Fitzgibbon, "Kinectfusion: Real-time dense surface mapping and tracking," in *International Symposium on Mixed and Augmented Reality*, 2011.

[217] F. Endres, J. Hess, N. Engelhard, J. Sturm, D. Cremers, and W. Burgard, "An evaluation of the RGB-D SLAM system," in *IEEE International Conference on Robotics and Automation, ICRA 2012,* May 14–18, 2012, St. Paul, Minnesota, USA, IEEE, pp. 1691–1696, 2012. DOI: 10.1109/ICRA.2012.6225199.

[218] O. Kähler, V. A. Prisacariu, C. Y. Ren, X. Sun, P. H. S. Torr, and D. W. Murray, "Very high frame rate volumetric integration of depth images on mobile devices," *IEEE Trans. Vis. Comput. Graph.*, vol. 21, no. 11, pp. 1241–1250, 2015. DOI: 10.1109/TVCG.2015.2459891.

[219] O. Kähler, V. A. Prisacariu, and D. W. Murray, "Real-time large-scale dense 3D reconstruction with loop closure," in *Computer Vision—ECCV 2016—14th European Conference, Proceedings, Part VIII,* Oct. 11–14, 2016, Amsterdam, The Netherlands, pp. 500–516, 2016.

[220] R. Brooks, "Visual map making for a mobile robot," in *Proceedings. 1985 IEEE International Conference on Robotics and Automation*, IEEE, vol. 2, pp. 824–829, 1985.

[221] D. Kortenkamp and T. Weymouth, "Topological mapping for mobile robots using a combination of sonar and vision sensing," in *AAAI*, vol. 94, pp. 979–984, 1994.

[222] G. Giralt, R. Chatila, and M. Vaisset, "An integrated navigation and motion control system for autonomous multisensory mobile robots," in *Autonomous Robot Vehicles*, New York: Springer, 1983, pp. 420–443.

[223] A. Kosaka and A. C. Kak, "Fast vision-guided mobile robot navigation using model-based reasoning and prediction of uncertainties," *CVGIP: Image Understanding*, vol. 56, no. 3, pp. 271–329, 1992.

[224] C. Fennema and A. R. Hanson, "Experiments in autonomous navigation," in *[1990] Proceedings. 10th International Conference on Pattern Recognition*, IEEE, vol. 1, pp. 24–31, 1990.

[225] H. Choset and K. Nagatani, "Topological simultaneous localization and mapping (SLAM): Toward exact localization without explicit localization," *IEEE Transactions on Robotics and Automation*, vol. 17, no. 2, pp. 125–137, 2001.

[226] A. Levin and R. Szeliski, "Visual odometry and map correlation," in *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004*, IEEE, vol. 1, 2004.

[227] P. Newman and K. Ho, "SLAM-loop closing with visually salient features," in *Proceedings of the 2005 IEEE International Conference on Robotics and Automation*, IEEE, pp. 635–642, 2005.

[228] S. Se, D. G. Lowe, and J. J. Little, "Vision-based global localization and mapping for mobile robots," *IEEE Transactions on Robotics*, vol. 21, no. 3, pp. 364–375, 2005.

[229] M. Cummins and P. Newman, "Fab-map: Probabilistic localization and mapping in the space of appearance," *The International Journal of Robotics Research*, vol. 27, no. 6, pp. 647–665, 2008.

[230] E. Garcia-Fidalgo and A. Ortiz, "Vision-based topological mapping and localization methods: A survey," *Robotics and Autonomous Systems*, vol. 64, pp. 1–20, 2015.

[231] S. Lowry, N. Sünderhauf, P. Newman, J. J. Leonard, D. Cox, P. Corke, and M. J. Milford, "Visual place recognition: A survey," *IEEE Transactions on Robotics*, vol. 32, no. 1, pp. 1–19, Feb. 2016.

[232] B. Kuipers and Y.-T. Byun, "A robot exploration and mapping strategy based on a semantic hierarchy of spatial representations," *Robotics and Autonomous Systems*, vol. 8, no. 1–2, pp. 47–63, 1991.

[233] S. Thrun, "Learning metric-topological maps for indoor mobile robot navigation," *Artificial Intelligence*, vol. 99, no. 1, pp. 21–71, 1998.

[234] S. Simhon and G. Dudek, "A global topological map formed by local metric maps," in *Proceedings. 1998 IEEE/RSJ International Conference on Intelligent Robots and Systems. Innovations in Theory, Practice and Applications (Cat. No. 98CH36190)*, IEEE, vol. 3, pp. 1708–1714, 1998.

[235] M. Bosse, P. Newman, J. Leonard, M. Soika, W. Feiten, and S. Teller, "An atlas framework for scalable mapping," in *2003 IEEE International Conference on Robotics and Automation (Cat. No. 03CH37422)*, IEEE, vol. 2, pp. 1899–1906, 2003.

[236] N. Tomatis, I. Nourbakhsh, and R. Siegwart, "Hybrid simultaneous localization and map building: A natural integration of topological and metric," *Robotics and Autonomous Systems*, vol. 44, no. 1, pp. 3–14, 2003.

[237] C. Mei, G. Sibley, and P. Newman, "Closing loops without places," in *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, IEEE, pp. 3738–3744, 2010.

[238] L. Riazuelo, M. Tenorth, D. Di Marco, M. Salas, D. Gálvez-López, L. Mösenlechner, L. Kunze, M. Beetz, J. D. Tardós, L. Montano, and J. M. Montiel, "RoboEarth semantic mapping: A cloud enabled knowledge-based approach," *IEEE Transactions on Automation Science and Engineering*, vol. 12, no. 2, pp. 432–443, 2015.

[239] D. Kochanov, A. Ošep, J. Stückler, and B. Leibe, "Scene flow propagation for semantic mapping and object discovery in dynamic street scenes," in *IEEE International Conference on Intelligent Robots and Systems*, vol. 2016-Nov., no. 3, pp. 1785–1792, 2016.

[240] Y. Yue, C. Zhao, R. Li, C. Yang, J. Zhang, M. Wen, Y. Wang, and D. Wang, "A hierarchical framework for collaborative probabilistic semantic mapping," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 2020.

[241] A. Rottmann, Ó. M. Mozos, C. Stachniss, and W. Burgard, "Semantic place classification of indoor environments with mobile robots using boosting," in *Proceedings of the National Conference on Artificial Intelligence*, vol. 3, no. 1998, pp. 1306–1311, 2005.

[242] M. Luperto and F. Amigoni, "Exploiting structural properties of buildings towards general semantic mapping systems," *Advances in Intelligent Systems and Computing*, vol. 302, pp. 375–387, 2016.

[243] Y. Freund and R. E. Schapire, "A desicion-theoretic generalization of on-line learning and an application to boosting," in *European Conference on Computational Learning Theory*, Springer, pp. 23–37, 1995.

[244] C. Stachniss, Ó. M. Mozos, and W. Burgard, "Speeding-up multi-robot exploration by considering semantic place information," in *Proceedings—IEEE International Conference on Robotics and Automation*, vol. 2006, no. May, pp. 1692–1697, 2006.

[245] R. Goeddel and E. Olson, "Learning semantic place labels from occupancy grids using CNNs," *IEEE International Conference on Intelligent Robots and Systems*, vol. 2016-Nov., pp. 3999–4004, 2016.

[246] N. Suenderhauf, F. Dayoub, S. McMahon, B. Talbot, R. Schulz, P. Corke, G. Wyeth, B. Upcroft, and M. Milford, "Place categorization and semantic mapping on a mobile robot," in *Proceedings—IEEE International Conference on Robotics and Automation*, vol. 2016-Jun., pp. 5729–5736, 2016.

[247] Y. Liao, S. Kodagoda, Y. Wang, L. Shi, and Y. Liu, "Understand scene categories by objects: A semantic regularized scene classifier using convolutional neural networks," in *Proceedings—IEEE International Conference on Robotics and Automation*, vol. 2016-Jun., pp. 2318–2325, 2016.

[248] C. Premebida, D. R. Faria, and U. Nunes, "Dynamic Bayesian network for semantic place classification in mobile robotics," *Autonomous Robots*, vol. 41, no. 5, pp. 1161–1172, 2017.

[249] C. Premebida, D. R. Faria, F. A. Souza, and U. Nunes, "Applying probabilistic mixture models to semantic place classification in mobile robotics," in *IEEE International Conference on Intelligent Robots and Systems*, vol. 2015-Dec., no. Oct., pp. 4265–4270, 2015.

[250] M. Mancini, S. R. Bulò, E. Ricci, and B. Caputo, "Learning deep NBNN representations for robust place categorization," *IEEE Robotics and Automation Letters*, vol. 2, no. 3, pp. 1794–1801, 2017.

[251] M. Mancini, S. R. Bulo, B. Caputo, and E. Ricci, "Robust place categorization with deep domain generalization," *IEEE Robotics and Automation Letters*, vol. 3, no. 3, pp. 2093–2100, 2018.

[252] K. Zheng and A. Pronobis, "From pixels to buildings: End-to-end probabilistic deep networks for large-scale semantic mapping," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, pp. 3511–3518, 2019.

[253] A. Ranganathan, "PLISS: Labeling places using online changepoint detection," *Autonomous Robots*, vol. 32, no. 4, pp. 351–368, 2012.

[254] A. Ranganathan and J. Lim, "Visual place categorization in maps," in *IEEE International Conference on Intelligent Robots and Systems*, pp. 3982–3989, 2011.

[255] E. Fazl-Ersi and J. K. Tsotsos, "Histogram of oriented uniform patterns for robust place recognition and categorization," *International Journal of Robotics Research*, vol. 31, no. 4, pp. 468–483, 2012.

[256] J. Stückler and S. Behnke, "Model learning and real-time tracking using multi-resolution surfel maps," in *Twenty-Sixth AAAI Conference on Artificial Intelligence*, 2012.

[257] Z. Wei, W. Chen, and J. Wang, "3D semantic map-based shared control for smart wheelchair," in *International Conference on Intelligent Robotics and Applications*, vol. 7507 LNAI, no. Part 2, pp. 41–51, 2012.

[258] M. Gunther, T. Wiemann, S. Albrecht, and J. Hertzberg, "Building semantic object maps from sparse and noisy 3D data," in *IEEE International Conference on Intelligent Robots and Systems*, no. 287752, pp. 2228–2233, 2013.

[259] S. Vasudevan and R. Siegwart, "Bayesian space conceptualization and place classification for semantic maps in mobile robotics," *Robotics and Autonomous Systems*, vol. 56, no. 6, pp. 522–537, 2008.

[260] A. Dame, V. A. Prisacariu, C. Y. Ren, and I. Reid, "Dense reconstruction using 3D object shape priors," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1288–1295, 2013.

[261] K. Tateno, F. Tombari, and N. Navab, "When 2.5D is not enough: Simultaneous reconstruction, segmentation and recognition on dense SLAM," in *Proceedings—IEEE International Conference on Robotics and Automation*, vol. 2016-Jun., pp. 2295–2302, 2016.

[262] J. Dong, X. Fei, and S. Soatto, "Visual-inertial-semantic scene representation for 3D object detection," in *Proceedings—30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, vol. 2017-Jan., pp. 3567–3577, 2017.

[263] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 779–788, 2016.

[264] J. G. Rogers and H. I. Christensen, "A conditional random field model for place and object classification," in *Proceedings—IEEE International Conference on Robotics and Automation*, pp. 1766–1772, 2012.

[265] T. T. Pham, M. Eich, I. Reid, and G. Wyeth, "Geometrically consistent plane extraction for dense indoor 3D maps segmentation," in *IEEE International Conference on Intelligent Robots and Systems*, vol. 2016-Nov., pp. 4199–4204, 2016.

[266] J. McCormac, R. Clark, M. Bloesch, A. Davison, and S. Leutenegger, "Fusion++: Volumetric object-level SLAM," in *Proceedings—2018 International Conference on 3D Vision, 3DV 2018*, pp. 32–41, 2018.

[267] W. J. Scheirer, A. de Rezende Rocha, A. Sapkota, and T. E. Boult, "Toward open set recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 7, pp. 1757–1772, 2012.

[268] F. Li and H. Wechsler, "Open set face recognition using transduction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 11, pp. 1686–1697, 2005.

[269] P. J. Phillips, P. Grother, and R. Micheals, "Evaluation methods in face recognition," in *Handbook of Face Recognition*, London: Springer, pp. 551–574, 2011.

[270] W. J. Scheirer, L. P. Jain, and T. E. Boult, "Probability models for open set recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 11, pp. 2317–2324, 2014.

[271] A. Bendale and T. Boult, "Towards open world recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1893–1902, 2015.

[272] A. Bendale and T. E. Boult, "Towards open set deep networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1563–1572, 2016.

[273] F. Furrer, T. Novkovic, M. Fehr, A. Gawel, M. Grinvald, T. Sattler, R. Siegwart, and J. Nieto, "Incremental object database: Building 3D models from multiple partial observations," in *IEEE International Conference on Intelligent Robots and Systems*, pp. 6835–6842, 2018.

[274] M. Grinvald, F. Furrer, T. Novkovic, J. J. Chung, C. Cadena Lerma, R. Siegwart, and J. Nieto, "Volumetric instance-aware semantic mapping and 3D object discovery," *IEEE Robotics and Automation Letters*, vol. 4, no. 3, pp. 3037–3044, 2019.

[275] K. Tateno, F. Tombari, and N. Navab, "Real-time and scalable incremental segmentation on dense SLAM," in *IEEE International Conference on Intelligent Robots and Systems*, vol. 2015-Dec., pp. 4465–4472, 2015.

[276] V. Vineet, O. Miksik, M. Lidegaard, M. Nießner, S. Golodetz, V. A. Prisacariu, O. Kähler, D. W. Murray, S. Izadi, P. Pérez, and P. H. Torr, "Incremental dense semantic stereo fusion for large-scale semantic scene reconstruction," in *Proceedings—IEEE International Conference on Robotics and Automation*, vol. 2015-Jun., no. Jun., pp. 75–82, 2015.

[277] Y. Nakajima and H. Saito, "Efficient object-oriented semantic mapping with object detector," *IEEE Access*, vol. 7, pp. 3206–3213, 2019.

[278] Q. H. Pham, B. S. Hua, D. T. Nguyen, and S. K. Yeung, "Real-time progressive 3D semantic segmentation for indoor scenes," in *Proceedings—2019 IEEE Winter Conference on Applications of Computer Vision, WACV 2019*, pp. 1089–1098, 2019.

[279] M. Runz and L. Agapito, "Co-fusion: Real-time segmentation, tracking and fusion of multiple objects," in *Proceedings—IEEE International Conference on Robotics and Automation*, pp. 4471–4478, 2017.

[280] M. Runz, M. Buffier, and L. Agapito, "MaskFusion: Real-time recognition, tracking and reconstruction of multiple moving objects," in *Proceedings of the 2018 IEEE International Symposium on Mixed and Augmented Reality, ISMAR 2018*, pp. 10–20, Jan. 2019.

[281] Y. Di, H. Morimitsu, Z. Lou, and X. Ji, "A unified framework for piecewise semantic reconstruction in dynamic scenes via exploiting superpixel relations," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 2020.

[282] Z. S. Hashemifar, K. W. Lee, N. Napp, and K. Dantu, "Consistent cuboid detection for semantic mapping," in *Proceedings—IEEE 11th International Conference on Semantic Computing, ICSC 2017*, pp. 526–531, 2017.

[283] L. Nicholson, M. Milford, and N. Suenderhauf, "QuadricSLAM: Dual quadrics from object detections as landmarks in object-oriented SLAM," *IEEE Robotics and Automation Letters*, vol. 4, no. 1, pp. 1–8, 2019.

[284] M. Hosseinzadeh, Y. Latif, T. Pham, N. Suenderhauf, and I. Reid, "Structure aware SLAM using quadrics and planes," in *Asian Conference on Computer Vision*, Cham, Switzerland: Springer, pp. 410–426, 2018.

[285] M. Hosseinzadeh, K. Li, Y. Latif, and I. Reid, "Real-time monocular object-model aware sparse SLAM," in *2019 International Conference on Robotics and Automation (ICRA)*, IEEE, pp. 7123–7129, 2019.

[286] P. Drews, P. Núñez, R. Rocha, M. Campos, and J. Dias, "Novelty detection and 3D shape retrieval using superquadrics and multi-scale sampling for autonomous mobile robots," in *Proceedings—IEEE International Conference on Robotics and Automation*, pp. 3635–3640, 2010.

[287] P. Kaiser, M. Grotz, E. E. Aksoy, M. Do, N. Vahrenkamp, and T. Asfour, "Validation of whole-body loco-manipulation affordances for pushability and liftability," in *IEEE-RAS International Conference on Humanoid Robots*, vol. 2015-Dec., pp. 920–927, 2015.

[288] S. Blumenthal, H. Bruyninckx, W. Nowak, and E. Prassler, "A scene graph based shared 3D world model for robotic applications," in *Proceedings—IEEE International Conference on Robotics and Automation*, pp. 453–460, 2013.

[289] E. E. Aksoy, A. Abramov, F. Wörgötter, and B. Dellen, "Categorizing object-action relations from semantic scene graphs," in *Proceedings—IEEE International Conference on Robotics and Automation*, pp. 398–405, 2010.

[290] M. Grotz, P. Kaiser, E. E. Aksoy, F. Paus, and T. Asfour, "Graph-based visual semantic perception for humanoid robots," in *IEEE-RAS International Conference on Humanoid Robots*, pp. 869–875, 2017.

[291] U.-H. Kim, J.-M. Park, T.-J. Song, and J.-H. Kim, "3-D scene graph: A sparse and semantic representation of physical environments for intelligent agents," *IEEE Transactions on Cybernetics*, pp. 1–13, 2019.

[292] Z. Zeng, Z. Zhou, Z. Sui, and O. C. Jenkins, "Semantic robot programming for goal-directed manipulation in cluttered scenes," in *Proceedings—IEEE International Conference on Robotics and Automation*, pp. 7462–7469, 2018.

[293] D. Shuey, D. Bailey, and T. P. Morrissey, "PHIGS: A standard, dynamic, interactive graphics interface," *IEEE Computer Graphics and Applications*, vol. 6, no. 8, pp. 50–57, 1986.

[294] S. Blumenthal and H. Bruyninckx, "Towards a domain specific language for a scene graph based robotic world model," in *Proceedings of the Fourth International Workshop on Domain-Specific Languages and Models for Robotic Systems (DSLRob)*, 2013. [Online]. Available: http://arxiv.org/abs/1408.0200.

[295] I. Bozcan and S. Kalkan, "COSMO: Contextualized scene modeling with Boltzmann machines," *Robotics and Autonomous Systems*, vol. 113, pp. 132–148, 2019.

[296] B. Pan, H. Cai, D.-A. Huang, K.-H. Lee, A. Gaidon, E. Adeli, and J. C. Niebles, "Spatio-temporal graph for video captioning with knowledge distillation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10 870–10 879, 2020.

[297] A. Flint, C. Mei, D. Murray, and I. Reid, "Growing semantically meaningful models for visual SLAM," in *Proc. IEEE Computer Vision and Pattern Recognition*, 2010.

[298] A. Flint, C. Mei, I. Reid, and D. Murray, "A dynamic programming approach to reconstructing building interiors," in *Proc. 11th European Conf. on Computer*, 2010.

[299] A. Flint, I. Reid, and D. Murray, "Manhattan scene understanding using monocular, stereo, and 3d features," in *Proc. 13th IEEE Int. Conf. on Computer Vision*, 2011.

[300] A. Kundu, Y. Li, F. Dellaert, F. Li, and J. M. Rehg, "Joint semantic segmentation and 3D reconstruction from monocular video," in *European Conference on Computer Vision*, pp. 703–718, 2014.

[301] A. Hermans, G. Floros, and B. Leibe, "Dense 3D semantic mapping of indoor scenes from RGB-D images," in *Proceedings—IEEE International Conference on Robotics and Automation,* IEEE, pp. 2631–2638, Sep. 2014.

[302] J. Stückler, B. Waldvogel, H. Schulz, and S. Behnke, "Dense real-time mapping of object-class semantics from RGB-D video," *Journal of Real-Time Image Processing*, vol. 10, no. 4, pp. 599–609, 2015.

[303] X. Li and R. Belaroussi, "Semi-dense 3d semantic mapping from monocular slam," *arXiv preprint arXiv:1611.04144*, 2016.

[304] T. Whelan, S. Leutenegger, R. F. Salas-Moreno, B. Glocker, and A. J. Davison, "ElasticFusion: Dense SLAM without a pose graph," in *Robotics: Science and Systems*, vol. 11, 2015.

[305] Y. Xiang and D. Fox, "DA-RNN: Semantic mapping with data associated recurrent neural networks," in *Robotics: Science and Systems*, vol. 13, 2017.

[306] S. Yang, Y. Huang, and S. Scherer, "Semantic 3D occupancy mapping through efficient high order CRFs," in *IEEE International Conference on Intelligent Robots and Systems*, vol. 2017-Sep., pp. 590–597, 2017.

[307] C. Hazirbas, L. Ma, C. Domokos, and D. Cremers, "FuseNet: Incorporating depth into semantic segmentation via fusion-based CNN architecture," in *Asian Conference on Computer Vision*, pp. 213–218, 2016. [Online]. Available: http://link.springer.com/10.1007/978-3-319-54181-5.

[308] S. Gupta, R. Girshick, P. Arbeláez, and J. Malik, “Learning rich features from RGB-D images for object detection and segmentation,” in *European Conference on Computer Vision*, Cham, Switzerland: Springer, pp. 345–360, 2014.
[309] L. Ma, J. Stuckler, C. Kerl, and D. Cremers, “Multi-view deep learning for consistent semantic mapping with RGB-D cameras,” in *IEEE International Conference on Intelligent Robots and Systems*, 2017.
[310] J. Engel, T. Schöps, and D. Cremers, “LSD-SLAM: Large-scale direct monocular SLAM,” in *European Conference on Computer Vision*, Cham, Switzerland: Springer, pp. 834–849, 2014.
[311] C. Weerasekera, Y. Latif, R. Garg, and I. Reid, “Dense monocular reconstruction using surface normals,” in *Proc. International Conference on Robotics and Automation*, 2017.
[312] A. Nüchter and J. Hertzberg, “Towards semantic maps for mobile robots,” *Robotics and Autonomous Systems*, vol. 56, no. 11, pp. 915–926, 2008.
[313] X. Xiong and D. Huber, “Using context to create semantic 3D models of indoor environments,” in *British Machine Vision Conference, BMVC 2010—Proceedings*, pp. 1–11, 2010.
[314] L. Morreale, A. Romanoni, M. Matteucci, and P. di Milano, “Dense 3D visual mapping via semantic simplification,” in *2019 International Conference on Robotics and Automation (ICRA)*, IEEE, pp. 6891–6897, 2019.
[315] Z. Zhang, “Microsoft kinect sensor and its effect,” *IEEE Multimedia*, vol. 19, no. 2, pp. 4–10, 2012.
[316] D. Wolf, J. Prankl, and M. Vincze, “Fast semantic segmentation of 3D point clouds using a dense CRF with learned parameters,” in *Proceedings—IEEE International Conference on Robotics and Automation*, vol. 2015-Jun., no. Jun., pp. 4867–4873, 2015.
[317] D. Maturana and S. Scherer, “Voxnet: A 3d convolutional neural network for real-time object recognition,” in *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, pp. 922–928, 2015.

[318] H. Riemenschneider, A. Bódis-Szomorú, J. Weissenberg, and L. Van Gool, "Learning where to classify in multi-view semantic segmentation," in *European Conference on Computer Vision*, Cham, Switzerland: Springer, 2014.

[319] I. Armeni, O. Sener, A. R. Zamir, H. Jiang, I. Brilakis, M. Fischer, and S. Savarese, "3D semantic parsing of large-scale indoor spaces," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 1534–1543, 2016.

[320] M. Bell and D. Gausebeck, "Capturing and aligning multiple 3-dimensional scenes," Nov. 4, US Patent 8,879,828, Nov. 2014.

[321] M. De Deuge, A. Quadros, C. Hung, and B. Douillard, "Unsupervised feature learning for classification of outdoor 3d scans," in *Australasian Conference on Robitics and Automation*, vol. 2, p. 1, 2013.

[322] Z. Wu, S. Song, A. Khosla, F. Yu, L. Zhang, X. Tang, and J. Xiao, "3d shapenets: A deep representation for volumetric shapes," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1912–1920, 2015.

[323] R. Socher, B. Huval, B. Bath, C. D. Manning, and A. Y. Ng, "Convolutional-recursive deep learning for 3d object classification," in *Advances in Neural Information Processing Systems*, pp. 656–664, 2012.

[324] A. Quadros, J. P. Underwood, and B. Douillard, "An occlusion-aware feature for range images," in *2012 IEEE International Conference on Robotics and Automation*, IEEE, pp. 4428–4435, 2012.

[325] S. Song and J. Xiao, "Sliding shapes for 3d object detection in RGB-D images," in *European Conference on Computer Vision*, vol. 2, p. 7, 2014.

[326] L. A. Alexandre, "3D object recognition using convolutional neural networks with transfer learning between input channels," in *Intelligent Autonomous Systems 13*, Cham, Switzerland: Springer, 2016, pp. 889–898.

[327] I. Lenz, H. Lee, and A. Saxena, "Deep learning for detecting robotic grasps," *The International Journal of Robotics Research*, vol. 34, no. 4–5, pp. 705–724, 2015.

[328] B. Li, T. Zhang, and T. Xia, "Vehicle detection from 3d lidar using fully convolutional network," in *Robotics: Science and Systems*, 2016.

[329] N. Höft, H. Schulz, and S. Behnke, "Fast semantic segmentation of RGB-D scenes with gpu-accelerated deep neural networks," in *Joint German/Austrian Conference on Artificial Intelligence (Künstliche Intelligenz)*, Cham, Switzerland: Springer, pp. 80–85, 2014.

[330] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Computation*, vol. 18, no. 7, pp. 1527–1554, 2006.

[331] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D u-net: Learning dense volumetric segmentation from sparse annotation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Cham, Switzerland: Springer, pp. 424–432, 2016.

[332] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," in *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[333] C. R. Qi, H. Su, M. Nießner, A. Dai, M. Yan, and L. J. Guibas, "Volumetric and multi-view cnns for object classification on 3d data," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5648–5656, 2016.

[334] L. Tchapmi, C. Choy, I. Armeni, J. Gwak, and S. Savarese, "Segcloud: Semantic segmentation of 3d point clouds," in *2017 International Conference on 3D Vision (3DV)*, IEEE, pp. 537–547, 2017.

[335] S. Song, F. Yu, A. Zeng, A. X. Chang, M. Savva, and T. Funkhouser, "Semantic scene completion from a single depth image," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1746–1754, 2017.

[336] A. Dai, C. R. Qi, and M. Nießner, "Shape completion using 3D-encoder-predictor CNNs and shape synthesis," in *Proceedings—30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, vol. 2017-Jan., pp. 6545–6554, 2017.

[337] A. Dai, D. Ritchie, M. Bokeloh, S. Reed, J. Sturm, and M. Niebner, "ScanComplete: Large-scale scene completion and semantic segmentation for 3D scans," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 4578–4587, 2018.

[338] G. Riegler, A. O. Ulusoy, and A. Geiger, "OctNet: Learning deep 3D representations at high resolutions deep learning for 3D data shape classification," *Cvpr*, 2017.

[339] D. Meagher, "Geometric modeling using octree encoding," *Computer Graphics and Image Processing*, vol. 19, no. 2, pp. 129–147, 1982.

[340] B. Graham, "Sparse 3D convolutional neural networks," in *Proceedings of the British Machine Vision Conference (BMVC)*, pp. 150.1–150.9, BMVA Press, Sep. 2015. DOI: 10.5244/C.29.150.

[341] M. Engelcke, D. Rao, D. Z. Wang, C. H. Tong, and I. Posner, "Vote3deep: Fast object detection in 3d point clouds using efficient convolutional neural networks," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, pp. 1355–1361, 2017.

[342] B. Graham, M. Engelcke, and L. van der Maaten, "3d semantic segmentation with submanifold sparse convolutional networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 9224–9232, 2018.

[343] B. Graham, "Spatially-sparse convolutional neural networks," in *BMVA Symposium on Deep Learning for Computer Vision*, 2016.

[344] D. Z. Wang and I. Posner, "Voting for voting in online point cloud object detection," in *Robotics: Science and Systems*, vol. 1, 2015.

[345] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, pp. 315–323, 2011.

[346] V. Jampani, M. Kiefel, and P. V. Gehler, "Learning sparse high dimensional filters: Image filtering, dense crfs and bilateral neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4452–4461, 2016.

[347] Y. Li, S. Pirk, H. Su, C. R. Qi, and L. J. Guibas, "FPNN: Field probing neural networks for 3d data," in *Advances in Neural Information Processing Systems*, pp. 307–315, 2016.

[348] H. Su, S. Maji, E. Kalogerakis, and E. Learned-Miller, "Multi-view convolutional neural networks for 3d shape recognition," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 945–953, 2015.

[349] M. Tatarchenko, J. Park, V. Koltun, and Q.-Y. Zhou, "Tangent convolutions for dense prediction in 3d," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3887–3896, 2018.

[350] R. Klokov and V. Lempitsky, "Escape from cells: Deep kd-networks for the recognition of 3d point cloud models," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 863–872, 2017.

[351] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "PointNet++: Deep hierarchical feature learning on point sets in a metric space," in *Advances in Neural Information Processing Systems.* Nips, 2017.

[352] F. Engelmann, T. Kontogianni, A. Hermans, and B. Leibe, "Exploring spatial context for 3D semantic segmentation of point clouds," in *Proceedings—2017 IEEE International Conference on Computer Vision Workshops, ICCVW 2017*, vol. 2018-Jan., pp. 716–724, 2018.

[353] B.-S. Hua, M.-K. Tran, and S.-K. Yeung, "Pointwise convolutional neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 984–993, 2018.

[354] W. Wu, Z. Qi, and L. Fuxin, "Pointconv: Deep convolutional networks on 3d point clouds," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 9621–9630, 2019.

[355] X. Ye, J. Li, H. Huang, L. Du, and X. Zhang, "3d recurrent neural networks with context fusion for point cloud semantic segmentation," in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 403–417, 2018.

[356] Q. Huang, W. Wang, and U. Neumann, "Recurrent slice networks for 3d segmentation of point clouds," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2626–2635, 2018.

[357] G. Te, W. Hu, A. Zheng, and Z. Guo, "Rgcnn: Regularized graph cnn for point cloud segmentation," in *Proceedings of the 26th ACM International Conference on Multimedia*, pp. 746–754, 2018.

[358] Y. Wang, Y. Sun, Z. Liu, S. E. Sarma, M. M. Bronstein, and J. M. Solomon, "Dynamic graph CNN for learning on point clouds," *ACM Transactions on Graphics (TOG)*, vol. 38, no. 5, pp. 1–12, 2019.

[359] Y. Yang, C. Feng, Y. Shen, and D. Tian, "Foldingnet: Point cloud auto-encoder via deep grid deformation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 206–215, 2018.

[360] M. Zamorski, M. Zikeba, P. Klukowski, R. Nowak, K. Kurach, W. Stokowiec, and T. Trzcinski, "Adversarial autoencoders for compact representations of 3D point clouds," *Computer Vision and Image Understanding*, vol. 193, Art. no. 102921, 2020.

[361] M. Simonovsky and N. Komodakis, "Dynamic edge-conditioned filters in convolutional neural networks on graphs," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3693–3702, 2017.

[362] X. Qi, R. Liao, J. Jia, S. Fidler, and R. Urtasun, "3D graph neural networks for RGBD semantic segmentation," in *Proceedings of the IEEE International Conference on Computer Vision*, vol. 2018-Jul., pp. 726–732, 2017.

[363] W. Zeng and T. Gevers, “3dcontextnet: Kd tree guided hierarchical learning of point clouds using local and global contextual cues,” in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018.

[364] L. Landrieu and M. Simonovsky, “Large-scale point cloud semantic segmentation with superpoint graphs,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4558–4567, 2018.

[365] J. L. Bentley, “Multidimensional binary search trees used for associative searching,” *Communications of the ACM*, vol. 18, no. 9, pp. 509–517, 1975.

[366] J. Li, B. M. Chen, and G. Hee Lee, “So-net: Self-organizing network for point cloud analysis,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 9397–9406, 2018.

[367] T. Kohonen, “The self-organizing map,” in *Proceedings of the IEEE*, vol. 78, no. 9, pp. 1464–1480, 1990.

[368] R. Mur-Artal and J. D. Tardós, “ORB-SLAM2: An open-source SLAM system for monocular, stereo, and RGB-D cameras,” *IEEE Transactions on Robotics*, vol. 33, no. 5, pp. 1255–1262, 2017.

[369] J. L. Schonberger and J.-M. Frahm, “Structure-from-motion revisited,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4104–4113, 2016.

[370] C. Kong and S. Lucey, “Deep non-rigid structure from motion,” in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1558–1567, 2019.

[371] S. Vijayanarasimhan, S. Ricco, C. Schmid, R. Sukthankar, and K. Fragkiadaki, “SFM-net: Learning of structure and motion from video,” *arXiv preprint arXiv:1704.07804*, 2017.

[372] K. Simonyan and A. Zisserman, “Two-stream convolutional networks for action recognition in videos,” in *Advances in Neural Processing Systems*, pp. 568–576, 2014.

[373] I. Laptev, M. Marszalek, C. Schmid, and B. Rozenfeld, "Learning realistic human actions from movies," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8, 2008.

[374] M. D. Rodriguez, J. Ahmed, and M. Shah, "Action MACH: A spatio-temporal maximum average correlation height filter for action recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8, 2008.

[375] K. Soomro, A. R. Zamir, and M. Shah, "UCF101: A dataset of 101 human actions classes from videos in the wild," *CoRR*, 2012. [Online]. Available: http://arxiv.org/abs/1212.0402.

[376] J. Carreira and A. Zisserman, "Quo vadis, action recognition? A new model and the kinetics dataset," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 6299–6308, Jul. 2017.

[377] G. Gkioxari and J. Malik, "Finding action tubes," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 759–768, 2015.

[378] G. B. Fabian Caba Heilbron, E. Victor, and J. C. Niebles, "ActivityNet: A large-scale video benchmark for human activity understanding," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 961–970, 2015.

[379] D. Damen, H. Doughty, G. M. Farinella, S. Fidler, A. Furnari, E. Kazakos, D. Moltisanti, J. Munro, T. Perrett, W. Price, and M. Wray, "Scaling egocentric vision: The EPIC-kitchens dataset," in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 720–736, 2018.

[380] H. Wang, A. Kläser, C. Schmid, and C. Liu, "Action recognition by dense trajectories," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3169–3176, 2011.

[381] A. Gaidon, Z. Harchaoui, and C. Schmid, "Actom sequence models for efficient action detection," in *CVPR 2011*, IEEE, pp. 3201–3208, 2011.

[382] A. Gaidon, Z. Harchaoui, and C. Schmid, "Activity representation with motion hierarchies," *International Journal of Computer Vision*, vol. 107, no. 3, pp. 219–238, 2014.

[383] B. Fernando, E. Gavves, J. Oramas, A. Ghodrati, and T. Tuytelaars, “Rank pooling for action recognition,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016.

[384] B. Fernando and S. Gould, “Learning end-to-end video classification with rank-pooling,” in *Proceedings of the International Conference on Machine Learning*, pp. 1187–1196, 2016.

[385] B. Fernando, P. Anderson, M. Hutter, and S. Gould, “Discriminative hierarchical rank pooling for activity recognition,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1924–1932, 2016.

[386] H. Bilen, B. Fernando, E. Gavves, A. Vedaldi, and S. Gould, “Dynamic image networks for action recognition,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3034–3042, 2016.

[387] D. Ramanan and D. A. Forsyth, “Automatic annotation of everyday movements,” in *Advances in Neural Processing Systems*, 2003.

[388] C. Wang, Y. Wang, and A. L. Yuille, “An approach to pose-based action recognition,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 915–922, 2013.

[389] D. C. Luvizon, D. Picard, and H. Tabia, “2d/3d pose estimation and action recognition using multitask deep learning,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5137–5146, 2018.

[390] S. Park and J. K. Aggarwal, “Semantic-level understanding of human actions and interactions using event hierarchy,” *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, vol. 2004-Jan., no. Jan., 2004,

[391] H. T. Cheng, M. Griss, P. Davis, J. Li, and D. You, “Towards zero-shot learning for human activity recognition using semantic attribute sequence model,” in *UbiComp 2013—Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, no. 2, pp. 355–358, 2013.

[392] H. T. Cheng, F. T. Sun, M. Griss, P. Davis, J. Li, and D. You, "NuActiv: Recognizing unseen new activities using semantic attribute-based learning," in *MobiSys 2013—Proceedings of the 11th Annual International Conference on Mobile Systems, Applications, and Services*, pp. 361–374, 2013.

[393] K. Ramirez-Amaro, M. Beetz, and G. Cheng, "Automatic segmentation and recognition of human activities from observation based on semantic reasoning," in *IEEE International Conference on Intelligent Robots and Systems*, no. Iros, pp. 5043–5048, 2014.

[394] K. Ramirez-Amaro, T. Inamura, E. Dean-León, M. Beetz, and G. Cheng, "Bootstrapping humanoid robot skills by extracting semantic representations of human-like activities from virtual reality," in *IEEE-RAS International Conference on Humanoid Robots*, vol. 2015-Feb., pp. 438–443, 2015.

[395] K. Ramirez-Amaro, E. Dean-Leon, and G. Cheng, "Robust semantic representations for inferring human co-manipulation activities even with different demonstration styles," in *IEEE-RAS International Conference on Humanoid Robots*, vol. 2015-Dec., pp. 1141–1146, 2015.

[396] K. Ramirez-Amaro, E. S. Kim, J. Kim, B. T. Zhang, M. Beetz, and G. Cheng, "Enhancing human action recognition through spatio-temporal feature learning and semantic rules," in *IEEE-RAS International Conference on Humanoid Robots*, vol. 2015-Feb., no. Feb., pp. 456–461, 2015.

[397] K. Ramirez-Amaro, M. Beetz, and G. Cheng, "Transferring skills to humanoid robots by extracting semantic representations from observations of human activities," *Artificial Intelligence*, vol. 247, pp. 95–118, 2017. DOI: 10.1016/j.artint.2015.08.009.

[398] K. Ramirez-Amaro, H. N. Minhas, M. Zehetleitner, M. Beetz, and G. Cheng, "Added value of gaze-exploiting semantic representation to allow robots inferring human behaviors," *ACM Transactions on Interactive Intelligent Systems*, vol. 7, no. 1, pp. 1–30, 2017.

[399] K. Ramirez-Amaro, E. Dean-Leon, F. Bergner, and G. Cheng, “A semantic-based method for teaching industrial robots new tasks,” *KI—Kunstliche Intelligenz*, vol. 33, no. 2, pp. 117–122, 2019. DOI: 10.1007/s13218-019-00582-5.

[400] J. S. Yoon, Z. Li, and H. S. Park, “3D semantic trajectory reconstruction from 3D pixel continuum,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 5060–5069, 2018.

[401] H. J. Chang, T. Fischer, M. Petit, M. Zambelli, and Y. Demiris, “Kinematic structure correspondences via hypergraph matching,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4216–4225, 2016.

[402] H. J. Chang, T. Fischer, M. Petit, M. Zambelli, and Y. Demiris, “Learning kinematic structure correspondences using multi-order similarities,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 12, pp. 2920–2934, 2018.

[403] P. Wang, L. Sun, A. F. Smeaton, C. Gurrin, and S. Yang, “Computer vision for lifelogging: Characterizing everyday activities based on visual semantics,” *Computer Vision For Assistive Healthcare*, pp. 250–282, 2018.

[404] J. J. Gibson, *The Ecological Approach to Visual Perception: Classic Edition*. New York and London: Psychology Press, 2014. First published in 1979.

[405] P. Fitzpatrick, G. Metta, L. Natale, S. Rao, and G. Sandini, “Learning about objects through action-initial steps towards artificial cognition,” in *2003 IEEE International Conference on Robotics and Automation (Cat. No. 03CH37422)*, IEEE, vol. 3, pp. 3140–3145, 2003.

[406] E. Şahin, M. Çakmak, M. R. Doğar, E. Uğur, and G. Üçoluk, “To afford or not to afford: A new formalization of affordances toward affordance-based robot control,” *Adaptive Behavior*, vol. 15, no. 4, pp. 447–472, 2007.

[407] M. R. Dogar, M. Cakmak, E. Ugur, and E. Sahin, “From primitive behaviors to goal-directed behavior using affordances,” in *2007 IEEE/RSJ International Conference on Intelligent Robots and Systems*, IEEE, pp. 729–734, 2007.

[408] N. Krüger, C. Geib, J. Piater, R. Petrick, M. Steedman, F. Wörgötter, A. Ude, T. Asfour, D. Kraft, D. Omrčen, A. Agostini, and R. Dillmann, "Object–action complexes: Grounded abstractions of sensory–motor processes," *Robotics and Autonomous Systems*, vol. 59, no. 10, pp. 740–757, 2011.

[409] E. E. Aksoy, A. Abramov, J. Dörr, K. Ning, B. Dellen, and F. Wörgötter, "Learning the semantics of object-action relations by observation," *International Journal of Robotics Research*, vol. 30, no. 10, pp. 1229–1249, 2011.

[410] M. J. Aein, E. E. Aksoy, M. Tamosiunaite, J. Papon, A. Ude, and F. Worgotter, "Toward a library of manipulation actions based on semantic object-action relations," in *IEEE International Conference on Intelligent Robots and Systems*, pp. 4555–4562, 2013.

[411] E. E. Aksoy, M. Tamosiunaite, and F. Wörgötter, "Model-free incremental learning of the semantics of manipulation actions," *Robotics and Autonomous Systems*, vol. 71, pp. 118–133, 2015.

[412] Y. Yang, Y. Li, C. Fermuller, and Y. Aloimonos, "Robot learning manipulation action plans by 'watching' unconstrained videos from the world wide web," in *Twenty-Ninth AAAI Conference on Artificial Intelligence*, 2015.

[413] D. Kappler, L. Y. Chang, N. S. Pollard, T. Asfour, and R. Dillmann, "Templates for pre-grasp sliding interactions," *Robotics and Autonomous Systems*, vol. 60, no. 3, pp. 411–423, 2012.

[414] K. M. Varadarajan and M. Vincze, "Afrob: The affordance network ontology for robots," in *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, IEEE, pp. 1343–1350, 2012.

[415] M. A. Roa and R. Suárez, "Grasp quality measures: Review and performance," *Autonomous Robots*, vol. 38, no. 1, pp. 65–88, 2015.

[416] A. Bicchi and V. Kumar, "Robotic grasping and contact: A review," in *Proceedings 2000 ICRA. Millennium Conference. IEEE International Conference on Robotics and Automation. Symposia Proceedings (Cat. No. 00CH37065)*, IEEE, vol. 1, pp. 348–353, 2000.

[417] J. Bohg, A. Morales, T. Asfour, and D. Kragic, "Data-driven grasp synthesis—A survey," *IEEE Transactions on Robotics*, vol. 30, no. 2, pp. 289–309, 2013.

[418] R. Detry, D. Kraft, O. Kroemer, L. Bodenhagen, J. Peters, N. Krüger, and J. Piater, "Learning grasp affordance densities," *Paladyn, Journal of Behavioral Robotics*, vol. 2, no. 1, pp. 1–17, 2011.

[419] R. Detry, D. Kraft, A. G. Buch, N. Krüger, and J. Piater, "Refining grasp affordance models by experience," in *2010 IEEE International Conference on Robotics and Automation*, IEEE, pp. 2287–2293, 2010.

[420] A. Zeng, K. Yu, S. Song, D. Suo, E. Walker Jr., A. Rodriguez, and J. Xiao, "Multi-view self-supervised deep learning for 6D pose estimation in the Amazon picking challenge," in *ICRA*, 2017.

[421] A. Morales, T. Asfour, P. Azad, S. Knoop, and R. Dillmann, "Integrated grasp planning and visual object localization for a humanoid robot with five-fingered hands," in *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*, IEEE, pp. 5663–5668, 2006.

[422] J. Tremblay, T. To, B. Sundaralingam, Y. Xiang, D. Fox, and S. Birchfield, "Deep object pose estimation for semantic robotic grasping of household objects," in *Conference on Robot Learning (CoRL)*, 2018.

[423] C. Papazov, S. Haddadin, S. Parusel, K. Krieger, and D. Burschka, "Rigid 3D geometry matching for grasping of known objects in cluttered scenes," *The International Journal of Robotics Research*, vol. 31, no. 4, pp. 538–553, 2012.

[424] A. Collet, M. Martinez, and S. S. Srinivasa, "The moped framework: Object recognition and pose estimation for manipulation," *The International Journal of Robotics Research*, vol. 30, no. 10, pp. 1284–1306, 2011.

[425] H. Dang and P. K. Allen, "Semantic grasping: Planning robotic grasps functionally suitable for an object manipulation task," in *IEEE International Conference on Intelligent Robots and Systems*, pp. 1311–1317, 2012.

[426] U. Hillenbrand and M. A. Roa, "Transferring functional grasps through contact warping and local replanning," in *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, IEEE, pp. 2963–2970, 2012.

[427] R. Detry, C. H. Ek, M. Madry, J. Piater, and D. Kragic, "Generalizing grasps across partly similar objects," in *2012 IEEE International Conference on Robotics and Automation*, IEEE, pp. 3791–3797, 2012.

[428] R. Detry, C. H. Ek, M. Madry, and D. Kragic, "Learning a dictionary of prototypical grasp-predicting parts from grasping experience," in *2013 IEEE International Conference on Robotics and Automation*, IEEE, pp. 601–608, 2013.

[429] A. Ten Pas and R. Platt, "Localizing handle-like grasp affordances in 3d point clouds," in *Experimental Robotics*, Cham, Switzerland: Springer, pp. 623–638, 2016.

[430] H. O. Song, M. Fritz, D. Goehring, and T. Darrell, "Learning to detect visual grasp affordance," *IEEE Transactions on Automation Science and Engineering*, vol. 13, no. 2, pp. 798–809, 2015.

[431] L. Manuelli, W. Gao, P. Florence, and R. Tedrake, "kPAM: Keypoint affordances for category-level robotic manipulation," in *International Symposium on Robotics Research (ISRR)*, 2019.

[432] D. Morrison, P. Corke, and J. Leitner, "Closing the loop for robotic grasping: A real-time, generative grasp synthesis approach," in *Robotics: Science and Systems (RSS)*, 2018.

[433] D. Morrison, P. Corke, and J. Leitner, "Learning robust, real-time, reactive robotic grasping," *The International Journal of Robotics Research*, 2019.

[434] J. Mahler, J. Liang, S. Niyaz, M. Laskey, R. Doan, X. Liu, J. A. Ojea, and K. Goldberg, "Dex-net 2.0: Deep learning to plan robust grasps with synthetic point clouds and analytic grasp metrics," in *Robotics: Science and Systems (RSS)*, 2017.

[435] J. Mahler, M. Matl, V. Satish, M. Danielczuk, B. DeRose, S. McKinley, and K. Goldberg, "Learning ambidextrous robot grasping policies," *Science Robotics*, 2019.

[436] A. Ten Pas, M. Gualtieri, K. Saenko, and R. Platt, "Grasp pose detection in point clouds," *The International Journal of Robotics Research (IJRR)*, 2017.

[437] A. Billard and D. Kragic, "Trends and challenges in robot manipulation," *Science*, vol. 364, no. 6446, p. eaat8414, 2019.

[438] D. Morrison, A. W. Tow, M. Mctaggart, R. Smith, N. Kelly-Boxall, S. Wade-Mccue, J. Erskine, R. Grinover, A. Gurman, T. Hunn, D. Lee, A. Milan, T. Thanh Pham, G. Rallos, A. Razjigaev, T. Rowntree, V. Kumar, Z. Zhuang, C. Lehnert, I. Reid, P. Corke, and J. Leitner, "Cartman: The low-cost cartesian manipulator that won the amazon robotics challenge," in *International Conference on Robotics and Automation (ICRA)*, 2018.

[439] A. Zeng, S. Song, K.-T. Yu, E. Donlon, F. R. Hogan, M. Bauza, D. Ma, O. Taylor, M. Liu, E. Romo, N. Fazeli, F. Alet, N. Chavan Dafle, R. Holladay, I. Morona, P. Qu Nair, D. Green, I. Taylor, W. Liu, T. Funkhouser, and A. Rodriguez, "Robotic pick-and-place of novel objects in clutter with multi-affordance grasping and cross-domain image matching," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, pp. 1–8, 2018.

[440] M. Schwarz, C. Lenz, G. M. García, S. Koo, A. S. Periyasamy, M. Schreiber, and S. Behnke, "Fast object learning and dual-arm coordination for cluttered stowing, picking, and packing," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, pp. 3347–3354, 2018.

[441] D. Guo, T. Kong, F. Sun, and H. Liu, "Object discovery and grasp detection with a shared convolutional neural network," in *2016 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, pp. 2038–2043, 2016.

[442] E. Jang, S. Vijayanarasimhan, P. Pastor, J. Ibarz, and S. Levine, "End-to-end learning of semantic grasping," in *Conference on Robot Learning*, pp. 119–132, 2017.

[443] L. Verhagen, H. C. Dijkerman, W. P. Medendorp, and I. Toni, "Cortical dynamics of sensorimotor integration during grasp planning," *Journal of Neuroscience*, vol. 32, no. 13, pp. 4508–4519, 2012.

[444] A. Myers, C. L. Teo, C. Fermüller, and Y. Aloimonos, "Affordance detection of tool parts from geometric features," in *2015 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, pp. 1374–1381, 2015.

[445] A. Nguyen, D. Kanoulas, D. G. Caldwell, and N. G. Tsagarakis, "Detecting object affordances with convolutional neural networks," in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, pp. 2765–2770, 2016.

[446] M. Kokic, J. A. Stork, J. A. Haustein, and D. Kragic, "Affordance detection for task-specific grasping using deep learning," in *2017 IEEE-RAS 17th International Conference on Humanoid Robotics (Humanoids)*, IEEE, pp. 91–98, 2017.

[447] D. Song, C. H. Ek, K. Huebner, and D. Kragic, "Multivariate discretization for Bayesian network structure learning in robot grasping," in *2011 IEEE International Conference on Robotics and Automation*, IEEE, pp. 1944–1950, 2011.

[448] D. Song, C. H. Ek, K. Huebner, and D. Kragic, "Task-based robot grasp planning using probabilistic inference," *IEEE Transactions on Robotics*, vol. 31, no. 3, pp. 546–561, 2015.

[449] M. Hjelm, C. H. Ek, R. Detry, and D. Kragic, "Learning human priors for task-constrained grasping," in *International Conference on Computer Vision Systems*, Cham, Switzerland: Springer, pp. 207–217, 2015.

[450] P. Ardón, È. Pairet, S. Ramamoorthy, and K. S. Lohan, "Towards robust grasps: Using the environment semantics for robotic object affordances," in *Proceedings of the AAAI Fall Symposium on Reasoning and Learning in Real-World Systems for Long-Term Autonomy*, pp. 5–12, 2018.

[451] R. Detry, J. Papon, and L. Matthies, "Task-oriented grasping with semantic and geometric scene understanding," in *IEEE International Conference on Intelligent Robots and Systems*, vol. 2017-Sep., pp. 3266–3273, 2017.

[452] N. Blodow, L. C. Goron, Z.-C. Marton, D. Pangercic, T. Rühr, M. Tenorth, and M. Beetz, "Autonomous semantic mapping for robots performing everyday manipulation tasks in kitchen environments," in *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*, IEEE, pp. 4263–4270, 2011.
[453] A. Cosgun and H. I. Christensen, "Context-aware robot navigation using interactively built semantic maps," *Paladyn, Journal of Behavioral Robotics*, vol. 9, no. 1, pp. 254–276, 2018.
[454] C. Galindo and A. Saffiotti, "Inferring robot goals from violations of semantic knowledge," *Robotics and Autonomous Systems*, vol. 61, no. 10, pp. 1131–1143, 2013.
[455] T. Mar, V. Tikhanoff, and L. Natale, "What can i do with this tool? self-supervised learning of tool affordances from their 3-d geometry," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 10, no. 3, pp. 595–610, 2017.
[456] K. Fang, Y. Zhu, A. Garg, A. Kurenkov, V. Mehta, L. Fei-Fei, and S. Savarese, "Learning task-oriented grasping for tool manipulation from simulated self-supervision," *The International Journal of Robotics Research*, vol. 39, no. 2–3, pp. 202–216, 2020.
[457] S. Tellex, R. Knepper, A. Li, D. Rus, and N. Roy, "Asking for help using inverse semantics," in *Robotics: Science and Systems*, 2014.
[458] S. Tellex, T. Kollar, S. Dickerson, M. R. Walter, A. G. Banerjee, S. Teller, and N. Roy, "Understanding natural language commands for robotic navigation and mobile manipulation," in *Twenty-Fifth AAAI Conference on Artificial Intelligence*, 2011.
[459] Z. Gong and Y. Zhang, "Temporal spatial inverse semantics for robots communicating with humans," in *Proceedings—IEEE International Conference on Robotics and Automation*, pp. 4451–4458, 2018.
[460] M. Pomarlan and J. Bateman, "Robot program construction via grounded natural language semantics and simulation robotics track," in *Proceedings of the International Joint Conference on Autonomous Agents and Multiagent Systems, AAMAS*, vol. 2, pp. 857–864, 2018.

[461] J. Savage, D. A. Rosenblueth, M. Matamoros, M. Negrete, L. Contreras, J. Cruz, R. Martell, H. Estrada, and H. Okada, "Semantic reasoning in service robots using expert systems," *Robotics and Autonomous Systems*, vol. 114, pp. 77–92, 2019.

[462] S. Niekum, S. Chitta, A. G. Barto, B. Marthi, and S. Osentoski, "Incremental semantically grounded learning from demonstration," in *Robotics: Science and Systems*, Berlin, Germany, vol. 9, 2013.

[463] P. Sermanet, K. Xu, and S. Levine, "Unsupervised perceptual rewards for imitation learning," in *Robotics: Science and Systems*, 2017.

[464] S. Nikolaidis, R. Ramakrishnan, K. Gu, and J. Shah, "Efficient model learning from joint-action demonstrations for human–robot collaborative tasks," in *2015 10th ACM/IEEE International Conference on Human–Robot Interaction (HRI)*, pp. 189–196, 2015.

[465] S. Hemachandra, M. R. Walter, S. Tellex, and S. Teller, "Learning spatial-semantic representations from natural language descriptions and scene classifications," in *Proceedings—IEEE International Conference on Robotics and Automation*, pp. 2623–2630, 2014.

[466] W. Yang, X. Wang, A. Farhadi, A. Gupta, and R. Mottaghi, "Visual semantic navigation using scene priors," in *ICLR*, 2019.

[467] C. Stachniss, D. Hahnel, and W. Burgard, "Exploration with active loop-closing for fastSLAM," in *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (IEEE Cat. No. 04CH37566)*, IEEE, vol. 2, pp. 1505–1510, 2004.

[468] L. Carlone, J. Du, M. K. Ng, B. Bona, and M. Indri, "An application of kullback-leibler divergence to active SLAM and exploration with particle filters," in *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, IEEE, pp. 287–293, 2010.

[469] L. Carlone, J. Du, M. K. Ng, B. Bona, and M. Indri, "Active SLAM and exploration with particle filters using kullback-leibler divergence," *Journal of Intelligent and Robotic Systems*, vol. 75, no. 2, pp. 291–311, 2014.

[470] S. Lemaignan, M. Warnier, E. A. Sisbot, A. Clodic, and R. Alami, "Artificial cognition for social human–robot interaction: An implementation," *Artificial Intelligence*, vol. 247, pp. 45–69, 2017.

[471] E. A. Sisbot, L. F. Marin-Urias, R. Alami, and T. Simeon, "A human aware mobile robot motion planner," *IEEE Transactions on Robotics*, vol. 23, no. 5, pp. 874–883, 2007.

[472] H. Kidokoro, T. Kanda, D. Brščić, and M. Shiomi, "Simulation-based behavior planning to prevent congestion of pedestrians around a robot," *IEEE Transactions on Robotics*, vol. 31, no. 6, pp. 1419–1431, 2015.

[473] R. Schulz, A. Glover, M. J. Milford, G. Wyeth, and J. Wiles, "Lingodroids: Studies in spatial cognition and language," in *2011 IEEE International Conference on Robotics and Automation*, IEEE, pp. 178–183, 2011.

[474] T. Fischer and Y. Demiris, "Computational modelling of embodied visual perspective-taking," *IEEE Transactions on Cognitive and Developmental Systems*, 2019.

[475] T. Fischer and Y. Demiris, "Markerless perspective taking for humanoid robots in unconstrained environments," in *Proceedings of the IEEE International Conference on Robotics and Automation*, pp. 3309–3316, 2016.

[476] C. Breazeal, M. Berlin, A. Brooks, J. Gray, and A. L. Thomaz, "Using perspective taking to learn from ambiguous demonstrations," *Robotics and Autonomous Systems*, vol. 54, no. 5, pp. 385–393, 2006.

[477] P. Anderson, Q. Wu, D. Teney, J. Bruce, M. Johnson, N. Sünderhauf, I. D. Reid, S. Gould, and A. van den Hengel, "Vision-and-language navigation: Interpreting visually-grounded navigation instructions in real environments," in *CVPR*, pp. 3674–3683, 2018.

[478] D. Fried, R. Hu, V. Cirik, A. Rohrbach, J. Andreas, L. Morency, T. Berg-Kirkpatrick, K. Saenko, D. Klein, and T. Darrell, "Speaker-follower models for vision-and-language navigation," in *NeurIPS*, pp. 3318–3329, 2018.

[479] H. Tan, L. Yu, and M. Bansal, "Learning to navigate unseen environments: Back translation with environmental dropout," in *NAACL*, pp. 2610–2621, 2019.

[480] C.-Y. Ma, J. Lu, Z. Wu, G. AlRegib, Z. Kira, R. Socher, and C. Xiong, "Self-monitoring navigation agent via auxiliary progress estimation," in *ICLR*, 2019.

[481] C. Ma, Z. Wu, G. AlRegib, C. Xiong, and Z. Kira, "The regretful agent: Heuristic-aided navigation through progress estimation," in *CVPR*, pp. 6732–6740, 2019.

[482] L. Ke, X. Li, Y. Bisk, A. Holtzman, Z. Gan, J. Liu, J. Gao, Y. Choi, and S. S. Srinivasa, "Tactical rewind: Self-correction via backtracking in vision-and-language navigation," in *CVPR*, pp. 6741–6749, 2019.

[483] Y. Qi, Q. Wu, P. Anderson, X. Wang, W. Wang, C. Shen, and A. v. d. Hengel, "Reverie: Remote embodied visual referring expression in real indoor environments," in *CVPR*, 2020.

[484] S. Kim and I. S. Kweon, "Simultaneous place and object recognition using collaborative context information," *Image and Vision Computing*, vol. 27, no. 6, pp. 824–833, 2009.

[485] R. Luo, S. Piao, and H. Min, "Simultaneous place and object recognition with mobile robot using pose encoded contextual information," in *Proceedings—IEEE International Conference on Robotics and Automation*, pp. 2792–2797, 2011.

[486] A. Oliva and A. Torralba, "Building the gist of a scene: The role of global image features in recognition," *Progress in Brain Research*, vol. 155, pp. 23–36, 2006.

[487] N. Atanasov, M. Zhu, K. Daniilidis, and G. J. Pappas, "Localization from semantic observations via the matrix permanent," *International Journal of Robotics Research*, vol. 35, no. 1–3, pp. 73–99, 2016.

[488] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part-based models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 9, pp. 1627–1645, 2009.

[489] L. Weng, B. Soheilian, and V. Gouet-Brunet, "Semantic signatures for urban visual localization," in *Proceedings—International Workshop on Content-Based Multimedia Indexing*, IEEE, vol. 2018 -Sep., pp. 1–6, 2018.

[490] W. Zhang, G. Liu, and G. Tian, "A coarse to fine indoor visual localization method using environmental semantic information," *IEEE Access*, vol. 7, pp. 21 963–21 970, 2019.

[491] F. Han, S. E. Beleidy, H. Wang, C. Ye, and H. Zhang, "Learning of Holism-Landmark graph embedding for place recognition in long-term autonomy," *IEEE Robotics and Automation Letters*, vol. 3, no. 4, pp. 3669–3676, 2018.

[492] S. Ardeshir, A. R. Zamir, A. Torroella, and M. Shah, "GIS-assisted object detection and geospatial localization," in *European Conference on Computer Vision*, Cham, Switzerland: Springer, vol. 8694 LNCS, pp. 602–617, 2014.

[493] F. Castaldo, A. Zamir, R. Angst, F. Palmieri, and S. Savarese, "Semantic cross-view matching," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pp. 9–17, 2015.

[494] O. Pink, "Visual map matching and localization using a global feature map," in *2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, IEEE, pp. 1–7, 2008.

[495] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Journal of Basic Engineering*, vol. 82, no. 1, pp. 35–45, 1960.

[496] A. Armagan, M. Hirzer, P. M. Roth, and V. Lepetit, "Learning to align semantic segmentation and 2.5 d maps for geolocalization," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3425–3432, 2017.

[497] A. Mousavian and J. Kosecka, "Semantic image based geolocation given a map," George Mason University Fairfax United States, Tech. Rep., 2016. [Online]. Available: http://arxiv.org/abs/1609.00278.

[498] J.-I. Meguro, T. Murata, H. Nishimura, Y. Amano, T. Hasizume, and J.-I. Takiguchi, "Development of positioning technique using omni-directional IR camera and aerial survey data," in *2007 IEEE/ASME International Conference on Advanced Intelligent Mechatronics*, IEEE, pp. 1–6, 2007.

[499] J.-C. Bazin, I. Kweon, C. Demonceaux, and P. Vasseur, "Dynamic programming and skyline extraction in catadioptric infrared images," in *2009 IEEE International Conference on Robotics and Automation*, IEEE, pp. 409–416, 2009.

[500] T. Stone, M. Mangan, P. Ardin, and B. Webb, "Sky segmentation with ultraviolet images can be used for navigation," in *Robotics: Science and Systems*, 2014.

[501] T. Stone, D. Differt, M. Milford, and B. Webb, "Skyline-based localisation for aggressively manoeuvring robots using uv sensors and spherical harmonics," in *2016 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, pp. 5615–5622, 2016.

[502] S. Ramalingam, S. Bouaziz, P. Sturm, and M. Brand, "Skyline2gps: Localization in urban canyons using omni-skylines," in *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, IEEE, pp. 3816–3823, 2010.

[503] O. Saurer, G. Baatz, K. Köser, L. Ladický, and M. Pollefeys, "Image based geo-localization in the ALPs," *International Journal of Computer Vision*, vol. 116, no. 3, pp. 213–225, 2016.

[504] E. Pepperell, P. Corke, and M. Milford, "Routed roads: Probabilistic vision-based place recognition for changing conditions, split streets and varied viewpoints," *The International Journal of Robotics Research*, vol. 35, no. 9, pp. 1057–1179, 2016.

[505] N. Jacobs, S. Satkin, N. Roman, R. Speyer, and R. Pless, "Geolocating static cameras," in *2007 IEEE 11th International Conference on Computer Vision*, IEEE, pp. 1–6, 2007.

[506] X. Yu, S. Chaturvedi, C. Feng, Y. Taguchi, T. Y. Lee, C. Fernandes, and S. Ramalingam, "VLASE: Vehicle localization by aggregating semantic edges," in *IEEE International Conference on Intelligent Robots and Systems*, pp. 3196–3203, 2018.

[507] H. Jegou, F. Perronnin, M. Douze, J. Sánchez, P. Perez, and C. Schmid, "Aggregating local image descriptors into compact codes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 9, pp. 1704–1716, 2011.

[508] Z. Yu, C. Feng, M.-Y. Liu, and S. Ramalingam, "Casenet: Deep category-aware semantic edge detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5964–5973, 2017.

[509] G. Singh and J. Košecká, "Semantically guided geo-location and modeling in urban environments," in *Large-Scale Visual Geo-Localization*, Cham, Switzerland: Springer, 2016, pp. 101–120.

[510] X. Ren and J. Malik, "Learning a classification model for segmentation," in *Proceedings Ninth IEEE International Conference on Computer Vision*, vol. 1, pp. 10–17, 2003. DOI: 10.1109/ICCV.2003.1238308.

[511] S. Gould, R. Fulton, and D. Koller, "Decomposing a scene into geometric and semantically consistent regions," in *2009 IEEE 12th International Conference on Computer Vision*, IEEE, pp. 1–8, 2009.

[512] Y. Latif, R. Garg, M. Milford, and I. Reid, "Addressing challenging place recognition tasks using generative adversarial networks," in *ICRA*, 2018.

[513] H. Porav, W. Maddern, and P. Newman, "Adversarial training for adverse conditions: Robust metric localisation using appearance transfer," in *ICRA*, 2018.

[514] A. Anoosheh, T. Sattler, R. Timofte, M. Pollefeys, and L. Van Gool, "Night-to-day image translation for retrieval-based localization," in *ICRA*, 2019.

[515] W. Churchill and P. Newman, "Experience-based navigation for long-term localisation," *The International Journal of Robotics Research*, vol. 32, no. 14, pp. 1645–1661, 2013.

[516] A.-D. Doan, Y. Latif, T.-J. Chin, Y. Liu, T.-T. Do, and I. Reid, "Scalable place recognition under appearance change for autonomous driving," in *ICCV*, 2019.

[517] M. Dymczyk, S. Lynen, T. Cieslewski, M. Bosse, R. Siegwart, and P. Furgale, "The gist of maps-summarizing experience for lifelong localization," in *ICRA*, 2015.

[518] D. Anguelov, C. Dulong, D. Filip, C. Frueh, S. Lafon, R. Lyon, A. Ogale, L. Vincent, and J. Weaver, "Google street view: Capturing the world at street level," *Computer*, vol. 43, no. 6, pp. 32–38, 2010.

[519] J. Geyer, Y. Kassahun, M. Mahmudi, X. Ricou, R. Durgesh, A. S. Chung, L. Hauswald, V. H. Pham, M. Mühlegg, S. Dorn, T. Fernandez, M. Jänicke, S. Mirashi, C. Savani, M. Sturm, O. Vorobiov, M. Oelker, S. Garreis, and P. Schuberth, "A2d2: Audi autonomous driving dataset," *arXiv preprint arXiv:2004.06320*, 2020.

[520] J. Houston, G. Zuidhof, L. Bergamini, Y. Ye, A. Jain, S. Omari, V. Iglovikov, and P. Ondruska, "One thousand and one hours: Self-driving motion prediction dataset," *arXiv preprint arXiv:2006.14480*, 2020.

[521] F. Warburg, S. Hauberg, M. López-Antequera, P. Gargallo, Y. Kuang, and J. Civera, "Mapillary street-level sequences: A dataset for lifelong place recognition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2626–2635, 2020.

[522] N. Jacobs, W. Burgin, N. Fridrich, A. Abrams, K. Miskell, B. H. Braswell, A. D. Richardson, and R. Pless, "The global network of outdoor webcams: Properties and applications," in *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pp. 111–120, 2009.

[523] S. Garg, "Robust visual place recognition under simultaneous variations in viewpoint and appearance," Ph.D. dissertation, Queensland University of Technology, Queensland, Australia, 2019.

[524] C. Toft, C. Olsson, and F. Kahl, "Long-term 3D localization and pose from semantic labellings," in *Proceedings—2017 IEEE International Conference on Computer Vision Workshops, ICCVW 2017*, vol. 2018-Jan., pp. 650–659, 2018.

[525] S. Garg, A. Jacobson, S. Kumar, and M. Milford, “Improving condition- and environment-invariant place recognition with semantic place categorization,” in *IEEE International Conference on Intelligent Robots and Systems*, IEEE, vol. 2017-Sep., pp. 6863–6870, 2017.

[526] M. J. Milford and G. F. Wyeth, “SeqSLAM: Visual route-based navigation for sunny summer days and stormy winter nights,” in *2012 IEEE International Conference on Robotics and Automation*, IEEE, pp. 1643–1649, 2012.

[527] T. Naseer, G. L. Oliveira, T. Brox, and W. Burgard, “Semantics-aware visual localization under challenging perceptual conditions,” in *Proceedings—IEEE International Conference on Robotics and Automation*, pp. 2614–2620, 2017.

[528] T.-Y. Lin, S. Belongie, and J. Hays, “Cross-view image geolocalization,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 891–898, 2013.

[529] S. Garg, N. Suenderhauf, and M. Milford, “Semantic-geometric visual place recognition: A new perspective for reconciling opposing views,” *International Journal of Robotics Research*, 2019.

[530] R. Arandjelovic, P. Gronat, A. Torii, T. Pajdla, and J. Sivic, “Netvlad: CNN architecture for weakly supervised place recognition,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5297–5307, 2016.

[531] L. Wang, L. Zhao, G. Huo, R. Li, Z. Hou, P. Luo, Z. Sun, K. Wang, and C. Yang, “Visual semantic navigation based on deep learning for indoor mobile robots,” *Complexity*, vol. 2018, 2018.

[532] J. Martínez-Gómez, V. Morell, M. Cazorla, and I. García-Varea, “Semantic localization in the PCL library,” *Robotics and Autonomous Systems*, vol. 75, part B, pp. 641–648, 2016.

[533] C. L. Zitnick and P. Dollár, “Edge boxes: Locating object proposals from edges,” in *European Conference on Computer Vision*, Cham, Switzerland: Springer, pp. 391–405, 2014.

[534] N. Sünderhauf, S. Shirazi, A. Jacobson, F. Dayoub, E. Pepperell, B. Upcroft, and M. Milford, “Place recognition with convnet landmarks: Viewpoint-robust, condition-robust, training-free,” in *Proceedings of Robotics: Science and Systems XII*, 2015.

[535] S. Cascianelli, G. Costante, E. Bellocchio, P. Valigi, M. L. Fravolini, and T. A. Ciarfuglia, "A robust semi-semantic approach for visual localization in urban environment," in *IEEE 2nd International Smart Cities Conference: Improving the Citizens Quality of Life, ISC2 2016—Proceedings*, IEEE, pp. 1–6, 2016.

[536] Z. Seymour, K. Sikka, H.-P. Chiu, S. Samarasekera, and R. Kumar, "Semantically-aware attentive neural embeddings for image-based visual localization," in *Proceedings of the British Machine Vision Conference (BMVC)*, 2019.

[537] N. Sünderhauf, S. Shirazi, F. Dayoub, B. Upcroft, and M. Milford, "On the performance of convnet features for place recognition," in *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, pp. 4297–4304, 2015.

[538] S. Garg, N. Suenderhauf, and M. Milford, "Don't look back: Robustifying place categorization for viewpoint-and condition-invariant place recognition," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, pp. 3645–3652, 2018.

[539] T. Fiolka, J. Stückler, D. A. Klein, D. Schulz, and S. Behnke, "Distinctive 3D surface entropy features for place recognition," in *2013 European Conference on Mobile Robots*, IEEE, pp. 204–209, 2013.

[540] R. Cupec, E. K. Nyarko, D. Filko, A. Kitanov, and I. Petrović, "Place recognition based on matching of planar surfaces and line segments," *The International Journal of Robotics Research*, vol. 34, no. 4–5, pp. 674–704, 2015.

[541] T. Cieslewski, E. Stumm, A. Gawel, M. Bosse, S. Lynen, and R. Siegwart, "Point cloud descriptors for place recognition using sparse visual information," in *2016 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, pp. 4830–4836, 2016.

[542] S. Garg, V. Babu, T. Dharmasiri, S. Hausler, N. Suenderhauf, S. Kumar, T. Drummond, and M. Milford, "Look no deeper: Recognizing places from opposing viewpoints under varying scene appearance using single-view depth estimation," in *IEEE International Conference on Robotics and Automation (ICRA), 2019*, 2019.

[543] F. Maffra, L. Teixeira, Z. Chen, and M. Chli, "Real-time wide-baseline place recognition using depth completion," *IEEE Robotics and Automation Letters*, vol. 4, no. 2, pp. 1525–1532, 2019.

[544] M. Vankadari, S. Garg, A. Majumder, S. Kumar, and A. Behera, "Unsupervised monocular depth estimation for night-time images using adversarial domain feature adaptation," in *16th European Conference on Computer Vision (ECCV), 2020*, 2020.

[545] F. Taubner, F. Tschopp, T. Novkovic, R. Siegwart, and F. Furrer, "LCD–line clustering and description for place recognition," *arXiv preprint arXiv:2010.10867*, 2020.

[546] A. Torii, R. Arandjelovic, J. Sivic, M. Okutomi, and T. Pajdla, "24/7 place recognition by view synthesis," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1808–1817, 2015.

[547] M. Larsson, E. Stenborg, C. Toft, L. Hammarstrand, T. Sattler, and F. Kahl, "Fine-grained segmentation networks: Self-supervised segmentation for improved long-term visual localization," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 31–41, 2019.

[548] G. Neuhold, T. Ollmann, S. Rota Bulo, and P. Kontschieder, "The mapillary vistas dataset for semantic understanding of street scenes," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 4990–4999, 2017.

[549] N. Sünderhauf, O. Brock, W. Scheirer, R. Hadsell, D. Fox, J. Leitner, B. Upcroft, P. Abbeel, W. Burgard, M. Milford, and P. Corke, "The limits and potentials of deep learning for robotics," *The International Journal of Robotics Research*, vol. 37, no. 4–5, pp. 405–420, 2018.

[550] D. Maturana and S. Scherer, "3d convolutional neural networks for landing zone detection from lidar," in *2015 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, pp. 3471–3478, 2015.

[551] Y. Zeng, Y. Hu, S. Liu, J. Ye, Y. Han, X. Li, and N. Sun, "Rt3d: Real-time 3-d vehicle detection in lidar point cloud for autonomous driving," *IEEE Robotics and Automation Letters*, vol. 3, no. 4, pp. 3434–3440, 2018.

[552] J. Dai, Y. Li, K. He, and J. Sun, "R-fcn: Object detection via region-based fully convolutional networks," in *Advances in Neural Information Processing Systems*, pp. 379–387, 2016.

[553] D. Bobkov, S. Chen, R. Jian, M. Z. Iqbal, and E. Steinbach, "Noise-resistant deep learning for object classification in three-dimensional point clouds using a point pair descriptor," *IEEE Robotics and Automation Letters*, vol. 3, no. 2, pp. 865–872, 2018.

[554] E. Wahl, U. Hillenbrand, and G. Hirzinger, "Surflet-pair-relation histograms: A statistical 3D-shape representation for rapid classification," in *Fourth International Conference on 3-D Digital Imaging and Modeling, 2003. 3DIM 2003. Proceedings*, IEEE, pp. 474–481, 2003.

[555] R. B. Rusu, N. Blodow, and M. Beetz, "Fast point feature histograms (FPFH) for 3D registration," in *2009 IEEE International Conference on Robotics and Automation*, IEEE, pp. 3212–3217, 2009.

[556] B. Drost, M. Ulrich, N. Navab, and S. Ilic, "Model globally, match locally: Efficient and robust 3D object recognition," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE, pp. 998–1005, 2010.

[557] W. Wohlkinger and M. Vincze, "Ensemble of shape functions for 3d object classification," in *2011 IEEE International Conference on Robotics and Biomimetics*, IEEE, pp. 2987–2992, 2011.

[558] T. Birdal and S. Ilic, "Point pair features based object detection and pose estimation revisited," in *2015 International Conference on 3D Vision*, IEEE, pp. 527–535, 2015.

[559] W. Tan, H. Liu, Z. Dong, G. Zhang, and H. Bao, "Robust monocular SLAM in dynamic environments," in *2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, IEEE, pp. 209–218, 2013.

[560] B. Mu, S.-Y. Liu, L. Paull, J. Leonard, and J. P. How, "SLAM with objects using a nonparametric pose graph," in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, pp. 4602–4609, 2016.

[561] V. Gaudillière, G. Simon, and M.-O. Berger, "Perspective-2-ellipsoid: Bridging the gap between object detections and 6-dof camera pose," *IEEE Robotics and Automation Letters*, vol. 5, no. 4, pp. 5189–5196, 2020.

[562] Y. Feldman and V. Indelman, "Towards self-supervised semantic representation with a viewpoint-dependent observation model," in *2020 Robotics: Science and Systems Workshop on Self-Supervised Robot Learning*, 2020.

[563] K. Ok, K. Liu, K. Frey, J. P. How, and N. Roy, "Robust object-based SLAM for high-speed autonomous navigation," in *2019 International Conference on Robotics and Automation (ICRA)*, IEEE, pp. 669–675, 2019.

[564] Z. Liao, W. Wang, X. Qi, and X. Zhang, "RGB-D object SLAM using quadrics for indoor environments," *Sensors*, vol. 20, no. 18, p. 5150, 2020.

[565] D. Hoiem, Y. Chodpathumwan, and Q. Dai, "Diagnosing error in object detectors," in *European Conference on Computer Vision*, Berlin and Heidelberg, Germany: Springer, pp. 340–353, 2012.

[566] D. Z. Wang, I. Posner, and P. Newman, "What could move? finding cars, pedestrians and bicyclists in 3d laser data," in *2012 IEEE International Conference on Robotics and Automation*, IEEE, pp. 4038–4044, 2012.

[567] Y. Wang, W.-L. Chao, D. Garg, B. Hariharan, M. Campbell, and K. Q. Weinberger, "Pseudo-lidar from visual depth estimation: Bridging the gap in 3d object detection for autonomous driving," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 8445–8453, 2019.

[568] Y. You, Y. Wang, W.-L. Chao, D. Garg, G. Pleiss, B. Hariharan, M. Campbell, and K. Q. Weinberger, "Pseudo-lidar++: Accurate depth for 3D object detection in autonomous driving," in *International Conference on Learning Representations*, 2019.

[569] J. M. U. Vianney, S. Aich, and B. Liu, "Refinedmpl: Refined monocular pseudolidar for 3D object detection in autonomous driving," *arXiv preprint arXiv:1911.09712*, 2019.

[570] R. Qian, D. Garg, Y. Wang, Y. You, S. Belongie, B. Hariharan, M. Campbell, K. Q. Weinberger, and W.-L. Chao, "End-to-end pseudo-lidar for image-based 3D object detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5881–5890, 2020.

[571] D. J. MacKay, "A practical Bayesian framework for backpropagation networks," *Neural Computation*, vol. 4, no. 3, pp. 448–472, 1992.

[572] Y. Gal and Z. Ghahramani, "Dropout as a Bayesian approximation: Representing model uncertainty in deep learning," in *International Conference on Machine Learning*, pp. 1050–1059, 2016.

[573] B. Lakshminarayanan, A. Pritzel, and C. Blundell, "Simple and scalable predictive uncertainty estimation using deep ensembles," in *Advances in Neural Information Processing Systems*, pp. 6402–6413, 2017.

[574] W. J. Maddox, P. Izmailov, T. Garipov, D. P. Vetrov, and A. G. Wilson, "A simple baseline for Bayesian uncertainty in deep learning," in *Advances in Neural Information Processing Systems*, pp. 13 153–13 164, 2019.

[575] D. Miller, L. Nicholson, F. Dayoub, and N. Sünderhauf, "Dropout sampling for robust object detection in open-set conditions," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, pp. 1–7, 2018.

[576] D. Hall, F. Dayoub, J. Skinner, H. Zhang, D. Miller, P. Corke, G. Carneiro, A. Angelova, and N. Sünderhauf, "Probabilistic object detection: Definition and evaluation," in *The IEEE Winter Conference on Applications of Computer Vision*, pp. 1031–1040, 2020.

[577] Z. Wang, D. Feng, Y. Zhou, W. Zhan, L. Rosenbaum, F. Timm, K. Dietmayer, and M. Tomizuka, "Inferring spatial uncertainty in object detection," *arXiv preprint arXiv:2003.03644*, 2020.

[578] A. Harakeh, M. Smart, and S. L. Waslander, "Bayesod: A Bayesian approach for uncertainty estimation in deep object detectors," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, pp. 87–93, 2020.

[579] P.-Y. Huang, W.-T. Hsu, C.-Y. Chiu, T.-F. Wu, and M. Sun, "Efficient uncertainty estimation for semantic segmentation in videos," in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 520–535, 2018.

[580] M. Rottmann and M. Schubert, "Uncertainty measures and prediction quality rating for the semantic segmentation of nested multi resolution street scene images," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2019.

[581] Q. M. Rahman, N. Sünderhauf, and F. Dayoub, "Performance monitoring of object detection during deployment," *arXiv preprint arXiv:2009.08650*, 2020.

[582] H. Blum, P.-E. Sarlin, J. Nieto, R. Siegwart, and C. Cadena, "The Fishyscapes benchmark: Measuring blind spots in semantic segmentation," in *ICCV Workshops*, 2019.

[583] A. Pronobis, O. Martínez Mozos, B. Caputo, and P. Jensfelt, "Multi-modal semantic place classification," *International Journal of Robotics Research*, vol. 29, no. 2–3, pp. 298–320, 2010.

[584] A. Pronobis and P. Jensfelt, "Hierarchical multi-modal place categorization," in *Proceedings of the European Conference on Mobile Robots*, pp. 1–6, 2011.

[585] O. M. Mozos, H. Mizutani, R. Kurazume, and T. Hasegawa, "Categorization of indoor places using the Kinect sensor," *Sensors (Switzerland)*, vol. 12, no. 5, pp. 6695–6711, 2012.

[586] H. Jung, O. M. Mozos, Y. Iwashita, and R. Kurazume, "Local N-ary patterns: A local multi-modal descriptor for place categorization," *Advanced Robotics*, vol. 30, no. 6, pp. 402–415, 2016.

[587] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Gray scale and rotation invariant texture classification with local binary patterns," in *European Conference on Computer Vision*, vol. 1842, pp. 404–420, 2000.

[588] D. Ramachandram and G. W. Taylor, "Deep multimodal learning: A survey on recent advances and trends," *IEEE Signal Processing Magazine*, vol. 34, no. 6, pp. 96–108, 2017.

[589] D. Feng, C. Haase-Schütz, L. Rosenbaum, H. Hertlein, C. Glaeser, F. Timm, W. Wiesbeck, and K. Dietmayer, "Deep multi-modal object detection and semantic segmentation for autonomous driving: Datasets, methods, and challenges," *IEEE Transactions on Intelligent Transportation Systems*, 2020.

[590] X. Chen, K. Kundu, Y. Zhu, A. G. Berneshawi, H. Ma, S. Fidler, and R. Urtasun, "3d object proposals for accurate object class detection," in *Advances in Neural Information Processing Systems*, pp. 424–432, 2015.

[591] X. Chen, H. Ma, J. Wan, B. Li, and T. Xia, "Multi-view 3d object detection network for autonomous driving," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1907–1915, 2017.

[592] J. Ku, M. Mozifian, J. Lee, A. Harakeh, and S. L. Waslander, "Joint 3d proposal generation and object detection from view aggregation," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, pp. 1–8, 2018.

[593] S. Vora, A. H. Lang, B. Helou, and O. Beijbom, "Pointpainting: Sequential fusion for 3d object detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4604–4612, 2020.

[594] M. Liang, B. Yang, S. Wang, and R. Urtasun, "Deep continuous fusion for multi-sensor 3d object detection," in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 641–656, 2018.

[595] Z. Wang, W. Zhan, and M. Tomizuka, "Fusing bird's eye view lidar point cloud and front view camera image for 3d object detection," in *2018 IEEE Intelligent Vehicles Symposium (IV)*, IEEE, pp. 1–6, 2018.

[596] C. R. Qi, W. Liu, C. Wu, H. Su, and L. J. Guibas, "Frustum pointnets for 3d object detection from RGB-D data," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 918–927, 2018.

[597] Z. Yang, Y. Sun, S. Liu, X. Shen, and J. Jia, "IPOD: Intensive point-based object detector for point cloud," *arXiv preprint arXiv:1812.05276*, 2018.

[598] Z. Wang and K. Jia, "Frustum convnet: Sliding frustums to aggregate local point-wise features for amodal," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, pp. 1742–1749, 2019.

[599] J. Krumm and D. Rouhana, "Placer: Semantic place labels from diary data," in *UbiComp 2013—Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 41–54, 2013.

[600] J. Krumm, D. Rouhana, and M. W. Chang, "Placer++: Semantic place labels beyond the visit," in *2015 IEEE International Conference on Pervasive Computing and Communications, PerCom 2015*, pp. 11–19, 2015.

[601] M. Lv, L. Chen, Z. Xu, Y. Li, and G. Chen, "The discovery of personally semantic places based on trajectory data mining," *Neurocomputing*, vol. 173, pp. 1142–1153, 2016.

[602] R. Roor, J. Hess, M. Saveriano, M. Karg, and A. Kirsch, "Sensor fusion for semantic place labeling," in *VEHITS 2017—Proceedings of the 3rd International Conference on Vehicle Technology and Intelligent Transport Systems*, p. 1, 2017.

[603] U. Weiss and P. Biber, "Semantic place classification and mapping for autonomous agricultural robots," in *IEEE International Conference on Robotics and Automation, Workshop on Semantic Mapping and Autonomous Knowledge Acquisition*, 2010.

[604] Roomba i Series, iRobot, [Online]. Available: https://www.irobot.com.au/roomba/i-series.

[605] RS Model, Robomow Friendly House, [Online]. Available: https://usa.robomow.com/products/rs-model/.

[606] Zodiac Robotic Cleaners VX42, Fluidra Group, [Online]. Available: https://www.zodiac.com.au/vx42.

[607] AUV Sentry, Woods Hole Oceanographic Institution, [Online]. Available: https://www.whoi.edu/what-we-do/explore/underwater-vehicles/sentry/.

[608] C. L. Kaiser, D. R. Yoerger, J. C. Kinsey, S. Kelley, A. Billings, J. Fujii, S. Suman, M. Jakuba, Z. Berkowitz, C. R. German, and A. D. Bowen, "The design and 200 day per year operation of the autonomous underwater vehicle sentry," in *2016 IEEE/OES Autonomous Underwater Vehicles (AUV)*, IEEE, pp. 251–260, 2016.

[609] Mars Curiosity Rover, Nasa Science Mars Exploration Program, [Online]. Available: https://mars.nasa.gov/msl/home/.

[610] R. Manning and W. L. Simon, *Mars Rover Curiosity: An Inside Account from Curiosity's Chief Engineer*. Washington, DC: Smithsonian Institution, 2014.

[611] C. Wang, A. V. Savkin, R. Clout, and H. T. Nguyen, "An intelligent robotic hospital bed for safe transportation of critical neurosurgery patients along crowded hospital corridors," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 23, no. 5, pp. 744–754, 2014.

[612] P. Alves-Oliveira, S. Petisca, F. Correia, N. Maia, and A. Paiva, "Social robots for older adults: Framework of activities for aging in place with robots," in *International Conference on Social Robotics*, Cham, Switzerland: Springer, pp. 11–20, 2015.

[613] Da Vinci, Intuitive, [Online]. Available: https://www.intuitive.com/en-us/products-and-services/da-vinci/systems/sp.

[614] J. C. Byrn, S. Schluender, C. M. Divino, J. Conrad, B. Gurland, E. Shlasko, and A. Szold, "Three-dimensional imaging improves surgical performance for both novice and experienced operators using the da vinci robot system," *The American Journal of Surgery*, vol. 193, no. 4, pp. 519–522, 2007.

[615] M. Hanheide, D. Hebesberger, and T. Krajník, "The when, where, and how: An adaptive robotic info-terminal for care home residents," in *Proceedings of the 2017 ACM/IEEE International Conference on Human–Robot Interaction*, pp. 341–349, 2017.

[616] C. Fattal, V. Leynaert, I. Laffont, A. Baillet, M. Enjalbert, and C. Leroux, "Sam, an assistive robotic device dedicated to helping persons with quadriplegia: Usability study," *International Journal of Social Robotics*, vol. 11, no. 1, pp. 89–103, 2019.
[617] Nao, Softbank Robotics, [Online]. Available: https://www.softbankrobotics.com/emea/en/nao.
[618] Pepper, Softbank Robotics, [Online]. Available: https://www.softbankrobotics.com/emea/en/pepper.
[619] Vector, Anki, [Online]. Available: https://anki.com/en-us/vector.html.
[620] Beam, Suitable Technologies, Inc., [Online]. Available: https://suitabletech.com/.
[621] R. Pinillos, S. Marcos, R. Feliz, E. Zalama, and J. Gómez-García-Bermejo, "Long-term assessment of a service robot in a hotel environment," *Robotics and Autonomous Systems*, vol. 79, pp. 40–57, 2016.
[622] N. Kejriwal, S. Garg, and S. Kumar, "Product counting using images with application to robot-based retail stock assessment," in *2015 IEEE International Conference on Technologies for Practical Robot Applications (TePRA)*, IEEE, pp. 1–6, 2015.
[623] T. Kanda, M. Shiomi, Z. Miyashita, H. Ishiguro, and N. Hagita, "A communication robot in a shopping mall," *IEEE Transactions on Robotics*, vol. 26, no. 5, pp. 897–913, 2010.
[624] H.-M. Gross, H. Boehme, C. Schroeter, S. Müller, A. König, E. Einhorn, C. Martin, M. Merten, and A. Bley, "Toomas: Interactive shopping guide robots in everyday use-final implementation and experiences from long-term field trials," in *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, IEEE, pp. 2005–2012, 2009.
[625] S. Kumar, S. Garg, and N. Kejriwal, "A tea-serving robot for office environment," in *ISR/Robotik 2014; 41st International Symposium on Robotics*, Munich, Germany: VDE, pp. 1–6, 2014.
[626] G. Q. Huang, M. Z. Chen, and J. Pan, "Robotics in ecommerce logistics," *HKIE Transactions*, vol. 22, no. 2, pp. 68–77, 2015.
[627] Starship, Starship Technologies, [Online]. Available: https://www.starship.xyz/.

[628] Robots—Your guide to the world of robots, IEEE, [Online]. Available: https://robots.ieee.org/robots/.

[629] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? The kitti vision benchmark suite," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, pp. 3354–3361, 2012.

[630] D. Du, Y. Qi, H. Yu, Y. Yang, K. Duan, G. Li, W. Zhang, Q. Huang, and Q. Tian, "The unmanned aerial vehicle benchmark: Object detection and tracking," in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 370–386, 2018.

[631] P. Zhu, L. Wen, X. Bian, H. Ling, and Q. Hu, "Vision meets drones: A challenge," *arXiv preprint arXiv:1804.07437*, 2018.

[632] Y. Lyu, G. Vosselman, G.-S. Xia, A. Yilmaz, and M. Y. Yang, "UAVid: A semantic segmentation dataset for UAV imagery," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 165, pp. 108–119, 2020.

[633] F. Rottensteiner, G. Sohn, M. Gerke, J. D. Wegner, U. Breitkopf, and J. Jung, "Results of the ISPRS benchmark on urban object detection and 3D building reconstruction," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 93, pp. 256–271, 2014.

[634] M. Campos-Taberner, A. Romero-Soriano, C. Gatta, G. Camps-Valls, A. Lagrange, B. Le Saux, A. Beaupere, A. Boulch, A. Chan-Hon-Tong, S. Herbin, H. Randrianarivo, M. Ferecatu, M. Shimoni, G. Moser, and D. Tuia, "Processing of extremely high-resolution lidar and RGB data: Outcome of the 2015 IEEE GRSS data fusion contest–part a: 2-d contest," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no. 12, pp. 5547–5559, 2016.

[635] C. Debes, A. Merentitis, R. Heremans, J. Hahn, N. Frangiadakis, T. van Kasteren, W. Liao, R. Bellens, A. Pižurica, S. Gautama, W. Philips, S. Prasad, Q. Du, and F. Pacifici, "Hyperspectral and lidar data fusion: Outcome of the 2013 GRSS data fusion contest," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 6, pp. 2405–2418, 2014.

[636] I. Demir, K. Koperski, D. Lindenbaum, G. Pang, J. Huang, S. Basu, F. Hughes, D. Tuia, and R. Raska, "Deepglobe 2018: A challenge to parse the earth through satellite images," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, IEEE, pp. 172–181, 2018.

[637] I. Nigam, C. Huang, and D. Ramanan, "Ensemble knowledge transfer for semantic segmentation," in *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, IEEE, pp. 1499–1508, 2018.

[638] B. Benjdira, Y. Bazi, A. Koubaa, and K. Ouni, "Unsupervised domain adaptation using generative adversarial networks for semantic segmentation of aerial images," *Remote Sensing*, vol. 11, no. 11, p. 1369, 2019.

[639] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in Neural Information Processing Systems*, pp. 2672–2680, 2014.

[640] F. S. Leira, T. A. Johansen, and T. I. Fossen, "Automatic detection, classification and tracking of objects in the ocean surface from UAVs using a thermal camera," in *2015 IEEE Aerospace Conference*, IEEE, pp. 1–10, 2015.

[641] J. Zheng, X. Cao, B. Zhang, Y. Huang, and Y. Hu, "Bi-heterogeneous convolutional neural network for UAV-based dynamic scene classification," in *2017 Integrated Communications, Navigation and Surveillance Conference (ICNS)*, IEEE, 5B4–1, 2017.

[642] J. C. van Gemert, C. R. Verschoor, P. Mettes, K. Epema, L. P. Koh, and S. Wich, "Nature conservation drones for automatic localization and counting of animals," in *European Conference on Computer Vision*, Cham, Switzerland: Springer, pp. 255–270, 2014.

[643] A. Rivas, P. Chamoso, A. González-Briones, and J. M. Corchado, "Detection of cattle using drones and convolutional neural networks," *Sensors*, vol. 18, no. 7, p. 2048, 2018.

[644] E. Bondi, F. Fang, M. Hamilton, D. Kar, D. Dmello, J. Choi, R. Hannaford, A. Iyer, L. Joppa, M. Tambe, and R. Nevatia, "Spot poachers in action: Augmenting conservation drones with automatic detection in near real time," in *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[645] Z. Huang, C. Fu, L. Wang, C. Shang, and B. Yang, "Construction site management and control technology based on UAV visual surveillance," *Automation and Instrumentation*, no. 4, p. 28, 2018.

[646] Wing, Wing Aviation LLC, [Online]. Available: https://wing.com/.

[647] J. R. Kellner, J. Armston, M. Birrer, K. Cushman, L. Duncanson, C. Eck, C. Falleger, B. Imbach, K. Kral, M. Krucek, J. Trochta, T. Vrška, and C. Zgraggen, "New opportunities for forest remote sensing through ultra-high-density drone lidar," *Surveys in Geophysics*, vol. 40, no. 4, pp. 959–977, 2019.

[648] F. Sadeghi and S. Levine, "Cad2rl: Real single-image flight without a single real image," in *Robotics: Science and Systems*, 2017.

[649] A. Loquercio, A. I. Maqueda, C. R. Del-Blanco, and D. Scaramuzza, "Dronet: Learning to fly by driving," *IEEE Robotics and Automation Letters*, vol. 3, no. 2, pp. 1088–1095, 2018.

[650] D. Gandhi, L. Pinto, and A. Gupta, "Learning to fly by crashing," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, pp. 3948–3955, 2017.

[651] C. Devin, A. Gupta, T. Darrell, P. Abbeel, and S. Levine, "Learning modular neural network policies for multi-task and multi-robot transfer," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, pp. 2169–2176, 2017.

[652] I. Clavera, D. Held, and P. Abbeel, "Policy transfer via modularity and reward guiding," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, pp. 1537–1544, 2017.

[653] M. Mueller, A. Dosovitskiy, B. Ghanem, and V. Koltun, "Driving policy transfer via modularity and abstraction," in *Conference on Robot Learning*, pp. 1–15, 2018.

[654] A. Loquercio, E. Kaufmann, R. Ranftl, A. Dosovitskiy, V. Koltun, and D. Scaramuzza, "Deep drone racing: From simulation to reality with domain randomization," *IEEE Transactions on Robotics*, vol. 36, no. 1, pp. 1–14, 2019.

[655] E. Kaufmann, M. Gehrig, P. Foehn, R. Ranftl, A. Dosovitskiy, V. Koltun, and D. Scaramuzza, "Beauty and the beast: Optimal methods meet learning for drone racing," in *2019 International Conference on Robotics and Automation (ICRA)*, IEEE, pp. 690–696, 2019.

[656] D. Palossi, A. Loquercio, F. Conti, E. Flamand, D. Scaramuzza, and L. Benini, "A 64-mw DNN-based visual navigation engine for autonomous nano-drones," *IEEE Internet of Things Journal*, vol. 6, no. 5, pp. 8357–8371, 2019.

[657] R. Spica, E. Cristofalo, Z. Wang, E. Montijano, and M. Schwager, "A real-time game theoretic planner for autonomous two-player drone racing," *IEEE Transactions on Robotics*, vol. 36, no. 5, pp. 1389–1403, 2020.

[658] N. Mandel, F. V. Alvarez, M. Milford, and F. Gonzalez, "Towards simulating semantic onboard UAV navigation," in *Proceedings of the 2020 IEEE Aerospace Conference*, IEEE, pp. 1–15, 2020. [Online]. Available: https://eprints.qut.edu.au/136148/.

[659] J. Boyd, "A discourse on winning and losing [briefing slides]," *Maxwell Air Force Base, AL: Air University Library (Document No. MU 43947)*, 1987.

[660] J. Boyd, *A Discourse on Winning and Losing*. Alabama: Air University Press, Curtis E. LeMay Center for Doctrine Development and Education, 2018.

[661] D. Cavaliere, S. Senatore, M. Vento, and V. Loia, "Towards semantic context-aware drones for aerial scenes understanding," in *2016 13th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, IEEE, pp. 115–121, 2016.

[662] D. Cavaliere, V. Loia, A. Saggese, S. Senatore, and M. Vento, "Semantically enhanced UAVs to increase the aerial scene understanding," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 49, no. 3, pp. 555–567, 2017.

[663] M.-J. Jeon, H.-K. Park, B. Jagvaral, H.-S. Yoon, Y.-G. Kim, and Y.-T. Park, "Relationship between UAVs and ambient objects with threat situational awareness through grid map-based ontology reasoning," *International Journal of Computers and Applications*, pp. 1–16, 2019.

[664] W. He, Z. Li, and C. P. Chen, "A survey of human-centered intelligent robots: Issues and challenges," *IEEE/CAA Journal of Automatica Sinica*, vol. 4, no. 4, pp. 602–609, 2017.

[665] J. Cheng, H. Cheng, M. Q.-H. Meng, and H. Zhang, "Autonomous navigation by mobile robots in human environments: A survey," in *2018 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, IEEE, pp. 1981–1986, 2018.

[666] K. J. Vänni and S. E. Salin, "A need for service robots among health care professionals in hospitals and housing services," in *International Conference on Social Robotics*, Cham, Switzerland: Springer, pp. 178–187, 2017.

[667] M. Niemelä, A. Arvola, and I. Aaltonen, "Monitoring the acceptance of a social service robot in a shopping mall: First results," in *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human–Robot Interaction*, pp. 225–226, 2017.

[668] I. P. Tussyadiah and S. Park, "Consumer evaluation of hotel service robots," in *Information and Communication Technologies in Tourism 2018*, Cham, Switzerland: Springer, 2018, pp. 308–320.

[669] A. Korchut, S. Szklener, C. Abdelnour, N. Tantinya, J. Hernandez-Farigola, J. C. Ribes, U. Skrobas, K. Grabowska-Aleksandrowicz, D. Szczesniak-Stanczyk, and K. Rejdak, "Challenges for service robots—requirements of elderly adults with cognitive impairments," *Frontiers in Neurology*, vol. 8, p. 228, 2017.

[670] M. A. V. J. Muthugala and A. B. P. Jayasekara, "A review of service robots coping with uncertain information in natural language instructions," *IEEE Access*, vol. 6, pp. 12 913–12 928, 2018.

[671] A. Rosete, B. Soares, J. Salvadorinho, J. Reis, and M. Amorim, "Service robots in the hospitality industry: An exploratory literature review," in *International Conference on Exploring Services Science*, Cham, Switzerland: Springer, pp. 174–186, 2020.

[672] M. Bakri, A. Ismail, M. Hashim, and M. Safar, "A review on service robots: Mechanical design and localization system," in *IOP Conference Series: Materials Science and Engineering*, Pulau Pinang, Malaysia: IOP Publishing, vol. 705, 2019.

[673] T. Mettler, M. Sprenger, and R. Winter, "Service robots in hospitals: New perspectives on niche evolution and technology affordances," *European Journal of Information Systems*, vol. 26, no. 5, pp. 451–468, 2017.

[674] N. Savela, T. Turja, and A. Oksanen, "Social acceptance of robots in different occupational fields: A systematic literature review," *International Journal of Social Robotics*, vol. 10, no. 4, pp. 493–502, 2018.

[675] J. Aaen, "Robots in welfare services: A systematic literature review," in *IRIS/SCIS Conference 2018*, 2018.

[676] A. Tuomi, I. Tussyadiah, and J. Stienmetz, "Service robots and the changing roles of employees in restaurants: A cross cultural study," *E-Review of Tourism Research*, vol. 17, no. 5, 2019.

[677] D. B. Gracia, C. Flavián, L. V. Casaló, and J. J. Schepers, "Robots or frontline employees?: Exploring customers' attributions of responsibility and stability after service failure or success," *Journal of Service Management*, vol. 31, no. 2, pp. 267–289, 2020.

[678] N. Walker, Y. Jiang, M. Cakmak, and P. Stone, "Desiderata for planning systems in general-purpose service robots," in *International Conference on Automated Planning and Scheduling (ICAPS) Workshop on Planning and Robotics (PlanRob)*, 2019.

[679] Care-o-bot 4, Fraunhofer Institute for Manufacturing Engineering and Automation, [Online]. Available: https://www.care-o-bot.de/en/care-o-bot-4.html.

[680] PR2, Willow Garage, [Online]. Available: http://www.willowgarage.com/pages/pr2/overview.

[681] Robot Operating System (ROS), Open Robotics, [Online]. Available: https://www.ros.org/.

[682] C. Do and W. Burgard, "Accurate pouring with an autonomous robot using an rgb-d camera," in *International Conference on Intelligent Autonomous Systems*, Cham, Switzerland: Springer, pp. 210–221, 2018.

[683] D. Berenson, P. Abbeel, and K. Goldberg, "A robot path planning framework that learns from experience," in *2012 IEEE International Conference on Robotics and Automation*, IEEE, pp. 3671–3678, 2012.

[684] SPOT, Boston Dynamics, [Online]. Available: https://www.bostondynamics.com/spot.

[685] N. Bellotto and H. Hu, "Multisensor-based human detection and tracking for mobile service robots," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 39, no. 1, pp. 167–181, 2008.

[686] J. Kim, J. Seol, S. Lee, S.-W. Hong, and H. I. Son, "An intelligent spraying system with deep learning-based semantic segmentation of fruit trees in orchards," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 2020.

[687] Z. Zeng, A. Röfer, and O. C. Jenkins, "Semantic linking maps for active visual object search," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 2020.

[688] L. Sharma and P. K. Garg, *From Visual Surveillance to Internet of Things: Technology and Applications.* New York: CRC Press, 2019.

[689] R. Jones, "Visual surveillance technologies," *The Routledge Handbook of Technology, Crime and Justice,* New York: Routledge, p. 436, 2017.

[690] T. Bouwmans and B. García-García, "Visual surveillance of natural environments: Background subtraction challenges and methods," in *From Visual Surveillance to Internet of Things: Technology and Applications,* New York: CRC Press, p. 67, 2019.

[691] I. Bouchrika, "A survey of using biometrics for smart visual surveillance: Gait recognition," in *Surveillance in Action*, Cham, Switzerland: Springer, 2018, pp. 3–23.

[692] S. K. Kumaran, D. P. Dogra, and P. P. Roy, "Anomaly detection in road traffic using visual surveillance: A survey," *arXiv preprint arXiv:1901.08292*, 2019.

[693] I. E. Olatunji and C.-H. Cheng, "Video analytics for visual surveillance and applications: An overview and survey," in *Machine Learning Paradigms*, Cham, Switzerland: Springer, 2019, pp. 475–515.

[694] H. Singh, "Applications of intelligent video analytics in the field of retail management: A study," in *Supply Chain Management Strategies and Risk Assessment in Retail Environments*, Pennsylvania: IGI Global, 2018, pp. 42–59.

[695] R. K. Tripathi, A. S. Jalal, and S. C. Agrawal, "Suspicious human activity recognition: A review," *Artificial Intelligence Review*, vol. 50, no. 2, pp. 283–339, 2018.

[696] U. Knauer, M. Himmelsbach, F. Winkler, F. Zautke, K. Bienefeld, and B. Meffert, "Application of an adaptive background model for monitoring honeybees," in *Proceedings of Visualization, Imaging and Image Processing (VIIP),* Benidorm, Spain: Acta Press, 2005.

[697] P. Dickinson, C. Qing, S. Lawson, and R. Freeman, "Automated visual monitoring of nesting seabirds," in *Workshop on Visual Observation and Analysis of Animal and Insect Behaviour, Istanbul*, Citeseer, 2010.

[698] C. Spampinato, S. Palazzo, and I. Kavasidis, "A texton-based kernel density estimation approach for background modeling under extreme conditions," *Computer Vision and Image Understanding*, vol. 122, pp. 74–83, 2014.

[699] Y. Iwatani, K. Tsurui, and A. Honma, "Position and direction estimation of wolf spiders, pardosa astrigera, from video images," in *2016 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, IEEE, pp. 1966–1971, 2016.

[700] T. Bouwmans and B. García-García, "Visual surveillance of human activities: Background subtraction challenges and methods," in *From Visual Surveillance to Internet of Things: Technology and Applications*, New York: CRC Press, p. 43, 2019.

[701] N. C. Sarkar, A. Bhaskar, Z. Zheng, and M. P. Miska, "Microscopic modelling of area-based heterogeneous traffic flow: Area selection and vehicle movement," *Transportation Research Part C: Emerging Technologies*, vol. 111, pp. 373–396, 2020.

[702] G. Monteiro, J. Marcos, M. Ribeiro, and J. Batista, "Robust segmentation process to detect incidents on highways," in *International Conference Image Analysis and Recognition*, Berlin and Heidelberg, Germany: Springer, pp. 110–121, 2008.

[703] J. Cho, J. Park, U. Baek, D. Hwang, S. Choi, S. Kim, and K. Kim, "Automatic parking system using background subtraction with CCTV environment international conference on control, automation and systems (ICCAS 2016)," in *2016 16th International Conference on Control, Automation and Systems (ICCAS)*, IEEE, pp. 1649–1652, 2016.

[704] K. Tarrit, J. Molleda, G. A. Atkinson, M. L. Smith, G. C. Wright, and P. Gaal, "Vanishing point detection for visual surveillance systems in railway platform environments," *Computers in Industry*, vol. 98, pp. 153–164, 2018.

[705] I. Bouchrika, J. N. Carter, and M. S. Nixon, "Towards automated visual surveillance using gait for identity recognition and tracking across multiple non-intersecting cameras," *Multimedia Tools and Applications*, vol. 75, no. 2, pp. 1201–1221, 2016.

[706] H.-M. Hu, Q. Guo, J. Zheng, H. Wang, and B. Li, "Single image defogging based on illumination decomposition for visual maritime surveillance," *IEEE Transactions on Image Processing*, vol. 28, no. 6, pp. 2882–2897, 2019.

[707] S. Rashmi and K. Rangarajan, "Rule based visual surveillance system for the retail domain," in *Proceedings of International Conference on Cognition and Recognition*, Singapore: Springer, pp. 145–156, 2018.

[708] M. H. Khan, F. Li, M. S. Farid, and M. Grzegorzek, "Gait recognition using motion trajectory analysis," in *International Conference on Computer Recognition Systems*, Cham, Switzerland: Springer, pp. 73–82, 2017.

[709] L. Zheng, Z. Bie, Y. Sun, J. Wang, C. Su, S. Wang, and Q. Tian, "Mars: A video benchmark for large-scale person re-identification," in *European Conference on Computer Vision*, Cham, Switzerland: Springer, pp. 868–884, 2016.

[710] J. Xu, R. Zhao, F. Zhu, H. Wang, and W. Ouyang, "Attention-aware compositional network for person re-identification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2119–2128, 2018.

[711] Y. Rao, J. Lu, and J. Zhou, "Attention-aware deep reinforcement learning for video face recognition," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 3931–3940, 2017.

[712] I. Masi, Y. Wu, T. Hassner, and P. Natarajan, "Deep face recognition: A survey," in *2018 31st SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, IEEE, pp. 471–478, 2018.

[713] M. Sharif, F. Naz, M. Yasmin, M. A. Shahid, and A. Rehman, "Face recognition: A survey," *Journal of Engineering Science and Technology Review*, vol. 10, no. 2, pp. 166–177, 2017.

[714] M. Kristan, A. Leonardis, J. Matas, M. Felsberg, R. Pflugfelder, L. C. Zajc, T. Vojir, G. Bhat, A. Lukezic, A. Eldesokey, G. Fernandez, A. Garcia-Martin, A. Iglesias-Arias, A. Aydin Alatan, A. Gonzalez-Garcia, A. Petrosino, A. Memarmoghadam, A. Vedaldi, A. Muhic, A. He, A. Smeulders, A. G. Perera, B. Li, B. Chen, C. Kim, C. Xu, C. Xiong, C. Tian, C. Luo, C. Sun, C. Hao, D. Kim, D. Mishra, D. Chen, D. Wang, D. Wee, E. Gavves, E. Gundogdu, E. Velasco-Salido, F. S. Khan, F. Yang, F. Zhao, F. Li, F. Battistone, G. De Ath, G. R. K. S. Subrahmanyam, G. Bastos, H. Ling, H. K. Galoogahi, H. Lee, H. Li, H. Zhao, H. Fan, H. Zhang, H. Possegger, H. Li, H. Lu, H. Zhi, H. Li, H. Lee, H. J. Chang, I. Drummond, J. Valmadre, J. S. Martin, J. Chahl, J. Y. Choi, J. Li, J. Wang, J. Qi, J. Sung, J. Johnander, J. Henriques, J. Choi, J. van de Weijer, J. R. Herranz, J. M. Martinez, J. Kittler, J. Zhuang, J. Gao, K. Grm, L. Zhang, L. Wang, L. Yang, L. Rout, L. Si, L. Bertinetto, L. Chu, M. Che, M. E. Maresca, M. Danelljan, M.-H. Yang, M. Abdelpakey, M. Shehata, M. Kang, N. Lee, N. Wang, O. Miksik, P. Moallem, P. Vicente-Monivar, P. Senna, P. Li, P. Torr, P. M. Raju, Q. Ruihe, Q. Wang, Q. Zhou, Q. Guo, R. Martin-Nieto, R. K. Gorthi, R. Tao, R. Bowden, R. Everson, R. Wang, S. Yun, S. Choi, S. Vivas, S. Bai, S. Huang, S. Wu, S. Hadfield, S. Wang, S. Golodetz, T. Ming, T. Xu, T. Zhang, T. Fischer, V. Santopietro, V. Struc, W. Wei, W. Zuo, W. Feng, W. Wu, W. Zou, W. Hu, W. Zhou, W. Zeng, X. Zhang, X. Wu, X.-J. Wu, X. Tian, Y. Li, Y. Lu, Y. W. Law, Y. Wu, Y. Demiris, Y. Yang, Y. Jiao, Y. Li, Y. Zhang, Y. Sun, Z. Zhang, Z. Zhu, Z.-H. Feng, Z. Wang, and Z. He, "The sixth visual object tracking VOT2018 challenge results," in *European Conference on Computer Vision Workshops*, pp. 3–53, 2018.

[715] A. Milan, L. Leal-Taixé, I. Reid, S. Roth, and K. Schindler, "Mot16: A benchmark for multi-object tracking," *arXiv preprint arXiv:1603.00831*, 2016.

[716] A. Gaidon, Q. Wang, Y. Cabon, and E. Vig, "Virtual worlds as proxy for multi-object tracking analysis," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4340–4349, 2016.

[717] S. Garg, E. Hassan, S. Kumar, and P. Guha, "A hierarchical frame-by-frame association method based on graph matching for multi-object tracking," in *International Symposium on Visual Computing*, Springer, pp. 138–150, 2015.

[718] P. Dendorfer, H. Rezatofighi, A. Milan, J. Shi, D. Cremers, I. Reid, S. Roth, K. Schindler, and L. Leal-Taixé, "Mot20: A benchmark for multi object tracking in crowded scenes," *arXiv preprint arXiv:2003.09003*, 2020.

[719] R. K. Tripathi, A. S. Jalal, and S. C. Agrawal, "Abandoned or removed object detection from visual surveillance: A review," *Multimedia Tools and Applications*, vol. 78, no. 6, pp. 7585–7620, 2019.

[720] S. Grigorescu, B. Trasnea, T. Cocias, and G. Macesanu, "A survey of deep learning techniques for autonomous driving," *Journal of Field Robotics*, vol. 37, no. 3, pp. 362–386, 2019.

[721] H. Yin and C. Berger, "When to use what data set for your self-driving car algorithm: An overview of publicly available driving datasets," in *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*, IEEE, pp. 1–8, 2017.

[722] J. Janai, F. Güney, A. Behl, and A. Geiger, "Computer vision for autonomous vehicles: Problems, datasets and state of the art," *Foundations and Trends® in Computer Graphics and Vision*, vol. 12, no. 1–3, pp. 1–308, 2020.

[723] K. Bengler, K. Dietmayer, B. Farber, M. Maurer, C. Stiller, and H. Winner, "Three decades of driver assistance systems: Review and future perspectives," *IEEE Intelligent Transportation Systems Magazine*, vol. 6, no. 4, pp. 6–22, 2014.

[724] H. Caesar, V. Bankiti, A. H. Lang, S. Vora, V. E. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan, and O. Beijbom, "Nuscenes: A multimodal dataset for autonomous driving," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11 621–11 631, 2020.

[725] P. Sun, H. Kretzschmar, X. Dotiwalla, A. Chouard, V. Patnaik, P. Tsui, J. Guo, Y. Zhou, Y. Chai, B. Caine, V. Vasudevan, W. Han, J. Ngiam, H. Zhao, A. Timofeev, S. Ettinger, M. Krivokon, A. Gao, A. Joshi, Y. Zhang, J. Shlens, Z. Chen, and D. Anguelov, "Scalability in perception for autonomous driving: Waymo open dataset," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2446–2454, 2020.

[726] A. Karpathy, "AI for Full-Self Driving," *Matroid, 5th Annual Scaled Machine Learning Conference*, [Online]. Available: https://www.youtube.com/watch?v=hx7BXih7zx8.

[727] M. Milford, S. Garg, and J. Mount, "P1-007: How automated vehicles will interact with road infrastructure now and in the future," iMove, QUT and Queensland Government, Tech. Rep., Jan. 2020. [Online]. Available: https://imoveaustralia.com/wp-content/uploads/2020/02/P1-007-Milestone-6-Final-Report-Second-Revision.pdf.

[728] M. Fausten, Autonomous driving Interview with Michael Fausten, Bosch Global, [Online]. Available: https://www.bosch.com/stories/autonomous-driving-interview-with-michael-fausten/.

[729] X. Chen, K. Kundu, Z. Zhang, H. Ma, S. Fidler, and R. Urtasun, "Monocular 3d object detection for autonomous driving," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2147–2156, 2016.

[730] Z. Cao, E. Bıyık, W. Z. Wang, A. Raventos, A. Gaidon, G. Rosman, and D. Sadigh, "Reinforcement learning based control of imitative policies for near-accident driving," in *Proceedings of Robotics: Science and Systems XVI*, 2020.

[731] H. Yi, S. Shi, M. Ding, J. Sun, K. Xu, H. Zhou, Z. Wang, S. Li, and G. Wang, "Segvoxelnet: Exploring semantic context and depth-aware features for 3D vehicle detection from point cloud," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 2020.

[732] B. Li, "3d fully convolutional network for vehicle detection in point cloud," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, pp. 1513–1518, 2017.

[733] M. Liang, B. Yang, Y. Chen, R. Hu, and R. Urtasun, "Multi-task multi-sensor fusion for 3d object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7345–7353, 2019.

[734] W. Zeng, W. Luo, S. Suo, A. Sadat, B. Yang, S. Casas, and R. Urtasun, "End-to-end interpretable neural motion planner," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 8660–8669, 2019.

[735] A. Sadat, M. Ren, A. Pokrovsky, Y.-C. Lin, E. Yumer, and R. Urtasun, "Jointly learnable behavior and trajectory planning for self-driving vehicles," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, pp. 3949–3956, 2019.

[736] D. Frossard and R. Urtasun, "End-to-end learning of multi-sensor 3d tracking by detection," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, pp. 635–642, 2018.

[737] K. Mangalam, H. Girase, S. Agarwal, K.-H. Lee, E. Adeli, J. Malik, and A. Gaidon, "It is not the journey but the destination: Endpoint conditioned trajectory prediction," in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 759–776, Aug. 2020.

[738] B. Liu, E. Adeli, Z. Cao, K.-H. Lee, A. Shenoi, A. Gaidon, and J. C. Niebles, "Spatiotemporal relationship reasoning for pedestrian intent prediction," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 3485–3492, 2020.

[739] S. Casas, W. Luo, and R. Urtasun, "Intentnet: Learning to predict intention from raw sensor data," in *Conference on Robot Learning*, pp. 947–956, 2018.

[740] D. Oh, D. Ji, C. Jang, Y. Hyun, H. S. Bae, and S. Hwang, "Segmenting 2k-videos at 36.5 fps with 24.3 gflops: Accurate and lightweight realtime semantic segmentation network," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 2020.

[741] C. Zhang, W. Luo, and R. Urtasun, "Efficient convolutions for real-time semantic segmentation of 3d point clouds," in *2018 International Conference on 3D Vision (3DV)*, IEEE, pp. 399–408, 2018.

[742] B. Yang, W. Luo, and R. Urtasun, "Pixor: Real-time 3d object detection from point clouds," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7652–7660, 2018.

[743] W. Luo, B. Yang, and R. Urtasun, "Fast and furious: Real time end-to-end 3d detection, tracking and motion forecasting with a single convolutional net," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3569–3577, 2018.

[744] Y. Zhu, C. Li, and Y. Zhang, "Online camera-lidar calibration with sensor semantic information," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 2020.

[745] X. Wei, I. A. Bârsan, S. Wang, J. Martinez, and R. Urtasun, "Learning to localize through compressed binary maps," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 10 316–10 324, 2019.

[746] I. A. Barsan, S. Wang, A. Pokrovsky, and R. Urtasun, "Learning to localize using a lidar intensity map," in *CoRL*, pp. 605–616, 2018.

[747] M. Bai, G. Mattyus, N. Homayounfar, S. Wang, S. K. Lakshmikanth, and R. Urtasun, "Deep multi-sensor lane detection," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, pp. 3102–3109, 2018.

[748] G. Máttyus and R. Urtasun, "Matching adversarial networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 8024–8032, 2018.

[749] S. Wang, M. Bai, G. Mattyus, H. Chu, W. Luo, B. Yang, J. Liang, J. Cheverie, S. Fidler, and R. Urtasun, "Torontocity: Seeing the world with a million eyes," in *2017 IEEE International Conference on Computer Vision (ICCV)*, IEEE, pp. 3028–3036, 2017.

[750] N. Homayounfar, W.-C. Ma, S. Kowshika Lakshmikanth, and R. Urtasun, "Hierarchical recurrent attention networks for structured online maps," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3417–3426, 2018.

[751] J. Liang, N. Homayounfar, W.-C. Ma, S. Wang, and R. Urtasun, "Convolutional recurrent network for road boundary extraction," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 9512–9521, 2019.

[752] J. Liang and R. Urtasun, "End-to-end deep structured models for drawing crosswalks," in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 396–412, 2018.

[753] G. Máttyus, W. Luo, and R. Urtasun, "Deeproadmapper: Extracting road topology from aerial images," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 3438–3446, 2017.

[754] B. Yang, M. Liang, and R. Urtasun, "Hdnet: Exploiting hd maps for 3d object detection," in *Conference on Robot Learning*, pp. 146–155, 2018.

[755] W.-C. Ma, I. Tartavull, I. A. Bârsan, S. Wang, M. Bai, G. Mattyus, N. Homayounfar, S. K. Lakshmikanth, A. Pokrovsky, and R. Urtasun, "Exploiting sparse semantic hd maps for self-driving vehicle localization," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, pp. 5304–5311, 2019.

[756] M. Yazdi and T. Bouwmans, "New trends on moving object detection in video images captured by a moving camera: A survey," *Computer Science Review*, vol. 28, pp. 157–177, 2018.

[757] S. W. Stoughton, "Police body-worn cameras," *NCL Rev.*, vol. 96, p. 1363, 2017.

[758] A. Saeed, A. Abdelkader, M. Khan, A. Neishaboori, K. A. Harras, and A. Mohamed, "Argus: Realistic target coverage by drones," in *Proceedings of the 16th ACM/IEEE International Conference on Information Processing in Sensor Networks*, pp. 155–166, 2017.

[759] C. Bracken-Roche, “Domestic drones: The politics of verticality and the surveillance industrial complex,” *Geographica Helvetica*, vol. 71, no. 3, pp. 167–172, 2016.

[760] D. B. Mesquita, R. F. dos Santos, D. G. Macharet, M. F. Campos, and E. R. Nascimento, “Fully convolutional siamese autoencoder for change detection in UAV aerial images,” *IEEE Geoscience and Remote Sensing Letters*, 2019.

[761] S. Minaeian, “Effective visual surveillance of human crowds using cooperative unmanned vehicles,” in *2016 Winter Simulation Conference (WSC)*, IEEE, pp. 3674–3675, 2016.

[762] M. H. M. Saad, “Room searching performance evaluation for the jagabottm indoor surveillance robot,” *KnE Engineering*, 2016.

[763] S. Meghana, T. V. Nikhil, R. Murali, S. Sanjana, R. Vidhya, and K. J. Mohammed, “Design and implementation of surveillance robot for outdoor security,” in *2017 2nd IEEE International Conference on Recent Trends in Electronics, Information and Communication Technology (RTEICT)*, IEEE, pp. 1679–1682, 2017.

[764] I. Kostavelis and A. Gasteratos, “Robots in crisis management: A survey,” in *International Conference on Information Systems for Crisis Response and Management in Mediterranean Countries*, Cham, Switzerland: Springer, pp. 43–56, 2017.

[765] M. Fuchida, S. Chikushi, A. Moro, A. Yamashita, and H. Asama, “Arbitrary viewpoint visualization for disaster response robots,” *Journal of Advanced Simulation in Science and Engineering*, vol. 6, no. 1, pp. 249–259, 2018. DOI: 10.15748/jasse.6.249.

[766] A. Asokan, A. J. Pothen, and R. K. Vijayaraj, “Armatron—A wearable gesture recognition glove: For control of robotic devices in disaster management and human rehabilitation,” in *2016 International Conference on Robotics and Automation for Humanitarian Applications (RAHA)*, IEEE, pp. 1–5, 2016.

[767] S. Dey, A. Bhattacharyya, and A. Mukherjee, “Semantic data exchange between collaborative robots in fog environment: Can coap be a choice?” In *2017 Global Internet of Things Summit (GIoTS)*, IEEE, pp. 1–6, 2017.

[768] S. Lee, D. Har, and D. Kum, "Drone-assisted disaster management: Finding victims via infrared camera and lidar sensor fusion," in *2016 3rd Asia-Pacific World Congress on Computer Science and Engineering (APWC on CSE)*, IEEE, pp. 84–89, 2016.

[769] M. Billinghurst, A. Clark, and G. Lee, "A survey of augmented reality," *Foundations and Trends in Human-Computer Interaction*, vol. 8, no. 2–3, pp. 73–272, 2015.

[770] C. Arth, R. Grasset, L. Gruber, T. Langlotz, A. Mulloni, and D. Wagner, "The history of mobile augmented reality," *arXiv preprint arXiv:1505.01319*, 2015.

[771] R. T. Azuma, "The most important challenge facing augmented reality," *Presence: Teleoperators and Virtual Environments,* vol. 25, no. 3, pp. 234–238, 2016.

[772] D. Van Krevelen and R. Poelman, "A survey of augmented reality technologies, applications and limitations," *International Journal of Virtual Reality*, vol. 9, no. 2, pp. 1–20, 2010.

[773] J. L. Bacca Acosta, S. M. Baldiris Navarro, R. Fabregat Gesa, S. Graf, and Kinshuk, "Augmented reality trends in education: A systematic review of research and applications," *Journal of Educational Technology and Society*, vol. 17, no. 4, pp. 133–149, 2014.

[774] J. Grubert, T. Langlotz, and R. Grasset, "Augmented reality browser survey," Institute for Computer Graphics and Vision, University of Technology Graz, Technical Report, vol. 1101, p. 37, 2011.

[775] L. Chen, T. W. Day, W. Tang, and N. W. John, "Recent developments and future challenges in medical mixed reality," in *2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, IEEE, pp. 123–135, 2017.

[776] F. Cosentino, N. John, and J. Vaarkamp, "An overview of augmented and virtual reality applications in radiotherapy and future developments enabled by modern tablet devices," *Journal of Radiotherapy in Practice*, vol. 13, no. 3, p. 350, 2014.

[777] M. Kersten-Oertel, P. Jannin, and D. L. Collins, "The state of the art of visualization in mixed reality image guided surgery," *Computerized Medical Imaging and Graphics*, vol. 37, no. 2, pp. 98–112, 2013.

[778] A. Meola, F. Cutolo, M. Carbone, F. Cagnazzo, M. Ferrari, and V. Ferrari, "Augmented reality in neurosurgery: A systematic review," *Neurosurgical Review*, vol. 40, no. 4, pp. 537–548, 2017.

[779] R. Wang, Z. Geng, Z. Zhang, and R. Pei, "Visualization techniques for augmented reality in endoscopic surgery," in *International Conference on Medical Imaging and Augmented Reality*, Cham, Switzerland: Springer, pp. 129–138, 2016.

[780] S. Bernhardt, S. A. Nicolau, L. Soler, and C. Doignon, "The status of augmented reality in laparoscopic surgery as of 2016," *Medical Image Analysis*, vol. 37, pp. 66–90, 2017.

[781] H. Monkman and A. W. Kushniruk, "A see through future: Augmented reality and health information systems.," in *ITCH*, pp. 281–285, 2015.

[782] E. Zhu, A. Hadadgar, I. Masiello, and N. Zary, "Augmented reality in healthcare education: An integrative review," *PeerJ*, vol. 2, e469, 2014.

[783] K. Kim, M. Billinghurst, G. Bruder, H. B.-L. Duh, and G. F. Welch, "Revisiting trends in augmented reality research: A review of the 2nd decade of ismar (2008–2017)," *IEEE Transactions on Visualization and Computer Graphics*, vol. 24, no. 11, pp. 2947–2962, 2018.

[784] F. Zhou, H. B.-L. Duh, and M. Billinghurst, "Trends in augmented reality tracking, interaction and display: A review of ten years of ismar," in *2008 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*, IEEE, pp. 193–202, 2008.

[785] R. Azuma, Y. Baillot, R. Behringer, S. Feiner, S. Julier, and B. MacIntyre, "Recent advances in augmented reality," *IEEE Computer Graphics and Applications*, vol. 21, no. 6, pp. 34–47, 2001.

[786] R. T. Azuma, "A survey of augmented reality," *Presence: Teleoperators and Virtual Environments*, vol. 6, no. 4, pp. 355–385, 1997.

[787] I. E. Sutherland, "A head-mounted three dimensional display," in *Proceedings of the December 9–11, 1968, Fall Joint Computer Conference, Part I*, pp. 757–764, 1968.

[788] R. Azuma, B. Hoff, H. Neely, and R. Sarfaty, "A motion-stabilized outdoor augmented reality system," in *Proceedings IEEE Virtual Reality (Cat. No. 99CB36316)*, IEEE, pp. 252–259, 1999.

[789] S. Feiner, B. MacIntyre, T. Höllerer, and A. Webster, "A touring machine: Prototyping 3D mobile augmented reality systems for exploring the urban environment," *Personal Technologies*, vol. 1, no. 4, pp. 208–217, 1997.

[790] Y. Baillot, D. Brown, and S. Julier, "Authoring of physical models using mobile computers," in *Proceedings Fifth International Symposium on Wearable Computers*, IEEE, pp. 39–46, 2001.

[791] T. Höllerer, S. Feiner, T. Terauchi, G. Rashid, and D. Hallaway, "Exploring mars: Developing indoor and outdoor user interfaces to a mobile augmented reality system," *Computers and Graphics*, vol. 23, no. 6, pp. 779–785, 1999.

[792] W. Piekarski and B. H. Thomas, "Tinmith-metro: New outdoor techniques for creating city models with an augmented reality wearable computer," in *Proceedings Fifth International Symposium on Wearable Computers*, IEEE, pp. 31–38, 2001.

[793] B. Thomas, V. Demczuk, W. Piekarski, D. Hepworth, and B. Gunther, "A wearable computer system with augmented reality to support terrestrial navigation," in *Digest of Papers. Second International Symposium on Wearable Computers (Cat. No. 98EX215)*, IEEE, pp. 168–171, 1998.

[794] K. Satoh, "Townwear: An outdoor wearable MR system with high-precision registration," in *Proc. 2nd Int. Symp. on Mixed Reality, 2001*, 2001.

[795] R. Behringer, "Registration for outdoor augmented reality applications using computer vision techniques and hybrid sensors," in *Proceedings IEEE Virtual Reality (Cat. No. 99CB36316)*, IEEE, pp. 244–251, 1999.

[796] M. Ribo, P. Lang, H. Ganster, M. Brandner, C. Stock, and A. Pinz, "Hybrid tracking for outdoor augmented reality applications," *IEEE Computer Graphics and Applications*, vol. 22, no. 6, pp. 54–63, 2002.

[797] L. Vacchetti, V. Lepetit, and P. Fua, "Combining edge and texture information for real-time accurate 3d camera tracking," in *Third IEEE and ACM International Symposium on Mixed and Augmented Reality*, IEEE, pp. 48–56, 2004.

[798] G. Klein and T. Drummond, "Robust visual tracking for non-instrumental augmented reality," in *The Second IEEE and ACM International Symposium on Mixed and Augmented Reality, Proceedings*, IEEE, pp. 113–122, 2003.

[799] G. Klein and T. Drummond, "Sensor fusion and occlusion refinement for tablet-based AR," in *Third IEEE and ACM International Symposium on Mixed and Augmented Reality*, IEEE, pp. 38–47, 2004.

[800] G. Reitmayr and T. W. Drummond, "Going out: Robust model-based tracking for outdoor augmented reality," in *2006 IEEE/ACM International Symposium on Mixed and Augmented Reality*, IEEE, pp. 109–118, 2006.

[801] S. Izadi, D. Kim, O. Hilliges, D. Molyneaux, R. Newcombe, P. Kohli, J. Shotton, S. Hodges, D. Freeman, A. Davison, and A. Fitzgibbon, "Kinectfusion: Real-time 3D reconstruction and interaction using a moving depth camera," in *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology*, pp. 559–568, 2011.

[802] R. F. Salas-Moreno, B. Glocken, P. H. Kelly, and A. J. Davison, "Dense planar SLAM," in *2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, IEEE, pp. 157–164, 2014.

[803] T. Whelan, R. F. Salas-Moreno, B. Glocker, A. J. Davison, and S. Leutenegger, "Elasticfusion: Real-time dense SLAM and light source estimation," *The International Journal of Robotics Research*, vol. 35, no. 14, pp. 1697–1716, 2016.

[804] M. B. Ibáñez, Á. Di Serio, D. Villarán, and C. D. Kloos, "Experimenting with electromagnetism using augmented reality: Impact on flow student experience and educational effectiveness," *Computers and Education*, vol. 71, pp. 1–13, 2014.
[805] Webcam-social-shopper, Zugara, [Online]. Available: http://zugara.com/virtual-dressing-room-technology/webcam-social-shopper.
[806] IKEA Place Demo AR App, Inter IKEA Systems B.V., [Online]. Available: https://newsroom.inter.ikea.com/gallery/video/ikea-place-demo-ar-app/a/c7e1289a-ca7e-4cba-8f65-f84b57e4fb8d.
[807] Charlotte Tilbury Westfield London Magic Mirror, Holition, [Online]. Available: https://vimeo.com/188396641.
[808] Pokemon Go, Niantic, The Pokemon Company, [Online]. Available: https://pokemongolive.com/en/.
[809] Take off to your next destination with Google Maps, Google, [Online]. Available: https://www.blog.google/products/maps/take-your-next-destination-google-maps/.
[810] H. D. Zeman, G. Lovhoiden, C. Vrancken, J. Snodgrass, and J. A. Delong, Projection of subsurface structure onto an object's surface, Dec. 13, US Patent 8,078,263, 2011.
[811] S. Caraiman, A. Stan, N. Botezatu, P. Herghelegiu, R. G. Lupu, and A. Moldoveanu, "Architectural design of a real-time augmented feedback system for neuromotor rehabilitation," in *2015 20th International Conference on Control Systems and Computer Science*, IEEE, pp. 850–855, 2015.
[812] T. R. Coles, N. W. John, D. Gould, and D. G. Caldwell, "Integrating haptics with augmented reality in a femoral palpation and needle insertion training simulation," *IEEE Transactions on Haptics*, vol. 4, no. 3, pp. 199–209, 2011.
[813] K. Abhari, J. S. Baxter, E. C. Chen, A. R. Khan, T. M. Peters, S. de Ribaupierre, and R. Eagleson, "Training for planning tumour resection: Augmented reality and human factors," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 6, pp. 1466–1477, 2014.

[814] A. Elmi-Terander, H. Skulason, M. Söderman, J. Racadio, R. Homan, D. Babic, N. van der Vaart, and R. Nachabe, "Surgical navigation technology based on augmented reality and integrated 3D intraoperative imaging: A spine cadaveric feasibility and accuracy study," *Spine*, vol. 41, no. 21, E1303, 2016.

[815] A. Marmol, A. Banach, and T. Peynot, "Dense-arthroSLAM: Dense intra-articular 3-d reconstruction with robust localization prior for arthroscopy," *IEEE Robotics and Automation Letters*, vol. 4, no. 2, pp. 918–925, 2019.

[816] L. Chen, W. Tang, N. W. John, T. R. Wan, and J. J. Zhang, "SLAM-based dense surface reconstruction in monocular minimally invasive surgery and its application to augmented reality," *Computer Methods and Programs in Biomedicine*, vol. 158, pp. 135–146, 2018.

[817] L. Chen, W. Tang, and N. W. John, "Real-time geometry-aware augmented reality in minimally invasive surgery," *Healthcare Technology Letters*, vol. 4, no. 5, pp. 163–167, 2017.

[818] H. Kato and M. Billinghurst, "Marker tracking and HMD calibration for a video-based augmented reality conferencing system," in *Proceedings 2nd IEEE and ACM International Workshop on Augmented Reality (IWAR'99)*, IEEE, pp. 85–94, 1999.

[819] ArtoolkitX, Realmax Inc., [Online]. Available: http://www.artoolkitx.org/.

[820] Vuforia Engine Developer Portal, PTC Inc., [Online]. Available: https://developer.vuforia.com/.

[821] ARCore, Google Developers, [Online]. Available: https://developers.google.com/ar.

[822] EasyAR, VisionStar Information Technology (Shanghai) Co., Ltd., [Online]. Available: https://www.easyar.com/.

[823] Glass, Google, [Online]. Available: https://www.google.com/glass/start/.

[824] HoloLens2, Microsoft, [Online]. Available: https://www.microsoft.com/en-us/hololens.

[825] M. Z. Zia, L. Nardi, A. Jack, E. Vespa, B. Bodin, P. H. Kelly, and A. J. Davison, "Comparative design space exploration of dense and semi-dense SLAM," in *2016 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, pp. 1292–1299, 2016.

[826] B. Bodin, L. Nardi, M. Z. Zia, H. Wagstaff, G. Sreekar Shenoy, M. Emani, J. Mawer, C. Kotselidis, A. Nisbet, M. Lujan, B. Franke, P. Kelly, and M. O'Boyl, "Integrating algorithmic parameters into benchmarking and design space exploration in 3D scene understanding," in *Proceedings of the 2016 International Conference on Parallel Architectures and Compilation*, pp. 57–69, 2016.

[827] L. Nardi, B. Bodin, S. Saeedi, E. Vespa, A. J. Davison, and P. H. Kelly, "Algorithmic performance-accuracy trade-off in 3d vision applications using hypermapper," in *2017 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW)*, IEEE, pp. 1434–1443, 2017.

[828] S. Saeedi, B. Bodin, H. Wagstaff, A. Nisbet, L. Nardi, J. Mawer, N. Melot, O. Palomar, E. Vespa, T. Spink, C. Gorgovan, A. Webb, J. Clarkson, E. Tomusk, T. Debrunner, K. Kaszyk, P. Gonzalez-de-Aledo, A. Rodchenko, G. Riley, C. Kotselidis, B. Franke, M. F. P. O'Boyle, A. J. Davison, P. H. J. Kelly, M. Lujan, and S. Furber, "Navigating the landscape for real-time localization and mapping for robotics and virtual and augmented reality," *Proceedings of the IEEE*, vol. 106, no. 11, pp. 2020–2039, 2018.

[829] L. Nardi, D. Koeplinger, and K. Olukotun, "Practical design space exploration," in *2019 IEEE 27th International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems (MASCOTS)*, IEEE, pp. 347–358, 2019.

[830] P. Dudek and P. J. Hicks, "A general-purpose processor-per-pixel analog simd vision chip," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 52, no. 1, pp. 13–20, 2005.

[831] A. Moini, *Vision Chips*, vol. 526. Boston, MA: Springer Science and Business Media, 1999.

[832] Á. Zarándy, *Focal-Plane Sensor-Processor Chips*. Budapest, Hungary: Springer Science and Business Media, 2011.

[833] S. J. Carey, D. R. Barr, and P. Dudek, "Low power high-performance smart camera system based on scamp vision sensor," *Journal of Systems Architecture*, vol. 59, no. 10, pp. 889–899, 2013.

[834] J. N. Martel and P. Dudek, "Vision chips with in-pixel processors for high-performance low-power embedded vision systems," in *ASR-MOV Workshop, CGO*, vol. 6, p. 14, 2016.

[835] M. Wong, S. Saeedi, and P. H. Kelly, "Analog vision-neural network inference acceleration using analog simd computation in the focal plane," Ph.D. dissertation, Master's thesis, Imperial College London-Department of Computing, London, 2018.

[836] C. Mead, "Neuromorphic electronic systems," in *Proceedings of the IEEE*, vol. 78, no. 10, pp. 1629–1636, 1990.

[837] K. Aizawa, "Computational sensors–vision vlsi," *IEICE Transactions on Information and Systems*, vol. 82, no. 3, pp. 580–588, 1999.

[838] J. Ruiz-Rosero, G. Ramirez-Gonzalez, and R. Khanna, "Field programmable gate array applications—A scientometric review," *Computation*, vol. 7, no. 4, p. 63, 2019.

[839] S. Mittal, "A survey of FPGA-based accelerators for convolutional neural networks," *Neural Computing and Applications*, pp. 1–31, 2020.

[840] A. G. Blaiech, K. B. Khalifa, C. Valderrama, M. A. Fernandes, and M. H. Bedoui, "A survey and taxonomy of FPGA-based deep learning accelerators," *Journal of Systems Architecture*, vol. 98, pp. 331–345, 2019.

[841] T. Wang, C. Wang, X. Zhou, and H. Chen, "A survey of FPGA based deep learning accelerators: Challenges and opportunities," *arXiv preprint arXiv:1901.04988*, pp. 1–10, 2018.

[842] A. Shawahna, S. M. Sait, and A. El-Maleh, "FPGA-based accelerators of deep learning networks for learning and classification: A review," *IEEE Access*, vol. 7, pp. 7823–7859, 2018.

[843] S. I. Venieris, A. Kouris, and C.-S. Bouganis, "Toolflows for mapping convolutional neural networks on FPGAs: A survey and future directions," *ACM Computing Surveys (CSUR)*, vol. 51, no. 3, pp. 1–39, 2018.

[844] C. Koch and B. Mathur, "Neuromorphic vision chips," *IEEE Spectrum*, vol. 33, no. 5, pp. 38–46, 1996.

[845] N. Wu, "Neuromorphic vision chips," *Science China Information Sciences*, vol. 61, no. 6, 2018.

[846] C. D. Schuman, T. E. Potok, R. M. Patton, J. D. Birdwell, M. E. Dean, G. S. Rose, and J. S. Plank, "A survey of neuromorphic computing and neural networks in hardware," *arXiv preprint arXiv:1705.06963*, 2017.

[847] R. Tapiador-Morales, A. Rios-Navarro, J. P. Dominguez-Morales, D. Gutierrez-Galan, M. Dominguez-Morales, A. Jimenez-Fernandez, and A. Linares-Barranco, "Event-based row-by-row multi-convolution engine for dynamic-vision feature extraction on fpga," in *2018 International Joint Conference on Neural Networks (IJCNN)*, IEEE, pp. 1–7, 2018.

[848] J. M. Cruz-Albrecht, M. W. Yung, and N. Srinivasa, "Energy-efficient neuron, synapse and STDP integrated circuits," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 6, no. 3, pp. 246–256, 2012.

[849] G. Tang, A. Shah, and K. P. Michmizos, "Spiking neural network on neuromorphic hardware for energy-efficient unidimensional SLAM," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, pp. 4176–4181, 2019.

[850] Y. Huang-Yu, H.-P. Huang, Y.-C. Huang, and C.-C. Lo, "Flyintel—A platform for robot navigation based on a brain-inspired spiking neural network," in *2019 IEEE International Conference on Artificial Intelligence Circuits and Systems (AICAS)*, IEEE, pp. 219–220, 2019.

[851] M. Beyeler, N. Oros, N. Dutt, and J. L. Krichmar, "A gpu-accelerated cortical neural network model for visually guided robot navigation," *Neural Networks*, vol. 72, pp. 75–87, 2015.

[852] T. C. Stewart, A. Kleinhans, A. Mundy, and J. Conradt, "Serendipitous offline learning in a neuromorphic robot," *Frontiers in Neurorobotics*, vol. 10, p. 1, 2016.

[853] A. Sengupta, Y. Ye, R. Wang, C. Liu, and K. Roy, "Going deeper in spiking neural networks: Vgg and residual architectures," *Frontiers in Neuroscience*, vol. 13, p. 95, 2019.

[854] F. Xing, Y. Yuan, H. Huo, and T. Fang, "Homeostasis-based CNN-to-SNN conversion of inception and residual architectures," in *International Conference on Neural Information Processing*, Cham, Switzerland: Springer, pp. 173–184, 2019.

[855] Y. Cao, Y. Chen, and D. Khosla, "Spiking deep convolutional neural networks for energy-efficient object recognition," *International Journal of Computer Vision*, vol. 113, no. 1, pp. 54–66, 2015.

[856] S. Song, S. P. Lichtenberg, and J. Xiao, "SUN RGB-D: A RGB-D scene understanding benchmark suite," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 7–12-Jun., no. Figure 1, pp. 567–576, 2015.

[857] B. S. Hua, Q. H. Pham, D. T. Nguyen, M. K. Tran, L. F. Yu, and S. K. Yeung, "SceneNN: A scene meshes dataset with annotations," in *Proceedings—2016 4th International Conference on 3D Vision, 3DV 2016*, pp. 92–101, 2016.

[858] A. Chang, A. Dai, T. Funkhouser, M. Halber, M. Niebner, M. Savva, S. Song, A. Zeng, and Y. Zhang, "Matterport3D: Learning from RGB-D data in indoor environments," in *2017 International Conference on 3D Vision (3DV)*, IEEE, pp. 667–676, 2017.

[859] K. Mangalam, E. Adeli, K.-H. Lee, A. Gaidon, and J. C. Niebles, "Disentangling human dynamics for pedestrian locomotion forecasting with noisy supervision," in *The IEEE Winter Conference on Applications of Computer Vision*, pp. 2784–2793, 2020.

[860] R. Fan, H. Wang, P. Cai, and M. Liu, "Sne-roadseg: Incorporating surface normal information into semantic segmentation for accurate freespace detection," in *European Conference on Computer Vision*, Cham, Switzerland: Springer, pp. 340–356, 2020.

[861] P. Pandey, A. Kumar, and A. Prathosh, "Unsupervised domain adaptation for semantic segmentation of NIR images through generative latent search," in *European Conference on Computer Vision*, Cham, Switzerland: Springer, 2020.

[862] J. R. Ruiz-Sarmiento, C. Galindo, and J. Gonzalez-Jimenez, "Robot@Home, a robotic dataset for semantic mapping of home environments," *International Journal of Robotics Research,* vol. 36, no. 2, pp. 131–141, 2017.

[863] H. Yu, Z. Yang, L. Tan, Y. Wang, W. Sun, M. Sun, and Y. Tang, "Methods and datasets on semantic segmentation: A review," *Neurocomputing*, vol. 304, pp. 82–103, 2018.

[864] mrgloom, Awesome Semantic Segmentation, GitHub, [Online]. Available: https://github.com/mrgloom/awesome-semantic-segmentation.

[865] A. Dai, A. X. Chang, M. Savva, M. Halber, T. Funkhouser, and M. Nießner, "ScanNet: Richly-annotated 3D reconstructions of indoor scenes," in *Proceedings—30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, vol. 2017-Jan., pp. 2432–2443, 2017.

[866] T. Hackel, N. Savinov, L. Ladicky, J. D. Wegner, K. Schindler, and M. Pollefeys, "Semantic3D.Net: A new large-scale point cloud classification benchmark," *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 4, no. 1W1, pp. 91–98, 2017.

[867] I. Armeni, S. Sax, A. R. Zamir, and S. Savarese, "Joint 2d-3d-semantic data for indoor scene understanding," *arXiv preprint arXiv:1702.01105*, 2017.

[868] Q. Wang, Z. Mao, B. Wang, and L. Guo, "Knowledge graph embedding: A survey of approaches and applications," *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, no. 12, pp. 2724–2743, 2017.

[869] H. Paulheim, "Knowledge graph refinement: A survey of approaches and evaluation methods," *Semantic Web*, vol. 8, no. 3, pp. 489–508, 2017.

[870] D. L. McGuinness and F. Van Harmelen, "Owl web ontology language overview," *W3C Recommendation*, vol. 10, no. 10, p. 2004, 2004.

[871] D. Martin, M. Burstein, J. Hobbs, O. Lassila, D. McDermott, S. McIlraith, S. Narayanan, M. Paolucci, B. Parsia, T. Payne, E. Sirin, N. Srinivasan, and K. Sycara, "Owl-s: Semantic markup for web services," *W3C Member Submission*, vol. 22, no. 4, 2004.

[872] K. Janowicz, A. Haller, S. J. Cox, D. Le Phuoc, and M. Lefrançois, "Sosa: A lightweight ontology for sensors, observations, samples, and actuators," *Journal of Web Semantics*, vol. 56, pp. 1–10, 2019.

[873] E. Prestes, J. L. Carbonera, S. R. Fiorini, V. A. Jorge, M. Abel, R. Madhavan, A. Locoro, P. Goncalves, M. E. Barreto, M. Habib, A. Chibani, S. Gérard, Y. Amirat, and C. Schlenoff, "Towards a core ontology for robotics and automation," *Robotics and Autonomous Systems*, vol. 61, no. 11, pp. 1193–1204, 2013.

[874] C. Schlenoff, S. Balakirsky, M. Uschold, R. Provine, and S. Smith, "Using ontologies to aid navigation planning in autonomous vehicles," *The Knowledge Engineering Review*, vol. 18, no. 3, pp. 243–255, 2003.

[875] M. K. Habib and S. Yuta, "Map representation of a large in-door environment with path planning and navigation abilities for an autonomous mobile robot with its implementation on a real robot," *Automation in Construction*, vol. 2, no. 2, pp. 155–179, 1993.

[876] S. Balakirsky and C. Scrapper, "Knowledge representation and planning for on-road driving," *Robotics and Autonomous Systems*, vol. 49, no. 1–2, pp. 57–66, 2004.

[877] A. Chella, M. Cossentino, R. Pirrone, and A. Ruisi, "Modeling ontologies for robotic environments," in *Proceedings of the 14th International Conference on Software Engineering and Knowledge Engineering*, pp. 77–80, 2002.

[878] T. Barbera, J. Albus, E. Messina, C. Schlenoff, and J. Horst, "How task analysis can be used to derive and organize the knowledge for the control of autonomous vehicles," *Robotics and Autonomous Systems*, vol. 49, no. 1–2, pp. 67–78, 2004.

[879] S. Wood, "Representation and purposeful autonomous agents," *Robotics and Autonomous Systems*, vol. 49, no. 1–2, pp. 79–90, 2004.

[880] S. L. Epstein, "Metaknowledge for autonomous systems," in *Proceedings of AAAI Spring Symposium on Knowledge Representation and Ontology for Autonomous Systems*, pp. 61–68, 2004.

[881] H. Jung, J. M. Bradshaw, S. Kulkarni, M. Breedy, L. Bunch, P. Feltovich, R. Jeffers, M. Johnson, J. Lott, N. Suri, W. Taysom, G. Tonti, and A. Uszok, "An ontology-based representation for policy-governed adjustable autonomy," in *Proceedings of the AAAI Spring Symposium on Knowledge Representation and Ontology for Autonomous Systems*, 2004.

[882] C. Schlenoff and E. Messina, "A robot ontology for urban search and rescue," in *Proceedings of the 2005 ACM Workshop on Research in Knowledge Representation for Autonomous Systems*, pp. 27–34, 2005.

[883] S. Dhouib, N. Du Lac, J.-L. Farges, S. Gerard, M. Hemaissia-Jeannin, J. Lahera-Perez, S. Millet, B. Patin, and S. Stinckwich, "Control architecture concepts and properties of an ontology devoted to exchanges in mobile robotics," in *6th National Conference on Control Architectures of Robots*, 2011.

[884] J. Hallam and H. Bruyninckx, "An ontology of robotics science," in *European Robotics Symposium 2006*, Berlin and Heidelberg, Germany: Springer, pp. 1–14, 2006.

[885] R. Davis, H. Shrobe, and P. Szolovits, "What is a knowledge representation?" *AI Magazine*, vol. 14, no. 1, pp. 17–33, 1993.

[886] R. Wood, P. Baxter, and T. Belpaeme, "A review of long-term memory in natural and synthetic systems," *Adaptive Behavior*, vol. 20, no. 2, pp. 81–103, 2012.

[887] T. Fischer, C. Moulin-Frier, M. Petit, G. Pointeau, J.-Y. Puigbo, U. Pattacini, S. C. Low, D. Camilleri, P. Nguyen, M. Hoffmann, H. J. Chang, M. Zambelli, A.-L. Mealier, A. Damianou, G. Metta, T. J. Prescott, Y. Demiris, P. Ford Dominey, and P. F. M. J. Verschure, "Dac-h3: A proactive robot cognitive architecture to acquire and express knowledge about the world and the self," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 10, no. 4, pp. 1005–1022, 2018.

[888] S. Lemaignan, R. Ros, L. Mösenlechner, R. Alami, and M. Beetz, "Oro, a knowledge management platform for cognitive architectures in robotics," in *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, IEEE, pp. 3548–3553, 2010.

[889] M. Tenorth and M. Beetz, "Knowrob—Knowledge processing for autonomous personal robots," in *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, IEEE, pp. 4261–4266, 2009.

[890] R. Gayathri and V. Uma, "Ontology based knowledge representation technique, domain modeling languages and planners for robotic path planning: A survey," *ICT Express*, vol. 4, no. 2, pp. 69–74, 2018.

[891] C. Galindo, A. Saffiotti, S. Coradeschi, P. Buschka, J.-A. Fernandez-Madrigal, and J. González, "Multi-hierarchical semantic maps for mobile robotics," in *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*, IEEE, pp. 2278–2283, 2005.

[892] P. F. Patel-Schneider, M. Abrahams, L. A. Resnick, D. L. McGuinness, and A. Borgida, "Neoclassic reference manual: Version 1.0," Artificial Intelligence Principles Research Department, AT&T Labs Research, 1996.

[893] H. Zender, O. M. Mozos, P. Jensfelt, G.-J. Kruijff, and W. Burgard, "Conceptual spatial representations for indoor mobile robots," *Robotics and Autonomous Systems*, vol. 56, no. 6, pp. 493–502, 2008.

[894] M. Tenorth, L. Kunze, D. Jain, and M. Beetz, "Knowrob-map-knowledge-linked semantic object maps," in *2010 10th IEEE-RAS International Conference on Humanoid Robots*, IEEE, pp. 430–435, 2010.

[895] D. B. Lenat, "Cyc: A large-scale investment in knowledge infrastructure," *Communications of the ACM*, vol. 38, no. 11, pp. 33–38, 1995.

[896] R. Gupta, M. J. Kochenderfer, D. Mcguinness, and G. Ferguson, "Common sense data acquisition for indoor mobile robots," in *AAAI*, pp. 605–610, 2004.

[897] G. Mohanarajah, D. Hunziker, R. D'Andrea, and M. Waibel, "Rapyuta: A cloud robotics platform," *IEEE Transactions on Automation Science and Engineering*, vol. 12, no. 2, pp. 481–493, 2014.

[898] L. Kunze, T. Roehm, and M. Beetz, "Towards semantic robot description languages," in *2011 IEEE International Conference on Robotics and Automation*, IEEE, pp. 5589–5595, 2011.

[899] 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, 2017.

[900] T. Kulvicius, I. Markelic, M. Tamosiunaite, and F. Wörgötter, "Semantic image search for robotic applications," in *22nd International Workshop on Robotics in Alpe-Adria-Danube Region*, Portorož, Slovenia: Jožef Stefan Institute, pp. 1–8, 2013.

[901] E. N. Alemu and J. Huang, "Healthaid: Extracting domain targeted high precision procedural knowledge from on-line communities," *Information Processing and Management*, vol. 57, no. 6, 2020.

[902] D. Leidner, C. Borst, and G. Hirzinger, "Things are made for what they are: Solving manipulation tasks by using functional object classes," in *2012 12th IEEE-RAS International Conference on Humanoid Robots (Humanoids 2012)*, IEEE, pp. 429–435, 2012.

[903] H. L. Younes and M. L. Littman, "Ppddl1. 0: An extension to pddl for expressing planning domains with probabilistic effects," Techn. Rep. CMU-CS-04-162, vol. 2, p. 99, 2004.

[904] A. S. Bauer, P. Schmaus, F. Stulp, and L. Daniel, "Probabilistic effect prediction through semantic augmentation and physical simulation," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 2020.

[905] M. Tenorth, U. Klank, D. Pangercic, and M. Beetz, "Web-enabled robots," *IEEE Robotics and Automation Magazine*, vol. 18, no. 2, pp. 58–68, 2011.

[906] M. Beetz, U. Klank, I. Kresse, A. Maldonado, L. Mösenlechner, D. Pangercic, T. Rühr, and M. Tenorth, "Robotic roommates making pancakes," in *2011 11th IEEE-RAS International Conference on Humanoid Robots*, IEEE, pp. 529–536, 2011.

[907] M. Tamosiunaite, I. Markelic, T. Kulvicius, and F. Wörgötter, "Generalizing objects by analyzing language," in *2011 11th IEEE-RAS International Conference on Humanoid Robots*, IEEE, pp. 557–563, 2011.

[908] J. Kuffner, "Cloud-enabled humanoid robots," in *2010 10th IEEE-RAS International Conference on Humanoid Robots (Humanoids),* Nashville TN, United States, Dec., 2010.

[909] K. Goldberg and R. Siegwart, *Beyond Webcams: An Introduction to Online Robots*. Cambridge, MA: MIT Press, 2002.

[910] K. Goldberg, M. Mascha, S. Gentner, N. Rothenberg, C. Sutter, and J. Wiegley, "Desktop teleoperation via the world wide web," in *Proceedings of 1995 IEEE International Conference on Robotics and Automation*, IEEE, vol. 1, pp. 654–659, 1995.

[911] G. McKee, "What is networked robotics?" In *Informatics in Control Automation and Robotics*, Berlin and Heidelberg, Germany: Springer, 2008, pp. 35–45.

[912] S. Davidson, "Open-source hardware," *IEEE Design and Test of Computers*, vol. 21, no. 5, pp. 456–456, 2004.

[913] K. Ashton, "That "internet of things' thing," *RFID Journal*, vol. 22, no. 7, pp. 97–114, 2009.

[914] L. Atzori, A. Iera, and G. Morabito, "The internet of things: A survey," *Computer Networks*, vol. 54, no. 15, pp. 2787–2805, 2010.

[915] K. Goldberg and B. Kehoe, "Cloud robotics and automation: A survey of related work," EECS Department, University of California, Berkeley, Tech. Rep. UCB/EECS-2013-5, 2013.

[916] B. Kehoe, S. Patil, P. Abbeel, and K. Goldberg, "A survey of research on cloud robotics and automation," *IEEE Transactions on Automation Science and Engineering*, vol. 12, no. 2, pp. 398–409, 2015.

[917] J. Wan, S. Tang, H. Yan, D. Li, S. Wang, and A. V. Vasilakos, "Cloud robotics: Current status and open issues," *IEEE Access*, vol. 4, pp. 2797–2807, 2016.

[918] G. Hu, W. P. Tay, and Y. Wen, "Cloud robotics: Architecture, challenges and applications," *IEEE Network*, vol. 26, no. 3, pp. 21–28, 2012.

[919] L. Carlone, M. K. Ng, J. Du, B. Bona, and M. Indri, "Rao-blackwellized particle filters multi robot SLAM with unknown initial correspondences and limited communication," in *2010 IEEE International Conference on Robotics and Automation*, IEEE, pp. 243–249, 2010.

[920] A. Doucet, N. de Freitas, K. Murphy, and S. Russell, "Rao-blackwellised particle filtering for dynamic Bayesian networks," in *Proceedings of the Sixteenth Conference on Uncertainty in Artificial Intelligence*, pp. 176–183, 2000.

[921] A. Chowdhury, S. Mukherjee, and S. Banerjee, "An approach towards survey and analysis of cloud robotics," in *Robotic Systems: Concepts, Methodologies, Tools, and Applications*, Pennsylvania: IGI Global, 2020, pp. 1979–2002.

[922] O. Saha and P. Dasgupta, "A comprehensive survey of recent trends in cloud robotics architectures and applications," *Robotics*, vol. 7, no. 3, p. 47, 2018.

[923] W. Chen, Y. Yaguchi, K. Naruse, Y. Watanobe, K. Nakamura, and J. Ogawa, "A study of robotic cooperation in cloud robotics: Architecture and challenges," *IEEE Access*, vol. 6, pp. 36 662–36 682, 2018.

[924] A. Koubâa, "Service-oriented software architecture for cloud robotics," in *Encyclopedia of Robotics.* Ed. by M. Ang, O. Khatib, and B. Siciliano, Berlin and Heidelberg, Germany: Springer, 2020.

[925] S. Chinchali, A. Sharma, J. Harrison, A. Elhafsi, D. Kang, E. Pergament, E. Cidon, S. Katti, and M. Pavone, "Network offloading policies for cloud robotics: A learning-based approach," in *Robotics: Science and Systems*, 2019.

[926] E. Fosch-Villaronga and C. Millard, "Cloud robotics law and regulation: Challenges in the governance of complex and dynamic cyber–physical ecosystems," *Robotics and Autonomous Systems*, vol. 119, pp. 77–91, 2019.

[927] A. Botta, L. Gallo, and G. Ventre, "Cloud, fog, and dew robotics: Architectures for next generation applications," in *2019 7th IEEE International Conference on Mobile Cloud Computing, Services, and Engineering (MobileCloud)*, IEEE, pp. 16–23, 2019.

[928] G. Klein and D. Murray, "Parallel tracking and mapping for small AR workspaces," in *2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality*, IEEE, pp. 225–234, 2007.

[929] M. Bloesch, J. Czarnowski, R. Clark, S. Leutenegger, and A. J. Davison, "CodeSLAM—learning a compact, optimisable representation for dense visual SLAM," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* 2560–2568, 2018.

[930] L. Riazuelo, J. Civera, and J. M. Montiel, "C2tam: A cloud framework for cooperative tracking and mapping," *Robotics and Autonomous Systems*, vol. 62, no. 4, pp. 401–413, 2014.

[931] P. Zhang, H. Wang, B. Ding, and S. Shang, "Cloud-based framework for scalable and real-time multi-robot SLAM," in *2018 IEEE International Conference on Web Services (ICWS)*, IEEE, pp. 147–154, 2018.

[932] S. Choudhary, L. Carlone, C. Nieto, J. Rogers, Z. Liu, H. I. Christensen, and F. Dellaert, "Multi robot object-based SLAM," in *International Symposium on Experimental Robotics*, Cham, Switzerland: Springer, pp. 729–741, 2016.

[933] B. Kim, M. Kaess, L. Fletcher, J. Leonard, A. Bachrach, N. Roy, and S. Teller, "Multiple relative pose graphs for robust cooperative mapping," in *2010 IEEE International Conference on Robotics and Automation*, IEEE, pp. 3185–3192, 2010.

[934] D. Fox, J. Ko, K. Konolige, B. Limketkai, D. Schulz, and B. Stewart, "Distributed multirobot exploration and mapping," in *Proceedings of the IEEE*, vol. 94, no. 7, pp. 1325–1339, 2006.

[935] H. Liu, S. Liu, and K. Zheng, "A reinforcement learning-based resource allocation scheme for cloud robotics," *IEEE Access*, vol. 6, pp. 17 215–17 222, 2018.

[936] A. Koubâa and B. Qureshi, "Dronetrack: Cloud-based real-time object tracking using unmanned aerial vehicles over the internet," *IEEE Access*, vol. 6, pp. 13 810–13 824, 2018.

[937] A. Koubâa, B. Qureshi, M.-F. Sriti, A. Allouch, Y. Javed, M. Alajlan, O. Cheikhrouhou, M. Khalgui, and E. Tovar, "Dronemap planner: A service-oriented cloud-based management system for the internet-of-drones," *Ad Hoc Networks*, vol. 86, pp. 46–62, 2019.

[938] M. Montemerlo, S. Thrun, D. Koller, and B. Wegbreit, "Fast-SLAM 2.0: An improved particle filtering algorithm for simultaneous localization and mapping that provably converges," in *IJCAI*, pp. 1151–1156, 2003.

[939] S. S. Ali, A. Hammad, and A. S. T. Eldien, "Fastslam 2.0 tracking and mapping as a cloud robotics service," *Computers and Electrical Engineering*, vol. 69, pp. 412–421, 2018.

[940] H. Wu, X. Wu, Q. Ma, and G. Tian, "Cloud robot: Semantic map building for intelligent service task," *Applied Intelligence*, vol. 49, no. 2, pp. 319–334, 2019.

[941] Apache CloudStack, The Apache Software Foundation, [Online]. Available: http://cloudstack.apache.org/.

[942] J. Martin, S. May, S. Endres, and I. Cabanes, "Decentralized robot-cloud architecture for an autonomous transportation system in a smart factory," in *SEMANTICS Workshops*, 2017.

[943] L. Ballotta, L. Schenato, and L. Carlone, "From sensor to processing networks: Optimal estimation with computation and communication latency," in *21st IFAC World Congress*, 2020.

[944] D. Roy, P. Panda, and K. Roy, "Tree-CNN: A hierarchical deep convolutional neural network for incremental learning," *Neural Networks*, vol. 121, pp. 148–160, 2020.

[945] J. Zhang, J. Zhang, S. Ghosh, D. Li, S. Tasci, L. Heck, H. Zhang, and C.-C. J. Kuo, "Class-incremental learning via deep model consolidation," in *The IEEE Winter Conference on Applications of Computer Vision*, pp. 1131–1140, 2020.

[946] S. Garg and M. Milford, "Fast, compact and highly scalable visual place recognition through sequence-based matching of overloaded representations," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 2020.

[947] W.-C. Chang, F. X. Yu, Y.-W. Chang, Y. Yang, and S. Kumar, "Pre-training tasks for embedding-based large-scale retrieval," in *2020 International Conference on Learning Representations (ICLR)*, 2020.

[948] Z. Ye, G. Zhang, and H. Bao, "Efficient covisibility-based image matching for large-scale sfm," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 2020.

[949] Y. Chen, S. Shen, Y. Chen, and G. Wang, "Graph-based parallel large scale structure from motion," *Pattern Recognition*, 2020.

[950] H. Yang and L. Carlone, "One ring to rule them all: Certifiably robust geometric perception with outliers," in *Advances in Neural Information Processing Systems*, 2020.

[951] H. Yang, P. Antonante, V. Tzoumas, and L. Carlone, "Graduated non-convexity for robust spatial perception: From non-minimal solvers to global outlier rejection," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 1127–1134, 2020.

[952] A. Kapitonov, S. Lonshakov, I. Berman, E. C. Ferrer, F. P. Bonsignorio, V. Bulatov, and A. Svistov, "Robotic services for new paradigm smart cities based on decentralized technologies," *Ledger*, vol. 4, no. 1, 2019.

[953] T.-H. Kim, C. Ramos, and S. Mohammed, "Smart city and IoT," *Future Generation Computer Systems*, vol. 76, pp. 159–162, 2017.

[954] C. Breazeal, "Role of expressive behaviour for robots that learn from people," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 364, no. 1535, pp. 3527–3538, 2009.

[955] C. Fang, X. Ding, C. Zhou, and N. Tsagarakis, "A 2 ML: A general human-inspired motion language for anthropomorphic arms based on movement primitives," *Robotics and Autonomous Systems*, vol. 111, pp. 145–161, 2019.

[956] J. Or, "Towards the development of emotional dancing humanoid robots," *International Journal of Social Robotics*, vol. 1, no. 4, p. 367, 2009.

[957] E. A. Jochum and J. Derks, "Tonight we improvise!: Real-time tracking for human-robot improvisational dance," in *Proceedings of the 6th International Conference on Movement and Computing*, pp. 1–11, 2019.

[958] H. Miwa, K. Itoh, M. Matsumoto, M. Zecca, H. Takanobu, S. Rocella, M. C. Carrozza, P. Dario, and A. Takanishi, "Effective emotional expressions with expression humanoid robot we-4rii: Integration of humanoid robot hand rch-1," in *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (IEEE Cat. No. 04CH37566)*, IEEE, vol. 3, pp. 2203–2208, 2004.

[959] Y.-H. Lai and S.-H. Lai, "Emotion-preserving representation learning via generative adversarial network for multi-view facial expression recognition," in *2018 13th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2018)*, pp. 263–270, 2018.

[960] A. Giambattista, L. Teixeira, H. Ayanoğlu, M. Saraiva, and E. Duarte, "Expression of emotions by a service robot: A pilot study," in *International Conference of Design, User Experience, and Usability*, Cham, Switzerland: Springer, pp. 328–336, 2016.

[961] B. Talbot, F. Dayoub, P. Corke, and G. Wyeth, "Robot navigation in unseen spaces using an abstract map," *IEEE Transactions on Cognitive and Developmental Systems*, 2020.

[962] J. Moon and B.-H. Lee, "Object-oriented semantic graph based natural question generation," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 2020.