sensors

MDPI

*Review*

# From SLAM to Situational Awareness: Challenges and Survey

**Hriday Bavle [1,*], Jose Luis Sanchez-Lopez [1], Claudio Cimarelli [1], Ali Tourani [1] and Holger Voos [1,2]**

[1] Interdisciplinary Center for Security Reliability and Trust (SnT), University of Luxembourg, 1855 Luxembourg, Luxembourg; joseluis.sanchezlopez@uni.lu (J.L.S.-L.); claudio.cimarelli@uni.lu (C.C.); ali.tourani@uni.lu (A.T.); holger.voos@uni.lu (H.V.)

[2] Department of Engineering, Faculty of Science, Technology, and Medicine (FSTM), University of Luxembourg, 1359 Luxembourg, Luxembourg

* Correspondence: hriday.bavle@uni.lu

**Abstract:** The capability of a mobile robot to efficiently and safely perform complex missions is limited by its knowledge of the environment, namely the *situation*. Advanced reasoning, decision-making, and execution skills enable an intelligent agent to act autonomously in unknown environments. Situational Awareness (SA) is a fundamental capability of humans that has been deeply studied in various fields, such as psychology, military, aerospace, and education. Nevertheless, it has yet to be considered in robotics, which has focused on single compartmentalized concepts such as sensing, spatial perception, sensor fusion, state estimation, and Simultaneous Localization and Mapping (SLAM). Hence, the present research aims to connect the broad multidisciplinary existing knowledge to pave the way for a complete SA system for mobile robotics that we deem paramount for autonomy. To this aim, we define the principal components to structure a robotic SA and their area of competence. Accordingly, this paper investigates each aspect of SA, surveying the state-of-the-art robotics algorithms that cover them, and discusses their current limitations. Remarkably, essential aspects of SA are still immature since the current algorithmic development restricts their performance to only specific environments. Nevertheless, Artificial Intelligence (AI), particularly Deep Learning (DL), has brought new methods to bridge the gap that maintains these fields apart from the deployment to real-world scenarios. Furthermore, an opportunity has been discovered to interconnect the vastly fragmented space of robotic comprehension algorithms through the mechanism of *Situational Graph (S-Graph)*, a generalization of the well-known scene graph. Therefore, we finally shape our vision for the future of robotic situational awareness by discussing interesting recent research directions.

**Keywords:** simultaneous localization and mapping; scene understanding; scene graphs; mobile robots

check for updates

**Citation:** Bavle, H.; Sanchez-Lopez, J.L.; Cimarelli, C.; Tourani, A.; Voos, H. From SLAM to Situational Awareness: Challenges and Survey. *Sensors* **2023**, *23*, 4849. https://doi.org/10.3390/s23104849

Academic Editor: Andrey V. Savkin

Received: 10 April 2023
Revised: 27 April 2023
Accepted: 13 May 2023
Published: 17 May 2023

## 1. Introduction

The robotics industry is experiencing an exponential growth, embarking on newer technological advancements and applications. Mobile robots have gained interest from a commercial perspective due to their capabilities to replace or aid humans in repetitive or dangerous tasks [1]. Already, many industrial and civil-related applications employ mobile robots [2]. For example, industrial machines and underground mines' inspections, surveillance and road-traffic monitoring, civil engineering, agriculture, healthcare, search, and rescue interventions in extreme environments, e.g., natural disasters, for exploration and logistics [3].

On one hand, mobile robots can be controlled in manual teleoperation or semi-autonomous mode with constant human intervention in the loop. Furthermore, teleoperated mobile robots can be apprehended using applications such as augmented reality (AR) [4], ref. [5] to enhance human–robot interactions. On the other hand, in fully autonomous mode, a robot performs an entire mission based on its understanding of the environment given only a few commands [6]. Remarkably, autonomy reduces the costs and risks while increasing productivity and is the goal of current research to solve the main challenges that it raises [7]. Unlike the industrial scenario, where autonomous agents can act in a controlled environment, mobile robots operate in the dynamic, unstructured, and cluttered world domain with little or zero previous knowledge of the scene structure.

Up to now, the robotics community has focused chiefly on research areas such as sensing, perception, sensor fusion, data analysis, state estimation, localization and mapping, i.e., Simultaneous Localization and Mapping (SLAM), and Artificial Intelligence (AI) applied to various image processing problems, in a compartmentalized manner. Figure 1 shows the mentioned targets' data obtained from *Scopus* abstract and citation database. However, autonomous behavior entails understanding the situation encompassing multiple interdisciplinary aspects of robotics, from perception, control, and planning to human–robot interaction. Although SA [8] is a holistic concept widely studied in fields such as psychology, the military, and in aerospace, it has been barely considered in robotics. Notably, Endsley [9] formally defined SA in the 1990s as "*the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning and the projection of their status shortly*," which remains valid to date [10]. Hence, we turn this definition into the perspective of mobile robotics to derive a unified field of research that garners all the aspects required by an autonomous system.

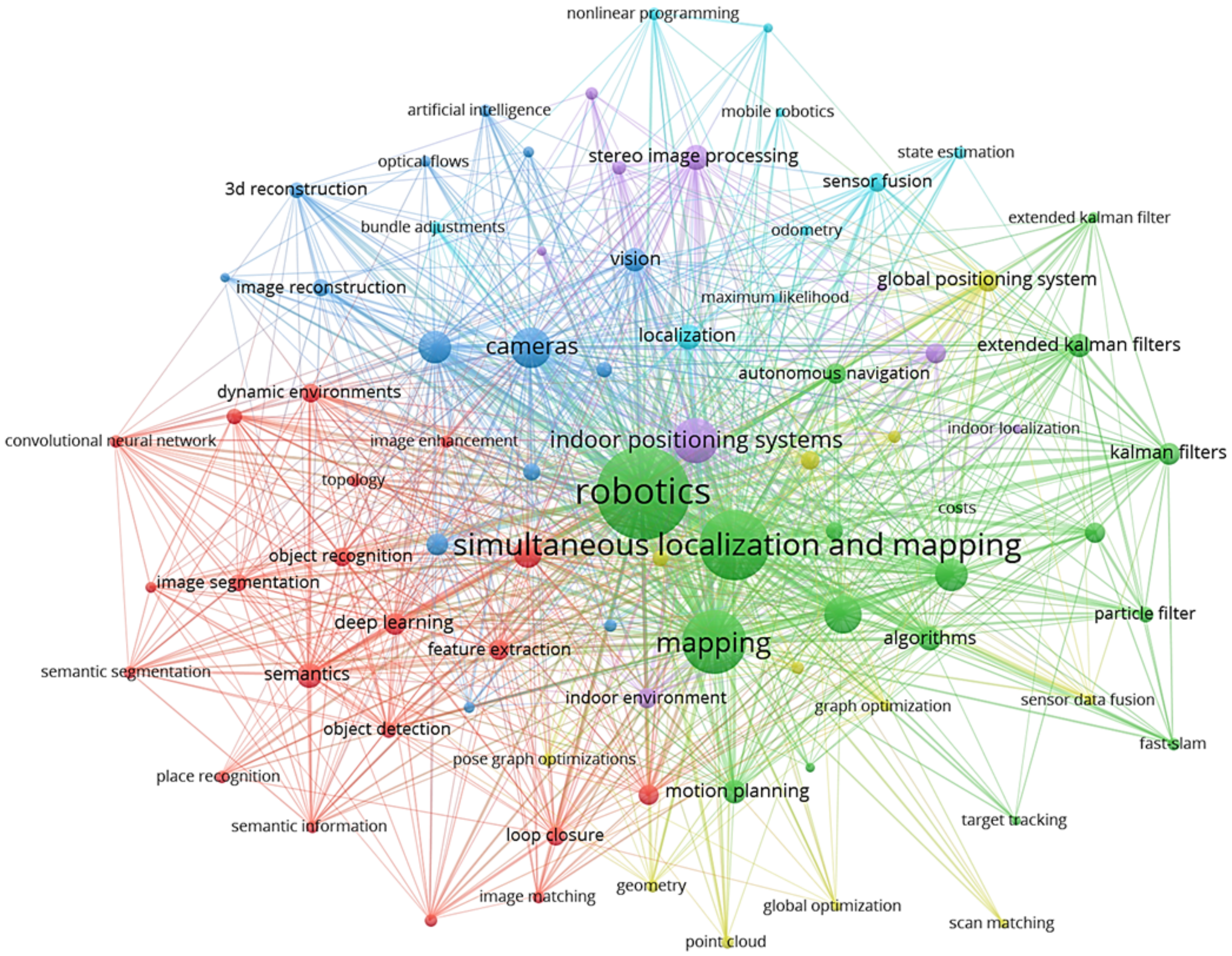


**Figure 1.** Scopus database since 2015 covering the research in *Robotics* and *SLAM*. All the works focused on independent research areas which could be efficiently encompassed in one field of Situational Awareness for robots.

Therefore, a robot's Situational Awareness (SA) system must continuously acquire new observations of the surroundings, understand its essential elements and make complex reasoning, and project the world state into a possible future outcome to make decisions and execute actions that would let it accomplish its goals. Accordingly, we depict in Figure 2 a general architecture of SA that divides the specific competence areas into three layers, with an increasing level of intelligence. Thus, we raise the following research question:

- *What are the components of a robot's situational awareness system?*

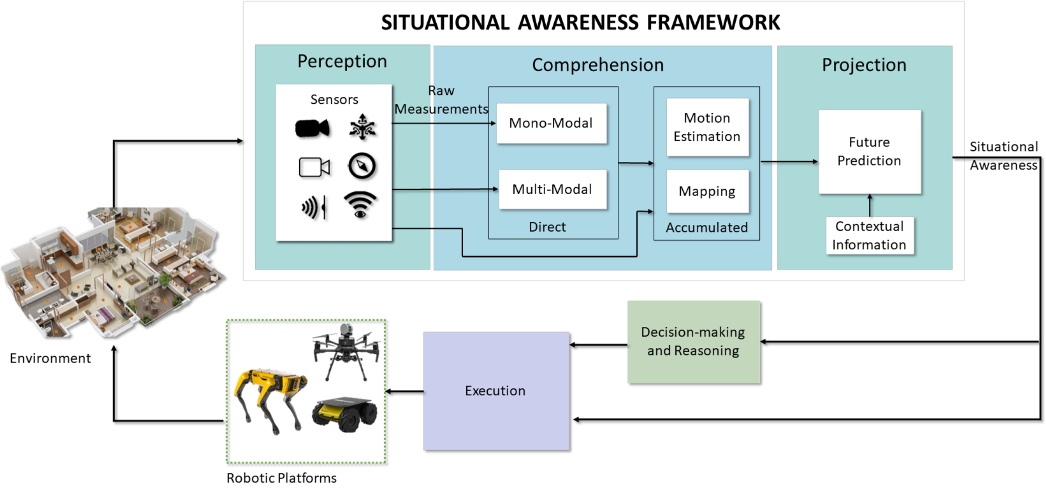


**Figure 2.** The proposed Situational Awareness system architecture for autonomous mobile robots. We break it into its principal components, namely perception, comprehension, and projection, and show how they are connected.

To answer the mentioned question, we characterize SA by its three main parts, for which we propose a description that delimits their scope and defines their purpose:

The **perception of the situation** consists of the acquisition of exteroceptive information, i.e., of the surroundings, such as visual light intensity or distance, and proprioceptive information, i.e., of the internal values of the robot, such as velocity or temperature, and information on the situation. Sensors provide raw measurements that must be transformed to acquire actual knowledge or may directly inform the robot about its state with little processing. For instance, active range sensors provide a distance to objects by well-defined models. On the contrary, the pixel intensity values of a camera, other than being distorted by unknown parameters, depend on complex algorithms to extract meaningful depth, which is still ongoing research. Considering this, multiple sensor modalities are essential to perceive complementary details of the situation, e.g., the acceleration of the robot and the metric scale, the visible light intensity and its changes, and compensate for low performance in different conditions, e.g., low light, light-transparent materials, or fast motion. Hence, perception includes the array of sensors that give each robot specific attributes and those algorithms that augment the amount of information at the disposal of successive layers. Furthermore, as cameras are the predominant source of mostly latent environmental features, image processing algorithms are necessary to gain insights into the situation from these sensors. We include such basic algorithms, which usually require a single image frame, into the next layer of direct comprehension.

The **comprehension of the situation** extends from understanding the current perception, considering the possible semantic relationships, to build a short-term understanding using perceptual observation at a given time instant, referred to as *direct situational comprehension*, or a long-term model that includes the knowledge acquired in the past, namely, the *accumulated situational comprehension*. Multiple abstract relationships can be created to link concepts in a structured model of the situation, such as geometric (e.g., the shape of the objects), semantic (e.g., the type and functionality of the objects), topological (e.g., the order in the space), ontological (e.g., the hierarchy of commonsense concepts), dynamic (e.g., the motion between the objects), or stochastic (e.g., to include uncertainty information). In addition, the comprehension of the situation is affected by mechanisms such as attention that are controlled by the decision-making and control processes (e.g., looking for a particular object in a room vs. getting a global overview of a room).

The **projection of the situation** into the future is essential for decision-making processes, and a higher level of comprehension facilitates this ability. A more profound understanding of the environmental context, which includes information such as the robot's position, velocity, pose, and any static or dynamic obstacles in the surrounding area, can lead to a more accurate projection model. The projection process involves forecasting the future states of both the ego-agent and external agents to predict behaviors and interactions, enabling the robot to adapt its actions to achieve its goals effectively.

The rest of this paper aims to delve into those research questions naturally developed as a consequence of the exposed SA topic:

- *What has been achieved so far, and what challenges remain?*
- *What could the future direction of Situational Awareness be?*

Thus, by reviewing the current state-of-the-art methods for mobile robots that may fall into perception, comprehension, and projection, we aim to study the broad field of Situational Awareness as one and understand the advancement and limitations of its components. Then, we discuss in which direction we envision the research will address the remaining challenges and bridge the gap that divides robots from mature, intelligent autonomous systems.

We summarize the main contributions of this paper as:

- Comprehensive review of the state-of-the-art approaches: we conduct a thorough analysis of the latest research related to enhancing situational awareness for mobile robotic platforms, covering computer vision, deep learning, and SLAM techniques.
- Identification and analysis of the challenges: we classify and discuss the reviewed approaches according to the proposed definition of situational awareness for mobile robots and highlight their current limitations for achieving complete autonomy in mobile robotics.
- Proposals for future research directions: we provide valuable insights and suggestions for future research directions and open problems that need to be addressed to develop efficient and effective situational awareness systems for mobile robotic platforms.

## 2. Situational Perception

Situational perception enables robots to perceive their known state as well as the situation around them using a single or a combination of onboard proprioceptive and exteroceptive sensors. The continuous technological advances regarding chip developments have made many sensors suited for use onboard mobile robots [11], as they come with a small form factor and possibly low power consumption. The primary sensor suite of the average robot can count on a wide array of devices, such as Inertial Measurement Units (IMUs), magnetometers, barometers, and Global Navigation Satellite Systems (GNSS) receivers, e.g., for the typical Global Positioning System (GPS) satellite constellation. Sensors such as IMUs, which can be utilized by robotic platforms to measure their attitude, angular velocities, and linear accelerations, are cheap and lightweight, making them ideal for running onboard the platforms, though the performance of these sensors can degrade over time due to the accumulation of errors coming from white Gaussian noise [12]. Magnetometers are generally integrated within an IMU sensor, measuring the accurate heading of the robotics platform relative to the Earth's magnetic field. The sensor measurements from a magnetometer can be corrupted when the robot navigates in environments with constant magnetic fields interfering with the Earth's magnetic field. Barometers can be utilized by flying mobile robots to measure their altitude changes through measured pressure changes, but they suffer from bias and random noise in measurements in indoor environments due to ground/ceiling effects [13]. Wheel encoders are utilized by ground mobile robots to measure the velocities of the platform and obtain its relative position. GNSS receivers, as well as their higher-precision variants, such as Real-Time Kinematic (RTK) or differential GNSS, provide reliable position measurements of robots in a global frame of reference relative to the Earth in outdoor environments. However, these sensors can work reliably in uncluttered outdoor environments with multiple satellites connected or within a direct line of sight with the RTK base station [14].

The adoption of cameras as exteroceptive sensors in robotics has become increasingly prevalent due to their ability to provide a vast range of information in a compact and cost-effective manner [15]. In particular, RGB cameras, including monocular cameras, have been widely used in robotics as primary sensor as it provides the robots with colored images which can be further processed to extract meaningful information from their environment. Additionally, cameras with depth information, such as stereo or RGB-D cameras, have emerged as a dominant sensor type in robotics given that they provide the robot with additional capabilities of perceiving the depth of the different objects within the environment. As such, the use of standard cameras is expected to continue playing a crucial role in developing advanced robotic systems. These standard cameras suffer from the disadvantage of motion blur in the presence of a rapid motion of the robot, and the perceived quality can degrade as the robot navigates in changing lighting conditions.

In robotics, RGB and RGB-D (i.e., with depth) cameras are thus complemented by infrared (IR) cameras, also referred to as thermal cameras when detecting long-wave radiation, for gaining extended visibility during nighttime or adverse weather conditions. These sensors can provide valuable information not detectable by human eyes or traditional cameras, such as heat signatures and thermal patterns. By incorporating thermal and IR cameras into the sensor suite, mobile robots can detect and track animated targets by following heat signatures, and navigating low-visibility environments, thus operating in a broader range of conditions. These specialized sensors can significantly enhance a robotic system's situational awareness and overall performance.

Neuromorphic vision sensors [16], also known as event cameras [17], such as the Dynamic Vision Sensor (DVS) [18], overcome the limitations of standard cameras by encoding pixel intensity changes rather than an absolute brightness value and providing very high dynamic ranges as well as no motion blur during rapid motions. However, due to the asynchronous nature of event cameras, measurements of the situations are only provided in case of variations in the perceived scene brightness that are often caused by the motion of the sensor itself. Hence, they can measure a sparse set of points, usually in correspondence with edges. To perceive a complete dense representation of the environment, such sensors onboard mobile robots are typically combined with the traditional pixels of RGB cameras, as in the case of the Dynamic and Active-Pixel Vision Sensor (DAVIS) [19] or Asynchronous Time-Based Image Sensor (ATIS) [20] cameras. Nevertheless, algorithms have also been proposed to reconstruct traditional intensity frames by integrating events over time to facilitate the reuse of preexisting image processing approaches [21] or even produce a high-frame-rate video by interpolating new frames [22].

Ranging sensors, such as small-factor solid-state Light Detection and Rangings (LIDARs) or ultrasound sensors, are the second most dominant group of employed exteroceptive sensors onboard mobile robots. One-dimensional LIDARs and ultrasound sensors are used mainly in flying mobile robots to measure their flight altitude but only measure limited information about their environments, while ground mobile robots can utilize the sensor to measure the distance to nearest object. Two- and three-dimensional LIDARs accurately perceive the surroundings in 360°, and the newer technological advancements have reduced their size and weight. However, utilizing these sensors onboard small-sized robotic platforms is still not feasible, and the high acquisition cost hampers the adoption of this sensor by the broad commercial market. Even for autonomous cars, a pure-vision system, which may include event cameras, is often more desirable from an economic perspective.

Frequency-modulated continuous wave (FMCW) radio detection and ranging (RADAR) systems transmit a continuous waveform with a changing frequency over time. This changing frequency creates a frequency sweep, or chirp, continuously transmitted and reflected off objects in the radar's field of view. An FMCW radar can determine the detected objects' range, velocity, and angle by measuring the frequency shift between the transmitted and received signals. Millimeter-wave (mmWave) radars use short-wavelength electromagnetic waves in the GHz range to obtain millimeter accuracy and are a valid alternative to LIDAR for range measurements in robotics [23]. Even though they have a lower angular resolution and more limited range than LIDAR, they offer a smaller form factor and a lower cost. Additionally, they can estimate the speed of objects by leveraging the Doppler effect. Nevertheless, mmWave radars can detect transparent surfaces that are challenging to see with LIDAR. As such, they have become an attractive option for robotic applications where cost, form factor, and the detection of transparent surfaces are crucial.

A Radiofrequency (RF) signal is another technology based on signal processing that allows a robot to infer its global position by estimating its distance with one or multiple base stations. Differently from GPS, RF signals may be able to provide positioning information in indoor environments as well, even though range measurements require it to be fused with other sources of motion estimation, e.g., from an IMU. Contrary to mmWave radars, RF-based localization or mapping is far less precise, but newer technology such as 5G promises superior performance. However, a drawback of these approaches is that they require synchronization between the antenna and the receiver for computing a time of arrival (TOA) and possibly line-of-sight (LOS) communication, especially when only one antenna is available. Other RF measurements, such as the time difference of arrival (TDOA), allow the release of the synchronization requirement or the computation of the position in a non-line-of-sight (NLOS) scenario by extracting information from matrices of channel state information (CSI). Kabiri et al. [24] provide an exhaustive review of RF-based localization methods and give an outlook on current challenges and future research directions.

Table 1 summarizes the different sensors used onboard mobile robots with their individual limitations. Given the multifaceted characteristics of the available sensors, relying on a monomodal perception does not guarantee a safe robot deployment in real-world settings. Consequently, multimodal perception is often preferred at the cost of more complex solutions to properly fuse and time-synchronize the measurements from multiple sensors. Nevertheless, it is essential to perceive complementary information of the situation and to build a complete state of the autonomous agent, e.g., the acceleration of a robot, the visual light intensity of the environment, its global position, or the distance with obstacles, and compensate for low performance in different conditions, e.g., dark rooms or low-light, transparent materials, or non-Lambertian surfaces.

**Table 1.** Different types of sensors utilized onboard mobile robots for situational perception.

| Classification | Sensor | Measurement | Mobile Robotic Platforms | Limitations | Examples |
|---|---|---|---|---|---|
| Proprioceptive | IMU | • Velocities<br>• Accelerations<br>• Yaw angle (with a magnetometer) | Indoor/outdoor robots | • Bias<br>• Gaussian noise<br>• Drifts rapidly | MPU-6050 |
| | GPS | • Absolute position | Outdoor robots | • Unreliable measurements in cluttered environments. | u-blox NEO-M8N |
| | Barometer | • Altitude from atmospheric pressure | Indoor/outdoor aerial robots | • Bias<br>• Gaussian noise | Bosch BMP280 |
| | Robot encoders | • Relative position<br>• Velocity | Indoor/outdoor ground robots | • Slippage<br>• Error accumulation | US Digital E4P |
| | RF Receiver | • Absolute position | Indoor/outdoor robots | • Prone to interference<br>• Limited range | DecaWave DWM1000 |
| Exteroceptive | RGB camera | • Visible light | Indoor/outdoor robots | • Motion blur<br>• Degradation in poor light conditions | IDS uEye LE |
| | RGB-D Camera | • Visible light<br>• Depth from IR structured light | Indoor/outdoor robots | • Limited and noisy range<br>• Errors in reflective/transparent surfaces | Intel Realsense D435 |
| | IR Camera | • Infrared radiation | Indoor/outdoor robots | • Limited information<br>• Prone to atmospheric interference<br>• Infrared cannot pass through glass or water | FLIR Lepton |
| | Event camera | • Brightness log-intensity changes | Indoor/outdoor robots | • Requires motion of camera or objects<br>• Absolute brightness not measured directly<br>• Not easy to purchase | DAVIS 346 or SONY IMX636ES |
| | LIDAR | • Metric distances and angle of scene points | Indoor/outdoor robots | • Prone to atmospheric interference<br>• Degradation in reflective and transparent surfaces | Velodyne VLP-16 |
| | MmWave FMCW RADAR | • Metric distances and angle of scene points<br>• Objects' speed | Indoor/Outdoor Robots | • Limited range and field of view<br>• Possible low angular and distance resolution<br>• Multipath propagation effect and ghost targets | AWR6843AOP |

The traditional approach to designing robots involves tailoring their sensor selection and configuration to the specific task they are intended to perform. However, this approach may not be sufficient for a humanlike perception capable of adapting to diverse external situations. To overcome this limitation, it is necessary to equip robots with a standard, versatile sensor suite that can provide detailed and accurate information about their surroundings and dynamic elements. This sensor suite, coupled with advanced processing algorithms (explained in Sections 3 and 4), can enable robots to perceive their environment like humans, irrespective of the application they are designed for.

## 3. Direct Situational Comprehension

Some research works focus on transforming the complex raw measurements provided by sensors into more tractable information with different levels of abstraction, i.e., feature extraction for an accurate scene understanding, without building a complex long-term model of the situation. Table 2 provides a summary of the presented direct comprehension methods with their key limitations while using onboard mobile robots. Direct situational comprehension based on sensor modalities can be divided into two main categories, as described below.

### *3.1. Monomodal*

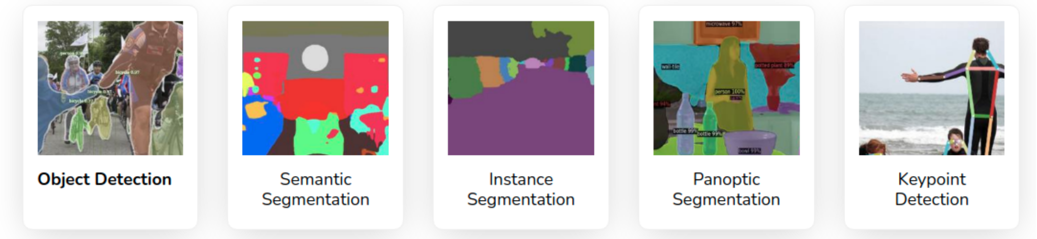
These algorithms utilize a single sensor source to extract useful environmental information. The two primary sensor modalities used in robotics are *vision-based sensors* and *range-based sensors* for the rich and plentiful amount of information in their scene observations.

Vision-based comprehension started with the early works of Viola and Jones [25] presenting an object-based detector for face detection using *Haar-like features* and *Adaboost feature classification*. Following works for visual detection and classification tasks such as [26–30] utilized well-known image features, e.g., Scale-Invariant Feature Transform (SIFT) [31], Speeded-Up Robust Features (SURF) [32], Histogram of Oriented Gradients (HOG) [33], along with Support Vector Machine (SVM)-based classifiers [34]. The mentioned methods focused on extracting only a handful of helpful information from the environment, such as pedestrians, cars, and bicycles, showing degraded performance in difficult lighting conditions and occlusions.

**Table 2.** Summary of the algorithms for the direct situational comprehension SA module. DL refers to methods leveraging deep learning.

| Modality | Sensor | Method | DL | Limitations | References |
|---|---|---|---|---|---|
| Monomodal | RGB | Feature detection | ✗ | • Sensitive to illumination changes<br>• Higher false positives<br>• Lower robustness in the presence of occlusions | [25–34] |
| | | Object detection | ✓ | • Higher computation cost<br>• Larger training dataset<br>• Sensitive to occlusions<br>• No instance segmentation | [35] |
| | | Semantic segmentation | ✓ | • Higher computation cost<br>• Larger training dataset<br>• No instance segmentation | [36–44] |
| | | Panoptic segmentation | ✓ | • Higher computation cost<br>• Larger training dataset<br>• Slower inference time | [45,46] |
| | | 2D Scene graphs | ✓ | • Limited to 2D spatial information<br>• Limited temporal information | [47–49] |
| | Thermal | Object detection | ✗ | • Limited applicability<br>• Higher false positives | [50] |
| | | Object detection | ✓ | • Limited applicability<br>• Limited datasets | [51,52] |
| | Event | Object detection | ✓ | • Trained over limited data<br>• Limited validation in the presence of occlusions | [53,54] |
| | | Semantic segmentation | ✓ | • Trained over limited data<br>• Limited to only a few semantic objects | [54] |
| | LIDAR | Object detection | ✗ | • Detection of fewer semantic entities<br>• Lower robustness in the presence of outliers and occlusions | [55–57] |
| | | Semantic segmentation | ✓ | • Limited to fewer semantic entities<br>• Higher computational cost<br>• Lower accuracy in indoor environments | [58–66] |
| Multimodal | RGB + Depth | Object detection | ✗ | • Higher false positives and negatives<br>• Limited range | [67,68] |
| | | Object detection | ✓ | • Limited range and limited to low-range applications<br>• High computation cost<br>• Lower inference time<br>• Lack of generalizability over nontrained semantic entities | [69–73] |
| | RGB + Thermal | Semantic Segmentation | ✓ | • Limited real world testing<br>• Mostly limited to outdoor environments<br>• Lower object detection accuracy | [74–77] |
| | RGB + Event | Semantic segmentation | ✓ | • Tested only on outdoor datasets | [78] |
| | RGB + LIDAR | Object detection | ✓ | • Limited accuracy in indoor environments<br>• No temporal history of detected objects for efficient tracking | [79–83] |

With the establishment of DL in computer vision and image processing for robot vision, recent algorithms in the literature robustly extract the scene information utilizing Convolutional Neural Networks (CNNs) in the presence of different lighting conditions and occlusions. In computer vision, different types of DL-based methods exist based on the kind of scene-extracted information. Algorithms such as *Mask-RCNN* [36], *RetinaNet* [37], *TensorMask* [38], *TridentNet* [39], and *Yolo* [35] perform detection and classification of several object instances, and they either provide a bounding box around the object or perform a pixelwise segmentation for each object instances. Other algorithms such as [40–44] perform a dense semantic segmentation, being able to extract all relevant information from the scene. Additionally, Kirillov et al. [45], Cheng et al. [46] aimed to detect and categorize all object instances in an image, regardless of their size or position, through a panoptic segmentation while still maintaining a semantic understanding of the scene. This task is particularly challenging because it requires integrating pixel-level semantics and instance-level information. Figure 3 showcases different image segmentation algorithms. Two-dimensional *scene graphs* [47,48] could then connect the semantic elements detected by the panoptic segmentation in a knowledge graph that let one reason about relationships. Moreover, this knowledge graph facilitates the inference of single behaviors and interactions among the participants in a scene, animate or inanimate [49].



**Object Detection**

Semantic Segmentation

Instance Segmentation

Panoptic Segmentation

Keypoint Detection

**Figure 3.** Deep-learning-based computer vision algorithms for monomodal scene understanding. Copyright to [84].

To overcome the limitations of the visible spectrum in the absence of light, thermal infrared sensors have been researched to augment situational comprehension. For instance, one of the earlier methods [50] found humans in nighttime images by extracting thermal shape descriptors that were then processed by *Adaboost* to identify positive detection. In contrast, newer methods [51,52] utilize deep CNNs on thermal images for identifying different objects in the scene, such as humans, bikes, and cars. Though research in the field of event-based cameras for scene understanding is not yet broad, some works such as [85] present an approach for dynamic object detection and tracking using event streams, whereas [53] present an asynchronous CNN for detecting and classifying objects in real time. *Ev-SegNet* [54] is an approach that introduced one of the first semantic segmentation pipelines based on event-only information.

Range-based comprehension methods with earlier works such as [55] and ref. [56] presented object detection for range images from 3D LIDAR using an SVM for object classification. However, the authors in [57] utilized range information to identify the terrain around the robot and objects and used an SVMs to classify each category. Nowadays, deep learning is also playing a fundamental role in scene understanding using range information. Some techniques utilize CNNs for analyzing range measurements translated into camera frames by projecting the 3D points onto an abstract image plane. For example, *Rangenet++* [58,59], *SqueezeSeg* [60], and *SqueezeSegv2* [61] project the 3D point-cloud information onto 2D range-based images for performing the scene understanding tasks. The above-mentioned methods argue that CNN-based algorithms can be directly applied to range images without using expensive 3D convolution operators for point cloud data. Others apply CNNs directly on the point cloud information for maximizing the preservation of spatial information. Approaches such as *PointNet* [62] (Figure 4a), *PointNet++* [63], *TangentConvolutions* [64], *DOPS* [65], and *RandLA-Net* [66] perform convolutions directly over the 3D point cloud data to semantically label the point cloud measurements.

### *3.2. Multimodal*

The fusion of multiple sensors for situational comprehension allows algorithms to increase their accuracy by observing and characterizing the same environment quantity but with different sensor modalities [86]. Algorithms combining RGB and depth information have been widely researched due to the easy availability of the sensors publishing RBG-D information. González et al. [67] studied and presented the improvement of the fusion of multiple sensor modalities (RGB and depth images), numerous image cues, and various image viewpoints for object detection, whereas Lin et al. [68] combined 2D segmentation and 3D geometry understanding methods to provide contextual information for classifying the categories of the objects and identifying the scene in which they are placed. Several algorithms classifying and estimating the pose of objects using CNNs, such as [69], *PoseCNN* [70], *DenseFusion* [71–73], rely extensively on RBG-D information. These methods are primarily employed for object manipulation tasks, using robotic manipulators fixed on static platforms or mobile robots.

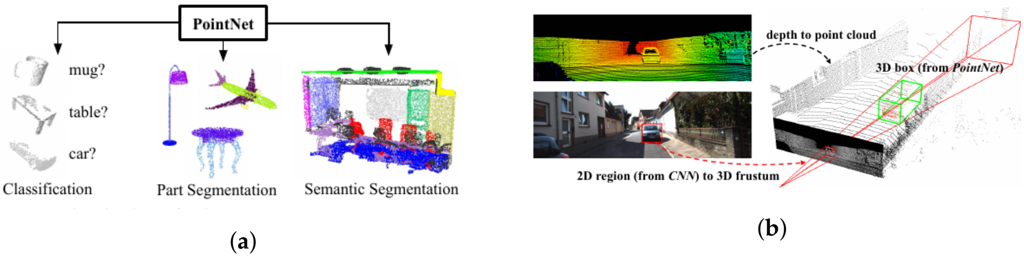


**Figure 4.** Monomodal and multimodal scene understanding algorithms, (**a**) *PointNet* algorithm using only LIDAR measurements. Copyright to [62]. (**b**) *Frustrum PointNets* algorithm combining RGB and LIDAR measurements improving the accuracy of *PointNet*. Copyright to [79].

Alldieck et al. [87] fused RGB and thermal images from a video stream using contextual information to access the quality of each image stream to combine the information from the two sensors accurately, whereas methods such as *MFNet* [74], *RTFNet* [75], *PST900* [76], and *FuseSeg* [77] combined the potential of RGB images along with thermal images using CNN architectures for the semantic segmentation of outdoor scenes, providing accurate segmentation results even in the presence of degraded lighting conditions. Zhou et al. [88] proposed *ECFFNet* to perform the fusion of RGB and thermal images at the feature level, which provided complementary information, effectively improving object detection in different lighting conditions. Spremolla et al. [89], Mogelmose et al. [90] performed a fusion of RGB, depth, and thermal camera computing descriptors in all three image spaces and fused them in a weighted average manner for efficient human detection.

Dubeau et al. [91] fused the information from an RGB and depth sensor with an event-based camera cascading the output of a deep Neural Network (NN) based on event frames with the output from a deep NN for RBG-D frames for a robust pose tracking of high-speed moving objects. *ISSAFE* [78] is another approach that combines event-based CNN with an RGB-based CNN using an attention mechanism to perform semantic segmentation of a scene, utilizing the event-based information to stabilize the semantic segmentation in the presence of high-speed object motions.

To improve situational comprehension using 3D point cloud data, methods have been presented that combine information extracted over RGB images with their 3D point cloud data to accurately identify and localize the objects in the scene. *Frustrum PointNets* [79] (Figure 4b) performed 2D detection over RGB images which were projected to a 3D viewing frustum from which the corresponding 3D points were obtained, to which a *PointNet* [62] (Figure 4a) was applied for object instance segmentation, and an *amodal* bounding box regression was performed. Methods such as *AVOD* [80,81] extract features from both RGB and 3D point clouds projected to a bird's eye view and fuse them to provide 3D bounding boxes for several object categories. *MV3D* [82] extract features from RGB images and 3D point cloud data from the front view as well as a bird's eye view to fuse them in a Region of Interest (RoI) pooling, predicting the bounding boxes as well as the object class. *PointFusion* [83] employs an RGB and 3D point cloud fusion architecture which is unseen and object-specific and can work with multiple sensors providing depth.

*Direct situational comprehension* algorithms only provide the representation of the environment at a given time instant and mostly discard the previous information, not creating a long-term map of the environment. In this regard, the extracted knowledge can thus be transferred to the subsequent layer of *accumulated situational comprehension*.

## 4. Accumulated Situational Comprehension

A greater challenge consists of building a long-term multiabstraction model of the situation, including past information. Even small errors not considered at a particular time instant can cause a high divergence between the state of the robot and the map estimate over time. To simplify the explanation, we divided this section into three subsections, namely, *motion estimation*, *motion estimation and mapping*, and *mapping*.

### *4.1. Motion Estimation*

The motion estimation component is responsible for estimating the state of the robot directly, using the sensor measurements from single/multiple sources and the inference provided by the *direct situational comprehension* component (see Section 3.1). While some motion estimation algorithms only use real-time sensor information to estimate the robot's state, others estimate the robot's state inside a pregenerated environment map. Early methods estimated the state of the robot based on filtering-based sensor fusion techniques such as an *Extended Kalman Filter (EKF)*, an *Unscented Kalman Filter (UKF)*, and *Monte Carlo Localization (MCL)*. Methods such as those in [92,93] use *MCL*, providing a probabilistic hypothesis of the state of the robot directly, using the range measurements from a range sensor. Anjum et al. [94] performed a *UKF* based fusion of several sensor measurements such as gyroscopes, accelerometers, and wheel encoders to estimate the motion of the robot. Kong et al. [95], Teslic et al. [96] performed an EKF based fusion of odometry from robot wheel encoders and measurements from a prebuilt map of line segments to estimate the robot state, whereas Chen et al. [97] used a prebuilt map of corner features. Ganganath and Leung [98] presented both UKF and MCL approaches for estimating the pose of the robot using wheel odometry measurements and a sparse prebuilt map of visual markers detected with an RGB-D camera. In contrast, Kim and Kim [99] presented a similar approach using ultrasound distance measurements with respect to an ultrasonic transmitter.

The simplified mathematical models are subject to several assumptions that limited earlier motion estimation methods. Newer methods try to improve these limitations by providing mathematical improvements over the earlier methods and accounting for delayed measurements between different sensors, such as the UKF developed by Lynen et al. [100] and the EKF by Sanchez-Lopez et al. [101], which compensate for time-delayed measurements in an iterative nature for a quick convergence to the actual state. Moore and Stouch [102] presented an EKF/UKF algorithm, well-known in the robotics community, which can take an arbitrary number of heterogeneous sensor measurements for the estimation of the robot state. Wan et al. [103] used an improved version of a Kalman filter called the *error-state Kalman filter*, which used measurements from RTK GPS, LIDAR, and IMU for a robust state estimation. Liu et al. [104] presented a *multi-innovation UKF (MI-UKF)*, which utilized a history of innovations in the update stage to improve the accuracy of the state estimate; it fused IMU, encoder, and GPS data and estimated the slip error components of the robot.

The motion estimation of robots using a Moving-Horizon Estimation (MHE) has also been studied in the literature where methods such as in [105] fuse wheel odometry and LIDAR measurements using an MHE scheme to estimate the state of the robot, claiming a robustness over any outliers in the LIDAR measurements. Liu et al. [106], Dubois et al. [107] studied a *multirate MHE* sensor fusion algorithm to account for sensor measurements obtained at different sampling rates. Osman et al. [108] presented a generic MHE-based sensor fusion framework for multiple sensors with different sampling rates, compensating for missed measurement, outlier rejection, and satisfying real-time requirements.

Recently, motion estimation algorithms of mobile robots using *factor-graph*-based approaches have also been extensively studied as they have the potential to provide a higher accuracy. Factor graphs can encode either the entire history of the robot state or go back up to a fixed time, i.e., fixed-lag smoothing methods, capable of handling different sensor measurements in terms of nonlinearity and varying frequencies optimally and intuitively (see Figure 5). Ranganathan et al. [109] presented one of the first graph-based approaches using *square-root fixed-lag smoothing* [110], for fusing information from odometry, visual, and GPS sensors, whereas Indelman et al. [111] presented an improved fusion based on an incremental smoothing approach, *iSAM2* [112], fusing IMU, GPS and visual measurements from a stereocamera setup. The methods presented in [113,114] utilized sliding-window factor graphs for estimating the robot's state by fusing several wheel odometry sources along with global pose sources. Mascaro et al. [115] also presented a sliding-window factor graph combining visual odometry information, IMU, and GPS information to estimate the drift between the local odometry frame with respect to the global frame, instead of directly estimating the robot state. Qin et al. [116] presented a generic factor graph-based framework for fusing several sensors. Each sensor served as a factor connected with the robot's state, quickly adding them to the optimization problem. Li et al. [117] proposed a novel graph-based framework for sensor fusion that combined data from a stereo visual–inertial navigation system, i.e., S-VINS, and multiple GNSS sources in a semitightly coupled manner. The S-VINS output was an initial input to the position estimate derived from the GNSS system in challenging environments where GNSS data are limited. By integrating these two data sources, the framework improved the robot's global pose estimation accuracy.

The *motion estimation* algorithms, as illustrated in Figure 5, do not simultaneously create a map of the environment, limiting their environmental knowledge, which has led to research on simultaneous *motion estimation and mapping* algorithms described in the following subsection.

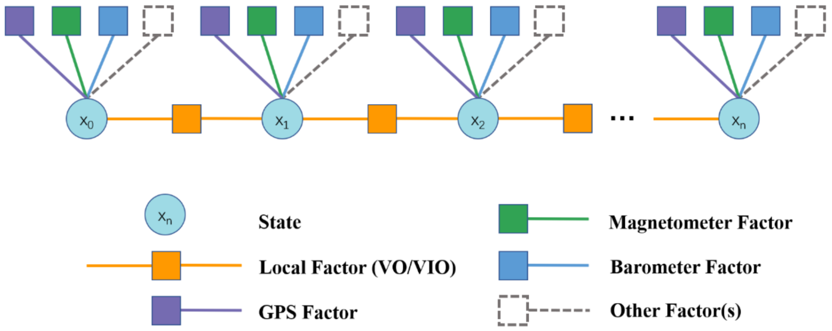


**Figure 5.** Localization factor graph used for estimating the robot state fusing multiple sensor measurements. Copyright to [116].

## 4.2. Motion Estimation and Mapping

This section covers the approaches which estimate not only the robot motion given the sensor measurements but also the map of the environment, i.e., they model the scene in which the robot navigates. These approaches are commonly known as SLAM, which is one the widely researched topics in the robotics industry [118], as it enables a robot to use scene modeling without the requirement of prior maps and in applications where initial maps cannot be obtained easily. Vision and LIDAR sensors are the two primary exteroceptive sensors used in SLAM for map modeling [15,119]. As in the case of *motion estimation* methods, SLAM can be performed using a single-sensor modality or using information from different sensor modalities and combining it with scene information extracted from the *direct situational comprehension* module (see Section 3). SLAM algorithms have a subset of algorithms that do not maintain the entire map of the environment and do not perform stages of *loop closure* called *odometry estimation algorithms*, where Visual Odometry (VO) becomes a subset of visual SLAM (VSLAM) and LIDAR odometry a subset of LIDAR SLAM. Tables 3 and 4 provide a brief summary of the above-presented approaches highlighting the different datasets used in their validation along with their limitations.

### 4.2.1. Filtering

Earlier SLAM approaches such as in [120–122] applied an EKF to estimate the robot pose by simultaneously adding/updating the landmarks observed by the robots. However, these methods were quickly discarded as their computational complexity increased with the number of landmarks, and they did not efficiently handle nonlinear measurements [123]. Accordingly, *FastSLAM 1.0* and *FastSLAM 2.0* [124] were proposed as improvements to

EKF-SLAM, which combined particle filters to calculate the trajectory of the robot with individual EKFs for landmark estimation. These techniques also suffered from the limitations of sample degeneracy when sampling the proposal distribution and problems with particle depletion.

#### 4.2.2. Metric Factor Graphs

Modern SLAM, as described in [118], has moved to a more robust and intuitive representation of the state of the robot along with sensor measurements, as well as the environment map to create factor graphs as presented in [110,112,125–127]. Factor-graph-based SLAM, based on the type of map used for the environmental representation and optimization, can be divided into *metric* and *metric–semantic* factor graphs.

A metric map encodes the understanding of the scene at a geometric level (e.g., lines, points, and planes), which is utilized by a SLAM algorithm to model the environment. *Parallel tracking and mapping (PTAM)* was one of the first feature-based monocular algorithms which split the tracking of the camera in one thread and the mapping of the key points in another, performing a batch optimization for optimizing both the camera trajectory and the mapped 3D points. Similar extensions to the *PTAM* framework are *ORB-SLAM* [128] and *REMODE* [129] which create a semidense 3D geometric map of the environment while estimating the camera trajectory. As an alternative to feature-based methods, direct methods use the image intensity values instead of extracting features to track the camera trajectory even in featureless environments such as semidense direct VO, called *DSO* [130] and *LDSO* [131], improving the *DSO* by adding loop closure into the optimization pipeline, whereas *LSD-SLAM* [132], *DPPTAM* [133], and *DSM* [134] perform a direct monocular SLAM tracking camera trajectory along with building a semidense model of the environment. Methods have also been presented that combine the advantages of both feature-based and intensity-based methods, such as *SVO* [135] performing high-speed semi-direct *VO*, *CPA-SLAM* [136], and *loosely coupled semidirect SLAM* [137] utilizing image intensity values for optimizing the local structure and image features to optimize the key-frame poses.

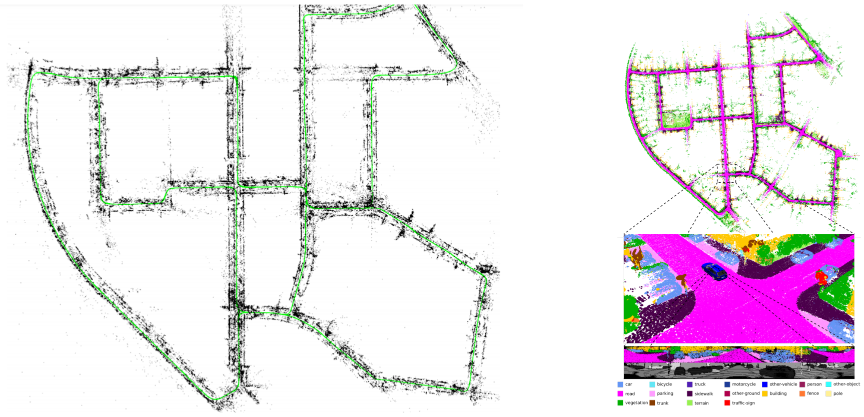
Deep Learning models may be used effectively to learn from data to estimate the motion from sequential observations. Hence, their online prediction could be better before initializing the factor-graph optimization problem closer to the correct solution [138,139]. *MagicVO* [140] and *DeepVO* [141] study supervised end-to-end pipelines to learn monocular *VO* from data not requiring complex formulations and calculations for several stages, such as feature extraction and matching, keeping the VO implementation concise and intuitive. There are also some supervised approaches such as *LIFT-SLAM* [142], *RWT-SLAM* [143], and [144,145] that utilize deep neural networks for improved feature/descriptor extraction. Alternatively, unsupervised approaches [146–148] exploit the brightness constancy assumption between frames in close temporal proximity to derive a self-supervised photometric loss. The methods have gained momentum, enabling the learning from unlabeled videos and continuously adapting the DL models to newly seen data [149,150]. Nevertheless, monocular visual-only methods suffer from the considerable limitation of being unable to estimate the metric scale directly and accurately track the robot poses in the presence of pure rotational or rapid/acrobatic motion. *RAUM-VO* [151] mitigates the rotational drift by integrating an unsupervised learned pose with the motion estimated with a frame-to-frame epipolar method [152].

To overcome these limitations, cameras are combined with other sensors, for example, synchronizing them with an IMU, giving rise to the research line working on monocular *Visual–Inertial Odometry (VIO)*. Methods such as *OKVIS* [153], *SVO-Multi* [154], *VINS-mono* [155], *SVO+GTSAM* [156], *VI-DSO* [157], *BASALT* [158] are among the most outstanding examples. Delmerico and Scaramuzza [159] benchmarked all the open-source *VIO* algorithms and compared their performance on computationally demanding embedded systems. Furthermore, *VINS-fusion* [160] and *ORB-SLAM2* [161] (see Figure 6a) provide a complete framework capable of fusing either *monocular*, *stereo*, or *RGB-D* cameras with an IMU to improve the overall tracking accuracy of the algorithms. *ORB-SLAM3* [162] presents improvement over *ORB-SLAM2* by performing even multimap SLAM using different visual sensors along with an IMU.

Methods have been presented that perform thermal inertial odometry for performing autonomous missions using robots in visually challenging environments [163–166]. The authors in *TI-SLAM* [167] not only performed thermal inertial odometry but also provided a complete SLAM back end with thermal descriptors for loop closure detection. Mueggler et al. [168] presented a continuous-time integration of event cameras with IMU measurements, improving by almost a factor of four the accuracy over event-only *EVO* [169]. Ultimate SLAM [170] combines RGB cameras with event cameras along with IMU information to provide a robust SLAM system in high-speed camera motions.

LIDAR odometry and SLAM for creating metric maps have been widely researched in robotics to create metric maps of the environment such as *Cartographer* [171] and *Hector-SLAM* [172], performing a complete SLAM using 2D LIDAR measurements, and *LOAM* [173] and *FLOAM* [174] providing a *parallel LIDAR odometry and mapping* technique to simultaneously compute the LIDAR velocity while creating accurate 3D maps of the environment. SUMA [175] improves the performance over *LOAM* using dense projective ICP over surfel-based maps. To further improve the accuracy, techniques have been presented which combine vision and LIDAR measurement as in *LIDAR-monocular visual odometry (LIMO)* [176] and *LVI-SLAM* [177], combining monocular image tracking with precise depth estimates from LIDAR measurements for motion estimation. Methods such as *LIRO* [178] and *VIRAL-SLAM* [179] couple additional measurements such as *ultrawide band (UWB)* with visual and IMU sensors for robust pose estimation and map building. Other methods such as *HDL-SLAM* [180] and *LIO-SAM* [181] tightly couple IMU, LIDAR, and GPS measurements for globally consistent maps.

While significant progress has been demonstrated using *metric SLAM* techniques, one of the limitations of these methods is the lack of information extracted from the metric representation, such as (1) a lack of *semantic knowledge of the environment*, (2) an *inefficiency in identifying static and moving objects*, and (3) an *inefficiency in distinguishing different object instances*.



(**a**) (**b**)

**Figure 6.** (**a**) Three-dimensional feature map of the environment created using *ORB-SLAM2*. Copyright to [161]. (**b**) The same environment is represented with a 3D semantic map using *SUMA++*, providing richer information to better understand the environment around the robot. Copyright to [182].

**Table 3.** Summary of the significant VSLAM validated over public datasets.

| Classification | Sensors | Method | Dataset | Limitations |
|---|---|---|---|---|
| Metric factor graphs | • RGB (Mono) | ORB-SLAM [128] | The New College Dataset [183] | • Difficulty in estimating pure rotations<br>• Higher error in low-texture environment<br>• Suffer from motion bias<br>• Scale uncertainty |
| | | DSO [157], LDSO [131] | TUM RGB-D [184], TUM-Mono [185], EuRoC Mav [186], Kitti Odometry [187] | • Require constant illumination<br>• Require photometric calibration for improved results<br>• Scale uncertainty |
| | | LSD-SLAM [132], DPPTAM [133] | TUM RGB-D [184] | |
| | | DSM [134] | EuRoC Mav [186] | |
| | | Semi-Direct VO [137] | EuRoC Mav [186], TUM Mono [185] | • Require accurate initialization<br>• Scale uncertainty |
| | | MagicVO [140], DeepVO [141] | Kitti Odometry [187] | • Higher computational resources<br>• Less accurate than the classical VO counterparts |
| | • RGB (mono)<br>• RGB (stereo)<br>• RGB-D | ORB-SLAM2 [161] | TUM RGB-D [184], EuRoC Mav [186], Kitti Odometry [187] | • Lower accuracy in high-speed motions<br>• Lower accuracy in low-feature, low-lighting environments<br>• Erroneous loop closures in low-feature/similar environments |
| | • RGB (mono)<br>• RGB (stereo)<br>• RGB-D<br>• RGB + IMU | ORB-SLAM3 [162] | TUM VI [188], EuRoC Mav [186] | • Lower accuracy in low-feature, low-lighting environments<br>• Erroneous loop closures in low-feature/similar environments |
| | • RGB + IMU | VINS-Mono [160] | EuRoC Mav [186] | • Requires good initial estimate<br>• Lower accuracy in low-feature, low-lighting environments |
| | | SVO-Multi [154] | EuRoC Mav [186], TUM RGB-D [184], ICL-NUIM [189] | • Requires robust initialization<br>• Flat ground plane assumption causing inaccuracies over nonplanar surfaces |
| | • RGB-D | CPA-SLAM [136] | TUM RGB-D [184], ICL-NUIM [189] | • Lower accuracy in case of inaccurate planar detection |
| Metric–semantic factor graphs | • RGB (Mono) | Monocular Object SLAM [190] | TUM RGB-D [184] | • Pregenerated object database |
| | | QuadricSLAM [191] | TUM RGB-D [184] | • Assumption of scene representation as quadrics<br>• Quadric computation computationally expensive<br>• Scale uncertainty |
| | | CubeSLAM [192] | TUM RGB-D [184], ICL-NUIM [189] | • Lower accuracy in case of higher errors in cuboid detection<br>• Scale uncertainty |
| | • RGB (Mono)<br>• RGB (Stereo)<br>• RGB-D | DynaSLAM [190] | TUM RGB-D [184], Kitti Odometry [187] | • Filter out useful dynamic key points<br>• No topological relationships between the dynamic–static entities |
| | • RGB-D | VDO-SLAM [193] | Kitti Odometry [187], Oxford Multimotion [194] | • No topological relations between the dynamic–static entities in the optimization graph |
| | • RGB (Stereo) + IMU | Kimera [195] | EuRoC Mav [186] | • Computationally expensive planar mesh generation<br>• No topological constraints between the semantic entities |

**Table 4.** Summary of the significant LIDAR-based SLAM validated over public datasets.

| Classification | Sensors | Method | Dataset | Limitations |
|---|---|---|---|---|
| Metric Factor Graphs | • LIDAR (2D) | Cartographer [171] | Deutsches Museum [171] | • Scan matching can present inaccuracies in cluttered/dynamic environments<br>• Loop closure can present inaccuracies in environments with similar structure |
| | • LIDAR (3D) | LOAM [173], FLOAM [174] | Kitti [187] | • Inaccuracies in nonstructured environments (without planar/edge features)<br>• Inaccuracies in the presence of dynamic objects<br>• No explicit appearance-based loop closure |
| | | SUMA [175] | Kitti [187] | • Require 3D LIDAR model<br>• Errors in loop closures in similar environments<br>• Validated mostly on outdoor urban environments |
| | • LIDAR (3D) + RGB (Mono) | LIMO [176] | Kitti [187] | • Inaccuracies in low-texture environments<br>• Degradation of performance during high-speed motions<br>• No loop closure |
| | • LIDAR (3D) + IMU + GPS | HDL-SLAM [180] | Kitti [187] | • Scan matching can present inaccuracies in cluttered/dynamic environments<br>• Optimization graph contains only robot poses and no environmental landmarks<br>• Inaccurate loop closure in similar structured environments |
| Metric–semantic factor graphs | • LIDAR (3D) | LeGO-LOAM [196] | Kitti [187] | • High dependence on ground plane<br>• Inaccuracies in the presence of features extracted from dynamic objects |
| | | SA-LOAM [171] | Kitti [187], Semantic-Kitti [197], Ford Campus [198] | • Limited accuracy in indoor environments<br>• Degradation in loop closure in case of noisy semantic detection |
| | | SUMA++ [182] | Kitti [187], Semantic-Kitti [197] | • Limited to outdoor urban environments<br>• Rely on accurate LIDAR model |

#### 4.2.3. Metric–Semantic Factor Graphs

As explained in Section 3, the advancements in *direct situational comprehension* techniques have enabled a higher-level understanding of the environments around the robot, leading to the evolution of *metric–semantic SLAM* overcoming the limitations of traditional *metric SLAM* and providing the robot with the capabilities of human-level reasoning. Several approaches to address these solutions have been explored, which are discussed in the following.

*Object-based metric–semantic SLAM* builds a map of the instances of the different detected object classes on the given input measurements. The pioneer works *SLAM++* [199] and [190] created a graph using camera pose measurements and the objects detected from previously stored database to jointly optimize the camera and the object poses. Following these methods, many object-based *metric–semantic SLAM* techniques were presented, such as [191,192,200–205] not requiring a previously stored database and jointly optimizing the camera poses, 3D geometric landmarks, as well as the semantic object landmarks. LIDAR-based metric–semantic SLAM techniques such as LeGO-LOAM [196] extract planar/edges features from semantics such as ground planes to improve the performance over metric SLAM *LOAM* [173]. *SA-LOAM* [206] utilizes semantically segmented 3D LIDAR measurements to generate a semantic graph for robust loop closures. The primary sources of inaccuracies of these techniques are due to an extreme dependence on the existence of objects, as well as (1) the *uncertainty in object detection*, (2) the *partial views of the objects which are still not handled efficiently*, and (3) *no consideration of the topological relationship between the objects*. Moreover, most of the previously presented approaches cannot handle dynamic objects. Research works on filtering dynamic objects from the scene, such as *DynaSLAM* [207], or adding dynamic objects to the graph, such as *VDO-SLAM* [193] and *RDMO-SLAM* [208], reduce the influence of the dynamic objects on the robot pose estimate obtained from the optimized graph. Nevertheless, they cannot handle complex dynamic environments and only generate a sparse map without topological relationships between these dynamic elements.

*SLAM with a metric–semantic map* augments the output metric map given by SLAM algorithms with semantic information provided by *scene understanding* algorithms, as in [209–211] or with *SemanticFusion* [212], *Kimera* [195], and *Kimera-Multi* [213]. These methods assume a static environment around the robot; thus, the quality of the *metric–semantic map* of the environment can degrade in the presence of moving objects in the background. Another limitation of these methods is that they do not utilize useful semantic information from the environment to improve the robot's pose estimation and thus the map quality.

*SLAM with semantics to filter dynamic objects* utilizes the available semantic information of the input images provided by the *scene understanding* module only to filter badly conditioned objects (i.e., moving objects) from pictures given to the SLAM algorithms, as in [214–216] for image-based approaches or *SUMA++* [182] (see Figure 6b) for a LIDAR-based approach. Although these methods increase the accuracy of the SLAM system by filtering moving objects, they neglect the rest of the semantic information from the environment to improve the robot's pose estimation.

### 4.3. Mapping

This section covers the recent works which focus only on the complex high-level representations of the environment. Most of these methods assume the SLAM problem to be solved and focus only on the scene representation. An ideal environmental representation must be efficient concerning the required resources, capable of reasonably estimating regions not directly observed, and flexible enough to perform reasonably well in new environments without any significant adaptations. Table 5 summarizes the main mapping methods described above with their generated map types and limitations.

Occupancy mapping is a method for constructing an environment map in robotics. It involves dividing the environment into a grid of cells, each representing a small portion of the space. The occupancy of a cell represents the likelihood of that cell being occupied by an obstacle or not. Initially, all cells in the map can be considered unknown or unoccupied. As the robot moves and senses the environment, the occupancy of cells is updated based on sensor data. One of the most popular approaches in this category is Octomap [217]. It represents the grid of cells through a hierarchical structure that allows a more efficient query of the occupancy probability in a specific location.

The adoption of Signed-Distance Field (SDF)-based approaches in robotics is well-established to represent the robot's surroundings [218] and enable a planning of a safe trajectory towards the mission goal [219]. In general, an SDF is a three-dimensional function that maps points of a metric space to the distance to the nearest surface. SDFs can represent distances in any number of dimensions, define complex geometries and shapes with an arbitrary curvature, and are widely used in computer graphics. However, a severe limitation of *SDFs* is that they can only represent watertight surfaces, i.e., surfaces that divide the space between inside and outside [220].

An SDF has two main variations, Euclidean Signed-Distance Field (ESDF) and Truncated Signed-Distance Field (TSDF), which usually apply to a discretized space made of voxels. On the one hand, an ESDF gives the distance to the closest obstacle for free voxels and the opposite for occupied ones. They have been used for mapping in *FIESTA* [221], where the authors exploited the property of direct modeling free space for collision checking and the gradient information for planning [222], while dramatically improving their efficiency. On the other hand, a TSDF relies on the projective distance, which is the length measured along the sensor ray from the camera to the observed surface. The distances are calculated only within a specific radius around the surface boundary, known as the truncation radius [223]. This helps improve computational efficiency and reduce storage requirements while accurately reconstructing the observed scene. TSDFs have been demonstrated in multiple works such as *Voxgraph* [224], *Freetures* [225], *Voxblox++* [226], or the more recent *Voxblox-Field* [227]. They can create and maintain globally consistent volumetric maps that are lightweight enough to run on computationally constrained platforms and demonstrate that the resulting representation can navigate unknown environments. A panoptic segmentation was rated with TSDFs by Narita et al. [228] for labeling each voxel semantically while differentiating between *stuff*, e.g., the background wall and floor, from *things*, e.g., movable objects. Furthermore, Schmid et al. [229] leveraged pixelwise semantics to maintain temporal consistency and detect changes in the map caused by movable objects, hence surpassing the limitations of a static environment assumption.

*Implicit neural representations (INR)* (sometimes also referred to as coordinate-based representations) are a novel way to parameterize signals of all kinds, even environments parameterized as 3D points clouds, voxels, or meshes. With this in mind, *scene representation networks (SRNs)* [230] have been proposed as a continuous scene representation that encodes both geometry and appearance and can be trained without any 3D supervision. It has been shown that *SRNs* generalize well across scenes, can learn geometry and appearance priors, and are helpful for novel view synthesis, few-shot reconstruction, joint shape, and appearance interpolation in the unsupervised discovery of nonrigid models. In [231], a new approach was presented, capable of modeling signals with fine details and accurately capturing their spatial and temporal derivatives. Based on periodic activation functions, that approach demonstrated that the resulting neural networks referred to as *sinusoidal representation networks (SIRENs)* were well suited for representing complex signals, including 3D scenes.

*Neural radiance fields (NeRF)* [232] exploit the framework of *INR*s to render realistic 3D scenes by a differential process that takes as input a ray direction and predicts the color and density of the scene's structure along that ray. Sucar et al. [233] pioneered the first application of *NeRF* to SLAM for representing the knowledge of the 3D structure inside the weights of a deep NN. For their promising results, that research prospect attracted numerous following works that continuously improved the fidelity of the reconstructions and the possibility of updating the knowledge of the scene while maintaining previously stored information [234–238].

Differently from the previous dense environment representation methods, which are helpful for autonomous navigation, sparser scene representations also exist, such as *point clouds* and *surfel maps*, which are more commonly used for more straightforward tasks such as localization. Remarkably, a *surfel*, i.e., a surface element, is defined by its position in 3D space, the surface normal, and other attributes such as color and texture. Their use has been extensively explored in recent LIDAR-based SLAM to efficiently represent a 3D map that can be performed as a consequence of optimization following revisited places, i.e., loop closure [239,240].

Three-dimensional *scene graphs* have also been researched to represent a scene, such as in [241–243], which build a model of the environment, including not only metric and semantic information but also essential topological relationships between the objects of the environment. They can construct an environmental graph spanning an entire building, including the semantics of objects (class, material, and shape), rooms, and the topological relationships between these entities. However, these methods are executed offline and require an available 3D mesh of the building with the registered RGB images to generate the *3D scene graphs*. Consequently, they can only work in static environments.

Dynamic Scene Graphs (DSGs) (see Figure 7) are an extension of the aforementioned *scene graphs* to include dynamic elements (e.g., humans) of the environment in an actionable representation of the scene that captures geometry and semantics [244]. Rosinol et al. [245] presented the first method to build a DSG automatically using the input of a VIO [195]. Furthermore, it allowed the tracking of the pose of humans and optimized the mesh based on the deformation of the space induced by detected loop closures. Although these results were promising, their main drawback was that the DSG was built offline, and the VIO first created a 3D-mesh-based semantic map fed to the dynamic scene generator. Consequently, the SLAM did not use these topological relationships to improve the accuracy of the spatial reconstruction of the robot trajectory. Moreover, except for humans, the remaining topological relationships were considered purely static, e.g., chairs or other furniture were fixed to the first detection location.

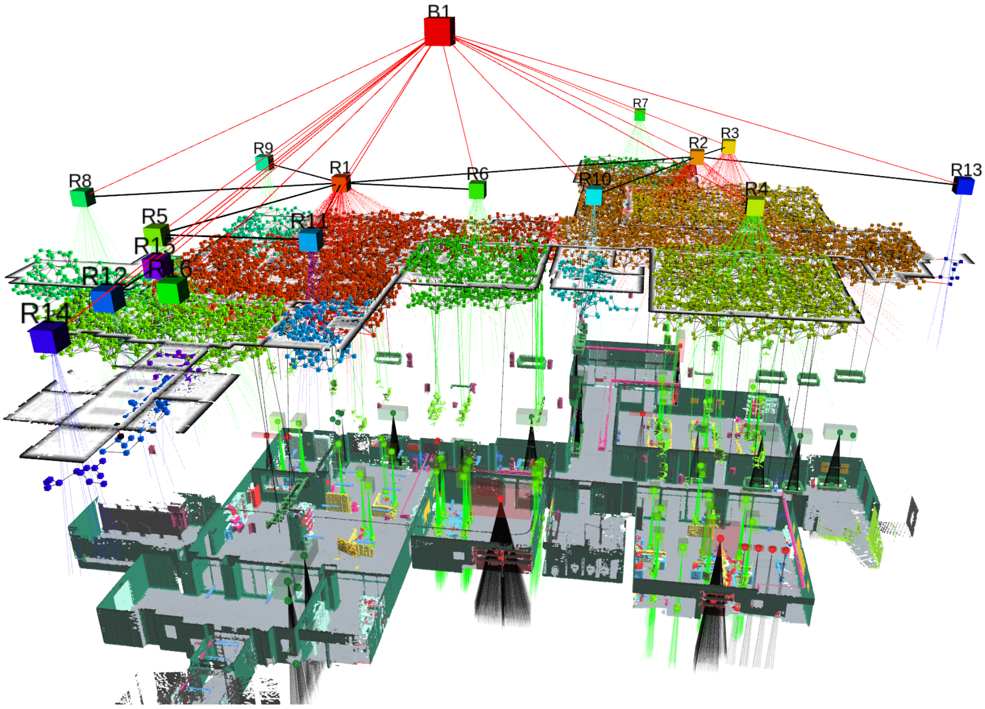


**Figure 7.** A Dynamic Scene Graph generated by [245] with a multilayer abstraction of the environment. Copyright to [245].

**Table 5.** Summary of the significant types of mapping algorithms and their limitations.

| **Mapping Type** | **Sensors** | **Methods** | **Limitations** |
|---|---|---|---|
| Occupancy maps | • RGB-D<br>• 2D LIDAR<br>• 3D LIDAR | Octomap [217] | • Cannot represent the exact shape and orientation of objects<br>• Increased complexity in map querying with increase map size<br>• No semantics |
| ESDF and TSDF | • RGB-D<br>• 3D LIDAR | Voxblox [223] | • No Semantics<br>• ESDF map updates can present errors during loop closures |
| | | Voxgraph [224] | • No semantics<br>• Degradation of the esdf map quality in the presence of noisy odometry estimates |
| | | Voxblox++ [226] | • Degradation of the esdf map quality in the presence of noisy odometry estimates<br>• Computationally expensive semantic detection<br>• Degradation in map quality with noise in the semantic detection |
| NeRF | • RGB-D | iMap [233], Urban Radiance Fields [246], Mega-NeRF [247] | • No semantics (but potentially learnable)<br>• Computationally expensive<br>• Needs to handle catastrophic forgetting while integrating new knowledge |
| Surfel maps | • RGB-D<br>• 3D LIDAR | ElasticFusion [239], SurfelMeshing [248], Other [240] | • Sparse representation<br>• Cannot represent continuous surfaces<br>• Less useful for path planning and obstacle avoidance |
| 3D Scene Graphs | • RGB-D | 3D DSG [241], Hydra [249] | • Validated mainly in indoor scenarios<br>• Handle few dynamic objects in the scene, such as humans |

More recently, Hydra [249] has implemented the scene graph construction into a real-time capable system relying on a highly parallelized architecture. Moreover, it can optimize an embedded deformation graph online, after a loop closure detection. Remarkably, the information in the graph allows the creation of descriptors based on histograms of objects and visited places that can be matched robustly with previously seen locations.

Therefore, DSGs, although in their infancy stage, are shown to be a practical decision-making tool that enables robots to perform autonomous tasks. For example, Ravichandran et al. [250] demonstrated how they could be used for learning a trajectory policy by turning a DSG into a graph observation that served as input to a Graph Neural Network (GNN). A DSG may also be used for planning challenging robotic tasks, as proposed in the Taskography benchmark [251].

Lastly, one of the main features of a DSG is the possibility to perform queries and predictions of the future evolution of the scene based on dynamic models linked with the agents or physical elements [244]. In addition, an even more intriguing property is their application to scene change detection or to the newly formalized *semantic scene variability estimation* task, which sets as a goal the prediction of long-term variation in location, semantic attributes, and topology of the scene objects [252]. This property has only been explored by scratching the surface of its potential application. Still, it already grasps our vision of a comprehension layer that produces the knowledge required by the projection and prediction of future states.

## 5. Situational Projection

In robotics, the projection of the situation is essential for reasoning and the execution of a planned mission [253]. The comprehended information can be projected in the future to predict the future state of the robot by using a dynamic model [101] as well as the dynamic entities in the environment. In order to predict the future state of a robot, the projection component requires more effort in producing models that can forecast the dynamic agents' behavior and how the scene is affected by changes that shift its appearance over time. Remarkably, numerous research areas address specific forecast models, the interactions between agents, and the surrounding environment's evolution. Below, we give an overview of the most prominent ones that we deem more related to the robotic SA concept.

### *5.1. Behavior Intention Prediction*

Behavior intention prediction (BIP) focuses on developing methods and techniques to enable autonomous agents, such as robots, to predict the intentions and future behaviors of humans and other agents they interact with. This research is essential for effective communication, collaboration, and decision-making. BIP typically involves integrating information from multiple sources, such as visual cues, speech, and contextual information, e.g., coming from the comprehension layer. This research has numerous applications, including human–robot collaboration, autonomous driving, and healthcare. Especially regarding the Autonomous Vehicle (AV) application, this topic has gained importance among researchers and is widely studied due to safety concerns. However, we argue that the outcome of its investigation may apply to other tasks implying an interaction among a multitude of agents.

To define the AV BIP task, we refer to the recent survey [254] that distinguishes various research topics related to understanding the driving scene under a unified taxonomy. The whole problem is then defined by analyzing on a timeline the events happening on the road scenario and the decision-making factors that lead to specific outcomes.

Scene contextual factors, such as traffic rules, uncertainties, and interpretation of goals, are crucial for inferring the interaction among road actors [255] and the safety of the current driving policy [256]. Specifically, interaction may be due to social behavior or physical events such as obstacles or dynamic clues, e.g., traffic lights, that influence the decision of the driver [257]. Multimodal perception is exploited to infer whether pedestrians are about to cross [258] or vehicles to change lanes [259]. Mostly, recent solutions rely on DL models such as CNNs [260,261], Recurrent Neural Networks (RNNs) [257,262], GNNs [263], or on the transformer attention mechanism, which can estimate the crossing intention using only pedestrian bounding boxes as input features [264]. Otherwise, causality relations are studied by explainable AI models to make risk assessment more intelligible [265]. Lastly, simulation tools of road traffic and car driving, such as CARLA [266], can be used as forecasting models provided that mechanisms to adapt synthetically generated data to the reality are put in place [267].

BIP then requires fulfilling the task of predicting the trajectory of the agents. For an AV, the input to the estimation is represented by the historical sequence of coordinates of all traffic participants and possibly other contextual information, e.g., velocity. The task is then to generate a plausible progression of the future position of other pedestrian vehicles. Methods for predicting human motion have been exhaustively surveyed by Rudenko et al. [268], and regarding vehicles, by Huang et al. [269], who classified the approaches into four main categories: physics-based, machine learning, deep learning, and reinforcement learning. Moreover, the authors determined the various contextual factors that may constitute additional inputs for the algorithms similar to those previously described. Finally, they acknowledged that complex deep learning architectures were the de facto solution for real-world implementation for their performance.

Additionally, DL allows for multimodal outputs, i.e., the generation of a diverse trajectory with an associated probability, and for multitask learning, i.e., simultaneously producing a likelihood of a specific behavior. Behavior prediction is, in fact, a separate task, more concerned with assigning to the road participants an intention of performing a particular action. In the recent literature, we can find reviews of approaches specific to understanding the behaviors of vehicles [270] and pedestrians [271]. Behavior prediction is also related to forecasting the occurrence of accidents. This capability is a highly demanded skill in many industrial scenarios.

## 6. Discussion

In the previously presented sections, we thoroughly reviewed the state-of-the-art techniques offered by the scientific community to improve the overall intelligence of autonomous robotic systems. Importing the knowledge from psychology to robotics, we showed that a situational awareness perspective in robotics can efficiently classify these presented state-of-the-art techniques in an organized and multilayered manner. Consequently, we addressed the research questions posed at the beginning:

Through the literature review, we found a gap between the presented approaches to provide a unified and complete Situational Awareness for the robots to understand and reason about the environment in order to perform a mission autonomously and closely to how human beings would. To this end, we proposed an ideal model of the robotic SA system, which, per our mentioned conventions, would be divided into three subsections. The **perception layer** should consist of a multimodal sensor suite for accurate environmental perception. The **comprehension layer** may bear methods from *direct situational comprehension* and *accumulated situational comprehension* to improve the robot's ego-awareness of its state, such as its pose, but also model the external factors with which it interacts, e.g., objects and the environment 3D structure, in the form a metric–semantic topological scene graphs. Finally, the **projection layer**, which still has few connections with the underlying perception and comprehension and is usually treated on a standalone basis, would add forecasting models to predict the future state of the robot as well as the dynamic environmental elements.

The progress in AI and DL has been pivotal in enhancing the robot's comprehension of a situation, as depicted by previous research. Despite significant progress in mobile robots' *direct situational comprehension*, a versatile and standardized sensor suite must be developed to handle environmental challenges, such as meteorological changes. Additionally, integrating these algorithms efficiently into scene modeling frameworks remains a challenge.

The scene graph presented in the related works is a widely adapted term in computer vision [272] to describe the relationship between objects in a scene with a structured representation between entities and predicates usually built from visual information. However, we saw the current scene graphs used in mobile robotics were insufficient to address complex autonomous tasks, such as multimodal open-set queryable maps for navigation [273,274]. Therefore, starting from the scene graph concept, we introduced the *S-Graph* as a knowledge graph that emphasized the ability to store the entire representation of the situation, comprising the currently perceived aspects of the scene, their comprehension, the integration with previous records or possibly also external sources from a standardized ontology [275], and the prediction of the future by the projection of the entities through their models.

Hence, we set the *S-Graph* (see Figure 8) as a future target of the evolution of current scene graphs, which adds a hierarchy of conceptual layers that contribute to including prior knowledge of the situation while maintaining their formulation. Furthermore, the current implementation of the *S-Graph* [276,277], which is still in its initial stage, stresses the practical characteristic of using the created entity relationships, e.g., topological aspect, to obtain an optimized answer on the state of an autonomous agent, e.g., the robot's pose.

We believe this approach can accelerate progress and improve the autonomy of mobile robots.

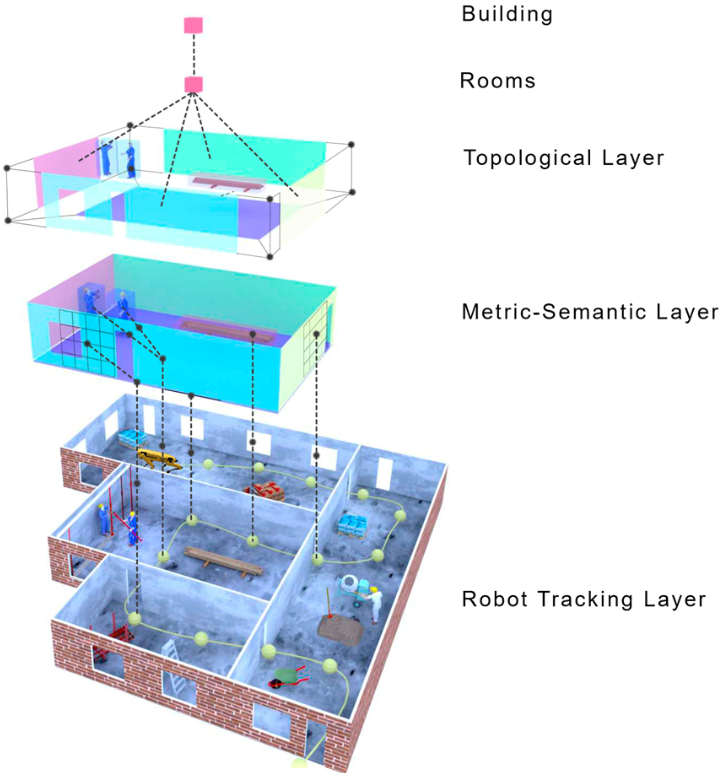


**Figure 8.** Proposed *S-Graph*. The graph is divided into three sublayers: the tracking layer, which tracks the sensor measurements as creates a local key-frame map containing its respective sensor measurements; the metric–semantic layer creates a metric–semantic map using the local key frames; the topological layer consists of the topological connections between the elements in a given area and the rooms that connect planar features.

## 7. Conclusions

In this paper, we argued that Situational Awareness is an essential capability of humans that has been studied in several different fields but has barely been considered in robotics. Instead, robotic research has focused on ideas in a diversified manner, such as sensing, perception, localization, and mapping. Thus, as a direct line of future work, we proposed a three-layered Situational Awareness framework composed of perception, comprehension, and projection. To this end, we provided a thorough literature review of the state-of-the-art techniques for improving robotic intelligence. Then, we reorganized them in a more structured, layered perception, comprehension, and projection format.

Finally, we conclude by providing appropriate answers to the earlier research question.

- *What has been achieved so far, and what challenges remain?*
  Given the advancements in AI and DL, we notice an improved comprehension layer by evaluating state-of-the-art algorithms. Comparing the initial approaches relying on heuristics and heavily engineered processing, current algorithms can solve complex tasks requiring generalization and adaptation in dynamic environments. Nevertheless, the algorithms follow a compartmentalized approach impeding a unified SA for mobile robots. Remarkably, forecasting the future situation is also in its infancy and relies on perfect data from the perception and comprehension layers to demonstrate meaningful results.
- *What could the future direction of Situational Awareness be?*
  We argue that after analyses of these algorithms, a situational awareness perspective can steer robots towards a faster achievement of their tasks, by comprising multimodal hierarchical *S-Graphs* generating a metric–semantic topological map of its environment as well as improving the robot's pose uncertainty in it. We foresee the *S-Graph* will be characterized by a tighter coupling of situational projection, perception, and comprehension, to complete the transition from static world assumptions to natural dynamic environments.

**Author Contributions:** Conceptualization, H.B., J.L.S.-L. and C.C.; methodology H.B., J.L.S.-L. and C.C.; formal analysis, H.B., C.C. and A.T.; investigation, H.B., J.L.S.-L., C.C. and A.T.; resources, H.V.; writing—original draft preparation, H.B., C.C. and A.T.; writing—review and editing, H.B., J.L.S.-L., C.C., A.T. and H.V.; supervision, H.V.; project administration, J.L.S.-L.; funding acquisition, J.L.S.-L. and H.V. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was partially funded by the Fonds National de la Recherche of Luxembourg (FNR), under projects C19/IS/13713801/5G-Sky, by European Commission Horizon 2020 programme, under project SESAME (grant agreement no. 101017258) and by a partnership between the Interdisciplinary Center for Security Reliability and Trust (SnT) of the University of Luxembourg and Stugalux Construction S.A. For the purpose of Open Access, the author has applied a CC BY public copyright license to any author-accepted manuscript version arising from this submission. All authors have read and agreed to the published version of the manuscript.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Tzafestas, S.G. Mobile robot control and navigation: A global overview. *J. Intell. Robot. Syst.* **2018**, *91*, 35–58.
2. Dzedzickis, A.; Subačiūtė-Žemaitienė, J.; Šutinys, E.; Samukaitė-Bubnienė, U.; Bučinskas, V. Advanced Applications of Industrial Robotics: New Trends and Possibilities. *Appl. Sci.* **2021**, *12*, 135. https://doi.org/10.3390/app12010135.
3. Siciliano, B.; Khatib, O. (Eds.) *Springer Handbook of Robotics*; Springer: Berlin/Heidelberg, Germany, 2008. https://doi.org/10.1007/978-3-540-30301-5.
4. Makhataeva, Z.; Varol, H.A. Augmented Reality for Robotics: A Review. *Robotics* **2020**, *9*, 21. https://doi.org/10.3390/robotics9020021.
5. Minaee, S.; Liang, X.; Yan, S. Modern Augmented Reality: Applications, Trends, and Future Directions. *arXiv* **2022**, arXiv:2202.09450.
6. Siegwart, R.; Nourbakhsh, I.R.; Scaramuzza, D. *Introduction to Autonomous Mobile Robots*; MIT Press: Cambridge, MA, USA, 2011.
7. Wong, C.; Yang, E.; Yan, X.T.; Gu, D. Autonomous robots for harsh environments: A holistic overview of current solutions and ongoing challenges. *Syst. Sci. Control Eng.* **2018**, *6*, 213–219. https://doi.org/10.1080/21642583.2018.1477634.
8. Salas, E. *Situational Awareness*; Routledge: Abingdon, UK, 2017.
9. Endsley, M.R. Toward a Theory of Situation Awareness in Dynamic Systems. *Hum. Factors* **1995**, *37*, 32–64.
10. Munir, A.; Aved, A.; Blasch, E. Situational Awareness: Techniques, Challenges, and Prospects. *AI* **2022**, *3*, 55–77. https://doi.org/10.3390/ai3010005.
11. Rubio, F.; Valero, F.; Llopis-Albert, C. A review of mobile robots: Concepts, methods, theoretical framework, and applications. *Int. J. Adv. Robot. Syst.* **2019**, *16*, 1729881419839596.
12. K., N.; A. G., S.; Mathew, J.; Sarpotdar, M.; Suresh, A.; Prakash, A.; Safonova, M.; Murthy, J. Noise modeling and analysis of an IMU-based attitude sensor: improvement of performance by filtering and sensor fusion. In *Advances in Optical and Mechanical Technologies for Telescopes and Instrumentation II*; SPIE: Edinburgh, UK, 2016. https://doi.org/10.1117/12.2234255.

13. Sabatini, A.; Genovese, V. A Stochastic Approach to Noise Modeling for Barometric Altimeters. *Sensors* **2013**, *13*, 15692–15707.
14. Zimmermann, F.; Eling, C.; Klingbeil, L.; Kuhlmann, H. Precise Positioning of Uavs—Dealing with Challenging Rtk-Gps Measurement Conditions during Automated Uav Flights. In *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*; ISPRS: Hannover, Germany, 2017; Volume 42W3, pp. 95–102. https://doi.org/10.5194/isprs-annals-IV-2-W3-95-2017.
15. Tourani, A.; Bavle, H.; Sanchez-Lopez, J.L.; Voos, H. Visual SLAM: What Are the Current Trends and What to Expect? *Sensors* **2022**, *22*, 9297.
16. Indiveri, G.; Douglas, R. Neuromorphic vision sensors. *Science* **2000**, *288*, 1189–1190.
17. Gallego, G.; Delbruck, T.; Orchard, G.M.; Bartolozzi, C.; Taba, B.; Censi, A.; Leutenegger, S.; Davison, A.; Conradt, J.; Daniilidis, K.; et al. Event-based Vision: A Survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *44*, 154–180. https://doi.org/10.1109/TPAMI.2020.3008413.
18. Lichtsteiner, P.; Posch, C.; Delbruck, T. A 128× 128 120 dB 15 micro-sec Latency Asynchronous Temporal Contrast Vision Sensor. *IEEE J. Solid-State Circuits* **2008**, *43*, 566–576. https://doi.org/10.1109/JSSC.2007.914337.
19. Brandli, C.; Berner, R.; Yang, M.; Liu, S.C.; Delbruck, T. A 240 × 180 130 dB 3 µs Latency Global Shutter Spatiotemporal Vision Sensor. *IEEE J. Solid-State Circuits* **2014**, *49*, 2333–2341. https://doi.org/10.1109/JSSC.2014.2342715.
20. Posch, C.; Matolin, D.; Wohlgenannt, R. A QVGA 143 dB Dynamic Range Frame-Free PWM Image Sensor With Lossless Pixel-Level Video Compression and Time-Domain CDS. *IEEE J. Solid-State Circuits* **2011**, *46*, 259–275. https://doi.org/10.1109/JSSC.2010.2085952.
21. Rebecq, H.; Ranftl, R.; Koltun, V.; Scaramuzza, D. Events-to-Video: Bringing Modern Computer Vision to Event Cameras. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019.
22. Rebecq, H.; Ranftl, R.; Koltun, V.; Scaramuzza, D. High Speed and High Dynamic Range Video with an Event Camera. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *43*, 1964–1980.
23. Venon, A.; Dupuis, Y.; Vasseur, P.; Merriaux, P. Millimeter wave FMCW radars for perception, recognition and localization in automotive applications: A survey. *IEEE Trans. Intell. Veh.* **2022**, *7*, 533–555.
24. Kabiri, M.; Cimarelli, C.; Bavle, H.; Sanchez-Lopez, J.L.; Voos, H. A Review of Radio Frequency Based Localisation for Aerial and Ground Robots with 5G Future Perspectives. *Sensors* **2023**, *23*, 188. https://doi.org/10.3390/s23010188.
25. Viola, P.; Jones, M. Rapid object detection using a boosted cascade of simple features. In Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2001, Kauai, HI, USA, 8–14 December 2001; Volume 1, p. I. https://doi.org/10.1109/CVPR.2001.990517.
26. Nguyen, T.; Park, E.A.; Han, J.; Park, D.C.; Min, S.Y. Object Detection Using Scale Invariant Feature Transform. In *Genetic and Evolutionary Computing*; Pan, J.S., Krömer, P., Snášel, V., Eds.; Springer International Publishing: Cham, Switzerland, 2014; pp. 65–72.
27. Li, Q.; Wang, X. Image Classification Based on SIFT and SVM. In Proceedings of the 2018 IEEE/ACIS 17th International Conference on Computer and Information Science (ICIS), Singapore, 6–8 June 2018; pp. 762–765. https://doi.org/10.1109/ICIS.2018.8466432.
28. Kachouane, M.; Sahki, S.; Lakrouf, M.; Ouadah, N. HOG based fast human detection. In Proceedings of the 2012 24th International Conference on Microelectronics (ICM), Algiers, Algeria, 16–20 December 2012. https://doi.org/10.1109/icm.2012.6471380.
29. Enzweiler, M.; Gavrila, D.M. Monocular Pedestrian Detection: Survey and Experiments. *IEEE Trans. Pattern Anal. Mach. Intell.* **2009**, *31*, 2179–2195. https://doi.org/10.1109/TPAMI.2008.260.
30. Messelodi, S.; Modena, C.M.; Cattoni, G. Vision-based bicycle/motorcycle classification. *Pattern Recognit. Lett.* **2007**, *28*, 1719–1726. https://doi.org/10.1016/j.patrec.2007.04.014.
31. Lowe, D.G. Distinctive Image Features from Scale-Invariant Keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110. https://doi.org/10.1023/B:VISI.0000029664.99615.94.
32. Bay, H.; Tuytelaars, T.; Van Gool, L. SURF: Speeded Up Robust Features. In *Computer Vision—ECCV 2006*; Leonardis, A., Bischof, H., Pinz, A., Eds.; Springer: Berlin/Heidelberg, Germany, 2006; pp. 404–417.
33. Dalal, N.; Triggs, B. Histograms of oriented gradients for human detection. In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 20–25 June 2005; Volume 1, pp. 886–893. https://doi.org/10.1109/CVPR.2005.177.
34. Hearst, M.; Dumais, S.; Osuna, E.; Platt, J.; Scholkopf, B. Support vector machines. *IEEE Intell. Syst. Their Appl.* **1998**, *13*, 18–28. https://doi.org/10.1109/5254.708428.
35. Bochkovskiy, A.; Wang, C.Y.; Liao, H.Y.M. YOLOv4: Optimal Speed and Accuracy of Object Detection. *arXiv* **2020**, arXiv:2004.10934.
36. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask R-CNN. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 2980–2988. https://doi.org/10.1109/ICCV.2017.322.
37. Lin, T.Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal Loss for Dense Object Detection. *arXiv* **2018**, arXiv:1708.02002.
38. Chen, X.; Girshick, R.; He, K.; Dollár, P. TensorMask: A Foundation for Dense Object Segmentation. *arXiv* **2019**, arXiv:1903.12174.
39. Li, Y.; Chen, Y.; Wang, N.; Zhang, Z. Scale-Aware Trident Networks for Object Detection. *arXiv* **2019**, arXiv:1901.01892.
40. Long, J.; Shelhamer, E.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation. *arXiv* **2015**, arXiv:1411.4038.

41. Chen, L.C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. *arXiv* **2018**, arXiv:1802.02611.
42. Kirillov, A.; Wu, Y.; He, K.; Girshick, R. PointRend: Image Segmentation as Rendering. *arXiv* **2020**, arXiv:1912.08193.
43. Poudel, R.P.K.; Liwicki, S.; Cipolla, R. Fast-SCNN: Fast Semantic Segmentation Network. *arXiv* **2019**, arXiv:1902.04502.
44. Chen, L.C.; Papandreou, G.; Schroff, F.; Adam, H. Rethinking Atrous Convolution for Semantic Image Segmentation. *arXiv* **2017**, arXiv:1706.05587.
45. Kirillov, A.; Girshick, R.; He, K.; Dollár, P. Panoptic Feature Pyramid Networks. *arXiv* **2019**, arXiv:1901.02446.
46. Cheng, B.; Collins, M.D.; Zhu, Y.; Liu, T.; Huang, T.S.; Adam, H.; Chen, L.C. Panoptic-DeepLab: A Simple, Strong, and Fast Baseline for Bottom-Up Panoptic Segmentation. *arXiv* **2020**, arXiv:1911.10194.
47. Xu, D.; Zhu, Y.; Choy, C.B.; Fei-Fei, L. Scene Graph Generation by Iterative Message Passing. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017.
48. Zareian, A.; Karaman, S.; Chang, S.F. Bridging Knowledge Graphs to Generate Scene Graphs. In *Computer Vision – ECCV 2020*; Springer International Publishing: Berlin/Heidelberg, Germany, 2020; pp. 606–623. https://doi.org/10.1007/978-3-030-58592-1_36.
49. Suhail, M.; Mittal, A.; Siddiquie, B.; Broaddus, C.; Eledath, J.; Medioni, G.; Sigal, L. Energy-based learning for scene graph generation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 13936–13945.
50. Wang, W.; Zhang, J.; Shen, C. Improved human detection and classification in thermal images. In Proceedings of the 2010 IEEE International Conference on Image Processing, Hong Kong, China, 26–29 September 2010; pp. 2313–2316. https://doi.org/10.1109/ICIP.2010.5649946.
51. Krišto, M.; Ivasic-Kos, M.; Pobar, M. Thermal Object Detection in Difficult Weather Conditions Using YOLO. *IEEE Access* **2020**, *8*, 125459–125476. https://doi.org/10.1109/ACCESS.2020.3007481.
52. Ippalapally, R.; Mudumba, S.H.; Adkay, M.; R., N.V.H. Object Detection Using Thermal Imaging. In Proceedings of the 2020 IEEE 17th India Council International Conference (INDICON), New Delhi, India, 10–13 December 2020; pp. 1–6. https://doi.org/10.1109/INDICON49873.2020.9342179.
53. Cannici, M.; Ciccone, M.; Romanoni, A.; Matteucci, M. Asynchronous Convolutional Networks for Object Detection in Neuromorphic Cameras. *arXiv* **2019**, arXiv:1805.07931.
54. Alonso, I.; Murillo, A.C. EV-SegNet: Semantic Segmentation for Event-based Cameras. *arXiv* **2018**, arXiv:1811.12039.
55. Stiene, S.; Lingemann, K.; Nuchter, A.; Hertzberg, J. Contour-Based Object Detection in Range Images. In Proceedings of the Third International Symposium on 3D Data Processing, Visualization, and Transmission (3DPVT'06), Chapel Hill, NC, USA, 14–16 June 2006; pp. 168–175. https://doi.org/10.1109/3DPVT.2006.46.
56. Himmelsbach, M.; Mueller, A.; Lüttel, T.; Wünsche, H.J. LIDAR-based 3D object perception. In Proceedings of the 1st International Workshop on Cognition for Technical Systems, Munich, Germany, 6–8 October 2008; Volume 1.
57. Kragh, M.; Jørgensen, R.N.; Pedersen, H. Object Detection and Terrain Classification in Agricultural Fields Using 3D Lidar Data. In *Computer Vision Systems*; Nalpantidis, L., Krüger, V., Eklundh, J.O., Gasteratos, A., Eds.; Springer International Publishing: Cham, Switzerland, 2015; pp. 188–197.
58. Milioto, A.; Vizzo, I.; Behley, J.; Stachniss, C. RangeNet ++: Fast and Accurate LiDAR Semantic Segmentation. In Proceedings of the 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Macau, China, 3–8 November 2019; pp. 4213–4220. https://doi.org/10.1109/IROS40897.2019.8967762.
59. Lyu, Y.; Huang, X.; Zhang, Z. Learning to Segment 3D Point Clouds in 2D Image Space. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 12255–12264. Available online: http://xxx.lanl.gov/abs/2003.05593 (accessed on 10 April 2023).
60. Wu, B.; Wan, A.; Yue, X.; Keutzer, K. SqueezeSeg: Convolutional Neural Nets with Recurrent CRF for Real-Time Road-Object Segmentation from 3D LiDAR Point Cloud. *arXiv* **2017**, arXiv:1710.07368.
61. Wu, B.; Zhou, X.; Zhao, S.; Yue, X.; Keutzer, K. SqueezeSegV2: Improved Model Structure and Unsupervised Domain Adaptation for Road-Object Segmentation from a LiDAR Point Cloud. *arXiv* **2018**, arXiv:1809.08495.
62. Qi, C.R.; Su, H.; Mo, K.; Guibas, L.J. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. *arXiv* **2017**, arXiv:1612.00593.
63. Qi, C.R.; Yi, L.; Su, H.; Guibas, L.J. PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space. *arXiv* **2017**, arXiv:1706.02413.
64. Tatarchenko, M.; Park, J.; Koltun, V.; Zhou, Q. Tangent Convolutions for Dense Prediction in 3D. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; pp. 3887–3896. Available online: http://xxx.lanl.gov/abs/1807.02443 (accessed on 10 April 2023).
65. Najibi, M.; Lai, G.; Kundu, A.; Lu, Z.; Rathod, V.; Funkhouser, T.; Pantofaru, C.; Ross, D.; Davis, L.S.; Fathi, A. DOPS: Learning to Detect 3D Objects and Predict their 3D Shapes. *arXiv* **2020**, arXiv:2004.01170.
66. Hu, Q.; Yang, B.; Xie, L.; Rosa, S.; Guo, Y.; Wang, Z.; Trigoni, N.; Markham, A. RandLA-Net: Efficient Semantic Segmentation of Large-Scale Point Clouds. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020. Available online: http://xxx.lanl.gov/abs/1911.11236 (accessed on 10 April 2023).

67. González, A.; Vázquez, D.; López, A.M.; Amores, J. On-Board Object Detection: Multicue, Multimodal, and Multiview Random Forest of Local Experts. *IEEE Trans. Cybern.* **2017**, *47*, 3980–3990. https://doi.org/10.1109/TCYB.2016.2593940.
68. Lin, D.; Fidler, S.; Urtasun, R. Holistic Scene Understanding for 3D Object Detection with RGBD Cameras. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Sydney, NSW, Australia, 1–8 December 2013.
69. Schwarz, M.; Milan, A.; Periyasamy, A.S.; Behnke, S. RGB-D object detection and semantic segmentation for autonomous manipulation in clutter. *Int. J. Robot. Res.* **2018**, *37*, 437–451. https://doi.org/10.1177/0278364917713117.
70. Xiang, Y.; Schmidt, T.; Narayanan, V.; Fox, D. PoseCNN: A Convolutional Neural Network for 6D Object Pose Estimation in Cluttered Scenes. *arXiv* **2018**, arXiv:1711.00199.
71. Wang, C.; Xu, D.; Zhu, Y.; Martin-Martin, R.; Lu, C.; Fei-Fei, L.; Savarese, S. DenseFusion: 6D Object Pose Estimation by Iterative Dense Fusion. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019.
72. Wang, H.; Sridhar, S.; Huang, J.; Valentin, J.; Song, S.; Guibas, L.J. Normalized Object Coordinate Space for Category-Level 6D Object Pose and Size Estimation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019.
73. Lin, Y.; Tremblay, J.; Tyree, S.; Vela, P.A.; Birchfield, S. Multi-view Fusion for Multi-level Robotic Scene Understanding. *arXiv* **2021**, arXiv:2103.13539.
74. Ha, Q.; Watanabe, K.; Karasawa, T.; Ushiku, Y.; Harada, T. MFNet: Towards real-time semantic segmentation for autonomous vehicles with multi-spectral scenes. In Proceedings of the 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Vancouver, BC, 2076 , 24–28 September 2017; pp. 5108–5115. https://doi.org/10.1109/IROS.2017.8206396.
75. Sun, Y.; Zuo, W.; Liu, M. RTFNet: RGB-Thermal Fusion Network for Semantic Segmentation of Urban Scenes. *IEEE Robot. Autom. Lett.* **2019**, *4*, 2576–2583. https://doi.org/10.1109/LRA.2019.2904733.
76. Shivakumar, S.S.; Rodrigues, N.; Zhou, A.; Miller, I.D.; Kumar, V.; Taylor, C.J. PST900: RGB-Thermal Calibration, Dataset and Segmentation Network. In Proceedings of the 2020 IEEE International Conference on Robotics and Automation (ICRA), Paris, France, 31 May–31 August 2020; pp. 9441–9447. https://doi.org/10.1109/ICRA40945.2020.9196831.
77. Sun, Y.; Zuo, W.; Yun, P.; Wang, H.; Liu, M. FuseSeg: Semantic Segmentation of Urban Scenes Based on RGB and Thermal Data Fusion. *IEEE Trans. Autom. Sci. Eng.* **2021**, *18*, 1000–1011. https://doi.org/10.1109/TASE.2020.2993143.
78. Zhang, J.; Yang, K.; Stiefelhagen, R. ISSAFE: Improving Semantic Segmentation in Accidents by Fusing Event-based Data. *arXiv* **2020**, arXiv:2008.08974.
79. Qi, C.R.; Liu, W.; Wu, C.; Su, H.; Guibas, L.J. Frustum PointNets for 3D Object Detection from RGB-D Data. *arXiv* **2018**, arXiv:1711.08488.
80. Ku, J.; Mozifian, M.; Lee, J.; Harakeh, A.; Waslander, S.L. Joint 3D Proposal Generation and Object Detection from View Aggregation. In Proceedings of the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Madrid, Spain, 1–5 October 2018; pp. 1–8. https://doi.org/10.1109/IROS.2018.8594049.
81. Liang, M.; Yang, B.; Wang, S.; Urtasun, R. Deep Continuous Fusion for Multi-Sensor 3D Object Detection. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018.
82. Chen, X.; Ma, H.; Wan, J.; Li, B.; Xia, T. Multi-View 3D Object Detection Network for Autonomous Driving. *arXiv* **2017**, arXiv:1611.07759.
83. Xu, D.; Anguelov, D.; Jain, A. PointFusion: Deep Sensor Fusion for 3D Bounding Box Estimation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018.
84. Papers With Code. Available online: https://paperswithcode.com/area/computer-vision (accessed on 10 April 2023).
85. Mitrokhin, A.; Fermüller, C.; Parameshwara, C.; Aloimonos, Y. Event-Based Moving Object Detection and Tracking. In Proceedings of the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Madrid, Spain, 1–5 October 2018; pp. 1–9. https://doi.org/10.1109/IROS.2018.8593805.
86. Hall, D.; Llinas, J. An introduction to multisensor data fusion. *Proc. IEEE* **1997**, *85*, 6–23. https://doi.org/10.1109/5.554205.
87. Alldieck, T.; Bahnsen, C.H.; Moeslund, T.B. Context-Aware Fusion of RGB and Thermal Imagery for Traffic Monitoring. *Sensors* **2016**, *16*, 1947. https://doi.org/10.3390/s16111947.
88. Zhou, W.; Guo, Q.; Lei, J.; Yu, L.; Hwang, J.N. ECFFNet: Effective and Consistent Feature Fusion Network for RGB-T Salient Object Detection. *IEEE Trans. Circuits Syst. Video Technol.* **2021**, *32*, 1224–1235. https://doi.org/10.1109/TCSVT.2021.3077058.
89. Spremolla, I.R.; Antunes, M.; Aouada, D.; Ottersten, B.E. RGB-D and Thermal Sensor Fusion-Application in Person Tracking. In *Proceedings of the 11th Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications—Volume 3: VISAPP*; SciTePress: Rome, Italy, 2016; pp. 610–617.
90. Mogelmose, A.; Bahnsen, C.; Moeslund, T.B.; Clapes, A.; Escalera, S. Tri-modal Person Re-identification with RGB, Depth and Thermal Features. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, Portland, OR, USA, 23–28 June 2013.
91. Dubeau, E.; Garon, M.; Debaque, B.; Charette, R.d.; Lalonde, J.F. RGB-D-E: Event Camera Calibration for Fast 6-DOF object Tracking. In Proceedings of the 2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), Ipojuca, Brasil, 9–13 November 2020; pp. 127–135. https://doi.org/10.1109/ISMAR50242.2020.00034.

92. Dellaert, F.; Fox, D.; Burgard, W.; Thrun, S. Monte Carlo localization for mobile robots. In Proceedings of the 1999 IEEE International Conference on Robotics and Automation (Cat. No.99CH36288C), Detroit, MI, USA, 10–15 May 1999; Volume 2, pp. 1322–1328. https://doi.org/10.1109/ROBOT.1999.772544.
93. Thrun, S.; Fox, D.; Burgard, W.; Dellaert, F. Robust Monte Carlo localization for mobile robots. *Artif. Intell.* **2001**, *128*, 99–141. https://doi.org/10.1016/S0004-3702(01)00069-8.
94. Anjum, M.L.; Park, J.; Hwang, W.; Kwon, H.i.; Kim, J.h.; Lee, C.; Kim, K.s.; "Dan" Cho, D.i. Sensor data fusion using Unscented Kalman Filter for accurate localization of mobile robots. In Proceedings of the ICCAS 2010, Gyeonggi-do, Republic of Korea, 27–30 October 2010; pp. 947–952. https://doi.org/10.1109/ICCAS.2010.5669779.
95. Kong, F.; Chen, Y.; Xie, J.; Zhang, G.; Zhou, Z. Mobile Robot Localization Based on Extended Kalman Filter. In Proceedings of the 2006 6th World Congress on Intelligent Control and Automation, Dalian, China 21–23 June 2006; Volume 2, pp. 9242–9246. https://doi.org/10.1109/WCICA.2006.1713789.
96. Teslic, L.; skrjanc, I.; Klanvcar, G. EKF-Based Localization of a Wheeled Mobile Robot in Structured Environments. *J. Intell. Robot. Syst.* **2011**, *62*, 187–203.
97. Chen, L.; Hu, H.; McDonald-Maier, K. EKF Based Mobile Robot Localization. In Proceedings of the 2012 Third International Conference on Emerging Security Technologies, Lisbon, Portugal, 5–7 September 2012; pp. 149–154. https://doi.org/10.1109/EST.2012.19.
98. Ganganath, N.; Leung, H. Mobile robot localization using odometry and kinect sensor. In Proceedings of the 2012 IEEE International Conference on Emerging Signal Processing Applications, IEEE, Las Vegas, NV, USA, 12–14 January 2012; pp. 91–94.
99. Kim, S.J.; Kim, B.K. Dynamic Ultrasonic Hybrid Localization System for Indoor Mobile Robots. *IEEE Trans. Ind. Electron.* **2013**, *60*, 4562–4573. https://doi.org/10.1109/TIE.2012.2216235.
100. Lynen, S.; Achtelik, M.W.; Weiss, S.; Chli, M.; Siegwart, R. A robust and modular multi-sensor fusion approach applied to MAV navigation. In Proceedings of the 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems, Tokyo, Japan, 3–7 November 2013; pp. 3923–3929. https://doi.org/10.1109/IROS.2013.6696917.
101. Sanchez-Lopez, J.L.; Arellano-Quintana, V.; Tognon, M.; Campoy, P.; Franchi, A. Visual Marker based Multi-Sensor Fusion State Estimation. *IFAC-PapersOnLine* **2017**, *50*, 16003–16008. https://doi.org/10.1016/j.ifacol.2017.08.1911.
102. Moore, T.; Stouch, D.W. A Generalized Extended Kalman Filter Implementation for the Robot Operating System. In Proceedings of the IAS, Pedova, Italy, 15–18 July 2014.
103. Wan, G.; Yang, X.; Cai, R.; Li, H.; Zhou, Y.; Wang, H.; Song, S. Robust and Precise Vehicle Localization Based on Multi-Sensor Fusion in Diverse City Scenes. In Proceedings of the 2018 IEEE International Conference on Robotics and Automation (ICRA), Brisbane, Australia, 21–26 May 2018. https://doi.org/10.1109/icra.2018.8461224.
104. Liu, F.; Li, X.; Yuan, S.; Lan, W. Slip-Aware Motion Estimation for Off-Road Mobile Robots via Multi-Innovation Unscented Kalman Filter. *IEEE Access* **2020**, *8*, 43482–43496. https://doi.org/10.1109/ACCESS.2020.2977889.
105. Kimura, K.; Hiromachi, Y.; Nonaka, K.; Sekiguchi, K. Vehicle localization by sensor fusion of LRS measurement and odometry information based on moving horizon estimation. In Proceedings of the 2014 IEEE Conference on Control Applications (CCA), Juan Les Antibes, France, 8–10 October 2014; pp. 1306–1311. https://doi.org/10.1109/CCA.2014.6981509.
106. Liu, A.; Zhang, W.A.; Chen, M.Z.Q.; Yu, L. Moving Horizon Estimation for Mobile Robots With Multirate Sampling. *IEEE Trans. Ind. Electron.* **2017**, *64*, 1457–1467. https://doi.org/10.1109/TIE.2016.2611458.
107. Dubois, R.; Bertrand, S.; Eudes, A. Performance Evaluation of a Moving Horizon Estimator for Multi-Rate Sensor Fusion with Time-Delayed Measurements. In Proceedings of the 2018 22nd International Conference on System Theory, Control and Computing (ICSTCC), Sinaia, Romania, 8–10 October 2018; pp. 664–669. https://doi.org/10.1109/ICSTCC.2018.8540711.
108. Osman, M.; Mehrez, M.W.; Daoud, M.A.; Hussein, A.; Jeon, S.; Melek, W. A generic multi-sensor fusion scheme for localization of autonomous platforms using moving horizon estimation. *Trans. Inst. Meas. Control* **2021**, *43*, 3413–3427. https://doi.org/10.1177/01423312211011454.
109. Ranganathan, A.; Kaess, M.; Dellaert, F. Fast 3D pose estimation with out-of-sequence measurements. In Proceedings of the 2007 IEEE/RSJ International Conference on Intelligent Robots and Systems, San Diego, CA, USA, 29 October–2 November 2007; pp. 2486–2493. https://doi.org/10.1109/IROS.2007.4399318.
110. Dellaert, F.; Kaess, M. Square Root SAM: Simultaneous Localization and Mapping via Square Root Information Smoothing. *Int. J. Robot. Res.* **2006**, *25*, 1181–1203. https://doi.org/10.1177/0278364906072768.
111. Indelman, V.; Williams, S.; Kaess, M.; Dellaert, F. Factor graph based incremental smoothing in inertial navigation systems. In Proceedings of the 2012 15th International Conference on Information Fusion, Singapore, 9–12 July 2012; pp. 2154–2161.
112. Kaess, M.; Johannsson, H.; Roberts, R.; Ila, V.; Leonard, J.J.; Dellaert, F. iSAM2: Incremental smoothing and mapping using the Bayes tree. *Int. J. Robot. Res.* **2012**, *31*, 216–235. https://doi.org/10.1177/0278364911430419.
113. Merfels, C.; Stachniss, C. Pose fusion with chain pose graphs for automated driving. In Proceedings of the 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Daejeon, Republic of Korea, 9–14 October 2016; pp. 3116–3123. https://doi.org/10.1109/IROS.2016.7759482.
114. Merfels, C.; Stachniss, C. Sensor Fusion for Self-Localisation of Automated Vehicles. *PFG—J. Photogramm. Remote Sens. Geoinf. Sci.* **2017**, *85*, 113–126. https://doi.org/10.1007/s41064-017-0008-1.

115. Mascaro, R.; Teixeira, L.; Hinzmann, T.; Siegwart, R.; Chli, M. GOMSF: Graph-Optimization Based Multi-Sensor Fusion for robust UAV Pose estimation. In Proceedings of the 2018 IEEE International Conference on Robotics and Automation (ICRA), Brisbane, Australia, 21–26 May 2018; pp. 1421–1428. https://doi.org/10.1109/ICRA.2018.8460193.
116. Qin, T.; Cao, S.; Pan, J.; Shen, S. A General Optimization-based Framework for Global Pose Estimation with Multiple Sensors. *arXiv* **2019**, arXiv:1901.03642.
117. Li, X.; Wang, X.; Liao, J.; Li, X.; Li, S.; Lyu, H. Semi-tightly coupled integration of multi-GNSS PPP and S-VINS for precise positioning in GNSS-challenged environments. *Satell. Navig.* **2021**, *2*, 1. https://doi.org/10.1186/s43020-020-00033-9.
118. Cadena, C.; Carlone, L.; Carrillo, H.; Latif, Y.; Scaramuzza, D.; Neira, J.; Reid, I.; Leonard, J.J. Past, Present, and Future of Simultaneous Localization and Mapping: Toward the Robust-Perception Age. *IEEE Trans. Robot.* **2016**, *32*, 1309–1332. https://doi.org/10.1109/TRO.2016.2624754.
119. Chen, W.; Shang, G.; Ji, A.; Zhou, C.; Wang, X.; Xu, C.; Li, Z.; Hu, K. An Overview on Visual SLAM: From Tradition to Semantic. *Remote Sens.* **2022**, *14*. https://doi.org/10.3390/rs14133010.
120. Lu, F.; Milios, E. Globally Consistent Range Scan Alignment for Environment Mapping. *Auton. Robot.* **1997**, *4*, 333–349.
121. Leonard, J.J.; Feder, H.J.S. A Computationally Efficient Method for Large-Scale Concurrent Mapping and Localization. In *Robotics Research*; Hollerbach, J.M., Koditschek, D.E., Eds.; Springer: London, UK, 2000; pp. 169–176.
122. Guivant, J.; Nebot, E. Optimization of the simultaneous localization and map-building algorithm for real-time implementation. *IEEE Trans. Robot. Autom.* **2001**, *17*, 242–257. https://doi.org/10.1109/70.938382.
123. Bailey, T.; Nieto, J.; Guivant, J.; Stevens, M.; Nebot, E. Consistency of the EKF-SLAM Algorithm. In Proceedings of the 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems, Beijing, China, 9–13 October 2006; pp. 3562–3568. https://doi.org/10.1109/IROS.2006.281644.
124. Thrun, S.; Montemerlo, M.; Koller, D.; Wegbreit, B.; Nieto, J.; Nebot, E. Fastslam: An efficient solution to the simultaneous localization and mapping problem with unknown data association. *J. Mach. Learn. Res.* **2004**, *4*, 380–407.
125. Folkesson, J.; Christensen, H.I. Graphical SLAM for Outdoor Applications. *J. Field Robot.* **2007**, *24*, 51–70. https://doi.org/10.1002/rob.20174.
126. Olson, E.; Leonard, J.; Teller, S. Fast iterative alignment of pose graphs with poor initial estimates. In Proceedings of the 2006 IEEE International Conference on Robotics and Automation (ICRA), Orlando, FL, USA, 15–19 May 2006; pp. 2262–2269.
127. Thrun, S.; Montemerlo, M. The Graph SLAM Algorithm with Applications to Large-Scale Mapping of Urban Structures. *Int. J. Robot. Res.* **2006**, *25*, 403–429. https://doi.org/10.1177/0278364906065387.
128. Mur-Artal, R.; Montiel, J.M.M.; Tardos, J.D. ORB-SLAM: A Versatile and Accurate Monocular SLAM System. *IEEE Trans. Robot.* **2015**, *31*, 1147–1163. https://doi.org/10.1109/tro.2015.2463671.
129. Pizzoli, M.; Forster, C.; Scaramuzza, D. REMODE: Probabilistic, monocular dense reconstruction in real time. In Proceedings of the 2014 IEEE International Conference on Robotics and Automation (ICRA), Honkong, China, 31 May–5 June 2014; pp. 2609–2616. https://doi.org/10.1109/ICRA.2014.6907233.
130. Engel, J.; Sturm, J.; Cremers, D. Semi-dense Visual Odometry for a Monocular Camera. In Proceedings of the 2013 IEEE International Conference on Computer Vision, Sydney, Australia, 1–8 December 2013; pp. 1449–1456. https://doi.org/10.1109/ICCV.2013.183.
131. Gao, X.; Wang, R.; Demmel, N.; Cremers, D. LDSO: Direct Sparse Odometry with Loop Closure. In Proceedings of the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Madrid, Spain, 1–5 October 2018; pp. 2198–2204. https://doi.org/10.1109/IROS.2018.8593376.
132. Engel, J.; Schöps, T.; Cremers, D. LSD-SLAM: Large-Scale Direct Monocular SLAM. In *Computer Vision—ECCV 2014*; Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T., Eds.; Springer International Publishing: Cham, Switzerland, 2014; pp. 834–849.
133. Concha, A.; Civera, J. DPPTAM: Dense piecewise planar tracking and mapping from a monocular sequence. In Proceedings of the 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Hamburg, Germany, 28 September–2 October 2015; pp. 5686–5693. https://doi.org/10.1109/IROS.2015.7354184.
134. Zubizarreta, J.; Aguinaga, I.; Montiel, J.M.M. Direct Sparse Mapping. *IEEE Trans. Robot.* **2020**, *36*, 1363–1370. https://doi.org/10.1109/tro.2020.2991614.
135. Forster, C.; Pizzoli, M.; Scaramuzza, D. SVO: Fast semi-direct monocular visual odometry. In Proceedings of the 2014 IEEE International Conference on Robotics and Automation (ICRA), Hongkong, China, 31 May–5 June 2014; pp. 15–22. https://doi.org/10.1109/ICRA.2014.6906584.
136. Ma, L.; Kerl, C.; Stückler, J.; Cremers, D. CPA-SLAM: Consistent plane-model alignment for direct RGB-D SLAM. In Proceedings of the 2016 IEEE International Conference on Robotics and Automation (ICRA), IEEE, Stockholm, Sweden, 16–21 May 2016; pp. 1285–1291.
137. Lee, S.H.; Civera, J. Loosely-Coupled Semi-Direct Monocular SLAM. *IEEE Robot. Autom. Lett.* **2019**, *4*, 399–406. https://doi.org/10.1109/LRA.2018.2889156.
138. Yang, N.; Stumberg, L.v.; Wang, R.; Cremers, D. D3vo: Deep depth, deep pose and deep uncertainty for monocular visual odometry. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA 13–19 June 2020; pp. 1281–1292.
139. Carlone, L.; Tron, R.; Daniilidis, K.; Dellaert, F. Initialization techniques for 3D SLAM: A survey on rotation estimation and its use in pose graph optimization. In Proceedings of the 2015 IEEE International Conference on Robotics and Automation (ICRA), IEEE, Seattle, WA, USA, 26–30 May 2015; pp. 4597–4604.

140. Jiao, J.; Jiao, J.; Mo, Y.; Liu, W.; Deng, Z. MagicVO: End-to-End Monocular Visual Odometry through Deep Bi-directional Recurrent Convolutional Neural Network. *arXiv* **2018**, arXiv:1811.10964.
141. Wang, S.; Clark, R.; Wen, H.; Trigoni, N. DeepVO: Towards end-to-end visual odometry with deep Recurrent Convolutional Neural Networks. In Proceedings of the 2017 IEEE International Conference on Robotics and Automation (ICRA), Singapore, 29 May–3 June 2017. https://doi.org/10.1109/icra.2017.7989236.
142. Bruno, H.M.S.; Colombini, E.L. LIFT-SLAM: A deep-learning feature-based monocular visual SLAM method. *Neurocomputing* **2021**, *455*, 97–110.
143. Peng, Q.; Xiang, Z.; Fan, Y.; Zhao, T.; Zhao, X. RWT-SLAM: Robust visual SLAM for highly weak-textured environments. *arXiv* **2022**, arXiv:2207.03539.
144. Naveed, K.; Anjum, M.L.; Hussain, W.; Lee, D. Deep introspective SLAM: Deep reinforcement learning based approach to avoid tracking failure in visual SLAM. *Auton. Robot.* **2022**, *46*, 705–724.
145. Sun, Y.; Hu, J.; Yun, J.; Liu, Y.; Bai, D.; Liu, X.; Zhao, G.; Jiang, G.; Kong, J.; Chen, B. Multi-objective location and mapping based on deep learning and visual slam. *Sensors* **2022**, *22*, 7576.
146. Godard, C.; Aodha, O.M.; Firman, M.; Brostow, G.J. Digging into self-supervised monocular depth estimation. In Proceedings of the IEEE International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 Novemebr 2019; pp. 3827–3837. https://doi.org/10.1109/ICCV.2019.00393.
147. Zhou, T.; Brown, M.; Snavely, N.; Lowe, D.G. Unsupervised learning of depth and ego-motion from video. In Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR, Honolulu, HI, USA, 21–26 July 2017, pp. 6612–6621. https://doi.org/10.1109/CVPR.2017.700.
148. Li, R.; Wang, S.; Long, Z.; Gu, D. UnDeepVO: Monocular Visual Odometry Through Unsupervised Deep Learning. In Proceedings of the IEEE International Conference on Robotics and Automation, Brisbane, Australia, 21–25 May 2018; pp. 7286–7291. https://doi.org/10.1109/ICRA.2018.8461251.
149. Vödisch, N.; Cattaneo, D.; Burgard, W.; Valada, A. Continual slam: Beyond lifelong simultaneous localization and mapping through continual learning. In *Robotics Research*; Springer Berlin/Heidelberg, Germany, 2023; pp. 19–35.
150. Zhang, J.; Sui, W.; Wang, X.; Meng, W.; Zhu, H.; Zhang, Q. Deep online correction for monocular visual odometry. In Proceedings of the 2021 IEEE International Conference on Robotics and Automation (ICRA), IEEE, Xi'an, China, 30 May–5 June 2021; pp. 14396–14402.
151. Cimarelli, C.; Bavle, H.; Sanchez-Lopez, J.L.; Voos, H. RAUM-VO: Rotational Adjusted Unsupervised Monocular Visual Odometry. *Sensors* **2022**, *22*, 2651. https://doi.org/10.3390/s22072651.
152. Kneip, L.; Lynen, S. Direct optimization of frame-to-frame rotation. In Proceedings of the IEEE International Conference on Computer Vision, Sydney Australia, 1–8 December 2013; pp. 2352–2359.
153. Leutenegger, S.; Lynen, S.; Bosse, M.; Siegwart, R.; Furgale, P. Keyframe-based visual–inertial odometry using nonlinear optimization. *Int. J. Robot. Res.* **2015**, *34*, 314–334. https://doi.org/10.1177/0278364914554813.
154. Forster, C.; Zhang, Z.; Gassner, M.; Werlberger, M.; Scaramuzza, D. SVO: Semidirect Visual Odometry for Monocular and Multicamera Systems. *IEEE Trans. Robot.* **2017**, *33*, 249–265. https://doi.org/10.1109/TRO.2016.2623335.
155. Qin, T.; Li, P.; Shen, S. VINS-Mono: A Robust and Versatile Monocular Visual-Inertial State Estimator. *IEEE Trans. Robot.* **2018**, *34*, 1004–1020. https://doi.org/10.1109/TRO.2018.2853729.
156. Forster, C.; Carlone, L.; Dellaert, F.; Scaramuzza, D. On-Manifold Preintegration for Real-Time Visual–Inertial Odometry. *IEEE Trans. Robot.* **2017**, *33*, 1–21. https://doi.org/10.1109/TRO.2016.2597321.
157. Von Stumberg, L.; Usenko, V.; Cremers, D. Direct Sparse Visual-Inertial Odometry Using Dynamic Marginalization. In Proceedings of the 2018 IEEE International Conference on Robotics and Automation (ICRA), Brisbane, Australia, 21–26 May 2018. https://doi.org/10.1109/icra.2018.8462905.
158. Usenko, V.; Demmel, N.; Schubert, D.; Stuckler, J.; Cremers, D. Visual-Inertial Mapping With Non-Linear Factor Recovery. *IEEE Robot. Autom. Lett.* **2020**, *5*, 422–429. https://doi.org/10.1109/lra.2019.2961227.
159. Delmerico, J.; Scaramuzza, D. A Benchmark Comparison of Monocular Visual-Inertial Odometry Algorithms for Flying Robots. In Proceedings of the 2018 IEEE International Conference on Robotics and Automation (ICRA), Brisbane, Australia, 21–26 May 2018; pp. 2502–2509. https://doi.org/10.1109/ICRA.2018.8460664.
160. Qin, T.; Shen, S. Online Temporal Calibration for Monocular Visual-Inertial Systems. *arXiv* **2018**, arXiv:1808.00692.
161. Mur-Artal, R.; Tardós, J.D. ORB-SLAM2: an Open-Source SLAM System for Monocular, Stereo and RGB-D Cameras. *IEEE Trans. Robot.* **2017**, *33*, 1255–1262.
162. Campos, C.; Elvira, R.; Rodriguez, J.J.G.; M. Montiel, J.M.; D. Tardos, J. ORB-SLAM3: An Accurate Open-Source Library for Visual, Visual–Inertial, and Multimap SLAM. *IEEE Trans. Robot.* **2021**, *37*, 1874–1890. https://doi.org/10.1109/tro.2021.3075644.
163. Khattak, S.; Papachristos, C.; Alexis, K. Keyframe-based Direct Thermal–Inertial Odometry. In Proceedings of the 2019 International Conference on Robotics and Automation (ICRA), Montreal, QC, Canada, 20–24 May 2019. https://doi.org/10.1109/icra.2019.8793927.
164. Dang, T.; Tranzatto, M.; Khattak, S.; Mascarich, F.; Alexis, K.; Hutter, M. Graph-based subterranean exploration path planning using aerial and legged robots. *J. Field Robot.* **2020**, *37*, 1363–1388. https://doi.org/10.1002/rob.21993.

165. Dang, T.; Mascarich, F.; Khattak, S.; Nguyen, H.; Nguyen, H.; Hirsh, S.; Reinhart, R.; Papachristos, C.; Alexis, K. Autonomous Search for Underground Mine Rescue Using Aerial Robots. In Proceedings of the 2020 IEEE Aerospace Conference, Big Sky, MT, USA, 7–14 March 2020; pp. 1–8. https://doi.org/10.1109/AERO47225.2020.9172804.
166. Zhao, S.; Wang, P.; Zhang, H.; Fang, Z.; Scherer, S. TP-TIO: A Robust Thermal-Inertial Odometry with Deep ThermalPoint. In Proceedings of the 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Las Vegas, NV, USA, 25–29 October 2020; pp. 4505–4512. https://doi.org/10.1109/IROS45743.2020.9341716.
167. Saputra, M.R.U.; Lu, C.X.; de Gusmao, P.P.B.; Wang, B.; Markham, A.; Trigoni, N. Graph-based Thermal-Inertial SLAM with Probabilistic Neural Networks. *arXiv* **2021**, arXiv:2104.07196.
168. Mueggler, E.; Gallego, G.; Rebecq, H.; Scaramuzza, D. Continuous-Time Visual-Inertial Odometry for Event Cameras. *IEEE Trans. Robot.* **2018**, *34*, 1425–1440. https://doi.org/10.1109/TRO.2018.2858287.
169. Rebecq, H.; Horstschaefer, T.; Gallego, G.; Scaramuzza, D. EVO: A Geometric Approach to Event-Based 6-DOF Parallel Tracking and Mapping in Real Time. *IEEE Robot. Autom. Lett.* **2017**, *2*, 593–600. https://doi.org/10.1109/LRA.2016.2645143.
170. Vidal, A.R.; Rebecq, H.; Horstschaefer, T.; Scaramuzza, D. Ultimate SLAM? Combining Events, Images, and IMU for Robust Visual SLAM in HDR and High-Speed Scenarios. *IEEE Robot. Autom. Lett.* **2018**, *3*, 994–1001. https://doi.org/10.1109/LRA.2018.2793357.
171. Hess, W.; Kohler, D.; Rapp, H.; Andor, D. Real-time loop closure in 2D LIDAR SLAM. In Proceedings of the 2016 IEEE International Conference on Robotics and Automation (ICRA), Stockholm, Sweden, 16–21 May 2016; pp. 1271–1278. https://doi.org/10.1109/ICRA.2016.7487258.
172. Kohlbrecher, S.; von Stryk, O.; Meyer, J.; Klingauf, U. A flexible and scalable SLAM system with full 3D motion estimation. In Proceedings of the 2011 IEEE International Symposium on Safety, Security, and Rescue Robotics, Kyoto, Japan, 1–5 November 2011; pp. 155–160. https://doi.org/10.1109/SSRR.2011.6106777.
173. Zhang, J.; Singh, S. LOAM: Lidar Odometry and Mapping in Real-time. In *Robotics: Science and Systems*; University of California: Berkeley, CA, USA, 2014.
174. Wang, H.; Wang, C.; Chen, C.L.; Xie, L. F-LOAM : Fast LiDAR Odometry and Mapping. In Proceedings of the 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, Prague, Czech Republic, 27 September–1 October 2021. https://doi.org/10.1109/iros51168.2021.9636655.
175. Behley, J.; Stachniss, C. Efficient Surfel-Based SLAM using 3D Laser Range Data in Urban Environments. In *Robotics: Science and Systems*; University of California: Berkeley, CA, USA, 2018; Volume 2018, p. 59.
176. Gräter, J.; Wilczynski, A.; Lauer, M. LIMO: Lidar-Monocular Visual Odometry. *arXiv* **2018**, arXiv:1807.07524.
177. Shan, T.; Englot, B.; Ratti, C.; Rus, D. LVI-SAM: Tightly-coupled Lidar-Visual-Inertial Odometry via Smoothing and Mapping. *arXiv* **2021**, arXiv:2104.10831.
178. Nguyen, T.M.; Cao, M.; Yuan, S.; Lyu, Y.; Nguyen, T.H.; Xie, L. LIRO: Tightly Coupled Lidar-Inertia-Ranging Odometry. *arXiv* **2020**, arXiv:2010.13072.
179. Nguyen, T.M.; Yuan, S.; Cao, M.; Nguyen, T.H.; Xie, L. VIRAL SLAM: Tightly Coupled Camera-IMU-UWB-Lidar SLAM. *arXiv* **2021**, arXiv:2105.03296.
180. Koide, K.; Miura, J.; Menegatti, E. A portable three-dimensional LIDAR-based system for long-term and wide-area people behavior measurement. *Int. J. Adv. Robot. Syst.* **2019**, *16*, 1729881419841532. https://doi.org/10.1177/1729881419841532.
181. Shan, T.; Englot, B.; Meyers, D.; Wang, W.; Ratti, C.; Rus, D. LIO-SAM: Tightly-coupled Lidar Inertial Odometry via Smoothing and Mapping. *arXiv* **2020**, arXiv:2007.00258.
182. Chen, X.; Milioto, A.; Palazzolo, E.; Giguère, P.; Behley, J.; Stachniss, C. SuMa++: Efficient LiDAR-based Semantic SLAM. *arXiv* **2021**, arXiv:2105.11320.
183. Smith, M.; Baldwin, I.; Churchill, W.; Paul, R.; Newman, P. The New College Vision and Laser Data Set. *I. J. Robot. Res.* **2009**, *28*, 595–599. https://doi.org/10.1177/0278364909103911.
184. Sturm, J.; Engelhard, N.; Endres, F.; Burgard, W.; Cremers, D. A benchmark for the evaluation of RGB-D SLAM systems. In Proceedings of the 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems, Vilamoura-Algarve, Portugal, 7–12 October 2012; pp. 573–580. https://doi.org/10.1109/IROS.2012.6385773.
185. Engel, J.; Usenko, V.; Cremers, D. A photometrically calibrated benchmark for monocular visual odometry. *arXiv* **2016**, arXiv:1607.02555.
186. Burri, M.; Nikolic, J.; Gohl, P.; Schneider, T.; Rehder, J.; Omari, S.; Achtelik, M.W.; Siegwart, R. The EuRoC micro aerial vehicle datasets. *Int. J. Robot. Res.* **2016**, *35*, 1157–1163. https://doi.org/10.1177/0278364915620033.
187. Geiger, A.; Lenz, P.; Urtasun, R. Are we ready for autonomous driving? The KITTI vision benchmark suite. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012; pp. 3354–3361. https://doi.org/10.1109/CVPR.2012.6248074.
188. Schubert, D.; Goll, T.; Demmel, N.; Usenko, V.; Stueckler, J.; Cremers, D. The TUM VI Benchmark for Evaluating Visual-Inertial Odometry. In Proceedings of the International Conference on Intelligent Robots and Systems (IROS), Madrid, Spain, 1–5 October 2018.
189. Handa, A.; Whelan, T.; McDonald, J.; Davison, A.J. A benchmark for RGB-D visual odometry, 3D reconstruction and SLAM. In Proceedings of the 2014 IEEE International Conference on Robotics and Automation (ICRA), Hong Kong, China, 31 May–7 June 2014; pp. 1524–1531. https://doi.org/10.1109/ICRA.2014.6907054.
190. Gálvez-López, D.; Salas, M.; Tardós, J.D.; Montiel, J.M.M. Real-time Monocular Object SLAM. *arXiv* **2015**, arXiv:1504.02398.

191. Nicholson, L.; Milford, M.; Sünderhauf, N. QuadricSLAM: Constrained Dual Quadrics from Object Detections as Landmarks in Semantic SLAM. *arXiv* **2018**, arXiv:1804.04011.
192. Yang, S.; Scherer, S.A. CubeSLAM: Monocular 3D Object Detection and SLAM without Prior Models. *arXiv* **2018**, arXiv:1806.00557.
193. Zhang, J.; Henein, M.; Mahony, R.; Ila, V. VDO-SLAM: A Visual Dynamic Object-aware SLAM System. *arXiv* **2020**, arXiv: 2005.11052.
194. Judd, K.M.; Gammell, J.D. The Oxford Multimotion Dataset: Multiple SE(3) Motions With Ground Truth. *IEEE Robot. Autom. Lett.* **2019**, *4*, 800–807. https://doi.org/10.1109/lra.2019.2892656.
195. Rosinol, A.; Abate, M.; Chang, Y.; Carlone, L. Kimera: An Open-Source Library for Real-Time Metric-Semantic Localization and Mapping. In Proceedings of the 2020 IEEE International Conference on Robotics and Automation (ICRA), Paris, France, 31 May–31 August 2020; pp. 1689–1696. https://doi.org/10.1109/ICRA40945.2020.9196885.
196. Shan, T.; Englot, B. LeGO-LOAM: Lightweight and Ground-Optimized Lidar Odometry and Mapping on Variable Terrain. In Proceedings of the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Madrid, Spain, 1–5 October 2018; pp. 4758–4765. https://doi.org/10.1109/IROS.2018.8594299.
197. Behley, J.; Garbade, M.; Milioto, A.; Quenzel, J.; Behnke, S.; Stachniss, C.; Gall, J. SemanticKITTI: A Dataset for Semantic Scene Understanding of LiDAR Sequences. *arXiv* **2019**, arXiv:1904.01416.
198. Pandey, G.; McBride, J.R.; Eustice, R.M. Ford Campus vision and lidar data set. *Int. J. Robot. Res.* **2011**, *30*, 1543–1552. https://doi.org/10.1177/0278364911400640.
199. Salas-Moreno, R.F.; Newcombe, R.A.; Strasdat, H.; Kelly, P.H.; Davison, A.J. SLAM++: Simultaneous Localisation and Mapping at the Level of Objects. In Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, 23–28 June 2013; pp. 1352–1359. https://doi.org/10.1109/CVPR.2013.178.
200. Atanasov, N.; Zhu, M.; Daniilidis, K.; Pappas, G.J. Localization from semantic observations via the matrix permanent. *Int. J. Robot. Res.* **2016**, *35*, 73–99. https://doi.org/10.1177/0278364915596589.
201. Bowman, S.L.; Atanasov, N.; Daniilidis, K.; Pappas, G.J. Probabilistic data association for semantic SLAM. In Proceedings of the 2017 IEEE International Conference on Robotics and Automation (ICRA), Singapore, 29 May–3 June 2017; pp. 1722–1729. https://doi.org/10.1109/ ICRA.2017.7989203.
202. Lianos, N.; Schönberger, J.L.; Pollefeys, M.; Sattler, T. VSO: Visual Semantic Odometry. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018.
203. Doherty, K.; Baxter, D.; Schneeweiss, E.; Leonard, J. Probabilistic Data Association via Mixture Models for Robust Semantic SLAM. *arXiv* **2019**, arXiv:1909.11213.
204. Bavle, H.; De La Puente, P.; How, J.P.; Campoy, P. VPS-SLAM: Visual Planar Semantic SLAM for Aerial Robotic Systems. *IEEE Access* **2020**, *8*, 60704–60718. https://doi.org/10.1109/ACCESS.2020.2983121.
205. Sanchez-Lopez, J.L.; Castillo-Lopez, M.; Voos, H. Semantic situation awareness of ellipse shapes via deep learning for multirotor aerial robots with a 2D LIDAR. In Proceedings of the 2020 International Conference on Unmanned Aircraft Systems (ICUAS), Athens, Greece, 1–4 September 2020; pp. 1014–1023. https://doi.org/10.1109/ICUAS48674.2020.9214063.
206. Li, L.; Kong, X.; Zhao, X.; Li, W.; Wen, F.; Zhang, H.; Liu, Y. SA-LOAM: Semantic-aided LiDAR SLAM with Loop Closure. *arXiv* **2021**, arXiv:2106.11516.
207. Bescos, B.; Facil, J.M.; Civera, J.; Neira, J. DynaSLAM: Tracking, Mapping, and Inpainting in Dynamic Scenes. *IEEE Robot. Autom. Lett.* **2018**, *3*, 4076–4083. https://doi.org/10.1109/lra.2018.2860039.
208. Liu, Y.; Miura, J. RDMO-SLAM: Real-time Visual SLAM for Dynamic Environments using Semantic Label Prediction with Optical Flow. *IEEE Access* **2021**, *9*, 106981–106997. https://doi.org/10.1109/ACCESS.2021.3100426.
209. Mao, M.; Zhang, H.; Li, S.; Zhang, B. SEMANTIC-RTAB-MAP (SRM): A semantic SLAM system with CNNs on depth images. *Math. Found. Comput.* **2019**, *2*, 29–41.
210. Lai, L.; Yu, X.; Qian, X.; Ou, L. 3D Semantic Map Construction System Based on Visual SLAM and CNNs. In Proceedings of the IECON 2020 The 46th Annual Conference of the IEEE Industrial Electronics Society, IEEE, Singapore, 18–21 October 2020. https://doi.org/10.1109/iecon43393.2020.9254223.
211. Hempel, T.; Al-Hamadi, A. An online semantic mapping system for extending and enhancing visual SLAM. *Eng. Appl. Artif. Intell.* **2022**, *111*, 104830. https://doi.org/10.1016/j.engappai.2022.104830.
212. McCormac, J.; Handa, A.; Davison, A.; Leutenegger, S. SemanticFusion: Dense 3D semantic mapping with convolutional neural networks. In Proceedings of the 2017 IEEE International Conference on Robotics and Automation (ICRA), Singapore, 29 May–3 June 2017; pp. 4628–4635. https://doi.org/10.1109/ICRA.2017.7989538.
213. Tian, Y.; Chang, Y.; Arias, F.H.; Nieto-Granda, C.; How, J.; Carlone, L. Kimera-Multi: Robust, Distributed, Dense Metric-Semantic SLAM for Multi-Robot Systems. *IEEE Trans. Robot.* **2022**, *38*, 2022–2038. https://doi.org/10.1109/tro.2021.3137751.
214. Wang, Z.; Zhang, Q.; Li, J.; Zhang, S.; Liu, J. A Computationally Efficient Semantic SLAM Solution for Dynamic Scenes. *Remote Sens.* **2019**, *11*, 1363. https://doi.org/10.3390/rs11111363.
215. Liu, G.; Zeng, W.; Feng, B.; Xu, F. DMS-SLAM: A general visual SLAM system for dynamic scenes with multiple sensors. *Sensors* **2019**, *19*, 3714.
216. Li, A.; Wang, J.; Xu, M.; Chen, Z. DP-SLAM: A visual SLAM with moving probability towards dynamic environments. *Inf. Sci.* **2021**, *556*, 128–142.

217. Hornung, A.; Wurm, K.M.; Bennewitz, M.; Stachniss, C.; Burgard, W. OctoMap: An efficient probabilistic 3D mapping framework based on octrees. *Auton. Robot.* **2013**, *34*, 189–206.
218. Oleynikova, H.; Millane, A.; Taylor, Z.; Galceran, E.; Nieto, J.; Siegwart, R. Signed distance fields: A natural representation for both mapping and planning. In Proceedings of the RSS 2016 Workshop: Geometry and Beyond-Representations, Physics, and Scene Understanding for Robotics, University of Michigan, Ann Arbor, MI, USA, 18–22 June 2016.
219. Oleynikova, H.; Taylor, Z.; Siegwart, R.; Nieto, J. Safe local exploration for replanning in cluttered unknown environments for microaerial vehicles. *IEEE Robot. Autom. Lett.* **2018**, *3*, 1474–1481.
220. Chibane, J.; Mir, A.; Pons-Moll, G. Neural Unsigned Distance Fields for Implicit Function Learning. *arXiv* **2020**, arXiv: 2010.13938.
221. Han, L.; Gao, F.; Zhou, B.; Shen, S. FIESTA: Fast Incremental Euclidean Distance Fields for Online Motion Planning of Aerial Robots. In Proceedings of the 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, Venetian Macao, Macau, 3–8 November 2019. https://doi.org/10.1109/iros40897.2019.8968199.
222. Zucker, M.; Ratliff, N.; Dragan, A.D.; Pivtoraiko, M.; Klingensmith, M.; Dellin, C.M.; Bagnell, J.A.; Srinivasa, S.S. Chomp: Covariant hamiltonian optimization for motion planning. *Int. J. Robot. Res.* **2013**, *32*, 1164–1193.
223. Oleynikova, H.; Taylor, Z.; Fehr, M.; Siegwart, R.; Nieto, J. Voxblox: Incremental 3D Euclidean Signed Distance Fields for on-board MAV planning. In Proceedings of the 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, Vancouver, BC, Canada, 24–28 September 2017. https://doi.org/10.1109/iros.2017.8202315.
224. Reijgwart, V.; Millane, A.; Oleynikova, H.; Siegwart, R.; Cadena, C.; Nieto, J. Voxgraph: Globally Consistent, Volumetric Mapping Using Signed Distance Function Submaps. *IEEE Robot. Autom. Lett.* **2020**, *5*, 227–234. https://doi.org/10.1109/lra.2019.2953859.
225. Millane, A.; Oleynikova, H.; Lanegger, C.; Delmerico, J.; Nieto, J.; Siegwart, R.; Pollefeys, M.; Cadena, C. Freetures: Localization in Signed Distance Function Maps. *arXiv* **2020**, arXiv:2010.09378.
226. Grinvald, M.; Furrer, F.; Novkovic, T.; Chung, J.J.; Cadena, C.; Siegwart, R.; Nieto, J.I. Volumetric Instance-Aware Semantic Mapping and 3D Object Discovery. *arXiv* **2019**, arXiv:1903.00268.
227. Pan, Y.; Kompis, Y.; Bartolomei, L.; Mascaro, R.; Stachniss, C.; Chli, M. Voxfield: Non-Projective Signed Distance Fields for Online Planning and 3D Reconstruction. In Proceedings of the 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, Kyoto, Japan, 23–27 October 2022; pp. 5331–5338.
228. Narita, G.; Seno, T.; Ishikawa, T.; Kaji, Y. PanopticFusion: Online Volumetric Semantic Mapping at the Level of Stuff and Things. In Proceedings of the 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Macau, China, 3–8 November 2019; pp. 4205–4212. https://doi.org/10.1109/IROS40897.2019.8967890.
229. Schmid, L.; Delmerico, J.; Schönberger, J.; Nieto, J.; Pollefeys, M.; Siegwart, R.; Cadena, C. Panoptic Multi-TSDFs: A Flexible Representation for Online Multi-resolution Volumetric Mapping and Long-term Dynamic Scene Consistency. In Proceedings of the 2022 International Conference on Robotics and Automation (ICRA), Philadelphia, PA, USA, 23–27 May 2022; pp. 8018–8024. https://doi.org/10.1109/ICRA46639.2022.9811877.
230. Sitzmann, V.; Zollhöfer, M.; Wetzstein, G. Scene Representation Networks: Continuous 3D-Structure-Aware Neural Scene Representations. In Proceedings of the Advances in Neural Information Processing Systems, Vancouver, BC, Canada, 8 December 2019.
231. Sitzmann, V.; Martel, J.N.P.; Bergman, A.W.; Lindell, D.B.; Wetzstein, G. Implicit Neural Representations with Periodic Activation Functions. *arXiv* **2020**, arXiv:2006.09661.
232. Mildenhall, B.; Srinivasan, P.P.; Tancik, M.; Barron, J.T.; Ramamoorthi, R.; Ng, R. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. In Proceedings of the ECCV, Glasgow, UK, 23–28 August 2020.
233. Sucar, E.; Liu, S.; Ortiz, J.; Davison, A.J. iMAP: Implicit Mapping and Positioning in Real-Time. *arXiv* **2021**, arXiv:2103.12352.
234. Zhu, Z.; Peng, S.; Larsson, V.; Xu, W.; Bao, H.; Cui, Z.; Oswald, M.R.; Pollefeys, M. NICE-SLAM: Neural Implicit Scalable Encoding for SLAM. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 18–24 June 2022.
235. Rosinol, A.; Leonard, J.J.; Carlone, L. NeRF-SLAM: Real-Time Dense Monocular SLAM with Neural Radiance Fields. *arXiv* **2022**, arXiv:2210.13641.
236. Zhu, Z.; Peng, S.; Larsson, V.; Cui, Z.; Oswald, M.R.; Geiger, A.; Pollefeys, M. NICER-SLAM: Neural Implicit Scene Encoding for RGB SLAM. *arXiv* **2023**, arXiv:2302.03594.
237. Johari, M.M.; Carta, C.; Fleuret, F. ESLAM: Efficient Dense SLAM System Based on Hybrid Representation of Signed Distance Fields. *arXiv* **2022**, arXiv:2211.11704.
238. Kruzhkov, E.; Savinykh, A.; Karpyshev, P.; Kurenkov, M.; Yudin, E.; Potapov, A.; Tsetserukou, D. MeSLAM: Memory Efficient SLAM based on Neural Fields. In Proceedings of the 2022 IEEE International Conference on Systems, Man, and Cybernetics (SMC), IEEE, Prague, Czech Republic, 9–12 October 2022; pp. 430–435.
239. Whelan, T.; Leutenegger, S.; Salas-Moreno, R.; Glocker, B.; Davison, A. ElasticFusion: Dense SLAM without a pose graph. In *Robotics: Science and Systems*; Sapienza University of Rome: Rome, Italy, 2015.
240. Wang, K.; Gao, F.; Shen, S. Real-time scalable dense surfel mapping. In Proceedings of the 2019 International Conference on Robotics and Automation (ICRA), IEEE, Montreal, QC, Canada, 20–24 May 2019; pp. 6919–6925.
241. Armeni, I.; He, Z.; Gwak, J.; Zamir, A.R.; Fischer, M.; Malik, J.; Savarese, S. 3D Scene Graph: A Structure for Unified Semantics, 3D Space, and Camera. In Proceedings of the the IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Republic of Korea, 27 October–2 November 2019, pp. 5664-5673.

242. Wald, J.; Dhamo, H.; Navab, N.; Tombari, F. Learning 3D Semantic Scene Graphs from 3D Indoor Reconstructions. *arXiv* **2020**, arXiv:2004.03967.
243. Wu, S.C.; Wald, J.; Tateno, K.; Navab, N.; Tombari, F. SceneGraphFusion: Incremental 3D Scene Graph Prediction from RGB-D Sequences. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Nashville, TN, USA, 20–25 June 2021. https://doi.org/10.1109/cvpr46437.2021.00743.
244. Rosinol, A.; Gupta, A.; Abate, M.; Shi, J.; Carlone, L. 3D Dynamic Scene Graphs: Actionable Spatial Perception with Places, Objects, and Humans. *arXiv* **2020**, arXiv:2002.06289.
245. Rosinol, A.; Violette, A.; Abate, M.; Hughes, N.; Chang, Y.; Shi, J.; Gupta, A.; Carlone, L. Kimera: From SLAM to Spatial Perception with 3D Dynamic Scene Graphs. *arXiv* **2021**, arXiv:2101.06894.
246. Rematas, K.; Liu, A.; Srinivasan, P.P.; Barron, J.T.; Tagliasacchi, A.; Funkhouser, T.; Ferrari, V. Urban Radiance Fields. In Proceedings of the CVPR, New Orleans, LA, USA, 18–24 June 2022.
247. Turki, H.; Ramanan, D.; Satyanarayanan, M. Mega-nerf: Scalable construction of large-scale nerfs for virtual fly-throughs. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 19–24 June 2022; pp. 12922–12931.
248. Schöps, T.; Sattler, T.; Pollefeys, M. Surfelmeshing: Online surfel-based mesh reconstruction. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *42*, 2494–2507.
249. Hughes, N.; Chang, Y.; Carlone, L. Hydra: A Real-time Spatial Perception System for 3D Scene Graph Construction and Optimization. *arXiv* **2022**, arXiv:2201.13360.
250. Ravichandran, Z.; Peng, L.; Hughes, N.; Griffith, J.D.; Carlone, L. Hierarchical Representations and Explicit Memory: Learning Effective Navigation Policies on 3D Scene Graphs using Graph Neural Networks. In Proceedings of the 2022 International Conference on Robotics and Automation (ICRA), IEEE, Philadelphia, PA, USA, 23–27 May 2022. https://doi.org/10.1109/icra46639.2022.9812179.
251. Agia, C.; Jatavallabhula, K.; Khodeir, M.; Miksik, O.; Vineet, V.; Mukadam, M.; Paull, L.; Shkurti, F. Taskography: Evaluating robot task planning over large 3D scene graphs. In Proceedings of the Conference on Robot Learning, PMLR, Auckland, New Zealand, 14–18 December 2022; pp. 46–58.
252. Looper, S.; Rodriguez-Puigvert, J.; Siegwart, R.; Cadena, C.; Schmid, L. 3D VSG: Long-term Semantic Scene Change Prediction through 3D Variable Scene Graphs. *arXiv* **2022**, arXiv:2209.07896.
253. Castillo-Lopez, M.; Ludivig, P.; Sajadi-Alamdari, S.A.; Sánchez-López, J.L.; Olivares-Méndez, M.A.; Voos, H. A Real-Time Approach for Chance-Constrained Motion Planning with Dynamic Obstacles. *IEEE Robot. Autom. Lett.* **2020**, *5*, 3620 - 3625. https://doi.org/10.1109/LRA.2020.2975759.
254. Fang, J.; Wang, F.; Shen, P.; Zheng, Z.; Xue, J.; Chua, T.s. Behavioral intention prediction in driving scenes: A survey. *arXiv* **2022**, arXiv:2211.00385.
255. Rasouli, A.; Tsotsos, J.K. Autonomous Vehicles That Interact With Pedestrians: A Survey of Theory and Practice. *IEEE Trans. Intell. Transp. Syst.* **2020**, *21*, 900–918. https://doi.org/10.1109/tits.2019.2901817.
256. Guo, J.; Kurup, U.; Shah, M. Is it Safe to Drive? An Overview of Factors, Metrics, and Datasets for Driveability Assessment in Autonomous Driving. *IEEE Trans. Intell. Transp. Syst.* **2020**, *21*, 3135–3151. https://doi.org/10.1109/tits.2019.2926042.
257. Wang, W.; Wang, L.; Zhang, C.; Liu, C.; Sun, L. Social Interactions for Autonomous Driving: A Review and Perspectives. *Found. Trends® Robot.* **2022**, *10*, 198–376. https://doi.org/10.1561/2300000078.
258. Kwak, J.Y.; Ko, B.C.; Nam, J.Y. Pedestrian intention prediction based on dynamic fuzzy automata for vehicle driving at nighttime. *Infrared Phys. Technol.* **2017**, *81*, 41–51. https://doi.org/10.1016/j.infrared.2016.12.014.
259. Xing, Y.; Lv, C.; Wang, H.; Wang, H.; Ai, Y.; Cao, D.; Velenis, E.; Wang, F.Y. Driver Lane Change Intention Inference for Intelligent Vehicles: Framework, Survey, and Challenges. *IEEE Trans. Veh. Technol.* **2019**, *68*, 4377–4390. https://doi.org/10.1109/tvt.2019.2903299.
260. Fang, Z.; Lopez, A.M. Intention Recognition of Pedestrians and Cyclists by 2D Pose Estimation. *IEEE Trans. Intell. Transp. Syst.* **2020**, *21*, 4773–4783. https://doi.org/10.1109/tits.2019.2946642.
261. Izquierdo, R.; Quintanar, A.; Parra, I.; Fernandez-Llorca, D.; Sotelo, M.A. Experimental validation of lane-change intention prediction methodologies based on CNN and LSTM. In Proceedings of the 2019 IEEE Intelligent Transportation Systems Conference (ITSC), IEEE, Auckland, New Zealand, 27–30 October 2019. https://doi.org/10.1109/itsc.2019.8917331.
262. Rasouli, A.; Yau, T.; Rohani, M.; Luo, J. Multi-Modal Hybrid Architecture for Pedestrian Action Prediction. In Proceedings of the 2022 IEEE Intelligent Vehicles Symposium (IV), IEEE, Aachen, Germany, 5–9 June 2022. https://doi.org/10.1109/iv51971.2022.9827055.
263. Cadena, P.R.G.; Qian, Y.; Wang, C.; Yang, M. Pedestrian Graph +: A Fast Pedestrian Crossing Prediction Model Based on Graph Convolutional Networks. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 21050–21061. https://doi.org/10.1109/tits.2022.3173537.
264. Achaji, L.; Moreau, J.; Fouqueray, T.; Aioun, F.; Charpillet, F. Is attention to bounding boxes all you need for pedestrian action prediction? In Proceedings of the 2022 IEEE Intelligent Vehicles Symposium (IV), IEEE, Aachen, Germany, 4–9 June 2022. https://doi.org/10.1109/ iv51971.2022.9827084.
265. Li, C.; Chan, S.H.; Chen, Y.T. Who Make Drivers Stop? Towards Driver-centric Risk Assessment: Risk Object Identification via Causal Inference. In Proceedings of the 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, Las Vegas, NV, USA, 25–29 October 2020. https://doi.org/10.1109/iros45743.2020.9341072.

266. Dosovitskiy, A.; Ros, G.; Codevilla, F.; Lopez, A.; Koltun, V. CARLA: An open urban driving simulator. In Proceedings of the Conference on Robot Learning, PMLR, Mountain View, CA, USA, 13–15 November 2017; pp. 1–16.
267. Zhou, K.; Liu, Z.; Qiao, Y.; Xiang, T.; Loy, C.C. Domain Generalization: A Survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *45*, 4396–4415. https://doi.org/10.1109/tpami.2022.3195549.
268. Rudenko, A.; Palmieri, L.; Herman, M.; Kitani, K.M.; Gavrila, D.M.; Arras, K.O. Human Motion Trajectory Prediction: A Survey. *Int. J. Robot. Res.* **2020**, *39*, 895–935. https://doi.org/10.1177/0278364920917446.
269. Huang, Y.; Du, J.; Yang, Z.; Zhou, Z.; Zhang, L.; Chen, H. A Survey on Trajectory-Prediction Methods for Autonomous Driving. *IEEE Trans. Intell. Veh.* **2022**, *7*, 652–674. https://doi.org/10.1109/TIV.2022.3167103.
270. Mozaffari, S.; Al-Jarrah, O.Y.; Dianati, M.; Jennings, P.; Mouzakitis, A. Deep Learning-Based Vehicle Behavior Prediction for Autonomous Driving Applications: A Review. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 33–47. https://doi.org/10.1109/TITS.2020.3012034.
271. Ridel, D.; Rehder, E.; Lauer, M.; Stiller, C.; Wolf, D. A Literature Review on the Prediction of Pedestrian Behavior in Urban Scenarios. In Proceedings of the 2018 21st International Conference on Intelligent Transportation Systems (ITSC), Maui, HI, USA, 4–7 November 2018; pp. 3105–3112. https://doi.org/10.1109/ITSC.2018.8569415.
272. Chang, X.; Ren, P.; Xu, P.; Li, Z.; Chen, X.; Hauptmann, A. A Comprehensive Survey of Scene Graphs: Generation and Application. *IEEE Trans. Pattern Anal. Mach. Intell.* **2023**, *45*, 1–26. https://doi.org/10.1109/tpami.2021.3137605.
273. Huang, C.; Mees, O.; Zeng, A.; Burgard, W. Audio Visual Language Maps for Robot Navigation. *arXiv* **2023**, arXiv:2303.07522.
274. Jatavallabhula, K.M.; Kuwajerwala, A.; Gu, Q.; Omama, M.; Chen, T.; Li, S.; Iyer, G.; Saryazdi, S.; Keetha, N.; Tewari, A.; et al. ConceptFusion: Open-set Multimodal 3D Mapping. *arXiv* **2023**, arXiv:2302.07241.
275. Cornejo-Lupa, M.A.; Cardinale, Y.; Ticona-Herrera, R.; Barrios-Aranibar, D.; Andrade, M.; Diaz-Amado, J. OntoSLAM: An Ontology for Representing Location and Simultaneous Mapping Information for Autonomous Robots. *Robotics* **2021**, *10*, 125. https://doi.org/10.3390/robotics10040125.
276. Bavle, H.; Sanchez-Lopez, J.L.; Shaheer, M.; Civera, J.; Voos, H. Situational Graphs for Robot Navigation in Structured Indoor Environments. *IEEE Robot. Autom. Lett.* **2022**, *7*, 9107–9114. https://doi.org/10.1109/LRA.2022.3189785.
277. Bavle, H.; Sanchez-Lopez, J.L.; Shaheer, M.; Civera, J.; Voos, H. S-Graphs+: Real-time Localization and Mapping leveraging Hierarchical Representations. *arXiv* **2023**, arXiv:2212.11770.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.