A Survey of Simultaneous Localization and Mapping with an Envision in 6G Wireless Networks

Baichuan Huang[1,2*], Jun Zhao[2], Sheng Luo[1,3], Jingbin Liu [1]

1. State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China
2. School of Computer Science and Engineering, Nanyang Technological University, Singapore
3. Hi-Target Surveying Instrument Co. Ltd, Guangzhou, China
*   Corresponding Author, huangbaichuan@whu.edu.cn

Abstract: Simultaneous Localization and Mapping (SLAM) achieves the purpose of simultaneous positioning and map construction based on self-perception. The paper makes an overview in SLAM including Lidar SLAM, visual SLAM, and their fusion. For Lidar or visual SLAM, the survey illustrates the basic type and product of sensors, open source system in sort and history, deep learning embedded, the challenge and future. Additionally, visual inertial odometry is supplemented. For Lidar and visual fused SLAM, the paper highlights the multi-sensors calibration, the fusion in hardware, data, task layer. The open question and forward thinking end the paper. The contributions of this paper can be summarized as follows: the paper provides a high quality and full-scale overview in SLAM. It's very friendly for new researchers to hold the development of SLAM and learn it very obviously. Also, the paper can be considered as dictionary for experienced researchers to search and find new interested orientation.

Keywords: Survey, SLAM (Simultaneous Localization and Mapping), Lidar SLAM, Visual SLAM, Lidar and Vision Fused, User guidance.

1. Introduction

SLAM is the abbreviation of Simultaneous Localization and Mapping, which contains two main tasks, localization and mapping. It is a significant open problem in mobile robotics: to move precisely, a mobile robot must have an accurate environment map; however, to build an accurate map, the mobile robot's sensing locations must be known precisely [1]. In this way, simultaneous map building and localization can be seen to present a question of "which came first, the chicken or the egg?" (The map or the motion?)

In 1990, [2] firstly proposed the use of the EKF (Extended Kalman Filter) for incrementally estimating the posterior distribution over robot pose along with the positions of the landmarks. In fact, starting from the unknown location of the unknown environment, the robot locates its own position and attitude through repeated observation of environmental features in the movement process, and then builds an incremental map of the surrounding environment according to its own position, so as to achieve the purpose of simultaneous positioning and map construction. Localization is a very complex and hot point in recent years. The technologies of localization depend on environment and demand for cost, accuracy, frequency and robustness, which can be achieved by GPS (Global Positioning System), IMU (Inertial Measurement Unit), and wireless signal, etc.[3,4]. But GPS can only work well outdoors and IMU system has cumulative error [5]. The technology of wireless, as an active system, can't make a balance

between cost and accuracy. With the fast development, SLAM equipped with Lidar, camera, IMU and other sensors springs up in last years.

Begin with filter-based SLAM, Graph-based SLAM play a dominant role now. The algorithm derives from KF (Kalman Filter), EKF and PF (Particle Filter) to graph-based optimization. And single thread has been replaced by multi-thread. The technology of SLAM also changed from the earliest prototype of military use to later robot applications with the fusion of multi sensors.

The organization of this paper can be summarized as follows: in Section II, Lidar SLAM including Lidar sensors, open source Lidar SLAM system, deep learning in Lidar and challenge as well as future will be illustrated. Section III highlights the visual SLAM including camera sensors, different density of open source visual SLAM system, visual inertial odometry SLAM, deep learning in visual SLAM and future. In Section IV, the fusion of Lidar and vision will be demonstrated. Finally, the paper identifies several directions for future research of SLAM and provides high quality and full-scale user guide for new researchers in SLAM.

Table 1 Comparison of the different methods

Methods	Feature
Lidar SLAM	more stable and robust
Visual SLAM	cheaper
Lidar and Visual SLAM	more powerful

2. Lidar SLAM

In 1991, [1] used multiple servo-mounted sonar sensors and EKF filter to equip robots with SLAM system. Begin with sonar sensors, the birth of Lidar makes SLAM system more reliable and robustness.

2.1 Lidar Sensors

Lidar sensors can be divided into 2D Lidar and 3D Lidar, which are defined by the number of Lidar beams. In terms of production process, Lidar can also be divided into mechanical Lidar, hybrid solid-state Lidar like MEMS (micro-electro-mechanical) and solid-state Lidar. Solid-state Lidar can be produced by the technology of phased array and flash.

Velodyne: In mechanical Lidar, it has VLP-16, HDL-32E and HDL-64E. In hybrid solid-state Lidar, it has Ultra puck auto with 32E.

SLAMTEC: it has low cost Lidar and robot platform such RPLIDAR A1, A2 and R3.

Ouster: it has mechanical Lidar from 16 to 128 channels.

Quanergy: S3 is the first issued solid-state Lidar in the world and M8 is the mechanical Lidar. The S3-QI is the micro solid-state Lidar.

Ibeo: It has Lux 4L and Lux 8L in mechanical Lidar. Cooperated with Valeo, it issued a hybrid solid-state Lidar named Scala.

In the trend, miniaturization and lightweight solid state Lidar will occupied the market and be satisfied with most application. Other Lidar companies include but not limited to **sick, Hokuyo, HESAI, RoboSense, LeddarTech, ISureStar, benewake, Livox, Innovusion, Innoviz, Trimble, Leishen Intelligent System.**

2.2 Lidar SLAM System

Lidar SLAM system is reliable in theory and technology. [6] illustrated the theory in math about how to simultaneous localization and mapping with 2D Lidar based on probabilistic. Furthre, [7] make surveys about 2D Lidar SLAM system.

Table 2 Comparison of the different Lidar SLAM

Methods	Feature
2D Lidar SLAM	easier to positioning
3D Lidar SLAM	powerful to perceive

2.2.1 2D SLAM

Gmapping: it is the most used SLAM package in robots based on RBPF (Rao-Blackwellisation Partical Filter) method. It adds scan-match method to estimate the position [6,8]. It is the improved version with Grid map based on FastSLAM [9,10].

HectorSlam: it combines a 2D SLAM system and 3D navigation with scan-match technology and an inertial sensing system [11].

KartoSLAM: it is a graph-based SLAM system [12].

LagoSLAM: its basic is the graph-based SLAM,

which is the minimization of a nonlinear non-convex cost function [13].

CoreSLAm: it is an algorithm to be understood with minimum loss of performance [14].

Cartographer: it is a SLAM system from Google [15]. It adopted sub-map and loop closure to achieve a better performance in product grade. The algorithm can provide SLAM in 2D and 3D across multiple platforms and sensor configurations.

2.2.2 3D SLAM

Loam: it is a real-time method for state estimation and mapping using a 3D Lidar [16]. It also has back and forth spin version and continuous scanning 2D Lidar version.

Lego-Loam: it takes in point cloud from a Velodyne VLP-16 Lidar (placed horizontal) and optional IMU data as inputs. The system outputs 6D pose estimation in real-time and has global optimization and loop closure [17].

Cartographer: it supports 2D and 3D SLAM [15].

IMLS-SLAM: it presents a new low-drift SLAM algorithm based only on 3D LiDAR data based on a scan-to-model matching framework [18].

2.2.3 Deep Learning with Lidar SLAM

Feature & Detection: PointNetVLAD [19] allows end-to-end training and inference to extract the global descriptor from a given 3D point cloud to solve point cloud based retrieval for place recognition. **VoxelNet** [20] is a generic 3D detection network that unifies feature extraction and bounding box prediction into a single stage, end-to-end trainable deep network. Other work can be seen in **BirdNet** [21]. **LMNet** [22] describes an efficient single-stage deep convolutional neural network to detect objects and outputs an objectness map and the bounding box offset values for each point. **PIXOR** [23] is a proposal-free, single-stage detector that outputs oriented 3D object estimates decoded from pixel-wise neural network predictions. **Yolo3D** [24] builds on the success of the one-shot regression meta-architecture in the 2D perspective image space and extend it to generate oriented 3D object bounding boxes from LiDAR point cloud. **PointCNN** [25]

proposes to learn a X-transformation from the input points. The X-transformation is applied by element-wise product and sum operations of typical convolution operator. **MV3D** [26] is a sensory-fusion framework that takes both Lidar point cloud and RGB images as input and predicts oriented 3D bounding boxes. **PU-GAN** [27] presents a new point cloud upsampling network based on a generative adversarial network (GAN). Other similar work can be seen in this best paper in CVPR2018 but not limited to [28].

Recognition & Segmentation: In fact, the method of segmentation to 3D point cloud can be divided into Edge-based, region growing, model fitting, hybrid method, machine learning application and deep learning [29]. Here the paper focuses on the methods of deep learning. **PointNet** [30] designs a novel type of neural network that directly consumes point clouds, which has the function of classification, segmentation and semantic analysis. **PointNet++** [31] learns hierarchical features with increasing scales of contexts. **VoteNet** [32] constructs a 3D detection pipeline for point cloud as a end-to-end 3D object detection network, which is based on PointNet++. **SegMap** [33] is a map representation solution to the localization and mapping problem based on the extraction of segments in 3D point clouds. SqueezeSeg [34-36] are convolutional neural nets with recurrent CRF (Conditional random fields) for real-time road-object segmentation from 3d Lidar point cloud. **PointSIFT** [37] is a semantic segmentation framework for 3D point clouds. It is based on a simple module which extracts features from neighbor points in eight directions. **PointWise** [38] presents a convolutional neural network for semantic segmentation and object recognition with 3D point clouds. **3P-RNN** [39] is a novel end-to-end approach for unstructured point cloud semantic segmentation along two horizontal directions to exploit the inherent contextual features. Other similar work can be seen but not limited to **SPG** [40] and the review [29]. **SegMatch** [41] is a loop closure method based on the detection and matching of 3D segments. **Kd-Network** [42] is designed for 3D model recognition tasks and works with unstructured point

clouds. **DeepTemporalSeg** [43] propose a deep convolutional neural network (DCNN) for the semantic segmentation of a LiDAR scan with temporally consistency. **LU-Net** [44] achieve the function of semantic segmentation instead of applying some global 3D segmentation method. Other similar work can be seen but not limited to **PointRCNN** [45].

Localization: **L3-Net** [46] is a novel learning-based LiDAR localization system that achieves centimeter-level localization accuracy. **SuMa++** [47] computes semantic segmentation results in point-wise labels for the whole scan, allowing us to build a semantically-enriched map with labeled surfels and improve the projective scan matching via semantic constraints.

2.3 Challenge and Future

2.3.1 Cost and Adaptability

The advantage of Lidar is that it can provide 3D information, and it is not affected by night and light change. In addition, the angle of view is relatively large and can reach 360 degrees. But the technological threshold of Lidar is very high, which lead to long development cycle and unaffordable cost on a large scale. In the future, miniaturization, reasonable cost, solid state, and achieving high reliability and adaptability is the trend.

2.3.2 Low-Texture and Dynamic Environment

Most SLAM system can just work in a fixed environment but things change constantly. Besides, low-Texture environment like long corridor and big pipeline will make trouble for Lidar SLAM. [48] uses IMU to assist 2D SLAM to solve above obstacles. Further, [49] incorporates the time dimension into the mapping process to enable a robot to maintain an accurate map while operating in dynamical environments. How to make Lidar SLAM more robust to low-texture and dynamic environment, and how to keep map updated should be taken into consideration more deeply.

2.3.3 Adversarial Sensor Attack

Deep Neural Network is easily attacked by adversarial samples, which is also proved in

camera-based perception. But in Lidar-based perception, it is highly important but unexplored. By relaying attack, [50] firstly spoofs the Lidar with interference in output data and distance estimation. The novel saturation attack completely incapacitate a Lidar from sensing a certain direction based on Velodyne's VLP-16. [51] explores the possibility of strategically controlling the spoofed attack to fool the machine learning model. The paper regards task as an optimization problem and design modeling methods for the input perturbation function and the objective function., which improves the attack success rates to around 75%. The adversarial sensor attack will spoof the SLAM system based on Lidar point cloud, which is invisible as hardly found and defended. In the case, research on how to prevent the Lidar SLAM system from adversarial sensor attack should be a new topic.

3. Visual SLAM

As the development of CPU and GPU, the capability of graphics processing becomes more and more powerful. Camera sensors getting cheaper, more lightweight and more versatile at the same time. The past decade has seen the rapid development of visual SLAM. Visual SLAM using camera also make the system cheaper and smaller compare with Lidar system. Now, visual SLAM system can run in micro PC and embedded device, even in mobile devices like smart phones [52-56].

Visual SLAM includes collection of sensors' data such as camera or inertial measurement unit , Visual Odometry or Visual Inertial Odometry in front end, Optimization in back end, Loop closure in back end and Mapping [57]. Relocalization is the additional modules for stable and accurate visual SLAM [58].

In process of Visual Odometry, in addition to the method based on features or template matching, or correlation methods to determine the motion of the camera, there is another method relying on the Fourier-Mellin Transform [59]. [60] and [61] give the example in the environment with no distinct visual features when use the ground-facing camera.

3.1 Visual Sensors

The most used sensors that visual SLAM based are cameras. In detail, camera can be divided into monocular camera, stereo camera, RGB-D camera, event camera, etc.

Monocular camera: visual slam based on monocular camera have a scale with real size of track and map. That's say that the real depth can't be got by monocular camera, which called Scale Ambiguity [62]. The SLAM based on Monocular camera has to initialization, and face the problem of drift.

Stereo camera: stereo camera is a combination of two monocular camera but the distance called baseline between the two monocular camera is known. Although the depth can be got based on calibration, correction, matching and calculation, the process will be a waste of lost of resources.

RGB-D camera: RGB-D camera also called depth camera because the camera can output depth in pixel directly. The depth camera can be realized by technology of stereo, structure-light and TOF. The theory of Structure-light is that infrared laser emits some pattern with structure feature to the surface of object. Then the IR camera will collect the change of patter due to the different depth in the surface. TOF will measure the time of laser's flight to calculate the distance.

Event camera: [63] illustrates that instead of capturing images at a fixed rate, event camera measures per-pixel brightness changes asynchronously. Event camera has very high dynamic range (140 dB vs. 60 dB), high temporal resolution (in the order of us), low power consumption, and do not suffer from motion blur. Hence, event cameras can performance better than traditional camera in high speed and high dynamic range. The example of the event camera are Dynamic Vision Sensor [64-67], Dynamic Line Sensor [68], Dynamic and Active-Pixel Vision Sensor [69], and Asynchronous Time-based Image Sensor [70].

Next the product and company of visual sensors will be introduced:

Microsoft: Kinectc v1(structured-light), Kinect v2(TOF), Azure Kinect(with microphone and IMU).

Intel: 200 Series, 300 Series, Module D400 Series, D415(Active IR Stereo, Rolling shutter),

D435(Active IR Stereo, Global Shutter), D435i(D435 with IMU).

Stereolabs ZED: ZED Stereo camera(depth up to 20m).

MYNTAI: D1000 Series(depth camera), D1200(for smart phone), S1030 Series(standard stereo camera).

Occipital Structure: Structure Sensor(Suitable for ipad).

Samsung: Gen2 and Gen3 dynamic vision sensors and event-based vision solution [65].

Other depth camera can be listed as follows but not limited to **Leap Motion, Orbbec Astra, Pico Zense, DUO, Xtion, Camboard, IMI, Humanplus, PERCIPIO.XYZ, PrimeSense**. Other event camera can be listed as follows but not limited to **iniVation, AIT(AIT Austrian Institute of Technology), SiliconEye, Prophesee, CelePixel, Dilusense**.

3.2 Visual SLAM System

The method of utilizing information from image can be classified into direct method and feature based method. Direct method leads to semiDense and dense construction while feature based method cause sparse construction. Next, some visual slam will be introduced (ATAM7 is a visual SLAM toolkit for beginners [58]):

Table 3 Comparison of the different Visual SLAM

Methods	Feature
Sparse Vslam	positioning, faster
Semi-Dense Vslam	Balance
Dense Vslam	Reconstruction, slow

3.2.1 Sparse Visual SLAM

MonoSLAM: it (monocular) is the first real-time mono SLAM system, which is based on EKF [71].

PTAM: it (monocular) is the first SLAM system that parallel tracking and mapping. It firstly adopts Bundle Adjustment to optimize and concept of key frame [54,72]. The later version supports a trivially simple yet effective relocalization method [73].

ORB-SLAM: it (monocular) uses three threads: Tracking, Local Mapping and Loop Closing [52,74].

ORB-SLAM v2 [75] supports monocular, stereo, and RGB-D cameras. **CubemapSLAM** [76] is a SLAM system for monocular fisheye cameras based on ORB-SLAM. **Visual Inertial ORB-SLAM** [77,78] explains the initialization process of IMU and the joint optimization with visual information.

proSLAM: it (stereo) is a lightweight visual SLAM system with easily understanding [79].

ENFT-sfm: it (monocular) is a feature tracking method which can efficiently match feature point correspondences among one or multiple video sequences [80]. The updated version **ENFT-SLAM** can run in large scale.

OpenVSLAm: it (all types of cameras) [81] is based on an indirect SLAM algorithm with sparse features. The excellent point of OpenVSLAM is that the system supports perspective, fisheye, and equirectangular, even the camera models you design.

TagSLAM: it realizes SLAM with AprilTag fiducial markers [82]. Also, it provides a front end to the GTSAM factor graph optimizer, which can design lots of experiments.

Other similar work can be listed as follows but not limited to **UcoSLAM** [83].

3.2.2 SemiDense Visual SLAM

LSD-SLAM: it (monocular) proposes a novel direct tracking method which operates on Lie Algebra and direct method [84]. [85] make it supporting stereo cameras and [86] make it supporting omnidirectional cameras. Other similar work with omnidirectional cameras can be seen in [87].

SVO: it (monocular) is Semi-direct Visual Odoemtry [88]. It uses sparse model-based image alignment to get a fast speed. The update version is extended to multiple cameras, fisheye and catadioptric ones [78]. [78] gives detailed math proof about VIO. **CNN-SVO** [89] is the version of SVO with the depth prediction from a single-image depth prediction network.

DSO: it (monocular) [90,91] is a new work from the author of LSD-SLAM [84]. The work creates a visual odoemtry based on direct method and sparse method without detection and description of feature point.

EVO: it (Event camera) [92] is an event-based

visual odometry algorithm. Our algorithm is unaffected by motion blur and operates very well in challenging, high dynamic range conditions with strong illumination changes. Other semiDense SLAM based on event camera can be seen in [93]. Other VO (visual odometry) system based on event camera can be seen in [94,95].

3.2.3 Dense Visual SLAM

DTAM: it (monocular) can reconstruct 3D model in real time based on minimizing a global spatially regularized energy functional in a novel non-convex optimization framework, which is called direct method [96,97].

MLM SLAM: it (monocular) can reconstruct dense 3D model online without graphics processing unit (GPU) [98]. The key contribution is a multi-resolution depth estimation and spatial smoothing process.

Kinect Fusion: it (RGB-D) is almost the first 3D reconstruction system with depth camera [99,100].

DVO: it (RGB-D) proposes a dense visual SLAM method, an entropy-based similarity measure for keyframe selection and loop closure detection based g2o framework [101-103].

RGBD-SLAM-V2: it (RGB-D) can reconstruct accurate 3D dense model without the help of other sensors [104].

Kintinuous: it (RGB-D) is a visual SLAM system with globally consistent point and mesh reconstructions in real-time [105-107].

RTAB-MAP: it (RGB-D) supports simultaneous localization and mapping but it's hard to be basis to develop upper algorithm [108-110]. The latter version support both visual and Lidar SLAM [111].

Dynamic Fusion: it (RGB-D) presents the first dense SLAM system capable of reconstructing non-rigidly deforming scenes in real-time based Kinect Fusion [112]. **VolumeDeform** [113] also realizes real-time non-rigid reconstruction but not open source. The similar work can be seen in **Fusion4D** [114].

Elastic Fusion: it (RGB-D) is a real-time dense visual SLAM system capable of capturing comprehensive dense globally consistent surfel-based

maps of room scale environments explored using an RGB-D camera [115,116].

InfiniTAM: it (RGB-D) is a real time 3D reconstruction system with CPU in Linux, IOS, Android platform [55,117,118].

Bundle Fusion: it (RGB-D) supports robust tracking with recovery from gross tracking failures and re-estimates the 3D model in real-time to ensure global consistency [119].

KO-Fusion: it (RGB-D) [120] proposes a dense RGB-D SLAM system with kinematic and odometry measurements from a wheeled robot.

SOFT-SLAM: it (stereo) [121] can create dense map with the advantages of large loop closing, which is based on SOFT [122] for pose estimation.

Other works can be listed as follows but not limited to **SLAMRecon**, RKD-SLAM [123] and **RGB-D SLAM** [124]. **Maplab** [125], **PointNVSNet** [126], **MID-Fusion** [127] and **MaskFusion** [128] will introduced in next chapter.

3.2.4 Visual Inertial Odometry SLAM

The determination of visual slam is technically challenging. Monocular visual SLAM has problems such as necessary initialization, scale ambiguity and scale drift [129]. Although stereo camera and RGB-D camera can solve the problems of initialization and scale, some obstacles can't be ignored such as fast movement (solved with Global Shuttle or fisheye even panoramic camera), small field of view, large calculation, occlusion, feature loss, dynamic scenes and changing light. Recently, VIO (visual inertial odometry SLAM) becomes the popular research.

First of all, [130-132] start some try in VIO. [77,78] give the samples and math proof in visual-inertial odeometry. [133] use several rounds of visual-inertial bundle adjustment to make a robust initialization for VIO. Specially, tango [134], Dyson 360 Eye and hololens [135] are the real products of VIO and receive good feedback. In addition to this, ARkit (filter-based) from Apple, ARcore (filter-based) from Google, Inside-out from uSens are the technology of VIO. PennCOSYVIO [136] synchronizes data from a VI-sensor (stereo camera and IMU), two Project Tango hand-held devices, and three GoPro Hero 4 cameras and calibrates

intrinsically and extrinsically. Next some open source VIO system will be introduced [137]:

SSF: it (loosely-coupled, filter-based) is a time delay compensated single and multi sensor fusion framework based on an EKF [138].

MSCKF: it (tightly-coupled, filter-based) is adopted by Google Tango based on extended Kalman filter [139]. But the similar work called **MSCKF-VIO** [140] open the source.

ROVIO: it (tightly-coupled, filter-based) is an extended Kalman Filter with tracking of both 3D landmarks and image patch features [141]. It supports monocular camera.

OKVIS: it (tightly-coupled, optimization-based) is an open and classic Keyframe-based Visual-Inertial SLAM [130]. It supports monocular and stereo camera based sliding window estimator.

VINS: **VINS-Mono** (tightly-coupled, optimization-based) [53,142,143] is a real-time SLAM framework for Monocular Visual-Inertial Systems. The open source code runs on Linux, and is fully integrated with ROS. **VINS-Mobile** [144,145] is a real-time monocular visual-inertial odometry running on compatible iOS devices. Furthermore, **VINS-Fusion** supports multiple visual-inertial sensor types (GPS, mono camera + IMU, stereo cameras + IMU, even stereo cameras only). It has online spatial calibration, online temporal calibration and visual loop closure.

ICE-BA: it (tightly-coupled, optimization-based) presents an incremental, consistent and efficient bundle adjustment for visual-inertial SLAM, which performs in parallel both local BA over the sliding window and global BA over all keyframes, and outputs camera pose and updated map points for each frame in real-time [146].

Maplab: it (tightly-coupled, optimization-based) is an open, research-oriented visual-inertial mapping framework, written in C++, for creating, processing and manipulating multi-session maps. On the one hand, maplab can be considered as a ready-to-use visual-inertial mapping and localization system. On the other hand, maplab provides the research community with a collection of multi-session mapping tools that include map merging,

visual-inertial batch optimization, loop closure, 3D dense reconstruction [125].

Other solutions can be listed as follows but not limited to **VI-ORB** (tightly-coupled, optimization-based) [77] (the works by the author of ORB-SLAM, but not open source), StructVIO [147]. RKSLAM [148] can reliably handle fast motion and strong rotation for AR applications. Other VIO system based on event camera can be listed as follows but not limited to [149-151]. **mi-VINS** [152] uses multiple IMU, which can work if IMU sensor failures.

In VIO, visual images can enhance the inertial navigation algorithm. To deal with the correlation between the generated visual odometry and also about the multiframe visual odometry, [153] integrates the features tracked from all overlapping image frames by a sequential de-correlation the Kalman filter measurement update with fever computation resources consumption. The proposed method is referred as multi-frame visual odometry (MFVO) [154]. In the image-aided inertial integrated navigation, the relative positions of visual odometry are pairwise correlated in terms of time. The shaping filter proposed [155] uses Cholesky factors based on that the measurement noise is only correlated with the ones from the previous epoch.

VIO SLAM based on deep learning can be seen in [156]. It shows a network that performs visual-inertial odometry (VIO) without inertial measurement unit (IMU) intrinsic parameters or the extrinsic calibration between an IMU and camera. [157] provides a network to avoid the calibration between camera and IMU.

3.2.5 Deep Learning with Visual SLAM

Nowadays, deep learning plays a critical role in the maintenance of computer vision. As the development of visual SLAM, more and more focus are paid into deep learning with SLAM. The term "semantic SLAM" refers to an approach that includes the semantic information into the SLAM process to enhance the performance and representation by providing high-level understanding, robust performance, resource awareness, and task driven perception. Next, we will introduce the implement of

SLAM with semantic information in these aspects:

Feature & Detection: Pop-up SLAM (Monocular) [158] proposes real-time monocular plane SLAM to demonstrate that scene understanding could improve both state estimation and dense mapping especially in low-texture environments. The plane measurements come from a pop-up 3D plane model applied to each single image. [159] gets semantic key points predicted by a convolutional network (convnet). **LIFT** [160] can get more dense feature points than SIFT. **DeepSLAM** [161] has a significant performance gap in the presence of image noise when catch the feature points. **SuperPoint** [162] presents a self-supervised framework for training interest point detectors and descriptors suitable for a large number of multiple-view geometry problems in computer vision. [163] proposes to use the easy-to-labeled 2D detection and discrete viewpoint classification together with a light-weight semantic inference method to obtain rough 3D object measurements. **GCN-SLAM** [164] presents a deep learning-based network, GCNv2, for generation of key points and descriptors. [165] fuses information about 3D shape, location, and, if available, semantic class. **SalientDSO** [166] can realize visual saliency and environment perception with the aid of deep learning. [167] integrates the detected objects as the quadrics models into the SLAM system. **CubeSLAM** (Monocular) is a 3D Object Detection and SLAM system [168] based on cube model. It achieve object-level mapping, positioning, and dynamic object tracking. [169] combines the cubeSLAM (high-level object) and Pop-up SLAM (plane landmarks) to make map more denser, more compact and semantic meaningful compared to feature point based SLAM. **MonoGRNet** [170] is a geometric reasoning network for monocular 3D object detection and localization. Feature based on event camera can be seen but not limited to [171,172]. About the survey in deep learning for detection, [173] could be a good choice.

Recognition & Segmentation: **SLAM++** (CAD model) [174] presents the major advantages of a new 'object oriented' 3D SLAM paradigm, which takes full advantage in the loop of prior knowledge that

many scenes consist of repeated, domain-specific objects and structures. [175] combines the state-of-art deep learning method and LSD-SLAM based on video stream from a monocular camera. 2D semantic information are transferred to 3D mapping via correspondence between connective keyframes with spatial consistency. **Semanticfusion** (RGBD) [176] combines CNN (Convolutional Neural Network) and a state-of-the-art dense Simultaneous Localization and Mapping (SLAM) system, ElasticFusion [116] to build a semantic 3D map. [177] leverages sparse, feature-based RGB-D SLAM, image-based deep-learning object detection and 3D unsupervised segmentation. **MarrNet** [178] proposes an end-to-end trainable framework, sequentially estimating 2.5D sketches and 3D object shapes. **3DMV** (RGB-D) [179] jointly combines RGB color and geometric information to perform 3D semantic segmentation of RGB-D scans. **Pix3D** [180] study 3D shape modeling from a single image. **ScanComplete** [181] is a data-driven approach which takes an incomplete 3D scan of a scene as input and predicts a complete 3D model, along with per-voxel semantic labels. **Fusion++** [182] is an online object-level SLAM system which builds a persistent and accurate 3D graph map of arbitrary reconstructed objects. As an RGB-D camera browses a cluttered indoor scene, **Mask-RCNN** instance segmentations are used to initialise compact per-object Truncated Signed Distance Function (TSDF) reconstructions with object size dependent resolutions and a novel 3D foreground mask. **SegMap** [183] is a map representation based on 3D segments allowing for robot localization, environment reconstruction, and semantics extraction. **3D-SIS** [184] is a novel neural network architecture for 3D semantic instance segmentation in commodity RGB-D scans. **DA-RNN** [185] uses a new recurrent neural network architecture for semantic labeling on RGB-D videos. **DenseFusion** [186] is a generic framework for estimating 6D pose of a set of known objects from RGB-D images. Other work can be seen in CCNet [187]. To recognize based on event camera, [188-191] are the best paper to be investigated.

Recovery Scale: **CNN-SLAM** (Monocular) [192] estimates the depth with deep learning. Another work can be seen in **DeepVO** [193], **GS3D** [194] . **UnDeepVO** [195] can get the 6-DoF pose and the depth using a monocular camera with deep learning. Google proposes the work [196] that present a method for predicting dense depth in scenarios where both a monocular camera and people in the scene are freely moving based on unsupervised learning. Other methods to get real scale in Monocular can be seen in [197,198]. **GeoNet** [199] is a jointly unsupervised learning framework for monocular depth, optical flow and ego-motion estimation from videos. **CodeSLAM** [200] proposes a depth map from single image, which can be optimised efficiently jointly with pose variables. **Mono-stixels** [201] uses the depth, motion and semantic information in dynamic scene to estimate depth. **GANVO** [202] uses an unsupervised learning framework for 6-DoF pose and monocular depth map from unlabelled image, using deep convolutional Generative Adversarial Networks. **GEN-SLAM** [203] outputs the dense map with the aid of conventional geometric SLAM and the topological constraint in monocular. [204] proposes a training objective that is invariant to changes in depth range and scale. Other similar work can be seen in **DeepMVS** [205] and **DeepV2D** [206]. Based on event camera, depth estimation can be applied in monocular camera [207,208] and stereo camera [209].

Pose Output & Optimization: [210] is a stereo-VO under the synchronicity. [211] utilizes a CNN to estimate motion from optical flow. **PoseNet** [212] can get the 6-DOF pose from a single RGB image without the help of optimization. **VInet** (Monocular) [213] firstly estimates the motion in VIO, reducing the dependence of manual synchronization and calibration. **DeepVO** (Monocular) [214] presents a novel end-to-end framework for monocular VO by using deep Recurrent Convolutional Neural Networks (RCNNs). The similar work can be seen in **SFMlearner** [215] and **SFM-Net** [216]. **VSO** [217] proposes a novel visual semantic odometry (VSO) framework to enable medium-term continuous tracking of points using semantics. **MID-Fusion** (RGBD, dense point

cloud) [127] estimates the pose of each existing moving object using an object-oriented tracking method and associate segmented masks with existing models and incrementally fuse corresponding color, depth, semantic, and foreground object probabilities into each object model. Other similar works can be seen in **VidLoc** [218]. Besides, [219,220] are using event camera to output the ego-motion.

Long-term Localization: [221] formulates an optimization problem over sensor states and semantic landmark positions that integrates metric information, semantic information, and data associations. [222] proposes a novel unsupervised deep neural network architecture of a feature embedding for visual loop closure. [223] shows the semantic information is more effective than the traditional feature descriptors. **X-View** [224] leverages semantic graph descriptor matching for global localization, enabling localization under drastically different view-points. [225] proposes a solution that represents hypotheses as multiple modes of an equivalent non-Gaussian sensor model to determine object class labels and measurement-landmark correspondences. About the application based on event camera, [226] are worthy to be read.

Dynamic SLAM: RDSLAM [227] is a novel real-time monocular SLAM system which can robustly work in dynamic environments based on a novel online keyframe representation and updating method. **DS-SLAM** [228] is a SLAM system with semantic information based on optimized ORB-SLAM. The semantic information can make SLAM system more robust in dynamic environment. **MaskFusion** (RGB-D, dense point cloud) is a real-time, object-aware, semantic and dynamic RGB-D SLAM system [128] based on Mask R-CNN [229]. The system can label the objects with semantic information even in continuously and independent motion. The related work can be seen in **Co-Fusion** [230]. **Detect-SLAM** [231] integrates SLAM with a deep neural network based object detector to make the two functions mutually beneficial in an unknown and dynamic environment. **DynaSLAM** [232] is a SLAM system for monocular, stereo and RGB-D camera in dynamic environments with aid of static

map. **StaticFusion** [233] proposes a method for robust dense RGB-D SLAM in dynamic environments which detects moving objects and simultaneously reconstructs the background structure. The related work based on dynamic environment can be also seen in **RGB-D SLAM** [124] and [234-236].

Recently, some works utilizes deep-learning to dominate the whole process of SLAM. **SimVODIS** [237] can output the depth and the relative pose between frames, while detecting objects and segmenting the object boundaries.

3.3 Challenge and Future

3.3.1 Robustness and Portability

Visual SLAM still face some important obstacles like the illumination condition, high dynamic environment, fast motion, vigorous rotation and low texture environment. Firstly, global shutter instead of rolling shutter is fundamental to achieve accurate camera pose estimation. Event camera such as dynamic vision sensors is capable of producing up to one million events per second which is enough for very fast motions in high speed and high dynamic range. Secondly, using semantic features like edge, plane, surface features, even reducing feature dependencies, such as tracking with join edges, direct tracking, or a combination of machine learning may become the better choice. Thirdly, based mathematical machinery for SfM/SLAM, the precise mathematical formulations to outperform implicitly learned navigation functions over data is preferred.

The future of SLAM has can be expected that one is SLAM based on smart phones or embedded platforms such as UAV (unmanned aerial vehicle) and another is detailed 3D reconstruction, scene understanding with deep learning. How to balance real-time and accuracy is the vital open question. The solutions pertaining to dynamic, unstructured, complex, uncertain and large-scale environments are yet to be explored [238].

3.3.2 Multiple Sensors Fusion

The actual robots and hardware devices usually do not carry only one kind of sensor, and often a fusion of multiple sensors. For example, the current

research on VIO on mobile phones combines visual information and IMU information to realize the complementary advantages of the two sensors, which provides a very effective solution for the miniaturization and low cost of SLAM. **DeLS-3D** [239] design is a sensor fusion scheme which integrates camera videos, motion sensors (GPS/IMU), and a 3D semantic map in order to achieve robustness and efficiency of the system. There are sensors listed as follows but not limited to Lidar, Sonar, IMU, IR, camera, GPS, radar, etc. The choice of sensors is dependent on the environment and required type of map.

3.3.3 Semantics SLAM

In fact, humans recognize the movement of objects based on perception not the features in image. Deep learning in SLAM can realize object recognition and segmentation, which help the SLAM system perceive the surrounding better. Semantics SLAM can also do a favor in global optimization, loop closure and relocalization. [240]: Traditional approaches for simultaneous localization and mapping (SLAM) depend on geometric features such as points, lines (**PL-SLAM** [241], **StructSLAM** [242]), and planes to infer the environment structure. The aim of high-precision real-time positioning in large-scale scenarios could be achieved by semantics SLAM, which teaches robots perceive as humans.

3.3.4 Software & hardware

SLAM is not an algorithm but an integrated, complex technology [243]. It not only depend on software, but also hardware. The future SLAM system will focus in the deep combination of algorithm and sensors. Based on illustration above, the domain specific processors rather than general processor, integrated sensors module rather than separate sensor like just camera will show great potential. The above work make the developer focus on the algorithm and accelerate the release of real products.

4. Lidar and Visual SLAM System

4.1 Multiple Sensors Calibration

Camera & IMU: **Kalibr** [244] is a toolbox that solves the following calibration problems: Multiple camera calibration, Visual-inertial calibration (camera-IMU) and Rolling Shutter Camera calibration. **Vins-Fusion** [143] has online spatial calibration and online temporal calibration. **MSCKF-VIO** [140] also has the calibration for camera and IMU. **mc-VINS** [245] can calibrate the extrinsic parameters and time offset between all multiple cameras and IMU. Besides, **IMU-TK** [246][247] can calibrate internal parameter of IMU. Other work can be seen in [248]. [249] proposes a end to end network for monocular VIO, which fuses data from camera and IMU.

Camera & Depth: **BAD SLAM** [250] proposes a calibrated benchmark for this task that uses synchronized global shutter RGB and depth cameras.

Camera & Camera: **mcptam** [251] is a SLAM system using multi-camera. It can also calibrate the intrinsic and extrinsic parameters. **MultiCol-SLAM** [252] is a multi-fisheye camera SLAM. Besides, the updated version of SVO can also support multiple cameras. Other similar work can be seen in **ROVIO** [253].

Lidar & IMU: **LIO-mapping** [254] introduces a tightly coupled lidar-IMU fusion method. **Lidar-Align** is a simple method for finding the extrinsic calibration between a 3D Lidar and a 6-Dof pose sensor. Extrinsic calibration of Lidar can be seen in [255][256]. The doctoral thesis [257] illustrate the work of Lidar calibration.

Camera & Lidar: [258] introduces a probabilistic monitoring algorithm and a continuous calibration optimizer that enable camera-laser calibration online, automatically. **Lidar-Camera** [259] proposes a novel pipeline and experimental setup to find accurate rigid-body transformation for extrinsically calibrating a LiDAR and a camera using 3D-3D point correspondences. **RegNet** [260] is the first deep convolutional neural network (CNN) to infer a 6 degrees of freedom (DOF) extrinsic calibration between multi-modal sensors, exemplified using a scanning LiDAR and a monocular camera. **LIMO** [261] proposes a depth extraction algorithm from LIDAR measurements for camera feature tracks

and estimating motion. **CalibNet** [262] is a self-supervised deep network capable of automatically estimating the 6-DoF rigid body transformation between a 3D LiDAR and a 2D camera in real-time. The calibration tool from **Autoware** can calibrate the signal beam Lidar and camera. . Other work can be seen as follows but not limited to [263-265].

Other work like **SVIn2** [266] demonstrates an underwater SLAM system fusing Sonar, Visual, Inertial, and Depth Sensor, which is based on OKVIS. [267] proposes a new underwater camera-IMU calibration model and [268] detects underwater obstacle using semantic image Segmentation. **WiFi-SLAM** [269] demonstrates a novel SLAM technology with wireless signal named WiFi. [270] uses the mmWave to locate even the NLOS robots and [271,272] introduce more technique about localization with the aid of wirless signals. **KO-Fusion** [120] fuses visual and wheeled odometer. [273] uses a thermal camera with IMU in visually degraded environments e.g. darkness.

4.2 Lidar and Visual Fusion

Hardware layer: **Pandora** from HESAI is a software and hardware solution integrating 40 beams Lidar, five color cameras and recognition algorithm. The integrated solution can comfort developer from temporal and spatial synchronization. Understanding the exist of **CONTOUR** and **STENCIL** from **KAARTA** will give you a brainstorming.

Data layer: Lidar has sparse, high precision depth data and camera has dense but low precision depth data, which will lead to image-based depth upsampling and image-based depth inpainting/completion. [274] presents a novel method for the challenging problem of depth image upsampling. [275] relies only on basic image processing operations to perform depth completion of sparse Lidar depth data. With deep learning, [276] proposes the use of a single deep regression network to learn directly from the RGB-D raw data, and explore the impact of number of depth samples. [277] considers CNN operating on sparse inputs with an application to depth completion from sparse laser scan data. **DFuseNet** [278] proposes a CNN that is

designed to upsample a series of sparse range measurements based on the contextual cues gleaned from a high resolution intensity image. Other similar work can be seen as follows but not limited to [279][280]. **LIC-Fusion** [281] fuses IMU measurements, sparse visual features, and extracted LiDAR points.

Task layer: [282] fuses stereo camera and Lidar to perceive. [283] fuses radar, Lidar, and camera to detect and classify moving objects. Other traditional work can be seen but not limited to [284-286]. [287] can augment VO by depth information such as provided by RGB-D cameras, or from Lidars associated with cameras even if sparsely available. **V-Loam** [288] presents a general framework for combining visual odometry and Lidar odometry. The online method starts with visual odometry and scan matching based Lidar odometry refines the motion estimation and point cloud registration simultaneously. **VL-SLAM** [289] is concerned with the development of a system that combines an accurate laser odometry estimator, with algorithms for place recognition using vision for achieving loop detection. [290] aims at the tracking part of SLAM using an RGB-D camera and 2d low-cost LIDAR to finish a robust indoor SLAM by a mode switch and data fusion. **VIL-SLAM** [291] incorporates tightly-coupled stereo VIO with Lidar mapping and Lidar enhanced visual loop closure. [292] combines monocular camera images with laser distance measurements to allow visual SLAM without errors from increasing scale uncertainty. In deep learning, many methods to detect and recognize fusing data from camera and Lidar such as **PointFusion** [293], **RoarNet** [294], **AVOD** [295], **MV3D** [26], **FuseNet** [296]. Other similar work can be seen in [297]. Besides, [298] exploits both Lidar as well as cameras to perform very accurate localization with a an end-to-end learnable architecture. [299] fuses 3D Lidar and monocular camera.

4.3 Challenge and Future[300]

Data Association: the future of SLAM must integrate multi-sensors. But different sensors have different data types, time stamps, and coordinate system expressions, needed to be processed

uniformly. Besides, physical model establishment, state estimation and optimization between multi-sensors should be taken into consideration.

Integrated Hardware: at present, there is no suitable chip and integrated hardware to make technology of SLAM more easily to be a product. On the other hand, if the accuracy of a sensor degrades due to malfunctioning, off-nominal conditions, or aging, the quality of the sensor measurements (e.g., noise, bias) does not match the noise model. The robustness and integration of hardware should be followed. Sensors in front-end should have the capability to process data and the evolution from hardware layer to algorithm layer, then to function layer to SDK should be innovated to application.

Crowdsourcing: decentralized visual SLAM is a powerful tool for multi-robot applications in environments where absolute positioning systems are not available [301]. Co-optimization visual multi-robot SLAM need decentralized data and optimization, which is called crowdsourcing. The privacy in the process of decentralized data should come into attention. The technology of differential privacy [302][303] maybe do a favor.

High Definition Map: High Definition Map is vital for robots. But which type of map is the best for robots? Could dense map or sparse map navigate, positioning and path plan? A related open question for long-term mapping is how often to update the information contained in the map and how to decide when this information becomes outdated and can be discarded.

Adaptability, Robustness, Scalability: as we know, no SLAM system now can cover all scenarios. Most of it requires extensive parameter tuning in order to work correctly for a given scenario. To make robots perceive as humans, appearance-based instead of feature-based method is preferred, which will help close loops integrated with semantic information between day and night sequences or between different seasons.

Ability against risk and constraints: Perfect SLAM system should be failure-safe and failure-aware. It's not the question about relocalization or loop closure here. SLAM system must have ability to response to risk or failure. In the same time, an ideal SLAM solution should be able run on different platforms no matter the computational constraints of platforms. How to balance the accuracy, robustness and the limited resource is a challenging problem [137].

Application: the technology of SLAM has a wide application such as: large-scale positioning, navigation and 3D or semantic map construction, environment recognition and understanding, ground robotics, UAV, VR/AR/MR, AGV(Automatic Guided Vehicle), automatic drive, virtual interior decorator, virtual fitting room, immersive online game, earthquake relief, video segmentation and editing.

Open question: Will end-to-end learning dominate SLAM?

5. An Envision in 6G Wireless

Nowadays, 5G has been developed widely to communicate more quickly and massively [272]. But for robots and autonomous driving cars, the technology of SLAM need greater data rates and less latency that 5G can't afford. Unlike 100 Gbps of data rates for 5G, 6G can provide greater data rates due to the frequency in 100 GHz to 3 THz (terahertz).

THz is the last unexplored band in the radio frequency spectrum. Less than the THz, the radio bands are called microwave. The radio frequency of optical bands , which are regarded as visible light communications (VLC), are more than THz [304]. The 6G technology will need no supports such as multiple-input multiple-output (MIMO) in 5G represented as mmWave communications. As for the difference with VLC, 6G with the THz communications will not affected by the light changes and NLOS.

5.1 Lidar and Visual Fusion

For the advance of wireless communication system, industry and academic are urged to pay attention to the research of 6G. Next, we will introduce some advantages of 6G [305].

Low Latency: Less than 1 msec end-to-end latency;

Data rate: High data rates up to 1 Tbps;

Ultra-high bandwidth: Very broad frequency bands;

Energy save: Very high energy efficiency;

Ubiquitous connection: Enable to connect global network including the massive intelligent things and the emergence of smart surface and environment such as walls, roads even the whole buildings.

Intelligent network: AI and RISs make it smarter and beyond classical big data analytics and edge computation.

Ubiquitous connection: Enable to connect everything.

6G is supposed as the platforms to serve for communication, computation, and storage resources with the aid of AI [306].

5.2 6G in Simultaneous Localization and Mapping

SLAM can be divided into radio-based (such as satellite positioning, cellular and WiFi) and sensor-based (such as Lidar, IMU and camera) [307]. With the wireless technology, the technology of SLAM can be achieved by constructing the map of environment and projecting the angle and the time of arrival to estimate the locations from the users. The conventional mmWave method utilizing AoA-based positioning or a combination of path loss and AoA, and RSSI [271]. With the aid of Reconfigurable intelligent surfaces (RISs), localization and mapping can improve accuracy and extended physical coverage.

It's apparent that 6G with THz will create centimeter level accuracy even in NLOS environment and 6G will provide a network to sense and localization rather than a independent source [308] for SLAM. Cause the greater data rates, the massive computations can be conducted in remote device or machine, which relief the pressure of computer power in robots and autonomous driving cars.

In the future, THz will enable the robots and autonomous driving cars with new capability of sensing the gas, air quality, health detection, body scanning and so on. Plus, with the aid of THz, computer vision will be augmented to see the NLOS views, which will play a vital role in rescue and sensing. In the basic, the accuracy of positioning will be improved to sub-centimeter level and the map of surrounding environment will be constructed as 3D maps without any calibration and prior knowledge, which is hard to realized before. Moreover, 6G and AI will achieve excellent successes again in future digital society with the full connectivity demands.

References

[1] John J Leonard and Hugh F Durrant-Whyte. Simultaneous map build- ing and localization for an autonomous mobile robot. In Proceedings IROS'91: IEEE/RSJ International Workshop on Intelligent Robots and Systems' 91, pages 1442–1447. Ieee, 1991.

[2] Randall Smith, Matthew Self, and Peter Cheeseman. Estimating un- certain spatial relationships in robotics. In Autonomous robot vehicles, pages 167–193. Springer, 1990.

[3] Baichuan Huang, Jingbin Liu, Wei Sun, and Fan Yang. A robust indoor positioning method based on bluetooth low energy with separate channel information. Sensors, 19(16):3487, 2019.

[4] Jingbin Liu, Ruizhi Chen, Yuwei Chen, Ling Pei, and Liang Chen. iparking: An intelligent indoor location-based smartphone parking service. Sensors, 12(11):14612–14629, 2012.

[5] Jingbin Liu, Ruizhi Chen, Ling Pei, Robert Guinness, and Heidi Kuusniemi. A hybrid smartphone indoor positioning solution for mobile lbs. Sensors, 12(12):17208–17233, 2012.

[6] Sebastian Thrun, Wolfram Burgard, and Dieter Fox. Probabilistic robotics. MIT press, 2005.

[7] Joao Machado Santos, David Portugal, and Rui P Rocha. An evalu- ation of 2d slam techniques available in robot operating system. In 2013 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR), pages 1–6. IEEE, 2013.

[8] Giorgio Grisetti, Cyrill Stachniss, Wolfram Burgard, et al. Improved techniques for grid mapping with rao-blackwellized particle filters. IEEE transactions on Robotics, 23(1):34, 2007.

[9] Michael Montemerlo, Sebastian Thrun, Daphne Koller, Ben Wegbreit, et al. Fastslam: A factored solution to the simultaneous localization and mapping problem. *Aaai/iaai*, 593598, 2002.

[10] Michael Montemerlo, Sebastian Thrun, Daphne Koller, Ben Wegbreit, et al. Fastslam 2.0: An improved particle filtering algorithm for simultaneous localization and mapping that provably converges. In *IJCAI*, pages 1151–1156, 2003.

[11] Stefan Kohlbrecher, Oskar Von Stryk, Johannes Meyer, and Uwe Klingauf. A flexible and scalable slam system with full 3d motion estimation. In *2011 IEEE International Symposium on Safety, Security, and Rescue Robotics*, pages 155–160. IEEE, 2011.

[12] Kurt Konolige, Giorgio Grisetti, Rainer Ku ̈mmerle, Wolfram Burgard, Benson Limketkai, and Regis Vincent. Efficient sparse pose adjustment for 2d mapping. In *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 22–29. IEEE, 2010.

[13] Luca Carlone, Rosario Aragues, Jose ́ A Castellanos, and Basilio Bona. A linear approximation for graph-based simultaneous localization and mapping. *Robotics: Science and Systems VII*, pages 41–48, 2012.

[14]BSteuxandOTinySLAMElHamzaoui.Aslamalgor ithminlessthan 200 lines c-language program. *Proceedings of the Control Automation Robotics & Vision (ICARCV), Singapore*, pages 7–10, 2010.

[15] Wolfgang Hess, Damon Kohler, Holger Rapp, and Daniel Andor. Real- time loop closure in 2d lidar slam. In *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1271–1278. IEEE, 2016.

[16] Ji Zhang and Sanjiv Singh. Loam: Lidar odometry and mapping in real-time. In *Robotics: Science and Systems*, volume 2, page 9, 2014.

[17] Tixiao Shan and Brendan Englot. Lego-loam: Lightweight and ground- optimized lidar odometry and mapping on variable terrain. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 4758–4765. IEEE, 2018.

[18] Jean-Emmanuel Deschaud. Imls-slam: scan-to-model matching based on 3d data. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2480–2485. IEEE, 2018.

[19] Mikaela Angelina Uy and Gim Hee Lee. Pointnetvlad: Deep point cloud based retrieval for large-scale place recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4470–4479, 2018.

[20] Yin Zhou and Oncel Tuzel. Voxelnet: End-to-end learning for point cloud based 3d object detection. In *Proceedings of the IEEE Confer- ence on Computer Vision and Pattern Recognition*, pages 4490–4499, 2018.

[21] Jorge Beltra ́n, Carlos Guindel, Francisco Miguel Moreno, Daniel Cruzado, Fernando Garcia, and Arturo De La Escalera. Birdnet: a 3d object detection framework from lidar information. In *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, pages 3517–3523. IEEE, 2018.

[22]KazukiMinemura,HengfuiLiau,AbrahamMonrro y,andShinpeiKato. Lmnet: Real-time multiclass object detection on cpu using 3d lidar. In *2018 3rd Asia-Pacific Conference on Intelligent Robot Systems (ACIRS)*, pages 28–34. IEEE, 2018.

[23] Bin Yang, Wenjie Luo, and Raquel Urtasun. Pixor: Real-time 3d object detection from point clouds. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 7652–7660, 2018.

[24] Waleed Ali, Sherif Abdelkarim, Mahmoud Zidan, Mohamed Zahran, and Ahmad El Sallab. Yolo3d: End-to-end real-time 3d oriented object bounding box detection from lidar point cloud. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 0–0, 2018.

[25]YangyanLi,RuiBu,MingchaoSun,WeiWu,Xinhan Di,andBaoquan Chen. Pointcnn: Convolution on x-transformed points. In *Advances in Neural Information Processing Systems*, pages 820–830, 2018.

[26] Xiaozhi Chen, Huimin Ma, Ji Wan, Bo Li, and Tian Xia. Multi-view 3d object detection network for autonomous driving. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1907–1915, 2017.

[27] Ruihui Li, Xianzhi Li, Chi-Wing Fu, Daniel Cohen-Or, and Pheng- Ann Heng. Pu-gan: A point cloud upsampling adversarial network. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 7203–7212, 2019.

[28] Hang Su, Varun Jampani, Deqing Sun, Subhransu Maji, Evangelos Kalogerakis, Ming-Hsuan Yang, and Jan Kautz. Splatnet: Sparse lattice networks for point cloud processing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2530– 2539, 2018.

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

[30] Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. *arXiv preprint arXiv:1612.00593*, 2016.

[31] Charles R Qi, Li Yi, Hao Su, and Leonidas J Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *arXiv preprint arXiv:1706.02413*, 2017.

[32] Charles R Qi, Or Litany, Kaiming He, and Leonidas J Guibas. Deep hough voting for 3d object detection in point clouds. *arXiv preprint arXiv:1904.09664*, 2019.

[33] Renaud Dube, Andrei Cramariuc, Daniel Dugas, Juan Nieto, Roland Siegwart, and Cesar Cadena. SegMap: 3d segment mapping using data-driven descriptors. In *Robotics: Science and Systems (RSS)*, 2018.

[34] Bichen Wu, Alvin Wan, Xiangyu Yue, and Kurt Keutzer. Squeezeseg: Convolutional neural nets with recurrent crf for real-time road-object segmentation from 3d lidar point cloud. *ICRA*, 2018.

[35] Bichen Wu, Xuanyu Zhou, Sicheng Zhao, Xiangyu Yue, and Kurt Keutzer. Squeezesegv2: Improved model structure and unsupervised domain adaptation for road-object segmentation from a lidar point cloud. In *ICRA*, 2019.

[36] XiangyuYue,BichenWu,SanjitASeshia,KurtKe utzer,andAlbertoL Sangiovanni-Vincentelli. A lidar point cloud generator: from a virtual world to autonomous driving. In *ICMR*, pages 458–464. ACM, 2018.

[37] Mingyang Jiang, Yiran Wu, Tianqi Zhao, Zelin Zhao, and Cewu Lu. Pointsift: A sift-like network module for 3d point cloud semantic segmentation. *arXiv preprint arXiv:1807.00652*, 2018.

[38] Binh-Son Hua, Minh-Khoi Tran, and Sai-Kit Yeung. Pointwise convo- lutional neural networks. In *Computer Vision and Pattern Recognition (CVPR)*, 2018.

[39] Xiaoqing Ye, Jiamao Li, Hexiao Huang, Liang Du, and Xiaolin Zhang. 3d recurrent neural networks with context fusion for point cloud semantic segmentation. In *Proceedings of the European*

Conference on Computer Vision (ECCV), pages 403–417, 2018.

[40] Loic Landrieu and Martin Simonovsky. Large-scale point cloud semantic segmentation with superpoint graphs. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4558–4567, 2018.

[41] Renaud Dubé, Daniel Dugas, Elena Stumm, Juan Nieto, Roland Sieg- wart, and Cesar Cadena. Segmatch: Segment based place recognition in 3d point clouds. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pages 5266–5272. IEEE, 2017.

[42] Roman Klokov and Victor Lempitsky. Escape from cells: Deep kd- networks for the recognition of 3d point cloud models. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 863– 872, 2017.

[43] Ayush Dewan and Wolfram Burgard. Deeptemporalseg: Temporally consistent semantic segmentation of 3d lidar scans. *arXiv preprint arXiv:1906.06962*, 2019.

[44] Pierre Biasutti, Vincent Lepetit, Jean-Franois Aujol, Mathieu Brdif, and Aurlie Bugeau. Lu-net: An efficient network for 3d lidar point cloud semantic segmentation based on end-to-end-learned 3d features and u-net. 08 2019.

[45] Shaoshuai Shi, Xiaogang Wang, and Hongsheng Li. Pointrcnn: 3d ob- ject proposal generation and detection from point cloud. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–779, 2019.

[46] Lu Weixin, Zhou Yao, Wan Guowei, Hou Shenhua, and Song Shiyu. L3-net: Towards learning based lidar localization for autonomous driv- ing. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.

[47] Chen Xieyuanli, Milioto Andres, and Emanuelea Palazzolo. Suma++: Efficient lidar-based semantic slam. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2019.

[48] Zhongli Wang, Yan Chen, Yue Mei, Kuo Yang, and Baigen Cai. Imu- assisted 2d slam method for low-texture and dynamic environments. *Applied Sciences*, 8(12):2534, 2018.

[49] Aisha Walcott-Bryant, Michael Kaess, Hordur Johannsson, and John J Leonard. Dynamic pose graph slam: Long-term mapping in low dynamic environments. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1871–1878. IEEE, 2012.

[50] Hocheol Shin, Dohyun Kim, Yujin Kwon, and Yongdae Kim. Illusion and dazzle: Adversarial optical channel exploits against lidars for automotive applications. In *International Conference on Cryptographic Hardware and Embedded Systems*, pages 445–467. Springer, 2017.

[51] Yulong Cao, Chaowei Xiao, Benjamin Cyr, Yimeng Zhou, Won Park, Sara Rampazzi, Qi Alfred Chen, Kevin Fu, and Z Morley Mao. Adversarial sensor attack on lidar-based perception in autonomous driving. *arXiv preprint arXiv:1907.06826*, 2019.

[52] Raul Mur-Artal, Jose Maria Martinez Montiel, and Juan D Tardos. Orb-slam: a versatile and accurate monocular slam system. *IEEE transactions on robotics*, 31(5):1147–1163, 2015.

[53] Tong Qin, Peiliang Li, and Shaojie Shen. Vins-mono: A robust and versatile monocular visual-inertial state estimator. *IEEE Transactions on Robotics*, 34(4):1004–1020, 2018.

[54] Georg Klein and David Murray. Parallel tracking and mapping on a camera phone. In *2009 8th IEEE International Symposium on Mixed and Augmented Reality*, pages 83–86. IEEE, 2009.

[55] O. Ka ̈hler, V. A. Prisacariu, C. Y. Ren, X. Sun, P. H. S Torr, and D. W. Murray. Very High Frame Rate Volumetric Integration of Depth Images on Mobile Device. *IEEE Transactions on Visualization and Computer Graphics (Proceedings International Symposium on Mixed and Augmented Reality 2015*, 22(11), 2015.

[56] Simon Lynen, Torsten Sattler, Michael Bosse, Joel A Hesch, Marc Pollefeys, and Roland Siegwart. Get out of my lab: Large-scale, real- time visual-inertial localization. In *Robotics: Science and Systems*, volume 1, 2015.

[57] Xiang Gao, Tao Zhang, Yi Liu, and Qinrui Yan. *14 Lectures on Visual SLAM: From Theory to Practice*. Publishing House of Electronics Industry, 2017.

[58] Takafumi Taketomi, Hideaki Uchiyama, and Sei Ikeda. Visual slam algorithms: A survey from 2010 to 2016. *IPSJ Transactions on Computer Vision and Applications*, 9(1):16, 2017.

[59] B Srinivasa Reddy and Biswanath N Chatterji. An fft-based technique for translation, rotation, and scale-invariant image registration. *IEEE transactions on image processing*, 5(8):1266–1271, 1996.

[60] Tim Kazik and Ali Haydar Go ̈ktog ̆an. Visual odometry based on the fourier-mellin transform for a rover using a monocular ground-facing camera. In *2011 IEEE International Conference on Mechatronics*, pages 469–474. IEEE, 2011.

[61] Merwan Birem, Richard Kleihorst, and Norddin El-Ghouti. Visual odometry based on the fourier transform using a monocular ground- facing camera. *Journal of Real-Time Image Processing*, 14(3):637–646, 2018.

[62] Liu Haomin, Zhang Guofeng, and Bao hujun. A survy of monocular simultaneous localization and mapping. *Journal of Computer-Aided Design & Computer Graphics*, 28(6):855–868, 2016.

[63]GuillermoGallego,TobiDelbruck,GarrickOrchard ,ChiaraBartolozzi, and Davide Scaramuzza. Event-based vision: A survey. 2019.

[64] Patrick Lichtsteiner, Christoph Posch, and Tobi Delbruck. A 128x128 120db 15us latency asynchronous temporal contrast vision sensor. *IEEE journal of solid-state circuits*, 43(2):566–576, 2008.

[65] Bongki Son, Yunjae Suh, Sungho Kim, Heejae Jung, Jun-Seok Kim, Changwoo Shin, Keunju Park, Kyoobin Lee, Jinman Park, Jooyeon Woo, et al. 4.1 a 640× 480 dynamic vision sensor with a 9μm pixel and 300meps address-event representation. In *2017 IEEE International Solid-State Circuits Conference (ISSCC)*, pages 66–67. IEEE, 2017.

[66] Christoph Posch, Daniel Matolin, Rainer Wohlgenannt, Thomas Maier, and Martin Litzenberger. A microbolometer asynchronous dynamic vision sensor for lwir. *IEEE Sensors Journal*, 9(6):654–664, 2009.

[67] Michael Hofstatter, Peter Scho ̈n, and Christoph Posch. A sparc- compatible general purpose address-event processor with 20-bit l0ns- resolution asynchronous sensor data interface in 0.18 μm cmos. In *Proceedings of 2010 IEEE International Symposium on Circuits and Systems*, pages 4229–4232. IEEE, 2010.

[68] Christoph Posch, Michael Hofstatter, Daniel Matolin, Guy Vanstraelen, Peter Schon, Nikolaus Donath, and Martin Litzenberger. A dual-line optical transient sensor with on-chip precision time-stamp generation. In *2007 IEEE International Solid-State Circuits Conference. Digest of Technical Papers*, pages 500–618. IEEE, 2007.

[69] Christian Brandli, Raphael Berner, Minhao Yang, Shih-Chii Liu, and Tobi Delbruck. A 240× 180 130 db 3 μs latency global shutter spatiotemporal vision sensor. *IEEE Journal of Solid-State Circuits*, 49(10):2333–2341, 2014.

[70] Christoph Posch, Daniel Matolin, and Rainer Wohlgenannt. A qvga 143 db dynamic range frame-free pwm image sensor with lossless pixel-level video compression and time-domain cds. *IEEE Journal of Solid-State Circuits*, 46(1):259–275, 2010.

[71] Andrew J Davison, Ian D Reid, Nicholas D Molton, and Olivier Stasse. Monoslam: Real-time single camera slam. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (6):1052–1067, 2007.

[72] Georg Klein and David Murray. Parallel tracking and mapping for small ar workspaces. In *Proceedings of the 2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality*, pages 1– 10. IEEE Computer Society, 2007.

[73] Georg Klein and David Murray. Improving the agility of keyframe- based slam. In *European Conference on Computer Vision*, pages 802– 815. Springer, 2008.

[74] Ethan Rublee, Vincent Rabaud, Kurt Konolige, and Gary R Bradski. Orb: An efficient alternative to sift or surf. In *ICCV*, volume 11, page 2. Citeseer, 2011.

[75] Raul Mur-Artal and Juan D Tardós. Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras. *IEEE Transactions on Robotics*, 33(5):1255–1262, 2017.

[76] Yahui Wang, Shaojun Cai, Shi-Jie Li, Yun Liu, Yangyan Guo, Tao Li, and Ming-Ming Cheng. Cubemapslam: A piecewise-pinhole monocular fisheye slam system. In *Asian Conference on Computer Vision*, pages 34–49. Springer, 2018.

[77] Raúl Mur-Artal and Juan D Tardós. Visual-inertial monocular slam with map reuse. *IEEE Robotics and Automation Letters*, 2(2):796– 803, 2017.

[78] Christian Forster, Luca Carlone, Frank Dellaert, and Davide Scara- muzza. On-manifold preintegration for real-time visual–inertial odom- etry. *IEEE Transactions on Robotics*, 33(1):1–21, 2016.

[79] D.Schlegel,M.Colosi,andG.Grisetti.ProSLAM: GraphSLAMfrom a Programmer's Perspective. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1–9, 2018.

[80] Guofeng Zhang, Haomin Liu, Zilong Dong, Jiaya Jia, Tien-Tsin Wong, and Hujun Bao. Efficient non-consecutive feature tracking for robust structure-from-motion. *IEEE Transactions on Image Processing*, 25(12):5957–5970, 2016.

[81] Shinya Sumikura, Mikiya Shibuya, and Ken Sakurada. Openvslam: a versatile visual slam framework, 2019.

[82] Bernd Pfrommer and Kostas Daniilidis. Tagslam: Robust slam with fiducial markers. *arXiv preprint arXiv:1910.00679*, 2019.

[83] Rafael Munoz-Salinas and Rafael Medina-Carnicer. Ucoslam: Simul- taneous localization and mapping by fusion of keypoints and squared planar markers. *arXiv preprint arXiv:1902.03729*, 2019.

[84] Jakob Engel, Thomas Schöps, and Daniel Cremers. Lsd-slam: Large- scale direct monocular slam. In *European conference on computer vision*, pages 834–849. Springer, 2014.

[85] Jakob Engel, Jörg Stückler, and Daniel Cremers. Large-scale direct slam with stereo cameras. In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 1935–1942. IEEE, 2015.

[86] David Caruso, Jakob Engel, and Daniel Cremers. Large-scale direct slam for omnidirectional cameras. In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 141–148. IEEE, 2015.

[87] Jianfeng Li, Xiaowei Wang, and Shigang Li. Spherical-model-based slam on full-view images for indoor environments. *Applied Sciences*, 8(11):2268, 2018.

[88] Christian Forster, Zichao Zhang, Michael Gassner, Manuel Werl- berger, and Davide Scaramuzza. Svo: Semidirect visual odometry for monocular and multicamera systems. *IEEE Transactions on Robotics*, 33(2):249–265, 2016.

[89] ShingYanLoo,AliJahaniAmiri,SyamsiahMashohor,SaiHongTang, and Hong Zhang. Cnn-svo: Improving the mapping in semi-direct visual odometry using single-image depth prediction. *arXiv preprint arXiv:1810.01011*, 2018.

[90] Jakob Engel, Vladlen Koltun, and Daniel Cremers. Direct sparse odometry. *CoRR*, abs/1607.02565, 2016.

[91] Jakob Engel, Vladlen Koltun, and Daniel Cremers. Direct sparse odom- etry. *IEEE transactions on pattern analysis and machine intelligence*, 40(3):611–625, 2017.

[92] Henri Rebecq, Timo Horstschäfer, Guillermo Gallego, and Davide Scaramuzza. Evo: A geometric approach to event-based 6-dof parallel tracking and mapping in real time. *IEEE Robotics and Automation Letters*, 2(2):593–600, 2016.

[93] Yi Zhou, Guillermo Gallego, Henri Rebecq, Laurent Kneip, Hongdong Li, and Davide Scaramuzza. Semi-dense 3d reconstruction with a stereo event camera. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 235–251, 2018.

[94] David Weikersdorfer, Raoul Hoffmann, and Jörg Conradt. Simulta- neous localization and mapping for event-based vision systems. In *International Conference on Computer Vision Systems*, pages 133–142. Springer, 2013.

[95] David Weikersdorfer, David B Adrian, Daniel Cremers, and Jörg Conradt. Event-based 3d slam with a depth-augmented dynamic vision sensor. In *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pages 359–364. IEEE, 2014.

[96] Javier Civera, Andrew J Davison, and JM Martinez Montiel. Inverse depth parametrization for monocular slam. *IEEE transactions on robotics*, 24(5):932–945, 2008.

[97] Richard A Newcombe, Steven J Lovegrove, and Andrew J Davison. Dtam: Dense tracking and mapping in real-time. In *2011 international conference on computer vision*, pages 2320–2327. IEEE, 2011.

[98] W Nicholas Greene, Kyel Ok, Peter Lommel, and Nicholas Roy. Multi-level mapping: Real-time dense monocular slam. In *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pages 833–840. IEEE, 2016.

[99] Richard A Newcombe, Shahram Izadi, Otmar Hilliges, David Molyneaux, David Kim, Andrew J Davison, Pushmeet Kohli, Jamie Shotton, Steve Hodges, and Andrew W Fitzgibbon. Kinectfusion: Real- time dense surface mapping and tracking. In *ISMAR*, volume 11, pages 127–136, 2011.

[100]ShahramIzadi,DavidKim,OtmarHilliges,David Molyneaux,Richard Newcombe, Pushmeet Kohli, Jamie Shotton, Steve Hodges, Dustin Freeman, Andrew Davison, et al. Kinectfusion: real-time 3d recon- struction and interaction using a moving depth camera. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*, pages 559–568. ACM, 2011.

[101] Frank Steinbrücker, Jürgen Sturm, and Daniel Cremers. Real-time visual odometry from dense rgb-d images. In *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, pages 719–722. IEEE, 2011.

[102] Christian Kerl, Ju¨rgen Sturm, and Daniel Cremers. Robust odometry estimation for rgb-d cameras. In *2013 IEEE International Conference on Robotics and Automation*, pages 3748–3754. IEEE, 2013.

[103] Christian Kerl, Ju¨rgen Sturm, and Daniel Cremers. Dense visual slam for rgb-d cameras. In *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 2100–2106. IEEE, 2013.

[104] Felix Endres, Ju¨rgen Hess, Ju¨rgen Sturm, Daniel Cremers, and Wolfram Burgard. 3-d mapping with an rgb-d camera. *IEEE transactions on robotics*, 30(1):177–187, 2013.

[105] Thomas Whelan, Michael Kaess, Maurice Fallon, Hordur Johannsson, John J Leonard, and John McDonald. Kintinuous: Spatially extended kinectfusion. 2012.

[106] Thomas Whelan, Michael Kaess, Hordur Johannsson, Maurice Fallon, John J Leonard, and John McDonald. Real-time large-scale dense rgb- d slam with volumetric fusion. *The International Journal of Robotics Research*, 34(4-5):598–626, 2015.

[107] Thomas Whelan, Hordur Johannsson, Michael Kaess, John J. Leonard, and John Mcdonald. Robust real-time visual odometry for dense rgb-d mapping. In *IEEE International Conference on Robotics and Automation*, 2011.

[108] Mathieu Labbe and Franc¸ois Michaud. Online global loop closure detection for large-scale multi-session graph-based slam. In *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 2661–2666. IEEE, 2014.

[109] MM Labbe´ and F Michaud. Appearance-based loop closure detection in real-time for large-scale and long-term operation. *IEEE Transactions on Robotics*, pages 734–745.

[110] Mathieu Labbe´ and Franc¸ois Michaud. Memory management for real- time appearance-based loop closure detection. In *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1271–1276. IEEE, 2011.

[111] Mathieu Labbe´ and Franc¸ois Michaud. Rtab-map as an open-source lidar and visual simultaneous localization and mapping library for large-scale and long-term online operation. *Journal of Field Robotics*, 36(2):416–446, 2019.

[112] Richard A Newcombe, Dieter Fox, and Steven M Seitz. Dynamicfu- sion: Reconstruction and tracking of non-rigid scenes in real-time. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 343–352, 2015.

[113] Matthias Innmann, Michael Zollho¨fer, Matthias Nießner, Christian Theobalt, and Marc Stamminger. Volumedeform: Real-time volumetric non-rigid reconstruction. In *European Conference on Computer Vision*, pages 362–379. Springer, 2016.

[114] Mingsong Dou, Sameh Khamis, Yury Degtyarev, Philip Davidson, Sean Ryan Fanello, Adarsh Kowdle, Sergio Orts Escolano, Christoph Rhemann, David Kim, Jonathan Taylor, et al. Fusion4d: Real-time performance capture of challenging scenes. *ACM Transactions on Graphics (TOG)*, 35(4):114, 2016.

[115] Thomas Whelan, Stefan Leutenegger, R Salas-Moreno, Ben Glocker, and Andrew Davison. Elasticfusion: Dense slam without a pose graph. Robotics: Science and Systems, 2015.

[116] Thomas Whelan, Renato F Salas-Moreno, Ben Glocker, Andrew J Davison, and Stefan Leutenegger. Elasticfusion: Real-time dense slam and light source estimation. *The International Journal of Robotics Research*, 35(14):1697–1716, 2016.

[117] V A Prisacariu, O Ka¨hler, S Golodetz, M Sapienza, T Cavallari, P H S Torr, and D W Murray. InfiniTAM v3: A Framework for Large-Scale 3D

Reconstruction with Loop Closure. *arXiv pre-print arXiv:1708.00783v1*, 2017.

[118] Olaf Ka̍hler, Victor Adrian Prisacariu, and David W. Murray. Real-time large-scale dense 3d reconstruction with loop closure. In *Computer Vision - ECCV 2016 - 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part VIII*, pages 500– 516, 2016.

[119] Angela Dai, Matthias Nießner, Michael Zollo̍fer, Shahram Izadi, and Christian Theobalt. Bundlefusion: Real-time globally consistent 3d re-construction using on-the-fly surface re-integration. *ACM Transactions on Graphics 2017 (TOG)*, 2017.

[120] Charlie Houseago, Michael Bloesch, and Stefan Leutenegger. Ko- fusion: Dense visual slam with tightly-coupled kinematic and odo- metric tracking. In *2019 International Conference on Robotics and Automation (ICRA)*, pages 4054–4060. IEEE, 2019.

[121] Igor Cvis̆ic, Josip Cesic, Ivan Markovic, and Ivan Petrovic. Soft- slam: Computationally efficient stereo visual slam for autonomous uavs. *Journal of field robotics*, 2017.

[122] Igor Cvis̆ić and Ivan Petrović. Stereo odometry based on careful feature selection and tracking. In *2015 European Conference on Mobile Robots (ECMR)*, pages 1–6. IEEE, 2015.

[123] Haomin Liu, Chen Li, Guojun Chen, Guofeng Zhang, Michael Kaess, and Hujun Bao. Robust keyframe-based dense slam with an rgb-d camera. *arXiv preprint arXiv:1711.05166*, 2017.

[124] Weichen Dai, Yu Zhang, Ping Li, and Zheng Fang. Rgb-d slam in dynamic environments using points correlations. *arXiv preprint arXiv:1811.03217*, 2018.

[125] T. Schneider, M. T. Dymczyk, M. Fehr, K. Egger, S. Lynen, I. Gilitschenski, and R. Siegwart. maplab: An open framework for research in visual-inertial mapping and localization. *IEEE Robotics and Automation Letters*, 2018.

[126] Jing Xu Hao Su Rui Chen, Songfang Han. Point-based multi-view stereo network. *arXiv preprint arXiv:1908.04422*, 2019.

[127] Binbin Xu, Wenbin Li, Dimos Tzoumanikas, Michael Bloesch, Andrew Davison, and Stefan Leutenegger. Mid-fusion: Octree-based object- level multi-instance dynamic slam. *arXiv preprint arXiv:1812.07976*, 2018.

[128] M. Runz, M. Buffier, and L. Agapito. Maskfusion: Real-time recog- nition, tracking and reconstruction of multiple moving objects. In *2018 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 10–20, Oct 2018.

[129] Hauke Strasdat, J Montiel, and Andrew J Davison. Scale drift-aware large scale monocular slam. *Robotics: Science and Systems VI*, 2(3):7, 2010.

[130] Stefan Leutenegger, Simon Lynen, Michael Bosse, Roland Siegwart, and Paul Furgale. Keyframe-based visual–inertial odometry using non-linear optimization. *The International Journal of Robotics Research*, 34(3):314–334, 2015.

[131] Guoquan Huang, Michael Kaess, and John J Leonard. Towards consis- tent visual-inertial navigation. In *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pages 4926–4933. IEEE, 2014.

[132] Mingyang Li and Anastasios I Mourikis. High-precision, consistent ekf-based visual-inertial odometry. *The International Journal of Robotics Research*, 32(6):690–711, 2013.

[133] Carlos Campos, J. M. M. Montiel, and Juan D. Tardós. Fast and robust initialization for visual-inertial slam. *2019 International Conference on Robotics and Automation (ICRA)*, pages 1288–1294, 2019.

[134] Mark Froehlich, Salman Azhar, and Matthew Vanture. An investigation of google tango R tablet for low cost 3d scanning. In *ISARC. Pro- ceedings of the International Symposium on Automation and Robotics in Construction*, volume 34. Vilnius Gediminas Technical University, Department of Construction Economics , 2017.

[135] Mathieu Garon, Pierre-Olivier Boulet, Jean-Philippe Doironz, Luc Beaulieu, and Jean-Franc̦ois Lalonde. Real-time high resolution 3d data on the hololens. In *2016 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct)*, pages 189–191. IEEE, 2016.

[136] Bernd Pfrommer, Nitin Sanket, Kostas Daniilidis, and Jonas Cleveland. Penncosyvio: A challenging visual inertial odometry benchmark. In 2017 IEEE International Conference on Robotics and Automation (ICRA), pages 3847–3854. IEEE, 2017.

[137] Jeffrey Delmerico and Davide Scaramuzza. A benchmark comparison of monocular visual-inertial odometry algorithms for flying robots. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2502–2509. IEEE, 2018.

[138] Stephan M Weiss. *Vision based navigation for micro helicopters*. PhD thesis, ETH Zurich, 2012.

[139] Anastasios I Mourikis and Stergios I Roumeliotis. A multi-state con- straint kalman filter for vision-aided inertial navigation. In *Proceedings 2007 IEEE International Conference on Robotics and Automation*, pages 3565–3572. IEEE, 2007.

[140] Ke Sun, Kartik Mohta, Bernd Pfrommer, Michael Watterson, Sikang Liu, Yash Mulgaonkar, Camillo J Taylor, and Vijay Kumar. Robust stereo visual inertial odometry for fast autonomous flight. *IEEE Robotics and Automation Letters*, 3(2):965–972, 2018.

[141] Michael Bloesch, Sammy Omari, Marco Hutter, and Roland Siegwart. Robust visual inertial odometry using a direct ekf-based approach. In *2015 IEEE/RSJ international conference on intelligent robots and systems (IROS)*, pages 298–304. IEEE, 2015.

[142] Peiliang Li, Tong Qin, Botao Hu, Fengyuan Zhu, and Shaojie Shen. Monocular visual-inertial state estimation for mobile augmented reality. In *2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 11–21. IEEE, 2017.

[143] Tong Qin and Shaojie Shen. Online temporal calibration for monocular visual-inertial systems. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3662–3669. IEEE, 2018.

[144] Tong Qin and Shaojie Shen. Robust initialization of monocular visual- inertial estimation on aerial robots. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 4225– 4232. IEEE, 2017.

[145] Zhenfei Yang and Shaojie Shen. Monocular visual–inertial state esti- mation with online initialization and camera–imu extrinsic calibration. *IEEE Transactions on Automation Science and Engineering*, 14(1):39– 51, 2016.

[146] Haomin Liu, Mingyu Chen, Guofeng Zhang, Hujun Bao, and Yingze Bao. Ice-ba: Incremental, consistent and efficient bundle adjustment for visual-inertial slam. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1974–1982, 2018.

[147] Danping Zou, Yuanxin Wu, Ling Pei, Haibin Ling, and Wenxian Yu. Structvio: Visual-inertial odometry with structural regularity of man- made environments. *IEEE Transactions on Robotics*, 2019.

[148] Haomin Liu, Guofeng Zhang, and Hujun Bao. Robust keyframe-based monocular slam for augmented reality. In *2016 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 1–10. IEEE, 2016.

[149] Elias Mueggler, Guillermo Gallego, Henri Rebecq, and Davide Scara- muzza. Continuous-time visual-inertial odometry for event cameras. *IEEE Transactions on Robotics*, 34(6):1425–1440, 2018.

[150] Alex Zihao Zhu, Nikolay Atanasov, and Kostas Daniilidis. Event-based visual inertial odometry. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5816–5824. IEEE, 2017.

[151] Kaleb J Nelson. Event-based visual-inertial odometry on a fixed-wing unmanned aerial vehicle. Technical report, AIR FORCE INSTITUTE OF TECHNOLOGY WRIGHT-PATTERSON AFB OH WRIGHT- PATTERSON, 2019.

[152] Kevin Eckenhoff, Patrick Geneva, and Guoquan Huang. Sensor- failure-resilient multi-imu visual-inertial navigation. *2019 International Conference on Robotics and Automation (ICRA)*, pages 3542–3548, 2019.

[153] E Jared Shamwell, Kyle Lindgren, Sarah Leung, and William D Nothwang. Unsupervised deep visual-inertial odometry with online error correction for rgb-d imagery. *IEEE transactions on pattern analysis and machine intelligence*, 2019.

[154] Hongyun Lee, Matthew McCrink, and James W Gregory. Visual- inertial odometry for unmanned aerial vehicle using deep learning. In *AIAA Scitech 2019 Forum*, page 1410, 2019.

[155] Shichao Yang, Yu Song, Michael Kaess, and Sebastian Scherer. Pop- up slam: Semantic monocular plane slam for low-texture environments. In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 1222–1229. IEEE, 2016.

[156] Georgios Pavlakos, Xiaowei Zhou, Aaron Chan, Konstantinos G Derpa- nis, and Kostas Daniilidis. 6-dof object pose from semantic keypoints. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2011–2018. IEEE, 2017.

[157] Kwang Moo Yi, Eduard Trulls, Vincent Lepetit, and Pascal Fua. Lift: Learned invariant feature transform. In *European Conference on Computer Vision*, pages 467–483. Springer, 2016.

[158]Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. Toward geometric deep slam. *arXiv preprint arXiv:1707.07410*, 2017.

[159] Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. Su- perpoint: Self-supervised interest point detection and description. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 224–236, 2018.

[160] Peiliang Li, Qin Tong, and Shaojie Shen. Stereo vision-based semantic 3d object and ego-motion tracking for autonomous driving. In *Euro- pean Conference on Computer Vision*, 2018.

[161] Jiexiong Tang, Ludvig Ericson, John Folkesson, and Patric Jensfelt. Gcnv2: Efficient correspondence prediction for real-time slam. *arXiv preprint arXiv:1902.11046*, 2019.

[162] Margarita Grinvald, Fadri Furrer, Tonci Novkovic, Jen Jen Chung, Cesar Cadena, Roland Siegwart, and Juan Nieto. Volumetric instance- aware semantic mapping and 3d object discovery. *arXiv preprint arXiv:1903.00268*, 2019.

[163] Huai-Jen Liang, Nitin J Sanket, Cornelia Fermu ̈ller, and Yiannis Aloimonos. Salientdso: Bringing attention to direct sparse odometry. *IEEE Transactions on Automation Science and Engineering*, 2019.

[164] Mehdi Hosseinzadeh, Yasir Latif, Trung Pham, Niko Suenderhauf, and Ian Reid. Structure aware slam using quadrics and planes. In *Asian Conference on Computer Vision*, pages 410–426. Springer, 2018.

[165] Shichao Yang and Sebastian Scherer. Cubeslam: Monocular 3-d object slam. *IEEE Transactions on Robotics*, 2019.

[166] Shichao Yang and Sebastian Scherer. Monocular object and plane slam in structured environments. *IEEE Robotics and Automation Letters*, 4(4):3145–3152, 2019.

[167] Zengyi Qin, Jinglu Wang, and Yan Lu. Monogrnet: A geometric reasoning network for 3d object localization. *The Thirty-Third AAAI Conference on Artificial Intelligence (AAAI-19)*, 2019.

[168] Xavier Lagorce, Sio Hoi Ieng, and Ryad Benosman. Event-based features for robotic vision. In *IEEE/RSJ International Conference on Intelligent Robots & Systems*, 2013.

[169] Elias Mueggler, Chiara Bartolozzi, and Davide Scaramuzza. Fast event- based corner detection. In *BMVC*, 2017.

[170] Xiongwei Wu, Doyen Sahoo, and Steven CH Hoi. Recent advances in deep learning for object detection. *arXiv preprint arXiv:1908.03673*, 2019.

[171 Renato F. Salas-Moreno, Richard A. Newcombe, Hauke Strasdat, Paul H. J. Kelly, and Andrew J. Davison. Slam++: Simultaneous localisation and mapping at the level of objects. In *Computer Vision & Pattern Recognition*, 2013.

[172] Xuanpeng Li and Rachid Belaroussi. Semi-dense 3d semantic mapping from monocular slam. 2016.

[173] John McCormac, Ankur Handa, Andrew Davison, and Stefan Leuteneg- ger. Semanticfusion: Dense 3d semantic mapping with convolutional neural networks. In *2017 IEEE International Conference on Robotics and automation (ICRA)*, pages 4628–4635. IEEE, 2017.

[174] Niko Sunderhauf, Trung T. Pham, Yasir Latif, Michael Milford, and Ian Reid. Meaningful maps with object-oriented semantic mapping. In *IEEE/RSJ International Conference on Intelligent Robots & Systems*, 2017.

[175]Jiajun Wu, Yifan Wang, Tianfan Xue, Xingyuan Sun, Bill Freeman, and Josh Tenenbaum. Marrnet: 3d shape reconstruction via 2.5 d sketches. In *Advances in neural information processing systems*, pages 540–550, 2017.

[176] Angela Dai and Matthias Nießner. 3dmv: Joint 3d-multi-view pre- diction for 3d semantic scene segmentation. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 452–468, 2018.

[177] Xingyuan Sun, Jiajun Wu, Xiuming Zhang, Zhoutong Zhang, Chengkai Zhang, Tianfan Xue, Joshua B Tenenbaum, and William T Freeman. Pix3d: Dataset and methods for single-image 3d shape modeling. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

[178] Angela Dai, Daniel Ritchie, Martin Bokeloh, Scott Reed, Jürgen Sturm, and Matthias Nießner. Scancomplete: Large-scale scene completion and semantic segmentation for 3d scans. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4578– 4587, 2018.

[179] John McCormac, Ronald Clark, Michael Bloesch, Andrew J. Davison, and Stefan Leutenegger. Fusion++: Volumetric object-level slam. *2018 International Conference on 3D Vision (3DV)*, pages 32–41, 2018.

[180] Renaud Dubé, Andrei Cramariuc, Daniel Dugas, Juan Nieto, Roland Siegwart, and Cesar Cadena. Segmap: 3d segment mapping using data- driven descriptors. *arXiv preprint arXiv:1804.09557*, 2018.

[181] Ji Hou, Angela Dai, and Matthias Nießner. 3d-sis: 3d semantic instance segmentation of rgb-d scans. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4421–4430, 2019.

[182] Yu Xiang and Dieter Fox. Da-rnn: Semantic mapping with data associated recurrent neural networks. *arXiv preprint arXiv:1703.03098*, 2017.

[183] Chen Wang, Danfei Xu, Yuke Zhu, Roberto Mart́ın-Mart́ın, Cewu Lu, Li Fei-Fei, and Silvio Savarese. Densefusion: 6d object pose estimation by iterative dense fusion. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3343–3352, 2019.

[184] Zilong Huang, Xinggang Wang, Lichao Huang, Chang Huang, Yunchao Wei, and Wenyu Liu. Ccnet: Criss-cross attention for semantic segmentation. *arXiv preprint arXiv:1811.11721*, 2018.

[185] Evangelos Stromatias, Miguel Soto, Maŕıa Teresa Serrano Gotarredona, and Bernabé Linares Barranco. An event-based classifier for dynamic vision sensor and synthetic data. *Frontiers in Neuroscience, 11 (art́ıculo 360)*, 2017.

[186] Jean-Matthieu Maro and Ryad Benosman. Event-based gesture recog- nition with dynamic background suppression using smartphone com- putational capabilities. *arXiv preprint arXiv:1811.07802*, 2018.

[187] Saeed Afshar, Tara Julia Hamilton, Jonathan C Tapson, André van Schaik, and Gregory Kevin Cohen. Investigation of event-based mem- ory surfaces for high-speed detection, unsupervised feature extraction, and object recognition. *Frontiers in neuroscience*, 12:1047, 2018.

[188] Alejandro Linares-Barranco, Antonio Rios-Navarro, Ricardo Tapiador- Morales, and Tobi Delbruck. Dynamic vision sensor integration on fpga-based cnn accelerators for high-speed visual classification. *arXiv preprint arXiv:1905.07419*, 2019.

[189] Keisuke Tateno, Federico Tombari, Iro Laina, and Nassir Navab. Cnn- slam: Real-time dense monocular slam with learned depth prediction. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6243–6252, 2017.

[190] Vikram Mohanty, Shubh Agrawal, Shaswat Datta, Arna Ghosh, Vishnu Dutt Sharma, and Debashish Chakravarty. Deepvo: A deep learning approach for monocular visual odometry. *arXiv preprint arXiv:1611.06069*, 2016.

[191] Buyu Li, Wanli Ouyang, Lu Sheng, Xingyu Zeng, and Xiaogang Wang. Gs3d: An efficient 3d object detection framework for autonomous driving. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1019–1028, 2019.

[192] Ruihao Li, Sen Wang, Zhiqiang Long, and Dongbing Gu. Undeepvo: Monocular visual odometry through unsupervised deep learning. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 7286–7291. IEEE, 2018.

[193] Zhengqi Li, Tali Dekel, Forrester Cole, Richard Tucker, Noah Snavely, Ce Liu, and William T Freeman. Learning the depths of moving people by watching frozen people. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4521–4530, 2019.

[194] D. Frost, V. Prisacariu, and D. Murray. Recovering stable scale in monocular slam using object-supplemented bundle adjustment. *IEEE Transactions on Robotics*, 34(3):736–747, June 2018.

[195] Edgar Sucar and Jean Bernard Hayet. Bayesian scale estimation for monocular slam based on generic object detection for correcting scale drift. 2017.

[196] Zhichao Yin and Jianping Shi. Geonet: Unsupervised learning of dense depth, optical flow and camera pose. In *CVPR*, 2018.

[197] Michael Bloesch, Jan Czarnowski, Ronald Clark, Stefan Leutenegger, and Andrew J Davison. Codeslamlearning a compact, optimisable representation for dense visual slam. In *Proceedings*

of the IEEE Conference on Computer Vision and Pattern Recognition, pages 2560– 2568, 2018.

[198] Fabian Brickwedde, Steffen Abraham, and Rudolf Mester. Mono- stixels: monocular depth reconstruction of dynamic street scenes. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1–7. IEEE, 2018.

[199] Yasin Almalioglu, Muhamad Risqi U. Saputra, Pedro Porto Buarque de Gusmão, Andrew Markham, and Agathoniki Trigoni. Ganvo: Unsupervised deep monocular visual odometry and depth estimation with generative adversarial networks. *2019 International Conference on Robotics and Automation (ICRA)*, pages 5474–5480, 2018.

[200]Punarjay Chakravarty, Praveen Narayanan, and Tom Roussel. Gen- slam: Generative modeling for monocular simultaneous localization and mapping. *arXiv preprint arXiv:1902.02086*, 2019.

[201]KatrinLasinger,RenéRanftl,KonradSchindler,andVladlenKoltun. Towards robust monocular depth estimation: Mixing datasets for zero- shot cross-dataset transfer. *arXiv:1907.01341*, 2019.

[202] Po-Han Huang, Kevin Matzen, Johannes Kopf, Narendra Ahuja, and Jia-Bin Huang. Deepmvs: Learning multi-view stereopsis. In *Pro- ceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2821–2830, 2018.

[203] Jia Deng Zachary Teed. Deepv2d: Video to depth with differentiable structure from motion. *arXiv:1812.04605*, 2018.

[204] Germain Haessig, Xavier Berthelon, Sio-Hoi Ieng, and Ryad Benos- man. A spiking neural network model of depth from defocus for event- based neuromorphic vision. *Scientific reports*, 9(1):3744, 2019.

[205] Guillermo Gallego, Henri Rebecq, and Davide Scaramuzza. A unifying contrast maximization framework for event cameras, with applications to motion, depth, and optical flow estimation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3867–3876, 2018.

[206] Zhen Xie, Shengyong Chen, and Garrick Orchard. Event-based stereo depth estimation using belief propagation. *Frontiers in neuroscience*, 11:535, 2017.

[207] Kishore Reddy Konda and Roland Memisevic. Learning visual odom- etry with a convolutional network. In *VISAPP (1)*, pages 486–490, 2015.

[208] Gabriele Costante, Michele Mancini, Paolo Valigi, and Thomas A Ciarfuglia. Exploring representation learning with cnns for frame-to- frame ego-motion estimation. *IEEE robotics and automation letters*, 1(1):18–25, 2015.

[209] Alex Kendall, Matthew Grimes, and Roberto Cipolla. Posenet: A convolutional network for real-time 6-dof camera relocalization. In *Proceedings of the IEEE international conference on computer vision*, pages 2938–2946, 2015.

[210] Ronald Clark, Sen Wang, Hongkai Wen, Andrew Markham, and Niki Trigoni. Vinet: Visual-inertial odometry as a sequence-to-sequence learning problem. In *Thirty-First AAAI Conference on Artificial Intelligence*, 2017.

[211] Sen Wang, Ronald Clark, Hongkai Wen, and Niki Trigoni. Deepvo: Towards end-to-end visual odometry with deep recurrent convolutional neural networks. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2043–2050. IEEE, 2017.

[212] Tinghui Zhou, Matthew Brown, Noah Snavely, and David G Lowe. Unsupervised learning of depth and ego-motion from video. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1851–1858, 2017.

[213] Sudheendra Vijayanarasimhan, Susanna Ricco, Cordelia Schmid, Rahul Sukthankar, and Katerina Fragkiadaki. Sfm-net: Learning of structure and motion from video. *arXiv preprint arXiv:1704.07804*, 2017.

[214]Konstantinos Nektarios Lianos, Johannes L. Schnberger, Marc Polle- feys, and Torsten Sattler. Vso: Visual semantic odometry. In *European Conference on Computer Vision (ECCV)*, 2018.

[215] Ronald Clark, Sen Wang, Andrew Markham, Niki Trigoni, and Hongkai Wen. Vidloc: A deep spatio-temporal model for 6-dof video-clip relocalization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6856–6864, 2017.

[216] Guillermo Gallego, Christian Forster, Elias Mueggler, and Davide Scaramuzza. Event-based camera pose tracking using a generative event model. *arXiv preprint arXiv:1510.01972*, 2015.

[217] David Reverter Valeiras, Garrick Orchard, Sio-Hoi Ieng, and Ryad B Benosman. Neuromorphic event-based 3d pose estimation. *Frontiers in neuroscience*, 9:522, 2016.

[218] Sean L. Bowman, Nikolay Atanasov, Kostas Daniilidis, and George J. Pappas. Probabilistic data association for semantic slam. In *IEEE International Conference on Robotics & Automation*, 2017.

[219] Nate Merrill and Guoquan Huang. Lightweight unsupervised deep loop closure. *arXiv preprint arXiv:1805.07703*, 2018.

[220] Erik Stenborg, Carl Toft, and Lars Hammarstrand. Long-term visual localization using semantically segmented images. pages 6484–6490, 2018.

[221] Abel Gawel, Carlo Del Don, Roland Siegwart, Juan Nieto, and Cesar Cadena. X-view: Graph-based semantic multi-view localization. *IEEE Robotics and Automation Letters*, 3(3):1687–1694, 2018.

[222] Kevin Doherty, Dehann Fourie, and John Leonard. Multimodal seman- tic slam with probabilistic data association. In *2019 IEEE International Conference on Robotics and Automation (ICRA). IEEE*, 2019.

[223] Andrea Censi, Jonas Strubel, Christian Brandli, Tobi Delbruck, and Davide Scaramuzza. Low-latency localization by active led markers tracking using a dynamic vision sensor. In *2013 IEEE/RSJ Interna-tional Conference on Intelligent Robots and Systems*, pages 891–898. IEEE, 2013.

[224] Wei Tan, Haomin Liu, Zilong Dong, Guofeng Zhang, and Hujun Bao. Robust monocular slam in dynamic environments. In *2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 209–218. IEEE, 2013.

[225] Chao Yu, Zuxin Liu, Xin-Jun Liu, Fugui Xie, Yi Yang, Qi Wei, and Qiao Fei. Ds-slam: A semantic visual slam towards dynamic envi- ronments. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 1168–1174. IEEE, 2018.

[226] Waleed Abdulla. Mask r-cnn for object detection and instance segmen- tation on keras and tensorflow, 2017.

[227] Martin Ru ̈nz and Lourdes Agapito. Co-fusion: Real-time segmentation, tracking and fusion of multiple objects. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pages 4471–4478. IEEE, 2017.

[228] Fangwei Zhong, Wang Sheng, Ziqi Zhang, China Chen, and Yizhou Wang. Detect-slam: Making object detection and slam mutually beneficial. In *IEEE Winter Conference on Applications of Computer Vision*, 2018.

[229] Berta Bescos, José ́ M Fa ́cil, Javier Civera, and José ́ Neira. Dynaslam: Tracking, mapping, and inpainting in dynamic scenes. *IEEE Robotics and Automation Letters*, 3(4):4076–4083, 2018.

[230] Raluca Scona, Mariano Jaimez, Yvan R Petillot, Maurice Fallon, and Daniel Cremers. Staticfusion: Background reconstruction for dense rgb-d slam in dynamic environments. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1–9. IEEE, 2018.

[231] Zemin Wang, Qian Zhang, Jiansheng Li, Shuming Zhang, and Jingbin Liu. A computationally efficient semantic slam solution for dynamic scenes. *Remote Sensing*, 11(11):1363, 2019.

[232] Linhui Xiao, Jinge Wang, Xiaosong Qiu, Zheng Rong, and Xudong Zou. Dynamic-slam: Semantic monocular visual localization and mapping based on deep learning in dynamic environment. *Robotics and Autonomous Systems*, 117:1–16, 2019.

[233] Ioan Andrei Baˆrsan, Peidong Liu, Marc Pollefeys, and Andreas Geiger. Robust dense mapping for large-scale dynamic environments. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 7510–7517. IEEE, 2018.

[234] Se-Ho Kim Ue-Hwan Kim and Jong-Hwan Kim. Simvodis: Simultane- ous visual odometry, object detection, and instance segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence, Under Review*, 2019.

[235] Muhammad Sualeh and Gon-Woo Kim. Simultaneous localization and mapping in the epoch of semantics: A survey. *International Journal of Control, Automation and Systems*, 17(3):729–742, 2019.

[236] Peng Wang, Ruigang Yang, Binbin Cao, Wei Xu, and Yuanqing Lin. Dels-3d: Deep localization and segmentation with a 3d semantic map. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5860–5869, 2018.

[237] Nikolay Atanasov, Sean L Bowman, Kostas Daniilidis, and George J Pappas. A unifying view of geometry, semantics, and data association in slam. In *IJCAI*, pages 5204–5208, 2018.

[238]RubenGomez-Ojeda,Francisco-AngelMoreno,DavidZun˜iga-Noe˙l, Davide Scaramuzza, and Javier Gonzalez-Jimenez. Pl-slam: a stereo slam system through the combination of points and line segments. *IEEE Transactions on Robotics*, 2019.

[239] Huizhong Zhou, Danping Zou, Ling Pei, Rendong Ying, Peilin Liu, and Wenxian Yu. Structslam: Visual slam with building structure lines. *IEEE Transactions on Vehicular Technology*, 64(4):1364–1375, 2015.

[240] Sren Riisgaard and Morten Rufus Blas. Slam for dummies: A tutorial approach to simultaneous localization and mapping. Technical report, 2005.

[241] Joern Rehder, Janosch Nikolic, Thomas Schneider, Timo Hinzmann, and Roland Siegwart. Extending kalibr: Calibrating the extrinsics of multiple imus and of individual axes. In *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pages 4304–4311. IEEE, 2016.

[242] Kevin Eckenhoff, Patrick Geneva, Jesse Bloecker, and Guoquan Huang. Multi-camera visual-inertial navigation with online intrinsic and ex- trinsic calibration. *2019 International Conference on Robotics and Automation (ICRA)*, pages 3158–3164, 2019.

[243] A. Tedaldi, A. Pretto, and E. Menegatti. A robust and easy to implement method for imu calibration without external equipments. In *Proc. of: IEEE International Conference on Robotics and Automation (ICRA)*, pages 3042–3049, 2014.

[244] A. Pretto and G. Grisetti. Calibration and performance evaluation of low-cost imus. In *Proc. of: 20th IMEKO TC4 International Symposium*, pages 429–434, 2014.

[245] Mingyang Li, Hongsheng Yu, Xing Zheng, and Anastasios I Mourikis. High-fidelity sensor modeling

and self-calibration in vision-aided in- ertial navigation. In *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pages 409–416. IEEE, 2014.

[246] Changhao Chen, Stefano Rosa, Yishu Miao, Chris Xiaoxuan Lu, Wei Wu, Andrew Markham, and Niki Trigoni. Selective sensor fusion for neural visual-inertial odometry. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 10542–10551, 2019.

[247] Thomas Schops, Torsten Sattler, and Marc Pollefeys. Bad slam: Bundle adjusted direct rgb-d slam. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.

[248] Adam Harmat, Michael Trentini, and Inna Sharf. Multi-camera tracking and mapping for unmanned aerial vehicles in unstructured environments. *Journal of Intelligent & Robotic Systems*, 78(2):291– 317, 2015.

[249] Steffen Urban and Stefan Hinz. MultiCol-SLAM - a modular real-time multi-camera slam system. *arXiv preprint arXiv:1610.07336*, 2016.

[250] Michael Bloesch, Michael Burri, Sammy Omari, Marco Hutter, and Roland Siegwart. Iterated extended kalman filter based visual-inertial odometry using direct photometric feedback. *The International Journal of Robotics Research*, 36(10):1053–1072, 2017.

[251] Haoyang Ye, Yuying Chen, and Ming Liu. Tightly coupled 3d lidar inertial odometry and mapping. *arXiv preprint arXiv:1904.06993*, 2019.

[252] Deyu Yin, Jingbin Liu, Teng Wu, Keke Liu, Juha Hyyppä, and Ruizhi Chen. Extrinsic calibration of 2d laser rangefinders using an existing cuboid-shaped corridor as the reference. *Sensors*, 18(12):4371, 2018.

[253] Shoubin Chen, Jingbin Liu, Teng Wu, Wenchao Huang, Keke Liu, Deyu Yin, Xinlian Liang,

Juha Hyyppä; and Ruizhi Chen. Extrinsic calibration of 2d laser rangefinders based on a mobile sphere. *Remote Sensing*, 10(8):1176, 2018.

[254] Jesse Sol Levinson. *Automatic laser calibration, mapping, and local- ization for autonomous vehicles*. Stanford University, 2011.

[255] Jesse Levinson and Sebastian Thrun. Automatic online calibration of cameras and lasers. In *Robotics: Science and Systems*, volume 2, 2013.

[256] A. Dhall, K. Chelani, V. Radhakrishnan, and K. M. Krishna. LiDAR- Camera Calibration using 3D-3D Point correspondences. *ArXiv e- prints*, May 2017.

[257] Nick Schneider, Florian Piewak, Christoph Stiller, and Uwe Franke. Regnet: Multimodal sensor registration using deep neural networks. In *2017 IEEE intelligent vehicles symposium (IV)*, pages 1803–1810. IEEE, 2017.

[258] Johannes Graeter, Alexander Wilczynski, and Martin Lauer. Limo: Lidar-monocular visual odometry. 2018.

[259] Ganesh Iyer, J Krishna Murthy, K Madhava Krishna, et al. Calib- net: self-supervised extrinsic calibration using 3d spatial transformer networks. *arXiv preprint arXiv:1803.08181*, 2018.

[260] Faraz M Mirzaei, Dimitrios G Kottas, and Stergios I Roumeliotis. 3d lidar–camera intrinsic and extrinsic calibration: Identifiability and analytical least-squares-based initialization. *The International Journal of Robotics Research*, 31(4):452–467, 2012.

[261] Ryoichi Ishikawa, Takeshi Oishi, and Katsushi Ikeuchi. Lidar and camera calibration using motions estimated by sensor fusion odometry. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 7342–7349. IEEE, 2018.

[262] Jesse Levinson and Sebastian Thrun. Automatic online calibration of cameras and lasers. In *Robotics: Science and Systems*, 2013.

[263] Sharmin Rahman, Alberto Quattrini Li, and Ioannis Rekleitis. Svin2: An underwater slam system using sonar, visual, inertial, and depth sensor.

[264] Changjun Gu, Yang Cong, and Gan Sun. Environment driven under- water camera-imu calibration for monocular visual-inertial slam. *2019 International Conference on Robotics and Automation (ICRA)*, pages 2405–2411, 2019.

[265] Bilal Arain, Chris McCool, Paul Rigby, Daniel Cagara, and Matthew Dunbabin. Improving underwater obstacle detection using semantic image segmentation. *2019 International Conference on Robotics and Automation (ICRA)*, pages 9271–9277, 2019.

[266] Brian Ferris, Dieter Fox, and Neil D Lawrence. Wifi-slam using gaussian process latent variable models. In *IJCAI*, volume 7, pages 2480–2485, 2007.

[267] Mohammed Aladsani, Ahmed Alkhateeb, and Georgios C Trichopou- los. Leveraging mmwave imaging and communications for simul- taneous localization and mapping. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4539–4543. IEEE, 2019.

[268] Ojas Kanhere and Theodore S Rappaport. Position locationing for millimeter wave systems. In *2018 IEEE Global Communications Conference (GLOBECOM)*, pages 206–212. IEEE, 2018.

[269] Theodore S Rappaport, Yunchou Xing, Ojas Kanhere, Shihao Ju, Arjuna Madanayake, Soumyajit Mandal, Ahmed Alkhateeb, and Geor- gios C Trichopoulos. Wireless communications and applications above 100 ghz: Opportunities and challenges for 6g and beyond. *IEEE Access*, 7:78729–78757, 2019.

[270] Shehryar Khattak, Christos Papachristos, and Kostas Alexis. Keyframe- based direct thermalinertial odometry. *2019 International Conference on Robotics and Automation (ICRA)*, pages 3563–3569, 2019.

[271] David Ferstl, Christian Reinbacher, Rene Ranftl, Matthias Ru ther, and Horst Bischof. Image guided depth upsampling using anisotropic total generalized variation. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 993–1000, 2013.

[272] Jason Ku, Ali Harakeh, and Steven L Waslander. In defense of classical image processing: Fast depth completion on the cpu. In *2018 15th Conference on Computer and Robot Vision (CRV)*, pages 16–22. IEEE, 2018.

[273] Fangchang Mal and Sertac Karaman. Sparse-to-dense: Depth prediction from sparse depth samples and a single image. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1–8. IEEE, 2018.

[274] Jonas Uhrig, Nick Schneider, Lukas Schneider, Uwe Franke, Thomas Brox, and Andreas Geiger. Sparsity invariant cnns. In *2017 Interna- tional Conference on 3D Vision (3DV)*, pages 11–20. IEEE, 2017.

[275] Shreyas S Shivakumar, Ty Nguyen, Steven W Chen, and Camillo J Tay- lor. Dfusenet: Deep fusion of rgb and sparse depth information for im- age guided dense depth completion. *arXiv preprint arXiv:1902.00761*, 2019.

[276] Zhao Chen, Vijay Badrinarayanan, Gilad Drozdov, and Andrew Rabi- novich. Estimating depth from rgb and sparse sensing. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 167– 182, 2018.

[277] Abdelrahman Eldesokey, Michael Felsberg, and Fahad Shahbaz Khan. Propagating confidences

through cnns for sparse data regression. *arXiv preprint arXiv:1805.11913*, 2018.

[278] Xingxing Zuo, Patrick Geneva, Woosik Lee, Yong Liu, and Guoquan Huang. Lic-fusion: Lidar-inertial-camera odometry. *arXiv preprint arXiv:1909.04102*, 2019.

[279] Olivier Aycard, Qadeer Baig, Siviu Bota, Fawzi Nashashibi, Sergiu Nedevschi, Cosmin Pantilie, Michel Parent, Paulo Resende, and Trung-Dung Vu. Intersection safety using lidar and stereo vision sensors. In *2011 IEEE Intelligent Vehicles Symposium (IV)*, pages 863–869. IEEE, 2011.

[280] Ricardo Omar Chavez-Garcia and Olivier Aycard. Multiple sensor fusion and classification for moving object detection and tracking. *IEEE Transactions on Intelligent Transportation Systems*, 17(2):525– 534, 2015.

[281] Hyunggi Cho, Young-Woo Seo, BVK Vijaya Kumar, and Ragu- nathan Raj Rajkumar. A multi-sensor fusion system for moving object detection and tracking in urban driving environments. In *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1836–1843. IEEE, 2014.

[282] Tao Wang, Nanning Zheng, Jingmin Xin, and Zheng Ma. Integrating millimeter wave radar with a monocular vision sensor for on-road obstacle detection applications. *Sensors*, 11(9):8992–9008, 2011.

[283] Guowei Wan, Xiaolong Yang, Renlan Cai, Hao Li, Yao Zhou, Hao Wang, and Shiyu Song. Robust and precise vehicle localization based on multi-sensor fusion in diverse city scenes. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 4670–4677. IEEE, 2018.

[284] Ji Zhang, Michael Kaess, and Sanjiv Singh. Real-time depth enhanced monocular odometry. In *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 4973–4980. IEEE, 2014.

[285] Ji Zhang and Sanjiv Singh. Visual-lidar odometry and mapping: Low- drift, robust, and fast. In *2015 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2174–2181. IEEE, 2015.

[286] Yoshua Nava. *Visual-LiDAR SLAM with loop closure*. PhD thesis, Masters thesis, KTH Royal Institute of Technology, 2018.

[287] Yinglei Xu, Yongsheng Ou, and Tiantian Xu. Slam of robot based on the fusion of vision and lidar. In *2018 IEEE International Conference on Cyborg and Bionic Systems (CBS)*, pages 121–126. IEEE, 2018.

[288] Weizhao Shao, Srinivasan Vijayarangan, Cong Li, and George Kantor. Stereo visual inertial lidar simultaneous localization and mapping. *arXiv preprint arXiv:1902.10741*, 2019.

[289] Franz Andert, Nikolaus Ammann, and Bolko Maass. Lidar-aided cam- era feature tracking and visual slam for spacecraft low-orbit navigation and planetary landing. In *Advances in Aerospace Guidance, Navigation and Control*, pages 605–623. Springer, 2015.

[290] Danfei Xu, Dragomir Anguelov, and Ashesh Jain. Pointfusion: Deep sensor fusion for 3d bounding box estimation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 244–253, 2018.

[291] Kiwoo Shin, Youngwook Paul Kwon, and Masayoshi Tomizuka. Roar- net: A robust 3d object detection based on region approximation refinement. *arXiv preprint arXiv:1811.03818*, 2018.

[292] Jason Ku, Melissa Mozifian, Jungwook Lee, Ali Harakeh, and Steven Waslander. Joint 3d proposal generation and object detection from view aggregation. *IROS*, 2018.

[293] Caner Hazirbas, Lingni Ma, Csaba Domokos, and Daniel Cremers. Fusenet: Incorporating depth into semantic segmentation via fusion- based cnn architecture. In *Asian conference on computer vision*, pages 213–228. Springer, 2016.

[294] Zining Wang, Wei Zhan, and Masayoshi Tomizuka. Fusing birds eye view lidar point cloud and front view camera image for 3d object detection. In *2018 IEEE Intelligent Vehicles Symposium (IV)*, pages 1–6. IEEE, 2018.

[295] Ming Liang, Bin Yang, Shenlong Wang, and Raquel Urtasun. Deep continuous fusion for multi-sensor 3d object detection. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 641– 656, 2018.

[296] Shuo Gu, Tao Lu, Yigong Zhang, Jose M Alvarez, Jian Yang, and Hui Kong. 3-d lidar+ monocular camera: An inverse-depth-induced fusion framework for urban road detection. *IEEE Transactions on Intelligent Vehicles*, 3(3):351–360, 2018.

[297] Cesar Cadena, Luca Carlone, Henry Carrillo, Yasir Latif, Davide Scaramuzza, Jose ́ Neira, Ian Reid, and John J Leonard. Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age. *IEEE Transactions on robotics*, 32(6):1309– 1332, 2016.

[298] Titus Cieslewski, Siddharth Choudhary, and Davide Scaramuzza. Data- efficient decentralized visual slam. In *2018 IEEE International Con- ference on Robotics and Automation (ICRA)*, pages 2466–2473. IEEE, 2018.

[299] Cynthia Dwork. Differential privacy. *Encyclopedia of Cryptography and Security*, pages 338–340, 2011.

[300] Frank McSherry and Kunal Talwar. Mechanism design via differential privacy. In *FOCS*, volume 7, pages 94–103, 2007.

[301] Hadi Sarieddeen, Nasir Saeed, Tareq Y Al-Naffouri, and Mohamed- Slim Alouini. Next generation terahertz communications: A rendezvous of sensing, imaging and localization. *arXiv preprint arXiv:1909.10462*, 2019.

[302] Khaled B Letaief, Wei Chen, Yuanming Shi, Jun Zhang, and Ying- Jun Angela Zhang. The roadmap to 6g: Ai empowered wireless networks. *IEEE Communications Magazine*, 57(8):84–90, 2019.

[303] Walid Saad, Mehdi Bennis, and Mingzhe Chen. A vision of 6g wireless systems: Applications, trends, technologies, and open research problems. *IEEE network*, 2019.

[304] Henk Wymeersch, Jiguang He, Benoˆıt Denis, Antonio Clemente, and Markku Juntti. Radio localization and mapping with reconfigurable intelligent surfaces. *arXiv preprint arXiv:1912.09401*, 2019.

[305] Marco Giordani, Michele Polese, Marco Mezzavilla, Sundeep Rangan, and Michele Zorzi. Towards 6g networks: Use cases and technologies. *arXiv preprint arXiv:1903.12216*, 2019.

[306] Walid Saad, Mehdi Bennis, and Mingzhe Chen. A vision of 6g wireless systems: Applications, trends, technologies, and open research problems. *IEEE network*, 2019.

[307] Henk Wymeersch, Jiguang He, Benoˆıt Denis, Antonio Clemente, and Markku Juntti. Radio localization and mapping with reconfigurable intelligent surfaces. *arXiv preprint arXiv:1912.09401*, 2019.

[308] Marco Giordani, Michele Polese, Marco Mezzavilla, Sundeep Rangan, and Michele Zorzi. Towards 6g networks: Use cases and technologies. *arXiv preprint arXiv:1903.12216*, 2019.

Authors

Baichuan Huang has graduated with a master's degree in Computer Application Technology in the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University. He gained bachelor's degree in school of electronic information, Wuhan University in 2018. During the period of postgraduate, he was award for winning the National scholarship by Ministry of Education of the People's Republic of China.

His research focuses on indoor location with communication, Simultaneous localization and mapping (SLAM), Multi-view stereo reconstruction in 3D computer vision.

Jun Zhao (S'10-M'15) is currently an Assistant Professor in the School of Computer Science and Engineering at Nanyang Technological University (NTU) in Singapore. He received a PhD degree in Electrical and Computer Engineering from Carnegie Mellon University (CMU) in the USA (advisors: Virgil Gligor, Osman Yagan; collaborator: Adrian Perrig), affiliating with CMU's renowned CyLab Security \& Privacy Institute, and a bachelor's degree from Shanghai Jiao Tong University in China. Before joining NTU first as a postdoc with Xiaokui Xiao and then as a faculty member, he was a postdoc at Arizona State University as an Arizona Computing PostDoc Best Practices Fellow (advisors: Junshan Zhang, Vincent Poor). His research interests are Communications \& Networks, Security and AI. One of his first-authored papers was shortlisted for the best student paper award in IEEE International Symposium on Information Theory (ISIT) 2014, a prestigious conference in information theory.

Sheng Luo graduated from Information Engineering University and got his Ph.D. degree in 2014. Now he is a postdoctoral researcher in both Hi-Target Surveying Instrument Co. Ltd and State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University. His research focuses on UAV-borne Lidar applications on emergency surveying and mapping and 3D Laser SLAM technology.

Jingbin Liu is a Professor in Positioning and Navigation with State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing (LIESMARS) in Wuhan University, China. He received his Ph.D Degree in 2008 from Wuhan University. Dr.Liu worked as Specialist Research Scientist with Finnish Geospatial Research Institute (FGI, formerly known as Finnish Geodetic Institute) from 2008 to 2016. His interest areas include indoor and outdoor positioning, smartphone navigation, indoor mobile mapping, and GNSS/INS/SLAM integration technology.