This article has been accepted for publication in IEEE Transactions on Intelligent Vehicles. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIV.2022.3192102

A Survey on Map-Based Localization Techniques for Autonomous Vehicles

Athanasios Chalvatzaras, Ioannis Pratikakis, and Angelos A. Amanatiadis, *Senior Member, IEEE*

***Abstract*—Autonomous vehicles integrate complex software stacks for realizing the necessary iterative perception, planning, and action operations. One of the foundational layers of such stacks is the perception one which is comprised of localization, detection, and recognition algorithms for understanding the location and the driving environment around the vehicle. More precisely, localization aims to identify the location of the vehicle on a global coordinate system and is considered one of the most critical parts in the stack since its accuracy and robustness affects the subsequent algorithms of the perception layer and also the following planning and action layers. Due to the rapid and significant interest in self-driving cars, several localization techniques have been proposed with different directions and approaches. Algorithms using prior maps are currently considered the most accurate ones and found almost in all current self-driving car prototypes. Thus, in this paper, we categorize, discuss and analyze the state-of-the-art map-based localization techniques in an attempt to examine their potentials and limitations. We first present techniques and approaches that aim to match prior maps with on-board observations from different sensor modalities. We then review methods that handle the localization problem as a probabilistic one and finally, we also go through the emerging domain of deep-learning localization algorithms and examine their potential in self-driving cars. For all three categories, we provide comparison tables and necessary insights for the optimal localization system design based on different requirements, specifications, and sensor configurations.**

***Index Terms*—Autonomous vehicles, vehicle localization, high-definition maps, environment perception.**

I. Introduction

THE number of registered vehicles in the U.S continues its constant increase showing a 22% rise in the last twenty years, however, a notable parallel increase of injuries and fatalities in road accidents is also apparent with the human error being the leading factor behind these accidents [1]. Environmental consequences are also apparent, with the most notable of all the mammoth amount of carbon emissions to the atmosphere. It is widely accepted that autonomous vehicles can address these concerns by nullifying human intervention and by applying environment friendly patterns in driving behaviour. Thus, more and more researchers together with the automotive industry direct their efforts towards increasing vehicles' autonomy.

While different architectural approaches are followed by each autonomous vehicle vendor, they all share three common fundamental layers in their stack namely the Perception, the Planning and the Control. Localization is one of the foundational building blocks of the Perception layer, carrying out the task of identifying the location of the vehicle on a global coordinate system and then providing this estimation to the subsequent Perception layer algorithms. The broadness and the significant research effort of the robotics and autonomous vehicle research community on the localization topic has resulted in an extensive taxonomy. The autonomous vehicle localization problem deals with the global localization in a dynamic and semi-structured environment, where the initial position of the vehicle is roughly approximated and with several objects moving dynamically in its environment. The accuracy of localization algorithms determine the reliable operation of the subsequent algorithmic layers. Their robustness under uncertain driving and harsh environmental conditions (e.g. no road markings, fog, high traffic, heavy rain, snowfall, etc.), is critical because poor localization performance can result in poor estimations with unwanted consequences, such as lane departure, curb hitting, or even worse fatal accidents. Autonomous vehicles require lateral and longitudinal accuracy of less than 10cm, in order to reliably operate under any driving or environmental conditions [2].

The aforementioned requirements render localization a complex and crucial task, hence attracted several approaches aiming to its solution, divided into two main clusters: Simultaneous Localization And Mapping (SLAM) and Map-Based approaches. SLAM assumes no prior knowledge about the vehicle's operational environment, yet an online estimation is created to assist the localization process. SLAM was first introduced by Smith and Cheeseman [3] in 1986, and is still an open research topic in robotic applications and by extension in autonomous vehicles [4], [5]. On the other hand, Map-Based techniques rely on the prior existence of information about the static environment. In more detail, the localization process relies on prior maps of the environment, often referred as High Definition Maps (HD Maps) in the automotive industry. Their resolution reach centimeter-level accuracy, therefore they are currently in the spotlight of interest in the domain of field robotics and autonomous vehicle ecosystems.

Both clusters can be divided further into multiple subcategories, showcasing the vastness and complexity of the localization subject. In consequence, there is a plethora of research

This research has been co-financed by the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH-CREATE-INNOVATE (project code: T2EDK-02743).

A. Chalvatzaras is with the Production and Management Engineering Department, Democritus University of Thrace, Xanthi, GR67132, Greece (e-mail: achalvat@pme.duth.gr).

I. Pratikakis, is with the Electrical and Computer Engineering Department, Democritus University of Thrace, Xanthi, GR67132, Greece. (e-mail: ipratika@ee.duth.gr).

A. A. Amanatiadis is with the Production and Management Engineering Department, Democritus University of Thrace, Xanthi, GR67132, Greece (e-mail: aamanat@pme.duth.gr).

This article has been accepted for publication in IEEE Transactions on Intelligent Vehicles. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIV.2022.3192102

works proposing different techniques and algorithms trying to tackle the issue from the autonomous vehicle perspective. Each methodology differs depending on sensor configuration, whether they are using or not a prior map, the environment to be applied and several other operational factors. This paves the way to assort, analyze, and compare all approaches for better understanding of the field both in terms of conducted research but also in terms of system design and applicability.

This paper presents a survey that focus on state-of-the-art localization techniques that utilize prior maps in their pipeline. The proposed classification does not focus on the typical sensor taxonomy but on the core research method for achieving localization estimations. In more detail, the first category includes techniques and approaches that try to match prior maps with on-board observations from different sensor modalities. Subsequently, methods that handle the localization problem as a probabilistic one are reviewed and finally, methods based on the emerging domain of deep-learning are examined. Each set of methods has its potentials and limitations, and also performs better at certain scenarios, based on its mapping and core localization process. However, absolute accuracy comparisons in the context of map-based localization is currently inefficient mainly due to the absence of a universal benchmarking dataset, unlike for example the KITTI dataset [6] which is utilized for comparing the accuracy of SLAM-based works.

The remainder of this work is structured as follows: Section II presents the related work and the motivation of this survey. Section III analyzes the sensor modalities that are currently used in state-of-the-art map-based localization techniques in autonomous vehicles. Section IV is focused on Prior Maps and goes through the different layers of a HD Map, and how it is used in the context of autonomous vehicle localization. Section V is devoted in the analysis of the localization techniques existing in the literature, classified by the core technique of how on-board measurements are matched with the prior map for the localization estimate. A review of publicly available datasets for localization algorithms is presented in Section VI, while in Section VII, insights are given for optimal system design based on requirements, operating environments, sensor configurations and localization specifications. Finally, conclusion remarks are given in Section VIII, along with current challenges and future directions.

II. Related Work and Motivation

While there are several survey works in the literature, their insights and contribution are tightly coupled with the sensor configurations that are investigating. More precisely, in [7], authors categorize localization works by sensor classification, without referring to prior map utilization or not. In [8], the survey is limited to works that utilize only Light Detection And Range (LiDAR) sensors. Authors in [9], address extensively the subject from the SLAM perspective, hence including works that do not utilize prior maps and focusing only on techniques that build and reuse long-term maps. The surveys in [10] and [11] are focused on the navigation and routing aspect of vehicular map-matching methods and proposed novel and proper method classification but the all surveyed methods utilize only the lane level layer of prior maps. A more recent review [12], introduces a novel taxonomy of localization methods for autonomous vehicles but only for highway scenarios, while the review in [13] focuses mainly on the computational complexity of different SLAM localization techniques

This review differs from the aforementioned reviews mainly due to its focus on the continuous evolution and use of high-definition maps. The maps currently include several other features apart from the lane level layer [14], [15], allowing recent algorithms to rely and evolve on the advantages that these maps offer. These advantages are explained in our review allowing to understand the specific needs and performance in different driving environments. Hence, we first review the high-definition maps, the availability of different sensors and then perform a novel map-matching classification of recent algorithms based on the core matching techniques. This enables to address the problem of map-matching localization in its full extent and provide insights based on the reported accuracies.

This survey tries to offer an up-to-date and complete framework based on the recent progress in HD maps, sensors, datasets and existing algorithmic solutions for different driving scenarios. All surveyed works are discussed and accompanied by the used sensors, the map layers that utilize, the experimental environment that were tested and their reported accuracy, thus providing the necessary insights for optimal system design.

III. Sensor Modalities

To localize itself, the vehicle should first precept its surroundings by sensing the operation environment with the use of sensors. The several existing sensors that can be utilized to support the localization task can be divided into two distinct categories, active sensors which use an energy source to probe the environment, hence emit some form of energy and measure the level of the reflected energy to extract knowledge about the scene, and passive sensors, which on contrast measure energy already present in the environment. In the context of localization based on prior maps, sensor readings are used to build the map and also later localize the vehicle, by registering live measurements with prior map data.

A. Active Sensors

1) Radio Detection And Ranging (RADAR): RADARs have been used in the automotive industry for years and can be found in systems like Adaptive Cruise Control (ACC), blind-spot warning, collision warning, and collision avoidance. RADARs consist of a transmitter, which disperses radio frequency domain electromagnetic waves in the environment, and a receiver, which receives back these waves after they reflect on various objects (buildings, cars, etc.). Their operation is based on the Doppler effect, hence they can measure not only the distance but also the velocity of an object, which a crucial advantage for the automotive field. Radio waves can penetrate various materials, giving RADARs the advantage of better housing properties for discreet integration in the vehicle. RADARs entail also wide Field of Views (FoVs) and

TABLE I
AUTONOMOUS VEHICLE EXTEROCEPTIVE AND PROPRIOCEPTIVE LOCALIZATION SENSORS OVERVIEW

	Sensor	Raw output	Typical specifications	Advantages over other sensors	Disadvantages
Exteroceptive Sensors	Radio detection and ranging	longitudinal distance longitudinal velocity latitudinal distance	form factor: ◐ field of view: ◐ range: ● resolution: ○ cost: ◐ noise susceptibility: ●	resistance to weather conditions discreet vehicle integration velocity measurement	false positive detections
Exteroceptive Sensors	Light detection and ranging	point cloud	form factor: ● field of view: ● range: ● resolution: ◐ cost: ● noise susceptibility: ◐	discriminate object's reflectivity high accuracy	affected by weather conditions needs high computational power
Exteroceptive Sensors	Digital camera	image frames	form factor: ○ field of view: ◐ range: ◐ resolution: ● cost: ○ noise susceptibility: ○	rich semantic information discreet vehicle integration	affected by weather conditions needs high computational power
Proprioceptive sensors	Satellite-based radio	longitudinal position latitudinal position altitudinal position	form factor: ○ field of view: N/A range: N/A resolution: ○ cost: ○ noise susceptibility: ◐	lane level accuracy (RTK) initial position estimation	signal blockage multipath
Proprioceptive sensors	Inertial Measurement Unit	triaxial acceleration triaxial angular velocity	form factor: ○ field of view: ● range: N/A resolution: ● cost: ○ noise susceptibility: ◐	provide egomotion high-frequency updates	accumulated errors

●: High/Big ◐: Medium ○: Low/Small. N/A: Not applicable.

long-range capabilities, and are also unaffected by challenging weather conditions such as rain or fog. Besides the numerous advantages, there are two severe disadvantages. First, their relatively low resolution compared to other sensors, especially in the vertical direction, and second their sensitivity to reflectivities, hence when used in the automotive field, developers tend to disregard static objects and mainly use RADARs to estimate the velocity of the other road users.

2) Light Detection And Ranging (LiDAR): The LiDAR sensor operates by emitting infrared laser beams (around 900nm wavelength) to detect the distance between the sensor and the objects. In more detail, LiDAR's light source can cover a 360° FoV, where the emitted beams reflect on the objects in the environment and return to the sensor's receivers. The distance can be extrapolated by calculating the travel time of the beam. LiDARs' output is a point cloud, where every point represents a specific location in the environment, wherever a laser beam is reflected. Advanced LiDAR sensors also measure the reflectivity levels of the objects, further enhancing the already vast amount of information. Its achieved high resolution scanning is a great advantage of LiDARs over other sensors, but comes at the cost of high computational power requirement. Furthermore, they cannot estimate directly the velocity of surrounding objects and are affected by harsh weather conditions. Traditional LiDAR sensors are also pricey and chunky since they contain mechanically moving parts. Emerging Solid State LiDAR (SSL) technology can overcome the constraints regarding price and size due to their completely different philosophy of operation, built entirely on silicon chips with no moving parts, resulting in resilience to vibrations and smaller size. There are three variations of SSLs. One of those leverages the micro-electromechanical system (MEMS) technology in which a MEMS-based mirror moves and directs the beams. Optical Phased Array (OPA) SSLs deliver light pulses to different directions by just adjusting the array. Finally, Flash SSLs do not require direct light sources at all. Their camera-like operation delivers a flash and detects the whole area at once, but with limitations on FoV and sensing range.

B. Passive Sensors

1) Cameras: Cameras are one of the most well-known, cheap, and available sensors, with countless variations arising from different lenses and setups used. Fisheye, narrow and wide-angle lenses are among the most popular. Multi-camera setups can also provide depth information, as well as the velocities of objects. Cameras are a very appealing choice for the localization task because of their extremely high-resolution output, robustness to noise, and small form factor. On the other hand, their precision may be hindered by weather conditions, and require significant computational power on the edge for the computer vision algorithms.

2) Satellite-based Radio Sensors: Global Navigation Satellite Systems (GNSS), rely on satellites to estimate the exact

This article has been accepted for publication in IEEE Transactions on Intelligent Vehicles. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIV.2022.3192102

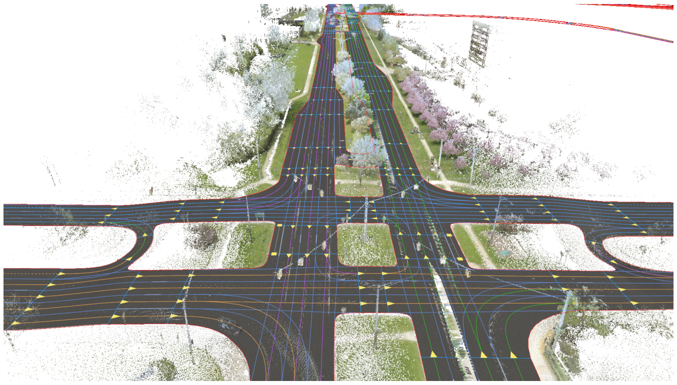


Fig. 1. A high definition map image including all layers illustrating the complexity of collected and processed data (Image source: HERE Technologies).

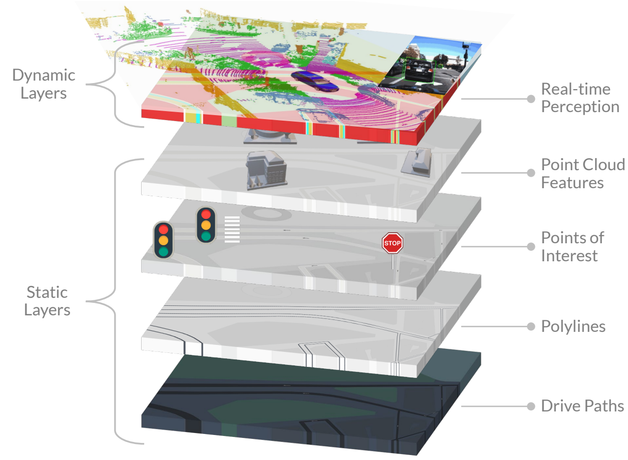


Fig. 2. Layer Stack in High Definition Maps. Only static layers are stored in the prior maps.

location of the vehicle. The most popular among them are the Global Positioning System (GPS) and the Galileo. The sensor receivers present an economic and popular solution, albeit their limitations of signal blockage and multipath render them unreliable and hinder their accuracy (~5 - 10m). Enhanced versions of GPS systems such as Differential GPS (DGPS), Assisted GPS (AGPS), and Real-Time Kinematic (RTK), can significantly improve this accuracy. DGPS exploits information not only from the sensor carried by the vehicle, but also from a fixed position infrastructure, and corrects the estimated position of the vehicle by calculating the difference between the estimated position of the fixed infrastructure and its beforehand known location. AGPS extracts the knowledge and relies on the availability of cellular networks to increase its availability and accuracy. Finally, RTK-GPS is more accurate than the two aforementioned systems leveraging dual-frequency GPS receivers and a base station with a prior known position. Its accuracy reaches centimeter-level, hence it can be used for vehicle localization. However, GPS systems need a direct line of sight with a clear sky to operate unerringly, namely, there will be signal loss in a tunnel, in a canyon, or even in a dense urban environments.

3) Inertial Measurement Units (IMU): An inertial measurement unit is a sensor comprised of an accelerometer, a gyroscope, and a magnetometer. With this setup, IMU can measure different quantities like velocity, acceleration, direction, specific force, and angular rate. Initial vehicle position knowledge can provide IMU the capability of a successful and precise vehicle ego-motion and position tracking in a GPS-challenged environment through dead reckoning methods. However, the main problem of IMUs is that their errors are accumulated over time and in consequence, the vehicle fails to localize itself precisely for long periods of time.

Table I lists all the aforementioned sensors, briefly presenting the benefits and drawbacks of each one. As it can be inferred, no sensor is perfect when it comes to the localization of autonomous vehicles, hence there is an urge to use those sensor modalities cooperatively. This can be achieved by sensor fusion, which combines sensory data from various sources, alleviating a great amount of uncertainty due to environmental conditions, or sensor operational weaknesses. Kalman Filters (KF) [16] are widely used when sensor fusion is necessary. Since autonomous vehicle's operating environment is highly non-linear, the Extended Kalman Filters (EKF) [17] and Unscented Kalman Filters (UKF) [18] can exploit Taylor expansion and unscented transform respectively, to linearize their highly non-linear operational space, and perform adequately the sensor fusion process.

IV. Leveraging Prior Maps for Achieving Centimeter-level Localization Accuracy

A. Prior Maps

The prior maps currently used for autonomous vehicle localization can be considered an enhanced version of the older digital maps which were introduced to support Advanced Driver Assistant Systems (ADAS). Their use in its current form can be traced back to the first Google car where a detailed map was recorded in a prior manual drive utilizing a roof-mounted laser scanner [19], and the Bertha Benz experimental vehicle [20], where such a map was the critical component for obtaining a comprehensive understanding of the complex traffic situations. The term High Definition Maps was introduced more recently and is currently widely accepted by the industry to characterize prior maps used for autonomous vehicle localization. An HD Map encapsulates static or even quasi-static information about the vehicle's operational environment with a high degree of precision and resolution, as shown in Fig. 1. It is synthesized from multiple source data and consists of multiple layers, such as the road model, the lane model, and point of interest and feature layer models [21], as illustrated in Fig. 2.

More precisely, the road model layer represents the road network in terms of driving paths without any additional information about lanes, traffic directions and signs. A layer above, the lane model is composed of the following parts: a precise geometrical lane model, lane attributes, traffic regulations, and lane connectivity. Finally, the maps are enriched with several more layers encoding feature-based and dense information and are being recorded in prior manual drives utilizing survey vehicles. Feature-based models are layers that reduce the operational space to a feature one from images or

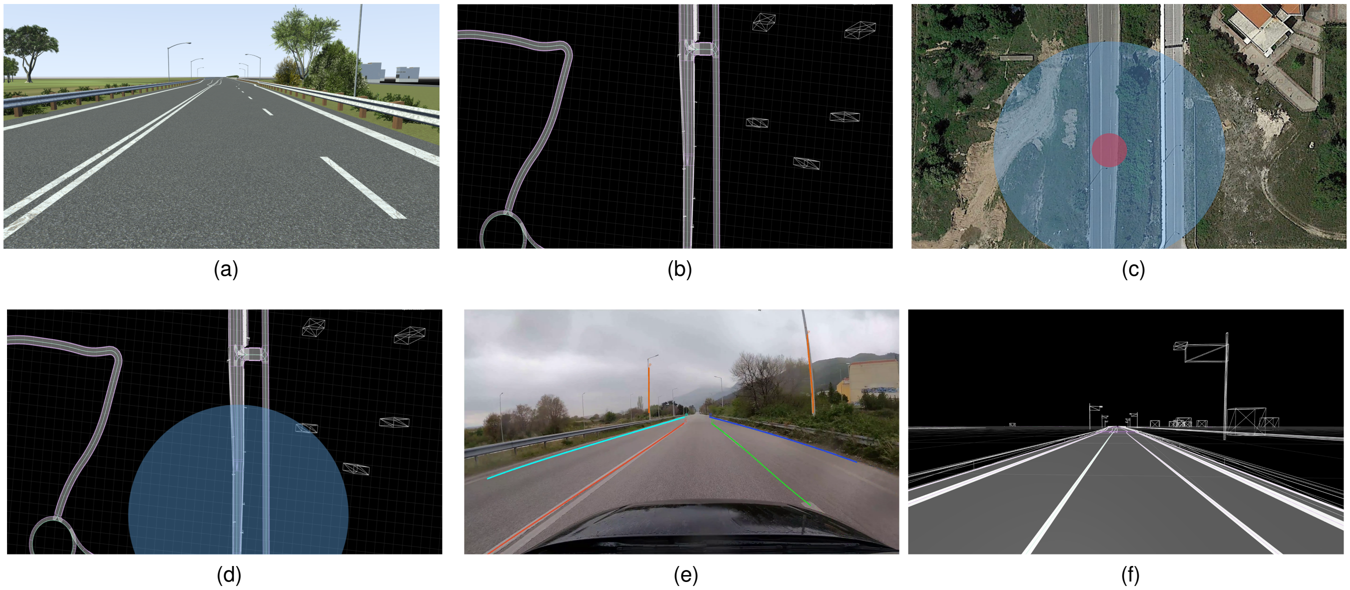


Fig. 3. Overview of the map-based localization pipeline. a) Map creation based on survey vehicle's off-line data; b) On-board stored HD map in OpenDRIVE format; c) Red circle: Initial estimation of the vehicle position based on IMU and GPS. Blue circle: Sensor perception range; d) Only the HD map landmarks within the blue circle range will contribute to the localization process; e) On-board live landmark detection; f) Matching process by associating the detected landmarks (Fig. 3e) to the prior map that contains the georeferenced road landmarks such as road lanes and poles.

TABLE II
AVAILABLE SOFTWARE SOLUTIONS FOR BUILDING CUSTOM HD MAPS

Software	Open Source	OpenDRIVE	LaneLet
RoadRunner [22]	✗	✓	✗
LGSVL [23]	✓	✓	✓
SUMO [24]	✓	✓	✗
MapToolbox [25]	✓	✗	✓
Orbit GT [26]	✗	✗	✗
DRIVE Map [27]	✗	✗	✗
VIRES VTD [28]	✗	✓	✗
TopoDot [29]	✗	✗	✗

Tick: feature satisfied. Cross: feature not satisfied.

extracted 3D landmarks, in contrast to dense models, which are created using sensors that introduce raw sensor information from RGB cameras, LiDARS, and RADARs, by scanning the environment and then saving the data in several outputs.

Navigation and planning tasks can also benefit from HD maps from additionally added dynamic layers to the already defined stack, containing information such as the location of other vehicles, pedestrians, and general data regarding traffic conditions. This amplified version of HD Maps is known as Local Dynamic Maps (LDM). They include additional layers of information, that contain transient dynamic data such as information about the weather, traffic light timings, and congestion levels, and layers incorporating highly dynamic data attached to Vehicle-to-Everything (V2X) communications, like the location of Road Side Units (RSUs) and other road users with such communication capabilities.

The HD maps' high resolution is the key component that allows localization algorithms to achieve the centimeter accuracy that is necessary in autonomous driving. For building such maps multiple data collection runs are performed a priory and the raw sensor data from several sensors is processed, transformed, joint and aligned mainly using SLAM algorithms and loop closure techniques [30], [31] for building the necessary consistent layers [32]. Several mapping and self-driving companies are producing such maps, however, currently there is little standardization between the various providers posing one of the future challenges in this field [33]. The ability however, to achieve these levels of accuracy does not come without a cost. The size of prior maps is exponentially increased in relation to higher accuracy requirements demanding more on-board storage and processing power. This results in the need for lighter maps based on simplified models yet, without sacrificing their accuracy levels. Additionally, since the road network and its properties is not static over time, the HD maps should be constantly updated over some period of time, rendering them a costly and difficult task [34].

For representing maps, several formats have emerged with the most popular ones being the Lanelet-based format, such as the Lanelet2 [35] which is more prominent in academia, and the OpenDrive [36] format which is prevalent in the industry. Due to their popularity there are tools available for conversion between the two formats [37], [38]. Currently, there are several available tools for building HD maps, with most of them being proprietary. A list of such available software tools is presented in Table II. Apart from the survey vehicle's collected data, aerial imagery and Geographic Information System (GIS) data can be also used for enriching or improving the accuracy of the created maps.

B. Map-based Localization Pipeline

Initially, the necessary HD maps are created utilizing specialized software for including all the georeferenced map attributes and localization layers from the survey vehicle's collected data, as shown in Fig. 3a. The map can be then loaded on-board or can be shared and updated through the evolving 5G/6G V2X technologies [39] using one of the available formatting standards, as shown in Fig. 3b. An initial location estimation is provided by the vehicle's GNSS and odometry sensors and is used as a reference point for selecting the HD map's landmarks that will contribute to the subsequent matching process. Since the matching process is the most computational intensive task, a search in a subpart of the map allows to achieve the necessary and critical real-time performance. An example of an initial GPS location estimation and the HD map's range of which its landmarks will be utilized are shown in Fig. 3c and Fig. 3d, respectively.

The perception stack will then utilize various perception and scene understanding techniques depending on the on-board sensors such as object detection and recognition [40], [41], semantic or instance segmentation [42], [43], panoptic segmentation [44] for detecting and localizing the landmarks in the vehicle's coordinate frame. Along with sensor fusion techniques [45], a complete vehicle's environment reconstruction can be achieved [46]. However, for accurate results, a static extrinsic calibration procedure should have been performed beforehand for estimating the rigid body transformations between all the sensors [47]. Additionally, self-calibration algorithms can be deployed for ensuring the correction of the calibration parameters based on the up-to-date sensor readings during the autonomous driving [48], [49].

The map matching step will then try to associate the detected, by the perception stack, landmarks, as shown in Fig. 3e, to the map elements of Fig. 3f. The task of the map matching can be seen in its general form as a method that optimizes N poses simultaneously between the detected landmarks in the vehicle's coordinate frame and the corresponding map landmarks of the map's coordinate frame through a cost function $J = J^o + J^l + J^p$, accounting for the odometry (o), the lanes (l), and the poles (p) of the general example of Fig. 3. The final state vector $\boldsymbol{p}^*$ which will allow to localize the vehicle can be obtained by optimizing the overall cost:

$$\boldsymbol{p}^* = \arg\min_{\boldsymbol{p}} \sum_{i\in\{o,l,p\}} \sum_{k=1}^{N} J^i(\boldsymbol{p}_k, \boldsymbol{z}_k^i, \boldsymbol{m}), \tag{1}$$

where $\boldsymbol{z}_k^i$ are the detected landmarks of the i class for pose $\boldsymbol{p}_k$ and $\boldsymbol{m}$ is the HD map.

Unlike map matching algorithms, the localization of the vehicle in probabilistic techniques, in its general form, is performed by forming a sufficiently accurate belief of the pose in a probabilistic way. This is achieved by estimating the probability distribution of the pose $\boldsymbol{p}_t$, for time step t on the condition that, along the map, all previous observations and vehicle controls are given.

$$bel(\boldsymbol{p}_t) = p(\boldsymbol{p}_t | \boldsymbol{z}_{1:t}, \boldsymbol{u}_{1:t}, \boldsymbol{m}), \tag{2}$$

where $\boldsymbol{z}_{1:t}$ are all the detected landmarks, $\boldsymbol{u}_{1:t}$ the controls of the vehicle, and $\boldsymbol{m}$ is the HD map.

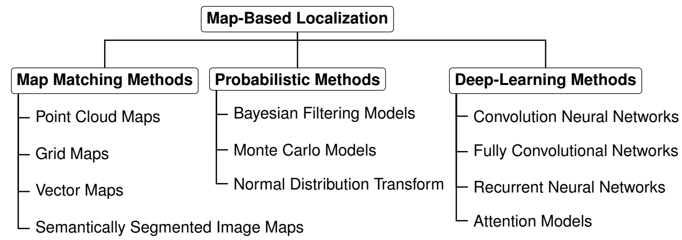


Fig. 4. Classification of map-based localization approaches based on their employed core model.

V. Localization Techniques

Considering the vastness of the localization problem and its severe effect on the performance of the whole overlying software stack leads unsurprisingly to numerous approaches trying to tackle it. In this paper, those works are classified based on their perspective on the mapping and the core localization process, as shown in Fig. 4. More precisely, those categories are map matching techniques, where aim to match prior maps with on-board observations from different sensor modalities, probabilistic techniques, where the localization problem is considered as a probabilistic one, and ultimately deep learning techniques, where the core localization task is handled by leveraging the emerging domain of deep learning. Each of the aforementioned classes is then decomposed in subclasses.

A. Map matching Techniques

Map matching techniques aim to match prior maps with on-board observations obtained by different sensor modalities, as shown in Fig. 5. LiDARs and cameras are the most popular sensors used in map matching techniques, but there are also cases where other sensor setups are used, like RADARs or even ultrasonic sensors. Map matching is most commonly executed by exploiting rigid registration algorithms such as Iterative Closest Point (ICP) [50] or Normal Distribution Transform (NDT) [51] when the data are in a point cloud format. Rigid transformation algorithms are very precise but require dense points thus they require high on-board computational power and disk storage. For this reason, only a proportion of points are used for the localization process. Point clouds are filtered, retaining only points with high information capacity, hence points depicting features, such as building facades, poles, curbs, road markings and lane markings. In this category of localization techniques the LiDAR sensor is the prominent selection for the mapping and localization process because of its precise range information and robustness to illumination changes.

Authors in [52] utilized a Mobile Mapping System (MMS) to acquire LiDAR scans from an urban environment of $2.8\ \text{km}^2$ and created an off-line feature map containing 1390 pole-like objects and 2006 building facades. During their experiment validation, the same vehicle equipped with a lower resolution laser scanner drove through the already mapped environment. The features extracted were matched with the prior feature map using a devised local pattern matching algorithm, abandoning the commonly used method of the Nearest Neighbor

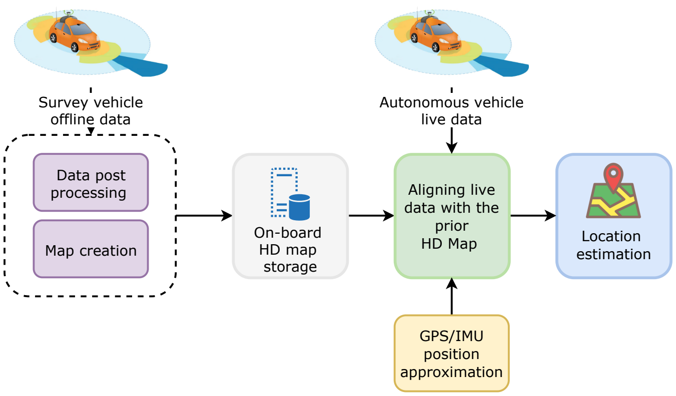


Fig. 5. Generic map matching localization pipeline with the following modules: map creation, storage, and location belief generation from map alignment.

(NN) approach. This technique reduced the rate of false positive matches from about 80% with NN approach to only about 10% since positions with low initial accuracy are likely to get wrong matches. Their experiments showcased an 89% of correct feature matches on their pre-recorded dataset with real-time capabilities, however the method would be insufficient in rural environments where features are sparse. Authors in [53], also used an MMS carrying a LiDAR for the mapping process and extracted features from the constructed map, but had a different philosophy on feature selection and on scan registration. They devised a multi-modal method, based on a proposed feature quantity to reliably and efficiently extract the scan area for localization. The quantity and nature of the extracted features were completely depended on the operation scene. For example, in high traffic urban areas buildings were chosen for feature extraction, because lane markings, traffic signs and road curbs are obscured most of the time. In those cases upper layers of the scans were utilized since the lower layers were disturbed by noise. On contrast in expressway situations they chose lane lines and traffic signs as the main features which were addressed in lower scan layers. The ICP algorithm was chosen for the registration between the live scans and the feature map. The reported accuracy in terms of average error with the best setting for the given environment were 0.438 m while the average registration time was 291.6 ms. The main advantage of this work is that it is evaluated under a lot of different environmental situations. On contrast the accuracy and the inference time of the method should be improved in order to be considered adequate for autonomous driving. In [54] a high performance MMS system was employed to map the surroundings of the autonomous vehicle. The authors devised a novel object detection algorithm to extract objects from the LiDAR data. With that approach, the size of the HD map was drastically reduced. The same algorithm was applied also on the live scan data, and as a final step the live data was registered against the prior map. The evaluation resulted to an average error around 0.30 m. The authors mention some improvements that can be implemented, such as evaluation on different environmental scenes other than urban and an increase in accuracy. Nevertheless, the method is faster than the other state-of-the-art methods based on keypoint selection.

Autonomous vehicles in urban environments suffer from localization failures due to poor or extreme illumination, bad weather conditions and appearance changes. For this reason, authors in [55] based their localization framework on semantic features found in urban scenes such as road markings, traffic lights, signs and poles. Since the repeated structures, and misdetections of such features, hinder the data association, a novel robust data association method was proposed based on local structural consistency, global pattern consistency and temporal consistency. In more detail, the contributions of this work beside the data association algorithm were a localization algorithm based on visual semantic features and a lightweight LiDAR prior map without the need of highly-accurate height information, and a factor graph optimization framework which tightly couples visual semantic measurements and odometry measurements. The evaluation was extensive and resulted to an accuracy of 0.43 m longitudinal average error, 0.12 m lateral average error and 0.11 deg average yaw error. The proposed method despite its extensive and concrete evaluation and its its sub-meter accuracy, has room for improvement concerning its inference time due to complex calculations. Also its robustness can be improved by the detection of more semantic features.

In [56] the ICP algorithm was also selected to register live LiDAR scans with the prior map. The map in that case was also height divided, because in different heights, different features of the surroundings are addressed. In dense urban environments higher scans are utilized due to their noise-free information, while lower scans sense a plethora of moving and dynamic objects, rendering their data noisy. In contrast, highway environments are featureless regarding the higher scans, but lower scans can provide sufficient information for the localization process. The key idea of this work is to register every live scan layer separately with its counterpart on the prior map, and finally fuse the registration results to obtain a robust localization result. Also a lane marking boundary database was created and used for the lower layer scan registration process. Their results, with a multiple layer matching with lane matching, showcased an average longitude error of 0.094 m, lateral error of 0.129 m, and yaw error of 0.292 deg. Despite the autonomous vehicle-adequate localization accuracy and the improvement of lateral position error, there are some improvements that can be implemented as mentioned by the authors. The first one is about further evaluations under different environments and weather conditions, and the second one is to utilize a conventional GNSS system and not an RTK one to assist their framework. The absence of static features in urban environments was also a concern for authors in [57], where localization experiments were conducted in a dense urban area. Road markings do not pose a good feature candidate for dense urban localization, because they are obscured most of the time during traffic. To surpass this issue, a custom vertical building corner extraction method was devised, exploiting data from LiDAR scans, since buildings are static and visible regardless the traffic conditions and are unlikely to radically change over time. In this work, also the ICP was selected as the registration algorithm between the

TABLE III
MAP MATCHING LOCALIZATION TECHNIQUES OVERVIEW

Technique	Prominent Sensor	Map Type	Experimental Environment	Reported Accuracy
Schlichting *et al.* [52]	LiDAR	Feature map containing pole-like objects and building facades	Urban	89% correct matches
Yoneda *et al.* [56]	LiDAR	Height-divided multi-layered feature map	Urban + Highway	Euclidean error: 0.438 m, Inference Time: 291.6 ms
Cao *et al.* [60]	LiDAR	Feature map containing poles, corners and walls	Urban	RMSE: 0.094 m lon, 0.129 m lat, 0.292 deg yaw
Im *et al.* [57]	LiDAR	Feature map containing extracted vertical building corners	Urban	2D RMS horizontal error: 0.138 m
Yoneda *et al.* [53]	LiDAR	Feature map containing mixture of features	Urban + Highway	RMS translation error: 0.12 m, Average angular error: 0.03 deg
Akai *et al.* [61]	LiDAR	Feature map containing road markings	Mountainous Rural Area	Average Distance Error: 0.102 m in rural area and 0.152 m in urban
Steinke *et al.* [62]	LiDAR	Feature map containing pole-like features extracted from Open Datasets	Urban	RMSE: 0.38 m lon, 0.08 m lat
Rohde *et al.* [58]	LiDAR	Feature map containing mixture of features	Urban	0.073 m lateral average error
PoseMap [63]	LiDAR	Feature map containing mixture of features	Urban	No specific accuracy results
Wolcott *et al.* [64]	Camera	Synthetic images generated from filtered point cloud map	Urban	RMSE: 0.19 m lon, 0.14 m lat
Lu *et al.* [65]	Camera	LiDAR road marking feature map	Urban	RMSE: 0.239 m lon, 0.595 m lat, 0.84 deg yaw
Kim *et al.* [66]	AVM Camera	Lane line map	Urban	RMSE: 0.26 m lon ,0.072 m lat
Asghar *et al.* [67]	Camera	Lane line grid map	Highway	Mean Absolute Error: 0.8 m lon, 0.25 m lat
LaneLoc [68]	Stereo camera	Feature map containing curbs and road markings	Track + Urban + Rural	Recidual accuracy of around 0.10 m. Missing ground truth.
Wu *et al.* [69]	Stereo camera	Feature map containing road markings	Highway	RMSE: 0.24 m lon, 0.13 m lat, 0.014 rad yaw
Ward *et al.* [70]	RADARs	RADAR point cloud map	Rural	RMSE: 0.37 m lon, 0.073 m lat
Cornick *et al.* [71]	LGPR	RADAR point cloud map	Highway	Euclidean error: 0.073 m
Schlichting *et al.* [72]	LiDAR	LiDAR point cloud map	Urban	RMSE: 1.4 m
Xiao *et al.* [73]	Camera	Vector map	Urban	RMSE: 0.24 m
Nagy *et al.* [54]	LiDAR	LiDAR feature point cloud map	Urban	Average error: 0.30 m
Wang *et al.* [55]	LiDAR	LiDAR feature point cloud map	Urban	Average error: 0.43 m lon, 0.12 m lat, 0.11 deg yaw
Petek *et al.* [74]	Camera	Lyft5 dataset	Urban	RMSE: 0.61 m lon, 0.25 m lat, 1.30 deg yaw
Hungar *et al.* [75]	LiDAR	LiDAR feature map	Urban	Absolute average error: 0.42 m
Pauls *et al.* [76]	Camera	Semantically segmented images	Urban	RMSE: 0.90 m lon, 0.11 m lat, 0.56 deg yaw

live scans and the pre-recorded map database. The accuracy translated as 2D Root Mean Square horizontal error was 0.138 m. The main advantages of this method are its great accuracy, that can be used when the roads are congested with traffic and that it relies on low size map. Nevertheless, a method for extracting more corners and with higher accuracy should be investigated, and the framework should also be evaluated under a larger and more verstile dataset in order to prove its robustness. In [58], the authors also tested their method in a dense urban environment with numerous dynamic obstacles and objects. They presented a novel pipeline, using Fourier-Mellin transformation to create a spectral prior map of the environment, and then register also the transformed live LiDAR scans on it. The initialization of the pipeline is provided by the approximated global pose estimation, provided by a consumer-grade GPS. Furthermore, a LiDAR-based odometry method was utilized, to enhance redundancy in vehicle motion estimation, and to provide one more estimation regarding vehicle's position. These two position approximations were combined in a Kalman filtering framework to minimize the uncertainty of the localization process. As a final step, the pipeline utilizes the cumulative sum test [59] to perform consistency checks and reject the inconsistent map matching results. The pipeline poses an accuracy of 0.12 m RMS translation error and 0.03 deg angular error. The framework achieved great accuracy results compared to other state of the art frameworks. Despite that, it is only tested in an urban environment with a small dataset and its inference time is not mentioned.

Authors in [60], tackled the complex urban localization problem by constructing a prior map with three different features, namely pole-like objects, walls, and corners. Pole-like objects such as trees and street lamps are not dense enough for data association but can be found in great numbers in urban areas. On the contrary, walls and corners provide dense information and are easily and reliably detected by LiDAR sensors. Live scan registration with the prior map was conducted by a novel pattern matching method and instead of traditional filtering techniques, a graphed-based optimization method is employed. Since the scenery in an urban environment can change rapidly, the map should be updated repeatedly to ensure the localization accuracy. There were conducted experiments in two separate test roads. One of them was surrounded by a rural environment and the proposed framework posed an accuracy of 0.102 m in terms of average distance error. In contrast, the other test road was surrounded by an urban environment, and large buildings inferred with the GPS signals dropping the accuracy of the framework to 0.152 m in terms of average distance error. The main advantages of this framework are its low cost comparing to other state of the art works, utilizing only a 3D LiDAR, its repeatability and high accuracy and its $100hz$ inference time. Nevertheless, it is only evaluated in feature-rich environments, using a small dataset of two trajectories. Authors in [63], proposed and tested in a period of eighteen months a long term localization framework where instead of a global map, feature bundles assisted the localization process. Each feature bundle is associated with a pose of the vehicle and can be updated dynamically considering changes in the environment and unexplored areas. This technique can be applied only with LiDAR sensors, since these extracted features are sparse and only LiDAR's large field of view can detect a significant amount of them. In order to make the localization procedure more robust, the authors also applied a SLAM solution to localize against the local history. Experiments were conducted to both urban and rural areas to prove the robustness of the method. There are no reported results in terms of accuracy, but the experiments were promising and claimed adequate localization accuracy at all cases. The evaluation process of the aforementioned framework is its core advantage. It is evaluated using two different platforms, with a large dataset, in a mix of environments for 18 months. However, the authors mentioned that its inference time can be improved by detecting and filtering out dynamic objects, trying to keep only static elements in the data.

LiDAR's data density was a concern for the authors in [72]. For this reason, their framework utilized only one 2D laser scanner with scan lines derived from a reference point cloud of the environment. As a next step, a neural network was employed to extract features describing the line shapes. Every line was fed to a k-means algorithm and its assigned cluster-id was stored in a reference graph. The localization is done via a sequence mining approach, where a sequence with a specific length is matched to the position with the highest correlation in the reference sequence. The evaluation showed that the algorithm has an accuracy of 1.4 m in terms of RMS error and a completeness of 99%. This very low cost framework with no GNSS information managed to localize successfully an autonomous vehicle with very few data. Its accuracy though can be improved by employing a change detection algorithm to reduce the errors and also utilize the intensity data values given by the LiDAR scans. In [61], an autonomous transportation system for the elderly was created for the rural mountainous areas of Japan. For the purpose of localization a LiDAR sensor was used to create a road marking map. Initially, a coarse localization is performed with a GPS and an IMU, and afterwards the position estimations were fused in an EKF setup, together with the position approximation extracted from the prior map, to achieve a centimeter level localization. The live LiDAR scans were registered with the prior map, exploiting the NDT algorithm. The experiment results showcased an accuracy of 0.38 m and 0.08 m in terms of average absolute longitudinal and lateral error respectively at average velocities of $45km/h$. Even though the framework achieves robust localization by fusing the NDT pose estimation with the dead reconing results with EKF even when NDT is disturbed, its accuracy is not autonomous driving-adequate. For this reasons the authors are currently developing a vehicle localization failure detection method to improve it. Authors in [75] also exploited LiDAR-detected features from the environment to assist their localization process. In contrast to the other methods in the literature, they selected non-semantic features and not landmarks like lane lines, road curbs etc. Their approach consists of three steps. At first they collect data and build their HD map, by extracting the aforementioned features from them. Afterwards, the live data scans are matched with the prior data to acquire a first position estimation. Finally, this result is fused together with measurements from odometry and GNSS to build a factor graph. The optimization of this factor graph produces the final pose estimation of the framework. The evaluation of the framework was executed by traveling the same trajectory for one and a half year to prove that its accuracy is unaffected by changes in the environment due to season changes. The accuracy of the method in terms of absolute average error was 0.42 m. The main advantage of this approach is that it is not relying on static landmarks in the operation environment, yet it is tested on a relatively small dataset, its accuracy needs to be improved in order to be considered as autonomous driving-adequate and as the authors mentioned the map generation process should be improved.

Mapping process is always a concern in the localization process since it is a complex and time-consuming procedure. For this reason authors in [62], used map data from Open Datasets for the city of Berlin. They chose features such as traffic signs, traffic lights, trees, walls, and corners since they are rigid to change during time, and also smoothly detectable by LiDAR sensors. Matching between live data and the prior map was performed by a custom geometric fingerprint algorithm, and the extracted pose estimation was fused together with the estimations generated from an IMU, a wheel odometer, and a GPS to assist the localization process in GPS-denied areas. The results pose a 0.073 m accuracy in terms of average lateral error in real world scenarios. The main advantages of this method are that there is no need to devise and execute a mapping process, it was tested in real world scenarios with an inference time of $44.7ms$ and finally there is no need for GNSS feed. However, an issue is that these datasets may be several years old, hence they are not updated.

While LiDARs provide dense information and their performance is invariant of environment's lighting conditions, are expensive and chunky sensors, increasing the cost of autonomous vehicles. For that reason, many researchers employed cameras to solve the localization problem. Cameras are among others, lightweight, small in size and affordable. Authors in [64], performed localization within a LiDAR point cloud map using only a monocular camera. LiDAR map, which was also augmented with surface reflectivities was build beforehand using pose-graph SLAM and the Generalized-ICP (GICP) algorithm [77]. Initially, synthetic images were generated from the LiDAR map with the assist of a consumer grade GPU around an estimated region obtained by GPS. Those synthetic images were compared with live camera frames by extracting their Normalized Mutual Information (NMI), providing a pose estimation. Finally, the pose provided by an IMU and the aforementioned map matching technique are fused in an EKF setup. The proposed method was compared with other state-of-the-art LiDAR-only automated vehicle localization algorithms. The comparison showcased that with a much cheaper sensor it achieved a closely similar error rate. To be more precise, in an urban environment the longitudinal RMS error was 19.1 cm and the lateral 14.3 cm. Even though the mapping process is complex in terms of computation, and the pipeline relies on GPU, the framework shows great potential because it relies only on a monocular camera. Moreover, it is evaluated on real world scenarios and the results are close to the state of the art but with a much cheaper prominent sensor. Also in [68], the map used for the localization process was created before hand by exploiting LiDAR sensors, and a variety of other sensors such as GPS, and IMU. Curbs and road markings were extracted to reduce the highly dense map to a lightweight feature map. In order to localize the vehicle, a stereo camera system and a map matching algorithm were utilized. Finally, the pose estimation extracted from the map matching algorithm was fused together with an estimation from an IMU in an EKF setup to acquire the final pose estimation. It is highlighted that the GNSS estimation was used only for the initialization of the localization process and it was not used further. The authors executed a long term evaluation of their framework on approximately 50 km of rural roads achieving a mean residual accuracy of around $10cm$. As an improvement the authors stated that they will work on improving the precision of the mapping process. They plan to replace the Velodyne LiDAR with a new camera system in order to benefit from the denser amount of information that it will introduce. Wu et al. [69] utilized stereo cameras as well to localize the vehicle using a feature 2D-NDT map containing traffic signs painted on the road, such as arrows, pedestrian crossings and speed limits. Corner points were extracted from these features, to reduce both space and computation power required to register the online frames to the map. Two novel methods were devised and tested for map matching, based on triangulation and linear transformation on birds eye view images, generating promising results. The evaluation was conducted in an urban environment, only on a straight trajectory, posing an accuracy in terms of RMS error of 13.7 cm concerning lateral position, 24.5 cm concerning longitudinal position and 0.014 rad concerning orientation angle. Despite its high accuracy, the authors mention that there are some improvements that can be implemented. More specifically, the positioning performance can be improved by accurately extracting MSER feature from occupancy grid maps, and the evaluation will be also conducted in curved roads and not only on straight trajectories. Road markings were also used by authors in [65], to assist the localization process. In more detail, they conceived a vision based approach, based on Chamfer Distance to match online detected road markings against a lightweight 3D map, constructed by LiDAR sensors beforehand. A non-linear optimization problem was formulated to estimate the final 6-DoF camera pose, taking into account also the vehicle odometry and the epipolar geometry constraints. The evaluation in an urban environment resulted on an accuracy in terms of RMS error of 0.239 m regarding longitude, 0.595 m regarding latitude and 0.84 deg concerning heading. The drawback of the framework is that it is not evaluated in environments with sparse landmarks. However, it relies in a lightweight map, and it achieves great accuracy utilizing data collected months apart. A three step localization framework was proposed in [66], based on the data introduced by an Around View Monitoring (AVM) camera. In contrast with the previously presented technique, authors used a feature map with lane lines as features. The real time localization algorithm consists of three steps. At first, the AVM camera detects the lane lines, and then ICP algorithm is utilized to provide a rigid transformation between the lane map and lanes obtained by the AVM camera. At last, the extracted pose estimation is fused together with pose estimations from vehicle's sensors utilizing an EKF setup. For higher accuracy, the covariance of the ICP is estimated using Haralick's method. Tests were performed on the Korean Automobile Testing and Research Institute and the ground truth map was generated using an RTK-DGPS. The reported lateral error was 0.072 m and the longitudinal error was 0.26 m. The proposed method was not tested under urban driving conditions, hence the authors are planning to test it in an urban environment. Also they are planning to conduct evaluation experiments under challenging situations such as low-visibility conditions, lack of longitudinal position correction over a long period and increase of non-linearity due to unusual maneuvering, and so on. ICP algorithm and a lane-level grid map was also employed by authors in [67]. The novelty of this method is the usage of two distinct map matching algorithms, each providing a distinct pose estimation. One algorithm is based on the ICP algorithm and the other on a decision-rule based approach to perform topological map matching. The final pose estimation of the vehicle was calculated by an EKF fusing both the estimations with GPS and dead reckoning. The framework was evaluated under real driving conditions and showcased a lateral accuracy of 0.25 m and 0.8 m longitudinal accuracy in terms of mean absolute error. Finally the authors mention some improvements that can be implemented. More specifically, they plan on employing an RTK GPS to define an HD map with well documented curved sections of the road and more detailed lane markings to further improve the accuracy and robustness of the localization approach. At last they will also add lane change assist to support the framework n situations where the vehicle overtakes or changes lane. In [73] the authors also avoided high-cost sensors, and devised a localization framework with only one camera. They have constructed a prior vector map and during the localization process registered live camera data against it. Semantically significant features were delicately extracted from the camera frames, to ease the alignment and the resiliency of the method. The evaluation was conducted on a trajectory on an urban road and resulted to an accuracy of 0.24 m in terms of RMSE. The method's advantage is that it utilizes low-cost sensors and lightweight prior maps with the trade-off of smaller accuracy in rather simple driving scenarios.

A monocular localization system was also proposed by authors in [74]. The system incorporates predicted uncertainties into a pose graph optimization framework. The uncertainties retain the robustness in challenging urban city scenes utilizing only sparse map features. In order to predict those uncertainties a novel multi-task uncertainty estimation method was developed. This method not only learned meaningful uncertainties for semantic segmentation, but also simultaneously detected objects in a single pass. The evaluation was conducted on the Lyft5 dataset, and concluded to an accuracy of 0.61 m in longitudinal direction, 0.25 m in lateral direction and 1.30 deg in heading direction in terms of RMSE. The current framework achieved a high level of accuracy with extremely sparse HD maps and low cost sensor setup in normal driving environments. Semantic segmentation and a monocular low cost setup were also utilized by authors in [76]. In more detail, they proposed a method based on an HD map built from semantically segmented images, inferred by a neural network. The challenges of data association and misdetections were solved by applying a distance transform on binary per-class images and a sliding-window pose graph optimization respectively. The evaluation showed that one of the most lightweight and flexible methods can achieve an accuracy adequate for autonomous driving. More specifically, the accuracy in terms of RMSE was 0.90 m in longitudinal direction, 0.11 m in lateral direction and 0.56 deg in heading direction. The authors are planning to implement some improvements with the most notable of all is the inclusion of a more advanced

This article has been accepted for publication in IEEE Transactions on Intelligent Vehicles. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIV.2022.3192102

motion model and GNSS information. They also plan to store all relevant traffic signs and single lane markings that will yield additional information in longitudinal direction. Finally, employing a better, but slower neural network will increase the accuracy of the framework.

Cameras have several advantages, however are also sensitive to lighting changes and often their performance is poor during night time operations. For that reason, several researchers utilized RADARs to perform the localization process, since RADARs are affordable, compact formatted and robust to adverse weather conditions. Authors in [70], localized their vehicle inside a RADAR map created before hand using low cost RADAR sensors. ICP algorithm was leveraged to register the live scans with the prior map. The final pose estimation was extracted by fusing the pose estimation from the aforementioned technique and the pose estimations from the other sensors carried by the vehicle such as GPS and IMU. For the evaluation a trucked equipped with the necessary sensors traveled an already traveled and mapped trajectory. The accuracy of the framework in terms of RMS error was 7.3 cm concerning lateral direction and 37.7 cm concerning longitudinal direction. The accuracy of the method is not autonomous driving-adequate, yet it used far less data and fewer computation that the other state-of-the-art methods. Finally, a unique approach proposed by authors in [71], where a vehicle carrying a Localizing Ground Penetrating RADAR (LGPR) created a map of the underground. Map matching is executed by a correlation maximizing optimization technique generating promising results in a highway environment. The advantage of LGPR is that senses a generally stable environment, however, it is not tested in an urban environment, where RADAR's sensitivity to noise may result in poor performance. The evaluation was conducted on a highway environment reporting an accuracy of 7.3 cm on terms of RMS. The main advantage of this method is that it is completely unaffected by weather conditions. However, the sensor setup is chunky and it is not possible to fit under every vehicle, and the framework is only tested in highway environment. Finally, the authors plan to fuse the position estimation of the LGPR with GNSS position estimation, and evaluate if the LGPR can penetrate underground waters.

To summarize, map matching techniques are well established in the state-of-the-art concerning the localization process. Their core benefits are minimal inference time, requirement of no complex computations and easy implementation. On the other hand, they require dense information in order to be robust and avoid failures. This happens because they cannot handle uncertainties. They also have high disk storage requirements, especially if there is no feature extraction from the raw data. Finally, they are not recommended for challenging environments and for environments with absence of features. In those cases there will be a large number of false positive registrations. In Table III, a brief overview of the discussed Map Matching localization techniques is presented.

B. Probabilistic Techniques

Probabilistic techniques model the core localization task as a probability problem as shown in their generic pipeline in Fig. 6. These methods entail significant advantages such as letting the observation and motion models unrestricted, introducing scaling capabilities since they are parallel-friendly, showcasing independence of system size, and are relatively easy to implement. On the other hand, they are computationally expensive and hard to debug, requiring also good initial location approximations. The two most common approaches are the NDT method and the particle filter one [78]. The NDT transforms discrete prior map data to a continuous Gaussian space, by fitting a number of Gaussian distributions to the data points [79] [80]. On the other hand particle filtering lifts the the Gaussian assumption, handles optimally arbitrary distributions, and does not restrict the motion model to follow a specific distribution [81] [82]. It is derived from the Bayes filter theory, which exploits the Bayes rule to handle uncertainty in challenging environments. Numerous robot and autonomous vehicle localization approaches rely on particle filtering [2] [78], and in that context may the term Monte Carlo Localization (MCL) be used, since Particle Filter is in its core a Monte Carlo method. Thus in this work, the terms particle filter and Monte Carlo Localization are used interchangeably.

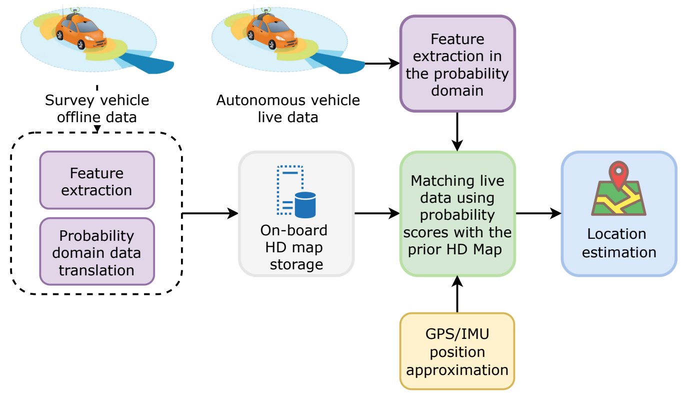


Fig. 6. Generic probabilistic localization pipeline with the following modules: feature extraction and domain translation, storage, and location belief generation using probability scores.

Both NDT and Particle filtering were utilized by the authors in [95]. They constructed a prior LiDAR road marking map with LiDAR intensity values, and devised an NDT matching based localization approach. The experimental environment was in a rural area, where absence of unique features caused NDT to fail often, hence a complementary particle filter framework approached to correct the pose estimation. The evaluation of the proposed method showed promising results in terms of average RMS error. More specifically the error in the longitudinal direction was 0.19 m and in the lateral direction 0.07 m. The localization errors were larger in open spaces where features where sparser and the NDT convergence was strongly relied on the initial pose. Road features are popular candidates for feature maps. In a series of works feature maps were constructed to assist the localization process containing explicitly road features. In more detail, authors in [91] constructed a curb map utilizing data from a LiDAR sensor and employed a robust Least Trimmed Squares method to calculate the curb parameters and for outlier rejection. The final pose

TABLE IV
PROBABILISTIC LOCALIZATION TECHNIQUES OVERVIEW

Technique	Prominent Sensor	Map Type	Experimental Environment	Reported Accuracy
OpenStreetSLAM [83]	Camera	OpenStreetMaps data	Urban	Average error: 5.19 m
Jo *et al.* [84]	Camera	Road marking map	Rural + Highway	RMSE euclidean: 0.53 m
Suhr *et al.* [85]	Camera	Road marking map	Urban	Mean euclidean error: 1.18 m
Stenborg *et al.* [86]	Camera	Semantically segmented feature map	Urban	Average error: 0.5 m
Elfring *et al.* [87]	Camera	Traffic sign map	Urban	Average error: 0.73 m lon, 0.77 m lat
Xu *et al.* [88]	Stereo Camera	Raw point cloud map	Urban	RMSE: 0.21 m lon, 0.11 m lat, 0.40 deg yaw
Wiest *et al.* [89]	LiDAR	Region landmark map	Rural	Average error: 0.05 m lon, 0.13 m lat, 0.05 deg yaw
Hata *et al.* [90]	LiDAR	Curb and road marking map	Track	Average lateral error: 0.31 m
Hata *et al.* [91]	LiDAR	Curb map	Track	Average lateral error: 0.52 m
Wolcott *et al.* [92]	LiDAR	Gaussian mixture map	Urban	Average error: 0.15 m lon, 0.10 m lat
Hata *et al.* [93]	LiDAR	Curb and road marking map	Track + Rural	Average error: 0.22 m lon, 0.27 m lat, 0.01 deg yaw
Wolcott *et al.* [94]	LiDAR	Gaussian mixture map	Urban	Average error: 0.15 m lon, 0.10 m lat
Akai *et al.* [95]	LiDAR	Road marking map	Rural	Average error: 0.19 m lon, 0.07 m lat
Suhr *et al.* [96]	LiDAR	Vertical features and road intensity map	Urban	Average error: 0.07 m lon, 0.07 m lat, 1.098 deg yaw
Weng *et al.* [97]	LiDAR	Pole feature map	Urban	Average error: 0.15 m lon, 0.057 m lat, 0.14 deg yaw
Wan *et al.* [98]	LiDAR	Occupancy grid Gaussian map	Urban	Average error: 0.032 m lon, 0.036 m lat
Ma *et al.* [99]	LiDAR	Sparse semantic map with lanes and traffic signs	Urban	Average error: 1.12 m lon, 0.05 m lat
Ghallabi *et al.* [100]	LiDAR	Lane line and road sign map	Track	Mean error: 1.49 m lon, 0.02 m lat
Ghallabi *et al.* [101]	LiDAR	Reflective objects	Track	Mean absoute error: 0.64 m
Li *et al.* [102]	LiDAR	LiDAR 2D grid map	Track	RMSE: 0.08 m lon, 0.06 m lat
Schaefer *et al.* [103]	LiDAR	Pole landmark map	Urban	RMSE: 0.11 m position, 0.21 deg angular
Schlichting *et al.* [104]	LiDAR	Mixture of features (pole-like, curbs, lane lines)	Urban + Highway	RMSE: 0.17 m lon, 0.18 m lat
Ahmed *et al.* [105]	LiDAR	Sparse point cloud LiDAR map	Urban + Highway	Average error: 0.114 m lon, 0.072 m lat
Wang *et al.* [106]	LiDAR	Double layer feature map (curbs and vertical features)	Urban	RMSE: 0.091 m lon, 0.077 m lat, 0.219 deg yaw
Chen *et al.* [107]	LiDAR	Point cloud LiDAR map	Urban	Average error: 0.81 m location, 1.74 deg angle error
Li *et al.* [108]	LiDAR	Point cloud LiDAR map	Urban	RMSE: 0.40 m lon, 0.30 m lat, 0.70 deg yaw
Kasmi *et al.* [109]	Camera	OpenStreetMaps + Lane Markings	Urban	Classification accuracy: 99% on 2-lane road
Chen *et al.* [110]	LiDAR	Pole and curb map	Urban + Highway	RMSE: 0.174 m position, 0.335 deg yaw

estimation was extracted utilizing a Monte Carlo Localization framework. The proposed method was evaluated on a track and resulted to an average lateral error of 0.52 m, including also the regions where curbs could not be detected. Despite its novelty and its high accuracy the framework requires some improvements in order to reach autonomous driving adequacy. As stated, an integration of a lane marking detection algorithm can improve the accuracy when curb features are absent. In [90], the same team of authors presented an extension of their previous work, where the LiDAR map was enriched with road markings detected by a novel Otsu thresholding algorithm, deriving an HD map which contained both road curbs and lane lines. With the same experimental setup, the accuracy of their method improved to 0.31 m average lateral error. It is clear that the addition of the lane marking information in pipeline dramatically improved the accuracy. Finally, in [93], they combined both their previous works and extended their evaluation method by also conducting experiments in a rural environment. On urban environments with high traffic, parked cars, and a variety of dynamic obstacles in the environment, road features are obstructed most of the time hindering the precision and sometimes even the success of localization process. The resulted accuracy in terms of average error on the track was 0.13 m on lateral direction, 0.20 m on longitudinal direction and 0.02 m concerning heading direction (angular). In contrast, in a rural environment the accuracy was lower and more specifically, 0.27 m on lateral direction, 0.22 m on longitudinal direction and 0.01 m concerning heading direction (angular). Finally the authors, evaluated their framework also in other environments, and they achieved promising results and high accuracy. A possible addition of wheel encoder data could improve the longitudinal accuracy. On the contrary, authors in [97] tested their particle filter-based localization framework in a congested urban environment. They devised a pole extraction framework from LiDAR point cloud data for both HD map construction and live data registration. Pole-like structures tend to stand out on such environments and as a result are great candidates for HD maps. The evaluation resulted in an accuracy of 0.15 m on the longitudinal direction, 0.057 m on the lateral direction and 0.14 deg on heading direction in terms of average error. The accuracy of the framework can be considered autonomous driving adequate, evaluated under challenging urban conditions. However, this accuracy may be hampered under scenarios where there is a lack of poles, such as large intersections or inside tunnels. A possible solution to this drawback would be to enrich the pipeline with more features such as edges of buildings and piers. The same pattern was followed by authors in [103] where they proposed an urban particle filter-based localization pipeline, assisted by a pole-like landmark LiDAR map containing a variety of features such as street lamps, traffic signs and tree trunks. In the context of that work, a novel pole detection algorithm was introduced, which also takes into account free space between the laser sensor and the endpoints. Their evaluation resulted to a reported RMS error accuracy concerning position of 0.11 m and 0.21 deg concerning heading direction. The aforementioned framework was evaluated under urban challenging driving conditions and produced high accuracy results. Nevertheless, the authors are planning to introduce two major extensions. First they plan to to fuse the separated mapping and localization modules into a single SLAM module and finally to explore how different sensor modalities will affect their pipeline. A pole-curb fusion vehicle localization system was proposed in [110], where cost functions for pole and curb were separately defined and then fused with a subsequent non-linear optimization method to obtain the vehicle location. Due to a novel curb representation and a global optimization method for tackling the data association problem of poles, satisfactory performance results were achieved in two datasets.

On the other hand, on highway environments there is an absence of pole-like features. For this reason, authors often choose other type of landmarks to create the HD map, like in [100] where a LiDAR point cloud map was constructed containing the lane lines and road signs. The localization process where handled by a particle filter approach and tested on a track without traffic and dynamic obstacles at various ego vehicle velocities. The main contribution was the named constraint update. More specific, a novel way of generating the particles, which relied on the previous lateral position of the vehicle and the road structure extracted from the map. Evaluation resulted on a mean longitudinal error of 1.49 m and lateral error of 0.02 m. The authors are considering to evaluate the framework also in more environments than highway and also add a camera sensor in the pipeline in order to integrate a visual-based localization. The core idea behind that is the reduction of the drift between two consecutive road signs. The same authors extended their work by also detecting and adding to their HD map database High Reflectivity Landmarks (HRL) with their LiDAR setup. Thus, in [101] apart from road signs and lane lines the HD map also contained lane markings, and guard rail reflectors. The matching between the live and the prior data was conducted utilizing the aforementioned particle filter framework. The evaluation resulted to a mean absolute error of 0.64 m. The main advantages of this method are its real-time capabilities due to the usage of a sparse geometry third-party map and its great accuracy results even with high velocities. Authors in [102] also presented a novelty concerning the mapping procedure. In more detail the point cloud is first translated to 3D occupancy grids and then the grids are projected to a plane perpendicular to the ground and parallel to the longitudinal direction. This produces a 2D grid map, which can reflect the depth of the road scene. For their map matching stage they employed a Monte Carlo framework, which provided the final localization estimation after fusing also data from IMU and wheel encoders. The performed evaluation proved the the robustness and precision of the method and resulted to average longitudinal error of 0.08 m and lateral error of 0.06 m. The main advantage of this method is that the storage requirements of the compressed road scene map substantially decreased compared to the raw point cloud. As future work, the authors plan to re-size the 2D grid self-adaptively based on the environment. Additionally, the grid size for a single compressed road scene map can also be adapted according to the objects in the map. Schlichting *et al.*, devised a novel LiDAR-based localization framework presented in [104]. The framework was assisted by a feature HD map containing a mixture of features, evaluated both in urban and highway environments. For the urban environments pole-like features, curbs and vertical planar object were utilized, while for highway ones lane lines and curbs. The final pose estimation was extracted from a particle filter approach, and then corrected by fusing it with the estimation from GPS/IMU in a Kalman filter based framework. The evaluation results were given as RMS error format. In the urban environment the longitudinal error was 0.17 m and the latitude error 0.18 m. Finally, the highway evaluation resulted to a 6.95 m longitudinal error and 0.12 m lateral error. As stated, a 3D solid state LiDAR and a more accurate GNSS can improve even more the accuracy and reliability of the method especially for increasing the longitudinal accuracy on highway environments. Kim et al. in [96], also tested their localization framework in multiple environments utilizing prior constructed LiDAR maps. In order to efficiently store, access and in general manage the dense information of the HD map, the authors introduced a management strategy based on road intensities and vertical features. For the final pose estimation a novel particle filter entropy based algorithm was proposed. The conducted tests proved the robustness of the approach resulting to a mean longitudinal error of 0.077 m, a mean lateral error of 0.073 m and a mean angular error of 1.098 deg. The vehicle successfully managed to perform an autonomous ride in a challenging urban environment. Nevertheless, the authors are planning to improve further their approach by fixing some communication issues and enhancing their sensor calibration techniques. Moreover, they will try to introduce a variety of sensors in their pipeline to improve the robustness. In [106], authors employed a double layered LiDAR point cloud prior map consisted of a bottom layer containing ground and curb features, and an upper layer containing vertical features. They also introduced a novel curb extraction algorithm based on collision detection, and a vertical projection based detection algorithm to detect vertical feature points. Both layers were combined and a Monte Carlo Localization framework was set up to extract the final pose estimation. The proposed framework showcased an accuracy of 0.077 m in lateral direction, 0.091 m in longitudinal direction and 0.219 deg in heading direction. The accuracy was calculated in terms of average error. The introduced double-layered feature map was extremely small in size compared with the traditional 3D point cloud maps, however it managed to produce high accuracy results. Additionally, the accuracy improved by introducing curb features into the pipeline.

A Monte Carlo Localization framework was also utilized by authors in [107], yet the main contribution of their work is a novel deep neural network approach to register point cloud LiDAR scans. More precisely, the trained network predicts the possibility of overlap between the live LiDAR scan and the scan of the pre-built HD map. The training of the model was based on the Overlap-Net [111]. The evaluation results were reported in terms of location error and yaw angle error having a value of 0.81 m and 1.74 deg respectively. The framework was compared with other state of the art frameworks and achieved similar performance, but with higher success rate and lower computational time. The datasets which the evaluation based on were collected in different seasons to prove the robustness of the approach. In order to overcome the density of LiDAR information and the challenges that it is introducing, authors in [105] constructed a sparse LiDAR point cloud map.

The main contribution of their work was a novel NDT-based Point-to-Distribution algorithm for live scan matching and the usage of sparser data. Despite the significant data reduction the accuracy and sampling rate of the framework remained competitive. The evaluation was conducted on two different rural trajectories. It resulted to a localization accuracy of 0.114 m of longitudinal average error, 0.072 m of lateral average error. The experiments also concluded that the approach significantly improved the localization and its robustness in outdoor AV environments, especially when the data of the HD map are sparse. As a great future improvement the authors are considering to introduce periodic updates of the stored map to handle dynamic environment changes. In [108] authors achieved to build a low-cost and high-speed localization framework with a decimeter-level accuracy. They constructed their map utilizing measurements from a 2D LiDAR and fused information of the vehicle's pose from high-accuracy GPS, IMU and odometry sensors. The map matching was executed utilizing a proposed hierarchical neighborhood-based method. Finally, an accelerated particle filtering algorithm was employed in order to extract the final (corrected) pose estimation by also adapting a hierarchical model point selection and logarithmic data point search, to perform the whole process in real time. The results of the evaluation conducted in an urban environment were promising, showing an accuracy of 0.40 m in longitudinal direction 0.11 m in lateral direction and 0.56 deg in heading direction in terms of RMS error. The main advantages of this method is its real-time fashion and that it is proven in different environments. However, there should be comparison with other state of the art approaches concerning map based techniques.

LiDARs are considered expensive sensors, require high computational power, and are big in size. For the aforementioned reasons numerous researchers try to avoid using LiDARs for the localization process. In [88], authors created a LiDAR point cloud map retaining also the intensity values, and later used it to localize a vehicle carrying only a stereo camera. The proposed framework produced synthetic images using the depth and reflectivity intensity values extrapolated from the prior LiDAR map. These synthetic images were registered with live camera images in a particle filter setup to obtain the final pose of the vehicle. The generated HD map was 3 years old, but the framework still managed to produce precise localization estimations even in challenging situations. The proposed method was tested in urban and rural environments and its average accuracy was around 0.21 m in longitudinal direction, 0.11 m in latitude direction and 0.40 deg in heading direction. Also in [85], an HD map was built beforehand by a survey vehicle. In this work, the authors exploited road marking depicted in the prior map to devise a localization framework utilizing low cost sensors, a low cost processing unit with an embedded processor, and a low-volume digital map where road markings are depicted by a minimum number of points. The framework registers the live camera images with the map and then fuses that information together with the information obtained from the other sensors (GPS, IMU) in a particle filter setup. The authors validated their framework in two separate challenging urban driving conditions achieving a mean Euclidean error of 1.18 and 1.69 m. The proposed framework is attractive from a mass production perspective due to its low cost sensors, low volume digital map, low computational complexity and its implementation on a cheap embedded system. A low cost single camera setup was also proposed in [87] by Elfring et. al. Specifically, they exploited a traffic sign map and proposed a novel method to compute likelihoods for traffic sign detection that purges the need of estimating 3D positions in the image domain. Additionally, a particle filter framework was proposed to estimate the vehicle position by fusing map registration pose estimation, GPS, IMU and wheel odometry information. The evaluation took place on a highway resulting to an accuracy of 0.77 m in lateral direction and 0.73 m in longitudinal direction, in terms of average error. The proposed work improved the state of the art concerning localization methods based on traffic signs while the main drawback remains its dependency on the availability of traffic signs.

Pole-like traffic signs are not reliable localization landmarks in a highway environment, since they are sparse and produce many false positive detections. For this reason, authors in [84] utilized a prior map containing road surface markers such as lane lines, stop lines and traffic sign markers. The aforementioned markers were detected by an AVM camera module and the final vehicle pose was extracted by a particle filter framework. One of the key novelties of this work, was the exact probabilistic modeling of the sensor noise to enhance the localization accuracy by correcting the pose estimation. The evaluation was executed in various test environments and resulted to a euclidean localization RMS error of 0.53 m. The main advantages of the aforementioned method are: the operation based on low-cost GPS and an AVM camera, the noise model of the RSM features is performed using data from real world driving conditions and finally the application of an around-view monitoring (AVM) system to the localization algorithm. However, the framework is unable to detect RSM perfectly in the degraded surface condition, and the authors are planning to integrate data from other sensors such as RADAR, LiDAR, etc. to compensate. Cameras were also utilized as the main localization sensors in [86], where the authors exploited the recent advances in semantic segmentation of images. They constructed a point cloud semantically segmented map, acquired live semantically segmented images, matched them on the map and the localization task was carried out by a devised particle filter setup. A comparison with a SIFT-based [112] approach, showed that the two methods performed on par with each other. There were 11 different localization trials with the best case resulting to an average error of 0.5 m. While the algorithm presents adequate performance, a segmentation algorithm training with data obtained during longer ranges would increase its robustness. Additionally, a combination of traditional feature point localization with semantic localization would be feasible with the proposed pipeline. Since HD map creation is a tedious and relatively hard procedure, authors in [83] tried to exploit information from OpenStreetMaps [113] to assist their localization framework. In more detail, they cumulated the drift from their vSLAM [114] algorithm by generating a new pose estimate from a map-based localization algorithm. An MCL framework was also devised to track the vehicles position inside the map. The visual odometry estimation was used as input to the motion model, and Chamfer matching [115] was exploited to align the live images to the map. The evaluation showed that the framework outperform state of the art concerning visual odometry approaches and it is has low infrastructural costs. The authors are planning to investigate of how to utilize all the different information that kind of maps provide. Authors in [109] also exploited OpenStreetMap data in order to build a highly-modular, camera-relied localization framework. In more detail, their framework consisted of three main layers: Road level, Ego-lane level and Lane level. Road level localization refers to which road the vehicle travels and achieved utilizing OpenStreetMaps datasets and a low-cost GPS receiver. Ego-lane level localization refers to which lane the vehicle travels on and was achieved by utilizing a recognition lane marking in an information-driven fashion. Finally, lane level localization refers to the exact spot of the vehicle on the lane in terms of lateral and longitudinal position. The later was achieved by employing the well known YOLO [116] detector in a probabilistic framework composed of a Bayesian Network and Hidden Markov Model. The evaluation showed that the model can achieve a 99% accuracy in terms of classification accuracy on a 2-lane road, 90.90% on a 3-lane road and 85.35% on a 4-lane road. Compared to other state-of-the-art Visual Odometry frameworks, the proposed method yielded a significantly better accuracy results of 5.19 m in terms of average error. The authors proved the robustness of their framework on different datasets under various weather situations. However, they are also planning to add detectors from different sensors as an improvement and work on updating the map databases when they provide false information. HD map management and size was also a concern for the authors in [89]. They utilized a map containing region descriptors extracted from grid maps, keeping the data volume of the produced map low. In order to achieve the descriptor extraction in an unsupervised fashion they used the Maximally Stable Extremal Regions (MSER) algorithm [117]. For the pose estimation, a particle filter was employed which associated MSER features from an online generated grid map with features of the digital map. Finally, the evaluation conducted in a rural environment, showed that the framework can precisely estimate the pose of the vehicle. Specifically, it yielded mean localization errors of 0.013 m in longitudinal direction, 0.05 m in lateral direction and 0.05 deg in heading direction. The method was not evaluated under complex urban scenarios, where lane marking data from camera sensor could be introduced in order to reduce the failures and improve method's robustness.

As previously mentioned, particle filter is a special case of Bayesian filtering methods. In [99], authors employed a Bayesian filter to obtain the final pose estimation. They composed a lightweight localization method, that does not require detailed knowledge about the appearance of the world. Towards achieving that goal, they utilized a sparse semantic LiDAR map with lane graphs and locations of traffic signs. In their Bayesian filter framework they fused GPS/IMU information and the pose extracted from the prior map. One of the main contributions of this work is that it takes up only 0.3% of the storage required by state-of-the-art other works. Also despite the low storage requirements the proposed framework is able to achieve a 0.05 m lateral accuracy and 1.12 m longitudinal accuracy in terms of average error. Despite being evaluated only in a highway environment, the framework achieves accuracy comparable to the state of the art by using a negligible percentage of storage data required by previous LiDAR intensity based approaches. Wan et al. in [98], modeled the environment with Gaussian distribution utilizing both the intensity of the LiDAR points and the height information. The localization framework was deployed and tested with a large fleet of cars, which conceived their final position estimation exploiting a Kalman filtering framework, built generally to readily fuse more sensors at various cost levels. Also the system adaptively utilized information from complementary sensors such as GNSS and IMU to achieve high localization accuracy and resilience in challenging situations. More precisely, the proposed method reached a lateral average error accuracy of 0.036 m and a longitudinal average error accuracy of 0.032 m. Those results were extracted by an evaluation on regular roads with adequate GNSS coverage. In [92], the structure of the environment was also modeled by utilizing Gaussian mixtures over the z-height. Specifically, their localization framework was assisted by a 2D LiDAR-created grid map with each cell holding a 1D Gaussian Mixture Model. The reason behind that modeling choice was that LiDAR intensity values and extracted features cannot perform optimally under all circumstances and weather conditions. Live scan registration on the prior map was performed by formulating a branch-and-bound search over multi-resolution, rasterized versions of the Gaussian Mixture Maps. The authors evaluated their method in an urban environment both with normal and harsh weather conditions (snow). Under normal driving conditions the results showed an RMS error accuracy of 0.15 m in the longitudinal direction and 0.10 m in the lateral direction. The framework kept its precision also under adverse weather conditions with accuracies of 0.18 m and 0.09 m in longitudinal and lateral direction, respectively. The main advantages of this framework is its rapid inference time due to Gaussian mixture maps, and can efficiently find the guaranteed optimal registration rather than finding the local optima like the modern scan matchers. The same authors extended their work in [94], where they reduced the data of large point clouds, utilizing GPUs rather than CPUs for the algorithm deployment and conducted extensive evaluations of their localization framework over large datasets of more than 500 km to prove its robustness. The evaluation showed that the framework maintained its precision even driving under extreme weather conditions and poorly textured roads. The results were on the same level as their previous work. Table IV includes a brief overview of the discussed probabilistic localization techniques along with the used map types, the experimental environments and the reported accuracies.

To recap, probabilistic techniques are widely implemented to assist the localization process. Their main advantage is that they let the observation model and motion model unrestricted and are independent of system size. Also they have scaling capabilities due to their parallel-friendly nature. Finally, they

TABLE V
DEEP LEARNING LOCALIZATION TECHNIQUES OVERVIEW

Technique	Prominent Sensor	Map Type	Experimental Environment	Reported Accuracy
DeLS-3D [118]	Camera	3D Semantic Map	Urban	Translation average error: 1.005 m and Rotational average error: 0.719 deg
LocNet [119]	LiDAR	Point Cloud Map	Urban	F_1 score: 0.381
DeepLocalization [120]	LiDAR + RADAR + Camera	Landmark Map	Urban + Rural	RMSE: 0.178 m lon, 0.170 m lat, 0.852 deg yaw. Inference time: 1758 ms
L3-Net [121]	LiDAR	Point Cloud Map	Urban	RMSE: 0.036 m lon, 0.032 m lat, 0,033 deg yaw
Barsan *et al.* [122]	LiDAR	LiDAR intensity map	Urban + Highway	Median Error: 0.043 m lon, 0.030 m lat
Liu *et al.* [123]	Camera	Google Street View	Urban	Recall: 98.2%, inference time: 42.6 ms
DenserNet [123]	Camera	Feature Map	Urban	Recall: 98.2%
Yan *et al.* [124]	Camera	Feature Map	Urban	Recall: 88.41%

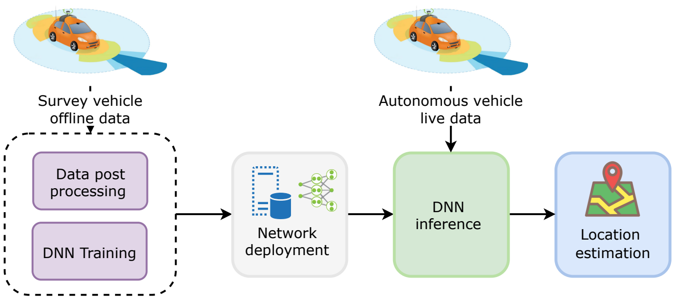


Fig. 7. Generic deep learning localization pipeline with the following modules: network training, network deployment, and location belief generation using network inference.

are easy to implement and can handle uncertainties masterfully. Hence, they are highly recommended for challenging environments. Their main drawback is that they require good initial location approximations and their high computational cost.

C. Deep Learning Techniques

In recent years, the domain of Artificial Intelligence (AI) penetrated almost every aspect of technology and computer science, with its subset of deep learning at the cutting edge. Deep learning is inspired by the structure of the human brain and operates by exploiting multilayered neural networks processing vasts amounts of data. The general pipeline is shown in Fig. 7, and representative example of neural network's utilization for localization process is the work proposed in [119]. In this work, the neural networks are used for the core localization task (e.g. matching the online measurement with the prior map) and also for the map creation step. More specifically a deep neural network is used to extract low-dimensional fingerprints from 3D LiDAR sensor readings transformed as rotational invariant representations. Subsequently, a particle filter localization framework is used for global metric localization. The prior map is constructed based on those fingerprints, and also the localization process relies on them. As a final step, the predicted pose is aligned with the original pose extracted from the global prior map with the ICP algorithm to furthermore improve the localization accuracy. The valuation of the aforementioned framework was not executed with a real vehicle, but with the utilization of the KITTI dataset on a specific sequence with the accuracy in terms of F_1 max score being 0.381. Despite its high accuracy and comparable results with the state of the art, the framework is planned to be evaluated under real driving situations as mentioned by the authors. Another approach exploiting deep neural networks is presented in [120], where landmarks extracted from the online data and from a prior map, are fed into two different deep neural network architectures. The inferred feature vectors are then concatenated and fed to a third neural network, outputting the position and the heading difference from the previously calculated pose. However, the initial pose should be known before hand from a GPS sensor, to initiate the pipeline and ultimately localize the vehicle effectively. The evaluation results were promising, showing that the framework is robust concerning the dynamic environment changes and around ten times faster than other state-of-the-art works, maintaining also the desired accuracy. The authors are considering as future work to make the framework independent of GPS measurements. More precisely, the accuracy in terms of RMS error was 0.178 m in the longitudinal direction, 0.170 m in the lateral direction, and 0.852 degrees in the heading direction. The mean inference time of the framework was 1.758 ms. In contrast, the work in [122] outputs a 3-DoF pose in each time step and not only the difference from the previous pose. More precisely, a deep neural network was trained to embed a prior LiDAR intensity map, and online LiDAR sweeps in a common space. During online localization process, live scan from the LiDAR is embedded in that space, and is slightly rotated for a fixed amount of times. Subsequently, the localization is achieved by exhaustively trying to match the live embedded scan, with an embedded scan from the prior map, by computing their cross-correlation. The results of the aforementioned work are really promising since the method generalizes well with different LiDAR types without the need to retrain, and considering the achieved centimeter level accuracy without intensity calibration. More specifically, the accuracy in terms of median error was 3.00 cm and 4.33 cm, in lateral and longitudinal direction, respectively. The evaluation was extensive utilizing a crafted dataset with more than $4000km$ of driving.

Two critical sub classes of deep neural networks are Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs). CNNs are used for process and classification of imagery and other high-resolution data, such as point clouds, whereas RNNs are exploited for producing predictive results to sequential data. In [118] researchers propose a system architecture for the localization task exploiting both CNNs

This article has been accepted for publication in IEEE Transactions on Intelligent Vehicles. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIV.2022.3192102

TABLE VI
LIST OF PUBLIC DATASETS

Dataset	HD Map	RTK GNSS	Data Provided	Annotation	Diversity	Environment
Lyft Level 5 [125]	✓	✗	LiDAR, Video, Vehicle Data	Bounding Box, Semantic Label	Illumination	Urban
NuScenes[1] [126]	✓	✗	LiDAR, Video, RADAR, GNSS, Vehicle Data	Bounding Box, Semantic Label	Weather, Night	Urban
PandaSet [127]	✗	✗	LiDAR, Image, GNSS, Vehicle Data	Bounding Box, Semantic Label	Illumination, Night	Urban
Argoverse [128]	✓	✗	LiDAR, Video, Image, Vehicle Data	Bounding Box, Semantic Label, Lane Marking	Season, Weather, Night	Urban
Waymo Open [129]	✗	✗	LiDAR, Image, GNSS, Vehicle Data	Bounding Box, Semantic Label	Illumination, Night	Urban
ApolloScape [130]	✗	✗	LiDAR, Video, Vehicle Data	Bounding Box, Semantic Label, Lane Marking	Illumination, Weather	Urban, Highway, Rural
UrbanLoco [131]	✗	✓	LiDAR, Video, Vehicle Data	N/A	Illumination	Urban
A2D2 [132]	✗	✗	LiDAR, Video, Vehicle Data	Bounding Box	Season	Urban
A*3D [133]	✗	✗	LiDAR, Image	Bounding Box	Weather, Night	Urban
Oxford RobotCar[2] [134]	✗	✓	LiDAR, Video, Vehicle Data	N/A	Illumination, Season, Weather	Urban
KITTI [135]	✗	✗	LiDAR, Video, Image, GNSS	Bounding Box, Semantic Label, Lane Marking	Illumination	Urban, Highway, Rural
Semantic KITTI [136]	✗	✗	LiDAR, Video, Image, GNSS	Bounding Box, Semantic Label, Lane Marking	Illumination	Urban, Highway, Rural
Cityscapes [137]	✗	✗	Video, Image, GNSS, Vehicle Data	Semantic Label	Season	Urban
Berkeley DeepDrive [138]	✗	✗	Video, Image, GNSS	Bounding Box, Semantic Label, Lane Marking	Season	Urban, Highway, Rural
Cirrus [139]	✗	✗	LiDAR, Image, GNSS	Bounding Box	Night	Urban, Highway
KAIST Multi-spectral [140]	✗	✗	LiDAR, Video, Thermal, GNSS	Bounding Box	Illumination, Night	Urban, Rural
DGL-MOTS [141]	✗	✗	Video, Image	Semantic Label	Illumination, Night	Urban, Highway, Rural

[1]Including NuPlan dataset [142]. [2]Including the real-time kinematic ground truth for the Oxford RobotCar Dataset [143].
Tick: feature satisfied. Cross: feature not satisfied.

and RNNs. Based on a GPS for coarse localization and on a 3D semantic point cloud map for fine localization, at first a stream of frames from the on-board vehicle camera are fed into a CNN jointly with rendered 2D semantic labeled map, outputting the translation and the rotation to align the images. After that, a RNN enhances the alignment estimation and finally a segment CNN generates a spatially more accurate and temporally more consistent result for the image stream. The average translation error was around 1.005 m and the average heading error was 0.719 degrees. Also in [121] modules of the traditional handcrafted pipelines for localization are replaced by deep neural networks. At first the authors extract a set of key-points evaluated by their linearities and scattering defined with the eigenvalues of the neighbors of a 3D point, and feed them in a Point-Net [144], to extract feature descriptors. The output of the Point-Net is regularized by 3D convolutions executed from a CNN, outputting the matching probabilities of all the dimensions and obtain the optimal estimation. Finally, a RNN encapsulates the temporal motion dynamics, which are traditionally modeled by filtering methods such as particle filters. The evaluation of this work was extensive under a variety of difference scenes such as dense urban environments, highways and rural areas. The localization accuracy in terms of RMS error in an urban setting was 0.036 m, 0.032 m and 0.033 degrees in longitudinal, lateral and heading directions, respectively. The aforementioned framework was evaluated with a year's gap between the collected mapping and testing data. CNNs were also utilized from authors in [145]. They devised a CNN architecture which aggregated feature maps at different semantic levels for image representations. Those maps improved the image retrieval accuracy by producing more keypoint features. Their method was flexible and can be applied on infrastructures with limited resources, without hindering the accuracy of the framework. The conducted evaluation was extensive and showcased remarkable results on the used datasets with their best accuracy in terms of recall being 98.2%. Another work by the same team was presented in [124], where they proposed a novel method that uses hierarchical features to close the semantic gap in feature learning. They have executed the attention fusion over the detected features to produce firm image representation for different scale landmarks. The aforementioned features were extracted by utilizing a CNN backbone architecture. The evaluation of their method proved that the framework is competitive among state-of-the-art methods, achieving a recall close to 88.41%.

Authors in [123] managed to localize a vehicle by using only Google Street View [146] images and three cameras mounted on the top of their experiment vehicle. The center of the framework is a three-flavour Visual Localization Network devised by the authors, based on MobileNet [147]. The network extracts features from Google Street View images and later saves them into a location library. Every location library entry consists of groups containing three feature images (left, center, right) and the respective vehicle location. The live data are matched with the prior data utilizing a novel feature voting algorithm. The framework achieved an accuracy of 98.2% in terms of recall and an inference time of 42.6 f/ms. In Table V, a summary of the discussed deep learning techniques for map-based localization is given along with the details.

To summarize, deep learning techniques are rapidly emerging in the field of autonomous vehicles, exploiting the potential of cameras and machine vision [148] along with the continuously increased on-board computational resources [149]. Their main advantage is that they are easily deployable, support parallel and distributed algorithms and their scalable nature. They can be also applied in dense or sparse point clouds obtained from automotive LiDARs or Radars [150], respectively. Their weaknesses are the complexity to build and train the initial model and the high computational cost in order to have a rapid inference time.

VI. DATASETS

Ideally, a dataset for training, testing and evaluating map-based localization algorithms should include the HD map and the $6D$ ground truth pose of the ego vehicle together with the synchronized available on-board sensor readings. However, as stated in previous sections, such datasets still do not exist in a unified and consistent form, leading to the inability of clear comparisons between the available algorithms and techniques. Unlike SLAM-based localization methods, most map-based localization works utilize their own platforms and datasets, reporting their accuracy and performance on their own tests. Currently, only the Lyft Level 5 AV Dataset [125] includes an HD map and the $6D$ localization ground truth, however, the odometry data is missing. In such cases, where critical information is missing, researchers can try to emulate the missing data from other available sensors reading, a task that is time consuming and not always very accurate. For example, in the case of Lyft 5 dataset, the missing odometry can be emulated by utilizing the ground truth poses to predict noisy odometry signals according to velocity-based motion models. In most datasets, where the HD map is missing, authors should incorporate SLAM-based and loop-closure techniques to construct the prior map, however, such constructed maps are far less precise from the pre-built HD maps. A list of all available open-source datasets for autonomous vehicles is shown in Table VI. It is clear that among all currently available datasets, only the NuScenes [126], the Argoverse [128] and the Lyft Level 5 [125] datasets include HD map data. Furthermore, real-time kinematic ground truth data is provided only in UrbanLoco [131] and Oxford RobotCar [143] datasets.

VII. Discussion and Future Directions

Each localization algorithmic approach has its advantages and limitations. More precisely, the map matching techniques present optimal performance in environments with dense features such as urban and highway driving, while in rural environments their performance is deteriorated. The advent of on-board high-performance computing systems allow them to run efficiently, however, high disk storage is still required. The main limitation of this category of localization algorithms is their inability to handle uncertainties and ambiguities. Their output validity is limited in cases where the driving environment has significantly changed against the prior map (long road constructions, new temporary road markings and deviations) and in cases where the raw sensor input is affected by weather and illumination conditions.

On the contrary, probabilistic based methods can handle more efficiently uncertainties, thus they are more suitable for challenging environments. This is mainly achieved from the fact that their observation and motion models are more unrestricted against those of map matching ones. Additionally, they present scaling capabilities, since they are parallel friendly and easy to implement algorithms. The backlash is that they generally require good initial position approximations and are more computational intensive algorithms. Furthermore, there are more prone to errors in cases of GNSS data loss.

Deep-learning methods are relative new methods and have not been explored intensively in real autonomous driving conditions. However, they present some advantages since they are more flexible and scalable. Furthermore, they can perform adequately with less sensor modalities since they can predict pose differences and global poses only with visual data. However, generally such methods are unaware of the inherent deep networks uncertainties, making them less effective in challenging environments.

It is arduous to devise, test, and deploy a versatile localization algorithm. Truly robust results should be delivered regardless environmental and weather conditions, hence a mix of on-board sensor modalities each one delivering the required levels of precision. For example, under heavy rain LiDAR scans would not have the required quality to localize the vehicle correctly, and under heavy sunlight cameras would also suffer poor performance. In addition, the operating environment changes over time, and more precisely every season has its own effects on it. In summer there is no such vivid and dense vegetation as spring, and in winter the landscape may be snowy. For the aforementioned reason, a prior map database consisting of numerous maps should exist, since localization with a single prior map will localize the vehicle adequately under certain environmental conditions, the performance may drop facing challenging circumstances.

The choice of features is also a crucial factor for localization accuracy, and in some cases its decisive even about its failure. Feature extraction from raw sensor data is important, not only because it reduces the size of the map and makes it easier to be stored and processed, but also because it renders the registration between live data and the map faster. In cities and generally dense urban areas with lots of traffic, road surface landmarks, curbs and lane lines as features are considered as a poor choice, since they are obscured most of the time. A decent choice would be pole like features such as tree trunks and lamp posts, and building facades. On contrast, in rural areas and highway environments the absence of these kind of features leads to the choice of road surface features, traffic signs and curbs.

Sensor modalities are also a vital aspect of the localization layer, because its accuracy is heavily influenced by the flawless introduction of environmental information in the system. The operational environment is a crucial aspect of this choice. For instance, in a city where the traffic is almost always high, and a lot of roaming objects in the scenery, RADAR sensors for the localization task would be a poor choice due to their noise sensitivity. Also in environments characterized by heavy rains and snowfalls, LiDARs may fail to localize the vehicle precisely, because environmental conditions directly affect their data quality. A sensible solution is to employ a mix of sensors, each of them introducing noiseless data regarding the operational conditions.

The density of sensor data are also a crucial factor of the localization system design. Sensors that introduce highly dense information such as cameras and LiDARs, consequently require high amount of computational resources on-board to process them. Many research efforts utilize high-end expensive computational devices on-board for deep neural network inference at the edge. Additionally, the emerging commercialization of the 5G technology will leverage the on-board requirements of high processing power. Future autonomous vehicles will entail higher cruise speeds and low-latency broadband connection to the mobile network almost everywhere. Vehicle-to-vehicle communications will further improve localization algorithms utilizing multiple vehicles' environment perception.

As a future guideline, the continuous technological evolution of sensors should also be considered. A notable case is the emerging technology of Solid State LiDAR sensor. SSLs can overcome the constraints regarding price and size compared with traditional LiDAR sensor setups due to their completely different philosophy of operation. They are built entirely on silicon chips with no moving parts, resulting in resilience to vibrations and smaller size, two very important factors for the self-driving domain. Additionally, cameras are also becoming smaller in size, more affordable and their resolution is constantly improving. Generally, it is that the evolving sensor size reduction will lead to different sensor placements and configurations, such as the emerging smart headlight integration [151].

Concerning the localization methods, hybrid architectures including initial estimations based on map matching methods and selection stages based on probabilistic methods will allow to deal with the drawbacks of each of the methods. By doing so, the map matching step will deal with the majority of the nominal cases while the ambiguous cases can be addressed with the probabilistic step, keeping the complexity of the overall algorithms reduced. Additionally, approaches that incorporate deep-learning methods for the feature detection together with map matching techniques for the location estimation step seem promising. In such methods, the uncertainties can be addressed by employing Monte Carlo dropout techniques, where the regression uncertainties can be quantified in the model predictions [152], [153].

Undoubtedly, whether it is smarter sensors or new algorithms, vision seems to be one of the most promising sensor modalities for the localization approach in the future. Additionally, since V2X communication will enable to exchange on-board sensor and GNSS measurements from the surrounding vehicles and infrastructure, several techniques such as double difference and weighted cooperation can be applied to optimize the localization accuracy by the available shared measurements [154]. Computer vision deep learning techniques can be even applied to wireless base stations with potential gains in localization accuracy by leveraging visual data to help overcome the beam selection and blockage prediction challenges [155].

Finally, during the course of this survey it was surprising that no rigid comparison results were presented against other map-based localization methods. This is mainly due to the absence of unified and concrete benchmarking datasets. Most of the research efforts were supported by their own datasets specifically built for a given area or campus with ground truth being measured using RTK-GNSS sensors. The need of the community for having a reliable and multi-purpose dataset is apparent, however, a constant effort for improving the currently available datasets is evident [143], [142]. Furthermore, these datasets should be diverse in terms of sensors, weather and environmental conditions including day and night driving. Significant efforts have been realized but in a simulation environment for the CARLA autonomous driving challenges.

VIII. Conclusion

Accurate localization is the basis of the autonomous vehicle software stack and its accuracy, reliability, and robustness is critical for the operation and acceptance of them. Perception, planning and control layers are depended on the precise operation of the localization layer, irrespective of harsh weather conditions, such as snow, fog, dust and GPS-denied areas. In order to enhance the localization accuracy and strengthen its performance, prior maps of the driving environment are necessary for stepping up the levels of autonomy in autonomous vehicles.

In this paper, a survey of the current state-of-the-art of map-based vehicle localization techniques utilizing several on-board sensor modalities was provided. These techniques were classified by their core localization process into map-matching, probabilistic and deep learning exploiting their advantages and disadvantages. Undoubtedly, no specific sensor configuration or localization approach can be considered the perfect candidate for the localization of autonomous vehicles in all weather and driving conditions. However, the insights of this paper can help the community to better understand, evaluate and improve their map-based localization frameworks.

Appendix
List of acronyms

ACC	Adaptive Cruise Control
ADAS	Advanced Driver Assistant Systems
AI	Artificial Intelligence
AVM	Around View Monitor
CNN	Convolutional Neural Network
DoF	Degrees of Freedom
EKF	Extended Kalman Filter
FoV	Field of View
GICP	Generalized Iterative Closest Point
GIS	Geographic Information System
GNSS	Global Navigation Satellite Systems
GPS	Global Positioning System
GPU	Graphics Processing Unit
HD Maps	High-Definition Maps
ICP	Iterative Closest Point
IMU	Inertial Measurement Unit
KF	Kalman Filter
LDM	Local Dynamic Map
LiDAR	Light Detection And Ranging
MCL	Monte Carlo Localization
MEMS	Micro-ElectroMechanical System
MMS	Mobile Mapping System
MSER	Maximally Stable Extremal Regions
NDT	Normal Distribution Transform
NMI	Normalized Mutual Information
NN	Nearest Neighbor
OPA	Optical Phased Array
RADAR	Radio Detection And Ranging
RMSE	Root Mean Square Error
RNN	Recurrent Neural Network
RSU	Road Side Unit
RTK	Real-Time Kinematic
SAE	Society of Automotive Engineers
SLAM	Simultaneous Localization And Mapping
SSL	Solid State LiDAR
UKF	Unscented Kalman Filter
V2X	Vehicle-to-Everything

REFERENCES

[1] National Center for Statistics and Analysis, "Overview of motor vehicle crashes in 2019," National Highway Traffic Safety Administration, Tech. Rep. DOT HS 813 060, Dec. 2020.

[2] J. Levinson and S. Thrun, "Robust vehicle localization in urban environments using probabilistic maps," in *IEEE International Conference on Robotics and Automation*, 2010, pp. 4372–4378.

[3] R. Smith and P. Cheeseman, "On the representation and estimation of spatial uncertainty," *The International Journal of Robotics Research*, vol. 5, no. 4, pp. 56–58, 1986.

[4] J. Zhang and S. Singh, "Visual-lidar odometry and mapping: low-drift, robust, and fast," in *IEEE International Conference on Robotics and Automation*, 2015, pp. 2174–2181.

[5] J. Deschaud, "IMLS-SLAM: Scan-to-model matching based on 3D data," in *IEEE International Conference on Robotics and Automation*, 2018, pp. 2480–2485.

[6] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The KITTI dataset," *The International Journal of Robotics Research*, vol. 32, no. 11, pp. 1231–1237, 2013.

[7] S. Kuutti, S. Fallah, K. Katsaros, M. Dianati, F. Mccullough, and A. Mouzakitis, "A survey of the state-of-the-art localization techniques and their potentials for autonomous vehicle applications," *IEEE Internet of Things Journal*, vol. 5, no. 2, pp. 829–846, 2018.

[8] M. Elhousni and X. Huang, "A survey on 3D LiDAR localization for autonomous vehicles," in *IEEE Intelligent Vehicles Symposium*, 2020, pp. 1879–1884.

[9] G. Bresson, Z. Alsayed, L. Yu, and S. Glaser, "Simultaneous localization and mapping: A survey of current trends in autonomous driving," *IEEE Transactions on Intelligent Vehicles*, vol. 2, no. 3, pp. 194–220, 2017.

[10] M. Kubicka, A. Cela, H. Mounier, and S.-I. Niculescu, "Comparative study and application-oriented classification of vehicular map-matching methods," *IEEE Intelligent Transportation Systems Magazine*, vol. 10, no. 2, pp. 150–166, 2018.

[11] P. Chao, Y. Xu, W. Hua, and X. Zhou, "A survey on map-matching algorithms," in *Australasian Database Conference*, 2020, pp. 121–133.

[12] J. Laconte, A. Kasmi, R. Aufrère, M. Vaidis, and R. Chapuis, "A survey of localization methods for autonomous vehicles in highway scenarios," *Sensors*, vol. 22, no. 1, p. 247, 2021.

[13] Y. Lu, H. Ma, E. Smart, and H. Yu, "Real-time performance-focused localization techniques for autonomous vehicle: A review," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 7, pp. 6082–6100, 2022.

[14] F. Li, P. Bonnifait, and J. Ibañez-Guzmán, "Map-aided dead-reckoning with lane-level maps and integrity monitoring," *IEEE Transactions on Intelligent Vehicles*, vol. 3, no. 1, pp. 81–91, 2018.

[15] A. Hansson, E. Korsberg, R. Maghsood, E. Nordén *et al.*, "Lane-level map matching based on hmm," *IEEE Transactions on Intelligent Vehicles*, vol. 6, no. 3, pp. 430–439, 2020.

[16] R. E. Kalman, "A New Approach to Linear Filtering and Prediction Problems," *Journal of Basic Engineering*, vol. 82, no. 1, pp. 35–45, 1960.

[17] S. J. Julier and J. K. Uhlmann, "New extension of the kalman filter to nonlinear systems," in *Signal processing, sensor fusion, and target recognition VI*, vol. 3068, 1997, pp. 182–193.

[18] E. Wan and R. Van Der Merwe, "The unscented kalman filter for nonlinear estimation," in *IEEE Adaptive Systems for Signal Processing, Communications, and Control Symposium*, 2000, pp. 153–158.

[19] M. M. Waldrop, "Autonomous vehicles: No drivers required," *Nature News*, vol. 518, no. 7537, p. 20, 2015.

[20] J. Ziegler, P. Bender, M. Schreiber, H. Lategahn, T. Strauss, C. Stiller, T. Dang, U. Franke, N. Appenrodt, C. G. Keller, E. Kaus, R. G. Herrtwich, C. Rabe, D. Pfeiffer, F. Lindner, F. Stein, F. Erbs, M. Enzweiler, C. Knppel, J. Hipp, M. Haueis, M. Trepte, C. Brenk, A. Tamke, M. Ghanaat, M. Braun, A. Joos, H. Fritz, H. Mock, M. Hein, and E. Zeeb, "Making bertha drivean autonomous journey on a historic route," *IEEE Intelligent Transportation Systems Magazine*, vol. 6, no. 2, pp. 8–20, 2014.

[21] R. Liu, J. Wang, and B. Zhang, "High definition map for automated driving: Overview and analysis," *Journal of Navigation*, vol. 73, no. 2, p. 324341, 2020.

[22] "Mathworks RoadRunner." [Online]. Available: https://www.mathworks.com/products/roadrunner

[23] G. Rong, B. H. Shin, H. Tabatabaee, Q. Lu, S. Lemke, M. Možeiko, E. Boise, G. Uhm, M. Gerow, S. Mehta *et al.*, "LGSVL simulator: A high fidelity simulator for autonomous driving," in *IEEE International Conference on Intelligent Transportation Systems*, 2020, pp. 1–6.

[24] M. Behrisch, L. Bieker, J. Erdmann, and D. Krajzewicz, "SUMO–Simulation of urban MObility: an overview," in *International Conference on Advances in System Simulation*, 2011.

[25] "AutoCore MapToolbox." [Online]. Available: https://github.com/autocore-ai/MapToolbox

[26] "Orbit GT." [Online]. Available: https://orbitgt.com/domains/mobile-mapping/

[27] "NVIDIA DRIVE map." [Online]. Available: https://www.nvidia.com/en-us/self-driving-cars/hd-mapping/

[28] K. Franke, "Volkswagen group: Leveraging VIRES VTD to design a cooperative driver assistance system," *Engineering Reality Magazine, Winter*, pp. 10–14, 2018.

[29] "TopoDOT." [Online]. Available: https://new.certainty3d.com

[30] L. Bampis, A. Amanatiadis, and A. Gasteratos, "Fast loop-closure detection using visual-word-vectors from image sequences," *The International Journal of Robotics Research*, vol. 37, no. 1, pp. 62–82, 2018.

[31] ——, "Encoding the description of image sequences: A two-layered pipeline for loop closure detection," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2016, pp. 4530–4536.

[32] K. Massow, B. Kwella, N. Pfeifer, F. Häusler, J. Pontow, I. Radusch, J. Hipp, F. Dölitzscher, and M. Haueis, "Deriving HD maps for highly automated driving from vehicular probe data," in *IEEE International Conference on Intelligent Transportation Systems*, 2016, pp. 1745–1752.

[33] A. Richter and M. Scholz, "Deploying guidelines and a simplified data model to provide real world geodata in driving simulators and driving automation," *Transportation research part F: traffic psychology and behaviour*, vol. 61, pp. 305–313, 2019.

[34] D. Pannen, M. Liebner, W. Hempel, and W. Burgard, "How to keep HD maps for automated driving up to date," in *IEEE International Conference on Robotics and Automation*, 2020, pp. 2288–2294.

[35] F. Poggenhans, J.-H. Pauls, J. Janosovits, S. Orf, M. Naumann, F. Kuhnt, and M. Mayr, "Lanelet2: A high-definition map framework for the future of automated driving," in *International Conference on Intelligent Transportation Systems*, 2018, pp. 1672–1679.

[36] M. Dupuis, M. Strobl, and H. Grezlikowski, "Opendrive 2010 and beyond–status and future of the de facto standard for the description of road networks," in *Driving Simulation Conference Europe*, 2010, pp. 231–242.

[37] M. Althoff, S. Urban, and M. Koschi, "Automatic conversion of road networks from opendrive to lanelets," in *IEEE International Conference on Service Operations and Logistics, and Informatics*, 2018, pp. 157–162.

[38] S. Maierhofer, M. Klischat, and M. Althoff, "Commonroad scenario designer: An open-source toolbox for map conversion and scenario creation for autonomous vehicles," in *IEEE International Intelligent Transportation Systems Conference*, 2021, pp. 3176–3182.

[39] I. Tomkos, D. Klonidis, E. Pikasis, and S. Theodoridis, "Toward the 6G network era: Opportunities and challenges," *IT Professional*, vol. 22, no. 1, pp. 34–38, 2020.

[40] Y. Cui, L. Yan, Z. Cao, and D. Liu, "Tf-blender: Temporal feature blender for video object detection," in *IEEE/CVF International Conference on Computer Vision*, 2021, pp. 8138–8147.

[41] D. Liu, Y. Cui, Y. Chen, J. Zhang, and B. Fan, "Video object detection for autonomous driving: Motion-aid feature calibration," *Neurocomputing*, vol. 409, pp. 1–11, 2020.

[42] B. Wu, X. Zhou, S. Zhao, X. Yue, and K. Keutzer, "Squeezesegv2: Improved model structure and unsupervised domain adaptation for road-object segmentation from a lidar point cloud," in *IEEE International Conference on Robotics and Automation*, 2019, pp. 4376–4382.

[43] D. Liu, Y. Cui, W. Tan, and Y. Chen, "Sg-net: Spatial granularity network for one-stage video instance segmentation," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 9816–9825.

[44] A. Milioto, J. Behley, C. McCool, and C. Stachniss, "Lidar panoptic segmentation for autonomous driving," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2020, pp. 8505–8512.

[45] Z. Wang, Y. Wu, and Q. Niu, "Multi-sensor fusion in automated driving: A survey," *IEEE Access*, vol. 8, pp. 2847–2868, 2019.

[46] H. Zhu, K.-V. Yuen, L. Mihaylova, and H. Leung, "Overview of environment perception for intelligent vehicles," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 10, pp. 2584–2601, 2017.

[47] G. Zamanakos, L. Tsochatzidis, A. Amanatiadis, and I. Pratikakis, "A cooperative LiDAR-camera scheme for extrinsic calibration," in *IEEE Image, Video, and Multidimensional Signal Processing Workshop*, 2022.

[48] Z. Zhang, C. Fu, C. Dong, C. Mertz, and J. M. Dolan, "Self-calibration of multiple lidars for autonomous vehicles," in *IEEE International Intelligent Transportation Systems Conference*, 2021, pp. 2897–2902.

[49] Y. Han, Y. Liu, D. Paz, and H. Christensen, "Auto-calibration method using stop signs for urban autonomous driving applications," in *IEEE International Conference on Robotics and Automation*, 2021, pp. 13 179–13 185.

[50] P. Besl and N. D. McKay, "A method for registration of 3-d shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 239–256, 1992.

[51] P. Biber and W. Strasser, "The normal distributions transform: a new approach to laser scan matching," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, vol. 3, 2003, pp. 2743–2748.

[52] A. Schlichting and C. Brenner, "Localization using automotive laser scanners and local pattern matching," in *IEEE Intelligent Vehicles Symposium Proceedings*, 2014, pp. 414–419.

[53] K. Yoneda, H. Tehrani, T. Ogawa, N. Hukuyama, and S. Mita, "Lidar scan feature for localization with highly precise 3-d map," in *IEEE Intelligent Vehicles Symposium Proceedings*, 2014, pp. 1345–1350.

[54] B. Nagy and C. Benedek, "Real-time point cloud alignment for vehicle localization in a high resolution 3d map," in *European Conference on Computer Vision Workshops*, 2018.

[55] H. Wang, C. Xue, Y. Zhou, F. Wen, and H. Zhang, "Visual semantic localization based on HD map for autonomous vehicles in urban scenarios," in *IEEE International Conference on Robotics and Automation*, 2021, pp. 11 255–11 261.

[56] K. Yoneda, C. Yang, S. Mita, T. Okuya, and K. Muto, "Urban road localization by using multiple layer map matching and line segment matching," in *IEEE Intelligent Vehicles Symposium Proceedings*, 2015, pp. 525–530.

[57] J.-H. Im, S.-H. Im, and G.-I. Jee, "Vertical corner feature based precise vehicle localization using 3D LiDAR in urban area," *Sensors*, vol. 16, no. 8, p. 1268, 2016.

[58] J. Rohde, B. Vlz, H. Mielenz, and J. M. Zllner, "Precise vehicle localization in dense urban environments," in *IEEE International Conference on Intelligent Transportation Systems*, 2016, pp. 853–858.

[59] F. Gustafsson, *Adaptive Filtering and Change Detection*. John Wiley & Sons, 2000.

[60] B. Cao, C.-N. Ritter, D. Ghring, and R. Rojas, "Accurate localization of autonomous vehicles based on pattern matching and graph-based optimization in urban environments," in *IEEE International Conference on Intelligent Transportation Systems*, 2020, pp. 1–6.

[61] N. Akai, L. Y. Morales, T. Yamaguchi, E. Takeuchi, Y. Yoshihara, H. Okuda, T. Suzuki, and Y. Ninomiya, "Autonomous driving based on accurate localization using multilayer LiDAR and dead reckoning," in *IEEE International Conference on Intelligent Transportation Systems*, 2017, pp. 1–6.

[62] N. Steinke, C.-N. Ritter, D. Goehring, and R. Rojas, "Robust LiDAR feature localization for autonomous vehicles using geometric fingerprinting on open datasets," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 2761–2767, 2021.

[63] P. Egger, P. V. K. Borges, G. Catt, A. Pfrunder, R. Siegwart, and R. Dub, "PoseMap: Lifelong, multi-environment 3D LiDAR localization," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2018, pp. 3430–3437.

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

[65] Y. Lu, J. Huang, Y.-T. Chen, and B. Heisele, "Monocular localization in urban environments using road markings," in *IEEE Intelligent Vehicles Symposium*, 2017, pp. 468–474.

[66] D. Kim, B. Kim, T. Chung, and K. Yi, "Lane-level localization using an AVM camera for an automated driving vehicle in urban environments," *IEEE/ASME Transactions on Mechatronics*, vol. 22, no. 1, pp. 280–290, 2017.

[67] R. Asghar, M. Garzn, J. Lussereau, and C. Laugier, "Vehicle localization based on visual lane marking and topological map matching," in *IEEE International Conference on Robotics and Automation*, 2020, pp. 258–264.

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

[69] T. Wu and A. Ranganathan, "Vehicle localization using road markings," in *IEEE Intelligent Vehicles Symposium*, 2013, pp. 1185–1190.

[70] E. Ward and J. Folkesson, "Vehicle localization with low cost radar sensors," in *IEEE Intelligent Vehicles Symposium*, 2016, pp. 864–870.

[71] M. Cornick, J. Koechling, B. Stanley, and B. Zhang, "Localizing ground penetrating RADAR: A step toward robust autonomous ground vehicle localization," *Journal of Field Robotics*, vol. 33, no. 1, pp. 82–102, 2016.

[72] A. Schlichting and U. Feuerhake, "Global vehicle localization by sequence analysis using lidar features derived by an autoencoder," in *IEEE Intelligent Vehicles Symposium*, 2018, pp. 656–661.

[73] Z. Xiao, D. Yang, T. Wen, K. Jiang, and R. Yan, "Monocular localization with vector HD map (MLVHM): A low-cost method for commercial IVs," *Sensors*, p. 1870, 2020.

[74] K. Petek, K. Sirohi, D. Bscher, and W. Burgard, "Robust monocular localization in sparse hd maps leveraging multi-task uncertainty estimation," in *IEEE International Conference on Robotics and Automation*, 2022.

[75] C. Hungar, S. Jrgens, D. Wilbers, and F. Kster, "Map-based localization with factor graphs for automated driving using non-semantic lidar features," in *IEEE International Conference on Intelligent Transportation Systems*, 2020, pp. 1–6.

[76] J.-H. Pauls, K. Petek, F. Poggenhans, and C. Stiller, "Monocular localization in hd maps by combining semantic segmentation and distance transform," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2020, pp. 4595–4601.

[77] A. Segal, D. Haehnel, and S. Thrun, "Generalized-icp." in *Robotics: science and systems*, vol. 2, no. 4, 2009, p. 435.

[78] W. Burgard, O. Brock, and C. Stachniss, *Map-Based Precision Vehicle Localization in Urban Environments*. The MIT Press, 2008, pp. 121–128.

[79] P. Biber and W. Strasser, "The normal distributions transform: a new approach to laser scan matching," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, vol. 3, 2003, pp. 2743–2748.

[80] A. Zaganidis, M. Magnusson, T. Duckett, and G. Cielniak, "Semantic-assisted 3D normal distributions transform for scan registration in environments with limited structure," in *IEEE International Conference on Intelligent Robots and Systems*, 2017, pp. 4064–4069.

[81] S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics (Intelligent Robotics and Autonomous Agents)*. The MIT Press, 2005.

[82] D. Fox, S. Thrun, W. Burgard, and F. Dellaert, "Particle filters for mobile robot localization," in *Sequential Monte Carlo methods in practice*. Springer, 2001, pp. 401–428.

[83] G. Floros, B. van der Zander, and B. Leibe, "OpenstreetSLAM: Global vehicle localization using openstreetmaps," in *IEEE International Conference on Robotics and Automation*, 2013, pp. 1054–1059.

[84] K. Jo, Y. Jo, J. K. Suhr, H. G. Jung, and M. Sunwoo, "Precise localization of an autonomous car based on probabilistic noise models of road surface marker features using multiple cameras," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 6, pp. 3377–3392, 2015.

[85] J. K. Suhr, J. Jang, D. Min, and H. G. Jung, "Sensor fusion-based low-cost vehicle localization system for complex urban environments," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 5, pp. 1078–1086, 2016.

[86] E. Stenborg, C. Toft, and L. Hammarstrand, "Long-term visual localization using semantically segmented images," in *IEEE International Conference on Robotics and Automation*, 2018, pp. 6484–6490.

[87] J. Elfring, S. Dani, S. Shakeri, H. Mesri, and J. W. van den Brand, "Vehicle localization using a traffic sign map," in *IEEE International Conference on Intelligent Transportation Systems*, 2020, pp. 1–6.

[88] Y. Xu, V. John, S. Mita, H. Tehrani, K. Ishimaru, and S. Nishino, "3D point cloud map based vehicle localization using stereo camera," in *IEEE Intelligent Vehicles Symposium*, 2017, pp. 487–492.

[89] J. Wiest, H. Deusch, D. Nuss, S. Reuter, M. Fritzsche, and K. Dietmayer, "Localization based on region descriptors in grid maps," in *IEEE Intelligent Vehicles Symposium Proceedings*, 2014, pp. 793–799.

[90] A. Hata and D. Wolf, "Road marking detection using LiDAR reflective intensity data and its application to vehicle localization," in *IEEE Conference on Intelligent Transportation Systems*, 2014, pp. 584–589.

[91] A. Y. Hata, F. S. Osorio, and D. F. Wolf, "Robust curb detection and vehicle localization in urban environments," in *IEEE Intelligent Vehicles Symposium Proceedings*, 2014, pp. 1257–1262.

[92] R. W. Wolcott and R. M. Eustice, "Fast LiDAR localization using multiresolution gaussian mixture maps," in *IEEE International Conference on Robotics and Automation*, 2015, pp. 2814–2821.

[93] A. Y. Hata and D. F. Wolf, "Feature detection for vehicle localization in urban environments using a multilayer LiDAR," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 2, pp. 420–429, 2016.

[94] R. W. Wolcott and R. M. Eustice, "Robust LiDAR localization using multiresolution gaussian mixture maps for autonomous driving," *The International Journal of Robotics Research*, vol. 36, no. 3, pp. 292–319, 2017.

[95] N. Akai, L. Y. Morales, E. Takeuchi, Y. Yoshihara, and Y. Ninomiya, "Robust localization using 3D NDT scan matching with experimentally determined uncertainty and road marker matching," in *IEEE Intelligent Vehicles Symposium*, 2017, pp. 1356–1363.

[96] H. Kim, B. Liu, C. Y. Goh, S. Lee, and H. Myung, "Robust vehicle localization using entropy-weighted particle filter-based data fusion of vertical and road intensity information for a large scale urban area," *IEEE Robotics and Automation Letters*, vol. 2, no. 3, pp. 1518–1524, 2017.

[97] L. Weng, M. Yang, L. Guo, B. Wang, and C. Wang, "Pole-based real-time localization for autonomous driving in congested urban scenarios," in *IEEE International Conference on Real-time Computing and Robotics*, 2018, pp. 96–101.

[98] G. Wan, X. Yang, R. Cai, H. Li, Y. Zhou, H. Wang, and S. Song, "Robust and precise vehicle localization based on multi-sensor fusion in diverse city scenes," in *IEEE International Conference on Robotics and Automation*, 2018, pp. 4670–4677.

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

[100] F. Ghallabi, G. El-Haj-Shhade, M.-A. Mittet, and F. Nashashibi, "LiDAR-based road signs detection for vehicle localization in an hd map," in *IEEE Intelligent Vehicles Symposium*, 2019, pp. 1484–1490.

[101] F. Ghallabi, M.-A. Mittet, G. El-Haj-Shhade, and F. Nashashibi, "Lidar-based high reflective landmarks (HRL)s for vehicle localization in an HD map," in *IEEE Intelligent Transportation Systems Conference*, 2019, pp. 4412–4418.

[102] L. Li, M. Yang, H. Li, C. Wang, and B. Wang, "Robust localization for intelligent vehicles based on compressed road scene map in urban environments," in *IEEE Transactions on Intelligent Vehicles*, 2022.

[103] A. Schaefer, D. Büscher, J. Vertens, L. Luft, and W. Burgard, "Long-term urban vehicle localization using pole landmarks extracted from 3-d lidar scans," in *European Conference on Mobile Robots*, 2019, pp. 1–7.

[104] A. Schlichting, F. Zachert, and D. Forouher, "Map matching for vehicle localization based on serial lidar sensors," in *IEEE Intelligent Transportation Systems Conference*, 2019, pp. 1257–1262.

[105] S. Z. Ahmed, V. B. Saputra, S. Verma, K. Zhang, and A. H. Adiwahono, "Sparse-3D LiDAR outdoor map-based autonomous vehicle localization," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2019, pp. 1614–1619.

[106] Z. Wang, J. Fang, X. Dai, H. Zhang, and L. Vlacic, "Intelligent vehicle self-localization based on double-layer features and multilayer LiDAR," *IEEE Transactions on Intelligent Vehicles*, vol. 5, no. 4, pp. 616–625, 2020.

[107] X. Chen, T. Läbe, L. Nardi, J. Behley, and C. Stachniss, "Learning an overlap-based observation model for 3D LiDAR localization," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2020, pp. 4602–4608.

[108] L. Li, M. Yang, L. Guo, C. Wang, and B. Wang, "Hierarchical neighborhood based precise localization for intelligent vehicles in urban environments," *IEEE Transactions on Intelligent Vehicles*, vol. 1, no. 3, pp. 220–229, 2016.

[109] A. Kasmi, J. Laconte, R. Aufrère, D. Denis, and R. Chapuis, "End-to-end probabilistic ego-vehicle localization framework," *IEEE Transactions on Intelligent Vehicles*, vol. 6, no. 1, pp. 146–158, 2020.

[110] G. Chen, F. Lu, Z. Li, Y. Liu, J. Dong, J. Zhao, J. Yu, and A. Knoll, "Pole-curb fusion based robust and efficient autonomous vehicle localization system with branch-and-bound global optimization and local grid map method," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 11, pp. 11 283–11 294, 2021.

[111] X. Chen, T. Lbe, A. Milioto, T. Rhling, O. Vysotska, A. Haag, J. Behley, and C. Stachniss, "Overlapnet: Loop closing for lidar-based slam," in *Robotics: Science and Systems XVI*, 2020.

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

[113] M. Haklay and P. Weber, "Openstreetmap: User-generated street maps," *IEEE Pervasive Computing*, vol. 7, no. 4, pp. 12–18, 2008.

[114] N. Karlsson, E. Di Bernardo, J. Ostrowski, L. Goncalves, P. Pirjanian, and M. E. Munich, "The vSLAM algorithm for robust localization and mapping," in *IEEE International Conference on Robotics and Automation*, 2005, pp. 24–29.

[115] H. Barrow, J. Tenenbaum, R. Bolles, and H. Wolf, "Parametric correspondence and chamfer matching: two new techniques for image matching," in *International Joint Conference on Artificial intelligence*, 1977, pp. 659–663.

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

[117] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide-baseline stereo from maximally stable extremal regions," *Image and vision computing*, vol. 22, no. 10, pp. 761–767, 2004.

[118] P. Wang, R. Yang, B. Cao, W. Xu, and Y. Lin, "DeLS-3D: Deep localization and segmentation with a 3d semantic map," in *IEEE Conference on Computer Vision and Pattern Recognition*, June 2018.

[119] H. Yin, L. Tang, X. Ding, Y. Wang, and R. Xiong, "LocNet: Global localization in 3d point clouds for mobile vehicles," in *IEEE Intelligent Vehicles Symposium*, 2018, pp. 728–733.

[120] N. Engel, S. Hoermann, M. Horn, V. Belagiannis, and K. Dietmayer, "DeepLocalization: Landmark-based self-localization with deep neural networks," in *IEEE Intelligent Transportation Systems Conference*, 2019, pp. 926–933.

[121] W. Lu, Y. Zhou, G. Wan, S. Hou, and S. Song, "L3-Net: Towards learning based lidar localization for autonomous driving," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 6389–6398.

[122] I. A. Barsan, S. Wang, A. Pokrovsky, and R. Urtasun, "Learning to localize using a LiDAR intensity map," in *Conference on Robot Learning*, 2018, pp. 605–616.

[123] D. Liu, Y. Cui, X. Guo, W. Ding, B. Yang, and Y. Chen, "Visual localization for autonomous driving: Mapping the accurate location in the city maze," in *International Conference on Pattern Recognition*, 2021, pp. 3170–3177.

[124] L. Yan, Y. Cui, Y. Chen, and D. Liu, "Hierarchical attention fusion for geo-localization," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2021, pp. 2220–2224.

[125] R. Kesten, M. Usman, J. Houston, T. Pandya, K. Nadhamuni, A. Ferreira, M. Yuan, B. Low, A. Jain, P. Ondruska *et al.*, "Lyft level 5 AV dataset 2019," 2019.

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

[127] P. Xiao, Z. Shao, S. Hao, Z. Zhang, X. Chai, J. Jiao, Z. Li, J. Wu, K. Sun, K. Jiang *et al.*, "Pandaset: Advanced sensor suite dataset for autonomous driving," in *IEEE International Intelligent Transportation Systems Conference*, 2021, pp. 3095–3101.

[128] M.-F. Chang, J. Lambert, P. Sangkloy, J. Singh, S. Bak, A. Hartnett, D. Wang, P. Carr, S. Lucey, D. Ramanan *et al.*, "Argoverse: 3d tracking and forecasting with rich maps," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 8748–8757.

[129] P. Sun, H. Kretzschmar, X. Dotiwalla, A. Chouard, V. Patnaik, P. Tsui, J. Guo, Y. Zhou, Y. Chai, B. Caine *et al.*, "Scalability in perception for autonomous driving: Waymo open dataset," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 2446–2454.

[130] X. Huang, X. Cheng, Q. Geng, B. Cao, D. Zhou, P. Wang, Y. Lin, and R. Yang, "The apolloscape dataset for autonomous driving," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2018, pp. 954–960.

[131] W. Wen, Y. Zhou, G. Zhang, S. Fahandezh-Saadi, X. Bai, W. Zhan, M. Tomizuka, and L.-T. Hsu, "Urbanloco: A full sensor suite dataset for mapping and localization in urban scenes," in *IEEE International Conference on Robotics and Automation*, 2020, pp. 2310–2316.

[132] J. Geyer, Y. Kassahun, M. Mahmudi, X. Ricou, R. Durgesh, A. S. Chung, L. Hauswald, V. H. Pham, M. Mühlegg, S. Dorn *et al.*, "A2D2: Audi autonomous driving dataset," *arXiv preprint arXiv:2004.06320*, 2020.

[133] Q.-H. Pham, P. Sevestre, R. S. Pahwa, H. Zhan, C. H. Pang, Y. Chen, A. Mustafa, V. Chandrasekhar, and J. Lin, "A* 3d dataset: Towards autonomous driving in challenging environments," in *IEEE International Conference on Robotics and Automation*, 2020, pp. 2267–2273.

[134] W. Maddern, G. Pascoe, C. Linegar, and P. Newman, "1 year, 1000 km: The oxford robotcar dataset," *The International Journal of Robotics Research*, vol. 36, no. 1, pp. 3–15, 2017.

[135] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The kitti dataset," *The International Journal of Robotics Research*, vol. 32, no. 11, pp. 1231–1237, 2013.

[136] J. Behley, M. Garbade, A. Milioto, J. Quenzel, S. Behnke, C. Stachniss, and J. Gall, "SemanticKITTI: A dataset for semantic scene understanding of lidar sequences," in *IEEE/CVF International Conference on Computer Vision*, 2019, pp. 9297–9307.

[137] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, "The Cityscapes dataset for semantic urban scene understanding," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2016, pp. 3213–3223.

[138] F. Yu, H. Chen, X. Wang, W. Xian, Y. Chen, F. Liu, V. Madhavan, and T. Darrell, "Bdd100k: A diverse driving dataset for heterogeneous multitask learning," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 2636–2645.

[139] Z. Wang, S. Ding, Y. Li, J. Fenn, S. Roychowdhury, A. Wallin, L. Martin, S. Ryvola, G. Sapiro, and Q. Qiu, "Cirrus: A long-range bi-pattern lidar dataset," in *IEEE International Conference on Robotics and Automation*, 2021, pp. 5744–5750.

[140] Y. Choi, N. Kim, S. Hwang, K. Park, J. S. Yoon, K. An, and I. S. Kweon, "Kaist multi-spectral day/night data set for autonomous and assisted driving," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 3, pp. 934–948, 2018.

[141] Y. Cui, Z. Cao, Y. Xie, X. Jiang, F. Tao, Y. V. Chen, L. Li, and D. Liu, "Dg-labeler and dgl-mots dataset: Boost the autonomous driving perception," in *IEEE/CVF Winter Conference on Applications of Computer Vision*, 2022, pp. 58–67.

[142] H. Caesar, J. Kabzan, K. S. Tan, W. K. Fong, E. Wolff, A. Lang, L. Fletcher, O. Beijbom, and S. Omari, "nuPlan: A closed-loop ml-based planning benchmark for autonomous vehicles," *arXiv preprint arXiv:2106.11810*, 2021.

[143] W. Maddern, G. Pascoe, M. Gadd, D. Barnes, B. Yeomans, and P. Newman, "Real-time kinematic ground truth for the oxford robotcar dataset," *arXiv preprint arXiv:2002.10152*, 2020.

[144] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, "Pointnet: Deep learning on point sets for 3D classification and segmentation," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 652–660.

[145] D. Liu, Y. Cui, L. Yan, C. Mousas, B. Yang, and Y. Chen, "Densernet: Weakly supervised visual localization using multi-scale feature aggregation," in *AAAI Conference on Artificial Intelligence*, vol. 35, no. 7, 2021, pp. 6101–6109.

[146] D. Anguelov, C. Dulong, D. Filip, C. Frueh, S. Lafon, R. Lyon, A. Ogale, L. Vincent, and J. Weaver, "Google street view: Capturing the world at street level," *IEEE Computer*, pp. 32–38, 2010.

[147] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, 2017.

[148] B. Ranft and C. Stiller, "The role of machine vision for intelligent vehicles," *IEEE Transactions on Intelligent vehicles*, vol. 1, no. 1, pp. 8–19, 2016.

[149] J. Nidamanuri, C. Nibhanupudi, R. Assfalg, and H. Venkataraman, "A progressive review-emerging technologies for adas driven solutions," *IEEE Transactions on Intelligent Vehicles*, 2021.

[150] J. Liu, W. Xiong, L. Bai, Y. Xia, T. Huang, W. Ouyang, and B. Zhu, "Deep instance segmentation with automotive radar detection points," *IEEE Transactions on Intelligent Vehicles*, 2022.

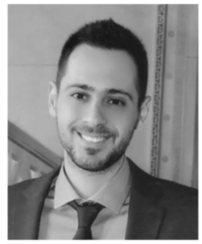
[151] Y.-P. Chang, C.-N. Liu, Z. Pei, S.-M. Lee, Y.-K. Lai, P. Han, H.-K. Shih, and W.-H. Cheng, "New scheme of lidar-embedded smart laser headlight for autonomous vehicles," *Optics express*, vol. 27, no. 20, pp. A1481–A1489, 2019.

[152] A. Kendall and Y. Gal, "What uncertainties do we need in bayesian deep learning for computer vision?" *Advances in neural information processing systems*, vol. 30, 2017.

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

[154] G. Zhang, W. Wen, and L.-T. Hsu, "A novel GNSS based V2V cooperative localization to exclude multipath effect using consistency checks," in *IEEE/ION Position, Location and Navigation Symposium*, 2018, pp. 1465–1472.

[155] M. Alrabeiah, A. Hredzak, and A. Alkhateeb, "Millimeter wave base stations with cameras: Vision-aided beam and blockage prediction," in *IEEE Vehicular Technology Conference*, 2020, pp. 1–5.



Athanasios Chalvatzaras received his Diploma and his MSc. (Hons.) from the Department of Informatics and Telecommunications, NKUA, Greece, in 2017 and 2019, respectively. He is currently pursuing his Ph.D. from the Department of Production and Management Engineering of the Democritus University of Thrace focusing his studies on Localization of Autonomous Vehicles and Machine Learning.

Ioannis Pratikakis is Professor in the Department of Electrical and Computer Engineering at Democritus University of Thrace in Xanthi, Greece. He received the Ph.D. degree in 3D Image analysis from the Electronics Engineering and Informatics department at Vrije Universiteit Brussel, Belgium, in January 1999. From March 1999 to March 2000, he joined IRISA in Rennes, France as an INRIA postdoctoral fellow. From January 2003 to June 2010, he was working as Adjunct Researcher at the Institute of Informatics and Telecommunications in the National Centre for Scientific Research "Demokritos", Athens, Greece. His research interests lie in machine learning, vision, image processing and graphics. He has published more than 200 papers in journals, book chapters and conference proceedings in the above areas for which he has received more than 8600 references with h-index:48 (source: Google Scholar). He has served as Guest editor in several journals including the International Journal of Computer Vision, The Visual Computer as well as the Computers & Graphics. He is member of the steering committee of EG Workshop on 3D Object Retrieval while he has been member of the Board of the Hellenic Artificial Intelligence Society for the period 2010-2014. He is Senior Member of the IEEE, member of the board of IEEE Greece Section (2013-2019) and member of the European Association for Computer Graphics (Eurographics).

Angelos Amanatiadis (S'00–GS'04–M'09–SM'16) is an Assistant Professor in the Department of Production and Management Engineering at Democritus University of Thrace (DUTH). He received his Diploma and his Ph.D. (Hons.) from the Department of Electrical and Computer Engineering, DUTH, Greece, in 2004 and 2009, respectively. He has published more than 70 scientific papers in the area of autonomous robotic systems, computer vision, machine learning, and real-time embedded systems. He was awarded as one of the Transactions "Outstanding Reviewers" in appreciation of outstanding service to the IEEE Instrumentation and Measurement Society. He has received the Stavros Niarchos Award for Promising Young Scientists and a State Scholarships Foundation (IKY) Postdoc scholarship.