Vision-Enhanced Low-Cost Localization In Crowdsourced Maps

GNSS-BASED POSITIONING

Vision-Enhanced Low-Cost Localization in Crowdsourced Maps

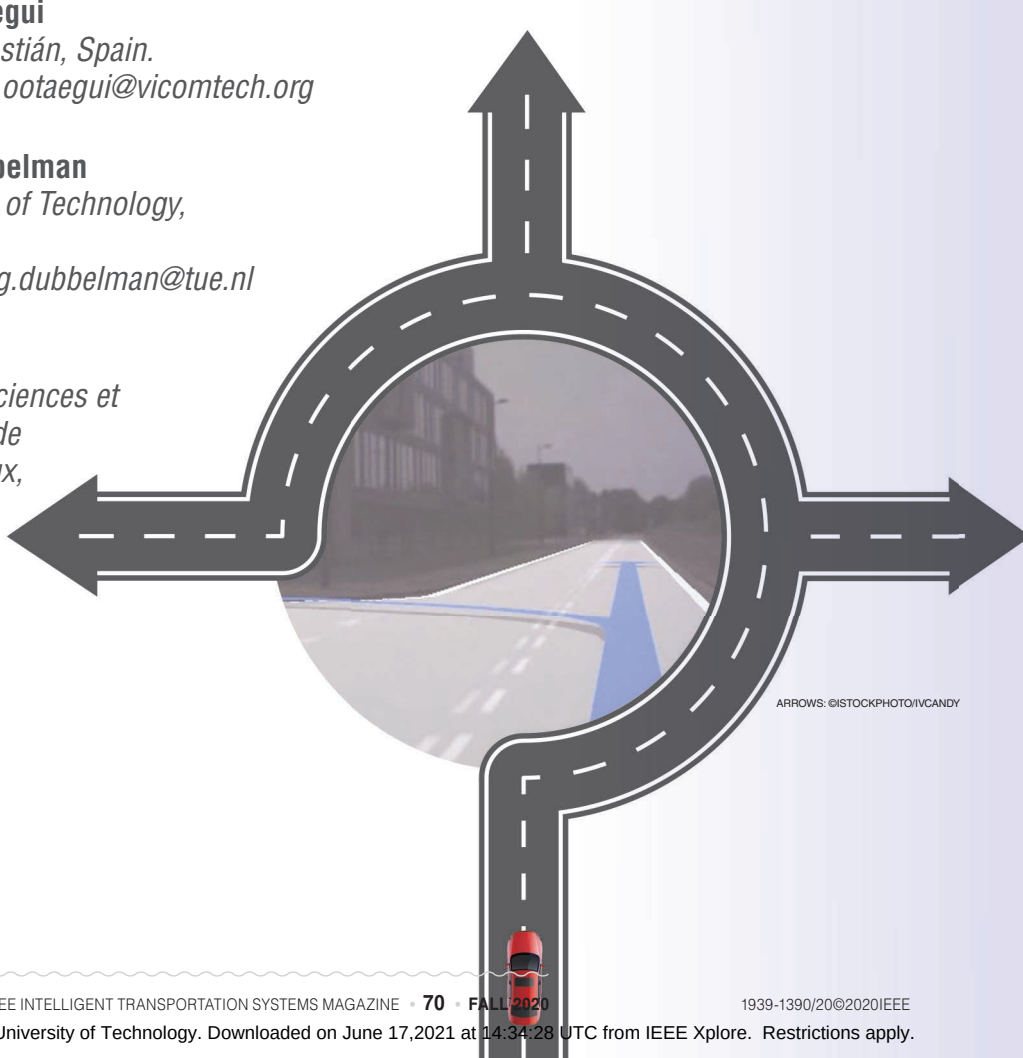
Benedict Flade and Julian Eggert
Are with the Honda Research Institute Europe GmbH, Offenbach, Germany.
Email: benedict.flade@honda-ri.de; julian.eggert@honda-ri.de

Axel Koppert
Is with the OHB Digital Solutions GmbH, Graz, Austria.
Email: axel.koppert@ohb-digital.at

Gorka Velez and Oihana Otaegui
Are with Vicomtech, San Sebastián, Spain.
Email: gvelez@vicomtech.org; ootaegui@vicomtech.org

Anweshan Das and Gijs Dubbelman
Are with Eindhoven University of Technology, Eindhoven, The Netherlands.
Email: anweshan.das@tue.nl; g.dubbelman@tue.nl

David Bétaille
Is with Institut Français des Sciences et Technologies des Transports de l'Aménagement et des Réseaux, Bouguenais, France.
Email: david.betaille@ifsttar.fr



ARROWS: ©ISTOCKPHOTO/IVCANDY

Digital Object Identifier 10.1109/MITS.2020.2994055
Date of current version: 11 June 2020

1939-1390/20©2020IEEE

***Abstract*—The lane-level localization of vehicles with low-cost sensors is a challenging task. In situations in which Global Navigation Satellite Systems (GNSSs) suffer from weak observation geometry or from the influence of reflected signals, the fusion of heterogeneous information presents a suitable approach for improving the localization accuracy. We propose a solution based on a monocular front-facing camera, a low-cost inertial measurement unit (IMU), and a single-frequency GNSS receiver. The sensor data fusion is implemented as a tightly coupled Kalman filter that corrects the IMU-based trajectory with GNSS observations while employing European Geostationary Overlay Service correction data. Further, we consider vision-based complementary data that serve as an additional source of information. In contrast to other approaches, the camera is not used to infer the motion of the vehicle, but rather for directly correcting the localization results under the usage of map information. More specifically, the so-called *camera-to-map alignment* is done by comparing virtual 3D views (candidates) created from projected map data with lane geometry features that are extracted from the camera image. One strength of the proposed solution is its compatibility with state-of-the-art map data, which are publicly available from different sources. We validate the approach on real-world data recorded in The Netherlands and show that it presents a promising and cost-efficient means to support future advanced driver assistance systems.**

Situation awareness is a crucial component for intelligent assistance systems and autonomous cars. Predicting the evolution of a traffic situation allows for the judgement of a critical situation and is the key to taking actions that mitigate danger. An accurate prediction requires assumptions about possible driving paths that can only be defined if surrounding cars are located and assigned to particular driving directions. A common approach is to generate a local dynamic map (LDM) that contains representations of both dynamic and static elements of the driving scene. However, one of the biggest challenges is achieving a localization accuracy sufficient to reliably assign vehicles to lanes stored in the map. Aside from being essential for situation awareness, tackling this challenge facilitates further technologies, such as lane-level navigation, which is why accurate localization can be seen as a key enabler for future advanced driver assistance systems (ADASs).

With an accuracy on the order of meters [1], standard Global Navigation Satellite Systems (GNSSs) meet the requirements for road-level navigation but are not suitable for localization within a lane. The requirements for lane-level navigation have been addressed within a prenormalization project with the title Satellite Positioning Performance Assessment for Road Transport (SaPPART) [2]. There, the horizontal accuracy requirements for lane-level navigation are specified to lie between several decimeters and one meter (95th percentile), which aligns with the accuracy targeted in our research. Differential GNSS techniques, including real-time kinematic (RTK) positioning, allow for the improvement of absolute accuracy beyond that level, but under specific conditions of use (e.g., in the presence of multipath effects, which are typical for urban environments) none of the local effects can be mitigated by using differential techniques.

To improve the accuracy and availability of the GNSS, it can be combined with additional technology based on other sensors. A common approach is to fuse inertial measurement units (IMUs) with the GNSS to obtain low-cost localization solutions. The IMU provides positioning when no GNSS signal is available and enables continuous localization. However, the accuracy of such systems is limited by the precision of the GNSS receiver. IMUs provide relative measurements that are accurate for a short period but drift over longer periods of time. The achievable accuracy is on the order of meters to centimeters (for RTK) while the drifts during satellite outages mainly depend on the grade of the inertial components [3].

The best results in terms of localization accuracy have been reported using lidar or radar-based techniques in combination with GNSSs and IMU [4]. However, lidar sensors are expensive and a single lidar can easily exceed the price of the vehicle itself.

Camera-based technology is more cost efficient and can be used in combination with the GNSS and IMUs to localize a vehicle in a map with lane-level accuracy [5], [6]. One common characteristic of such approaches is the requirement for suitable underlying map geometry data, which can be stored in an LDM [7]. The definition of what to store continuously progresses along with research investigations related to automated driving [8]. In terms of contents, vectors (points, polylines, splines, or clothoides) describe natural and artificial objects such as center lines, road markings, borders, landmarks, or traffic signs. Point clouds that aggregated into different vectors, allowing for detailed 3D representations of the environment, present an additional type of content [4].

> Global accuracy is crucial for any use of the GNSS positioning within map-based applications, especially in the case of cooperative entities that share information based on maps that come from different makers.

The positioning accuracy of these objects is twofold: locally, i.e., with respect to each other, and globally, i.e., with regard to a global reference system. Global accuracy is crucial for any use of the GNSS positioning within map-based applications, especially in the case of cooperative entities that share information based on maps that come from different makers.

A wide range of research actors generated high-definition (HD) maps with local and global subdecimeter accuracy. These maps, which used to be rare experimental products [5], [9]–[11], cover specific areas or kilometers of highways. Although map makers are extending the coverage of their HD map portfolio, their products focus on certain cities or major roads and are not globally available.

In mobile mapping, the key issue everyone is facing remains the accuracy of vehicle positioning. To date, most of the prototypes of industrial products do not document their global accuracy. There still are spots in time and space, particularly in deep urban centers, where even sophisticated GNSS approaches such as post-processing kinematic (PPK) combined with IMU are not sufficient, or worse, not reliable. High-resolution aerial images by national mapping agencies are another source of raw data, possibly used jointly [12], [13]. By exploiting such aerial images, it is possible to achieve submeter localization accuracy.

It is obvious that the localization of a vehicle in a map requires not only accurate absolute positioning with respect to a global reference frame but also a method for accurately localizing the vehicle relative to the map. In this context, map-relative localization can be divided into two classes: 1) lane assignment, which refers to the estimation of the current ego-lane index [14], and 2) map alignment, which estimates the map-relative ego-vehicle position, described in Cartesian or ellipsoidal coordinates [15]–[17].

Rabe et al. investigate particle filter-based lane assignments while approaching intersections [18]. Ballardini et al. use hidden Markov models and line-detection algorithms to perform a lane assignment on highways [19].

Hu and Uchimura present an approach for map alignment in which they generate potential virtual views based on digital map data that are compared to features detected in the camera image [20]. Candidate-based alignment approaches that employ lane-detection methods are presented in [21]. More recently, Caselitz et al. show how to employ a visual odometry-based system in combination with panoramic lidar maps [22]. In their 3D map alignment approach, they use a local bundle adjustment to reconstruct 3D points from image features that are then matched to the map data. Similar to us, they use a monocular camera for localization, with the difference that our research is based on 2D open-source map data. An approach based on lane markings to improve low-cost GPS, using a camera and open-source map data, is introduced by Lu et al. in [23].

In this article, we present a novel localization method based on two steps (see Figure 1). The first step covers the fusion of the GNSS and IMU to obtain a preliminary position estimation. We employ a multiconstellation GNSS approach (GPS and Galileo) augmented with the European Geostationary Overlay Service (EGNOS) corrections, which results in a meter accurate absolute position that is the starting point of the second step. In this second step, a camera-to-map alignment (C2MA) algorithm is used to realize map-relative localization.

What distinguishes our work from related research is the integration of a tightly coupled GNSS/IMU filter specifically tailored to the application using EGNOS and Galileo in combination with C2MA. The latter leverages the advantages of an LDM implementation while only requiring enhanced OpenStreetMap (OSM) data. One strength of our novel combination of components is the usage of visual alignment in the perspective view, which can be the basis for further technologies, such as augmented-reality ADAS.

The purpose of our method is to enable low-cost localization within a map that can be used, e.g., for situational awareness and decision making in driver assistance and autonomous driving applications. Our approach is low cost in two critical aspects: 1) the generation of the map, building upon available crowdsourced data instead of generating costly ad hoc 3D maps, and 2) the sensorization of the localization system, requiring only a GNSS receiver, an IMU, and a monocular camera. The advantage of this sensor setup is its availability in a wide range of current production vehicles that could be equipped with our system at almost no additional cost. There are no conceptual issues that hinder our approach to run on entry- and middle-class vehicles in the near future.

Methods

The location of an object can be defined with respect to an underlying map. In this case, the map serves as a reference frame, hereafter referred to as *map frame* (*m*-frame, superscript *m*), which can be accessed by using the map

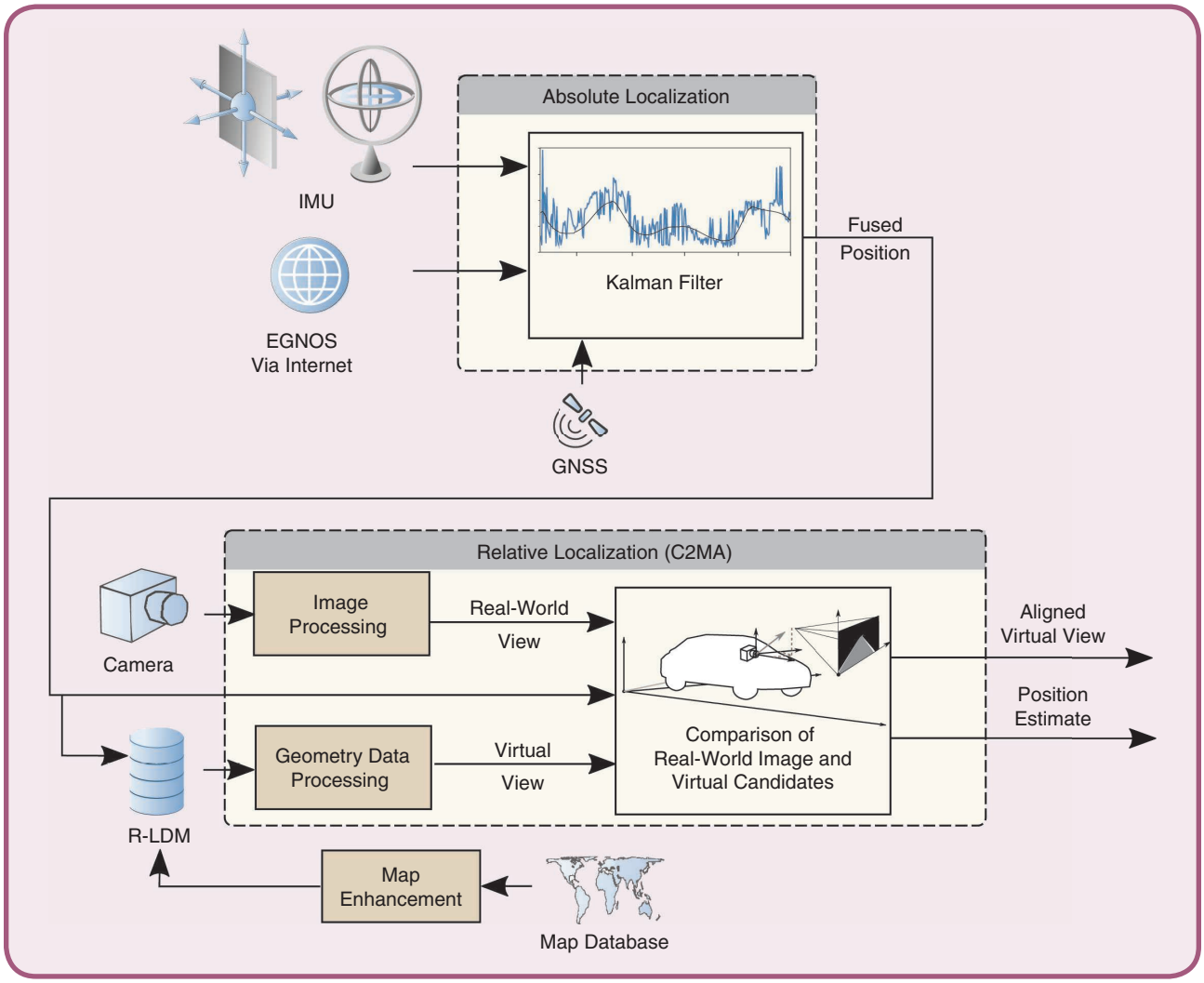


FIG 1 A pipeline of the localization approach. The first step is the fusion of GNSS, EGNOS, and IMU data. The second step is the camera-to-map alignment (C2MA) using publicly available road geometry information.

coordinates of the mapped features. Thus, any localization technique that uses the coordinates of mapped features is sensitive to the position in the map frame. Assuming that a map is locally accurate while not guaranteeing global accuracy, the coordinates of the objects in the m-frame may deviate from their corresponding coordinates determined by the GNSS-based methods. The GNSS provides positions in the terrestrial reference frame (e-frame, superscript e) that are accessible through satellite positions, which are given in the e-frame. Local map accuracy is characterized by transformations between the e-frame and the m-frame coordinate for neighboring features of the map.

Consider a GNSS-based position estimate in the e-frame $\boldsymbol{p}^e$ that will be transformed to a map-relative position estimate $\boldsymbol{p}^m$. Introducing a translation parameter $\boldsymbol{d}^e$, we can express the relationship between $\boldsymbol{p}^e$ and $\boldsymbol{p}^m$ as

$$\boldsymbol{p}^m = \boldsymbol{p}^e + \boldsymbol{d}^e. \tag{1}$$

This relation presents the basis for our map-relative localization procedure. The first step is the estimation of $\boldsymbol{p}^e$ using the GNSS and the second step is aligning this position with the map to find $\boldsymbol{d}^e$, with which we obtain $\boldsymbol{p}^m$.

Note that using vision-based map alignment starting from a GNSS-based position estimate $\boldsymbol{p}^e$, we observe $\boldsymbol{d}^e$ as the superposition of a bias in the GNSS-based estimate of $\boldsymbol{p}^e$ and of the map offset, which cannot be separated. Yet, when recognizing the possibility of a local shift between the m-frame and the e-frame, we can still estimate the map-relative positions using sensors that are sensitive to positions in at least one of the two frames, provided that $\boldsymbol{d}^e$ is observable with those sensors.

GNSS-Based Position Estimation

We obtain an estimate of $\boldsymbol{p}^e$ by exploiting a tightly coupled extended Kalman filter, which fuses the GNSS and IMU observations. The GNSS observations (pseudorange and

Doppler) are used to correct the predictions of the vehicle state $\boldsymbol{x}$ based on IMU measurements.

We choose the following state vector to describe the vehicle kinematics and sensor-specific states:

$$\boldsymbol{x} = \left[\boldsymbol{p}^e \;\; \boldsymbol{v}^l \;\; \boldsymbol{\psi} \;\; \boldsymbol{b}_a^b \;\; \boldsymbol{b}_g^b \;\; \boldsymbol{\delta t} \;\; \dot{\delta t}\right]^T. \tag{2}$$

The vehicle kinematics are described by their position in the e-frame $\boldsymbol{p}^e$ (expressed in Cartesian coordinates), the velocity $\boldsymbol{v}^l$ in the local-level frame (l-frame, north, east, down), and the Euler angles $\boldsymbol{\psi}$ (roll, pitch, yaw) extracted from $\boldsymbol{R}_b^l$. $\boldsymbol{R}_b^l$ is the rotation matrix describing the rotation between the axes of the body frame defined by the axes of the IMU (b-frame, superscript b) and the axes of the l-frame.

The IMU accelerometer and gyroscope are known for showing a bias that needs to be estimated during operation: $\boldsymbol{b}_a^b$ is the accelerometer bias and $\boldsymbol{b}_g^b$ is the gyroscope bias, and both are expressed in the b-frame. $\boldsymbol{\delta t}$ and $\dot{\delta t}$ are the clock biases and the clock drift of the GNSS receiver. For each GNSS s in use, a separate receiver clock error δt_s has to be estimated due to different time system realizations. The following subsequent differential equations describe the state kinematics:

$$\begin{aligned}
\dot{\boldsymbol{p}}^e &= \boldsymbol{R}_l^e \boldsymbol{v}^l \\
\dot{\boldsymbol{v}}^l &= \boldsymbol{R}_b^l(\boldsymbol{f}^b) + \boldsymbol{\gamma}^l(\boldsymbol{p}^e) \\
\dot{\boldsymbol{R}}_b^l &= [\boldsymbol{\omega}_{ib}^b]_\times \boldsymbol{R}_b^l \\
\dot{\boldsymbol{b}}_a^b &= \boldsymbol{w}_a \\
\dot{\boldsymbol{b}}_g^b &= \boldsymbol{w}_g \\
\dot{\delta t}_s &= \dot{\delta t} \\
\ddot{\delta t} &= w_t.
\end{aligned} \tag{3}$$

$\boldsymbol{f}^b$ presents the specific force and $\boldsymbol{\omega}_{ib}^b$ is the turn rate in the b-frame. $\boldsymbol{\gamma}^l(\boldsymbol{p})$ stands for the gravity at $\boldsymbol{p}^e$ expressed in the l-frame. $\boldsymbol{w}_a$ and $\boldsymbol{w}_g$ are the random noises affecting the sensor biases and w_t is the random noise of the GNSS receiver clock drift. Note that this is a simplified model, as we neglect the effect of the Earth's rotation.

An IMU is used to measure $\boldsymbol{f}^b$ and $\boldsymbol{\omega}_{ib}^b$. We model the measurements as

$$\begin{aligned}
\boldsymbol{f}^b &= \boldsymbol{f}_m^b - \boldsymbol{b}_a^b + \boldsymbol{w}_f \\
\boldsymbol{\omega}_{ib} &= \boldsymbol{\omega}_{ib,m}^b - \boldsymbol{b}_g^b + \boldsymbol{w}_{\omega,}
\end{aligned} \tag{4}$$

where $\boldsymbol{f}_m^b$ is the measured specific force, $\boldsymbol{\omega}_{ib,m}^b$ is the measured turn rate, $\boldsymbol{b}_a^b$ and $\boldsymbol{b}_g^b$ are the accelerometer and gyroscope biases, and $\boldsymbol{w}_f$ and $\boldsymbol{w}_\omega$ are the random measurement noises.

A GNSS receiver provides pseudorange, phase, and Doppler observations. Here, we consider low-cost receivers that provide single-frequency multi-GNSS observations. All necessary information for computing a position (satellite orbits, satellite clock corrections, and ionospheric model) is contained in the satellite signal. However, the quality of the broadcast data limits the achievable accuracy. A satellite-based augmentation system (SBAS) provides corrections to the broadcast data and an accurate ionosphere model, which enables meter-level accuracy for single-frequency users. EGNOS is the European SBAS. Its data can be received via either satellite or the Internet-based EGNOS Data Access Service (EDAS) [24].

The observations are influenced by the receiver–satellite geometry as well as by several other effects that need to be modeled. Thus, the following corrections are applied to the measured pseudoranges ρ.

- The satellite clock offset is computed from the broadcast data.
- The ionospheric delay for the L1 and E1 signals are computed from the EGNOS ionosphere model.
- The tropospheric delay is computed from the SBAS Minimum Operational Performance Standard troposphere model.

For the GPS pseudoranges, we also apply the long-term clock correction and the fast corrections provided by EGNOS. The satellite positions $\boldsymbol{p}_{\text{sat}}^e$ and velocities $\boldsymbol{v}_{\text{sat}}^e$ are computed from the broadcast ephemeris. For GPS satellites, we apply the EGNOS long-term orbit corrections. The Doppler observations $\dot{\rho}$ are corrected for the satellite clock drift δt_{sat}.

The observation models for the corrected pseudoranges $\rho_{\text{corr, sat}}$ and the corrected Doppler observations $\dot{\rho}_{\text{corr, sat}}$ are

$$\rho_{\text{corr, sat}} = \|\boldsymbol{p}_{\text{sat}}^e - \boldsymbol{p}^e\| + \delta t_s + n_{\rho,\text{sat}} \tag{5}$$

$$\dot{\rho}_{\text{corr, sat}} = (\boldsymbol{v}_{\text{sat}}^e - \boldsymbol{R}_l^e \boldsymbol{v}^l)^T \frac{\boldsymbol{p}_{\text{sat}}^e - \boldsymbol{p}^e}{\|\boldsymbol{p}_{\text{sat}}^e - \boldsymbol{p}^e\|} + \dot{\delta t} + n_{\dot{\rho},\text{sat}}, \tag{6}$$

where $n_{\rho,\text{sat}}$ and $n_{\dot{\rho},\text{sat}}$ are the measurement noises of the pseudorange and Doppler.

We use (3) to derive our dynamic system model for the Kalman filter. The system model uses the IMU measurements in (4) to predict the state. The measurement model is given by (5) and (6) and is used for correcting the prediction whenever a set of the GNSS observations arrives. An important aspect of the correction step is measurement selection and weighting. We determine the weights of the observations based on the carrier-to-noise ratio reported by the receiver. Furthermore, we exclude erroneous measurements using innovation gating.

Vision-Based C2MA

While the previous section focused on localization in a global frame, we now present the C2MA approach, which allows for localization with regard to a local frame, i.e., with respect to the map. To estimate such map-relative vehicle positions, the basic idea is to exploit visual cues detected within the camera image, such as lane markings, curbs, or road texture, which are then compared to road

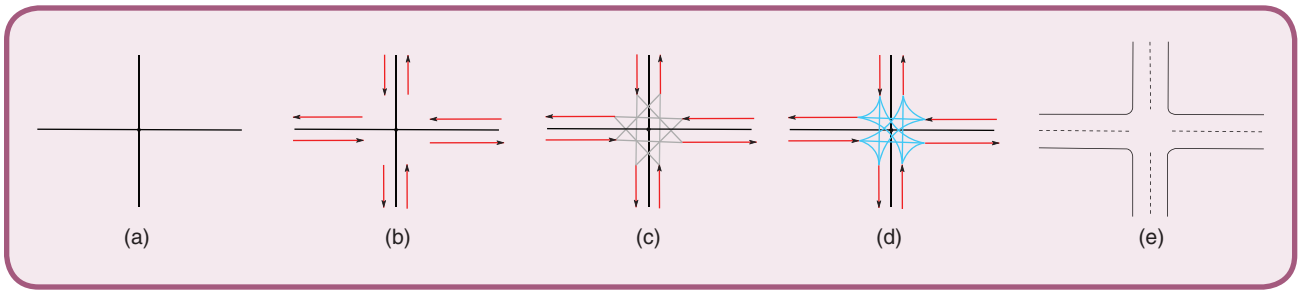


FIG 2 An enhancement of map data: (a) center of the road, (b) center of lanes, (c) topological connections, (d) geometric connections, and (e) full geometry.

geometry information derived from the map. The lower box of Figure 1 briefly summarizes the steps of the pipeline, as follows:

1) retrieval of visible map data around the GNSS estimate
2) generation of virtual candidates based on map data
3) processing of visual data from the camera
4) calculating orientation-based feature vectors for candidates and processed camera images
5) a best-match search based on a feature comparison between processed real-world data and virtual candidates.

The following sections offer brief insight into our world model, the image processing pipeline, and the candidate comparison. More detailed descriptions can be found in [25]–[28].

Environment Representation

One major component for our approach is map data. With regard to ADASs, the main demands on map data are consistency with the real world, frequent updates, high coverage, and easy editability. The three latter points are where crowdsourced and collaborative mapping projects excel. In our research, we choose map data from the community-developed OSM project [29]. However, the drawbacks of OSM are its low level of detail and its varying accuracy. The geometry data provided by the OSM project is mostly restricted to polylines that describe the center of a road [see Figure 2(a)]. Lane-level information, e.g., the number of lanes, is mostly added as a tag, but the geometric data of each individual lane or topological data on how lanes are connected is missing.

To use OSM for lane-level self-localization tasks, we need to enhance the geometry data. Knowing the number of lanes, we infer lane segment center lines based on the corresponding road center line [Figure 2(b)]. Then we connect the lane segments on the topological level and interpolate the junctions between two related lane segments on geometric level [Figure 2(c) and (d)]. As a final step, the right and left boundary polylines are added, which leads to full lane-level geometry [Figure 2(e)].

With ADAS moving from an ego-centered perspective toward systems that additionally consider the environment around the vehicle, a coherent world model must be implemented. In this context, we use our relational LDM (R-LDM) presented, e.g., in [27], which allows for receiving, integrating, storing, fusing, updating, and predicting ADAS-related data.

Similar to the LDM of the SAFESPOT integrated project [7], our LDM consists of four layers that group entities based on their level of dynamics, with road geometry information on its lowest layer. We want to emphasize that the static layer of our concept is not limited to enhanced OSM data, but can be used with any other forms of polyline-based map data.

Lane Detection–Based Feature Extraction

To align map and camera data, we need to extract visual cues from the camera image. Lane markings are useful visual cues since they are present and visible on most roads. Instead of explicitly mapping and storing each marking, we exploit the lane boundaries that have been generated in the map enhancement process. In early research, we applied common edge-detection algorithms, e.g., Canny edge detection, to the camera image [30]. However, such approaches tend to overestimate image features, which can lead to a large number of false positives. Therefore, we propose the use of an explicit method to detect lane markings that was initially presented in [26], which is based on an evolution of the classic top-hat filter.

In the camera image, road patches close to the camera cover more image pixels than equally sized patches further away. To counteract this, we construct a bird's-eye image using image-to-road homography, a technique also known as *inverse perspective mapping* (*IPM*) [31].

In the next step of lane-marking detection, the pairs of points that define the candidate lane marking pieces are transformed into the IPM domain, applying the image-to-plane homography to each of its points. Then, a connected-component analysis (CCA) is applied to group pairs into stripes. At this stage, stripes that do not correspond to lane markings are filtered out.

The output of the algorithm is a set of polylines in the camera image domain that define the detected lane markings and can be used as visual cues for the alignment

algorithm. Since most information value is contained in the lower part of the image, we define an area of interest by cutting the upper part of the camera image at position u (Figure 3). As long as the cut area contains no lane marking, this does not influence the results but improves the computation time.

Candidate Generation

The idea is to create a set of map geometry-based images, hereafter called *candidates* c, which are then compared to the real-world visual data obtained from a camera. Each candidate is a virtual perspective view of the map data, which all differ in their pose $\boldsymbol{\rho}_c$. Here, we generally define the pose

$$\boldsymbol{\rho}_c = (\boldsymbol{p}_c^m, \boldsymbol{\psi}_c) \tag{7}$$

as being a subset of the state or, more specifically, the composition of the candidate's lateral, longitudinal, and vertical position $\boldsymbol{p}_c^m$ in the map frame as well as its Euler angles $\boldsymbol{\psi}_c$ (roll, pitch, and yaw), resulting in six degrees of freedom.

Each candidate c can be described by its pose difference

$$\Delta\boldsymbol{\rho}_c = \boldsymbol{\rho}_c - \boldsymbol{\rho}_0, \tag{8}$$

with $\boldsymbol{\rho}_0$ acting as the so-called *candidate zero* that represents the presumed pose. This presumed pose is constructed on the GNSS-based position inputs and serves as an initial guess and to restrict the localization error.

The distribution of the candidates can be uniform, Gaussian, or chosen according to more sophisticated optimization strategies while its search range highly depends on the expected accuracy of the presumed pose. In other words, the worse the expectation, the larger the required search range. To reduce the potential search range, we need information on the current lane, such as the lane number, as far as this information is available. This is realized by tracking lane detections in consecutive frames.

Assuming that a vehicle is neither driving off the road nor on a lane that points to the opposite direction, we can set reasonable limits for the search range. However, the assumptions of being located on the road as well as the assumption of driving in the correct direction can be incorrect in certain situations.

Candidate Comparison and Parameter Estimation

Keeping in mind that the map data are based on enhanced open-source information, we target a robust alignment solution that does not necessarily require exact lane marking positions on the pixel level. Therefore, we choose to compare the virtual views and the preprocessed camera image based on the orientation of its features. This is done by applying the histogram of oriented gradients. Here, dominant orientations are detected in local spatial regions to capture edge structures while being invariant toward local geometric and photometric transformations. The result is a flattened feature vector v that is determined for each candidate, as well as for the camera image that describes the locally dominant orientation. To compare candidates with the camera image, we choose the cosine angle related similarity score

$$m_c = \frac{\boldsymbol{v}^{\mathrm{rwv}} \cdot \boldsymbol{v}_c^{\mathrm{vv}}}{\|\boldsymbol{v}^{\mathrm{rwv}}\| \cdot \|\boldsymbol{v}_c^{\mathrm{vv}}\|}, \tag{9}$$

with $\boldsymbol{v}^{\mathrm{rwv}}$ acting as the feature vector of the real-world view and $\boldsymbol{v}_c^{\mathrm{vv}}$ as the feature vector of the cth virtual view candidate.

Using the parameters of the best match, the initially assumed pose can be updated to obtain a map aligned pose. We want to stress that the pipeline allows for an alignment of up to six degrees of freedom. However, when assuming a flat environment, the vertical position as well as roll and pitch angles can be neglected.

Combining GNSS and C2MA

The final step is the combination of the GNSS-based absolute localization and the C2MA. The starting point of

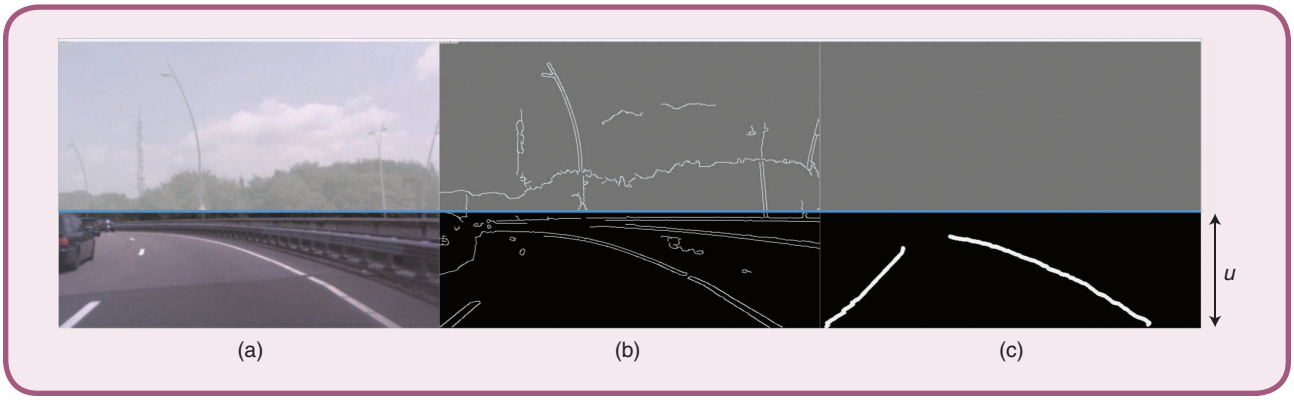


FIG 3 The highway situation (a) with its corresponding processed Canny edge (b) and lane-detection (c) images. The area of interest that is considered for the alignment is defined by u.

the alignment is the output of the Kalman filter (the top box of Figure 1). Leveraging the alignment, the translation parameter $\boldsymbol{d}^e$ can be interpreted as being the pose difference $\Delta\boldsymbol{\rho}_c$ that belongs to the candidate with the best similarity score.

$\boldsymbol{d}^e$ is applied to every output of the absolute localization module to determine a map-relative position using (1). The C2MA can be triggered either by every output of the Kalman filter or at a lower rate. If the rate of the filter output is higher than the update rate of the alignment, $\boldsymbol{d}^e$ is maintained until a newer $\Delta\boldsymbol{\rho}_c$ becomes available. The same strategy is applied if the alignment is not available, e.g., if the map situation or the image (obstruction of the relevant features) does not permit proper alignment.

To illustrate the concept and potential of the described approach, we use real-world recordings that allow us to evaluate the GNSS/IMU localization as well as the C2MA with a globally accurate map.

Experiments and Evaluation

To illustrate the concept and potential of the described approach, we use real-world recordings that allow us to evaluate the GNSS/IMU localization as well as the C2MA with a globally accurate map. To record the data sets, we use a vehicle equipped with a stereo camera composed of two PointGrey Firefly MV RGB cameras, a u-blox EVK-M8T GNSS receiver with a patch antenna, an Xsens MTi-1 IMU, and a 3G/4G Wi-Fi router to have Internet connectivity for receiving the EGNOS corrections via EDAS. Aligned with our low-cost claim, we only use one of the two installed cameras for the following experiments.

The data have been recorded using the middleware RTMaps to assure time synchronization of the data stream. We use RTMaps not only for record-and-replay testing, but also for online testing in the automobile, which allows us to reproduce similar algorithm behavior in the car and in the laboratory. We record raw GNSS observations with a rate of 2 Hz, turn rate and accelerations observed by the IMU at 100 Hz, and video at 60 fps. Those observation rates of the GNSS and IMU are chosen to capture the vehicle dynamics appropriately while keeping the computational load at a reasonable level. The GNSS/IMU filter processes the observations upon arrival and is operating at 100 Hz while the C2MA provides one position correction per second (1 Hz) during real-time execution.

Localization Accuracy in a Globally Accurate Map

For the following evaluation, we selected a data set of a 10-km drive recorded in the city of Eindhoven, The Netherlands. In this experiment, the C2MA is done frame by frame, i.e., every output from the GNSS/IMU fusion is aligned. A reference trajectory was obtained by using the high-performance GNSS/IMU system OXTS RT3000 in PPK processing.

Evaluating map-relative localization algorithms proves to be difficult, since there is no general method for finding ground truth for this task. Reference trajectories determined by the GNSS are not suitable as the ground truth for map-relative localization if the map is inaccurate with respect to the *e*-frame. Therefore, we expended effort on geometrically correcting the map (shifting lane centers and correcting lane widths) based on high-resolution satellite images.

As a baseline, we evaluate the accuracy of the tightly coupled GNSS/IMU fusion descriabed in the "GNSS-Based Position Estimation" section. The cumulative density function of the lateral errors along the trajectory is shown in Figure 4, where it is compared to the lateral errors of the positions computed by the GNSS receiver. It can be seen that the fusion approach leads to a significant improvement in the lateral accuracy. The plot of lateral errors (Figure 5) along the driven path shows several passages with degraded accuracy. In this exemplary evaluation, those passages are where the direct signal is blocked, e.g., due to trees, and only a reflected signal is received.

Figures 4 and 5 further show the reduced lateral errors when adding the C2MA step to the localization system. Overall, the 95th percentile of the lateral errors of the full system is 0.95 m with a maximum error of 1.53 m while the nonaligned system results in a 95th percentile of 1.05 m

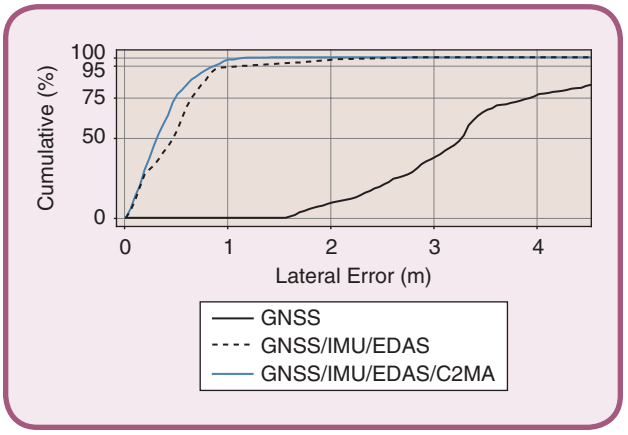


FIG 4 The cumulative density function of the lateral errors of the GNSS-only, GNSS/IMU/EDAS, and GNSS/IMU/EDAS/C2MA localization modules with reference to the PPK–GNSS trajectory.

with a maximum error of 3.07 m. Most crucial are the improvements in situations in which the nonaligned position estimate shows significant errors (e.g., at longitudinal positions of 1,000 and 2,000 m). Here, the alignment allows for significant improvement of the position estimate.

The accuracy obtained from the GNSS/IMU/EDAS/C2MA fusion meets the requirements for lane-level navigation discussed at the beginning of this article. The 95th percentile of the lateral errors has been improved to the submeter level. The maximum error (100th percentile) can be reduced to 1.5 m, which is on the order of half the width of a lane [32] and can be seen as the minimum requirement for correct lane assignment [33].

Aligned Virtual View

In addition to position estimation, our approach natively enables further applications, such as augmented reality or similar human–machine interfaces. Since the C2MA approach visually aligns road geometry data with features in the camera image, any other map-related object can be projected onto the camera image as well.

Figure 6 shows two examples of augmented real-world views. In both images, the best virtual candidate of the C2MA is depicted with additional information overlaid. On the left, the output of the alignment is used to deliver navigation-related information such as the so-called *path horizon*, which is a set of possible paths. Further, we can provide driver assistance by displaying a virtual wall that serves as a visual separation between two lanes, which are shown on the right. The color or transparency of the virtual wall can convey additional information, e.g., by correlating with the distance of the ego-vehicle to the lane boundary.

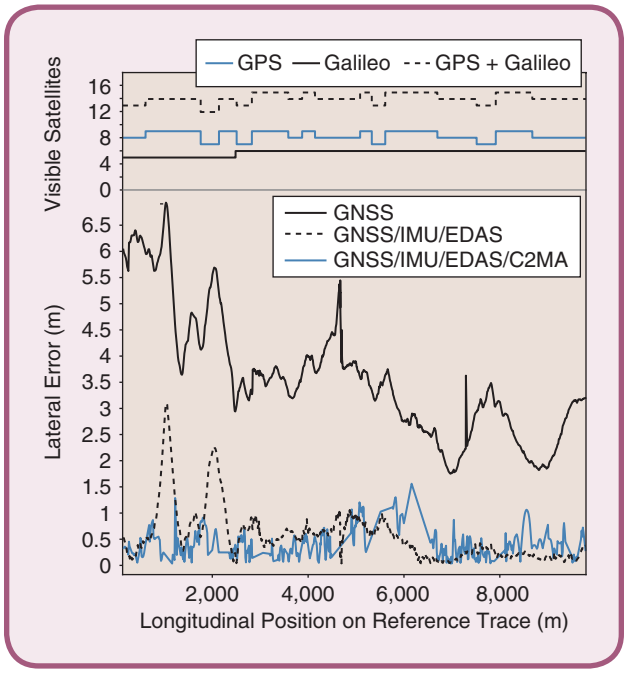


FIG 5 A comparison of the lateral error of the GNSS-only, GNSS/IMU/EDAS, and GNSS/IMU/EDAS/C2MA localization modules in a peri-urban environment. At the top of the figure, the number of visible GPS and Galileo satellites is shown for each position of the trace.

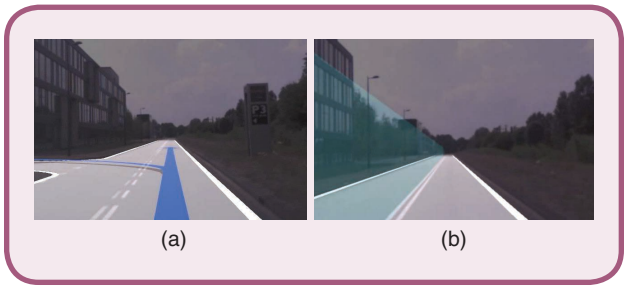


FIG 6 The application examples for the C2MA: (a) path horizon and (b) virtual wall.

Conclusion and Outlook

This article presents a novel two-step localization method based on the fusion of GNSS and IMU observations and the alignment of images from a monocular camera with open-source map geometry. The real-world capabilities of the GNSS/IMU/EDAS subsystem have already been proven during end-user lane-level navigation tests in Barcelona, Spain, within the frame of the EU INLANE project [34]. The C2MA can be understood as an add-on that was tested on real-world recordings while currently reaching a performance level of 1 Hz. Since it is implemented in Python and runs completely on the CPU, there is substantial room for optimization.

However, our first tests already reveal the potential of the alignment step to support localization, especially in situations in which satellite-based solutions struggle. We conclude that the accuracy of our map, as well as the quality of the C2MA, is suitable for improving the accuracy of the localization result with respect to the reference trajectory. The results of the evaluation lie within the range of the targeted accuracy requirements for lane-related applications.

Since the improvement is already visible in a favorable GNSS environment, we expect the improvements for more difficult conditions to be even better. However, such environments pose additional challenges. One weakness of our two-step approach is a situation in which GNSS/IMU errors cannot be corrected since the camera-to-map alignment is not available, e.g., when map geometry (complicated intersections) or the vision (occlusion of the relevant features) do not permit proper alignment. Another challenge is given in situations in which degraded GNSS accuracy leads to ambiguous alignment results.

These challenges point to future research directions. Temporal filtering of the alignment results can help to solve or avoid ambiguous situations. Using the alignment result directly as an observation in the Kalman

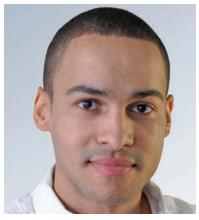
filter can enable an estimation of errors in the GNSS observations, as shown in [5], which leads to improved localization accuracy. What also remains to be evaluated is the influence of nonflat road surface on the alignment result. Currently, we are neither considering road bumps nor any other kind of hilly environment. Neglecting height information in the map data also affects special road structures, such as bridges, since lane markings of upper or lower roads cannot be separated in such cases.

Despite those remaining challenges, we believe that we have designed an innovative solution that enables localization on the lane level. This is achieved with low-cost sensors and without the restraint of requiring proprietary or HD map data. The full potential of our system is revealed when considering how this lane-accurate localization, a consistent LDM, and the set of map-aligned images can be used in combination. While lane-accurate localization within the LDM is crucial for situational awareness of the vehicle, the generated augmented real-world view can be used to support the driver in the current driving situation.

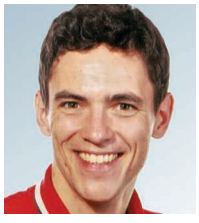
Acknowledgment

Map data is copyrighted by OpenStreetMap contributors, licensed under an Open Database License, and is available from http://www.openstreetmap.org. This work was supported by the EU's Horizon 2020 INLANE project, under grant 687458.

About the Authors



Benedict Flade studied simulation and control of mechatronic systems. He received the master's degree from TU Darmstadt, Germany. Since 2016, he has been a scientist with Honda Research Institute Europe GmbH. He is also involved in the Horizon2020 INLANE project, which is funded by the European GNSS Agency. His research interests include the fields of digital cartography, vehicle localization, computer vision, sensor fusion, and risk estimation.



Axel Koppert studied satellite geodesy and navigation. He earned his Dipl.-Ing. degree from Graz University of Technology, Austria. In 2015, he joined TeleConsult Austria GmbH, which was renamed to OHB Digital Solutions GmbH in 2019. His research focuses on robust algorithms for GNSS-based positioning and sensor data fusion for a variety of applications, including automotive and unmanned aerial vehicles. He works on positioning systems for the H2020 projects INLANE and TransSec that are funded by the European GNSS Agency.

Gorka Velez earned his M.Sc. degree in electronic engineering from the University of Mondragon, Spain, in 2007 and his Ph.D. degree from the University of Navarra, Spain, in 2012. He is a senior researcher with the Intelligent Transportation Systems (ITS) and Engineering Department, Vicomtech, Spain. He is the technical coordinator of the H2020 INLANE project funded by the European GNSS Agency. His research focuses on applying machine learning technologies on the ITS and industrial sectors.

Anweshan Das earned his master's degree in automation and robotics from TU Dortmund, Germany. He is a full-time Ph.D. candidate in the Electrical Engineering Department, TU Eindhoven. He is involved in the H2020 INLANE and Prystine projects. His research aims to improve vehicle localization systems (both absolute and map-based) and 3D map-making techniques using graph-based optimization.

David Bétaille earned his M.Sc. degree in robotics engineering from Ecole Centrale de Nantes in 1992, his Ph.D. degree in geodesy and navigation from University College London in 2004, and his accreditation to supervise research (HDR) from Nantes University in 2014. He is the director of research with the Components and Systems Department, French Institute of Science and Technology for Transport, Development and Networks. His research activity focuses on positioning and map matching using satellite systems combined with dead reckoning and digital enhanced-map data for intelligent transportation systems applications. He is a Member of the IEEE.

Gijs Dubbelman earned his B.Sc. degree in information and communication technology, his M.Sc. degree (cum laude) in artificial intelligence, and his Ph.D. degree in intrinsic statistical techniques for robust pose estimation from the University of Amsterdam. He is an assistant professor with Eindhoven University of Technology. In 2011 and 2012, he was with the Field Robotics Center at Carnegie Mellon's Robotics Institute, where he researched 3D computer vision for autonomous robots.

Oihana Otaegui earned her M.Sc. degree in electronic engineering and her doctoral degree in acquisition and tracking for satellite navigation from the University of Navarre, Spain, in 1999 and 2005, respectively. She manages the Intelligent Transportation Systems Department at Vicomtech and coordinates several projects, including the H2020 INLANE, VI-DAS, and Cloud-LSVA projects. She has authored publications in the areas of satellite-based localization and computer vision. She is a member of the Basque Academy for Science, Letters and Arts.

Julian Eggert earned his Ph.D. degree in physics from the Technical University of Munich, Germany. In 1999, he joined Honda Research and Development, Germany, and, in 2003, the Honda Research Institute, where he is a chief scientist and leader of projects in the area of artificial cognitive systems with applications in automotive and robotics domains. His research interests include generative models for perception, large-scale models for visual processing, and scene analysis, as well as semantic environment models for context-embedded reasoning, situation classification, risk prediction, and behavior planning.

References

[1] V. Gikas and H. Perakis, "Rigorous performance evaluation of smartphone GNSS/IMU sensors for its applications," *Sensors*, vol. 16, no. 8, p. 1240, 2016. doi: 10.3390/s16081240.

[2] F. Peyret and P.-Y. Gilliéron, "SaPPART guidelines, performance assessment of positioning terminal," IFSTTAR, Lyon, France, COST Action TU 1302, SaPPART, pp. 1–47, 2018. [Online]. Available: https://infoscience.epfl.ch/record/254957

[3] D. Betaille, F. Peyret, O. Nouvel, W. Vigneau, and H. Secretan, "How to produce a reference trajectory for studying GNSS errors in urban environments," in *Proc. European Navigation Conf.*, 2008.

[4] S. Kuutti et al., "A survey of the state-of-the-art localization techniques and their potentials for autonomous vehicle applications," *IEEE Internet Things J.*, vol. 5, no. 2, pp. 829–846, Apr. 2018. doi: 10.1109/JIOT.2018.2812300.

[5] Z. Tao, P. Bonnifait, V. Fremont, J. Ibanez-Guzman, and S. Bonnet, "Road-centered map-aided localization for driverless cars using single-frequency GNSS receivers," *J. Field Robot.*, vol. 34, no. 5, pp. 1010–1033, Aug. 2017. doi: 10.1002/rob.21708.

[6] H. Cai, Z. Hu, G. Huang, D. Zhu, and X. Su, "Integration of GPS, monocular vision, and high definition (HD) map for accurate vehicle localization," *Sensors*, vol. 18, no. 10, 2018. doi: 10.3390/s18103270.

[7] C. Zott, S. Y. Yuen, C. L. Brown, C. Bartels, Z. Papp, and B. Netten, "SAFESPOT Local Dynamic Map: Context-dependent view generation of a platform's state & environment," in *Proc. 15th World Congr. Intelligent Transport Systems and ITS America's 2008 Annu. Meeting*, pp. 1–12.

[8] H. Seif and X. Hu, "Autonomous driving in the iCity—HD maps as a key challenge of the automotive industry," *Engineering*, vol. 2, no. 2, pp. 159–162, June 2016. doi: 10.1016/J.ENG.2016.02.010.

[9] D. Betaille and R. Toledo-Moreo, "Creating enhanced maps for lane-level vehicle navigation," *IEEE Trans. ITS*, vol. 11, no. 4, pp. 786–798, Dec. 2010. doi: 10.1109/TITS.2010.2050689.

[10] P. Czerwionka, M. Wang, and F. Wiesel, "Optimized route network graph as map reference for autonomous cars operating on German autobahn," in *Proc. IEEE Int. Conf. Robotics and Automation (ICRA)*, 2011, pp. 78–83. doi: 10.1109/ICARA.2011.6144860.

[11] P. Bender, J. Ziegler, and C. Stiller, "Lanelets: Efficient map representation for autonomous driving," in *Proc. IEEE Intelligent Vehicles Symp. (IV)*, 2014, pp. 420–425. doi: 10.1109/IVS.2014.6856487.

[12] O. Tournaire, B. Soheilian, and N. Paparoditis, "Towards a sub-decimetric georeferencing of ground-based mobile mapping systems in urban areas: Matching ground-based and aerial-based imagery using roadmarks," in *Proc. ISPRS Commission I Symp.*, 2006.

[13] G. Mattyus, S. Wang, S. Fidler, and R. Urtasun, "HD maps: Fine-grained road segmentation by parsing ground and aerial images," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, June 2016, pp. 3611–3619. doi: 10.1109/CVPR.2016.393.

[14] D. Svensson and J. Sorstedt, "Ego lane estimation using vehicle observations and map information," in *Proc. IEEE Intelligent Vehicles Symp. (IV)*, June 2016, pp. 909–914. doi: 10.1109/IVS.2016.7535496.

[15] H. Lategahn and C. Stiller, "City GPS using stereo vision," in *Proc. IEEE Int. Conf. Vehicular Electronics and Safety (ICVES)*, 2012, pp. 1–6. doi: 10.1109/ICVES.2012.6294279.

[16] A. Schindler, "Vehicle self-localization with high-precision digital maps," in *Proc. IEEE Intelligent Vehicles Symp. Workshops (IV Workshops)*, June 2013, pp. 134–139. doi: 10.1109/IVWorkshops.2013.6615239.

[17] M. Schreiber, C. Knoppel, and U. Franke, "LaneLoc: Lane marking based localization using highly accurate maps," in *Proc. IEEE Intelligent Vehicles Symp. (IV)*, June 2013, pp. 449–454. doi: 10.1109/IVS.2013.6629509.

[18] J. Rabe, M. Necker, and C. Stiller, "Ego-lane estimation for lane-level navigation in urban scenarios," in *Proc. IEEE Intelligent Vehicles Symp. (IV)*, June 2016, pp. 896–901. doi: 10.1109/IVS.2016.7535494.

[19] A. L. Ballardini, D. Cattaneo, R. Izquierdo, I. Parra, M. A. Sotelo, and D. G. Sorrenti, "Ego-lane estimation by modeling lanes and sensor failures," in *Proc. IEEE 20th Int. Conf. Intelligent Transportation Systems (ITSC)*, Oct. 2017, pp. 1–7. doi: 10.1109/ITSC.2017.8317834.

[20] Z. Hu and K. Uchimura, "Real-time data fusion on tracking camera pose for direct visual guidance," in *Proc. IEEE Intelligent Vehicles Symp. (IV)*, June 2004, pp. 842–847. doi: 10.1109/IVS.2004.1336494.

[21] S. Zabihi, S. Beauchemin, E. D. Medeiros, and M. Bauer, "Lane-based vehicle localization in urban environments," in *Proc. IEEE Int. Conf. Vehicular Electronics and Safety (ICVES)*, Nov. 2015, pp. 225–230. doi: 10.1109/ICVES.2015.7396922.

[22] T. Caselitz, B. Steder, M. Ruhnke, and W. Burgard, "Monocular camera localization in 3D LiDAR maps," in *Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems (IROS)*, Oct. 2016, pp. 1926–1931. doi: 10.1109/IROS.2016.7759304.

[23] W. Lu, S. A. Rodriguez F, E. Seignez, and R. Reynaud, "Lane marking-based vehicle localization using low-cost GPS and open source map," *Unmanned Syst.*, vol. 3, no. 4, pp. 239–251, 2015. doi: 10.1142/S2301385015400014.

[24] "EGNOS Data Access Service (EDAS): Service definition document," European GNSS Agency, Prague, Czechia, 2014.

[25] B. Flade, M. Nieto, G. Velez, and J. Eggert, "Lane detection based camera to map alignment using open-source map data," in *Proc. IEEE 21st Int. Conf. Intelligent Transportation Systems (ITSC)*, 2018, pp. 890–897. doi: 10.1109/ITSC.2018.8569304.

[26] M. Nieto, L. García, O. Senderos, and O. Otaegui, "Fast multi-lane detection and modeling for embedded platforms," in *Proc. 26th European Signal Processing Conf. (EUSIPCO 2018)*, 2018, pp. 1032–1036. doi: 10.23919/EUSIPCO.2018.8553262.

[27] J. Eggert, D. A. Salazar, T. Puphal, and B. Flade, "Driving situation analysis with relational local dynamic maps (R-LDM)," in *Proc. FAST-zero Symp.*, Sept. 2017.

[28] F. Damerow, Y. Li, T. Puphal, B. Flade, and J. Eggert, "Intersection warning system for occlusion risks using relational local dynamic maps," *IEEE Intell. Transp. Syst. Mag.*, vol. 10, no. 4, pp. 47–59, Sept. 2018. doi: 10.1109/MITS.2018.2867525.

[29] OpenStreetMap, "Contributors." Accessed on: Mar. 2, 2019. [Online]. Available: https://www.openstreetmap.org

[30] G. Cao, F. Damerow, B. Flade, M. Helmling, and J. Eggert, "Camera to map alignment for accurate low-cost lane-level scene interpretation," in *Proc. IEEE 19th Int. Conf. Intelligent Transportation Systems (ITSC)*, Nov. 2016, pp. 498–504. doi: 10.1109/ITSC.2016.7795601.

[31] M. Bertozzi and A. Broggi, "Gold: A parallel real-time stereo vision system for generic obstacle and lane detection," *IEEE Trans. Image Process.*, vol. 7, no. 1, pp. 62–81, 1998. doi: 10.1109/83.650851.

[32] L. Hall, R. Powers, D. Turner, W. Brilon, and J. Hal, "Overview of cross section design elements," in *Proc. Int. Symp. Highway Geometric Design Practices*, 1998, pp. 12:1–12:12.

[33] R. Van Bree, P. Buist, C. Tiberius, B. Arem, and V. Knoop, "Lane identification with real time single frequency precise point positioning: A kinematic trial," in *Proc. 24th Int. Tech. Meeting Satellite Division Institute Navigation (ION GNSS 2011)*, Jan. 2011 vol. 1, pp. 314–323.

[34] inLane. Accessed on: Mar. 2, 2019. [Online]. Available: https://inlane.eu/

ITS