Vision-Enhanced Low-Cost Localization in Crowdsourced Maps

Benedict Flade, Axel Koppert, Gorka Isasmendi, Anweshan Das, David Bétaille, Gijs Dubbelman, Oihana Otaegui, Julian Eggert

2020

Preprint:

This is an accepted article published in IEEE Intelligent Transportation Systems Magazine. The final authenticated version is available online at: https://doi.org/10.1109/MITS.2020.2994055

Vision-Enhanced Low-Cost Localization in Crowdsourced Maps

Benedict Flade[1], Axel Koppert[2], Gorka Velez[3], Anweshan Das[4], David Bétaille[5], *Member, IEEE,* Gijs Dubbelman[4], Oihana Otaegui[3], and Julian Eggert[1]

***Abstract*—Lane-level localization of vehicles with low-cost sensors is a challenging task. In situations in which Global Navigation Satellite Systems (GNSS) suffer from weak observation geometry or 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, correcting the IMU-based trajectory with GNSS observations while employing EGNOS correction data. Furthermore, we consider vision-based complementary data that serves as an additional source of information. In contrast to other approaches, the camera is not used for inferring the motion of the vehicle but for directly correcting the localization results under 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, extracted from the camera image. One strength of the proposed solution is its compatibility with state-of-the-art map data, which is publicly available from different sources. We validate the approach on real world data recorded in The Netherlands and show that it presents a promising cost-efficient means to support future advanced driver assistance systems.**

***Index Terms*—Localization, GNSS, Camera, IMU, Lane Detection, OpenStreetMap, ADAS, Perception**

I. Introduction

SITUATION awareness is a crucial component for intelligent assistance systems and autonomous cars. Predicting the evolution of the traffic situation allows for the judgement of critical situation and is the key to taking actions that mitigate the danger. An accurate prediction requires assumptions on possible driving paths that can only be defined if surrounding cars are located and assigned to particular driving directions. A common approach is the generation of 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 the achievement of a localization accuracy sufficient to reliably assign vehicles to lanes stored in the map. Apart 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 (ADAS).

With an accuracy on the order of meters [1], standard Global Navigation Satellite Systems (GNSS) meet the requirements for road-level navigation but they are not suitable for localization within a lane. The requirements for lane-level navigation have been addressed within the pre-normalization SaPPART EU COST action [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 in specific conditions of use (e.g. in presence of multipath effects, which are typical for urban environments) none of the local effects can be mitigated using differential techniques.

To improve the accuracy and availability of GNSS, it can be combined with additional technology based on other sensors. A common approach is the fusion of Inertial Measurement Units (IMU) with GNSS to obtain low-cost localization solutions. The IMU provides positioning when no GNSS signal is available and enables continuous localization. Yet, the accuracy of such systems is limited by the accuracy of the GNSS receiver. IMUs provide relative measurements which 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 drift during satellite outages mainly depends 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 GNSS 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 GNSS and IMU to localize a vehicle in a map with lane-level accuracy [5], [6]. One common characteristic of such approaches is the requirement of suitable underlying map geometry data which can be stored in a Local Dynamic Map [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 centerlines, road markings, borders, landmarks

[1] Honda Research Institute (HRI) Europe GmbH, Carl-Legien-Str. 30, 63073 Offenbach, Germany {benedict.flade,julian.eggert}@honda-ri.de
[2] OHB Digital Solutions GmbH, Rettenbacher Straße 22, 8044 Graz, Austria axel.koppert@ohb-digital.at
[3] Vicomtech, Mikeletegi 57, Donostia - San Sebastián 20009, Spain {gvelez,ootaegui}@vicomtech.org
[4] Eindhoven University of Technology, 5612 AZ Eindhoven, The Netherlands {anweshan.das,g.dubbelman}@tue.nl
[5] Institut Français des Sciences et Technologies des Transports de l'Aménagement et des Réseaux (IFSTTAR), Allée des Ponts et Chaussées, 44344 Bouguenais, France david.betaille@ifsttar.fr

DOI: 10.1109/MITS.2020.2994055

or traffic signs. Point clouds aggregated into different vectors, allowing for detailed 3D representations of the environment, present an additional type of content [4].

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 GNSS positioning within map-based applications, especially in the case of cooperative entities that share information based on maps coming from different makers.

A wide range of research actors generated high definition (HD) maps with local and global sub-decimeter accuracy. These maps, which used to be rare experimental products [9], [10], [11], [5], cover specific areas or kilometers of highways. Even though 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 anyone is facing remains the accuracy of vehicle positioning. To date most of the prototypes of industrial products do not document their global accuracy. There are still spots in time and space, particularly in deep urban centers, where even sophisticated GNSS approaches such as Post Processing Kinematics (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], [16], [17].

Rabe et al. investigate particle filter based lane assignment while approaching intersections [18] while Ballardini et al. use Hidden Markov Models and line detection algorithms in order 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 is compared to features detected in the camera image [20]. Candidate-based alignment approaches employing 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 local bundle adjustment in order 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. A lane marking based approach to improve low-cost GPS using a camera and open-source map data is introduced by Lu et al. in [23].

In this paper, we present a novel localization method based on two steps (see Fig. 1). The first step covers the fusion of GNSS and IMU to obtain a preliminary position estimation. We employ a multi-constellation GNSS approach (GPS and Galileo) augmented with 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 Local Dynamic Map implementation while only requiring enhanced OSM data. One strength of our novel combination of components is the usage of visual alignment in 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 situation awareness and decision making in driver assistance and autonomous driving applications. Our approach is low-cost on 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 the availability in a wide range of nowadays 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.

The remainder of this paper is structured as follows. In section II, we present our methods for absolute positioning, for camera to map alignment and for their combination. This is followed by section III in which we evaluate our approach with data recorded in the Netherlands. To complement the paper, we summarize the results and give a short outlook in section IV.

II. 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, hereinafter referred to as map frame (m-frame, superscript m), that can be accessed by using the map coordinates of mapped features. Thus any localization technique making use of coordinates of mapped features is sensitive to the position in the map frame. Assuming a map that 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 GNSS-based methods. GNSS provides positions in the terrestrial reference frame (e-frame, superscript e) 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 shall 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:

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

DOI: 10.1109/MITS.2020.2994055

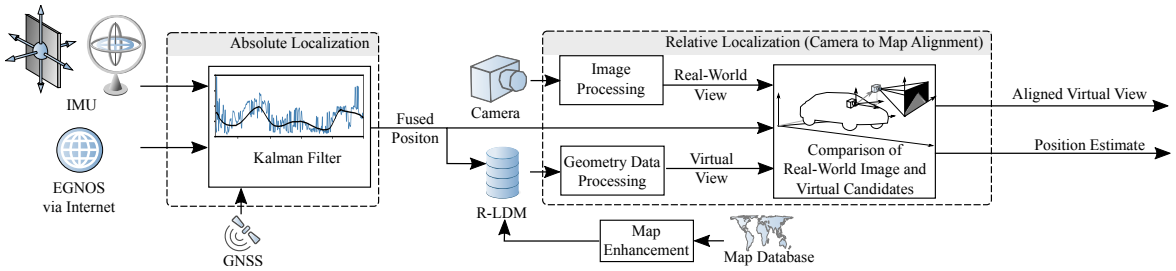


Fig. 1. Pipeline of the localization approach. First step: Fusion of GNSS, EGNOS and IMU data. Second step: Camera to map alignment (C2MA) under usage of publicly available road geometry information.

This relation presents the basis for our map-relative localization procedure. The first step is the estimation of $\boldsymbol{p}^e$ using GNSS. The second step is the alignment of this position with the map in order 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 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.

A. GNSS-based Position Estimation

We obtain an estimate of $\boldsymbol{p}^e$ by exploiting a tightly-coupled Extended Kalman Filter, which fuses 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 for describing the vehicle kinematics and sensor-specific states:

$$\boldsymbol{x} = \begin{bmatrix} \boldsymbol{p}^e & \boldsymbol{v}^l & \boldsymbol{\psi} & \boldsymbol{b}_a^b & \boldsymbol{b}_g^b & \boldsymbol{\delta t} & \dot{\delta t} \end{bmatrix}^T \tag{2}$$

The vehicle kinematics are described by its 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 which needs to be estimated during operation: $\boldsymbol{b}_a^b$ is the accelerometer bias and $\boldsymbol{b}_g^b$ is the gyroscope bias, 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. 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, 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 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, ionospheric model) is contained in the satellite signal. However, the quality of the broadcast data limits the achievable accuracy. Satellite-based augmentation systems (SBAS) provide corrections to the broadcast data and an accurate ionosphere model, enabling meter-level accuracy for single-frequency users. EGNOS is the European SBAS. Its data can be received via satellite or via Internet (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 computed from the broadcast data,
- the ionospheric delay for the signals L1 and E1, respectively, computed from the EGNOS ionosphere model,
- the tropospheric delay, computed from the SBAS MOPS troposphere model.

DOI: 10.1109/MITS.2020.2994055

For the GPS pseudoranges we also apply the long-term clock correction and the fast corrections provided by EGNOS. The satellite positions $\boldsymbol{p}^e_{sat}$ and velocities $\boldsymbol{v}^e_{sat}$ 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 $\rho_{corr,sat}$ and the corrected Doppler observations $\dot{\rho}_{corr,sat}$ are

$$\rho_{corr,sat} = \|\boldsymbol{p}^e_{sat} - \boldsymbol{p}^e\| + \delta t_s + n_{\rho,sat} \quad (5)$$

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

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

We use (3) to derive our dynamic system model for the Kalman filter. The system model uses the IMU measurements (4) for predicting the state. The measurement model is given by (5) and (6) and is used for correcting the prediction whenever a set of 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.

B. Vision-based Camera to Map Alignment

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. In order 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 geometry information derived from the map. The right box of Fig. 1 briefly summarizes the steps of the pipeline.

1) Retrieval of visible map data around GNSS estimate
2) Generation of virtual candidates based on map data
3) Processing of visual data from the camera
4) Calculation of orientation-based feature vector for candidates and processed camera image
5) Best-match search based on feature comparison between processed real world data and virtual candidates

The following sub-sections serve to give a brief insight into our world model, the image processing pipeline and the candidate comparison. A more detailed description can be found in [25], [26], [27] and [28].

1) Environment Representation: One major component for our approach is map data. With regard to advanced driver assistance systems, the main demands on map data are consistency with the real-world, frequent updates, high coverage and an easy editability. The 3 latter points are where crowdsourced and collaborative mapping projects excel. In our research, we choose map data from the community developed OpenStreetMap (OSM) project [29]. However, the drawbacks of OSM are its low level of detail and its varying accuracy. The geometry data provided by OSM is mostly restricted to polylines that describe the center of a road, see Fig. 2 a). Lane-level information, e.g. the number of lanes, is mostly added as a tag, but geometric data of each individual lane or topological data on how lanes are connected is missing.

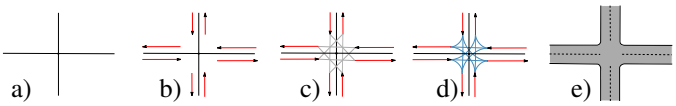


Fig. 2. Enhancement of map data. a) Center of the road. b) Center of lanes. c) Topological connections. d) Geometric connections. e) Full geometry.

In order 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 centerlines based on the corresponding road centerline, see Fig. 2 b). Then we connect the lane segments on topological level and interpolate the junctions between two related lane segments on geometric level, see Fig. 2 c) and d). As a final step, right and left boundary polylines are added, leading to full lane-level geometry, see Fig. 2 e).

With ADAS moving from an ego-centered perspective towards systems that additionally consider the environment around the vehicle, a coherent world model needs to be implemented. In this context, we use our Relational Local Dynamic Map (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 being 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.

2) Lane Detection Based Feature Extraction: In order 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, as e.g. Canny edge detection, to the camera image [30]. However, such approaches tend to overestimate image features, which can lead to a high amount of false positives. Therefore, we propose the use of an explicit lane marking detection method, 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 pixel than equally sized patches further away. To counteract this, we construct a bird's-eye image using an image-to-road homography, a technique also known as Inverse Perspective Mapping (IPM) [31].

In the next step of the 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 lane marking detection algorithm is a set of polylines in the camera image domain that define the

DOI: 10.1109/MITS.2020.2994055

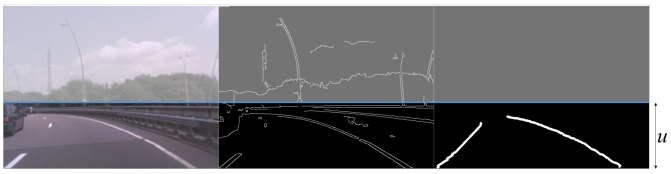


Fig. 3. Highway situation with its corresponding processed Canny edge (center) and lane detection (right) images. The area of interest which is considered for the alignment is defined by u.

detected lane markings and which 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, see Fig. 3. As long as the cut area contains no lane marking, this is not influencing the results but improving the computation time.

3) Candidate Generation: The idea is to create a set of map geometry based images, hereinafter called candidates c, that are then compared to 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, yaw), resulting in six degrees of freedom.

Each candidate c can be described by its pose difference

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

with $\boldsymbol{\rho}_0$ being the so-called candidate zero which is representing the presumed pose. This presumed pose is based 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. Or in other words, the worse the expectation, the larger the required search range. In order 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 outside of the road, nor on a lane that points to the opposite direction, we can set reasonable limits for the search range. Yet, the assumptions of being located on the road as well as the assumption of driving in the correct direction can be wrong in certain situations.

4) Candidate Comparison and Parameter Estimation: Keeping in mind that the map data is based on enhanced open-source information, we target a robust alignment solution which does not necessarily require exact lane marking positions on pixel-level. For this reason, we choose to compare the virtual views and the pre-processed camera image based on the orientation of its features. This is done by applying the histogram of oriented gradients (HOG). Here, dominant orientations are detected in local spatial regions in order to capture edge structures while being invariant towards local geometric and photometric transformations. The result is a flattened feature vector $\boldsymbol{v}$ that is determined for each candidate plus 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 $\boldsymbol{v}^{\mathrm{rwv}}$ being the feature vector of the real-world view and $\boldsymbol{v}_c^{\mathrm{vv}}$ being the feature vector of the c-th virtual view candidate.

Using the parameters of the best match, the initially assumed pose can be updated in order to obtain a map aligned pose. We want to stress that the pipeline allows for an alignment of up to 6 degrees of freedom. However, when assuming a flat environment, the vertical position as well as roll and pitch angles can be neglected.

C. Combining GNSS and Camera to Map Alignment

The final step is the combination of the GNSS-based absolute localization and the camera to map alignment. The starting point of the alignment is the output of the Kalman filter (Fig. 1). Leveraging the alignment, the translation parameter $\boldsymbol{d}^e$ can be interpreted as being the pose difference $\boldsymbol{\Delta\rho}_c$ that belongs to the candidate with the best similarity score.

$\boldsymbol{d}^e$ is applied to every output of the absolute localization module in order to determine a map-relative position using (1). The camera to map alignment 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 $\boldsymbol{\Delta\rho}_c$ 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 a proper alignment.

III. Experiments and Evaluation

To illustrate the concept and the 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. For recording 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 WiFi 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 for assuring time synchronization of the data stream. We use RTMaps not only for record-and-replay testing, but also for online testing in the car. This allows us to reproduce similar behaviour of the algorithms 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 GNSS and IMU

DOI: 10.1109/MITS.2020.2994055

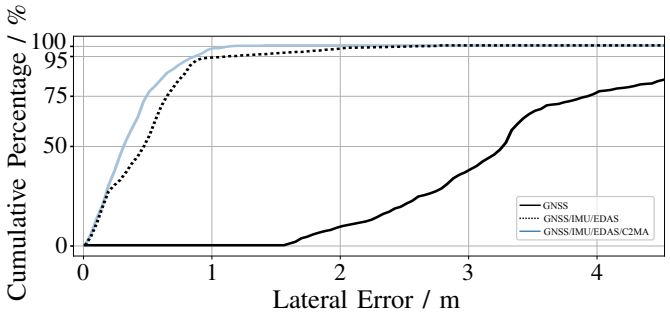


Fig. 4. Cumulative density function of the lateral errors of the GNSS-only, GNSS/IMU/EDAS and GNSS/IMU/EDAS/C2MA localization modules with reference to the PPK-GNSS trajectory.

are chosen to capture the vehicle dynamics appropriately while keeping the computational load at a reasonable level. The GNSS/IMU filter processes the observations at arrival and is operating at 100 Hz while the camera to map alignment provides one position correction per second (1 Hz) during real-time execution.

A. 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 camera to map alignment is done frame by frame, i.e. every output from the GNSS/IMU fusion is aligned. A reference trajectory has been obtained using the high-performance GNSS/IMU system OXTS RT3000 in PPK processing.

Evaluating map-relative localization algorithms proves to be difficult as there is no general method for finding ground truth for this task. Reference trajectories determined by GNSS are not suitable as ground truth for map-relative localization if the map is inaccurate with respect to the e-frame. Therefore we spent 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 described in section II-A. The cumulative density function of lateral errors along the trajectory is shown in Fig. 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 lateral accuracy. The plot of lateral errors (Fig. 5) along the driven path shows several passages with degraded accuracy. In this exemplary evaluation, those are passages where the direct signal is blocked, e.g. due to trees, and only a reflected signal is received.

Fig. 5 and 4 furthermore show the reduction of the 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 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 1000 m and 2000 m). Here, the alignment allows for significant improvement of the position estimate.

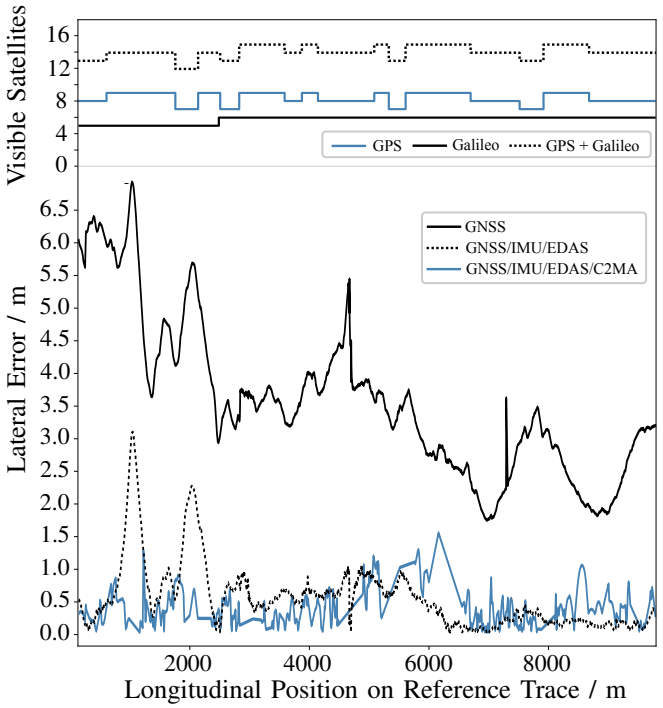


Fig. 5. Comparison of the lateral error of the GNSS-only, GNSS/IMU/EDAS and GNSS/IMU/EDAS/C2MA localization modules in peri-urban environment. In the upper part of the figure, the number of visible GPS and Galileo satellites is shown for each position of the trace.

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

B. 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 to camera image as well.

Fig. 6 shows two examples of augmented real-world views. In both images, the best virtual candidate of the camera to map alignment 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. Furthermore, we can provide driver assistance by displaying a virtual wall that serves as a visual separation between two lanes, 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.

IV. Conclusions and Outlook

This paper presents a novel two-step localization method based on a) the fusion of GNSS and IMU observations and

DOI: 10.1109/MITS.2020.2994055

Fig. 6. Application examples for the camera to map alignment: Path horizon (left) and virtual wall (right).

b) 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-project INLANE [34]. The camera to map alignment can be understood as an add-on that has been tested on real-world recordings while currently reaching a performance of 1 Hz. Being implemented in Python and running completely on the CPU, there is substantial room for optimization.

Yet, our first tests already reveal the potential of the alignment step to support localization, especially in situation in which satellite based solutions struggle. We conclude that the accuracy of our map as well as the quality of the camera to map alignment 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 our targeted accuracy requirements for driving lane-related applications.

As 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 are situation in which GNSS/IMU errors cannot be corrected since the camera to map align is not available, e.g. when map geometry (complicated intersections) or the vision (occlusion of the relevant features) do not permit a 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], leading to an improved localization accuracy. What also remains to be evaluated is the influence of non-flat 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, which enables localization on lane-level. This is achieved with low-cost sensors and without the restraint of requiring proprietary or high definition 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. On the one hand lane-accurate localization within the LDM is crucial for situational awareness of the vehicle. On the other hand, if the camera images are overlaid with aligned map information, the generated augmented real-world view can be used for supporting the driver in the current driving situation.

ACKNOWLEDGMENT

Map data © OpenStreetMap contributors, licensed under the Open Database License (ODbL) and available from http://www.openstreetmap.org. This work has been supported by the European Union's Horizon 2020 project *INLANE*, under the grant agreement number 687458.

REFERENCES

[1] V. Gikas and H. Perakis, "Rigorous performance evaluation of smartphone gnss/imu sensors for its applications," *Sensors*, vol. 16, no. 8, 2016.

[2] F. Peyret and P.-Y. Gilliron, "Sappart guidelines, performance assessment of positioning terminal," pp. 1–47, 2018, publication from the COST Action TU 1302, SaPPART.

[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 Conference*, 2008.

[4] 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, Apr. 2018.

[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," *Journal of Field Robotics*, vol. 34, no. 5, pp. 1010–1033, Aug. 2017.

[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.

[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," *Proc. 15th World Congr. Intelligent Transport Systems and ITS Americas 2008 Annu. Meeting*, pp. 1–12, Nov. 2008.

[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, Jun. 2016.

[9] D. Betaille and R. Toledo-Moreo, "Creating enhanced maps for lane-level vehicle navigation," *IEEE Transactions on ITS*, vol. 11, no. 4, pp. 786–798, Dec. 2010.

[10] P. Czerwionka, M. Wang, and F. Wiesel, "Optimized route network graph as map reference for autonomous cars operating on german autobahn," in *IEEE International Conference on Robotics and Automation (ICRA) 2011*. IEEE, 2011.

[11] P. Bender, J. Ziegler, and C. Stiller, "Lanelets: Efficient map representation for autonomous driving," in *IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2014.

[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 *ISPRS Commission I Symposium*, 2006.

[13] G. Mattyus, S. Wang, S. Fidler, and R. Urtasun, "Hd maps: Fine-grained road segmentation by parsing ground and aerial images," in *IEEE Conference on Computer Vision and Pattern Recognition 2016*. Computer Vision Foundation, Jun. 2016.

[14] D. Svensson and J. Sorstedt, "Ego lane estimation using vehicle observations and map information," in *2016 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, Jun. 2016, pp. 909–914.

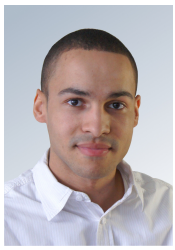
[15] H. Lategahn and C. Stiller, "City GPS using stereo vision," *2012 IEEE International Conference on Vehicular Electronics and Safety, ICVES 2012*, pp. 1–6, 2012.

[16] A. Schindler, "Vehicle self-localization with high-precision digital maps," in *2013 IEEE Intelligent Vehicles Symposium Workshops (IV Workshops)*. IEEE, Jun. 2013, pp. 134–139.

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

DOI: 10.1109/MITS.2020.2994055

[18] J. Rabe, M. Necker, and C. Stiller, "Ego-lane estimation for lane-level navigation in urban scenarios," in *2016 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, Jun. 2016, pp. 896–901.
[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 *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, Oct. 2017, pp. 1–7.
[20] Z. Hu and K. Uchimura, "Real-time data fusion on tracking camera pose for direct visual guidance," in *2004 IEEE Intelligent Vehicles Symposium (IV)*, Jun. 2004, pp. 842–847.
[21] S. Zabihi, S. Beauchemin, E. D. Medeiros, and M. Bauer, "Lane-based vehicle localization in urban environments," in *2015 IEEE International Conference on Vehicular Electronics and Safety (ICVES)*. IEEE, Nov. 2015, pp. 225–230.
[22] T. Caselitz, B. Steder, M. Ruhnke, and W. Burgard, "Monocular camera localization in 3D LiDAR maps," in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, Oct. 2016, pp. 1926–1931.
[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 Systems*, vol. 03, no. 04, pp. 239–251, 2015.
[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 *2018 IEEE 21st International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2018.
[26] M. Nieto, L. Garca, O. Senderos, and O. Otaegui, "Fast multi-lane detection and modeling for embedded platforms," in *26th European Signal Processing Conference (EUSIPCO 2018)*. IEEE, 2018.
[27] J. Eggert, D. A. Salazar, T. Puphal, and B. Flade, "Driving situation analysis with relational local dynamic maps (r-ldm)," in *FAST-zero Symposium*. Society of Automotive Engineers of Japan, Sep. 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 Intelligent Transportation Systems Magazine*, vol. 10, no. 4, pp. 47–59, Sep. 2018.
[29] OpenStreetMap Contributers, "www.openstreetmap.org."
[30] Gaoya Cao, F. Damerow, B. Flade, M. Helmling, and J. Eggert, "Camera to map alignment for accurate low-cost lane-level scene interpretation," in *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, Nov. 2016, pp. 498–504.
[31] M. Bertozzi and A. Broggi, "Gold: a parallel real-time stereo vision system for generic obstacle and lane detection," *IEEE Transactions on Image Processing*, vol. 7, no. 1, pp. 62–81, 1998.
[32] L. Hall, R. Powers, D. Turner, W. Brilon, and J. Hal, "Overview of cross section design elements," in *International Symposium on 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," *24th International Technical Meeting of the Satellite Division of the Institute of Navigation 2011, ION GNSS 2011*, vol. 1, pp. 314–323, Jan. 2011.
[34] INLANE consortium, "www.inlane.eu."



Benedict Flade studied simulation and control of mechatronic systems and received his masters degree from TU Darmstadt, Germany. Since 2016, he is employed as a scientist at the Honda Research Institute Europe GmbH. Furthermore, he is involved in the Horizon2020 project INLANE, funded by the European GNSS Agency (GSA). His research interests cover the fields of digital cartography, vehicle localization, computer vision, sensor fusion and risk estimation.

Axel Koppert studied Satellite Geodesy and Navigation in Darmstadt and Graz and received 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 UAV). He works on positioning systems for the GSA-funded H2020 projects INLANE and TransSec.

Gorka Velez received the M.Sc. degree in Electronic Engineering from the University of Mondragon (Spain) in 2007, and the Ph.D. degree from the University of Navarra (Spain) in 2012. He currently works as a Senior Researcher at the Intelligent Transportation Systems (ITS) and Engineering Department of Vicomtech (Spain). His research is focused on applying machine learning technologies on the ITS and industrial sectors. He is the technical coordinator of the H2020 project INLANE funded by the European GNSS Agency.

Anweshan Das received his masters degree in Automation and Robotics from TU Dortmund, Germany. He is a full-time Ph.D. at the MPS/VCA group of the Electrical Engineering Department at TU Eindhoven. He is involved in the H2020 projects INLANE and Prystine. His research aims to improve vehicle localization systems (both absolute and map-based localization) and 3D map making techniques using graph-based optimization.

David Bétaille is director of research with the Components and Systems Department of the French Institute of science and technology for transport, development and networks (IFSTTAR). His activity is related to positioning and map-matching using satellite systems combined with dead reckoning and digital enhanced-map data for ITS applications. M.Sc. robotics engineering from Ecole Centrale de Nantes, 1992. Ph.D. geodesy and navigation from Univ. College London, 2004. Accreditation to supervise research (HDR) from Nantes Univ., 2014.

Gijs Dubbelman is assistant professor with Eindhoven University of Technology. He obtained his B.Sc. degree in Information and Communication Technology and his M.Sc. degree cum laude in Artificial Intelligence from the University of Amsterdam. In 2011, he obtained his Ph.D. from the same university on the topic of intrinsic statistical techniques for robust pose estimation. In 2011 and 2012, he was a member of the Field Robotics Center of Carnegie Mellons 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 on Acquisition and Tracking for Satellite Navigation from the University of Navarre (Spain) in 1999 and 2005 respectively. She is currently in charge of the ITS department at Vicomtech and she is the coordinator of several projects, including the H2020 projects INLANE, VI-DAS and Cloud-LSVA. She has published in the areas of satellite-based localization and computer vision. Since 2016, she is member of the Basque Academy for Science, Letters and Arts.

Julian Eggert received his Ph.D. degree in physics from the Technical University of Munich, Germany. In 1999, he joined Honda R&D (Germany) and in 2003 the Honda Research Institute, where he is currently in the position of a Chief Scientist and leader of projects in the area of artificial cognitive systems with applications in car and robotics domains. His fields of interest are 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.

DOI: 10.1109/MITS.2020.2994055