Honda Research Institute EU
Honda Research Institute Europe GmbH
https://www.honda-ri.de/

Camera to Map Alignment for Accurate Low-Cost Lane-Level Scene Interpretation

Gaoya Cao, Benedict Flade, Markus Helmling, Florian Damerow, Julian Eggert

2016

Preprint:

This is an accepted article published in 19th International IEEE Conference on Intelligent Transportation Systems. The final authenticated version is available online at: https://doi.org/10.1109/ITSC.2016.7795601

Camera to Map Alignment for Accurate Low-Cost Lane-Level Scene Interpretation

Gaoya Cao[1†], Florian Damerow[1†], Benedict Flade[2†], Markus Helmling[1†] and Julian Eggert[2†]

***Abstract*— To be able to predict the evolution of the driving context, estimate the expected risks and plan future behavior alternatives, it is crucial to know where traffic participants can go and where they will most likely go. Consequently, for future Advanced Driver Assistance Systems (ADAS), precise, lane-accurate localization of the ego- as well as the other vehicles is a key technology. The proposed standard solutions to lane-accurate localization are usually based on expensive detailed 3D maps and precise absolute position estimation sensors. To the contrary, in this paper we propose to improve the localization of scene items based on state-of-the-art map data, combined with a coarse and cheap position estimation as e.g. provided by standard GNSS. From the map data, we infer the structure of the contextual road geometry, and align it with the road view(s) as provided by a front camera. This results in an improved *relative* positioning of the sensed items on the map structures, allowing a better scene interpretation. The alignment occurs by best-match search based on a feature comparison between the real road view from the camera and virtually generated road views based on the map considering different assumed ego-vehicle positions. On standard road scenes, we validate the approach, and show that it can be used as a cheap means to support intelligent ADAS and improve localization.**

I. Introduction

A necessary functional component of Adaptive Driving Assistance Systems (ADAS) is the proper representation and interpretation of the immediate driving environment, including road structure as well as the sorrounding traffic participants. In future ADAS, predictive technologies will enable the anticipation of the possible paths and behaviors of the ego- and other vehicles to a larger extent, supporting risk-based warnings, behavior support or even autonomous driving.

It will then be crucial to be able to predict where other vehicles will go. A key technology required for high-confidence predictions is the consistent estimation of the vehicles positions with respect to the road structure. In some cases, the road structure will determine the most probable future path of a vehicle, as it is the case if the vehicle is moving on a turning, entrance or exit lane. In other cases, observation about a vehicle's behavior with respect to the lane markings is a strong indicator of a maneuver intention, such as changing lane, overtaking or initiating a turn at an intersection. The consequence is that, since lanes and the relative position of vehicles with respect to lanes are such strong "semantic" indicators for vehicle behavior, an estimation of position of a vehicle relative to the road structure in form of lanes is an important prerequisite for intelligent, context-aware ADAS.

One way to achieve an accurate alignment between road structure and sensed traffic participants is to independently increase both the absolute accuracy of map data *and* the absolute accuracy of the vehicle position estimation, e.g. up to a sub-lane or even cm level. The long-term roadmap targeted by several map-makers (see e.g. [1], [2]) addresses the first point, and more accurate position estimation technologies are under investigation (see e.g. [3]). Both the map aggregation as well as the position estimation have to rely on the same ("absolute") coordinate system, and offset errors should be avoided. If the position and orientation settings of sensors in the vehicle are known, this can then be used to calculate where sensed vehicles are on the road structure.

However, in current off-the-shelf systems the conditions for such an accurate alignment on an absolute basis are not given yet. Available digital road maps do not contain detailed road geometry but only road directions and intersection points, in form of nodes and links. Errors of both the map data as well as the position estimations are in the range of several meters, and several unknown parameters (e.g. in the location of the sensors within the car) and sources of error lead to offsets between map and position estimation data.

Nevertheless, sensors on vehicles such as cameras provide snapshots of the current road geometry which can be compared with geometry data derived from map information. This would allow for an alignment and for an improved relative positioning of other vehicles in the sensed scene, leading in the optimal case to a correction of errors and offsets and to an improvement of vehicle to lane assignments.

In this paper, we therefore investigate a system based on a map-to-sensor alignment for lane-accurate position estimation using currently available technologies only. We use standard map data, augment it by geometry information to provide a "virtual" view of the road structure from an assumed camera position, and compare it with the sensed camera view. A search of the most appropriate camera position that explaines best the sensed scene then leads to the current position of the ego-vehicle with respect to the road structure. The remainder of the paper is organised as follows. In section III, we explain the details of our approach. In section IV, we show and explain cases where our approach

† The authors contributed equally to this work

[1]Gaoya Cao, Florian Damerow and Markus Helmling are with the Control Methods and Robotics Lab, Technical University of Darmstadt, 64283 Darmstadt, Germany caogaoya@gmail.com, florian.damerow@rmr.tu-darmstadt.de, markushelmling@outlook.de

[2]Benedict Flade and Julian Eggert are with the Honda Research Institute (HRI) Europe, Carl-Legien-Str. 30, 63073 Offenbach, Germany benedict.flade@honda-ri.de, julian.eggert@honda-ri.de

©2016 IEEE. Personal use of this material is permitted.
Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.
DOI: 10.1109/ITSC.2016.7795601

succeeds and where it has difficulties. Furthermore, we show the results and statistical properties of a larger investigation on driven routes from [4]. Finally, in section V we close the paper with a discussion on future steps.

II. Related Work

Absolute positioning for navigation is a nowadays massively available technology. Low cost GNSS-receivers based on GPS, GLONASS, GALILEO or similar systems, are available in a wide range of devices such as smartphones, tablets, car-integrated devices or standalone navigation systems. Especially in the automotive sector, satellite navigation systems have become an integral part of everyday life and form the basis for tasks such as navigation or autonomous driving. The current generation of navigation systems allows road level navigation which is sufficient for basic routing and guidance. For the next generation of navigation systems and ADAS functions such as lane level navigation, lane-change assistance or fully autonomous driving, a sub-lane positioning accuracy will be required. Different data fusion techniques which combine GNSS data with Dead Reckoning (DR) systems are being investigated (see e.g. [5],[6]), however these provide mainly improvements when the GNSS signal is not available but do not increase the overall absolute accuracy substantially. Other approaches like Differential GPS provide accurate localization, but are neither widely available nor do they provide a guarantee that the alignment with the map data will not contain other errors.

As an alternative, techniques from the domain of visual simultaneous localization and mapping (vSLAM) are being investigated. In this case, a special database is generated based on special sensor recordings like cameras or LiDARS, and on a second pass, the generated data is compared, correlated and matched to the new recordings (see e.g. the "RoadDNA" announcements from TomTom, [7]). Similarly, Levinson et al. [8] combine SLAM with GPS, IMU, wheel odometry and LIDAR data, achieving a localization accuracy within the 10 cm range. The advantage of SLAM-based approaches is that the generation process is controlled so that the matching can be simplified; the disadvantage is that dedicated, detailed mapping data has to be generated which is currently not broadly available, and that the involved vehicles need to comply with certain, eventually expensive sensor specifications given by the generation process.

To avoid the drawback of dedicated map data generation, standard map data has been fused with recordings from visual sensors (cameras) by a number of researchers. In [9], the authors combine OSM data with a GIS database which is enriched with additional geometrical data in order to model a 3-dimensional road shape. The shape is projected onto the camera image and is used as a starting point for a more detailed road shape estimation. Differently to the proposal of this paper, the target is drivable road shape segmentation, so that there is no active alignment for relative position estimation. The authors of [10] also address the problem of map matching, with the target of improving the accuracy of vehicle lateral localization. A two-dimensional parabola road network model is introduced to describe the road geometry, which works well for smooth curve geometries.

III. Camera to Map Alignment

The general approach for camera to map alignment is illustrated in Fig. 1. Starting from a rough GNSS-based position, virtual 3D street views, close to this position, are generated using digital map data. Each view is generated with a different pose, which is defined by a shift and rotation in all 6 degrees of freedom, relative to the original GNSS-position. The digital map used in this work originates from a publicly available navigational map and is enhanced by the estimation of a more detailed, geometric representation of the road structure (see section III-A).

The target is to find the virtual view that fits best to the actually sensed view, provided by an on-board camera system. As a result, the corresponding pose of this view can be employed to correct the discrepancy between GNSS-based map localization and camera image. A key component is the calculation of a similarity score between each virtual view and the camera image. For this purpose, we use histogram of orientated gradients (HOG) as descriptors and a cosine similarity score. The individual steps are described in detail in the following sections III-*A* to *C*.

A. Map Representation

The presented approach focuses on improving a coarse vehicle position estimation, acquired by a GNSS sensor, with the help of publicly available map data, e.g. from OpenStreetMap, to arrive at a lane-precise map localization. This localization can then be employed for tasks requiring such a level of accuracy, spanning from lane precise navigation up to precise trajectory prediction and planning or maneuver detection. The approach copes with uncertainties inherent in both the positional data from the GNSS sensor as well as from the map data of low accuracy.

The original digital map data mainly provides navigational information. As presented in Fig. 2a), this mainly involves the rough course of the road, represented by the road centerlines and intersection nodes, defining the connections between roads. However, no explicit information about how the individual lanes are connected or about the geometric shape of the road is currently available from the map data. Additional information, such as number of lanes, speed limit and type of road is generally tagged to the corresponding road segments.

The geometric road information, especially at intersections, is neither sufficient to generate a virtual view of the road with an accuracy suitable for the camera to map alignment nor for a precise localization on lane level. Therefore, in a first step, shown in Fig. 2b), we infer further semantic and geometric information using heuristics. In particular, we estimate the connections at intersections on the lane level. Furthermore, in Fig. 2c) we estimate the actual geometric road shape by combining the previously inferred information with an estimate of the lane and road width.

©2016 IEEE. Personal use of this material is permitted.
Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.
DOI: 10.1109/ITSC.2016.7795601

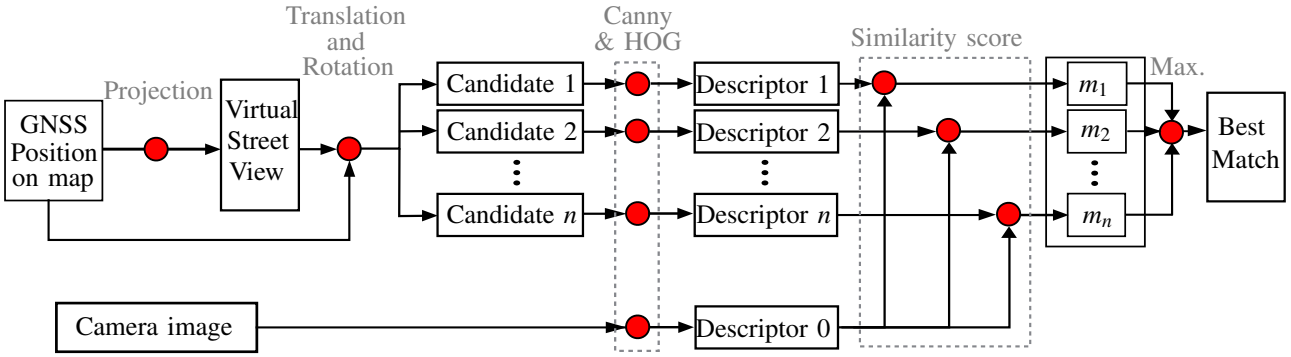

Fig. 1: Camera to map alignment approach. First, virtual street views (candidates) are generated at different poses (6 degrees of freedom) near the original GNSS-position. Then, a similarity score between each virtual view and the actual camera image is calculated, using Canny Edge Detection in a preprocessing step and HOG as descriptors. The camera pose of the best matching candidate can be employed relative to the original GNSS-position.

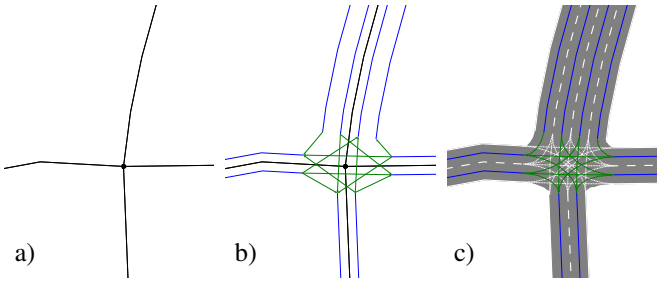

Fig. 2: Extension of publicly available digital map data a) Generally provided geometric information. b) Inferred semantic information on the lane level. c) Enrichment and estimation of detailed geometric road structures.

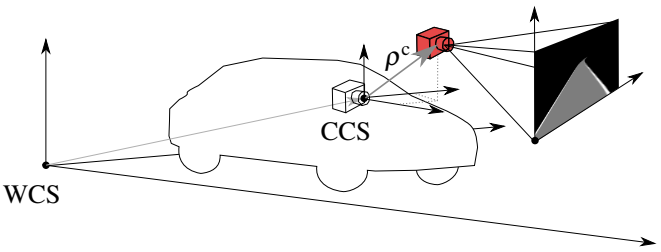

Fig. 3: The pose ρ^c, relative to the car coordinate system (CCS) is used to generate the virtual views.

The enhanced digital map data is stored in a graph, which contains the full approximate road geometry information, comprising road surfaces, all lanes and how they connect, as well as parametrized center and boundary lines.

B. View Generation

In order to align the camera image with the digital map, a virtual 3D street view has to be generated. A possible virtual street view, hereinafter called candidate, is obtained by simulating the view of the road shape given by the enhanced digital map data as seen from the position of a virtual camera.

Every candidate is defined by its pose ρ^c in relation to the coordinate system of the car. In order to locate the car in the map, a transformation from the world coordinate system (WCS) into the car coordinate system (CCS) has to be performed. For means of simplification, the CCS is assigned to the position of the real camera (see Fig. 3).

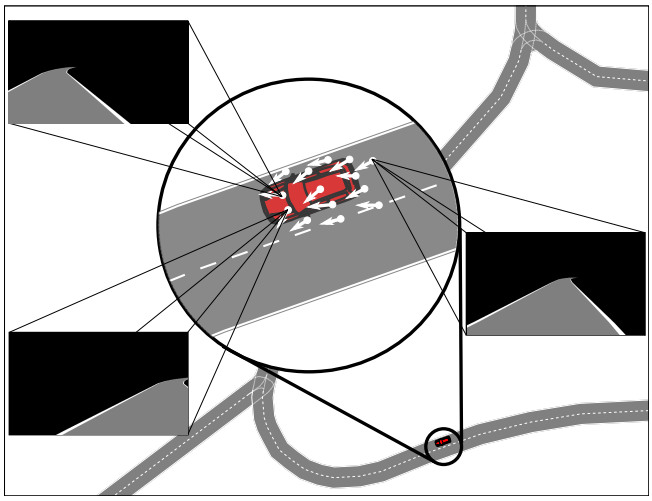

Fig. 4: Example of potential candidates (white arrows), with three of them being zoomed in. Figure idea similar to [12].

Allowing six degrees of freedom, every possible candidate can now be described by $\rho^c = (s^c_{\text{lat}}, s^c_{\text{lon}}, s^c_{\text{vert}}, \alpha^c_{\text{yaw}}, \alpha^c_{\text{pitch}}, \alpha^c_{\text{roll}})$, consisting of a lateral, a longitudinal and a vertical shift as well as a rotation around all three axis.

When choosing candidates, the rough GNSS position serves as a starting point for the virtual view generation. The candidate distribution can be uniform or Gaussian but also sampling-based approaches like e.g. RANSAC [11] can be used. Fig. 4 shows an example of potential candidates (white arrows), whith three of them being zoomed in.

C. Visual Alignment

1) Histogram of Oriented Gradients: Generally, roads and other road geometries are described by the positions and orientations of their composing boundaries and edges. The basic idea of the Histogram of Oriented Gradients (HOG) is to describe the dominant orientations of local spatial regions in an image, using histograms of gradients. Being applied very successfully in the field of human detection, we now use the HOG algorithm for map to camera matching [13]. With the Histogram of Oriented Gradients, it is possible

©2016 IEEE. Personal use of this material is permitted.
Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.
DOI: 10.1109/ITSC.2016.7795601

to capture edge structures, such as road boundaries, while being invariant towards local geometric and photometric transformations. Due to this invariance, our approach does not rely on a highly accurate map data and is robust towards minor modelling errors as long as the basic road geometry matches. More complex orientation based filters have been investigated, such as Gabor Filters of different orientations, but the performance did not change significantly so that we chose the HOG due to its simplicity.

In our approach, we apply the HOG on the camera image as well as on each generated virtual view candidate. As a preprocessing step we use the Canny edge detection algorithm [14] in order to enhance relevant structures, with an lower threshold of 30 and an upper threshold of 80.

An image normalization algorithm (gamma compression) is used on the camera image to reduce the influence of illumination effects. As long as the deviations between the road shape in the map and the actual road shape are small compared to the HOG spatial or orientation bin size, mapping inaccuracies do not play a significant role. Regarding the gradient computation, we use the kernels $[-1,0,1]$ and $[-1,0,1]^T$ for horizontal and vertical direction. For the voting, we consider cells of 8×8 pixels and 4 orientations that represent unsigned gradients and are evenly spaced over 0 - π. In the normalization step, which groups cells into larger spatial blocks, we choose rectangular blocks (R-HOG) with a size of 2×2 cells per block. The result is a matrix consisting of the orientation descriptors of the blocks.

2) Candidate Comparison: The aim of the alignment algorithm is a comparison of the real camera view with a set of generated street views $\mathscr{S} := \{\rho_1^c, \rho_2^c, ..., \rho_n^c\}$, descibed by their pose. In this context, we formulate a comparison score valuing similar local orientations. The resulting descriptor matrix is flattened into a 1D-vector and then compared using

$$m = \frac{\mathbf{v}_c \cdot \mathbf{v}_i}{||\mathbf{v}_c|| \cdot ||\mathbf{v}_i||}, \tag{1}$$

with $\mathbf{v}_c$ and $\mathbf{v}_i$ representing the HOG descriptors of the candidate and the real world image.

IV. Results on different scenarios

In our evaluation, we use the KITTI dataset which comprises, among other data, front-camera images from a PointGray Flea2 color camera as well as trace coordinates from inertial and GPS navigation system OXTS RT 3003 [4]. The map data used in the evaluation process is obtained from the OpenStreetMap project [15], enhanced by additional details such as lane connections and detailed geometry, see section III-A.

Table I shows the parameters used in the evaluation. As the position of the GPS sensor differs from the camera position, an offset of 0.26 m lateral and 1.08 m longitudinal has also been considered.

TABLE I: Parameters of the evaluation

Parameter	Value
Image size	1242×375 pixels
Candidate distribution	Uniform
Camera perspective angle	30°
Camera height offset above ground	1.65 m

A. Single Picture Alignment

In order to validate the camera to map alignment we first evaluated single images. The camera data, acquired from the KITTI dataset is associated with its GPS Data (longitude, latitude, heading), which is transformed into cartesian map coordinates. Fig. 6 shows a selection of camera images overlayed with the best match of the alignment. From the first row, it can be seen that roads with detectable markings result in a successful match. In the result images, the superposed white lines are gained from the virtual view. Even when dealing with environmental influences such as shadows, true matches can be achieved. Regarding the first image, tree shadows on the left side of the road as well as cars on the right side of the road make a lane detection difficult. Nevertheless, the alignment can be performed since a dominant orientation is still recognizable. The transition from asphalt to grass provides a dominant orientation, leading to a high matching value. Also in the case of hard-to-identify lane markings, the alignment is successful as long as a dominant orientation can be determined.

Especially at intersections, shown in the second row of Fig. 6, an alignment presents a major challenge. Different to other approaches, with our algorithm, intersections do not only lead to a lateral but also to a longitudinal alignment. In the left picture, the vehicle has already entered the crossroad. Based on the corner and road boundary of the opposite side of the intersection, the candidate matches the camera image, even if the curb on the right side of the opposite road is hardly recognizable.

Another intersection example is depicted in the center image of Fig. 6, second row, in which the opposite part of the road is mostly hidden by cars, trees and a house. The alignment can still be performed based on the corner.

The right picture of the second row shows an intersection in which the left road is downhill and additionally mostly hidden by trees. Even if a variation in the roll angle is considered, due to using a flat map, it can be seen that the road shape of the left road is not represented properly. This is caused by the lack of road height information in the candidate generation step, thus leading to problems in hilly environments. Nonetheless, the candidate can still be aligned at the two adjacent borders, allowing a lateral and longitudinal relative localization.

Even if our approach profits from several advantages of the HOG such as a robustness towards small local geometric transformations, there are still factors leading to problems from which a selection is shown in the last row of Fig. 6. Despite the robustness, heavy deviations in the map from the true road shape can impede an alignment. Sources of errors are parking or moving cars limiting the view (left image), neighbouring lanes as well as other dominant edges, such as train rails or fences distracting the HOG (center image). Missing lane markings or a surrounding that can hardly be

©2016 IEEE. Personal use of this material is permitted.
Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.
DOI: 10.1109/ITSC.2016.7795601

distinguished from the road or an incorrectly estimated lane width remain a problem to be tackled as well (right image). In all those cases, the success of the alignment relies on a smart choice of the candidate range (see section IV-C).

B. Trace Evaluation

In the localization process, several factors can lead to a misalignment. The following evaluation aims to test the camera to map alignment algorithm on the basis of the GPS trace, obtained from the KITTI dataset.

For each GNSS-position, 155 uniformly distributed candidates are generated with a candidate position range of [-1.5 m, 1.5 m] in lateral and [-0.1 m, 0.1 m] in longitudinal direction. As visible in Fig. 5, the GPS trace is located outside of the road area in the digital map, i.e. it exhibits a noticeable error. For aligning the camera to the map, the lateral displacement of the candidate with the highest similarity to the real world image is taken to calculate an updated point in the map.

Position A (top view) illustrates an example of successful alignment. It can be seen, that the map matches the one-way road, benefiting from dominant curbs. Nevertheless, some misalignments may occur, such as at Position B (bottom view). In this case, the left boundary of the road in conjunction with the vehicle garage on the left side of the image is mistaken as road boundary, leading to a wrong batch. To avoid such a misalignment, a reasonable choice of candidate range is of high importance, investigated in the following section. Apart from the outliers, a lateral offset can be recognized after a certain period of time, indicating a constant error either in the map or GPS-data. Determining this offset allows us to consider this source of error in the subsequent alignments.

C. Statistical evaluation

For evaluating the lateral stability of our approach, we superimposed the GPS Trace from KITTI, taken as ground truth, with white noise in the range of $\sigma_{\text{GPS}} \in [0.1\text{ m}, 2.0\text{ m}]$ with a step length of 0.1 m. In this context, we analyzed the lateral distribution of the alignment result at 497 GPS points from several traces with respect to ground truth. A drawback is a possible systematic offset, caused by error sources such as the map or the gps receiver, which was taken into account but not eliminated completely.

Fig. 7 shows a match distribution, provided that all candidates are chosen within a defined range around the assumed GPS position. The plot at the top left shows the distribution for a range of [-0.3 m, 0.3 m]. As expected, for a small noise ($\sigma_{\text{GPS}} < 0.3$ m), the majority of the matches is within a region of a few centimeters around ground truth. For a noise exceeding 1.0 m, the matches are spread, indicating a high number of misalignments, consistent with the fact that a small search radius cannot lead to successful alignments if the GPS signal noise is to big and thus the actual camera view is not covered by any candidate.

In the right plot of the first row, a candidate range of $[-1.5\text{ m}, 1.5\text{ m}]$ is chosen. For all values of σ_{GPS}, the majority of the matches is located close to the assumed ground truth. Nevertheless, two further maxima of the distribution become slightly visible for values of $\sigma_{\text{GPS}} > 1.0$ m.

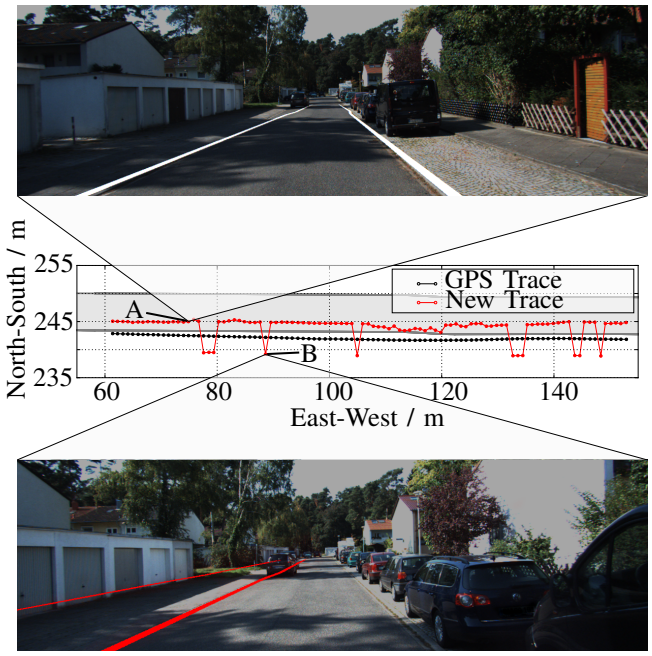

Fig. 5: Trace evaluation. Center: Trace corrected by alignment (red dots) and original GPS trace (black dots). Top (A): Correct alignment. Bottom (B): Misaligned position estimation.

Regarding a candidate range of $[-3\text{ m}, 3\text{ m}]$ (bottom left), the majority of the matches is close to the assumed ground truth for all values of σ_{GPS} but the two additional maxima become clearly recognizable. Those sidelobes can be traced back to neighbouring lanes or crash barriers which are leading to misalignments.

In order to achieve a high matching rate without sidelobes in the case of noisy signal, several simulations with an adaptive candidate range were analyzed, assuming that information about the standard deviation of the position is available. The candidate range is chosen depending on the standard deviation of the noise. Best results could be achieved with a candidate range of $[-\sigma_{\text{GPS}}, \sigma_{\text{GPS}}]$, shown in the plot at lower right. The result is a high number of successful alignments for a precise GPS position with low noise and still acceptable results for a noisy GPS signal with $\sigma_{\text{GPS}} > 1.0$ m. For comparison, the 4 cases are compared in the plot of Fig. 8, showing the match rate with regard to GPS noise.

From the preceding plots, it can be concluded that even for large noise in the GPS trace, the candidates will match the lane geometry on the premise that the standard deviation is known. On the one hand, if the candidate range increases, the probability of matching on adjacent lanes rises. on the other hand, for a small candidate range, the alignment succeeds only for a weak noise.

In oder to evaluate the results quantatively, the match rate r is proposed. The potential match, determined by the

©2016 IEEE. Personal use of this material is permitted.
Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.
DOI: 10.1109/ITSC.2016.7795601



Fig. 6: Single images of camera to map alignment with successful alignments in white and misalignments in red.

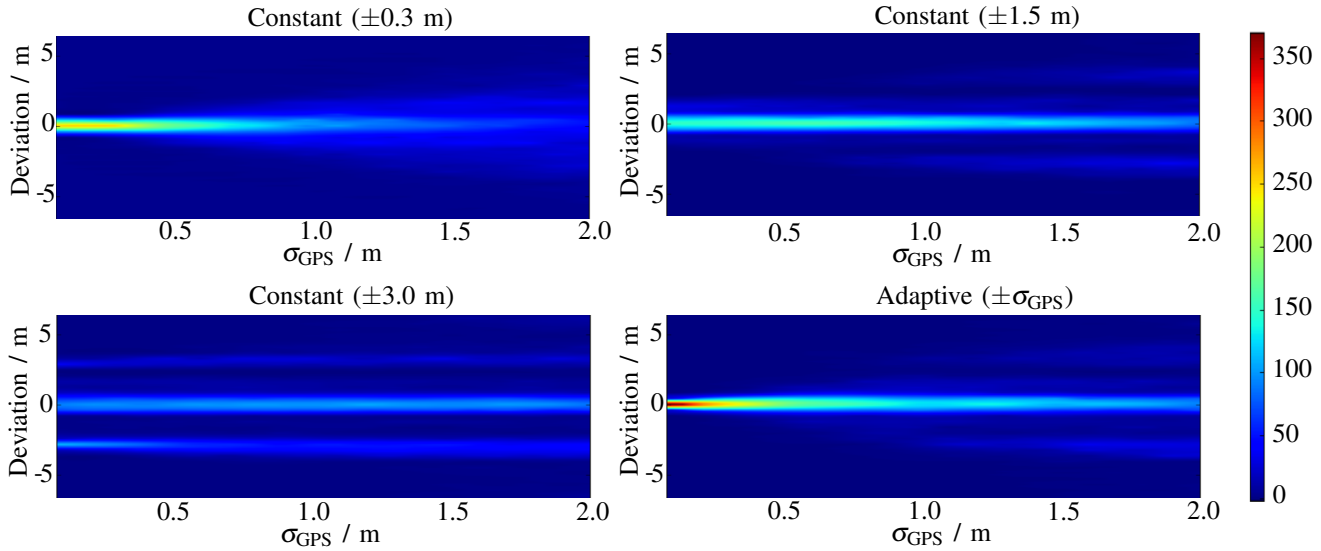

Fig. 7: Statistical evaluation of the alignment result at 497 GPS points using different candidate ranges.

alignment algorithm, is classified as a true-match for

$$d^{\mathrm{c}}_{\mathrm{gt,lat}} - \mathrm{offset}_{\mathrm{lat}} < d_{\mathrm{thresh}}, \tag{2}$$

with $d^{\mathrm{c}}_{\mathrm{gt,lat}}$ being the lateral distance to ground truth and d_{thresh} a chosen threshold. An offset has to be taken into account that may be caused by the map or GPS receiver. For the threshold, we choose 0.5 m. Due to unawareness of the actual offset, good matches might be falsely flagged as bad matches, influencing the results. However this is a shortcoming of the chosen database and is not caused by the presented method. More detailed results can be achieved by using an online estimation of the offset which is fed back into the alignment algorithm.

The match rate r is defined by the fraction

$$r = \frac{n_{\text{true-match}}}{n_{\text{no-match}}}, \tag{3}$$

while $n_{\text{true-match}}$ and $n_{\text{no-match}}$ describe the amount of true matches and misalignments.

Fig. 8 serves to compare the match rate of different candidate ranges as functions of the standard deviation of the noise σ_{GPS} of the GPS position. The match rate of candidate range $[-3\ \mathrm{m}, 3\ \mathrm{m}]$ changes insignificantly and remains on a constant level as σ_{GPS} changes. For decreasing candidate ranges, the probability of matching non-ego-lanes becomes smaller but a small range may bare the risk of not containing the actual match. Regarding a candidate range of $[-0.3\ \mathrm{m}, 0.3\ \mathrm{m}]$ (red), the match rate is greater than 80% for $\sigma_{\mathrm{GPS}} < 0.5$ m but declines rapidly with growing σ_{GPS}, resulting in a worse performance than the 3 m range for $\sigma_{\mathrm{GPS}} > 0.8$ m. The plot confirms the good results of the adaptive candidate range [-σ_{GPS}, σ_{GPS}] which is outperforming the constant candidate ranges.

V. Conclusion and Outlook

In this work we introduced a novel position estimation approach within a framework of camera to map alignment which uses the Histogram of Oriented Gradients for the alignment score. Simulations show that the HOGs present a suitable basis for the localization of scene items without the restrain of requiring a high end GNSS-receiver, expensive cameras or highly detailed map data. Promising results could be shown on regular roads as well as at intersections. By evaluating dominant local orientations, the approach is robust towards minor map-modelling errors as long as the basic

©2016 IEEE. Personal use of this material is permitted.
Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.
DOI: 10.1109/ITSC.2016.7795601

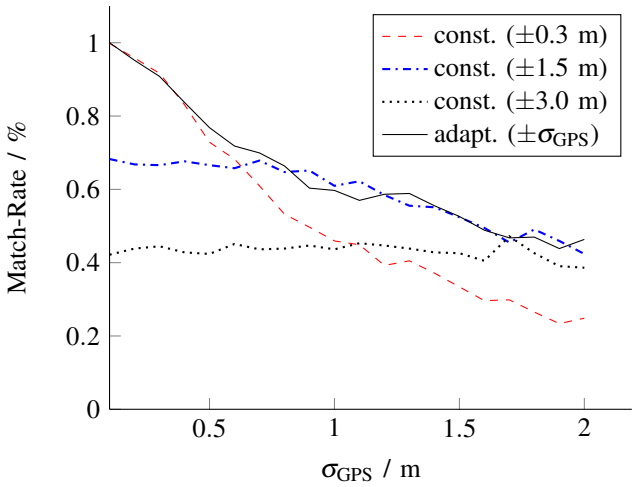

Fig. 8: Match rate of the alignment with a matching-threshold of 0.5 m.

geometry of the scene items matches.

Currently, strong shadows present a problem as they complicate the detection of dominant edges. The texture of the lane markers or road borders that are covered by shadows is weakened while the edges of the shadow areas lead to descriptors of high values in their orientation. Approaches such as presented by Finlayson et al. in [16] reduce shadow influences by minimizating the entropy and thus may help to improve the presented alignment approach.

The performance of camera to map alignment depends crucially on the chosen candidate range. An adaptive candidate range turned out to be an important ingredient for a robust alignment, so that the variance of the position estimation (e.g. from GNSS) should be considered.

In addition, several misalignments are due to estimation outliers which could be reduced by filtering approaches. Future work will concentrate on including plausibility assumptions based on a velocity model or inertial measurement unit data with regard to an outlier detection. Priors such as the color and homogeneity information of the road surface may be taken into account.

Up to now, exclusively the shape of the road boundary has been used as landmarks, but also any other kind of landmark, included in a digital map, can be used with our alignment approach. It remains to be evaluated which other landmarks may improve the results.

ACKNOWLEDGMENTS

Map data copyrighted OpenStreetMap contributors 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] HERE, "HERE introduces HD maps for highly automated vehicle testing," accessed 15-June-2016. [Online]. Available: http://360.here.com/2015/07/20/here-introduces-hd-maps-for-highly-automated-vehicle-testing/

[2] Bloomberg, "Even When Drivers Become Obsolete, Zenrin Plans to Sell Car Maps," accessed 15-June-2016. [Online]. Available: http://www.bloomberg.com/news/articles/2016-03-13/race-to-guide-self-driving-cars-is-getting-another-competitor

[3] Atlatec, "http://www.atlatec.de," accessed 15-June-2016.

[4] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets Robotics: The KITTI Dataset," *International Journal of Robotics Research*, 2013.

[5] R. Toledo-Moreo, D. Betaille, F. Peyret, and J. Laneurit, "Fusing GNSS, Dead-Reckoning, and Enhanced Maps for Road Vehicle Lane-Level Navigation," *IEEE Journal of Selected Topics in Signal Processing*, vol. 3, no. 5, pp. 798–809, oct 2009.

[6] R. Toledo-Moreo, D. Betaille, and F. Peyret, "Lane-Level Integrity Provision for Navigation and Map Matching With GNSS, Dead Reckoning, and Enhanced Maps," *IEEE Transactions on Intelligent Transportation Systems*, vol. 11, no. 1, pp. 100–112, mar 2010.

[7] TomTom, "TomTom launches RoadDNA: Precise localisation technology for automated driving," accessed 15-June-2016. [Online]. Available: http://corporate.tomtom.com/releasedetail.cfm?ReleaseID=931327

[8] J. Levinson, M. Montemerlo, and S. Thrun, "Map-Based Precision Vehicle Localization in Urban Environments." in *Robotics: Science and Systems*, Atlanta, GA, USA, jun 2007.

[9] J. M. Alvarez, F. Lumbreras, T. Gevers, and A. M. López, "Geographic Information for Visionbased Road Detection," in *IEEE Intelligent Vehicle Symposium (IV)*, 2010, pp. 621–626.

[10] Z. Hu and K. Uchimura, "Real-time data fusion on tracking camera pose for direct visual guidance," in *Intelligent Vehicles Symposium, IEEE*, jun 2004, pp. 842–847.

[11] M. Fischler and R. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, jun 1981.

[12] J. M. Alvarez, A. M. Lopez, T. Gevers, and F. Lumbreras, "Combining Priors, Appearance, and Context for Road Detection," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 3, pp. 1168–1178, jun 2014.

[13] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 1. IEEE, 2005, pp. 886–893.

[14] J. Canny, "A Computational Approach to Edge Detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-8, no. 6, pp. 679–698, nov 1986.

[15] OpenStreetMapContributers, "www.openstreetmap.org."

[16] G. D. Finlayson, M. S. Drew, and C. Lu, "Entropy Minimization for Shadow Removal," *International Journal of Computer Vision*, vol. 85, no. 1, pp. 35–57, oct 2009.

©2016 IEEE. Personal use of this material is permitted.
Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.
DOI: 10.1109/ITSC.2016.7795601