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

Strategies for Improving Camera to Map Alignment

Johannes Silberbauer, Benedict Flade, Stephan Hasler, Malte Probst , Julian Eggert

2017

Preprint:

This is an accepted article published in Workshop New Challenges in Neural Computation (NC2). The final authenticated version is available online at: https://doi.org/[DOI not available]

Strategies for Improving Camera to Map Alignment

J. Silberbauer[1], B. Flade[2], S. Hasler[2], M. Probst[2], J. Eggert[2]

[1] Technical University Darmstadt
[2] Honda Research Institute Europe GmbH

Abstract. Accurate localization of the ego vehicle is a key requirement for contemporary and feature advanced driver assistance systems (ADAS). Map relative localization can be achieved by aligning monocular front camera images to low precision map data. An initial approach uses an iterative strategy to match histogram of oriented gradients (HoG) features. We propose two new strategies for improving this approach: First, we replace the original feature extraction step by HoG computation on a semantic segmentation of the camera image. Secondly, we train an end-to-end convolutional neural network (CNN) to directly predict the correct alignment. We evaluate the approach on a data set recorded on rural roads in German cities on which both approaches significantly improve alignment accuracy. Furthermore we show that an end-to-end learning approach can successfully be used in this context allowing the alignment to be performed in a single forward pass. In this context we also present current challenges like obtaining accurate ground truth data.

1 Introduction

Future ADAS attempt to offer advanced functionality such as risk prediction or lane level navigation. Therefore knowledge of the ego vehicles position and orientation relative to the map is required. Standard absolute localization approaches relying on GNSS fail since they do not account for errors related to the map i.e. an offset to real world coordinates. To circumvent this problem, some approaches for map relative localization have been investigated, i.e. [1]. While offering highly accurate localization such strategies mostly require maps build with expensive equipment. As an attractive alternative Cao et al. [2] show that monocular camera images and low precision maps can be used for inexpensive self-localization in the context of ADAS. This paper presents strategies for improving this original approach.

2 Improvement Strategies

The algorithm from Cao et al. aims at correcting the coarse position estimation provided by GNSS. To that end a perspective view of the map (a candidate image) is created. To determine a position correction multiple candidate images



(a) (b) (c) (d)

Fig. 1: Candidates for the camera image in (a) from the KITTI data set [4].

with different offsets from the GNSS position are generated (Figure 1). The corrected position is the initial GNSS position plus the offset minimizing the cosine distance between HoG features [3] extracted from both the candidate and the camera image. We propose two improvements to the original approach which we evaluate focusing on lateral translations. As a first strategy we want to reduce the effects of dominant non-road edges in the camera image by partly replacing the feature extraction step of the original algorithm. This is done by using a pre-trained segmentation model presented in [5] to predict the road area in the camera image before applying HoG. As a second approach we will eliminate the iterative nature of the algorithm by directly learning the correct alignment. For that we use a CNN that takes candidate and camera image as input. The output is the lateral translation d that aligns candidate to camera image.

3 Experimental Setting and Results

We evaluate all three approaches on a data set of 540 positions from two KITTI streams [4] split into 80% train and 20% test subsets and use the KITTI positioning data as ground truth. For each image in the training set 200 candidate images are rendered with offsets $d_r \in [-2m, 2m]$. The goal of the evaluation is to predict these offsets. For both HoG based approaches we determine the offset by optimizing across 31 lateral translations ($d_h \in [-3m, 3m]$) around the starting position. The CNN approach predicts the offset directly. We consider the mean absolute (mae) and root mean squared error (rmse) in lateral translation with respect to ground truth. Figure 2 shows the predictions for all three approaches on the test set. The HoG baseline algorithm results in an mae/rmse of 0.79m/1.37m, the HoG approach with segmentation produces 0.33m/0.55m and the CNN results in 0.2m/0.3m.

4 Conclusion and Outlook

The presented results demonstrate that both approaches significantly improve alignment accuracy. The better performance of the CNN is likely due to the fact that the data set has a low diversity of driving situations suggesting that the model is over fitting to situations contained therein. More specifically this means that the model is likely to perform worse on other data then the current data set. However, this remains to be tested and the current results still provide indications for the general feasibility of the approach. A challenge is that current ground truth data is not always accurate which causes the model to learn

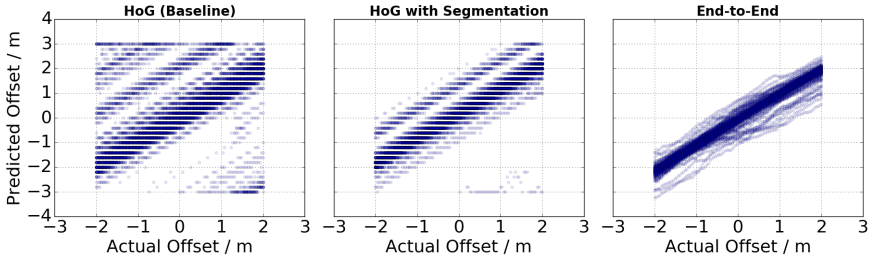

Fig. 2: Performance of the three algorithms on the test set.

a visually imprecise alignment. In future work we will test the approach on a data set containing more diverse driving situations. However, obtaining accurate ground truth is a major challenge which might require manual generation. Furthermore we will extend the evaluation beyond a single lateral degree of freedom where we would like to evaluate the algorithm's run time as well. In that case we expect alignment to be dramatically slower for the HoG approaches needing to optimize offsets in multiple dimensions i.e. longitudinal, lateral and altitude. In contrast to that the CNN approach would still be able to predict the correct offset in a single forward pass. This has been demonstrated in other domains [6], but still remains to be proven here which is a challenging problem because visual alignment might be ambiguous i.e. regarding longitudinal translation on a straight road.

Acknowledgments

This work has been supported by the European Union's Horizon 2020 project *INLANE*, under the grant agreement number 687458.

References

1. Yan Lu, Jiawei Huang, Yi-Ting Chen, and Bernd Heisele. Monocular localization in urban environments using road markings. In *IV, 2017 IEEE*, pages 468–474. IEEE, 2017.
2. G. 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 ITSC*, pages 498–504, Nov 2016.
3. N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *2005 IEEE CVPR*, volume 1, pages 886–893 vol. 1, June 2005.
4. Andreas Geiger, Philip Lenz, Christoph Stiller, and Raquel Urtasun. Vision meets robotics: The KITTI dataset. *IJRR*, 2013.
5. M. Teichmann, M. Weber, M. Zoellner, R. Cipolla, and R. Urtasun. Multinet: Real-time joint semantic reasoning for autonomous driving. *arXiv preprint arXiv:1612.07695*, 2016.
6. S. Miao, Z. J. Wang, and R. Liao. A CNN regression approach for real-time 2D/3D registration. *IEEE Transactions on Medical Imaging*, 35(5):1352–1363, May 2016.