

Mapping and Localization using Surround View

Marc Sons¹, Martin Lauer², Christoph G. Keller³ and Christoph Stiller²

Abstract—Intelligent vehicles heavily rely on robust and accurate self-localization. Global navigation satellite systems (GNSS) are not reliable in urban environments due to multipath and shadowing effects. Vision-based localization offers a promising alternative.

We present a high-precision six degrees of freedom self-localization method using multiple cameras covering the surrounding environment. First, a point feature map is created using images from a previous pass of the area to map. Thereafter, the map is used for high-precision localization in real-time. While localization, a rough prior estimate of the current pose is used to shrink the search space for feature matching by projecting mapped landmarks into current images. Then, stored observations of the projected landmarks are matched to actual observations and the egopose is estimated by back-projection error minimization. Thereby, our map structure provides mapped landmarks efficiently towards localization with multiple cameras. In real-world experiments we show that our approach provides reliable localization results while passing the mapped area in arbitrary orientation.

I. INTRODUCTION

Current intelligent transportation systems require robust and accurate self-localization in a multitude of scenarios. Common approaches couple inertial measurement units (IMU) and GNSS reach centimeter accuracy merely in open sky environments. Despite promising approaches using spinning laser scanners [16], recent vision-based approaches ([1], [5], [20]) show highly accurate results. Using cameras is particularly suitable for intelligent vehicles since they are cheap and easy to integrate. However, a major drawback of these approaches is the requirement to drive accurately on the mapped track to provide reliable localization results. Localization fails due to small deviations of the vehicle's orientation or even while driving slightly beside the track. A point feature based approach [1] is fused with a lane-marking based localization [12] and an IMU to estimate the egopose during the memorial Bertha-Benz drive [3]. Here, two supplementary approaches are combined to provide reliable localization in inner-city scenarios as well as in poorly structured areas. From a practical point of view, maintaining two approaches and combining them cause additional effort. More favorable would be to maintain merely one vision-based localization system which performs in both scenarios. Here, we present a high-precision vision-based localization method covering six degrees of freedom using multiple cameras, which span a surrounding field of view (FOV) in total. We localize relatively to a map created fully automatically

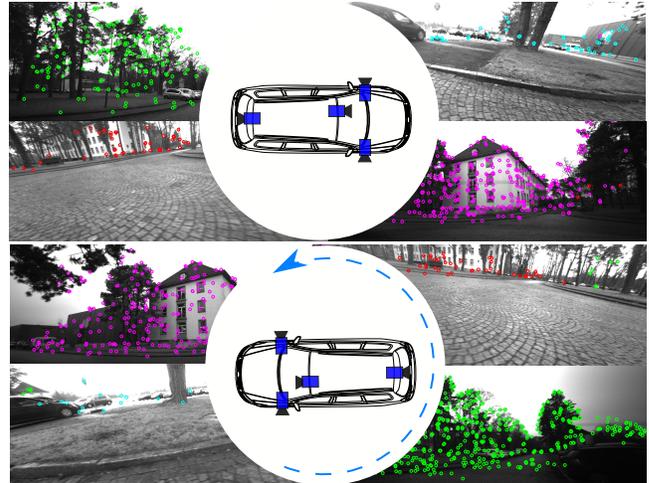


Fig. 1. Visualization of matches which support egopose estimate using four cameras(image top left \rightarrow front camera, image top right \rightarrow left camera, image bottom left \rightarrow right camera, image bottom right \rightarrow rear camera). The color of a point encodes the camera seen the landmark while mapping. The vehicle is driving towards mapping direction in the upper scenario whereas in bottom the vehicle is driving in opposite direction. Obviously, landmarks mapped in the front/left camera are matched in the rear/right camera and vice versa. The number of inlier matches is similar and pose estimation has same precision as in the upper scenario ($\sim 5\text{cm}$).

from images of a previously pass of the same track. Mapping as well as localization do not require further measurements from GPS, IMU, odometers or other sensors. For mapping, salient image points are matched and tracked over time in all cameras. Through bundle adjustment, the tracks are utilized to reconstruct vehicle poses and 3D-landmarks. One major contribution is the dynamic grid map, which provides the mapped landmarks and their related features and corresponding observation positions during localization. This grid structure utilizes the properties of the surround view benefitly towards localization with multiple cameras and is independent of the size of the mapped area. Our localization approach is based on the principles described in [1]. First, a rough prior estimate of the egopose is estimated using place recognition [10] and visual odometry [17]. The prior estimate is used to select nearby landmarks and match them to current image observations near to the projection of the landmarks in image space. Finally, the egopose is estimated by minimization of the back-projection error of the matches. In difference to [1], our landmark projection scheme enables to match mapped landmarks to observations of each camera. Due to the surround view, our map provides landmark observations from all directions and therefore, we are able to localize independently of the

¹with the Forschungszentrum Informatik(FZI), Karlsruhe, Germany

²with the Institute of Measurement and Control, Karlsruhe Institute of Technology(KIT), Karlsruhe, Germany

³with the Daimler AG, Research & Development, Sindelfingen, Germany

orientation of the vehicle. Thereby, our novel map structure implicitly ensures that more relevant landmarks are preferred and sampling of the landmarks is equally distributed in orientation and distance independent of the surrounding environment. By this, accuracy and robustness of the localization is increased. Furthermore, using surround view increases the availability of localization in poorly structured areas.

We show in real-world experiments that we are able to localize precisely while passing the mapped track in arbitrary orientation. Fig. 1 shows successfully matched landmarks to features from four cameras. The color of the points encodes the camera which saw the landmarks while mapping. Obviously, the number of matches in the upper scenario is similar to the bottom one, where the vehicle is driving in opposite direction. Our experiments show that it is possible to localize with similar success rate and accuracy while passing the track in opposite mapping direction.

II. RELATED WORK

The presented work is strongly related to recent work on vision-based Simultaneous Localization and Mapping (SLAM) [4], [6], [11]. SLAM is about computing a map of an unknown area while localizing simultaneously within this map concurrently. Hereby, the egopose is jointly estimated with all detected landmarks. Therefore, the complexity increases during runtime. To bypass this issue and enable real-time localization for automated driving in city scaled areas, it has been established to estimate merely the egopose in a previously acquired map [3].

Schreiber et al. [12] uses a lane-markings map for localization. Stereo vision is used to compute virtual bird-view images of the groundplane and lane-markings are extracted from them. All detected lane-markings are referenced to reconstructed poses from visual odometry. During localization, current detected markings are matched to the map. Using lane markings for localization provides an accurate lateral position estimate, however, the longitudinal accuracy is poor. *Agrawal et al.* uses stereo vision coupled with a low cost GPS receiver in [18]. Rougher outdoor scenarios have been addressed by the same group in [19]. Motion estimates from visual odometry are fused with gps measurements to avoid drift. In inner-city scenarios where GPS produces highly inaccurate measurements and fails frequently, a reliable localization can not be made available. In [1], [10] different versions of topologic localization approaches are presented. The topologic map stores image signatures to corresponding vehicle poses. The best fitting map pose is estimated in real-time by comparing map signatures with the signature of the current image while localization. However, these approaches merely give a rough estimate of the global position. *Lategahn et al.* [1] uses a topologic localization to initialize egopose estimation through image point features, which are associated to landmarks from a previously generated map. The egopose is estimated precisely by minimizing back-projection error of the landmark-image point correspondences. In order to reconstruct a correctly scaled trajectory, stereo vision is used while mapping whereas a single camera

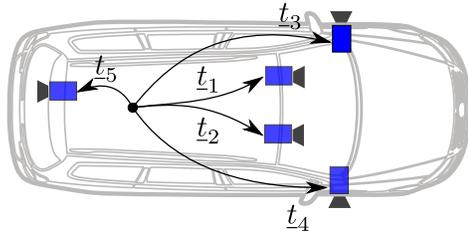


Fig. 2. Each camera is geometrically referenced through a 3D-transformation t_i to a rig coordinate system.

is used for localization.

The presented work extends the methods described in [1] towards the use of multiple cameras for localization and mapping. Due to their map structure and matching scheme, it is necessary to drive accurately on the track to localize successfully. Furthermore, the mapped landmarks are used without considering their quality and distribution in the map. Our novel map structure and localization approach based on the same principle methods overcome these flaws and utilizes the surround view measurements efficiently, which enables localization while passing the mapped area in arbitrary vehicle orientation.

III. PREREQUISITES

Before mapping and localization is explained in detail, we briefly describe the sensor setup, measurement process and vision front end.

A. Camera Setup

We assume a vehicle with several solidly mounted gray-scale cameras. There are no constraints on the number, position, orientation and type of the cameras. However, it is advantageously to cover a large range of sensing around the vehicle for localization purposes and additionally, having overlapping FOVs for mapping. All cameras are jointly triggered with a particular frequency. Hence, images of all cameras have equal timestamps. Finally, a jointly calibration of all cameras is assumed which provides an affine transformation $t \in \mathbb{R}^6$ for each camera referring to a rig coordinate system (see Fig. 2). Furthermore, the calibration provides a function

$$\pi_i(l) = z, \quad (1)$$

which maps a landmark $l \in \mathbb{R}^3$ from real-world to a point $z \in \mathbb{R}^2$ in the image plane of the i -th camera [15].

B. Vision Frontend

For mapping as well as for localization, a corner-/blob-detector [17] is used to extract key points from the recorded images and a local image patch descriptor is used to describe each key point. While mapping, key points are matched between subsequent images from each camera and between images from suitable different cameras at the same timestamp, which requires that the corresponding cameras have overlapping FOVs. All correspondences are tracked over time. During localization, point features from the current

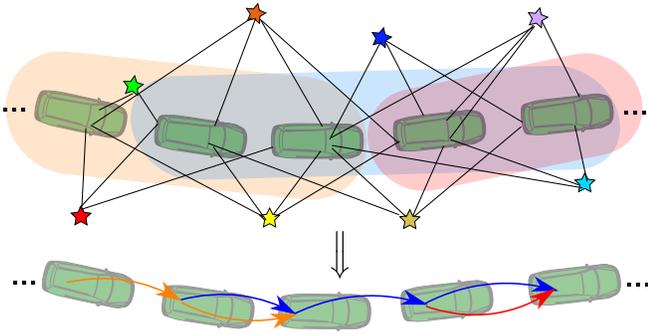


Fig. 3. Top: Problem (3) is divided into multiple spatial overlapping windows (colored areas). Landmarks (stars) are observed from different cameras and vehicle poses. Bottom: The derived pose differences (colored arrows) from the windowed bundle adjustment jointly form a pose adjustment problem along the entire track. The color of an arrow indicates the related adjustment window.

image are matched to observations which are stored together with corresponding 3D-landmarks in the map. In both cases, a reliable feature matching is required, e.g. the used descriptor should not be sensitive to image distortions due to viewpoint and illumination changes. We use the DIRD-descriptor [2] since it is robust against illumination changes which is particularly important for localization purposes. We match those key points of which the $L1$ -norm of the feature vector difference is minimal. Furthermore, a threshold and a uniqueness check is performed on the matching costs to decide whether a match is correct. To fulfill real-time constraints and to decrease the error matching rate, the search space in the image is spatially constrained through regions of interest (ROIs) and epipolar geometry [9]. Furthermore, matching is efficiently performed using Single-Instruction-Multiple-Data(SIMD) instructions.

IV. MAPPING

To create the map, we assume that the area to map is passed once and images of this track are recorded. The set of images recorded from all cameras at a particular time define a single vehicle pose $\underline{p} \in \mathbb{R}^6$, whereby a pose is an affine transformation referring to the origin of the map. Each landmark \underline{l} in the map is induced by a set $Z = \{\underline{z}_1, \dots, \underline{z}_n\}$ of matched key points. Equation (1) defines the relationship

$$\pi_i((\underline{p}^{-1} \otimes \underline{t}_i) \cdot \underline{l}) = \underline{z} \quad (2)$$

between a landmark \underline{l} , a corresponding key point \underline{z} and a vehicle pose \underline{p} from which the observation was generated. In (2), the map referred landmark is firstly transformed into the camera coordinate system and then projected into the image plane. Thereby, the $\otimes : \mathbb{R}^6 \times \mathbb{R}^6 \rightarrow \mathbb{R}^6$ -operator concatenates two affine transformations and the $\cdot : \mathbb{R}^6 \times \mathbb{R}^3 \rightarrow \mathbb{R}^3$ -operator transforms a 3D-point.

Estimating the map

$$\arg \min_{P, L} \left\| \sum_{i=1}^N \sum_{\underline{z}_{j,k} \in Z_i} \pi_k((\underline{p}_j^{-1} \otimes \underline{t}_k) \cdot \underline{l}_i) - \underline{z}_{j,k} \right\|^2 \quad (3)$$

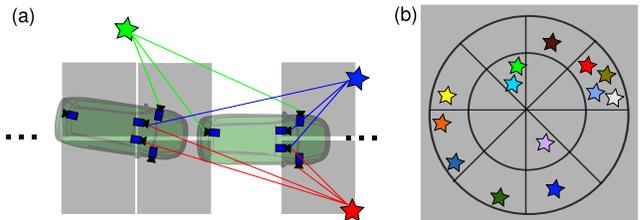


Fig. 4. (a) The landmarks (stars) are stored together with their corresponding features and observation poses in grid cells (gray tiles). A landmark is stored in all cells where one of their observation poses (blue cameras) is located inside. Several observations of a landmark are stored together when they are located in the same cell. (b) All landmarks stored in one cell are divided into disjoint sections of a two-dimensional polar grid depending on the landmarks' positions.

corresponds to find all landmark positions $L = \{\underline{l}_1, \dots, \underline{l}_N\}$ and vehicle poses $P = \{\underline{p}_1, \dots, \underline{p}_M\}$ that best explain the measurements $T = \{Z_1, \dots, Z_N\}$. Thereby, landmark $\underline{l}_i \in L$ corresponds to keypoint set $Z_i \in T$. The indices j and k of keypoint $\underline{z}_{j,k} \in Z_i$ refer to the k -th camera and vehicle pose \underline{p}_j from which $\underline{z}_{j,k}$ was measured.

Equation (3) is a nonlinear least squares problem, which could be solved using the Levenberg-Marquardt algorithm [7]. To compute the map in reasonable time, we use a sparse matrix solver [14], which efficiently exploits the block diagonal structure of the linearized information matrix of (3). Furthermore, a suitable loss function [14] is applied to residual evaluation to become robust against sporadically outlier measurements. To initialize (3), we are using visual odometry.

Mapping city-scaled areas is the most common case for automated driving. Here, the complexity of problem (3) exceeds computation and memory bounds on average desktop computers. To assure bounded complexity, we divide the map into spatially overlapping windows and solve (3) for each window. Then, sequential pose differences $\underline{\Delta}_{j \rightarrow j+1} = \underline{p}_j^{-1} \otimes \underline{p}_{j+1}$ are extracted from each window. The set D of extracted pose differences from all windows induces a posegraph adjustment problem [8], which scales in the number of vehicle pose differences $|D|$ instead in the number of landmarks $|T|$. Since $|D| \ll |T|$, it is feasible to optimize posegraphs of very large areas. Fig. 3 illustrates this approach. Thereafter, the optimized vehicle poses \hat{P} are kept fixed and each landmark $\underline{l} \in L$ can be triangulated separately.

A. Map Structure

For robust feature matching during localization (see section V), it is required to provide landmarks from the map whose related observations are nearby to the current vehicle pose in real-time. For this purpose, each landmark $\underline{l}_i \in L$ is stored together with its corresponding set of observations $O_i = \{(\underline{f}_1, \underline{o}_1), \dots, (\underline{f}_n, \underline{o}_n)\}$ in a planar grid since we assume that the vehicle is moving in a plane approximately. Here, \underline{f}_k denotes the feature vector of one of the observations of \underline{l}_i and \underline{o}_k the corresponding pose of the camera in the map. Each cell can be loaded from disk separately and

stores landmark-observation-feature tuples $(l_i, (f_k, o_k) \in O_i)$ whose observation pose o_k is located inside. (see Fig. 4 (a)). The cell dimension is a user-defined parameter.

All landmarks within one cell are further divided into disjoint sections of a two-dimensional polar grid depending on the position of the landmark (see Fig. 4 (b)). The angular and distance resolution of the polar grid are user-defined parameters. Hence, each polar section comprises $M \geq 0$ landmarks in general. Furthermore, within each section the landmarks can be sorted depending on a particular importance measure, e.g. the number of related mapping observations. While localization, the sections are processed sequentially. If $m \leq M$ landmarks of a section are matched successfully to an observation, all untouched $M - m$ other landmarks are skipped and the next polar section is treated. The parameter m is user-defined. Thus, even if the landmarks in the map are not equally distributed due to an inhomogenous environment, the polar grid structure ensures that the landmarks are provided almost evenly distributed. Additionally, landmarks with higher importance are implicitly preferred, which improves accuracy and robustness of the localization. Furthermore, the localization is independent of the number of landmarks stored in the map and, therefore, computational complexity is bounded and adjustable through m and the resolution of the polar grid.

V. LOCALIZATION

Next, the localization method is described. We do not assume that the camera setup is the same as used while mapping. However, it is advantageous to use a similar setup to facilitate feature matching. Our localization approach comprises two major steps. The first step is to compute a rough prediction $q_p \in \mathbb{R}^6$ of the current egopose. This prior estimate is used to select nearby landmark cells from the feature grid (see section IV). The landmarks from the selected cells are then processed in the order as described in section IV-A. Each candidate landmark is projected into current images of all cameras using (2) and q_p . When a projection is valid, a ROI in the image space around the projection is determined. The observation feature of the landmark, which was mapped nearest to q_p is then compared to image features within the ROI. This search space restriction robustifies and speeds up feature matching. This is illustrated in Fig. 5. It is conceivable to determine the size of the ROIs dependent on the uncertainty of q_p , however, we set the size constant to a user defined parameter since we found no considerable influence of uncertainty adapting ROIs during our experiments. Due to this projection scheme, it is possible to match candidate landmarks to key points in the images of all cameras. Finally, the resulting high-precision egopose

$$\hat{q} = \arg \min_q \left\| \sum_{i=1}^L \pi_k((q^{-1} \otimes t_k) \cdot l_i) - z_i \right\|^2 \quad (4)$$

is estimated with thus determined landmark-key point matches $\{(l_1, z_1), \dots, (l_L, z_L)\}$, whereby k determines the camera from which z_i originated. Problem (4) is solved

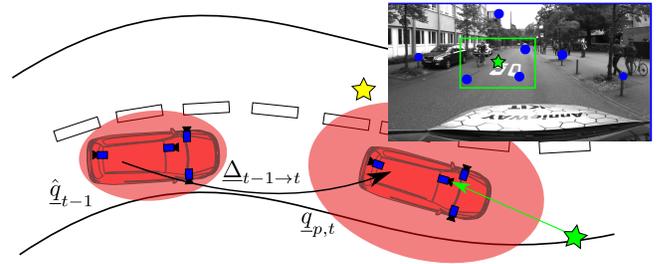


Fig. 5. Visual odometry is used to determine a rough prediction $q_{p,t}$ for the current localization step. The green landmark from a nearby cell of the landmark grid map is projected into one of the images. The feature of the landmark which is nearest to $q_{p,t}$ is matched to current key points (blue points) within the ROI (green box) around the projection of the landmark. The yellow landmark is not matched since it could not be projected into one of the images. The ellipses depict uncertainty of the final estimate \hat{q}_{t-1} of the previous successful localization step, and the rough prediction $q_{p,t} = \hat{q}_{t-1} \otimes \Delta_{t-1 \rightarrow t}$ respectively.

using the Levenberg-Marquardt algorithm. Furthermore, (4) is applied within a RANSAC scheme [21] to become robust against outlier matches. The optimization is initialized with q_p .

It remains to explain how we determine q_p . For this, topological localization [1] is used at the very first localization step, whereby the topologic map is created from the mapping drive images and their corresponding map poses (see section IV). After the first successful localization step, a motion estimate $\Delta_{t \rightarrow t+1} \in \mathbb{R}^6$ from visual odometry is used to predict $q_{p,t+1} = \hat{q}_t \otimes \Delta_{t \rightarrow t+1}$ for the following localization step $t+1$ (see Fig. 5).

VI. EXPERIMENTS

Our vehicle is equipped with five gray-value cameras. All cameras are jointly triggered with a frequency of 10 Hz for mapping and localization. The placement of the cameras at the vehicle is similar to Fig. 2. The two cameras in front and the rear camera have a FOV of approximately 110° . They are mounted behind the windshield whereas the two cameras to the sides have wide angle lenses with a FOV of approximately 175° and are mounted at the front mudguard. The average image resolution of all images is 1500×500 pixels after undistortion.

To demonstrate the capabilities of our surround view localization, we mapped different test tracks on a traffic-reduced area. Most parts of the mapped area are structured poorly. We evaluate our localization while passing these tracks in different directions several days after mapping. For our first experiment, we mapped a meander-shaped trajectory (red curve in Fig. 6). The map consists of 324 poses and 572278 landmarks whose back projection residual is less than 3 pixels, each. We match features consecutively in each camera and between images of the two front cameras. All matches are further tracked jointly. For windowed bundle adjustment, we divide the map into windows with a length of 20 poses and let overlap 5 poses at the begin and end of each window. The cell dimension of the grid is $1\text{m} \times 1\text{m}$. The angular and distance resolution of the polar grid of each cell is 1.5°

and 0.3m, respectively. While localization, we step to the next polar region after obtaining $m = 1$ successful match or matching 10 landmarks unsuccessfully from the current region.

To show that we are able to provide reliable localization results while passing the mapped track from arbitrary directions, we localized in real-time (10Hz) while driving a meander-shaped trajectory orthogonally to the map (purple curve in Fig. 6) and, furthermore, while driving randomly (black curve in Fig. 6) within the mapped area. We localized successfully within the yellow highlighted areas in Fig. 6. Thereby, a single localization step is successful when the absolute number of inlier matches of the RANSAC and the inlier-matches-ratio exceed certain thresholds and the recent history of localization estimates is sufficiently smooth. To our knowledge, it is highly unlikely that a consecutive sequence of single-shot localization estimates which fulfill such criteria is rendered from benign coincidences. Furthermore, we compared the recorded images with a visualization of the localization estimates on the map to get a qualitative feedback of the correctness. Fig. 6 shows clearly that we are able to localize whenever the vehicle is close to the mapped track independent of its orientation. While driving off the track, we integrate motion increments from visual odometry. Fig. 7 shows quantitative results of the localization drives. The upper chart shows the localization ratio, the number of successfully localized steps divided by the overall number of steps. The middle and the lower chart show the average accuracy in position and rotation of all successful localization steps. To determine the accuracy of the localization, we fit smooth vehicle model trajectories to the estimated localization poses. The fits incorporate data from a vehicle odometer, which provides reliable measurements of the angular rate and the driven speed of the vehicle. For evaluation, the single shot localization estimates are compared to the fitted data. The right two bars in all charts of Fig. 7 show the localization results of the orthogonally meander pass (M) and the random drive (R). The localization ratio is low since the vehicle is driving off the mapped track most of the time. The rotational- and positional accuracy of these runs (see Fig. 7) are below 1° and 0.3m, respectively.

Fig. 6 shows the mapped track (blue) of our second experiment. The arrows show the driving direction while mapping. The map is computed equally to the meander-shaped map and consists of 597 vehicle poses and 1128237 landmarks. Fig. 6 illustrates the low drift of the map since the track fits excellently to the topview images from OpenStreetMap. The loop is closed accurately without incorporating GPS data or loop closure constraints during mapping. In this experiment, we compare the localization performance while passing the mapped track (blue curve in Fig. 6) in forward (F)- and backward (B) direction and while driving in both directions approximately 2m beside the track (FO, BO) to resemble driving on next unmapped lane. The left four bars in Fig. 7 show quantitative results of these experiments. The localization ratio of all passes is > 0.9 . Furthermore, there are no significant differences in the accuracy between all

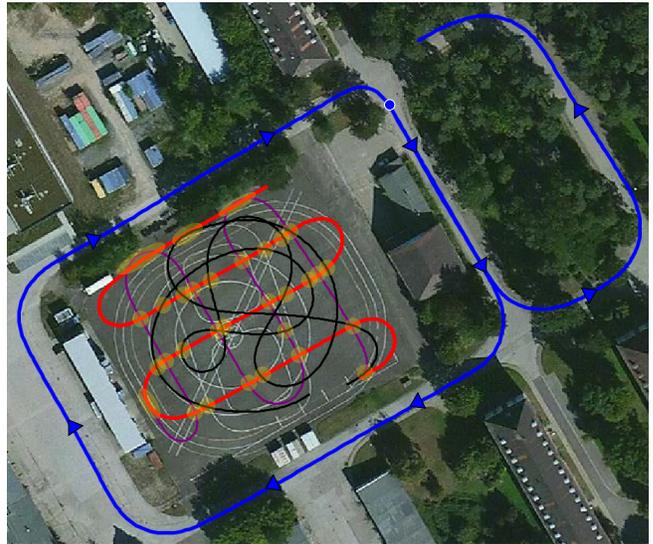


Fig. 6. Aerial image of our experimental area. The mapped positions of the first experiment are shown in blue. The arrows show the driving direction while mapping and the circle show the starting point. The loop is closed accurately without including GPS data or loop closure constraints which show the low drift error qualitatively. The meander-shaped red trajectory shows the mapped track on a free area for the second experiment. The purple and black curve show the localized positions of two localization runs. We successfully localize whenever the map track is passed orthogonally.

passes. The rotational- and positional uncertainties of all runs are $< 0.2^\circ$ and $< 0.07\text{m}$, respectively. The experiments show that we are able to localize with the same precision while driving in opposite mapping direction and while driving $\sim 2\text{m}$ beside the track in both directions.

It is noticeable from Fig. 7 that the localization accuracies of the latter experiments are significantly better. While localizing the random (R) and orthogonally meander-shaped pass (M) in the meander map, landmarks mapped in the front- and rear images are matched to observations to the side cameras and vice versa. However, during our latter experiments we drove mainly in direct and opposite mapping direction. Here, landmarks mapped in the rear-/left cameras are mainly matched to observations from the front-/right cameras and vice versa. Since the mounting height, orientation and the type of lenses of the side cameras are significantly different to the front and rear cameras, the absolute number of correctly matched landmarks is reduced. Hence, the localization uncertainty is increased. At this point, a descriptor which is more robust to such image transformations or a more convenient camera setup could overcome this flaw.

VII. CONCLUSION

Within this work, we presented a localization method based on a pre-built map using several cameras mounted at the vehicle. No additional sensors are required for mapping and localization. We show how our grid-based map provides the mapped features efficiently towards feature-based localization using a surround view setup. The presented map structure implicitly prefers more relevant landmarks and ensures a beneficial selection of landmarks around the vehicle.

Due to the landmark projection it is possible to match a landmark to an image observation of every camera. Hence, we are able to localize even if the vehicle's orientation is arbitrary different to direction of the map.

We reinforce the capabilities of our approach in real-world experiments and showed that we are able to localize high-precisely and independent of the driving direction. We further show that it is possible to provide reliable egopose estimates while passing the mapped track orthogonally and while driving beside the track. The presented localization method overcomes the limitations of existing vision-based approaches, which require driving similar to the mapped track to localize reliably.

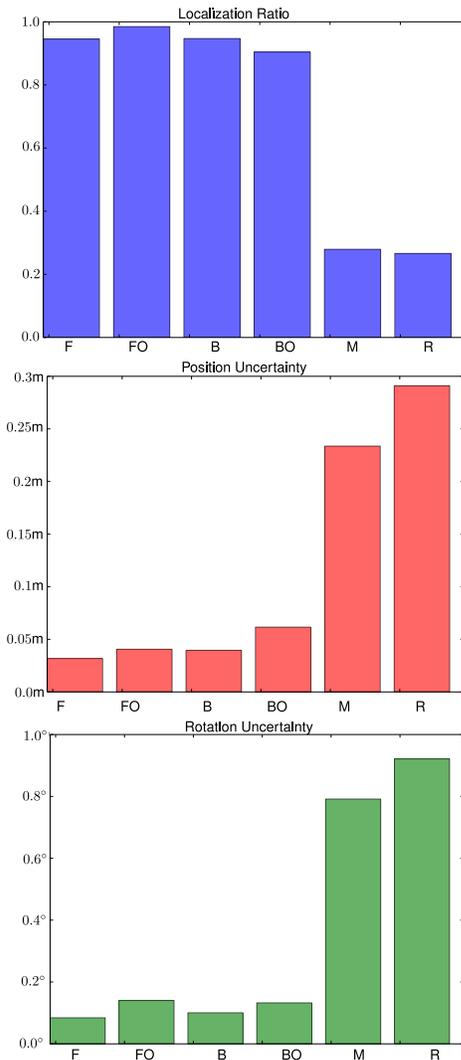


Fig. 7. Average results of the localization experiments. The four left bars in each chart show localization results while driving along the blue map track in Fig. 6. While driving in mapping direction on (F) and beside the track (FO) we got similar localization ratios and accuracies as driving in opposite direction on (B) and beside the map track (BO). The right bars show the results of the second localization experiment while driving an orthogonally meander-shaped track (M) and randomly within the mapped area. Here the localization ratio is significantly reduced since we are off the mapping track at most timestamps. The accuracies are worse since the landmarks mapped in the front and rear cameras are matched to features in the side cameras and vice versa.

REFERENCES

- [1] H. Lategahn and C. Stiller, *Vision-only localization*, IEEE Transactions on Intelligent Transportation Systems, 15(3), 1246-1257, 2014
- [2] H. Lategahn, J. Beck and C. Stiller, *Dird is an illumination robust descriptor* In 2014 IEEE Intelligent Vehicles Symposium Proceedings (pp. 756-761), IEEE, 2014
- [3] J. Ziegler, P. Bender, M. Schreiber, H. Lategahn, T. Strauss, C. Stiller, ... and E. Kaus, *Making Bertha drive-an autonomous journey on a historic route*, IEEE Intelligent Transportation Systems Magazine, 6(2), 8-20, 2014
- [4] T. Bailey, and H. Durrant-Whyte, *Simultaneous localization and mapping (SLAM): Part II*, IEEE Robotics & Automation Magazine, 13(3), 108-117, 2006
- [5] J. Ziegler, H. Lategahn, M. Schreiber, C. G. Keller, C. Knoppel, J. Hipp, M. Haueis and C. Stiller, *Video based localization for Bertha*, In Intelligent Vehicles Symposium Proceedings (pp. 1231-1238), 2014 IEEE
- [6] M. Bosse, P. Newman, J. Leonard and S. Teller, *Simultaneous localization and map building in large-scale cyclic environments using the Atlas framework*, The International Journal of Robotics Research, 23(12), 1113-1139, 2004
- [7] G. Grisetti, H. Strasdat, K. Konolige, W. Burgard, *g2o: A general framework for graph optimization*, In IEEE International Conference on Robotics and Automation, 2011
- [8] M. Sons, H. Lategahn, C. G. Keller and C. Stiller, *Multi trajectory pose adjustment for life-long mapping*, In 2015 IEEE Intelligent Vehicles Symposium (IV) (pp. 901-906), IEEE, 2015
- [9] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*, Cambridge university press, 2003
- [10] H. Badino, D. Huber and T. Kanade, *Visual topometric localization*, In Intelligent Vehicles Symposium (IV), (pp. 794-799), IEEE, 2011
- [11] J. McDonald, M. Kaess, C. Cadena, J. Neira and J. J. Leonard *6-DOF multi-session visual SLAM using anchor nodes*, 2011
- [12] M. Schreiber, C. Knöppel and U. Franke, *Laneloc: Lane marking based localization using highly accurate maps*, In Intelligent Vehicles Symposium (IV), (pp. 449-454), IEEE, 2013
- [13] F. Fraundorfer and D. Scaramuzza, *Visual odometry: Part II: Matching, robustness, optimization, and applications*, IEEE Robotics & Automation Magazine, 19(2), 78-90, 2012
- [14] S. Agarwal, N. Snavely, M. S. Seitz and R. Szeliski, *Bundle adjustment in the large*, In European conference on computer vision (pp. 29-42), Springer Berlin Heidelberg, 2010
- [15] T. Strauss, *Kalibrierung von Multi-Kamera-Systemen-Kombinierte Schätzung von intrinsischem Abbildungsverhalten der einzelnen Kameras und deren relativer Lage zueinander ohne Erfordernis sich überlappender Sichtbereiche*, Doctoral dissertation, Karlsruhe, Karlsruher Institut für Technologie (KIT), Diss., 2015
- [16] F. Moosmann and C. Stiller, *Velodyne slam*, In Intelligent Vehicles Symposium (IV), (pp. 393-398), IEEE, 2011
- [17] A. Geiger, J. Ziegler and C. Stiller, *Stereoscan: Dense 3d reconstruction in real-time*, In Intelligent Vehicles Symposium (IV), (pp. 963-968), IEEE, 2011
- [18] M. Agrawal and K. Konolige, *Real-time localization in outdoor environments using stereo vision and inexpensive gps*, In Pattern Recognition, ICPR 2006, 18th International Conference on, vol.3 (pp. 1063-1068), IEEE, 2006
- [19] K. Konolige, M. Agrawal and J. Sola, *Large-scale visual odometry for rough terrain*, Robotics Research (pp. 201-212) 2011
- [20] H. Chiu, et al. *Sub-meter vehicle navigation using efficient pre-mapped visual landmarks*, In Intelligent Transportation Systems (ITSC), 2016 IEEE 19th International Conference on (pp. 505-512), IEEE, 2016
- [21] M. A. Fischler and R. C. Bolles *Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography*, In Communications of the ACM 24(6) (pp. 381-395), 1981