Constrained Multi Camera Calibration for Lane Merge Observation

Kai Cordes and Hellward Broszio
VISCODA GmbH, Hannover, Germany

Keywords: Multi Camera Calibration, Lane Merge, Multi View, Vehicle Localization.

Abstract: For the trajectory planning in autonomous driving, the accurate localization of the vehicles is required. Accurate localizations of the ego-vehicle will be provided by the next generation of connected cars using 5G. Until all cars participate in the network, un-connected cars have to be considered as well. These cars are localized via static cameras positioned next to the road. To achieve high accuracy in the vehicle localization, the highly accurate calibration of the cameras is required. Accurately measured landmarks as well as a priori knowledge about the camera configuration are used to develop the proposed constrained multi camera calibration technique. The reprojection error for all cameras is minimized using a differential evolution (DE) optimization strategy. Evaluations on data recorded on a test track show that the proposed calibration technique provides adequate calibration accuracy while the accuracies of reference implementations are insufficient.

1 INTRODUCTION

Automated driving is regarded as the most promising technology for improving road safety and efficiency in the future (Fallgren et al., 2018). In the early phases of partially automated driving, the driver is required to constantly monitor the environment to be able to take back the control of the vehicles whenever the need arises. In the future, fully automated driving systems will allow the driver to remain completely out of the loop. Vehicles are expected to take over the complete driving task.

1.1 Automated Driving, Cooperative Maneuvers

As part of the automated driving tasks, a vehicle should be able to perform appropriate maneuvers, such as automatic lane changing whenever needed. For this, the cooperation with nearby vehicles is crucial. Cooperative maneuvers enhance the safety and help the vehicles getting through difficult traffic situations. Vehicle-to-anything (V2X) communication will ensure the distribution of information using a new generation of mobile communication technology. A major need is the accurate localization of the vehicles (Fernández Barciela et al., 2017). The localization of a vehicle is a main part of current research in projects, such as 5GCAR[1].

[1]http://www.5gcar.eu

In addition to their self-localization capabilities, vehicles are localized using external sensors, such as cameras positioned nearby the road. This is crucial for the integration phase, where only a subset of vehicles is equipped with self-localizing and communicating technology, the *connected* vehicles. The *unconnected* vehicles are localized with a multi camera system positioned near to the road. One important application scenario for the joint localization of connected and unconnected vehicles is the lane merge (Brahmi et al., 2018).

1.2 Lane Merge Coordination

In the lane merge, one vehicle merges into a group of vehicles driving on the motorway as shown in Figure 1. The goal is the coordination of driving trajectories among a group of vehicles to improve the traffic safety and efficiency. A subject vehicle is coordinated with remote vehicles driving on the main lane in order to merge smoothly and safely into the lane without collisions and with minimal impact on the traffic flow. The trajectory recommendations are computed based on road user properties such as position, heading, and speed, continuously transmitted by the connected cars. It is necessary that the system considers unconnected, i.e. non-communicating road users as well. On the one hand, it cannot be assumed that every road user is connected to the network. On the other, remotely monitoring incorporates redundant or even additional information to the system such as hig-

Cordes, K. and Broszio, H.
Constrained Multi Camera Calibration for Lane Merge Observation.
DOI: 10.5220/0007387805290536
In *Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2019)*, pages 529-536
ISBN: 978-989-758-354-4

(a) Lane merge sketch (b) Video capture (c) Camera view

Figure 1: Sketch and video capture of the lane merge scenario using a mobile crane. One of the camera views while the lane merge is performed is shown on the right.

her localization accuracy. This serves the goal of a smooth lane merge without collisions and with minimal impact on the traffic flow. For coordination and vehicle feedback of trajectory recommendations and the lane merge itself, a minimal viewing distance is required, depending on the speed of the vehicles (Luo et al., 2016). Thus, for the external observation of the lane merge, more than one camera is needed. The observation camera system is positioned next to the road. The estimated vehicle data is sent to the lane merge coordination entity via a cellular network. This entity plans the cooperative maneuver and distributes corresponding instructions to connected vehicles, while the behavior of unconnected vehicles is predicted and considered (Brahmi et al., 2018).

Since the positional accuracy is of key importance, a highly-accurate calibration is required. The proposed approach incorporates geometric knowledge about the scene and about the installation of the cameras. The scene knowledge consists of accurately measured landmarks on the road. For the multi camera system, the height above the ground and relative distances between the cameras are measured. This information in incorporated in a joint optimization of all cameras.

1.3 Related Work

The accuracy demands for the camera calibration in tracking applications depends on the tracking scenario. While people trackers focus on the reliable tracking in the 2D image plane (Milan et al., 2016; Leal-Taixé et al., 2017), vehicle tracking for cooperative maneuvers requires highly accurate cameras for two reasons: (1) tracking accuracy in 3D is of special importance since collisions are to be avoided, and (2) larger viewing distances are expected which increases the demand for accurate 2D-3D correspondences.

In vehicle observation, most camera calibration approaches make use of the road marking and assume planar roads in the field of view. Tang et al. (Tang et al., 2018; Tang et al., 2017) use manually selected lines on road markings for the calibration of the cameras. The UA-DETRAC benchmark (Wen et al., 2015) neglects camera calibration and focusses on 2D detections with large variations regarding the vehicle models, the recording conditions (different viewpoints and weather conditions), and the vehicle density.

In contrast to these approaches, we focus on a very specific setting, the lane merge. Since the goal is the automated lane merge coordination, we observe one car merging into the main lane of a multi-lane road where several other cars are driving. One of the cars on the main lane opens a gap for the incoming car. The recorded data is expected to provide valuable information to learn trajectory recommendations for the lane merge coordination entity. Since the setup requires large viewing distances, accurate camera parameters are required to fulfill the accuracy demand for the localization of the vehicles.

We provide the following contributions:

- Practical solution for a suitable multi camera setup dedicated to the lane merge observation scenario
- Incorporation of easily measurable metrics of the camera setup
- Evaluations with data recorded on a test track show the accuracy improvements of the proposed techniques.

In the following Section 2, the data recording setup is briefly described. In Section 3, the proposed multi camera calibration approach is explained in detail. Section 4 shows experimental results while Section 5 concludes this paper.

2 LANE MERGE OBSERVATION

The data for the lane merge is recorded on a test track which provides two lanes of approximately 100m

(a) CAM_1 (b) CAM_2 (c) CAM_3 (d) CAM_4

Figure 2: Video streams for the three sets ($SET^{(1)}$, $SET^{(2)}$, $SET^{(3)}$ from top to bottom). The temporal synchronization for a set $SET^{(i)}$ is done manually. In one row, one point in time is shown for each set. The images demonstrate the lane merge application. The camera calibration is done using landmarks as shown in Figure 4 using images without vehicles.

length for the main traffic and the acceleration entry lane for the merging car. The video capturing is done with four video cameras attached to a mobile crane (cf. Figure 1). The cameras capture 1920×1080 pixels at 50 fps. The temporal synchronization is done manually in a post processing step using a video editing tool. For determining the synchronization, a previously defined signal (headlight flash) is recorded. Using the manually determined signal in the video sequences, all the video streams are synchronized as shown in Figure 2. Four cars drive on the main lanes while one car on the entry lane merges into the traffic.

3 MULTI CAMERA CALIBRATION

For the calibration, intrinsic and extrinsic camera parameters are required. For each camera, the radial distortion is computed in a preprocessing step using manually selected lines in the images (Thormählen et al., 2003). For the computation of the extrinsic parameters and the focal length (7 parameters), in contrast to existing approaches in benchmark generation (Tang et al., 2018; Tang et al., 2017), landmarks on the road and their GPS positions are used. The GPS positions are determined using a D-GPS sensor with RTK precision, which provides a localization accuracy of approximately 2cm (RTK-DGPS: real time kinematics differential global positioning system). The landmarks are selected such that (1) they are well distributed in the region of interest and (2) many of them are visible in each camera view. Their positions are chosen on edges of the road markings (fifteen positions) to ease the re-identification in the camera images.

Additional geometrics of the camera setup are collected using a laser scan tool which measures the height of the camera (distance to the ground plane) and relative distances between cameras. The camera setup (example snapshot shown in Figure 3) is tested for three different positions of the mobile crane with differently mounted cameras to capture data sets with different viewpoints.

The proposed calibration procedure leading to accurate cameras is described in the following sections. It minimizes the reprojection error as defined in Section 3.1 uses an evolutionary optimizer shown in Section 3.2, and is capable of incorporating a priori known geometric constraints given in Section 3.3.

Figure 3: Panorama snapshot of three out of four cameras from the observers perspective. The calibration technique incorporates known distances between cameras. The distance between the top cameras is 77cm, the distance between the top left and the bottom right camera is 127cm. The basket size is 120×60cm, its height is 120cm.

3.1 Cost Function

Using homogeneous coordinates, the 3D-2D correspondence of object point $\mathbf{P}_j \in \mathbb{R}^4$ and feature point $\mathbf{p}_{j,k} \in \mathbb{R}^3$ is given by the camera projection matrix $\mathtt{A}_k \in \mathbb{R}^{3\times4}$:

$$\mathbf{p}_{j,k} = \mathtt{A}_k \mathbf{P}_j \tag{1}$$

The standard technique for camera calibration is the resectioning method (Tsai, 1987; Hartley and Zisserman, 2003), referred to as single camera optimization.

3.1.1 Single Camera Optimization

For optimizing each camera $\mathtt{A}_k$ independently, the 3D points $\mathbf{P}_j$ with known 3D coordinates are projected into the camera planes, resulting in 2D positions $\mathbf{p}_j$. The known 2D representations $\hat{\mathbf{p}}_{j,k}$ of the landmarks are manually selected in the images. The squared distances $d(\hat{\mathbf{p}}_j, \mathtt{A}_k\mathbf{P}_j)^2$ determine the reprojection error:

$$\varepsilon_k = \sum_{j=1}^{J_k} d(\hat{\mathbf{p}}_{j,k}, \mathtt{A}_k\mathbf{P}_j)^2 \tag{2}$$

The minimization of equation (2) with an appropriate initialization for the camera matrix $\mathtt{A}_k$ gives the calibration result for each camera. A camera matrix is built using the 7 parameters (C_x, C_y, C_z) (global coordinate position) (pan, tilt, roll)-angles and the focal length f (Hartley and Zisserman, 2003). The distance $(d(.))^2$ is only computed if the 3D point is projected into the visible region of the camera k resulting in different numbers of points J_k for each camera.

3.1.2 Multi Camera Optimization

Multi camera calibration enables the joint optimization of parameters, such as the knowledge that two cameras have the same focal length. As shown in (Cordes et al., 2015), these additional costraints improve the parameter estimation. The reprojection error is then determined as:

$$\varepsilon = \sum_{k=1}^{K} \sum_{j=1}^{J_k} d(\hat{\mathbf{p}}_{j,k}, \mathtt{A}_k\mathbf{P}_j)^2 \tag{3}$$

In our application, $K = 4$ cameras and up to $J_k = 15$ points are used, depending on the visibility of the landmarks in each camera.

3.1.3 Constrained Multi Camera Optimization

Additional constraints for the calibration technique are given by the camera setup, e.g. measured distances between two cameras or the height of a camera above the ground plane. In each of our camera setups, two cameras have nearly the same height above the ground plane. Several of these constraints can be easily exploited using spherical coordinates as shown in Section 3.3. The global optimization for all parameters including special constraints measured in the test scenario is done using evolutionary computation, cf. Section 3.2.

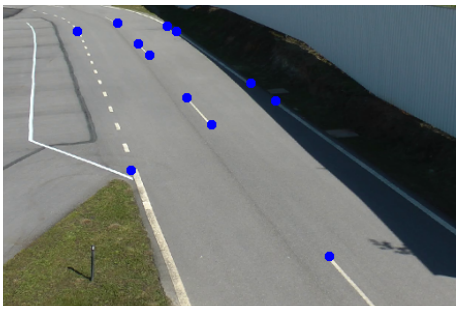


Figure 4: Visualization of 12 of the 15 landmarks used for the camera calibration.

3.2 Differential Evolution Optimization

For the minimization of the the cost function (3), the Differential Evolution (DE) algorithm (Price et al., 2005) is used. It is known as an efficient global optimization method for continuous problem spaces. In our application 4 cameras with 7 parameters each (position, orientation, focal length) are employed. The number of estimated parameters determine the search space dimension.

DE includes an adaptive range scaling for the generation of solution candidates. This enables global search in the case where the solution candidate vectors are spread in the search space and the mean difference vector is large. In the case of a converging population the mean difference vector becomes smaller. This enables efficient fine tuning towards the end of the optimization process (Cordes et al., 2009).

For better convergence, the extended DERSF (DE with Random Scale Factor) method proposed in (Das et al., 2005) is used. It leads to a wider distribution of the candidate vectors and, thus, improves the search. Since the dimension of the search space is high (cf. Section 4.1), spreading the population helps in achieving the global minimum of the cost function.

3.3 Incorporation of Geometric Constraints

To limit the search space, additional constraints are included in the optimization. This leads to faster convergence and a higher probability of achieving the global minimum of the cost function. Therefore, distances between cameras are provided. The incorpora-

tion is done using spherical coordinates. The mapping from Euclidean coordinates to spherical coordinates $(x,y,z)^t \mapsto (r,\theta,\varphi)^t$, $r \in \mathbb{R}^+, \theta \in [0;\pi], \varphi \in [0;2\pi]$, is:

$$\begin{pmatrix} r \\ \theta \\ \varphi \end{pmatrix} = \begin{pmatrix} \sqrt{x^2+y^2+z^2} \\ \arccos(\frac{z}{r}) \\ \arctan 2(y,x) \end{pmatrix} \quad (4)$$

The reverse mapping $(r,\theta,\varphi)^t \mapsto (x,y,z)^t$ is:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} r \cdot \sin(\theta) \cdot \cos(\varphi) \\ r \cdot \sin(\theta) \cdot \sin(\varphi) \\ r \cdot \cos(\theta) \end{pmatrix} \quad (5)$$

We parameterize the first camera position $\mathbf{C}_{\text{MAIN}}$ in Euclidean coordinates and determine all other camera positions $\mathbf{C_i}$ relative to the first one with spherical coordinates. Then, r in equations (4) and (5) corresponds to the camera distance $||\mathbf{C_i} - \mathbf{C}_{\text{MAIN}}||_2$. For example, the distance between the two top cameras in Figure 3 is $r = 77$cm. Incorporating the relative distances for each of the four cameras decreases the number of estimated parameters by 3.

From equation (5), it follows that points with the same height above the ground plane (and positive orientation) lead to a relative angle θ of $\theta = \frac{\pi}{2}$. Thus, for the camera CAM_2, the value θ can be set to $\theta = \frac{\pi}{2}$, because CAM_1 and CAM_2 are installed at the same height h (cf. Figure 3). This decreases the number of estimated parameters by 1.

Both options, the relative camera distance and setting the same height of CAM_1 and CAM_2, are explored in Section 4 in experiments X_3, X_4, leading to the proposed calibration technique. The incorporation of geometric constraints in the calibration greatly improves its accuracy.

4 EXPERIMENTAL RESULTS

For the evaluation, three different camera sets $\text{SET}^{(1)}$, $\text{SET}^{(2)}$, $\text{SET}^{(3)}$ are tested. Each set consists of four cameras $\text{CAM}_1, \ldots, \text{CAM}_4$ attached to the basket of a mobile crane as visualized in Figure 3. For $\text{SET}^{(1)}$, $\text{SET}^{(2)}$ the relative camera positions are approximately the same since only the crane basket height is changed. The camera orientations were adjusted to provide an appropriate field of view for each camera. For $\text{SET}^{(3)}$, the mobile crane adopted a new position and height. All cameras were removed and attached to new positions at the basket of the crane. Images of the cameras while the vehicles perform a lane merge are shown in Figure 2. For the calibration, images without vehicles are used as shown in Figure 4. The images have a resolution of 1920×1080.

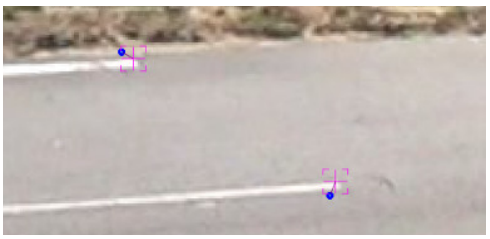


Figure 5: Example for the reprojection error (here ≈ 12 px for both reprojected points shown in blue): The image depicts a small part of the calibration image of $\text{SET}^{(3)}$, CAM_3.

For all cameras, position, orientation, and focal length are estimated. Since the cameras CAM_1 and CAM_2 are identical, used with equal zoom factor, the focal length of these cameras share the same estimation value.

4.1 Setup

For the evaluation, four camera estimation experiments $X_i, i = 1,\ldots,4$ are defined, two for single camera calibration and two for multi camera calibration. The single camera calibration (X_1, X_2) is explained in Section 3.1.1, the multi camera calibration (X_3, X_4) is explained in Section 3.1.2. A priori known information is incorporated as explained in Section 3.3.

X_1: Estimation of all 7 parameters (position, orientation, focal length) for each camera k independently

X_2: Estimation of 6 parameters - (C_x, C_y), orientation, and focal length - with known ground truth height $C_z = h_{gt}$ for each camera k independently

X_3: Estimation of all parameters of the four cameras using the information that (1) CAM_1 and CAM_2 have the same (unknown) height and (2) CAM_1 and CAM_2 have the same focal length (since these cameras are identical), leading to $28-2=26$ parameters

X_4: Estimation like X_3, additionally incorporating the information of the relative distances r_{gt} between CAM_1 and the other cameras, leading to $28-2-3=23$ parameters

The relative camera distances r_{gt} and the camera heights h_{gt} are measured after installation of the cameras using a laser measure tool.

All approaches receive the same input data which is the 3D positions of the 15 calibration markers and their manually selected 2D positions in the images. The radial distortion coefficients are determined in a preprocessing step.

The multi camera calibration approaches X_3, X_4 get a very coarse initialization, such as viewing direction towards the road and a square of 10m $\times$ 10m

Table 1: Reprojection errors for single and the proposed multi camera optimization for $\mathbf{SET}^{(1)}$. The distances are measured in pel.

	CAM_1	CAM_2	CAM_3	CAM_4	$\bar{\varepsilon}$
J_k	12	7	12	15	
ε_1	5.04	20.22	5.16	9.21	9.91
ε_2	6.59	20.33	7.13	9.22	10.82
ε_3	8.16	19.12	4.43	6.55	9.57
ε_4	8.09	19.20	4.86	6.81	9.74

Table 2: Resulting camera heights h_i and relative distances r_i for single and the proposed multi camera optimization for $\mathbf{SET}^{(1)}$. Their mean error is $\bar{e}$. The distances are measured in m.

	CAM_1	CAM_2	CAM_3	CAM_4	$\bar{e}$
h_1	8.09	7.85	7.13	7.05	0.30
r_1	0.00	1.81	1.38	1.73	0.44
h_2	set to ground truth h_{gt}				–
r_2	0.00	1.97	1.02	1.28	0.45
h_3	7.59	7.59	7.12	7.11	0.14
r_3	0.00	0.93	1.26	1.56	0.27
h_4	7.57	7.57	6.94	7.11	**0.10**
r_4	0.00	set to ground truth r_{gt}			–
h_{gt}	7.63	7.63	6.65	7.10	–
r_{gt}	0.00	0.77	1.62	1.27	–

for the (C_x, C_y)-position while the single camera optimization techniques X_1, X_2, are fairly well initialized to achieve the best possible minimum of the cost function. This is done with manual interaction. An example for reprojected points $\hat{\mathbf{p}}_{j,k}$ and their reprojection error is given in Figure 5.

Since the DE optimization used for X_3, X_4 is based on randomized generation of candidate vectors and a random initialization (cf. Section 4.3), mean values of 100 results are reported. The DE optimization employs common parameters (Price et al., 2005). The number of particles is set to 300, the number of generations is 15000. The computation time is about 4 minutes on our I5 2.5 GHz notebook using unoptimized code which is still appropriate for the calibration of all cameras.

4.2 Accuracy Evaluation

The results are given for each of the three sets independently. We report the reprojection errors for $\mathrm{SET}^{(1)}$, $\mathrm{SET}^{(2)}$, and $\mathrm{SET}^{(3)}$ in Table 1, Table 3, and Table 5, respectively. The comparisons with ground truth measurements for the three sets are shown in the Table 2, Table 4, and Table 6, respectively. The heights h_{gt} of the cameras are determined with a laser measure tool which provides high accuracy (error $\approx$ 1mm), e.g. the height of the reference camera CAM_1 for $\mathrm{SET}^{(1)}$ is $h_{gt} = 7.63$m (ground truth), its height for $\mathrm{SET}^{(2)}$ is $h_{gt} = 5.42$m.

Table 3: Reprojection errors for single and the proposed multi camera optimization for $\mathbf{SET}^{(2)}$. The distances are measured in pel.

	CAM_1	CAM_2	CAM_3	CAM_4	$\bar{\varepsilon}$
J_k	13	14	13	14	
ε_1	4.55	18.59	4.21	8.86	9.05
ε_2	6.66	18.79	6.75	8.86	10.27
ε_3	6.20	13.69	3.61	6.39	7.80
ε_4	6.21	14.03	4.14	6.58	7.74

Table 4: Resulting camera heights h_i and relative distances r_i for $\mathbf{SET}^{(2)}$. Their mean error is $\bar{e}$. The distances are measured in m. The relative camera distances r_{gt} are equal to those in $\mathbf{SET}^{(1)}$.

	CAM_1	CAM_2	CAM_3	CAM_4	$\bar{e}$
h_1	5.76	5.27	4.77	4.90	0.21
r_1	0.00	1.12	1.37	1.62	0.32
h_2	set to ground truth h_{gt}				–
r_2	0.00	0.98	0.96	1.29	0.30
h_3	5.39	5.39	4.77	4.89	0.10
r_3	0.00	1.04	1.03	1.41	0.24
h_4	5.45	5.45	4.65	4.83	**0.08**
r_4	0.00	set to ground truth r_{gt}			–
h_{gt}	5.42	5.42	4.44	4.89	–
r_{gt}	0.00	0.77	1.62	1.27	–

The experiments X_i as reported in Section 4.1 lead to reprojection errors ε_i (cf. Tables 1, 3, and 5). For simplicity, the mean errors $\bar{\varepsilon}$ are computed with the same weight for each camera k, not regarding the numbers of points J_k. For the comparison with ground truth measurements, the estimated camera heights h_i and the relative camera distances r_i are shown (cf. Tables 2, 4, and 6). These values are most demonstrative to show the metrics and errors of the calibration results. Experiment X_i lead to results for camera height h_i and relative camera distances r_i. The entries h_{gt} and r_{gt} at the bottom of each of the tables depict the measured ground truth values for camera height and relative distance.

In all evaluations $\mathrm{SET}^{(1)}$, $\mathrm{SET}^{(2)}$, and $\mathrm{SET}^{(3)}$, the mean reprojection error $\bar{\varepsilon}$ tends to increase when more constraints are added to the optimization, i.e. $\bar{\varepsilon}$ for experiment X_2 is always larger than $\bar{\varepsilon}$ for experiment X_1. The same holds for experiment X_4 compared to experiment X_3 (for $\mathrm{SET}^{(2)}$, the results are comparable).

But, the estimation for the relative distance r_2 in experiment X_2 tend to better results for all three sets. It follows that the reprojection error does not provide

Table 5: Reprojection errors for single and the proposed multi camera optimization for **SET**$^{(3)}$. The distances are measured in pel.

	CAM_1	CAM_2	CAM_3	CAM_4	$\bar{\varepsilon}$
J_k	9	6	4	13	
ε_1	8.75	7.06	0.77	10.16	6.69
ε_2	10.19	8.05	5.02	10.71	8.49
ε_3	11.40	9.02	7.26	10.73	9.60
ε_4	14.26	7.79	6.71	10.55	9.83

Table 6: Resulting camera heights h_i and relative distances r_i for single and the proposed multi camera optimization for **SET**$^{(3)}$. Their mean error is $\bar{e}$. The distances are measured in m.

	CAM_1	CAM_2	CAM_3	CAM_4	$\bar{e}$
h_1	4.49	6.69	4.21	4.48	1.12
r_1	0.00	14.18	8.67	3.33	7.98
h_2	set to ground truth h_{gt}				–
r_2	0.00	7.96	7.44	3.35	5.50
h_3	5.03	5.03	5.26	4.92	0.44
r_3	0.00	1.83	4.34	1.69	1.87
h_4	5.48	5.48	5.70	5.32	**0.09**
r_4	0.00	set to ground truth r_{gt}			–
h_{gt}	5.51	5.51	5.53	5.44	–
r_{gt}	0.00	0.32	1.40	0.52	–

a good measure for the comparison of the accuracy. Still, the results for r_2 are far from acceptable with a mean error of $\bar{e} = 0.45$m for SET$^{(1)}$ or even $\bar{e} = 5.50$m for SET$^{(3)}$. To subsume, the error for the single camera calibration regarding camera height and relative camera distance is surprisingly large. The reason is the large uncertainty in the direction of the optical axis of the camera. Variations of the camera position in this direction does not affect the reprojection error much. The small depth in the constellation of the landmarks in SET$^{(3)}$ (side view of the road) make this set the most challenging among these three.

The experimental setup X_3 improves the accuracy results (r_3, h_3) significantly, leading to $\bar{e} = 0.27$m for the mean relative camera distance for SET$^{(1)}$ and $\bar{e} = 1.87$m for SET$^{(3)}$. The results for the estimated camera height show acceptable values for SET$^{(1)}$ and SET$^{(2)}$ ($\bar{e} = 0.14$m and 0.10m), but large errors for SET$^{(3)}$ ($\bar{e} = 0.44$m). The mean relative camera distance error of 1.87m in SET$^{(3)}$ is still not acceptable.

The best results are achieved in experiment setup X_4 leading to error distances of 0.10m, 0.08m, and 0.09m for the three camera sets. We can conclude that X_4 provides the only usable solution for highly accurate camera calibration.

4.3 Convergence Evaluation

To evaluate the robustness of the optimization, the most challenging scenario SET$^{(3)}$ is examined. in Figure 6, we show mean and standard deviation of the camera height for the experiments X_3 and X_4.

The proposed method X_4 using the relative distances between the cameras (cf. Section 4.1 for details) provides a stable and accurate solution while the experiment X_3 comes to uncertain results. The mean of the camera heights lead to an error of $\bar{e} = 0.44$m (cf. Table 6). The proposed approach using the experimental setup X_4 has a mean error of $\bar{e} = 0.09$m.

5 CONCLUSIONS

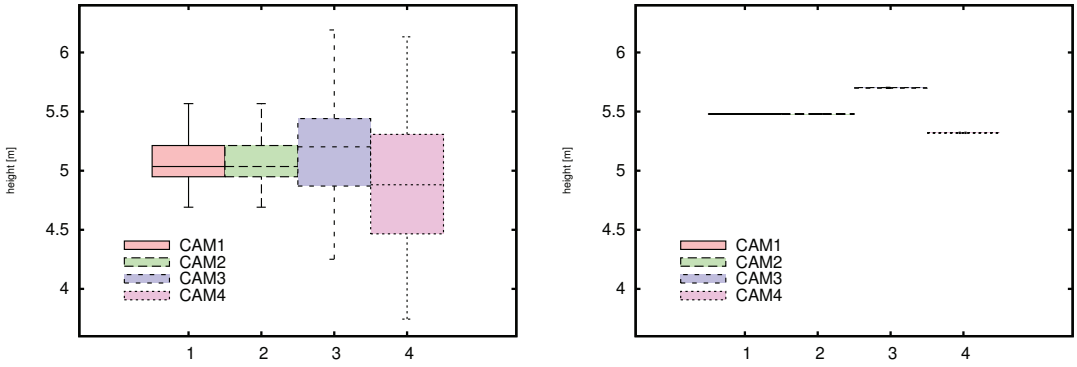
For the targeted application of vehicle observation during a lane merge, a constrained multi camera calibration technique is designed. For the application, high localization accuracy is required.

The proposed approach incorporates additional constraints such as the relative distances between cameras in the optimization. The distance measurements can be done very easily during the camera installation. This makes the presented approach a very useful part of the camera installation for the observation task.

The optimization procedure minimizes the reprojection error of known landmark positions on the road. The global optimization is based on evolutionary computation and provides suitable convergence behaviour on all test sets. For the accuracy evaluation, the resulting camera heights are compared. The proposed approach provides a mean error in the camera height below 10cm for cameras installed at a height of 5.4-7.6m observing landmarks with a distance of up to 100m.

ACKNOWLEDGEMENTS

This work has been performed in the framework of the H2020 5GCAR project co-funded by the EU. The authors would like to acknowledge the contributions of their colleagues from 5GCAR.



(a) Experiment X_3: camera heights of SET$^{(3)}$ (b) Experiment X_4: camera heights of SET$^{(3)}$

Figure 6: Mean and standard deviation for 100 evaluations of the camera height for experiments X_3 (a) and X_4 (b) (cf. Section 4.1). Here, the most challenging SET$^{(3)}$ is shown. For X_4, much smaller standard deviations are achieved.

REFERENCES

Brahmi, N., Frye, T., Martiñan, D., Dafonte, P., Abad, X., Sáez, M., Cellarius, B., Gangakhedkar, S., Cao, H., Sequeira, L., Berger, K., Theillaud, R., Otterbach, J., Grudnitsky, A., Villeforceix, B., Allio, S., Lefebvre, M., Odinot, J.-M., Tiphene, J., Servel, A., Fernández Barciela, A. E., Broszio, H., Cordes, K., and Abbas, T. (2018). 5GCAR Demonstration guidelines. https://5gcar.eu/. [Online; accessed 20-October-2018].

Cordes, K., Hockner, M., Ackermann, H., Rosenhahn, B., and Ostermann, J. (2015). WM-SBA: Weighted multi-body sparse bundle adjustment. In *IAPR International Conference on Machine Vision Applications (MVA)*, pages 162–165.

Cordes, K., Mikulastik, P., Vais, A., and Ostermann, J. (2009). Extrinsic calibration of a stereo camera system using a 3D CAD model considering the uncertainties of estimated feature points. In *European Conference on Visual Media Production (CVMP)*, pages 135–143.

Das, S., Konar, A., and Chakraborty, U. K. (2005). Two improved differential evolution schemes for faster global search. In *Annual Conference on Genetic and Evolutionary Computation*, GECCO 05, pages 991–998. ACM.

Fallgren, M., Dillinger, M., Li, Z., Vivier, G., Abbas, T., Alonso-Zarate, J., Mahmoodi, T., Alli, S., Svensson, T., and Fodor, G. (2018). On selected V2X technology components and enablers from the 5GCAR project. In *IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, pages 1–5.

Fernández Barciela, A. E., Servel, A., Tiphene, J., Fallgren, M., Sun, W., Brahmi, N., Ström, E., Svensson, T., Bernardez, D., Zarate, J. A., Kousaridas, A., Boban, M., Dillinger, M., Condoluci, M., Mahmoodi, T., Li, Z., Otterbach, J., Lefebvre, M., Vivier, G., Abbas, T., and Wingård, P. (2017). 5GCAR scenarios, use cases, requirements and KPIs. https://5gcar.eu/. [Online; accessed 20-October-2018].

Hartley, R. I. and Zisserman, A. (2003). *Multiple View Geometry*. Cambridge University Press, second edition.

Leal-Taixé, L., Milan, A., Schindler, K., Cremers, D., Reid, I. D., and Roth, S. (2017). Tracking the trackers: An analysis of the state of the art in multiple object tracking. *arXiv:1704.02781*.

Luo, Y., Xiang, Y., Cao, K., and Li, K. (2016). A dynamic automated lane change maneuver based on vehicle-to-vehicle communication. *Transportation Research Part C: Emerging Technologies*, 62:87–102.

Milan, A., Leal-Taixé, L., Reid, I. D., Roth, S., and Schindler, K. (2016). MOT16: A benchmark for multi-object tracking. *arXiv:1603.00831*.

Price, K. V., Storn, R., and Lampinen, J. A. (2005). *Differential Evolution - A Practical Approach to Global Optimization*. Natural Computing Series. Springer, Berlin, Germany.

Tang, Z., Wang, G., Liu, T., Lee, Y., Jahn, A., Liu, X., He, X., and Hwang, J. (2017). Multiple-kernel based vehicle tracking using 3D deformable model and camera self-calibration. *arXiv:1708.06831*.

Tang, Z., Wang, G., Xiao, H., Zheng, A., and Hwang, J.-N. (2018). Single-camera and inter-camera vehicle tracking and 3D speed estimation based on fusion of visual and semantic features. In *IEEE Conference on Computer Vision and Pattern Recognition Workshop (CVPRw)*, pages 108–115.

Thormählen, T., Broszio, H., and Wassermann, I. (2003). Robust line-based calibration of lens distortion from a single view. *Proceedings of Mirage*, pages 105–112.

Tsai, R. (1987). A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf tv cameras and lenses. *IEEE Journal on Robotics and Automation*, 3(4):323–344.

Wen, L., Du, D., Cai, Z., Lei, Z., Chang, M., Qi, H., Lim, J., Yang, M., and Lyu, S. (2015). UA-DETRAC: A new benchmark and protocol for multi-object detection and tracking. *arXiv:1511.04136*.