2018 IEEE Intelligent Vehicles Symposium (IV)
Changshu, Suzhou, China, June 26-30, 2018

Continuous Stereo Self-Calibration on Planar Roads

Georg R. Mueller, Patrick Burger and Hans-Joachim Wuensche*

Abstract— **This paper presents an algorithm for continuous online estimation of the twelve degrees-of-freedom (12-DoF) extrinsic calibration of a stereo camera system for an autonomous car. An Extended Kalman Filter (EKF) recursively estimates the stereo camera calibration by tracking salient points in 3D space that are visible in both cameras. All extrinsic parameters of the stereo camera calibration are only observable under translation and two independent rotations of the vehicle. However, when driving on urban, paved roads the vehicle performs only limited pitch or roll movement. An analysis of the Fisher information matrix reveals that in these situations especially the installation height is difficult to estimate. For these scenarios a further measurement model is added to the algorithm that utilizes the homography of salient points. This homography is induced by the planar road surface in consecutive camera images. Recognition of the road surface, when the position and orientation of the cameras is unknown, can be difficult. Therefore, a convolutional neural network (CNN) is developed that segments camera images pixel-wisely into three classes: 'road', 'static (other)' and 'potentially dynamic'. Only salient points that are segmented as 'road' are considered for the homography measurement model. Points that are segmented as 'potentially dynamic' are not taken into account for the calibration algorithm. The structure of the CNN has been chosen carefully to enable segmentation of camera images on a mid-range GPU on-board of our autonomous vehicle. An evaluation of the extended algorithm, based on a recorded dataset, shows a considerably faster estimation of the installation height.**

I. Introduction

In recent years, more and more cars have been equipped with stereo cameras to enable advanced driver assistance systems (ADAS). Once calibrated, they can provide a colored, high resolution, 3D reconstruction of the scene observed by both sensors.

Stereo cameras are usually calibrated offline [1], [2] to determine intrinsic (camera internal) and extrinsic parameters, that describe the six degrees-of-freedom (6-DoF) relative pose between the sensors and the 6-DoF relative pose to a common robot coordinate system (CoS). While the intrinsic calibration for cameras with fixed focal lengths usually does not change over time, especially the extrinsic parameters can be affected by drift. This drift can be caused by mechanical stress or temperature variations and reduces the quality of the 3D reconstruction.

To ensure a maintance-free operation, various algorithms for stereo online calibration have been developed. These algorithms can determine a valid calibration during operation of the robot. Many algorithms [3], [4] rely on epipolar geometry from which the relative pose between two sensors of a stereo system can be determined up to scale. Knorr et al. [5] exploit the homography constraint induced by a ground plane to calculate the relative pose between multiple cameras.

Especially for stereo vision the relative pose between the sensors of a stereo system has to be determined with very high accuracy to enable a valid rectification of the stereo images. This relative pose can be also determined up to scale using bundle adjustment from a sequence of images. This algorithm calculates 3D positions of salient points that are detected in multiple camera images. In an optimization these 3D positions and the relative pose between the camera images are refined to match the measured projections on the sensor. Carrera et al. [6] showed that with bundle adjustment and visual SLAM the relative pose between multiple sensors on a mobile robot can be determined precisely up to scale. An extension by Heng et al. [7] recovers this scale from wheel-odometry data and in addition a 5-DoF extrinsic calibration from the vehicle reference coordinate system to one of the cameras of their multi-camera system. Due to planar motion of the vehicle their algorithm cannot determine the height of the camera setup which is thus measured by hand. A disadvantage of algorithms that use bundle adjustment is the high computational complexity. Numerous 3D point positions and camera poses over many frames need to be optimized.

In contrast to these approaches, a continuous stereo camera calibration algorithm has been introduced by Dang et al. [8]. They applied an Iterated Extended Kalman Filter (IEKF) to recursively estimate the relative pose between the sensors up to scale and the motion of the vehicle. However, this algorithm does not determine the 6-DoF transformation from a robot CoS to the CoS of the stereo camera.

In [9] we have introduced an algorithm that continuously estimates the complete 12-DoF extrinsic calibration of a stereo system. An Unscented Kalman Filter (UKF) recursively determines the calibration and the position of salient points that are visible in both camera images in 3D. It is assumed that all points are static and therefore can be predicted with the measured inverse vehicle's motion. However, this expectation restricts the use cases of this algorithm. Thus, in [10] we extended this algorithm by a CNN to segment camera images into static and potentially dynamic pixels. Only pixels segmented as static are considered for calibration. This enables the algorithm to estimate the stereo camera calibration even in dynamic, urban scenarios.

However, when driving on urban roads, the installation height of the stereo camera cannot be estimated within

*All authors are with the Institute for Autonomous Systems Technology (TAS), Bundeswehr University Munich, 85577 Neubiberg, Germany, Contact author email: {georg.mueller,patrick.burger,jw}@unibw.de

short time. In this paper, we extend our previous algorithm with a homography measurement model that enables a fast and robust estimation of the camera installation height. A similar homography based method has been introduced by [11]. However, their algorithm estimates only the 3-DoF orientation of a monocular camera on a vehicle along with the vehicle's velocity by observing salient points of the road surface in consecutive frames.

Robust detection of salient points that solely belong to the road surface is critical for the homography measurement. For this we introduce a CNN to segment camera images to robustly detect pixels on the road surface.

This paper is structured as follows: In Section II we provide a brief introduction of our algorithm for continuous stereo camera calibration. In the following Section III the homography measurement model is introduced. The developed CNN for segmenting camera images is described in Section IV. In Section V the performance of this extended algorithm is compared to our previous algorithm from [10]. Finally in Section VI we draw a conclusion and provide insight into further work.

II. Recursive Stereo Calibration Algorithm

In [10] we have introduced an algorithm for continuous stereo camera calibration in urban scenarios. An EKF is applied to recursively estimate the 12-DoF extrinsic calibration of a stereo camera system by tracking salient points in 3D that are visible in both cameras. The EKF's state vector contains six states that describe the relative transformation from a vehicle CoS $\mathbf{F}_{\text{veh}}$, located at the vehicle's center of mass, to the left camera of the stereo system $\mathbf{F}_{\text{cam}_\text{l}}$

$$\mathbf{x}_{\text{cam}_\text{l}} = [x_{\text{cam}_\text{l}}, y_{\text{cam}_\text{l}}, z_{\text{cam}_\text{l}}, \Phi_{\text{cam}_\text{l}}, \Theta_{\text{cam}_\text{l}}, \Psi_{\text{cam}_\text{l}}]^\mathsf{T} \tag{1}$$

and six states that represent the transformation from the CoS of the left camera to the CoS of the right camera $\mathbf{F}_{\text{cam}_\text{r}}$

$$\mathbf{x}_{\text{cam}_\text{r}} = [x_{\text{cam}_\text{r}}, y_{\text{cam}_\text{r}}, z_{\text{cam}_\text{r}}, \Phi_{\text{cam}_\text{r}}, \Theta_{\text{cam}_\text{r}}, \Psi_{\text{cam}_\text{r}}]^\mathsf{T}. \tag{2}$$

For each tracked point the filter's state vector is extended by

$$\mathbf{x}_{\text{p}_i} = \left[x_{\text{p}_i}, y_{\text{p}_i}, z_{\text{p}_i}\right]^\mathsf{T}, \tag{3}$$

which describes the 3D position of a point in the CoS of the left camera. Thus, the filter's state vector size depends on the number of tracked points:

$$\mathbf{x} = \left[\mathbf{x}_{\text{cam}_\text{l}}^\mathsf{T}, \mathbf{x}_{\text{cam}_\text{r}}^\mathsf{T}, \mathbf{x}_{\text{p}_1}^\mathsf{T}, \ldots, \mathbf{x}_{\text{p}_N}^\mathsf{T}\right]^\mathsf{T} \tag{4}$$

The EKF continuously estimates the 12-DoF extrinsic calibration by tracking the 3D position of each observed point while the vehicle moves.

The Kalman criterion can be used to analyze the observability of all filter states of the EKF, which is required to enable the filter to estimate the states over time. It shows full observability can be achieved if the vehicle moves forward between three time steps and is rotated around two independent axes. In addition at least two points have to be tracked in 3D by the EKF at all times.

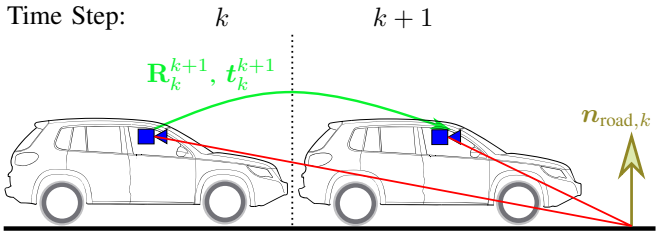


Fig. 1: The homography induced by the planar road is used to estimate the installation height z_{cam_l} and pitch angle $\Theta_{\text{cam}_\text{l}}$ of the stereo system.

This algorithm has proven to be capable of determining an accurate camera calibration. While high yaw rates can be measured particularly when cornering, the vehicle's pitching movements occur only to a limited extent, especially in urban environments.

The information content of a measurement for the system states of the EKF can be determined from the Fisher information matrix $\mathbf{J}$ which can be computed according to [12] as

$$\mathbf{J} = (\mathbf{\Phi}_{k-1}^{-1})^\mathsf{T}\mathbf{J}_{k-1}\mathbf{\Phi}_{k-1}^{-1} + \underbrace{\mathbf{C}_k^\mathsf{T}\mathbf{R}_k^{-1}\mathbf{C}_k}_{\mathbf{M}}. \tag{5}$$

The last term $\mathbf{M}$ describes the information of a new measurement. $\mathbf{M}$ is calculated from the EKF's state transition matrix $\mathbf{\Phi}_{k-1}$, the EKF's measurement jacobian matrix $\mathbf{C}_k$ and the measurement noise covariance matrix $\mathbf{R}_k$ (see [9]). The information content of the measurements for the filter's system states can be analyzed from the eigenvalues and corresponding eigenvectors of the matrix $\mathbf{M}$.

An investigation of $\mathbf{M}$ under realistic vehicle movements shows that only limited information about the translational degrees of freedom y_{cam_l} and z_{cam_l} can be obtained from the measurements. In addition, if there is no pitch or roll movement of the vehicle, the measurements do not provide any information about the installation height z_{cam_l} of the stereo system. Thus, the EKF can hardly estimate this parameter. This results in a slow convergence in comparison to the other states, which was also shown in [10].

III. Homography-Aided Camera Height Estimation

To ensure the installation height can be determined reliably in any driving situation, an additional measurement model is introduced. In the absence of pitch and roll motion it can be assumed that the vehicle drives on a planar surface as shown in Fig. 1. A point $\boldsymbol{p}_k$ in the left camera image at time step k belonging to this plane can be predicted into the next camera image using the homography matrix $\mathcal{H}$:

$$\underline{\boldsymbol{p}}_{k+1} = \mathcal{H}_k^{k+1}\underline{\boldsymbol{p}}_k \tag{6}$$

For this prediction, the point is transformed into homogeneous coordinates $\underline{\boldsymbol{p}}_k$. The homography matrix can be

calculated with the known intrinsic calibration $\mathbf{K}$ of the left camera to

$$\mathcal{H}_k^{k+1} = \mathbf{K}\left(\mathbf{R}_k^{k+1} - \boldsymbol{t}_k^{k+1}\boldsymbol{n}_k^{\mathsf{T}}\right)\frac{1}{d_k}\ \mathbf{K}^{-1}. \tag{7}$$

In equations (6) and (7), the index for the left camera's coordinate system has been omitted for better readability. The matrix $\mathbf{R}_k^{k+1}$ describes the rotation and the vector $\boldsymbol{t}_k^{k+1}$ the translation of the camera from time step k to $k+1$.

These can be taken from the homogeneous transformation matrix (HTM) $\mathbf{H}_{\text{cam}_\text{l},k}^{\text{cam}_\text{l},k+1}$, which is derived from the measured vehicle movement $\mathbf{H}_{\text{veh},k}^{\text{veh},k+1}$ and the filter's currently estimated camera calibration, which can also be expressed by $\mathbf{H}_{\text{veh},k+1}^{\text{cam}_\text{l},k+1}$ and $\mathbf{H}_{\text{cam}_\text{l},k}^{\text{veh},k}$:

$$\mathbf{H}_{\text{cam}_\text{l},k}^{\text{cam}_\text{l},k+1} = \mathbf{H}_{\text{veh},k+1}^{\text{cam}_\text{l},k+1}\mathbf{H}_{\text{veh},k}^{\text{veh},k+1}\mathbf{H}_{\text{cam}_\text{l},k}^{\text{veh},k}. \tag{8}$$

The vector $\boldsymbol{n}_k$ is the normal vector of the plane in the left camera coordinate system. This can be obtained from the normal vector of the plane $\boldsymbol{n}_{\text{road},k} = (0,0,1)^{\mathsf{T}}$ which can be transformed using homogeneous coordinates with

$$\underline{\boldsymbol{n}}_{\text{cam}_\text{l},k} = \mathbf{H}_{\text{road},k}^{\text{cam}_\text{l},k}\ \underline{\boldsymbol{n}}_{\text{road},k} = \mathbf{H}_{\text{veh},k}^{\text{cam}_\text{l},k}\mathbf{H}_{\text{road},k}^{\text{veh},k}\ \underline{\boldsymbol{n}}_{\text{road},k} \tag{9}$$

into the coordinate system of the left camera. The transformation $\mathbf{H}_{\text{road},k}^{\text{veh},k}$ from the plane to the vehicle coordinate system is determined from CAD data. The height and the pitch angle from the $\mathbf{F}_{\text{veh}}$ to the $\mathbf{F}_{\text{road}}$ is calculated from vehicle damper measurements or set constant if not available. It is assumed that the origin of $\mathbf{F}_{\text{road}}$ is located directly below the origin of $\mathbf{F}_{\text{veh}}$. From $\mathbf{H}_{\text{road},k}^{\text{cam}_\text{l},k}$ the scalar distance d_k from the plane to the camera can be determined as well.

This homography based measurement model is used exclusively to determine the installation height z_{cam_l} and the pitch angle of the stereo system $\Theta_{\text{cam}_\text{r}}$. In the filter correction step, it is therefore added to support the estimation of these states only. The remaining EKF states of the stereo camera calibration remain unaffected by this measurement model, as they can already be determined satisfactory with the existing method.

Salient points visible in the left camera image at time step k and $k+1$ serve as measurements for the homography measurement model as shown in Fig. 1. For a valid homography measurement, it is crucial that these points are located on the road surface on which the vehicle is driving on.

IV. Segmentation of camera images

In [10] we have presented an algorithm for continuous stereo camera calibration in urban scenarios, which utilizes a CNN to segment RGB camera images on a pixel level. This output is further converted into a binary image to distinguish between static and potentially dynamic pixels. Only pixels labeled as static are used for the calibration algorithm then. This is crucial for calibration in an urban scenario where many salient points are non-static, as tracking dynamic points distorts the calibration.

The CNN used for segmentation is based on a FCN-8s [13] and was trained on the Cityscapes Dataset [14] to segment RGB camera images into the 19 classes provided in the dataset. From the segmented image, only salient points labeled as 'road' are selected for the EKF update using the homography measurement model.

CNN Layer	Stride	Padding	Kernel Size	Output Size
conv1	4	100	11	96
pool1	2	0	3	
conv2	1	2	5	256
pool2	2	0	3	
conv3	1	1	3	384
conv4	1	1	3	384
conv5	1	1	3	256
pool5	2	0	3	
conv6	1	0	6	1024
conv7	1	0	1	1024
upscore	32	0	63	3

TABLE I: Structure of the TAS-tinyNet.

In order to segment a 1468 pix × 496 pix RGB image with the 10 Hz frame rate of the stereo camera system a high-end GPU is required. However, for mobile applications on our autonomous vehicle platform MuCAR-4 (Munich Cognitive Autonomous Robot Car 4th Generation) only a low-power, mid-range GPU is available. Since the segmented image is only used for masking, it makes sense to design a CNN that allows direct segmentation into potentially dynamic pixels (e.g. car, person, etc.), road pixels and all other static pixels (e.g. sidewalk, building, etc.). This reduces the size of the network structure and the network's inference time significantly.

Derived from the BVLC AlexNet, which is a modified variant of the well-known AlexNet [15], a simplified CNN 'TAS-tinyNet' was designed as shown in Table I. In contrast to the original AlexNet, the fully connected layers fc6 and fc7 have been exchanged for the purpose of segmentation to convolutional layers conv6 and conv7 with a reduced output size.

This network has been trained using the Cityscapes Dataset. Since the layers at the beginning are identical to the layers of the AlexNet, the trained weights of the AlexNet can be transferred to the TAS-tinyNet for initialization. However, the training of this network failed and resulted in a very noisy segmented image. To overcome this problem, the network was first trained on the Cityscapes Dataset to segment RGB images into 19 classes as given in the dataset. For this the output size of the upscore layer was set to 19. In a further step we used this pretrained network to segment into 3 classes – 'road', 'static (other)' and 'potentially dynamic'. To do this, we reduced the output size of the upscore layer to 3 and grouped the labels of the Cityscapes Dataset for training. The performance of the trained network is analyzed on the validation data provided in the Cityscapes Dataset. Table II shows the evaluation which is based on the standard Jaccard index. This is also known as the intersection-over-union metric (IoU).

A comparison of our trained FCN-8s 19-class output from [10] and the TAS-tinyNet 3-class output is visualized

Class	Per Class Acc.	Per Class IoU
Road	0.979	0.952
Pot. dynamic	0.896	0.818
Static (other)	0.961	0.927

TABLE II: Evaluation of per-class IoU of the TAS-tinyNet on the validation data of the Cityscapes Dataset.

in Fig. 2. The FCN-8s shows a more detailed segmented image in contrast to the TAS-tinyNet, but requires 2685 MiB and 470 ms on a mid-range Nvidia GeForce 1050Ti GPU for inference. As the stereo camera captures images at 10 Hz, it is not possible to calculate a segmented image with the FCN-8s for every new image, so that synchronization is lost. Although the segmentation of the TAS-tinyNet is not as detailed, the quality of the segmented output image is still sufficient for the purpose of calibration. In addition, the TAS-tinyNet requires only 431 MiB and can process the same RGB-image within 55 ms, so the algorithm can be applied to our autonomous vehicle MuCAR-4.

For reliable calibration, it is important that no salient points belonging to dynamic objects are tracked. To ensure this in a robust way, the area of pixels marked as potentially dynamic is expanded using dilation. By this more pixels are excluded from the calibration process. In addition, only points on the road surface must be used for the homography measurement model. For this purpose the region of pixels that are segmented as 'road' is reduced by an erosion operation, to avoid false segmented pixels on the transition between classes in the image. The resulting gaps are finally filled as 'static (other)'.

V. Evaluation

The algorithm has been tested in various scenarios in Munich on our autonomous platform MuCAR-4. MuCAR-4 is based on a Volkswagen Tiguan and perceives the environment with a color stereo system mounted on a rotateable platform behind the windshield. An image showing the platform can be found in [9]. When the platform is rotated, the relative pose of the cameras remains unaffected. A high-end inertial navigation system (INS) and odometry sensors of our autonomous car MuCAR-4 provide precise measurements of the vehicle's motion.

The extended algorithm presented in this paper, is compared to the algorithm from [10] based on a dataset recorded in Munich. The driven trajectory is of 3.1 km length, as visualized in Fig. 3. All roads on this path are paved, so there is only limited pitch and roll motion of the vehicle. During recording, the view of the planar road was occasionally hidden by cars, cyclists or pedestrians.

Both algorithms are initialized with the same initial falsified calibration estimate as shown in Table III. In Fig. 4 the convergence of both algorithms is shown. Red visualizes an offline determined calibration. Cyan shows the calibration estimate of the homography-aided estimator, with the corresponding 3σ standard deviation in light blue. In black

a) RGB Input

b) FCN-8s

c) TAS-tinyNet

Fig. 2: Segmentation of a RGB camera image. a) RGB input image. b) FCN-8s segmentation. c) TAS-tinyNet segmentation.

Fig. 3: The path driven for evaluation of the calibration algorithms is shown in red. At the position marked in yellow the camera platform was rotated around the yaw axis by 2°.

Parameter	Initial Error	Parameter	Initial Error
x_{cam_l}	0.3 m	x_{cam_r}	–0.03 m
y_{cam_l}	0.3 m	y_{cam_r}	–0.03 m
z_{cam_l}	–0.3 m	z_{cam_r}	–0.03 m
$\Phi_{\text{cam}_\text{l}}$	3°	$\Phi_{\text{cam}_\text{r}}$	0.1°
$\Theta_{\text{cam}_\text{l}}$	3°	$\Theta_{\text{cam}_\text{r}}$	–0.1°
$\Psi_{\text{cam}_\text{l}}$	3°	$\Psi_{\text{cam}_\text{r}}$	0.1°

TABLE III: Initial errors in the stereo camera calibration estimate.

the estimator from [10] is displayed. The corresponding 3σ standard deviation is shown in gray.

It can be seen, that the algorithm with the additional homography measurement estimates the installation height z_{cam_l} of the stereo camera system significantly faster. In addition, the reduced variance in the installation height leads to a faster convergence of the parameter y_{cam_l}. Since the homography is only determined in the left camera image, the estimation of the relative pose $\mathbf{x}_{\text{cam}_\text{r}}$ between the cameras is not affected and can still be estimated robustly as shown in Fig. 4b.

Between the periods $t_{1,\text{stop}} = 371\,\text{s}$ to $t_{1,\text{start}} = 405\,\text{s}$ and $t_{2,\text{stop}} = 468\,\text{s}$ to $t_{2,\text{start}} = 517\,\text{s}$ the vehicle had to stop due to red traffic lights. These time points are marked in Fig. 4 with vertical gray lines. Only a limited amount of states can be observed while standing still. Thus, there is no filter update performed during these periods and the system states remain constant. Since the filter only predicts the system states, the state covariances slightly increase due to a minimal Gaussian process noise, which is assumed for the system states while standing still. This can be seen especially in the parameters $\Theta_{\text{cam}_\text{l}}$ and $\Psi_{\text{cam}_\text{l}}$.

At time $t_{3,\text{rot}} = 533.3\,\text{s}$ the stereo platform is rotated by 2°. This is visualized in Fig. 4 by a vertical yellow line. The purpose of this change is to show how the filters react to a modification in the calibration. Therefore, this change is not provided as input to the filters. It can be seen in Fig. 4a that both algorithms can detect this change in $\Psi_{\text{cam}_\text{l}}$ and converge within 7 s.

VI. Conclusion

Nowadays, autonomous cars are often equipped with stereo cameras. In [10] we have presented an algorithm that continuously estimates the complete 12-DoF extrinsic stereo camera calibration online. However, especially when driving on urban, paved roads not every parameter is observable. Analysis of the Fisher information matrix reveals that the installation height is hard to observe without significant pitch or roll motion

For these situations, the algorithm is extended with a homography measurement model. It is assumed that the vehicle drives on a planar surface, which is a viable assumption while driving on paved roads. An evaluation of the extended algorithm on a recorded dataset reveals that the installation height of the stereo camera system can be estimated in much shorter time and with significantly reduced variance.

For the homography measurement, salient points on the road surface ahead of the vehicle are tracked between consecutive frames. However, the recognition of the road surface can be very difficult, especially during calibration when the orientation of the camera is unknown. In addition the view of the road can be occluded by other traffic participants, e.g. cars driving ahead. For this purpose a CNN is designed and trained to segment RGB camera images on a pixel level into three classes: 'road', 'potentially dynamic' and 'static (other)'. From the CNN segmented output, only salient points that are segmented as road are considered for the homography measurement. Salient points that are labeled as 'potentially dynamic' are excluded from the calibration process, since tracking these points falsifies the calibration estimation.

The structure of the introduced CNN has been chosen carefully to enable segmentation of RGB images within 10 Hz on a mid-range GPU which is integrated in our autonomous platform MuCAR-4. This is necessary to keep synchronized with the frame-rate of the stereo camera system.

For motion estimation of the vehicle, currently a high-end INS is used. In the future we want to enhance the algorithm to enable calibration of the stereo system based on wheel odometry only. In addition we want to train CNNs also for challenging weather conditions (e.g. snow, heavy rain) which were currently not provided in the training dataset.

References

[1] Z. Zhang, "A Flexible New Technique for Camera Calibration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 11, pp. 1330–1334, 2000.

[2] L. E. Kruger, C. Wohler, A. Wurz-Wessel, and F. Stein, "In-Factory Calibration of Multiocular Camera Systems," in *Photonics Europe*. International Society for Optics and Photonics, 2004, pp. 126–137.

[3] H. Longuet-Higgins, "A Computer Algorithm for Reconstructing a Scene From Two Projections," *Nature*, vol. 293, pp. 133–135, 1981.

[4] P. Hansen, H. Alismail, P. Rander, and B. Browning, "Online Continuous Stereo Extrinsic Parameter Estimation," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognition (CVPR)*, Providence, RI, USA, 2012, pp. 1059–1066.

[5] M. Knorr, W. Niehsen, and C. Stiller, "Online Extrinsic Multi-Camera Calibration Using Ground Plane Induced Homographies," in *Proc. IEEE Intelligent Vehicles Symp. (IV)*. Gold Coast City, Australia: IEEE, 2013, pp. 236–241.

[6] G. Carrera, A. Angeli, and A. J. Davison, "SLAM-Based Automatic Extrinsic Calibration of a Multi-Camera Rig," in *Proc. IEEE Int. Conf. Robotics and Automation (ICRA)*, Shanghai, China, 2011, pp. 2652–2659.

[7] L. Heng, B. Li, and M. Pollefeys, "CamOdoCal: Automatic Intrinsic and Extrinsic Calibration of a Rig with Multiple Generic Cameras and Odometry," in *Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Syst. (IROS)*, Tokyo, Japan, 2013, pp. 1793–1800.

[8] T. Dang, C. Hoffmann, and C. Stiller, "Continuous Stereo Self-Calibration by Camera Parameter Tracking," *IEEE Trans. Image Process.*, vol. 18, no. 7, pp. 1536–1550, 2009.

[9] G. R. Mueller and H.-J. Wuensche, "Continuous Extrinsic Online Calibration for Stereo Cameras," in *Proc. IEEE Intelligent Vehicles Symp. (IV)*, Gothenburg, Sweden, Jun. 2016, pp. 966–971.

[10] ——, "Continuous Stereo Camera Calibration in Urban Scenarios," in *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*, Yokohama, Japan, Oct. 2017, pp. 2298–2303.

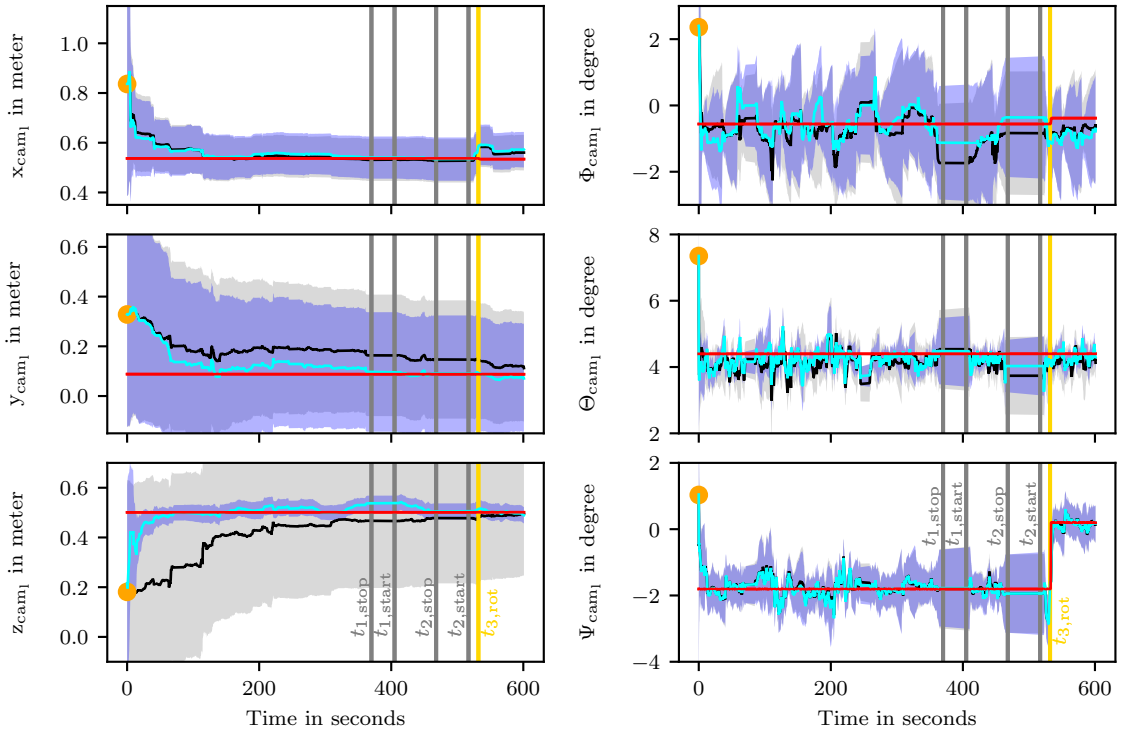
[11] M. Miksch, B. Yang, and K. Zimmermann, "Automatic Extrinsic Camera Self-Calibration Based on Homography and Epipolar Geometry," in *Proc. IEEE Intelligent Vehicles Symp. (IV)*, San Diego, CA, USA, Jun. 2010, pp. 832–839.

[12] J. H. Taylor, "The Cramer-Rao Estimation Error Lower Bound Computation for Deterministic Nonlinear Systems," in *1978 IEEE Conference on Decision and Control including the 17th Symposium on Adaptive Processes*, Jan 1978, pp. 1178–1181.

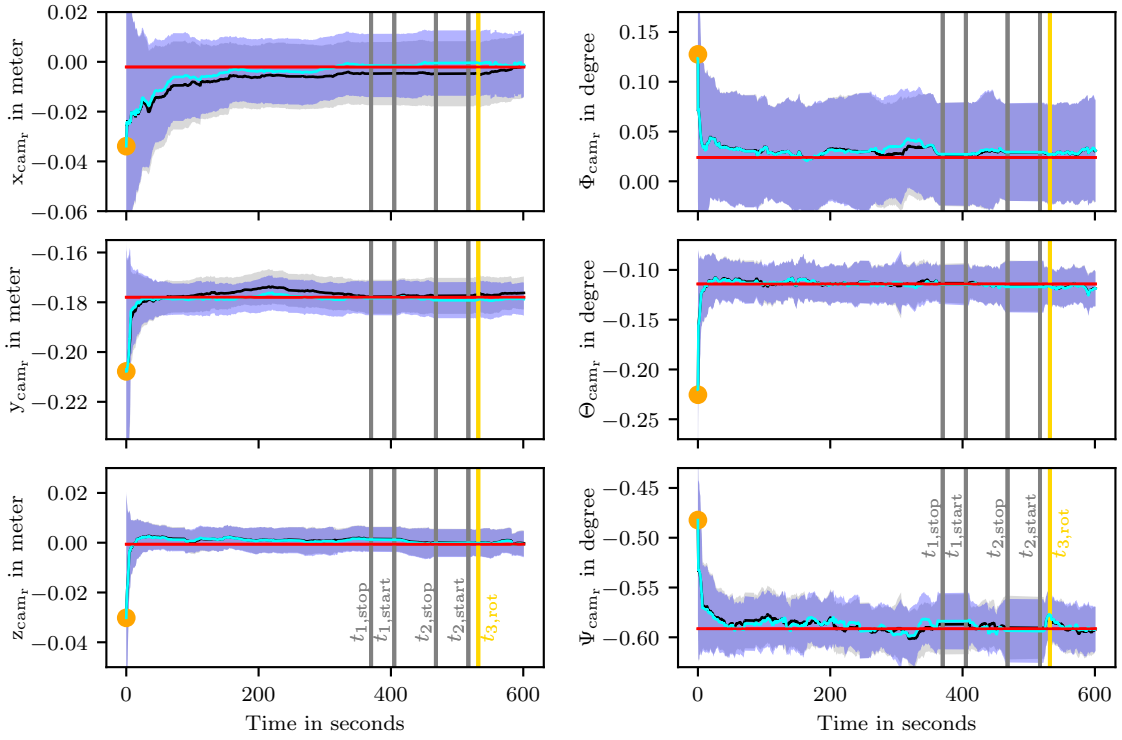
[13] J. Long, E. Shelhamer, and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognition (CVPR)*, Boston, MA, USA, Jun. 2015, pp. 3431–3440.

[14] M. Cordts, M. Omran *et al.*, "The Cityscapes Dataset for Semantic Urban Scene Understanding," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016, pp. 3213–3223.

[15] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in *Proc. of the 25th International Conference on Neural Information Processing Systems - Volume 1*, Lake Tahoe, NV, USA, 2012, pp. 1097–1105.



(a) Continuous estimation of the stereo camera extrinsic parameters $\mathbf{x}_{cam_l}$.



(b) Estimation of the relative pose between the stereo cameras $\mathbf{x}_{cam_r}$.

Fig. 4: Evaluation of the stereo online calibration algorithms with and without the homography measurement model. Both algorithms are initialized with a falsified guess marked in orange. The ground truth, determined offline, is shown in red. The original estimation from [10] without the homography measurement model is visualized in black and the corresponding 3σ error region in light gray. The improved calibration estimate supported by the homography measurement model is shown in cyan, the corresponding 3σ error region is visualized in light blue.