Knorr, Moritz ; Niehsen, Wolfgang ; Stiller, Christoph: Online extrinsic multi-camera calibration using ground plane induced homographies. In: IEEE Intelligent Vehicles Symposium. Gold Coast, Australia, June 2013, pp. 236–241

Online Extrinsic Multi-Camera Calibration Using Ground Plane Induced Homographies

Moritz Knorr[1], Wolfgang Niehsen[1], *Senior Member, IEEE*, and Christoph Stiller[2], *Senior Member, IEEE*

***Abstract*— This paper presents an approach for online estimation of the extrinsic calibration parameters of a multi-camera rig. Given a coarse initial estimate of the parameters, the relative poses between cameras are refined through recursive filtering. The approach is purely vision based and relies on plane induced homographies between successive frames. Overlapping fields of view are not required. Instead, the ground plane serves as a natural reference object. In contrast to other approaches, motion, relative camera poses, and the ground plane are estimated simultaneously using a single iterated extended Kalman filter. This reduces not only the number of parameters but also the computational complexity. Furthermore, an arbitrary number of cameras can be incorporated. Several experiments on synthetic as well as real data were conducted using a setup of four synchronized wide angle fisheye cameras, mounted on a moving platform. Results were obtained, using both, a planar and a general motion model with full six degrees of freedom. Additionally, the effects of uncertain intrinsic parameters and nonplanar ground were evaluated experimentally.**

I. INTRODUCTION AND RELATED WORK

Multi-camera systems are often preferred over single cameras as they capture in general more information and help to resolve ambiguities in the observed scene or motion. For egomotion estimation, as an example, Chen et al. [4] showed that significantly better estimates can be achieved with multiple cameras. The calibration of a multi-camera rig on the other hand is more complex and time consuming. While cameras can be calibrated intrinsically one at a time and beforehand, in order to determine the extrinsic calibration parameters, the cameras have to be attached to the rig. In order to reduce the effort of calibration and to be able to compensate for changes during runtime, online estimation of the extrinsic calibration parameters is desirable. In contrast to offline calibration methods, online calibration cannot rely on calibration patterns with known appearance and geometry and need to work in natural environments. Promising solutions based on different assumptions have been proposed in the recent past. For stereo applications with overlapping fields of view, Dang et al. [5] presented an approach which continuously estimates the intrinsic and extrinsic calibration parameters of an active stereo rig using an iterated extended Kalman filter with a robust innovation step. However, the requirement for overlapping fields of view is often not met. For extrinsic calibration, Esquivel et al. [6] proposed an approach which is also applicable in case of totally disjoint fields of view. Starting from the hand-eye calibration problem, a solution for the estimation of relative orientations is derived. Then, relative position and scale are estimated and thereafter refined using nonlinear optimization.

Carrera et al. [3] as well as Lebraly et al. [9] extend this work by using a specific bundle adjustment which considers the rigid coupling between cameras. In the approach of Carrera et al. [3], each camera uses SLAM (simultaneous localization and mapping) to determine a map of feature points separately. An initialization for bundle adjustment is then found through robust fusion of the individual maps. In order to ensure overlap between the maps, a set of programmed motions is performed. In contrast, Lebraly et al. [9] use a linear initialization based on the solution proposed by Esqivel et al. [6]. Both approaches use a minimal parameterization of the camera rig, and describe the motion only in the global coordinate system of a master or reference camera. The approach, presented in this paper adopts the minimal parameterization concept.

Recently, Pagel [12] presented an approach for relative pose estimation between two cameras with non-overlapping fields of view. A sparse bundle adjustment is used to estimate the motion trajectories and 3D point positions within a sliding time window. The trajectories as well as the corresponding scales are then refined using a bundle adjustment type of algorithm. However, the author states that the approach is not suitable for planar motions.

Planar motions represent a special case and cause some approaches to fail. In case of planar motion, there is only one rotation axis. Hence, the decoupled estimation of the relative orientation, such as in [6], based on rotations only, is not possible. The work of Ruland et al. [15] is dedicated to this case. They use the Ackermann steering model and given motion parameters to estimate the 2D position of a camera with respect to the vehicle coordinate frame. The orientation of the camera with respect to the ground plane is assumed to be known. This, however, is not restrictive as several methods for orientation estimation exist. As an example, Miksch et al. [11] use plane induced homographies in combination with epipolar geometry to estimate the camera orientation from linear motions.

Pagel and Willersinn [13] extended an earlier approach to the case of planar motion. Their approach is most closely related to the one presented in this paper, as it also relies on Kalman filtering and the ground plane for calibration.

[1]Moritz Knorr and Wolfgang Niehsen are with Robert Bosch GmbH, Corporate Research, Computer Vision Research Lab, D-31134 Hildesheim, Germany, {moritzmichael.knorr, wolfgang.niehsen}@de.bosch.com

[2]Christoph Stiller is with Institut für Mess- und Regelungstechnik, University of Karlsruhe, Geramny, stiller@kit.edu

However, it is by far more complex. Their concept is based on continuous parameter estimation, propagation and fusion. Several robust iterated extended Kalman filters, as presented in [5], are used to filter the motion of each camera, the ground plane and relative poses. The egomotion estimation itself is based on the epipolar and trifocal constraint in combination with the minimization of the projection error of previously triangulated points. Unfortunately, results were only obtained for a two camera setup and synthetic data.

Our main contribution is that we seek to estimate the parameters of the moving camera rig by observing corresponding features in successive frames, which are related by plane induced homographies, using a *single* iterated extended Kalman filter [1], thus reducing the overall number of state parameters and computational complexity. The motion of the camera rig, the relative poses of the cameras with respect to a global coordinate system that is affixed to a master camera, and the ground plane are estimated simultaneously. The filter is described in more detail in Section IV. The motion of the camera rig is described in the coordinate system of the master camera. The motion of the remaining cameras can be determined using the rigid coupling and estimated extrinsic calibration parameters.

Although using the ground plane is restrictive since it requires visibility, detection, and planarity in the considered vicinity, it establishes a reference object which can be captured by all cameras even in the case of disjoint fields of view. Furthermore, it allows for direct estimation of the relative motion scales and will also work in the special case of planar or almost planar motion. Kitt et al. [8], for example, use constraints, imposed by the ground plane, to prevent bundle adjustment from long term scale drift. From the related work, we could identify two different motion models, planar motion and unconstrained or 3D general motion. This paper presents comparative experimental studies on approaches based on both models. More details on the motion models are given in Section II. The setup for experimental evaluation consists of four synchronized wide angle fisheye cameras. In Section III it is shown how homographies can be combined with projection models other than the pinhole camera model. The projection model as well as the intrinsic calibration toolbox of Mei [10] were used. The cameras were calibrated intrinsically offline and beforehand.

In our experiments, corresponding features in consecutive frames are determined using the feature detector by Rosten and Drummond [14] in combination with the feature descriptor by Calonder et al. [2]. Corresponding features are then found by comparison of Hamming distances. Besides the comparative evaluation of both motion models on real and synthetic data, the effect of uncertain intrinsic parameters and the nonplanarities in the scene are determined experimentally. The evaluation can be found in Section V.

II. MOTION MODELS

A. GROUND PLANE INDUCED HOMOGRAPHY

The goal of this work is to determine the relative poses of $N-1$ cameras with respect to a master camera. For camera i this relative pose transformation $\Delta\mathbf{T}^i$ can be written as a homogeneous matrix [7]

$$\Delta\mathbf{T}^i = \begin{bmatrix} \Delta\mathbf{R}^i & \Delta\mathbf{C}^i \\ \mathbf{0}^T & 1 \end{bmatrix}, \tag{1}$$

where $\Delta\mathbf{R}^i \in \mathrm{SO}(3)$ is an orthonormal rotation matrix and $\Delta\mathbf{C}^i \in \mathbb{R}^{3\times 1}$ is a displacement vector. Hence, it involves the full six degrees of freedom of a rigid transformation in 3D. Fig. 1 depicts this relationship. The relative motion of the master camera m from frame $k-1$ to frame k is

$$\mathbf{T}_k^m = \begin{bmatrix} \mathbf{R}_k^m & \mathbf{C}_k^m \\ \mathbf{0}^T & 1 \end{bmatrix}. \tag{2}$$

According to [7] we can express the ground plane induced homography at frame k for the master camera as

$$\mathbf{H}_k^m = \mathbf{R}_k^m - \mathbf{C}_k^m (\mathbf{n}_{k-1}^m)^T / h_{k-1}^m, \tag{3}$$

where $\mathbf{n}_{k-1}^m$ is the normal of the plane in the preceding camera coordinate system (CCS), and h_{k-1}^m is the corresponding height above ground. The homography describes the mapping of any point of the normalized image plane at time $k-1$ onto the normalized image plane at time k

$$\mathbf{p}' \simeq \mathbf{H}_k^m \mathbf{p}. \tag{4}$$

Here "$\simeq$" denotes equality up to scale since we are using homogeneous point coordinates. For the ith camera, the motion between consecutive time steps can be determined using the relative pose and motion with respect to the master camera according to (1) and (2)

$$\mathbf{T}_k^i = (\Delta\mathbf{T}^i)^{-1} \mathbf{T}_k^m \Delta\mathbf{T}^i. \tag{5}$$

Fig. 2 depicts this relationship between master camera and camera i. The plane normal in the coordinate system of camera i is

$$\mathbf{n}_{k-1}^i = (\Delta\mathbf{R}^i)^T \mathbf{n}_{k-1}^m, \tag{6}$$

and the respective height is

$$h_{k-1}^i = h_{k-1}^m - (\Delta\mathbf{R}^i \mathbf{n}_{k-1}^i)^T \Delta\mathbf{C}^i. \tag{7}$$

The plane induced homography $\mathbf{H}_k^i$ can then be determined by replacing the respective terms in (3) by (5) to (7).

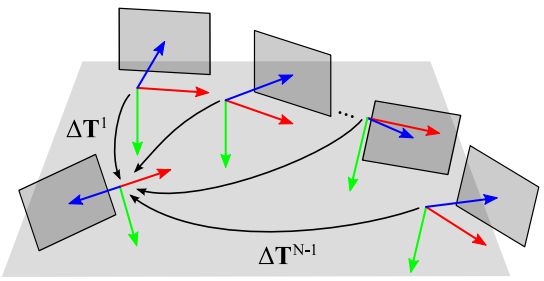

Fig. 1. Schematic representation of the multi-camera rig. Several camera coordinate systems are related to the coordinate system of the master camera via relative poses.

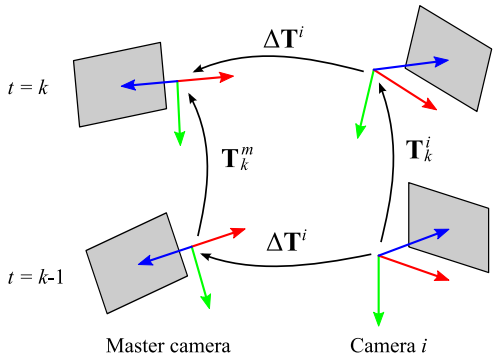

Fig. 2. Geometric relationship of the master camera and camera i for time instances $k-1$ and k.

B. PLANAR MOTION

In the previous section, the transformation between consecutive time steps was introduced for the general case with six degrees of freedom (2). In case of planar motion, the corresponding transformation has only three degrees of freedom, i.e. the rotation of the camera about the ground plane normal, the translation direction within the plane and the length of the translation vector. A change in height or a rotation about a different axis are not intended. The rotation about the ground plane normal is given by

$$\widetilde{\mathbf{R}}_k^m = \mathbf{R}(\mathbf{n}^m, \widetilde{\omega}_k^m), \tag{8}$$

where the tilde indicates the planar case and $\widetilde{\omega}_k^m$ denotes the angular velocity. The translation vector $\widetilde{\mathbf{C}}_k^m$ is expressed with respect to the projection $\mathbf{x}_\perp^m$ of the camera's x axis onto the ground plane

$$\widetilde{\mathbf{C}}_k^m = \mathbf{R}(\mathbf{n}^m, \widetilde{\alpha}_k^m)\ \frac{\mathbf{x}_\perp^m}{||\mathbf{x}_\perp^m||_2} ||\widetilde{\mathbf{C}}_k^m||_2. \tag{9}$$

Here $\widetilde{\alpha}_k^m$ denotes the angle between projected x axis and $\widetilde{\mathbf{C}}_k^m$. Fig. 3 depicts this relationship. Translation and rotation as defined by (8) and (9) can then be used to determine the camera motion (2) and the homography (3).

C. SYSTEM PARAMETERIZATION

To avoid ambiguities and inconsistencies, a minimal parameterization is chosen. All rotation matrices, except for (8), are described using angle-axis representation [7]. The relative motion in case of the planar motion model is expressed by the angular velocity $\widetilde{\omega}_k^m$, the angle $\widetilde{\alpha}_k^m$, and velocity $||\widetilde{\mathbf{C}}_k^m||_2$

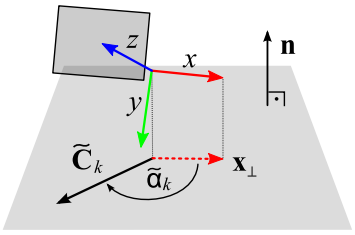

Fig. 3. Planar motion. The translation vector is expressed via the angle to the projected x axis of the camera and the length of the translation.

as described in the previous section. The ground plane is defined by the distance to the master camera and the normal using spherical coordinates. Unfortunately, we are only able to estimate the calibration parameters and motion up to scale. Hence, to avoid ambiguities, the height above ground is fixed in case of the planar motion model, and the distance of the master to an arbitrary other camera is fixed in the case of the general motion model. The fixed values are set to their corresponding known values, so that the estimation results and the ground truth can be compared metrically. The ground plane normal is not assumed to be known and will therefore be estimated in both cases. The motion of the master camera and the plane are then described using nine or five free parameters, depending on the model. The relative poses are described using six parameters. For the general 3D motion model, one of the relative camera poses has only five parameters due to the fixed distance. The total number of parameters, depending on the number of cameras N, is $9+5+6\cdot(N-2)$ for the general 3D, and $5+6\cdot(N-1)$ for the planar motion model.

III. CAMERA PROJECTION MODEL

The homography, as given in (3) describes the mapping between normalized image planes. For pinhole cameras, the transformation between the image plane and the normalized image plane is given by an intrinsic calibration matrix $\mathbf{K}$ [7]. Pre- and post-multiplying (3) with the calibration matrix and its inverse yields a new homography $\mathbf{KHK}^{-1}$ which directly relates image planes. However, for different camera models, the transformation becomes, in general, nonlinear. For the calibration of the wide angle fisheye cameras, as used in the experimental evaluation, the projection model and calibration toolbox of Mei [10] were used. The model describes the transformation between a world point $\mathbf{X}_W = (X_W, Y_W, Z_W)^T$ and an image point $\mathbf{p}$, as summarized in the following. First, the world point is projected onto the unit sphere

$$\mathbf{X}_S = \frac{\mathbf{X}_W}{||\mathbf{X}_W||_2}. \tag{10}$$

Then, a new reference frame is chosen, which allows projecting all world points to finite pixel positions. After normalization, radial and tangential distortions are corrected. The resulting point is then projected into the image plane using a general calibration matrix. The projection of a point on the unit sphere into the image plane is then given by

$$\mathbf{p} = \kappa(\mathbf{X}_S, \mathbf{k}), \tag{11}$$

where $\mathbf{k}$ comprises the intrinsic calibration parameters and $\kappa(\cdot)$ is the combination of projections described above. The inverse projection is given by $\mathbf{X}_S = \kappa^{-1}(\mathbf{p}, \mathbf{k})$. The predicted position of a feature in camera i at time k is then

$$\hat{\mathbf{p}}' = \kappa(\mathbf{H}_k^i \kappa^{-1}(\mathbf{p}, \mathbf{k}^i), \mathbf{k}^i), \tag{12}$$

where $\mathbf{H}_k^i$ is the ground plane induced homography (3) for camera i. The calibration toolbox [10] also provides an approximate covariance matrix of the intrinsic parameters which is obtained through forward propagation [7]. It was

used to evaluate the effect of uncertain intrinsic parameters on the calibration results (Section V).

IV. RECURSIVE FILTERING

We seek to estimate the parameters of the moving camera rig by observing corresponding features in successive frames in each view respectively. Kalman filters have proven to work well for similar parameter estimation problems [5]. The motion and ground plane parameters, as well as the relative pose parameters are associated with a single state vector of a dynamic system which evolves, corresponding to a discrete time stochastic system [1]

$$\mathbf{x}_k = \mathbf{f}(\mathbf{x}_{k-1}) + \mathbf{v}_{k-1} \tag{13}$$

with associated measurements

$$\mathbf{z}_k = \mathbf{m}(\mathbf{x}_k) + \mathbf{w}_k. \tag{14}$$

The process noise $\mathbf{v}_k$ and the measurement noise $\mathbf{w}_k$ are assumed to be zero mean, white, mutually uncorrelated, and additive [1]. The respective covariance matrices are $\mathbf{Q}_k$ and $\mathbf{R}_k$. The Kalman filter is initialized with state $\hat{\mathbf{x}}_0^+$, as described in Section II-C, and state covariance $\mathbf{P}_0^+$.

The measurements are provided by a feature matching algorithm. It determines the position $\mathbf{p}$ of a feature in one frame and the position $\overline{\mathbf{p}}'$ of the corresponding feature in the subsequent frame. Due to limited accuracy, the position in the subsequent frame will in general not correspond to the same physical point as in the preceding frame. Instead, only a perturbed position $\mathbf{p}'$ is provided as observation. The position prediction of a feature in camera i at time k, $\hat{\mathbf{p}}'$, is carried out by (12) in the previous section. The measured and predicted x and y positions of the features in the subsequent frame are stacked to vectors

$$\mathbf{z}_k^i = \left(p'_{1,x}, p'_{1,y}, p'_{2,x}, \ldots\right)_{i,k} \tag{15}$$

$$\hat{\mathbf{z}}_k^i = \left(\hat{p}'_{1,x}, \hat{p}'_{1,y}, \hat{p}'_{2,x}, \ldots\right)_{i,k}, \tag{16}$$

respectively. The measurement vector and the predicted measurement vector are then obtained by combining the measurements and predictions of all cameras

$$\mathbf{z}_k = (\mathbf{z}_k^m, \mathbf{z}_k^1, \ldots, \mathbf{z}_k^{N-1})^T \tag{17}$$

$$\hat{\mathbf{z}}_k = (\hat{\mathbf{z}}_k^m, \hat{\mathbf{z}}_k^1, \ldots, \hat{\mathbf{z}}_k^{N-1})^T, \tag{18}$$

where the superscript m denotes the master camera. The state and measurement prediction function are then given by

$$\hat{\mathbf{x}}_k^- = \mathbf{f}(\hat{\mathbf{x}}_{k-1}^+) \tag{19}$$

$$\hat{\mathbf{z}}_k = \mathbf{m}(\hat{\mathbf{x}}_k^-), \tag{20}$$

where the superscripts minus and plus denote prediction and measurement update respectively. Equation (20) combines the feature position prediction (12) for all cameras. The respective homographies are determined using the predicted state parameters (19) and the equations given in Section II. For simplicity, the dependence on the intrinsic camera parameters $\mathbf{k}$ as well as the feature positions in the preceding frame are omitted. Equation (19) describes, in case of the general motion model, the transformation of the plane normal into the subsequent camera coordinate system as well as the change in height above ground. For the planar motion model, the function becomes an identity matrix. At least one of the functions (19) and (20) is nonlinear. Therefore, an iterated extended Kalman filter, according to [1], is used. The prediction covariance is then

$$\mathbf{P}_k^- = \mathbf{F}_{k-1}\mathbf{P}_{k-1}^+\mathbf{F}_{k-1}^T + \mathbf{Q}_k, \tag{21}$$

where

$$\mathbf{F}_{k-1} = \left.\frac{\partial \mathbf{f}(\mathbf{x})}{\partial \mathbf{x}}\right|_{\mathbf{x}=\hat{\mathbf{x}}_{k-1}^+}. \tag{22}$$

In an extended Kalman filter the predicted state (19) is used to determine the measurement prediction (20). In case the state prediction is already erroneous, the effect on the measurement prediction might be even more significant due to the linearization around the prediction. It is therefore proposed to use a *relinearization of the measurement equation* [1] which yields an approximate maximum a posteriori estimate (MAP) of the true state when assuming Gaussian noise. Starting from the predicted state estimate ${}^0\hat{\mathbf{x}}_k^+ = \hat{\mathbf{x}}_k^-$, the Jacobi matrix of (20)

$${}^j\mathbf{M}_k = \left.\frac{\partial \mathbf{m}(\mathbf{x})}{\partial \mathbf{x}}\right|_{\mathbf{x}={}^j\hat{\mathbf{x}}_k^+}, \tag{23}$$

as well as the Kalman gain and the updated state estimate

$${}^j\mathbf{W}_k = \mathbf{P}_k^- \, {}^j\mathbf{M}_k^T \left({}^j\mathbf{M}_k \mathbf{P}_k^- \, {}^j\mathbf{M}_k^T + \mathbf{R}_k\right)^{-1} \tag{24}$$

$${}^{j+1}\hat{\mathbf{x}}_k^+ = \hat{\mathbf{x}}_k^- + {}^j\mathbf{W}_k \left(\mathbf{z}_k - \mathbf{m}({}^j\hat{\mathbf{x}}_k^+) + {}^j\mathbf{M}_k \left({}^j\hat{\mathbf{x}}_k^+ - \hat{\mathbf{x}}_k^-\right)\right). \tag{25}$$

are computed iteratively. The superscript j denotes the iteration index. After J iterations the final updated state estimate and updated covariance are

$$\mathbf{P}_k^+ = \left(\mathbf{I} - {}^J\mathbf{W}_k \, {}^J\mathbf{M}_k\right)\mathbf{P}_k^- \tag{26}$$

$$\hat{\mathbf{x}}_k^+ = {}^J\hat{\mathbf{x}}_k^+, \tag{27}$$

where $\mathbf{I}$ is the identity matrix. The explicit matrix inversion in (24) can be avoided, using sequential processing as described in [1]. To illustrate the filter structure, Fig. 4 visualizes an exemplary Jacobi matrix (23) for the general motion model.

V. EXPERIMENTS AND EVALUATION

A. EXPERIMENTAL SETUP

The camera rig used for experimental evaluation consists of four synchronized megapixel wide angle fisheye cameras, mounted on a moving platform. The maximum distance between any two cameras is approximately four meters. Three sequences of lengths 420 to 1045 frames, including three to five turns, have been captured at low speeds. Corresponding features in successive frames have been determined, using the feature detector by Rosten and Drummond [14] and the feature descriptor by Calonder et al. [2]. Matching is then carried out by comparing the absolute and relative Hamming distance in a fixed search window. This approach provides corresponding feature points at pixel accuracy. Including the

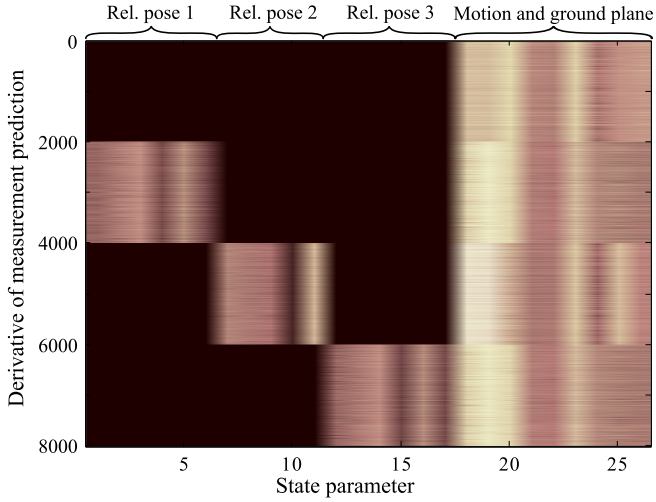

Fig. 4. Jacobi matrix of the measurement prediction with respect to the state parameters (absolute values in logarithmic scale). Zero elements are indicated in black, higher values are brighter. Four views with 1000 feature correspondences were used respectively. The relative pose of camera 2 is parameterized by five parameters only as the distance between this camera and the master camera is fixed (general 3D motion case).

uncertainties in the intrinsic parameters, a standard deviation of approximately 0.5 pixels in each dimension was assumed. The synthetic sequences are supposed to resemble the characteristics of the first real sequence (`Real 1`) with the difference that

- the first synthetic sequence (`Synth. 1`) contains only planar motions of the multi-camera rig,
- the second synthetic sequence (`Synth. 2`) contains additionally pitching and rolling during turns, and
- in the third synthetic sequence (`Synth. 3`), the mobile platform moves on a concave surface.

The first two synthetic sequences are used to compare the two motion models, whereas the third one is used to determine the effect of nonplanarities on the calibration results. In the real world sequences, few non-rigid objects were close to the moving platform and all cameras were able to capture a part of the ground plane at any time. The ground truth was determined using calibration pattern boards, additional cameras, and bundle adjustment.

B. PREPROCESSING

The iterated extended Kalman filter is highly sensitive to outliers, i.e. feature points on non-rigid objects or off the ground plane. Therefore, in a first step, the measurement prediction (20) was used to eliminate features with significant deviations form the predicted image positions. Then, the predicted state estimate (19) was used to determine the horizon. Features located above or close to the horizon were discarded. Finally, we used a RANSAC (random sample consensus) algorithm [7] to determine an inlier set which was then further processed by the Kalman filter. The homography hypotheses were generated from randomly drawn samples and the feature pairs were classified using the transfer error and a fixed threshold.

C. QUANTITATIVE EVALUATION

Given the ground truth relative pose $\Delta\mathbf{T}^i$ and the estimated relative pose $\Delta\hat{\mathbf{T}}^i$ of camera i, the residual transformation $\mathbf{T}^i_\Delta = (\Delta\hat{\mathbf{T}}^i)^{-1}\Delta\mathbf{T}^i$ is computed. From this, the mean position and angular errors are determined as

$$e_P = \frac{1}{N-1}\sum_{i=1}^{N-1} ||\mathbf{C}^i_\Delta||_2 \tag{28}$$

$$e_A = \frac{1}{N-1}\sum_{i=1}^{N-1} \mathtt{acos}\left(\left(\mathtt{tr}(\mathbf{R}^i_\Delta) - 1\right)/2\right) \tag{29}$$

respectively. From the ground truth, 20 sets of perturbed initial relative pose parameters were generated. Additional 20 sets were generated with four times the position and angular error. Results were then obtained, using the initial parameter sets and the real and synthetic sequences. For reference, further 20 runs were performed with ground truth initialization. Table I and II show the mean errors before and after estimation. The mean initial errors are given in the first row, respectively. The deviation of the mean initial position errors for both motion models is due to different fixed parameters, distance and height, respectively. The initial angular errors are not affected. In total, 720 runs were performed.

Results show that the position estimates are significantly better for the general 3D motion model, except in case of pure synthetic planar motion. While the general motion approach is only slightly affected by rolling and pitching, it strongly affects the estimation results with the planar motion model. This can be explained by the change in height and orientation with respect to the ground plane during turns. Ruland et al. [15] also observed a strong dependency between errors in camera height and estimation errors. Both approaches appear to be very sensitive to nonplanarities, as can be seen in the last row in both tables respectively. For the real sequences, the accuracy of the orientation estimation is

TABLE I

MEAN POSITION AND ANGULAR ERRORS (GENERAL MOTION)

	Mean error for different initializations		
Sequence	298.2mm (5.28°)	76.5mm (1.32°)	0.0mm (0.00°)
Real 1	37.1mm (0.44°)	23.7mm (0.44°)	22.3mm (0.43°)
Real 2	80.7mm (0.43°)	26.8mm (0.43°)	25.3mm (0.44°)
Real 3	22.4mm (0.29°)	13.9mm (0.29°)	13.2mm (0.29°)
Synth. 1	28.4mm (0.06°)	7.5mm (0.02°)	2.6mm (0.02°)
Synth. 2	53.5mm (0.06°)	11.1mm (0.03°)	3.6mm (0.03°)
Synth. 3	54.4mm (0.24°)	14.6mm (0.23°)	9.7mm (0.23°)

TABLE II

MEAN POSITION AND ANGULAR ERRORS (PLANAR MOTION)

	Mean error for different initializations		
Sequence	317.9mm (5.28°)	82.6mm (1.32°)	0.0mm (0.00°)
Real 1	89.4mm (0.46°)	53.4mm (0.48°)	48.6mm (0.50°)
Real 2	134.5mm (0.38°)	43.5mm (0.30°)	27.4mm (0.30°)
Real 3	84.5mm (0.26°)	41.4mm (0.26°)	34.9mm (0.26°)
Synth. 1	71.8mm (0.20°)	15.8mm (0.05°)	2.7mm (0.01°)
Synth. 2	99.9mm (0.43°)	31.7mm (0.33°)	21.7mm (0.30°)
Synth. 3	106.0mm (0.64°)	48.7mm (0.55°)	44.0mm (0.53°)

roughly the same for both approaches. The remaining mean angular errors are approximately 0.26 to 0.48 degrees. The remaining mean position errors are approximately 14 to 27 millimeters in case of the general motion approach and 41 to 53 millimeters in case of the planar motion approach, when initialized with the medium level of perturbation.

To illustrate the accuracy of the egomotion estimation, a reconstruction of the ground plane of the first real sequence is given in Fig. 5. The close-up shows the overlapping region at start and end of the sequence. Ground truth was used here for initialization. The ghosting artifacts are caused by errors in the egomotion estimation.

D. INTRINSIC PARAMETERS

In order to assess the influence of intrinsic calibration, sets of perturbed intrinsic parameters have been generated according to $\mathbf{k}_{perturbed} \sim \mathcal{N}(\mathbf{k}, \Sigma_{\mathbf{kk}})$ using the intrinsic parameters $\mathbf{k}$ and the approximate covariance matrix $\Sigma_{\mathbf{kk}}$, see Section III. The algorithm for general motion was then initialized with ground truth extrinsic, and ground truth and perturbed intrinsic parameters. Table III shows the results for the first synthetic sequence. It can be seen that the uncertainty in the intrinsic parameters has a larger impact on orientations than on positions.

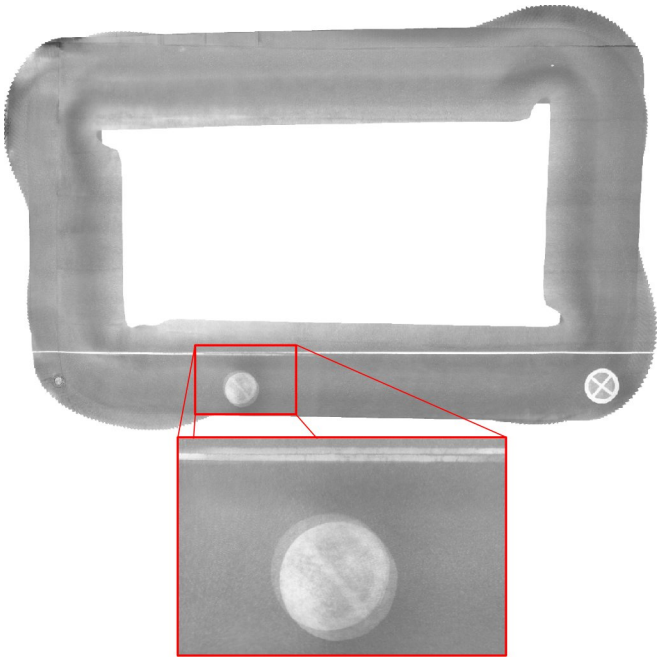

Fig. 5. Reconstruction of ground plane of the first real sequence and close-up of the overlapping region at sequence start and end. The ghosting artifacts are caused by errors in the motion estimation. The reconstructed area is approximately 40 meters across.

TABLE III
INFLUENCE OF INTRINSIC PARAMETERS

	Mean angular and position error and respective standard deviations	
Ground truth parameters	2.63mm (0.55mm)	0.0228° (0.0055°)
Perturbed parameters	3.65mm (0.95mm)	0.0438° (0.0128°)

VI. CONCLUSION AND FUTURE WORK

In this paper an online approach for extrinsic calibration of a multi-camera rig was presented. We have exploited ground plane induced homographies to impose constraints on the extrinsic calibration of all cameras. Rather than applying Kalman filters individually to each camera pose, we have introduced a joint Kalman filter whose state vector comprises extrinsic parameters of all cameras. Two motion models were compared. Except for the special case of pure synthetic planar motion, the general 3D motion approach was always superior. Furthermore, The results indicate that a violation of the planar ground assumption has a strong effect on the estimation results. Therefore, future work will focus on the relaxation of the planarity assumption.

REFERENCES

[1] Y. Bar-Shalom and L. Xiao-Rong, *Estimation and Tracking: Principles, Techniques and Software*. Artech House Boston, 1993.

[2] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, "Brief: Binary robust independent elementary features," in *Proceedings of the 11th European conference on Computer vision: Part IV*. Berlin, Heidelberg: Springer-Verlag, 2010, pp. 778–792.

[3] G. Carrera, A. Angeli, and A. Davison, "Slam-based automatic extrinsic calibration of a multi-camera rig," in *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, may 2011, pp. 2652 –2659.

[4] Y.-S. Chen, L.-G. Liou, Y.-P. Hung, and C.-S. Fuh, "Three-dimensional ego-motion estimation from motion fields observed with multiple cameras," *Pattern Recognition*, vol. 34, no. 8, pp. 1573 – 1583, 2001.

[5] T. Dang, C. Hoffmann, and C. Stiller, "Continuous stereo self-calibration by camera parameter tracking," *Image Processing, IEEE Transactions on*, vol. 18, no. 7, pp. 1536 –1550, july 2009.

[6] S. Esquivel, F. Woelk, and R. Koch, "Calibration of a multi-camera rig from non-overlapping views," in *Proceedings of the 29th DAGM conference on Pattern recognition*. Berlin, Heidelberg: Springer-Verlag, 2007, pp. 82–91.

[7] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge University Press, 2004.

[8] B. M. Kitt, J. Rehder, A. D. Chambers, M. Schonbein, H. Lategahn, and S. Singh, "Monocular visual odometry using a planar road model to solve scale ambiguity," in *Proc. European Conference on Mobile Robots*, September 2011.

[9] P. Lebraly, E. Royer, O. Ait-Aider, C. Deymier, and M. Dhome, "Fast calibration of embedded non-overlapping cameras," in *Robotics and Automation, 2011 IEEE International Conference on*, may 2011, pp. 221 –227.

[10] C. Mei, "Omnidirectional calibration toolbox." [Online]. Available: http://www.robots.ox.ac.uk/~cmei/Toolbox.html

[11] M. Miksch, B. Yang, and K. Zimmermann, "Automatic extrinsic camera self-calibration based on homography and epipolar geometry," in *Intelligent Vehicles Symposium, 2010 IEEE*, june 2010, pp. 832 –839.

[12] F. Pagel, "Motion adjustment for extrinsic calibration of cameras with non-overlapping views," in *Computer and Robot Vision, 2012 Ninth Conference on*, may 2012, pp. 94 –100.

[13] F. Pagel and D. Willersinn, "Extrinsic camera calibration in vehicles with explicit ground estimation," in *Int. Workshop on Intelligent Transportation*, 2011.

[14] E. Rosten and T. Drummond, "Machine learning for high-speed corner detection," in *Proceedings of the 9th European conference on Computer Vision - Volume Part I*. Berlin, Heidelberg: Springer-Verlag, 2006, pp. 430–443.

[15] T. Ruland, H. Loose, T. Pajdla, and L. Krueger, "Hand-eye autocalibration of camera positions on vehicles," in *Intelligent Transportation Systems, 2010 13th International IEEE Conference on*, sept. 2010, pp. 367 –372.