

Motion Adjustment for Extrinsic Calibration of Cameras with Non-overlapping Views

Frank Pagel

Video Exploitation Systems VID

Fraunhofer Institute of Optronics, System Technologies and Image Exploitation IOSB

Karlsruhe, Germany

frank.pagel@iosb.fraunhofer.de

Abstract—This paper addresses the issue of calibrating multiple cameras on a mobile platform. Due to decreasing sensor prices and increasing processing performance, the use of multiple cameras in vehicles becomes an attractive possibility for environment perception. To avoid restrictions relating to the camera arrangement, we focus on non-overlapping camera configurations. Hence, we resign the usage of corresponding features between the cameras. The hand-eye calibration technique based on visual odometry is basically able to solve this problem by exploiting the cameras' motions. However, this technique suffers from inaccuracies in motion estimation. Especially the absolute magnitudes of the translational velocities of each camera are essential for a successful calibration.

This contribution presents a novel approach to solve the hand-eye calibration problem for two cameras on a mobile platform with non-overlapping fields of view. The so-called motion adjustment simultaneously estimates the extrinsic parameters up to scale as well as the relative motion magnitudes. Results with simulated and real data are presented.

Keywords—motion adjustment, non-overlapping views, hand-eye calibration, multi-camera calibration

I. INTRODUCTION AND PROBLEM DESCRIPTION

Over the last years multi-camera applications in vehicles have become more and more popular as the costs of the sensor production decreased drastically. Multiple sensors can be used to cover a wider field of view or scene reconstruction tasks [2]. However, in many cases it is not possible to cover the full environment or to guarantee overlapping fields of view due to restrictions in design, energy consumption or physical capacity.

To be able to use the full range of possibilities that go along with such a sensor setup, the relative adjustment of the sensors must be known. The extrinsic parameters between each camera can be described as an Euclidean transformation. Then, the informations that are extracted from the single cameras can be merged and referenced in a common coordinate system. This might be very useful for reconstruction, object detection or attention guidance tasks. 3D scene reconstruction with multiple cameras is a growing field of research [10]. Considering modern structure from motion (SFM) techniques, single cameras can also be used for egomotion estimation [9], [13] as well as for dense reconstruction [11].

Common calibration techniques like for stereo cameras [3] fail because of the non-overlapping FOV. The cameras do not see the same scene and hence no corresponding image features can be used. This paper addresses the calibration without using any pattern or known scene structure.

Although pattern-based calibration methods for non-overlapping cameras in vehicles exist, such methods are difficult to realize. In [7] traffic signs are used as calibration patterns, seen by different cameras at different times. A matching between cameras at different times causes different object appearances and needs storage and time management respectively. Furthermore, for general motions, surroundings and camera adjustments, correspondences between frames cannot be guaranteed at any time (imagine for example a left and a right looking camera in a vehicle). [1] and [8] calibrated a multi-camera rig on a mobile robot by performing a bundle adjustment between several views and different robot positions. However, they performed their calibration in a single room and hence in a strictly restricted area, to guarantee corresponding objects between the frames at different platform positions.

Another solution comes from the robotics community, called hand-eye calibration [6], [17]. Such a solution is presented in [4] for a hand-held camera rig. In [16] a planar approach is proposed for extrinsic calibration of a two camera system with non-overlapping views and fixed camera height in a parking scenario at low speeds.

In practice we suffer from two circumstances: First, in a vehicle the rotational part of the motion transformations is rather small. The motion estimation is vision-based and therefore needs corresponding features between the frames, so that we only use continuous motion estimations. And second, it appears to be difficult to estimate the absolute magnitude of the motion's translation (see Fig. 1). But this magnitude, especially the difference between the velocities of two camera modules, is crucial for the calibration. The basic assumption behind the hand-eye calibration is that the distance between two cameras stays the same, no matter how the rig moves. Without the knowledge of the relative translational motion of the cameras the constraint is mostly useless.

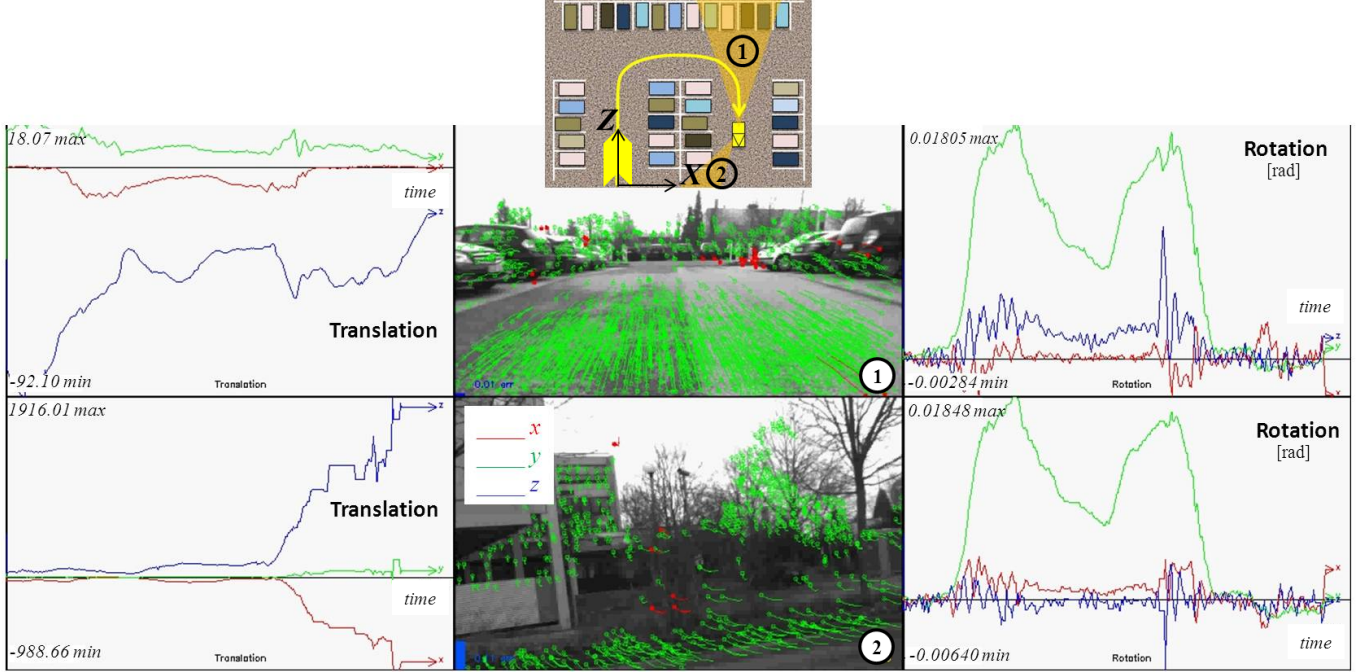


Figure 1. Estimated motion transformation parameters from our test sequence on a parking lot. Center top: Sketch of the scene and the vehicle-mounted cameras. The vehicle performs a long U-turn. Center: Frames from the two cameras with optical flow vectors. Left: Continuously estimated translation parameters t_x, t_y, t_z . Right: Continuously estimated rotation parameters r_x, r_y, r_z . The translation is initially estimated only up to scale and for the rest of the sequence relative to the previous time step, and hence unitless.

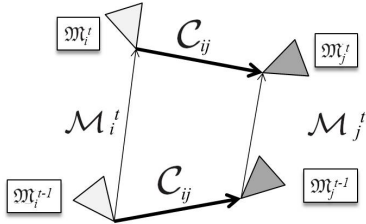


Figure 2. Basic geometric constellation for a 2-camera rig. \mathbf{C} is the extrinsic calibration matrix, \mathcal{M} is the camera motion.

The approach that handles these difficulties is the so-called motion adjustment. Motion adjustment simultaneously estimates the extrinsic parameters as well as the motion scales and is designed of being used for both offline and online calibration tasks.

Section II introduces the geometric model and basic algorithmic techniques. The motion adjustment approach and its implementation are presented in Section III. Experimental results with both simulated and real data are shown in Section IV. Finally, section V gives a conclusion and an outlook to future work.

II. GEOMETRIC MODEL

In this Section we shortly present the parameters that are necessary to describe the complete geometric structure of a moving camera rig. Both the motion of a single camera

and the relative position of two cameras can be considered as a Euclidean transformation. The transformation between two camera modules \mathfrak{M}_i and \mathfrak{M}_j at time t is given by the transformation matrix

$$\mathbf{C}_{ij} = \begin{bmatrix} \mathbf{R}_{ij} & \mathbf{t}_{ij} \\ \mathbf{0}^T & 1 \end{bmatrix}_{4 \times 4}. \quad (1)$$

$\mathbf{R}(r_x, r_y, r_z) \in \mathbb{R}^{3 \times 3}$ is a rotation matrix with $\mathbf{R}^T \mathbf{R} = \mathbf{R} \mathbf{R}^T = \mathbf{I}$ and $\mathbf{t} = [t_x, t_y, t_z]^T$ is a translation vector. Hence, the inverse is given by $\mathbf{C}_{ij}^{-1} = \mathbf{C}_{ji}$. The Euler angles are represented by the vector $\mathbf{r} = [r_x, r_y, r_z]^T$. \mathbf{C} is the extrinsic transformation with its parameters \mathbf{t} and \mathbf{r} . The XZ -plane is considered as the ground plane and Y the longitudinal axis of the world coordinate system.

The motion of camera \mathfrak{M}_i between two time steps t and $t+1$ is given by

$$\mathcal{M}_{i_t} = \begin{bmatrix} \mathbf{\Omega}_{i_t} & \mathbf{v}_{i_t} \\ \mathbf{0}^T & 1 \end{bmatrix}_{4 \times 4}, \quad (2)$$

with rotation matrix $\mathbf{\Omega}(\omega_x, \omega_y, \omega_z) \in \mathbb{R}^{3 \times 3}$ and translation vector $\mathbf{v} = [v_x, v_y, v_z]^T$ and the rotation parameters $\omega = [\omega_x, \omega_y, \omega_z]^T$. These relations are shown in Fig. 2.

Alternatively, the rotation matrices \mathbf{R} and $\mathbf{\Omega}$ can be represented by quaternions $\mathbf{q} = [q_w, q_x, q_y, q_z]$. A rotation of a 3D point \mathbf{X} is then given by the quaternion multiplication $\mathbf{q} \mathbf{X} \bar{\mathbf{q}} \equiv \mathbf{R} \mathbf{X}$, where $\bar{\mathbf{q}}$ is the conjugate quaternion of \mathbf{q} . Quaternions will be useful in Section III-C.

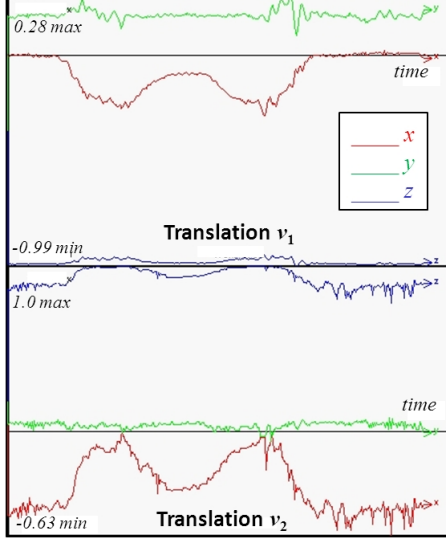


Figure 3. Scaled translation values of the motion parameters from Fig. 1 with $\|\mathbf{v}_t\| = 1$.

The extrinsic parameters \mathbf{R} and \mathbf{t} between \mathfrak{M}_i and \mathfrak{M}_j can be expressed with respect to the cameras' motions (Fig. 2)

$$\mathcal{M}_i = \mathbf{C}_{ij}^{-1} \mathcal{M}_j \mathbf{C}_{ij} \quad (3)$$

which leads us with Eqn. 1 and 2 to the basic hand-eye calibration equation

$$[\mathbf{I} - \boldsymbol{\Omega}_i] \mathbf{t}_{ij} + \mathbf{R}_{ij} \mathbf{v}_j - \mathbf{v}_i = \mathbf{0} \quad (4)$$

in dependence of the extrinsic translation \mathbf{t}_{ij} and rotation \mathbf{R}_{ij} .

III. SPARSE MOTION ADJUSTMENT

Fundamental for performing a motion-based hand-eye calibration is the motion estimation of the cameras. Consider, for example, a vehicle driving around the corner with two cameras mounted, one on the left and one on the right side. Then surely the distance covered by the outer camera is longer than the distance of the inner camera. This is also the reason why odometric data from the vehicle is only little useful for hand-eye calibrations, as it does not cover the differences of the cameras' motion scales.

In our case, this is done using a standard bundle adjustment approach [9] by minimizing the projection error of a sparse set of points. The transformations as well as the scene coordinates are estimated within a block of Γ consecutive frames. In our case, the intrinsic parameters are assumed to be known.

Despite the fact that the motion parameters were predicted temporally (relative to the previous $\Gamma - 1$ estimations), the estimation of the camera modules' velocities can become unreliable over time as seen in Fig.1. However, the estimation of the translational motion is not entirely useless. As

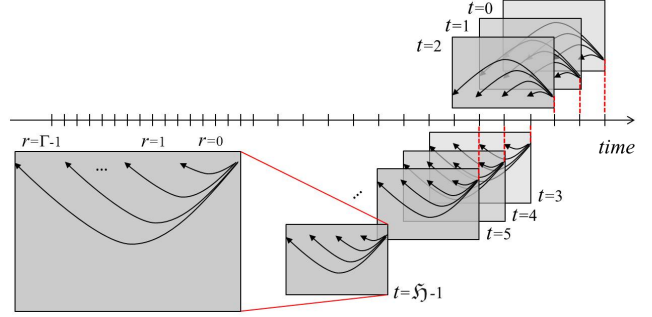


Figure 4. Bundle adjustment is calculated for a motion block of size Γ , motion adjustment is calculated over history of \mathfrak{J} motion blocks.

can be seen in Fig. 3, the scaled translation vectors reflect the car's motion adequately.

The simple idea of motion adjustment is not only estimating the extrinsic parameters but also the scale factors of the normalized motion translations simultaneously. This leads us to the extended hand-eye calibration equation

$$[\mathbf{I} - \boldsymbol{\Omega}_{i_t}] \mathbf{t}_{ij} + \kappa_t \mathbf{R}_{ij} \mathbf{v}_{j_t}^s - \mathbf{v}_{i_t}^s = \mathbf{0} \quad (5)$$

with $\kappa_t \in \mathbb{R}$, $\mathbf{v}^s = \frac{\mathbf{v}}{\|\mathbf{v}\|}$ and $t = 1, \dots, \mathfrak{J}$. Eq. 5 can then be minimized with respect to \mathbf{R}_{ij} , \mathbf{t}_{ij} and κ_t with standard non-linear optimization techniques. Due to the scaled translational motion, the extrinsic translation can also only be estimated up to scale.

It is well known and can easily be seen in Eq. 5 that rotational motion $\boldsymbol{\Omega}_{i_t} \neq \mathbf{I}$ is necessary to be able to estimate \mathbf{t}_{ij} [17]. Assuming \mathbf{R}_{ij} to be known, we still need at least two different motions to solve Eq. 5 for \mathbf{t}_{ij} and κ_t . Hence κ_t is estimated for a motion block of size $\Gamma > 1$. Within such a block the relative translational scales must be preserved. Fortunately, this is the case with a motion estimation via bundle adjustment with a consecutive block of frames. Furthermore, to preserve full rank of the linear equation system in Eq. 5, the rotation matrix $\boldsymbol{\Omega}$ need to vary within a motion block Γ , which means that moving in circles is an insufficient motion.

A. Efficient Implementation

The problem of scaled motion adjustment is equivalent to the bundle adjustment used for simultaneous motion and structure estimation. So, according to [9] motion adjustment can be quite efficiently implemented. Because of the relation to the solving of sparse bundle adjustment, we speak of sparse motion adjustment.

Basically, motion adjustment for two camera modules $\mathfrak{M}_i, \mathfrak{M}_j$ can be solved using a standard non-linear Levenberg-Marquardt optimization [12]. The problem can be formally described by the state vector \mathbf{x} and the mea-

surement vector z

$$\begin{aligned} \mathbf{x} &= [\mathbf{t}_{ij}^T, \mathbf{r}_{ij}^T, \boldsymbol{\kappa}_{ij}^T]^T \\ \mathbf{z} &= [\dots, \mathbf{v}_{i_{tr}}^s, \omega_{i_{tr}}^T, \mathbf{v}_{j_{tr}}^s, \omega_{j_{tr}}^T, \dots]^T \\ \text{with } &t = 0, \dots, \mathfrak{H} - 1, \\ &r = 0, \dots, \Gamma - 1 \end{aligned}$$

and $\boldsymbol{\kappa}_{ij} = [\dots, \kappa_{ij_t}, \dots]^T$. $\mathcal{M}(\mathbf{v}_{i_{tr}}^s, \omega_{i_{tr}})$ is the r -th motion in the t -th history block of \mathfrak{M}_i with scaled translation $\mathbf{v}_{i_{tr}}^s$. From Eq. 5 we get the constraint function

$$\mathbf{h}(\mathbf{x}, \mathbf{z}) = [\dots, \mathbf{h}_{tr}(\mathbf{x}, \mathbf{z})^T, \dots]^T =: \boldsymbol{\epsilon} \quad (6)$$

with $\boldsymbol{\epsilon} = [\dots, \epsilon_{tr}, \dots]_{(\mathfrak{H} \cdot \Gamma \times 1)}^T = [\dots, \boldsymbol{\epsilon}_t^T, \dots]^T$ and

$$\mathbf{h}_{tr}(\mathbf{x}, \mathbf{z}) = \left\| [\mathbf{I} - \boldsymbol{\Omega}_{i_{tr}}] \mathbf{t}_{ij} + \kappa_t \mathbf{R}_{ij} \mathbf{v}_{j_{tr}}^s - \mathbf{v}_{i_{tr}}^s \right\|^2 \stackrel{!}{=} 0. \quad (7)$$

Consider the Levenberg-Marquardt normal equation [12]

$$[\mathbf{J}^T \mathbf{J} - \rho \mathbf{I}] \begin{bmatrix} \boldsymbol{\delta}_C \\ \boldsymbol{\delta}_\kappa \end{bmatrix} = \mathbf{J}^T \boldsymbol{\epsilon} \quad (8)$$

where $\boldsymbol{\delta}_C$ and $\boldsymbol{\delta}_\kappa$ are the update vectors for the extrinsic and the scale parameters respectively. Then, the jacobian $\frac{\partial \mathbf{h}}{\partial \mathbf{x}}$ looks like

$$\mathbf{J} = \begin{bmatrix} \mathbf{B}_1 & \mathbf{s}_1 & \mathbf{0} \\ \vdots & \ddots & \\ \mathbf{B}_{\mathfrak{H}} & \mathbf{0} & \mathbf{s}_{\mathfrak{H}} \end{bmatrix} \quad (9)$$

with

$$\mathbf{B}_t = \begin{bmatrix} \frac{\partial \mathbf{h}_{t,0}}{\partial \mathbf{t}} & \frac{\partial \mathbf{h}_{t,0}}{\partial \mathbf{r}} \\ \vdots & \vdots \\ \frac{\partial \mathbf{h}_{t,\Gamma-1}}{\partial \mathbf{t}} & \frac{\partial \mathbf{h}_{t,\Gamma-1}}{\partial \mathbf{r}} \end{bmatrix} \quad \text{and} \quad \mathbf{s}_t = \begin{bmatrix} \frac{\partial \mathbf{h}_{C,t,0}}{\partial \kappa_t} \\ \vdots \\ \frac{\partial \mathbf{h}_{C,t,\Gamma-1}}{\partial \kappa_t} \end{bmatrix}. \quad (10)$$

Using Eqn. 9 and 10, the left side term of Eq. 8 becomes

$$\begin{aligned} &[\mathbf{J}^T \mathbf{J} - \rho \cdot \mathbf{I}] =: \begin{bmatrix} \mathbf{U} & \mathbf{W} \\ \mathbf{W}^T & \mathbf{V} \end{bmatrix} \\ &= \left[\begin{array}{c|ccc} \left[\sum_{t=0}^{\mathfrak{H}-1} \mathbf{B}_t^T \mathbf{B}_t \right] - \rho & \mathbf{B}_0^T \mathbf{s}_0 & \dots & \mathbf{B}_{\mathfrak{H}-1}^T \mathbf{s}_{\mathfrak{H}-1} \\ \hline \mathbf{s}_0^T \mathbf{B}_0 & \mathbf{s}_0^T \mathbf{s}_0 - \rho & & \\ \vdots & & \ddots & \mathbf{0} \\ \mathbf{s}_{\mathfrak{H}-1}^T \mathbf{B}_{\mathfrak{H}-1} & \mathbf{0} & & \mathbf{s}_{\mathfrak{H}-1}^T \mathbf{s}_{\mathfrak{H}-1} - \rho \end{array} \right] \end{aligned} \quad (11)$$

So the normal equation can be written as

$$\begin{bmatrix} \mathbf{U} - \mathbf{WV}^{-1} \mathbf{W}^T & \mathbf{0} \\ \mathbf{W}^T & \mathbf{V} \end{bmatrix} \begin{bmatrix} \boldsymbol{\delta}_C \\ \boldsymbol{\delta}_\kappa \end{bmatrix} = \begin{bmatrix} \boldsymbol{\epsilon}_C - \mathbf{WV}^{-1} \boldsymbol{\epsilon}_\kappa \\ \boldsymbol{\epsilon}_\kappa \end{bmatrix} \quad (12)$$

with

$$\boldsymbol{\epsilon}_C = \sum_{t=0}^{\mathfrak{H}-1} \mathbf{B}_t^T \boldsymbol{\epsilon}_t \quad \text{and} \quad \boldsymbol{\epsilon}_\kappa = [\dots, \mathbf{s}_t^T \boldsymbol{\epsilon}_t, \dots]^T. \quad (13)$$

As \mathbf{V} is diagonal and $\mathbf{U} - \mathbf{WV}^{-1} \mathbf{W}^T$ is positive definit, Eq. 12 can be solved efficiently for $\boldsymbol{\delta}_C$ and $\boldsymbol{\delta}_\kappa$ with a Cholesky decomposition.

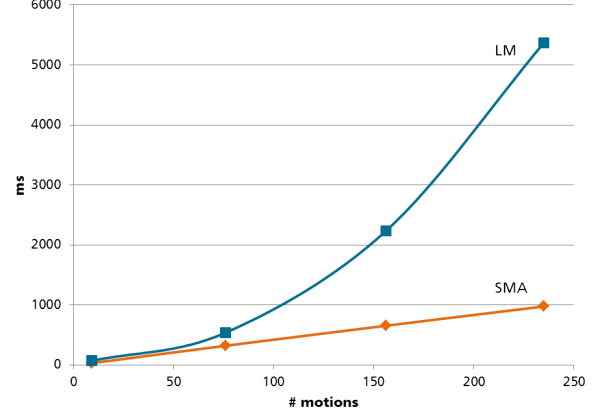


Figure 5. Comparison of runtimes in ms between standard LM-approach and sparse motion adjustment computation with respect to the considered number of motions \mathfrak{H} .

B. Covariance Estimation

To have a measure of the remaining uncertainty the covariance matrix $\boldsymbol{\Sigma}_x$ of the extrinsics \mathbf{x} is calculated according to the well known Kalman equations

$$\boldsymbol{\Sigma}_x = \boldsymbol{\Sigma}_0 - \boldsymbol{\Sigma}_0 \mathbf{J}^T [\mathbf{J} \boldsymbol{\Sigma}_0 \mathbf{J}^T + \sigma^2 \mathbf{I}]^{-1} \mathbf{J} \boldsymbol{\Sigma}_0 \quad (14)$$

where $\boldsymbol{\Sigma}_0 = \text{Diag}(\sigma_t^2, \sigma_t^2, \sigma_t^2, \sigma_r^2, \sigma_r^2, \sigma_r^2, \sigma_\kappa^2, \dots, \sigma_\kappa^2)$ is the initial noise of \mathbf{x} , $\mathbf{J} = \frac{\partial \mathbf{h}}{\partial \mathbf{x}}$ and

$$\sigma^2 = \frac{1}{\mathfrak{H} \cdot \Gamma} \sum_{t=1}^{\mathfrak{H}} \sum_{r=1}^{\Gamma} \mathbf{h}_{tr}(\mathbf{x}, \mathbf{z}) \quad (15)$$

is the variance of the remaining residual (note that the residual function \mathbf{h}_{tr} is already square).

C. Initialization of Camera Orientations

For purely translational motions of two camera modules $\mathfrak{M}_i, \mathfrak{M}_j$, e. g. $\boldsymbol{\Omega}_{\{i,j\}} = \mathbf{I}$, the basic equation 5 becomes

$$\mathbf{R} \mathbf{v}_i - \mathbf{v}_j = \mathbf{0}. \quad (16)$$

Note that both motion scales are equal, $\kappa_i = \kappa_j$, and can therefore be omitted in Eq. 16. Substituting \mathbf{R} by its quaternion \mathbf{q} , we get

$$\begin{aligned} \mathbf{0} &= \mathbf{q} \mathbf{v}_i^q \bar{\mathbf{q}} - \mathbf{v}_j^q \\ &= \mathbf{q} \mathbf{v}_i^q - \mathbf{v}_j^q \mathbf{q} \\ &= \mathbf{Q}_{\mathbf{v}_i}^< \mathbf{q} - \mathbf{Q}_{\mathbf{v}_j}^> \mathbf{q} \\ &= \underbrace{[\mathbf{Q}_{\mathbf{v}_i}^< - \mathbf{Q}_{\mathbf{v}_j}^>]}_{=: \mathbf{A}} \mathbf{q} \end{aligned} \quad (17)$$

with

$$\mathbf{Q}_{\mathbf{v}_i}^< = \begin{bmatrix} 0 & -v_{x_i} & -v_{y_i} & -v_{z_i} \\ v_{x_i} & 0 & v_{z_i} & -v_{y_i} \\ v_{y_i} & -v_{z_i} & 0 & v_{x_i} \\ v_{z_i} & v_{y_i} & -v_{x_i} & 0 \end{bmatrix},$$

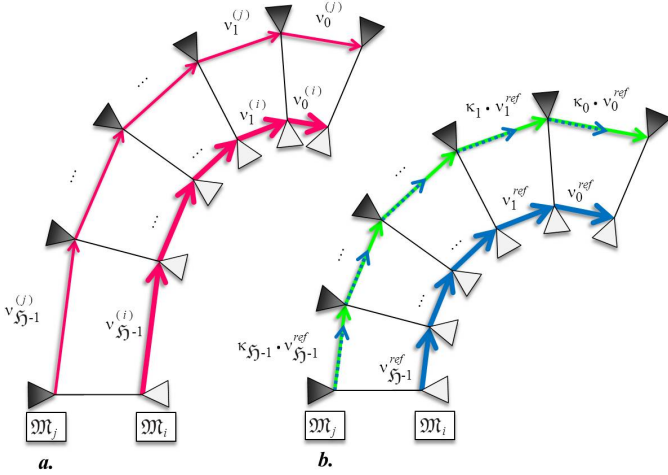


Figure 6. a. Original trajectory (red) of two camera modules $\mathfrak{M}_i, \mathfrak{M}_j$ with unscaled velocity magnitudes $v_t^{(i,j)}$. b. First, the translation vectors are scaled to a reference magnitude v_t^{ref} (blue/blue dotted), e.g. $v_t^{ref} = 1$. Afterwards the translation of \mathfrak{M}_j is rescaled to $\kappa \cdot v_t^{ref}$ (green) using the scales from the motion adjustment. The scales κ_t guarantee consistent motions with respect to the relative position \mathbf{t}_{ij} and orientation \mathbf{R}_{ij} of the cameras.

$$\mathbf{Q}_{v_j}^> = \begin{bmatrix} 0 & -v_{x_j} & -v_{y_j} & -v_{z_j} \\ v_{x_j} & 0 & -v_{z_j} & v_{y_j} \\ v_{y_j} & v_{z_j} & 0 & -v_{x_j} \\ v_{z_j} & -v_{y_j} & v_{x_j} & 0 \end{bmatrix}$$

and $\mathbf{v}^q = (0, v_x, v_y, v_z)$. The orientation \mathbf{q} is then given by the eigenvector of the smallest eigenvalue of $\mathbf{A}^T \mathbf{A}$ [4].

IV. EXPERIMENTAL RESULTS

From Eqn. 4 and 5 we see that without any rotational motion, e.g. $\mathbf{\Omega} = \mathbf{I}$, we are unable to solve for the extrinsic translation \mathbf{t} . This implies that for purely planar motions where both cameras have the same rotation axes, which street vehicles usually perform, it is impossible to solve for the translational height difference t_y between the cameras [17]. In that case we suggest a separate estimation of the ground plane [15].

For our experiments we assumed the simpler case of a 2D motion with $v_y = \omega_x = \omega_z = 0$, where we assume the ground plane to be the XZ -world plane. Hence, only 3 extrinsic parameters need to be estimated, namely t_x, t_z and the rotation angle r_y . A 2D motion can be derived from a planar 3D motion by rotating all motions of \mathfrak{M}_i with a rotation matrix \mathbf{R}_{2D_i} that fulfills

$$\boldsymbol{\alpha}_i = \mathbf{R}_{2D_i} \cdot [0, 1, 0]^T. \quad (18)$$

Here, $\boldsymbol{\alpha}_i$ is the rotation axis of \mathfrak{M}_i with $\mathbf{\Omega}_{i_t} \boldsymbol{\alpha}_i = \boldsymbol{\alpha}_i$.

A. Simulation

To investigate the dependency between motion and relative camera positions, we ran simulations with two cameras

moving in a plane, assuming the relative orientation is known. As motion we chose a linearly increasing rotation. Knowing the ground truth, we could calculate a direct error measure as the relative difference between the estimated translation vectors

$$E = \frac{\|\mathbf{t}^* - \hat{\mathbf{t}}\|}{\|\mathbf{t}^*\|} \quad (19)$$

where \mathbf{t}^* is the ground truth value of the camera position and $\hat{\mathbf{t}}$ is the estimation result. As we simulated a 2D problem, e.g. a motion in the $X - Z$ -ground plane, we were only interested in the estimation of the two translational parameters t_x, t_z . All $\kappa_{i=1, \dots, 5}$ were initialized with 1 and $\hat{\mathbf{t}}_{init} = [0, 0]^T$. Input data for the sparse motion adjustment calculation were the scaled motion history blocks of the reference camera in the center and the second camera, positioned within a square around the reference camera.

Fig. 7 shows two examples of the resulting error heat maps for two differently scaled motions (but both with the same rotational motion). Surprisingly, the extrinsic position of the two cameras could not be calculated equally well at all positions around the center. We could observe that in cases with bad estimation results the scale factors κ_t could not be determined correctly.

Although we unfortunately cannot specify such “bad” cases *a priori*, we are able to detect them by analyzing the resulting covariance matrix (Eq. 14). A principal component analysis of the covariance matrix $\boldsymbol{\Sigma}_x$ reveals the remaining uncertainty in the parameter space. Significant eigenvalues of $\boldsymbol{\Sigma}_x$ with a strong direction along the translational parameter axes indicate a major remaining uncertainty of t_x, t_z and little uncertainty in the κ_t -parameter space. Whereas high values of E came with small eigenvalues along all directions in the parameter space, which indicates that the remaining uncertainty is more or less equally partitioned among the translational as well as the scale parameters.

B. Real Data Experiment

We tested the sparse motion adjustment in a planar application. A bird’s eye view was generated for two cameras $\mathfrak{M}_1, \mathfrak{M}_2$ in a vehicle based on the homography Knowing the Euclidean transformation $[\mathbf{R}_w \mathbf{t}_w]$ between camera and world coordinate system for a reference camera module \mathfrak{M}_1 (for example from an offline calibration), we can determine the homography \mathbf{H}_{BEV} that maps an image point \mathbf{x}_2 from \mathfrak{M}_2 to its corresponding point \mathbf{x}_w on the world plane. The columns of \mathbf{R}_G may be given by $[c_1 c_2 c_3]$ and its rows by $[r_1^T r_2^T r_3^T]^T$. The homography in dependence of the extrinsic parameters between camera \mathfrak{M}_2 and \mathfrak{M}_1 is then given by

$$\mathbf{H}_{BEV} = \mathbf{K}_2 \left[\mathbf{R}_{21} + \frac{\mathbf{t}_{21} \mathbf{n}^T}{d} \right] \mathbf{H}_w \quad (20)$$

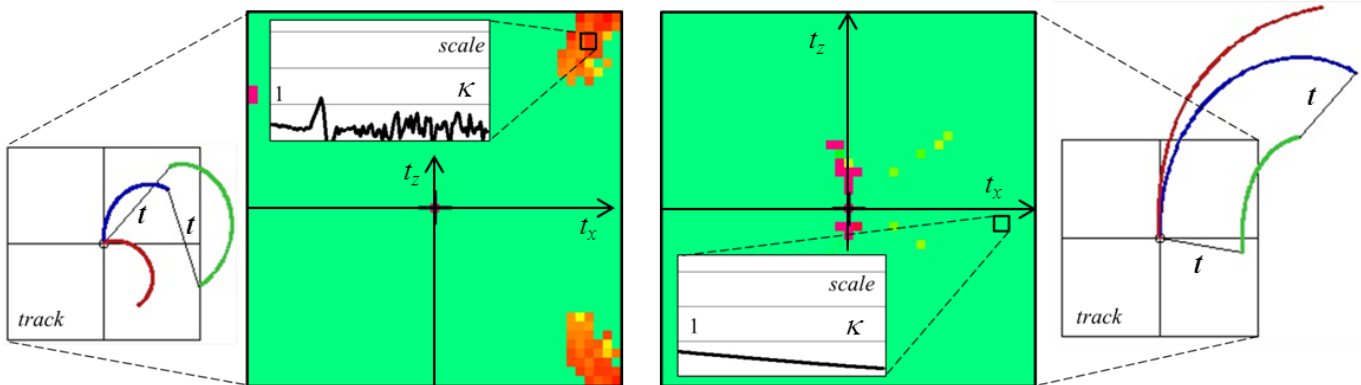


Figure 7. Error maps in dependence of the camera positions for differently scaled motions. The reference camera \mathfrak{M}_1 is always in the center, the position of the second camera \mathfrak{M}_2 is within a square around the reference camera. Red means $E > 1$, green is $E = 0$. The blue track is the trajectory of \mathfrak{M}_1 (each motion transformation is scaled to $\|\mathbf{v}_{t,0}\| = \nu_{ref}$). The red track is the trajectory of the motion of \mathfrak{M}_2 with relative scales $\hat{\kappa}_t = 1$ and $\hat{t}_x = \hat{t}_z = 0$. The blue and the red motions are the input data for the motion adjustment computation. The green track is the ground truth of the resulting trajectory of \mathfrak{M}_2 with the correct relative scales κ_t^* and $\hat{t}_x = t_x^*, \hat{t}_z = t_z^*$. Left: Error map for a motion with small scale ($\nu_{ref} = 0.5$) and a detail view of the estimated motion scales with a resulting big error E . Right: Error map for a motion with big scale ($\nu_{ref} = 1.5$) and a detail view of the estimated motion scales with a small error E . The rotational motion in both heatmaps is the same with linearly increasing values of $\omega_y \in [0.01, 0.02]$ and $\mathfrak{J} = 100, \Gamma = 6$.

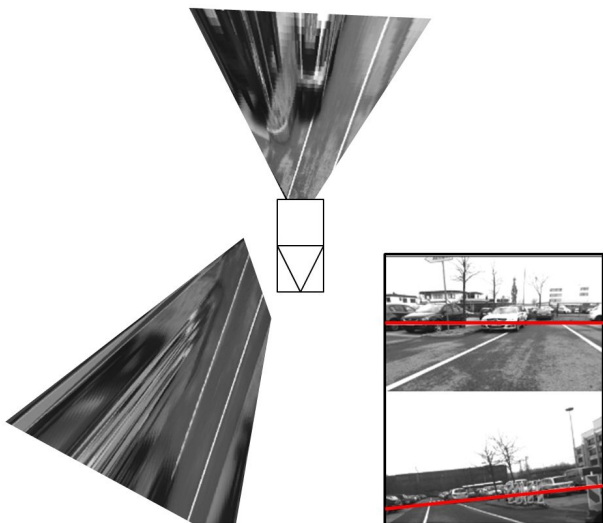


Figure 8. Screen shot of a generated bird's eye view from the test sequence. The two camera images are projected onto the ground plane with respect to the extrinsic parameters resulting from the motion adjustment. Bottom right: The two original images with horizontal line. It can be seen that the lines on the road are almost colinear.

with the intrinsic matrix \mathbf{K}_2 of \mathfrak{M}_2 , d is the second element of \mathbf{t}_w , $\mathbf{n} = \mathbf{r}_2$ and $\mathbf{H}_w = [\mathbf{c}_1 \mathbf{c}_3 \mathbf{t}_w]$ (for more details see [5]).

Fig. 8 shows a still image from the transformed test sequence as seen in Fig. 1 and 3 respectively. The orientation of the cameras as well as the aligned lane markers give a qualitative impression of the calibration result. There are remaining inaccuracies due to the lack of knowledge regarding the distance of the cameras from the ground.

Fig. 9 shows the computed scale factors for the rotational motions. It can be seen how the scaling fits the two tra-

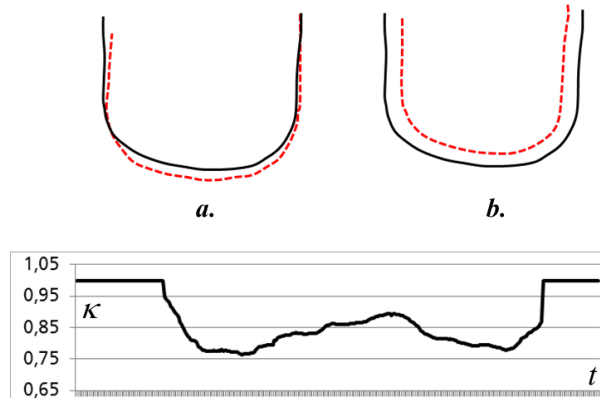


Figure 9. a. The two trajectories of the cameras with normalized translation vectors (black: reference \mathfrak{M}_1 , red dotted: slave camera module \mathfrak{M}_2). b. Rescaled trajectory after motion adjustment. Bottom: Diagram of the estimated scale factors $\hat{\kappa}_t$ for every rotational motion transformation $\mathcal{M}_{2,t}$.

jectories. The scaling entails the distances between the two camera trajectories to be $\approx \|\mathbf{t}\|$ at every time step t .

V. CONCLUSION AND FUTURE WORK

In this paper, a novel alternative computation of the hand-eye calibration principle for the case of a motion-based calibration of multiple cameras in vehicles has been proposed. The motion scales and the extrinsic camera parameters are estimated simultaneously, hence, the technique is called motion adjustment due to its relation to the well-known bundle adjustment. As could be shown with simulated and real data, motion adjustment is able to determine the extrinsic alignment up to scale. Although the experiments were performed for the planar case (as the motion of road vehicles is mainly planar), motion adjustment is capable of estimating

the full extrinsic 3D transformation up to scale. For increasing the robustness of the motion adjustment in online approaches, the motion adjustment will be embedded into a Kalman filter. Future work aims to integrate the motion adjustment into a global multi camera framework [14] with the purpose of continuously calibrating non-overlapping cameras in vehicles.

ACKNOWLEDGMENT

The author would like to thank Daimler AG, Germany, for kindly providing the video data sequences.

REFERENCES

- [1] G. Carrera, A. Angeli, A.J. Davison, "SLAM-based automatic extrinsic calibration of a multi-camera rig", *IEEE International Conference on Robotics and Automation ICRA*, 2011.
- [2] B. Clipp, J.H. Kim, J.M. Frahm, M. Pollefeys, R.I. Hartley, "Robust 6DOF Motion Estimation for Non-Overlapping, Multi-Camera Systems", in *Proceedings of the IEEE Workshop on Applications of Computer Vision*, 2008.
- [3] T. Dang, C. Hoffmann and C. Stiller, "Continuous stereo self-calibration by camera parameter tracking", *IEEE Transactions on Image Processing*, 2009.
- [4] S. Esquivel, F. Woelk, R. Koch, "Calibration of a Multi-camera Rig from Non-overlapping Views", *DAGM-Symposium*, 2007.
- [5] R.I. Hartley, A. Zisserman, "Multiple View Geometry in Computer Vision", *Cambridge University Press*, 2nd ed., 2004.
- [6] R. Horaud, F. Dornaika, "Hand-eye calibration", *International Journal of Robotics Research*, 1995.
- [7] B. Lamprecht, S. Rass, S. Fuchs, K. Kyamakya, "Extrinsic Camera Calibration for an On-Board Two Camera System Without Overlapping Field of View", in *Proceedings of the IEEE Intelligent Transportation Systems Conference ITSC*, 2007.
- [8] P. Lebraly, E. Royer, O. Ait-Aider, C. Deymier, M. Dhome, "Fast Calibration of Embedded Non-Overlapping Cameras", *IEEE International Conference on Robotics and Automation ICRA*, 2011.
- [9] M.I.A. Lourakis, A.A. Argyros, "SBA: A Software Package for Generic Sparse Bundle Adjustment", *ACM Transactions on Mathematical Software*, 2009.
- [10] A. F. Mordohai, A. Akbarzadeh, J. M. Frahm, P. Mordohai, C. Engels, D. Gallup, P. Merrell, M. Phelps, S. Sinha, B. Talton, L. Wang, Q. Yang, H. Stewenius, R. Yang, G. Welch, H. Towles, D. Nister, M. Pollefeys, "Towards Urban 3D Reconstruction From Video", in *3DPVT*, 2006.
- [11] R.A. Newcombe, A.J. Davison, "Live Dense Reconstruction with a Single Moving Camera", in *Proceedings of IEEE Conference of Computer Vision and Pattern Recognition CVPR*, 2010.
- [12] W. Press, S. Teukolsky, W. Vetterling, B. Flannery, "Numerical Recipes in C", 2nd ed. *Cambridge, UK: Cambridge University Press*, 1992.
- [13] F. Pagel, "Robust Monocular Egomotion Estimation Based on an IEKF", in *Proceedings of the IEEE IEEE Canadian Conference on Computer and Robot Vision CRV*, 2009.
- [14] F. Pagel, D. Willersinn, "Motion-based Online Calibration for Non-overlapping Camera Views", in *Proceedings of the IEEE Intelligent Transportation Systems Conference ITSC*, 2010.
- [15] F. Pagel, D. Willersinn, "Extrinsic Camera Calibration in Vehicles with Explicit Ground Estimation", *Int. Workshop on Intelligent Transportation WIT*, 2011.
- [16] T. Ruland, H. Loose, T. Pajdla, L. Kruger, "Hand-Eye Auto-calibration of Camera Positions on Vehicles", in *Proceedings of the IEEE Intelligent Transportation Systems Conference ITSC*, 2010.
- [17] R. Y. Tsai, R. K. Lenz, "A New Technique for Fully Autonomous and Efficient 3D Robotics Hand/Eye Calibration", in *IEEE Transactions on Robotics and Automation*, 1989.