Extrinsic Multi-Sensor Calibration For Mobile Robots Using the Gauss-Helmert Model

Kaihong Huang Cyrill Stachniss

***Abstract*— Most state estimation procedures in mobile robotics require information about the locations of the individual sensors on the platform. This is especially the case when estimating geometric properties about the environment or the robot. In this paper, we present an novel approach to extrinsic, multi-sensor calibration. Our approach seeks to find the extrinsic parameters by maximizing the consistency of the motion estimate that is computed from different sensing modalities. Our approach formulates constraints between the movements of the sensors using the Gauss-Helmert model. This allows us to compute a more accurate solution than the ordinary least squares approach. We implemented our approach and tested it using multiple sensors as well as when using a motion capture studio. The experiments presented in this paper show that our approach is able to accurately determine the extrinsic configuration of each sensor and is robust enough for typically applications in mobile robotics.**

I. Introduction

Most mobile robots perform some form of state estimation such as localization, mapping, SLAM, or exploration. For most of such tasks, it is important to know where the individual sensors are mounted on the robot. The task of determining the position and orientation of a sensor is often referred to extrinsic calibration. Without such calibration information, a lot of the estimation tasks, especially those related to computing geometric models such as SLAM, do not work properly and/or provide suboptimal results. Such extrinsic calibration is also important when the information from multiple sensors have to be fused.

In this paper, we address the problem of estimating the extrinsic parameters of multiple sensors relative to each other or with respect to a base frame of a mobile robot. Our calibration approach is based on the robot's motion and does not rely on external markers, calibration patterns, or a known environment. We assume that each sensing modality allows for estimating a (relative) trajectory of the sensor given the observations. This holds for cameras by using visual odometry, for laser scanners or RGB-D cameras through scan matching, and obviously for odometry as well as GPS receivers. We do not address calibration for sensors that do not have this property such as a bumper sensor.

Our approach exploits constraints between the motions of individual sensors and formulates the resulting error minimization problem using the Gauss-Helmert model [18]. This allows us to exploit constraints resulting from the fact that the sensors are rigidly mounted on the platform while being able to handle noisy observations and compute a weighted least squares solution. This leads to a statistically sound estimation providing accurate extrinsic calibration parameters for multiple sensors and has the potential to yield better results than traditional least squares.

The main contribution of this paper is a novel motion-based calibration algorithm for multiple sensors using the Gauss-Helmert formulation. Our approach makes the assumptions that the sensors are rigidly attached to the robot and that a relative trajectory estimate can be obtained from each sensor. Under these assumptions, our approach (i) accurately determines the extrinsic calibration parameters, (ii) requires an initial guess with an accuracy that can be obtained through a direct approach, (iii) obtains more accurate results compared to ordinary least squares using the Gauss-Markov model, and (iv) can be executed in a reasonable amount of time to be useful for real world applications. These four claims are backed up through the paper and its experimental evaluation.

II. Related Work

The problem of multi-sensor extrinsic calibration addressed in this paper is strongly related to the hand-eye calibration problem [14], [15], [17]. Here, the unknown transformation between a robotic arm and a rigidly mounted camera is estimated given a series of relative pose measurements. Earlier work by Shiu and Ahmad [14] describes this problem as solving a homogeneous transform equation in the form of $AX = XB$, which relates the relative motions (A, B) of two devices to the sensors' relative transformation (X). Several efficient closed form solutions are available, by decoupling the rotation and translation estimation [14], [17], or by using dual-quaternions formulation to allow for the estimation of rotation and translation simultaneously [4]. The direct solutions are comparably simple and fast to compute, but they are also sensitive to noise. To improve robustness and accuracy, Horaud and Dornaika [5] propose to jointly optimize rotation and translation with nonlinear optimization. A subsequent approach [15] proposes a metric on the special Euclidean group SE(3) and a corresponding error model for nonlinear optimization. Finally, Zhao *et al.* [19] propose using a L_∞ cost function and utilize convex optimization approach instead of common least squares formulation.

Beside general hand eye problem, there are also works focused on special calibration problems. Driven by the needs

All authors are with the University of Bonn, Institute of Geodesy and Geoinformation, Bonn, Germany.

This work has partly been supported by the DFG under the grant number FOR-1015: Mapping on Demand as well as by the European Commission under the Horizon 2020 framework programme under grant agreement no. 645403. Furthermore, we received support by the China Scholarship Council CSC.

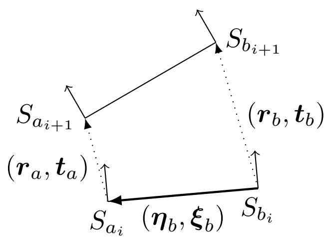

Fig. 1: Relative movement at time i. Given rotation and translation measurements $(\boldsymbol{r}_a, \boldsymbol{t}_a)$ and $(\boldsymbol{r}_b, \boldsymbol{t}_b)$ from two sensors, our task is to find an optimal estimation of parameter $(\boldsymbol{\eta}_b, \boldsymbol{\xi}_b)$, which is the relative orientation and displacement of sensor A in B's frame.

of information fusion between camera and odometry, camera-odometry calibration for cars and wheeled mobile robots is one of the many [7], [8], [3], [13]. Explicit motion-based sensor calibration has also been addressed by Taylor and Nieto [16] and also combined with the SLAM problem [9]. Guo *et al.* [7] proposed a two-step analytical least squares solution to estimate rotation and translation separately, without using any special hardware or known landmarks. The approach by Heng *et al.* follow the formulation of Guo *et al.* [7], but only use the direct solution as initial guess to an extensive least squares estimation problem including camera and odometry poses, intrinsic and extrinsic parameters, as well as 3D scene points. In contrast to that, Schneider *et al.* [13] reported an online recursive estimation approach using Unscented Kalman filter.

To the best of our knowledge, no other work exists that bases the optimization for motion-based calibration on the Gauss-Helmert model, which jointly optimize the model parameter and observations together, thereby taking the observation noise into full account.

III. Motion-Based Extrinsic Calibration

The goal of motion-based extrinsic sensor calibration is to determine the position and orientation of each sensor without markers and only based on the motion of the platform. The calibration is done with respect ot a reference point, whereas such reference point can be the robot's odometry frame or one of the sensors.

Consider a robot that is equipped with M sensors, which are rigidly attached to the robot. We index these sensors with $a, b, \ldots, m$ and each sensor is assumed to provide a (noisy) relative motion estimate of the sensor. More precisely, at each time-step i, we have a set of relative pose measurements consisting of angle-axis rotation $\boldsymbol{r}$ and translation $\boldsymbol{t}$ for each sensor:

$$\{(\boldsymbol{r}_{ai}, \boldsymbol{t}_{ai}), \ldots, (\boldsymbol{r}_{mi}, \boldsymbol{t}_{mi}) \mid \mathsf{R}(\boldsymbol{r}) \in \mathrm{SO}(3), \boldsymbol{t} \in \mathbb{R}^3\}, \quad (1)$$

with $i = 1, \ldots, N$. Each motion is expressed relatively in an ego-centric frame for each sensor, i.e. the sensor pose of the previous timestep.

The objective is to estimate the fixed but unknown transformations between all the sensors. Without loss of generality, we define the sensor a as the base sensor and estimate for each other sensor $s \in \{b, \ldots, m\}$ the relative rotation and translation $(\boldsymbol{\eta}_s, \boldsymbol{\xi}_s)$ of S_{a_i} under the frame of S_{b_i}.

$$\{(\boldsymbol{\eta}_s, \boldsymbol{\xi}_s) \mid \mathsf{R}(\boldsymbol{\eta}_s) \in \mathrm{SO}(3),\ \boldsymbol{\xi}_s \in \mathbb{R}^3,\ s = b...m\} \quad (2)$$

If we choose S_a to be the base/odometry frame of the robot, the tuples $(\boldsymbol{\eta}_s, \boldsymbol{\xi}_s)$ are a solution to the extrinsic multi-sensor calibration problem. A point in S_a, written as ${}^a\boldsymbol{p}$, can be transfered to the sensor frame S_b by:

$${}^b\boldsymbol{p} = \mathsf{R}(\boldsymbol{\eta}_b)^a\boldsymbol{p} + \boldsymbol{\xi}_b. \quad (3)$$

Fig. 1 (left) depicts the geometry for this problem. Successive poses of a sensor pair form a virtual quadrilateral or vector loop, which leads to the hand-eye problem equation:

$$\boldsymbol{\xi}_b + \mathsf{R}(\boldsymbol{\eta}_b)\boldsymbol{t}_a - \mathsf{R}(\boldsymbol{r}_b)\boldsymbol{\xi}_b - \boldsymbol{t}_b = \mathit{0}_3 \quad (4)$$

$$\mathsf{R}(\boldsymbol{\eta}_b)\mathsf{R}(\boldsymbol{r}_a)\mathsf{R}(\boldsymbol{\eta}_b)^\mathsf{T}\mathsf{R}(\boldsymbol{r}_b)^\mathsf{T} = \mathit{I}_3. \quad (5)$$

IV. Problem of Ordinary Least Squares for Motion-Based Calibration

The solution provided by closed form method are usually far from perfect due to the decoupling and simplified noise assumption (see [6] p.179). It is therefore necessary to employ an iterative refinement process by means of least squares estimation, which not only jointly optimizes $\boldsymbol{\xi}, \boldsymbol{\eta}$ but also take the full uncertainty relations into account.

A widely used method is to iteratively minimize a weighted sum of a squared error function, e.g. [7], [15]. In our case it is:

$$\operatorname*{argmin}_{\boldsymbol{\eta}, \boldsymbol{\xi}} \sum_i \|\boldsymbol{c}_{ti}\|^2_{w_{ti}} + \|\boldsymbol{c}_{ri}\|^2_{w_{ri}} \quad (6)$$

with

$$\boldsymbol{c}_{ti} := (\mathit{I}_3 - \mathsf{R}(\boldsymbol{r}_{bi}))\boldsymbol{\xi} + \mathsf{R}(\boldsymbol{q})\boldsymbol{t}_{ai} - \boldsymbol{t}_{bi} \quad (7)$$

$$\boldsymbol{c}_{ri} := \log\left(\mathsf{R}(\boldsymbol{\eta})\mathsf{R}(\boldsymbol{r}_{ai})\mathsf{R}(\boldsymbol{\eta})^\mathsf{T}\mathsf{R}(\boldsymbol{r}_{bi})^\mathsf{T}\right). \quad (8)$$

where $\|\boldsymbol{c}\|^2_w := \boldsymbol{c}^\mathsf{T}\mathsf{W}\boldsymbol{c}$ and W being a positive definite weight matrix. The form of $\boldsymbol{c}_{ri}$ may varies, alternatives can be a Frobenius norm of a full rotation matrix, or of a quaternion [15].

A minimizing of this cost function is easy to implement with the help of popular off-the-self optimizer such as g2o[10] or Ceres[1] and therefore it is prominent in the current literature of calibration problems. An overlooked point in most works are the values of the weight matrices $\mathsf{W}_{ri}, \mathsf{W}_{ti}$ in this formula. And as we will see in the following discussion, there is a hidden assumption in this optimization model, which will degrade the estimation accuracy when the measurement noise-level is high, and therefore only suitable for low noise situations.

Eq. (6)-(8) is a approach based on the Gauss-Markov model (GM for short). The principle behind the GM model is the "Gauss-Markov" theorem, which says the least squares estimate gives no bias and has minimal variance if the functional relation $\boldsymbol{l} = \boldsymbol{f}(\boldsymbol{x})$ is linear and the weights are chosen to be $\mathsf{W} = \Sigma_{ll}^{-1}$. The Gauss-Markov theorem also holds approximately for nonlinear function $\boldsymbol{f}$ if the variances of the observations are small compared to the second derivatives of the functions (see [6] p.79). For the GM model it is important that the observations ($\boldsymbol{l}$) has to be explicit functions of the unknown parameters ($\boldsymbol{x}$), to guarantee statistical optimality.

To fit in the Gauss-Markov theorem, we see that the GM model is actually using the residual vectors $\boldsymbol{c}_{ti}$ and $\boldsymbol{c}_{ri}$ as their observation entities $\boldsymbol{l}$ rather than the original measurements $\{\boldsymbol{r}_{si}, \boldsymbol{t}_{si}\}$. So the weight matrix W_{ri}, W_{ti} should therefore take the inverse of the *covariance matrix of the residual vectors*, i.e. $W_{ri} = \Sigma_{c_r}^{-1}$, $W_{ti} = \Sigma_{c_t}^{-1}$. Since $\boldsymbol{c}_{ti}$ and $\boldsymbol{c}_{ri}$ are functions of $\{\boldsymbol{r}_{si}, \boldsymbol{t}_{si}\}$, we can apply error propagation to Eq. (8)-(7) and obtain

$$\Sigma_{c_r} \approx J_r \Sigma_{rr} J_r^\mathsf{T}, \quad \Sigma_{c_t} \approx J_t \Sigma_{tt} J_t^\mathsf{T} + J_r \Sigma_{rr} J_r^\mathsf{T} \tag{9}$$

where $J_r := \frac{\partial \boldsymbol{c}_{ri}}{\partial \boldsymbol{r}_i}$ and $J_t := \frac{\partial \boldsymbol{c}_{ti}}{\partial \boldsymbol{t}_i}$ are the Jacobians, and Σ_{tt}, Σ_{rr} are the covariance matrices of the measurement noise. Given the correct weight matrices, we can proceed with minimizing Eq. (6) and refine $\boldsymbol{x}$ iteratively.

An issue to point out is that *($\boldsymbol{r}, \boldsymbol{t}$) is fixed during the whole estimation process and so are the Jacobians and the linearized model*. If there are significant errors in the measurement, they will persist and result in a degraded estimation accuracy. This degradation is what we observed in the real world application and thus motivated the work of this paper, its effect is confirmed by our experiment in Sec. VI.

V. Gauss-Helmert-Based Approach to Multi-Sensor Calibration

To tackle the aforementioned degradation problem in high noise situation, we propose to formulate the estimation problem using the Gauss-Helmert model [18], which makes corrections to not only the unknown parameters but also the observations at each iteration.

The Gauss-Helmert model is more general than the GM model, it allow us to handle constraints between observation $\boldsymbol{l}$ and unknown parameters $\boldsymbol{x}$ by implicit functions of the form $\boldsymbol{g}(\boldsymbol{x}, \boldsymbol{l}) = \boldsymbol{0}$. It is closely related to the total-least-squares method or the errors-in-variables method developed in statistics. Instead of minimizing residuals in the constraints, the Gauss-Helmert model strictly enforces the constraints and makes corrections to not only the unknown parameters but also the observations. The objective is then to minimize the corrections to the observation while estimating the unknown parameters.

Gauss-Markov and Gauss-Helmert model are equivalent for problems in which observations are explicit functions of the unknown parameters, e.g $\boldsymbol{l} = \boldsymbol{f}(\boldsymbol{x})$, because in those cases the corrections to the parameters $\boldsymbol{x}$ are essentially corrections to the observation and so all the constraints are automatically satisfied. However, Eq. (4) and Eq. (5) for $(\boldsymbol{\xi}, \boldsymbol{\eta})$ are obviously not in this explicit form.

A. Model for Multiple Sensors

Most related works consider only the formulation for two sensor, assuming we can simply apply the 2-sensor case to each sensor pair in case there are more than two sensors. But as we will see in the following discussion, explicitly formulating the multi-sensor case will give us the insight that there are actually interactions between sensor pairs, which is a good reason for us to introduce high-accuracy external measurement in the calibration process, which will help improving the overall accuracy.

First, we denote the unknown parameters collectively as $\boldsymbol{x}$ and known observations as $\boldsymbol{l}$:

$$\boldsymbol{x} := \begin{bmatrix} \boldsymbol{\eta}_b \\ \vdots \\ \boldsymbol{\eta}_m \\ \boldsymbol{\xi}_b \\ \vdots \\ \boldsymbol{\xi}_m \end{bmatrix}, \quad \boldsymbol{l}_i := \begin{bmatrix} \boldsymbol{r}_{ai} \\ \vdots \\ \boldsymbol{r}_{mi} \\ \boldsymbol{t}_{ai} \\ \vdots \\ \boldsymbol{t}_{mi} \end{bmatrix} \quad i = 1, \ldots, N \tag{10}$$

again N is the number of measurement groups/motion segments considered. From Eq. (4), we can formulate the following constraints for the translation:

$$\boldsymbol{g}_1(\boldsymbol{x}, \boldsymbol{l}) := \begin{bmatrix} (I_3 - R(\boldsymbol{r}_b))\boldsymbol{\xi}_b + R(\boldsymbol{\eta}_b)\boldsymbol{t}_a - \boldsymbol{t}_b \\ \vdots \\ (I_3 - R(\boldsymbol{r}_m))\boldsymbol{\xi}_m + R(\boldsymbol{\eta}_m)\boldsymbol{t}_a - \boldsymbol{t}_m \end{bmatrix} = \boldsymbol{0} \tag{11}$$

and for rotation[1]

$$\boldsymbol{g}_2(\boldsymbol{x}, \boldsymbol{l}) := \begin{bmatrix} R(\boldsymbol{\eta}_b) & -I_3 & & \\ \vdots & & \ddots & \\ R(\boldsymbol{\eta}_m) & & & -I_3 \end{bmatrix} \begin{bmatrix} \boldsymbol{r}_a \\ \vdots \\ \boldsymbol{r}_m \end{bmatrix} = \boldsymbol{0}. \tag{12}$$

Expressed as one function, this leads to

$$\boldsymbol{g}(\boldsymbol{x}, \boldsymbol{l}) := \begin{bmatrix} \boldsymbol{g}_1(\boldsymbol{x}, \boldsymbol{l}) \\ \boldsymbol{g}_2(\boldsymbol{x}, \boldsymbol{l}) \end{bmatrix} = \boldsymbol{0}. \tag{13}$$

Due to the existence of noise, Eq. (13) generally does not hold for the "raw" measurements (now called $\boldsymbol{l}_i^0$). Thus, corrections on the measurements (and not only to the parameters as in the Gauss-Markov model) must be made so that they can fulfill the constraints. This can be expressed through

$$\boldsymbol{g}(\boldsymbol{x}^*, \boldsymbol{\epsilon}_i + \boldsymbol{l}_i^0) = \boldsymbol{0}, \tag{14}$$

with the corrections $\boldsymbol{\epsilon}_i$.

The goal is to find the optimal $\boldsymbol{x}^*$ through minimizing these corrections while the constraints being fulfilled and the uncertainty of the observations expressed though Σ_{ll}^{-1} being taken into account:

$$\boldsymbol{x}^* = \underset{\boldsymbol{x}, \{\boldsymbol{\epsilon}_i\}}{\operatorname{argmin}} \frac{1}{2} \sum_i^N \|\boldsymbol{\epsilon}_i\|^2_{\Sigma_{ll}^{-1}}$$
$$\text{s.t}\ \boldsymbol{g}(\boldsymbol{x}, \boldsymbol{\epsilon}_i + \boldsymbol{l}_i^0) = \boldsymbol{0},\ \forall i. \tag{15}$$

Similar to ordinary least squares, we cannot solve the nonlinear problem directly, we must linearize it around an initial value and solve it iteratively.

[1] For sensors that provide no rotation measurement, like GPS, their constraints in $\boldsymbol{g}_2$ are simply omitted. We can still recover their relative pose parameter ($\boldsymbol{\eta}$) from $\boldsymbol{g}_1$ constraint, which aligns the translation measurement from different sensors. For example, when in pure translation $\|\boldsymbol{r}_b\| = 0$, $\boldsymbol{g}_1$ becomes $R(\boldsymbol{\eta}_b)\boldsymbol{t}_a = \boldsymbol{t}_b$ and has the exact form of $\boldsymbol{g}_2$.

Assume that in the k-th iteration, we have the estimated measurements as well as parameters $\boldsymbol{l}_i^k, \boldsymbol{x}^k$ and update them by

$$\boldsymbol{l}_i^{k+1} = \Delta \boldsymbol{l}_i + \boldsymbol{l}_i^k \quad \text{and} \quad \boldsymbol{x}^{k+1} = \Delta \boldsymbol{x} + \boldsymbol{x}^k. \tag{16}$$

Eq. (14) can be approximated around $(\boldsymbol{x}^k, \boldsymbol{l}_i^k)$ through linearization by:

$$\boldsymbol{g}(\boldsymbol{x}^k, \boldsymbol{l}_i^k) + \mathsf{A}_i^k \Delta \boldsymbol{x} + \mathsf{B}_i^k \Delta \boldsymbol{l}_i = \boldsymbol{0}, \tag{17}$$

with $\mathsf{A}_i^k, \mathsf{B}_i^k$ being the Jacobians with respect to the parameters (A_i^k) and to the observations (B_i^k). Thus, after linearization, our optimization problem in Eq. (15) becomes

$$\begin{aligned} &\underset{\{\Delta \boldsymbol{x}, \Delta \boldsymbol{l}_i\}}{\operatorname{argmin}} \frac{1}{2} \sum_i \|\Delta \boldsymbol{l}_i + \boldsymbol{\epsilon}_i^k\|^2_{\Sigma_{ll}^{-1}} \\ &\text{s.t. } \boldsymbol{g}(\boldsymbol{x}^k, \boldsymbol{l}_i^k) + \mathsf{A}_i^k \Delta \boldsymbol{x} + \mathsf{B}_i^k \Delta \boldsymbol{l}_i = \boldsymbol{0}, \quad \forall i. \end{aligned} \tag{18}$$

The terms $\boldsymbol{\epsilon}_i^k := \boldsymbol{l}_i^k - \boldsymbol{l}_i^0$ are the corresponding k-th iteration *measurement-errors*. The solution to Eq. (18) for the k-th iteration is:

$$\Delta \boldsymbol{x} = \sum_i \left(\sum_i \mathsf{A}_i^\mathsf{T} \mathsf{W}_i \mathsf{A}_i \right)^{-1} \mathsf{A}_i^\mathsf{T} \mathsf{W}_i \boldsymbol{c}_i \tag{19}$$

$$\Delta \boldsymbol{l}_i = \Sigma_{ll} \mathsf{B}_i^\mathsf{T} \mathsf{W}_i (\boldsymbol{c}_i - \mathsf{A}_i \Delta \boldsymbol{x}) - \boldsymbol{\epsilon}_i^k \tag{20}$$

with

$$\mathsf{W}_i := (\mathsf{B}_i \Sigma_{ll} \mathsf{B}_i^\mathsf{T})^{-1} \text{ and } \boldsymbol{c}_i := -\boldsymbol{g}(\boldsymbol{x}^k, \boldsymbol{l}_i^k) + \mathsf{B}_i \boldsymbol{\epsilon}_i^k. \tag{21}$$

This allows us to update the estimate

$$\boldsymbol{x}^{k+1} := \boldsymbol{x}^k + \Delta \boldsymbol{x}, \quad \boldsymbol{l}_i^{k+1} := \boldsymbol{l}_i^k + \Delta \boldsymbol{l}_i \tag{22}$$

and repeat the process until convergence. Finally, the theoretical precision of the optimal $\boldsymbol{x}^*$ is given by

$$\Sigma_{\boldsymbol{xx}} := \left(\sum_i \mathsf{A}_i^\mathsf{T} (\mathsf{B}_i \Sigma_{ll} \mathsf{B}_i^\mathsf{T})^{-1} \mathsf{A}_i \right)^{-1}. \tag{23}$$

The solution obtained here is *best* and *unbiased*, given the observation errors are normally distributed, see [2]. The term best means that the solution has minimum variance compared to all other quadratic-based unbiased estimators. A detail discussion of the necessary and sufficient conditions for the existence of a unique solution for both the observation correction as well as the estimated parameter can be found at [11].

B. Exploiting Inter-Sensor Constraints

The constraints in Eq. (12) have a simpler form than those in Eq. (5) due to the angle-axis representation of our approach. Note that Eq. (12) not only provides the information to estimate the $\boldsymbol{\eta}$, but also an inter-sensor constraints on the rotation measurement between sensors, due to:

$$\theta_i := \|\boldsymbol{r}_{bi}\| = \|\mathsf{R}_b \boldsymbol{r}_{ai}\| = \|\boldsymbol{r}_{ai}\| = \cdots = \|\boldsymbol{r}_{mi}\|. \tag{24}$$

This means that rotations within the same time step i will be corrected to have the same magnitude θ_i, averaged by their uncertainty level. This provide a very valuable insight of how to improve the overall calibration accuracy. This can be done by introducing accurate rotational measurement, even if it is only temporary for the calibration.

For this reason, we can use an temporary external measurement with a high rotational accuracy while performing calibration.

C. Global Optimality for Multiple Sensors

In the constraint formulation in Eq. (11) and Eq. (12), we connect all the sensors ($s = b, \ldots, m$) to a base sensor (a) in a star network fashion and thereby form one joint optimization problem. Despite the fact that the network is not fully connected, we claim that this does not impact optimality. Let us refer to sensor a as the root and sensors $s = b, \ldots, m$ as leaves. As shown below, we can proof that once the root-to-leaf constraints are fulfilled, all leaf-to-leaf constraints are fulfilled automatically. Therefore, the star topology does not impact optimality.

To be more clear, we extent the notation used so far and denote the original parameter $(\boldsymbol{\eta}_s, \boldsymbol{\xi}_s)$ as $(\boldsymbol{\eta}_{sa}, \boldsymbol{\xi}_{sa})$ for $s = b, \ldots, m$ as this allow us to specify the (unnecessary) constraints between leaves.

After the optimization, the root-to-leaf constraints are fulfilled be definition. Thus, for all sensors $s = b, \ldots, m$ holds:

$$\boldsymbol{\xi}_{sa} + \mathsf{R}(\boldsymbol{\eta}_{sa}) \boldsymbol{t}_a - \mathsf{R}(\boldsymbol{r}_s) \boldsymbol{\xi}_{sa} - \boldsymbol{t}_s = 0 \tag{25}$$

$$\mathsf{R}(\boldsymbol{\eta}_{sa}) \mathsf{R}(\boldsymbol{r}_a) \mathsf{R}^\mathsf{T}(\boldsymbol{\eta}_{sa}) \mathsf{R}^\mathsf{T}(\boldsymbol{r}_s) = \mathsf{I}_3. \tag{26}$$

Without loss of generality, let us consider the leaf nodes m and b. If we set

$$\begin{aligned} \mathsf{R}(\boldsymbol{\eta}_{bm}) &= \mathsf{R}(\boldsymbol{\eta}_{ba}) \mathsf{R}^\mathsf{T}(\boldsymbol{\eta}_{ma}) \\ \boldsymbol{\xi}_{bm} &= \boldsymbol{\xi}_{ba} - \mathsf{R}(\boldsymbol{\eta}_{bm}) \boldsymbol{\xi}_{ma}, \end{aligned}$$

being the transformation between b and m, then we can prove that following equations are fulfilled:

$$\boldsymbol{\xi}_{bm} + \mathsf{R}(\boldsymbol{\eta}_{bm}) \boldsymbol{t}_m - \mathsf{R}(\boldsymbol{r}_b) \boldsymbol{\xi}_{bm} - \boldsymbol{t}_b = 0 \tag{27}$$

$$\mathsf{R}(\boldsymbol{\eta}_{bm}) \mathsf{R}(\boldsymbol{r}_m) \mathsf{R}^\mathsf{T}(\boldsymbol{\eta}_{bm}) \mathsf{R}^\mathsf{T}(\boldsymbol{r}_b) = \mathsf{I}_3. \tag{28}$$

First, we proof that Eq. (28) holds. We start from Eq. (26) and set s to b and m respectively to obtain

$$\begin{aligned} \mathsf{R}(\boldsymbol{r}_b) \mathsf{R}(\boldsymbol{\eta}_{ba}) &= \mathsf{R}(\boldsymbol{\eta}_{ba}) \mathsf{R}(\boldsymbol{r}_a) \\ \mathsf{R}^\mathsf{T}(\boldsymbol{\eta}_{ma}) \mathsf{R}(\boldsymbol{r}_m) &= \mathsf{R}(\boldsymbol{r}_a) \mathsf{R}^\mathsf{T}(\boldsymbol{\eta}_{ma}). \end{aligned}$$

Thus, we start from the left hand side of Eq. (28) and obtain

$$\begin{aligned} &\mathsf{R}(\boldsymbol{\eta}_{bm}) \mathsf{R}(\boldsymbol{r}_m) - \mathsf{R}(\boldsymbol{r}_b) \mathsf{R}(\boldsymbol{\eta}_{bm}) \\ &= \mathsf{R}(\boldsymbol{\eta}_{ba}) \mathsf{R}^\mathsf{T}(\boldsymbol{\eta}_{ma}) \mathsf{R}(\boldsymbol{r}_m) - \mathsf{R}(\boldsymbol{r}_b) \mathsf{R}(\boldsymbol{\eta}_{ba}) \mathsf{R}^\mathsf{T}(\boldsymbol{\eta}_{ma}) \\ &= \mathsf{R}(\boldsymbol{\eta}_{ba}) \mathsf{R}(\boldsymbol{r}_a) \mathsf{R}^\mathsf{T}(\boldsymbol{\eta}_{ma}) - \mathsf{R}(\boldsymbol{\eta}_{ba}) \mathsf{R}(\boldsymbol{r}_a) \mathsf{R}^\mathsf{T}(\boldsymbol{\eta}_{ma}) \\ &= \boldsymbol{0}. \end{aligned}$$

Thus, the first part has been proven.

Second, we continues with Eq. (27) starting with Eq. (25) and obtain similar to above

$$\begin{aligned} \boldsymbol{\xi}_{ba} - \mathsf{R}(\boldsymbol{r}_b) \boldsymbol{\xi}_{ba} - \boldsymbol{t}_b &= -\mathsf{R}(\boldsymbol{\eta}_{ba}) \boldsymbol{t}_a \\ \boldsymbol{t}_m - \boldsymbol{\xi}_{ma} &= \mathsf{R}(\boldsymbol{\eta}_{ma}) \boldsymbol{t}_a - \mathsf{R}(\boldsymbol{r}_m) \boldsymbol{\xi}_{ma}. \end{aligned}$$

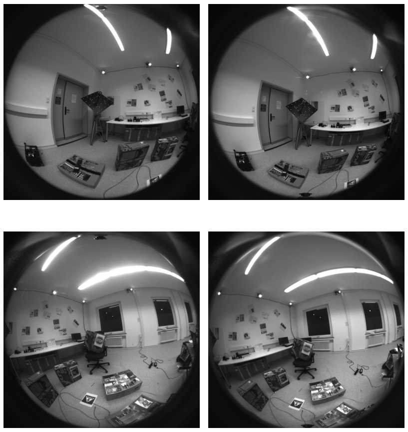

Fig. 2: Examples stereo images pairs used for calibration.

We start from the left hand side of Eq. (27) and expand $\boldsymbol{\xi}_{bm}$, which leads to

$$
\begin{aligned}
&\boldsymbol{\xi}_{bm} - R(\boldsymbol{r}_b)\boldsymbol{\xi}_{bm} + R(\boldsymbol{\eta}_{bm})\boldsymbol{t}_m - \boldsymbol{t}_b \\
&= [\boldsymbol{\xi}_{ba} - R(\boldsymbol{\eta}_{bm})\boldsymbol{\xi}_{ma}] + R(\boldsymbol{\eta}_{bm})\boldsymbol{t}_m - R(\boldsymbol{r}_b)[\boldsymbol{\xi}_{ba} - R(\boldsymbol{\eta}_{bm})\boldsymbol{\xi}_{ma}] - \boldsymbol{t}_b \\
&= [\boldsymbol{\xi}_{ba} - R(\boldsymbol{r}_b)\boldsymbol{\xi}_{ba} - \boldsymbol{t}_b] + R(\boldsymbol{\eta}_{bm})[\boldsymbol{t}_m - \boldsymbol{\xi}_{ma}] + R(\boldsymbol{r}_b)R(\boldsymbol{\eta}_{bm})\boldsymbol{\xi}_{ma} \\
&= -R(\boldsymbol{\eta}_{ba})\boldsymbol{t}_a + R(\boldsymbol{\eta}_{bm})[R(\boldsymbol{\eta}_{ma})\boldsymbol{t}_a - R(\boldsymbol{r}_m)\boldsymbol{\xi}_{ma}] + R(\boldsymbol{r}_b)R(\boldsymbol{\eta}_{bm})\boldsymbol{\xi}_{ma} \\
&= [R(\boldsymbol{r}_b)R(\boldsymbol{\eta}_{bm}) - R(\boldsymbol{\eta}_{bm})R(\boldsymbol{r}_m)]\boldsymbol{\xi}_{ma} \\
&= \mathbf{0}.
\end{aligned}
$$

Thus, the second part has been proved. As a result, the fulfillment of the root-to-leaf constraints automatically satisfies all leaf-to-leaf constraints and thus they do not need to be modeled explicitly and thus the star topology does not impact optimality.

VI. Experimental Results

The experiments are designed to show the capabilities of our method and to support our key claims. We claim that our approach (i) accurately determines the extrinsic calibration parameters, (ii) requires an initial guess with a level of accuracy that can be provided by a direct method, (iii) obtains more accurate results compared to ordinary least squares using the Gauss-Markov model, and (iv) can be executed in a reasonable amount of time to be useful for real world applications. We perform the evaluations on own real-world as well as simulated datasets to support these claims.

A. Real World and Simulated Data

To evaluate our approach, we use simulated as well as real world data. For the real world setup, we use a UAV with two stereo camera pairs (called A and B later on), one pointing forward and one pointing backwards. Example images of this calibration data can be found in Fig. 2. In addition to that, we place markers of a motion capture studio on the UAV to simulate an additional sensor that has be calibrated.

In the experiment, we recorded a total of $N = 1655$ relative motion-segments estimated through our own visual odometry pipeline [12]. Around 25 of the segments are outliers as the visual odometry lost track. The maximum rotation angle of the inlier data is $7.6°$ and the data is recorded at 20 fps. Fig. 3 depicts a histogram of the rotation angle and translation magnitude of the gathered data.

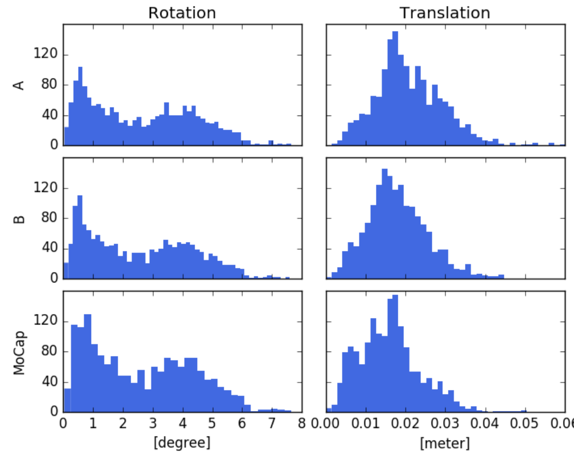

Fig. 3: Histogram of the rotation and translation magnitude of our real world dataset. Left column shows the rotation ($\|\boldsymbol{r}\|$, in degree), and right column shows the translation ($\|\boldsymbol{t}\|$, in meters). From top to bottom, each rows represent stereo pair A, pair B and the motion capture studio respectively.

TABLE I: Standard deviation of the measurement noise

	σ_r [°]	σ_t [mm]
Stereo pair A	0.0286	2
Stereo pair B	0.0286	3
Mocap	0.573	0.2

We first obtain an initial guess by the SVD-based direct method (e.g, [7]) then run our approach to obtain an improved solution as well as the corresponding theoretical precision. The covariance matrices are heuristically set to $\Sigma_{rr} = \sigma_r I_3$ and $\Sigma_{tt} = \sigma_t I_3$ with σ_r, σ_t given in Tab. I. We also assume the noise are independent and identically distributed random variables.

B. Accuracy Comparison

The first set of experiments is designed to show that our approach computes the calibration parameters more accurately than an SVD-based solution and than a typical least squares solution based on the Gauss-Markov model.

We first performed the calibration on the real world data, and our GH approach converged at the 4th iterations. The theoretical precision given by our approach is depicted in Tab. II. They are the square roots of the main diagonal of the covariance matrix in Eq. (23).

As this is a real world experiment, the ground truth is unfortunately not available, so we cannot make an ground truth comparison. But judging from the measurement-residual distribution (shown in Fig. 4) given by our model after the estimation, it is clear that i) the Gaussian noise assumption holds for our dataset to larger extent, ii) our constraint model is correct and there is no bias in the estimation, otherwise the residual histograms will not be symmetrically centered around zero. Therefore we have a good reason to believe that the theoretical precision given by our approach is plausible. To provide a more quantitative experiments, we performed

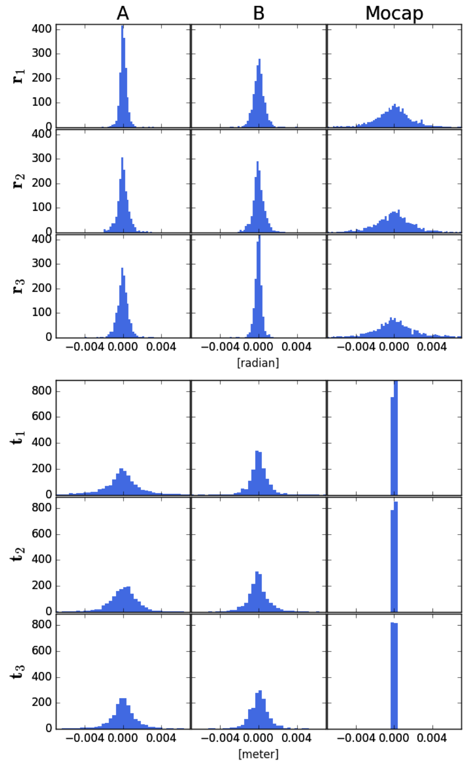

Fig. 4: Distribution of measurement-residuals $\boldsymbol{\epsilon}$ after the refinement by our approach. Columns from left to right are for stereo A, B and Mocap respectively. Rows from top to bottom are the 6 measurement components, namely $\boldsymbol{r}_1, \boldsymbol{r}_2, \boldsymbol{r}_3$ for rotation and $\boldsymbol{t}_1, \boldsymbol{t}_2, \boldsymbol{t}_3$ for translation. It is clear that the Gaussian noise assumption holds for our real world dataset to a large extent, and our constraint model is correct since there is no bias.

the analysis of the accuracy in simulation, which is the second experiment.

In the simulation we generated in total 30,000 experiments (1000 per noise level), with noise levels starting with the values shown in Tab. I and scaled them with a factor varying from 1 to 30. With a factor of 30, the rotation error of the motion measurement can be as much as $\pm 25°$ according to the 3σ principle, which is quite high compared to real motion sensors.

The methods we compared are: (i) the SVD-based direct solution (called SVD), (ii) least squares estimator based the Gauss-Markov model (called GM), and (iii) our Gauss-Helmert based approach (called GH). Both GH and GM are using Eq. (11) and Eq. (12) as constraint functions (or residual vectors). The initial guess to GH and GM are perturbed ground-truth values by uniform additive noise. For reasons of comparison, we also provided a ground truth initialization. The metric for comparison is the averaged Root-Mean-Square-Error (RMSE) of the estimated 6 rotation parameters and 6 translation parameters from 1,000 trials.

The result of the simulation is depicted in Fig. 5. From this plot, we can draw several conclusions. First, the GH and GM perform always better than SVD. The only advantage of the SVD is that it is a direct solution and requires no initial

TABLE II: Precision of the estimated parameters $\boldsymbol{\xi}, \boldsymbol{\eta}$

	σ_{ξ_1}	σ_{ξ_2}	σ_{ξ_3}	σ_{η_1}	σ_{η_2}	σ_{η_3}
Stereo B	2.8 mm	1.90 mm	2.8 mm	0.041°	0.032°	0.033°
Mocap	1.43 mm	1.97 mm	1.13 mm	0.165°	0.191°	0.202°

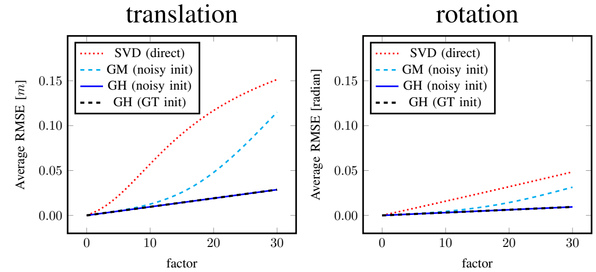

Fig. 5: Accuracy comparison through Monte-Carlo simulation. The x-axis show the factor by which we scale the input noise for the translational (left) and rotational (right) component. The plots show the RMSE for the direct SVD approach, the GM as well as GH model using a noisy initialization, and finally the GH model initialized with the ground truth as the initial guess. The GH model outperforms the other approaches as expected and performs identical if initialized with noisy values or the ground truth ones.

guess. Second, the GH and GM approaches produce identical results as long as the noise-level is small. Third, as the noise level increases, the accuracy of the GM solution degrades while our GH solution does not. The error of the GH solution grows linearly with the increased input noise, which is the expected theoretical result. Thus, we can conclude that over the spectrum of all situations, our GH approach performs best. At the highest noise-levels, GH is better than GM by 75% in translation (0.0287 vs. 0.1150), and by 71% in rotation (0.0095 vs. 0.0314) in this simulation setup. Only for situation with low noise (eg, a factor smaller than three), GH and GM show the same performance as can be expected given our explanation in Sec. IV. Finally, we can see that the GH approach produces the same results if using the GT for initialization or a noisy variant of it.

C. Robustness with Respect to the Initial Guess

This experiment is designed to show that our approach is robust enough to use an initial values from a direct approach In the previous experiments, we compared a ground truth initialization to perturbed values for the sake of comparison. But in reality, however, we have to obtain the initial guess either by manual measure or by utilizing a direct approach. In this experiment, we perform a similar experiment as before but with the initial guess coming from the SVD method, i.e. without any knowledge about the true configuration. We compare these to the results, which are obtained when using the ground truth as the initial guess. The results are depicted in Fig. 6 and show that the RMSE curves are identical for both initializations, hence we conclude that our approach is robust enough to be used in combination with SVD for providing the start value for the optimization.

D. Runtime

The last set of experiments in designed to illustrate the computation requirements of our approach and to show that

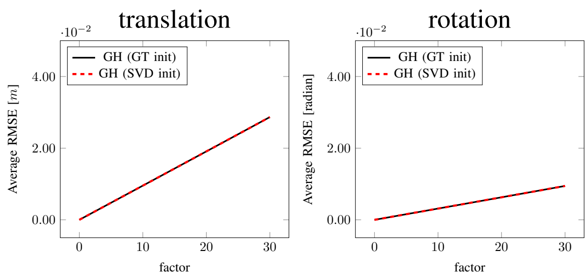

Fig. 6: Accuracy in relation to the initial guess for different noise levels on the translational and roitational component. In all cases, the initialized via the SVD result or with the ground truth produces the same results. Thus, we conclude that we can safely use SVD to initialize our GH optimization.

real world calibration can be done easily. For our real world calibration, we considered 1,630 poses extracted from each of the three sensors. The timings of our Python code running on a i5 notebook computer are approx. 6 s for the overall approach. Around 600 ms is used for computing the initial solution using SVD and around 1.2 s-1.5 s is required per iteration. These timing involves all computations, except the motion estimation from the sensors itself, in this case the visual odometry. At least in our setup, computing the visual odometry takes longer than the calibration itself.

E. Summary

In summary, our real world and simulation experiments suggest that our Gauss-Helmert based calibration approach outperforms the direct SVD-based approach as well as the ordinary least squares solution that is based on the Gauss-Markov model. The use of the Gauss-Helmert model with constraints allows us to provide better results than all other approach at high noise levels and produces the identical result than the Gauss-Markov model for small noise levels. We furthermore showed that our approach is robust enough with respect to the initial guess provided by the direct SVD approach. In real world settings, the optimization is faster than the time needed to compute a visual odometry and we required 6 s in a i5 notebook to calibrate 3 sensors with 1,630 poses each. Thus, we supported our claims with this experimental evaluation.

VII. Conclusion

In this paper, we presented a novel approach to compute the extrinsic parameters of multiple sensors installed on a moving platform. Our approach defines constraints between the motions of the individual sensors given that they are rigidly mounted in the robot and then computes the extrinsic parameters in a statistically sound way using the Gauss-Helmert model. This allows us to successfully determine the relative transformations between the origins of the sensor coordinate systems as the robot's base. We implemented and evaluated our approach in simulations as well as on real data. The experiments suggest that our approach can accurately determine the extrinsic parameters of the individual sensors under realistic conditions. We provided comparisons to a direct SVD approach as well as to the ordinary Gauss-Markov least squares estimation and furthermore supported all claims made in this paper through our evaluation.

References

[1] S. Agarwal, K. Mierle, and Others. Ceres solver. http://ceres-solver.org.

[2] A. R. Amiri-Simkooei. Least-squares variance component estimation: theory and gps applications. 2007.

[3] Y. Chang, X. Lebegue, and J. K. Aggarwal. Calibrating a mobile camera's parameters. *Pattern Recognition*, 26(1):75–88, 1993.

[4] K. Daniilidis. Hand-eye calibration using dual quaternions. *The International Journal of Robotics Research*, 18(3):286–298, 1999.

[5] F. Dornaika and R. Horaud. Simultaneous robot-world and hand-eye calibration. *IEEE transactions on Robotics and Automation*, 14(4):617–622, 1998.

[6] W. Förstner and B. P. Wrobel. *Photogrammetric Computer Vision*. Springer, 2016.

[7] C. X. Guo, F. M. Mirzaei, and S. I. Roumeliotis. An analytical least-squares solution to the odometer-camera extrinsic calibration problem. In *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*, 2012.

[8] L. Heng, B. Li, and M. Pollefeys. Camodocal: Automatic intrinsic and extrinsic calibration of a rig with multiple generic cameras and odometry. In *Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, 2013.

[9] R. Kümmerle, G. Grisetti, and W. Burgard. Simultaneous calibration, localization, and mapping. In *Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, pages 3716–3721, 2011.

[10] R. Kümmerle, G. Grisetti, H. Strasdat, K. Konolige, and W. Burgard. g2o: A general framework for graph optimization. In *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*, pages 3607–3613, 2011.

[11] F. Neitzel and B. Schaffrin. On the gauss–helmert model with a singular dispersion matrix where bq is of smaller rank than b. *Journal of computational and applied mathematics*, 291:458–467, 2016.

[12] J. Schneider, C. Eling, L. Klingbeil, H. Kuhlmann, W. Förstner, and C. Stachniss. Fast and effective online pose estimation and mapping for uavs. In *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*, pages 4784–4791, 2016.

[13] S. Schneider, T. Luettel, and H. Wuensche. Odometry-based online extrinsic sensor calibration. In *Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, pages 1287–1292, 2013.

[14] Y. C. Shiu and S. Ahmad. Calibration of wrist-mounted robotic sensors by solving homogeneous transform equations of the form ax=xb. *IEEE Tran. on Rob. and Aut.*, 5(1):16–29, Feb 1989.

[15] K. H. Strobl and G. Hirzinger. Optimal hand-eye calibration. In *Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, 2006.

[16] Z. Taylor and J. Nieto. Motion-based calibration of multimodal sensor arrays. In *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*, pages 4843–4850, 2015.

[17] R. Y. Tsai and R. K. Lenz. Real time versatile robotics hand/eye calibration using 3d machine vision. In *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*, 1988.

[18] H. Wolf. Das geodätische Gauß-Helmert-Modell und seine Eigenschaften. pages 103:41–43, 1978.

[19] Z. Zhao. Hand-eye calibration using convex optimization. In *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*, pages 2947–2952, 2011.