

# Robust Extrinsic Calibration of Multiple Stationary Laser Range Finders

Jan Quenzel<sup>1</sup>, Nils Papenberg<sup>2</sup> and Sven Behnke<sup>1</sup>

**Abstract**—Laser range finders are used in industrial safety and surveillance applications, e.g. to track persons. The relative position between two sensors is needed for the correct operation of such a system. In order to reduce costs, sensors are positioned to have minimal overlap while extending the combined field of view. The necessary calibration is often done in a time-consuming manual process. To automate this, moving objects are tracked in the laser scans and used to find the relative sensor poses.

We present a novel approach to pairwise calibration using shared observations in a RANSAC-based fashion to estimate the relative transformation between two sensors. Afterwards, we apply robust pose graph optimization that deals with possibly faulty pairwise estimates. Our method uses  $\ell_1$ -norm minimization in the tangential space of the rotation matrices under transitivity constraints and, thus, does not need a further initial guess to obtain the final sensor poses. The complete system is evaluated with simulated and real data in very challenging situations. To emphasize the robustness of the proposed calibration, we show results using a very simple non-robust tracking. The overall accuracy is further improved when ellipse fitting of the tracked targets is used.

## I. INTRODUCTION

Laser range finders (LRF) contactless measure distances to objects with high precision. As a result, they are often used in industrial safety applications or for monitoring open spaces. The sensors are positioned statically in the environment. Walls, columns, and other obstacles restrict the field of view achievable with individual sensors. Using multiple sensors addresses this limitation, while ensuring either redundancy or increasing the covered area, which is necessary in open spaces.

During deployment, the sensors need to be configured for their specific application. If the relative sensor poses are known, this step can be simplified, e.g., protective fields can be taught once to all sensors simultaneously. Furthermore, the sensor data can be preprocessed directly on the sensor instead of central processing. The bandwidth can be reduced further by sending only object tracks, thus allowing to use a larger number of sensors.

The relative poses are often not available, due to missing or incorrect maps. Hence, the calibration of a LRF network is in practice usually done manually. This is an elaborate

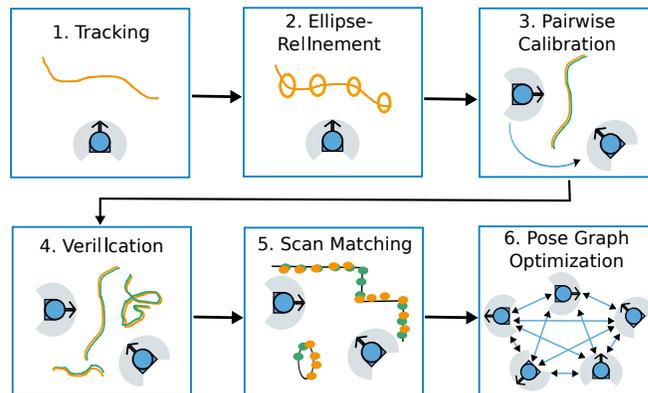


Fig. 1. Overview of the proposed system. An arbitrary tracking can be used in Step 1. The tracking is refined using ellipse-fitting (Step 2). The pairwise calibration (Step 3) calculates a transformation hypothesis for every track combination which is verified with all other tracks (Step 4). Then scan matching (Step 5) is used to correctly align walls. In the final Step 6, the sensor poses are generated using pose graph optimization.

and costly process. Calibrating in an automatic manner is preferable and faster, especially when the production line setup changes due to different requirements. Creating a correct map with a mobile robot using SLAM [1] is possible but increases the system costs and some areas might not be accessible. Sensor localization could be done by scan matching techniques but these are prone to getting stuck in local minima. This occurs particularly in featureless, symmetric, self-similar, and occluded environments, especially with minimal and non overlapping regions. Moving the sensors oftentimes reduces ambiguity, but this is not possible for static sensors.

Instead, we track moving objects in the vicinity of sensors. The observed tracks are used to determine the pairwise relative poses between sensors. We do not make the assumption that only one person or object is moving through the scene. On the contrary, we use multiple tracks between two sensors for verification, enabling us to distinguish between apparent and actually matching tracks. Assuming only one moving target would simplify the calibration and prolong calibration but is not always possible when other preparations take place or surrounding production lines are partially suspended.

After pairwise calibration, we represent the sensor network as a pose graph, where the first sensor is fixed at the origin of the reference coordinate frame. We assume that the graph is strongly connected. This representation allows the usage of Graph-SLAM algorithms[2], which perform non-linear optimization. These methods need a good initial guess and

This work was supported by SICK AG.

<sup>1</sup>Jan Quenzel and Sven Behnke are with the Autonomous Intelligent Systems Group, University of Bonn, 53113, Germany {quenzel, behnke}@ais.uni-bonn.de

<sup>2</sup>Nils Papenberg is with the Fraunhofer MEVIS Project Group Image Registration and with the Institute of Mathematics and Image Computing, University of Luebeck, Luebeck, 23562, Germany nils.papenberg@mic.uni-luebeck.de

are sensitive to outliers. To address this issue, we follow the Multiple Rotation Averaging approach [2] and optimize rotations first since translations highly depend on rotations, whereas rotations only depend linearly on translations. We propose a new robust method called LIMC. Our algorithm does not require a similar accurate starting point and works directly with all given pairwise rotation estimates. The rotations are constrained to be inverse and transitive consistent. Transformation into rotations tangential space allows for an easy representation as a linear system. This enables the usage of robust  $\ell_1$ -norm minimization. The sensor positions are then calculated in a least-squares fashion with fixed rotations.

In the following, we will restrict ourselves to the extrinsic calibration of the planar relative transform between sensors since for many industrial applications the planar case is sufficient.

The key contributions of our paper are:

- robust object-based pairwise calibration using multiple tracks for verification,
- robust Multiple Rotation Averaging without requiring an initial guess using  $\ell_1$ -norm optimization in tangential space of rotations,
- evaluation in a wide variety of scenarios with multiple moving persons with little to no overlap.

## II. RELATED WORK

Several methods have been proposed for extrinsic LRF calibration, based on feature extraction within the environment. For example Moghadam et al. [3] use lines, and Gong et al. [4] exploit arbitrary trihedrons, which are formed by two adjacent walls and the ceiling or floor. Similar, Choi et al. [5] use two orthogonal walls for pairwise calibration. Instead, Fernandez-Moral et al. [6] utilize a common planar surface visible from different orientations, whereas Brscic et al. [7] use poles with attached reflectors as artificial features. The drawback for these methods is often either the required overlapping field of view or the necessary availability of features.

More general methods only require incremental egocentric motion estimates. For example Brookshire & Teller [8] apply non-linear least squares (LS) to find a good solution. Instead, Schneider et al. [9] adapt an Unscented Kalman Filter (U-KF) for online calibration. Kümmerle et al. [1] even include the calibration parameters into their SLAM-formulation, while Taylor et al. [10] utilize the motion estimates in combination with observations similar to bundle adjustment.

For static sensors, these methods are not applicable. Instead, object-tracking can be used to determine the relative transformations between sensors. Sasaki et al. [11] apply a Least Median of Squares method on estimated object centers for pairwise calibration. Transformations with a mean squared error below a predefined threshold are then accepted. Instead, Glas et al. [12] match observations based on their velocity profile. Common observations are then used to create pairwise constraints, which are filtered by a KF. The final sensor position is obtained by Mass-Spring-Relaxation.

Later the authors improved the quality by tracking groups of people [13]. Recently, Glas et al. extended their approach to 3D [14] by using head-detections from RGB-D data in combination with horizontal laser range finders. In comparison, Schenk et al. [15] demonstrate superior results using LS-matching on observed positions for constraint generation. After relaxation, the calibration was further improved by a final ICP alignment on the static background scans. These object-based approaches assume a high overlap between all sensors, while the underlying tracking models assume people as dynamic objects. The calculations are carried out on a single central system.

A different approach is followed by Röwekämper et al. [16]. The existence of overlap is determined by simultaneous measurements on a dynamic object. This introduces a limit of only one moving calibration target. The pairwise estimate for overlapping sensors is generated using RANSAC, which is further incorporated into a factor graph to produce a joint initial guess. Afterwards, the graph is extended to include landmarks. A pose configuration is determined that minimizes the landmark reprojection error. The usefulness is limited due to the restriction of only one moving calibration target.

Instead, a combination of the above approaches is possible to obtain a robust system by replacing relaxation with a graph optimization. Outliers in the pairwise estimates lead to false sensor poses. Hence, the research focuses strongly on robustness to outliers. Carlone et al. [2] state that distinguishing between inlier and outlier measurements is unobservable. Therefore, they try to find the maximal set of inliers using mixed-integer programming and  $\ell_1$ -norm minimization to solve a 2D-Graph-SLAM optimisation problem. Wang & Singer [17] use alternating direction method of multipliers to minimize the difference between measured and estimated rotations. Chatterjee et al. [18] minimize the same  $\ell_1$ -norm using the Lie-algebra of rotations and Iterative Reweighted LS (IRLS). Similarly, the Maximum Likelihood estimator by Boumal et al. [19] operates on the rotation manifold while adjusting weights to distinguish wrong from correct measurements. Arrigoni et al. [20] extend the spectral relaxation by Arie-Nachimson et al. [21] from the rotation manifold  $SO(n)$  to the rigid transformation manifold  $SE(n)$ . For outlier rejection, IRLS is employed with a Cauchy weight function.

Many of the above techniques rely on a good initialization due to the non-linearity of the underlying methods. In our work we do not require such an initial guess.

## III. EXTRINSIC CALIBRATION

Our pairwise calibration can be thought of as a combination of Röwekämper RANSAC [16] and Schenk et al. [15] object-based LS approach with an additional verification step. Afterwards, we employ robust MRA which does not need an initialisation. First, we will introduce our notation and define necessary entities.

### A. Pose Graph

Let  $G = (V, E)$  be a directed connected graph called the pose- or measurement-graph. Each vertex represents the sensor pose in the reference coordinate system:

$$\begin{aligned} V &= \{v_i \in SE(n) | i = 1, \dots, N\}, \\ SE(n) &= \{(\mathbf{R}, \mathbf{t}) : \mathbf{R} \in SO(n), \mathbf{t} \in \mathbb{R}^n\}, \\ SO(n) &= \{\mathbf{R} \in \mathbb{R}^{n \times n} : \mathbf{R}^\top \mathbf{R} = \mathbf{I}, \det(\mathbf{R}) = +1\}. \end{aligned}$$

The rotation matrix  $\mathbf{R}$  describes the orientation and vector  $\mathbf{t}$  the position. Since we are only interested in planar extrinsic calibration, we restrict ourselves to  $n = 2$ , but the extension to 3D would be straightforward. A directed edge  $(i, j) \in E$  exists, if the visible areas of sensors  $i$  and  $j$  overlap and measuring the relative position  $g_{i,j}$  is possible.  $g_{i,j}$  maps from local coordinate frame  $i$  to local coordinate frame  $j$ . Given two sensor positions  $v_i$  and  $v_j$  the relative position  $g_{i,j} = (R_{i,j}, t_{i,j})$  can be calculated by:

$$\mathbf{R}_{i,j} = \mathbf{R}_i^\top \mathbf{R}_j, \quad (1)$$

$$\mathbf{t}_{i,j} = \mathbf{R}_i^\top (\mathbf{t}_j - \mathbf{t}_i). \quad (2)$$

Since the sensor positions are unknown, we fix  $v_1$  at the origin of the reference frame. To distinguish between estimated and measured relative position, the indices  $i, j$  are used for the first and  $r, s$  for the latter.

### B. Object Tracking

Any object tracking method can be used to generate the necessary tracks. Here, we first subtract the background to distinguish range measurements belonging to static elements like walls from dynamic objects. Close measurements on dynamic objects are clustered into segments. For each segment, the center of mass (CoM) is tracked over time by an Extended Kalman Filter under the assumption of a constant velocity. If multiple segments are candidates for one track, we associate the geometrically closest. This object tracking can be executed directly on the used SICK LRF. At the cost of reliability, tracking enables the user to make no assumption about the object shape. Tracking issues, like changes of CoM depending on the view poses, vanishing, reappearing, and near objects, present a challenge for a robust calibration process. A list of the tracked object positions and velocities is sent to the central system and collected for a predefined time period.

The collected tracks can be further processed by fitting an ellipse into the object points. We employ the method by Fitzgibbon et al. [22]. The parameters for the major and minor axis varies between consecutive scans. Therefore, we non-linearly fit a second ellipse with fixed major  $a = 0.25$  cm and minor radius  $b = 0.15$  cm in a second step. We are only interested in the center point  $\mathbf{t}$  of the object. Therefore, we minimize the geometrical distance of a scan point  $\mathbf{S}_i$  to its corresponding point under angle  $\theta_i$  on the ellipse in

parametric form:

$$\mathbf{x}^* = \arg \min_{[\theta, \alpha, \mathbf{t}]} \|\mathbf{f}(\mathbf{x})\|_2^2, \quad (3)$$

$$f_i(\mathbf{x}) = \mathbf{S}_i - \left[ \mathbf{R}(\alpha) \cdot \begin{pmatrix} a \cdot \cos(\theta_i) \\ b \cdot \sin(\theta_i) \end{pmatrix} + \mathbf{t} \right]. \quad (4)$$

The values for  $a, b$  are experimentally evaluated to fit a human at upper body height. The center point of the ellipse is again tracked via an EKF to improve the tracks accuracy. This accommodates for the error introduced by the CoM, resulting in more accurate and stable tracks.

### C. Preprocessing for Track Association

To achieve online calibration capabilities, often a fixed interval length is used. An estimate is generated after the first interval and further refined each time. Instead, we calculate the interval length dynamically to circumvent problems with vanishing, appearing, and near objects. Whenever a change in visible objects is detected, a new time period starts. In contrast to existing methods, we do not only use Kalman filtering, but use Kalman smoothing. The RTS-Smoother [23] is applied on each time period to smooth the track estimates. This is useful since the filtered estimate incorporates only measurements from the past and present at each time step, while smoothed estimates use all available measurements.

### D. Pairwise Extrinsic Calibration and Verification

We are given the smoothed positions  $\mathbf{t}_p^l, \mathbf{t}_q^l$  for two objects  $p, q$ , where  $p$  and  $q$  were observed simultaneously with sensor  $r$  and  $q$ , respectively. To obtain a first hypothesis  $h_{r,s}$  for the rigid transformation  $g_{r,s} = (\mathbf{R}, \mathbf{t}) \in SE(n)$ , we wish to minimize the distance:

$$\rho(\mathbf{R}, \mathbf{t}) = \sum_l \min \left\{ \left\| (\mathbf{R} \cdot \mathbf{t}_p^l + \mathbf{t}) - \mathbf{t}_q^l \right\|_2^2, r_{max}^2 \right\}. \quad (5)$$

Here,  $r_{max}$  is a constant representing the maximum sensor range that should reduce the influence of disappearing objects. To be robust against outliers, we employ a RANSAC variant called MSAC in combination with LS-optimization as in [24] for minimization. We calculate such a hypothesis for each track combination  $(p, q)$  and for each time interval. The resulting set of hypothesis are then verified on all time intervals. To find the best hypothesis, we use the Mahalanobis distance  $d_S$  to incorporate the uncertainty  $\Sigma$  of the tracked position:

$$d_S(\mathbf{a}, \mathbf{b}) = \sqrt{(\mathbf{a} - \mathbf{b})^\top [\Sigma_a + \Sigma_b]^{-1} (\mathbf{a} - \mathbf{b})}, \quad (6)$$

$$d_{i,j} = \sum_l d_m(\mathbf{R} \cdot \mathbf{t}_i^l + \mathbf{t}, \mathbf{t}_j^l), \quad (7)$$

$$E(\mathbf{R}, \mathbf{t}) = \sum_{k,i,j=1}^{L,|P|,|Q|} w_{i,j}^k \cdot \min \{d_{i,j}^k, r_{max}\}. \quad (8)$$

Here,  $w_{i,j} = \{0, 1\}$  represent the association between two tracks. Each  $w_{i,j}^k$  is calculated via nearest neighbour. One should note that this association  $w_{i,j}^k$  between tracks is not static in time but is changed to cope with vanishing and

reappearing objects. The pair distance  $d_{i,j}$  is again bounded above by the maximal LRF range  $r_{max}$ .

The key insight is that multiple tracks will give a similar transformation. The error  $\rho$  will be small when applied to other correctly associated tracks. At the same time, the error  $\rho$  will be large for wrong transforms from incorrect associated tracks, even though they might have a small MSE, a metric often used by other methods. Therefore,  $E$  will be small for correct transformations and large for wrong hypothesis, thus penalizing wrong associations. Naturally, we set  $g_{r,s}$  to be the hypothesis with minimal error  $E$ . Again, we calculate a least-squares hypothesis given all simultaneous observed tracks over the complete dataset without explicitly associating objects per time period but for every time stamp between sensor  $r$  and  $s$ . Therefore, multiple tracks are incorporated into a single transform.

With respect to the static environment, the resulting transformation can still be inaccurate due to variations of the tracked object. This effect can be reduced by further alignment of static structures between two views. Schenk et al. only used ICP on the background after the final positioning of all sensors. Instead, we employ a point-to-line-variant of ICP [25] already on the pairwise relative poses.

### E. Robust Multiple Rotation Averaging

The relative pairwise positions  $g_{r,s}$  allow for a pose graph formulation. Since we only obtain the imperfect estimates  $g_{r,s}$  for  $g_{i,j}$ , we want to obtain the set of sensor positions that minimizes these imperfections:

$$\arg \min_{v_1, \dots, v_N \in SE(n)} \sum_{i,j} d_R(\mathbf{R}_{i,j}, \mathbf{R}_{r,s})^2 + d_t(\mathbf{t}_{i,j}, \mathbf{t}_{r,s})^2. \quad (9)$$

Here,  $d_R$  denotes the angular difference or geodesic distance between the rotations  $\mathbf{R}_{i,j}$  and  $\mathbf{R}_{r,s}$ , while  $d_t$  is the Euclidean distance between  $\mathbf{t}_{i,j}$  and  $\mathbf{t}_{r,s}$ .

This problem is non-linear and depends on a good initial guess to converge to the global maximum. A typical strategy [2] involves Multiple Rotation Averaging. At first, the translations are ignored and (9) is solved for the rotations only. Estimating the translation then becomes an easy-to-solve linear problem. Thereby, no weighting between rotational and translational error is necessary.

All pairwise rotations  $\mathbf{R}_{i,j}$  can be combined to a single matrix  $\mathbf{H}$  for which the following relationship holds:

$$\mathbf{H} = \mathbf{R}^T \mathbf{R} \text{ with } \mathbf{R} \in \mathbb{R}^{n \times nN}. \quad (10)$$

Often  $\mathbf{R}$  is either directly or iteratively calculated, starting with an initial guess from odometry in SLAM. Instead, we optimize on all pairwise rotations  $\mathbf{R}_{r,s}$ . An advantage is that no initial guess is needed, since we know

$$\mathbf{R}_{i,j} \approx \mathbf{R}_{r,s} \quad (11)$$

and we already calculated  $\mathbf{R}_{r,s}$ .

To incorporate the above relationship (10), we constrain the pairwise rotations to be inverse and transitive consistent:

$$\mathbf{I} = \mathbf{R}_{i,j} \cdot \mathbf{R}_{j,i} = \mathbf{R}_{j,i} \cdot \mathbf{R}_{i,j}, \quad (12)$$

$$\mathbf{R}_{i,j} = \mathbf{R}_{i,k} \cdot \mathbf{R}_{k,j}. \quad (13)$$

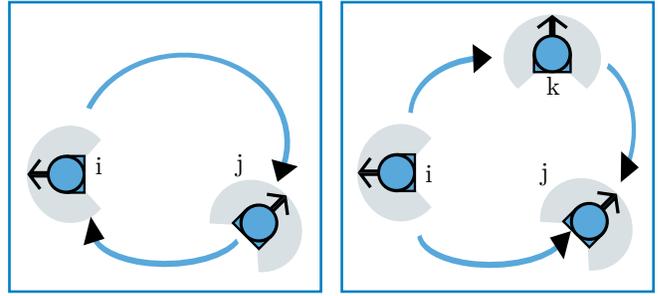


Fig. 2. The transform from sensor  $i$  to  $j$  is constrained to be the inverse of  $j$  to  $i$ . For the transitivity we want  $g_{k,j} \circ g_{i,k}$  to be the same as  $g_{i,j}$ .

The intuition for the transitivity constraint (13) is the following, illustrated in Fig. 2. When starting at sensor  $i$ , going over to sensor  $k$  and moving towards sensor  $j$ , we expect to be at the same position compared to moving directly from sensor  $i$  to  $j$ . Similarly, for the inverse consistency when going from sensor  $i$  to  $j$  and back, we should be at the initial position.

These constraints on their own do not allow for efficient optimization, since the problem remains non-convex. Like the authors in [18], we use the tangential space  $T_{\mathbf{R}_{r,s}}(SO(n))$  of the rotation matrices  $\mathbf{R}_{r,s}$  to relax the problem. The transformation into tangential space is done via the logarithm map between the Lie-group  $SO(n)$  and the Lie-algebra  $\mathfrak{so}(n)$ :

$$\log_{\mathbf{R}} : SO(n) \Rightarrow T_{\mathbf{R}}(SO(n)), \quad (14)$$

$$T_{\mathbf{R}}(SO(n)) = \{\mathbf{R}\mathbf{W} : \mathbf{W} \in \mathfrak{so}(n)\}, \quad (15)$$

$$\mathfrak{so}(n) = \{\mathbf{W} = -\mathbf{W}^T\}. \quad (16)$$

The vectorized representation of  $T_{\mathbf{R}_{r,s}}(SO(n))$  allows to express the above constraints as linear equations. For  $n = 2$  and  $n = 3$  follows:

$$\mathbf{W} = \begin{bmatrix} 0 & -w_3 & w_2 \\ w_3 & 0 & -w_1 \\ -w_2 & w_1 & 0 \end{bmatrix} \Leftrightarrow \mathbf{w} = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} \in \mathbb{R}^3, \quad (17)$$

$$\mathbf{W} = \begin{bmatrix} 0 & -w_1 \\ w_1 & 0 \end{bmatrix} \Leftrightarrow w = w_1 \in \mathbb{R}^1. \quad (18)$$

Constraints (11), (12) and (13) become:

$$\mathbf{R}_{i,j} \approx \mathbf{R}_{r,s} \Leftrightarrow \mathbf{w}_{i,j} \approx \mathbf{w}_{r,s} \quad (19)$$

$$\mathbf{I} = \mathbf{R}_{i,j} \cdot \mathbf{R}_{j,i} \Leftrightarrow \mathbf{0} = \mathbf{w}_{i,j} + \mathbf{w}_{j,i}, \quad (20)$$

$$\mathbf{R}_{i,j} = \mathbf{R}_{k,j} \cdot \mathbf{R}_{i,k} \Leftrightarrow \mathbf{w}_{i,j} = \mathbf{w}_{k,j} + \mathbf{w}_{i,k}. \quad (21)$$

Reorganizing the equations leads to the form  $\mathbf{A}\mathbf{x} \approx \mathbf{b}$ , where  $\mathbf{x}$  contains the stacked  $\mathbf{w}_{i,j}$ . Matrix  $\mathbf{A}$  is very sparse and all non-zero entries are either +1 or -1. The vector  $\mathbf{b}$  consists of the vectorised pairwise rotations  $\mathbf{w}_{r,s}$  and  $\mathbf{0}$  for the inverse and transitivity constraints. One has to keep the cyclic nature of rotations in mind. Even though  $\mathbf{w}_{i,j} = 2\pi$  is equal to  $\mathbf{w}_{j,i} = 0$  the constraint  $\mathbf{w}_{i,j} = \mathbf{w}_{j,i}$  does not hold. Therefore, normalization for  $\mathbf{w}_{r,s}$  is necessary. For the transitivity, it suffices to set the corresponding entries of  $\mathbf{b}$  to  $k \cdot \pi$  with  $k \in \{-1, 0, 1\}$ .

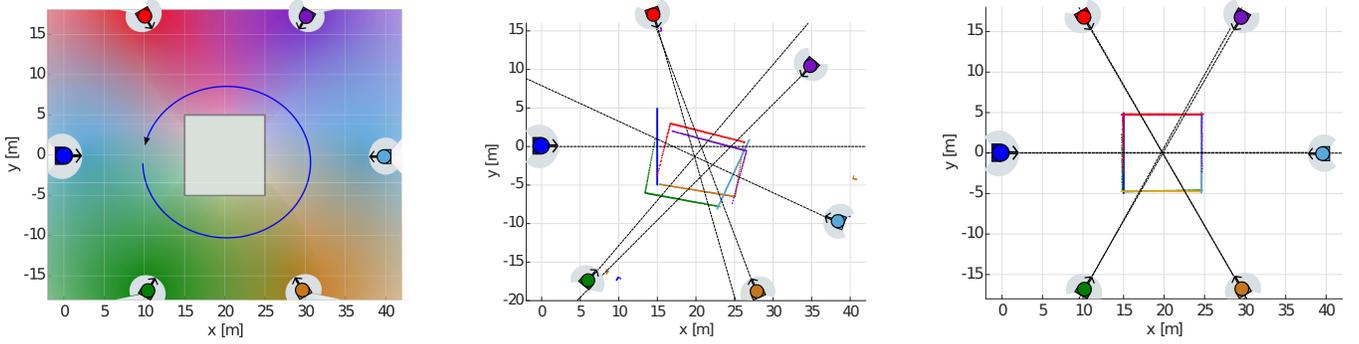


Fig. 3. The simulated setup is shown on the left with six inward-facing LRFs. A central cuboid restricts the measurement range and provides static background. Each field of view is color-coded like the corresponding LRF. The movement direction of the cylinder is indicated by the blue line. The difference between non-robust (middle,[21]) and our robust pose graph optimization (right) is emphasized by the black lines, indicating the sensor orientation. Ideally, the orientation lines of opposing sensors should coincide.

#### Algorithm 1: LIMC

```

// measurements:  $\mathbf{w}_{i,j} = \mathbf{w}_{r,s}$ 
 $\mathbf{A}_1 = \mathbf{I}; \mathbf{b}_1 = \mathbf{0};$ 
foreach  $(i, j) \in E$  do
     $\mathbf{A}_{i,j} = \mathbf{I};$ 
     $\mathbf{b}_{i,j} = \text{normalize}(\log(\mathbf{R}_{r,s}));$ 
     $\mathbf{x}'_{i,j} = \mathbf{b}_{i,j};$ 
end
 $\mathbf{A}_m = \mathbf{A}; \mathbf{b}_m = \mathbf{b};$ 
// symmetry:  $\mathbf{w}_{i,j} - \mathbf{w}_{j,i} = \mathbf{0}$ 
 $\mathbf{A}_s = [\dots, \mathbf{I}_{i,j}, \dots, -\mathbf{I}_{j,i}, \dots];$ 
 $\mathbf{b}_s = \mathbf{0};$ 
// transitivity:  $\mathbf{w}_{i,k} + \mathbf{w}_{k,j} - \mathbf{w}_{i,j} = \mathbf{0}$ 
 $\mathbf{A}_t = [\dots, \mathbf{I}_{i,k}, \dots, \mathbf{I}_{k,j}, \dots, -\mathbf{I}_{i,j}, \dots];$ 
 $\mathbf{b}_t = \mathbf{A}_t \mathbf{b}_m - \text{normalize}(\mathbf{A}_t \mathbf{b}_m);$ 
// solve  $\mathbf{A} \cdot \mathbf{x} = \mathbf{b}$ , starting at  $\mathbf{x}'$ 
 $\mathbf{A} = [\mathbf{A}_m, \mathbf{A}_s, \mathbf{A}_t]^\top;$ 
 $\mathbf{b} = [\mathbf{b}_m, \mathbf{b}_s, \mathbf{b}_t]^\top;$ 
 $\mathbf{x}^* = \arg \min_{\mathbf{x}} \|\mathbf{A} \cdot \mathbf{x} - \mathbf{b}\|_{\ell_1};$ 
// axis-angle to rotation matrix
foreach  $(i, j) = [(1, 1), \dots, (N, N)]$  do
     $\hat{\mathbf{R}}_{i,j} = \exp(\mathbf{x}^*_{i,j});$ 
end

```

Due to noise, no unique solution  $\mathbf{A}\mathbf{x} = \mathbf{b}$  exists. Thus we seek the difference to be sparse and have a small  $\ell_1$ -norm [26]. The expectation is that the constraints (12), (13) hold, while the error is concentrated around the measurement equations (11). Using the  $\ell_1$ -norm reduces the influence of outliers in the measurement graph, compared to the standard LS-approach. Trust in pairwise measurements can be reflected by weighting both sides of the measurement equations, while missing measurements can be set to zero or completely removed.

After minimizing the  $\ell_1$ -residual, the stacked angles can be extracted from  $\mathbf{x}$  and mapped back towards  $SO(n)$ . The exponential-map  $\exp_{\mathbf{R}}$  uses the matrix exponential to transform from tangential space into rotation space. The resulting LIMC method is summarized in Algorithm 1.

The resulting rotations are so close to fulfil (10) that one could extract the upper  $n \times nN$  rows which correspond to  $\mathbf{R}$ . Alternatively, to ensure (10) we either use Spectral Relaxation [21] or Riemannian Gradient Descent [27]. Since the resulting transformations were very good, as we will show in the next section, we did not use any further non-linear optimization as in [27].

#### IV. EXPERIMENTS AND RESULTS

We evaluate our system with simulated and real datasets from horizontal scanning LRFs. For the simulated dataset, Fig. 3, we position six inward-facing sensors uniformly distributed on a circle with 60 m diameter. The sensors range is limited to 20 m, thus LRF on opposite sides have no overlap. To further restrict the common sight, we add a  $10 \times 10$  m cuboid in the center of the circle. To distinguish the influence of the underlying object tracking from the algorithm accuracy, we use a moving cylinder with 50 cm diameter.

We also tested the whole system in five challenging real-world situations and compare against manual calibration. The first two are conducted in a cluttered office environment of size  $16 \times 11$  m. Many range measurements are scattered on objects within the office and do not only occur on the walls or on distinctive features. We use six SICK TIM561 with 10 m measurement range and 15 Hz scan frequency. The sensors were connected via LAN with a PC and distributed to cover the room from different view points as shown in Fig.5. The sensors were time-synchronized using NTP. In the second setup, Fig. 4, the sensors are positioned in a typical Automated Guided Vehicle (AGV) scenario in a semi circle. We recorded six trajectories of one person moving through the scenes. In the distributed scene, we also tested up to four people simultaneously walking within the environment.

The third real test was performed in an industrial setting. Three LMS1xx and one LMS5xx are used with 20 m (50 m) measurement range, 25 Hz scan frequency and  $0.5^\circ$  angular resolution. Data transmission was done via 802.11g WLAN onto a central PC. The setup is illustrated in Fig. 6. Two sensors (LMS1xx) were positioned in a hallway and two are within two industrial halls of different width. Each sensor

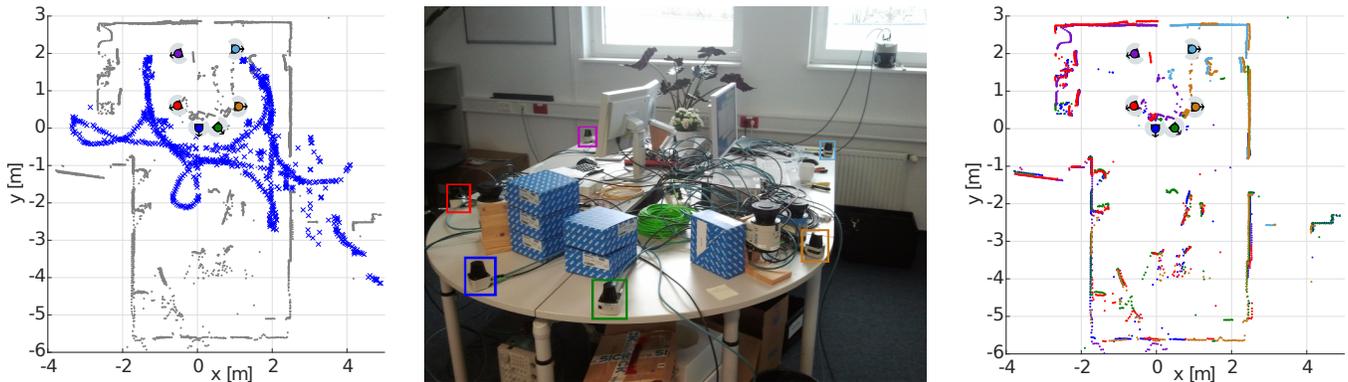


Fig. 4. Second real world scenario with six outward-facing TIM 561, center: Each LRF is marked by a colour-coded rectangle. Left: the observed movement is visible, where each object position is marked by a blue cross. Right: calibrated result. Each scan has the corresponding sensor color.

TABLE I

PERCENTAGE OF COMMONLY DISCARDED RELATIVE POSES WITH A MSE ABOVE 20 cm AS IN [15].

	simulated	office, semi- circle	office, distributed	industrial setting
discarded [%]	38.50	54.12	57.78	75.00

has a minimal overlap with at most two neighbours. The minimal cycle length is four. The high distance between the sensors represents the hardest challenge, since neighbours are up to 25 m away from each other. The view is mostly blocked by walls, restricting the benefit of ICP. This scenario does not represent a deployment scheme for industrial safety. The range of safety LRFs is typically specified to be less than 7 m. Instead we wanted to test the boundaries of our method. Again, we calibrated the system six times with one person and four times with two persons walking in opposite and the same direction. In all experiments, we used no fixed paths and an average capture time of one minute.

For the fourth test we deployed five S300 LRFs at upper body height to cover a L-shaped computer lab of size  $8 \times 7$  m. The sensors were scanning at a frequency of 12.5 Hz with an opening angle of  $270^\circ$ . In this scenario upto four people were walking randomly through the laboratory. Afterwards, we distributed the sensors to cover the corridors outside the lab.

To emphasize the effect of the verification step, Tab. I shows the percentage of pairwise estimates with a mean squared error above 20 cm. Around 50 % of the estimates are commonly discarded by other methods, despite the fact that the relative poses are correct.

Unfortunately, some details regarding the Mass-Spring-Relaxation in [15] are unclear, prohibiting a direct comparison. In addition, their datasets are not publicly available, hence we could not directly compare our methods. If we consider their hardest test cases (C, D), our setting is more challenging due to lower overlap, less common background and a higher minimal circle length.

A method, similar to Mass-Spring-Relaxation, is obtained

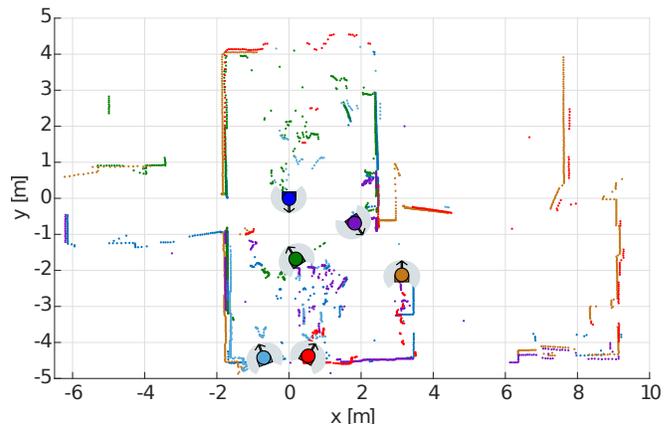


Fig. 5. A typical result for the distributed setup is shown in a cluttered office space of size  $16 \times 10$  m. Six laser range finders are positioned to maximize the combined field of view with redundancy. The laser scans are colour-coded to match the sensors.

via weighted Spectral-Relaxation [21] followed by weighted least-squares translation estimation. Thus, we will compare the robust approach against this purely weighted approach.

Weighting was not necessary on relative poses in L1MC except for the industrial scene where the outlier proportion approached 50 %. L1MC, like most MRA-methods [20], fails at that point when using uniform weights. The weights were calculated as proposed by Schenk et al. [15]. Table III reports the mean and median rotational and positional error for the four setups. We use the angular error  $d_R$  for the rotations and the Euclidean distance  $d_t$  for the positions. As expected, the non-robust approach is less accurate due to outliers in the relative poses. Weighting reduces the influence, but cannot completely remove it. On the contrary, the robust pose graph optimization calculated a good calibration, even with uniform weights. The results using the iterative L1RA [18] are not reported, since they coincide with our L1MC up to numerical errors. Thus, L1MC can be seen as a non-iterative L1RA that does not need an initial guess.

The CoM on the cylinder differs around 40 cm between two sensor poses in the simulated setup. Using the ellipse without refinement reduces the result drastically from 27 cm

TABLE II

COMPARISON OF DIFFERENT EXTRACTED FEATURES AND THE RESULTING ACCURACY ON THE SHOWN DATASET IN FIG. 4 WITHOUT ICP.

		our with LIMC		our w/o MRA <sup>1</sup>		w/o MRA <sup>1</sup> [15]		our with SR [21]	
		mean	median	mean	median	mean	median	mean	median
CoM	$d_R$ [°]	1.0776	0.5903	1.6635	0.9327	2.5689	1.6083	2.1988	1.7597
	$d_t$ [m]	0.1086	0.0823	0.2722	0.0582	0.3378	0.1452	0.1148	0.1031
Ellipse	$d_R$ [°]	0.6490	0.6242	1.3505	0.4965	2.3493	0.7765	0.6232	0.4662
	$d_t$ [m]	0.0599	0.0457	<b>0.0706</b>	0.0360	0.2056	0.1033	0.0605	0.0464
NLS-Ellipse	$d_R$ [°]	<b>0.3144</b>	<b>0.1870</b>	<b>0.7531</b>	<b>0.2748</b>	<b>1.6595</b>	<b>1.0169</b>	<b>0.5722</b>	<b>0.3434</b>
	$d_t$ [m]	<b>0.0406</b>	<b>0.0293</b>	0.1921	<b>0.0284</b>	<b>0.1727</b>	<b>0.0548</b>	<b>0.0404</b>	<b>0.0301</b>

TABLE III

RESULTS FOR THE SIMULATED AND FIVE REAL SETUPS OVER ALL TRIALS W/O UNIFORM WEIGHTS. <sup>2</sup>

robust		simulated <sup>3</sup>		semi-circle <sup>3</sup>		distributed <sup>3</sup>		industrial setting <sup>3</sup>		computer lab		corridor		
		mean	median	mean	median	mean	median	mean	median	mean	median	mean	median	
w/	yes	$d_R$ [°]	0.1125	0.1081	1.3706	0.9661	<b>0.6011</b>	<b>0.5006</b>	<b>0.6725</b>	<b>0.7228</b>	0.6801	0.6130	<b>2.4112</b>	<b>2.0338</b>
		$d_t$ [m]	0.0624	0.0561	0.1075	0.0409	0.0902	0.0764	<b>0.1861</b>	<b>0.1144</b>	0.0431	0.0313	<b>0.3425</b>	<b>0.1341</b>
w/	no	$d_R$ [°]	<b>0.1042</b>	<b>0.0945</b>	1.2371	1.0059	3.3158	1.3237	3.7149	1.4346	<b>0.3993</b>	<b>0.3416</b>	6.8653	3.6297
		$d_t$ [m]	0.0617	0.0550	<b>0.0581</b>	<b>0.0319</b>	0.1949	0.2174	3.1873	2.0526	0.0364	<b>0.0339</b>	1.5243	0.4993
w/o	yes	$d_R$ [°]	0.4900	0.5026	<b>1.2097</b>	1.0677	0.8498	0.5448	–	–	0.4036	0.4884	–	–
		$d_t$ [m]	<b>0.0144</b>	<b>0.0381</b>	0.0801	0.0604	<b>0.0891</b>	<b>0.0721</b>	–	–	<b>0.0353</b>	0.0427	–	–
w/o	no	$d_R$ [°]	7.6167	7.5707	1.6324	1.0117	2.4811	0.9973	–	–	0.4007	0.4672	–	–
		$d_t$ [m]	6.3353	6.3464	0.0947	0.0459	0.1569	0.1432	–	–	0.0353	0.0427	–	–

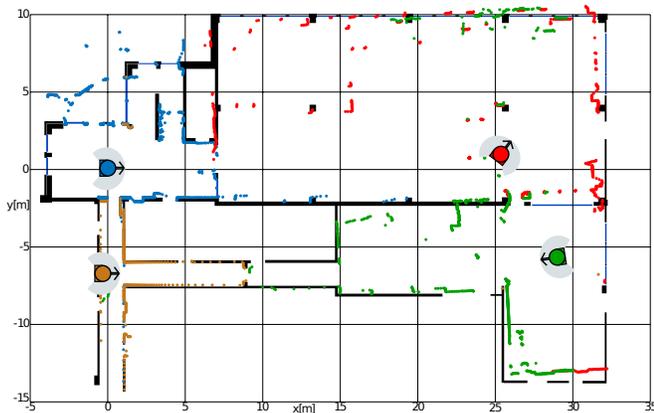


Fig. 6. The resulting calibration with our system in the industrial test setting. The scans are colour-coded by sensor. Two persons moved around in the same direction with varying distance. The area is  $30 \times 25$  m in size. The left and right sensors are 25 m apart.

to 7.7 cm (4 cm). The refinement kept the mean error fixed, but reduced the median error to 2.8 cm. The rotational error is reduced from  $1.66^\circ$  ( $0.93^\circ$ ) to  $1.3^\circ$  ( $0.37^\circ$ ) without LIMC.

One example for final sensor poses is shown in Fig. 4 (right) with overlaid laser scans in the second office scene. We obtain a mean (median) rotational error of  $0.3623^\circ$  ( $0.3100^\circ$ ) and 0.0434 m (0.0359 m), respectively, for the positional error.

A typical result for the industrial setting overlaid with the floor plan is shown in Fig. 6. Nearby poses are better calibrated than the more distant ones. The upper left pose is fixed on the origin while both sensors in the right half are rotated to the left by around  $1^\circ$ . Since the position depends on the orientation, this results in a higher positional error.

TABLE IV

COMPARISON FOR DIFFERENT NUMBER OF PEOPLE AND THE RESULTING ACCURACY ON THE COMPUTER LAB DATASET WITHOUT ICP.

number of people		our with LIMC		our with SR [21]	
		mean	median	mean	median
1	$d_R$ [°]	1.1193	1.0870	0.5044	0.4439
	$d_t$ [m]	0.0531	0.0443	0.0345	0.0378
2	$d_R$ [°]	0.9562	0.9064	0.4921	0.3589
	$d_t$ [m]	0.0407	0.0331	<b>0.0305</b>	<b>0.0140</b>
3	$d_R$ [°]	<b>0.2496</b>	<b>0.2754</b>	<b>0.2024</b>	<b>0.1594</b>
	$d_t$ [m]	0.0450	<b>0.0261</b>	0.0466	0.0326
4	$d_R$ [°]	0.3951	0.3196	0.3981	0.3243
	$d_t$ [m]	<b>0.0334</b>	0.0295	0.0340	0.0352

Adding a sensor between the upper ones could reduce this effect. We obtain a median positional error below 0.12 m and an angular accuracy below  $0.73^\circ$ . This is sufficient for many surveillance applications.

In first tests in the office environment, we observed the tendency of improving accuracy using multiple persons for calibration in smaller environments. After further evaluation, we found that more diverse walking paths in overlapping areas influences the obtained accuracy rather than the number of moving people. Tab. IV shows similar accuracy and walking paths independent of the number of moving people for the computer lab dataset, while the paths varied strongly

<sup>1</sup>Some relative position estimates are missing without multiple rotation averaging.

<sup>2</sup>The pose graph optimization failed with uniform weights for the industrial and corridor setting, due to the existence of  $\geq 50\%$  outliers.

<sup>3</sup>The denoted results were generated using only CoM for object tracking.

in the office environment.

The execution time of our Matlab implementation on a single core of an Intel Core-i7 CPU varies between seconds and some minutes, depending on the number of persons used for calibration. Most processing time is used in the non-linear ellipse fitting. In contrast, LIMC was done within milliseconds without exploiting sparseness.

## V. CONCLUSION

We presented a robust method to calculate the sensor positions using object observations. A RANSAC variant generates hypothesis for the relative poses between two sensors. A verification step with soft-assignments chooses the hypothesis that fits best the observed objects. Robust pose graph optimization is applied on the relative poses in form of Multiple Rotation Averaging. We optimize the relative rotations in their tangential space under inverse and transitivity constraints using the robust  $\ell_1$  norm. This does not require any initial guess and copes with missing rotations.

The experimental results suggest that our system is robust under difficult conditions and achieves state of the art accuracy.

## REFERENCES

- [1] R. Kümmerle, G. Grisetti, and W. Burgard, "Simultaneous parameter calibration, localization, and mapping," *Advanced Robotics*, 2012.
- [2] L. Carlone, A. Censi, and F. Dellaert, "Selecting good measurements via  $\ell_1$  relaxation: A convex approach for robust estimation over graphs," in *Proc. of the IEEE/RSJ Int. Conference on Intelligent Robots and Systems (IROS)*, 2014.
- [3] P. Moghadam, M. Bosse, and R. Zlot, "Line-based extrinsic calibration of range and image sensors," in *Proc. of the IEEE Int. Conference on Robotics and Automation (ICRA)*, 2013.
- [4] X. Gong, Y. Lin, and J. Liu, "3D LIDAR-Camera Extrinsic Calibration Using an Arbitrary Trihedron," *Sensors*, 2013.
- [5] Y. Bok, D. Choi, P. Vasseur, and I. Kweon, "Extrinsic calibration of non-overlapping camera-laser system using structured environment," in *Proc. of the IEEE/RSJ Int. Conference on Intelligent Robots and Systems (IROS)*, 2014.
- [6] E. Fernandez-Moral, V. Arevalo, and J. Gonzalez-Jimenez, "Extrinsic calibration of a set of 2D laser rangefinders," in *Proc. of the IEEE Int. Conference on Robotics and Automation (ICRA)*, 2015.
- [7] D. Brscic, T. Kanda, T. Ikeda, and T. Miyashita, "Person Tracking in Large Public Spaces Using 3-D Range Sensors," *IEEE Transactions on Human-Machine Systems*, 2013.
- [8] J. Brookshire and S. Teller, "Automatic calibration of multiple coplanar sensors," *Robotics: Science and Systems VII*, 2012.
- [9] S. Schneider, T. Luettel, and H. Wuensche, "Odometry-based online extrinsic sensor calibration," in *Proc. of the IEEE/RSJ Int. Conference on Intelligent Robots and Systems (IROS)*, 2013.
- [10] Z. Taylor and J. Nieto, "Motion-Based Calibration of Multimodal Sensor Arrays," in *Proc. of the IEEE Int. Conference on Robotics and Automation (ICRA)*, 2015.
- [11] T. Sasaki and H. Hashimoto, *Object Tracking for Calibration of Distributed Sensors in Intelligent Space*. InTech, 2011.
- [12] D. Glas, T. Miyashita, H. Ishiguro, and N. Hagita, "Automatic Position Calibration and Sensor Displacement Detection for Networks of Laser Range Finders for Human Tracking," in *Proc. of the IEEE/RSJ Int. Conference on Intelligent Robots and Systems (IROS)*, 2010.
- [13] D. Glas, F. Ferreri, T. Miyashita, H. Ishiguro, and N. Hagita, "Automatic calibration of laser range finder positions for pedestrian tracking based on social group detections," *Advanced Robotics*, 2014.
- [14] D. Glas, D. Brscic, T. Miyashita, and N. Hagita, "SNAPCAT-3D: Calibrating Networks of 3D Range Sensors for Pedestrian Tracking," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2015.
- [15] K. Schenk, A. Kolarow, M. Eisenbach, K. Debes, and H. Gross, "Automatic calibration of a stationary network of laser range finders by matching movement trajectories," in *Proc. of the IEEE/RSJ Int. Conference on Intelligent Robots and Systems (IROS)*, 2012.
- [16] J. Röwekämper, M. Ruhnke, B. Steder, W. Burgard, and G. Tipaldi, "Automatic extrinsic calibration of multiple laser range sensors with little overlap," in *Proc. of the IEEE Int. Conference on Robotics and Automation (ICRA)*, 2015.
- [17] L. Wang and A. Singer, "Exact and stable recovery of rotations for robust synchronization," *Information and Inference*, 2013.
- [18] A. Chatterjee and V. Govindu, "Efficient and robust large-scale rotation averaging," in *Proc. of the IEEE Int. Conference on Computer Vision (ICCV)*, 2013.
- [19] N. Boumal, A. Singer, and P. Absil, "Robust estimation of rotations from relative measurements by maximum likelihood," in *IEEE Annual Conference on Decision and Control (CDC)*, 2013.
- [20] F. Arrigoni, A. Fusiello, and B. Rossi, "Spectral Motion Synchronization in SE(3)," *CoRR*, 2015.
- [21] M. Arie-Nachimson, S. Kovalsky, I. Kemelmacher-Shlizerman, A. Singer, and R. Basri, "Global motion estimation from point matches," in *Second International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT)*, 2012.
- [22] A. Fitzgibbon, M. Pilu, and R. Fisher, "Direct least-squares fitting of ellipses," 1999.
- [23] S. Särkkä, *Bayesian filtering and smoothing*. Cambridge University Press, 2013.
- [24] P. Torr and A. Zisserman, "MLESAC: A new robust estimator with application to estimating image geometry," *Computer Vision and Image Understanding*, 2000.
- [25] A. Censi, "An ICP variant using a point-to-line metric," in *Proc. of the IEEE Int. Conference on Robotics and Automation (ICRA)*, 2008.
- [26] A. Yang, Z. Zihan, A. Balasubramanian, S. Sastry, and M. Yi, "Fast  $\ell_1$ -Minimization Algorithms for Robust Face Recognition," *Image Processing, IEEE Transactions on*, 2013.
- [27] L. Carlone, R. Tron, K. Daniilidis, and F. Dellaert, "Initialization techniques for 3D SLAM: a survey on rotation estimation and its use in pose graph optimization," in *Proc. of the IEEE Int. Conference on Robotics and Automation (ICRA)*, 2015.