*Article*

# Motion-Based Extrinsic Sensor-to-Sensor Calibration: Effect of Reference Frame Selection for New and Existing Methods

Tuomas Välimäki *[ID], Bharath Garigipati and Reza Ghabcheloo [ID]

Automation Technology and Mechanical Engineering, Tampere University, FI-33720 Tampere, Finland; reza.ghabcheloo@tuni.fi (R.G.)
* Correspondence: tuomas.valimaki@tuni.fi

**Abstract:** This paper studies the effect of reference frame selection in sensor-to-sensor extrinsic calibration when formulated as a motion-based hand–eye calibration problem. As the sensor trajectories typically contain some composition of noise, the aim is to determine which selection strategies work best under which noise conditions. Different reference selection options are tested under varying noise conditions in simulations, and the findings are validated with real data from the KITTI dataset. The study is conducted for four state-of-the-art methods, as well as two proposed cost functions for nonlinear optimization. One of the proposed cost functions incorporates outlier rejection to improve calibration performance and was shown to significantly improve performance in the presence of outliers, and either match or outperform the other algorithms in other noise conditions. However, the performance gain from reference frame selection was deemed larger than that from algorithm selection. In addition, we show that with realistic noise, the reference frame selection method commonly used in the literature, is inferior to other tested options, and that relative error metrics are not reliable for telling which method achieves best calibration performance.

**Keywords:** reference frame selection; sensor-to-sensor calibration; extrinsic calibration; hand–eye calibration; motion-based calibration

## 1. Introduction

With growing industry interest in automating many tasks performed with mobile working machines such as forklifts, excavators, or harvesters, easy-to-use sensor-to-sensor calibration is essential in enabling sensor fusion for the desired algorithms and functionality. Traditional sensor calibration methods rely on structured environments and/or calibration targets, meaning laborious setups for calibration data collection. Multiple targetless sensor-to-sensor calibration methods have been proposed to circumvent these issues, but most still require a shared field of view for sensors, which is not always possible or desirable, and limits the types of sensors that can be used.

Recently, many motion-based extrinsic calibration methods, that treat the process as a hand–eye calibration problem, have been proposed. Formulated in this manner, the calibration (i) does not require calibrated targets, (ii) does not require a shared field of view, and (iii) can work with multimodal sensors, requiring only the sensor trajectories. These properties make the hand–eye calibration approach especially suitable for mobile working machines that often incorporate multimodal sensor arrays that do not have overlapping fields of view. When no calibration targets are needed, the recalibration of the sensor array is also possible in field conditions, which is very important for large machines that cannot be easily taken back to shop for repairs.

In the traditional hand–eye calibration problem, with a camera mounted on a robotic arm, there is very little noise when compared with sensor trajectories generated using SLAM (Simultaneous Localization and Mapping) algorithms. What noise there is, is best modeled as additive Gaussian noise. SLAM trajectories, however, may contain noticeable drift and

outliers in addition to general noise. To the authors best knowledge, no existing literature studies how the effect of these different types of noise on the calibration performance could be mitigated through the selection of reference frames.

In this paper, we compare multiple state-of-the-art motion-based extrinsic calibration algorithms and data preprocessing under varying noise conditions. The first novelty comes from considering the effect of reference frame selection in calibration performance. In addition, we propose a nonlinear optimization method with outlier rejection that is shown to improve performance under some of the conditions. We show that the reference frame selection method commonly used in the literature is inferior in presence of noise typical for SLAM generated trajectories. In practice, since the ground truth is unknown, the only option to compare two methods is to compare which one reduces the relative error the most. However, using simulated data with known ground truth information, we show that the relative error metrics are not reliable for telling which method is better. The findings are demonstrated both using a large set of simulated data and the publicly available KITTI dataset [1]. We also provide general guidelines for reference frame selection in data preprocessing and propose directions for future research.

## 2. Related Work

Motion-based extrinsic sensor-to-sensor calibration has a close affinity to hand–eye calibration from manipulator literature, and many of the same algorithms apply. There are two main types of algorithms that solve the hand–eye calibration problem, separable algorithms that solve the orientation and translation parts in separate linear stages, and simultaneous algorithms that solve the complete nonlinear problem in a single stage. Separable solutions are generally simpler but more error-prone, since the errors from rotation estimation propagate to the translational part. Lately, also other methods to solve the hand–eye calibration have been proposed, such as factor graph optimization [2] and even neural networks [3].

In [4], the authors present a separable motion based calibration method for multimodal sensors. A coarse alignment is based on a modified Kabsch algorithm, that is then refined using sensor specific metrics for sensors with overlapping field of view. The method is later improved in [5]. Similarly, a lidar-to-camera calibration method based on trajectory matching for coarse alignment and a refinement step based on image features is presented in [6]. The separable course alignment is applicable regardless of sensor type, but fine-tuning is limited to cameras.

In [7], an early work on simultaneous solution, the problem is formulated using quaternions, and observability conditions are provided based on a new identification Jacobian. More recently, [8] presents a simultaneous general approach for hand–eye calibration for multiple cameras, based on optimization of atomic transformations dubbed ATOM. In [9], the authors build upon ATOM to allow multimodal sensors, but the method still necessitates a camera due to reprojection error being used as optimization metric. Both calibration methods are target-based and the sensors must share field of view.

A survey of state-of-the-art hand–eye calibration methods in the traditional robotic arm use case is presented in [10]. The authors also present their own simultaneous optimization method and propose multiple possible cost functions. The simultaneous nonlinear optimization method proposed in our paper is similar to the formulation of Xc1 in [10], but differs in using rotation vectors instead of quaternions. The MATLAB implementation provided by the authors of [10], however, does not effectively include rotational errors at all in the cost function, but instead the cost for the rotational error is the norm of a unit quaternion. Our reimplementation replicates the behaviour implemented by the original authors. Our paper also proposes a novel modification to the cost function to add outlier rejection without resorting to a computationally heavy RANSAC [11] framework.

Iterative simultaneous optimization methods are shown to perform better than closed-form solutions, which have difficulties handling noisy or inconsistent data, in the traditional robotic arm use case [8,10,12]. However, to the authors best knowledge, no previous work

discusses the data preprocessing choices, especially reference selection, and their effect on calibration performance under different noise conditions.

## 3. Method

### *3.1. Calibration*

We introduce a notation, where we represent homogeneous transformations as $\mathbf{T}$, indicating the source and target frames as sub- and super-indices respectfully, e.g., $_{w1}\mathbf{T}^{s1_i}$ represents the transformation from the world frame associated with sensor 1 to the sensor pose at time instance $i$. The corresponding rotation and translation parts are referred to as $\{\mathbf{R}, t\}$ respectfully, using the same sub- and super-indexing.

Consider two sensors, $s1$ and $s2$, installed on the same rigid body. When the body moves, the sensor trajectories are recorded e.g., using suitable SLAM algorithms. As presented in Figure 1, we can now form the relationship

$$\left(_{w1}\mathbf{T}^{s1_i}\right)^{-1}{}_{w1}\mathbf{T}^{s1_j}{}_{s1}\mathbf{T}^{s2} = {}_{s1}\mathbf{T}^{s2}\left(_{w2}\mathbf{T}^{s2_i}\right)^{-1}{}_{w2}\mathbf{T}^{s2_j}, \tag{1}$$

or equivalently

$$_{s1_i}\mathbf{T}^{s1_j}{}_{s1}\mathbf{T}^{s2} = {}_{s1}\mathbf{T}^{s2}{}_{s2_i}\mathbf{T}^{s2_j}, \tag{2}$$

which is the well known hand–eye calibration formulation $\mathbf{AX} = \mathbf{XB}$. Here, $\mathbf{A}$ and $\mathbf{B}$ comprise the relative transformations between two frames in the respective sensor trajectories and $\mathbf{X}$ is the unknown sensor-to-sensor transformation $_{s1}\mathbf{T}^{s2}$. Notice that the equality is valid for any pair of $(i, j)$ from the data, and we can solve the group of equations formed from the set of all selected pairs $S = \{(i, j), \dots\}$. The relative transformations are typically calculated between two consecutive poses, i.e., $i = j - 1$, but this need not be the case as covered in Section 3.4. We will also show that different selection methods result in more accurate calibration.
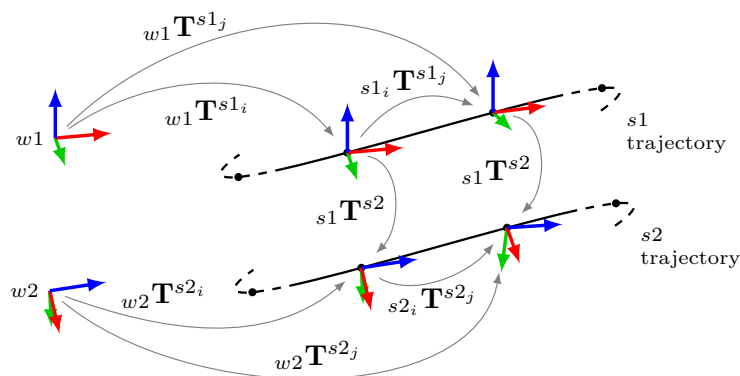


**Figure 1.** Graphical representation of the motion-based sensor-to-sensor calibration concept, where pose $i$ is used as the reference frame for pose $j$.

We analyzed several state-of-the-art calibration methods, namely those presented by Ali et al. [10], Park et al. [6], Taylor and Nieto [4], and Zhuang and Qu [7]. For [4,6], only the coarse alignment is applicable, and from [10] we use the cost Xc1. The algorithms are later referred to by only the first author. In addition, we propose our own nonlinear optimization method detailed in the following section.

### *3.2. Proposed Optimization*

We directly minimize the nonlinear relation in (2). The unknowns for the optimization are $\{\theta, t\}$, where $\theta$ is the rotation vector corresponding with the axis-angle representation of $_{s1}\mathbf{R}^{s2}$, and the indices are dropped from $_{s1}t^{s2}$ for the sake of brevity.

We use two cost functions, in which

$$\mathbf{M}_{i,j} := {}_{s1_i}\mathbf{T}^{s1_j}[\theta, t]_T - [\theta, t]_T{}_{s2_i}\mathbf{T}^{s2_j}, \tag{3}$$

where the $[\ ]_T$ notation represents the homogeneous transformation formed from the parameters. The first cost is

$$\{\theta, t\} = \arg\min_{\theta, t} \sum_{(i,j) \in S} \left\| \mathbf{M}_{i,j} \right\|_2^2, \tag{4}$$

later referred to as direct non-linear (DNL), and the second one is

$$\{\theta, t, \alpha\} = \arg\min_{\theta, t, \alpha} \sum_{(i,j) \in S} \left( \alpha_{i,j} \left\| \mathbf{M}_{i,j} \right\|_2^2 + (1 - \alpha_{i,j})c \right)$$

$$\text{subject to:} \quad \begin{aligned} 0 \leq \alpha_{i,j} \leq 1, \forall i, j \\ \sum_{i,j} \alpha_{i,j} \geq d \end{aligned} \quad , \tag{5}$$

later referred to as direct non-linear with outlier rejection (DNLO). DNLO adds outlier rejection by incorporating a weighting vector $\alpha$ into the decision variables. The rejection can be tuned by appropriately selecting parameters $c$ and $d$. During optimization, if the error term $\left\| \mathbf{M}_{i,j} \right\|_2^2 > c$, the corresponding weight $\alpha_{i,j}$ will tend to 0 and the error is regarded an outlier. Therefore, $c$ defines a threshold for error tolerance. Similarly, if $\left\| \mathbf{M}_{i,j} \right\|_2^2 < c$ then $\alpha_{i,j}$ will tend to 1. The sum over $\alpha$ can therefore be regarded as the proportion of inliers, and $d$ defines the minimum proportion of weighted poses to keep. Our implementation uses Ipopt as part of CasADi [13] for the minimization.

### 3.3. Error Metrics

For evaluation purposes, we adopt the four error metrics used in [10]. The first two are derived from the relation (2), the first being mean relative translation error

$$e_{rt} = \frac{1}{|S|} \sum_{(i,j) \in S} \left\| {}_{s1_i}\mathbf{R}^{s1_j}{}_{s1}t^{s2} + {}_{s1_i}t^{s1_j} - \right.$$

$$\left. {}_{s1}\mathbf{R}^{s2}{}_{s2_i}t^{s2_j} + {}_{s1}t^{s2} \right\|_2, \tag{6}$$

and the second being mean relative rotation error

$$e_{rR} = \frac{1}{|S|} \sum_{(i,j) \in S} \left\| \left[ ({}_{s1}\mathbf{R}^{s2}{}_{s2_i}\mathbf{R}^{s2_j})^{-1}{}_{s1_i}\mathbf{R}^{s1_j}{}_{s1}\mathbf{R}^{s2} \right]_\theta \right\|_2, \tag{7}$$

where the notation $[\ ]_\theta$ represents the rotation vector formed from the rotation matrix and $|S|$ is the cardinality of set $S$.

The remaining two are defined as the difference from the known ground truth values $\{{}_{s1}\mathbf{R}^{s2}_{gt}, {}_{s1}t^{s2}_{gt}\}$, namely, the absolute translation error

$$e_{at} = \left\| {}_{s1}t^{s2}_{gt} - {}_{s1}t^{s2} \right\|_2, \tag{8}$$

and the absolute rotation error

$$e_{aR} = \left\| \left[ {}_{s1}\mathbf{R}^{s2^{-1}}{}_{s1}\mathbf{R}^{s2}_{gt} \right]_\theta \right\|_2. \tag{9}$$

### 3.4. Reference Selection

In addition to multiple calibration algorithms, we study different ways to select the pairs $(i, j)$ for the relative transformations needed in the hand–eye calibration problem introduced in Section 3.1. The studied reference frames for the relative coordinates are presented in Figure 2 and are namely,

- Case $A$: where all poses are w.r.t. the first pose of the trajectory, i.e., $i = 0$. The method is expected to be particularly susceptible to drift due to the error constantly increasing the farther away the current pose is from the start of the trajectory.
- Case $B1$: where the reference for relative movement is the previous pose in trajectory, i.e., $i = j - 1$. This is the *de facto* reference used throughout previous studies. Keeping the transformations small reduces susceptibility to drift, but might prove vulnerable to noise. If the transformations between the poses are too small w.r.t. the noise, i.e., the signal-to-noise ratio worsens, the calibration performance likely suffers.
- Case $Bn$: a broader definition of the previous case where the reference is the $n^{\text{th}}$ previous pose, i.e., $i = j - n$. Freedom in choosing $n$, provides a way to trade off between sensitivity to noise and drift. The difference to just keeping every $n^{\text{th}}$ pose from the original trajectory is, that as opposed to drastically reducing the number of poses, only $n - 1$ first poses are dropped.
- Case $Cn$: a keyframe based approach inspired by [14], where the trajectory is divided into segments of equal length $n$. All poses in each segment are w.r.t. the first frame of the segment, the so called keyframe $k_s$, i.e., $i = k_s$. The method provides a mix of smaller and larger relative transformations that might prove useful in certain situations.
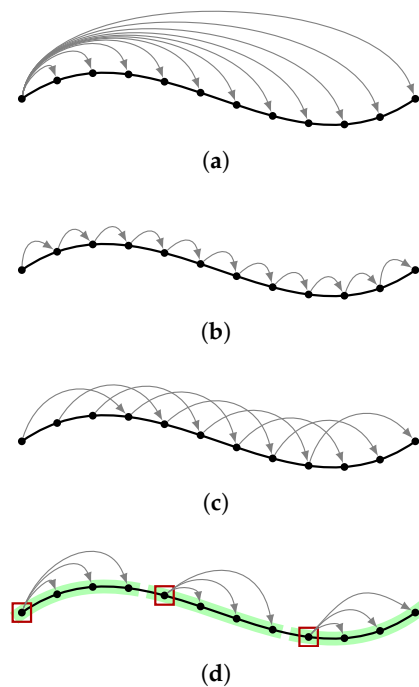


(a)

(b)

(c)

(d)

**Figure 2.** Different ways to define the reference frame for relative coordinates used in calibration. (**a**) $i = 0$, (**b**) $i = j - 1$, (**c**) $i = j - n$, and (**d**) $i = k_s$. In (**d**) the individual segments are highlighted in green with the corresponding keyframe $k_s$ marked with a red box.

Note that there needs to be a minimum of two poses in the resulting relative transformations to solve for the unknowns in (2), and there must be sufficient rotation and translation between the pairs $(i, j)$.

## 4. Simulation Experiments

### 4.1. Data Generation

For the purpose of simulation experiments, a random base trajectory is first generated for the primary sensor. The trajectory for the second sensor is then generated based on the sensor-to-sensor ground truth transformation $_{s1}\mathbf{T}_{gt}^{s2}$, after which both trajectories are subjected to different noise patterns to be expected in trajectories extracted using SLAM algorithms. The simulation experiments are based on 38 of such random trajectories.

The noise patterns used in the study were

- Gaussian noise with varying $\sigma^2$, where the trajectories are right-multiplied with $[\tilde{\mathbf{R}}, \tilde{t}]_T$. The rotation matrix $\tilde{\mathbf{R}}$ is formed from Euler angles drawn from $\mathcal{N}(0, 2a)$ and the translation components in $\tilde{t}$ are drawn from $\mathcal{N}(0, a)$, where $a = \sigma^2$ rad or m respectively.
- random jumps in $[x, y, z]$ -coordinates, where the translation components in $\tilde{t}$ are drawn from $\mathcal{N}(0, 0.02)$. The variance $0.02$ m was selected because it is considerably larger than the noise levels in the previous point
- drift on randomly selected principal axis with increasing severity determined by *drift rate*, i.e., how many meters per meter moved the error increases.

A range of each noise pattern is applied to all of the 38 generated base trajectories.

In addition to the above noise patterns individually, the algorithms were tested using a mixture of all three, which could be argued to be the most realistic case. For the mix of noises, the median values of the tested ranges were used for the noise components, namely $\sigma^2 = 5 \times 10^{-3}$ m or rad, 5% outliers, and drift rate $0.025$ m/m. Usually, SLAM algorithm performance is reported using absolute trajectory error (ATE) [15]. The average RMS ATE for various vision based algorithms reported in [16] ranges from $0.035$ m to $0.601$ m. The average RMS ATE for the mixed noise trajectories is $0.364$ m, so the selection of parameters can be considered reasonable. The authors did not manage to find a source for the typical composition of noise, and it is expected to vary based on the environment and sensors involved.

### 4.2. Results

We ran the calibration on the generated simulation data using all six calibration methods in the comparison. Each calibration method was tested using six different reference frame selection methods $A$, $B1$, $B5$, $B10$, $C5$, and $C10$, where the labels refer to the methods presented in Section 3.4. The calibration results are compared using the error metrics presented in Section 3.3. Throughout the experiments, the parameters $c = 0.01$ and $d = \frac{|S|}{2}$ were used for DNLO. The values were empirically chosen using the simulation data.

#### 4.2.1. Comparing Reference Selection Methods

Due to limited space to present the extensive results, we provide the comparison of all reference selection methods only for our DNLO in Figure 3, and later compare the best performing variants for each algorithm. The extensive results for all methods and algorithms are provided as Supplementary Materials. The graphs present the mean and 95% confidence intervals of the calibration errors for all 38 trajectories, except for the mixed noise case, where we present the median and quartiles.
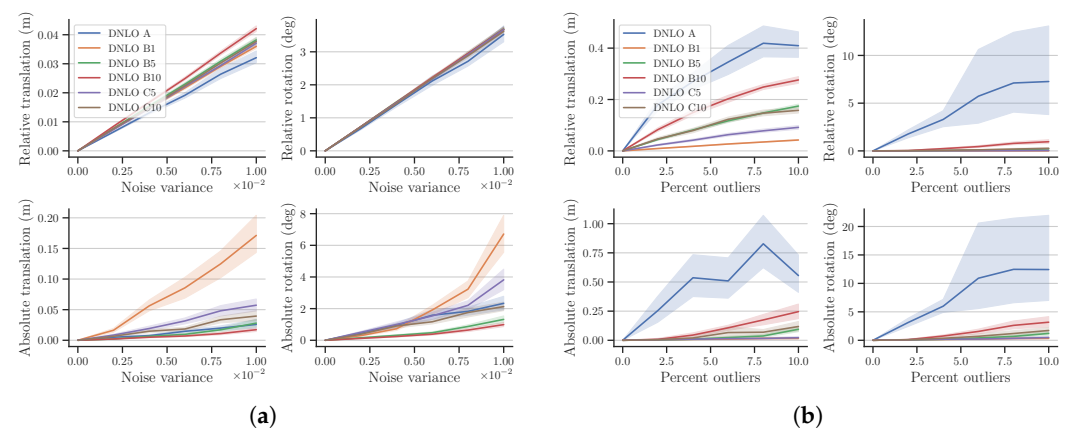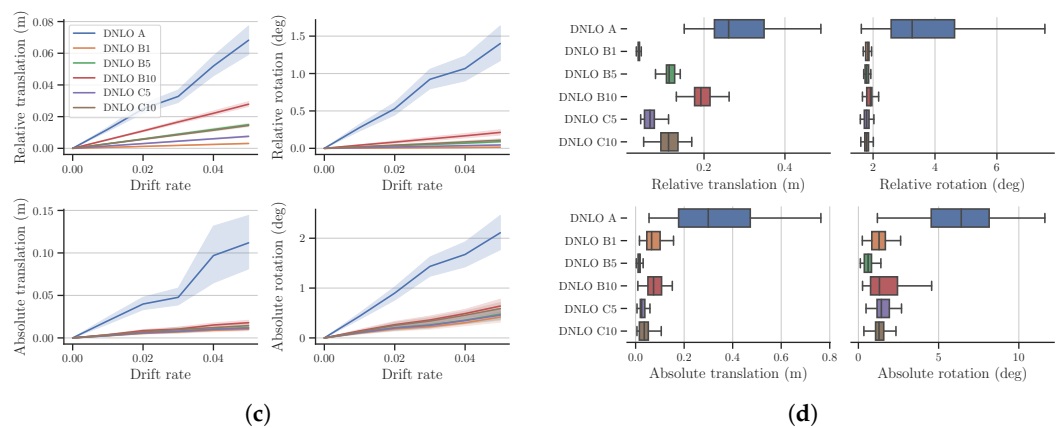


**Figure 3.** *Cont.*

(c)            (d)

**Figure 3.** Calibration errors on simulation data for all reference selection methods on DNLO: (**a**) added Gaussian noise, (**b**) added outliers, (**c**) added drift, and (**d**) mixed noise. Plots (**a**) through (**c**) display the mean and 95% confidence intervals, whereas the boxplot (**d**) shows the median and quartiles.

Firstly, it can be verified that without noise, the reference selection strategy makes no difference. Further, the results mostly follow expectations where longer reference distances are better with added noise, but worse with added outliers and drift. However, it is worth noticing that preprocessing option *A*, providing the largest transformations, is not the best option for even Gaussian noise, and noticeably worse for the other noise types. Option *A* likely makes the relative transformations too unbalanced for the calibration, as seen from Figure 4. The data are generated with a ground vehicle in mind, so the transformations are heavily partial to the *x*, *y*, and *yaw* components.
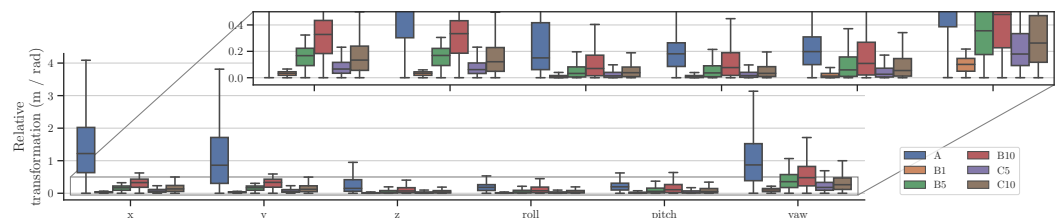


**Figure 4.** Median and quartiles of the absolute components of the resulting relative transformations when using the studied reference frame selection methods on the noiseless base trajectories of the simulation data.

For added outliers and drift, the performance of the selection strategies is ordered from those providing the smallest transformations to those providing the largest. In case of DNLO, the outliers are also more prominent when using the short reference *B1*, and therefore easier to reject. For added drift, the differences between the different strategies are minute, apart from option *A*. The mixed noise case in Figure 3d provides the most interesting notion, where *B1* is clearly best w.r.t. the relative errors, but not w.r.t. the absolute error, which is the true measure of calibration performance. This is noteworthy when selecting possible reference frames, as under normal operating conditions $_{s1}\mathbf{T}_{gt}^{s2}$ is unknown and one can only reason using the available relative error metrics. This means that the relative error metrics can not be used as a reliable selection criteria when choosing suitable reference frames.

### 4.2.2. Comparing Best Performing Algorithms

The best performing variant, w.r.t. $e_{at}$ (8), of each algorithm are compared in Figure 5 and follow a similar pattern as already discussed regarding DNLO. In the case of added Gaussian noise, preprocessing option *B10* performs best for all algorithms, with the exception of Zhuang [7]. The likely cause is the general reasoning that larger transformations between the points are less susceptible to noise.
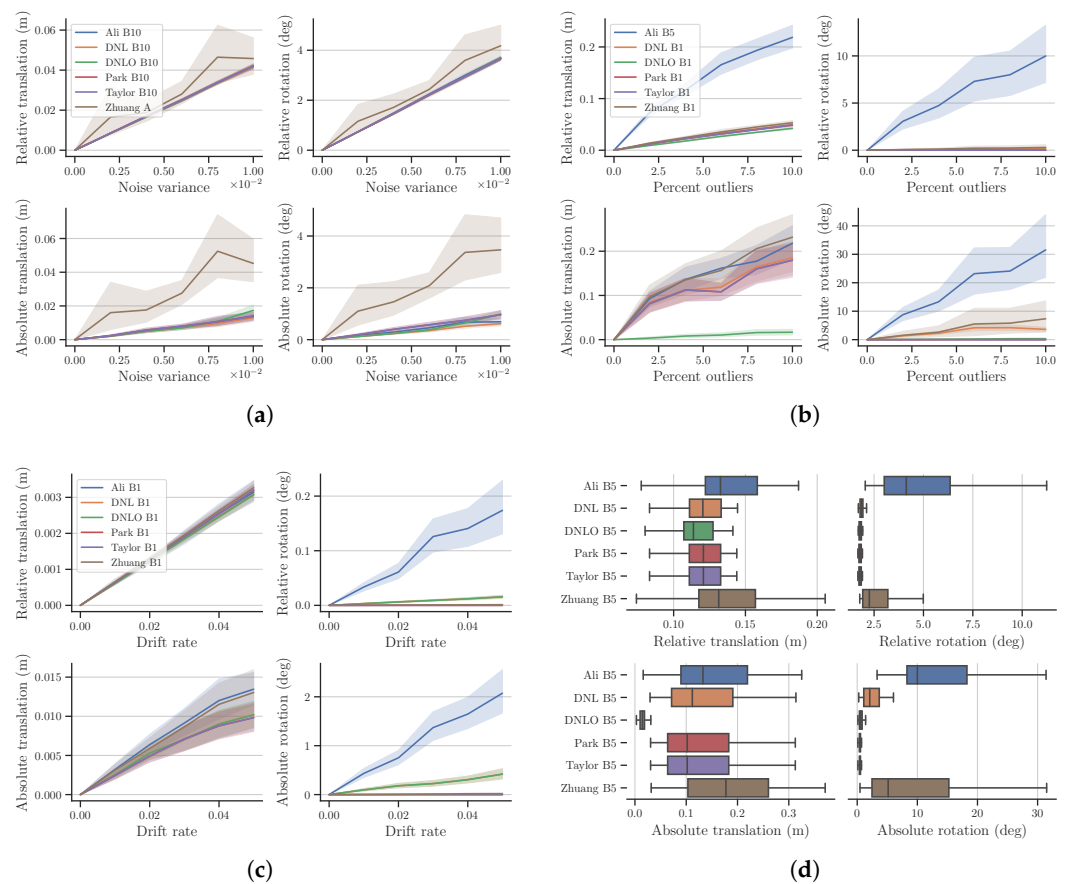
**Figure 5.** Calibration errors on simulation data for best performing variant of each algorithm: (**a**) added Gaussian noise, (**b**) added outliers, (**c**) added drift, and (**d**) mixed noise. Plots (**a**) through (**c**) display the mean and 95% confidence intervals, whereas the boxplot (**d**) shows the median and quartiles.

For added outliers, the *de facto* preprocessing *B*1 again provides the best results for most algorithms. However, it is clear that the added outlier rejection of DNLO outperforms the others. The poor performance of Ali [10] is likely explained by the rotation error not directly affecting the cost. The very small rotation errors in the *separable* algorithms Park [6] and Taylor [4], in both added outliers and drift, are due to the simulated noise only affecting the Cartesian coordinates. As the *separable* algorithms solve rotation first, it is unaffected by the noise. For added drift, there is again very little difference between the tested reference selection methods, apart from *A*, for any of the algorithms. Under the mixed noise conditions *B*5 performed best for all algorithms. This was to be expected, as it provides a balance between the short reference distance, favoring added outliers and drift, and the longer ones favoring added Gaussian noise. As with DNLO, upon selecting the possible reference frames, *B*1 appears better using the error metrics available.

## 5. Experiments on KITTI Data

To validate the findings of the simulation tests with real data, we perform experiments on the KITTI dataset [1], which is widely used and publicly available. The dataset, collected by driving around residential areas, contains data from two stereo cameras, a Velodyne HDL-64E lidar, and a GPS/INS system. We selected two drive sequences from the dataset, `2011_10_03_drive_0027`, a longer sequence with over 4000 frames, and `2011_09_30_drive_0027`, a shorter sequence with around 1000 frames. The first was selected as it is also used in [4], and the latter was randomly selected to validate that the hand–eye calibration method works with varying amounts of data.

As the ability to calibrate multimodal sensor-setups is one of the main draws of the demonstrated hand–eye calibration methods, we test both camera to lidar and camera to camera calibration. The calibration values provided in the dataset are used as ground truth values for computing the absolute error metrics.

### 5.1. Trajectory Generation

To generate the trajectories, the KITTI raw data was first exported into ROS bag format to allow the use of existing ROS implementations of the chosen SLAM algorithms. ORB SLAM3 [16] was used for stereo visual SLAM for both the grayscale and color stereo cameras available in the dataset, and HDL graph SLAM [17] for LIDAR-based SLAM, using the point clouds from Velodyne HDL-64E. The trajectories were generated based on the keyframe poses after the optimization step in both algorithms.

ORB SLAM3 is currently considered to be the state-of-the-art visual and visual-inertial SLAM system for stereo cameras. It is a bundle adjustment based SLAM that uses ORB features and descriptors matching based tracking in the visual front-end. Its multi-map system allows superior mid-term and long-term data association necessary for relocalization when the tracking is lost and loop closure detection. HDL graph SLAM is a pose graph-based method, where GICP based scan matching is used for generating odometry. It utilizes distance-based loop closure detection, where GICP is again used for scan matching, and loop closure matches as posegraph constraints.

### 5.2. Results

As with the simulation data, we focus on the best performing variant of each calibration algorithm. The calibration errors on camera to lidar calibration on the short residential data sequence are presented in Table 1. Zhuang [7], while having the lowest relative errors, converges to a wrong solution. All other algorithms perform best when provided with larger relative transformations, *B*5 or *B*10.

Similarly, the best performing variants for the camera to camera calibration of the long residential sequence, presented in Table 2, are those using the midlength steps in reference selection. For DNLO, the keyframe based selection *C*5 is marginally better than *B*5, which performs best for all other algorithms.

**Table 1.** Camera to lidar calibration on KITTI `2011_09_30_drive_0027`.

| Method | Relative Error | | Absolute Error | | Improvement Over B1 | |
|---|---|---|---|---|---|---|
| | $e_{rt}$ (m) | $e_{rR}$ (deg) | $e_{at}$ (m) | $e_{aR}$ (deg) | $e_{at}$ (m) | $e_{aR}$ (deg) |
| Ali B5 | 0.170 | 0.294 | 0.342 | 0.722 | 0.288 | −0.056 |
| DNL B5 | 0.170 | 0.293 | 0.342 | 0.721 | 0.288 | −0.052 |
| DNLO B10 | 0.271 | 0.488 | 0.202 | **0.232** | 0.128 | 0.209 |
| Park B10 | 0.325 | 0.473 | **0.183** | 0.849 | 0.435 | −0.102 |
| Taylor B10 | 0.325 | 0.473 | **0.183** | 0.849 | 0.435 | −0.102 |
| Zhuang B1 | **0.055** | **0.152** | 1.063 | 2.899 | - | - |

Best performing variants highlighted in bold.

**Table 2.** Camera to camera calibration on KITTI `2011_10_03_drive_0027`.

| Method | Relative Error | | Absolute Error | | Improvement Over B1 | |
|---|---|---|---|---|---|---|
| | $e_{rt}$ (m) | $e_{rR}$ (deg) | $e_{at}$ (m) | $e_{aR}$ (deg) | $e_{at}$ (m) | $e_{aR}$ (deg) |
| Ali B5 | 0.155 | 0.181 | **0.074** | 0.432 | 0.104 | 0.018 |
| DNL B5 | 0.155 | 0.181 | **0.074** | 0.432 | 0.104 | 0.018 |
| DNLO C5 | **0.076** | **0.149** | 0.159 | **0.345** | 0.034 | 0.071 |
| Park B5 | 0.157 | 0.181 | 0.078 | 0.351 | 0.111 | 0.075 |
| Taylor B5 | 0.157 | 0.181 | 0.078 | 0.351 | 0.111 | 0.075 |
| Zhuang B5 | 0.155 | 0.181 | **0.074** | 0.400 | 0.106 | −0.081 |

Best performing variants highlighted in bold.

It is noteworthy, that the performance gained by choosing a suitable reference frame in most cases outweighs the difference between the different calibration algorithms. The tests further demonstrate that using the *de facto* reference $i = j − 1$ is not the best option, but future work is needed to study how to best determine suitable spacing to the reference frame.

## 6. Conclusions

The aim of this paper was to study how different reference frame selection methods affect motion-based hand–eye sensor extrinsic calibration. With simulation experiments, we demonstrated that the type of noise (added Gaussian, outliers, or drift) present in the sensor trajectories affects which reference selection strategy performs best. Therefore, a single strategy optimal for all situations is difficult. The results mostly followed our hypotheses, where longer reference distances are better with added noise, but worse with added outliers and drift. However, notably the reference selection strategy typically used throughout previous studies was deemed to perform worse than others with realistic data. Moreover, this could only be observed with a known ground truth transformation and the relative error metrics available when selecting the reference frames can not be used as reliable selection criteria. These findings were validated with SLAM trajectories generated from the KITTI dataset [1]. In most cases, the calibration algorithm used had little impact on which reference selection method performed best, and the choice of reference frame selection had more impact on the calibration performance than the choice of the tested state-of-the-art calibration algorithm. This could be observed from experiments both with simulated and real world data. Further research is needed to determine how to best select the reference frames based on given trajectories.

We also proposed a cost for nonlinear optimization to mitigate the effect of outliers in the calibration. The method was evaluated with respect to state-of-the-art, and either matched or outperformed the other tested algorithms in most noise conditions.

**Supplementary Materials:** The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/s23073740/s1, Figure S1: Calibration errors for Ali [10] on simulation data for all reference selection methods: (**a**) added Gaussian noise, (**b**) added outliers, (**c**) added drift, and (**d**) mixed noise. Plots (**a**) through (**c**) display the mean and 95% confidence intervals, whereas the boxplot (**d**) shows the median and quartiles; Figure S2: Calibration errors for DNL on simulation data for all reference selection methods: (**a**) added Gaussian noise, (**b**) added outliers, (**c**) added drift, and (**d**) mixed noise. Plots (**a**) through (**c**) display the mean and 95% confidence intervals, whereas the boxplot (**d**) shows the median and quartiles; Figure S3: Calibration errors for Park [6] on simulation data for all reference selection methods: (**a**) added Gaussian noise, (**b**) added outliers, (**c**) added drift, and (**d**) mixed noise. Plots (**a**) through (**c**) display the mean and 95% confidence intervals, whereas the boxplot (**d**) shows the median and quartiles; Figure S4: Calibration errors for Taylor [4] on simulation data for all reference selection methods: (**a**) added Gaussian noise, (**b**) added outliers, (**c**) added drift, and (**d**) mixed noise. Plots (**a**) through (**c**) display the mean and 95% confidence intervals, whereas the boxplot (**d**) shows the median and quartiles; Figure S5: Calibration errors for Zhuang [7] on simulation data for all reference selection methods: (**a**) added Gaussian noise, (**b**) added outliers, (**c**) added drift, and (**d**) mixed noise. Plots (**a**) through (**c**) display the mean and 95% confidence intervals, whereas the boxplot (**d**) shows the median and quartiles; Table S1: Camera to lidar calibration on 2011_09_30_drive_0027 using Ali [10]; Table S2: Camera to lidar calibration on 2011_09_30_drive_0027 using DNL; Table S3: Camera to lidar calibration on 2011_09_30_drive_0027 using DNLO; Table S4: Camera to lidar calibration on 2011_09_30_drive_0027 using Park [6]; Table S5: Camera to lidar calibration on 2011_09_30_drive_0027 using Taylor [4]; Table S6: Camera to lidar calibration on 2011_09_30_drive_0027 using Zhuang [7]; Table S7: Camera to camera calibration on 2011_10_03_drive_0027 using Ali [10]; Table S8: Camera to camera calibration on 2011_10_03_drive_0027 using DNL; Table S9: Camera to camera calibration on 2011_10_03_drive_0027 using DNLO; Table S10: Camera to camera calibration on 2011_10_03_drive_0027 using Park [6]; Table S11: Camera to camera calibration on

`2011_10_03_drive_0027` using Taylor [4]; Table S12: Camera to camera calibration on `2011_10_03_drive_0027` using Zhuang [7].

## Abbreviations

The following abbreviations are used in this manuscript:

| | |
|---|---|
| SLAM | Simultaneous Localization and Mapping |
| DNL | Direct non-linear |
| DNLO | Direct non-linear with outlier rejection |
| RMS | Root-mean-square |
| ATE | Absolute trajectory error |

## References

1. Geiger, A.; Lenz, P.; Stiller, C.; Urtasun, R. Vision meets robotics: The KITTI dataset. *Int. J. Robot. Res.* **2013**, *32*, 1231–1237. [CrossRef]
2. Koide, K.; Menegatti, E. General Hand-Eye Calibration Based on Reprojection Error Minimization. *IEEE Robot. Autom. Lett.* **2019**, *4*, 1021–1028. [CrossRef]
3. Pachtrachai, K.; Vasconcelos, F.; Edwards, P.; Stoyanov, D. Learning to Calibrate—Estimating the Hand-eye Transformation Without Calibration Objects. *IEEE Robot. Autom. Lett.* **2021**, *6*, 7309–7316. [CrossRef]
4. Taylor, Z.; Nieto, J. Motion-based calibration of multimodal sensor arrays. In Proceedings of the 2015 IEEE International Conference on Robotics and Automation (ICRA), Seattle, WA, USA, 26–30 May 2015; pp. 4843–4850. [CrossRef]
5. Taylor, Z.; Nieto, J. Motion-Based Calibration of Multimodal Sensor Extrinsics and Timing Offset Estimation. *IEEE Trans. Robot.* **2016**, *32*, 1215–1229. [CrossRef]
6. Park, C.; Moghadam, P.; Kim, S.; Sridharan, S.; Fookes, C. Spatiotemporal Camera-LiDAR Calibration: A Targetless and Structureless Approach. *IEEE Robot. Autom. Lett.* **2020**, *5*, 1556–1563. [CrossRef]
7. Zhuang, H.; Qu, Z. A new identification Jacobian for robotic hand/eye calibration. *IEEE Trans. Syst. Man Cybern.* **1994**, *24*, 1284–1287. [CrossRef]
8. Pedrosa, E.; Oliveira, M.; Lau, N.; Santos, V. A General Approach to Hand-Eye Calibration Through the Optimization of Atomic Transformations. *IEEE Trans. Robot.* **2021**, *37*, 1619–1633. [CrossRef]
9. Pinto de Aguiar, A.S.; Riem de Oliveira, M.A.; Pedrosa, E.F.; Neves dos Santos, F.B. A Camera to LiDAR calibration approach through the optimization of atomic transformations. *Expert Syst. Appl.* **2021**, *176*, 114894. [CrossRef]
10. Ali, I.; Suominen, O.; Gotchev, A.; Morales, E.R. Methods for Simultaneous Robot-World-Hand-Eye Calibration: A Comparative Study. *Sensors* **2019**, *19*, 2837. [CrossRef] [PubMed]
11. Fischler, M.A.; Bolles, R.C. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM* **1981**, *24*, 381–395. [CrossRef]
12. Tabb, A.; Ahmad Yousef, K.M. Solving the robot-world hand–eye(s) calibration problem with iterative methods. *Mach. Vis. Appl.* **2017**, *28*, 569–590. [CrossRef]

13. Andersson, J.A.E.; Gillis, J.; Horn, G.; Rawlings, J.B.; Diehl, M. CasADi: A software framework for nonlinear optimization and optimal control. *Math. Program. Comput.* **2019**, *11*, 1–36. [CrossRef]

14. Schneider, T.; Li, M.; Cadena, C.; Nieto, J.; Siegwart, R. Observability-Aware Self-Calibration of Visual and Inertial Sensors for Ego-Motion Estimation. *IEEE Sensors J.* **2019**, *19*, 3846–3860. [CrossRef]

15. Sturm, J.; Engelhard, N.; Endres, F.; Burgard, W.; Cremers, D. A benchmark for the evaluation of RGB-D SLAM systems. In Proceedings of the 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems, Vilamoura-Algarve, Portugal, 7–12 October 2012; pp. 573–580. [CrossRef]

16. Campos, C.; Elvira, R.; Rodriguez, J.J.G.; M. Montiel, J.M.; D. Tardos, J. ORB-SLAM3: An Accurate Open-Source Library for Visual, Visual-Inertial, and Multimap SLAM. *IEEE Trans. Robot.* **2021**, *37*, 1874–1890. [CrossRef]

17. Koide, K.; Miura, J.; Menegatti, E. A portable three-dimensional LIDAR-based system for long-term and wide-area people behavior measurement. *Int. J. Adv. Robot. Syst.* **2019**, *16*, 1–16. [CrossRef]