Probabilistic Framework for Hand–Eye and Robot–World Calibration $AX = YB$

Junhyoung Ha, *Member, IEEE*

***Abstract*—Hand–eye and robot–world calibration is a problem in which the unknown homogeneous transformations X and Y must be estimated for a loop closure equation $AX = YB$ for a set of transformation measurement pairs $\{(A_i, B_i)\}$. Previous studies on $AX = YB$ have mainly relied on linear least-squares minimization followed by nonlinear iterative optimization for solution refinement to minimize the distances between A_iX and YB_i. However, these methods have not been fully clarified, particularly in terms of calibration dependence on the coordination of A, B, X, and Y along the system loop, as well as the underlying noise distributions of A_i and B_i. They also lack flexibility in the noise properties of individual measurements; thus, they cannot incorporate the relative reliability between measurements. To address these limitations, we propose a probabilistic framework for hand–eye and robot–world calibration. The proposed framework clarifies the unclear aspects of existing methods by revealing their underlying assumptions regarding system noise. Consequently, it identifies the applicability of distance minimization to a given calibration problem and provides the optimal coordination of transformations for distance minimization. For cases in which distance minimization is inapplicable, an iterative algorithm for the maximum likelihood estimation is proposed, whereby the different noise properties of individual measurements can be accounted for. An estimation uncertainty analysis is presented for the proposed iterative algorithm to quantify the expected estimation accuracy. The presented theories and the proposed algorithm are validated using a set of numerical and hardware experiments. The code for the iterative algorithm and the estimation uncertainty is available at https://github.com/hjhdog1/probabilisticAXYB.**

***Index Terms*—Frame calibration, hand–eye and robot–world calibration, hand–eye calibration, maximum likelihood estimation.**

I. Introduction

HAND–EYE calibration is a common problem in robotics and computer vision that arises when a robot manipulator and a camera are integrated with an unknown relative transformation [1], [2], [3]. Two versions of this problem exist: 1) the traditional formulation of hand–eye calibration $A_iX = XB_i$ and 2) simultaneous hand–eye and robot–world calibration $A_iX = YB_i$. Here, A_i, B_i, X, and Y are 4×4 homogeneous rigid-body transformation matrices that belong to the special Euclidean group $SE(3)$. In these problems, transformations X and Y are identified using the transformation matrix pairs $\{(A_i, B_i)\}$, which are obtained from camera measurements and robot transformations. The more general notions of these problems are "*one-frame sensor calibration*" and "two-frame sensor calibration" [4], respectively, based on the fact that loop closure equations appear not only for robots with cameras but also for any integration of two transformation measurement systems.

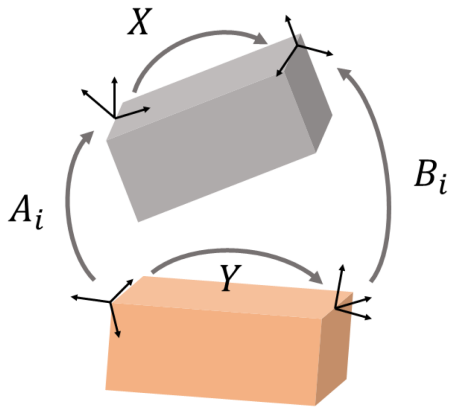


Fig. 1. Two rigid bodies incorporated with two measurement systems. The loop closure equation is formulated as $A_iX = YB_i$.

The two-frame sensor calibration problem is physically interpreted in terms of two rigid bodies in relative motion, each of which has two body frames [4], as illustrated in Fig. 1. In this case, X and Y are the unknown constant transformations defined on each rigid body, whereas A_i and B_i are measurement transformations that vary with respect to the relative motion of the two rigid bodies. The two-frame sensor calibration is reduced to a one-frame sensor calibration $A_iX = XB_i$ when the two rigid bodies are identical or when the loop closure equations of two measurement pairs (A_i, B_i) and (A_j, B_j) are merged to eliminate Y as $A_i^{-1}A_jX = XB_i^{-1}B_j$.

Extensive studies on the two-frame sensor calibration problem can be found in the literature. When no measurement noise exists in $\{(A_i, B_i)\}$, the algebraic solutions are those presented in [4], [5], [6]. In the presence of measurement noise,

Manuscript received 23 May 2022; revised 10 August 2022; accepted 24 August 2022. This work was supported in part by Korea Medical Device Development Fund grant funded by the Korea government (the Ministry of Science and ICT, the Ministry of Trade, Industry and Energy, the Ministry of Health & Welfare, the Ministry of Food and Drug Safety) under Grant RS-2020-KD000090 and in part by the Korea Institute of Science and Technology (KIST) Institutional Program under MIDAS2 Grant 2E31572. This paper was recommended for publication by Associate Editor J. Kelly and Editor F. Chaumette upon evaluation of the reviewers comments.

The author is with the Center for Healthcare Robotics, Artificial Intelligence and Robotics Institute, Korea Institute of Science and Technology, Seoul 02792, South Korea (e-mail: hjhdog1@gmail.com).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TRO.2022.3214350.

Digital Object Identifier 10.1109/TRO.2022.3214350

transformations X and Y that best fit the equation $A_iX = YB_i$ should be determined. This is generally formulated as an optimization problem using various objective functions and representations of the elements of $SE(3)$. In [3] and [5], the rotation parts of X and Y were represented using unit quaternions. Furthermore, a linear least-squares minimization was formulated with the quaternion representation. Dual quaternion representations were used for homogeneous transformations to simultaneously identify the rotation and translation components [7], whereas a method based on the Kronecker product was proposed in [8]. Moreover, global polynomial optimization was applied to this problem using convex linear matrix inequality relaxations [9]. Nonlinear iterative solvers were used in [4] and [10] to solve X and Y simultaneously for $AX = YB$. In [11], a least-squares solution to $AX = YB$ was derived for a relaxed orthogonality constraint. Reprojection error of calibration patterns was considered as an objective in [12], [13], [14]. A reprojection error-based approach was also presented in [15], where hand pose uncertainties were additionally accommodated. An intensive comparative study was presented in [14] to assess various objectives and algorithms.

Although many of these studies relied on linear least-squares minimization, in which an analytic minimizer was available, they posteriorly refined the linear solutions through iterative local optimization [5], [7], [10], [11] as follows:

$$\min_{X,Y\in SE(3)} \sum_i d(A_iX, YB_i)^2. \tag{1}$$

This optimization could also be directly solved using global optimization frameworks [4], [9]. In the above, $d(\cdot,\cdot)$ denotes the distance between the two transformations, which varies between the methods. The objective of this minimization is to further reduce the distance between the left and right sides of the loop closure equation $A_iX = YB_i$, which appears to be perfectly reasonable.

However, the true meaning of this minimization has not been fully clarified. For details, we refer to Fig. 1. Although the transformations A_i, B_i, X, and Y have already been coordinated in this figure, they can be coordinated in different ways. There are three other possible options for the coordination that also yield the identical loop closure equation as $A_iX = YB_i$, which can be imagined by inverting and/or swapping the transformations. Then, the following question arises: Which coordination gives the best calibration result? For further discussion, we examine a coordination that differs from that in Fig. 1. From the optimization perspective, selecting a different coordination is equivalent to swapping and inverting the transformations in the objective function, for example,

$$\min_{X,Y\in SE(3)} \sum_i d(B_iX^{-1}, Y^{-1}A_i)^2. \tag{2}$$

The distance $d(B_iX^{-1}, Y^{-1}A_i)$ is different from $d(A_iX, YB_i)$ because there is no bi-invariant distance on $SE(3)$ [16]. This suggests that four different optimizations are available, which all differ in terms of the resulting solutions. However, it has never been clarified which solution yields the best calibration result.

Another unclear aspect of the distance minimization approaches is the individual noise properties of the measurements. For example, suppose A_i is a noisy measurement, whereas B_i is relatively noiseless. We would like to trust B_i over A_i, but it is impossible to do so in the distance minimization framework. In fact, it is even unclear what assumptions on the system noise underlie the distance minimization.

To address these ambiguities and achieve optimal calibration results for various system noise configurations, we propose a probabilistic framework for hand–eye and robot–world calibration or, more generally, two-frame sensor calibration. The proposed framework differs from the previous probabilistic approaches [17], [18], [19], which probabilistically recover X and Y under missing correspondences between $\{A_i\}$ and $\{B_j\}$. Our study assumes known correspondences, with which the proposed framework is formulated to achieve the optimal calibration result based on the noise probabilistic distributions of individual measurements. A notable prior work includes [20], where the loop-closure error was selected as the likelihood function, and the optimal positional weight in the error function was estimated iteratively. In our study, we further investigate the system noise properties and aim to find the correct likelihood function that incorporates various system noise configurations. The main contributions of this study are as follows.

1) Three possible noise configurations in the two-frame sensor calibration problem are identified in the context of the proposed framework. Subsequently, we formulate the maximum likelihood estimation for each class of noise configuration.
2) The existing distance minimization approaches are clearly interpreted within the proposed framework by revealing their underlying assumptions on system noise; that is, A is noiseless, whereas B is noisy with right-translated isotropic noise. This finding not only helps identify the applicability of the distance minimization to a given calibration problem but also enables the optimal coordination of transformations for distance minimization approaches.
3) This work presents an iterative algorithm that maximizes the likelihood function using the exact analytic gradients, which is applicable to a system in which the assumptions of the distance minimization approaches are not valid. The rotational and positional noises of measurements A_i and B_i in our formulation are modeled in a general form as anisotropic noise. The proposed algorithm allows individual noise distributions for each measurement.
4) The estimation uncertainty covariance is derived to quantify the expected estimation accuracy of the proposed algorithm.

The proposed algorithm and theories were validated using both numerical and hardware experiments.

The remainder of this article is organized as follows. The three possible noise configurations in the two-frame sensor calibration problem are described in Section II. The maximum likelihood estimation for each noise configuration is formulated in Section III. Thereafter, the existing distance minimization approaches are explained in terms of the proposed probabilistic framework in Section IV. The iterative algorithm for the maximum likelihood

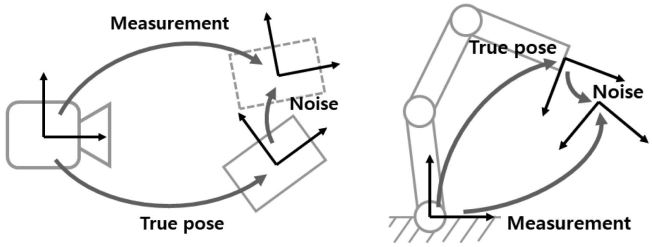


Fig. 2. Examples of noisy measurements. (a) Object tracking through the vision system. (b) Robot hand localization by robot kinematics.

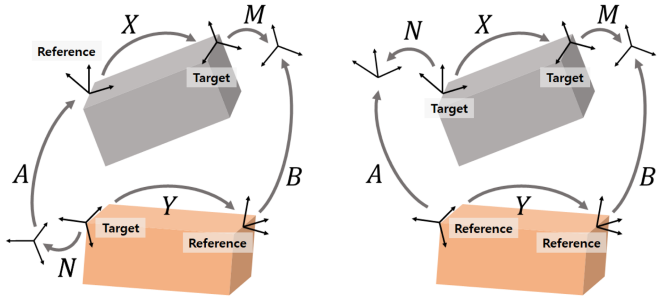


Fig. 3. Two possible configurations of measurement systems. (a) Each reference frame is on each rigid body. (b) Two reference frames are on the same rigid body. The transformations N and M represent the measurement noise.

estimation and its estimation uncertainty analysis are presented in Sections VI and VII, respectively, followed by numerical and hardware experiments in Sections VIII and IX, respectively. Section X concludes the article.

II. Noise Configurations in the Hand–Eye and Robot–World Calibration Problem

The measurements A_i and B_i are homogeneous rigid-body transformations that can be acquired by various devices. Fig. 2 shows a few examples of transformation measurements in the presence of measurement noise. When an object is localized through a vision system, as indicated in Fig. 2(a), a localization error exists, which is represented by the relative transformation from the true pose to the measured pose. In Fig. 2(b), transformation of the robot hand is acquired using robot kinematics. Pose errors also exist in this case because of the actuation error and unmodeled factors (e.g., joint and link compliance). The same situation arises for any system that estimates transformations, such as electromagnetic (EM) sensors. Throughout this article, these error transformations are considered as measurement noises, which are random variables drawn from their underlying probabilistic density functions (PDFs). The error properties are typically provided by the manufacturers as device specifications, or the error distributions for various vision systems and localization methods can be found in the literature [17], [21], [22], [23].

Prior knowledge of the noise properties can be utilized to obtain optimal estimations of X and Y. To do so, we first identify the possible noise configurations in the two-frame sensor calibration problem. We remark again that the calibration problem can be interpreted as two measurement systems working on and with two rigid bodies, as illustrated in Fig. 1. For each measurement system, let "reference frame" refer to the coordinate frame of a measurement system (e.g., the camera frame, robot base frame, and EM field frame) and let "target frame" refer to the frame under observation (e.g., the object body frame, robot hand frame, and EM sensor frame). Two possible configurations of the frames on the two rigid bodies exist: i) Each reference frame is on each rigid body and ii) the two reference frames are on the same rigid body. These two configurations are illustrated in Fig. 3(a) and (b), respectively. Note that the noise transformations $N, M \in SE(3)$ are present at the target frames, which are drawn from the PDFs $p(N)$ and $P(M)$.

In the presence of measurement noise, equation $AX = YB$ does not strictly hold. Alternatively, from Fig. 3, the two configurations i) and ii) can be expressed by the following strict equations:

$$NAX = YBM^{-1} \tag{3}$$

$$AN^{-1}X = YBM^{-1}. \tag{4}$$

As a special case of these configurations, one of the measurement systems can be considered noiseless. Without loss of generality, this case is established by coordinating the noisy measurement as B and setting $N = I$ in one of the above equations, which yields the following strict equation:

$$AX = YBM^{-1}. \tag{5}$$

This case is handled separately as the third noise configuration. Thus, the three noise configurations that are considered in this study are summarized as follows.

1) Noise configuration 1: Defined as (3) and Fig. 3(a).
2) Noise configuration 2: Defined as (4) and Fig. 3(b).
3) Noise configuration 3: Defined as (5). Fig. 3(a) and (b) corresponds to this noise configuration when $N = I$.

III. Probabilistic Formulation of Hand–Eye and Robot–World Calibration

The calibration problem can be formulated as a maximum likelihood estimation for each noise configuration presented in the previous section.

A. Noise Configuration 1: $N_iA_iX = YB_iM_i^{-1}$

Starting with noise configuration 1, (3) associated with a measurement pair (A_i, B_i) is given as follows:

$$N_iA_iX = YB_iM_i^{-1} = C_i \tag{6}$$

where $C_i \in SE(3)$ is a transformation variable introduced to express N_i and M_i in the following decoupled forms:

$$N_i = C_iX^{-1}A_i^{-1} \tag{7}$$

$$M_i = C_i^{-1}YB_i. \tag{8}$$

For given X, Y, and $\{C_i\}$, the likelihood function $L(X, Y, \{C_i\})$ is calculated by

$$L(X, Y, \{C_i\}) = \prod_i p(N_i)p(M_i). \tag{9}$$

The probabilistic hand–eye and robot–world calibration is then formulated as the maximization of (9) with respect to X, Y, and $\{C_i\}$; that is,

$$\max_{X,Y,C_i \in SE(3)} L(X, Y, \{C_i\}) \tag{10}$$

subject to (7) and (8), where $L(\cdot)$ is defined in (9).

B. Noise Configuration 2: $A_i N_i^{-1} X = Y B_i M_i^{-1}$

A similar maximization is derived for noise configuration 2, where the loop-closure equation is expressed as follows:

$$A_i N_i^{-1} X = Y B_i M_i^{-1} = C_i. \tag{11}$$

Compared to noise configuration 1, M_i remains identical to (8), whereas N_i is replaced with

$$N_i = X C_i^{-1} A_i. \tag{12}$$

The maximum likelihood estimation is then expressed as (10), where N_i and M_i are given in (12) and (8), respectively.

C. Noise Configuration 3: $A_i X = Y B_i M_i^{-1}$

Finally, we consider noise configuration 3 given in (5). Because the only random variable in this case is M_i, the maximum likelihood estimation is reduced to the following:

$$\max_{X,Y \in SE(3)} L(X, Y) \tag{13}$$

for the reduced likelihood function

$$L(X, Y) = \prod_i p(M_i) \tag{14}$$

where M_i is given by

$$M_i = X^{-1} A_i^{-1} Y B_i. \tag{15}$$

Note that C_i does not appear in this case because it was introduced to decouple N_i and M_i in the previous cases.

IV. Connection to Existing Methods

In this section, we demonstrate that the probabilistic calibration problem can be reduced to the existing distance minimization methods under a particular setting of the noise PDFs. This reveals the hidden assumptions in the distance minimization methods and aids in understanding the existing methods better for correct adaptation to real-world problems.

A. Assumptions on System Noise in Existing Methods

The distance minimization in previous works [4], [5], [7], [9], [10], [11] is formulated as

$$\min_{X,Y \in SE(3)} \sum_i d(A_i X, Y B_i)^2 \tag{16}$$

where $d(\cdot, \cdot)$ is a distance function on $SE(3)$, which varies between the methods. Most of the previous works use left-invariant distance functions such that

$$d(T_0 T_1, T_0 T_2) = d(T_1, T2) \tag{17}$$

for any $T_1, T_2 \in SE(3)$ and $T_0 \in SE(3)$. This is a natural property of a distance function such that the distance is coordinate invariant; that is, the distance is preserved with respect to a global rigid-body transformation.

Now, we refer to noise configuration 3 discussed in Section III-C, where A_i is noiseless and B_i is noisy. Furthermore, we assume that the probability distribution of the noise transformation M_i is unimodal and centered at I. This is a natural noise property of a valid sensor; unbiased measurements are acquired around the ground truth. More precisely, $p(M_i)$ is assumed to decrease exponentially with respect to the squared distance between I and M_i according to

$$p(M_i) = k \exp\left(-\frac{d(I, M_i)^2}{\sigma^2}\right) \tag{18}$$

where $\sigma \in \mathbb{R}$ is an arbitrary positive scalar, and $k \in \mathbb{R}$ is a normalizing constant that satisfies

$$\int_{SE(3)} k \exp\left(-\frac{d(I, T)^2}{\sigma^2}\right) dT = 1. \tag{19}$$

We remark that the integration of $SE(3)$ is discussed in Appendix A, based on which the random samplings on $SE(3)$ are derived for our later numerical experiments. Now, substituting (15) into $d(I, M_i)$ and applying the left invariance of the distance function yields

$$d(I, M_i) = d(I, X^{-1} A_i^{-1} Y B_i) = d(A_i X, Y B_i). \tag{20}$$

Thus, the likelihood (14) is reduced to

$$L(X, Y) = k^n \exp\left(-\frac{1}{2\sigma^2} \sum_i d(A_i X, Y B_i)^2\right) \tag{21}$$

where n is the number of measurement pairs. By taking $\log$ on the above equation, the maximum likelihood estimation becomes

$$\max_{X,Y \in SE(3)} -\frac{1}{2\sigma^2} \sum_i d(A_i X, Y B_i)^2 + n \log k \tag{22}$$

which is identical to distance minimization in (16) when the constant factors $\frac{1}{2\sigma^2}$ and $n \log k$ are ignored. This result confirms that the distance minimization method is the maximum likelihood estimation under the assumption that A_i is noiseless and that B_i is noisy, with the PDF of the noise transformation expressed by (18). This type of noise is referred to as "isotropic noise" in this article, as the noise PDF only depends on the distance between I and M_i but is invariant to the direction of M_i with respect to I.

B. Distance Function Parameter in Existing Methods

A typical form of the distance function in various previous works is the one that decouples rotation and translation

$$d(T_1, T_2)^2 = d_{SO(3)}(R_1, R_2)^2 + \epsilon \|p_1 - p_2\|^2 \tag{23}$$

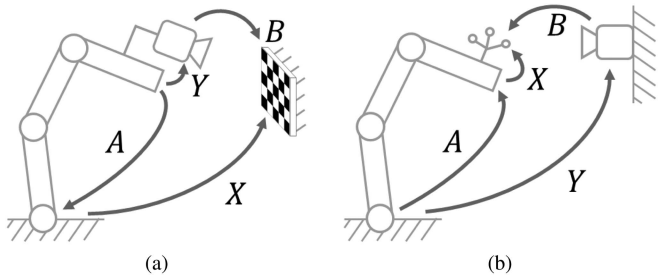


Fig. 4. Examples of combination of robot and measurement system. (a) Camera on robot hand. (b) Tracking system localizing the marker frame on the robot hand.

where $d_{SO(3)}(\cdot,\cdot)$ is a distance function on $SO(3)$, $\epsilon \in \mathbb{R}$ is a positive scalar, and $R_i \in SO(3)$ and $p_i \in \mathbb{R}^3$ for $i \in \{1,2\}$ denote the rotation and translation of the transformation $T_i \in SE(3)$, respectively. This distance becomes a left-invariant geodesic distance if $d_{SO(3)}(\cdot,\cdot)$ is the popular bi-invariant geodesic distance on $SO(3)$ [16], [24], [25]. In most previous studies [4], [5], [9], [10], [11] the Frobenius norm was used instead for the distance on $SO(3)$, that is, $d_{SO(3)}(R_1,R_2) = \|R_1 - R_2\|_F$, which is bounded equivalent to the bi-invariant geodesic distance on $SO(3)$ [26].

The only user parameter in the distance function (23) is the positional weight factor ϵ, whose physical meaning can be interpreted from the system noise perspective as follows: Substituting (23) to (18), the PDF $p(M_i)$ becomes

$$p(M_i) = k \exp\left(-\frac{d_{SO(3)}(I,R_{M_i})^2}{2\sigma^2}\right) \exp\left(-\frac{\epsilon}{2\sigma^2}\|p_{M_i}\|^2\right) \tag{24}$$

where $R_{M_i} \in SO(3)$ and $p_{M_i} \in \mathbb{R}^3$ are the rotation and translation of M_i, respectively. Given the above PDF, the rotational and translational noises are independent and isotropic, with standard deviations (STDs) of σ and $\sigma/\sqrt{\epsilon}$, respectively. Thus, ϵ represents the noise property of B, such that $1/\sqrt{\epsilon}$ is the STD of the translational noise relative to that of the rotational noise in B under the assumption that both rotational and translational noises are isotropic. Note that A is assumed to be noiseless.

C. Correct Adaptation of Existing Methods

Recall that there are always four possible choices for coordination of A, B, X, and Y for a given calibration problem. When a distance minimization approach is applied according to (16), it is implied that A is noiseless and B is noisy, the noise transformation of which is right-translated and isotropic. Thus, A, B, X, and Y should ideally be coordinated such that the noise configuration matches the implied assumption.

Two examples of $AX = YB$ are shown in Fig. 4. In Fig. 4(a), a camera is attached to the robot hand, and it captures a checkerboard. In Fig. 4(b), a set of markers is attached to the robot hand, and the marker frame is tracked using a tracking system. We assume that the robot kinematics is sufficiently accurate to be considered noiseless, and the measurements of the camera and tracker are relatively noisy. In these two examples, A must be coordinated to represent a noiseless transformation, that is, the robot transformation. However, the direction of A is yet to be determined for each case. Considering the assumption of the right-translated isotropic noise of B, it is natural for B to be coordinated such that its distal end points toward the target frame. Finally, all the transformations are coordinated accordingly, as shown in the figures.

In summary, the basic rule when using the distance minimization approach is to coordinate B as the dominantly noisy transformation with right-translated isotropic noise and coordinate A as the less noisy measurement. However, a challenging situation arises when both measurements contain similar levels of noise, in which case, the noise assumptions of the distance minimization approaches are no longer valid (e.g., the integration of two cameras). In Section VI, to handle this situation and, more generally, the situations where different noise distributions are given for individual A_i and B_i, an iterative algorithm is derived to maximize the likelihood function for given noise distributions of A_i and B_i.

V. Comparison to $\mathcal{A}X = X\mathcal{B}$ Formulation

The $AX = YB$ formulation reduces to the earlier $\mathcal{A}X = X\mathcal{B}$ formulation when Y is eliminated for two measurement pairs; two equations $A_iX = YB_i$ and $A_jX = YB_j$ merge to $A_i^{-1}A_jX = XB_i^{-1}B_j$. Then, the measurement pairs for $\mathcal{A}X = X\mathcal{B}$ are given as

$$\{(\mathcal{A}_{ij}, \mathcal{B}_{ij})\} = \{(A_i^{-1}A_j, B_i^{-1}B_j)\} \tag{25}$$

for $i \neq j$. Herein, we will discuss the $\mathcal{A}X = X\mathcal{B}$ formulation in two aspects: i) The noise assumption in distance minimization and ii) its probabilistic formulation.

The existing distance minimization approaches for $\mathcal{A}X = X\mathcal{B}$ [27], [28] can be expressed as

$$\min_{X \in SE(3)} \sum_{i \neq j} d(\mathcal{A}_{ij}X, X\mathcal{B}_{ij})^2. \tag{26}$$

In a similar manner as in the previous section, the above minimization with a left-invariant distance function can be interpreted as a maximum likelihood estimation under the assumption that $\mathcal{A}$ is noiseless and $\mathcal{B}$ contains right-translated isotropic noise. However, the actual measurement noise in $\mathcal{B}$ is more complicated than this; the noises are not independent between $\mathcal{B}_{ij}$ and $\mathcal{B}_{ik}$ because they share a common noise inherited from B_i. Noting that the likelihood function (9) is a product of individual noise PDFs because the noises are independent, the maximum likelihood estimation in this case will be different from (26) once the noise dependency is considered.

Another issue with the noise assumption is the right-translated isotropic noise. Note that we merged $A_iX = YB_i$ and $A_jX = YB_j$ as if they were strict equations. In practice, they do not hold because of measurement noises. In noise configuration 3, for example, the strict equations are given as $A_iX = YB_iM_i$ and $A_jX = YB_jM_j$, which merge to $A_i^{-1}A_jX = XB_i^{-1}M_i^{-1}M_jB_j$. Here, $\mathcal{B}_{ij}$ contains noise transformations between B_i^{-1} and B_j, which indicates that it is unrealistic to model the measurement noise to be simply right-translated isotropic. Along with the previously discussed noise dependency, this complicated noise property makes it challenging to find a correct probabilistic formulation.

Overall, in the $\mathcal{A}X = X\mathcal{B}$ formulation, i) the noise assumption of the distance minimization is difficult to meet, and ii) a probabilistic formulation is challenging for modeling the complicated noise property. Thus, we focused on the $AX = YB$ formulation in this study, while our experiments included a comparison of our approach to an existing $\mathcal{A}X = X\mathcal{B}$ approach [27].

VI. Gradient Ascent Algorithm for Log-Likelihood Maximization

Noise distribution was assumed to be isotropic in the distance minimization approaches. In this section, we further generalize the noise distribution to be anisotropic while remaining unimodal and use it to derive a gradient ascent algorithm for the maximum likelihood estimation. The translational part of the noise transformations is modeled using a multivariate Gaussian distribution in the following form:

$$p \sim \mathcal{N}(0_{3\times 1}, \Sigma_p) \tag{27}$$

where $\Sigma_p \in P(3)$ is the positive-definite covariance matrix. A similar extension is possible for the rotational part:

$$R \sim k \exp\left(-\frac{1}{2} w^T \Sigma_w^{-1} w\right) \tag{28}$$

where $w \in \mathbb{R}^3$ is defined by $[w] = \log R$, and $\Sigma_w \in P(3)$ is a 3×3 positive-definite matrix that plays a similar role to the covariance matrix in the multivariate Gaussian distribution. Here, the notation $[\cdot]$ denotes the 3×3 skew-symmetric matrix representation of a 3-D vector. In this case, k is the normalizing constant that satisfies

$$\int_{SO(3)} k \exp\left(-\frac{1}{2} w^T \Sigma_w^{-1} w\right) dR = 1 \tag{29}$$

where the integration measure dR is given in Appendix A as $dR = \frac{2-2\cos\|w\|}{\|w\|^2} dw_x dw_y dw_z$, as in (73). Notably, the PDF in (28) is different from the Gaussian distribution in the exponential coordinate

$$w \sim \mathcal{N}(0_{3\times 1}, \Sigma_w) \tag{30}$$

because the integration measure dR differs from $dw_x dw_y dw_z$. Further details can be found in Appendix B, in which a random sampling method for (28) is presented.

Given this noise model, the log-likelihood is obtained by taking log of (9), which yields

$$\begin{aligned}\log L(X, Y, \{C_i\}) = & -\frac{1}{2}\sum_i \Big(w_{N_i}^T \Sigma_{w_{N_i}}^{-1} w_{N_i} + p_{N_i}^T \Sigma_{p_{N_i}}^{-1} p_{N_i} \\ & + w_{M_i}^T \Sigma_{w_{M_i}}^{-1} w_{M_i} + p_{M_i}^T \Sigma_{p_{M_i}}^{-1} p_{M_i}\Big)\end{aligned} \tag{31}$$

where $w_{N_i} \in \mathbb{R}^3$ and $w_{M_i} \in \mathbb{R}^3$ are the rotational vectors of N_i and M_i, respectively, and $p_{N_i} \in \mathbb{R}^3$ and $p_{M_i} \in \mathbb{R}^3$ are the translational components of N_i and M_i, respectively; that is,

$$N_i = T(w_{N_i}, p_{N_i}),\ M_i = T(w_{M_i}, p_{M_i}) \tag{32}$$

with $T(\cdot,\cdot) \in SE(3)$ defined as

$$T(w, q) = \begin{bmatrix} \exp([w]) & q \\ 0_{1\times 3} & 1 \end{bmatrix} \tag{33}$$

for $w \in \mathbb{R}^3$ and $q \in \mathbb{R}^3$. Note that the noise covariances $\Sigma_{w_{N_i}}, \Sigma_{p_{N_i}}, \Sigma_{w_{M_i}}$, and $\Sigma_{p_{M_i}}$ represent the noise distributions of the individual measurements.

Subsequently, each iteration of the gradient ascent is given as

$$\begin{aligned} X &\leftarrow XT(\Delta w_X, \Delta q_X) \\ Y &\leftarrow YT(\Delta w_Y, \Delta q_Y) \\ C_i &\leftarrow C_i T(\Delta w_{C_i}, \Delta q_{C_i}) \end{aligned} \tag{34}$$

where $\Delta w_{(\cdot)} \in \mathbb{R}^3$ and $\Delta q_{(\cdot)} \in \mathbb{R}^3$ are scaled gradients in the rotation and translation of the subscripted variables, respectively. The gradients for each noise configuration are presented in the following subsections.

A. Noise Configuration 1: $N_i A_i X = Y B_i M_i^{-1} = C_i$

Here, the gradients of (31) with respect to X, Y, and C_i are derived for noise configuration 1. Note that as N_i and M_i are functions of X, Y, and C_i in (7) and (8), the algebraic derivation of the gradients of (31) involves complicated chain rule-based differentiation. Alternatively, the gradient of a function can be derived using the Taylor expansion. When a scalar-valued function $f(x)$ is defined for a variable x, the gradient $\partial f/\partial x$ is acquired as the coefficient of the first-order term in the Taylor expansion for a perturbation δx, as follows:

$$f(x + \delta x) = f(x) + \frac{\partial f}{\partial x}(x)\delta x + \cdots. \tag{35}$$

Note that a gradient derived in this manner is not an approximation but an exact formula because the first-order term of the Taylor expansion truly reflects the exact gradient.

In our case, $f(x)$ corresponds to the log-likelihood and x corresponds to $(X, Y, \{C_i\})$. For the Taylor expansion, suppose that perturbations occur in X by $XT(\delta w_X, \delta q_X)$, where $\delta w_X \in \mathbb{R}^3$ and $\delta q_X \in \mathbb{R}^3$ are the perturbations in rotation and translation, respectively, and Y and C_i are similarly perturbed. Again, $T(\cdot,\cdot)$ is defined in (33). The corresponding perturbations of N_i and M_i are denoted by $\delta w_{N_i}, \delta p_{N_i}, \delta w_{M_i} \in \mathbb{R}^3$, and $\delta p_{M_i} \in \mathbb{R}^3$ for rotations and translations of N_i and M_i, respectively. We refer to [29] for various examples of $SE(3)$ perturbations. Using the expressions of N_i and M_i in (32), in the presence of perturbations, (7) and (8) become

$$\begin{aligned} &T(w_{N_i} + \delta w_{N_i}, p_{N_i} + \delta p_{N_i}) \\ &\quad = C_i T(\delta w_{C_i}, \delta q_{C_i}) T(\delta w_X, \delta q_X)^{-1} X^{-1} A_i^{-1} \end{aligned} \tag{36}$$

$$\begin{aligned} &T(w_{M_i} + \delta w_{M_i}, p_{M_i} + \delta p_{M_i}) \\ &\quad = T(\delta w_{C_i}, \delta q_{C_i})^{-1} C_i^{-1} Y T(\delta w_Y, \delta q_Y) B_i. \end{aligned} \tag{37}$$

The Taylor expansion of the above equations involves expansion of the exponential map. As the explicit expansion up to the first-order term is sufficient for our purposes, let us expand the exponential map as follows:

$$\exp([w + \delta w]) = \exp([w]) + [\mathrm{dexp}_w \delta w] \exp([w]) + \cdots \tag{38}$$

where the differential of the exponential map $\mathrm{dexp}_w \in \mathbb{R}^{3\times 3}$ is presented in [30]. When $w = 0$, the above equation is reduced

to the Taylor expansion of the matrix exponential, that is,

$$\exp([\delta w]) = I + [\delta w] + \cdots . \tag{39}$$

Substituting (38) and (39) into (36) and (37) yields the following:

$$\begin{aligned}
\delta w_{N_i} &= \text{dexp}^{-1}_{w_{N_i}} R_{C_i} (\delta w_{C_i} - \delta w_X) + \cdots \\
\delta p_{N_i} &= [p_{C_i} - p_{N_i}] R_{C_i} (\delta w_{C_i} - \delta w_X) \\
&\quad + R_{C_i} (\delta q_{C_i} - \delta q_X) + \cdots \\
\delta w_{M_i} &= \text{dexp}^{-1}_{w_{M_i}} \left(-\delta w_{C_i} + R^T_{C_i} R_Y \delta w_Y\right) + \cdots \\
\delta p_{M_i} &= -[p_{M_i}] \left(-\delta w_{C_i} + R^T_{C_i} R_Y \delta w_Y\right) + \delta q_{C_i} \\
&\quad + R^T_{C_i} [p_Y - p_{C_i}] R_Y \delta w_Y + R^T_{C_i} R_Y \delta q_Y + \cdots .
\end{aligned} \tag{40}$$

The change in the log-likelihood (31) given these perturbations is calculated as follows:

$$\begin{aligned}
\delta \log L = -\sum_i \Big(& w^T_{N_i} \Sigma^{-1}_{w_{N_i}} \delta w_{N_i} + p^T_{N_i} \Sigma^{-1}_{p_{N_i}} \delta p_{N_i} \\
& + w^T_{M_i} \Sigma^{-1}_{w_{M_i}} \delta w_{M_i} + p^T_{M_i} \Sigma^{-1}_{p_{M_i}} \delta p_{M_i} \Big) + \cdots .
\end{aligned} \tag{41}$$

After substituting (40) into (41), the gradients of $\log L$ with respect to the rotational and translational displacement vectors of X, Y, and C_i can be obtained from the first-order terms as follows:

$$\begin{aligned}
\frac{\partial \log L}{\partial w_X} &= -\sum_i U_i \\
\frac{\partial \log L}{\partial w_Y} &= -\sum_i V_i R^T_{C_i} R_Y \\
\frac{\partial \log L}{\partial w_{C_i}} &= U_i + V_i - p^T_{M_i} \Sigma^{-1}_{p_{M_i}} [R_{C_i} (p_Y - p_{C_i})] \\
\frac{\partial \log L}{\partial q_X} &= \sum_i p_{N_i} \Sigma^{-1}_{p_{N_i}} R_{C_i} \\
\frac{\partial \log L}{\partial q_Y} &= -\sum_i p_{M_i} \Sigma^{-1}_{p_{M_i}} R^T_{C_i} R_Y \\
\frac{\partial \log L}{\partial q_{C_i}} &= -p_{N_i} \Sigma^{-1}_{p_{N_i}} R_{C_i} + p_{M_i} \Sigma^{-1}_{p_{M_i}}
\end{aligned} \tag{42}$$

where $U_i \in \mathbb{R}^{1\times 3}$ and $V_i \in \mathbb{R}^{1\times 3}$ are given by

$$\begin{aligned}
U_i &= -w^T_{N_i} \Sigma^{-1}_{w_{N_i}} \text{dexp}^{-1}_{w_{N_i}} R_{C_i} - p^T_{N_i} \Sigma^{-1}_{p_{N_i}} [p_{C_i} - p_{N_i}] R_{C_i} \\
V_i &= w^T_{M_i} \Sigma^{-1}_{w_{M_i}} \text{dexp}^{-1}_{w_{M_i}} - p^T_{M_i} \Sigma^{-1}_{p_{M_i}} [R^T_{C_i} R_Y p_{B_i}].
\end{aligned} \tag{43}$$

For nonzero $w \in \mathbb{R}^3$, the inverse of the differential of the exponential map is provided in [30] as

$$\text{dexp}^{-1}_w = I - \frac{1}{2}[w] + \frac{1}{\|w\|^2} \left(1 - \frac{\|w\|}{2} \cot \frac{\|w\|}{2}\right) [w]^2. \tag{44}$$

When $w = 0$, it is simply given as $\text{dexp}^{-1}_w = I$.

Note that the gradients associated with X and Y in (42) are summations over the n measurement pairs. To achieve similar convergence rates between (X, Y) and $\{C_i\}$, the gradient ascent in (34) is performed with normalized scales of gradients, which are expressed by $\Delta w_{(\cdot)} = \frac{1}{n} \epsilon_w \frac{\partial \log L}{\partial w_{(\cdot)}}$ and $\Delta q_{(\cdot)} = \frac{1}{n} \epsilon_q \frac{\partial \log L}{\partial q_{(\cdot)}}$ for X and Y as well as $\Delta w_{C_i} = \epsilon_w \frac{\partial \log L}{\partial w_{C_i}}$ and $\Delta q_{C_i} = \epsilon_q \frac{\partial \log L}{\partial q_{C_i}}$. Here, $\epsilon_w \in \mathbb{R}$ and $\epsilon_q \in \mathbb{R}$ are the step sizes for rotation and translation, respectively.

B. Noise Configuration 2: $A_i N_i^{-1} X = Y B_i M_i^{-1} = C_i$

For noise configuration 2, the noise transformations N_i and M_i are expressed in terms of X, Y, and C_i by (8) and (12), respectively. As in the first noise configuration, considering perturbations of X, Y, and C_i in (8) and (12) and substituting these equations into (31) yields the following gradients:

$$\begin{aligned}
\frac{\partial \log L}{\partial w_X} &= -\sum_i S_i \\
\frac{\partial \log L}{\partial w_{C_i}} &= S_i + V_i - p^T_{M_i} \Sigma^{-1}_{p_{M_i}} [R_{C_i} (p_Y - p_{C_i})] \\
\frac{\partial \log L}{\partial q_X} &= -\sum_i p_{N_i} \Sigma^{-1}_{p_{N_i}} R_X \\
\frac{\partial \log L}{\partial q_{C_i}} &= p_{N_i} \Sigma^{-1}_{p_{N_i}} R_X + p_{M_i} \Sigma^{-1}_{p_{M_i}}
\end{aligned} \tag{45}$$

where V_i is the same as in (43), and $S_i \in \mathbb{R}^{1\times 3}$ is defined as

$$S_i = w^T_{N_i} \Sigma^{-1}_{w_{N_i}} \text{dexp}^{-1}_{w_{N_i}} R_X - p^T_{N_i} \Sigma_{p_{N_i}} [R_X R^T_{C_i} (p_{A_i} - p_{C_i})]. \tag{46}$$

The gradients associated with Y remain the same as those described in the previous subsection. A detailed derivation is omitted because it is almost identical to the previous case.

C. Noise Configuration 3: $A_i X = Y B_i M_i^{-1}$

The final noise configuration is where A_i is noiseless and only B_i is noisy. The log-likelihood in this case is given by taking log in (14), which yields

$$\log L(X, Y) = -\frac{1}{2} \sum_i \left(w^T_{M_i} \Sigma^{-1}_{w_{M_i}} w_{M_i} + p^T_{M_i} \Sigma^{-1}_{p_{M_i}} p_{M_i} \right). \tag{47}$$

Considering perturbations in X and Y in the noise transformation M_i in (15), the gradients are again derived from first-order terms of the Taylor expansion of (47), as follows:

$$\begin{aligned}
\frac{\partial \log L}{\partial w_X} &= \sum_i \left(D_i - p^T_{M_i} \Sigma^{-1}_{p_{M_i}} [p_{M_i}] \right) \\
\frac{\partial \log L}{\partial w_Y} &= \sum_i \left(-D_i R_{M_i} R^T_{B_i} + p^T_{M_i} \Sigma^{-1}_{p_{M_i}} R_{M_i} R^T_{B_i} [p_{B_i}] \right) \\
\frac{\partial \log L}{\partial q_X} &= \sum_i p^T_{M_i} \Sigma^{-1}_{p_{M_i}} \\
\frac{\partial \log L}{\partial q_Y} &= -\sum_i p^T_{M_i} \Sigma^{-1}_{p_{M_i}} R_{M_i} R^T_{B_i}
\end{aligned} \tag{48}$$

where $D_i \in \mathbb{R}^{1\times 3}$ is defined as

$$D_i = w^T_{M_i} \Sigma^{-1}_{w_{M_i}} \text{dexp}^{-1}_{w_{N_i}}. \tag{49}$$

VII. Estimation Uncertainty

The transformations X and Y acquired through the maximization outlined in the previous section are estimations of the true transformations, of which the estimation uncertainty depends on the noise levels of A_i and B_i. The confidence of the estimations can be quantified by identifying this uncertainty. To do so, let X^*, Y^*, and C_i^* denote the maximum likelihood estimations of X, Y, and C_i; that is,

$$(X^*, Y^*, \{C_i^*\}) = \arg\min_{X,Y,C_i} L. \tag{50}$$

When no measurement noise exists, the estimations are identical to the true transformations. If measurement noise exists, the estimations deviate from the true transformations as follows:

$$\begin{aligned} X^* &= \tilde{X}T(w_X, q_X) \\ Y^* &= \tilde{Y}T(w_Y, q_Y) \\ C_i^* &= \tilde{C}_i T(w_{C_i}, q_{C_i}) \end{aligned} \tag{51}$$

where $\tilde{(\cdot)}$ represents the true transformations, and $w_{(\cdot)} \in \mathbb{R}^3$ and $q_{(\cdot)} \in \mathbb{R}^3$ denote the rotational and positional error vectors of the subscripted transformation, respectively. Let us assume that noises of A_i and B_i are small; thus, the resulting errors are also small. In this case, the error vectors $(w_{(\cdot)}, q_{(\cdot)})$ can be approximated in the tangent space of $SE(3)$, which is simply a six-dimensional vector space. In the following subsections, we derive the error covariance matrix as a function of the noise covariances for each noise configuration.

A. Noise Configuration 1: $N_i A_i X = Y B_i M_i^{-1} = C_i$

In noise configuration 1, the measurements (A_i, B_i) are expressed in terms of the true transformations $(\tilde{A}_i, \tilde{B}_i)$ and true noises $(\tilde{N}_i, \tilde{M}_i)$, as follows:

$$A_i = \tilde{N}_i^{-1}\tilde{A}_i,\ B_i = \tilde{B}_i\tilde{M}_i. \tag{52}$$

By substituting (51) and (52) into (7) and (8), we obtain

$$\begin{aligned} N_i &= \tilde{C}_i T(w_{C_i}, q_{C_i}) T(w_X, q_X)^{-1} \tilde{X}^{-1} \tilde{A}_i{}^{-1} \tilde{N}_i \\ M_i &= T(w_{C_i}, q_{C_i})^{-1} \tilde{C}_i{}^{-1} \tilde{Y} T(w_Y, q_Y) \tilde{B}_i \tilde{M}_i \end{aligned} \tag{53}$$

where $\tilde{C}_i = \tilde{A}_i\tilde{X} = \tilde{Y}\tilde{B}_i$. Given the small error vectors, the first-order approximations of the above equations are determined as follows:

$$\begin{aligned} w_{N_i} &= R_{\tilde{C}_i}(w_{C_i} - w_X) + w_{\tilde{N}_i} \\ p_{N_i} &= [p_{\tilde{C}_i}]R_{\tilde{C}_i}(w_{C_i} - w_X) + R_{\tilde{C}_i}(q_{C_i} - q_X) + p_{\tilde{N}_i} \\ w_{M_i} &= w_{C_i} + R_{\tilde{C}_i}^T R_{\tilde{Y}} w_Y + w_{\tilde{M}_i} \\ p_{M_i} &= q_{C_i} + [R_{\tilde{C}_i}^T (p_{\tilde{Y}} - p_{\tilde{C}_i})] R_{\tilde{C}_i}^T R_{\tilde{Y}} w_Y + p_{\tilde{M}_i}. \end{aligned} \tag{54}$$

where $w_{\tilde{N}_i}, p_{\tilde{N}_i}, w_{\tilde{M}_i} \in \mathbb{R}^3$, and $p_{\tilde{M}_i} \in \mathbb{R}^3$ denote the true rotational and positional noise vectors in $\tilde{N}_i$ and $\tilde{M}_i$, respectively; that is, $\tilde{N}_i = T(w_{\tilde{N}_i}, q_{\tilde{N}_i})$ and $\tilde{M}_i = T(w_{\tilde{M}_i}, q_{\tilde{M}_i})$. The above equations can be expressed using a single linear equation

$$\gamma = Q\alpha + \beta \tag{55}$$

where $\gamma = \left[s_{N_1}^T \dots s_{N_n}^T s_{M_1}^T \dots s_{M_n}^T\right]^T$, $\alpha = \left[s_X^T\ s_Y^T\ s_{C_1}^T \dots s_{C_n}^T\right]^T$, and $\beta = \left[s_{\tilde{N}_1}^T \dots s_{\tilde{N}_m}^T\ s_{\tilde{M}_1}^T \dots s_{\tilde{M}_m}^T\right]$ with $s_{(\cdot)} \in \mathbb{R}^6$ defined as $s_{(\cdot)} = [w_{(\cdot)}^T\ q_{(\cdot)}^T]^T$ for the subscripted variable. Matrix Q is a $12n \times (6n + 12)$ matrix, the elements of which are filled with the coefficients of the deviation vectors in (54). Then, the log-likelihood (31) is expressed using γ as $\log L = -\frac{1}{2}\gamma^T W^{-1}\gamma$, where W is a $12n \times 12n$ matrix in which the 3×3 diagonal blocks are $\Sigma_{w_{N_i}}, \Sigma_{p_{N_i}}, \Sigma_{w_{M_i}}$, and $\Sigma_{p_{M_i}}$ and the off-diagonal blocks are zero. Combined with (55), the log-likelihood becomes

$$\log L = -\frac{1}{2}\alpha^T Q^T W^{-1} Q\alpha - \beta^T W^{-1} Q\alpha - \frac{1}{2}\beta^T W^{-1}\beta. \tag{56}$$

Given the collection of true noises as β, the concatenated error vector α is the maximizer of the above equation of the form

$$\alpha = -(Q^T W^{-1} Q)^{-1} Q^T W^{-1}\beta. \tag{57}$$

As $\beta \sim \mathcal{N}(0, W)$, the PDF of α is expressed as $\alpha \sim \mathcal{N}(0, Z)$, where $Z = (Q^T W^{-1} Q)^{-1}$. Note that Z is the $(12 + 6n) \times (12 + 6n)$ covariance matrix of the error vectors of X^*, Y^*, and C_i^*. We are most interested in the first 12×12 block of Z, which represents the uncertainty covariance of X^* and Y^*.

B. Noise Configuration 2: $A_i N_i^{-1} X = Y B_i M_i^{-1} = C_i$

For noise configuration 2, only matrix Q is derived differently from the previous case. The elements of matrix Q are filled with coefficients of the deviation vectors in the following linear equations:

$$\begin{aligned} w_{N_i} &= R_{\tilde{X}}(w_X - w_{C_i}) + w_{\tilde{N}_i} \\ p_{N_i} &= [p_X]R_{\tilde{X}}(w_X - w_{C_i}) + R_{\tilde{X}}(q_X - q_{C_i}) + p_{\tilde{M}_i}. \end{aligned} \tag{58}$$

The equations associated with w_{M_i} and p_{M_i} remain the same as those in (54).

C. Noise Configuration 3: $A_i X = Y B_i M_i^{-1}$

For noise configuration 3, where A_i is noiseless, the only noise transformation M_i in (15) is expressed in terms of X^* and Y^* as follows:

$$M_i = X^{*-1} A_i^{-1} Y^* B_i. \tag{59}$$

By substituting (51), B_i of (52), and $A_i = \tilde{A}_i$ into the above equation, the first-order approximations of w_{M_i} and p_{M_i} are derived as follows:

$$\begin{aligned} w_{M_i} &= -w_X + R_{\tilde{B}_i}^T w_Y + w_{\tilde{M}_i} \\ p_{M_i} &= -q_X - R_{\tilde{B}_i}^T [p_{\tilde{B}_i}] w_Y + R_{\tilde{B}_i}^T q_Y + p_{\tilde{M}_i}. \end{aligned} \tag{60}$$

Because N_i and C_i^* are not present in this case, the vectors α, β, and γ are defined in smaller sizes: $\alpha = [s_X^T\ s_Y^T]^T \in \mathbb{R}^{12}$, $\beta = [\, s_{\tilde{M}_1}^T \ \dots\ s_{\tilde{M}_m}^T\,] \in \mathbb{R}^{6n}$, and $\gamma = [s_{M_1}^T \ \dots\ s_{M_n}^T]^T \in \mathbb{R}^{6n}$. Matrix Q is now a $6n \times 12$ matrix, the elements of which are filled with the coefficients of deviation vectors in (60). The covariance of α is a 12×12 matrix as $Z = (Q^T W^{-1} Q)^{-1}$, whereas matrix W is now a $6n \times 6n$ matrix in which the 3×3 diagonal blocks are $\Sigma_{w_{M_i}}$ and $\Sigma_{p_{M_i}}$.

VIII. Numerical Experiments

Four numerical experiments are presented in this section. The first experiment demonstrates the varying accuracy of an existing distance minimization approach with respect to differently coordinated A, B, X, and Y, where it is revealed that the coordination suggested by the proposed probabilistic framework yields the best calibration result. In the second experiment, the analysis of the estimation uncertainty presented in Section VII is validated by running the calibration numerous times and comparing the error vector distribution with analytically derived covariance matrices. In the third experiment, a performance comparison between the proposed method and the distance minimization approach is presented. The final experiment demonstrates the effect of incorrect selection of the noise configuration on the estimation accuracy.

The first subsection describes the generation of noiseless datasets, which were used in the experiments presented in the following subsections. The distance minimization, most frequently used for comparison, is that presented in [4], which was reported in [31] and [32] to be the most reliable for a small number of measurements and the most accurate compared to the methods in [5] and [11]. In addition, the methods in [7] and [8] were compared with the proposed method for their remarkable performance reported in [14]. The method in [27] was also chosen for comparison as a representative method based on the $\mathcal{A}X = X\mathcal{B}$ formulation.

A. Noiseless Data Generation

For the numerical experiments, multiple sets of noisy measurements (A_i, B_i) were synthetically generated together with $\tilde{X}$ and $\tilde{Y}$, that is, the true transformations of X and Y.

1) Generation of $\tilde{X}$ and $\tilde{Y}$: The transformations $\tilde{X}$ and $\tilde{Y}$ were first randomly generated by

$$\begin{aligned} R_{\tilde{X}}, R_{\tilde{Y}} &\sim U(R) \\ p_{\tilde{X}}, p_{\tilde{Y}} &\sim \mathcal{N}(p_{\tilde{A}_0}, I_{3\times 3}) \end{aligned} \tag{61}$$

where $R_{(\cdot)}$ and $p_{(\cdot)}$ are the rotation and translation of the corresponding transformation, respectively. In this case, $U(R)$ denotes a uniform distribution on $SO(3)$, a sampling method of which is described in Appendix B.

2) Generation of $\tilde{A}_i$ and $\tilde{B}_i$: A random transformation $\tilde{A}_0 \in SE(3)$ was sampled using $\tilde{X}$ and $\tilde{Y}$ in (61). Subsequently, the noiseless transformations $\tilde{A}_i$ were generated around $\tilde{A}_0$ as follows:

$$R_{\tilde{A}_i} = R_{\tilde{A}_0} R_i \tag{62}$$

$$p_{\tilde{A}_i} \sim \mathcal{N}\left(0_{1\times 3}, \frac{1}{2} I_{3\times 3}\right) \tag{63}$$

where $R_i \in SO(3)$ was sampled according to the distribution (28) with $\Sigma_w = I_{3\times 3}$. Appendix B presents a detailed sampling method for (28). Once $\{\tilde{A}_i\}$ were generated, the transformations $\{\tilde{B}_i\}$ were generated by

$$\tilde{B}_i = \tilde{Y}^{-1} \tilde{A}_i \tilde{X}. \tag{64}$$

B. Accuracy of Existing Method Regarding Coordination of A, B, X, and Y

The distance minimization approaches are equivalent to the maximum likelihood estimation if A is assumed to be noiseless and B is assumed to be noisy with right-translated isotropic noise. When a distance minimization approach is applied, the optimal calibration result will be obtained from the transformations A, B, X, and Y that are coordinated to respect these assumptions. To verify this, distance minimization was performed for the four possible coordinations of the transformations, between which the estimation accuracy was compared.

Synthetic noisy datasets were generated with noiseless A and noisy B. When generating each noisy dataset, a noiseless dataset is generated as explained in Section VIII-A, following which $\tilde{B}_i$ is corrupted with the measurement noise M_i as $B_i = \tilde{B}_i M_i$, where M_i is generated as follows:

$$p_{M_i} \sim \mathcal{N}\left(0_{3\times 1}, \Sigma_{p_{M_i}}\right) \tag{65}$$

$$R_{M_i} \sim k \exp\left(-\frac{1}{2} w_{M_i}^T \Sigma_{w_{M_i}}^{-1} w_{M_i}\right). \tag{66}$$

In the above, R_{M_i} and p_{M_i} are the rotation and translation of N_i, respectively, and w_{M_i} is given by $[w_{M_i}] = \log R_{M_i}$. The matrices $\Sigma_{p_{M_i}}$ and $\Sigma_{w_{M_i}}$ were set to $\Sigma_{p_{M_i}} = \Sigma_{w_{M_i}} = \epsilon I_{3\times 3}$, where $\epsilon = 0.05$. This value of ϵ corresponds to the STDs of 2.86° for the rotational noise and 0.05 for the translational noise. The transformation A_i remains noiseless as $A_i = \tilde{A}_i$.

Once a dataset $\{(A_i, B_i)\}$ is prepared, distance minimization [4] is applied by the following four minimizations:

$$\begin{aligned} &\min_{X,Y} \sum_i d\left(A_i X, Y B_i\right)^2 \\ &\min_{X,Y} \sum_i d\left(B_i^{-1} Y^{-1}, X^{-1} A_i^{-1}\right)^2 \\ &\min_{X,Y} \sum_i d\left(B_i X^{-1}, Y^{-1} A_i\right)^2 \\ &\min_{X,Y} \sum_i d\left(A_i^{-1} Y, X B_i^{-1}\right)^2. \end{aligned} \tag{67}$$

Note that only the first minimization corresponds to the correct coordination. Once the minimizations are solved, the accuracy of the estimations can be evaluated using the following error terms in rotation and translation:

$$\begin{aligned} e_{R_X} &= \| \log\left(R_X^T R_{\tilde{X}}\right)^\vee \|, \; e_{R_Y} = \| \log\left(R_Y^T R_{\tilde{Y}}\right)^\vee \| \\ e_{p_X} &= \|p_X - p_{\tilde{X}}\|, \; e_{p_Y} = \|p_Y - p_{\tilde{Y}}\| \end{aligned} \tag{68}$$

where the notation $^\vee$ denotes the vectorization of the skew-symmetric matrix, that is, $[w]^\vee = w$.

In this experiment, 100 calibration datasets were generated with a set size of $n = 20$. The means and STDs of errors for 100 calibrations are presented in Fig. 5, where the bars and ranges represent the mean and standard deviation, respectively. The coordination numbers 1, 2, 3, and 4 in the figure correspond to the first, second, third, and fourth minimizations in (67), respectively. The first coordination resulted in the best accuracy in both X and Y, thereby confirming the hidden assumptions in the

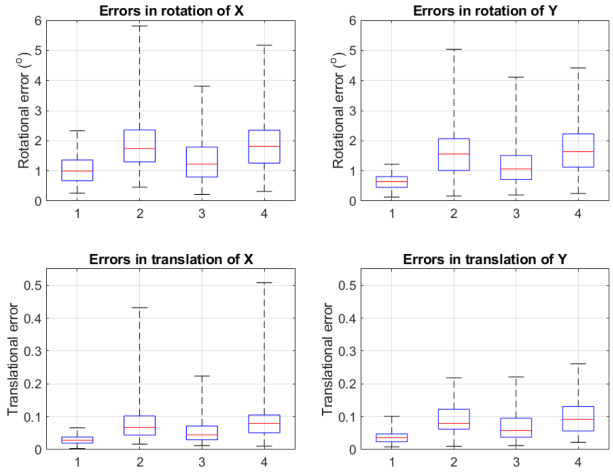


Fig. 5. Box plots of rotational and translational errors of X and Y of distance minimization for 100 numerical experiments. The red lines are the medians, edges of the boxes represent the 25th and 75th percentiles, and boundaries of the whiskers represent the minimum and maximum errors.

existing distance minimization approaches regarding the measurement noise; A is noiseless, and B contains right-translated isotropic noise.

C. Validation of Estimation Uncertainty

We compared the covariance matrix that was theoretically derived in Section VII with numerous experimental results and verified that the covariance was correctly derived. For this purpose, a single noiseless dataset was generated, and 3000 noisy datasets were generated by corrupting the same noiseless dataset for noise configuration 1 with $\Sigma_{w_{N_i}} = \Sigma_{p_{N_i}} = \Sigma_{w_{M_i}} = \Sigma_{p_{M_i}} = 0.05$. Subsequently, the proposed method was applied to each noisy dataset. The estimation errors were represented with two 6-dimensional vectors (w_X, q_X) and (w_Y, q_Y), which were acquired from

$$\tilde{X}^{-1}X^* = T(w_X, q_X)$$
$$\tilde{Y}^{-1}Y^* = T(w_Y, q_Y). \tag{69}$$

Again, X^* and Y^* are the estimations, and $T(\cdot,\cdot)$ is defined in (33). We selected four pairs of two dimensions among the six dimensions to visualize the error vectors of X on a 2-D plane. The 2-D error vectors are plotted in Fig. 6, together with the ellipsoids of the numerically computed and analytically derived covariance matrices. In practice, the true transformations $\tilde{X}, \tilde{Y}$, and $\tilde{C}_i$ are not available. The ellipsoids of the analytic covariance when the estimations X^*, Y^*, and C_i^* were used instead are also plotted in the figure. The three ellipsoids matched one another very well, which confirmed that the analytic covariance was derived correctly, and the use of the estimations to compute the analytic covariance was acceptable.

D. Comparison of Log-Likelihood Maximization and Existing Methods

The proposed method was compared with various approaches [4], [7], [8], [10], [27] for noise configurations 1 and 2,

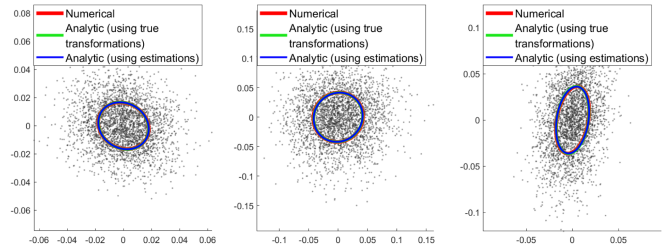


Fig. 6. Numerically computed error vectors (w_X, q_X) compared to analytically derived covariance ellipsoids on planes of two selected dimensions. (a) $w_{X,y}$ vs. $w_{X,z}$. (b) $q_{X,x}$ vs. $q_{X,y}$. (c) $w_{X,y}$ vs. $q_{X,z}$. The gray dots are the error vectors, the red ellipsoids represent the numerical covariances that were statistically computed using the error vectors, the green ellipsoids are the analytical covariances that were computed using the true transformations, and the blue ellipsoids are the analytic covariances that were determined using the estimated transformations.

where both A and B contain measurement noises. These methods are referred to as "Dist. Min.," "Li," "Shah," "Tabb," and "AX=XB," respectively, in the result figures. Note that "Dist. Min.," "Li," "Shah," and "Tabb" are based on the $AX = YB$ formulation, where "Li" and "Shah" are linear approaches, whereas "Dist. Min." and "Tabb" are nonlinear approaches. In addition to this experiment, the equivalence between the proposed method and distance minimization [4] for noise configuration 3 with isotropic noise was demonstrated.

To generate each noisy dataset for noise configuration 1, a noiseless dataset $\{(\tilde{A}_i, \tilde{B}_i)\}$ was first generated and was corrupted by $A_i = N_i^{-1}\tilde{A}_i$ and $B_i = \tilde{B}_i M_i$, where the noise transformations N_i and M_i were randomly sampled by

$$p_{N_i} \sim \mathcal{N}\left(0_{3\times1}, \Sigma_{p_{N_i}}\right)$$
$$R_{N_i} \sim k\exp\left(-\frac{1}{2}w_{N_i}^T \Sigma_{w_{N_i}}^{-1} w_{N_i}\right)$$
$$p_{M_i} \sim \mathcal{N}\left(0_{3\times1}, \Sigma_{p_{M_i}}\right)$$
$$R_{M_i} \sim k\exp\left(-\frac{1}{2}w_{M_i}^T \Sigma_{w_{M_i}}^{-1} w_{M_i}\right). \tag{70}$$

Here, R_{N_i} and R_{M_i} are the rotations, while p_{N_i} and p_{M_i} are the translations of N_i and M_i, respectively; the noise covariances were selected as $\Sigma_{w_{N_i}} = \Sigma_{p_{N_i}} = \Sigma_{w_{M_i}} = \Sigma_{p_{M_i}} = \epsilon I_{3\times3}$ with $\epsilon = 0.05$. Noisy datasets for noise configuration 2 were also generated by $A_i = \tilde{A}_i N_i$ and $B_i = \tilde{B}_i M_i$, where N_i and M_i were randomly sampled in the same manner as described above. For noise configuration 3, A_i and B_i were generated by $A_i = \tilde{A}_i$ and $B_i = \tilde{B}_i M_i$, respectively.

In the experiment, 100 noisy datasets were generated with a set size of $n = 20$ for each noise configuration. The means and STDs of the errors in X and Y for noise configuration 1 are shown in Fig. 7 and those for noise configuration 2 are presented in Fig. 8. In the figures, the bars and ranges represent the errors and standard deviations, respectively. The errors of the proposed algorithm are smaller than those of the compared approaches for both cases.

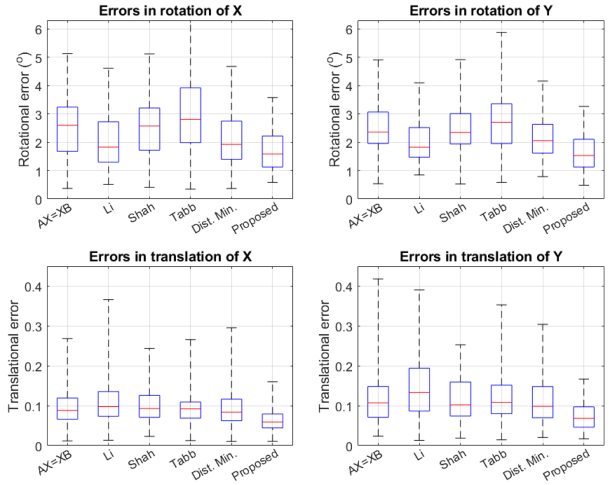


Fig. 7. Box plots of rotational and translational errors of X of the compared methods and the proposed algorithm for 100 numerical experiments on noise configuration 1. The red lines are the medians, edges of the boxes represent the 25th and 75th percentiles, and boundaries of the whiskers represent the minimum and maximum errors. The maximum error of Tabb's method in rotation of X is $9.78°$.

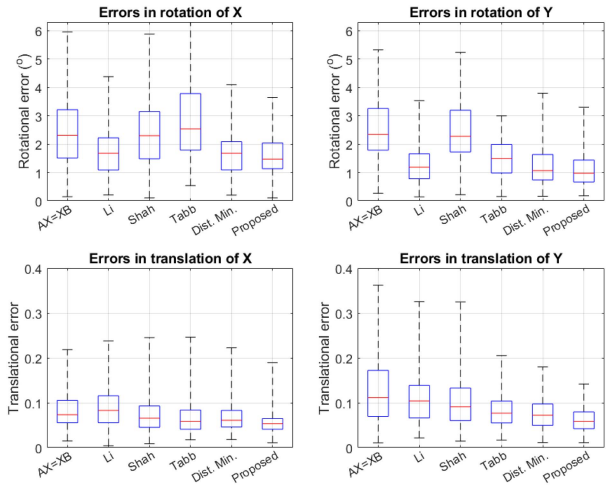


Fig. 8. Box plots of rotational and translational errors of X of the compared methods and the proposed algorithm for 100 numerical experiments on noise configuration 2. The red lines are the medians, the edges of the boxes represent the 25th and 75th percentiles, and the boundaries of the whiskers represent the minimum and maximum errors. The maximum error of Tabb's method in rotation of X is $7.19°$.

Given the isotropic noise in noise configuration 3, the proposed method is reduced to a geodesic distance minimization. The comparison in this case is presented in Fig. 9, which indicates that the two methods have almost identical errors. This result was expected based on the equivalence discussed in Section IV. In terms of the individual solutions, the rotations of X and Y differed by less than 0.01% of the mean rotational errors between the two methods, whereas the translations differed by only 0.0033% to 0.21% of the mean translational errors. This slight difference was due to the varying numerical criteria between the two methods and/or the bounded equivalence between the distance functions, which could be violated for large $d(A_iX, YB_i)$.

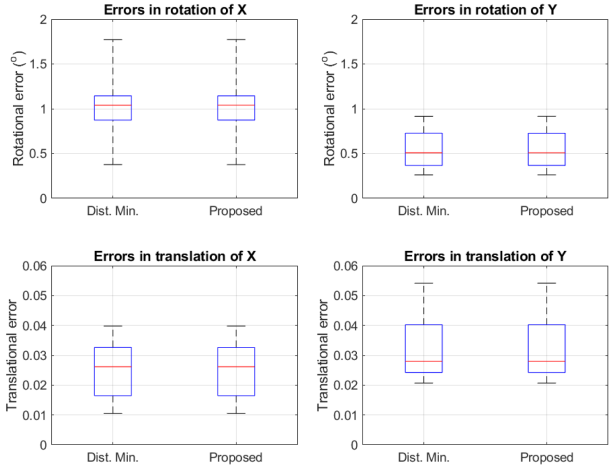


Fig. 9. Box plots of the rotational and translational errors of X of distance minimization and the proposed algorithm for 100 numerical experiments on noise configuration 3. The red lines are the medians, edges of the boxes represent the 25th and 75th percentiles, and boundaries of the whiskers represent the minimum and maximum errors.

In our MATLAB implementation, the average computational time of the proposed method for $n = 20$ was approximately 1.8 s, which is reasonable for the purpose of offline calibration. The computational time was observed to be almost linearly related to n. For example, when $n = 100$, the computational time was 8.3 s on average.

E. Estimation Bias Owing to Inaccurate Noise Configuration

Recalling that distance minimization is identical to the proposed method for noise configuration 3 given isotropic noise in B, the results of the previous experiment can also be interpreted as an accuracy comparison between the correct and inaccurate noise configurations. It may be expected that, even with inaccurate noise configuration, the estimations of X and Y would be sufficiently accurate, given a large number of measurement pairs. However, it can also be hypothesized that inaccurate noise configurations may result in biased estimations, even for a large number of measurement pairs. To determine which argument is correct, the calibration accuracy was evaluated for correct and inaccurate noise configurations with an increasing number of measurement pairs.

In this experiment, a noiseless dataset was initially generated with a set size of $n = 1000$ and was contaminated with noise as noise configuration 1, where $\Sigma_{w_{N_i}} = \Sigma_{p_{N_i}} = \Sigma_{w_{M_i}} = \Sigma_{p_{M_i}} = \epsilon I$ with $\epsilon = 0.05$. Thereafter, the distance minimization [4] and the proposed method with the correct noise configuration were applied 100 times for m randomly selected measurement pairs for each $m \in \{10, 20, \ldots, 90, 100, 200, \ldots, 1000\}$. The averaged calibration errors of X and Y for the 100 calibrations for each m are plotted in Figs. 10 and 11, respectively. Interestingly, the calibration errors decreased as m increased initially; however, the errors then became almost saturated as m increased further. The proposed method attained smaller saturated errors, particularly in the translations. This experiment can be considered a comparison of the calibration accuracy between the

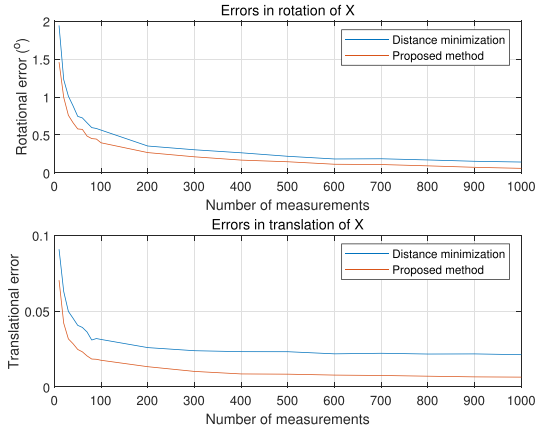


Fig. 10. Average calibration errors for X of distance minimization and proposed method in terms of the number of measurement pairs.

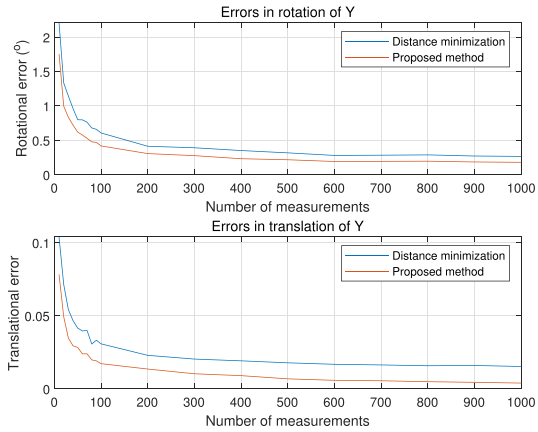


Fig. 11. Average calibration errors for Y of distance minimization and proposed method in terms of number of measurement pairs.

correct noise configuration (noise configuration 1) and incorrect noise configuration (noise configuration 3). The result suggests that an inaccurate noise configuration can produce a significant estimation bias, particularly in translations, even with numerous measurement pairs.

IX. Hardware Experiments

Two hardware experiments are presented in this article. In the first experiment, a camera attached to a robot hand captured a checkerboard. The second experiment involved two rigidly combined cameras that captured two different checkerboards. Both were formulated as $AX = YB$, whereas the noise configurations differed.

A. Camera on Robot Hand

In this experiment, a webcam (Logitech C920) was attached to the end effector of a UR3e manipulator [33]. A 7×6 checkerboard with a block size of 12.46 mm was captured by the webcam under 88 different manipulator configurations. To include various robot and camera poses, the measurements were collected by manually back-driving the robot. The intrinsic parameters

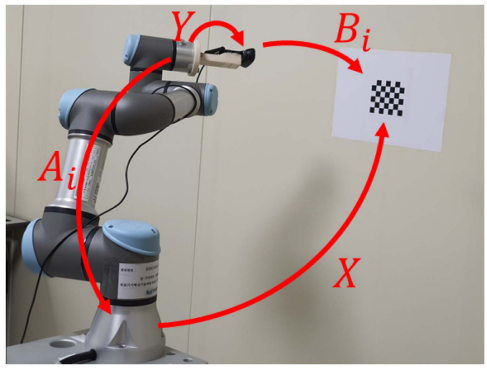


Fig. 12. Experimental setup for hand–eye calibration. A webcam attached to a UR3e manipulator captured a checkerboard.

of the camera were calibrated using the method of Zhang [34], and the checkerboard transformations were computed by solving the perspective-n-point problem using the OpenCV library. The experimental setup is shown in Fig. 12.

For the correct coordination of transformations, we refer to Section IV-C. Considering the small manipulation error of UR3e (i.e., ± 0.03 mm of repeatability error [33]), the correct coordination of A, B, X, and Y is shown in Fig. 12. With the correct coordination, distance minimization is identical to the proposed probabilistic method with isotropic covariance matrices. As this identity has already been validated in a numerical experiment, a comparison between the proposed method and distance minimization is not presented here. Instead, we investigate the difference in the calibration results between the four cases in which A, B, X, and Y are differently coordinated:

1) Case 1: The transformations were coordinated as in Fig. 12.
2) Case 2: A and B were swapped and inverted from case 1; thus, X and Y were also swapped and inverted from case 1.
3) Case 3: A and B were swapped from case 1; thus, each of X and Y in case 1 was inverted.
4) Case 4: Each of A and B was inverted from case 1; thus, X and Y in case 1 were swapped.

The distance minimization approach [4] was applied to the four cases, where the four cases corresponded to the four minimizations in (67). The positional weight factor in the distance function was set to $\frac{3\pi}{180}$, which weighed a rotational error of 1° as equivalent to a positional error of 3 mm.

When dealing with real data acquired in a hardware experiment, the ground truth transformations X and Y are generally unknown. Thus, although case 1 was expected to provide the best calibration accuracy, a direct comparison of the calibration accuracy between the cases was not possible. Instead, we estimated X and Y using all of the measurement pairs for each case and considered these to be the best estimations of that case. Subsequently, the coordination of A, B, X, and Y that yielded closer estimations of X and Y to the best estimations for smaller numbers of measurements could be considered superior.

The best estimations of X and Y for each case were acquired using all 88 measurement pairs. Thereafter, the estimations using subsets of the 88 measurement pairs were compared with the

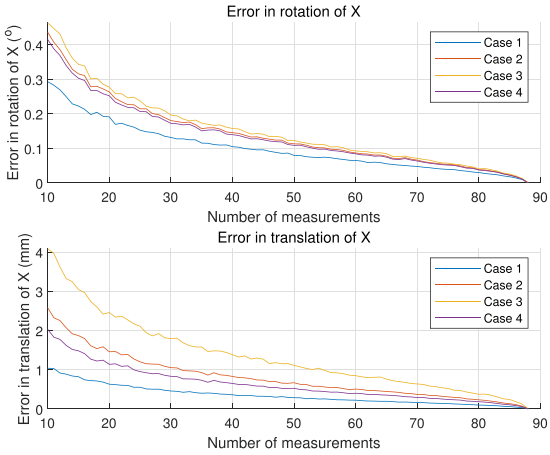


Fig. 13. Average difference from best estimations of X in cases 1, 2, 3, and 4 for different numbers of measurements of the robot hand and webcam.

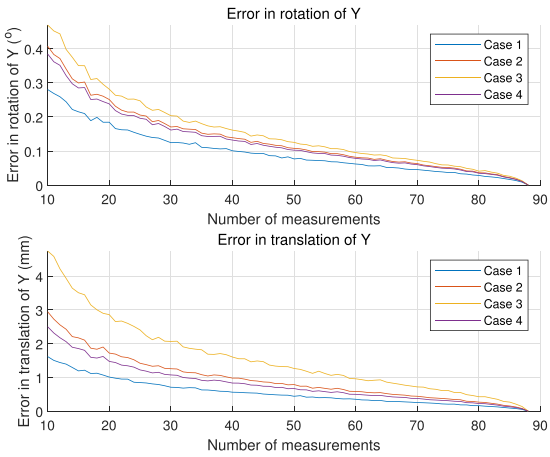


Fig. 14. Average difference from best estimations of Y in cases 1, 2, 3, and 4 for different numbers of measurements of the robot hand and webcam.

best estimations. The subsets were sampled for various set sizes between 10 and 88, with 500 random subsets generated for each subset size. Figs. 13 and 14 depict the results, where the differences between the best estimations and the estimations using the subsets were averaged over the 500 random subsets for each subset size. The figures show that case 1 yielded the closest estimations of X and Y to the best estimations, which implies that the correct coordination truly improves the calibration result.

B. Two Rigidly Combined Cameras

In this experiment, two webcams (Logitech C920) were rigidly combined, and each camera captured each of the two checkerboards, as illustrated in Fig. 15. As shown in the figure, the fields of view of the two cameras did not overlap each other. The calibration problem in this situation was formulated as $AX = YB$. The transformation from camera 1 to checkerboard 1 was measured as A_i, and the transformation from camera 2 to checkerboard 2 was measured as B_i, which is also depicted in the figure. Calibration of the camera and computation

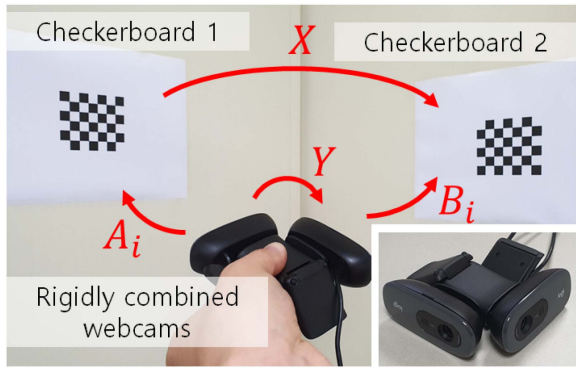


Fig. 15. Experimental setup with two rigidly combined webcams and two checkerboards. Each checkerboard was captured by each camera. The lower right image depicts combined webcams.

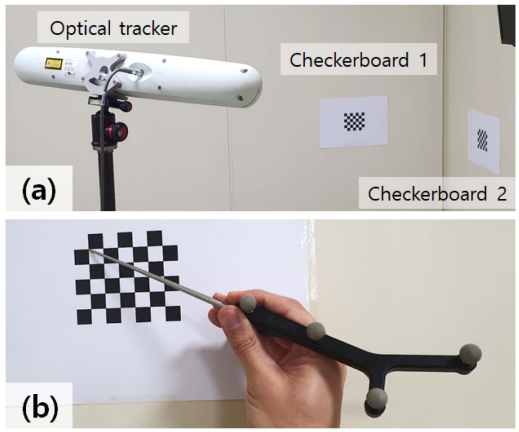


Fig. 16. Measurement of ground truth distances between two checkerboards. (a) Optical tracker placed facing checkerboards. (b) Probe with passive markers targeting corner points of checkerboards.

of the checkerboard transformation were performed using the OpenCV library. In total, 183 measurement pairs were collected manually by moving the combined cameras to capture the checkerboards from diverse angles. The transformations X and Y were calibrated using the proposed method and the distance minimization [4] using these measurement pairs. As discussed in Section II, the noise transformations were assumed to be present at the target frames, that is, the checkerboard frames. Accordingly, this case was handled as noise configuration 2 in the proposed method, where noise covariances of the camera measurements were set to $(1^\circ)^2 I$ for the rotation and $(3\text{ mm})^2$ for the position. To ensure consistency, the positional weight factor in the distance minimization approach was set to $\frac{3\pi}{180}$.

Although the ground truth for Y (the relative camera frame) was unavailable, the transformation X could be verified by measuring the distances between the corner points of the two checkerboards. An optical tracking system (NDI Polaris Spectra) and a probe equipped with four passive markers were used to collect accurate 3-D corner positions, as shown in Fig. 16. The tracking system that was used is known to have a tracking accuracy of 0.64 mm [35]. The tracking error was expected to be approximately 1 mm, including the manual probing error, which is a fraction of a millimeter.

TABLE I
ERRORS IN DISTANCES BETWEEN THE CORNER POINTS OF THE TWO CHECKERBOARDS

	Distance minimization	Proposed method
Mean error (mm)	4.83	0.50
Max. error (mm)	5.33	0.94
Min. error (mm)	4.57	0.19

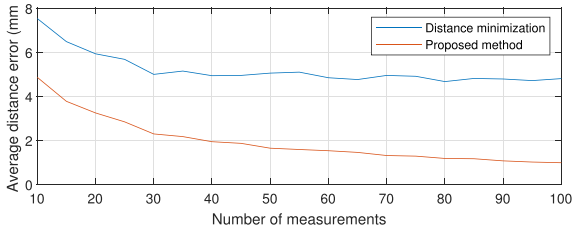


Fig. 17. Average error in the distances between corner points of the two checkerboards in terms of number of measurements used in calibration. The average errors were evaluated for $\{5, 10, 15, \ldots, 100\}$ measurements.

All four bounding corner points of each checkerboard were collected by averaging 100 consecutive tracking measurements, and the distances were computed for $16 (= 4 \times 4)$ pairs of corner points by pairing a point of checkerboard 1 with a point of checkerboard 2. The distances varied from 512.08 to 605.25 mm, which were compared with the distances computed using X estimated by the distance minimization and the proposed method. Table I presents a comparison in which the average error of the 16 distances was 0.50 mm for the proposed method and 4.83 mm for the distance minimization. The accuracy of the proposed method was superior to that of the distance minimization method by more than 4 mm on average, which is significant considering the expected optical tracking error (≈ 1 mm).

To demonstrate the performance for a smaller number of measurements, the two methods were executed multiple times using smaller datasets. Fig. 17 presents the average errors in the 16 distances of the proposed method and the distance minimization when using subsets of the 183 measurement pairs. The subsets were sampled with set sizes varying from 10 to 100. As a large number of possible subsets existed, the subsets were randomly sampled 500 times for each set size, and the errors were averaged. The proposed method yielded a smaller calibration error compared to the distance minimization method for any number of measurement pairs.

X. CONCLUSION

In this study, a probabilistic framework was proposed for hand–eye and robot–world calibration or, more generally, for two-frame sensor calibration. This suggests that the existing distance minimization approaches are maximum likelihood methods under the assumption that A is noiseless and B contains isotropic noise. This enables efficient coordination of A, B, X, and Y along the system loop for distance minimization approaches, which was validated using a set of numerical experiments and a hardware experiment in which a camera was combined with a robotic hand. An iterative algorithm for maximizing the likelihood of the measurements was presented, along with an estimation uncertainty analysis. The proposed algorithm and uncertainty analysis were validated using numerical experiments, and the algorithm was further validated with a hardware experiment using two cameras. The experiments demonstrated the importance of careful coordination of A, B, X, and Y for the distance minimization approaches and confirmed the superior accuracy of the proposed algorithm compared to various existing approaches.

APPENDIX A
INTEGRATION ON $SE(3)$

Consider an integration on $SE(3)$ of the form

$$\int f(T)dT \tag{71}$$

where $f(T)$ is a scalar function of $T \in SE(3)$. The integration measure dT is expressed as $dT = dRdp$, where dp and dR are the integration measures on $\mathbb{R}^3$ and $SO(3)$, respectively. The positional integration measure dp is given straightforwardly as $dp = dp_x dp_y dp_z$, where p_x, p_y, and p_z are the x-, y-, and z-positions, respectively, whereas the algebraic expression of dR depends on the parameterization of $R \in SO(3)$. For the ZXZ Euler angle representation $R = R_{ZXZ}(\alpha, \beta, \gamma)$, the integration measure dR is given by [18], [36] as

$$dR = \sin\beta d\alpha d\beta d\gamma. \tag{72}$$

For the exponential coordinate $w = [w_x\ w_y\ w_z]^T$, where $R = \exp([w])$, dR is expressed as [37]

$$dR = \frac{2 - 2\cos\|w\|}{\|w\|^2} dw_x dw_y dw_z. \tag{73}$$

When (72) is used, integration is performed over $\alpha, \gamma \in [0, 2\pi]$, $\beta \in [0, \pi]$, and $p_x, p_y, p_z \in \mathbb{R}$. When (73) is used, the integration is calculated over $w \in \mathcal{U}$ and $p_x, p_y, p_z \in \mathbb{R}$, where $\mathcal{U}$ is an open ball of radius π in $\mathbb{R}^3$. Equations (72) and (73) are also known as the volume forms of $SO(3)$, which relate the infinitesimal volume on $SO(3)$ to the infinitesimal volume in the parameter space (e.g., a cuboid in R^3 for (α, β, γ) or an open ball in R^3 for w). These volume forms can be exploited when expressing a given PDF on $SO(3)$ equivalently in the parameter space, which is explained in detail in Appendix B, where random samplings on $SO(3)$ are described.

APPENDIX B
RANDOM SAMPLING ON $SO(3)$

The numerical experiments in this study involved two random samplings on $SO(3)$, including i) uniform sampling $R \sim U(R)$ and ii) the random sampling described in (28). Random sampling algorithms for these two distributions are presented as follows.

A. Uniform Sampling by $R \sim U(R)$

Uniform sampling of $SO(3)$ is traditionally derived based on the subgroup algorithm presented in [38], [39]. Another possible approach is to exploit the volume form of $SO(3)$ in the exponential coordinate provided in (73). Note that when w in the exponential coordinate is sampled uniformly, the resulting samples on $SO(3)$ are not uniform. Imagine a PDF on

the exponential coordinate, $p(w)$, for which the corresponding distribution on $SO(3)$ is uniform. Subsequently,

$$\int U(R)dR = \int p(w)dw \tag{74}$$

holds for any open set on $SO(3)$ and the corresponding open set on the exponential coordinate, where $R = \exp([w])$. By substituting the volume form (73), the above equation becomes

$$\int U(R)\frac{2-2\cos\|w\|}{\|w\|^2}dw = \int p(w)dw. \tag{75}$$

As $U(R)$ is uniform (i.e., constant), the PDF $p(w)$ becomes proportional to $\frac{2-2\cos\|w\|}{\|w\|^2}$. Based on this fact, a random w is sampled from $p(w)$ by uniformly sampling w in an open ball of radius π and randomly taking or rejecting it with a probability proportional to $\frac{2-2\cos\|w\|}{\|w\|^2}$. Note that the supremum of $\frac{2-2\cos\|w\|}{\|w\|^2}$ is 1, and the detailed sampling algorithm is determined as follows:

1) Uniformly sample $w \in \mathbb{R}^3$ from $\{w|\|w\| < \pi\}$.
2) Uniformly sample $r \in \mathbb{R}$ from [0,1].
3) If $r < \frac{2-2\cos\|w\|}{\|w\|^2}$, return $R = \exp([w])$. Otherwise, we repeat from (1).

B. Sampling by $R \sim k\exp(-\frac{1}{2}w^T\Sigma_w^{-1}w)$

Given a PDF of $SO(3)$ expressed as $p(R) = k\exp(-\frac{1}{2}w^T\Sigma_w^{-1}w)$, where $R = \exp([w])$, let $p(w)$ denote the corresponding PDF on the exponential coordinate. Similar to the case of uniform sampling, PDF $p(w)$ satisfies

$$\int p(R)\frac{2-2\cos\|w\|}{\|w\|^2}dw = \int p(w)dw \tag{76}$$

for any open set on $SO(3)$ and the corresponding open set on the exponential coordinate. Again, $p(w)$ is proportional to $\frac{2-2\cos\|w\|}{\|w\|^2}\exp(-\frac{1}{2}w^T\Sigma_w^{-1}w)$. Thus, the sampling algorithm can be derived as follows.

1) Uniformly sample $w \in \mathbb{R}^3$ from $\{w|\|w\| < \pi\}$.
2) Uniformly sample $r \in \mathbb{R}$ from [0,1].
3) If $r < \frac{2-2\cos\|w\|}{\|w\|^2}\exp(-\frac{1}{2}w^T\Sigma_w^{-1}w)$, return $R = \exp([w])$. Otherwise, we repeat from (1).

REFERENCES

[1] Y. C. Shiu and S. Ahmad, "Calibration of wrist-mounted robotic sensors by solving homogeneous transform equations of the form $AX = XB$," *IEEE Trans. Robot. Automat.*, vol. 5, no. 1, pp. 16–29, Feb. 1989.

[2] R. Y. Tsai and R. K. Lenz, "A new technique for fully autonomous and efficient 3D robotics hand/eye calibration," *IEEE Trans. Robot. Automat.*, vol. 5, no. 3, pp. 345–358, Jun. 1989.

[3] H. Zhuang, Z. S. Roth, and R. Sudhakar, "Simultaneous robot/world and tool/flange calibration by solving homogeneous transformation equations of the form $AX = YB$," *IEEE Trans. Robot. Automat.*, vol. 10, no. 4, pp. 549–554, Aug. 1994.

[4] J. Ha, D. Kang, and F. C. Park, "A stochastic global optimization algorithm for the two-frame sensor calibration problem," *IEEE Trans. Ind. Electron.*, vol. 63, no. 4, pp. 2434–2446, Apr. 2016.

[5] F. Dornaika and R. Horaud, "Simultaneous robot–world and hand–eye calibration," *IEEE Trans. Robot. Automat.*, vol. 14, no. 4, pp. 617–622, Aug. 1998.

[6] D. Condurache and I.-A. Ciureanu, "A novel solution for $AX = YB$ sensor calibration problem using dual lie algebra," in *Proc. 6th Int. Conf. Control, Decis. Inf. Technol.*, 2019, pp. 302–307.

[7] A. Li, L. Wang, and D. Wu, "Simultaneous robot–world and hand–eye calibration using dual-quaternions and Kronecker product," *Int. J. Phys. Sci.*, vol. 5, no. 10, pp. 1530–1536, 2010.

[8] M. Shah, "Solving the robot–world/hand–eye calibration problem using the Kronecker product," *J. Mechanisms Robot.*, vol. 5, no. 3, 2013, Art. no. 031007.

[9] J. Heller, D. Henrion, and T. Pajdla, "Hand–eye and robot–world calibration by global polynomial optimization," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2014, pp. 3157–3164.

[10] A. Tabb and K. M. A. Yousef, "Solving the robot–world hand–eye (s) calibration problem with iterative methods," *Mach. Vis. Appl.*, vol. 28, no. 5, pp. 569–590, 2017.

[11] F. Ernst et al., "Non-orthogonal tool/flange and robot/world calibration," *Int. J. Med. Robot. Comput. Assist. Surg.*, vol. 8, no. 4, pp. 407–420, 2012.

[12] A. Malti, "Hand–eye calibration with epipolar constraints: Application to endoscopy," *Robot. Auton. Syst.*, vol. 61, no. 2, pp. 161–169, 2013.

[13] K. Koide and E. Menegatti, "General hand–eye calibration based on reprojection error minimization," *IEEE Robot. Automat. Lett.*, vol. 4, no. 2, pp. 1021–1028, Apr. 2019.

[14] I. Ali, O. Suominen, A. Gotchev, and E. R. Morales, "Methods for simultaneous robot–world–hand–eye calibration: A comparative study," *Sensors*, vol. 19, no. 12, 2019, Art. no. 2837.

[15] M. Ulrich and M. Hillemann, "Generic hand–eye calibration of uncertain robots," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2021, pp. 11 060–11 066.

[16] F. C. Park, "Distance metrics on the rigid-body motions with applications to mechanism design," *J. Mech. Des.*, vol. 117, no. 1, pp. 48–54, 1995.

[17] Y. Zheng, S. Sugimoto, and M. Okutomi, "ASPnP: An accurate and scalable solution to the perspective-n-point problem," *IEICE Trans. Inf. Syst.*, vol. 96, no. 7, pp. 1525–1535, 2013.

[18] M. K. Ackerman and G. S. Chirikjian, "A probabilistic solution to the ax= xb problem: Sensor calibration without correspondence," in *Proc. Int. Conf. Geometric Sci. Inf.*, 2013, pp. 693–701.

[19] Q. Ma, H. Li, and G. S. Chirikjian, "New probabilistic approaches to the $AX = XB$ hand–eye calibration without correspondence," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2016, pp. 4365–4371.

[20] K. H. Strobl and G. Hirzinger, "Optimal hand-eye calibration," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2006, pp. 4647–4653.

[21] J. Weng, T. S. Huang, and N. Ahuja, "Motion and structure from two perspective views: Algorithms, error analysis, and error estimation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 5, pp. 451–476, May 1989.

[22] A. Zabatani et al., "Intel realsense sr300 coded light depth camera," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 10, pp. 2333–2345, 2019.

[23] S. Li, C. Xu, and M. Xie, "A robust O(n) solution to the perspective-n-point problem," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 7, pp. 1444–1450, Jul. 2012.

[24] F. Park, A. Murray, and J. McCarthy, "Designing mechanisms for workspace fit," in *Computational Kinematics*. Berlin, Germany: Springer, 1993, pp. 295–306.

[25] F. C. Park and B. Ravani, "Bezier curves on Riemannian manifolds and lie groups with kinematics applications," *J. Mech. Des.*, vol. 117, no. 1, pp. 36–40, 1995.

[26] D. Q. Huynh, "Metrics for 3D rotations: Comparison and analysis," *J. Math. Imag. Vis.*, vol. 35, no. 2, pp. 155–164, 2009.

[27] F. C. Park and B. J. Martin, "Robot sensor calibration: Solving $AX = XB$ on the euclidean group," *IEEE Trans. Robot. Automat.*, vol. 10, no. 5, pp. 717–721, Oct. 1994.

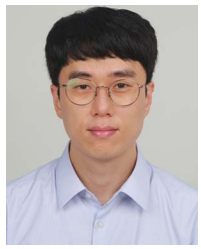
[28] I. Fassi and G. Legnani, "Hand to sensor calibration: A geometrical interpretation of the matrix equation $AX = XB$," *J. Robotic Syst.*, vol. 22, no. 9, pp. 497–506, 2005.

[29] J. Sola, J. Deray, and D. Atchuthan, "A micro lie theory for state estimation in robotics," 2018. [Online]. Available: https://arxiv.org/pdf/1812.01537.pdf

[30] J. Park and W.-K. Chung, "Geometric integration on euclidean group with application to articulated multibody systems," *IEEE Trans. Robot.*, vol. 21, no. 5, pp. 850–863, Oct. 2005.

[31] A. Noccaro, L. Raiano, G. Di Pino, and D. Formica, "Evaluation of hand-eye and robot-world calibration algorithms for TMS application," in *Proc. 7th IEEE Int. Conf. Biomed. Robot. Biomechatronics*, 2018, pp. 1115–1119.

[32] A. Noccaro, A. Mioli, M. D'Alonzo, M. Pinardi, G. Di Pino, and D. Formica, "Development and validation of a novel calibration methodology and control approach for robot-aided transcranial magnetic stimulation (TMS)," *IEEE Trans. Biomed. Eng.*, vol. 68, no. 5, pp. 1589–1600, May 2021.
[33] "e-Series from universal robots." [Online]. Available: https://arxiv.org/pdf/1812.01537.pdf
[34] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 11, pp. 1330–1334, Nov. 2000.
[35] R. Elfring, M. de la Fuente, and K. Radermacher, "Assessment of optical localizer accuracy for computer aided surgery systems," *Comput. Aided Surg.*, vol. 15, no. 1–3, pp. 1–12, 2010.
[36] H. Li, Q. Ma, T. Wang, and G. S. Chirikjian, "Simultaneous hand–eye and robot–world calibration by solving the $AX = YB$ problem without correspondence," *IEEE Robot. Automat. Lett.*, vol. 1, no. 1, pp. 145–152, Jan. 2016.
[37] F. C. Park, "The optimal kinematic design of mechanisms," Ph.D. dissertation, Harvard University, 1991.
[38] P. Diaconis and M. Shahshahani, "The subgroup algorithm for generating uniform random variables," *Probability Eng. Inf. Sci.*, vol. 1, no. 1, pp. 15–32, 1987.
[39] A. Yershova, S. Jain, S. M. Lavalle, and J. C. Mitchell, "Generating uniform incremental grids on SO(3) using the HOPF fibration," *Int. J. Robot. Res.*, vol. 29, no. 7, pp. 801–812, 2010.



Junhyoung Ha (Member, IEEE) received the B.S. and Ph.D. degrees in mechanical and aerospace engineering from the Seoul National University, Seoul, South Korea, in 2008 and 2015, respectively.

From 2015 to 2018, he was a Postdoctoral Researcher with Boston Children's Hospital, Harvard Medical School. Since 2019, he has been a Senior Researcher with the Center for Healthcare Robotics, AI and Robotics Institute, Korea Institute of Science and Technology (KIST), Seoul, South Korea. He is also currently an Associate Professor with the Division of AI and Robot, KIST School, University of Science and Technology (UST), Seoul, South Korea. His research interests include robot control, motion planning, and deep learning.