

Correspondence Matching and Time Delay Estimation for Hand-eye Calibration

Jin Wu , *Member, IEEE*, Ming Liu , *Senior Member, IEEE*, Chengxi Zhang  and Zebo Zhou 

Abstract—For visually aided industrial robots, the transformation between the mounted visual information system and the end-effector must be calibrated prior to the practical use. However, during the data acquisition stage, measurement uncertainties will lead to inevitable mismatched data. This paper proposes a novel correspondence matching method for the hand-eye calibration system of the type $AX = XB$. The correspondence matching refers to the problem of finding out the well-matched A and B and the relative time delay between them, from a series of measurements. A neat Lie algebra formulation has been obtained to cast the original problem into a high-dimensional point-cloud registration problem. This solution allows for the simultaneous solution of the hand-eye calibration, correspondence matching and time-delay estimation as well as the uncertainty description of the obtained results. Synthetic simulations verify the correctness of the proposed method. Experimental studies have also confirmed its practical feasibility on an automatic welding platform.

Index Terms—Hand-eye Calibration, Correspondence Matching, Welding Robot, Pose Estimation, Iterative Closest Point

I. INTRODUCTION

HAND-EYE calibration techniques have been first reported in the late 1980s by Shiu et al. [1] and Tsai [2]. The problem considered in the hand-eye calibration is to compute the transformation between the camera and the robot gripper so that measurements from the two parts can be unified to a common standard frame [3]. One breakthrough that early endeavors achieved is that the hand-eye calibration problem can be described precisely with the mathematical formulation of $AX = XB$ with A, B the known homogeneous transformations in two different frames respectively, and X being the unknown transformation as the task of the hand-eye calibration. Therefore, many further studies focus on more efficient

solution of such type of equation [4]. It has been found out that the problem $AX = XB$ can be decomposed into sub-systems containing rotation and translation respectively [5]. Thus a natural solution exists that the rotation and translation can be solved separately. This has been achieved by Shiu et al., Park et al. and etc. However, according to the nature that the rotation and translation may be coupled nonlinearly together in the equation $AX = XB$, the above derived solutions are not optimal. Then simultaneous solving methods have been introduced which include analytical algorithms [6] and numerical iterative solutions like Branch-and-Bound (BnB) [7].

For the hand-eye calibration problem $AX = XB$, one must first acquire measurements from different equipments to form the series of A and B . In practice, due to measurement uncertainties, each pair of the A and B will not always be well matched. In early reports, Tsai et al. have considered such issue in practical calibration tasks [2]. It is obtained that the mismatching due to the asynchronization and data outliers may induce severely bad calibration results. Tsai et al. concluded that some measurement techniques must be performed to enhance the calibration accuracy. However, some of Tsai's suggestions are not fully practical for all hand-eye calibration tasks. This lead to the study of the correspondence-free hand-eye calibration that incorporates the measurement uncertainty in a probabilistic formulation [8]. What needs to be pointed out is that such outlier-robust approach can not fully eliminate the data mismatching. Rather, they highly depend on the certain distribution of the measurements while finding a completely proper distribution will be another hard problem [9]. The BnB method [7] considers the inter-camera correspondence matching (i.e. from A or B) by including epipolar constraints into the optimization. The correspondence matching problem presented in this paper refers to selecting the most appropriate match between A and B . This problem has been recently studied by Furrer et al. who effectively exploit the correspondence matching and time-delay estimation by means of the random sample consensus (RANSAC) and the maximum likelihood estimation (MLE) [10]. However, the RANSAC method may not be very accurate since it requires a precise initial guess for the convergent performance. The accuracy of the refinement step in [10] is also highly determined by the quality of the initial guess. It also follows an assumption that the time delay is a constant, which may vary according to different sensor connections and sampling methods.

Another challenge for the correspondence matching and time delay estimation is the computational burden. As hand-eye calibration is non-convex in principle, various solutions

Manuscript received December 20, 2019; revised February 2, 2020 and April 4, 2020; accepted April 9, 2020. This work was supported in part by the National Natural Science Foundation of China under Grant 41604025, in part by the General Research Fund of Research Grants Council, Hong Kong, under Grant 11210017, and in part by the Early Career Scheme Project of Research Grants Council, Hong Kong, under Grant 21202816. The work of M. Liu was supported by the Shenzhen Science, Technology and Innovation Commission (SZSTI) under Grant JCYJ20160401100022706. The Associate Editor coordinating the review process was X X. (*Corresponding author: Ming Liu*)

J. Wu and M. Liu are with Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology, Hong Kong, China. (e-mail: jin_wu_uestc@hotmail.com; eelium@ust.hk)

C. Zhang is with Harbin Institute of Technology at Shenzhen, Shenzhen, China. (e-mail: dongfangxy@163.com)

Z. Zhou is with School of Aeronautics and Astronautics, University of Electronic Science and Technology of China, Chengdu, China. (e-mail: klinsmann.zhou@gmail.com)

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TIM.2020.XXX

are time-consuming and thus making them much more computationally inefficient in the RANSAC-based matching. For instance, in practice, the stochastic rotation searching [11] takes about 14 seconds on a typical personal laptop while the alternative linear programming (ALP) can consume up to 130 ~ 190 seconds in the real experiments [12]. Then relying on such time bases, the RANSAC has to be computed iteratively for many loops until it converges, which leads to quite uncontrollable computational efficiency in the engineering applications. To overcome this common drawback, one must find a new mathematical tool to better describe the relationship between the measurements.

The proposed method described in this paper is based on our recent contribution [4] which solves the hand-eye calibration problem with rotation and translation simultaneously with the framework of 4D procrustes analysis (4DPA). The aim of the current paper is to give a simple and globally convergent framework for correspondence matching. The major innovations and contributions of the proposed work are:

- 1) The Lie algebra is invoked for parameterization of the problem. A new point-cloud registration framework is then obtained, which could be further refined by the iterative closest point (ICP) [13].
- 2) The correspondence matching will conduct outlier detection and removal automatically and thus enhance the hand-eye calibration accuracy.
- 3) The designed scheme considers the rotation and translation simultaneously so there will not be loss of accuracy from the aspect of the rotation-translation coupling.
- 4) The time-delay estimation can then be flexibly integrated with the correspondence matching in a simultaneous manner. By virtue of the employed ICP framework, it is able for us to obtain the uncertainty descriptions of the obtained hand-eye calibration results. A visual welding robotic platform is utilized for the validation of the effectiveness of the proposed approach.

The remainder of this paper is organized as follows: Section II introduces our novel method and related derivations. Synthetic simulations and related experimental results using the visual welding robot have been reported in the Section III. This paper ends with the concluding remarks and discussion of future works, which will be presented in Section IV.

II. PROPOSED ALGORITHM

A. Correspondence Matching

The hand-eye calibration problem involved in this paper can be expressed as follows

$$\arg \min_{\mathbf{X} \in \mathbb{O}} \mathcal{L} = \sum_{i=1}^N \|\mathbf{A}_i \mathbf{X} - \mathbf{X} \mathbf{B}_i\|^2 \quad (1)$$

where $\{\mathbf{A}_i, \mathbf{B}_i\}$ for $i = 1, 2, \dots, N$ are series of pose measurements and \mathbf{X} is the transformation to be estimated which belongs to a certain manifold \mathbb{O} . In existing studies, \mathbb{O} can be the special orthogonal group $SO(3)$ and the special Euclidean group $SE(3)$. Our recent study shows that when $\mathbf{X} \in SE(3)$, the problem can be cast into an equivalent

one on the special orthogonal group $SO(4)$ [4]. As $SO(3)$ and $SO(4)$ are certain subsets of the Lie group, this paper solves the correspondence matching of \mathbf{A}_i and \mathbf{B}_i on a general special orthogonal group $SO(n)$ such that the constraints $\mathbf{X}^\top \mathbf{X} = \mathbf{I}, \det(\mathbf{X}) = 1$ will be met. The correspondence matching problem can be described mathematically via

$$\arg \min_{\mathbf{X} \in SO(n)} \sum_{i=1}^N \left[\sum_{\mathbf{A}_j \in \mathcal{A}_i, \mathbf{B}_k \in \mathcal{B}_i} \|\mathbf{A}_j \mathbf{X} - \mathbf{X} \mathbf{B}_k\|^2 \right] \quad (2)$$

where $\mathcal{A}_i, \mathcal{B}_i$ are subsets of the neighborhood measurements around \mathbf{A}_i and \mathbf{B}_i , with the size of M . The optimization is based on [14] that extends the hand-eye calibration to a more complex optimization. Unfortunately, according to the nonlinearity of \mathbf{A}, \mathbf{B} and \mathbf{X} , the above problem is non-convex. The challenge remaining to us is to seek an approach that turns (2) into a globally solvable problem. For any rotation matrix \mathbf{X} on $SO(n)$, there is always a $\frac{n(n-1)}{2} \times 1$ vector \mathbf{x} corresponds to it such that $\exp(\mathbf{x}_\times) = \mathbf{X}$ which is called the exponential mapping. The \mathbf{x}_\times denotes a mapping from the $\frac{n(n-1)}{2} \times 1$ vector to the $n \times n$ rotation matrix which can be found in [15]. The wedge operation denotes an inverse from \mathbf{x}_\times to \mathbf{x} so that $\mathbf{x}_\times^\wedge = \mathbf{x}$. \mathbf{x} is also called the Lie algebra of the rotation matrix \mathbf{X} . The inverse of the exponential mapping is the logarithmic mapping such that

$$\log[\exp(\mathbf{x}_\times)] = \mathbf{x}_\times \quad (3)$$

According to the existing results, when $\mathbf{X} \in SO(3)$, the following identities hold

$$\boldsymbol{\alpha}_i = \mathbf{X} \boldsymbol{\beta}_i \quad (4)$$

for the noise-free case, where $\boldsymbol{\alpha}_i = [\log \mathbf{A}_i]^\wedge$ and $\boldsymbol{\beta}_i = [\log \mathbf{B}_i]^\wedge$. Park et al. then formulates the estimation of \mathbf{X} into the point-cloud registration such that [5]

$$\arg \min_{\mathbf{X} \in SO(3)} \sum_{i=1}^N \|\boldsymbol{\alpha}_i - \mathbf{X} \boldsymbol{\beta}_i\|^2 \quad (5)$$

whose closed-form solution is presented in [16]. However, when \mathbf{X} is on $SO(n)$, there will be no such a relation in (4). The reason is that the sizes of the $\boldsymbol{\alpha}_i, \boldsymbol{\beta}_i$ and \mathbf{X} are not consistent. From the equality of hand-eye calibration, it has been obtained that

$$\log \mathbf{A}_i = \log \mathbf{X} \mathbf{B}_i \mathbf{X}^\top \quad (6)$$

which implies that $\log \mathbf{A}_i$ and $\log \mathbf{B}_i$ should have the same norm as \mathbf{X} plays a norm-preserving role, according to the Lie theory [5]. Thus, it is able to rewrite the hand-eye equality into

$$\boldsymbol{\alpha}_i = \mathbf{G} \boldsymbol{\beta}_i \quad (7)$$

where \mathbf{G} is an orthonormal matrix but not properly on $SO(n)$, which follows *only* that $\mathbf{G}^\top \mathbf{G} = \mathbf{I}$. The matrix \mathbf{G} is in the quadratic form of the elements of \mathbf{X} and can be evaluated

$$\mathbf{G} = \begin{pmatrix} x_{34}x_{43} - x_{33}x_{44} & x_{32}x_{44} - x_{34}x_{42} & x_{34}x_{41} - x_{31}x_{44} & x_{33}x_{42} - x_{32}x_{43} & x_{31}x_{43} - x_{33}x_{41} & x_{32}x_{41} - x_{31}x_{42} \\ x_{23}x_{44} - x_{24}x_{43} & x_{24}x_{42} - x_{22}x_{44} & x_{21}x_{44} - x_{24}x_{41} & x_{22}x_{43} - x_{23}x_{42} & x_{23}x_{41} - x_{21}x_{43} & x_{21}x_{42} - x_{22}x_{41} \\ x_{14}x_{43} - x_{13}x_{44} & x_{12}x_{44} - x_{14}x_{42} & x_{14}x_{41} - x_{11}x_{44} & x_{13}x_{42} - x_{12}x_{43} & x_{11}x_{43} - x_{13}x_{41} & x_{12}x_{41} - x_{11}x_{42} \\ x_{24}x_{33} - x_{23}x_{34} & x_{22}x_{34} - x_{24}x_{32} & x_{24}x_{31} - x_{21}x_{34} & x_{23}x_{32} - x_{22}x_{33} & x_{21}x_{33} - x_{23}x_{31} & x_{22}x_{31} - x_{21}x_{32} \\ x_{13}x_{34} - x_{14}x_{33} & x_{14}x_{32} - x_{12}x_{34} & x_{11}x_{34} - x_{14}x_{31} & x_{12}x_{33} - x_{13}x_{32} & x_{13}x_{31} - x_{11}x_{33} & x_{11}x_{32} - x_{12}x_{31} \\ x_{14}x_{23} - x_{13}x_{24} & x_{12}x_{24} - x_{14}x_{22} & x_{14}x_{21} - x_{11}x_{24} & x_{13}x_{22} - x_{12}x_{23} & x_{11}x_{23} - x_{13}x_{21} & x_{12}x_{21} - x_{11}x_{22} \end{pmatrix} \quad (9)$$

with various symbolic computation engines like MATLAB and Mathematica. For instance, if $\mathbf{X} \in SO(3)$, we have

$$\mathbf{G} = \begin{pmatrix} x_{23}x_{32} - x_{22}x_{33} & x_{21}x_{33} - x_{23}x_{31} & x_{22}x_{31} - x_{21}x_{32} \\ x_{12}x_{33} - x_{13}x_{32} & x_{13}x_{31} - x_{11}x_{33} & x_{11}x_{32} - x_{12}x_{31} \\ x_{13}x_{22} - x_{12}x_{23} & x_{11}x_{23} - x_{13}x_{21} & x_{12}x_{21} - x_{11}x_{22} \end{pmatrix} \quad (8)$$

where x_{ij} denotes the entry of \mathbf{X} in the i -th row and j -th column. The expression of \mathbf{G} for that $\mathbf{X} \in SO(4)$ is (9). Here, using symbolic engine, one can verify $\mathbf{G} = -\mathbf{X}$ and $\det(\mathbf{G}) = -1$. For the case of $\mathbf{X} \in SO(4)$, we have $\det(\mathbf{G}) = 1$.

When $\det(\mathbf{G}) = -1$, one has $\det(-\mathbf{G}) = 1$ for $\text{mod}(n, 2) = 0$, $\det(-\mathbf{G}) = -1$ for $\text{mod}(n, 2) = 1$ and $(-\mathbf{G})^\top(-\mathbf{G}) = \mathbf{I}$ i.e. $-\mathbf{G} \in SO(n)$. Denoting $\mathbf{Y} = \det(\mathbf{G})\mathbf{G} \in SO(n)$ yields

$$[\det(\mathbf{G})] \boldsymbol{\alpha}_i = \mathbf{Y}\boldsymbol{\beta}_i \quad (10)$$

Then (2) can be cast into

$$\arg \min_{\mathbf{Y} \in SO(n)} \sum_{i=1}^N \left[\sum_{\mathbf{A}_j \in \mathcal{A}_i, \mathbf{B}_k \in \mathcal{B}_i} \|\det(\mathbf{G}) \boldsymbol{\alpha}_j - \mathbf{Y}\boldsymbol{\beta}_k\|^2 \right] \quad (11)$$

where $\det(\mathbf{G})$ is the predetermined sign of \mathbf{G} using the symbolic engine. The optimization in (11) actually takes the form of the ICP that has received extensive research attention during the past 30 years. Although (11) is still non-convex, the globally optimal solutions have been recently reported, such as the Go-ICP [17]. Alternatively, (11) can also be solved via the semidefinite programming [14]. Therefore such a solution in (11) is much more accurate, robust and controllable than that in (2). Note that (11) is in the form of ICP so the uncertainty description (covariance matrix) of \mathbf{Y} can be obtained using the inverse of the Hessian, i.e., an approximation of the Fisher information matrix.

B. Solving \mathbf{X} from \mathbf{Y}

Note that the correspondence matching problem is shifted to compute \mathbf{Y} rather than \mathbf{X} . That is to say \mathbf{X} does not need to be computed. However as there is an one-to-one quadratic mapping from \mathbf{Y} to \mathbf{X} , solving \mathbf{X} from \mathbf{Y} belongs to the quadratic polynomial optimization, which is quite simple in engineering practice. Therefore, the presented approach not only solves the correspondence matching problem, but also gives a novel quadratic optimization approach to the hand-eye calibration, which can be conveniently solved via common methods like gradient descent and Levenberg-Marquardt algorithm (LMA) [18]. However, iterative methods may be not so computationally efficient in practice and may suffer from local optimum. In the following contents, an efficient analytical method is derived.

Solving \mathbf{X} from \mathbf{Y} can be shifted into solving \mathbf{X} from \mathbf{G} . Given a known matrix \mathbf{G} , the shown structures in (8) and (9) indicate that each element of \mathbf{G} is in the quadratic form of \mathbf{X} . Denoting $\mathbf{x} = \text{vec}(\mathbf{X})$, we have

$$\mathbf{g} = \text{vec}(\mathbf{G}) = \mathbf{J}\mathbf{x} \quad (14)$$

where \mathbf{g} is the vectorization of \mathbf{G} and \mathbf{J} represents a matrix in the linear form of elements of \mathbf{x} . In particular, \mathbf{J} denotes a matrix proportional to the Jacobian of \mathbf{g} with respect to \mathbf{x} . Thus, \mathbf{J} can be called as the pseudo Jacobian for the sake of convenience. The gradient descent algorithm then looks like

$$\mathbf{x}_p = \mathbf{x}_{p-1} - \gamma_p \frac{\mathbf{J}^\top (\mathbf{J}\mathbf{x}_{p-1} - \mathbf{g})}{\|\mathbf{J}^\top (\mathbf{J}\mathbf{x}_{p-1} - \mathbf{g})\|} \quad (15)$$

where p denotes the iteration index and γ_p is a positive gain (step length) for converging search. For $\mathbf{X} \in SO(3)$, the pseudo Jacobian \mathbf{J} is shown in (12), while for $\mathbf{X} \in SO(4)$, the expression is given in (13). Note that, the iteration (15) is locally optimal so a good analytical solution gives perfect initial guess, which has been discussed in [4].

$$\mathbf{J} = \begin{pmatrix} 0 & 0 & 0 & 0 & -x_{33} & x_{23} & 0 & x_{32} & -x_{22} \\ 0 & 0 & 0 & x_{33} & 0 & -x_{13} & -x_{32} & 0 & x_{12} \\ 0 & 0 & 0 & -x_{23} & x_{13} & 0 & x_{22} & -x_{12} & 0 \\ 0 & x_{33} & -x_{23} & 0 & 0 & 0 & 0 & -x_{31} & x_{21} \\ -x_{33} & 0 & x_{13} & 0 & 0 & 0 & x_{31} & 0 & -x_{11} \\ x_{23} & -x_{13} & 0 & 0 & 0 & 0 & -x_{21} & x_{11} & 0 \\ 0 & -x_{32} & x_{22} & 0 & x_{31} & -x_{21} & 0 & 0 & 0 \\ x_{32} & 0 & -x_{12} & -x_{31} & 0 & x_{11} & 0 & 0 & 0 \\ -x_{22} & x_{12} & 0 & x_{21} & -x_{11} & 0 & 0 & 0 & 0 \end{pmatrix} \quad (12)$$

$$\mathbf{J} = \begin{pmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -x_{44} & x_{34} & 0 & 0 & x_{43} & -x_{33} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & x_{44} & 0 & -x_{24} & 0 & -x_{43} & 0 & x_{23} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -x_{44} & 0 & 0 & x_{14} & x_{43} & 0 & 0 & -x_{13} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -x_{34} & x_{24} & 0 & 0 & x_{33} & -x_{23} & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & x_{34} & 0 & -x_{14} & 0 & -x_{33} & 0 & x_{13} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -x_{24} & x_{14} & 0 & 0 & x_{23} & -x_{13} & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & x_{44} & -x_{34} & 0 & 0 & 0 & 0 & 0 & 0 & -x_{42} & x_{32} \\
0 & 0 & 0 & 0 & 0 & -x_{44} & 0 & x_{24} & 0 & 0 & 0 & 0 & 0 & 0 & x_{42} & 0 \\
0 & 0 & 0 & 0 & x_{44} & 0 & 0 & -x_{14} & 0 & 0 & 0 & 0 & -x_{42} & 0 & 0 & x_{12} \\
0 & 0 & 0 & 0 & 0 & x_{34} & -x_{24} & 0 & 0 & 0 & 0 & 0 & 0 & -x_{32} & x_{22} & 0 \\
0 & 0 & 0 & 0 & -x_{34} & 0 & x_{14} & 0 & 0 & 0 & 0 & 0 & x_{32} & 0 & -x_{12} & 0 \\
0 & 0 & 0 & 0 & x_{24} & -x_{14} & 0 & 0 & 0 & 0 & 0 & 0 & -x_{22} & x_{12} & 0 & 0 \\
0 & 0 & -x_{44} & x_{34} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & x_{41} & -x_{31} \\
0 & x_{44} & 0 & -x_{24} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -x_{41} & 0 \\
-x_{44} & 0 & 0 & x_{14} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & x_{41} & 0 & 0 & -x_{11} \\
0 & -x_{34} & x_{24} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & x_{31} & -x_{21} & 0 \\
x_{34} & 0 & -x_{14} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -x_{31} & 0 & x_{11} & 0 \\
-x_{24} & x_{14} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & x_{21} & -x_{11} & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & -x_{43} & x_{33} & 0 & 0 & x_{42} & -x_{32} & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & x_{43} & 0 & -x_{23} & 0 & -x_{42} & 0 & x_{22} & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & -x_{43} & 0 & 0 & x_{13} & x_{42} & 0 & 0 & -x_{12} & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & -x_{33} & x_{23} & 0 & 0 & x_{32} & -x_{22} & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & x_{33} & 0 & -x_{13} & 0 & -x_{32} & 0 & x_{12} & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & -x_{23} & x_{13} & 0 & 0 & x_{22} & -x_{12} & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & x_{43} & -x_{33} & 0 & 0 & 0 & 0 & 0 & 0 & -x_{41} & x_{31} & 0 & 0 & 0 & 0 \\
0 & -x_{43} & 0 & x_{23} & 0 & 0 & 0 & 0 & 0 & 0 & x_{41} & 0 & -x_{21} & 0 & 0 & 0 \\
x_{43} & 0 & 0 & -x_{13} & 0 & 0 & 0 & 0 & -x_{41} & 0 & 0 & x_{11} & 0 & 0 & 0 & 0 \\
0 & x_{33} & -x_{23} & 0 & 0 & 0 & 0 & 0 & 0 & -x_{31} & x_{21} & 0 & 0 & 0 & 0 & 0 \\
-x_{33} & 0 & x_{13} & 0 & 0 & 0 & 0 & 0 & x_{31} & 0 & -x_{11} & 0 & 0 & 0 & 0 & 0 \\
x_{23} & -x_{13} & 0 & 0 & 0 & 0 & 0 & 0 & -x_{21} & x_{11} & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & -x_{42} & x_{32} & 0 & 0 & x_{41} & -x_{31} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & x_{42} & 0 & -x_{22} & 0 & -x_{41} & 0 & x_{21} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
-x_{42} & 0 & 0 & x_{12} & x_{41} & 0 & 0 & -x_{11} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -x_{32} & x_{22} & 0 & 0 & x_{31} & -x_{21} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
x_{32} & 0 & -x_{12} & 0 & -x_{31} & 0 & x_{11} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
-x_{22} & x_{12} & 0 & 0 & x_{21} & -x_{11} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{pmatrix} \quad (13)$$

C. Time-Delay Estimation

The transformations \mathbf{A}_i and \mathbf{B}_i are sampled in time series so the the sequential data indicates the motion of the camera or the gripper. Although \mathbf{A}_i and \mathbf{B}_i all own the same sampling index i , the actual timestamps may be very unsynchronized. The time-delay estimation is to determine how much \mathbf{A}_i has lagged or leaded \mathbf{B}_i in the timeline. Due to the continuous motion of the gripper and camera, it is able for us to restore the motion trajectory by the B-spline on Lie group [19]. Then we can reconstruct intense neighborhoods of \mathbf{A}_i and \mathbf{B}_i by the motion prediction or smoothing. The sizes of such neighborhoods are decided by the value of the time-delay. For engineering purposes, we can choose a relative big value and a small step length to construct the neighborhoods. Note that, with more points, the ICP computes more slowly and the step length can not be too tiny that generates too much points. Therefore, the time delay τ also participates in the final form of ICP, whose refined expression is

$$\arg \min_{\mathbf{Y} \in SO(n)} \sum_{i=1}^N \left[\sum_{\mathbf{A}_j \in \mathcal{A}_i(\Gamma, l), \mathbf{B}_k \in \mathcal{B}_i} \|[\det(\mathbf{G})] \alpha_j - \mathbf{Y} \beta_k\|^2 \right] \quad (16)$$

where $\mathcal{A}_i(\Gamma, l)$ is the reconstructed neighborhoods of \mathcal{A}_i for optimal searching with step length l and time uncertainty span of Γ . Here \mathcal{B}_i is fixed to determine the relative time delay τ and \mathcal{A}_i can also be fixed in the same manner. The developed algorithm reconstructs two point clouds with points in $\mathcal{A}_i(\tau, l)$ and \mathcal{B}_i . The time delay is estimated by $\tau = \Delta t_{j,k}$ which denotes the time span between the indices j and k . Then the ICP will estimate the rigid transformation until the loss function in (16) converges. The practice that takes the time delay into the account for the temporal calibration between inertial measurement unit (IMU) and camera has been validated by Kelly et al. [20]. Therefore, (16) can be used for convergent time delay estimation. Also, there is no constraint that τ should be constant, so the developed method is more applicable to real devices with heterogeneous measurement and transmission uncertainties.

III. EXPERIMENTAL RESULTS

The symbolic derivation and numerical evaluation of the proposed method in this paper, have been made open-source on https://github.com/zarathustr/hand_eye_corr. We herein introduce the synthetic simulations and real experiments to illustrate the superiority of the proposed scheme.

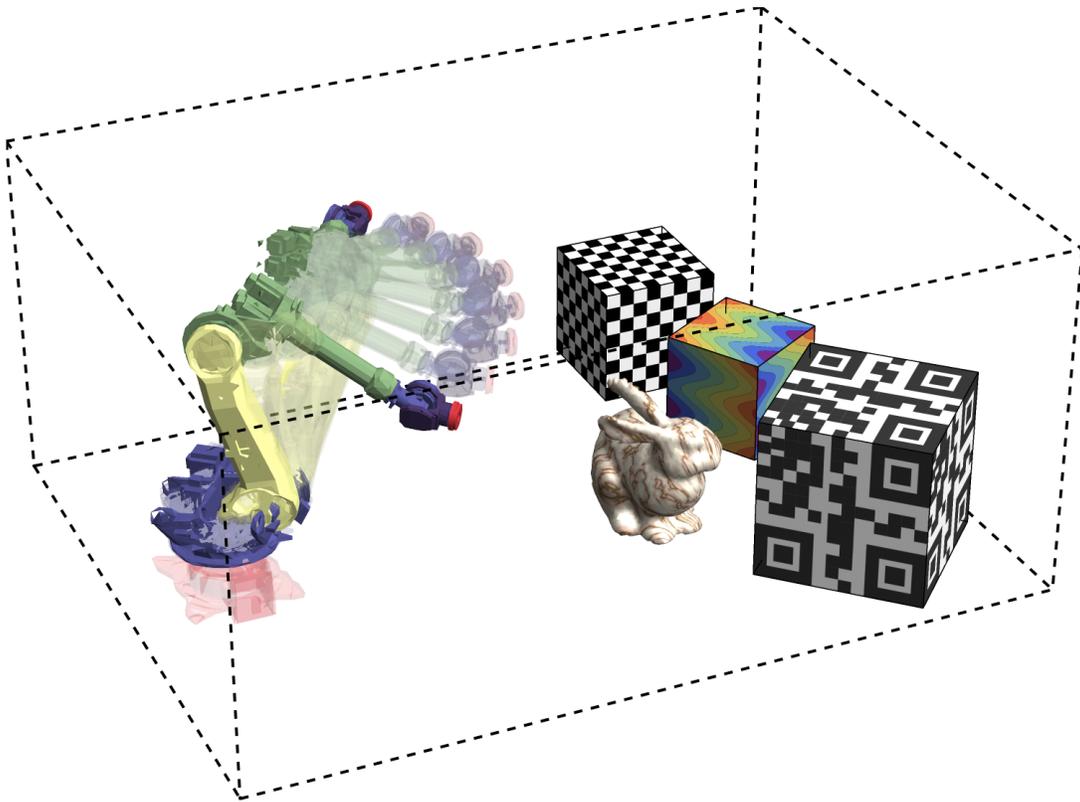


Fig. 1: The synthetic experiment setup.

A. Synthetic Simulations

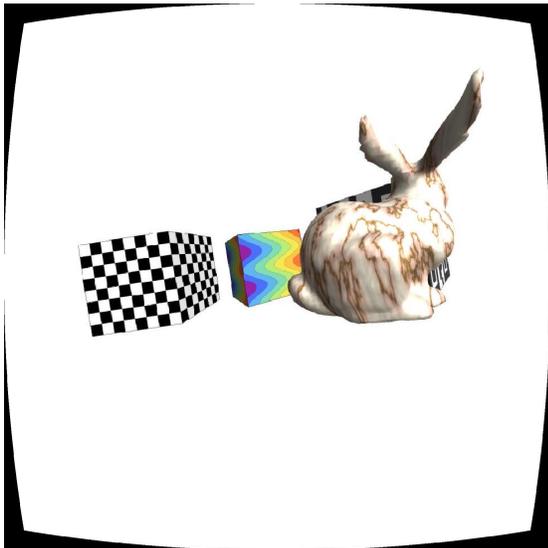


Fig. 2: A sample captured image from the fisheye camera.

In this sub-section, we conduct two synthetic simulations using the scene setup in Fig. 1. There are various simulated objects like the 3-D rabbit model from Stanford [21], a rainbow box and two boxes with Quick Response (QR) code and checkerboard textures. A Motoman ES-165D 3-D model has been invoked for simulation of the robotic manipulator. The fisheye camera of the resolution 1920×1080 is connected

to the end-effector, which is shown in dark red in Fig. 1, with the equidistant field-of-view (FOV) of 120° . A motion planning method for visually guided robots has been employed to generate the continuous motion of the robotic manipulator [22]. The ideal extrinsic parameter between the camera and the end-effector is given by the following homogeneous transformation in $SE(3)$

$$\mathbf{T}_{\text{true}} = \begin{pmatrix} \mathbf{R}_{\text{true}} & \mathbf{t}_{\text{true}} \\ \mathbf{0} & 1 \end{pmatrix} \quad (17)$$

where

$$\mathbf{R}_{\text{true}} = \begin{pmatrix} 0.98894 & 0.14595 & 0.02623 \\ -0.14712 & 0.98786 & 0.04987 \\ -0.01863 & -0.05317 & 0.99841 \end{pmatrix} \quad (18)$$

$$\mathbf{t}_{\text{true}} = (0.27950, 1.81155, -0.19203)^\top \quad (\text{dm})$$

The sample of captured image using the fisheye camera is shown in Fig. 2. The image distortion has been corrected using the offline calibration via [23]. To obtain measurements of the type $\mathbf{AX} = \mathbf{XB}$, the relative transformations have to be computed [24]. The relative transformations of the robotic end-effector is computed using the readings from the simulated angle encoders. We use the epipolar geometry to compute the relative pose between two camera frames [25]. Fig. 3 shows the corrected images and the matched features in the camera frames. Note that, here the speeded-up robust features (SURF) has been employed for the feature extraction [26].

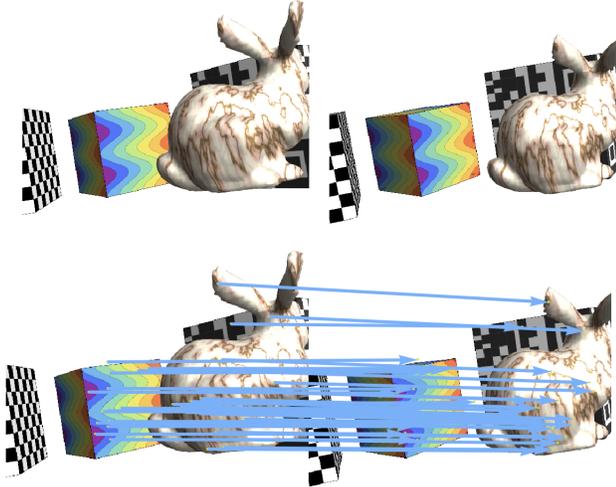


Fig. 3: The undistorted images and the matched features.

The $SE(3)$ poses are converted to the hand-eye rotations on $SO(4)$ via the mapping \mathcal{F} [4]

$$T = \begin{pmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0} & 1 \end{pmatrix} \xrightarrow[\mathcal{F}^{-1}]{\mathcal{F}} \mathbf{X} = \begin{pmatrix} \mathbf{R} & \varepsilon \mathbf{t} \\ -\varepsilon \mathbf{t}^T \mathbf{R} & 1 \end{pmatrix} \quad (19)$$

where $\mathbf{R} \in SO(3)$ is the rotation while $\mathbf{t} \in \mathbb{R}^3$ denotes the translation vector; ε denotes a scaling factor that couples 3-D rotation and translation into an $SO(4)$ rotation matrix. Orthonormalizing \mathbf{X} in (19) produces the hand-eye problem $\mathbf{A}\mathbf{X} = \mathbf{X}\mathbf{B}$ on $SO(4)$ [27]. Two categories of simulations are conducted, they are:

- 1) The first simulation consists of a constant time delay of 0.3s for camera with respect to the robotic manipulator.
- 2) The second simulation considers time-varying time delays.

We set Γ to 0.9s indicating a large searching bound of the time delay. For the first experiment, 6203 camera poses are captured for hand-eye calibration. We first use the analytical solution in [4] and then refine it using the proposed method in this paper. In Fig. 4, mean root loss function values of (1) where $\mathbb{O} = SE(3)$ are shown. There are two independent processes involved: one contains only correspondence matching while another also owns time delay estimation. The convergence rates of the ICP indicate that the two processes are convergent. The BnB method [7] and the 4DPA method [4] are invoked for comparison with the proposed method. Both BnB and 4DPA do not consider the time delay, the estimated extrinsic parameters are

$$\begin{aligned} \mathbf{R}_{\text{BnB}} &= \begin{pmatrix} 0.91112 & 0.12094 & 0.39399 \\ -0.10401 & 0.99251 & -0.06414 \\ -0.39880 & 0.01746 & 0.91688 \end{pmatrix} \\ \mathbf{t}_{\text{BnB}} &= (-0.04893, 2.59782, 0.44775)^\top \quad (\text{dm}) \\ \mathbf{R}_{\text{4DPA}} &= \begin{pmatrix} 0.90824 & -0.28481 & 0.30658 \\ 0.36727 & 0.89367 & -0.25781 \\ -0.20056 & 0.34675 & 0.91627 \end{pmatrix} \\ \mathbf{t}_{\text{4DPA}} &= (-0.15772, 1.06920, -0.68103)^\top \quad (\text{dm}) \end{aligned} \quad (20)$$

They all contain large rotation and translation errors with respect to the true values. The proposed method with correspondence matching and time delay estimation owns the following estimate

$$\mathbf{R}_{\text{proposed}} = \begin{pmatrix} 0.98684 & 0.14595 & 0.06956 \\ -0.14939 & 0.98766 & 0.04712 \\ -0.06182 & -0.05689 & 0.99646 \end{pmatrix} \quad (21)$$

$$\mathbf{t}_{\text{proposed}} = (0.20721, 1.75934, -0.18335)^\top \quad (\text{dm})$$

which is very close to the true values in (17). The process in

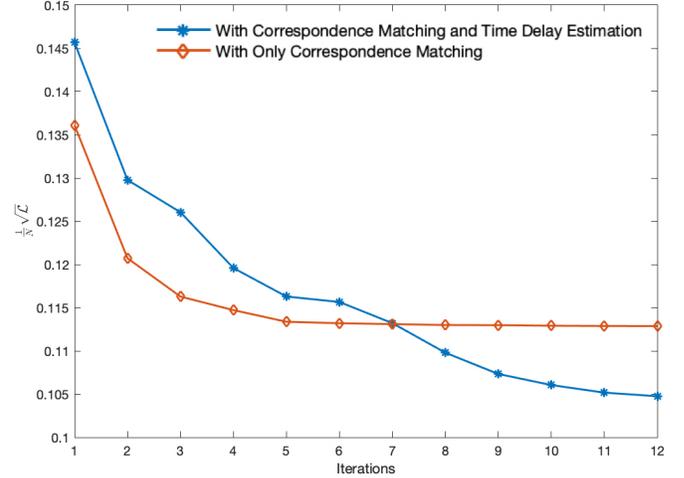


Fig. 4: Convergence of the ICP for correspondence matching and time-delay estimation.

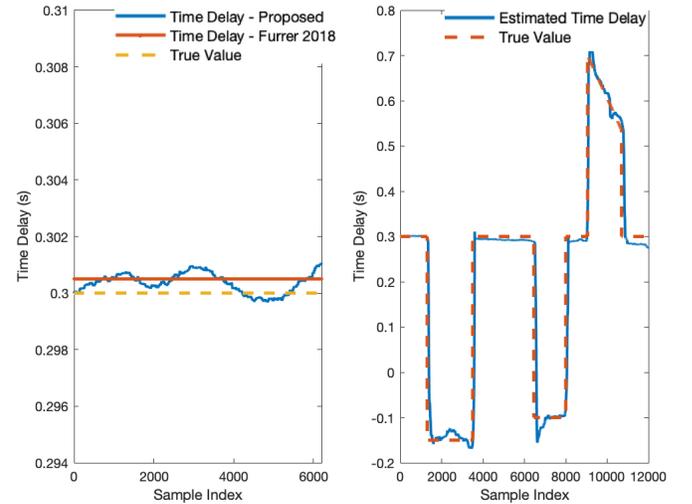


Fig. 5: Estimated time delays.

the presence of time delay estimation is more accurate than the one with correspondence matching only. It is also noticed that, time delay estimation also consumes a few more iterations to converge to the optimum. The settled true value of the time delay along with the estimates from [10] and the proposed method are shown in the left sub-figure of Fig. 5. The second experiment presents 12041 camera poses and the ideal time delay has been generated via the true values in the right sub-figure of Fig. 5 where step impulse and trapezoidal varying

delays are presented. The method [10] can only estimate the constant time delay and for our algorithm, the computed time delay is varying but also fits the ideal constant time delay. The method [10] can not handle time-varying time delay estimation. The proposed method, however, gives perfect time delay estimation, compared with the true values. The two synthetic experiments indicate that proposed method is correct, convergent and more efficient than the previous representative in the presence of time-varying delays.

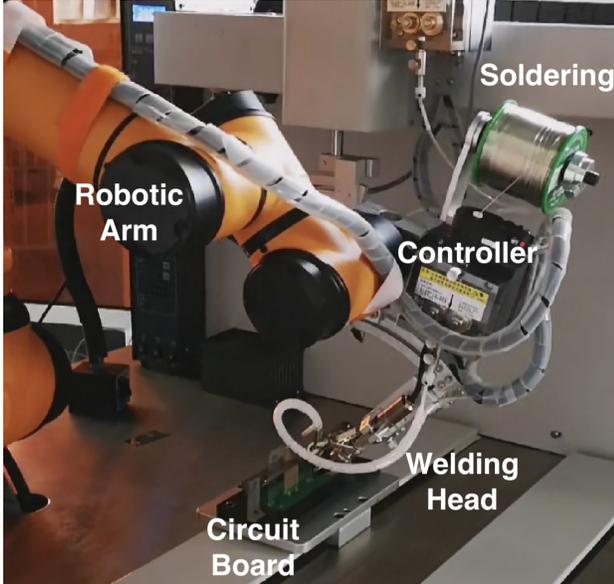


Fig. 6: The visual welding robotic platform.

TABLE II: Specifications of the robotic manipulator

| | |
|-------------------------------|---------------------|
| Product Name | AUBO i5+ |
| Degree of Freedom | 6 |
| Payload Mass | 5kg |
| Positioning Accuracy | $\pm 0.05\text{mm}$ |
| Maximum End-effector Velocity | 2.8m/s |
| Transmission Protocol | CAN Bus |
| Limits of Joints | $\pm 175^\circ$ |
| Maximum Velocity of Joints | 150°/s |
| Data Output Frequency | 100Hz |
| Operating System | Ubuntu Linux 16.04 |

B. Real Experiment

A visual welding robotic device is employed to conduct the in-run hand-eye calibration. The welding head is firmly



Fig. 7: Details of the employed visual welding head.

attached to a camera that gazes at the circuit board for visual welding (See Fig. 6). The specifications of the adopted robotic manipulator are listed in Table II. The camera estimates the pose relative to the circuit board by reconstructing the pose between the 3D point cloud from circuit 3D design file and the observed 2D feature points in the camera frame, using the efficient perspective-n-point (EPnP) method [28]. The camera captures the image at 30fps with the resolution of 1280×720 and the feature extraction is conducted by the binary robust invariant scalable keypoints (BRISK) [29] for fast computation speed. Fig. 7 shows another perspective of the relationship between the employed camera and welding head. The pose of the welding head is the end-effector and its pose relative to the base of the robotic arm has been given by accumulating transformations of each joint involved. The data streams of the camera and the welding head are transmitted separately via two wires and the correspondence matching and the time delay estimation are required. The experiment is conducted by generating welding trajectory via the welding software to accomplish an automatic welding task. The poses of the camera and robotic manipulator are shown in Fig. 8. The circuit board is firmly mounted on the table allowing for stable feature extraction. The globally optimal Go-ICP [17] is used for correspondence matching via (16). The higher dimensional ICP problem is divided into several sub-problems of 3D ICP using the Go-ICP. We set a window size of 200 to store the

TABLE I: Hand-eye Calibration Performances of the Visual Welding Task

| Window Size | BnB [7] (Inter-Camera Correspondence) | | 4DPA [4] (No Correspondence) | | Furrer et al. [10] (Correspondence/Constant Delay) | | Proposed (Correspondence/Varying Delay) | |
|-------------|--|--------------------------|---------------------------------|--------------------------|---|--------------------------|--|--|
| | Error | Time (s) | Error | Time (s) | Error | Time (s) | Error | Time (s) |
| 50 | 2.3066×10^{-02} | 7.8293×10^{-01} | 2.6870×10^{-02} | 6.5401×10^{-03} | 2.2352×10^{-02} | 9.8206×10^{-01} | 1.9297×10^{-02} | 3.5712×10^{-01} |
| 100 | 2.0146×10^{-02} | 9.2361×10^{-01} | 2.2909×10^{-02} | 7.1221×10^{-03} | 1.9754×10^{-02} | 2.8982×10^{-00} | 1.6358×10^{-02} | 4.7934×10^{-01} |
| 150 | 1.8876×10^{-02} | $1.0788 \times 10^{+00}$ | 1.9861×10^{-02} | 7.9034×10^{-03} | 1.7411×10^{-02} | 6.7310×10^{-00} | 1.5277×10^{-02} | 5.8120×10^{-01} |
| 200 | 1.6453×10^{-02} | $1.3654 \times 10^{+00}$ | 1.8673×10^{-02} | 8.5417×10^{-03} | 1.5288×10^{-02} | $1.2845 \times 10^{+01}$ | 1.2769×10^{-02} | 6.7905×10^{-01} |
| 250 | 1.4032×10^{-02} | $1.8871 \times 10^{+00}$ | 1.6634×10^{-02} | 9.4708×10^{-03} | 1.2960×10^{-02} | $1.9472 \times 10^{+01}$ | 9.9761×10^{-03} | 7.6633×10^{-01} |
| 300 | 1.0913×10^{-02} | $2.7074 \times 10^{+00}$ | 1.3965×10^{-02} | 1.0633×10^{-02} | 1.0012×10^{-02} | $2.3012 \times 10^{+01}$ | 8.7654×10^{-03} | 8.9145×10^{-01} |
| 350 | 9.1623×10^{-03} | $3.6853 \times 10^{+00}$ | 1.1603×10^{-02} | 1.2354×10^{-02} | 8.7203×10^{-03} | $3.1094 \times 10^{+01}$ | 7.4410×10^{-03} | 9.6781×10^{-01} |
| 400 | 7.7726×10^{-03} | $4.9337 \times 10^{+00}$ | 9.4578×10^{-03} | 1.6785×10^{-02} | 7.1138×10^{-03} | $3.6835 \times 10^{+01}$ | 6.8982×10^{-03} | $1.0466 \times 10^{+00}$ |

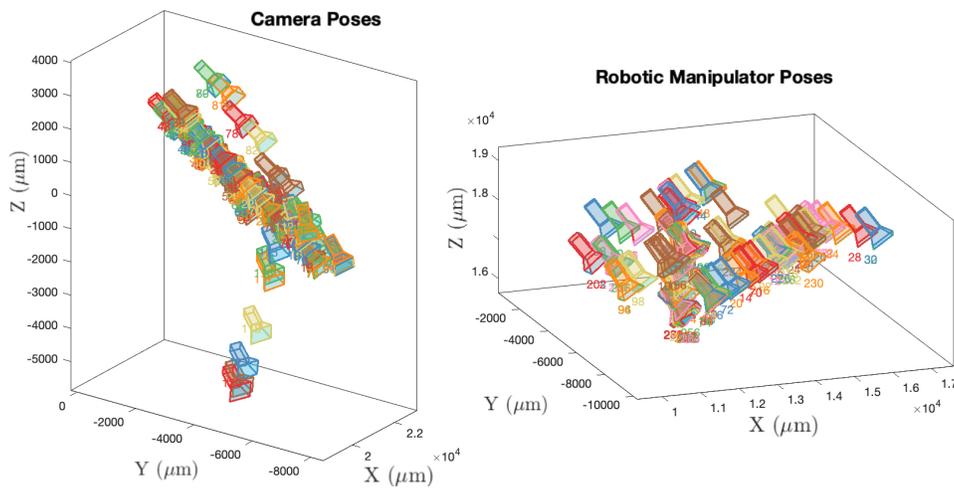


Fig. 8: The poses of camera and the robotic manipulator during the experiment.

most recent poses for real-time hand-eye calibration.

In the next experimental part, the BnB approach [7] and our recent 4DPA method [4] are compared with the proposed method. In the BnB, the correspondence matching is achieved iteratively with epipolar constraints. Different window sizes are set to evaluate the performances of various algorithms. All the methods are implemented using the C++ programming language for fairness. Also, all the algorithms are simplified to the best of our knowledge. A real-time Linux-based industrial computer has been used for data sampling which ensures the efficiency of data transmission. The gathered statistics are shown in Table I. The error is defined as the mean sum error in (1) where $\odot = SE(3)$. It is shown that using the proposed method, the hand-eye calibration error has been reduced. The proposed method can be considered as the improvement to the 4DPA method [4]. With growing window sizes, the measurement quantities increase as well, which results in better hand-eye calibration accuracy. However, with more measurements, the BnB will also need more constraints for the global solution. Thus the computation will cost much more time on the balancing between minimizing the objective and satisfying the conditions. Among all the compared algorithms, the 4DPA consumes the least computation time but obtains the worst calibration accuracy. Since 4DPA does not own correspondence matching, there are some mismatched correspondences. This problem is then fixed using the proposed method in this paper, which takes less time than the BnB to converge. As the $SO(4)$ parameterization in 4DPA has been shown to be efficient and accurate [4], the proposed method further improves the accuracy by removing the outliers. Due to the nature of the Lie algebra employed in the previous derivations, the proposed method can maintain the shape of the special orthogonal group $SO(n)$ and thus achieves smaller errors than BnB. The mean reprojection errors of all directions in the camera world frame (unit: mm) are computed using the results from various representatives are summarized in Table III. We can also observe that the reprojection error of the proposed method is less than the others, which verifies the effectiveness of the developed scheme.

TABLE III: The Reprojection Errors Derived from Calibration Results

| Algorithm | Reprojection Error (mm) |
|--------------------|-------------------------|
| BnB [7] | 0.36722 |
| 4DPA [4] | 0.58293 |
| Furrer et al. [10] | 0.28119 |
| Proposed | 0.085257 |

C. Discussion

From the simulation results, we can see that the correspondence and time delay are vital to the final accuracy of the hand-eye calibration. The proposed method is capable of estimating best correspondences and varying time delay in a simultaneous manner. The fitness of the derived solution from proposed method with ground truth in the simulation study shows that the correspondence matching is successful and the time delay is accurately estimated. In the experimental study of the hand-eye calibration of a visual welding robot, we show that the mismatched correspondences and time delay actually take place in the asynchronous data transmission. It is worthy mentioning that for any real scene, one can hardly give the ground truth of the hand-eye parameter \mathbf{X} . Therefore the only two fair approaches for comparison would be:

- 1) The loss function values of the summation of hand-eye residuals shown in (1).
- 2) The reprojection error in the camera world frame (expressed in mean distance error).

Note that these two indices are invoked for comparative evaluation of the experimental study while the proposed method owns satisfactory accuracy. It is also shown that due to extensive matching, the proposed method costs more time than ad-hoc correspondence-free ones like 4DPA [4].

IV. CONCLUSIONS

A novel correspondence matching and time delay estimation algorithm for the hand-eye calibration has been proposed. Throughout Lie algebra and symbolic computation, a new

point-cloud registration framework is obtained. By virtue of the derived registration, it is convenient for us to convert the original hard non-convex problem into an ICP-based problem, which can then be solved via globally optimal ICP. The experimentations have validated the superiority of the accuracy and computational performances of the proposed method against recent representatives.

ACKNOWLEDGEMENTS

This work was supported in part by the National Natural Science Foundation of China under Grant 41604025, in part by the General Research Fund of Research Grants Council, Hong Kong, under Grant 11210017, and in part by the Early Career Scheme Project of Research Grants Council, Hong Kong, under Grant 21202816. The work of M. Liu was supported by the Shenzhen Science, Technology and Innovation Commission (SZSTI) under Grant JCYJ20160401100022706.

REFERENCES

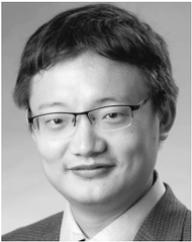
- [1] Y. C. Shiu and S. Ahmad, "Calibration of Wrist-Mounted Robotic Sensors by Solving Homogeneous Transform Equations of the Form $AX = XB$," *IEEE Trans. Robot. Autom.*, vol. 5, no. 1, pp. 16–29, 1989.
- [2] R. Y. Tsai and R. K. Lenz, "A New Technique for Fully Autonomous and Efficient 3D Robotics Hand/Eye Calibration," *IEEE Trans. Robot. Autom.*, vol. 5, no. 3, pp. 345–358, 1989.
- [3] X. Chen, C. Wang, W. Zhang, K. Lan, and Q. Huang, "An Integrated Two-Pose Calibration Method for Estimating Head-eye Parameters of a Robotic Bionic Eye," *IEEE Trans. Instrum. Meas.*, 2019. DOI: [10.1109/TIM.2019.2917235](https://doi.org/10.1109/TIM.2019.2917235)
- [4] J. Wu, Y. Sun, M. Wang, and M. Liu, "Hand-eye Calibration: 4D Procrustes Analysis Approach," *IEEE Trans. Instrum. Meas.*, 2019. DOI: [10.1109/TIM.2019.2930710](https://doi.org/10.1109/TIM.2019.2930710)
- [5] F. C. Park and B. J. Martin, "Robot Sensor Calibration: Solving $AX = XB$ on the Euclidean Group," *IEEE Trans. Robot. Autom.*, vol. 10, no. 5, pp. 717–721, 1994.
- [6] D. Condurache and A. Burlacu, "Orthogonal Dual Tensor Method for Solving the $AX = XB$ Sensor Calibration Problem," *Mech. Mach. Theory*, vol. 104, pp. 382–404, 2016.
- [7] J. Heller, M. Havlena, and T. Pajdla, "Globally Optimal Hand-Eye Calibration Using Branch-and-Bound," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 5, pp. 1027–1033, 2016.
- [8] Q. Ma, H. Li, and G. S. Chirikjian, "New Probabilistic Approaches to the $AX = XB$ Hand-eye Calibration without Correspondence," in *IEEE ICRA*, pp. 4365–4371, 2016.
- [9] H. Nguyen and Q.-C. Pham, "On the Covariance of X in $AX = XB$," *IEEE Trans. Robot.*, vol. 34, no. 6, pp. 1651–1658, 2018.
- [10] F. Furrer, M. Fehr, T. Novkovic, H. Sommer, I. Gilitschenski, and R. Siegwart, "Evaluation of Combined Time-Offset Estimation and Hand-Eye Calibration on Robotic Datasets," *FSR*, Springer, 2018.
- [11] 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, 2016.
- [12] Z. Zhao, "Simultaneous Robot-World and Hand-eye Calibration by the Alternative Linear Programming," *Pattern Recogn. Lett.*, vol. 127, pp. 174–180, 2019.
- [13] J. Wu, M. Liu, Z. Zhou, and R. Li, "Fast Symbolic 3-D Registration Solution," *IEEE Trans. Autom. Sci. Eng.*, 2019. DOI: [10.1109/TASE.2019.2942324](https://doi.org/10.1109/TASE.2019.2942324)
- [14] S. Ahmed, E. C. Kerrigan, and I. M. Jaimoukha, "A Semidefinite Relaxation-Based Algorithm for Robust Attitude Estimation," *IEEE Trans. Signal Process.*, vol. 60, no. 8, pp. 3942–3952, 2012.
- [15] J. Wu, M. Wang, H. Fourati, M. Liu, and H. Li, "Generalized n-Dimensional Registration: Theory and Application," 2020.
- [16] K. S. Arun, T. S. Huang, and S. D. Blostein, "Least-Squares Fitting of Two 3-D Point Sets," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-9, no. 5, pp. 698–700, 1987.
- [17] J. Yang, H. Li, D. Campbell, and Y. Jia, "Go-ICP: A Globally Optimal Solution to 3D ICP Point-Set Registration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 11, pp. 2241–2254, 2016.
- [18] J. Wu, Z. Zhou, H. Fourati, R. Li, and M. Liu, "Generalized Linear Quaternion Complementary Filter for Attitude Estimation From Multi-sensor Observations: An Optimization Approach," *IEEE Trans. Autom. Sci. Eng.*, vol. 16, no. 3, pp. 1330–1343, 2019.
- [19] H. Sommer, J. R. Forbes, R. Siegwart, and P. Furgale, "Continuous-Time Estimation of Attitude Using B-Splines on Lie Groups," *J. Guid. Control. Dyn.*, vol. 39, no. 2, pp. 242–261, 2016.
- [20] J. Kelly, N. Roy, and G. S. Sukhatme, "Determining the Time Delay between Inertial and Visual Sensor Measurements," *IEEE Trans. Robot.*, vol. 30, no. 6, pp. 1514–1523, 2014.
- [21] V. Krishnamurthy and M. Levoy, "Fitting Smooth Surfaces to Dense Polygon Meshes," in *ACM SIGGRAPH*, vol. 96, 1996, pp. 313–324.
- [22] F. Aghili, "A Prediction and Motion-Planning Scheme for Visually Guided Robotic Capturing of Free-floating Tumbling Objects with Uncertain Dynamics," *IEEE Trans. Robot.*, vol. 28, no. 3, pp. 634–649, 2012.
- [23] K. Kanatani, "Calibration of Ultrawide Fisheye Lens Cameras by Eigenvalue Minimization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 4, pp. 813–822, 2012.
- [24] K. Daniilidis, "Hand-Eye Calibration Using Dual Quaternions," *Int. J. Rob. Res.*, vol. 18, no. 3, pp. 286–298, 1999.
- [25] Z. Zhang, "Determining the Epipolar Geometry and Its Uncertainty: A Review," *Int. J. Comput. Vis.*, vol. 27, no. 2, pp. 161–195, 1998.
- [26] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up Robust Features (SURF)," *Comput. Vis. Image Understand.*, vol. 110, no. 3, pp. 346–359, 2008.
- [27] J. Wu, "Optimal Continuous Unit Quaternions from Rotation Matrices," *J. Guid. Control. Dyn.*, vol. 42, no. 4, pp. 919–922, 2019.
- [28] V. Lepetit, F. Moreno-Noguer, and P. Fua, "EPnP: An Accurate O(n) Solution to the PnP Problem," *Int. J. Comput. Vis.*, vol. 81, no. 2, pp. 155–166, 2009.
- [29] S. Leutenegger, M. Chli, and R. Siegwart, "BRISK: Binary Robust Invariant Scalable Keypoints," in *IEEE ICCV*, pp. 2548–2555, 2011.



Jin Wu (S'15-M'17) was born in May, 1994 in Zhenjiang, China. He received the B.S. Degree from University of Electronic Science and Technology of China, Chengdu, China. He has been a research assistant in Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology since 2018. His research interests include robot navigation, multi-sensor fusion, automatic control and mechatronics. He is a co-author of over 50 technical papers in representative journals and conference proceedings of IEEE, AIAA, IET

and etc. Mr. Jin Wu received the outstanding reviewer award for ASIAN JOURNAL OF CONTROL. One of his papers published in IEEE TRANSACTIONS ON AUTOMATION SCIENCE AND ENGINEERING was selected as the ESI Highly Cited Paper by ISI Web of Science during 2017 to 2018. He is a member of IEEE.

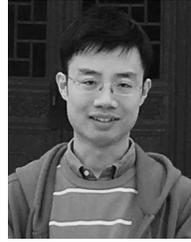
Mr. J. Wu started software programming using C++ since 2004. He has been in the UAV industry from 2012 and has launched two companies ever since. During 2019 to 2020, He was with Tencent Robotics X, Shenzhen, China, as a Localization Algorithm Engineer. Currently, he is a Robotics Engineer at Shenzhen Unity Drive, Shenzhen, China. He is the inventor of the world's first ultra low-cost circuit for 3-D and higher dimensional registration. He is currently in charge of some open projects of robotics and aerospace engineering from several State Key Laboratories in China.



Ming Liu (S'12-M'13-SM'18) received the B.A. degree in automation from Tongji University, Shanghai, China, in 2005, and the Ph.D. degree from the Department of Mechanical and Process Engineering, ETH Zürich, Zürich, Switzerland, in 2013, supervised by Prof. Roland Siegwart. During his master's study with Tongji University, he stayed one year with the Erlangen-Nürnberg University and Fraunhofer Institute IISB, Erlangen, Germany, as a Master Visiting Scholar.

He is currently with the Electronic and Computer Engineering, Computer Science and Engineering Department, Robotics Institute, The Hong Kong University of Science and Technology, Hong Kong, as an Assistant Professor. He is also a founding member of Shanghai Swing Automation Ltd., Co. He is coordinating and involved in NSF Projects and National 863-Hi-TechPlan Projects in China. His research interests include dynamic environment modeling, deep-learning for robotics, 3-D mapping, machine learning, and visual control.

Dr. Liu was a recipient of the European Micro Aerial Vehicle Competition (EMAV'09) (second place) and two awards from International Aerial Robot Competition (IARC'14) as a Team Member, the Best Student Paper Award as first author for MFI 2012 (IEEE International Conference on Multisensor Fusion and Information Integration), the Best Paper Award in Information for IEEE International Conference on Information and Automation (ICIA 2013) as first author, the Best Paper Award Finalists as co-author, the Best RoboCup Paper Award for IROS 2013 (IEEE/RSJ International Conference on Intelligent Robots and Systems), the Best Conference Paper Award for IEEE-CYBER 2015, the Best Student Paper Finalist for RCAR 2015 (IEEE International conference on Real-time Computing and Robotics), the Best Student Paper Finalist for ROBIO 2015, the Best Student Paper Award for IEEE-ICAR 2017, the Best Paper in Automation Award for IEEE-ICIA 2017, twice the innovation contest Chunhui Cup Winning Award in 2012 and 2013, and the Wu Wenjun AI Award in 2016. He was the Program Chair of IEEEERCAR 2016 and the Program Chair of International Robotics Conference in Foshan 2017. He was the Conference Chair of ICVS 2017. He has published many popular papers in top journals including IEEE TRANSACTIONS ON ROBOTICS (TRO) and INTERNATIONAL JOURNAL OF ROBOTICS RESEARCH (IJRR). Dr. Liu is currently an Associate Editor for IEEE ROBOTICS AND AUTOMATION LETTERS, IET CYBER-SYSTEMS AND ROBOTICS, INTERNATIONAL JOURNAL OF ROBOTICS AND AUTOMATION. He is a Senior Member of IEEE.



Chengxi Zhang was born in Feb, 1990, Shandong, China. He received his B.S. and M.S. degrees in microelectronics and solid-state electronics at Harbin Institute of Technology, China, in 2012 and 2015. He received his Ph.D. in control science and engineering at Shanghai Jiao Tong University, in 2019. Currently, he is a post-doc researcher at School of Electronic and Information Engineering, Harbin Institute of Technology, Shenzhen campus, China. His research interests are earth mapping plan design, embedded system software and hardware design, distribute information fusion, cyber-physical system control, event-based control.



Zebo Zhou was born in November, 1982 in Yongchuan, Chongqing, China. He received the B.Sc. and M.Sc. degrees in School of Geodesy and Geomatics from Wuhan University, Wuhan, China, in 2004 and 2006, respectively, and the Ph.D. degree from the College of Surveying and Geoinformatics, Tongji University, Shanghai, China in 2009. He was a visiting fellow with the Surveying & Geospatial Engineering Group, within the School of Civil & Environmental Engineering, University of New South Wales, Australia in 2009 and 2015. He is currently an associate professor with the School of Aeronautics and Astronautics, University of Electronic Science and Technology of China, Chengdu, China. His research interests include GNSS navigation and positioning, GNSS/INS integrated navigation, multi-sensor fusion.

Prof. Zhou has been in charge of projects of National Natural Science Foundation of China and has taken part in the National 863 High-tech Founding of China. He served as a Guest Editor of several special issues published on INTERNATIONAL JOURNAL OF DISTRIBUTED SENSOR NETWORKS and ASIAN JOURNAL OF CONTROL. He has been presenting related works on the annual conference of the institute of navigation (ION), the annual Chinese satellite navigation conference (CSNC) for several times and has received the Best Paper Awards in these conferences.