sensors

MDPI

Article

# Simultaneous Robot–World and Hand–Eye Calibration without a Calibration Object

**Wei Li [1], Mingli Dong [2], Naiguang Lu [1,2,*], Xiaoping Lou [2] and Peng Sun [1,2]**

[1] Institute of Information Photonics and Optical Communications, Beijing University of Posts and Telecommunications, Beijing 100876, China; liweikilary@bupt.edu.cn (W.L.); sunpeng@bistu.edu.cn (P.S.)

[2] Key Laboratory of the Ministry of Education for Optoelectronic Measurement Technology and Instrument, Beijing Information Science and Technology University, Beijing 100192, China; dongml@bistu.edu.cn (M.D.); louxiaoping@bistu.edu.cn (X.L.)

* Correspondence: nglv2002@sina.com; Tel.: +86-10-8242-6892

Received: 9 August 2018; Accepted: 5 November 2018; Published: 15 November 2018

**Abstract:** An extended robot–world and hand–eye calibration method is proposed in this paper to evaluate the transformation relationship between the camera and robot device. This approach could be performed for mobile or medical robotics applications, where precise, expensive, or unsterile calibration objects, or enough movement space, cannot be made available at the work site. Firstly, a mathematical model is established to formulate the robot-gripper-to-camera rigid transformation and robot-base-to-world rigid transformation using the Kronecker product. Subsequently, a sparse bundle adjustment is introduced for the optimization of robot–world and hand–eye calibration, as well as reconstruction results. Finally, a validation experiment including two kinds of real data sets is designed to demonstrate the effectiveness and accuracy of the proposed approach. The translation relative error of rigid transformation is less than 8/10,000 by a Denso robot in a movement range of 1.3 m × 1.3 m × 1.2 m. The distance measurement mean error after three-dimensional reconstruction is 0.13 mm.

**Keywords:** robot–world calibration; hand–eye calibration; calibration object; Kronecker product; sparse bundle adjustment

## 1. Introduction

With the progress of robot-vision-system advanced technology, it is necessary to evaluate the geometric relationships among the robot, sensors, and a reference frame. This problem is usually called "robot–sensor calibration", and it has been an active area of research for almost 40 years [1]. As research has progressed, the applications of robot–sensor calibration have extended into many domains, such as automobile assembly, robot navigation, and endoscopic surgery. As reported previously [2], the most widespread mathematical representations for the robot–sensor calibration problem can all be grouped into two categories: $AX = XB$ and $AX = ZB$.

The first class, and the most common robot–sensor calibration problem, is hand–eye calibration $AX = XB$, which was proposed by Tsai et al. [3] and Shiu et al. [4]. The earliest solution strategy estimated the rotation and translation with respect to homogeneous transformation $X$ separately [5,6]. However, it was found that such a method would produce rotation error spread in the process of the translation estimation. In later strategies, both the rotation and translation with respect to homogeneous transformation $X$ are solved simultaneously [7–9]. The above calibration methods solve the hand–eye relationship with different parametric approaches, such as the quaternion, dual quaternion, and Kronecker product, which are all inseparable from a known calibration object. However, there are many situations in which using an accurately-manufactured calibration object is not convenient, or is not possible at all. Indeed, due to restrictions in limited onboard weight or strictly sterile conditions, it may be inadvisable to use a calibration object in applications such as mobile robotics or endoscopy surgery. Thus, later, an approach for getting rid of the calibration object based on the structure from motion (SFM) technique was proposed by Andreff et al. [10], and this method—also named "extended hand–eye calibration"—could handle a wider range of problems. Subsequently, a similar approach was presented in [11], where a scale factor was included into quaternion and dual quaternion formulation. Ruland et al. [12] proposed a branch-and-bound parameter space search method for this extended hand–eye calibration problem, which guaranteed the global optimum of rotational and translational components with respect to a cost function based on reprojection errors. In [13,14], Heller et al. firstly utilized second order cone programming (SOCP) to calculate the hand–eye relationship and scale factor based on the angular reprojection error, and then exploited a branch-and-bound approach to minimize an objective function based on the epipolar constraint. However, this branch-and-bound search process was very time intensive. Soon afterwards, Zhi et al. [15] proposed an improved iterative approach to expedite the calculation speed concerning the above extended hand–eye calibration problem. Recently, with some consideration for the asynchrony of different sensors with respect to sampling rates and processing time by an online system, Li et al. [16] presented a probabilistic approach to solve the correspondence problem of data pairs ($A_i$, $B_i$). However, this method has not been tested with a real robotic system. Due to the narrow range of motion allowed by the surgical instrument in minimally-invasive surgery, Pachtrachai et al. [17] replaced planar calibration object with the CAD models of surgical tools in the process of hand–eye calibration; thus, the instrument 3D pose tracking problem has to be addressed in advance.

The second class of robot–sensor calibration problems is the form $AX = ZB$, which was first derived by Zhuang et al. [18]. This equation allowed the simultaneous estimation of the transformations from the robot-base coordinates to the world frame $Z$, and from the robot-gripper coordinate to the camera coordinate $X$. There are also two ways to approach the robot–world and hand–eye calibration problem. In the first, the rotation and translation components associated with $X$ and $Z$ are calculated separately based on dual quaternion and Kronecker product [19,20]. In the second, the rotation and translation components are computed simultaneously based on the quaternion and Kronecker product [21,22]. Additionally, in order to obtain a globally-optimal solution, Heller et al. [23] utilized convex linear matrix inequality (LMI) relaxations to simultaneously solve robot–world and hand–eye relationships. Very recently, in [24,25], Tabb et al. proposed a bundle adjustment-based approach, which is similar to the bundle adjustment partition of our algorithm. However, the main difference is that Tabb's approach works based on a chessboard target, which our approach does not need.

To the best of the authors' knowledge, all approaches for robot–world and hand–eye calibration are implemented with an external calibration object. However, it is necessary to further research the solving of the robot–world and hand–eye calibration problem without a calibration object, which is named "extended robot–world and hand–eye calibration" in this paper. Our work on this problem is motivated by two particular situations. The first is the use of a robot-mounted camera for multi-view, high-quality reconstruction of general objects by a rescue or endoscopic robot, where the reconstruction outcomes depend on the feature-matching accuracy instead of a 2D chessboard target. The second situation is the use of the same robot-mounted camera for large-scale digital photogrammetry under industrial conditions, such as aircraft and shipbuilding assembly sites. In this situation, in view of the limit of single measurement range, the measurement surface is segmented into several small parts. A robot with two linear guides is used to define and record the placement of the optical measurement system in front of the measurement surface. The imaging system, based on retroreflective targets (RRTs), is mounted on the robot gripper as an end effector, and non-experts can be allowed to complete the calibration and acquire the three-dimensional (3D) coordinates of the target points attached to the measurement surface from a remote location.

For these particular situations, an extended robot–world and hand–eye calibration approach without a calibration target is proposed for a robotic visual measurement system. At first, our approach improves the $AX = ZB$ mathematical model by supposing that different camera poses comprise up to an unknown scale factor, and propose a fast linear method to give an initial estimate to the calibration equation. Then, we combine space intersection and sparse bundle adjustment to refine the robot–world and hand–eye transformation relationship, as well as 3D reconstruction, simultaneously. Finally, we demonstrate the effectiveness, correctness, and reliability of our approach with relevant synthetic and real data experiments.

## 2. Problem Formulation

### 2.1. Initial Estimate

Supposing that we have an arbitrary position of the robotic system, from Figure 1, we can define:

$$AX = ZB \tag{1}$$

The homogeneous transformation matrix $A$ is obtained by calibrating extrinsic camera parameters with respect to a fixed calibration object. The homogeneous transformation matrix $B$ is computed using the internal-link forward kinematics of the robot arm. $X$ is the robot-gripper-to-camera rigid transformation, which is always constant, as the camera is rigidly mounted on the robot gripper, and $Z$ is the robot-base-to-world rigid transformation.

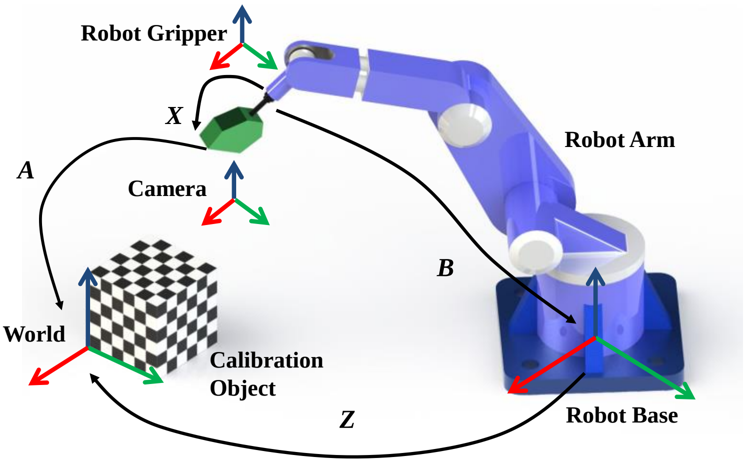


**Figure 1.** The robotic system of robot–world and hand–eye calibration.

Now, let $R_A$, $R_B$, $R_X$ and $R_Z \in SO(3)$ denote the respective 3 × 3 rotational matrices of $A$, $B$, $X$ and $Z$. Let $t_A$, $t_B$, $t_X$, and $t_Z$ denote the respective 3 × 1 translational vectors, which are measured using the same scale unit. Equation (1) can be easily decomposed into a rotational matrix equation and translational vector equation:

$$R_A R_X = R_z R_B, R_A t_X + t_A = R_z t_B + t_z \tag{2}$$

If there is no 3D calibration object, such as in 2D-to-3D correspondences, we have to use SFM to estimate camera poses based on 2D-to-2D correspondences only. However, due to the lack of a given scale factor, SFM can reconstruct the structure of the scene and the camera poses up to an unknown scale factor. Of course, we can introduce an explicit scaling factor to the robot–world and hand–eye calibration equation, with reference to Andreff [10]. Equation (2) can be transformed into

$$R_A R_X = R_z R_B, R_A t_X + \alpha t_A = R_z t_B + t_z \tag{3}$$

The Equations (3) can be used to formulate an objective function $f(\cdot)$ for non-linear optimization, which is based on the objective function for standard robot–world and hand–eye calibration proposed by [21]:

$$\begin{aligned} f(q_x, q_z, t_X, t_Z, \alpha) &= \lambda_1 \sum_{i=1}^{N} \left\| Q(q_{A_i}) q_x - W(q_z) q_{B_i} \right\|^2 + \lambda_2 \sum_{i=1}^{N} \left\| W(q_{A_i})^T Q(q_{A_i}) t_X + \alpha t_{A_i} - W(q_z)^T Q(q_z) t_{B_i} - t_Z \right\|^2 \\ &+ \lambda_3 (1 - q_x \overline{q}_x)^2 + \lambda_4 (1 - q_z \overline{q}_z)^2 \end{aligned} \tag{4}$$

where $W(q)^{\mathrm{T}}Q(q)$ is an orthogonal matrix for quaternion $q$, and the parameters $\lambda_1$ through $\lambda_4$ are regularization factors (e.g., $\lambda_1 = \lambda_2 = 1$ and $\lambda_3 = \lambda_4 = 10^6$). In addition to scale factor $\alpha$, the rotations and translations associated with $X$ and $Z$ can be estimated simultaneously by solving Equation (4).

$$W(q)^T Q(q) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & q_0^2 + q_1^2 - q_2^2 - q_3^2 & 2(q_1 q_2 - q_0 q_3) & 2(q_1 q_3 + q_0 q_2) \\ 0 & 2(q_1 q_2 + q_0 q_3) & q_0^2 - q_1^2 + q_2^2 - q_3^2 & 2(q_2 q_3 - q_0 q_1) \\ 0 & 2(q_1 q_3 - q_0 q_2) & 2(q_2 q_3 + q_0 q_1) & q_0^2 - q_1^2 - q_2^2 + q_3^2 \end{bmatrix}$$

Referring to [20], we can also obtain the separable solutions to the robot–world and hand–eye calibration problem by Kronecker product. Since $R_A$ and $R_B$ are both an orthogonal matrix, the orientation component of Equation (3) can also be represented as:

$$\begin{pmatrix} nI & -\sum_{j=1}^{n} R_{B_i} \otimes R_{A_i} \\ -\sum_{j=1}^{n} R_{B_i}^T \otimes R_{A_i}^T & nI \end{pmatrix} \begin{pmatrix} vec(R_Z) \\ vec(R_X) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \tag{5}$$

Those vectors of Equation (5) can efficiently be computed by singular value decomposition (SVD). The symbol $\otimes$ denotes the Kronecker product, and the column vector operator $vec$ reorders the coefficients of a $(m \times n)$ matrix $A$ into an $mn$ vector $vec(A) = (a_{11}, \ldots, a_{1n}, a_{21}, \ldots, a_{mn})$ [26]. Once $R_Z$ is calculated by Equation (5), $t_X$, $t_Z$ is the solution to the linear system:

$$\begin{bmatrix} R_A & -I_{3\times3} & t_A \end{bmatrix} \begin{bmatrix} t_X \\ t_Z \\ \alpha \end{bmatrix} = R_Z t_B \tag{6}$$

The solution to $t_X$, $t_Z$ and $\alpha$ can be easily determined by least square technique. However, the variety of the additional scale factor $\alpha$ will bring instability into Equation (3). To overcome this problem, we propose a novel solution through eliminating $\alpha$ based on the Kronecker product. We define $t_A{}^*$ as a skew-symmetric matrix corresponding to $t_A$, which can be denoted as

$$t_A{}^* = \begin{bmatrix} 0 & -t_3 & t_2 \\ t_3 & 0 & -t_1 \\ -t_2 & t_1 & 0 \end{bmatrix}$$

Since the scale factor $\alpha$ has no influence on the computation of rotation, the rotational part of Equation (3) is the same, and the translational part of Equation (3) is multiplied on both sides by the skew-symmetric $t_A{}^*$. Obviously, $t_A{}^*\, t_A = [0, 0, 0]^{\mathrm{T}}$, and the new equation can be formulated as follows:

$$R_A R_X R_B{}^T = R_z\,, \; t_A{}^* R_A t_X = t_A{}^* R_z t_B + t_A{}^* t_z \tag{7}$$

By using the Kronecker product theory, and if $AXB = C$ for an unknown matrix $X$, then the equation can be rewritten as a linear system:

$$vec(AXB) = \left(B^T \otimes A\right) vec(X) = vec(C)$$

Thus, Equation (7) can be reconstituted into

$$\begin{bmatrix} R_B \otimes R_A & -I_9 & 0_{9\times 3} & 0_{9\times 3} \\ 0_{3\times 9} & {t_B}^T \otimes {t_A}^* & -{t_A}^* R_A & {t_A}^* \end{bmatrix} \begin{bmatrix} vec(R_X) \\ vec(R_Z) \\ t_X \\ t_Z \end{bmatrix} = \begin{bmatrix} 0_{18\times 1} \\ 0_{6\times 1} \end{bmatrix} \tag{8}$$

Obviously, the solution of the linear system (8) can be solved by SVD, and since $R_X$ and $R_Z$ are rotational matrices, there is a proportionality constraint in that the $R_X$ and $R_Z$ have a determinant value of 1. Thus, the unique solution can be determined. Supposing that the solution of the linear system (8) is proportional to the right singular vector v corresponding to the minimum singular value, the resulting $R_X$ and $R_Z$ can be estimated as

$$R_X = \omega V_X, R_Z = \varphi V_Z \tag{9}$$

where $V_X = vec^{-1}(\mathrm{v}_{1:9})$, $V_Z = vec^{-1}(\mathrm{v}_{10:18})$, $vec^{-1}$ is defined as the inverse operator to $vec$, and the proportionality constants are

$$\omega = \mathrm{sign}(V_X)\det(V_X)^{-\frac{1}{3}}, \varphi = \mathrm{sign}(V_Z)\det(V_Z)^{-\frac{1}{3}}$$

Therefore, the calculated robot–world and hand–eye translation vectors are

$$t_X = \omega vec(\mathrm{v}_{19:21}), t_Z = \varphi vec(\mathrm{v}_{22:24}) \tag{10}$$

However, the calculated matrices $R_X$ and $R_Z$ may be not strictly orthogonal due to noise. Therefore, to ensure that they are indeed rotations, it is necessary to re-orthogonalize the computed rotation matrices.

### 2.2. Data Selection

SFM is a general method for obtaining camera poses from image correspondences, and mainly consists of feature point detection, feature point matching, camera pose calibration, and reconstruction. Given two view feature points that are coarse matching, there are a significant number of outliers in the estimated transformations of camera poses, and these outliers will inevitably affect the accuracy of the initial estimate for extended robot–world and hand–eye calibration. RANSAC [27] is a simple but robust algorithm for outlier removal, which has been used widely in computer vision. In this section, we utilize it to enhance robustness of the initial estimate. Referring to the RANSAC method [15], we randomly select a certain number of two view image correspondences and solve the extended robot–world and hand–eye calibration equation by the linear system (8). Firstly, as three pairs of camera pose solutions are just enough to determine the unique robot–world and hand–eye transformation [20], three pairs of camera orientation results are treated as the minimum number required for this sample. Then, we predict $\hat{A}_i$ using Equation (2):

$$R_{\hat{A}_i} = R_z R_{B_i} {R_X}^T, \quad t_{\hat{A}_i} = R_z t_{B_i} + t_z - R_{\hat{A}_i} t_X \tag{11}$$

So, the rotation error $e_R$ can be defined as follows:

$$e_R = \left\| R_{A_i} - R_{\hat{A}_i} \right\|_2 \tag{12}$$

Because the predicted translation $t_{\hat{A}_i}$ and original translation $t_{A_i}$ may not be calculated based on the same scale factor, the translation error $e_t$ is defined as follows:

$$e_t = \left| \left\| t_{\hat{A}_i} \right\|_2 - \frac{\left\langle t_{A_i}, t_{\hat{A}_i} \right\rangle}{\left\| t_{A_i} \right\|_2} \right| \tag{13}$$

In addition, we combine the rotation and translation errors as the total error. Considering that the translation unit is always set to millimeters, in order to balance the rotation and translation errors, we scale the translation error by 0.01, so the total error $e$ is

$$e = e_R + 0.01 * e_t \tag{14}$$

Finally, we calculate the total error $e$ for all valid random samples, and determine the largest set of consistent pairs. In this section, we let the error threshold $e$ be 0.01, and the maximum outlier ratio be 50%. It should be considered that this selection process is just an initial estimate. There is no need to spend substantial amounts of time for minor accuracy improvement, so we stop RANSAC when the maximum iteration limit reaches 100.

### *2.3. Sparse Bundle Adjustment*

Following the initial estimation for robot–world and hand–eye transformation by the Kronecker product, which is solved by Equations (9) and (10), we employ bundle adjustment to jointly refine the robot–world, hand–eye transformations, and the reconstruction results simultaneously. Bundle adjustment is almost invariably solved as the last step of feature-based 3D reconstruction algorithms and motion estimation computer vision algorithms to obtain optimal solutions. Generally speaking, the goal in bundle adjustment is to minimize the overall reprojection error between the observed and predicted image points. The mathematical expression can be depicted as below: assume that $m$ 3D points are seen in $n$ views, and let $x_{ij}$ indicate the projection of the $i$th point on the $j$th image. Assume also that $\lambda_{ij}$ is equal to 1 if the $i$th point can be observed on the $j$th image, otherwise it is equal to 0. Moreover, assume that $A_j$ is the rigid homogeneous transformation from the $j$th image frame to the world frame and that $G_i$ is the predicted 3D $i$th point by space intersection, and let $P_j(\cdot)$ be the predicted projection matrix of the $j$th image, including camera-intrinsic parameters. The bundle adjustment model minimizes the reprojection error with respect to all 3D points and camera parameters, specifically:

$$\min_{P_j, A_j, G_i} \sum_{i=1}^{m} \sum_{j=1}^{n} \lambda_{ij} \| x_{ij} - P_j(A_j^{-1} G_i) \|_2 \tag{15}$$

Problems that are substantially similar to problem (15) can typically be tackled with non-linear least-squares optimization routines such as the Levenberg–Marquardt or Gauss–Newton approaches. Conventional bundle adjustment methods solve the normal equations repeatedly with complexity $O(n^3)$ in the number of unknown parameters for each iteration. However, substantial time-saving can be achieved by taking advantage of the sparse block structure contained in the normal equation [28]. In this way, a software implementation of sparse bundle adjustment is proposed by Lourakis and Argyros [29].

In our experiment, we utilize their implementation to solve the extended robot–world and hand–eye calibration problem. In order to refine the initial guess of $X$ and $Z$ using sparse bundle adjustment, the homogeneous transformation $A_j(\alpha)$ up to an unknown scale factor is substituted by the inverse Equation (1) $A_j = ZB_jX^{-1}$, because the robot arm pose $B_j$, which is calibrated before delivery, can provide the real metric units. Then, the point 3D initial coordinates can be calculated by space intersection or triangulation. Finally, the sparse bundle adjustment method optimizes the robot–world transformation $Z$, hand–eye transformation $X$, and target point 3D coordinates $G_i$ simultaneously, while keeping the robot motions $B_j$ and camera-intrinsic parameters constant. Specifically, the sparse bundle adjustment model can be rewritten as:

$$\min_{X, Z, G_i} \sum_{i=1}^{m} \sum_{j=1}^{n} \lambda_{ij} \| x_{ij} - P_j(X B_j^{-1} Z_j^{-1} G_i) \|_2 \tag{16}$$

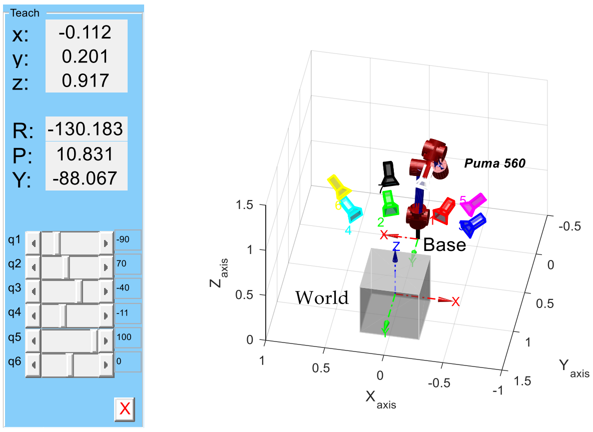
Note that the robot–world $Z$ and hand–eye $X$ transformations consist of 6 rotation parameters and 6 translation parameters, while each point consists of 3 position parameters. The total number of minimization parameters in Equation (16) equals $3m + 12$. According to specific needs, we can set a termination condition for iteration. The iterations are terminated when the estimated robot–world translation changes by less than $10^{-3}$ mm, or the reconstruction 3D points changes by less than $10^{-3}$ mm, compared to that of the last iteration, or reaches the maximum limit of iterations, which is ten in this paper.

## 3. Experiments

This section validates the proposed method for the extended robot–world and hand–eye calibration problem both on synthetic and real datasets. For the data comparison, with some considerations, one could not expect that the method without a calibration object would obtain results as accurate as the method with a calibration object. In this paper, our main purpose is that the estimation of the robot–world and hand–eye transformation is feasible without a calibration object. We refer to the means of data comparison of previous extended hand–eye calibration methods, such as those presented by Nicolas Andreff [10], Jochen Schmidt [11], and Jan Heller [13]. We present an experimental evaluation of the extended robot–world and hand–eye methods, in which the estimation of rotation, translation, and scale factor can be formulated using the Kronecker product [20], or quaternions [21], or reprojection error [25], and a standard robot–world and hand–eye calibration method [25] with chessboard pattern calibration was used as an approximate truth-value, since no ground truth is available to compare accuracy between different methods. For convenience, in the following experiments, the labels "Dornaika" and "Shah" stand for the estimation of rotation, translation, and scale factor using the quaternions Equation (4) or Kronecker product Equation (5), respectively. The label "KPherwc" stands for the proposed initial calibration method based on the Kronecker product Equation (8), and the label "BAherwc" stands for the proposed optimization approach based on sparse bundle adjustment Equation (16). VisualSFM [30]—a state-of-the-art, open-source SFM implementation—was used to obtain the camera poses for a general object in real-data experiments. All method results were obtained using a hybrid MATLAB 9.0 and C++ reference implementation, and we conducted the methods on an Intel Core i7-8750H processor running Linux.

### 3.1. Experiments with Synthetic Data

In order to simulate the actual process of robot motions, considering that PUMA560 is the most classic robot arm kinematics model, and that this robot has been well studied and its parameters are very well known—it has been described as the "white rat" of robotics research [31]—referring to Zhuang [18], we used PUMA560 robot kinematics modeling and a camera imaging model to build a synthetic scene and a virtual camera. As shown in Figure 2, a red PUMA560 robot arm was constantly in movement with a different-colored camera attached to the end gripper. A synthetic scene consisting of 50 3D points was generated randomly into a gray cube with side length 0.5 m, and 8 different virtual camera poses set such that the cameras were faced approximately to the center of the cube were created. The intrinsic parameters of the virtual camera and Denavit–Hartenberg (DH) parameters of the PUMA 560 robot are separately listed in Tables 1 and 2.



Teach

x: -0.112
y: 0.201
z: 0.917

R: -130.183
P: 10.831
Y: -88.067

q1 -90
q2 70
q3 -40
q4 -11
q5 100
q6 0

**Figure 2.** Schematic diagram of the synthetic experiment using the PUMA560 model.

**Table 1.** Intrinsic parameters of the virtual camera for the synthetic experiment.

| Intrinsic Parameter | Image Resolution | Focal Length | Principle Point Offsets | Affine Distortion | Radial Distortion and Decentering Distortion |
|---|---|---|---|---|---|
| Value | 4288 × 2848 pixels | 20 mm | (0.1, 0.1) mm | 0 | 0 |

**Table 2.** Denavit–Hartenberg parameters of the PUMA560 robot for the synthetic experiment.

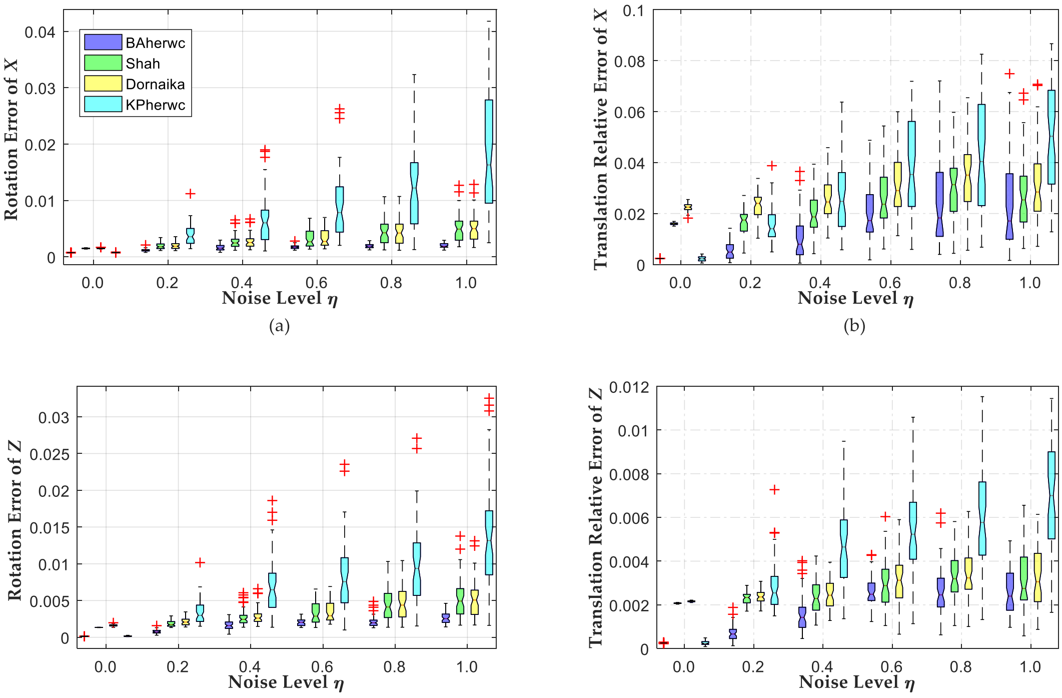
| Joint | $q_i$/(°) | $d_i$/m | $a_i$/m | $\alpha_i$/(°) | Offset/(°) |
|---|---|---|---|---|---|
| 1 | $q_1$ | 0 | 0 | 0 | $\sigma_1$ |
| 2 | $q_2$ | 0.2435 | 0 | −90 | $\sigma_2$ |
| 3 | $q_3$ | −0.0934 | 0.4318 | 0 | $\sigma_3$ |
| 4 | $q_4$ | 0.4331 | −0.0203 | 90 | $\sigma_4$ |
| 5 | $q_5$ | 0 | 0 | −90 | $\sigma_5$ |
| 6 | $q_6$ | 0 | 0 | 90 | $\sigma_6$ |

To test the performance of different methods against projection noise, the simulated data were conducted with the synthetic scene and a virtual camera. The scene 3D points were projected into the image plane after each position movement, but the projection points would be neglected if they were outside the image plane. In order to qualitatively analyze and evaluate the results of the synthetic experiment, we defined the error evaluation criteria associated with rotation and translation as follows:

$$e_R = \|\widetilde{R} - R\|_2 \quad e_t = \frac{\|\widetilde{t} - t\|_2}{\|\widetilde{t}\|_2}$$

where $\widetilde{R}$ represents the true rotation, $R$ represents the estimated rotation, $\widetilde{t}$ represents the true translation, and $t$ represents the estimated translation. In the synthetic experiment, since the nominal value for the robot–world and hand–eye transformation can be set up in advance, there is no need to use a standard robot–world and hand–eye calibration method [25] as an approximate truth-value. We set $\|\widetilde{t}_X\|_2 = 0.1$ m and $\|\widetilde{t}_Z\|_2 = 1$ m. The entire experiment is a four-step process. Firstly, considering that real-world feature point extraction is generally expected to have accuracy within 1 pixel, projection points in the synthetic experiment were corrupted by 6 different levels of Gaussian noise in the image domain with a standard deviation $\eta \in [0, 1]$ pixel and a step of 0.2 pixel. Secondly, according to the synthetic scene, we defined the nominal value for the hand–eye transformation $\widetilde{X}$ and the robot-to-world transformation $\widetilde{Z}$ with constant translation $\widetilde{t}_X$ and $\widetilde{t}_Z$. Thirdly, we calculated a sequence of camera positions based on space resection, and the corresponding robot motions were calculated by $B = Z^{-1}AX$. Considering that the noise of robot motion is determined after production, we added a constant noise ($\sigma$ = 0.025 mm) to robot joint offset. Finally, we performed the homogeneous transformations $X$ and $Z$ with the above four different methods and compared their rotation and translation errors in the presence of various noise levels. For each noise level, 50 repeated experiments were done with randomly generated sets of data, and the final value was the mean of all 50 errors.

Figure 3 illustrates the rotation and translation errors for each noise level using the boxplot. Clearly, our optimization method (“BAherwc”) exhibits the best behavior both in rotation and translation estimation of the transformation $X$ and $Z$, whereas the proposed initial calibration method (“KPherwc”) performs worst under noise conditions; thus, it is extremely effective to refine the initial calibration results by follow-up sparse bundle adjustment. Meanwhile, the translation relative errors estimated by “Shah” are slightly better than those estimated by “Dornaika”. This is a result of the “Dornaika” method calculating the rotation and translation transformations regarding $X$ and $Z$ all in the same step. Due to noise, the estimated rotations may not be accurate representations of the rotation matrices, and thus, a set of nonlinear constraints have to be made for the rotation matrices; meanwhile, the estimated translations are not updated with the orthogonal restriction, which causes the larger positional errors that are illustrated in Figure 3.



(a)

(b)

(c)

(d)

**Figure 3.** Error of estimated rotation and translation against different noise levels $\eta$: (**a**,**b**) The rotation and translation errors with regard to hand–eye transformation $X$; (**c**,**d**) The rotation and translation errors with regard to robot–world transformation $Z$.

### 3.2. Experiments with Real Datasets

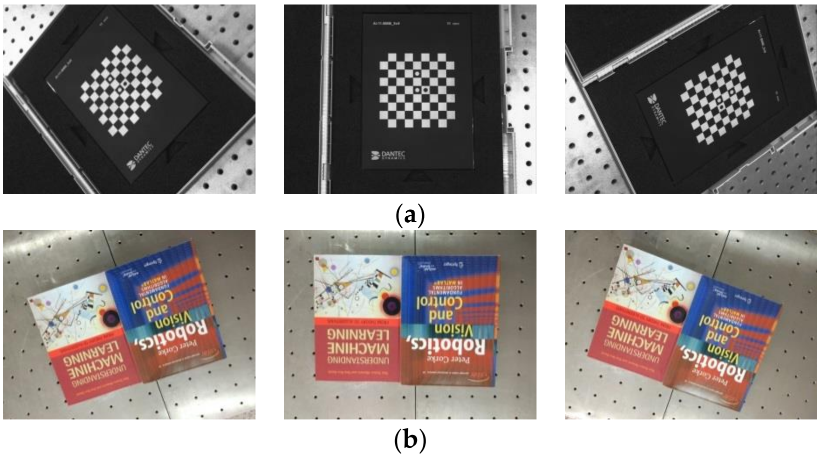
In this experiment, a Denso VS-6577GM serial 6-DOF robot arm with a Nikon D300s digital SLR camera and an AF NIKKOR 20 mm lens was used to acquire the real data. Since no ground truth gripper–camera transformation is available in the real data, it is difficult to give direct error results about the computed robot–world and hand–eye transformation, such as for the synthetic data. Therefore, it is desirable to measure the quality of the calibration results between the camera and robot device in some indirect way. In the rest of this section, we arranged two different scenes to complete the accuracy assessment: scene A, with a general object, was used to show the general applicability of the proposed method compared to the standard robot–world and hand–eye calibration approaches, and scene B, a photogrammetric retro-reflective target was used to improve the feature point-locating accuracy and decrease the false match rate for calibrating the extrinsic camera. Before the experiment, we used [32] to calibrate the camera together with seven parameters of lens, so the images were undistorted prior to being further used in order to improve sparse bundle adjustment results.

#### 3.2.1. Dataset A

With dataset A, our main purpose is not to prove how high the accuracy of our method is, but to demonstrate the feasibility of estimating the robot–world and hand–eye transformation without a calibration object in a general scene. Two image sets were required for the performance of the different methods in real-world conditions, as shown in Figure 4. Some consideration for the absence of a ground truth is available in the real-data experiment. We cannot give errors between the real robot–world transformation and the computed one, just like in the synthetic experiment. Since the method with a calibration object can usually obtain more accurate results than the method without a calibration object, a chessboard pattern was firstly used for solving robot–world $Y_{\text{bar}}$ and hand–eye $X_{\text{bar}}$ transformation simultaneously by the Tabb method [25], which could be assumed to give an approximate true value for the present. Afterwards, we removed the chessboard pattern, and used books as the object instead. We used the above "Dornaika", "Shah", "KPherwc", and "BAherwc" methods to calculate the homogeneous transformation $X_{\text{scene}}$ and $Y_{\text{scene}}$ with the general object of books. Finally, we compared their results to the approximate true value $X_{\text{bar}}$ and $Y_{\text{bar.}}$ The errors of robot–world and hand–eye relationships are defined as follows:

$$E_X = \|X_{\text{bar}} - X_{\text{scene}}\|_2 \quad E_Y = \|Y_{\text{bar}} - Y_{\text{scene}}\|_2$$

Figure 5a shows that the robot gripper carrying the camera took a series of photos around the center of the books. The positions of the gripper were adjusted with ten different locations, and it was ensured that the entirety of the books were in the view in every frame. The camera was set to manual mode, and images of 4288 × 2848 pixels were taken using a PC remote control. After all the photos were taken, a fast open-source SFM implementation was used to obtain the camera pose $A_i(\alpha)$, and the robot motion transformation $B_i$ was obtained from the known gripper-to-robot-base transformations. Then we computed robot–world and hand–eye transformation using the above four methods. Figure 5b shows the resulting 3D model output from bundle adjustment, containing 49,352 points in space, and the poses of all the ten cameras. Due to a high number of correspondences, only every hundredth member of the set of corresponding image points was used in our experiment.


(**a**)

(**b**)

**Figure 4.** Sample images of calibration scenarios taken by the camera mounted on the gripper of the robot: (**a**) Chessboard pattern scene; (**b**) Books scene.

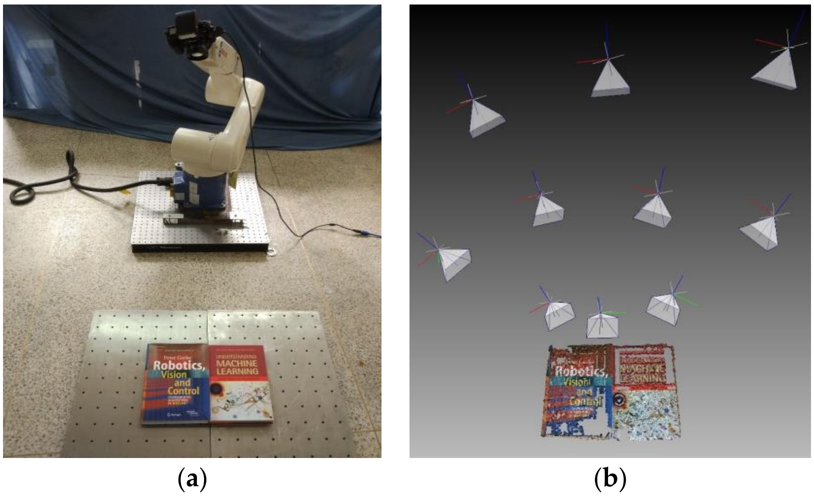


**Figure 5.** General object data set experiment: (**a**) Denso robot arm with Nikon camera; (**b**) 3D model output after bundle adjustment.

Table 3 summarize the results obtained with the two image sets mentioned above. Compared with other similar methods, it can be seen that our "BAherwc" method is nearest to the results of Tabb method [25] based on the chessboard pattern calibration. This is because in the "BAherwc" method, it was initialized by the results from the "KPherwc" method; then, the reprojection error is directly minimized, like in Tabb reprojection [25]. On the other hand, in the "Dornaika" and "Shah" method, the variety of the scale factor will bring instability into the solution of the robot–world and hand–eye transformation during the SFM implementation. Of course, one could not expect to obtain results as accurate as with Tabb's standard calibration. However, depending on the different application, the advantages of the proposed extended method may outweigh this drawback. It is especially true for mobile robotics or endoscopy setups that we have in mind, where robot–world and hand–eye calibration has to be performed under specific situations, due to the restrictions in limited onboard weight or the strict sanitary conditions. In order to achieve a rough qualitative analysis, we also measured the translation from the gripper to the camera lens center by hand with the known mechanical structure of the gripper and join parts, which is approximated to [0, 58, 66] mm. The estimated translation by our "BAherwc" approach is [0.183, 57.326, 64.910] mm, which is close to the result of the previous physical measurement, showing the validity of the obtained results.

**Table 3.** Error comparison for the general object data set without a chessboard pattern as benchmark (Unit: mm).

| Approach | Dornaika | Shah | KPherwc | BAherwc |
|---|---|---|---|---|
| Hand–eye transformation error $E_X$ | 3.945 | 2.337 | 3.409 | 1.145 |
| robot–world transformation error $E_Y$ | 6.001 | 3.751 | 4.544 | 1.808 |

### 3.2.2. Dataset B

In dataset B, our main purpose is to provide a mobile benchmark for large-scale digital photogrammetry under industrial conditions, which needs a robot to move along the guide rail to complete multi-station stitching measures. In order to reduce noise disturbance caused by SFM, we used photogrammetric retro-reflective targets (RRTs) to obtain accurate feature point matching. RRTs consist of a dense arrangement of small glass beads (Figure 6a bottom-right), as the name would suggest, which have good retro-reflective performance. The reflected light intensity in the light source direction is up to hundreds of times larger than the general diffuse reflection target. Thus, it is easy to obtain a subpixel level locating accuracy of feature points in the complex background image. As indicated in Figure 6a, dozens of RRTs and two yellow invar alloy scale bars $S_1$ and $S_2$

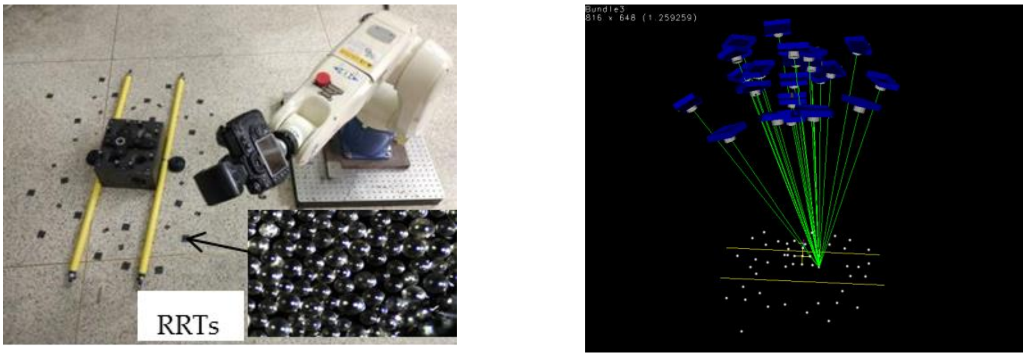
constructed a photogrammetric control field, and two 6 mm diameter coded RRTs were rigidly fixed on the yellow scale bar end, for which the distance had been accurately measured by a laser interferometer. Furthermore, coded RRTs can be encoded using specific pattern of distribution, which can actualize the automatic image matching of corresponding points. Then, the robot gripper carrying the camera took a series of photos around the center of the photogrammetric field, ensuring that the entire RRTs were in the camera view in every frame. Afterwards, we used the Hartley 5-point minimal relative pose method [33] and photogrammetry bundle adjustment to calibrate and optimize the extrinsic camera parameters.

Figure 6b shows the distribution of camera pose and RRTs. After bundle adjustment, the cameras were moved to 20 different poses faced to the RRTs and scale bars, and a Denso robot was moved in volume of 1.3 m × 1.3 m × 1.2 m. In view of photogrammetric relative orientation yields a high precision camera poses $A_i(i = 1, \ldots, 20)$, we solved hand–eye transformation $X$ and the robot-to-world transformation $Z$ by means of three methods (the "Dornaika", "Shah", and our "BAherwc" method) based on existing camera pose $A_i$ and robot motion $B_i$. Then, the predictive camera poses $\hat{A}_i$ can be inverse-computed with Equation (1):

$$\hat{A}_i = ZBX^{-1}$$

In this section, the discrepancy between $\hat{A}_i$ and $A_i$ is supposed to an accuracy assessment basis of robot–world and hand–eye calibration. Considering the difference of scale factor between $\hat{t}_{A_i}$ and $t_{A_i}$, all translations are normalized beforehand, and the mean errors of all motions (from 1 to 20) are computed in rotation and translation. The rotation and translation relative errors are described as:

$$e_R = \|R_{A_i} - \hat{R}_{A_i}\|_2 \quad e_t = \left\| \frac{t_{A_i}}{\|t_{A_i}\|_2} - \frac{\hat{t}_{A_i}}{\|\hat{t}_{A_i}\|_2} \right\|_2$$



(**a**) Photogrammetric control field (**b**) Distribution of camera poses and RRTs

**Figure 6.** Photogrammetric scene data set experiment: (**a**) Photogrammetric control field; (**b**) Distribution of camera pose and target points.

Comparisons of the accuracy in rotation and translation for photogrammetric scene data set are provided in Table 4. It can be seen that our method "BAherwc" is almost to half an order of magnitude better than the other methods with regard to both in rotation and translation estimation. The rotation error is less than 5/10,000, and translation relative error is less than 8/10,000. Our optimization method "BAherwc" outperforms the "Dornaika" and "Shah" methods, mainly because the feature extraction and matching accuracy of retroreflective targets is significantly higher than that of the general targets used by SFM; this is an expected behavior, as the "BAherwc" method, by minimizing overall reprojection error, depends on the feature extraction accuracy. This experiment might have confirmed the validity of our approach for the calibration of transformation parameters between the camera and the robot device based on the digital photogrammetric system.

**Table 4.** Error comparison in rotation and translation for the photogrammetric scene dataset.

| Approach | Rotation Error $e_R$ | Translation Error $e_t$ |
|---|---|---|
| Dornaika | 0.0023 | 0.0033 |
| Shah | 0.0015 | 0.0017 |
| BAherwc | 0.00047 | 0.00076 |

Since the two invar alloy scale bars, $S_1$ and $S_2$, provided both feature correspondences and a distance measurement reference, we evaluated our "BAherwc" approach in the relative accuracy of 3D reconstruction by using distance measurement. The average distance measurement errors of the two scale bars are given in the following Table 5. For comparison, $\hat{S}_1$ and $\hat{S}_2$ are defined as calculated values based on our "BAherwc" method as reconstruction result byproduct. The distance measurement errors are described as:

$$e_s = \left| S_i - \hat{S}_i \right|, i = 1, 2$$

**Table 5.** The average distance measurement errors of scale bars (Unit: mm)

| Scale Bar | Nominal Value | Measurement Value | Distance Measurement Error $e_s$ |
|---|---|---|---|
| $S_1$ | 1096.037 | 1095.906 | 0.131 |
| $S_2$ | 1096.057 | 1095.923 | 0.134 |

Finally, to show the iterative process of our bundle adjustment method, Figure 7 illustrates the distance estimation error variances of the scale bars $S_1$ and $S_2$ at each iteration. One can see that although the initial reconstruction results are clearly inaccurate, the reconstruction errors after finite iteration still converge, and the final differences between the nominal value and measurement value of the two scale bars are close to 0.1 mm. Given that offline photogrammetry systems offer the highest precision and accuracy levels, the precision of feature point measurement can be as high as 1/50 of a pixel, yielding typical measurement precision on the object in the range of 1:100,000 to 1:200,000 [34], with the former corresponding to 0.01 mm for an object of 1 m in size. The absolute accuracy of length measurements is generally 2–3 times less (e.g., about 0.025 mm for a 1-m long object) than the precision of object point coordinates, which expresses the relative accuracy of 3D shape reconstruction. The relative accuracy of reconstruction by our bundle adjustment method is also influenced by the robot arm, and it can be improved by follow-up photogrammetric network design.

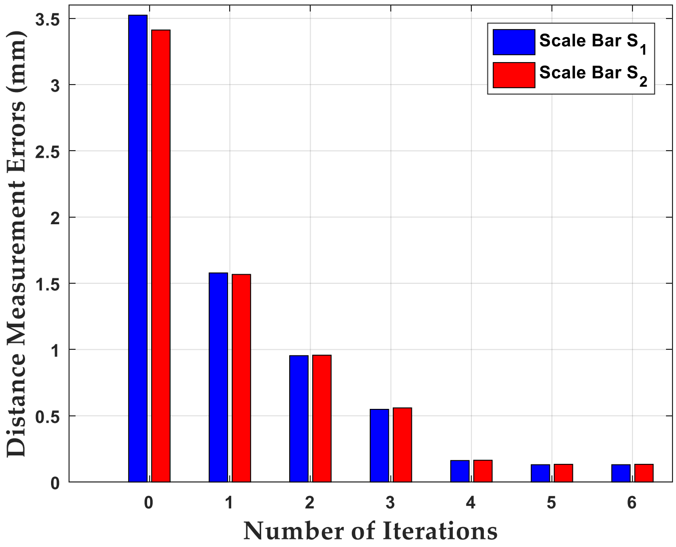


**Figure 7.** Distance estimation error iteration at each iteration by bundle adjustment.

## 4. Conclusions

In this paper, we present an extended approach for robot–world and hand–eye calibration without the need for a calibration object. In order to obtain the calibration data, we use two kinds of extrinsic calibrations: one for computer vision system with SFM, and the other for the digital photogrammetric system with RRTs. These two calibration methods can both estimate the extrinsic camera parameters lacking a known scales factor. Meanwhile, the robot end gripper pose is computed using the manipulator's forward kinematics, whose parameters are generally supposed to be known. Then, we use a fast initial estimation for extended robot–world and hand–eye calibration based on the Kronecker product. After the initial guess, to further improve the calibration results, we used sparse bundle adjustment to optimize the robot–world and hand–eye transformation relationship along with reconstruction. Finally, to evaluate and verify the feasibility of the proposed method, four accuracy assessment solutions were designed in the synthetic-data and real-data experiments. It is shown that our "BAherwc" approach can maintain a certain accuracy and robustness without a calibration object under the lower noise disturbance, and the Denso VS-6577GM, rigidly mounted to the floor, can obtain relatively reliable reconstruction results for follow-up photogrammetry stitching measures. In the future, we will move the industrial robot along the guide rail to expand the measurement range of the calibration procedures.

**Author Contributions:** Conceptualization, W.L. and P.S.; Methodology, W.L.; Software, P.S.; Validation, N.L. and P.S.; Formal Analysis, W.L.; Investigation, W.L.; Resources, M.D.; Data Curation, W.L.; Writing—Original Draft Preparation, W.L.; Writing—Review & Editing, W.L. and P.S.; Visualization, W.L.; Supervision, N.L. and M.D.; Project Administration, X.L.; Funding Acquisition, M.D. and X.L.

**Funding:** This research was funded by National Natural Science Foundation of China (No. 51475046, No. 51475047).

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Du, H.; Chen, X.; Xi, J. Development and Verification of a Novel Robot-Integrated Fringe Projection 3D Scanning System for Large-Scale Metrology. *Sensors* **2017**, *17*, 2886. [CrossRef] [PubMed]
2. Shah, M.; Eastman, R.D.; Hong, T. An overview of robot–sensor calibration methods for evaluation of perception systems. In Proceedings of the Workshop on Performance Metrics for Intelligent Systems, College Park, MD, USA, 20–22 March 2012; pp. 15–20.
3. Tsai, R.Y.; Lenz, R.K. A new technique for fully autonomous and efficient 3D robotics hand/eye calibration. *IEEE Trans. Robot. Autom.* **1989**, *5*, 345–358. [CrossRef]
4. Shiu, Y.C.; Ahmad, S. Calibration of wrist-mounted robotic sensors by solving homogeneous transform equations of the form AX = XB. *IEEE Trans. Robot. Autom.* **1989**, *5*, 16–29. [CrossRef]
5. Park, F.C.; Martin, B.J. Robot sensor calibration: Solving AX = XB on the Euclidean group. *IEEE Trans. Robot. Autom.* **1994**, *10*, 717–721. [CrossRef]
6. Andreff, N.; Horaud, R.; Espiau, B. On-line hand–eye calibration. In Proceedings of the Second International Conference on 3-D Digital Imaging and Modeling, Ottawa, ON, Canada, 8 October 1999; pp. 430–436.
7. Horaud, R.; Dornaika, F. Hand–eye calibration. *Int. J. Robot. Res.* **1995**, *14*, 195–210. [CrossRef]
8. Daniilidis, K. Hand–eye calibration using dual quaternions. *Int. J. Robot. Res.* **1999**, *18*, 286–298. [CrossRef]
9. Zhao, Z. Hand–eye calibration using convex optimization. In Proceedings of the 2011 IEEE International Conference on Robotics and Automation (ICRA), Shanghai, China, 9–13 May 2011; pp. 2947–2952.
10. Andreff, N.; Horaud, R.; Espiau, B. Robot hand–eye calibration using structure-from-motion. *Int. J. Robot. Res.* **2001**, *20*, 228–248. [CrossRef]
11. Schmidt, J.; Vogt, F.; Niemann, H. Calibration-free hand-eye calibration: A structure-from-motion approach. In *DAGM.*; Springer: Berlin/Heidelberg, Germany, 2005; pp. 67–74.
12. Ruland, T.; Pajdla, T.; Krüger, L. Globally optimal hand–eye calibration. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Providence, RI, USA, 16–21 June 2012; pp. 1035–1042.

13. Heller, J.; Havlena, M.; Sugimoto, A.; Pajdla, T. Structure-from-motion based hand–eye calibration using L∞ minimization. In Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Colorado Springs, CO, USA, 20–25 June 2011; pp. 3497–3503.
14. Heller, J.; Havlena, M.; Pajdla, T. Globally optimal hand–eye calibration using branch-and-bound. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *38*, 1027–1033. [CrossRef] [PubMed]
15. Zhi, X.; Schwertfeger, S. Simultaneous hand–eye calibration and reconstruction. In Proceedings of the 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Vancouver, BC, Canada, 24–28 September 2017; pp. 1470–1477.
16. Li, H.; Ma, Q.; Wang, T.; Chirikjian, G.S. Simultaneous hand–eye and robot–world calibration by solving the $ AX = YB $ problem without correspondence. *IEEE Robot. Autom. Lett.* **2016**, *1*, 145–152. [CrossRef]
17. Pachtrachai, K.; Allan, M.; Pawar, V. Hand–eye calibration for robotic assisted minimally invasive surgery without a calibration object. In Proceedings of the 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Daejeon, Korea, 9–14 October 2016; pp. 2485–2491.
18. Zhuang, H.; Roth, Z.S.; Sudhakar, R. Simultaneous robot/world and tool/flange calibration by solving homogeneous transformation equations of the form AX = YB. *IEEE Trans. Robot. Autom.* **1994**, *10*, 549–554. [CrossRef]
19. Hirsh, R.L.; DeSouza, G.N.; Kak, A.C. An iterative approach to the hand–eye and base-world calibration problem. In Proceedings of the 2001 IEEE International Conference on Robotics and Automation (ICRA), Seoul, Korea, 21–26 May 2001; pp. 2171–2176.
20. Shah, M. Solving the robot–world/hand–eye calibration problem using the Kronecker product. *J. Mech. Robot.* **2013**, *5*, 031007–1–031007-7. [CrossRef]
21. Dornaika, F.; Horaud, R. Simultaneous robot–world and hand–eye calibration. *IEEE Trans. Robot. Autom.* **1998**, *14*, 617–622. [CrossRef]
22. Li, A.; Wang, L.; Wu, D. Simultaneous robot–world and hand–eye calibration using dual-quaternions and Kronecker product. *Int. J. Phys. Sci.* **2010**, *5*, 1530–1536.
23. Heller, J.; Henrion, D.; Pajdla, T. Hand–eye and robot–world calibration by global polynomial optimization. In Proceedings of the 2014 IEEE International Conference on Robotics and Automation (ICRA), Hong Kong, China, 31 May 2014; pp. 3157–3164.
24. Tabb, A.; Khalil, M.; Yousef, A. Parameterizations for reducing camera reprojection error for robot–world hand–eye calibration. In Proceedings of the 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Hamburg, Germany, 28 September–2 October 2015; pp. 3030–3037.
25. Tabb, A.; Khalil, M. Solving the robot–world hand–eye (s) calibration problem with iterative methods. *Mach. Vis. Appl.* **2017**, *28*, 569–590. [CrossRef]
26. Neudecker, H. A note on Kronecker matrix products and matrix equation systems. *SIAM J. Appl. Math.* **1969**, *17*, 603–606. [CrossRef]
27. Fischler, M.A.; Bolles, R.C. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM* **1981**, *24*, 381–395. [CrossRef]
28. Zhang, L.; Koch, R. Structure and motion from line correspondences: Representation, projection, initialization and sparse bundle adjustment. *J. Vis. Commun. Image Represent.* **2014**, *25*, 904–915. [CrossRef]
29. Lourakis, M.A.; Argyros, A. SBA: A software package for generic sparse bundle adjustment. *ACM Trans. Math. Softw.* **2009**, *36*, 1–30. [CrossRef]
30. Wu, C. Towards linear-time incremental structure from motion. In Proceedings of the 2013 IEEE International Conference on 3D Vision (3DV), Seattle, WA, USA, 29 June–1 July 2013; pp. 127–134.
31. Corke, P. *Robotics, Vision and Control.: Fundamental Algorithms in MATLAB*®, 2nd ed.; Springer: Heidelberg, Germany, 2017; pp. 133–170. ISBN 978-3-642-20143-1.
32. Sun, P.; Lu, N.G.; Dong, M.L. Modelling and calibration of depth-dependent distortion for large depth visual measurement cameras. *Opt. Express* **2017**, *25*, 9834–9847. [CrossRef] [PubMed]

33. Hartley, R.; Li, H. An efficient hidden variable approach to minimal-case camera motion estimation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *34*, 2303–2314. [CrossRef] [PubMed]
34. Luhmann, T. Close range photogrammetry for industrial applications. *ISPRS J. Photogramm. Remote Sens.* **2010**, *65*, 558–569. [CrossRef]