Journal of Network Intelligence
Taiwan Ubiquitous Information
 ISSN 2414-8105 (Online)
Volume **10**, Number **2**, May **2025**

An Enhanced Robust Hand-Eye Calibration Model Adapted to Noise Information

Qinyue Tong

School of Aeronautics and Astronautics
Zhejiang University, Hangzhou 310027, China
22224030@zju.edu.cn

Yangming Zheng*

School of Aeronautics and Astronautics
Zhejiang University, Hangzhou 310027, China
zymsun2002@zju.edu.cn

Hao Luo

School of Aeronautics and Astronautics
Zhejiang University, Hangzhou 310027, China
luohao@zju.edu.cn

Zhe-Ming Lu

School of Aeronautics and Astronautics
Zhejiang University, Hangzhou 310027, China
zheminglu@zju.edu.cn

*Corresponding author: Yangming Zheng

Received March 30, 2024; revised August 30, 2024; accepted December 1, 2024.

Abstract. *The hand-eye calibration refers to the determination of the transformation relationship between a robot's hand and a camera. Previous research has predominantly addressed the hand-eye problem from a mathematical standpoint, yielding hand-eye matrices that exhibit satisfactory precision in ideal numerical simulations but often fail to adapt to the environmental noise encountered in practical applications. To attain a highly robust and precise hand-eye matrix, we propose an enhanced robust hand-eye calibration model adapted to noise information (ERHEC-NI). ERHEC-NI firstly obtains an initial estimate of the hand-eye matrix by the Tsai method. Then it fine-tunes and optimizes the initial value by incorporating the noise information from a small number of sample points in the environment, thus enhancing the result's robustness. We evaluate the performance of ERHEC-NI through numerical simulations and experimental evaluations. The results demonstrate its superiority over existing relevant approaches.*
Keywords: Robot, Hand-eye calibration, Noise information, Tsai

1. **Introduction.** In recent years, the rapid advancements in robotics technology and computer vision have propelled Visual Guided Robotics (VGR) [1] to become an integral component of artificial intelligence and global industrial development [2]. With visual sensors, VGR systems are capable of perceiving and comprehending their surrounding environments, enabling autonomous operation, intelligent manufacturing, quality control, human-robot collaboration, and augmented reality [3] applications. In the field of artificial intelligence, the visual capabilities of VGR systems are paramount for achieving

autonomous intelligence and facilitating interaction with the environment. However, in order to enable robots to accurately perceive and comprehend objects and scenes in their surroundings, it is imperative to establish a precise transformation relationship between the robot's hand and the camera. This relationship facilitates crucial tasks such as visual guidance, object grasping, and target localization. The process of determining the precise transformation relationship between the robot's end-effector and the camera is commonly referred to as hand-eye calibration [4].

In the realm of industrial development, hand-eye calibration plays a pivotal role in intelligent manufacturing and quality control within the domain of computer vision. By establishing a precise calibration of the relationship between a robot's hand and the camera, it becomes feasible to achieve streamlined production processes and accurate object grasping, thereby enhancing both production efficiency and product quality. Moreover, hand-eye calibration ensures the safety and coordination between robots and operators in human-robot collaborative scenarios, facilitating the joint completion of intricate tasks. Within the domains of augmented reality and virtual reality, hand-eye calibration constitutes a critical technological underpinning, providing indispensable support for immersive virtual experiences and the realization of augmented reality effects. In recent years, the application of hand-eye calibration techniques in the field of medical instruments has also gained an increasing development [5, 6, 7, 8].

However, accurately estimating the precise transformation relationship between a robot's end-effector and a camera presents significant challenges for engineers and researchers. Shiu and Ahmad [9] were the pioneers who mathematically formulated the hand-eye calibration problem as a mathematical solvability issue represented by the equation $AX = XB$ (where X represents the hand-eye matrix). They employed a linear closed-loop solution to address this equation. Subsequently, a host of methods have been proposed by scholars to solve the hand-eye calibration equation [10]. Tsai et al. [11] introduced an approach based on the concept of axis-angle transformation to solve the hand-eye equation. Park et al. [12] proposed a separable solution strategy, wherein the rotation matrix is firstly obtained using the principles of least squares, followed by the computation of the translation matrix using the resultant rotation matrix.

The aforementioned methods employ a separable computational framework to sequentially solve for the rotation matrix and translation vector in the hand-eye matrix. The estimation of both rotation and translation parameters is achieved through linear algebraic methods. While these methods often yield satisfactory outcomes in ideal numerical simulation environments, their practical application is subject to the introduction of noise stemming from camera calibration and the motion of the robotic arm. Consequently, the linearization process of the rotation equation within the two-step approach becomes impractical for real-world scenarios. Furthermore, the two-step methodology tends to propagate errors from rotation estimation to translation estimation, thereby exhibiting limited robustness when confronted with noise in real-world scenarios.

To better accommodate the inherent noise in real-world scenarios and enhance the precision of hand-eye calibration, numerous scholars have proposed to utilize nonlinear optimization methods to simultaneously determine the rotational and translational components within the hand-eye matrix. These methodologies include approaches such as Inria [13], Kronecker [14], and Dual Tensor [15]. Furthermore, a number of approaches adopting the principles of global optimization have been introduced [16]. These methods have demonstrated enhanced robustness compared to two-step computational solution, and they are capable of yielding more accurate outcomes in hand-eye calibration. However, their implementation often entails complexity, potentially impeding computational efficiency. What is more, optimization methods may not guarantee convergence, and the

resultant solutions frequently encounter challenges associated with local minima, thereby compromising the stability of the algorithms.

Regardless of whether researchers employ a step-wise or synchronous methodology, their focus predominantly revolves around the exploration and investigation of approaches to effectively solve the mathematical problem of the hand-eye equation $AX = XB$. The evaluation of the results is accomplished by decomposing the hand-eye equation and conducting a comprehensive analysis from both rotational and translational aspects. However, when applying the hand-eye matrix in practical scenarios, it is often subjected to various degrees of noise originating from multiple sources within the system, including cameras, robotic arms, and operational environments. Although traditional solution methods may yield satisfactory results at the theoretical level, their practical accuracy is significantly compromised when confronted with such diverse noise, thereby lacking the desired robustness. Consequently, the key challenge in the further advancement of hand-eye calibration techniques lies in obtaining a hand-eye matrix that possesses both robustness and noise resilience.

In order to address the aforementioned issues, we propose an enhanced robust hand-eye relationship estimation model called ERHEC-NI that incorporates noise information into the hand-eye matrix. This model not only preserves the stability of traditional algorithms but also adapts to the noise in the scene, exhibiting excellent robustness and accuracy. ERHEC-NI initially utilizes the Tsai to obtain an initial value for the hand-eye matrix. To incorporate noise information into the hand-eye matrix, we combine a small amount of real sample data with the noise information and optimize the initial value of the hand-eye matrix through an objective function. Furthermore, unlike the traditional evaluation metrics shown in Equation (6) and (7), we assess the quality of the hand-eye matrix based on the error between the true three-dimensional coordinates of target points and the calculated coordinates. ERHEC-NI effectively integrates the noise information from the scene into the hand-eye matrix, thereby enhancing the stability, robustness, and accuracy of the algorithm in practical applications. We validate the feasibility of this method through both numerical simulations and experimental validations.

2. **Related Work.**

2.1. **Camera Calibration.**

The research findings suggest that camera calibration is not independent of hand-eye calibration, as a precise hand-eye relationship is often contingent upon a robust camera calibration result [4]. Camera calibration refers to the process of establishing the geometric relationship between real-world target points and their corresponding image points by utilizing two-dimensional plane information in conjunction with a limited amount of three-dimensional spatial information. The Zhang calibration method [17] has played a significant role in the development of camera calibration techniques. Zhang proposed the adoption of a chessboard pattern as the calibration target for analysis and processing, enabling the derivation of both the intrinsic and extrinsic camera parameters. Moreover, alternative calibration targets, such as solid circles [18], concentric circular rings, and various novel templates [19], have been introduced in recent years. In this study, we will employ the Zhang calibration method for the camera calibration.

2.2. **Hand-Eye Calibration.**

Ever since the appearance of the hand-eye equation $AX = XB$, scholars have embarked on a comprehensive exploration and investigation of this problem. In addition to the aforementioned approaches, various other methodologies have been proposed, including Dual Quaternions [20, 21], Screw Motions [22], and more. Each of these methods possesses distinct advantages and limitations at the mathematical level. However, they often fail to adequately consider the noise interference within the scene,

leading to a lack of robustness and stability, thereby impeding their practical applicability and achieving satisfactory results.

Furthermore, scholars have also introduced the concept of minimizing re-projection errors to solve the problem of hand-eye calibration [23]. This method offers promising prospects by mitigating the impact of noise on the algorithm and enhancing solution accuracy. However, it primarily considers noise arising from camera calibration, and limits its comprehensive evaluation to the mathematical level. Moreover, re-projection error-based techniques often necessitate specific constraints on the camera's field of view, thereby diminishing their generalizability.

To overcome these limitations, [24] proposed a novel approach that utilizes measured values in the world coordinate system to determine the geometric relationship between the robot and the world, subsequently establishing the hand-eye matrix. But this method still requires further refinement to achieve the desired levels of accuracy and robustness.

With the advancement of neural network technology, researchers have begun incorporating neural network techniques into hand-eye calibration [25, 26]. The utilization of artificial neural networks in hand-eye calibration can be viewed as establishing a mapping relationship between the hand coordinates in the robot's base coordinate system and the image coordinates corresponding to the calibration object. By training the model, the desired hand coordinates can be obtained from the observed object positions captured by the camera. One advantage of this approach is its ability to operate without prior knowledge of camera parameters or pose estimation. This stems from the remarkable capacity of artificial neural networks to generalize nonlinear relationships among variables, rendering them suitable for handling noisy data. However, it is worth noting that training neural networks often necessitates a substantial amount of data, thereby rendering each hand-eye calibration task excessively laborious.

3. **Problem Statement.**

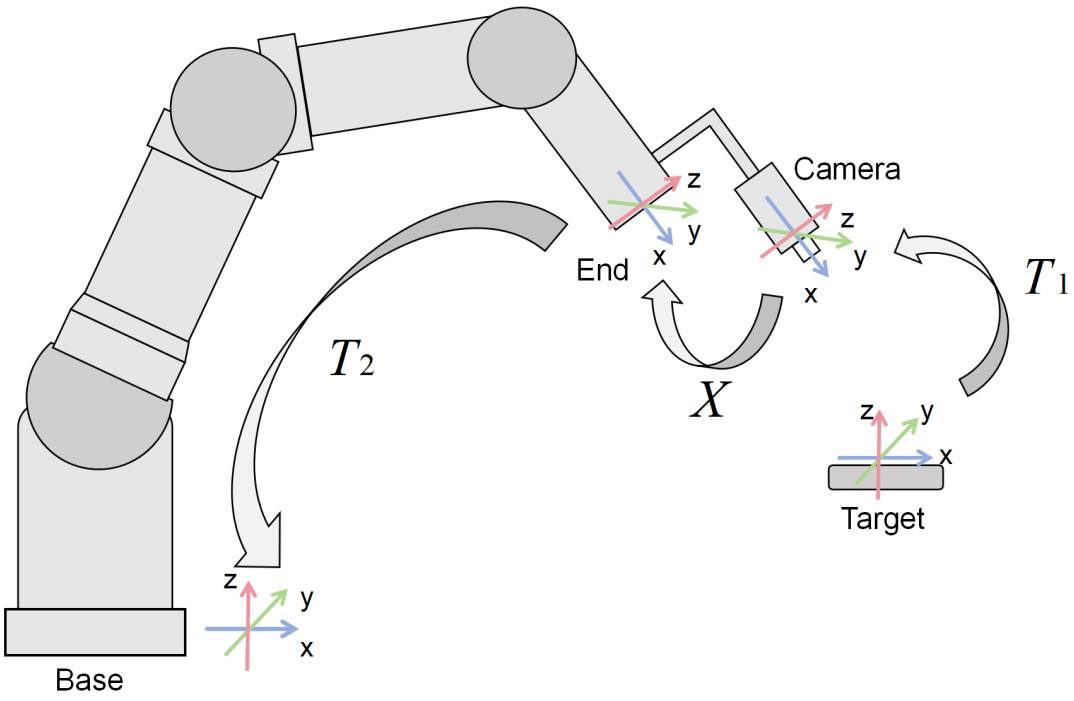

FIGURE 1. Coordinate conversion relationship of hand-eye calibration

3.1. **Mathematical Modeling.** Hand-eye calibration aims to determine the transformation matrix between the end-effector and the camera. Figure 1 illustrates the scenario of a single calibration, where T_1, X, and T_2 represent the transformation matrices between the world coordinate system and the camera coordinate system, the camera coordinate system and the robotic arm's end-effector coordinate system, and the robotic arm's end-effector coordinate system and the base coordinate system, respectively. Given the fixed relative position between the calibration board and the robotic arm base, when the robotic arm is moved, the relationship between adjacent poses satisfies the following Equation (1):

$$T_2 X T_1 = T_2' X T_1', \tag{1}$$

where the variables T_1' and T_2' represent the transformation relationships between different coordinate axes after the robotic arm is moved.

Based on matrix transformations, Equation (2) can be derived as followed:

$$T_2'^{-1} T_2 X = X T_1' T_1^{-1}. \tag{2}$$

Consider the left side of the Equation (2) as A and the right side as B, which is the classical equation $AX = XB$ for hand-eye calibration. Decoupling $AX = XB$ yields:

$$A = \begin{bmatrix} R_A & t_A \\ 0 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} R_B & t_B \\ 0 & 1 \end{bmatrix}, \quad X = \begin{bmatrix} R_X & t_X \\ 0 & 1 \end{bmatrix}, \tag{3}$$

where R and t respectively represent the rotation matrix and translation vector.

By substituting the decoupled results back into the equation $AX = XB$, we obtain Equation (4) and (5):

$$R_A R_X = R_X R_B, \tag{4}$$

$$R_A t_X + t_A = R_X t_B + t_X. \tag{5}$$

Upon observing Equation (4) and (5), it is evident that solving for X requires a minimum of two sets of data, which means we need to perform at least three transformations on the robotic arm. However, in practical applications, to enhance the accuracy, multiple transformations of the robotic arm are typically performed to obtain the result.

By deriving Equation (6) and (7) from Equation (4) and (5) respectively, these two equations are commonly employed by researchers as metrics for evaluating the quality of the hand-eye matrix.

$$e_R = \frac{1}{n} \sum_{i=1}^{n} ||R_A^i R_X - R_X R_B^i||_F, \tag{6}$$

$$e_t = \frac{1}{n} \sum_{i=1}^{n} ||R_A^i t_X - R_X t_B^i - t_X + t_A^i||_2. \tag{7}$$

3.2. **The Conversation of 2D to 3D.** In order to establish the transformation relationship between a two-dimensional image and a three-dimensional space, we can refer to the work presented in [17], where Equation (8) can be derived as followed:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & u_0 & 0 \\ 0 & f_y & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R_{T_1} & t_{T_1} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X_W \\ Y_W \\ Z_W \\ 1 \end{bmatrix}, \tag{8}$$

where u and v represent pixel coordinates, X_W, Y_W, and Z_W represent world coordinates, f_x, f_y, u_0, and v_0 represent parameters in the intrinsic matrix, T_1 has the same transformation relationship as defined in Section 3.1, and Z_c represents depth information.

From the transformation relationship shown in Figure 1, we can obtain Equation (9) as followed:

$$\begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = T_2 X T_1 \begin{bmatrix} X_W \\ Y_W \\ Z_W \\ 1 \end{bmatrix}, \tag{9}$$

where x, y, and z represent the coordinates in the base coordinate system, X_W, Y_W, and Z_W represent the coordinates in the world coordinate system, and T_2, X, and T_1 are defined the same as in Section 3.1.

By combining Equation (8) and (9), we can derive Equation (10) as followed:

$$Z_c \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & u_0 & 0 \\ 0 & f_y & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R_x & t_x \\ 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} R_{T_2} & t_{T_2} \\ 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}, \tag{10}$$

where R_X and t_X represent the rotation and translation parts of the hand-eye matrix, T_2 has the same transformation relationship as defined in Section 1.1, and x, y, and z represent the coordinates of the target point in the base coordinate system.

4. **Model.** To effectively accommodate the noise inherent in the scene, we propose a model named ERHEC-NI. ERHEC-NI takes a set of calibration data and a small amount of sample point with noise information as input and outputs a highly robust and accurate hand-eye relationship matrix. This model achieves the complete process of target grasping by a monocular vision robot while obtaining a highly precise hand-eye matrix. The architecture of the model is illustrated in Figure 2. The yellow Pose module in Figure 2 represents the end-effector pose information corresponding to each calibration image. The pink Pre-Data module represents the prepared sample point information, where each sample point's information includes its three-dimensional coordinates P in the base coordinate system, its pixel coordinates p in the image, its depth information Z_c, and the corresponding end-effector pose G of the image. Therefore, the Pre-Data module contains n sets of data, corresponding to n sample points, and the i-th set of the data contains $\{P_i, p_i, Z_{c_i}, G_i\}$.

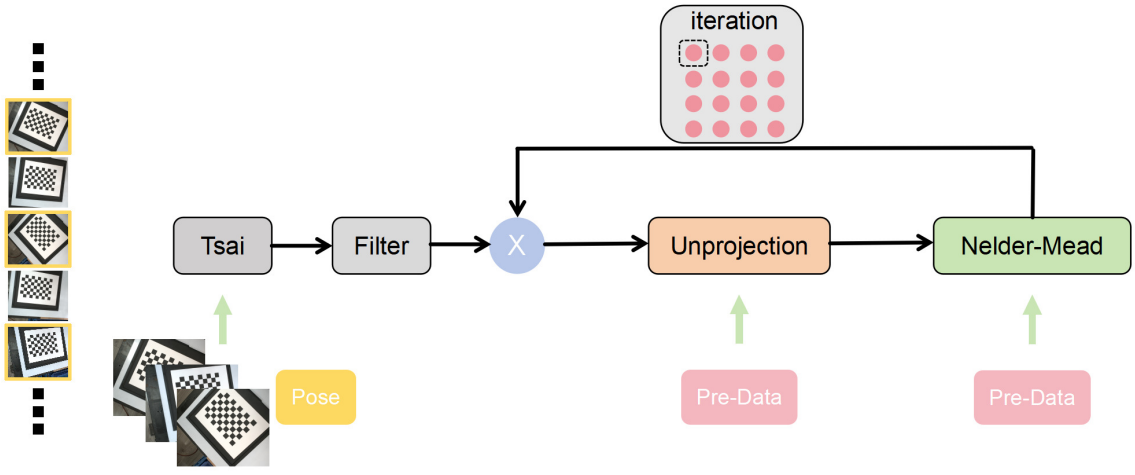

FIGURE 2. Framework of ERHEC-NI model

4.1. **Key Frame Selection.** In order to obtain a better initial value for the hand-eye matrix, we aim to identify a set of suitable keyframes as input to the model. To provide enough motion parallax and maintain the algorithm stability, we select the input frames based on a specified maximum and minimum relative rotation angles (R_{max} and R_{min}) and maximum and minimum relative translation distances (t_{max} and t_{min}) between adjacent frames. The keyframes, along with their corresponding end-effector pose information, are then passed to the Tsai module, resulting in the initial value X of the hand-eye matrix denoted by the blue circle (the initial value X needs to be obtained through the Filter module first, which will be introduced in Section 4.2). In the Tsai module, we integrate the Zhang calibration method to perform camera calibration, further enhancing the accuracy and reliability of the hand-eye calibration process.

4.2. **Initial Value Filtering.** In order to provide a higher-quality initial value for the subsequent optimization of the hand-eye matrix, we perform an initial filtering of the matrix within the filter module. The filtering approach employed here is computing the coordinates of the top-left corner of the chessboard pattern (assumed to be the origin of the world coordinate system, denoted as $W_0 = [0, 0, 0, 1]$) in the coordinate system of the robotic arm base. The calculation method is illustrated by Equation (9). Subsequently, the mean square deviation of the values along each axis is computed, resulting in e_x, e_y, and e_z as shown in Equations (11), (12), and (13) respectively.

$$e_x = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}, \tag{11}$$

$$e_y = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2}, \tag{12}$$

$$e_z = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(z_i - \bar{z})^2}, \tag{13}$$

where e_x, e_y, and e_z respectively represent the errors along the three axes, x_i, y_i, and z_i respectively represent the coordinates in the base coordinate system of the top-left corner points of the calibration board in the i-th calibration image.

Through subsequent experiments, it can be concluded that we prefer to obtain an initial value of the hand-eye matrix with an error of less than 3mm for each axis. We acknowledge that such evaluation metrics still have limitations in fully considering the presence of noise. However, compared to the traditional evaluation metrics represented by Equations (6) and (7), these metrics provide a more geometrically intuitive perspective.

4.3. **Unprojection.** The purpose of the orange Unprojection module in Figure 2 is to compute the coordinates of the sample points from the Pre-data module in the base coordinate system, given the current hand-eye matrix X. By inputting the data from the Pre-data module, it achieves the transformation of target points from a two-dimensional plane to a three-dimensional space. The specific implementation process is demonstrated by Equation (10).

The Unprojection module is designed to compute the predicted three-dimensional coordinates of the sample points under the initial value X, for the purpose of subsequent comparison with their actual three-dimensional coordinate values.

4.4. **Noise Information Fusion.** In our research, we establish the ground truth hand-eye matrix as the one that accurately captures the transformation relationship between the pixel coordinates of target points and their corresponding true three-dimensional coordinates within a specific scene. To obtain the results that closely approximate the ground truth, we optimize and fine-tune the initial value X by a small set of sample point data with noise information. This approach aims to refine the initial value of the hand-eye matrix by incorporating the noise characteristics into the initial value X, thereby enhancing its alignment with the ground truth.

In the Nelder-Mead module depicted in green, we employ the Nelder-Mead algorithm [27] to systematically optimize each parameter of the hand-eye matrix X by the objective function. The optimization process follows the way illustrated by the iteration module in Figure 2, wherein each iteration updates a parameter values of X that minimize the objective function. The objective function to be minimized is presented in Equation (14):

$$\min f(x) = \frac{1}{n}\sum_{i=1}^{n}(\hat{x}_i - x_i)^2 + \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2 + \frac{1}{n}\sum_{i=1}^{n}(\hat{z}_i - z_i)^2, \tag{14}$$

where $[\hat{x}, \hat{y}, \hat{z}]$ represents the true three-dimensional coordinates of the sample points. The calculated three-dimensional coordinates $[x, y, z]$ are derived through Equation (10), where the hand-eye matrix X is explicitly manifested.

In the process of optimizing the parameters of X, our decision to individually optimize each parameter is motivated by the purpose to address potential issues such as local optima and computational inefficiency commonly encountered in multi-variable optimization problems. We consider the 12 parameters in the rotation and translation components of X as a cohesive unit, constituting a single optimization round. To enhance precision, we suggest conducting multiple rounds of optimization for the hand-eye matrix X. This iterative approach helps to refine the parameter values and leads to improved accuracy in the calibration process.

It is apparent that our approach does not involve an exhaustive analysis of the diverse sources of noise within the hand-eye system, nor does it attempt to address them individually. Instead, we adopt a goal-oriented methodology that uses a small set of sample points to effectively incorporate the noise information present in the system into the initial value of the hand-eye matrix. Finally, we obtain a result of hand-eye calibration that exhibits both high robustness and precision.

5. **Experiments.** In this section, we conducted numerical simulations and experimental validations successively to assess the feasibility of ERHEC-NI.

5.1. **Evaluation Metrics.** The traditional metrics presented in Equations (6) and (7) only provide a mathematical evaluation of the results obtained from different methods for solving the hand-eye equation $AX = XB$. While this evaluation approach holds mathematical rigor, it solely consider the noise introduced during the estimation of the external parameters in the camera calibration. It fails to consider the various sources of noise that the hand-eye matrix may encounter during practical applications, such as the noise introduced by the calibration tool coordinate system, the noise arising from the estimation of the camera's intrinsic parameter matrix, the noise from the depth estimation of target points, the noise introduced by the motion of the robotic arm and so on. Consequently, the direct application of high-quality hand-eye matrices obtained through traditional evaluation metrics often fails to yield desirable results in real-world scenarios.

Adopting a goal-oriented approach, our primary task is not to find a optimal solution of the mathematical equation $AX = XB$ under specific constraints, but rather to obtain a

hand-eye matrix that holds great robustness and noise resilience in real-world application scenarios. Thus, we define the ground truth hand-eye matrix as the one that accurately captures the transformation relationship between the pixel coordinates of target points and their corresponding true three-dimensional coordinates within a specific scene.

We propose a novel evaluation metric that is goal-oriented and incorporates real-world noise information to evaluate the quality of the hand-eye matrix. Our approach needs to obtain the pixel coordinates of the target points in the image and utilizing Equation (10) to estimate their three-dimensional coordinates $[x, y, z]$ in the base coordinate system. Subsequently, we compare these estimated coordinates with the actual three-dimensional coordinates $[\hat{x}, \hat{y}, \hat{z}]$. The specific implementation, as presented in Equation (15), needs to define the error of different hand-eye matrices as the average of the errors along the three axes.

$$e_X = \frac{1}{n} \sum_{i=1}^{n} \frac{(|\hat{x}_i - x_i| + |\hat{y}_i - y_i| + |\hat{z}_i - z_i|)}{3}, \tag{15}$$

where e_X represents the error corresponding to a certain hand-eye matrix X.

5.2. **Numerical Simulations.** We conducted numerical simulations to verify the efficacy of the proposed approach, which optimizes and fine-tunes the initial values of the hand-eye matrix by using a small number of sample points with noise information. Furthermore, we analyzed the sensitivity of the hand-eye matrix to different levels of noise before and after the optimization.

In our simulation experiments, we generated a random 3×3 orthogonal matrix R_X and a 3×1 vector t_X, which served as a hand-eye matrix X estimated by traditional methods. The sources of noise in the scene are often complex and difficult to quantify. Considering that the method proposed in this paper fundamentally incorporates all the noise information into the hand-eye matrix, we directly introduced Gaussian noise to the hand-eye matrix X during the stage of noise injection. Specifically, we added Gaussian noise with a mean of 0 and a standard deviation of $0.01d$ to each parameter in the rotation and translation components of X, where d represents the noise level. Consequently, the resulting $\hat{X}$, influenced by the introduced noise, was regarded as the ground truth of hand-eye matrix, while the noise-free X served as the initial value of the hand-eye matrix.

Fifty pixel points were randomly generated, out of which eight were selected for the optimization of the hand-eye matrix X. The resulting refined hand-eye matrix was denoted as X'. All pixel points were considered as data samples for evaluating the performance of X'. To capture more comprehensive trends, we repeated the aforementioned procedure 500 times, introducing noise with the same distribution but distinct data each time. The evaluation metric for the simulation experiments was determined by computing the average error of the fifty points under various hand-eye matrix computations, as shown in Equation (15). The remaining parameters involved in the calculations were also randomly generated.

Subsequently, we set the noise level to $d = 1, 2, 3, ..., 10$. By repeatedly performing the above calculation steps, we can obtain the error distribution of the hand-eye matrix before and after optimization under different noises. This analysis is depicted in Figure 3, where the errors corresponding to different noise levels are computed as the average of 500 experimental trials.

Based on the analysis of Figure 3, two important conclusions can be learned: (1) Compared to the initial hand-eye matrix, the optimization process in this study markedly improves the precision of hand-eye matrix in the situation of diverse noise sources. (2)

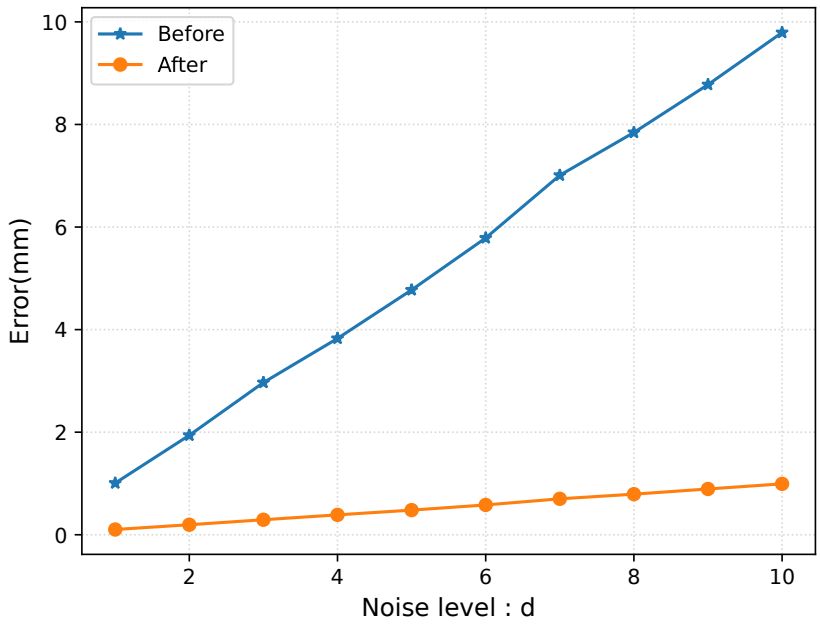

FIGURE 3. Accuracy of hand-eye matrix before and after optimization under different noise influences

Facing progressively escalating noise disturbances, the error growth of the optimized hand-eye matrix is significantly lower than that of the initial hand-eye matrix, thereby exhibiting enhanced robustness and stability properties.

In order to provide a more detailed representation of the numerical values in the simulation experiments, we have included Table 1, which specifically pertains to a noise level of $d = 5$.

TABLE 1. Accuracy of hand-eye matrix before and after optimization

Method	Maximum Error (in mm)	Minimum Error (in mm)	Average Error (in mm)	Standard Deviation (in mm)
Before	12.980	0.485	4.998	2.079
After	**1.413**	**0.068**	**0.505**	**0.214**

As demonstrated in Figure 3 and Table 1, the concept of refining the initial value of the hand-eye matrix by using a small number of sample points with noise information proves to be an effective approach. Furthermore, the resulting hand-eye matrix obtained through this method exhibits enhanced robustness and higher precision.

5.3. **Experimental Validations.** In this section, we conducted a series of experimental validations using a robotic arm to assess the feasibility of ERHEC-NI. The experimental setup consisted of the ROKAE robotic arm, HIKVISION camera, and an infrared rangefinder. The calibration board employed followed a standardized 8×6 chessboard pattern, with each individual square measuring 30mm in length. The experimental scenario is depicted in Figure 4, where the robotic arm's end-effector was equipped with a precision needle.

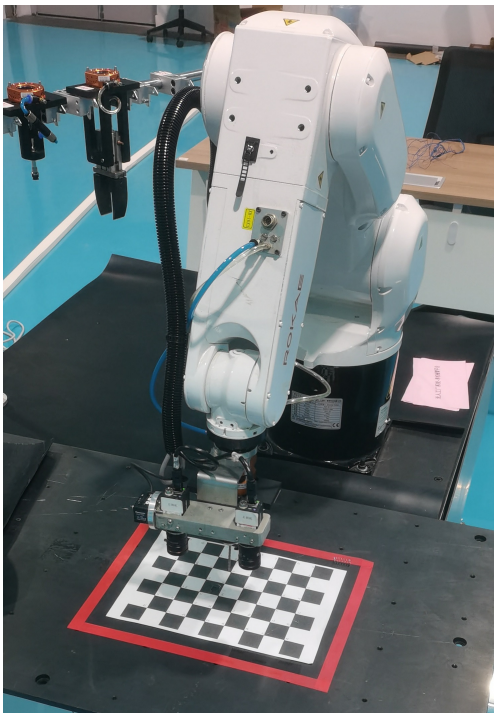

FIGURE 4. Experimental setting for robotic arm

5.3.1. *Experimental Details.* Before the experiments, several preparatory steps need to be performed. Firstly, it is necessary to calibrate the tool coordinate system at the end-effector of the robotic arm, ensuring that the vertex of the tool TCP (Tool Center Point) serves as the reference vertex for the end-effector coordinate system. Subsequently, the robotic arm is manipulated to capture images. Simultaneously obtaining the calibration images, we select eight sample points to serve as preparatory data for optimizing the initial values of the hand-eye matrix. These sample points' pixel coordinates, actual coordinates, and the corresponding pose information of the image need to be recorded. The method to obtain the actual three-dimensional coordinates of the target points is illustrated in Figure 5, needing to manipulate robotic arm's TCP vertex to make contact with the target points, thereby capturing the TCP vertex's three-dimensional coordinates in the base coordinate system of the robotic arm. Finally, employing a similar sampling approach, we select fifty uniformly distributed sample points within the operational workspace of the robotic arm, which will serve as the dataset for the subsequent error analysis.

To achieve more generalizable outcomes, we conducted five calibration experiments on the robotic arm. Each experiment involved capturing 10-15 images. The positional information of the robotic arm's end-effector was recorded during image acquisition, and the parameters A_i in the hand-eye calibration equation $AX = XB$ were computed by multiplying the end-effector pose matrices from consecutive images. Simultaneously, the camera's extrinsic matrices corresponding to each image were obtained by the Zhang calibration method. Similarly, the other parameter B_i in the hand-eye equation was determined by multiplying the extrinsic matrices from consecutive images. The hand-eye matrices were computed using six different methods: Tsai, Inria, Navy, Dual Quaternion, ACRW, and ERHEC-NI. Subsequently, we computed the errors and conducted a comprehensive analysis of the results.

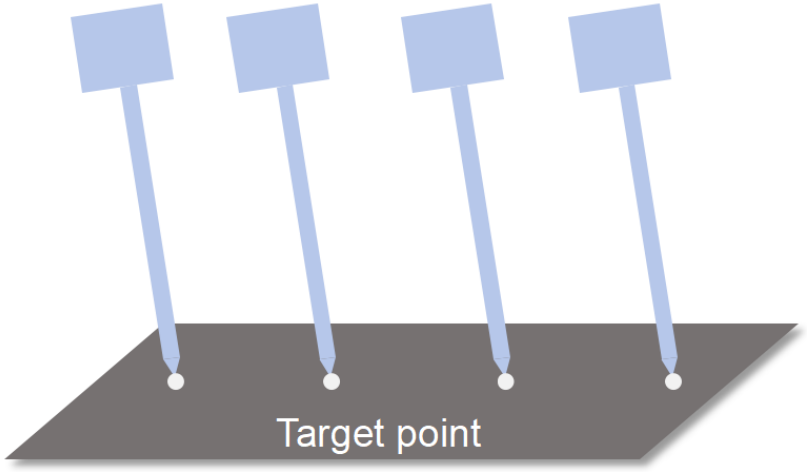

FIGURE 5. The way of obtaining the three-dimensional coordinates of target points

5.3.2. *Experimental Results.* In our study, we conducted comparative experiments between ERHEC-NI and five other methods. The experimental results are presented in Figure 6 and Table 2.

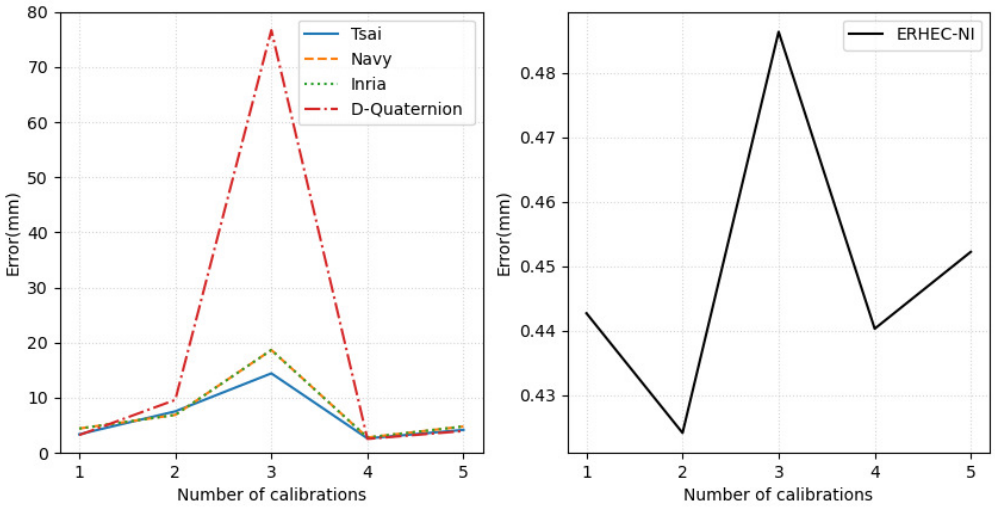

FIGURE 6. Accuracy of different methods in different calibration batches

In accordance with the evaluation metrics defined in Section 5.1, we define the error of each method as the average of the errors of all the points to assess the performance of different methods. Figure 6 illustrates the errors of ERHEC-NI and four traditional methods in different calibration groups. The experimental results show the conclusions below: (1) Compared to other algorithms, ERHEC-NI shows significant improvement in accuracy when confronted with environmental noise. The error is reduced from the millimeter level to below 0.5 millimeters. (2) ERHEC-NI exhibits enhanced stability, as it can yield precise hand-eye matrices even when initialized with suboptimal values, thanks to its fine-tuning module. (3) The stability of the four traditional algorithms is highly dependent on the quality of the calibration data samples, and Tsai demonstrating

relatively better stability. This is the reason why we choose Tsai as the hand-eye matrix parameter initialization module.

To present the experimental data more effectively, we selected the fourth group of experiments and created Table 3. In addition, we included the ACRW method for comparison. It can be observed that while the ACRW method considers the noise in the scene and shows some improvement in accuracy, it does not achieve the same level of precision as ERHEC-NI.

Table 2. Accuracy of different methods

Method	Maximum Error (in mm)	Minimum Error (in mm)	Average Error (in mm)	Standard Deviation (in mm)
Tsai [11]	4.661	1.539	2.642	0.841
Navy [12]	5.168	0.811	2.807	1.095
Inria [13]	5.168	0.811	2.808	1.095
Dual Quaternions [20]	4.349	0.721	2.553	0.896
ACRW [24]	1.849	0.799	1.283	0.388
ERHEC-NI	**1.060**	**0.047**	**0.440**	**0.258**

The experimental results presented in Table 3 further validate our ideas, demonstrating that ERHEC-NI exhibits better accuracy, robustness, and stability when faced with real-world noise scenarios. These characteristics make ERHEC-NI well-suited for practical applications, as it can be deployed more directly and efficiently, meeting the demands of real-world use. However, it is imperative to acknowledge that our model entails certain complexities in terms of algorithmic implementation, as it involves additional steps compared to conventional approaches. Nonetheless, it is worth noting that the time-related limitations of our methodology are unlikely to exert detrimental effects, given that hand-eye calibration is not a task necessitating frequent repetition.

5.4. **Discussion.** ERHEC-NI uses the Tsai method to obtain an initial estimate of the hand-eye matrix. It then employs the idea of incorporating noise information into the initial estimate to fine-tune and optimize the hand-eye matrix, thereby achieving a closer value to the ground truth. From both the principles and applications of calibration, it is evident that the accuracy of hand-eye calibration algorithms primarily relies on the accuracy of the robotic arm's motion and camera extrinsic calibration. It is also influenced by factors such as the precision of the depth of target points, camera intrinsic parameters, and unquantifiable noise in the scene. Essentially, ERHEC-NI couples all the noise of the hand-eye system onto the hand-eye matrix. This enables the hand-eye matrix to adapt to the noise generated by all hardware components and scene factors, leading to an overall improvement in system accuracy through self-learning and adaptive optimization of the hand-eye matrix. This approach mitigates the significant reliance on the quality of calibration information in previous methods and addresses the difficulty in handling noise. Moreover, ERHEC-NI is characterized by low cost, high efficiency, and better suitability for practical applications.

6. **Conclusions.** In this paper, we focus on the problem of how to obtain a hand-eye matrix with good robustness. Previous methods were unable to effectively handle various noise that interfere with the hand-eye matrix in practical applications. To address this issue, we propose a model named ERHEC-NI, which obtains a hand-eye matrix with enhanced robustness by fusing noise information. Specifically, in this work, we first obtain an initial value of the hand-eye matrix using the Tsai method. Then, we incorporate the

noise information of sample points into the hand-eye matrix, aiming to refine the initial estimate and obtain an approximation closer to the ground truth. Results from numerical simulations and experimental validations demonstrate the superiority of the ERHEC-NI method compared to other algorithms.

Acknowledgment. This work was partially supported by Ningbo Science and Technology Innovation 2025 major project under grants 2020Z106 and 2023Z040. The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

REFERENCES

[1] F. Aghili, "A prediction and motion-planning scheme for visually guided robotic capturing of free-floating tumbling objects with uncertain dynamics," *IEEE Transactions on Robotics*, vol. 28, no. 3, pp. 634–649, 2012.

[2] A. O. Fernandes, L. F. Moreira, and J. M. Mata, "Machine vision applications and development aspects," in *2011 9th IEEE International Conference on Control and Automation (ICCA)*. IEEE, 2011, pp. 1274–1278.

[3] A. Y. Nee, S. Ong, G. Chryssolouris, and D. Mourtzis, "Augmented reality applications in design and manufacturing," *CIRP Annals*, vol. 61, no. 2, pp. 657–679, 2012.

[4] R. Horaud and F. Dornaika, "Hand-eye calibration," *The International Journal of Robotics Research*, vol. 14, no. 3, pp. 195–210, 1995.

[5] S. Thompson, D. Stoyanov, C. Schneider, K. Gurusamy, S. Ourselin, B. Davidson, D. Hawkes, and M. J. Clarkson, "Hand–eye calibration for rigid laparoscopes using an invariant point," *International Journal of Computer Assisted Radiology and Surgery*, vol. 11, pp. 1071–1080, 2016.

[6] Z. Zhang, L. Zhang, and G.-Z. Yang, "A computationally efficient method for hand–eye calibration," *International Journal of Computer Assisted Radiology and Surgery*, vol. 12, pp. 1775–1787, 2017.

[7] D. Briese, C. Niebler, and G. Rose, "Hand-eye calibration using dual quaternions in medical environment," in *Medical Imaging 2014: Image-Guided Procedures, Robotic Interventions, and Modeling*, vol. 9036. SPIE, 2014, pp. 488–493.

[8] Y. Liu, J. Zhang, Z. She, A. Kheradmand, and M. Armand, "Gbec: Geometry-based hand-eye calibration," *arXiv preprint arXiv:2404.05884*, 2024.

[9] Y. C. Shiu and S. Ahmad, "Calibration of wrist-mounted robotic sensors by solving homogeneous transform equations of the form ax= xb," 1987.

[10] M. Shah, R. D. Eastman, and T. Hong, "An overview of robot-sensor calibration methods for evaluation of perception systems," in *Proceedings of the Workshop on Performance Metrics for Intelligent Systems*, 2012, pp. 15–20.

[11] R. Y. Tsai and R. K. Lenz, "Real time versatile robotics hand/eye calibration using 3d machine vision," in *Proceedings. 1988 IEEE International Conference on Robotics and Automation.* IEEE, 1988, pp. 554–561.

[12] F. C. Park and B. J. Martin, "Robot sensor calibration: solving ax = xb on the euclidean group," *IEEE Transactions on Robotics and Automation*, vol. 10, no. 5, pp. 717–721, 1994.

[13] F. Dornaika and R. Horaud, "Simultaneous robot-world and hand-eye calibration," *IEEE Transactions on Robotics and Automation*, vol. 14, no. 4, pp. 617–622, 1998.

[14] R.-h. Liang and J.-f. Mao, "Hand-eye calibration with a new linear decomposition algorithm," *Journal of Zhejiang University-SCIENCE A*, vol. 9, no. 10, pp. 1363–1368, 2008.

[15] D. Condurache and A. Burlacu, "Orthogonal dual tensor method for solving the ax= xb sensor calibration problem," *Mechanism and Machine Theory*, vol. 104, pp. 382–404, 2016.

[16] J. Heller, M. Havlena, and T. Pajdla, "A branch-and-bound algorithm for globally optimal hand-eye calibration," in *2012 IEEE Conference on Computer Vision and Pattern Recognition.* IEEE, 2012, pp. 1608–1615.

[17] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 11, pp. 1330–1334, 2000.

[18] L. Bu, H. Huo, X. Liu, and F. Bu, "Concentric circle grids for camera calibration with considering lens distortion," *Optics and Lasers in Engineering*, vol. 140, p. 106527, 2021.

[19] X. Chen, X. Song, J. Wu, Y. Xiao, Y. Wang, and Y. Wang, "Camera calibration with global lbp-coded phase-shifting wedge grating arrays," *Optics and Lasers in Engineering*, vol. 136, p. 106314, 2021.

[20] K. Daniilidis and E. Bayro-Corrochano, "The dual quaternion approach to hand-eye calibration," in *Proceedings of 13th International Conference on Pattern Recognition*, vol. 1. IEEE, 1996, pp. 318–322.

[21] X. Wang, H. Sun, C. Liu, and H. Song, "Dual quaternion operations for rigid body motion and their application to the hand–eye calibration," *Mechanism and Machine Theory*, vol. 193, p. 105566, 2024.

[22] H. H. Chen, "A screw motion approach to uniqueness analysis of head-eye geometry," in *Proceedings. 1991 IEEE Computer Society Conference on Computer Vision and Pattern Recognition.* IEEE Computer Society, 1991, pp. 145–146.

[23] K. Koide and E. Menegatti, "General hand–eye calibration based on reprojection error minimization," *IEEE Robotics and Automation Letters*, vol. 4, no. 2, pp. 1021–1028, 2019.

[24] M. Dinham and G. Fang, "A low cost hand-eye calibration method for arc welding robots," in *2009 IEEE International Conference on Robotics and Biomimetics (ROBIO)*. IEEE, 2009, pp. 1889–1893.

[25] K. Zhou, X. Huang, S. Li, and G. Li, "Convolutional neural network-based pose mapping estimation as an alternative to traditional hand–eye calibration," *Review of Scientific Instruments*, vol. 94, no. 6, 2023.

[26] J. Hua and L. Zeng, "Hand–eye calibration algorithm based on an optimized neural network," in *Actuators*, vol. 10, no. 4. MDPI, 2021, p. 85.

[27] R. Kshirsagar, S. Jones, J. Lawrence, and J. Tabor, "Optimization of tig welding parameters using a hybrid nelder mead-evolutionary algorithms method," *Journal of Manufacturing and Materials Processing*, vol. 4, no. 1, p. 10, 2020.