

Article

A Likelihood-Based Pose Estimation Method for Robotic Arm Repeatability Measurement Using Monocular Vision

Peng Zhang, Jiatian Li , Jiayin Liu, Feng He and Yiheng Jiang

Faculty of Land and Resources Engineering, Kunming University of Science and Technology, Kunming 650093, China; 20232101047@stu.kust.edu.cn (P.Z.); 20233101016@stu.kust.edu.cn (J.L.); 20222201097@stu.kust.edu.cn (F.H.); rayjj@stu.kust.edu.cn (Y.J.)

* Correspondence: 20080010@kust.edu.cn

Abstract

Repeatability accuracy is a key performance metric for robotic arms. To address limitations in existing monocular vision-based measurement methods, this study proposes a likelihood-based pose estimation approach. Our method first obtains initial pose estimates through optimized likelihood estimation, then iteratively refines depth information. By modeling the statistical characteristics of multiple observed poses, we derive a global theoretical pose. Within this framework, two-dimensional feature points are backprojected into three-dimensional space to form an observed point cloud. The error between this observed cloud and the theoretical feature point cloud is computed using the Iterative Closest Point (ICP) algorithm, enabling accurate quantification of repeatability accuracy. Based on 30 repeated trials at each of five target poses, the proposed method achieved repeatability positioning accuracy of 0.0115 mm, 0.0121 mm, 0.0068 mm, 0.0162 mm, and 0.0175 mm at the five poses, respectively, with a mean value of 0.0128 mm and a standard deviation of 0.0038 mm across the poses. Compared with two existing monocular vision-based methods, it demonstrates superior accuracy and stability, achieving average accuracy improvements of 0.79 mm and 1.06 mm, respectively, and reducing the standard deviation by over 85%.

Keywords: repeatability; maximum likelihood estimation; pose estimation; Cramér–Rao bound; Iterative Closest Point (ICP)



Academic Editor: Alexander Ferrein

Received: 20 October 2025

Revised: 14 November 2025

Accepted: 20 November 2025

Published: 20 November 2025

Citation: Zhang, P.; Li, J.; Liu, J.; He, F.; Jiang, Y. A Likelihood-Based Pose Estimation Method for Robotic Arm Repeatability Measurement Using Monocular Vision. *Sensors* **2025**, *25*, 7089. <https://doi.org/10.3390/s25227089>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In the field of intelligent manufacturing and automated production, repeatability is a core metric for evaluating the performance of robotic arms and other equipment. Its measurement accuracy directly impacts product quality and production efficiency [1–3]. High-precision repeatability measurement is fundamental to ensuring the reliability of processes such as assembly, welding, and inspection.

Currently, robotic arm repeatability measurement methods are categorized into laser-based and vision-based approaches according to the sensor type. Laser measurement techniques [4–7] capture coordinate data by detecting reflected laser spots on target surfaces, demonstrating significant advantages in specific scenarios due to their high measurement accuracy. However, this approach suffers from limitations, including high equipment costs, poor environmental adaptability, and the requirement for specialized targets. The complex installation and calibration procedures make it difficult to meet the practical demands of complex and dynamic industrial environments. In contrast, visual measurement methods offer superior scene adaptability, more convenient deployment methods, and lower

cost investment. Within the visual measurement technology framework, pose estimation constitutes a critical component for achieving high-precision measurement. Diverse pose estimation methods provide multiple technical pathways for visual repeatability accuracy measurement. Current research primarily focuses on three major directions: camera calibration methods, non-iterative methods, and iterative methods. Camera calibration methods establish a transformation relationship between the image coordinate system and the world coordinate system to achieve accurate measurement of the end-effector pose of a robotic arm. R.A. Bobby et al. [8] proposed a phased monocular camera calibration method and applied it to robotic arm repeatability measurement tasks. Wu et al. [9] introduced a vision-based robotic motion measurement approach, enhancing measurement accuracy and stability through monocular feature matrices and singular value decomposition. Non-iterative methods rely on analytical geometric constraints, achieving measurement through direct pose parameter solutions. To enhance algorithmic efficiency and accuracy, ref. [10] proposed the EPnP model, and multiple improved variants were subsequently derived, including RPnP [11], DLS [12], OPnP [13], SRPnP [14], and the deep learning-based EPro-PnP model [15]. However, measuring robotic repeatability accuracy under monocular vision represents a unique task paradigm. Its core objective involves performing multiple observations of the same scene in a static state and evaluating their consistency. This task focuses more on maximizing the measurement accuracy and stability of single-frame images without relying on temporal sequence information. Consequently, the widely recognized baseline methods remain the classical PnP algorithms and their improved variants. Among these, Gao et al. [16] proposed an image-based target localization method for collaborative robots, utilizing a monocular vision system combined with a PnP algorithm to measure robotic arm repeatability accuracy. Luo et al. [17] employed a coplanar P4P algorithm to determine target feature point coordinates in the camera coordinate system, integrating dual quaternions to solve the pose transformation matrix and thereby calculating repeatability accuracy. The iterative method transforms the PnP problem into an optimization objective function, solving the pose through iterative steps. Chen et al. [18] proposed a robotic arm positioning measurement system combining monocular vision with P4P and OI algorithms [19] that employs P4P results as initial values for OI algorithm optimization. However, most existing visual measurement methods scarcely consider the statistical properties of observational data and lack global optimization mechanisms for multi-frame data.

To address this, we propose a likelihood-based pose method for achieving high-precision repeatability measurement. This method first optimizes the estimated single-frame pose, then employs a likelihood-based pose estimation framework: based on maximum likelihood estimation theory, it transforms the pose problem of solving PnP for a stationary target observed multiple times into a problem of observing a slightly moving target multiple times through random signal modeling. Subsequently, the Iterative Closest Point (ICP) algorithm [20] computes displacement vectors between 3D–3D matched point pairs. The minimum envelope sphere radius is then determined using the progressive algorithm with a preference for the farthest point. This enables precise quantification of repeatability accuracy, overcoming the limitation of traditional single-frame measurement methods that cannot perform global optimization. Furthermore, by calculating 3D–3D point cloud deviations, more accurate and stable repeatability accuracy is obtained.

2. Materials and Methods

Repeatability accuracy measures the degree of pose dispersion across multiple responses to identical commands, with its core challenge lying in quantifying end-effector pose repeatability. Based on the aforementioned theory, this study employs the eye-in-

hand model. By establishing a rigid mechanical connection between the visual sensor and the end-effector, a fixed coordinate system transformation relationship is constructed. Subsequently, by calculating the repeatability of the visual sensor's pose, an indirect measurement of the end-effector's repeatability accuracy is achieved. The overall process is illustrated in Figure 1.

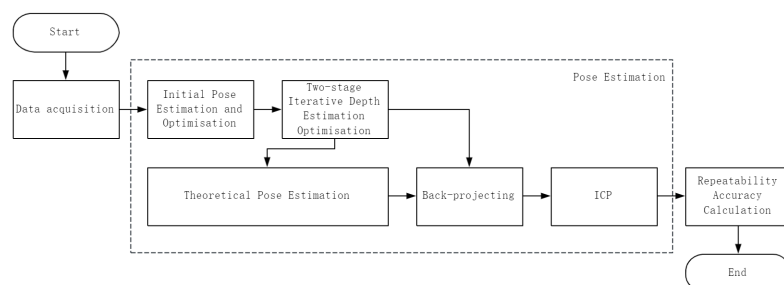


Figure 1. Algorithm flowchart.

The overall process comprises three modules: data acquisition, pose estimation, and repeatability accuracy calculation. The data acquisition module utilizes an industrial camera to capture image data, establishing the foundational geometric data for subsequent computations. The pose estimation module employs a weighted likelihood method to refine the initial pose estimates from PnP, mitigating the impact of observation outliers caused by image noise. Based on the optimized pose, a two-stage iterative algorithm solves single-frame depth information, explicitly enhancing estimation stability in the depth direction; by constructing a random signal model and employing maximum likelihood theory, it fuses multi-frame optimized poses to derive a globally optimal pose solution as the theoretical pose; Within the theoretical pose framework, depth information that is iteratively optimized is utilized to backproject two-dimensional feature points onto the world coordinate system, forming a three-dimensional observation point cloud under theoretical pose observation. The ICP algorithm is then employed to compute the registration error between the observation point cloud and the theoretical three-dimensional feature point cloud. The repeatability accuracy calculation module determines the final repeatability accuracy result by computing the minimum envelope sphere radius of multi-frame registration errors.

2.1. Initial Pose Estimation and Optimization

Within the data acquisition module, image data from a calibration plate is captured using an industrial camera. The calibration plate image acquisition protocol adheres to the repeatability accuracy data collection method specified in the national standard ISO 9283 [21]. Position the calibration plate within the workspace of the robotic arm under test. Using the teach pendant, program the control sequence to write five distinct poses into the controller, designated as P1, P2, P3, P4, and P5, respectively. Command the robotic arm to traverse repeatedly between two points at normal operating speed, capturing image data upon reaching each target pose. Each image acquisition sequence is repeated 30 times. The detailed image acquisition workflow is illustrated in Figure 2.

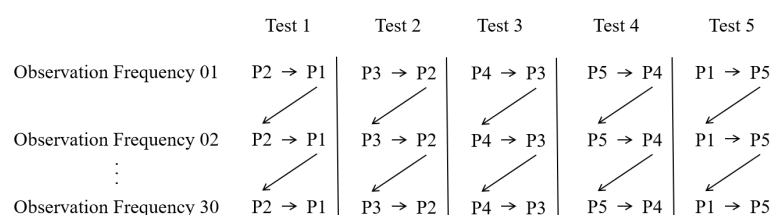


Figure 2. Image acquisition flowchart.

To convert abstract image information into quantifiable pose data, the PnP algorithm is employed for initial camera pose estimation. Processing images acquired from a single observation yields n pairs of 2D–3D matched feature points. For the k -th observation, the coordinates of the i -th feature point in the world coordinate system are denoted as $P_i^w(k) = [X_i^w(k), Y_i^w(k), Z_i^w(k)]^T$, and its pixel coordinates in the image coordinate system are $P_i^{img}(k) = [u_i^k, v_i^k]^T$, where $i = 1, 2, \dots, n$. The transformation relationship can be expressed using the ideal pinhole camera model:

$$\begin{cases} u_i^k = f_x \frac{r_1^k P_i^w(k) + t_x^k}{r_3^k P_i^w(k) + t_z^k} + c_x \\ v_i^k = f_y \frac{r_2^k P_i^w(k) + t_y^k}{r_3^k P_i^w(k) + t_z^k} + c_y \end{cases} \quad (1)$$

$Cam = \{f_x, f_y, c_x, c_y\}$ denotes the camera intrinsic parameters, obtainable through calibration experiments and representing fixed values. To quantify the precision of the calibration, we report the maximum re-projection error across all calibration images, which is 0.11 pixels. This low value confirms the accuracy and reliability of the estimated intrinsic parameters $Cam = \{f_x, f_y, c_x, c_y\}$ for our subsequent visual odometry computations. $t^k = [t_x^k, t_y^k, t_z^k]^T$ is the translation vector, and $R^k = [r_1^k, r_2^k, r_3^k]^T$ denotes the rotation matrix, where $k = 1, 2, \dots, K$, and t^k and R^k are initial pose parameters to be estimated, characterizing the camera's pose during the k -th observation.

As images in practical applications often contain noise, the two-dimensional feature point $P_i^{img}(k)$ observed at the k -th observation is typically accompanied by observation noise $\epsilon(k)$. Since the initial pose estimation is based on the PnP algorithm, which is sensitive to noise in two-dimensional feature points, this noise introduces bias into the estimated initial pose t^k and R^k , thereby affecting subsequent theoretical pose estimation. To safeguard the accuracy of subsequent theoretical pose estimation, nonlinear optimization was performed on the initial pose t^k and R^k . Currently, many nonlinear optimization-based VO systems explicitly assume that visual feature measurement noise is Gaussian white noise [22,23]. Based on this, it is assumed that at the k -th observation, the observation noise $\epsilon(k)$ follows a Gaussian distribution with mean zero and variance σ^2 , and its projection process can be expressed as

$$P_i^{img}(k) = \pi(R^k, t^k, P_i^w(k)) + \epsilon(k) \quad (2)$$

where $\pi(R^k, t^k, P_i^w(k))$ denotes the projection function of the ideal pinhole camera model. At the k -th observation, to account for uncertainty in feature point extraction, a weight $w(k) = [w_1^k, w_2^k, \dots, w_n^k]^T$ is introduced for each 2D feature point to mitigate the impact of outliers on pose estimation results. This paper employs an improved IGG 3 function to determine the weights, which exhibits favorable robustness. The weight of the i -th feature point can be expressed as

$$w_i^k = \begin{cases} 1 & |e_i| \leq k_0 \\ \frac{k_0}{|e_i|} \left(\frac{k_1 - |e_i|}{k_1 - k_0} \right)^2 & k_0 < |e_i| < k_1 \\ 0 & |e_i| \geq k_1 \end{cases} \quad (3)$$

where k_0 and k_1 are harmonic coefficients, and e_i is the standardized residual:

$$\begin{cases} e_i = \frac{e'_i}{\sigma_0} \\ \sigma_0 = 1.4826 \text{median}(e'_i) \end{cases} \quad (4)$$

where e_i is the reprojection error for each 2D feature point, e_0 is the normalization factor, and $\text{median}(e'_i)$ is the median of e_i .

As the two-dimensional feature points are mutually independent, the joint probability density for n such points is expressed as in Equation (5). Taking the natural logarithm of Equation (5) yields the likelihood function, as shown in Equation (6). From Equation (6), it follows that maximizing the likelihood function is equivalent to minimizing the following function, thereby enabling the solution of the optimized camera pose, as depicted in Equation (7).

$$P = \prod_{i=1}^n P(P_i^{\text{img}}(k) | \pi(R^k, t^k, P_i^w(k)), w(k), \sigma^2) \quad (5)$$

$$\begin{aligned} \ln L(P_i^{\text{img}}(k) | \pi(R^k, t^k, P_i^w(k)), w(k), \sigma^2) \\ = - \sum_{i=1}^n w(k) \ln(2\pi\sigma^2) - \frac{1}{\sigma^2} \sum_{i=1}^n w(k) (P_i^{\text{img}}(k) - \pi(R^k, t^k, P_i^w(k)))^2 \end{aligned} \quad (6)$$

$$[\tilde{R}^k, \tilde{t}^k] = \min \left\{ \sum_{i=1}^n w(k) (P_i^{\text{img}}(k) - \pi(R^k, t^k, P_i^w(k)))^2 \right\} \quad (7)$$

where $\tilde{R}^k = [\tilde{r}_1^k, \tilde{r}_2^k, \tilde{r}_3^k]^T$ denotes the optimized rotation matrix and $\tilde{t}^k = [\tilde{t}_x^k, \tilde{t}_y^k, \tilde{t}_z^k]^T$ denotes the optimized translation vector. Through iteration, the optimized camera pose $\tilde{R}^k$ and $\tilde{t}^k$ can be obtained. This completes the initial pose estimation and optimization.

2.2. Two-Stage Iterative Depth Estimation Optimization

Although the optimized camera pose filters out the influence of image noise on pose estimation, it can not resolve the issue of missing depth information in monocular visual measurement. Therefore, using the optimized camera pose as the initial value, a two-stage iterative algorithm is employed to optimize the depth information. From the ideal pinhole camera model, rays emanating from the camera's optical center as the model's origin simultaneously traverse both the feature image point and the three-dimensional feature point. The unit ray vector h_i^k can be expressed as

$$h_i^k = \frac{1}{\sqrt{(u_i^k - c_x)^2 + (v_i^k - c_y)^2 + f^2}} \begin{bmatrix} u_i^k - c_x \\ v_i^k - c_y \\ f \end{bmatrix} \quad (8)$$

According to the principles of perspective projection, the coordinate of the feature point in the camera coordinate system, $P_i^c(k) = [X_i^c(k), Y_i^c(k), Z_i^c(k)]^T$, and $P_i^w(k)$ should simultaneously lie on the ray unit vector h_i^k . Their coordinates in the camera coordinate system can be expressed as

$$P_i^c(k) = d_i^k h_i^k \quad (9)$$

where d_i^k denotes the depth-of-field factor for feature point i during the k -th observation. By the collinearity principle, the collinearity equation yields

$$\begin{aligned} P_i^c(k) &= (h_i^k)^T P_i^c(k) h_i^k \\ &= (h_i^k)^T (R_{cw}^k P_i^w(k) + t_{cw}^k) h_i^k \end{aligned} \quad (10)$$

where R_{cw}^k and t_{cw}^k denote the rotation matrix and translation vector, respectively. Due to measurement errors between camera coordinates and true coordinates, the error function is formulated as

$$E(R_{cw}^k, t_{cw}^k) = \sum_{i=1}^n [(R_{cw}^k P_i^w(k) + t_{cw}^k) - (h_i^k)^T (R_{cw}^k P_i^w(k) + t_{cw}^k) h_i^k]^2 \quad (11)$$

Based on Equation (11), the rotation matrix R_{cw}^k is initialized using $\tilde{R}^k$. After fixing the rotation matrix, the partial derivative with respect to the translation vector is taken to optimize the translation vector $\tilde{t}_{cw}^k$ and the depth of field $d_{cw}^i(k)$:

$$\tilde{t}_{cw}^k = (nI - \sum_{i=1}^n h_i^k (h_i^k)^T)^{-1} \sum_{i=1}^n (h_i^k (h_i^k)^T - I) R_{cw}^k P_i^w(k) \quad (12)$$

$$d_{cw}^i(k) = (h_i^k)^T (R_{cw}^k P_i^w(k) + \tilde{t}_{cw}^k) \quad (13)$$

Using the estimated depth of field, the point cloud in the camera coordinate system is then reconstructed, yielding the reconstructed camera point cloud coordinates $\tilde{P}_i^c(k)$:

$$\tilde{P}_i^c(k) = d_{cw}^i(k) h_i^k \quad (14)$$

Given the optimized translation vector and depth, the translation vector and depth can be fixed while optimizing the rotation matrix. Due to the orthogonality of the rotation matrix, its optimization problem is transformed into a least-squares problem under orthogonal constraints. Therefore, the optimized rotation matrix $\tilde{R}_{cw}^k$ is solved using the Umeyama algorithm:

$$\tilde{R}_{cw}^k = \arg \min_{R_{cw}^k} \sum_{i=1}^n \|d_{cw}^i(k) h_i^k - (R_{cw}^k P_i^w(k) + \tilde{t}_{cw}^k)\|^2 \quad (15)$$

The error equation is constructed using the optimized pose and depth, and the maximum number of iterations is set to 100, with an iteration threshold of 0.1 mm:

$$E(R_{cw}^k, t_{cw}^k) = \sum_{i=1}^n \|d_{cw}^i(k) h_i^k - (\tilde{R}_{cw}^k P_i^w(k) + \tilde{t}_{cw}^k)\|^2 \quad (16)$$

By alternately iterating Equations (12)–(16) until convergence, the high-precision depth and camera pose are obtained.

2.3. Reproducibility Error Propagation Model

As shown in Figure 3, the repeatability positioning accuracy of a robotic arm is typically defined as the dispersion of the end-effector's pose after multiple repetitive motions. In essence, it calculates the dispersion of spatial discrete points, which can be quantified using the minimum enclosing sphere radius of the point set. When the robotic arm is equipped with a camera, since the end-effector and the camera are rigidly connected, their poses are related only by a fixed hand-eye transformation matrix. However, as repeatability positioning accuracy reflects the dispersion of poses, the camera pose can be used as a substitute for the end-effector pose in repeatability calculations. The eye-in-hand model is illustrated in Figure 4.

Current methods for monocular vision-based repeatability accuracy measurement predominantly rely on single-frame PnP solutions. These approaches statistically analyze pose distributions from multiple independent single-frame estimations to quantify robotic arm repeatability. However, due to the inherent scale ambiguity in monocular vision, these methods suffer from significantly compromised accuracy along the Z-axis. To address this fundamental limitation, we develop an error model of repetitive motion. The assumption that the robotic arm's repeatability error follows a normal distribution is widely accepted in the field [24,25]. Based on this premise, the error equation is constructed as follows:

$$s(k) = t + e(k) \quad (17)$$

where $s(k)$ represents the k -th observation, t denotes the spatial reference position that the repetitive motion attempts to reach (which remains constant regardless of the number of observations), and $e(k)$ is the error of a single observation and follows a normal distribution. The spatial reference position t can be estimated by constructing the error equation. The above formulation indicates that the observational data consists of the reference position and the observational error. Now, based on the spatial reference position t , a new hypothesis is proposed: all observations are assumed to be taken at the spatial reference position t . Under this hypothesis, the observational data containing errors $e(k)$ are back-projected to generate three-dimensional feature points, which inherently incorporate the observational errors $e(k)$. This process integrates multi-frame information to transfer the repeatability error $e(k)$ onto the point cloud, making the repeatability error $e(k)$ explicitly quantifiable through point cloud registration.

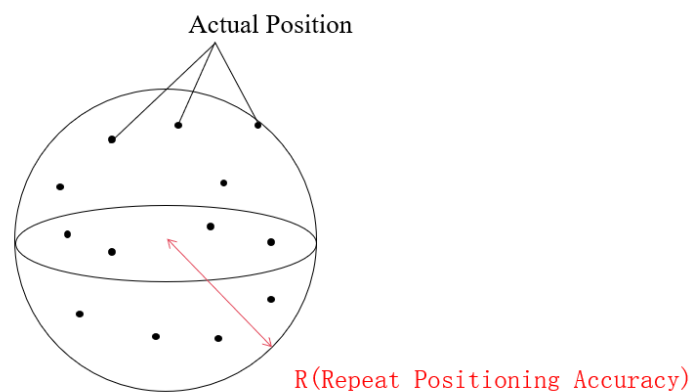


Figure 3. Repeatability positioning accuracy.

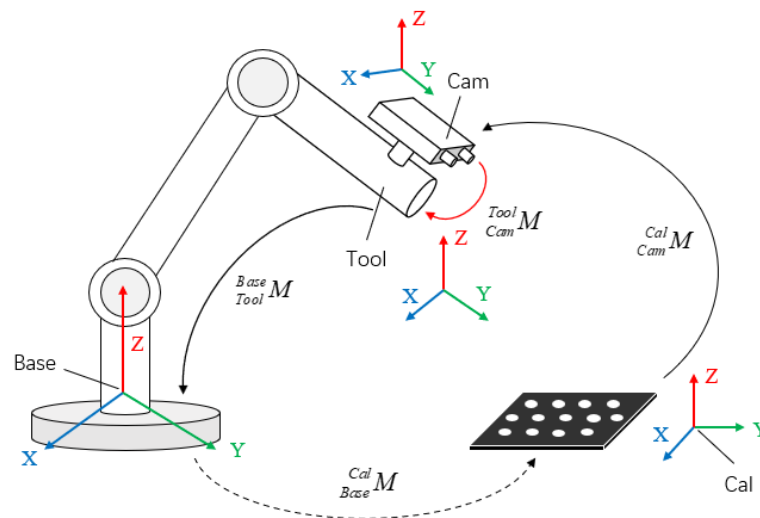


Figure 4. Eye-in-hand model.

2.4. Theoretical Pose Estimation

As the camera performs repetitive motion, under ideal conditions, the optimized camera pose $\hat{R}_{cw}^k$ for each frame should be identical to $\hat{t}_{cw}^k$. However, in practical operation, each observation by the robotic arm introduces random errors $n(k)$ due to factors such as motor control fluctuations, minute variations in transmission clearance, and environmental noise. Therefore, a single observation value can be expressed as the sum of the observed value $m(\eta)$ at the theoretical pose η and the error $n(k)$:

$$\tilde{P}_i^{img}(k) = m(\eta) + n(k) \quad (18)$$

$$\begin{cases} \tilde{P}_i^{img}(k) = \pi(\tilde{R}_{cw}^k, \tilde{t}_{cw}^k, P_i^w(k)) \\ m(\eta) = \pi(\eta, P_i^w(k)) \end{cases} \quad (19)$$

where $\tilde{P}_i^{img}(k)$ denotes the k -th observation value; $n(k)$ represents the k -th error, which follows a Gaussian distribution with mean zero and variance δ^2 . The errors from each observation are mutually independent. The theoretical pose η can be expressed as

$$\begin{cases} \eta = [R_{true}|t_{true}] \\ R_{true} = [R_1, R_2, R_3]^T \\ t_{true} = [T_x, T_y, T_z]^T \end{cases} \quad (20)$$

where the theoretical pose η comprises a rotational component R_{true} and a translational component t_{true} . Given that the theoretical pose η remains constant over K repetitive motions, while errors $n(k)$ arise from random factors, the K observations can be modeled as a random signal: a single observation value can be expressed as the sum of the observed value $m(\eta)$ at the theoretical pose η and the error $n(k)$. Since error $n(k)$ follows a Gaussian distribution, the maximum likelihood method can be employed to estimate the theoretical pose η . Considering the Gaussian nature of the error, maximizing the likelihood function is equivalent to minimizing the sum of squares of the reprojection errors between all observed values and the theoretical true values $m(\eta)$ [26]. Therefore, the global optimization problem for determining the theoretical pose η can be formulated as

$$\begin{aligned} \eta &= \arg \min_{\eta} \sum_{k=1}^K \|\tilde{P}_i^{img}(k) - m(\eta)\|^2 \\ &= \sum_{k=1}^K [\tilde{P}_i^{img}(k) - m(\eta)]^T [\tilde{P}_i^{img}(k) - m(\eta)] \end{aligned} \quad (21)$$

where $\|\tilde{P}_i^{img}(k) - m(\eta)\|$ denotes the Euclidean distance between the k -th observation $\tilde{P}_i^{img}(k)$ and the theoretical true value $m(\eta)$. Solving this nonlinear least-squares problem yields the theoretical pose η . This process effectively fuses information from K observations, mitigates the impact of random errors in individual observations, and estimates the theoretical pose η via maximum likelihood estimation.

2.5. Cramér–Rao Bound

For parameter estimation problems, the choice of estimation criterion directly determines the performance of the estimate. To understand the potential performance of parameter estimation, the Cramér–Rao bound (CRB) is adopted as the evaluation benchmark. According to optimal estimation theory, the CRB represents the lower variance limit for all unbiased estimators of a parameter and can be expressed as

$$C(\eta) \geq C_{CR}(\eta) \quad (22)$$

In the equation, $C(\eta)$ represents the variance of the parameter estimation error, and $C_{CR}(\eta)$ denotes the Cramér–Rao bound, which can be obtained by inverting the Fisher information matrix J :

$$C_{CR}(\eta) = J^{-1} \quad (23)$$

Based on the aforementioned stochastic signal model, the normalized likelihood function can be derived according to maximum likelihood theory:

$$L(\eta) = -\ln[\det(Q)] - \frac{1}{K} \sum_{k=1}^K \{[\tilde{P}_i^{img}(k) - m(\eta)]^T Q^{-1} [\tilde{P}_i^{img}(k) - m(\eta)]\} \quad (24)$$

$$Q = \delta^2 I \quad (25)$$

where $\det(\cdot)$ denotes the matrix determinant and Q is the covariance matrix. Based on the likelihood function derived from the above equation, the general form of the Fisher information matrix can be obtained as

$$\begin{aligned} [J]_{i,j} &= E\left[\frac{\partial L(\eta)}{\partial \eta_i} \frac{\partial L(\eta)}{\partial \eta_j}\right] \\ &= K\{tr[Q^{-1} \frac{\partial Q}{\partial \eta_i} Q^{-1} \frac{\partial Q}{\partial \eta_j}]\} + 2K[\frac{\partial m(\eta)}{\partial \eta_i} Q^{-1} \frac{\partial m(\eta)}{\partial \eta_j}] \end{aligned} \quad (26)$$

In the pose estimation problem, since the first term in the above equation is independent of the parameters to be estimated, and the noise is assumed to be isotropic Gaussian, with Q being the covariance matrix of the random variables, it follows that the Fisher information matrix for the pose estimation problem can be expressed as

$$[J]_{i,j} = \frac{2K}{\delta^2} \left[\frac{\partial m^T(\eta)}{\partial \eta_i} \frac{\partial m(\eta)}{\partial \eta_j} \right] \quad (27)$$

Consequently, the Cramér–Rao bound of the parameters to be estimated can be expressed as

$$C_{CR}(\eta) = \frac{\delta^2}{2K} \left[\frac{\partial m^T(\eta)}{\partial \eta_i} \frac{\partial m(\eta)}{\partial \eta_j} \right]^{-1} \quad (28)$$

As can be seen from the above equation, the Cramér–Rao bound for parameter estimation decreases linearly with the increase in the number of observations and increases linearly with the increase in noise power.

2.6. Repeatability Accuracy Calculation

In existing methods for calculating repeatability accuracy using monocular vision, the PnP algorithm is typically employed for camera pose estimation and optimization, with repeatability accuracy calculated directly from the optimized pose. Although the PnP algorithm can directly solve for poses using 2D–3D matched point pairs, the ICP algorithm exhibits lower estimation uncertainty and achieves more precise pose estimation when depth value information is precisely known [27]. Therefore, this paper introduces the ICP algorithm for calculating repeatability accuracy.

As the ICP is a point cloud registration algorithm, only the pose of each observation and the theoretical pose are currently known; therefore, a new observation model must be constructed to incorporate the ICP. This leads to the following hypothesis: assuming the poses of K observations are all theoretical poses η , if the 2D image point coordinates remain unchanged, the calibration board must be moved to ensure that the relative pose remains constant. Based on the principle of constant relative pose, it follows that the depth information will not change. Consequently, by backprojecting the observations into three-dimensional space, the coordinates of the three-dimensional feature points on the repositioned calibration plate can be obtained, yielding the corresponding point pairs. The coordinate pairs $\tilde{P}_i^w(k)$ and $P_i^w(k)$ form corresponding point pairs:

$$\tilde{P}_i^w(k) = \pi^{-1}(\tilde{P}_i^{img}(k), d_{cw}^i(k), \eta) \quad (29)$$

where $\pi^{-1}(\tilde{P}_i^{img}(k), d_{cw}^i(k), \eta)$ is the backprojection function for the ideal pinhole camera model. This transfers the robotic arm error from the arm's pose to the calibration plate's pose. The robotic arm error $n(k)$ can then be quantified by registering the corresponding point pairs using the point-to-point ICP algorithm:

$$n(k) = [\Delta R^k, \Delta t^k] = \arg \min_{\Delta R^k, \Delta t^k} \sum_{i=1}^n \|(\Delta R^k \tilde{P}_i^{w}(k) + \Delta t^k) - P_i^w(k)\|^2 \quad (30)$$

where ΔR^k denotes rotational deviation and Δt^k denotes translational deviation. The maximum number of iterations and the convergence threshold for the ICP algorithm are set to 10^5 and 10^{-8} . This transformation converts the problem of solving the pose of a stationary target PnP through multiple observations into solving the relative pose of a slightly moving target through multiple ICP-matched point clouds. If the translational deviation component Δt^k obtained from each ICP iteration is regarded as a discrete point in three-dimensional space, after K repetitions of the motion, the set of translational deviation points can be obtained:

$$\tau = \{\Delta t^1, \Delta t^2, \dots, \Delta t^K\} \quad (31)$$

By determining the radius of the minimum enveloping sphere for this point set, the repeatability accuracy of the robotic arm can be quantified.

3. Results

The experimental setup comprised a robotic arm, an industrial camera, and a circular calibration plate. The robotic arm was an Elfin-Pro E03-Pro collaborative robot featuring six axes of motion and weighing 18 kg, with a maximum payload of 3 kg and a working range of 590 mm. According to the manufacturer's datasheet, it has a repeatability positioning accuracy of ± 0.02 mm. All experiments were conducted in a temperature-controlled environment (23 ± 2 °C) to ensure that the robot reached thermal equilibrium, thereby minimizing potential accuracy drift due to temperature variations. The camera was an SVersion-HJ-RGBD-90 industrial camera with a resolution of 3120×4096 pixels. The calibration board was a CGB-020 5×4 dot calibration board, as illustrated in Figure 5. Our computer was configured with an Intel Core i7-12700K CPU and an NVIDIA GeForce RTX 3060 GPU. All image data in the experiments use pixels as the unit, while all vector data are expressed in millimeters (mm).

First, the algorithm's efficacy was validated. The algorithm validation experiment primarily assessed initial pose optimization and theoretical pose estimation. This paper employed weighted maximum likelihood to optimize the initial pose. To validate the effectiveness of the improved IGG 3 function weight-allocation strategy, optimization was conducted using both maximum likelihood (ML) and weighted maximum likelihood (WML) under identical initial pose conditions. Performance was assessed by statistically evaluating the reprojection error, with the results presented below in Table 1.

Table 1. Effectiveness analysis results.

	Minimum/Pixels	Maximum/Pixels	Mean/Pixels	Standard Deviation/Pixels
WML	0.0937	0.4940	0.2835	0.1402
ML	0.1395	3.9224	0.5072	0.6016

The experimental results demonstrate that incorporating the improved IGG 3 function weight-allocation strategy effectively reduced the reprojection error. Specifically, the maximum error of WML decreased by 87.4% compared to ML, proving that the modified

IGG 3 function dynamically adjusted weights to suppress outlier effects and mitigate the impact of inherent salt-and-pepper noise on pose estimation outcomes. The mean error and standard deviation of WML improved by 44.11% and 76.69%, respectively, compared to ML. This indicates that the introduced IGG 3 function weight-allocation strategy enables the optimization objective to focus more on high-confidence matching points, significantly enhancing the accuracy and consistency of pose estimation and demonstrating that the algorithm delivers stable gains under normal operating conditions while maintaining robustness in extreme scenarios.

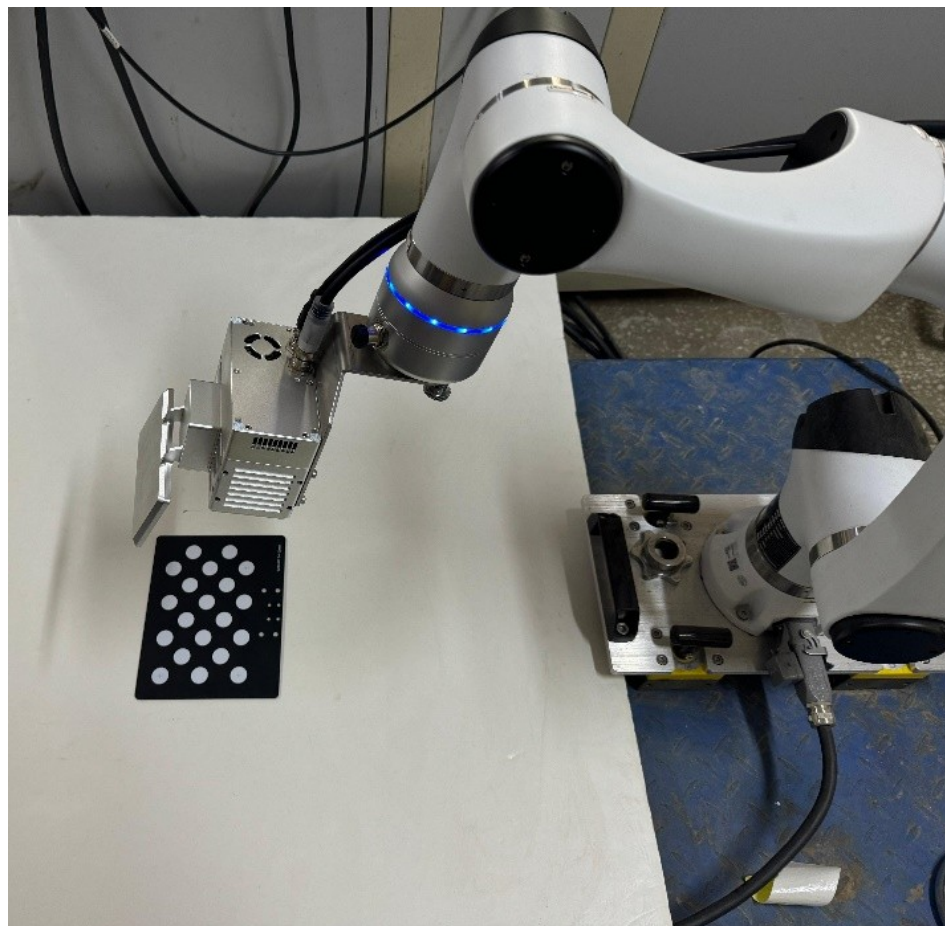


Figure 5. Experimental setup.

This paper constructed a random signal model based on the statistical characteristics of errors and employed the maximum likelihood method to estimate theoretical poses. To validate the effectiveness of the maximum likelihood estimated theoretical poses and assess whether the estimation results can serve as the system's theoretical poses, the theoretical boundary was calculated. The proposed method was compared with the Cramér–Rao bound using the mean reprojection error as the evaluation metric. The results are shown in Table 2 and Figure 6.

Table 2. Performance evaluation findings.

	1/Pixels	2/Pixels	3/Pixels	4/Pixels	5/Pixels
Ours	0.0383	0.0363	0.0417	0.0473	0.0432
Cramér–Rao bound	0.0116	0.0277	0.0346	0.0336	0.0193

Experimental results demonstrate that across five independent replicate experiments, the accuracy of the maximum likelihood global optimization algorithm approached that

of the Cramér–Rao bound. The maximum difference between this method and the Cramér–Rao bound was 0.0267 pixels, with the minimum difference being 0.0086 pixels. Both differences remained within 0.03 pixels, proving that the maximum likelihood method based on the random signal model can effectively estimate theoretical poses.

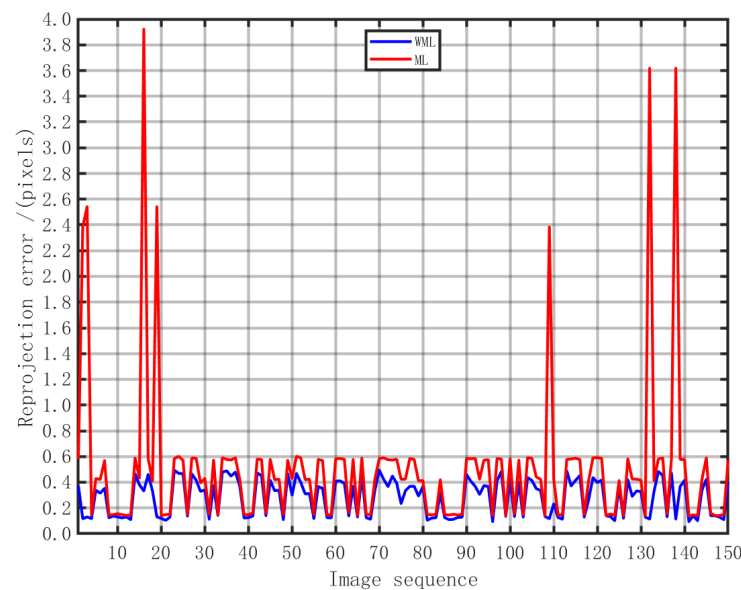


Figure 6. Error comparison chart.

Second, simulation experiments analyzed factors influencing the pose estimation module algorithm. As the initial pose estimation employed the PnP algorithm, to identify key factors affecting the initial pose optimization algorithm, the proposed method was compared with the classical PnP improved algorithm [10–14] and the orthogonal iteration algorithm [19]. In the experiments, a scenario with known camera intrinsic parameters was simulated. The synthetic virtual camera resolution was assumed to be 640×480 pixels with a fixed focal length of 800 pixels. Three-dimensional feature points were randomly generated, distributed on a plane with coordinates ranging from $[-2, 2] \text{ m} \times [-2, 2] \text{ m}$. Two-dimensional feature points were projected using an ideal pinhole camera model. Employing a controlled variable method, the number of reference points N and noise level ω were analyzed as study variables. For each variable value, 500 simulation experiments were conducted. Algorithm performance was evaluated by calculating the mean and median of the rotational and translational errors. The variable parameter settings are detailed in Table 3, with the experimental results presented in Figures 7 and 8.

Table 3. Variable parameter settings.

	Minimum	Maximum	Step
N	4	20	1
ω/pixels	0	5	0.5

The experimental results demonstrate that the error of all algorithms decreased as the number of reference points increased, while it rose with increasing noise levels. Furthermore, at a fixed noise level, accuracy gradually converged once the number of reference points reached a certain threshold. As EPnP is a non-iterative linear algorithm, it is sensitive to noise, making it difficult to guarantee accuracy when few feature points are available. DLS disregards geometric constraints, making it prone to singularities during computation and resulting in overall instability. Although OI is an iterative method, its accuracy and

stability deteriorated when reference points were scarce or near singularity, frequently trapping it in local optima. The RPnP, OPnP, and SRPnP algorithms all exhibited high precision; however, as RPnP is a suboptimal algorithm, its accuracy was marginally lower than that of OPnP and SRPnP. In contrast, the proposed method enhances accuracy by using a weighted maximum likelihood approach to optimize initial poses, improving the extraction precision of two-dimensional feature points. It constructs a random signal model and employs maximum likelihood estimation for theoretical pose estimation, thereby preventing estimation results from becoming trapped in local optima. Consequently, it outperformed the comparison methods in both accuracy and stability.

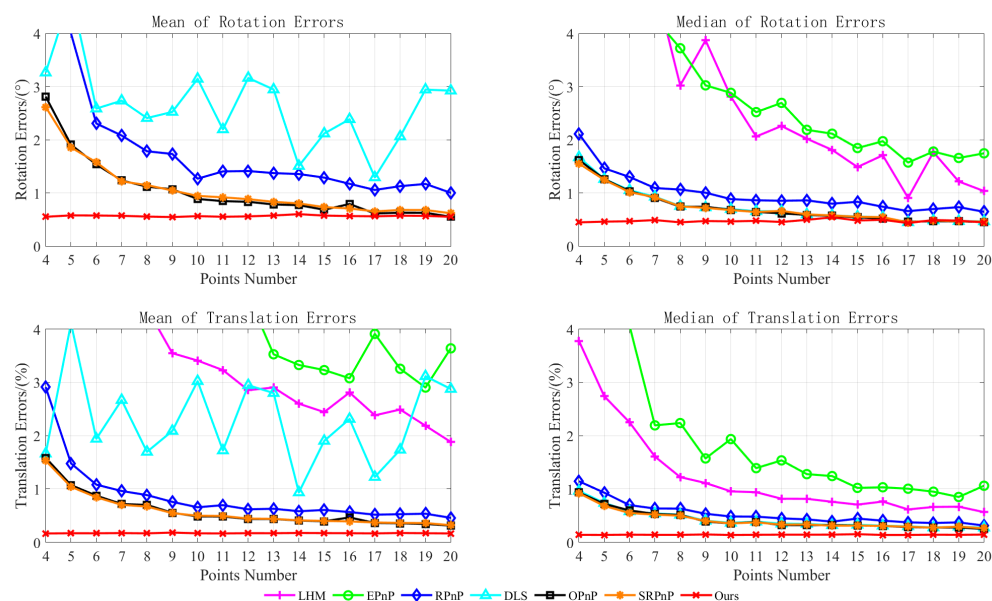


Figure 7. Variations in reference point count.

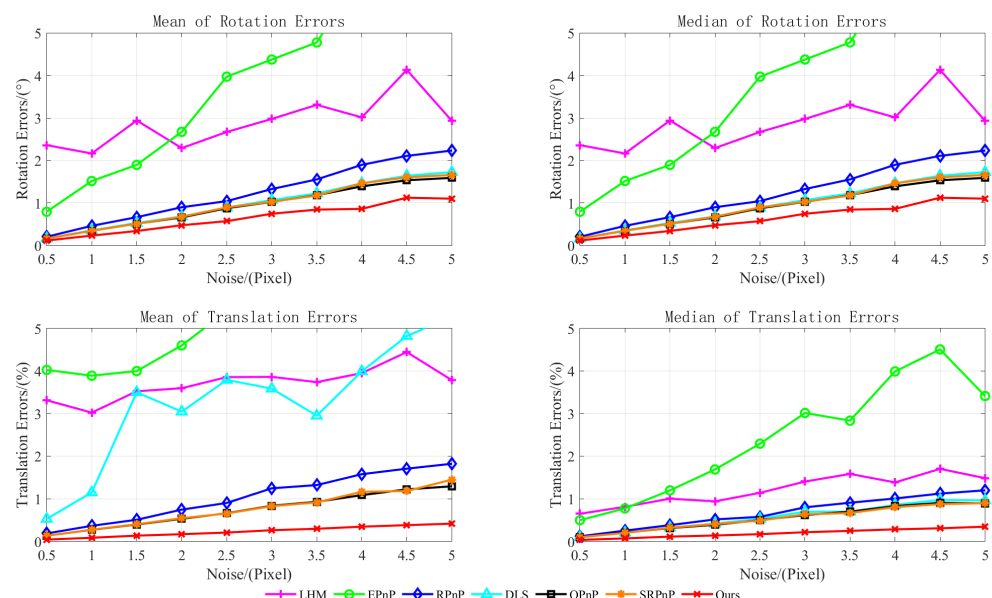


Figure 8. Variations in noise level.

Finally, a comparative analysis was conducted. This analysis primarily encompassed module comparisons and comparisons with similar methods. Within the robotic arm's workspace, an industrial camera mounted on the robotic arm captured 150 calibration plate images. A portion of the dataset images is shown in Figure 9.



Figure 9. Calibration plate image data.

The above images represent calibration plate photographs captured under five different poses: P1–P5. To quantitatively evaluate the impact of each module’s algorithm on the final repeatability measurement accuracy, a comparative experiment was designed using 30 datasets from the P1 pose. The comparison schemes were as follows:

Scheme A: Basic PnP (unoptimized);

Scheme B: A + WML (initial optimization);

Scheme C: B + two-stage iteration + ML (global optimization);

Scheme D: C + ICP (proposed method).

Scheme A performs the PnP solution directly after data acquisition, using the first frame’s pose as the baseline to compute the relative positional relationships for subsequent frames, thereby determining repeatability accuracy. Scheme B optimizes the PnP solution from Scheme A before calculating relative positions and determining repeatability accuracy. Scheme C builds upon Scheme B by employing a two-stage iterative algorithm to optimize depth and pose. It uses a maximum likelihood method based on a random signal model to estimate theoretical poses. Using these theoretical poses as reference, it calculates each frame’s relative position to the theoretical pose to determine repeatability accuracy. Scheme D represents the method proposed in this paper. Building upon Scheme C, it constructs a three-dimensional homographic point cloud based on the theoretical pose and the optimized depth field. This enables the introduction of the ICP algorithm to compute repeatability accuracy. The specific comparative experimental results are presented in Table 4.

Table 4. Module comparative analysis results.

	A	B	C	D
Repeatability/mm	17.9465	1.2730	0.9792	0.0115

The comparative experimental results demonstrated that the proposed method achieved repeatability accuracy of 0.0115 mm, significantly outperforming alternative approaches. The specific analysis was as follows: Scheme A exhibited repeatability accuracy of 17.9465 mm, markedly higher than that of other approaches. This is primarily because single-frame PnP solutions are susceptible to observational noise interference, leading to substantial positional estimation deviations and poor stability. If the inaccurate single-frame pose from the initial frame is directly used as the reference for calculating relative poses, the relative poses incorporate gross errors, thereby causing repeatability accuracy measurements to fail. Following the introduction of WML optimization in Scheme B, repeatability accuracy improved by 93%. Although this scheme effectively mitigated the impact of gross errors on the results by modeling the statistical characteristics of observational noise, thereby significantly enhancing single-frame pose stability, it failed to resolve the inaccuracy in the depth calculation under the coplanar configuration of monocular visual features. This compromised the accuracy of the three-dimensional repeatability results. Scheme C builds upon Scheme B by employing a two-stage iterative algorithm. It optimizes depth and relative pose based on collinearity constraints and estimates the theoretical pose via a maximum likelihood method grounded in a random signal model. This

resolved depth estimation inaccuracies and enhanced repeatability accuracy. However, PnP methods reliant solely on 2D–3D matching face inherent limitations. The proposed method incorporates the ICP algorithm into Scheme C, achieving significantly higher accuracy than that scheme. This demonstrates that the ICP registration scheme for homonymous point clouds—based on the theoretical pose η and the optimized depth field backprojection—effectively overcomes the projection error sensitivity inherent in 2D–3D matching, yielding more precise repeatability accuracy.

To demonstrate the superiority of the proposed method, we conducted a comparative analysis against existing monocular vision-based repeatability accuracy calculation methods. Given identical input image data, we compared the repeatability accuracy results of our method with those from [16,18]. The specific repeatability accuracy calculations and corresponding error analysis are presented below.

Through the analysis of the data in Tables 5 and 6, the method proposed in this paper significantly outperformed the comparative methods in repetitive positioning accuracy measurement, with measurement values within 0.0175 mm and a mean repeatability positioning accuracy of 0.0128 mm, which is on the same order of magnitude as the factory-calibrated accuracy (± 0.02 mm) of the test robot, indicating that the proposed method has minimal measurement error and is capable of accurately capturing the true performance of the robotic arm. Further analysis of error statistics indicates that the proposed method surpassed the comparative methods in both mean error and standard deviation metrics, demonstrating small error fluctuations in the measurement results and no significant outlier error points. Specifically, the method in [16], based on the PnP algorithm, is sensitive to feature point distribution and calculates repetitive positioning accuracy through single-frame pose estimation results. Although it optimizes the extracted 2D feature points during image point extraction to avoid the influence of gross errors, resulting in relatively good stability, it does not consider the impact of depth estimation errors on the results. Since this algorithm is similar to Scheme B, its accuracy can only be controlled at the millimeter level. While the PnP + OI combined method proposed in [18] improves overall accuracy through nonlinear optimization, it relies on 2D–3D matching-based PnP methods that possess inherent limitations. In this paper, an improved IGG3 weighting function is adopted to suppress feature point outliers, a two-stage iterative strategy is employed to optimize depth and pose, and a maximum likelihood method based on a random signal model is constructed to estimate the theoretical pose. The theoretical pose is then used to backproject 3D points as observations, combined with the ICP algorithm to break through the accuracy limits of 2D–3D matching. The introduction of the ICP algorithm indeed increased computational cost by an average of 8 s, as reported in Table 6. This trade-off is justified through an analysis of algorithmic complexity and the fundamental objectives of this work. The complexity of the ICP algorithm is dominated by the nearest-neighbor search. Given point clouds of size N and M , the per-iteration complexity of ICP using a KD-Tree is $O(N \log M)$, leading to an overall complexity of $O(kN \log M)$ for k iterations, which is a manageable polynomial-time complexity. Considering that the primary objective of this study is high-accuracy 3D repeatability assessment (as opposed to high-speed online control), and that traditional monocular methods suffer from inherent depth ambiguity, the fixed computational cost is warranted. The ICP registration elevates the measurement from the 2D image space to a metric 3D space, fundamentally resolving the scale ambiguity issue. This is considered a critical advancement for measurement tasks where precision is paramount.

Table 5. Repeatability positioning accuracy calculation results.

	P1/mm	P2/mm	P3/mm	P4/mm	P5/mm
[16]	1.0569	1.0898	1.0303	1.0702	1.1041
[18]	0.8049	0.8088	0.7567	0.8107	0.8331
Our method	0.0115	0.0121	0.0068	0.0162	0.0175

Table 6. Error and computational efficiency analysis.

	Mean/mm	Standard Deviation/mm	Time Cost/s
[16]	1.0703	0.0257	3.3454
[18]	0.8028	0.0286	7.1082
Our method	0.0128	0.0038	11.1390

4. Conclusions

This paper addresses critical issues in monocular vision-based robotic arm repeatability measurement, including missing statistical assumptions in observational data, non-global optimization of pose estimation, and insufficient depth estimation accuracy. It proposes a quantitative method based on likelihood pose estimation. By optimizing single-frame poses from PnP estimates, constructing a random signal model, and employing the approach to fuse multi-frame observations for theoretical pose estimation, this approach overcomes the accuracy limitations of traditional 2D–3D matching through optimized depth information and ICP algorithms. Experimental validation demonstrates that the proposed method achieves repeatability accuracy consistently within 0.0068–0.0175 mm, with a standard deviation of 0.0038 mm. This represents improvements of 85% and 87% over comparative methods, offering enhanced precision and stability. It provides novel theoretical and methodological pathways for applying visual measurement techniques in robotic arm performance evaluation. Although the proposed method performs well in controlled environments with stable lighting and no occlusions, variable illumination and partial occlusion in industrial settings remain critical practical challenges. In future work, we will focus on enhancing the robustness of the approach. Specific directions include investigating illumination-invariant feature descriptors and developing robust estimation algorithms capable of identifying and rejecting outlier observations caused by occlusions. The ultimate goal is to adapt this measurement system for deployment in more complex industrial field environments.

Author Contributions: Conceptualization, P.Z. and J.L. (Jiatian Li); Methodology, P.Z. and J.L. (Jiatian Li); Software, P.Z.; Validation, P.Z. and Y.J.; Writing—original draft, P.Z.; Writing—review and editing, P.Z., J.L. (Jiatian Li), J.L. (Jiayin Liu) and F.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China under grant number 41561082.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in the study are included in the article; further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Slamani, M.; Nubiola, A.; Bonev, I. Assessment of the positioning performance of an industrial robot. *Ind. Robot. Int. J.* **2012**, *39*, 57–68. [\[CrossRef\]](#)
- Deng, F.; Li, K.; Hu, X.; Jiang, H.; Huang, F. Life calculation of angular contact ball bearings for industrial robot RV reducer. *Ind. Lubr. Tribol.* **2019**, *71*, 826–831. [\[CrossRef\]](#)
- Yu, Z.; Wan, J.; Hao, Z.; Kou, L. Measuring the pose repeatability accuracy of the industrial robot end-effector based on the ISSA-IGCF-IHT method. *Meas. Sci. Technol.* **2024**, *35*, 115022. [\[CrossRef\]](#)
- Ma, N.; Su, C.; Zhang, X. Research on fast field calibration method of industrial robot. In Proceedings of the 2015 Chinese Automation Congress (CAC), Wuhan, China, 27–29 November 2015; IEEE: New York, NY, USA, 2015; pp. 443–447.
- Tian, W.; Xu, M.; Zhang, X.; Guo, X.; Wang, L.; Huang, T. Repeatability prediction of 6-DOF hybrid robot based on equivalent error model of actuated joint. *Measurement* **2023**, *207*, 112377. [\[CrossRef\]](#)
- Yin, F.; Wang, L.; Tian, W.; Zhang, X. Kinematic calibration of a 5-DOF hybrid machining robot using an extended Kalman filter method. *Precis. Eng.* **2023**, *79*, 86–93. [\[CrossRef\]](#)
- Wang, Z.; Zhang, R.; Keogh, P. Real-time laser tracker compensation of robotic drilling and machining. *J. Manuf. Mater. Process.* **2020**, *4*, 79. [\[CrossRef\]](#)
- Boby, R.A.; Saha, S.K. Single image based camera calibration and pose estimation of the end-effector of a robot. In Proceedings of the 2016 IEEE International Conference on Robotics and Automation (ICRA), Stockholm, Sweden, 16–21 May 2016; IEEE: New York, NY, USA, 2016; pp. 2435–2440.
- Wu, L.; Deng, X.; Wang, Y.; Du, X.; Xiong, X.; Jiang, B. Research on vision-based robot planar motion measurement method. *J. Braz. Soc. Mech. Sci. Eng.* **2023**, *45*, 216. [\[CrossRef\]](#)
- Lepetit, V.; Moreno-Noguer, F.; Fua, P. EP n P: An accurate O (n) solution to the P n P problem. *Int. J. Comput. Vis.* **2009**, *81*, 155–166. [\[CrossRef\]](#)
- Li, S.; Xu, C.; Xie, M. A robust O (n) solution to the perspective-n-point problem. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *34*, 1444–1450. [\[CrossRef\]](#) [\[PubMed\]](#)
- Hesch, J.A.; Roumeliotis, S.I. A direct least-squares (DLS) method for PnP. In Proceedings of the 2011 International Conference on Computer Vision, Barcelona, Spain, 6–13 November 2011; IEEE: New York, NY, USA, 2011; pp. 383–390.
- Zheng, Y.; Kuang, Y.; Sugimoto, S.; Astrom, K.; Okutomi, M. Revisiting the pnp problem: A fast, general and optimal solution. In Proceedings of the IEEE International Conference on Computer Vision, Sydney, Australia, 1–8 December 2013; pp. 2344–2351.
- Wang, P.; Xu, G.; Cheng, Y.; Yu, Q. A simple, robust and fast method for the perspective-n-point problem. *Pattern Recognit. Lett.* **2018**, *108*, 31–37. [\[CrossRef\]](#)
- Chen, H.; Wang, P.; Wang, F.; Tian, W.; Xiong, L.; Li, H. Epro-pnp: Generalized end-to-end probabilistic perspective-n-points for monocular object pose estimation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *47*, 2781–2790.
- Gao, H.; Zhang, H.; Jiang, Y.; Sun, J.; Yu, J. A localization method of manipulator towards achieving more precision control. *Comput. Intell.* **2024**, *40*, e12600. [\[CrossRef\]](#)
- Luo, Z.; Cai, Z.; Jiang, W.; Yang, L. Endoscopic Attitude Measurement Method for Industrial Robot Based on Monocular Vision. *Modul. Mach. Tool Autom. Manuf. Tech.* **2020**, *11*, 60–64.
- Chen, G.; Xu, D.; Yang, P. High precision pose measurement for humanoid robot based on PnP and OI algorithms. In Proceedings of the 2010 IEEE International Conference on Robotics and Biomimetics, Tianjin, China, 14–18 December 2010; IEEE: New York, NY, USA, 2010; pp. 620–624.
- Lu, C.P.; Hager, G.D.; Mjolsness, E. Fast and globally convergent pose estimation from video images. *IEEE Trans. Pattern Anal. Mach. Intell.* **2000**, *22*, 610–622. [\[CrossRef\]](#)
- Zhang, Z. Iterative point matching for registration of free-form curves and surfaces. *Int. J. Comput. Vis.* **1994**, *13*, 119–152. [\[CrossRef\]](#)
- ISO 9283; Manipulating Industrial Robots: Performance Criteria and Related Testing Methods. ISO: Geneva, Switzerland, 1998.
- Leutenegger, S.; Lynen, S.; Bosse, M.; Siegwart, R.; Furgale, P. Keyframe-based visual-inertial odometry using nonlinear optimization. *Int. J. Robot. Res.* **2015**, *34*, 314–334. [\[CrossRef\]](#)
- Mourikis, A.I.; Roumeliotis, S.I. A multi-state constraint Kalman filter for vision-aided inertial navigation. In Proceedings 2007 IEEE International Conference on Robotics and Automation, Rome, Italy, 10–14 April 2007; IEEE: New York, NY, USA, 2007; pp. 3565–3572.
- Zhang, D.; Liang, H.; Li, X.a.; Jia, X.; Wang, F. Kinematic calibration of industrial robot using Bayesian modeling framework. *Reliab. Eng. Syst. Saf.* **2025**, *253*, 110543. [\[CrossRef\]](#)
- Nubiola, A.; Bonev, I.A. Absolute calibration of an ABB IRB 1600 robot using a laser tracker. *Robot. Comput.-Integr. Manuf.* **2013**, *29*, 236–245. [\[CrossRef\]](#)

-
26. Qu, Y.; Zhang, C.; Liu, Y. Maximum likelihood pose estimation using machine vision. *J. Appl. Opt.* **2019**, *40*, 253–258. [[CrossRef](#)]
 27. Ma, X.; Liang, X. Analysis on pose estimation uncertainty and observation parametrization for RGB-D cameras. *Robot* **2021**, *43*, 54–65.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.