


RESEARCH ARTICLE

Increased plane identification precision with stereo identification

Junjie Ji and Jing-Shan Zhao 

State Key Laboratory of Tribology, Department of Mechanical Engineering, Tsinghua University, Beijing, 100084, P. R. China
Corresponding author: Jing-Shan Zhao; Email: jingshanzhao@mail.tsinghua.edu.cn

Received: 1 December 2022; **Revised:** 25 March 2023; **Accepted:** 25 April 2023; **First published online:** 19 June 2023

Keywords: stereo camera; plane parameters identification; calibration; region of interests

Abstract

Stereo vision allows machines to perceive their surroundings, with plane identification serving as a crucial aspect of perception. The accuracy of identification constrains the applicability of stereo systems. Some stereo vision cameras are cost-effective, compact, and user-friendly, resulting in widespread use in engineering applications. However, identification errors limit their effectiveness in quantitative scenarios. While certain calibration methods enhance identification accuracy using camera distortion models, they rely on specific models tailored to a camera's unique structure. This article presents a calibration method that is not dependent on any particular distortion model, capable of correcting plane position and orientation identified by any algorithm, provided that the identification error is biased. A high-precision mechanical calibration platform is designed to acquire accurate calibration data while using the same detected material in real measurement scenarios. Experimental comparisons confirm the efficacy of plane pose correction on PCL-RANSAC, with the average relative error of distance reduced by 5.4 times and the average absolute error of angle decreasing by 41.2%.

1. Introduction

Identification of position and orientation parameters of specific objects is a crucial task in machine vision. 3D cameras generate point clouds containing pose information of structured objects, with measurement and character identification being the primary functions of 3D vision.

Stereo vision, a type of 3D imaging system, captures information simultaneously from different views, containing 3D object information. Stereo vision styles include monocular vision [1–4], binocular vision [5–9], and multiview stereo vision [10–12], the latter being a more complex binocular variant.

RGB-D cameras, typically based on binocular vision principles [13, 14], are widely used in robots [15, 16], drones [17, 18], industrial production [19, 20] and more. Although their refresh rate is sometimes lower than LiDAR, RGB-D systems offer color information, providing additional scene details while remaining low-cost and easy to set up.

Construction robots often encounter plane identification problems, such as those found in wall painting robots [21], floor surface profiling robots [22], ground plane detection [23–25], and surface reconstruction [26, 27]. Methods for plane extraction and parameter identification include random sample consensus (RANSAC) and its variants. PCL-RANSAC, a RANSAC-based method [28] offers a mature and stable plane identification algorithm [26].

In quantitative applications, identification precision is of utmost importance. Some measurements of RGB-D cameras are known to have biased errors [29–32], but a priori information can be adopted to increase accuracy [33]. A priori information is obtained from other sources but not by current measuring instruments.

Calibration methods collect a priori information for reducing the measurement error. Zhang [34] reported a camera calibration method that models radial lens distortion. Darwish *et al.* [35] proposed

a method to calibrate each error source of the RGB-D camera. Li *et al.* [36] introduced a calibration method for plane fitting by constructing a plane fitting error model on an RGB-D system. Feng *et al.* [37] proposed a high-precision method for identifying petroleum pipeline interfaces by using camera calibration. Fuersattel *et al.* [38] presented a calibration algorithm based on the least squares method to increase the plane fitting precision. However, these calibration methods rely on specific error models and do not use the same detected objects in application, leading to potential discrepancies between calibration and application environments that may cause unpredictable errors.

Interpolation methods can effectively fit unknown models if sample points are accurate. Mechanical measurement methods are typically reliable and precise, with high-precision encoders ensuring platform accuracy. However, these methods often depend on a particular camera distortion model to obtain sufficiently accurate sample points. This article proposes an interpolation-based calibration method that is independent of any specific distortion models, addressing existing limitations. The method involves gathering accurate pose mapping rules offline and then applying these relations for pose correction during online use. Although initial pre-gathered relations are discrete in pose space, a continuous mapping relation is formed using interpolation. The method is ultimately applied to a construction robot to validate its improved precision in brick placement on a wall.

2. Calibration method proposition

2.1. Overview of plane pose correction

Suppose there is a bad-plane-pose space, which includes low-precision plane poses identified by a general algorithm from a stereo system, and a fine-plane-pose space, which includes all accurate poses of the actual plane. The main idea of the proposed calibration method is to find a single mapping from the bad-plane-pose space to the fine-plane-pose space. Furthermore, for generality, the method should accommodate situations where pose errors are irregular. Here, the pose error represents the difference between the plane pose in the bad-plane-pose space and its corresponding pose in the fine-plane-pose space.

Errors can be approximately regular within local subspaces, while global errors are irregular. Interpolation equations are derived within local subspaces to build the mapping from the initial plane pose to the accurate plane pose. A segmentation strategy is proposed for the entire space to obtain a group of subspaces.

Accurate interpolation points are required for building precise mapping relations. A high-precision mechanical platform is designed for gathering accurate data, and its geometry is analyzed. Accurate plane poses are obtained using a region-of-interest method that takes the geometry of the mechanical platform into account.

The flowchart of applying the proposed method is shown in Fig. 1.

The method comprises two parts: offline calibration and online application, as shown in Fig. 1. In offline calibration, the camera to be used in the online scenario is fixed on the calibration platform. The platform can adjust the relative plane pose between the camera and the plane sample. By pairing initial plane poses and accurate plane poses, mapping relations are formed as a result of the offline calibration. In the online application, after obtaining the initial plane pose using conventional methods, the mapping relations correct the pose to a high-precision one. The procedures to obtain the initial plane pose for both offline and online identification should be the same.

2.2. Mapping from a bad-plane-pose space to the fine-plane-pose space

2.2.1. Plane pose mapping relations

In this article, the issue of plane detection is examined within the camera coordinate system, as illustrated in Fig. 2. The camera coordinate system is constituted by a right-handed Cartesian coordinate system, with its origin O_C situated at the camera's center. The z -axis extends from O_C towards the scene, while

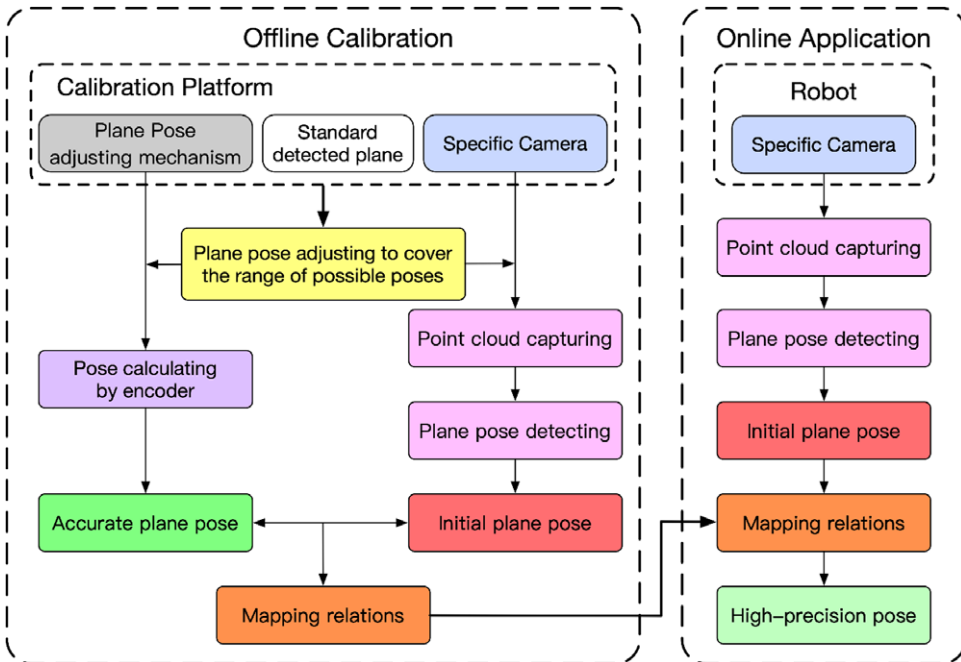


Figure 1. Flowchart of using calibration methodology.

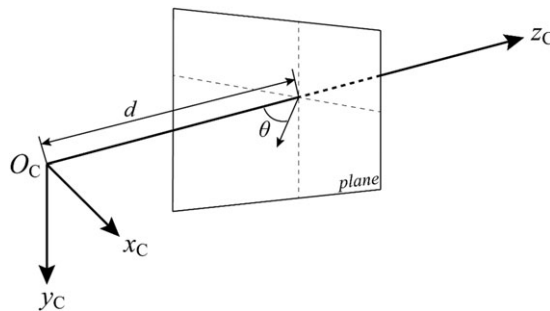


Figure 2. Camera coordinate system and the detected plane.

x -axis extends from O_C to the right, parallel to the camera’s horizontal direction. The y -axis extends from O_C downwards, parallel to the camera’s vertical direction.

Suppose there is a point situated within the scene, and the vision system provides an estimation of this point’s position. This estimation is inherently biased and contains a systematic error. Eq. (1) represents the biased estimation for an arbitrary point.

$$E_p(\hat{p}) = p + \Delta p \tag{1}$$

where p signifies the point’s coordinates, \hat{p} represents an estimate of p , $E_p(\hat{p})$ denotes the expected value of the estimation of \hat{p} . As the estimate is biased, the expectation is equivalent to the summation of the true value and an offset Δp .

Likewise, it is postulated that the estimate of the plane’s pose parameter is biased as well, as shown in Eq. (2).

$$E_c(\hat{c}) = c + \Delta c \tag{2}$$

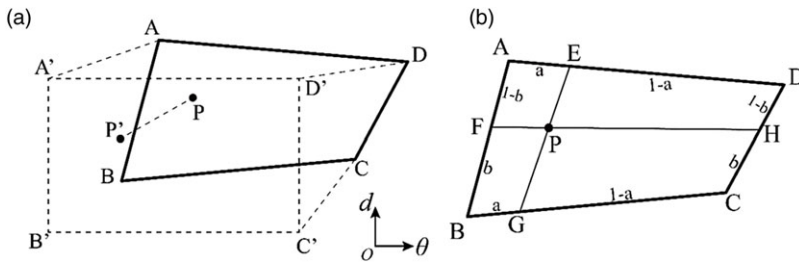


Figure 3. Mapping from the bad-plane-pose space to the fine-plane-pose space.

where c represents the plane’s coefficients vector, while \hat{c} denotes the vector’s estimation. The calibration’s objective is to identify the offset Δc to obtain an accurate estimation $E_c(\hat{c})$.

Figure 2 displays a plane within the camera coordinate system. The plane’s pose, encompassing both position and orientation, is parameterized by the distance, d , and inclination angle θ . The distance d is defined as the length between the camera’s center and the plane along the z -axis. The inclination angle θ is defined as the angle between the plane’s normal vector and the z -axis.

Eq. (3) expresses the plane’s equation.

$$a_1x + a_2y + a_3z + a_4 = 0 \tag{3}$$

The defined distance defined can be expressed by Eq. (4).

$$d = -\frac{a_4}{a_3} \tag{4}$$

The defined inclination angle can be expressed by Eq. (5)

$$\theta = \arccos\left(\frac{a_3}{\sqrt{a_1^2 + a_2^2 + a_3^2}}\right) \tag{5}$$

The plane’s pose can be described by an ordered pair (θ, d) , which corresponds to coordinates in a two-dimensional orthogonal coordinate system. In this system, one dimension is the inclination angle θ , and the other is the distance d . The calibration’s objective is to establish a function concerning the plane’s pose, which can be expressed by Eq. (6).

$$(\hat{\theta}, \hat{d}) = f[(\theta, d)] \tag{6}$$

In Eq. (6), the pose coordinates (θ, d) indicate the initial pose in the bad-plane-pose space while $(\hat{\theta}, \hat{d})$ represent the higher precision plane pose estimation in the fine-plane-pose space. The function f should be an injective function, which means that for any $f(\theta_1, d_1) = f(\theta_2, d_2)$, there is $\theta_1 = \theta_2, d_1 = d_2$.

To formulate the function f , one must gather sufficient discrete mapping relations and employ an interpolation method to constitute a continuous mapping function from the bad-plane-pose space to the fine-plane-pose space.

2.2.2. Interpolation method for establishing mapping relations

The gathered discrete mapping relations comprise a set of initial poses, $\{(\theta_i, d_i)\}$, and a corresponding set of accurate poses $\{(\theta_i^*, d_i^*)\}$. Figure 3(a) shows four gathered initial poses, $A(\theta_1, d_1) B(\theta_2, d_2) C(\theta_3, d_3) D(\theta_4, d_4)$, alongside their respective accurate poses $A'(\theta_1^*, d_1^*) B'(\theta_2^*, d_2^*) C'(\theta_3^*, d_3^*) D'(\theta_4^*, d_4^*)$. The point P symbolizes the pose (θ_p, d_p) acquired in real-time within a bad-plane-pose space. The corrected pose estimation $P'(\hat{\theta}_p, \hat{d}_p)$ in the fine-plane-pose space can be determined as follows.

The corrected pose $P'(\hat{\theta}_p, \hat{d}_p)$ and the initial pose $P(\theta_p, d_p)$ are interconnected by the intermediate variables, a and b . As depicted in Fig. 3(b), the points E, G lie on segments AD and BC , respectively,

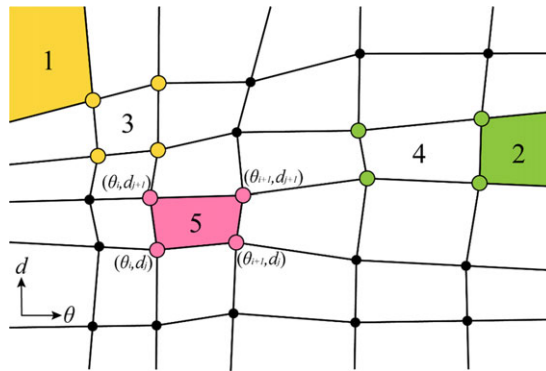


Figure 4. Entire pose space composing all mapped subspaces.

dividing the segments AD, BC with the same ratio a . Similarly, points F, H are situated on the segments BA and CD , dividing these segments with the same ratio b . Segments EG and FH intersect at point P . By performing the same operation within the quadrilateral $A'B'C'D'$ using the same ratios a and b , then the intersection point P' signifies the corrected pose. The intermediate variables, a and b , can be obtained according to Eq. (7)

$$[\theta_p \ d_p] = [1-a \ a] \begin{bmatrix} b & 1-b & 0 & 0 \\ 0 & 0 & 1-b & b \end{bmatrix} \begin{bmatrix} \theta_1 & d_1 \\ \theta_2 & d_2 \\ \theta_3 & d_3 \\ \theta_4 & d_4 \end{bmatrix} \tag{7}$$

Subsequently, substituting the ratios a and b into the Eq. (7) yields Eq. (8).

$$[\hat{\theta}_p \ \hat{d}_p] = [1-a \ a] \begin{bmatrix} b & 1-b & 0 & 0 \\ 0 & 0 & 1-b & b \end{bmatrix} \begin{bmatrix} \theta_1^* & d_1^* \\ \theta_2^* & d_2^* \\ \theta_3^* & d_3^* \\ \theta_4^* & d_4^* \end{bmatrix} \tag{8}$$

The estimation of parameter pair with enhanced precision $(\hat{\theta}_p, \hat{d}_p)$ is now ascertainable according to the Eqs. (7) and (8).

2.2.3. Plane pose space partitioning strategy

According to Eqs. (7) and (8), a minimum of four pairs of gathered poses are necessary to establish a local mapping, as illustrated in Fig. 3. The entire plane pose space comprises a number of local subspaces, as demonstrated in Fig. 4.

Figure 4 presents the abstract diagram of the bad-plane-pose space, where solid points represent the gathered initial pose data $\{(\theta_i, d_i)\}$. By connecting neighbored points, a series of quadrilaterals is formed. The distortion of quadrilateral reflects the irregular initial pose identification error. Each quadrilateral corresponds to a single interpolation mapping function.

In Fig. 4, some subspaces are surrounded by less than four vertices and are denoted by regions marked with numbers 1 and 2, referred to as corner subspaces and edge subspaces, respectively. Excluding edge and corner subspaces, the remainder is designated as internal subspaces. The mapped functions of edge and corner subspaces can be constituted by their nearest internal subspaces, represented by subspaces marked with numbers 3 and 4 in Fig. 4.

Assuming the number of the gathered points matrix in the row is n_d , and the number of the gathered points matrix in the column is n_θ . The entire $\theta - d$ space is divided into $(n_\theta + 1) \times (n_d + 1)$ grids. This

including four corner subspaces, $2 \times (n_\theta - 1 + n_d - 1)$ edge subspaces, and $(n_\theta - 1) \times (n_d - 1)$ internal subspaces.

The method for identifying the corresponding subspace containing the to-be-corrected pose $P(\theta_p, d_p)$ is as follows. Convert the initial pose $P(\theta_p, d_p)$ into the homogeneous vector $\bar{P} = [\theta_p \ d_p \ 1]^T$. A judgment vector $J = [J_1 \ J_2 \ J_3 \ J_4]^T$ is defined by Eq. (9)

$$J = Q\bar{P} \quad (9)$$

where Q denotes a representative matrix for a quadrilateral, defined by Eq. (10).

$$Q = \begin{bmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \\ a_4 & b_4 & c_4 \end{bmatrix} \quad (10)$$

Each row of the representative matrix Q in Eq. (10) signifies the coefficients of the line equation, representing one edge of the corresponding quadrilateral subspace in the pose coordinate system, as shown in Eq. (11).

$$a_i\theta + b_id + c_i = 0 \quad (11)$$

The initial pose point $P(\theta_p, d_p)$ is located within the quadrilateral subspace only when each element of the judgment vector J is positive. Upon determining the inclusion relationship between an initial pose $P(\theta_p, d_p)$ and its corresponding subspace is obtained, the corrected pose estimation $P'(\hat{\theta}_p, \hat{d}_p)$ can be ascertained according to Eqs. (7) and (8).

2.3. Calibration platform and geometry analysis

A high-precision mechanical calibration platform is designed to gather high-quality mapping relations. The detected object sample and the camera device are affixed to the platform, which adjusts the relative pose between the camera and the plane. A to-be-corrected relative pose estimated by the vision system and an accurate relative pose supplied by the platform constitutes one pose mapping relation. The platform alters the relative poses to encompass the entire range of relative poses as extensively as possible.

The calibration platform, as depicted in Fig. 5, is a three-degree-of-freedom platform with two translational degrees and one rotational degree. The rotation degree facilitates changing the detected plane's inclination angle θ while the horizontal translational degree adjusts the distance d between the detected plane and the camera. The supplementary vertical translation degree is employed solely for adjusting the camera's height during initial platform assembly. High-precision encoders on the platform guarantee the accuracy of the acquired relative poses, with the camera situated on the vertical slide.

To attain heightened calibration precision, the target plane's pose must be precisely known. Ideally, the sample had better be the actual detected object in the applied scenario. As shown in Fig. 5, a rectangular entity is positioned on the rotating table, with the center axis of the rotating table aligning with the block's center axis. The entity's front surface serves as the plane to-be-detected.

Each calibration initiation begins with an initialization process to ensure calibration accuracy. The platform propels the camera forward until its front surface aligns with the block's surface. To verify the alignment of the camera with the block's surface, a fragile piece of paper is placed between the two surfaces and the horizontal slider adjusted until the paper is neither too taut nor too slack.

Nonetheless, the distance value derived from the horizontal slide cannot directly represent the distance between the camera and the target plane. Due to camera's optical origin offset and the block's thickness, the distance must be compensated. Figure 6 illustrates the geometry principle of the distance compensation from a top view.

The distance d_E represents the value read directly from the horizontal slide. The offset s signifies the lateral offset between the camera's optical origin and the symmetrical plane. The distance d_1 and d_2

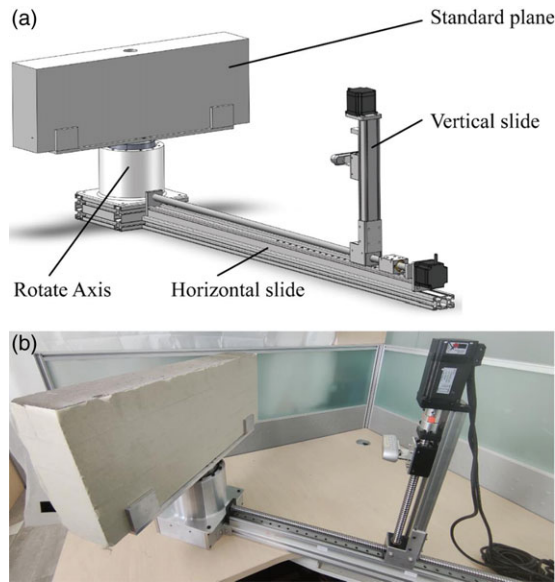


Figure 5. Prototype of mechanical calibration platform.

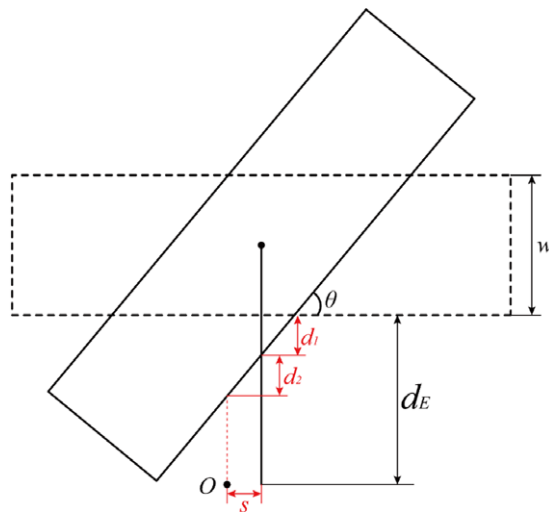


Figure 6. Distance compensation of calibration platform.

denote the compensation resulting from the block’s rotation and the lateral offset s . The final modified distance can be expressed by Eq. (12).

$$d^* = d_E - d_1 - d_2 = d_E - \frac{w}{2} \left(\frac{1}{\cos \theta} - 1 \right) - s \tan \theta \tag{12}$$

The platform adjusts the relative poses according to the pre-defined distance list, $\{D_j | j = 1, 2, \dots, n_D\}$, and angle list $\{\theta_i | i = 1, 2, \dots, n_\theta\}$. Initially, the horizontal slide distance is set to D_1 , and the inclination angle is set to each value within the angles list sequentially. Subsequently, the distance is set to the remaining values in the distance list and the angle changes are repeated. Ultimately, $n_D \times n_\theta$ mapping relations are gathered from the calibration.

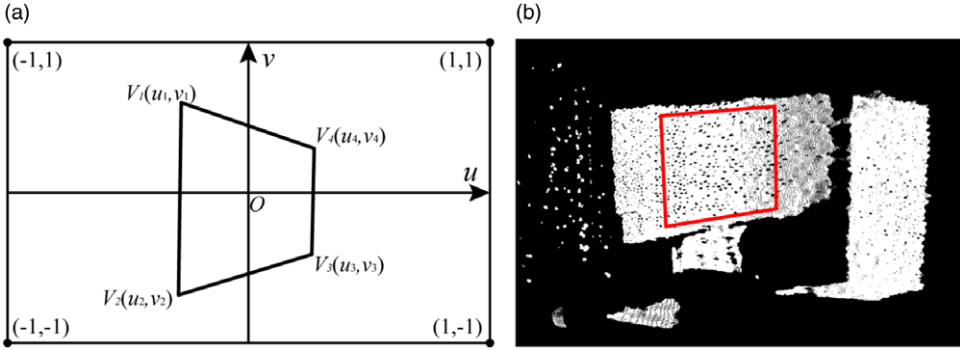


Figure 7. Planar coordinate system arrangement of camera view.

2.4. Preprocessing of raw point cloud data

Typically, a stereo camera can directly generate point clouds through methods based on binocular disparity or other principles. The random sample consensus (RANSAC) algorithm is an effective iterative method to identify planes from point clouds.

Prior to plane identification, pre-filtering points that may belong to the plane point cloud can eliminate numerous outlier points. If an excessive number of outliers exist beyond the target plane within the entire point cloud, they might influence the plane identification outcome. The region-of-interest (ROI) method is effective for filtering purposes.

A planar image coordinate system (refer to Fig. 7) is established to represent the ROI filter employed on the calibration platform.

At a specific instant, there is one point, $P(x_p, y_p, z_p)$, lies within the camera coordinate system, and belongs to the real-time point cloud. Its corresponding point, $P(u_p, v_p)$, in the defined image coordinate system is defined as

$$\begin{bmatrix} u_p \\ v_p \end{bmatrix} = \begin{bmatrix} \frac{2}{\theta_h} \arctan & 0 & 0 \\ 0 & \frac{2}{\theta_v} \arctan & 0 \end{bmatrix} \begin{bmatrix} \frac{x_p}{z_p} \\ \frac{y_p}{z_p} \\ 1 \end{bmatrix} \tag{13}$$

where θ_h denotes the horizontal field of view and θ_v signifies the vertical field of view. The operator *arctan* indicates the arc tangent operation. If a point is visible in the view field, the value ranges of its coordinates in the image coordinate system defined by Eq. (13) are $u_p \in [-1, 1]$ and $v_p \in [-1, 1]$.

Suppose four known vertices of ROI in the camera coordinate system are $V_1(x_1, y_1, z_1)$, $V_2(x_2, y_2, z_2)$, $V_3(x_3, y_3, z_3)$, and $V_4(x_4, y_4, z_4)$.

$$\begin{bmatrix} u_1 & u_2 & u_3 & u_4 \\ v_1 & v_2 & v_3 & v_4 \end{bmatrix} = \begin{bmatrix} \frac{2}{\theta_h} \arctan & 0 & 0 \\ 0 & \frac{2}{\theta_v} \arctan & 0 \end{bmatrix} \begin{bmatrix} \frac{x_1}{z_1} & \frac{x_2}{z_2} & \frac{x_3}{z_3} & \frac{x_4}{z_4} \\ \frac{y_1}{z_1} & \frac{y_2}{z_2} & \frac{y_3}{z_3} & \frac{y_4}{z_4} \\ 1 & 1 & 1 & 1 \end{bmatrix} \tag{14}$$

In Eq. (14), $V_1(u_1, v_1)$, $V_2(u_2, v_2)$, $V_3(u_3, v_3)$, $V_4(u_4, v_4)$ are the corresponding coordinates in the defined image coordinate system.

Divide the quadrilateral ROI, $V_1V_2V_3V_4$, into two triangles $\Delta V_1V_2V_3$ and $\Delta V_1V_3V_4$. If a point situates within the ROI, it must lie in one of the triangles. The algorithm for determining whether point

$P(x_p, y_p, z_p)$ is located within the triangle $\Delta V_1 V_2 V_3$ is expressed in Eq. (15).

$$\begin{bmatrix} u_1 \tan(\theta_h/2) & u_2 \tan(\theta_h/2) & u_3 \tan(\theta_h/2) \\ v_1 \tan(\theta_v/2) & v_2 \tan(\theta_v/2) & v_3 \tan(\theta_v/2) \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} J_1 \\ J_2 \\ J_3 \end{bmatrix} = \begin{bmatrix} x_p \\ y_p \\ z_p \end{bmatrix} \tag{15}$$

where the vector $[J_1 \ J_2 \ J_3]^T$ represents the judgment vector. If each element of vector $[J_1 \ J_2 \ J_3]^T$ is positive, the point $P(x_p, y_p, z_p)$ lies inside this triangle. If any element of the vector is negative, then the other triangle of the ROI, $\Delta V_1 V_3 V_4$, should be examined.

For the proposed calibration platform, the ROI can be chosen as a rectangle affixed on the detected surface. Along with the changing of relative distance and inclination angle, real-time coordinates of vertices can be computed by using Eq. (16)

$$V_{ROI}^* = T_2 M T_1 V_{ROI}^0 \tag{16}$$

In Eq. (16), V_{ROI}^* represents the real-time coordinates of ROI vertices, as defined by Eq. (17)

$$V_{ROI}^* = [V_1^* \ V_2^* \ V_3^* \ V_4^*] \tag{17}$$

In Eq. (16), V_{ROI}^0 signifies the coordinates of ROI vertices in the initial condition, as defined by Eq. (18).

$$V_{ROI}^0 = [V_1^0 \ V_2^0 \ V_3^0 \ V_4^0] = \begin{bmatrix} -\frac{l_1}{2} & -\frac{l_1}{2} & \frac{l_1}{2} & \frac{l_1}{2} \\ \frac{l_2}{2} & -\frac{l_2}{2} & -\frac{l_2}{2} & \frac{l_2}{2} \\ c_0 & c_0 & c_0 & c_0 \\ 1 & 1 & 1 & 1 \end{bmatrix} \tag{18}$$

where l_1 denotes the length of the rectangle and l_2 represents the height of the rectangle. The initial state of the platform is defined as the status $\theta = 0, d = 0$. The symbol c_0 denotes the displacement between the camera’s optical origin and the front surface.

In Eq. (16), the matrix T_1 represents a transformation matrix, as shown in Eq. (19)

$$T_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -\frac{w}{2} - c_0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{19}$$

where w denotes the thickness of the sample entity.

In Eq. (16), the matrix M represents a transformation matrix, as depicted in Eq. (20)

$$M = \begin{bmatrix} \cos \theta & 0 & -\sin \theta & 0 \\ 0 & 1 & 0 & 0 \\ \sin \theta & 0 & \cos \theta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{20}$$

where θ denotes the relative inclination angle.

In Eq. (16), the matrix T_2 represents a transformation matrix, as shown in Eq. (21)

$$T_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & d^* + \frac{w}{2} + c_0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (21)$$

where d^* denotes the modified distance defined by Eq. (12).

3. Experimental validation

3.1. Devices overview

Intel RealSense™ D435i (abbreviated as D435i throughout this article) is a compact, low-cost, consumer-grade, binocular stereo camera. RealSense series camera are popular in various applications. Keselman *et al.* [39] discussed the performances and limitations of RealSense cameras. Zhu *et al.* [6] employed the D435i for safety monitoring of solitary individuals. Huynh *et al.* [40] utilized D435i for estimating robot poses. Rong *et al.* [41] used it for recognizing oyster mushrooms in auto-harvesting.

Most D435i applications are qualitative, such as morphology detection or color feature recognition. The official document for the D435i states a relative error of 2% of the distance. Among the ten recently published papers that mentioned D435i in their abstracts [29–32, 41–46], only four utilize the camera's measurement function. When measuring the length of the grape clusters [30], the error ranges from around –20 to 20 mm. Neupane *et al.* [29] conducted an experiment measuring ceramics, PTFE, and fruits, with measurement residuals for the D435i ranging from 5 to 240 mm with measurement distance varying from 400 to 4000 mm. Measurements on the sewers [31] show errors from around 40 to 130 mm, while fluid surface measurements [32] reveal errors from approximately 10 to 60 mm. These precision values are appraised reading from the figure reported in the cited papers. Some cameras mentioned in these papers are actually D435 model, which differs from D435i only in the presence of an inertial measurement unit.

The experiment in this section utilizes D435i to demonstrate the extent to which precision can be increased using the existing method. Improved precision enables the adoption of the D435i in quantitative and high-precision-required scenarios.

The geometric parameters of D435i are as follows: the lateral offset of optical origin to the symmetrical plane is $s = 17.5$ mm; the longitudinal offset of optical origin to the camera surface is $c_0 = 4.2$ mm; the horizontal field of view is $\theta_h = 86^\circ$; the vertical field of view is $\theta_v = 57^\circ$.

The encoders in the slides and the rotation joint of the calibration platform ensure accuracy during the fine pose data gathering process. The linear translation stage employs a closed-loop stepper system. The slide encoder's resolution is 5 $\mu\text{m}/\text{pulse}$. And the absolute translation error is, at most, 0.03 mm according to the slide's official statement.

The resolution for the rotation motor's encoder is 19 bits, corresponding to 524,288 counts per revolution or 0.00069° per count. Typically, the position error is larger than the numerical resolution. According to the official statement, the motor's maximum absolute position error is 0.05°.

3.2. Calibration data acquisition

A RANSAC-based method implemented in PCL Library (PCL-RANSAC) [28] serves as the experimented plane identification method to be calibrated. The PCL Library is an open project for point cloud processing and is widely used in research [26, 47, 48]. In the experiments, PCL-RANSAC threshold is set to 0.005 m.

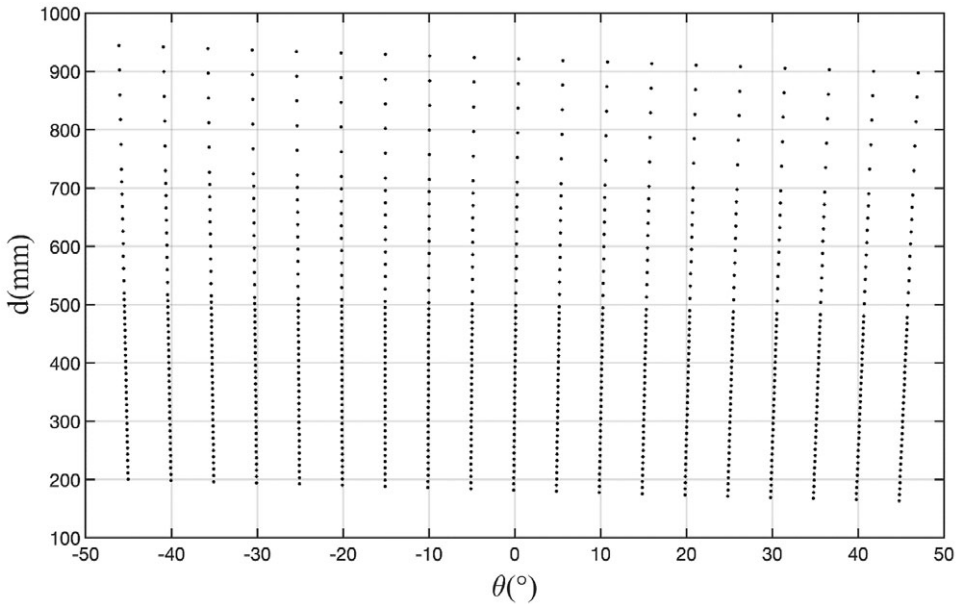


Figure 8. Calibration data of initial poses in bad-plane-pose space.

A list of standard distances and angles in calibration can be found in Appendix A. At each set of distance and inclination angle values, 20 sets of point cloud are captured. As random error cannot be ignored in single measurement, averaging 20 measurements to reduce random errors.

Following the previously introduced calibration processes, the calibration data of initial poses in bad-plane-pose space is shown in Fig. 8.

In Fig. 8, each solid point represents an initially detected pose obtained by the stereo system, with the distribution of the points appearing distorted.

To display the difference between initial pose points and the accurate ones, segments connecting each initial pose point to its corresponding accurate pose point are shown in Fig. 9.

Each segment in Fig. 9 represents a mapping relation from the initial pose to the accurate pose. Then, using the interpolation method represented in Section 2, any online detected initial pose can be corrected to a more precise one.

3.3. Validation on correcting the pose identified by PCL-RANSAC

Comparisons are made between the poses identified by PCL-RANSAC and the results after correcting by using the proposed method.

First, a comparison of distance identification is made. The inclination of the plane angle is fixed at $\theta = 0$ in this comparison experiment. Tested distance values are randomly chosen every 10 mm from 260 to 700 mm.

Figure 10 displays the comparison of the identified distances before and after the calibration mapping. The red polylines represent the initial data from PCL-RANSAC. Random errors cause fluctuation around an approximate linear increasing trend. The identification error increases with the detected distance. After calibrating the initial distance, errors are reduced to a much lower level, represented by the blue polyline shown in Fig. 10.

Figure 11 displays the relationship between the ratio of error and distance, illustrating the correlation between the error and distance. The results show that the ratio increases along with the distance increasing before calibration. However, after calibration, the ratio does not exhibit significant changes.

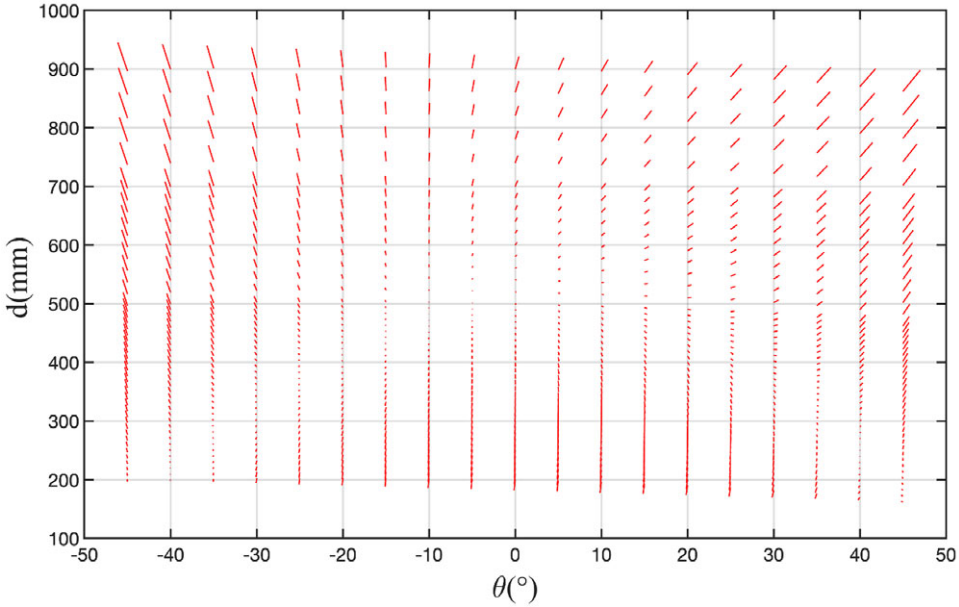


Figure 9. Mapping relations from initial poses to corrected poses.

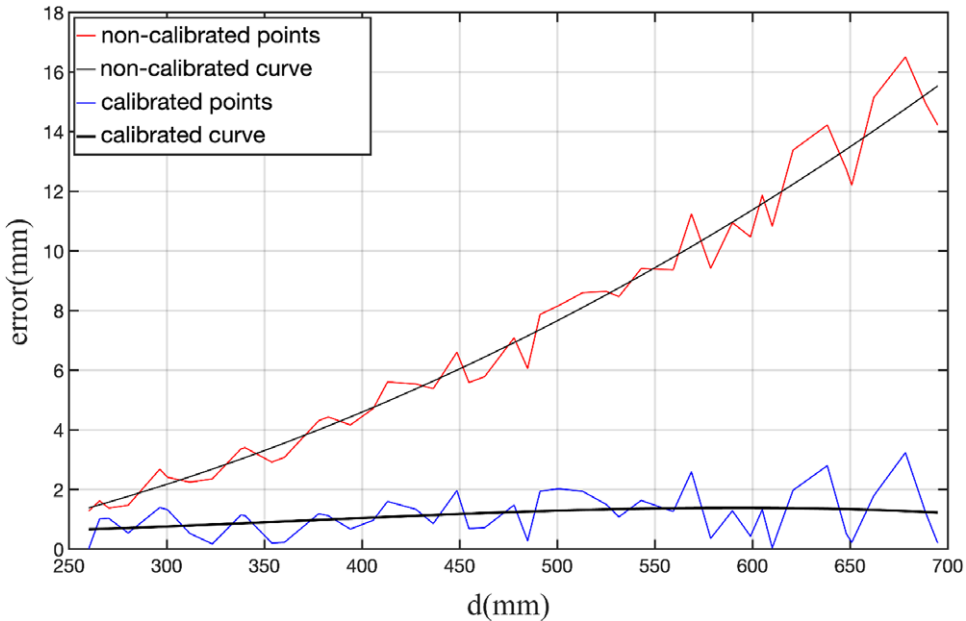


Figure 10. Comparison before and after calibration with varied plane distances.

Data points with errors ten percent larger than the observed distance are considered outliers and ignored. The results of the distance correction experiment are shown in Table B1, Appendix B.

In addition to distance experiments, a comparison on both distance and inclination angle identification is conducted. Distances are chosen randomly every 10 mm, from 200 to 800 mm. The inclination angles are chosen randomly every 10° from -45° to 45°.

Figure 12 shows the comparison between calibrated and non-calibrated poses in the $\theta - d$ coordinate system.

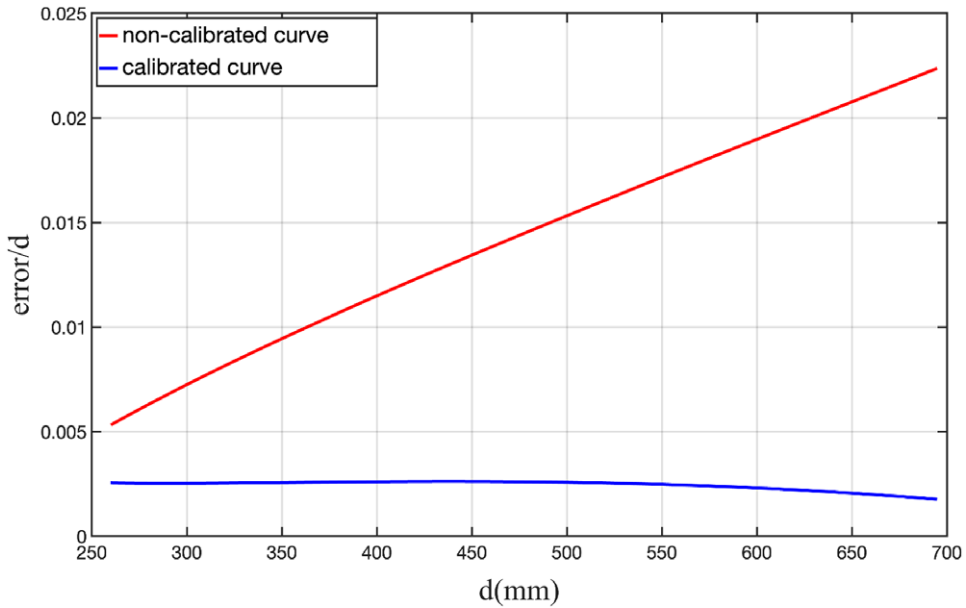


Figure 11. The relationship between the ratio of error to observing distance and observing distance.

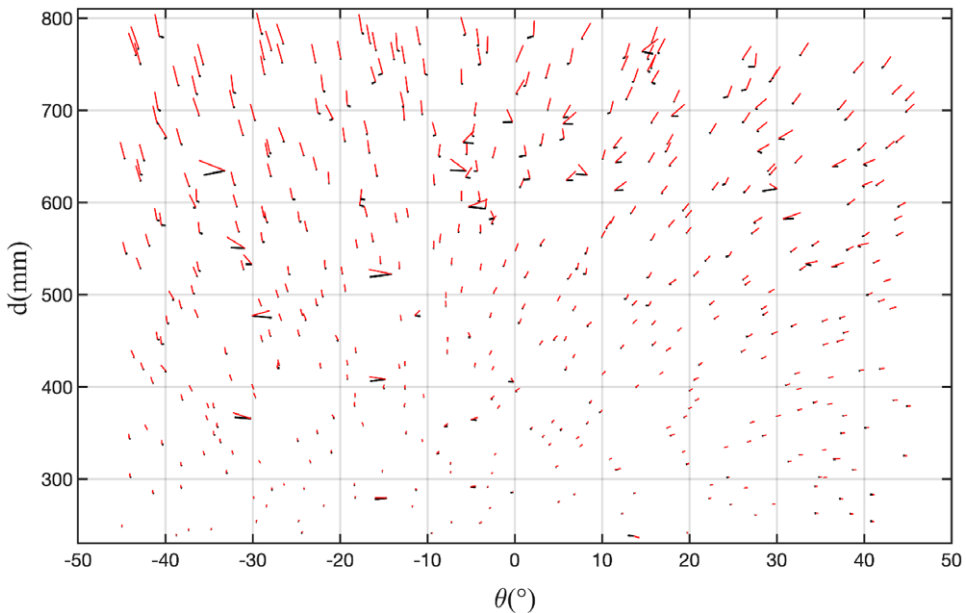


Figure 12. Comparison before and after calibration with varied distances and angles.

In Fig. 12, red segments connect the accurate pose points to the non-calibrated pose points, while black segments connect the accurate pose points to the calibrated ones. For clarity, Fig. 13 illustrates the calibrated segments only.

From Figs. 12 and 13, the comparison clearly shows a significant reduction in errors. However, in Fig. 13, some segments are still noticeable compared to others. These segments, which are nearly parallel to the θ axis, indicate that the angle errors of these points do not achieve good results. Nonetheless, compared to Fig. 12, the original data for these samples are already unformatted compared to the other

Table I. Results of distance and angle correction experiment.

<i>l</i>	Mean absolute error of distance (mm)	Mean relative error of distance (%)	Mean absolute error of angle (°)
PCL-RANSAC	7.350	1.292	0.4299
After calibration	0.9091	0.2378	0.2530

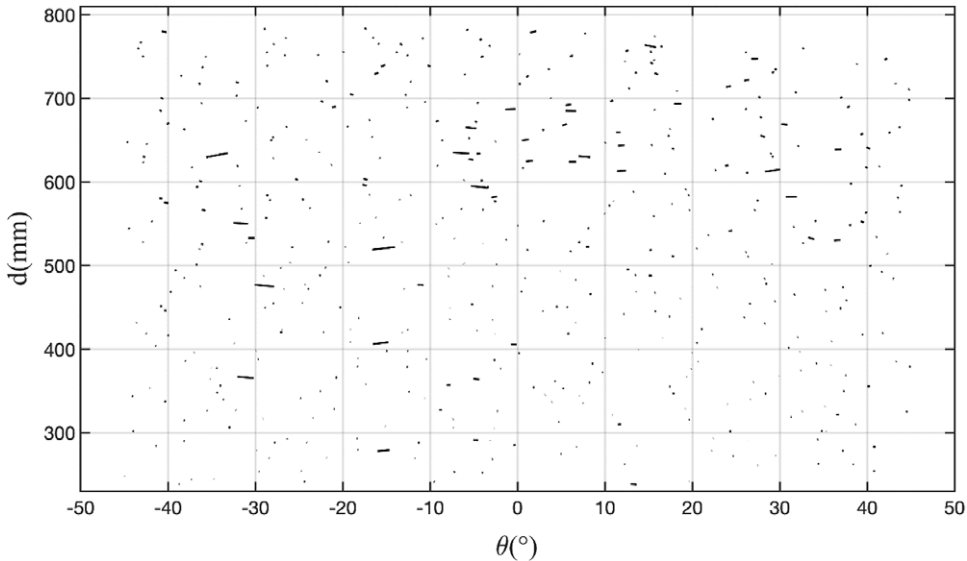


Figure 13. Lines connected accurate pose points and calibrated pose points.

data. These unusual error samples are likely caused by random noises and fluctuations in the point cloud. Most samples are calibrated to a low-error value, even including a few abnormal samples.

The results of the distance and angle correction experiment are shown in Table I. The mean absolute error of distances, the mean relative error of distances, and the mean absolute error of angle are listed. Before calculating the averages, absolute operations are performed on each original value.

According to the results in Table I, this calibration significantly improves the precision of distance and angle identification over PCL-RANSAC.

4. Application to construction robotics

A construction robot discussed herein, a mobile manipulator, comprises a 6-degree-of-freedom robotic arm, an elevation mechanism, and a wheeled chassis, endowing it with redundant mobility capabilities suitable for construction tasks. The robotic arm is responsible for carrying and positioning bricks, while the wheeled chassis ensures ample workspace for the construction robot.

The aforementioned calibration method for plane parameter identification is utilized during brick wall construction processes. The primary objective is to accurately position each brick.

Assuming a brick wall comprises n_1 layers and n_2 bricks per layer. The pose information for a brick is denoted by vector $\mathbf{b}_{ij} = [x_{ij}, y_{ij}, z_{ij}, a_{ij}, b_{ij}, c_{ij}]^T$. The first three components represent the spatial coordinates of the brick’s center, and the remaining three signify the front surface’s normal vector.

Matrix \mathbf{B}^* , the target matrix, encapsulates the construction task and contains the desired pose for each brick, as articulated in Eq. (22).

$$\mathbf{B}^* = \begin{bmatrix} \mathbf{b}_{11}^* & \cdots & \mathbf{b}_{1n_2}^* \\ \cdots & \cdots & \cdots \\ \mathbf{b}_{n_1 1}^* & \cdots & \mathbf{b}_{n_1 n_2}^* \end{bmatrix} \tag{22}$$

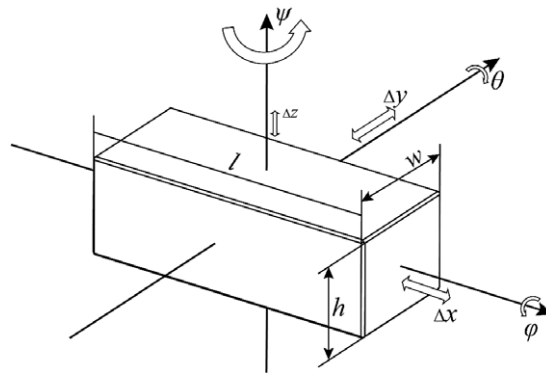


Figure 14. Brick's position and orientation.

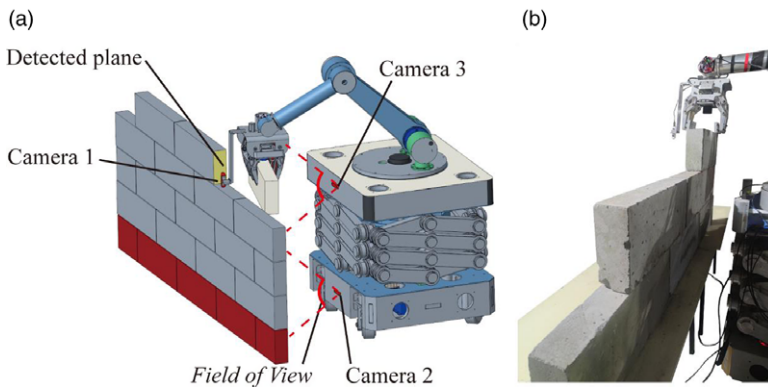


Figure 15. Construct the brick wall with a mobile manipulator. (a) overall view of the robot system. (b) prototype experiment.

The construction of an autonomous robot involves iteratively retrieving bricks from storage and accurately positioning them in the designated area until all bricks have been placed.

Accurate pose detections are performed for bricks on the wall, as illustrated in Fig. 14. All bricks share uniform dimensions, l , d , and h correspond to length, width, and the height. Brick's pose is expressed through yaw(ψ), pitch(θ), and roll(ϕ). Given that the brick's bottom surface rests atop the foundation wall, it remains approximately parallel to the ground. Consequently, yaw is emphasized during pose detection, while pitch and roll are comparatively negligible, as they are constrained by the alignment of the bottom and upper surfaces. Similarly, the displacements Δx and Δy are prioritized over Δz , as demonstrated in Fig. 14.

A multi-camera stereo system is then established to reduce constructing error. Three cameras, positioned on the robot as shown in Fig. 15, ensure the placing precision. Camera 1, mounted on the robot's gripper, detects the side planes of existing bricks, while the other two cameras, situated at the robot's front, monitor the caught brick's outer surface and brick wall's outer surface. The robot identifies the foundation wall's pose while placing each brick, thus ensuring precise placement.

The proposed calibration method is applied to automate brick wall construction using the robot, achieving a flatness performance below 4 mm.

5. Results and discussion

The calibration process may be perceived as a rectification procedure for any plane poses identification algorithm. Subsequent to plane pose correction, precision is enhanced relative to the initial plane pose.

Validations are conducted employing a widely used plane pose identification algorithm, PCL-RANSAC. As indicated by the experimental outcomes in Table I, the mean absolute distance error diminishes from 7.350 to 0.9091 mm. The mean relative error of distance declines from 1.292 to 0.2378%. The mean absolute angle error decreases from 0.4299° to 0.2530°. The angle identification error has no notable correlation to the true pose angle. Thus, the mean relative angle error is not enumerated in Table I.

The experimental results validate that the proposed calibration method and mechanical platform can augment the precision. Relative to the method reported in ref. [35], the proposed calibration method demonstrates superior performance at short detected distances. The relative error reported in ref. [35] is 0.867% at 0.8 m, -1.346% at 0.802 m, -0.520% at 0.947 m, and 0.298% at 1.110 m. Moreover, the relative error reported in ref. [36] is 0.49% at 1.23 m. Conversely, our results reveal a mean relative distance error of 0.2378% within 1 m.

No specific camera distortion model is required in our calibration method, unlike the method proposed in refs. [35] and [36]. The bilinear interpolation method can fit any camera distortion model, provided the system error of plane pose identification remains continuous at varying distances and angles.

Additionally, the mechanical platform shows crucial for enhanced calibration. The interpolation-based method demands extensive data collection to better fit the error functions. The mechanical platform can swiftly move and accurately position itself, facilitating calibration completion in a minimal timeframe.

As depicted in Fig. 10, a biased residual of approximately 2 mm persists after calibration. This may be attributable to suboptimal initial calibration data collection. As demonstrated in Fig. 9, the original calibration data points are unevenly distributed. Moreover, despite the mechanism resolution sufficiency for precise pose calculation, installation errors of components and the detected plane sample's nonideality may engender errors.

6. Conclusions

This article presents a calibration method that does not rely on particular camera distortion models for plane identification. A high-precision, three-degree-of-freedom mechanical calibration platform is devised to perform high-precision calibration data gathering tasks. The platform gathers mapping relations between low-precision plane poses derived from the stereo system and accurate plane poses obtained from the platform. By employing the interpolation method, any real-time acquired plane pose can be rectified to a more precise one by utilizing pre-gathered mapping relations. Experimental comparisons validate the plane pose correction's efficacy on PCL-RANSAC. The mean absolute distance error reduces from 7.350 to 0.9091 mm, the mean relative distance error diminishes from 1.292 to 0.2378%, and the mean absolute error of angle reduces from 0.4299° to 0.2530°. This calibration method can be applied to any plane parameters identification algorithm, as long as the initially identified pose exhibits a biased error. Simultaneously, this method can be employed in other plane detection scenarios except for the plane pose detection of brick surfaces.

Author contributions. Junjie Ji conceived and designed the study. Junjie Ji conducted analysis and data gathering. Junjie Ji wrote the article. Jing-Shan Zhao revised the manuscript and provided supervision. All authors read and approved the final manuscript.

Financial support. This work was supported in part by 2020GQI1003, Guoqiang Research Institute of Tsinghua University.

Competing interests. The authors declare none.

Ethical standards. Not applicable.

References

- [1] C. Song, M. Niu, Z. Liu, J. Cheng, P. Wang, H. Li and L. Hao, "Spatial-temporal 3D dependency matching with self-supervised deep learning for monocular visual sensing," *Neurocomputing* **481**, 11–21 (2022). doi: [10.1016/j.neucom.2022.01.074](https://doi.org/10.1016/j.neucom.2022.01.074).
- [2] Z. Wang, X. Li, X. Zhang, Y. Bai and C. Zheng, "An attitude estimation method based on monocular vision and inertial sensor fusion for indoor navigation," *IEEE Sens. J.* **21**(23), 27051–27061 (2021). doi: [10.1109/JSEN.2021.3119289](https://doi.org/10.1109/JSEN.2021.3119289).
- [3] Y. Li, J. Li, Q. Yao, W. Zhou and J. Nie, "Research on predictive control algorithm of vehicle turning path based on monocular vision," *Processes* **10**(2), 417 (2022). doi: [10.3390/pr10020417](https://doi.org/10.3390/pr10020417).
- [4] Q. Lu, H. Zhou, Z. Li, X. Ju, S. Tan and J. Duan, "Calibration of five-axis motion platform based on monocular vision," *Int. J. Adv. Manuf. Technol.* **118**(9–10), 3487–3496 (2022). doi: [10.1007/s00170-021-07402-x](https://doi.org/10.1007/s00170-021-07402-x).
- [5] Z. Zimiao, X. Kai, W. Yanan, Z. Shihai and Q. Yang, "A simple and precise calibration method for binocular vision," *Meas. Sci. Technol.* **33**(6), 065016 (2022). doi: [10.1088/1361-6501/ac4ce5](https://doi.org/10.1088/1361-6501/ac4ce5).
- [6] L. Zhu, Y. Zhang, Y. Wang and C. Cheng, "Binocular vision positioning method for safety monitoring of solitary elderly," *Comput. Mater. Continua* **71**(1), 593–609 (2022). doi: [10.32604/cmc.2022.022053](https://doi.org/10.32604/cmc.2022.022053).
- [7] L. Fang, Z. Guan and J. Li, "Automatic roadblock identification algorithm for unmanned vehicles based on binocular vision," *Wireless Commun. Mobile Comput.* **2021**, 1–7 (2021). doi: [10.1155/2021/3333754](https://doi.org/10.1155/2021/3333754).
- [8] K. Bonnen, J. S. Matthis, A. Gibaldi, M. S. Banks, D. M. Levi and M. Hayhoe, "Binocular vision and the control of foot placement during walking in natural terrain," *Sci. Rep.* **11**(1), 20881 (2021). doi: [10.1038/s41598-021-99846-0](https://doi.org/10.1038/s41598-021-99846-0).
- [9] J. Zhao and R. S. Allison, "The role of binocular vision in avoiding virtual obstacles while walking," *IEEE Trans. Visual. Comput. Graphics* **27**(7), 3277–3288 (2021). doi: [10.1109/TVCG.2020.2969181](https://doi.org/10.1109/TVCG.2020.2969181).
- [10] H. Chen and W. Cui, "A comparative analysis between active structured light and multi-view stereo vision technique for 3D reconstruction of face model surface," *Optik* **206**, 164190 (2020). doi: [10.1016/j.ijleo.2020.164190](https://doi.org/10.1016/j.ijleo.2020.164190).
- [11] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein and R. Szeliski, "A Comparison and Evaluation of Multi-View Stereo Reconstruction Algorithms," *In: 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, New York, NY, USA, vol. 1 (2006) pp. 519–528. doi: [10.1109/CVPR.2006.19](https://doi.org/10.1109/CVPR.2006.19).
- [12] L. Liu, N. Deng, B. Xin, Y. Wang, W. Wang, Y. He and S. Lu, "Objective evaluation of fabric pilling based on multi-view stereo vision," *J. Textile Inst.* **112**(12), 1986–1997 (2021). doi: [10.1080/00405000.2020.1862479](https://doi.org/10.1080/00405000.2020.1862479).
- [13] F. Duan and Q. Zhang, "Stereoscopic image feature indexing based on hybrid grid multiple suffix tree and hierarchical clustering," *IEEE Access* **8**, 23531–23541 (2020). doi: [10.1109/ACCESS.2020.2970123](https://doi.org/10.1109/ACCESS.2020.2970123).
- [14] Y. Long, Y. Wang, Z. Zhai, L. Wu, M. Li, H. Sun and Q. Su, "Potato volume measurement based on RGB-D camera," *IFAC-PapersOnLine* **51**(17), 515–520 (2018). doi: [10.1016/j.ifacol.2018.08.157](https://doi.org/10.1016/j.ifacol.2018.08.157).
- [15] T. Kim, M. Kang, S. Kang and D. Kim, "Improvement of Door Recognition Algorithm Using Lidar and RGB-D Camera for Mobile Manipulator," *In: 2022 IEEE Sensors Applications Symposium (SAS)*, Sundsvall, Sweden (2022) pp. 1–6. doi: [10.1109/SAS54819.2022.9881249](https://doi.org/10.1109/SAS54819.2022.9881249).
- [16] B. Shen, X. Lin, G. Xu, Y. Zhou and X. Wang, "A Low Cost Mobile Manipulator for Autonomous Localization and Grasping," *In: 2021 5th International Conference on Robotics and Automation Sciences (ICRAS)*, Wuhan, China (2021) pp. 193–197. doi: [10.1109/ICRAS52289.2021.9476294](https://doi.org/10.1109/ICRAS52289.2021.9476294).
- [17] K. Backman, D. Kulic and H. Chung, "Learning to assist drone landings," *IEEE Robot. Autom. Lett.* **6**(2), 3192–3199 (2021). doi: [10.1109/LRA.2021.3062572](https://doi.org/10.1109/LRA.2021.3062572).
- [18] M. C. P. Santos, L. V. Santana, A. S. Brandao and M. Sarcinelli-Filho, "UAV Obstacle Avoidance Using RGB-D System," *In: 2015 International Conference on Unmanned Aircraft Systems (ICUAS)*, Denver, CO, USA (2015) pp. 312–319. doi: [10.1109/ICUAS.2015.7152305](https://doi.org/10.1109/ICUAS.2015.7152305).
- [19] S. Back, J. Kim, R. Kang, S. Choi and K. Lee, "Segmenting Unseen Industrial Components In A Heavy Clutter Using RGB-D Fusion And Synthetic Data," *In: 2020 IEEE International Conference on Image Processing (ICIP)*, Abu Dhabi, United Arab Emirates (2020) pp. 828–832. doi: [10.1109/ICIP40778.2020.9190804](https://doi.org/10.1109/ICIP40778.2020.9190804).
- [20] D. Damen, A. Gee, W. Mayol-Cuevas and A. Calway, "Egocentric Real-Time Workspace Monitoring Using an RGB-D Camera," *In: 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems, Vilamoura-Algarve, Portugal* (2012) pp. 1029–1036. doi: [10.1109/IROS.2012.6385829](https://doi.org/10.1109/IROS.2012.6385829).
- [21] M. T. Sorour, M. A. Abdellatif, A. A. Ramadan and A. A. Abo-Ismael, "Development of roller-based interior wall painting robot," *5*(11) (2011).
- [22] S. Wilson, J. Potgieter and K. M. Arif, "Robot-assisted floor surface profiling using low-cost sensors," *Remote Sens-BASEL* **11**(22), 2626 (2019). doi: [10.3390/rs11222626](https://doi.org/10.3390/rs11222626).
- [23] L. Chen, J. Zhou and X. Chu, "A novel ground plane detection method using an RGB-D sensor," *IOP Conf. Ser. Mater. Sci. Eng.* **646**(1), 012049 (2019). doi: [10.1088/1757-899X/646/1/012049](https://doi.org/10.1088/1757-899X/646/1/012049).
- [24] X. Liu, L. Zhang, S. Qin, D. Tian, S. Ouyang and C. Chen, "Optimized LOAM using ground plane constraints and segMatch-based loop detection," *Ah S Sens.* **19**(24), 5419 (2019). doi: [10.3390/s19245419](https://doi.org/10.3390/s19245419).
- [25] M. Guo, L. Zhang, X. Liu, Z. Du, J. Song, M. Liu, X. Wu and X. Huo, "3D Lidar SLAM Based on Ground Segmentation and Scan Context Loop Detection," *In: 2021 IEEE 11th Annual International Conference on CYBER Technology in Automation, Control, and Intelligent Systems (CYBER)*, Jiaxing, China (2021) pp. 692–697. doi: [10.1109/CYBER53097.2021.9588285](https://doi.org/10.1109/CYBER53097.2021.9588285).
- [26] C. Fotsing, N. Menadjou and C. Bobda, "Iterative closest point for accurate plane detection in unorganized point clouds," *Autom. Constr.* **125**, 103610 (2021). doi: [10.1016/j.autcon.2021.103610](https://doi.org/10.1016/j.autcon.2021.103610).
- [27] M. W. Ryu, S. M. Oh, M. J. Kim, H. H. Cho, C. B. Son and T. H. Kim, "Algorithm for generating 3D geometric representation based on indoor point cloud data," *Appl. Sci.* **10**(22), 8073 (2020). doi: [10.3390/app10228073](https://doi.org/10.3390/app10228073).

- [28] R. B. Rusu and S. Cousins, “3D is here: Point Cloud Library (PCL),” *In: 2011 IEEE International Conference on Robotics and Automation*, Shanghai, China (2011) pp. 1–4. doi: [10.1109/ICRA.2011.5980567](https://doi.org/10.1109/ICRA.2011.5980567).
- [29] C. Neupane, A. Koirala, Z. Wang and K. B. Walsh, “Evaluation of depth cameras for use in fruit localization and sizing: Finding a successor to kinect v2,” *Agronomy* **11**(9), 1780 (2021). doi: [10.3390/agronomy11091780](https://doi.org/10.3390/agronomy11091780).
- [30] Y. Peng, S. Zhao and J. Liu, “Segmentation of overlapping grape clusters based on the depth region growing method,” *Electronics* **10**(22), 2813 (2021). doi: [10.3390/electronics10222813](https://doi.org/10.3390/electronics10222813).
- [31] C. H. Bahnsen, A. S. Johansen, M. P. Philipsen, J. W. Henriksen, K. Nasrollahi and T. B. Moeslund, “3D sensors for sewer inspection: A quantitative review and analysis,” *Ah S Sens.* **21**(7), 2553 (2021). doi: [10.3390/s21072553](https://doi.org/10.3390/s21072553).
- [32] D. B. Bung, B. M. Crookston and D. Valero, “Turbulent free-surface monitoring with an RGB-D sensor: The hydraulic jump case,” *J. Hydraul. Res.* **59**(5), 779–790 (2021). doi: [10.1080/00221686.2020.1844810](https://doi.org/10.1080/00221686.2020.1844810).
- [33] M. Parvis, “Using a-priori information to enhance measurement accuracy,” *Measurement* **12**(3), 237–249 (1994). doi: [10.1016/0263-2241\(94\)90030-2](https://doi.org/10.1016/0263-2241(94)90030-2).
- [34] Z. Zhang, “A flexible new technique for camera calibration,” *IEEE Trans. Pattern Anal. Mach. Intell.* **22**(11), 1330–1334 (2000). doi: [10.1109/34.888718](https://doi.org/10.1109/34.888718).
- [35] W. Darwish, W. Li, S. Tang, B. Wu and W. Chen, “A robust calibration method for consumer grade RGB-D sensors for precise indoor reconstruction,” *IEEE Access* **7**, 8824–8833 (2019). doi: [10.1109/ACCESS.2018.2890713](https://doi.org/10.1109/ACCESS.2018.2890713).
- [36] Y. Li, W. Li, W. Darwish, S. Tang, Y. Hu and W. Chen, “Improving plane fitting accuracy with rigorous error models of structured light-based RGB-d sensors,” *Remote Sens.* **12**(2), 320 (2020). doi: [10.3390/rs12020320](https://doi.org/10.3390/rs12020320).
- [37] W. Feng, Z. Liang, J. Mei, S. Yang, B. Liang, X. Zhong and J. Xu, “Petroleum pipeline interface recognition and pose detection based on binocular stereo vision,” *Processes* **10**(9), 1722 (2022). doi: [10.3390/pr10091722](https://doi.org/10.3390/pr10091722).
- [38] P. Fuerstattel, S. Placht, A. Maier and C. Riess, “Geometric primitive refinement for structured light cameras,” *Mach. Vis. Appl.* **29**(2), 313–327 (2018). doi: [10.1007/s00138-017-0901-z](https://doi.org/10.1007/s00138-017-0901-z).
- [39] L. Keselman, J. I. Woodfill, A. Grunnet-Jepsen and A. Bhowmik, “Intel(R) RealSense(TM) Stereoscopic Depth Cameras,” *In: 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Honolulu, HI, USA (2017) pp. 1267–1276. doi: [10.1109/CVPRW.2017.167](https://doi.org/10.1109/CVPRW.2017.167).
- [40] B.-P. Huynh and Y.-L. Kuo, “Dynamic filtered path tracking control for a 3RRR robot using optimal recursive path planning and vision-based pose estimation,” *IEEE Access* **8**, 174736–174750 (2020). doi: [10.1109/ACCESS.2020.3025952](https://doi.org/10.1109/ACCESS.2020.3025952).
- [41] J. Rong, P. Wang, Q. Yang and F. Huang, “A field-tested harvesting robot for oyster mushroom in greenhouse,” *Agronomy* **11**(6), 1210 (2021). doi: [10.3390/agronomy11061210](https://doi.org/10.3390/agronomy11061210).
- [42] Z. Li, X. Tian, X. Liu, Y. Liu and X. Shi, “A two-stage industrial defect detection framework based on improved-yOLOv5 and optimized-inception-resnetV2 models,” *Appl. Sci.* **12**(2), 834 (2022). doi: [10.3390/app12020834](https://doi.org/10.3390/app12020834).
- [43] T. Schlett, C. Rathgeb and C. Busch, “Deep learning-based single image face depth data enhancement,” *Comput. Vis. Image Understanding* **210**, 103247 (2021). doi: [10.1016/j.cviu.2021.103247](https://doi.org/10.1016/j.cviu.2021.103247).
- [44] V. Tadic, A. Odry, E. Burkus, I. Kecskes, Z. Kiraly, M. Klincsik, Z. Sari, Z. Vizvari, A. Toth and P. Odry, “Painting path planning for a painting robot with a realSense depth sensor,” *Appl. Sci.* **11**(4), 1467 (2021). doi: [10.3390/app11041467](https://doi.org/10.3390/app11041467).
- [45] H. Zeng, B. Wang, X. Zhou, X. Sun, L. Huang, Q. Zhang and Y. Wang, “TSFE-net: Two-stream feature extraction networks for active stereo matching,” *IEEE Access* **9**, 33954–33962 (2021). doi: [10.1109/ACCESS.2021.3061495](https://doi.org/10.1109/ACCESS.2021.3061495).
- [46] P. Oščádal, D. Heczko, A. Vysocký, J. Mlotek, P. Novák, I. Virgala, M. Sukop and Z. Bobovský, “Improved pose estimation of aruco tags using a novel 3D placement strategy,” *Ah S Sens.* **20**(17), 4825 (2020). doi: [10.3390/s20174825](https://doi.org/10.3390/s20174825).
- [47] M. Miknis, R. Davies, P. Plassmann and A. Ware, “Near Real-Time Point Cloud Processing Using the PCL,” *In: 2015 International Conference on Systems, Signals and Image Processing (IWSSIP)*, London, UK (2015) pp. 153–156. doi: [10.1109/IWSSIP.2015.7314200](https://doi.org/10.1109/IWSSIP.2015.7314200).
- [48] D. Holz, A. E. Ichim, F. Tombari, R. B. Rusu and S. Behnke, “Registration with the point cloud library: A modular framework for aligning in 3-D,” *IEEE Robot. Autom. Mag.* **22**(4), 110–124 (2015). doi: [10.1109/MRA.2015.2432331](https://doi.org/10.1109/MRA.2015.2432331).

Appendix A

During the calibration data collection process by a mechanical platform, distance values are designated at 200 mm, 210 mm, 220 mm, 230 mm, 240 mm, 250 mm, 260 mm, 270 mm, 280 mm, 290 mm, 300mm, 310 mm, 320 mm, 330 mm, 340 mm, 350 mm, 360 mm, 370 mm, 380 mm, 390 mm, 400mm, 410 mm, 420 mm, 430 mm, 440 mm, 450 mm, 460 mm, 470 mm, 480 mm, 490 mm, 500 mm, 520mm, 540 mm, 560 mm, 580 mm, 600 mm, 620 mm, 640 mm, 660 mm, 680 mm, 700 mm, 740 mm, 780mm, 820 mm, 860 mm, and 900 mm. Inclination angle values are designated at -45° , -40° , -35° , -30° , -25° , -20° , -15° , -10° , -5° , 0° , 5° , 10° , 15° , 20° , 25° , 30° , 35° , 40° , and 45° .

Appendix B

Table B1 enumerates the results of the distance correction experiment.

Table B1. Results of distance correction experiment.

NO.	True distances (mm)	PCL-RANSAC		After calibration	
		Absolute error (mm)	Relative error (%)	Absolute error (mm)	Relative error (%)
1	260.00	1.28	0.491	0.008	0.003
2	265.62	1.62	0.609	1.014	0.382
3	270.32	1.38	0.509	1.033	0.382
4	280.24	1.47	0.526	0.539	0.192
5	296.42	2.68	0.905	1.396	0.471
6	300.39	2.41	0.803	1.321	0.440
7	311.77	2.25	0.720	0.528	0.169
8	323.32	2.36	0.729	0.174	0.054
9	337.97	3.36	0.995	1.144	0.338
10	340.11	3.40	1.001	1.124	0.330
11	353.79	2.92	0.825	0.205	0.058
12	360.26	3.08	0.854	0.227	0.063
13	377.82	4.31	1.140	1.188	0.314
14	382.75	4.43	1.158	1.119	0.292
15	394.01	4.16	1.057	0.672	0.171
16	405.82	4.72	1.163	0.971	0.239
17	413.07	5.61	1.357	1.602	0.388
18	427.58	5.54	1.295	1.335	0.312
19	436.48	5.38	1.233	0.857	0.196
20	448.62	6.60	1.472	1.965	0.438
21	454.76	5.59	1.229	0.688	0.151
22	462.80	5.78	1.250	0.720	0.156
23	477.88	7.09	1.483	1.479	0.309
24	484.84	6.06	1.250	0.280	0.058
25	491.18	7.87	1.603	1.942	0.395
26	500.72	8.17	1.631	2.025	0.404
27	512.98	8.60	1.676	1.940	0.378
28	525.09	8.65	1.647	1.499	0.286
29	531.57	8.47	1.593	1.079	0.203
30	542.97	9.41	1.733	1.634	0.301
31	559.40	9.37	1.675	1.259	0.225
32	568.73	11.24	1.977	2.594	0.456
33	578.59	9.42	1.628	0.361	0.062
34	589.72	10.95	1.857	1.286	0.218
35	598.98	10.47	1.748	0.432	0.072
36	605.01	11.87	1.962	1.325	0.219
37	610.11	10.83	1.775	0.042	0.007
38	620.75	13.38	2.156	1.974	0.318
39	638.33	14.23	2.229	2.799	0.439

Table B1. *Continued.*

NO.	True distances (mm)	PCL-RANSAC		After calibration	
		Absolute error (mm)	Relative error (%)	Absolute error (mm)	Relative error (%)
40	648.08	12.74	1.965	0.512	0.079
41	650.86	12.21	1.876	0.226	0.035
42	662.11	15.15	2.288	1.791	0.270
43	678.34	16.51	2.434	3.236	0.477
44	688.68	14.95	2.171	1.225	0.178
45	694.86	14.23	2.048	0.203	0.029
Avg.	474.75	7.38	1.416	1.133	0.244