Precision Camera Calibration Using Known Target Motions Along Three Perpendicular Axes

Thomas E. Shkurti[a*], **Ammar Nahari**[a], **Wyatt Newman**[b], **M. Cenk Çavuşoğlu**[b]
[a]Department of Computer & Data Sciences, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, OH, USA, 44106
[b]Department of Electrical & Computer Engineering, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, OH, USA, 44106

Abstract. This paper presents a new method for precision intrinsic calibration of pinhole-model cameras. We refer to this method as "3-Axis". The algorithm employs a target moving with known displacements along 3 perpendicular axes, to decrease the number of unknown terms that must be determined. This method is described in detail along with new evaluation strategies for comparing the accuracy of calibration algorithms. The 3-Axis approach is then compared to the current state of the art in simulated and physical settings, and is shown to exceed it in accuracy for an equal number of samples acquired.

Keywords: Camera Calibration, Photogrammetry, Robotics.

*Thomas E. Shkurti, tes77@case.edu

1 Introduction

Intrinsic camera calibration is a necessary step before employing photogrammetry,[1] modeling,[2] and image-processing[3] techniques. The fitness of a calibration process can be evaluated as a relation between the amount and quality of calibration data fed into the calibration solver, and the accuracy with which it then estimates the intrinsic optical parameters of the camera.[4,5] Here, "quality" is an abstract measure of how accurately one can determine or enforce parameters of the calibration data (for instance the exact size of a fiducial, or its position in 3D space).[5] Generally, data with greater quality is more time-consuming and costly to collect. Similarly, it is more time-consuming and costly to acquire a numerically larger data set of a given quality.

This paper discusses a new intrinsic calibration method ("3-Axis") based on known movements of the fiducial along three orthogonal axes. This data collection system is intended to allow rapid, automated collection of high-quality data using commonly available equipment such as cutting mills, 3D printers, or robotic arms. Additionally, we propose a new and detailed method of comparing the accuracy of calibration methods using simulated data with a known ground truth.

We begin with an overview of the current state of the art, and identify the ROS-Industrial calibration system[6] and Zhang's calibration algorithm[7] as standards of performance to which 3-Axis will be compared. Next, we describe in detail the operation of the calibration algorithm. We then propose experimental methods of comparison between ROS-Industrial, Zhang's algorithm, and 3-Axis, using both simulated and real cameras. Finally, we present the performance metrics for all three systems, and conclude that 3-Axis is more likely to perform better than ROS-Industrial or Zhang's algorithm for data sets of any given size and quality.

2 Related Studies

Camera calibration methods are typically divided into three categories: traditional, self-calibrating, and active-vision:[8–10]

- **Traditional** methods involve imaging artificial calibration targets deliberately placed in the environment. Some or all of the target's physical parameters are known, as is some or all information relating to its position in the scene. The assumption is that the camera parameters discovered in this artificial environment will remain the same when the calibration target is removed and the camera is used to image other objects in another environment.
- **Self-calibration** methods do not require an artificial calibration target with known parameters. Rather, the camera gathers calibration information from the same visual environment where it will be used. This, of course, requires assumptions about what features will be found in the environment, but in absolute terms the camera calibrator has *no control* over the contents of the images.

- **Active-sensing** methods are similar to self-calibration methods, but the camera calibrator has control over the positioning of the camera (returning some control over the contents of the images).

3-Axis is a traditional calibration method- traditional methods tend to be the most accurate,[10] and are the focus of this review.

In a 2019 systematic review,[10] Long and Dongri identify three primary traditional calibration algorithms: the DLT algorithm by Abdel-Aziz and Karara,[11] the two-step algorithm by Tsai,[12] and the planar pattern algorithm by Zhang.[7] A 2002 review by Salvi, Armangué, and Batlle[4] compares Tsai's algorithm to DLT-based algorithms and finds Tsai's to be more accurate; while a 2014 review by Li et. al.[5] finds Zhang's algorithm to be comparable in accuracy to Tsai's. Indeed, it is Zhang's algorithm that is used in common computer vision toolkits such as OpenCV[13,14] and Matlab,[15,16] and most commonly overall.[17] It often serves as a baseline against which to compare other experimental calibration algorithms.[18–23] Tsai's algorithm, however, lacks such a commonly-used library implementation.[24]

Much development post 2019 has focused on calibration using few or single images;[25–28] calibration of atypical camera types or cameras in combination with other sensors;[29–33] or improvement of target fiducial detection.[34–36] However, alternative methods of data collection and processing in Zhang-like algorithms continue to be explored:

- Peng and Sturm[37] have created a utility to suggest advantageous calibration target positions for Zhang's algorithm (which are determined arbitrarily by a human operator in the normal use case), although human intervention is still required to move the target to these poses and the positioning is not precise.

- Gunen et. al.[38] examine and improve the optimization solvers used in Zhang's classical algorithm, as well as other common calibration algorithms.
- Jiang et. al.[19] replace the physical calibration target with a virtual target displayed on a screen in front of the camera, allowing for more precise and automated target movement with a more complicated projection model.
- Chen, Yang, and Pan[39] also employ a dynamic target generated on a screen- this one displays linear patterns and is moved to two different spatial locations with a precisely-known displacement, allowing for the independent calculation of distortion, principal point, and focal length parameters.
- Jin and Yang[22] employ a secondary calibration target viewed from a single position to estimate distortion, as a prelude to full-model calibration.
- Juarez-Salazar, Zhang, and Diaz-Ramirez[40] propose an alternative pinhole camera model more suited for cameras with high radial distortion, which is calibrated with classical checkerboard targets.
- Sun, Cheng, and Fan[41] propose a method employing a target made of two opaque cylinders with known radius, length, and position with respect to each other, placed at arbitrary locations in the camera image.
- Liu, Zhao, and Kou[42] combine traditional rectilinear targets with circular ones, to perform calibration using conic asymptotes.
- Yang, Chen, and Yu[23] propose a system similar to 3-Axis, but involving a target moving along a sled with only a single dimension of displacement.

Finally, particular attention is paid to the ROS-Industrial camera calibration toolkit.[6,43] This calibration approach also employs a target moving at known displacements along a single axis of motion, assumed to be close to perpendicular to the image plane of the camera and centered in the middle of the field of view. It is able to collect data automatically, and interfaces with the Robot Operating System; the precision of the calibration data produced exceeds that of the standard ROS implementation of Zhang's algorithm.[18,43,44] For this reason, both ROS-Industrial and the ROS implementation of Zhang's calibration algorithm were chosen as the state of the art against which to compare 3-Axis's performance.

3 Problem Formulation

Camera calibration attempts to find descriptive optical parameters for a projection model $\mathcal{P}$ of a physical camera. The most commonly used are the pinhole projection model (focal length F_x and F_y, and image center C_x and C_y), and Brown distortion model (radial distortion parameters k_1, k_2, k_3, and tangential distortion parameters p_1 and p_2).[18] Based on these parameters, $\mathcal{P}$ can map an arbitrary point (x, y, z) in 3D metric space to a point (u, v) in 2D pixel space:

$$(u, v) = \mathcal{P}(x, y, z) \tag{1}$$

$$u_{planar} = \frac{x}{z} \quad v_{planar} = \frac{y}{z} \tag{2}$$

$$
\begin{aligned}
r &= \|u_{planar}, v_{planar}\| \\
u_{radial} &= u_{planar}\left(k_1 \cdot r^2 + k_2 \cdot r^4 + k_3 \cdot r^6\right) \\
v_{radial} &= v_{planar}\left(k_1 \cdot r^2 + k_2 \cdot r^4 + k_3 \cdot r^6\right) \\
u_{tangential} &= 2\ p_1 \cdot u_{planar} \cdot v_{planar} + p_2\left(r^2 + 2\ u_{planar}{}^2\right) \\
v_{tangential} &= 2\ p_2 \cdot u_{planar} \cdot v_{planar} + p_1\left(r^2 + 2\ y_{planar}{}^2\right) \\
u_{distorted} &= u_{planar} + u_{radial} + u_{tangential} \\
v_{distorted} &= v_{planar} + v_{radial} + v_{tangential}
\end{aligned}
\tag{3}
$$

$$
\begin{aligned}
u &= F_x \cdot u_{distorted} + C_x \\
v &= F_y \cdot v_{distorted} + C_y
\end{aligned}
\tag{4}
$$

The optimization of the projection model compares a (typically large) set of fiducial points captured in actual camera images (here referred to as (U_i, V_i)) with points projected from the fiducial's physical location (X_i, Y_i, Z_i) by $\mathcal{P}$, and attempts to minimize the Euclidean distance between them:

$$
\arg\min_{\mathcal{P}} \sum_i \left(\left|\left|(U_i, V_i) - \mathcal{P}(X_i, Y_i, Z_i)\right|\right| \right) \tag{5}
$$

Here, a simple sum of Euclidean norms is used as the aggregate distance measure, although other aggregations (median, mean, sum of squares, etc.) and other distance measures (for instance, taxicab distance) could be applied. This differs from Tsai's two-step approach,[12] as the distortion and projection characteristics are optimized concurrently.

The fiducial position (X_i, Y_i, Z_i) can be alternatively represented by a 3D rigid transform from the camera frame to the fiducial frame $\mathbf{G_{CF}}_i$:

$$\underset{\mathcal{P}}{\arg\min} \sum_i \left(\left|\left|(U_i, V_i) - \mathcal{P}\left(\mathbf{G_{CF}}_i\right)\right|\right| \right) \tag{6}$$

However, the position of a given fiducial in metric space with respect to the camera is often not known to any precision: the camera is a three-dimensional volumetric object with an optical center somewhere in its interior. Indeed, determining some parameters of this optical center is one of the objectives *of* calibration. Therefore, this naive method introduces three unknown parameters and only two known parameters for every point, in addition to the parameters within $\mathcal{P}$. The result is an underspecified and unsolvable optimization problem:

$$\underset{\mathcal{P},\, \mathbf{G_{CF}}_i}{\arg\min} \sum_i \left(\left|\left|(U_i, V_i) - \mathcal{P}\left(\mathbf{G_{CF}}_i\right)\right|\right| \right) \tag{7}$$

Zhang's algorithm addresses this problem by including multiple fiducials on a single target of known dimensions. This allows the decomposition of $\mathbf{G_{CF}}_i$ into a set of fiducials on the same target j and target positions k:

$$\mathbf{G_{CF}}_{j,k} = \mathbf{G_{CT}}_k \cdot \boxed{\mathbf{G_{TF}}_j} \tag{8}$$

In Eq. (8), $\mathbf{G_{CT}}_k$ represents the position of the multi-fiducial target at position k, a parameter that Zhang's algorithm does not assume to be known. $\boxed{\mathbf{G_{TF}}_j}$ is the position of fiducial j *on* the target; the $\boxed{\text{box}}$ notation indicates that it is known prior to the calibration. Eq. (7) then becomes

$$\underset{\mathcal{P},\,\mathbf{G_{CT}}_k}{\arg\min} \sum_{j,k} \left(\left| \left| (U_{j,k}\ , V_{j,k}) - \mathcal{P}\left(\mathbf{G_{CT}}_k \cdot \boxed{\mathbf{G_{TF}}_j} \right) \right| \right| \right) \tag{9}$$

So long as the number of fiducial positions increases more rapidly than the number of target positions (that is to say, so long as the target has more than one fiducial visible on it), Eq. (9) produces a solvable optimization problem.

Typically, the measure of quality for a camera calibration model is the reprojection error **RE**.[4] This is simply the minimal final cost returned by the optimization system in the course of estimating $\mathcal{P}$:

$$\mathbf{RE} = \min_{\mathcal{P}} \sum_{i} \left(\left| \left| (U_i, V_i) - \mathcal{P}(\mathbf{G_{CF}}_i) \right| \right| \right) \tag{10}$$

The RE is then typically normalized by the number of data points used in the calibration.

This is often the only *possible* measurement of calibration quality when calibrating real cameras, as the purpose of calibration is to identify a ground truth otherwise unknown.

4 Methods

Our calibration method employs an alternative decomposition of $\mathbf{G_{CF}}_i$, employing a precision-controllable 3D movement system to produce more known parameters:

$$\mathbf{G_{CF}}_i = \mathbf{G_{CM}}_i \cdot \boxed{\mathbf{G_{ME}}_i} \cdot \mathbf{G_{ET}}_i \cdot \boxed{\mathbf{G_{TF}}_i} \tag{11}$$

where

- $\boxed{\mathbf{G}_{\mathbf{TF}i}}$ is the transformation between the origin of the target and a fiducial point on the target. Flat, two-dimensional targets can easily be constructed with known dimensions (and known positions for the center of each fiducial) using a commercial desktop printer. This term differs for each fiducial in a given target position, but is always known.

- $\mathbf{G}_{\mathbf{ET}i}$ is the transformation from the origin of the target, to the moving end of a precision-controllable, three-axis device such as a CNC machine, mill, or robotic arm. While it may be possible to affix a target to such a device in a known transformation, this was not assumed as a requirement for the calibration process. Therefore, it is assumed that the parameter must be solved for during calibration.

- $\boxed{\mathbf{G}_{\mathbf{ME}i}}$ is the transformation between the tip and the origin of the 3-axis device. This is the nominal position requested of the device, and thus precisely known, albeit different for each target acquisition.

- $\mathbf{G}_{\mathbf{CM}i}$ is the transformation from the "center" of the 3-axis device, to the "center" of the camera. Since both of these mechanisms are irregular solid devices with "origin points" determined by their physical and optical properties, these points are not considered physically meaningful and certainly not externally measurable. Therefore, this parameter must be found during the process of calibration.

It is of particular interest that the two estimated terms, $\mathbf{G}_{\mathbf{CM}}$ and $\mathbf{G}_{\mathbf{ET}}$, are consistent throughout the entire data set. This is *untrue* for the classical Zhang's algorithm- there, a different target-camera transform must be estimated for each physical location of the target.[7] This alteration changes the number of non-intrinsic terms needing to be optimized for a data set of n points, from

$\mathcal{O}(n)$ to a constant 12 (the minimal representation of two rigid transformations as 3 displacements and 3 Euler angles each).

ROS-Industrial also employs a constraint to ensure a constant number of optimized non-intrinsics by moving the target along a single axis in the approximate center of the camera's field of view.[18,43] However, the assumption of travel along the exact view center is difficult to enforce in reality;[44] and a comparatively smaller number of data points are collected near the edge of the image, which makes it more difficult to calculate distortion parameters that have a greater effect (in terms of absolute motion in pixels) near the edges of the image.[18]

Tsai's algorithm does not employ a single projection model, but rather a two-step process wherein rough focal information is determined assuming an undistorted image, and distortion parameters are calculated subsequently. Additionally, implementations of the algorithm assume that multiple targets will be present in one single image in a known configuration,[12,24] instead of a single target being moved and imaged subsequently.

Next, a redundant free parameter can be identified and removed from the optimization by expanding Eq. (11) into translation and rotation terms and taking advantage of certain cancellations:

$$\begin{aligned}
\mathbf{G_{CF}} &= \mathbf{R_{CM}}\left(\boxed{\mathbf{G_{ME}}}\cdot\mathbf{G_{ET}}\cdot\boxed{\mathbf{G_{TF}}}+\mathbf{T_{CM}}\right)\\
&= \mathbf{R_{CM}}\left(\boxed{\mathbf{R_{ME}}}\left(\mathbf{G_{ET}}\cdot\boxed{\mathbf{G_{TF}}}+\boxed{\mathbf{T_{ME}}}\right)+\mathbf{T_{CM}}\right)\\
&= \mathbf{R_{CM}}\left(\boxed{\mathbf{R_{ME}}}\left(\mathbf{R_{ET}}\left(\boxed{\mathbf{G_{TF}}}+\mathbf{T_{ET}}\right)+\boxed{\mathbf{T_{ME}}}\right)+\mathbf{T_{CM}}\right)
\end{aligned} \tag{12}$$

$\boxed{\mathbf{G_{ME}}}$ is a known transformation that is always a pure translation, being the position of the 3D motion device with respect to its own origin. Therefore, the $\boxed{\mathbf{R_{ME}}}$ term is known to be identity and can be removed:

$$\begin{aligned}\mathbf{G_{CF}} &= \mathbf{R_{CM}}\left(\mathbf{I_{3\times3}}\left(\mathbf{R_{ET}}\left(\boxed{\mathbf{G_{TF}}}+\mathbf{T_{ET}}\right)+\boxed{\mathbf{T_{ME}}}\right)+\mathbf{T_{CM}}\right)\\ &= \mathbf{R_{CM}}\left(\mathbf{R_{ET}}\left(\boxed{\mathbf{G_{TF}}}+\mathbf{T_{ET}}\right)+\boxed{\mathbf{T_{ME}}}+\mathbf{T_{CM}}\right)\\ &= \mathbf{R_{CM}}\left(\mathbf{R_{ET}}\cdot\boxed{\mathbf{G_{TF}}}+\mathbf{R_{ET}}\cdot\mathbf{T_{ET}}+\boxed{\mathbf{T_{ME}}}+\mathbf{T_{CM}}\right)\end{aligned} \tag{13}$$

Note that the term $\mathbf{R_{ET}} \cdot \mathbf{T_{ET}}$ is an unknown rotation, applied to an *unknown translation which appears nowhere else in the formula*. Additionally, both $\mathbf{R_{ET}}$ and $\mathbf{T_{ET}}$ are (like all of the unknown terms in the model) constant across all data points in the set. Therefore, for *any* value of $\mathbf{R_{ET}}$, there exists some equally unknown but equally constant 3-vector $\mathbf{P_{ET}}$ such that $\mathbf{P_{ET}} = \mathbf{R_{ET}} \cdot \mathbf{T_{ET}}$. Substitution into (13) gives

$$\begin{aligned}\mathbf{G_{CF}} &= \mathbf{R_{CM}}\left(\mathbf{R_{ET}}\cdot\boxed{\mathbf{G_{TF}}}+\mathbf{P_{ET}}+\boxed{\mathbf{T_{ME}}}+\mathbf{T_{CM}}\right)\\ &= \mathbf{R_{CM}}\left(\mathbf{R_{ET}}\cdot\boxed{\mathbf{G_{TF}}}+\mathbf{P_{ET}}+\mathbf{T_{CM}}+\boxed{\mathbf{T_{ME}}}\right)\end{aligned} \tag{14}$$

Now, the two unknown 3-vectors $\mathbf{P_{ET}}$ and $\mathbf{T_{CM}}$ appear nowhere else in the model. The sum of two unknown 3-vectors is, of course, another unknown 3-vector- we will call it $\mathbf{P_{CT}}$. Another substitution into (14) gives

$$\mathbf{G_{CF}} = \mathbf{R_{CM}}\left(\mathbf{R_{ET}}\cdot\boxed{\mathbf{G_{TF}}}+\mathbf{P_{CT}}+\boxed{\mathbf{T_{ME}}}\right) \tag{15}$$

Although this formulation is no longer as intuitive or physically meaningful as Eq.(11), it eliminates a set of redundant free parameters; and reduces the total dimensionality of the search space from 12 (2 unknown translations and 2 unknown rotations, each in 3 dimensions) to 9 (2 unknown rotations and 1 unknown translation). This is distinct from the position model used in Tsai's algorithm, which assumes only one aggregate transformation; that from the camera to a scene where every target point is at a known location with respect to every other.[12] This simpler model cannot accommodate multi-image calibration data where the rotation between the target and the object moving the target is not fixed or known.

Although the 3-Axis algorithm can operate on any type of fiducial using any error-minimizing solver, we chose to employ circle grid targets[13] and the Ceres optimization solver,[45] because these were the methods used by ROS-Industrial.[43]

Finally, consideration is given to alternative measures of calibration quality analysis. When calibrating *simulated* cameras that render images of a scene *in silico*, the ground truth is known. It is thus possible to derive an "*Actual* Reprojection Error" statistic (**ARE**) that distinguished between the calibrated projection model $\mathcal{P}_C$ and the ground-truth projection model $\mathcal{P}_G$.

This is done by generating a set of 3D points S; and performing pinhole camera projection on them using$\mathcal{P}_C$ and then $\mathcal{P}_G$. The average Euclidean distance (in pixels) between a point projected using the ground-truth intrinsics, and the corresponding point projected using the calibrated intrinsics, is the ARE for that calibration attempt:

$$\mathbf{ARE} = \sum_{i \in S} \big(\| \mathcal{P}_C(X_i, Y_i, Z_i) - \mathcal{P}_G(X_i, Y_i, Z_i) \| \big) \tag{16}$$

5 Experiments and Results

5.1 Simulation Experiment Design

In order to determine the effectiveness of the 3-Axis calibration system compared to the current state of the art, tests were undertaken using synthetic images generated using the Gazebo simulation engine.[46,47] One such image is shown in Figure 1, with the simulation environment producing it shown in Figue 2.

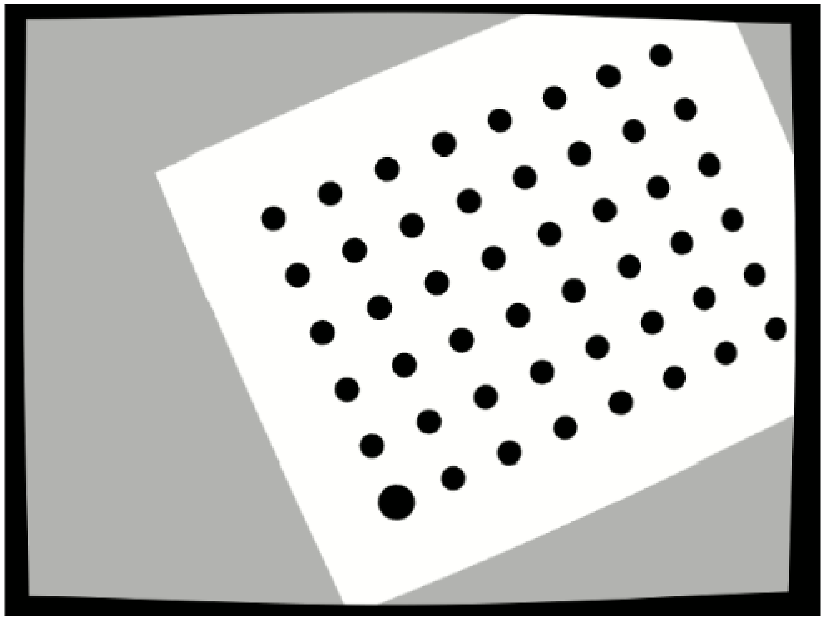

Fig 1 Simulated camera image from Figure 2. The black padding at the edges of the image is the result of barrel distortion added after the initial image render. As this image was used to collect data for Zhang's algorithm, the target is also given a psuedorandom roll, pitch, and yaw which differs for each position.

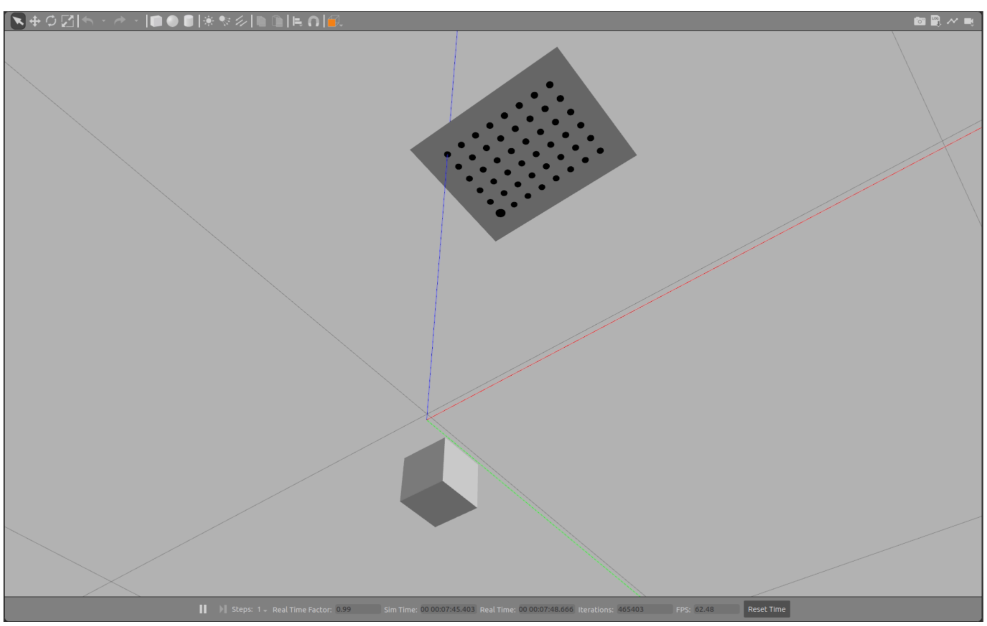

Fig 2 Simulation environment for generating target data. The camera (indicated by three colored axes) faces upward along the Z (blue) axis to the simulated target. The target can be deleted and spawned at any position in 3D space, to simulate the action of a precision 3-axis movement system.

Each synthetic calibration data set was generated containing 9024 fiducial points (a 6×8 target imaged at 188 different positions). For the ROS-Industrial data set, all of the calibration target positions were at the same, non-orthogonal angle to the camera. For the data set to be used by Zhang's method, each target position additionally included a pseudorandom pitch, yaw, and roll independently uniformly distributed between +20°and -20°. Repeated calibrations were run using less and less of this data set (pseudorandomly decimated in increments of 480 fiducial points, equivalent to randomly removing 10 images per iteration) to establish a relationship between data set size (larger data sets being more time-intensive to acquire in a real situation) and performance.

Additionally, 4 different sources of measurement imprecision were injected into the data with known, variable magnitudes in 20 increments:

- **Proportional scale of the target**, as a linear scaling of its actual dimensions versus nominal dimensions used to calculate the $\boxed{\mathbf{G_{TF}}}$ parameter given to the calibration software. *Range*: 90% – 110%
- **Detection error**, as Gaussian error added to the detected pixel position (U_i, V_i) of each fiducial. *Range*: $\sigma = 0$px – $\sigma = 10$px
- **Target motion scaling**, as a linear scaling of the target's actual positions in 3D space compared to the nominal positions $\boxed{\mathbf{G_{ME}}_i}$ provided to the calibration solver. Range: 90% – 110%
- **Target motion imprecision**, as Gaussian error added to the target's position in 3D space (compared to the nominal positions $\boxed{\mathbf{G_{ME}}_i}$ provided to the calibration solver). Range $\sigma =$ 0mm – $\sigma = 10$mm

The target detection error and target mis-scaling error sources are applicable to all three algorithms under test (Zhang, ROS-Industrial, and 3-Axis). Since Zhang's algorithm makes no assumptions as to the position of the target, the position noise and mis-scaling error sources were only applicable to ROS-Industrial and 3-Axis. This resulted in the generation of a total of 10 different 20×20 sets of calibration attempts (four each from 3-Axis and ROS-I, and two from Zhang's algorithm). For each attempt, the stated reprojection error (RE) and actual reprojection error with respect to the simulation ground truth (ARE) were computed for each attempt in the set. The ARE was calculated using the same sequence of 1000 points uniformly pseudorandomly distributed through a 1-meter cubic volume for all tests.

5.2 Simulation Experiment Results

Figure 3 shows which calibration algorithm produced a lower *actual* reprojection error (ARE) for each source of introduced data error, arranged according to number of target images included and the introduced data error magnitude. Similarly, Figure 4 shows which calibration algorithm produced a lower *reported* reprojection error. Descriptive statistics for these measurements are given in Table 2. Finally, Table 1 covers the mean variance of the estimated intrinsics from their ground-truth values for each calibration method, under each source of error.

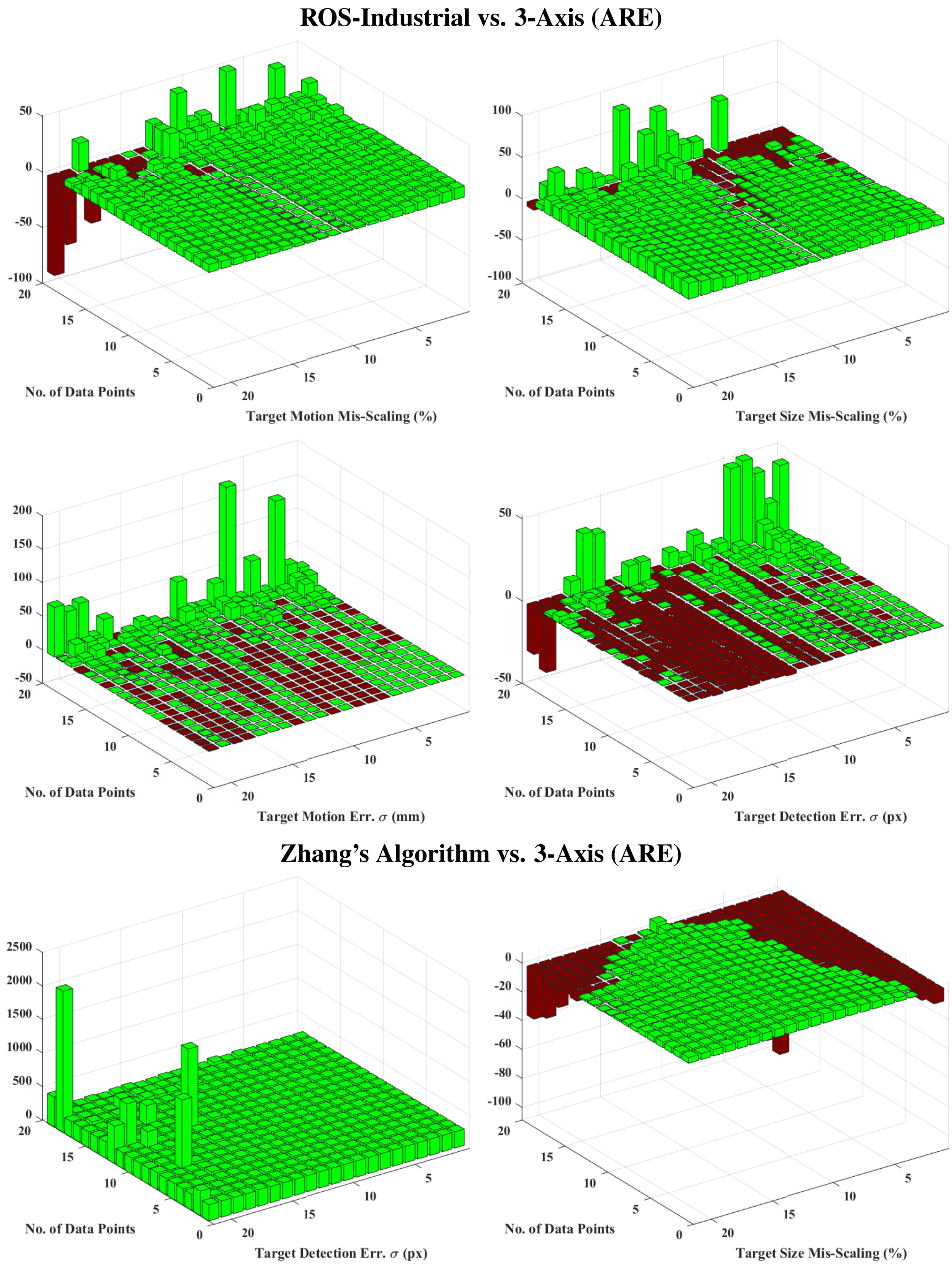

Fig 3 *Actual* reprojection error results of simulated data with different introduced flaws, arranged with respect to number of data points and introduced error magnitude. Light green bars indicate 3-Axis's ARE was lower than the competitor algorithm's in that configuration; dark red bars indicate 3-Axis's ARE was higher.

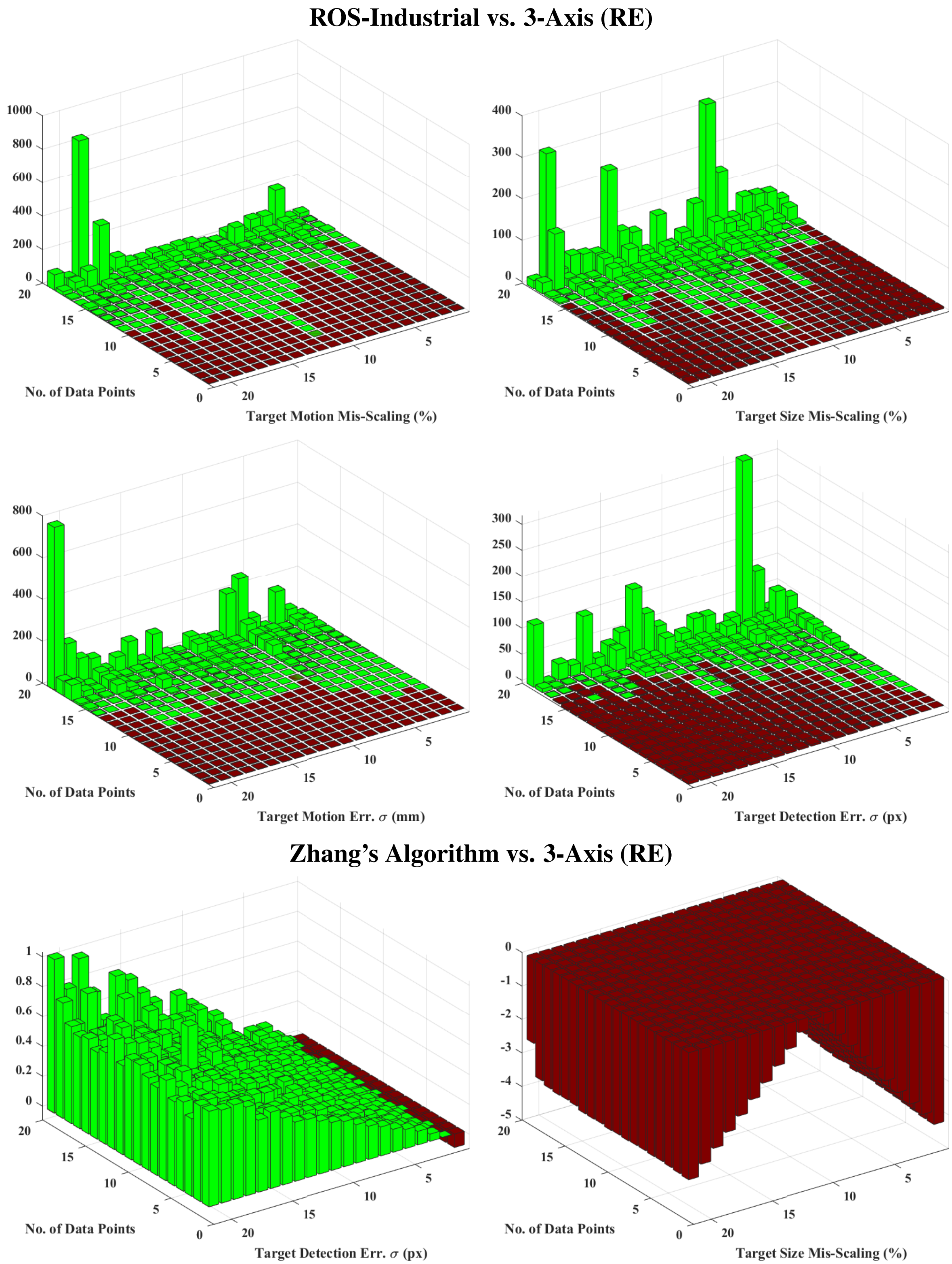

Fig 4 *Stated* reprojection error results of simulated data with different introduced flaws, arranged with respect to number of data points and introduced error magnitude. Light green bars indicate 3-Axis's RE was lower than the competitor algorithm's in that configuration; dark red bars indicate 3-Axis's RE was higher.

Table 1 Difference between median intrinsics as estimated by calibration methods, and ground truth, under different introduced calibration error sources. As Zhang's algorithm does not use known target positions, no data was collected or results computed for it using the Movement Imprecision and Movement Mis-Scaling error sources. Additionally, this implementation of Zhang's Algorithm did not return a k_3 term.

3-Axis vs. ROS-I

Introduced Flaw	*Movement Imprecision*	*Movement Mis-Scaling*	*Detection Error*	*Target Mis-Scaling*
F_x **Error Improvement**	0.9896	4.2556	0.7135	22.815
F_y **Error Improvement**	1.0554	28.642	0.6677	23.491
C_x **Error Improvement**	0.0316	-0.1699	-0.4672	-2.3580
C_y **Error Improvement**	-0.1247	-0.3356	-0.3138	-0.3952
k_1 **Error Improvement**	0.0518	0.0174	0.0324	0.0317
k_2 **Error Improvement**	0.3384	0.2593	0.1933	0.2731
k_3 **Error Improvement**	-0.0280	-0.3124	-0.0850	-0.1152
p_1 **Error Improvement**	0.0079	0.007	0.0077	0.0065
p_2 **Error Improvement**	0.7615	0.8026	0.6656	0.62070

3-Axis vs. Zhang

Introduced Flaw	*Movement Imprecision*	*Movement Mis-Scaling*	*Detection Error*	*Target Mis-Scaling*
F_x **Error Improvement**	*N/A*	*N/A*	7.9141	25.078
F_y **Error Improvement**	*N/A*	*N/A*	8.5009	25.030
C_x **Error Improvement**	*N/A*	*N/A*	0.5527	7.9029
C_y **Error Improvement**	*N/A*	*N/A*	0.2214	6.9560
k_1 **Error Improvement**	*N/A*	*N/A*	-0.0241	0.0173
k_2 **Error Improvement**	*N/A*	*N/A*	-0.0681	0.0989
k_3 **Error Improvement**	*N/A*	*N/A*	*N/A*	*N/A*
p_1 **Error Improvement**	*N/A*	*N/A*	-0.0012	0.0023
p_2 **Error Improvement**	*N/A*	*N/A*	-0.0004	0.0032

3-Axis performs better in ARE measures than either Zhang's algorithm or ROS-I, although the difference is more extreme with Zhang's algorithm while ROS-I performed comparatively better. In particular, Zhang's algorithm performed much more poorly when subjected to target detection errors, while ROS-I performed best under that same condition (indeed, the detection error test was the only one where it beat the reprojection error of 3-Axis in a majority of comparisons). In general, ROS-I was more competitive with 3-Axis when the error was one of random perturbation than when the error was one of scaling.

Stated reprojection error does does differ from *actual* calibration quality as measured by ARE in many tests. It is generally *lower* in 3-Axis than in ROS-I when the introduced error quality is smaller in magnitude, and the number of data points is low; and higher in the converse cases. When

Table 2 Actual Reprojection Error (ARE) under each of the four injected error sources, in pixels. The first section of the table provides the mean ARE for each calibration method over 400 trials; the second section provides the difference in mean ARE between 3-Axis and the control calibration methods; and the third section provides the percentage (out of 400 individual trials) where 3-Axis performed better than the control method. As Zhang's algorithm does not use known target positions, no data was collected or results computed for it using the Movement Imprecision and Movement Mis-Scaling error sources.

Introduced Flaw	*Movement Imprecision*	*Movement Mis-Scaling*	*Detection Error*	*Target Mis-Scaling*
3-Axis Mean ARE (px)	1.23	9.49	2.87	7.75
ROS-I Mean ARE (px)	4.29	14.72	3.38	14.18
Zhang Mean ARE (px)	*N/A*	*N/A*	258.16	7.483
Improvement vs. ROS-I	3.05	5.22	0.51	6.43
Improvement vs. Zhang	*N/A*	*N/A*	255.30	-0.268
Cases Where 3-Axis Beat ROS-I	64.16%	94.23%	49.12%	91.97%
Cases Where 3-Axis Beat Zhang	*N/A*	*N/A*	100%	71.67%

comparing 3-Axis against Zhang's algorithm, 3-Axis improves on RE as detection error increases, and is always higher when the target is incorrectly scaled. This higher stated reprojection error is actually useful for a calibration algorithm, as it provides a more reliable indication that there is a flaw in the data being acquired (as opposed to generating an *in*accurate calibration with a reported reprojection error comparable to that of an *accurate* calibration).

5.3 Physical Experiment Design

Calibration was also performed on a physical device, specifically the surgical endoscope of a da Vinci® IS-1200 Surgical Robotic System.[48] This device was selected because of its compact size and relatively high distortion, and difficulties previously encountered in attaining any reliable calibration from it.[44]

As there is no ground truth available for this physical device, the quality of the calibration must be evaluated by stated reprojection error and consistency of results across calibration runs. Five calibrations were performed, using a preexisting 3-axis controllable sled mechanism moved to 153 valid detection points for C3 and ROS-I. The calibration target contained 9 by 8 circular fiducials, for a total of 11016 data points in each calibration. Another 153 images were acquired using the same target, moved in a single line along the Z-axis; these serve as a data set for the ROS-Industrial calibration system. The sled mechanism is shown in Figure 6, and one of the images produced by the endoscope camera in Figure 5. Finally, the first 153 valid random target positions from the simulation data set were used as references to manually position the physical target (with added rotations, which could not be performed using the sled system), and establish another 11016 data point set that was used by Zhang's algorithm. These acquisitions were also repeated 5 times each.

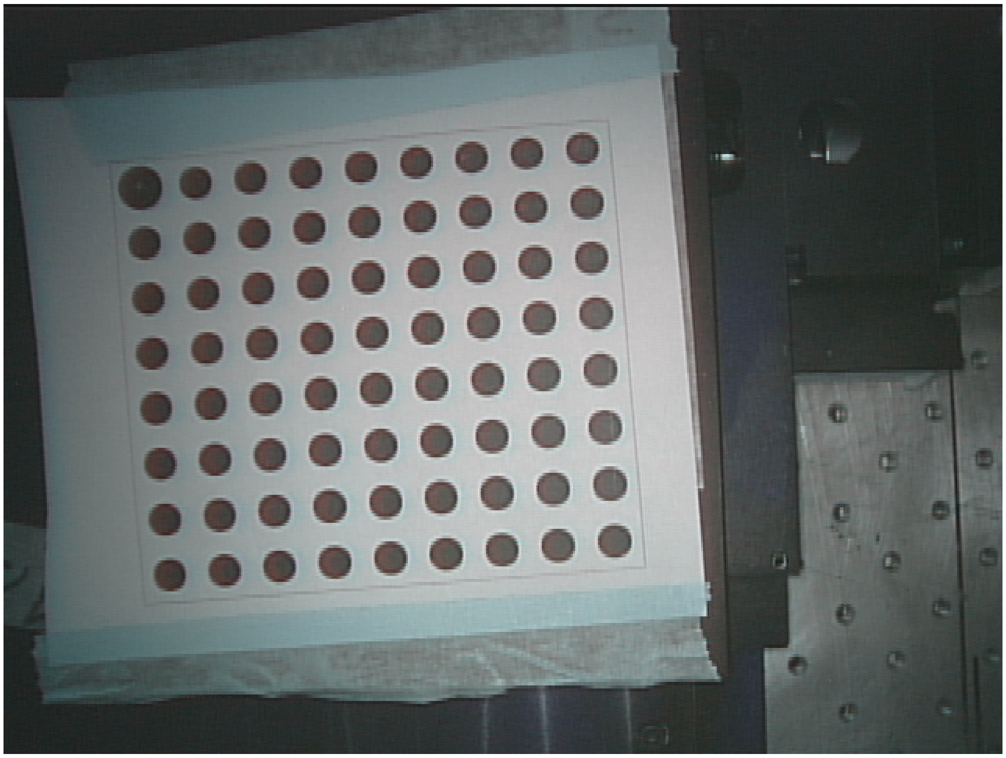

Fig 5 Example calibration image taken with the endoscope camera.

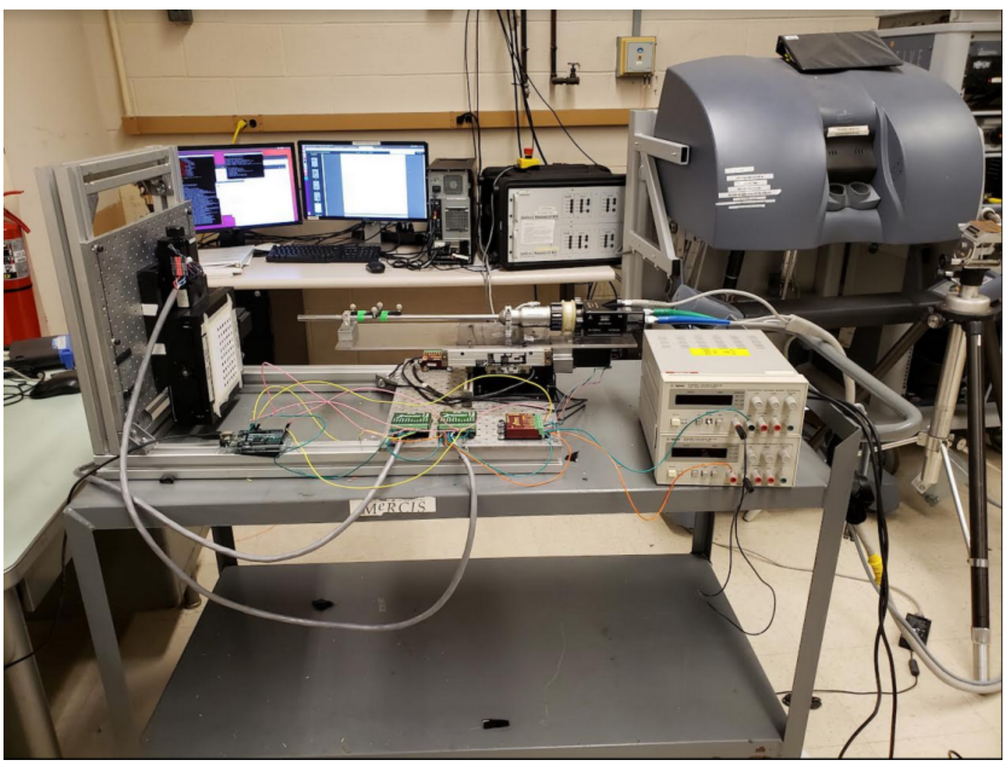

Fig 6 Calibration target movement system for endoscope (silver cylinder in the center of the image) in 3-Axis and ROS-I calibration. This benchtop device functions similarly to a CNC machine.

5.4 Physical Experiment Results

Table 3 displays the reprojection errors of all five tests for the 3-Axis, ROS-I, and Zhang calibration systems. Table 4 shows comparative variance statistics for each intrinsic parameter of interest, for each algorithm. ROS-I did not converge for three of the five tests, although in the two cases where it did converge it produced reprojection errors that were lower than any of those reported by 3-Axis. Similarly, in two cases Zhang's algorithm produced reprojection errors that were an order of magnitude higher than the other three, although even those were under one pixel. 3-Axis, however, consistently reported sub-pixel reprojection errors for all five tests.

Table 3 Reprojection errors under physical calibration (RPE, in pixels).

Test #	*3-Axis*	*ROS-I*	*Zhang*
1	0.9762	0.1917	4.2612
2	0.5984	84.48	3.1412
3	0.7502	0.2430	0.2482
4	0.6567	400.0	0.2905
5	0.5554	115.9	0.2891
Avg.	0.7074	120.10	1.6461

Table 4 Standard deviation statistics for intrinsics discovered under physical calibration. The studied implementation of Zhang's algorithm did not estimate a k_3 term.

Term	*3-Axis* σ	*ROS-I* σ	*Zhang* σ
F_x	0.4800	24.85	343.4
F_y	0.7908	33.40	345.9
C_x	2.707	2.298	27.34
C_y	9.602	0.7787	12.01
k_1	0.01277	0.1840	0.1263
k_2	0.1734	0.7897	0.5303
k_3	0.3704	0.006109	*N/A*
p_1	0.001947	0.005730	0.005310
p_2	0.002442	0.3967	0.004089

3-Axis is more consistent than ROS-I for 6 out of the 9 test parameters, and more consistent than Zhang's algorithm for 6 out of 8 (this implementation of Zhang's algorithm does not report the k_3 term).

The simulation results discussed above suggest that reported reprojection error is not always a particularly reliable measure of actual calibration quality, but some inferences can still be drawn from the results. In particular, despite the ROS-I and 3-Axis data sets both being acquired using the same mechanical device moved to the same positions, ROS-I was not *reliably* able to converge on camera intrinsics with a subpixel reported RE. Additionally, while Zhang's algorithm did not produce as large of spikes in reprojection error as ROS-I, it still demonstrated "good" and "bad" calibrations when fed similar data sets. 3-Axis, on the other hand, behaved consistently over all five calibration attempts.

6 Conclusions

We discussed a new intrinsic calibration method ("3-Axis") based on known movements of the fiducial along three orthogonal axes. We began with an overview of the current state of the art, and identified the ROS-Industrial calibration system[6] and Zhang's calibration algorithm[7] as standards of performance to which 3-Axis would be compared. Next, we described in detail the operation of the calibration algorithm. We then proposed experimental methods of comparison between ROS-Industrial, Zhang's algorithm, and 3-Axis, using both simulated and real cameras. Finally, we presented the resultant performance metrics for all systems. We concluded that 3-Axis is more likely to provide more accurate calibration results than either ROS-Industrial or Zhang's algorithm, for a calibration data set of any given size and quality. Additionally, it produces more *consistent* calibration results than ROS-Industrial or Zhang's algorithm for any given real-world data set. Finally, 3-Axis is shown to produce *stated* reprojection errors that more closely match the *actual* reprojection errors of the algorithm than ROS-I or Zhang's algorithm. This means that the performance of the algorithm can be assessed more accurately in cases where it is not possible to compare the discovered values to a ground truth. Continued validation of the work will occur as it is used in functional calibrations of different camera hardware.

The 3-Axis utility itself is available at https://github.com/cwru-robotics/3d-calibration.

Scripts and assets used to perform automatic simulation tests, as well as the data produced, are available at https://github.com/cwru-robotics/comparative-calibration.

Software used to drive and interface with the endoscope custom mill platform is available at https://github.com/cwru-robotics/Calib-Sled

Disclosures

The authors declare no conflicts of interest.

Code, Data, and Materials Availability

Data underlying the results presented in this paper are available in Ref.[49]

Acknowledgments

This work was supported in part by the National Science Foundation under grants CISE IIS-1524363, CISE IIS-1563805, and SBE SES-2129072; and by Roadprintz, Inc.

References

1 O. Faugeras and O. A. Faugeras, *Three-dimensional computer vision: a geometric viewpoint*, MIT press (1993).

2 R.-J. Ahlers and J. Lu, “Stereoscopic vision-an application oriented overview,” in *Optics, Illumination, and Image Sensing for Machine Vision IV*, **1194**, 298–308, SPIE (1990).

3 A. Casals, *Sensor devices and systems for robotics*, vol. 52, Springer Science & Business Media (2012).

4 J. Salvi, X. Armangué, and J. Batlle, “A comparative review of camera calibrating methods with accuracy evaluation,” *Pattern Recognition* **35**(7), 1617–1635 (2002).

5 W. Li, T. Gee, H. Friedrich, *et al.*, “A Practical Comparison between Zhang’s and Tsai’s Calibration Approaches,” in *Proceedings of the 29th International Conference on Image and Vision Computing New Zealand*, *IVCNZ ’14*, 166–171, Association for Computing Machinery, (New York, NY, USA) (2014).

6 “ROS Industrial.”

7 Z. Zhang, “A flexible new technique for camera calibration,” *IEEE Transactions on pattern analysis and machine intelligence* **22**(11), 1330–1334 (2000).

8 W. Qi, F. Li, and L. Zhenzhong, “Review on camera calibration,” in *2010 Chinese Control and Decision Conference*, 3354–3358 (2010).

9 T. Luhmann, C. Fraser, and H.-G. Maas, “Sensor modelling and camera calibration for close-range photogrammetry,” *ISPRS Journal of Photogrammetry and Remote Sensing* **115**, 37–46 (2016). Theme issue ’State-of-the-art in photogrammetry, remote sensing and spatial information science’.

10 L. Long and S. Dongri, "Review of camera calibration algorithms," in *Advances in Computer Communication and Computational Sciences*, S. K. Bhatia, S. Tiwari, K. K. Mishra, *et al.*, Eds., 723–732, Springer Singapore, (Singapore) (2019).

11 Y. Abdel-Aziz and H. Karara, "Direct linear transformation into object space coordinates in close-range photogrammetry, in proc. symp. close-range photogrammetry," *Urbana-Champaign* , 1–18 (1971).

12 R. Y. Tsai, "An efficient and accurate camera calibration technique for 3D machine vision," in *CVPR'86*, (1986).

13 Opencv Dev Team, "findcirclesgrid," *Opencv API Reference* (2019).

14 Y. M. Wang, Y. Li, and J. B. Zheng, "A camera calibration technique based on OpenCV," in *The 3rd International Conference on Information Sciences and Interaction Sciences*, 403–406 (2010).

15 MathWorks, Inc., "What is camera calibration?," *Matlab Help Center* (2022).

16 G. D'Emilia and D. D. Gasbarro, "Review of techniques for 2D camera calibration suitable for industrial vision systems," *Journal of Physics: Conference Series* **841**, 012030 (2017).

17 K. Zhan, D. Fritsch, and J. F. Wagner, "Stability analysis of intrinsic camera calibration using probability distributions," *IOP Conference Series: Materials Science and Engineering* **1048**, 012010 (2021).

18 G. Chiou, "Reducing the variance of intrinsic camera calibration results in the ros camera calibration package," Master's thesis, University of Texas at San Antonio (2017).

19 J. Jiang, L. Zeng, B. Chen, *et al.*, "An accurate and flexible technique for camera calibration," *Computing* **101**(12), 1971–1988 (2019).

20 J.-H. Chuang, C.-H. Ho, A. Umam, *et al.*, “Geometry-based camera calibration using closed-form solution of principal line,” *IEEE Transactions on Image Processing* **30**, 2599–2610 (2021).

21 X. Wang, Y. Zhao, and F. Yang, “Camera calibration method based on Pascal’s theorem,” *International Journal of Advanced Robotic Systems* **16**(3), 1729881419846406 (2019).

22 D. Jin and Y. Yang, “Using distortion correction to improve the precision of camera calibration,” *Optical Review* **26**(2), 269–277 (2019).

23 M. Yang, X. Chen, and C. Yu, “Camera calibration using a planar target with pure translation,” *Appl. Opt.* **58**, 8362–8370 (2019).

24 M. Tapper, P. J. McKerrow, and J. Abrantes, “Problems encountered in the implementation of tsai’s algorithm for camera calibration,” in *Proc. 2002 Australasian Conference on Robotics and Automation, Auckland*, 66–70, Citeseer (2002).

25 J. Zhang, H. Yu, H. Deng, *et al.*, “A robust and rapid camera calibration method by one captured image,” *IEEE Transactions on Instrumentation and Measurement* **68**(10), 4112–4121 (2019).

26 H. Zhu, Y. Li, X. Liu, *et al.*, “Camera calibration from very few images based on soft constraint optimization,” *Journal of the Franklin Institute* **357**(4), 2561–2584 (2020).

27 Z. Zhang, R. Zhao, E. Liu, *et al.*, “A single-image linear calibration method for camera,” *Measurement* **130**, 298–305 (2018).

28 M. Lopez, R. Mari, P. Gargallo, *et al.*, “Deep single image camera calibration with radial distortion,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (2019).

29 D. Su, A. Bender, and S. Sukkarieh, “Improved cross-ratio invariant-based intrinsic calibration of a hyperspectral line-scan camera,” *Sensors* **18**(6) (2018).

30 Z. Lu and L. Cai, “Camera calibration method with focus-related intrinsic parameters based on the thin-lens model,” *Opt. Express* **28**, 20858–20878 (2020).

31 J. Kümmerle and T. Kühner, “Unified intrinsic and extrinsic camera and LiDAR calibration under uncertainties,” in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, 6028–6034 (2020).

32 G. H. An, S. Lee, M.-W. Seo, *et al.*, “Charuco board-based omnidirectional camera calibration method,” *Electronics* **7**(12) (2018).

33 H. Duan, L. Mei, J. Wang, *et al.*, “A new imaging model of Lytro light field camera and its calibration,” *Neurocomputing* **328**, 189–194 (2019). Chinese Conference on Computer Vision 2017.

34 Y. Zhang, X. Zhao, and D. Qian, “Learning-based distortion correction and feature detection for high precision and robust camera calibration,” *IEEE Robotics and Automation Letters* **7**(4), 10470–10477 (2022).

35 S. Sels, B. Ribbens, S. Vanlanduit, *et al.*, “Camera calibration using Gray code,” *Sensors* **19**(2) (2019).

36 B. Chen and B. Pan, “Camera calibration using synthetic random speckle pattern and digital image correlation,” *Optics and Lasers in Engineering* **126**, 105919 (2020).

37 S. Peng and P. Sturm, “Calibration wizard: A guidance system for camera calibration based on modelling geometric and corner uncertainty,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, (2019).

38 M. A. Gunen, E. Besdok, P. Civicioglu, *et al.*, "Camera calibration by using weighted differential evolution algorithm: a comparative study with ABC, PSO, COBIDE, DE, CS, GWO, TLBO, MVMO, FOA, LSHADE, ZHANG and BOUGUET," *Neural Computing and Applications* **32**(23), 17681–17701 (2020).

39 C. Chen, F. Yang, and B. Pan, "Model-free method for intrinsic camera parameters calibration and lens distortion correction," *Optical Engineering* **60**(10), 104102 (2021).

40 R. Juarez-Salazar, J. Zheng, and V. H. Diaz-Ramirez, "Distorted pinhole camera modeling and calibration," *Appl. Opt.* **59**, 11310–11318 (2020).

41 J. Sun, X. Cheng, and Q. Fan, "Camera calibration based on two-cylinder target," *Opt. Express* **27**, 29319–29331 (2019).

42 X. Liu, Y. Zhao, and X. Kou, "Calibration of intrinsic camera parameters with a conic and its asymptotes," *Appl. Opt.* **60**, 10024–10034 (2021).

43 "Ros Industrial camera calibration."

44 T. Shkurti, "Simulation and control enhancements for the da Vinci surgical robot," Master's thesis, Case Western Reserve University (2019).

45 S. Agarwal, K. Mierle, and T. C. S. Team, "Ceres Solver," (2022).

46 C. Aguero, N. Koenig, I. Chen, *et al.*, "Inside the virtual robotics challenge: Simulating real-time robotic disaster response," *Automation Science and Engineering, IEEE Transactions on* **12**, 494–506 (2015).

47 N. Koenig and A. Howard, "Design and use paradigms for Gazebo, an open-source multi-robot simulator," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2149–2154, (Sendai, Japan) (2004).

48 P. Kazanzides, Z. Chen, A. Deguet, *et al.*, "An open-source research kit for the da Vinci® surgical system," in *IEEE International Conference on Robotics and Automation (ICRA)*, (2014).

49 "Comparative calibration data." https://github.com/cwru-robotics/comparative-calibration.

List of Figures

List of Tables

2 Actual Reprojection Error (ARE) under each of the four injected error sources, in pixels. The first section of the table provides the mean ARE for each calibration method over 400 trials; the second section provides the difference in mean ARE between 3-Axis and the control calibration methods; and the third section provides the percentage (out of 400 individual trials) where 3-Axis performed better than the control method. As Zhang's algorithm does not use known target positions, no data was collected or results computed for it using the Movement Imprecision and Movement Mis-Scaling error sources.

3 Reprojection errors under physical calibration (RPE, in pixels).

4 Standard deviation statistics for intrinsics discovered under physical calibration. The studied implementation of Zhang's algorithm did not estimate a k_3 term.