

Multi-camera calibration with pattern rigs, including for non-overlapping cameras: CALICO

Amy Tabb¹, Henry Medeiros², Mitchell J. Feldmann³, and Thiago T. Santos⁴

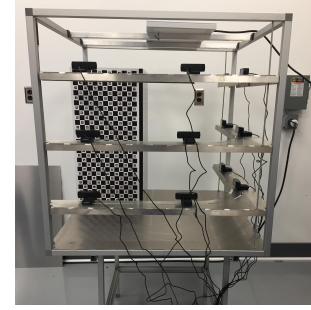
Abstract—This paper describes CALICO, a method for multi-camera calibration suitable for challenging contexts: stationary and mobile multi-camera systems, cameras without overlapping fields of view, and non-synchronized cameras. Recent approaches are roughly divided into infrastructure- and pattern-based. Infrastructure-based approaches use the scene’s features to calibrate, while pattern-based approaches use calibration patterns. Infrastructure-based approaches are not suitable for stationary camera systems, and pattern-based approaches may constrain camera placement because shared fields of view or extremely large patterns are required.

CALICO is a pattern-based approach, where the multi-calibration problem is formulated using rigidity constraints between patterns and cameras. We use a *pattern rig*: several patterns rigidly attached to each other or some structure. We express the calibration problem as that of algebraic and reprojection error minimization problems. Simulated and real experiments demonstrate the method in a variety of settings. CALICO compared favorably to Kalibr. Mean reconstruction accuracy error was ≤ 0.71 mm for real camera rigs, and ≤ 1.11 for simulated camera rigs. Code and data releases are available at [1] and <https://github.com/amy-tabb/calico>.¹

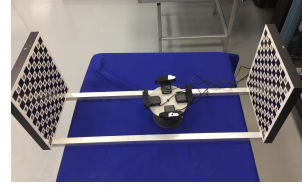
I. INTRODUCTION

Multi-camera calibration is necessary for a variety of activities, from human activity detection and recognition to reconstruction tasks. Internal parameters can typically be computed by waving a calibration target in front of cameras, and then using Zhang’s algorithm [2] to calibrate individual cameras. Determining cameras’ external parameters, or the relationships between cameras in a multi-camera system, may be a more difficult problem as methods for computing these relationships depend strongly on characteristics of the hardware and the arrangement of the cameras. For instance, the cameras’ shared field of view, level of synchronization, and mobility strongly influence the ease of multi-camera calibration or choice of methods available for calibration.

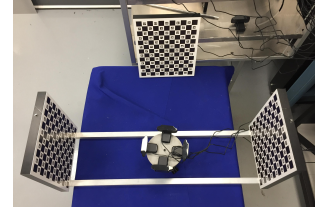
This work describes CALICO,² a method for multi-camera calibration. We assume that cameras can be triggered to acquire images of a stationary calibration rig, or the cameras are synchronized. Our method uses calibration objects based on one or more planar patterns [4] that allow for partial pattern detection, allows non-overlapping fields of view, and allows for mobile or stationary multi-camera systems. Updated



(a) Datasets Net-1, Net-2; *box* type



(b) Dataset Mult-1; *robot* type



(c) Dataset Mult-2; *robot* type

Fig. 1. **Best viewed in color.** Cameras and patterns for two types of real experiments, *box* and *robot*. The top row, (a) is an arrangement of cameras mounted on the periphery of a box where cameras point towards the box’s interior; the bottom row, (b) and (c) are multi-camera camera rigs where cameras point away from each other. In the top figure, there are 12 cameras. The pattern rig is moved within the workspace, and cameras acquire an image for every move. The bottom row shows a multicamera system, in (b) only two cameras are used while in (c) all four cameras are used. In the multicamera system, the multicamera system is moved and the patterns are stationary. Each of these experiment types are challenging to calibrate using existing methods. In the *box* type, cameras share a common field of view but cameras on one side of the box may not be able to view the same planar pattern as cameras on another side. In the *robot* configuration, there may be no shared field of view between cameras. CALICO approximately solves the set of constraints resulting from the camera, pattern, and time relationships in such datasets.

datasets and code described in this paper are at provided at [1] and <https://doi.org/10.5281/zenodo.3520865>. Prior versions of this work are available on arXiv [5].

Given these preliminaries, our contributions to the state-of-the-art consist of:

- 1) A formulation of the multi-camera calibration problem as a set of rigidity constraints imposed by the transformations between cameras and calibration patterns.
- 2) The multi-camera calibration problem is solved efficiently by iteratively solving for variables using closed-form methods, minimizing algebraic error over the the set of rigidity constraints, and minimization of reprojection errors.

II. RELATED WORK

Recent work on multi-camera and/or multi-sensor calibration can be divided into pattern- versus infrastructure-based

¹USDA-ARS-AFRS amy.tabb@usda.gov

²University of Florida hmedeiros@ufl.edu

³University of California-Davis mjfeldmann@ucdavis.edu

⁴Embrapa Digital Agriculture thiago.santos@embrapa.br

¹Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture. USDA is an equal opportunity provider and employer.

²CALICO: CALibration of asynchronous Camera netWOorks. Acronym generated by [3] for a prior version of this work.

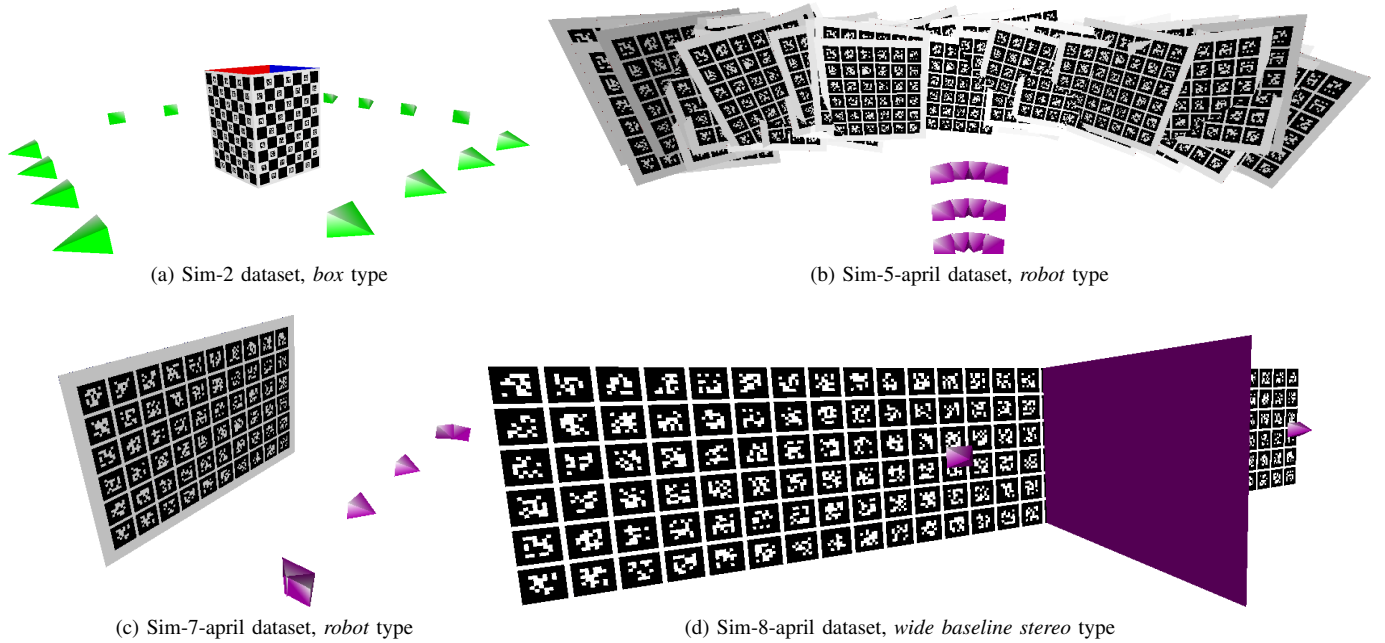


Fig. 2. **Examples of simulated datasets.** Arrangements of cameras for four of nine simulated datasets. (a) shows cameras as pyramids arranged on the edges of a square, and a pattern rig of four charuco patterns at one time step, a *box* type experiment. (b) shows cameras as pyramids and the placement of April Tag calibration patterns over all time steps, a *robot* type experiment. (c) has cameras from dataset Sim-7-april and one April Tag calibration pattern, another *robot* type experiment. (d) has the two cameras from Sim-8-april and a very large and long April Tag calibration pattern; there is a barrier so the cameras each view mutually exclusive portions of the pattern.

methods. Other differences between methods include requirements about fully or partially shared fields of view, mobile or stationary camera systems, whether camera synchronization is needed, and how the rigid constraints of a multi-camera system are represented and computed. Many works use existing solutions to the hand-eye and hand-eye, robot-world calibration methods, hand-eye, robot-world methods are reviewed in Shah *et al.* 2012 [6].

Pattern methods. Pattern-based methods require calibration patterns. Li *et al.* 2013 [7] describes a new planar pattern type that allows for partial pattern detection and pose estimation; non-overlapping views are possible, but at least two cameras need to see part of one pattern at each image acquisition. Liu *et al.* 2016 [8] (Caliber) do not specify the pattern type; instead the user designates the calibration rig and scene via a kinematic tree. The final step of both [7] and [8] is reprojection error minimization.

The Kalibr toolbox, Furgale *et al.* 2013 [9], May *et al.* 2013 [10], Maye 2014 [11], estimates rigidity constraints between cameras via an information theoretic measure. Kalibr requires the use of a calibration pattern that can be seen by pairs of cameras; the implementation released by the authors uses April Tags [12].

Infrastructure methods. Infrastructure methods use the sensed environment to calibrate a multi-camera system, usually through an existing Structure-from-Motion framework, such as COLMAP [13] or a version of SLAM. Esquivel *et al.* 2007 [14] and Lébraly *et al.* 2010 [15] create maps for individual cameras, then enforce rigidity constraints over time steps by closed-form hand-eye calibration solutions. Esquivel *et al.* 2007 [14] perform a global optimization for algebraic

error, while Lébraly *et al.* 2010 [15] minimize for reprojection error via bundle adjustment that incorporates the camera rig’s rigidity constraints. Carrera *et al.* 2011 [16] fuse maps from individual cameras and compute transformations between maps to estimate rigidity constraints, followed by a bundle adjustment with rigidity constraints similar to [15].

Lin *et al.* 2020 [17] is an infrastructure-based method that uses the 1D radial camera model [18] to estimate extrinsic parameters for each image in a sequence, up to an unknown forward translation (meaning the z component of a translation vector is unknown). Camera rig transformations from the map to calibration rig are estimated in a local assign-and-estimate loop, followed by stages of non-linear least squares minimization to refine extrinsic parameters, determine camera intrinsics and forward translation. A final step minimizes for reprojection error and optimizes all parameters (intrinsics, extrinsics, and rig poses).

Hybrid methods. Several methods use extra hardware for calibration in challenging settings. Mishra *et al.* 2022 [19] calibrate ceiling-mounted cameras using a mobile robot with its own camera and calibration pattern. Chen and Schwertfeger 2019 [20] calibrate a multi-sensor system of cameras and 3D lidar units with the aid of patterns and a camera-based motion capture system. A hand-held camera estimates a map using SLAM to calibrate a surveillance camera network in Pollok and Monari 2016 [21]. Robinson *et al.* 2017 [22] calibrate cameras with non-overlapping views by adding intermediate cameras such that a planar calibration pattern could be seen by pairs of cameras.

Our work, CALICO, formulates the calibration problem and rigidity constraints in a manner that is most similar to

Esquivel *et al.* 2007 [14] and Lébraly *et al.* 2010 [15] in that the transformation between frames is estimated. Instead of a camera-centered gauge and infrastructure-based approach as in those works, we use an object-centered gauge and a pattern-based approach. CALICO also shares some similarities with Liu *et al.* 2016 [8] in that multiple calibration targets may be used, and the technique is flexible in that stationary or mobile multi-camera systems may be calibrated. Our work is different from [8] in that we do not require users specify of a kinematic tree; in CALICO, the relationships between patterns and cameras are estimated as part of the method.

III. MULTI-CAMERA CALIBRATION PROBLEM FORMULATION

The multiple-camera calibration problem is the computation of relative poses between cameras, or, camera poses with respect to a coordinate system. These poses are represented by homogeneous transformation matrices (HTMs): 4×4 matrices composed of a 3×3 orthogonal rotation matrix $\mathbf{R}$ and a column vector $\mathbf{t}$ of size 3 representing the translation, or

$$\begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0}^T & 1 \end{bmatrix}. \quad (1)$$

We consider multi-camera calibration in a pattern-based, non-synchronized context. Camera poses are computed relative to a coordinate system; we assume internal camera calibration parameters are available or can be computed from an image dataset. Table I shows the notation used in this paper. There are n_c cameras, which are rigidly attached to each other, or rigid surfaces, and referred to as the **camera rig**. In our formulation, we want to compute the n_c transformations from a world coordinate system to each camera, HTMs $\mathbb{C}$.

There are n_p calibration patterns rigidly attached to each other, or rigid surfaces; the pattern set is referred to as the **pattern rig**.³ Each pattern in the set is distinguishable from other patterns, and images of a partial pattern can be used to estimate the camera pose with respect to the pattern.

All cameras capture an image at each of n_t time labels. Synchronized image acquisition is not required. After an image capture for all cameras, the pattern rig, camera rig, or both, are moved and images acquired to create data for the next time label. Variations to this data acquisition procedure are possible and discussed in Section V-D.

A. Representation of rigid constraints: constraint set $\mathbb{CS}$

We define camera HTMs $\mathbb{C}$ with respect to a world coordinate system w (discussed in Section III-C). We use extra unknown variables, pattern and time HTM sets $\mathbb{P}$ and $\mathbb{T}$. Variables from $\mathbb{C}$, $\mathbb{P}$, and $\mathbb{T}$ are indexed by $\mathcal{C}$, $\mathcal{P}$, and $\mathcal{T}$, respectively. $\mathbb{P}$'s HTMs represent the transformations between the world coordinate system w and the pattern rig's individual patterns; $\mathbb{T}$'s HTMs represent the transformations between the world coordinate system w to its pose at different time labels. While in the formal listing of $\mathbb{P}$ in Table I, line 10, time sub-

³It may be more standard to refer to the pattern rig as the calibration rig; we chose a name that had a different first letter to reduce possible confusion with camera rig.

TABLE I
NOTATION USED IN THIS PAPER.

	Notation	Meaning
1	n_c	Number of cameras
2	n_p	Number of patterns
3	n_t	Number of time labels
4	w	World coordinate frame
5	$\mathcal{C} = \{c_0, c_1, \dots, n_c - 1\}$	Set of camera indices
6	$\mathcal{P} = \{p_0, p_1, \dots, n_p - 1\}$	Set of pattern indices
7	$\mathcal{T} = \{t_0, t_1, \dots, n_t - 1\}$	Set of time label indices
8	$\mathbb{C} = \{c_i \mathbf{C}_w \in SE(3) c_i \in \mathcal{C}\}$	HTMs from world coordinate frame to cameras
9	$\mathbb{T} = \{w @ t_j \mathbf{T}_w \in SE(3) t_j \in \mathcal{T}\}$	HTMs from world coordinate frame to time labels
10	$\mathbb{P} = \{p_k @ t_j \mathbf{P}_{w @ t_j} \in SE(3) p_k \in \mathcal{P}, \{t_j, t_m\} \in \mathcal{T}, p_k @ t_j \mathbf{P}_{w @ t_j} = p_k @ t_m \mathbf{P}_{w @ t_m}\}$	HTMs from world coordinate system to patterns
11	$\mathbb{A} = \{c_i @ t_j \mathbf{A}_{w @ p_k}^{w @ t_j} \in SE(3) c_i \in \mathcal{C}, p_k \in \mathcal{P}, t_j \in \mathcal{T}\}$	HTMs relating pattern p_j to camera c_i at time t_k .
12	$c_i \mathbf{C} = c_i \mathbf{C}_w$	Abbreviated $\mathbb{C}$ notation
13	$t_j \mathbf{T} = t_j \mathbf{T}_w$	Abbreviated $\mathbb{T}$ notation
14	$p_k \mathbf{P} = p_k @ t_j \mathbf{P}_{w @ t_j}$	Abbreviated $\mathbb{P}$ notation
15	$c_i \mathbf{A}_{p_k}^{t_j} = c_i @ t_j \mathbf{A}_{w @ p_k}^{w @ t_j}$	Abbreviated $\mathbb{A}$ notation
16	$\mathbb{CS} = \{c_i \mathbf{C} = c_i \mathbf{A}_{p_k}^{t_j} p_k \mathbf{P} t_j \mathbf{T} c_i \in \mathcal{C}, p_k \in \mathcal{P}, t_j \in \mathcal{T}\}$	Constraint set for multi-camera calibration
17	$\mathbb{CS}' \subseteq \mathbb{CS}$	Constraint subset of $\mathbb{CS}$ where all $\mathbb{C}$, $\mathbb{P}$, $\mathbb{T}$ variables are initialized
18	$r_{ae} \in [0, 1]$	proportion of constraints added to the algebraic error minimization.
19	$r_{rp} \in [0, 1]$	proportion of constraints added to the reprojection error minimization.
20	ae	Total algebraic error over $\mathbb{CS}$
21	re	Total reprojection error over $\mathbb{CS}$
22	$rrmse$	Root reprojection mean square error over $\mathbb{CS}$
23	X	3D point on a calibration pattern
24	x	2D image point representing calibration pattern point X
25	$\mathbb{X}$	Set of world-to-image calibration pattern point pairs (X, x)
26	$\mathbb{Y}$	Set of reconstructed calibration pattern points

and superscripts are included, $|\mathbb{P}| = |\mathcal{P}|$; there is one HTM in $\mathbb{P}$ for every pattern in the pattern rig.

We also use observation HTMs $\mathbb{A}$ in our formulation; $\mathbb{A}$'s HTMs represent the pose of a camera with respect to a pattern at a particular time label. In other words, when camera c_i views pattern p_k at time t_j , the transformation from the coordinate system defined by pattern p_k to camera c_i is estimated through Perspective-n-Point (PnP) pose computation as is typical in single-camera calibration. The resulting pose is notated as $c_i @ t_j \mathbf{A}_{w @ p_k}^{w @ t_j}$.

With the unknown variables $\mathbb{C}$, $\mathbb{P}$, and $\mathbb{T}$, and observations $\mathbb{A}$, we create a set of constraints, $\mathbb{CS}$, with form

$$c_i \mathbf{C}_w = c_i @ t_j \mathbf{A}_{w @ p_k}^{w @ t_j} p_k @ t_j \mathbf{P}_{w @ t_j} t_j \mathbf{T}_w. \quad (2)$$

and this constraint is represented in a graph in Figure 3.

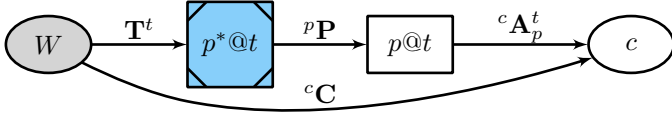


Fig. 3. Graphical representation of one constraint in $\mathcal{CS}$.

The notation above is quite long, so we use abbreviated forms for the remainder of this paper (Table I, lines 12-15). Then Eq. 2 is written as

$${}^{c_i}\mathbf{C} = {}^{c_i}\mathbf{A}_{p_k}^{t_j} {}^{p_k}\mathbf{P} {}^{t_j}\mathbf{T}. \quad (3)$$

Example 1. Consider a subset of constraints from a multi-camera calibration problem in Eqs. 4-7. Notice that camera c_0 has one observation at time t_0 in Eq. 4 and two observations at t_1 in Eqs. 5 and 6, as two patterns (p_0, p_1) were within camera c_0 's field of view and consequently two observations are generated. On the other hand, camera c_1 only observed p_1 at time t_1 .

$${}^{c_0}\mathbf{C} = {}^{c_0}\mathbf{A}_{p_0}^{t_0} {}^{p_0}\mathbf{P} {}^{t_0}\mathbf{T} \quad (4)$$

$${}^{c_0}\mathbf{C} = {}^{c_0}\mathbf{A}_{p_0}^{t_1} {}^{p_0}\mathbf{P} {}^{t_1}\mathbf{T} \quad (5)$$

$${}^{c_0}\mathbf{C} = {}^{c_0}\mathbf{A}_{p_1}^{t_1} {}^{p_1}\mathbf{P} {}^{t_1}\mathbf{T} \quad (6)$$

$${}^{c_1}\mathbf{C} = {}^{c_1}\mathbf{A}_{p_1}^{t_1} {}^{p_1}\mathbf{P} {}^{t_1}\mathbf{T} \quad (7)$$

B. Non-linear cost functions: algebraic and reprojection error

With Eq. 3 generated for every pattern detection in the data set, we consider the multi-camera calibration problem as an optimization problem: with initial solutions, we minimize the algebraic error of the constraint set (Eq. 8), and then refine the estimates for $\mathbb{C}, \mathbb{P}, \mathbb{T}$ by minimizing for reprojection error.

Algebraic error. The algebraic error cost function represents the difference between the left and right sides of a constraint, summed over all constraints in $\mathcal{CS}$,

$$\operatorname{argmin}_{\mathbb{C}, \mathbb{P}, \mathbb{T}} \sum_{\mathcal{CS}} \| {}^{c_i}\mathbf{C} - {}^{c_i}\mathbf{A}_{p_k}^{t_j} {}^{p_k}\mathbf{P} {}^{t_j}\mathbf{T} \|_F^2, \quad (8)$$

where $\|\cdot\|_F$ is the Frobenius norm.

Reprojection error. ${}^{c_i}\mathbf{A}_{p_k}^{t_j}$, estimated from image data, does not account for the problem's rigidity constraints. For the reprojection error cost function, we generate a HTM ${}^{c_i}\hat{\mathbf{A}}_{p_k}^{t_j}$ that represents the HTM relating pattern p_k to camera c_i at time t_j :

$${}^{c_i}\hat{\mathbf{A}}_{p_k}^{t_j} = {}^{c_i}\mathbf{C} {}^{t_j}\mathbf{T}^{-1} {}^{p_k}\mathbf{P}^{-1}, \quad (9)$$

similarly as [23], [24].

A three-dimensional point X on a calibration pattern and the corresponding two-dimensional point x in the image have the projective relationship in homogenous coordinates,

$$x = {}^{c_i}\hat{\mathbf{A}}_{p_k}^{t_j} X. \quad (10)$$

The detected image point that corresponds to X is $\tilde{x}$, and its reprojection error is $\|x - \tilde{x}\|^2$. We substitute and combine Eqs. 9 and 10 to represent total reprojection error over $\mathcal{CS}$:

$$re = \sum_{\mathcal{CS}} \sum_{(X, \tilde{x}) \in \mathbb{X}} \| {}^{c_i}\mathbf{C} ({}^{t_j}\mathbf{T})^{-1} ({}^{p_k}\mathbf{P})^{-1} X - \tilde{x} \|^2, \quad (11)$$

where $\mathbb{X}$ is the set of world-to-image calibration pattern point pairs (X, x) detected in one constraint in $\mathcal{CS}$.

As with the algebraic error, we represent the reprojection error as a minimization problem as well,

$$\operatorname{argmin}_{\mathbb{C}, \mathbb{P}, \mathbb{T}} \sum_{\mathcal{CS}} \sum_{(X, \tilde{x}) \in \mathbb{X}} \| {}^{c_i}\mathbf{C} ({}^{t_j}\mathbf{T})^{-1} ({}^{p_k}\mathbf{P})^{-1} X - \tilde{x} \|^2. \quad (12)$$

C. Gauge freedom of world coordinate frame w

The constraint set $\mathcal{CS}$ is indeterminate with respect to w . We fix the world coordinate system w to the coordinate system of a pattern $p^* \in \mathcal{P}$ at time label $t^* \in \mathcal{T}$, such that ${}^{p^*}\mathbf{P} = \mathbf{I}_4$ and ${}^{t^*}\mathbf{T} = \mathbf{I}_4$, where $\mathbf{I}_4$ is an identity matrix of size four. Our heuristic for choosing p^* and t^* is described in Section IV-C. This is an example of an *object-centered* gauge [25].

Example 2. Specifying w simplifies some of the constraints. Suppose $p^* = p_0$ and $t^* = t_1$. The example from above becomes

$${}^{c_0}\mathbf{C} = {}^{c_0}\mathbf{A}_{p_0}^{t_0} \mathbf{I}_4 {}^{t_0}\mathbf{T} \rightarrow {}^{c_0}\mathbf{C} = {}^{c_0}\mathbf{A}_{p_0}^{t_0} {}^{t_0}\mathbf{T} \quad (13)$$

$${}^{c_0}\mathbf{C} = {}^{c_0}\mathbf{A}_{p_0}^{t_1} \mathbf{I}_4 \mathbf{I}_4 \rightarrow {}^{c_0}\mathbf{C} = {}^{c_0}\mathbf{A}_{p_0}^{t_1} \quad (14)$$

$${}^{c_0}\mathbf{C} = {}^{c_0}\mathbf{A}_{p_1}^{t_1} {}^{p_1}\mathbf{P} \mathbf{I}_4 \rightarrow {}^{c_0}\mathbf{C} = {}^{c_0}\mathbf{A}_{p_1}^{t_1} {}^{p_1}\mathbf{P} \quad (15)$$

$${}^{c_1}\mathbf{C} = {}^{c_1}\mathbf{A}_{p_1}^{t_1} {}^{p_1}\mathbf{P} \mathbf{I}_4 \rightarrow {}^{c_1}\mathbf{C} = {}^{c_1}\mathbf{A}_{p_1}^{t_1} {}^{p_1}\mathbf{P}. \quad (16)$$

Since ${}^{c_0}\mathbf{A}_{p_0}^{t_1}$ is an observation and known, ${}^{c_0}\mathbf{C}$ can be initialized, and so on through the remaining constraints.

IV. METHOD DESCRIPTION

Algorithm 1 CALICO: Multi-camera calibration method.

Input: Set of constraints, $\mathcal{CS}$.

Output: Set of HTMs $\mathbb{C}, \mathbb{P}, \mathbb{T}$.

- 1: Determine intrinsic and extrinsic camera parameters with respect to visible patterns at all time labels, IV-A.
- 2: Test if the multi-camera system can be calibrated, IV-B.
- 3: Choose the gauge; find p^* and time t^* , IV-C.
- 4: **while** Any variable can be initialized **do**
- 5: Initialize individual, then variable pairs, IV-D.
- 6: Minimize algebraic error for all initialized variables,
- 7: Eq. 8, every r_{ae} iterations.
- 8: Refine HTMs $\mathbb{C}, \mathbb{P}, \mathbb{T}$ by minimizing reprojection error, Eq. 12; add batches of r_{rp} variables to the minimization problem.

We want to compute $\mathbb{C}$, with rigidity constraints represented by $\mathbb{P}, \mathbb{T}$ which are also unknown. We ultimately want to minimize reprojection error, Eq. 9. A brief sketch of the method is: fix the gauge to eliminate indeterminacy. Then, we initialize the values of unknown variables similarly to Example 2, using a local search; we find those constraints that have only one uninitialized variable remaining, and solve with a closed-form method. When no more constraints with only one variable initialized remain, we examine constraints with two uninitialized variables, and then solve for two variables with a closed-form method. We find approximate solutions to the non-linear least squares problems for algebraic error, Eq. 8 during the search-and-solve procedure. Finally, we minimize

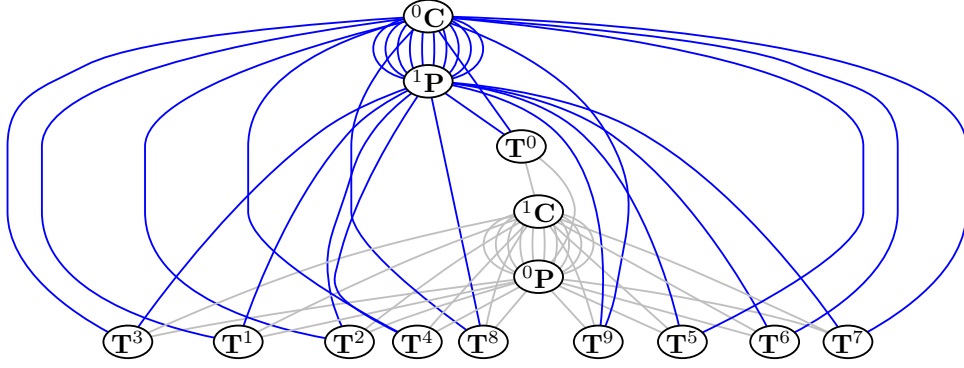


Fig. 4. The interaction graph for Dataset Mult-1. Edges connected to Camera ${}^0\mathbf{C}$ are blue, those connected to Camera ${}^1\mathbf{C}$ are light grey. $\mathbf{T}^0 = t^*$ and ${}^0\mathbf{P} = p^*$ in this dataset.

for reprojection error, Eq. 12, over the entire set of variables $\mathbb{CS}$.

Before these steps, though, we compute the internal and external camera calibration (${}^c_i \mathbf{A}_{p_k}^{t_j}$) for all cameras in the dataset. We test whether CALICO can calibrate the dataset. Alg. 1 describes the steps of the method.

A. Individual camera calibration

Individual cameras are calibrated for intrinsic and extrinsic camera calibration parameters. Each pattern detection triggers the generation of one constraint equation in $\mathbb{CS}$, Eq. 3. The extrinsic parameters – the HTMs from a pattern’s world coordinate system to the camera coordinate system – make up $\mathbb{A}$. Some images may contain pattern detections of more than one pattern at one time label; in these cases a constraint equation is generated and HTM ${}^c_i \mathbf{A}_{p_k}^{t_j}$ for each pattern detected. We use OpenCV’s implementation of Zhang’s calibration algorithm [2], [26] for this step.

B. Test ability to calibrate

We construct an *interaction graph* whose vertices are the variables in $\{\mathbb{C} \cup \mathbb{T} \cup \mathbb{P}\}$. For each constraint in $\mathbb{CS}$, edges are created between the variables in that constraint. We then compute the number of connected components in the graph. The entire camera rig can be calibrated with CALICO by a particular image dataset if there is exactly one connected component. An example from dataset Mult-1 is shown in Figure 4.

If the graph consists of multiple connected components, then the cameras corresponding to each component can be calibrated with respect to each other in the connected component but not with respect to cameras in different connected components.

C. Choose gauge

We choose the gauge and define the world coordinate system w by a pattern $p^* \in \mathcal{P}$ at time label $t^* \in \mathcal{T}$, such that the greatest number of variables can be initialized. The time

and pattern combination with the greatest frequency in $\mathbb{CS}$ is chosen as t^* and p^* , i.e.,

$$p^* = \underset{p_k \in \mathcal{P}}{\operatorname{argmax}} \sum \left| \left\{ {}^c_i \mathbf{C} = {}^c_i \mathbf{A}_{p_k}^{t_j} \mathbf{P}^{t_j} \mathbf{T} \mid {}^c_i \in \mathcal{C}, t_j \in \mathcal{T} \right\} \in \mathbb{CS} \right|, \quad (17)$$

or the pattern that has been observed the most times by any camera at any time label, and

$$t^* = \underset{t_j \in \mathcal{T}}{\operatorname{argmax}} \sum \left| \left\{ {}^c_i \mathbf{C} = {}^c_i \mathbf{A}_{p^*}^{t_j} \mathbf{P}^{t_j} \mathbf{T} \mid {}^c_i \in \mathcal{C} \right\} \in \mathbb{CS} \right|, \quad (18)$$

which is the time label corresponding to the largest number of observations of pattern p^* by any camera. The two HTMs corresponding to indices p^* and t^* are initialized to ${}^{p^*} \mathbf{P} = \mathbf{I}_4$ and ${}^{t^*} \mathbf{T} = \mathbf{I}_4$.

D. Initial solutions search and non-linear least squares minimization refinement

Initial solutions for $\mathbb{C}$, $\mathbb{P}$, and $\mathbb{T}$ are found through a local search-and-solve procedure (Sections IV-E and IV-F). We first solve constraints involving only one unknown variable and then two. Then, subsets of $\mathbb{CS}$ are used to create cost functions that approximately minimize non-linear least squares costs functions of algebraic and reprojection error, which further refines variables’ estimates.

First, we search through $\mathbb{CS}$ for constraints where $p_k = p^*$ and $t_j = t^*$ and initialize the camera HTMs for those constraints, as in Example 2. The number of initialized variables is at least three because of the way in which p^* and t^* were chosen; there is at least one constraint where $p_k = p^*$ and $t_j = t^*$ and consequently one camera variable that may be initialized.

Following the first search, we iterate through $\mathbb{CS}$ and handle constraints with only one uninitialized variable remaining. The variable’s HTM is initialized using closed-form methods discussed in Section IV-E. When there are no more constraints in $\mathbb{CS}$ having only one uninitialized variable, constraints having two uninitialized variables are solved using methods for robot-world, hand-eye calibration, discussed in Section IV-F. Then

we return to iterating through $\mathbb{CS}$ and handling constraints with only one uninitialized variable. We refer to the set of constraints where all three variables are initialized as $\mathbb{CS}'$.

Already-initialized variables' values are refined by minimization of algebraic error at several points during the search for uninitialized variables. Finally, we minimize reprojection error.

E. Initializing a single variable

Let $\mathbb{CS}^{(m)}$ be the set of elements of $\mathbb{CS}$ for which only one HTM variable $\mathbf{X}^{(m)}$ is unknown (at a given iteration, $\mathbf{X}^{(m)}$ could be either ${}^{c_i}\mathbf{C}$, ${}^{p_k}\mathbf{P}$, or ${}^{t_j}\mathbf{T}$). We solve Eq. 3 for $\mathbf{X}^{(m)}$ by rearranging the terms to the form

$$\mathbf{X}^{(m)}\mathcal{A} = \mathcal{B}, \quad (19)$$

where $\mathcal{A}$ and $\mathcal{B}$ are HTMs that are data (from $\mathbb{A}$), initialized variables, or multiplied combinations of HTMs from $\mathbb{A}$ and initialized variables within a constraint. $\mathbf{X}^{(m)}$ is the unknown transformation. If $|\mathbb{CS}^{(m)}| = 1$, we simply compute $\mathbf{X}^{(m)} = (\mathcal{A})^{-1}\mathcal{B}$. Otherwise, we solve for $\mathbf{X}^{(m)}$ with $\mathbb{CS}^{(m)}$ using Shah's closed-form method for registering two sets of six degrees-of-freedom data [27].

$\mathbf{X}^{(m)}$'s initial solution is refined by finding an approximate solution to

$$\operatorname{argmin}_{\mathbf{X}^{(m)}} \sum_{\mathbb{CS}^{(m)}} \left\| \mathbf{X}^{(m)}\mathcal{A} - \mathcal{B} \right\|_F^2 \quad (20)$$

using the Levenberg-Marquardt algorithm.

F. Initializing variable pairs

Similarly, let $\mathbb{CS}^{(m)}$ be the set of elements of $\mathbb{CS}$ HTMs for which both $\mathbf{X}^{(m)}$ and $\mathbf{Z}^{(m)}$ are unknown. In this case, a constraint with two unknown variables can be rearranged into the form of the robot-world, hand-eye calibration problem

$$\mathcal{A}\mathbf{X}^{(m)} = \mathcal{B}\mathbf{Z}^{(m)}, \quad (21)$$

where $\mathcal{A}$ and $\mathcal{B}$ are known HTMs. We solve for the unknown variables using Shah's closed-form method for the robot-world, hand-eye problem [28]. In some cases, $\mathcal{A}$ or $\mathcal{B}$ may be identity matrices. As with single variables, we refine this initial estimate by finding an approximate solution to

$$\operatorname{argmin}_{\mathbf{X}^{(m)}, \mathbf{Z}^{(m)}} \sum_{\mathbb{CS}^{(m)}} \left\| \mathcal{A}\mathbf{X}^{(m)} - \mathcal{B}\mathbf{Z}^{(m)} \right\|_F^2 \quad (22)$$

using the Levenberg-Marquardt algorithm.

G. Variable initialization order

At each iteration, there may be many choices of single variables or pairs of variables to solve for via the methods of sections IV-E and IV-F. Solving for single variables is always selected over solving for variable pairs. Beyond this choice, the variable initialization order is determined by a heuristic that prioritizes variables contained in the largest number of remaining constraints. That is, we select the HTM $\mathbf{X}^{(m)}$ that maximizes $|\mathbb{CS}^{(m)}|$. Ties are broken by choosing $\mathbb{C}$ variables first, then $\mathbb{P}$, and finally $\mathbb{T}$, and breaking further ties by choosing the variable with the smallest index first.

H. Minimization of algebraic error over $\mathbb{CS}'$

The variables' estimates are refined by minimizing for algebraic error, Eq. 8, over the set $\mathbb{CS}'$, using the Levenberg-Marquardt algorithm. This minimization is done after a user-specified proportion of iterations, r_{ae} , through the **while** loop at line 5 of Algorithm 1. For example, if r_{ae} is 0.20 and there are 100 constraints in $\mathbb{CS}$, minimization of Eq. 3 over $\mathbb{CS}'$ occurs when the number of passes through initializing one variable (IV-E) or variable pairs (IV-F) is 20, 40, 60, and so on.

I. Final refinement: minimization of reprojection error.

Finally, the variables' estimates are refined by minimizing for reprojection error, Eq. 12 using the Levenberg-Marquardt algorithm.. Similarly to the minimization of algebraic error, the reprojection error is refined over subsets of $\mathbb{CS}'$. We use a user-specified proportion of constraints, r_{rp} , to control when and how the variables are minimized. If $r_{rp} = 0.5$, then reprojection error is minimized for all variables in half of $\mathbb{CS}$; then the constraints in the second half of $\mathbb{CS}$ are added to the cost function and the minimization performed.

V. EXPERIMENTS

CALICO was evaluated in simulated and real-world experiments. First, we describe experimental details in Section V-D, introduce three evaluation metrics in Section V-A, and then describe datasets and results in Sections V-C and V-F, respectively. We use the implementation of the Levenberg-Marquardt method found in the software package Ceres [29]. The results shown in this paper were generated on a workstation with one 12-core 2.70GHz processor. OpenCV [26] was used to compute the camera calibration parameters.

The parameter, r_{ae} , indicating the proportion of variables included in subsequent runs of algebraic error minimization (Eq. 8) was set to 0.2 for all datasets. The parameter indicating the proportion of variables included in subsequent runs of the reprojection error minimization, r_{rp} was set to 0.5 for all datasets (Eq. 11).

A. Evaluation

We used three metrics for evaluation: algebraic error, reprojection root mean squared error, and reconstruction accuracy error.

1) *Algebraic error*: The algebraic error represents the fit of the estimated HTMs to $\mathbb{CS}$. It is given by

$$ae = \frac{1}{|\mathbb{CS}|} \sum_{\mathbb{CS}} \left\| {}^{c_i}\mathbf{C} - {}^{c_i}\mathbf{A}_{p_k}^{t_j} {}^{p_k}\mathbf{P} {}^{t_j}\mathbf{T} \right\|_F^2. \quad (23)$$

2) *The Reprojection Root Mean Squared Error (rrmse)*: The reprojection root mean squared error is

$$rrmse = \sqrt{\frac{1}{N} re}, \quad (24)$$

where $N = \sum_{\mathbb{CS}} |\mathbb{X}|$ is the total number of calibration pattern points observed and re is reprojection error from Equation 11.

3) *Reconstruction Accuracy Error*: Reconstruction accuracy error, *rae*, is used here in a similar way as in [24], to assess the method’s ability to reconstruct the three-dimensional location of calibration pattern points. Given detections of a pattern point in images over multiple cameras and times, the three-dimensional point that generated those pattern points is estimated by N-view triangulation. Reconstruction accuracy error (*rae*) is the Euclidean difference between estimated and ground truth world points. The ground truth world points consist of the coordinate system defined by the calibration pattern, so these point are known even in real settings.

The three-dimensional point X that generated the corresponding image points x can be found by solving the following minimization problem

$$\hat{Y} = \underset{X}{\operatorname{argmin}} \sum_{\text{CS}} \left\| {}^c\mathbf{C} (\mathbf{T}^t)^{-1} ({}^p\mathbf{P})^{-1} X - x \right\|^2. \quad (25)$$

Since Eq. 25 is a non-linear least squares problem, we require an initial solution for X . We use the direct linear transform (DLT) algorithm described in [30] as the ‘Linear-LS Method’ and the normalization method presented in [31]. This approach to triangulation is also described in [32].

Using an initial solution computed using the DLT method, we find an approximate solution for $\hat{Y}$ in Eq. 25 using the Levenberg-Marquardt algorithm. $\hat{Y}$ is found for all calibration pattern points detected in two or more constraints, generating the set $\mathbb{Y}$. The reconstruction accuracy error (*rae*) is the Euclidean distance between the estimated $\hat{Y}_q$ points and corresponding calibration object points X_q . We report the mean *rae* value over the set $\mathbb{Y}$:

$$rae = \frac{1}{|\mathbb{Y}|} \sum_{Y_q \in \mathbb{Y}} \left\| \hat{Y}_q - X_q \right\|. \quad (26)$$

B. Comparison with Kalibr

Part of our evaluation of CALICO includes a comparison with multi-sensor calibration method Kalibr. We use aruco markers in our laboratory, but Kalibr uses April tags [12], so we created simulated datasets with April tags for comparing Kalibr with CALICO. Another difference is that Kalibr requires one and only one calibration pattern, which sometimes imposes constraints on camera placement.

Our philosophy on comparison with Kalibr is that we are not interested in very small differences in accuracy between our method and Kalibr. There are some situations where Kalibr is not able to calibrate a dataset, and CALICO is. Since we are using others’ implementation, it is difficult to determine which characteristics of Kalibr calibrations are features of Kalibr as a theoretical method, and Kalibr’s implementation, and extensively engineering another’s implementation is beyond the scope of this paper.

Consequently, our comparison of CALICO with Kalibr is focussed on whether CALICO’s calibrations are equivalent to the comparable state-of-the-art method, Kalibr, in the situations where Kalibr is able to calibrate, and not on whether CALICO ‘beats’ Kalibr, and vice versa.

C. Datasets

Our experiments include simulated and real datasets. They can be divided into a four types: *box*, *robot*, *stereo*, and *wide baseline stereo*. The *box* type has cameras arranged on the perimeter of the box and pointing towards the interior of the box. The *robot* type consists of cameras lying on a cylinder, circle, or line and pointing outward. Stereo and wide-baseline stereo datasets have a pair of cameras; in wide-baseline stereo, individual cameras may view exclusive sets of individual aruco or April Tag pattern indices. Examples of these different dataset types are in Figure 2.

The datasets use either aruco markers [4] or April tags [12]. The real and simulated datasets are described in Tables II and III, respectively.

D. Pattern and camera configurations, data acquisition, and implementation details

We used a set of planar calibration targets created with chessboard-type aruco tags [4] generated using OpenCV [26], referred to as charuco patterns. The particular arrangement, and orientation, of the patterns is computed by CALICO. If a particular calibration pattern’s orientation can be determined and its pattern index detected, there is no restriction on the type of pattern used if the connections between individual calibration patterns is rigid. For example, one could mount patterns on laboratory walls to satisfy rigidity requirements. One could also use a single pattern as is typical for camera calibration if the camera’s field of view is shared.

Our real experiments’ data acquisition process is as follows. First, multiple images are acquired per camera to allow for internal parameter calibration. Then, the user places the calibration rig in view of at least one camera. The user indicates that the current acquisition is time label t_0 and acquires an image from all cameras. Then, the calibration rig is moved such that at least one or more cameras view a pattern, not necessarily the same pattern, the user indicates that the time label and images are written from all of the cameras. This process is continued until n_t .

The use of ‘time label t_0 ’ does not imply that the cameras are synchronized, but instead that the images are captured and labeled with the same time tag for the same position of the camera rig and pattern rigs. A mechanism for doing so may be implemented through a user interface that allows the user to indicate that all cameras should acquire images, assign them a common time tag, and report to the user when the capture is concluded.

Camera calibration for internal parameters is performed for each of the cameras independently. Individual patterns are uniquely identified through aruco or charuco tags [4]; cameras’ extrinsic parameters (rotation and translation) are computed with respect to the coordinate systems defined by the patterns recognized in the images. If two (or more) partial patterns are detected in the same image, that camera will have two (or more) extrinsic transformations defined for that time label, one for each of the patterns detected.

The minimal requirement on the number and type of images to collect is that the interaction graph consist of a

TABLE II
REAL DATASET DESCRIPTIONS. THESE DATASETS USE CHARUCO PATTERNS.

Dataset	type	CS	cameras	patterns	times (n_t)
Net-1	<i>box</i>	107	12	2	10
Net-2	<i>box</i>	211	12	2	20
Mult-1	<i>robot</i>	20	2	2	10
Mult-2	<i>robot</i>	20	2	2	10
Mult-3	<i>robot</i>	72	4	3	24

TABLE III
SIMULATED DATASET DESCRIPTIONS. DATASETS SIM-1 AND SIM-2 USE CHARUCO PATTERNS, WHILE THE REMAINDER USE APRIL TAGS.

Dataset	type	max. distance between cameras	CS	cameras	patterns	times (n_t)
Sim-1	<i>box</i>	3162.3	87	8	4	43
Sim-2	<i>box</i>	1677.1	472	16	4	37
Sim-3-april	<i>stereo</i>	103.08	81	2	1	41
Sim-4-april	<i>robot</i> ^a	210.81	665	12	1	100
Sim-5-april	<i>robot</i> ^a	210.81	659	12	1	100
Sim-6-april	<i>stereo</i>	375	109	2	1	55
Sim-7-april	<i>robot</i> ^b	750	45	6	1	20
Sim-8-april	wide baseline stereo ^c	1120.8	23	2	1	12
Sim-9-april	wide baseline stereo ^c	1120.8	23	2	1	12

^a The camera poses are the same for these two datasets; difference is in the pattern poses.

^b The six cameras lie on a line, with two cameras pointing towards $-z$, and the remaining four cameras pointing at ± 45 and ± 90 degrees from $-z$.

^c The camera poses are the same for these two datasets; in Sim-8-April, there is a partition in between the two cameras such that each camera views half of the pattern's April Tags.

single connected component (Section IV-B). In real calibration scenarios, we found that larger values of n_t produce better quality initial solutions than smaller values of n_t , as the quality of the initial solution (Section IV-D) benefits from additional constraints between variables. The solution quality produced with CALICO can be assessed via the error metrics we describe in the next section; if the error is too high, one can acquire more images, add them to the dataset, and recalibrate.

1) *Simulated experiments*: Simulated experiments were generated to allow assessment of CALICO to a ground truth, as well as to allow comparison to Kalibr. The simulated experiments include *box*, *robot*, *stereo*, and *wide baseline stereo* camera arrangements and are described in Table III. We used OpenGL to generate datasets.

Datasets Sim-1 and Sim-2 use charuco patterns are of the *box* type, representing arrangements similar to those used in motion-capture experiments, where cameras are mounted on the walls around a room. These datasets differ in camera density, number of cameras, distance from the pattern, and the resulting number of constraints.

The Kalibr multi-camera calibration method requires one and only one calibration pattern and its implementation provided by the authors require the use of April tags. Datasets Sim-3-april through Sim-9-april use April tags and use only one pattern. Sim-3-april is a two-camera simulated dataset; the cameras are roughly side-by-side, so it is of the *stereo* type. In Sim-4-april, there are twelve cameras in three rows of four

each, arranged around one quarter of a cylinder and pointing away from the cylinder's center axis. The calibration pattern is moved at 100 time steps in front of the cameras. Sim-5-april has the same arrangement of cameras as Sim-4-april, but random perturbations have been added to the pattern poses.

Sim-7-april is a *robot* type dataset, with camera arranged on a line. A criticism of CALICO has been that camera could be calibrated in some experiments by using a large enough pattern and Kalibr. Sim-8-april and Sim-9-april were generated to explore this criticism. Sim-8-april is a *wide baseline stereo* type dataset, with a very large pattern such that most images contain a portion of the pattern. Sim-9-april is the same dataset as Sim-8-april, except that there is a barrier such that one camera does not contain the set of April tags that are in the other cameras' images, and vice versa. While we wish to address the prior criticism of CALICO and explore Kalibr's performance in these situations, we reiterate our comparison philosophy from V-B that it may be possible to alter Kalibr's implementation such that Kalibr can calibrate these situations, but that kind of work is beyond the scope of this paper.

2) *Stationary multi-camera system*: A multi-camera system was constructed using low cost webcams, and arranged on two sides of a metal rectangular prism. The pattern rig is constructed of two charuco patterns rigidly attached to each other. To create datasets Net-1 and Net-2, the pattern rig was moved within the workspace. The stationary multi-camera system creates *box* datasets.

3) *Mobile multi-camera system*: Three multi-camera datasets were created from four cameras facing away from each other. In the first two datasets in this group, Mult-1 and Mult-2, only two cameras facing back-to-back were used for data acquisition. In the Mult-1 dataset, each camera in the camera rig acquired images from only one pattern. In Mult-2, acquisition was as for Mult-1, except that the camera rig was rotated 180 degrees such that each camera views the two patterns. Mult-3 uses all four cameras, and uses three calibration patterns. The mobile multi-camera system creates *robot* type datasets.

E. Ground truth evaluation and comparison experiment

For the nine Sim-# datasets, ground truth camera poses are available. We assess the performance of CALICO as well as Kalibr when relevant with respect to the ground truth by comparing the camera poses. First, the camera poses of both the ground truth and the results of a method are mapped to a common coordinate system, where ${}^{c_0}\mathbf{C}' = \mathbf{I}_4$. In other words, the first camera's HTM for the comparison ${}^{c_0}\mathbf{C}'$ is the identity matrix, and all the other camera HTMs are transformed to this coordinate system by post-multiplying by the inverse of the first camera's original HTM, ${}^{c_i}\mathbf{C}' = ({}^{c_i}\mathbf{C}) ({}^{c_0}\mathbf{C}^{-1})$.

The difference in rotation angle and translation distance are then compared between the transformed camera poses of the ground truth and results from CALICO or Kalibr. The denominator for averaging angle and distance is $|\mathbf{C}| - 1$ since there is no difference between the pose of the c_0^{th} cameras between CALICO or Kalibr and the ground truth.

We compared CALICO to the Kalibr multi-camera calibration toolbox [9], using Kalibr’s implementation on Github,⁴ and accessed this implementation through a Docker image from Stereo Labs.⁵

CALICO compared well to the ground truth, values in Table IV, with a maximum rotational error of 0.23° and translational error of 12.28 mm. In the case of 12.28 mm translational error for the Sim-3-april dataset, the internal parameters for one of the cameras were estimated differently than the ground truth, leading to a different resulting camera pose.

CALICO also compared well with Kalibr. For the Sim-3-april dataset, as mentioned in the previous paragraph, the calibration of internal and external parameters was propagated to the constraint set $\mathbb{CS}$. Since for this work we chose the first camera to align between the datasets so they could be compared, and this internal parameter difference happened on the second (of two) cameras, different choices of anchor camera would likely influence results. On the Sim-4-april dataset, Kalibr’s output converged but the solution was not usable; most of the twelve cameras are clustered around two cameras. For the Sim-5-april dataset, Kalibr was not able to converge. The camera rig in Sim-4-april and Sim-5-april have purely rotational motion about an axis, and from [11], [33] learned that this case is a degenerate motion for Kalibr and explain its behavior for those datasets. For the Sim-7-april and Sim-8-april datasets, Kalibr failed with the message that there was not enough mutual information. It seems that the Kalibr implementation we used requires cameras to not only image the same pattern, but also requires that cameras image a set of the same April tags.

F. Results and discussion

„

CALICO was used to calibrate the nine simulated, two stationary multi-camera, and three mobile multi-camera datasets. Quantitative results are shown in Table V. Results are in terms of the three metrics: algebraic error (ae), reprojection root mean squared error ($rrmse$), reconstruction accuracy error (rae). Run time is shown as well.

Qualitatively, we found that camera poses were similar to those expected from the configuration of the multi-camera datasets. For most of the datasets, the reconstructed calibration patterns are very close to the source patterns.

Quantitatively, the camera poses found with CALICO had subpixel $rrmse$ values, except for dataset Net-1. The higher $rrmse$ value of that experiment versus the others is perhaps explained by that experiment’s small number of time instants (10) versus a comparably large number of cameras (12), Table II. For all of the datasets, CALICO produced a less than 1.11 mm mean reconstruction accuracy error rae .

From Table V, algebraic error seems not well related to the quality of results that are important to reconstruction tasks like rae and $rrmse$. While algebraic error is used to generate initial solutions, reconstruction error may be high for views of the calibration patterns where the estimation of the pattern

TABLE IV
COMPARISON OF CALICO TO GROUND TRUTH AND KALIBR.

Dataset	CALICO		Kalibr	
	Rotation error	Translation error	Rotation error	Translation error
Sim-1	0.234°	8.565	did not compare	
Sim-2	0.062°	1.368	did not compare	
Sim-3-april	0.203°	12.282	0.167°	3.657
Sim-4-april	0.093°	6.280	30.714°	107.756
Sim-5-april	0.018°	0.759	no convergence	
Sim-6-april	0.037°	0.692	0.0257°	0.357
Sim-7-april	0.146°	3.605	not enough mutual info	
Sim-8-april	0.029°	2.73	not enough mutual info	
Sim-9-april	0.022°	0.563	0.006°	0.322

Mean rotation error (degrees) and average translation error (mm). Details in V-E.

Datasets Sim-1 and Sim-2 use charuco patterns and Kalibr reads April Tags, so comparison with Kalibr is not possible with these datasets.

TABLE V
RESULTS OF USING CALICO TO COMPUTE CAMERA POSE FOR NINE DATASETS.

Dataset	ae	$rrmse$ (pixels)	rae (mm)	run time (s)	
				load, internal calibration	multi- camera calibration
Net-1	3.818	1.107	0.382	80.310	0.219
Net-2	5.598	0.941	0.323	25.0	0.618
Mult-1	0.960	0.401	0.607	2.6	0.076
Mult-2	0.862	0.342	0.34	3.43	0.085
Mult-3	2.013	0.464	0.701	11.46	0.282
Sim-1	48.938	0.450	0.235	7.1	0.403
Sim-2	1.725	0.370	0.078	22.65	1.39
Sim-3-april	11.652	0.430	1.054	18.5	0.242
Sim-4-april	56.592	0.438	1.102	202.0	1.732
Sim-5-april	10.083	0.215	0.06	194.17	1.572
Sim-6-april	0.391	0.386	0.084	23.3	0.313
Sim-7-april	18.62	0.438	0.161	17.0	0.137
Sim-8-april	2.40	0.41	4.63	11.8	0.096
Sim-9-april	1.80	0.41	0.813	11.6	0.109

Algebraic error, ae (Eq. 23) has no units.

to camera transformation ${}^c\mathbf{A}_p^t$ is not reliable. The stationary multi-camera datasets have a high proportion of images in this category, so we hypothesize that this is why the algebraic error is higher for those datasets.

We also show run time in Table V divided into two categories: (a) tasks preliminary to multi-camera calibration, such as image loading and individual camera calibration, and (b) multi-camera calibration. We parallelize all portions of the process. The run time is dominated by the image loading and individual camera calibration tasks. The largest datasets – twelve cameras and 100 images each – had low run times of 3:24 minutes for Sim-4-april and 3:15 minutes for Sim-5-april.

VI. CONCLUSIONS

We presented CALICO, a method for multi-camera calibration suitable for stationary multi-camera systems, mobile multi-camera systems, multi-camera systems with non-overlapping fields of view, and/or non-synchronized systems. The method’s performance was demonstrated on fourteen

⁴<https://github.com/ethz-asl/kalibr>

⁵<https://hub.docker.com/r/stereolabs/kalibr>

datasets and compared to another multi-camera calibration method, Kalibr. CALICO uses one or more calibration patterns, and so does not depend on being able to move camera rigs to a richly textured scene or and does not require synchronization. We formulated the multi-camera calibration problem within this context as a set of rigidity constraints between cameras, pattern transformations, and time labels and minimize for reprojection error. We demonstrated that the multi-camera calibration part, which excludes data loading and individual camera calibration, was computed in less than 2 seconds for all datasets.

ACKNOWLEDGMENTS

A. Tabb was supported by USDA-ARS Project 8080-21000-032-00-D. T.T. Santos' work is supported by the São Paulo Research Foundation (FAPESP, grant 2017/19282-7).

REFERENCES

- [1] A. Tabb and M. J. Feldmann, "Data and code from: Multi-camera calibration with pattern rigs, including for non-overlapping cameras: Calico," Dec. 2023. [Online]. Available: <https://zenodo.org/doi/10.5281/zenodo.3520865>
- [2] Z. Zhang, "A flexible new technique for camera calibration," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 22, no. 11, pp. 1330–1334, Nov. 2000. [Online]. Available: <https://doi.org/10.1109/34.888718>
- [3] B. A. Cook, "ACRONYM: Acronym CReatiON for You and Me," *arXiv:1903.12180 [astro-ph]*, Mar. 2019, arXiv: 1903.12180. [Online]. Available: <http://arxiv.org/abs/1903.12180>
- [4] S. Garrido-Jurado, R. Muñoz-Salinas, F. Madrid-Cuevas, and M. Marín-Jiménez, "Automatic generation and detection of highly reliable fiducial markers under occlusion," *Pattern Recognition*, vol. 47, no. 6, pp. 2280–2292, Jun. 2014. [Online]. Available: <https://doi.org/10.1016/j.patcog.2014.01.005>
- [5] A. Tabb, H. Medeiros, M. J. Feldmann, and T. T. Santos, "Calibration of asynchronous camera networks: Calico," 2019.
- [6] M. Shah, R. D. Eastman, and T. Hong, "An overview of robot-sensor calibration methods for evaluation of perception systems," in *Proceedings of the Workshop on Performance Metrics for Intelligent Systems - PerMIS '12*. College Park, Maryland: ACM Press, 2012, p. 15. [Online]. Available: <http://dl.acm.org/citation.cfm?doid=2393091.2393095>
- [7] Li, Bo, L. Heng, K. Koser, and M. Pollefeys, "A multiple-camera system calibration toolbox using a feature descriptor-based calibration pattern," *IEEE*, Nov. 2013, pp. 1301–1307. [Online]. Available: <http://ieeexplore.ieee.org/document/6696517/>
- [8] A. Liu, S. Marschner, and N. Snavely, "Caliber: Camera Localization and Calibration Using Rigidity Constraints," *International Journal of Computer Vision*, vol. 118, no. 1, pp. 1–21, May 2016. [Online]. Available: <https://doi.org/10.1007/s11263-015-0866-1>
- [9] P. Furgale, J. Rehder, and R. Siegwart, "Unified temporal and spatial calibration for multi-sensor systems," in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2013, pp. 1280–1286.
- [10] J. Maye, P. Furgale, and R. Siegwart, "Self-supervised calibration for robotic systems," in *2013 IEEE Intelligent Vehicles Symposium (IV)*, Jun. 2013, pp. 473–480, iSSN: 1931-0587.
- [11] J. Maye, "Online Self-Calibration for Robotic Systems," Doctoral Thesis, ETH Zurich, 2014, accepted: 2017-06-13T23:36:32Z. [Online]. Available: <https://www.research-collection.ethz.ch/handle/20.500.11850/154673>
- [12] E. Olson, "AprilTag: A robust and flexible visual fiducial system," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, May 2011, pp. 3400–3407.
- [13] J. L. Schönberger and J.-M. Frahm, "Structure-from-Motion Revisited," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, NV, USA: IEEE, Jun. 2016, pp. 4104–4113. [Online]. Available: <https://doi.org/10.1109/CVPR.2016.445>
- [14] S. Esquivel, F. Woelk, and R. Koch, "Calibration of a Multi-camera Rig from Non-overlapping Views," in *Pattern Recognition*, F. A. Hamprecht, C. Schnörr, and B. Jähne, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, vol. 4713, pp. 82–91, series Title: Lecture Notes in Computer Science. [Online]. Available: http://link.springer.com/10.1007/978-3-540-74936-3_9
- [15] P. Lébraly, E. Royer, O. Ait-Aider, and M. Dhome, "Calibration of Non-Overlapping Cameras—Application to Vision-Based Robotics," in *Proceedings of the British Machine Vision Conference 2010*. Aberystwyth: British Machine Vision Association, 2010, pp. 10.1–10.12. [Online]. Available: <http://www.bmva.org/bmvc/2010/conference/paper10/index.html>
- [16] G. Carrera, A. Angeli, and A. J. Davison, "SLAM-based automatic extrinsic calibration of a multi-camera rig," in *2011 IEEE International Conference on Robotics and Automation*. Shanghai, China: IEEE, May 2011, pp. 2652–2659. [Online]. Available: <http://ieeexplore.ieee.org/document/5980294/>
- [17] Y. Lin, V. Larsson, M. Geppert, Z. Kukulova, M. Pollefeys, and T. Sattler, "Infrastructure-Based Multi-camera Calibration Using Radial Projections," in *Computer Vision – ECCV 2020*, ser. Lecture Notes in Computer Science, A. Vedaldi, H. Bischof, T. Brox, and J.-M. Frahm, Eds. Cham: Springer International Publishing, 2020, pp. 327–344.
- [18] S. Thirthala and M. Pollefeys, "Radial Multi-focal Tensors," *International Journal of Computer Vision*, vol. 96, no. 2, pp. 195–211, Jan. 2012. [Online]. Available: <https://doi.org/10.1007/s11263-011-0463-x>
- [19] S. Mishra, S. Nagesh, S. Mangani, G. Mills, P. Chakravarty, and G. Pandey, "Look Both Ways: Bidirectional Visual Sensing for Automatic Multi-Camera Registration," Aug. 2022, arXiv:2208.07362 [cs]. [Online]. Available: <http://arxiv.org/abs/2208.07362>
- [20] H. Chen and S. Schwertfeger, "Heterogeneous Multi-sensor Calibration based on Graph Optimization," in *2019 IEEE International Conference on Real-time Computing and Robotics (RCAR)*, Aug. 2019, pp. 158–163.
- [21] T. Pollok and E. Monari, "A visual SLAM-based approach for calibration of distributed camera networks," in *2016 13th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, Aug. 2016, pp. 429–437.
- [22] A. Robinson, M. Persson, and M. Felsberg, "Robust Accurate Extrinsic Calibration of Static Non-overlapping Cameras," in *Computer Analysis of Images and Patterns*, ser. Lecture Notes in Computer Science, M. Felsberg, A. Heyden, and N. Krüger, Eds. Cham: Springer International Publishing, 2017, pp. 342–353.
- [23] A. Tabb and K. M. Ahmad Yousef, "Parameterizations for reducing camera reprojection error for robot-world hand-eye calibration," in *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Sep. 2015, pp. 3030–3037. [Online]. Available: <http://doi.org/10.1109/IROS.2015.7353795>
- [24] A. Tabb and K. M. Ahmad Yousef, "Solving the robot-world hand-eye(s) calibration problem with iterative methods," *Machine Vision and Applications*, vol. 28, no. 5, pp. 569–590, Aug. 2017. [Online]. Available: <https://doi.org/10.1007/s00138-017-0841-7>
- [25] B. Triggs, P. F. McLauchlan, R. I. Hartley, and A. W. Fitzgibbon, "Bundle Adjustment — A Modern Synthesis," in *Vision Algorithms: Theory and Practice*, G. Goos, J. Hartmanis, J. van Leeuwen, B. Triggs, A. Zisserman, and R. Szeliski, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2000, vol. 1883, pp. 298–372, series Title: Lecture Notes in Computer Science. [Online]. Available: http://link.springer.com/10.1007/3-540-44480-7_21
- [26] G. Bradski, "The OpenCV Library," *Dr. Dobbs' Journal of Software Tools*, 2000.
- [27] M. Shah, "Comparing two sets of corresponding six degree of freedom data," *Computer Vision and Image Understanding*, vol. 115, no. 10, pp. 1355–1362, Oct. 2011. [Online]. Available: <https://doi.org/10.1016/j.cviu.2011.05.007>
- [28] —, "Solving the Robot-World/Hand-Eye Calibration Problem Using the Kronecker Product," *Journal of Mechanisms and Robotics*, vol. 5, no. 3, p. 031007, Jun. 2013. [Online]. Available: <https://doi.org/10.1115/1.4024473>
- [29] S. Agarwal, K. Mierle, and Others, "Ceres solver," <http://ceres-solver.org>.
- [30] R. I. Hartley and P. Sturm, "Triangulation," *Computer Vision and Image Understanding*, vol. 68, no. 2, pp. 146–157, Nov. 1997. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1077314297905476>
- [31] R. Hartley, "In defense of the eight-point algorithm," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 6, pp. 580–593, Jun. 1997. [Online]. Available: <http://ieeexplore.ieee.org/document/601246/>

- [32] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge University Press, ISBN: 0521540518, 2004.
- [33] J. Lee, D. Hanley, and T. Bretl, "Extrinsic calibration of multiple inertial sensors from arbitrary trajectories," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 2055–2062, 2022.