“©2020 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.”

Observability driven Multi-modal Line-scan Camera Calibration

Jasprabhjit Mehami†, Teresa Vidal-Calleja, and Alen Alempijevic

***Abstract*— Multi-modal sensors such as hyperspectral line-scan and frame cameras can be incorporated into a single camera system, enabling individual sensor limitations to be compensated. Calibration of such systems is crucial to ensure data from one modality can be related to the other. The best known approach is to capture multiple measurements of a known planar pattern, which are then used to optimize calibration parameters through non-linear least squares. The confidence in the optimized parameters is dependent on the measurements, which are contaminated by noise due to sensor hardware. Understanding how this noise transfers through the calibration is essential, especially when dealing with line-scan cameras that rely on measurements to extract feature points. This paper adopts a maximum likelihood estimation method for propagating measurement noise through the calibration, such that the optimized parameters are associated with an estimate of uncertainty. The uncertainty enables development of an active calibration algorithm, which uses observability to selectively choose images that improve parameter estimation. The algorithm is tested in both simulation and hardware, then compared to a naive approach that uses all images to calibrate. The simulation results for the algorithm show a drop of 26.4% in the total normalized error and 46.8% in the covariance trace. Results from the hardware experiments also show a decrease in the covariance trace, demonstrating the importance of selecting good measurements for parameter estimation.**

I. INTRODUCTION

Multi-modal vision sensing systems have become increasingly common in robotics for applications in dynamic or featureless environments. Combining sensing modalities in a complementary manner overcomes the limitations of individual sensors when working in difficult scenarios with poor lighting, occlusions and texture-less surfaces. The ubiquitous frame camera, which provides dense 2D spatial information, is often used for multi-frame operations such as visual odometry [1] or simultaneous localization and mapping (SLAM) [2], but is limited to the visible spectrum. A hyperspectral camera can see beyond the visible spectrum. Therefore, a frame and hyperspectral camera system can be set up to expand the useful features that can be captured from an environment. Before using such a camera system, camera calibration is a necessary process to ensure consistency between images and relate feature points from one camera to another. The calibration is performed by optimizing parameters through measurements of known feature points. These measurements exhibit uncertainty due to inherent random noises from the camera hardware. It is important to quantify this uncertainty, as this will determine how reliable the optimized parameters are.

J. Mehami, T. Vidal-Calleja, and A. Alempijevic are with the Centre for Autonomous Systems, Faculty of Engineering and IT, University of Technology Sydney, Ultimo, NSW 2007, Australia
† Jasprabhjit.Mehami@student.uts.edu.au

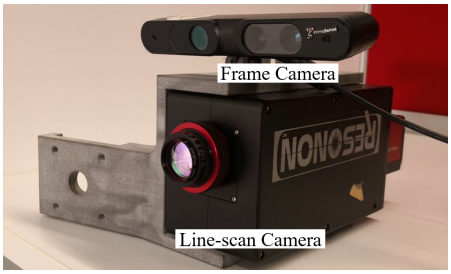


Fig. 1: Rigidly mounted camera system comprised of a Resonon Pika NIR-320 hyperspectral line-scan camera and a Primesense Carmine 1.09 frame camera.

Hyperspectral cameras typically capture data in a line-scan manner with a single spatial dimension. This mode of operation is preferred as the cameras can work in areas where high resolution and frame-rate is required [3][4][5]. The single spatial dimension makes calibration more difficult as the feature points seen by the camera are not easily recognizable. This makes it difficult to use multiple measurements of feature points across many images, which would ensure well optimized parameters. To overcome this issue, special patterns have been designed to work with line-scan cameras, where feature points can be extracted through the knowledge of the pattern's dimensions and the line-scan images [4][5][6]. These generally calibrated a line-scan camera using a single image and relied heavily on the accuracy of multiple separate planes.

A planar pattern, captured in multiple different poses, could be used for calibration if the pose can be estimated by external means such as assuming constant motion of either the pattern or camera [7], or using an inertial measurement unit (IMU) [8]. An intrinsically calibrated frame camera was also shown to be able to estimate the pose of a planar pattern, and then perform calibration with the line-scan camera as a single camera system [3][9][10]. In all these works, uncertainty analysis of the line-scan calibration had not been considered, but has been well established for frame camera calibration [11][2][12][13]. In such cases, measurement noise was propagated through each step of the calibration and the optimized parameters were returned with an associated error.

This paper describes a probabilistic calibration of a line-scan and frame camera system that employs an active calibration algorithm, which uses a metric to selectively choose new measurements that will add value to the optimized parameters. Uncertainty estimation plays a vital role as the measurements are not only used to perform the overall

calibration, but before this, they aid in estimating feature points that are seen by the line-scan camera from the planar pattern. These estimated feature points will now have an associated uncertainty. This uncertainty needs to be incorporated into the optimized parameters.

The singular spatial dimension of the line-scan camera also adds difficulty when capturing images of the pattern in different poses. Not all measurements will improve the optimized parameters. Some measurements may have insufficient parallax, due to poses being similar, which can degrade the optimization. A metric of observability, from control theory, is used to filter such measurements by incorporating the estimated uncertainty of the optimized parameters.

Calibration experiments using the algorithm are conducted in simulation and hardware. The experiments compare the algorithm to a naive approach that calibrates using all images. The optimized parameters found through the active approach exhibit lower errors and uncertainty. Therefore, calibrations which use measurements that show significant variation in pose, lead to well-optimized parameters.

II. Camera Models

A. Frame Camera

The pinhole camera model (PCM) for a 2D frame camera relates a 3D point $P\,(X, Y, Z)$ in the world coordinate system, to a 2D pixel location $p\,(u, v)$ on the image plane through the projection matrix $\mathbf{M} \in \mathbb{R}^{3\times4}$ which is a product of both the extrinsic and intrinsic parameters. The extrinsics is comprised of $\mathbf{R} \in \mathbb{R}^{3\times3}$ square rotational matrix from the SO_3 group, and the translation vector $\mathbf{t} \in \mathbb{R}^3$. The intrinsics include the focal lengths f_x and f_y and the optical centre $C\,(u_0, v_0)$. The PCM for a frame camera in inhomogeneous coordinates is defined as:

$$\begin{aligned} u &= u_0 + f_x \frac{r_{11}X + r_{12}Y + r_{13}Z + t_1}{r_{31}X + r_{32}Y + r_{33}Z + t_3} + \Delta u \\ v &= v_0 + f_y \frac{r_{21}X + r_{22}Y + r_{23}Z + t_2}{r_{31}X + r_{32}Y + r_{33}Z + t_3} + \Delta v\,, \end{aligned} \tag{1}$$

where (u, v) represents the distorted pixels seen on the image plane and $(\Delta u, \Delta v)$ are the deviations of these pixels due to distortions caused by a lens. $(r_{11}...r_{33})$ are the elements of $\mathbf{R}$ and (t_1, t_2, t_3) are the elements of $\mathbf{t}$. $\mathbf{R}$ can be parameterized using Euler angles (r_z, r_y, r_x). If the camera is assumed to be intrinsically calibrated and the distortion is removed from the images, the vector for the calibrated intrinsic parameters is $\boldsymbol{\theta}_{f_K} \in \mathbb{R}^4$ is defined as $\boldsymbol{\theta}_{f_K} = [u_0, v_0, f_x, f_y]$, and the vector for the unknown extrinsic parameters is $\boldsymbol{\theta}_{f_E} \in \mathbb{R}^6$ is defined as $\boldsymbol{\theta}_{f_E} = [t_1, t_2, t_3, r_z, r_y, r_x]$.

B. Line-scan Camera

A line-scan camera can be regarded as a special form of a frame camera where all pixels are located along $u = 0$ of the image plane. The extrinsics are unchanged, but the intrinsics are modified such that $u_0 = 0$ and $f_x = 1$. The distortion only occurs about the v direction and can be modelled as two order-radial $[K_1, K_2]$ and one-order tangential P_1 coefficients [14]. The formulation for the PCM of a line-scan camera in inhomogeneous coordinates is defined:

$$\begin{aligned} 0 &= \frac{r_{11}X + r_{12}Y + r_{13}Z + t_1}{r_{31}X + r_{32}Y + r_{33}Z + t_3} \\ v &= v_0 + f_y \frac{r_{21}X + r_{22}Y + r_{23}Z + t_2}{r_{31}X + r_{32}Y + r_{33}Z + t_3} + \Delta v\,, \end{aligned} \tag{2}$$

where $(0, v)$ represents the distorted pixels seen on the image line. The unknown intrinsic and extrinsic parameters can be combined into a vector $\boldsymbol{\theta}_l \in \mathbb{R}^{11}$ which is defined as $\boldsymbol{\theta}_l = [v_0, f_y, K_1, K_2, P_1, t_1, t_2, t_3, r_z, r_y, r_x]$.

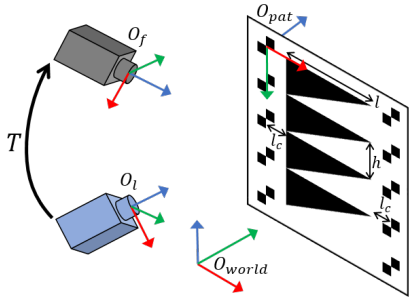


Fig. 2: The coordinate frames of the cameras, world and calibration pattern. The pattern is annotated with the required dimensions. $\mathbf{T}$ represents the unknown rigid transformation between the cameras.

C. Probabilistic Behaviour

Vision sensors exhibit uncertain behaviour which is seen as noise in the pixel locations on the image plane. Let us assume there is no motion of the cameras or the objects it is capturing. Therefore, the noise is only due to the internal hardware and electronics of the sensing circuity. This noise is assumed to be constant and independent for both components of any pixel in an image, and thus modelled as a zero-mean Gaussian with a variance of σ_u^2. For I images, where each image contains N corresponding features, the u and v components of each feature are stacked into a single measurement vector $\mathbf{u} \in \mathbb{R}^{2NI}$, where the noise is modelled as a zero-mean Gaussian with a diagonal covariance matrix $\boldsymbol{\Sigma}_u \in \mathbb{R}^{2NI\times2NI}$. These noisy measurements induce uncertainties in all calculated parameters, therefore the noise in the calibration parameters can also be modelled as a zero-mean Gaussian. The covariance matrices for the noise in the frame camera and line-scan camera parameters are defined as $\boldsymbol{\Sigma}_{f_K} \in \mathbb{R}^{4\times4}$, $\boldsymbol{\Sigma}_{f_E} \in \mathbb{R}^{6\times6}$ and $\boldsymbol{\Sigma}_l \in \mathbb{R}^{11\times11}$.

III. Calibration Procedure

The steps for calibrating the camera system follow the works of [3] and [9]. The frame camera is assumed to be intrinsically calibrated with Zhang's method [15]. The optimized calibration parameters for the line-scan camera $\widehat{\boldsymbol{\theta}}_l$ are found through the maximum likelihood estimation (MLE) by minimizing the reprojection error:

$$\widehat{\boldsymbol{\theta}}_l = \text{argmin}||\mathbf{u} - f(\boldsymbol{\theta}_l, \mathbf{x})||^2_{\boldsymbol{\Sigma}_u}\,, \tag{3}$$

where $\mathbf{x} \in \mathbb{R}^{3NI}$ is a vector of all 3D feature points seen by the line-scan camera in a common reference frame. The

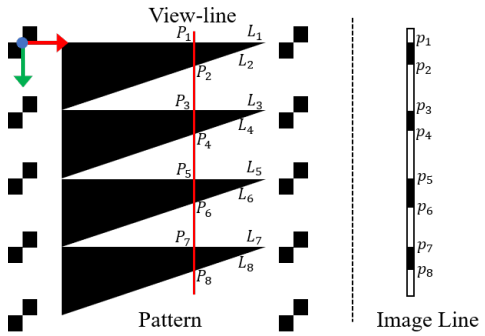


Fig. 3: Red view-line of the line-scan camera which intersects the pattern lines $(L_1 ... L_8)$ resulting in pattern feature points $(P_1 ... P_8)$. The corresponding image line with the pixel feature points $(p_1 ... p_8)$.

function f represents the non-linear projection model for the line-scan camera as detailed in (2). The problem is solved by non-linear least squares. This section will describe how $\mathbf{x}$ is determined and how the calibration is performed.

A. Pattern and Coordinate Frames

The cameras and the pattern are shown Fig. 2 with their respective coordinate frames and the required dimensions of the pattern. All poses are relative to the world coordinate frame O_{world}. The cameras are assumed to be rigidly mounted together via an unknown transformation $\mathbf{T}$. The calibration is performed by taking O_{pat} to be O_{world}.

The pattern is adopted from [9]. It is comprised of right-angled triangles which provide feature points for the line-scan camera and cross-corner markers to determine the pattern's pose. All points lie on the XY plane of the pattern. To avoid the calibration being under-determined, at least two different poses of the pattern must be captured, where the pattern contains at least three triangles and at least four cross-corner-markers.

B. Cross-ratio Point Correspondence

For the line-scan camera, difficulties arise in determining the location of the unknown view-lines on the pattern. The cross-ratio η has been employed to determine these view-lines [3][4][6][9]. It is a scale invariant relationship which links a set of four collinear points, with a different set, given that they are related by a projective transformation:

$$\eta = \frac{(c-a)(d-b)}{(c-b)(d-a)} = \frac{(C-A)(D-B)}{(C-B)(D-A)}\,. \tag{4}$$

The values (a, b, c, d) are image coordinates in pixels and (A, B, C, D) are pattern coordinates in world units. To determine the unknown feature points, their individual components are considered. Firstly, examining Fig. 3, the equations of the pattern lines $(L_1, L_2, ...L_N)$ w.r.t O_{pat} can be defined as:

$$\begin{aligned} Y &= h(i-1)/2 && i = (1,3,5,...N\text{–}1) \\ Y &= -Xh/l + h(2l_c + il)/2l && i = (2,4,6,...N) \end{aligned} \tag{5}$$

The following steps are taken to locate these points:

1) Y component of $(P_1, P_3, P_5, ...P_{N-1})$ are known.
2) Cross-ratio is used to determine Y component of $(P_2, P_4, P_6, ...P_N)$.
3) X component of $(P_2, P_4, P_6, ...P_N)$ is found using (5).
4) Least squares linear fit of $(P_2, P_4, P_6, ...P_N)$.
5) X component of $(P_1, P_3, P_5, ...P_{N-1})$ is found by substituting into the linear fit from the previous step.

C. Common Reference Frame

The extrinsic transformation of pose I of the pattern w.r.t to O_f is $\mathbf{E}_I \in \mathbb{R}^{4\times4}$. This is estimated through the Perspective-n-Point (PnP) algorithm using the known cross-corner markers. $\mathbf{E}_I$ transforms feature points from O_{world} to the common reference frame of O_f as follows:

$$\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}_f = \mathbf{E}_I \begin{bmatrix} X \\ Y \\ 0 \\ 1 \end{bmatrix}_{world} \tag{6}$$

D. Closed Form Solution

A closed form solution for the calibration can be determined using direct linear transformation (DLT) [3][9][10]. This solution does not include the distortion parameters, thus it is used as the initial guess for the optimization. DLT of (2) results in two over-determined homogeneous equations with matrices $\mathbf{A}_1 \in \mathbb{R}^{NI\times4}$ and $\mathbf{A}_{23} \in \mathbb{R}^{NI\times8}$ and solutions $\mathbf{m}_1 \in \mathbb{R}^4$ and $\mathbf{m}_{23} \in \mathbb{R}^8$ respectively. The closed form solution for the unknown parameters is obtained as such:

$$\begin{aligned} \mathbf{n}_1 &= \begin{bmatrix} \mathbf{m}_1(1) & \mathbf{m}_1(2) & \mathbf{m}_1(3) \end{bmatrix}^\top \\ \mathbf{n}_2 &= \begin{bmatrix} \mathbf{m}_{23}(1) & \mathbf{m}_{23}(2) & \mathbf{m}_{23}(3) \end{bmatrix}^\top \\ \mathbf{n}_3 &= \begin{bmatrix} \mathbf{m}_{23}(5) & \mathbf{m}_{23}(6) & \mathbf{m}_{23}(7) \end{bmatrix}^\top \\ \gamma_1 &= \pm 1/||\mathbf{n}_1|| \quad \gamma_3 = \pm 1/||\mathbf{n}_3|| \end{aligned} \tag{7}$$

Intrinsic, rotation and translation parameters:

$$\begin{aligned} v_0 &= \gamma_3^2(\mathbf{n}_2 \cdot \mathbf{n}_3) \quad f_y = ||\gamma_3^2(\mathbf{n}_2 \times \mathbf{n}_3)|| \\ \mathbf{r}_1 &= \gamma_1 \mathbf{n}_1 \quad \mathbf{r}_2 = -(\mathbf{r}_1 \times \mathbf{r}_3) \quad \mathbf{r}_3 = \gamma_3 \mathbf{n}_3 \\ t_1 &= \gamma_1 \mathbf{m}_1(4) \quad t_2 = \frac{\gamma_3 \mathbf{m}_{23}(4) - v_0 t_3}{f_y} \quad t_3 = \gamma_3 \mathbf{m}_{23}(8)\,. \end{aligned} \tag{8}$$

The vectors $\mathbf{r}_1$, $\mathbf{r}_2$ and $\mathbf{r}_3$ are the rows of the rotation matrix $\mathbf{R}$. Note that γ_1 and γ_3 each have two solutions which account for the mirror pose of $\mathbf{T}$, therefore their correct values need to be chosen that replicate the camera setup.

IV. Uncertainty Propagation

The uncertainty in the calibration arises from the pixel noise in $\mathbf{u}$. Therefore, this section describes the first-order covariance propagation used to estimate the uncertainty in $\boldsymbol{\theta}_l$ as demonstrated by [2].

A. Frame Camera

For the intrinsically calibrated frame camera, $\boldsymbol{\Sigma}_{f_E}$ is determined by propagating covariances through the PnP algorithm. The authors of [2] show that $\boldsymbol{\Sigma}_{f_E}$ is a combination of the feature noise introduced by $\boldsymbol{\Sigma}_u$ and noise in the intrinsics $\boldsymbol{\Sigma}_{f_K}$, which can be approximated by a scaled

unscented transform (UT) [16] to give $\mathbf{\Sigma}_{UT-f}$. Therefore, $\mathbf{\Sigma}_{f_E}$ for the frame camera is defined as such:

$$\mathbf{\Sigma}_{f_E} = (\mathbf{J}_f^\top \mathbf{\Sigma}_u^{-1} \mathbf{J}_f)^{-1} + \mathbf{\Sigma}_{UT-f}\,, \tag{9}$$

where $\mathbf{J}_f \in \mathbb{R}^{6\times 4}$ is the Jacobian of the function for the PnP algorithm w.r.t $\boldsymbol{\theta}_{f_K}$. This function was assumed to be (1) as the closed form for the PnP algorithm is difficult to express. $\mathbf{\Sigma}_{UT-f}$ requires nine sigma points with the corresponding weights that are determined using $\boldsymbol{\theta}_{f_K}$, $\mathbf{\Sigma}_{f_K}$ and UT scaling parameters [16].

B. Line-scan Camera

The uncertainty in the optimized line-scan camera parameters $\mathbf{\Sigma}_l$ takes a similar form to (9) and is referred to in the text as Ozog's method:

$$\mathbf{\Sigma}_l = (\mathbf{J}_l^\top \mathbf{\Sigma}_u^{-1} \mathbf{J}_l)^{-1} + \mathbf{\Sigma}_{UT-l}\,, \tag{10}$$

where $\mathbf{J}_l \in \mathbb{R}^{11\times 2NI}$ is the Jacobian of function f in (3) w.r.t $\boldsymbol{\theta}_l$. $\mathbf{\Sigma}_{UT-l}$ now arises from uncertainties in the world coordinates $\mathbf{x}$ given by the covariance $\mathbf{\Sigma}_x \in \mathbb{R}^{3NI\times 3NI}$. This covariance is required to calculate $\mathbf{\Sigma}_{UT-l}$, and is determined by propagating the covariance $\mathbf{\Sigma}_u$ through the process of obtaining $\mathbf{x}$.

Firstly, the feature points on the pattern were determined using the corresponding pixel locations on the image line. The steps for estimating the feature points required the cross-ratio (4), pattern line equations (5), and a linear least squares fit. This process takes in uncertain pixel coordinates $\mathbf{u}$ and outputs their uncertain world coordinates $[X, Y, 0]^\top_{world}$. Therefore, $\mathbf{\Sigma}_u$ gets propagated to these world coordinates. Then, the points are transformed to O_f via $\mathbf{E}_I$ per (6) to give $\mathbf{x}$. $\mathbf{E}_I$ is equivalent to $\boldsymbol{\theta}_{f_E}$ through a conversion from rotation matrix to Euler angles. For pose I, the uncertainty in $\boldsymbol{\theta}_{f_E}$ is $\mathbf{\Sigma}_{f_E}$ which is calculated in (9). Propagating both uncertainties in $[X, Y, 0]^\top_{world}$ and $\mathbf{\Sigma}_{f_E}$ gives $\mathbf{\Sigma}_x$. $\mathbf{\Sigma}_{UT-l}$ requires $2NI+1$ sigma points with the corresponding weights that are determined using $\mathbf{x}$, $\mathbf{\Sigma}_x$ and UT scaling parameters.

V. Active Calibration

Calibration should be performed with a variation of measurements that have sufficient parallax. We propose a technique to determine if new measurements will improve the estimation of the optimized parameters by looking at the observability. Firstly, the observation model for the camera system can be defined as such:

$$\mathbf{u} = f(\boldsymbol{\theta}_l, \mathbf{x}) + \nu_u + \nu_x\,. \tag{11}$$

The measurement noise of the observations ν_u is modelled as a zero-mean Gaussian with covariance $\mathbf{\Sigma}_u$. The feature points $\mathbf{x}$ are prior parameters that introduce noise ν_x, which is also modelled as a zero-mean Gaussian with a covariance $\mathbf{\Sigma}_{x_u} \in \mathbb{R}^{2NI\times 2NI}$. This covariance is propagated from $\mathbf{\Sigma}_x$ by calculating the Jacobian $\mathbf{J}_x$ of the function f in (3) w.r.t $\mathbf{x}$. The observability analysis for an optimization formulated as a non-linear least squares, can be performed by constructing the Fisher information matrix (FIM) [17][18]:

$$\mathbf{J} = \mathbf{J}_l^\top \mathbf{\Sigma}_T^{-1} \mathbf{J}_l\,. \tag{12}$$

Algorithm 1: Active calibration of the camera system

Input: $\mathbf{u}$, $\mathbf{x}$, $\mathbf{\Sigma}_u$, $\mathbf{\Sigma}_x$
Output: $\widehat{\boldsymbol{\theta}_l}$, $\widehat{\mathbf{\Sigma}_l}$

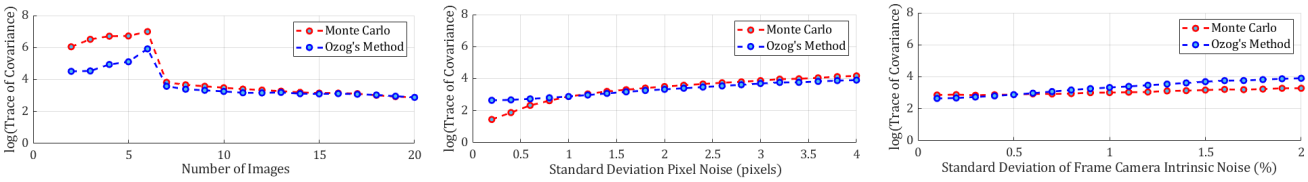
function Active Calibration
  $Current\_Images \leftarrow [1]$
  for $i = 2$ *to* I **do**
    $Current\_Images \leftarrow \text{Append}(i)$
    $\boldsymbol{\theta}_l \leftarrow \text{NonLinearCalibration}(\mathbf{u} - f(\mathbf{x}, \boldsymbol{\theta}_l))$
    $\mathbf{J}_l \leftarrow \frac{\partial f}{\partial \boldsymbol{\theta}_l}$
    $\mathbf{\Sigma}_l \leftarrow (\mathbf{J}_l^\top \mathbf{\Sigma}_u^{-1} \mathbf{J}_l)^{-1} + UT(\mathbf{x}, \mathbf{\Sigma}_x)$
    $\mathbf{J}_x \leftarrow \frac{\partial f}{\partial \mathbf{x}}$
    $\mathbf{\Sigma}_{x_u} \leftarrow \mathbf{J}_x^\top \mathbf{\Sigma}_x \mathbf{J}_x$
    $\mathbf{\Sigma}_T \leftarrow \mathbf{\Sigma}_{x_u} + \mathbf{\Sigma}_u$
    $\mathbf{J} \leftarrow \mathbf{J}_l^\top \mathbf{\Sigma}_T^{-1} \mathbf{J}_l$
    $(\lambda_1 ... \lambda_{11}) \leftarrow \text{EigenValues}(\mathbf{J})$
    if $\text{SumRelativeChange}(\lambda_1 ... \lambda_{11}) > 0$ **then**
      $\widehat{\boldsymbol{\theta}_l} \leftarrow \boldsymbol{\theta}_l$
      $\widehat{\mathbf{\Sigma}_l} \leftarrow \mathbf{\Sigma}_l$
      if SteadyStateReached() **then**
        StopCalibration()
    else
      $Current\_Images \leftarrow \text{Remove}(i)$

$\mathbf{J} \in \mathbb{R}^{11\times 11}$ and $\mathbf{\Sigma}_T \in \mathbb{R}^{2NI\times 2NI}$ is defined as $\mathbf{\Sigma}_T = \mathbf{\Sigma}_u + \mathbf{\Sigma}_{x_u}$. The calibration is observable if $\mathbf{J}$ is full rank which ensures the Cramér–Rao lower bound exists [19][20]. The eigenvalues $(\lambda_1 ... \lambda_{11})$ of $\mathbf{J}$ generally relate to the observability of each parameter in $\boldsymbol{\theta}_l$. Therefore, by ensuring that $(\lambda_1 ... \lambda_{11})$ are large, the lower bound of the uncertainty for each parameter will be small. This concept is used to evaluate new measurements by calculating a metric of the sum of relative change between the current and previous $(\lambda_1 ... \lambda_{11})$. By only keeping measurements where the relative sum is positive, we ensure improved individual parameter observability. This metric also provides the capability of terminating the calibration once a steady-state is achieved. The proposed algorithm is described in Algorithm 1.

VI. Experiments

A. Validation of Uncertainty Propagation

Experiments in simulation were carried out to confirm if the uncertainty $\mathbf{\Sigma}_l$ was correctly propagated from $\mathbf{\Sigma}_u$ and $\mathbf{\Sigma}_{f_K}$ using Ozog's method. A Monte Carlo simulation was performed with added pixel noise of SD σ_u and added noise in $\boldsymbol{\theta}_{f_K}$. The trace of $\mathbf{\Sigma}_l$ shows the results for changing number of images, changing σ_u, and changing added noise in $\boldsymbol{\theta}_{f_K}$, which can be seen in Fig. 4a, Fig. 4b and Fig. 4c respectively. The estimated uncertainty by Ozog's method is valid for both changing noises and number of images. Large differences between the plots are a collective result of ignoring higher-order-terms during the error propagation steps. Fig. 4a also shows that the uncertainty is high in the beginning likely due to insufficient information for calibration.



(a) Changing number of images (b) Changing pixel image noise (c) Changing frame camera intrinsic noise

Fig. 4: $\boldsymbol{\Sigma}_l$ estimated using Ozog's method was compared to a Monte Carlo simulation of 1000 runs. In (a), the trace is compared to a changing number of images, with fixed noises $\sigma_u = 1$ and the SD of the noise in $\boldsymbol{\theta}_{f_K}$ is 0.3% of its mean value. In (b), σ_u was varied while the SD of the noise in $\boldsymbol{\theta}_{f_K}$ is fixed. In (c), the SD of the noise in $\boldsymbol{\theta}_{f_K}$ is varied with fixed pixel noise.

B. Active Calibration in Simulation

Experiments in simulation were performed using 60 poses of the pattern where ground-truth was available. The algorithm was compared to a naive approach which calibrated using all images. The results for the total normalized absolute errors are shown in Fig. 5a, the trace of $\boldsymbol{\Sigma}_l$ is shown in Fig. 5b, and absolute errors in the optimized parameters are shown in the plots of Fig. 5c, Fig. 5d, and Fig. 5e. The algorithm chose the best 38 images out of the 60, where the last image was the 54$^{\text{th}}$ at which it reached steady state and terminated the calibration. When comparing the results for the final optimized parameters, the active run showed a 26.4% drop in the total normalized error and a 46.8% drop in the trace of $\boldsymbol{\Sigma}_l$. The time to complete the calibration was 76s for active and 19s for naive on the same computational hardware.

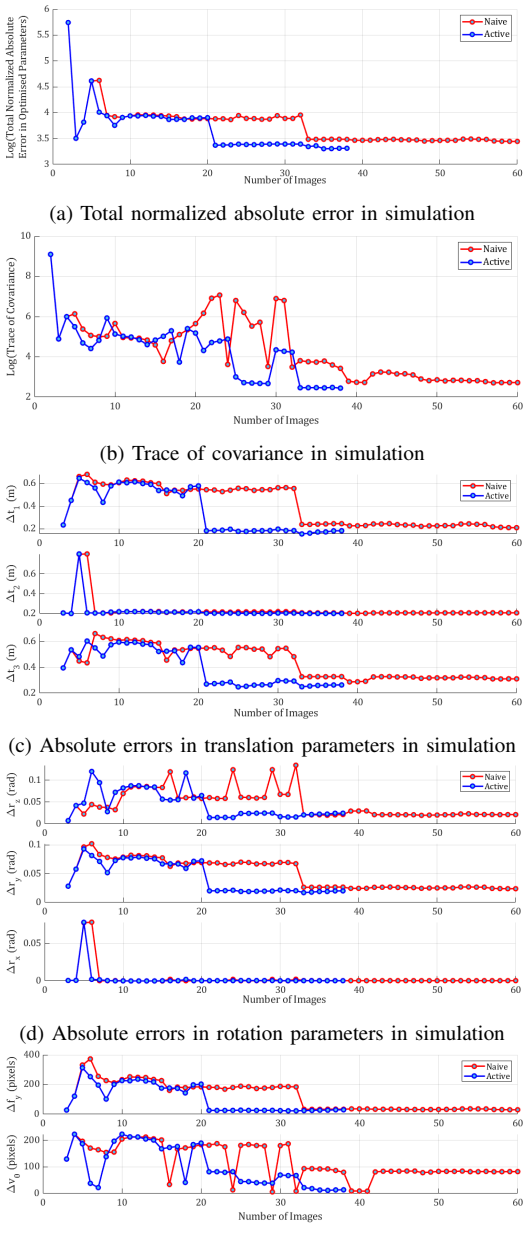
C. Active Calibration of Real Camera System

Active calibration of the camera system shown in Fig. 1 was performed. Naive and active runs were compared using 40 poses of the pattern. Fig. 6a shows the trace of $\boldsymbol{\Sigma}_l$ for both runs. The algorithm used the best 22 images from the 40 available. It did not reach a steady state so all 40 images were evaluated. When comparing the results for the final optimized parameters, the active run showed a 35.4% drop in the trace of $\boldsymbol{\Sigma}_l$. The active calibration parameters were then used to transform and project a hyperspectral view-line onto an image of the pattern as shown in Fig. 6b. The uncertainty in camera parameters is used to calculate the 1-σ error boundary of the pixel coordinates. The time to complete the calibration was 52s for active and 5s for naive on the same computational hardware.

D. Discussion of Calibration Experiments

The results of the experiments show an advantage to using the active calibration algorithm, although it is comparatively slower due to its iterative nature. The poor results of the naive runs can be attributed to calibrating using measurements with insufficient parallax. The error plots in Fig. 5c, Fig. 5d and Fig. 5e are roughly reflected by the total error evaluation metric in Fig. 5a. This evaluation metric gave equal weighting to all parameters, therefore, large variations in some parameters, such as the distortion coefficients, greatly influenced the plot.

The metric used in the active calibration algorithm only considers the sign of the relative change in eigenvalue sum. If the metric is positive, the latest measurement has likely



(a) Total normalized absolute error in simulation

(b) Trace of covariance in simulation

(c) Absolute errors in translation parameters in simulation

(d) Absolute errors in rotation parameters in simulation

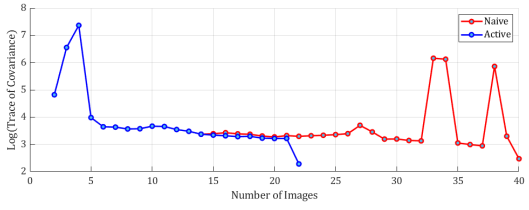
(e) Absolute errors in intrinsic parameters in simulation

Fig. 5: Active and naive runs were compared in simulation. The total normalized absolute errors in $\boldsymbol{\theta}_l$ is shown in (a), the trace of $\boldsymbol{\Sigma}_l$ is shown in (b), absolute errors in t_1, t_2, and t_3 are shown in (c), absolute errors in r_z, r_y, and r_x are shown in (d), and absolute errors in f_y and v_0 are shown in (e). For both runs, $\sigma_u = 1$ and the SD of the noise in $\boldsymbol{\theta}_{f_K}$ is 0.3% of its mean value.

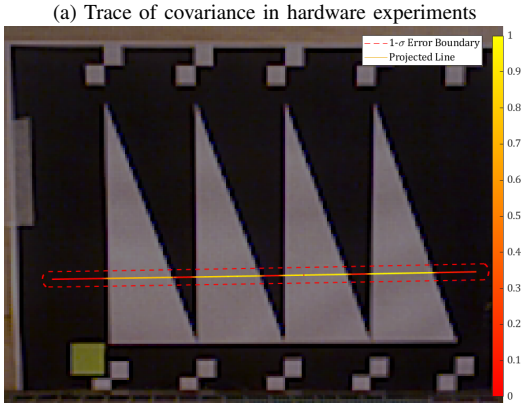
increased parameter observability, but the magnitude of the metric is disregarded. In terms of early termination, there may be a risk of ignoring later observations which may improve parameter estimation. The algorithm focuses on keeping measurements that increase the metric until the eigenvalues stabilize, at this point it assumes that no further measurements will improve the results. The algorithm may also terminate too early, using fewer than 10 images, which generally resulted in higher uncertainties as shown in Fig. 4a, Fig. 5b, and Fig. 6a. Therefore, a minimum number of poses should be set.

VII. Conclusions

A probabilistic calibration for a line-scan and frame camera system is proposed, which propagated noise in the pixel measurements to the optimized parameters. This method of uncertainty propagation provided the groundwork for an observability analysis of the optimization-based calibration, which was then exploited to evaluate new measurements based on their contribution towards improving parameter observability. The algorithm of active calibration demonstrated that not all new measurements were beneficial. Ignoring such measurements resulted in calibration parameters closer to ground-truth with lower uncertainties. The observability metric was also used to terminate the calibration once reaching a steady state value. Future work will consider using the calibrated camera system to estimate material properties, with a level of uncertainty, through the added features provided by the hyperspectral line-scan camera.



(a) Trace of covariance in hardware experiments



(b) Projected spectral view-line on pattern

Fig. 6: Active and naive runs were performed on a real camera system. The trace of $\boldsymbol{\Sigma}_l$ is shown in (a). For both runs, $\sigma_u = 1$ and the SD of the noise in $\boldsymbol{\theta}_{f_K}$ was retrieved from MATLAB's Camera Calibration toolbox. In (b), a hyperspectral view-line was projected to a frame camera with a 1-σ error boundary.

Acknowledgment

This paper is supported by funding from the Australian Government Department of Agriculture and Water Resources as part of its Rural R&D for Profit program, MLA grant number V.RDP.2005.

References

[1] G. Hu, S. Huang, L. Zhao, A. Alempijevic, and G. Dissanayake, "A robust RGB-D slam algorithm," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2012, pp. 1714–1719.

[2] P. Ozog and R. M. Eustice, "On the importance of modeling camera calibration uncertainty in visual SLAM," *Proceedings - IEEE International Conference on Robotics and Automation*, pp. 3777–3784, 2013.

[3] B. Sun, J. Zhu, L. Yang, S. Yang, and Z. Niu, "Calibration of line-scan cameras for precision measurement," *Applied optics*, vol. 55, no. 25, pp. 6836–6843, 2016.

[4] D. Li, G. Wen, B. Wei Hui, S. Qiu, and W. Wang, "Cross-ratio invariant based line scan camera geometric calibration with static linear data," *Optics and Lasers in Engineering*, vol. 62, pp. 119–125, 2014.

[5] C. A. Luna, M. Mazo, J. L. Lázaro, and J. F. Vázquez, "Calibration of line-scan cameras," *IEEE Transactions on Instrumentation and Measurement*, vol. 59, no. 8, pp. 2185–2190, 2010.

[6] D. Su, A. Bender, and S. Sukkarieh, "Improved cross-ratio invariant-based intrinsic calibration of a hyperspectral line-scan camera," *Sensors (Switzerland)*, vol. 18, no. 6, 2018.

[7] G. Di Leo, C. Liguori, A. Pietrosanto, and R. Lengu, "Uncertainty of line camera image based measurements," *IEEE International Conference on Instrumentation and Measurement Technology*, pp. 1–6, 2017.

[8] A. Wendel and J. Underwood, "Extrinsic parameter calibration for line scanning cameras on ground vehicles with navigation systems using a calibration pattern," *Sensors (Switzerland)*, vol. 17, no. 11, 2017.

[9] D. Li, G. Wen, and S. Qiu, "Cross-ratio–based line scan camera calibration using a planar pattern," *Optical Engineering*, vol. 55, no. 1, p. 014104, 2016.

[10] B. Hui, G. Wen, P. Zhang, and D. Li, "A novel line scan camera calibration technique with an auxiliary frame camera," *IEEE Transactions on Instrumentation and Measurement*, vol. 62, no. 9, pp. 2567–2575, 2013.

[11] L. Zhu, H. Luo, and X. Zhang, "Uncertainty and sensitivity analysis for camera calibration," *Industrial Robot*, vol. 36, no. 3, pp. 238–243, 2009.

[12] G. Di Leo and A. Paolillo, "Uncertainty evaluation of camera model parameters," *Conference Record - IEEE Instrumentation and Measurement Technology Conference*, pp. 598–603, 2011.

[13] R. Galego, A. Ortega, R. Ferreira, A. Bernardino, J. Andrade-Cetto, and J. Gaspar, "Uncertainty analysis of the DLT-Lines calibration algorithm for cameras with radial distortion," *Computer Vision and Image Understanding*, vol. 140, pp. 115–126, nov 2015.

[14] S. Fang, X. Xia, and Y. Xiao, "A calibration method of lens distortion for line scan cameras," *Optik*, vol. 124, no. 24, pp. 6749–6751, 2013.

[15] Z. Zhang, "A Flexible New Technique for Camera Calibration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, pp. 1330–1334, dec 2000.

[16] S. J. Julier, "The scaled unscented transformation," *Proceedings of the American Control Conference*, vol. 6, no. 2, pp. 4555–4559, 2002.

[17] Z. Wang and G. Dissanayake, "Observability analysis of SLAM using fisher information matrix," *2008 10th International Conference on Control, Automation, Robotics and Vision, ICARCV 2008*, no. December, pp. 1242–1247, 2008.

[18] D. Su, T. Vidal-Calleja, and J. V. Miro, "Asynchronous microphone arrays calibration and sound source tracking," *Autonomous Robots*, vol. 44, no. 2, pp. 183–204, 2020.

[19] F. Nobre and C. Heckman, "Learning to calibrate: Reinforcement learning for guided calibration of visual–inertial rigs," *The International Journal of Robotics Research*, vol. 38, no. 12-13, pp. 1388–1402, 2019.

[20] Y. Bar-Shalom, X.-R. Li, and T. Kirubarajan, "Estimation with applications to tracking and navigation," in *Estimation with Applications to Tracking and Navigation*. John Wiley & Sons, 2001, ch. 3, pp. 164–165.