2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
September 24–28, 2017, Vancouver, BC, Canada

Autonomous Active Calibration of a Dynamic Camera Cluster using Next-Best-View

Jason Rebello* Arun Das§ Steven Waslander†

Abstract— **Dynamic camera cluster (DCC) calibration determines a time-varying set of extrinsic calibration transformations between cameras in a multi-camera cluster, where one or more cameras are mounted on actuated mechanisms. In this paper, we present a novel active vision approach for DCC calibration, which directly reduces the parameter uncertainty by selecting calibration measurements using an information theoretic next-best-view policy. Our system automatically selects the next best measurement for the calibration by determining the optimal actuator inputs which minimize the predicted covariance of the extrinsic parameters. We show that our method is able to successfully estimate calibration parameters up to a user specified accuracy with no manual excitation. We test our method in simulation on a variety of actuated mechanisms and validate the results on a real 3 axis gimbal, and demonstrate our approach is able to achieve accurate calibrations using fewer measurement sets when compared to existing approaches.**

I. Introduction

Multi-Camera Cluster (MCC) localization accuracy is directly related to its motion through, and perception of, the environment. In static MCC configurations, where the extrinsic calibration is not time-varying, the robot's perception of its environment is directly coupled to its motion, thus, movements which cause occlusions in one or more cameras can degrade the pose estimation quality. The use of a dynamic camera cluster (DCC) in such situations is advantageous as it allows the robot to independently move the cluster cameras and track informative landmarks irrespective of the robot's motion [1]. However, in order to enable DCCs in a visual SLAM application, a precise, time-varying extrinsic calibration between the cluster cameras is required.

In [1], calibration of a DCC is performed by parameterizing the actuated mechanism between two cameras as the forward kinematics of a serial manipulator, then finding the optimal parameters using measurements to a fiducial target. Although the approach provides accurate calibration, it relies on manual selection of the joint angles, and requires sufficient measurement excitation from different actuator configurations to ensure accurate calibration results.

Manual selection of the measurement configurations for calibration is limiting in two main aspects. First, providing sufficient information to accurately estimate the transformation parameters requires collecting measurements over the full configuration space of the mechanism. Defining a set of collection points is not obvious, especially as the system's configuration space increases in complexity, and a poor selection of measurements will result in an inaccurate or degenerate calibration. Although the configuration space can be sampled to concurrently select measurement configurations for batch processing, such an approach becomes cumbersome and impractical as the dimension of the configuration space grows, and further does not guarantee that the selected samples will provide the sufficient measurement excitation required for accurate calibration. Second, manual selection of the mechanism inputs precludes automating the calibration. Automatic calibration is emerging as a crucial functionality for state-of-the-art applications, as robots deployed in the field and those which are mass produced require accurate calibration that can be performed by non-experts without human intervention.

In response to these limitations, we propose an *active perception* [2] method of calibration for DCCs, which is autonomous, and works on the principle of next-best-view [3] to select the actuator configurations such that measurements of the target from the resulting viewpoint locally minimize parameter uncertainty at each iteration. Our approach requires no human input after commencement of the calibration and terminates only when the desired accuracy in parameters is achieved. We validate our method in simulation and using a physical 3-axis gimbal, and demonstrate that our next-best-view approach is able to perform automatic DCC calibration using fewer images, when compared to random and discretization based sampling strategies.

II. Related Work

Active Vision systems [2] have the advantage of being able to manipulate the viewpoint of a camera in order to obtain maximum information from the environment, and are particularly useful in application such as Visual SLAM, where occlusions and limited field of view are prevailing factors [4]. These systems have also been applied to calibration problems, such as determination of the intrinsic parameters of the camera lens model [5], and determining the extrinsic transformations required to solve the hand-to-eye and hand-in-eye problems[6], [7] by actively moving the camera or a fiducial target.

The *hand-in-eye* problem consists of computing the extrinsic transformation between manipulator's end effector and camera, and the transformation from the manipulator's base frame to the fiducial target frame [8]. *Kinematic* calibration seeks to determine the manipulator's link parameters, and has been performed successfully using both visual sensors

*Ph.D. Candidate, Mechanical and Mechatronics Engineering, University of Waterloo; jrebello@uwaterloo.ca.

§Ph.D. Candidate, Mechanical and Mechatronics Engineering, University of Waterloo; adas@uwaterloo.ca.

†Associate Professor, Mechanical and Mechatronics Engineering, University of Waterloo; stevenw@uwaterloo.ca

978-1-5386-2682-5/17/$31.00 ©2017 IEEE

and LIDAR [9]. Although related to DCC calibration, Hand-Eye and Kinematic calibration approaches generally use only a *single* camera, where as the DCC calibration computes time-varying extrinsics of *multiple* cameras which are related through an actuated mechanism. Autonomous calibration of the hand-in-eye problem has been attempted, however, the existing approaches use a set of predefined motions for measurement collection, and determine some of hand-in-eye transformations using additional, non-camera-based sensing [10], [11]. Closely related to DCC calibration, an extended kinematic calibration has been proposed, which simultaneously estimates the camera parameters, kinematic parameters, and hand-eye relationship [12]. Although similar to active DCC calibration, the existing active calibration approaches use a set of *predefined* movements, while our proposed method actively calculates and selects the next best pose of the camera that results in the maximum reduction in parameter uncertainty.

Next-Best-View (NBV), in general, is the process of determining the next best camera location from which to collect measurements, in order to maximize an information metric that is specific to the task [3]. NBV approaches have been successfully applied to a variety of applications, such as visual servoing [13], 3D reconstruction [14], and monitoring complex industrial processes [15]. In order to formulate the NBV problem, many existing approaches discretize the configuration space, and select the next-best-view from a finite set of possible configurations. For example, assisted intrinsic camera calibration has also been accomplished by generating a discrete set of fiducial target poses, then suggesting the optimal target poses from the set which result in a high quality calibration [5]. Although promising, the efficacy of these existing approaches is heavily dependent on the discretization strategy. In contrast, our NVB formulation requires no discretization of the configuration space, and instead performs a continuous optimization over the configuration space in order to select the next-best-view.

III. Background and Notation

General Rigid Body Transformation: Let a point in 3D expressed in co-ordinate frame, $\mathcal{F}_x$ be denoted as $\mathbf{p}^x \in \mathbb{R}^3$. In order to transform points between co-ordinate frames, we will define the rigid body transformation from frame $\mathcal{F}_a$ to $\mathcal{F}_b$ as $\mathbf{T}_\tau^{b:a} \in \mathbb{SE}(3)$, where $\mathbf{T}_\tau^{b:a} : \mathbb{R}^3 \mapsto \mathbb{R}^3$ and $\tau = [r_x\, r_y\, r_z\, t_x\, t_y\, t_z]^T$ is a parameter vector which is used to construct $\mathbf{T}_\tau^{b:a}$. The first three components of the parameter vector, $r_x, r_y, r_z \in [0, 2\pi]$, represent 3-2-1 Euler angle rotations, while the last three parameters, $t_x, t_y, t_z \in \mathbb{R}^3$, represent the translation parameters along the respective axis.

Image Projections of 3D Points: The projection function, $\Psi(\mathbf{p}_i^c) : \mathbb{R}^3 \mapsto \mathbb{P}^2$, maps a point, $\mathbf{p}_i^c \in \mathbb{R}^3$ in a camera frame $\mathcal{F}_c$, $c \in \{d, s\}$, where d is the set of dynamic cameras and s is the set of static cameras in the DCC, to a pixel location on the 2D image plane, defined as $\Psi(\mathbf{p}_i^c) = [u_i^c\, v_i^c]^T$. Here, u_i and v_i are the pixel co-ordinates in the camera c image.

Denavit-Hartenburg Parameterization: We denote the DH parameters which describe the transformation from link frame $\mathcal{F}_{l-1}$ to link frame $\mathcal{F}_l$ on an actuated mechanism, as $\omega_l = [\theta_l, d_l, a_l, \alpha_l]^T$, where $\theta_l, \alpha_l \in [0, 2\pi]$ and $d, a \in \mathbb{R}$. A homogeneous rigid body transformation, $\mathbf{T}_\tau^{l:l-1} \in \mathbb{SE}(3)$, which describes the transformation between link frames $\mathcal{F}_{l-1}$ and $\mathcal{F}_l$, can be computed as a function of the DH parameters. For an overview of the procedure of applying the Denavit-Hartenburg Parameterization to a manipulator, the reader is referred to [16]. Note that our calibration approach can be generalized for any parametrization that allows for the computation of the mechanism's forward kinematics, and that DH parametrization was selected due to its prevalence within the robotics community.

Entropy of a Gaussian Distribution The Shannon entropy is a measure of the uncertainty of information content [17]. In the case where the probability density function of a continuous random variable, Y, is modelled as a Gaussian distribution, the Shannon entropy, in units of nats, is $h_e(\Sigma) = \frac{1}{2}\ln((2\pi e)^n\,|\Sigma|)$, where Σ is the covariance matrix of the distribution.

IV. Problem Formulation

In this section we present the calibration of a dynamic multi-camera cluster with one static camera, s, and one *dynamic camera*, d, which is mounted to an actuated mechanism. We use a static fiducial marker of known size and assign to it a coordinate frame $\mathcal{F}_t$. We first summarize the author's previous DCC calibration approach [1], and then describe our proposed method of next-best-view measurement selection used to automate the calibration process.

A. DCC Calibration Formulation

The aim of the calibration process is to determine the rigid body transformation $\mathbf{T}_{\Theta,\lambda}^{d:s}$ from the static camera frame, $\mathcal{F}_s$, to the dynamic camera frame, $\mathcal{F}_d$, where Θ is the set of *estimated* parameters which is used to build the rigid body transform, and $\lambda \in \mathbb{R}^L$ is the set of *measured* parameters which are available from either known inputs to the mechanism, or measured joint feedback. For the calibration, the transformation between cameras has the form $\mathbf{T}_{\Theta,\lambda}^{d:s} = \mathbf{T}_{\tau_d}^{d:e}\mathbf{T}_{\omega,\lambda}^{e:b}\mathbf{T}_{\tau_s}^{b:s}$, where $\mathbf{T}_{\tau_s}^{b:s}$ defines the transformation between the static camera and mechanism base frame, $\mathbf{T}_{\omega,\lambda}^{e:b}$ defines the transformation from the base frame of the mechanism to the end effector frame, and $\mathbf{T}_{\tau_d}^{d:e}$ defines the transformation from the end effector frame to the dynamic camera frame. Note that $\mathbf{T}_{\omega,\lambda}^{e:b}$ is a chain of transforms through the mechanism's links computed using its forward kinematics, and is a function of its DH parameters and control inputs.

For each instance where both the static and dynamic cameras capture measurements from the fiducial target, the pose of the camera with respect to the marker frame, $\mathbf{T}^{c:t}$, is determined by solving the perspective-n-point (PnP) problem [18]. We then define the i^{th} *measurement set* as $Z_i = \{P_i^s, P_i^d, Q_i^s, Q_i^d, \lambda_i\}$, where P_i^s, $P_i^d \in \mathbb{R}^3$ are the set of corresponding marker point positions defined in the frames of the static and dynamic cameras, respectively, and are easily computed using the known point positions in the target frame

and the transform to the camera frame, $\mathbf{T}^{c:t}$. The set of pixel measurements to the marker points, as observed by the static and dynamic cameras, are denoted by Q_i^s and Q_i^d $\in \mathbb{R}^2$, respectively, and λ_i is the set of joint angles for the mechanism at snapshot i.

Using the measurement set and the transformation between camera frames, we can now define the re-projection error between the measured marker point j in the static camera frame and the corresponding point measured in the dynamic camera frame, for measurement set i, as

$$e_j^d(\Theta, \lambda_i) = z_j^d - \Psi^d(\mathbf{T}_{\Theta,\lambda_i}^{d:s}\mathbf{p}_j^s) \tag{1}$$

where $z_j^d \in Q_i^d$ is the measurement of point j, from measurement set Q_i^d, observed in the dynamic camera, and $\mathbf{p}_j^s \in P_i^s$ is the 3D position of point j, from the point position set P_i^s, as observed from the static camera. Since both the actuated and static camera observe the same marker at each snapshot, we can similarly compute the error for points observed in the actuated frame and projected into the static frame as $e_j^s(\Theta, \lambda_i) = z_j^s - \Psi^s((\mathbf{T}_{\Theta,\lambda_i}^{d:s})^{-1}\mathbf{p}_j^d)$, where $z_j^s \in Q_i^s$ is the measurement of point j, from pixel measurement set Q_i^s, observed in the static camera, and $\mathbf{p}_j^d \in P_i^d$ is the 3D position of point j, from the point set P_i^d, as observed from the dynamic camera. The total squared re-projection error as a function of the estimation parameters, $\Lambda(\Theta) : \mathbb{R}^n \mapsto \mathbb{R}$ over all of the collected measurement sets, $\Gamma = \{Z_1, Z_2, \ldots, Z_k\}$, is defined as

$$\begin{aligned}\Lambda(\Theta) = \sum_{Z_i \in \Gamma} \sum_{j=1}^{|P_i^s|} & e_j^d(\Theta, \lambda_i)^T e_j^d(\Theta, \lambda_i) \\ & + e_j^s(\Theta, \lambda_i)^T e_j^s(\Theta, \lambda_i).\end{aligned} \tag{2}$$

Finally, an unconstrained optimization of equation (2) is performed in order to find the optimal calibration parameters, Θ^*, which minimize the total re-projection error over the set of collected measurements. Note that Θ^* consists of the parameters required to define the transformation from the static camera frame to the base frame of the manipulator, the DH parameters of the actuated mechanism, and finally the parameters corresponding to the transformation from the end effector frame to the moving camera frame.

B. Autonomous Calibration

While Section IV-A shows that it is possible to achieve a calibration of the dynamic camera cluster, the quality of the calibration is heavily dependent on the ability to collect an extensive set of measurements while providing sufficient excitation to the joints inputs, as the parameter estimates are highly sensitive to the selected measurement sets. Even if an exhaustive measurement set is collected, it does not guarantee accurate calibration after the optimization, as the relationship between manually collected measurements and uncertainty of the estimation parameters is unclear. We therefore present a method that seeks to find a next-best-view which locally minimizes calibration parameters' covariance with each successively collected measurement, until a user specified accuracy is achieved.

1) Parameter Initialization: The autonomous calibration process is initialized with a collection of M measurement sets, denoted as $\widetilde{Z}_{1:M}$, that are obtained by sampling the configuration space. Common strategies such as *simple*, *random*, *systematic*, or *cluster* [19] sampling can be used to generate the initial measurement set collection. Each measurement set Z_i is obtained from a sampled mechanism input λ_i. The prior mean and covariance on the estimated parameters are obtained by optimizing (2) using all the sampled measurements, $\widetilde{Z}_{1:M}$, and the resulting parameter estimate, $\widetilde{\Theta}_{1:M}$, produced by the optimization is assumed to be Gaussian distributed with covariance $\Sigma_{\widetilde{\Theta},\lambda_{1:M}}$.

2) Covariance Calculation: In order to compute the covariance of the parameters, $\Sigma_{\widetilde{\Theta},\lambda_{1:M}}$, the Jacobian of the measurement equation, $J_{\widetilde{\Theta},\lambda_{1:M}}$, is required, where

$$J_{\widetilde{\Theta},\lambda_{1:M}} = \begin{bmatrix} J_{\widetilde{\Theta},\lambda_1} & \cdots & J_{\widetilde{\Theta},\lambda_M} \end{bmatrix}^T. \tag{3}$$

Each row-block of (3), $J_{\widetilde{\Theta},\lambda_i}$, corresponds to the Jacobian contribution of the configuration associated with the i^{th} joint input λ_i, and can be calculated as, $J_{\widetilde{\Theta},\lambda_i} = J_{\Psi^c} J_{\widetilde{\Theta},\lambda_i}^{d:s}$, where $J_{\Psi^c} = \frac{\partial \Psi^c}{\partial \mathbf{p}_j^c}$ is the camera model dependent projection Jacobian for camera c, and $J_{\widetilde{\Theta},\lambda_i}^{d:s}$ is the Jacobian of the kinematic chain from the static camera to the dynamic camera. The Jacobian of the kinematic chain describes the perturbation of the j^{th} 3D point mapped from the static camera frame to the dynamic camera frame, with respect to the transformation parameters, $J_{\widetilde{\Theta},\lambda_i}^{d:s} = \frac{\partial \mathbf{p}_j^d}{\partial \widetilde{\Theta}}$, where $\mathbf{p}_j^d = (\mathbf{T}_{\widetilde{\Theta},\lambda_i}^{d:s})\mathbf{p}_j^s$. Finally, using a first order approximation of the Fisher information matrix, the parameter covariance is given as,

$$\Sigma_{\widetilde{\Theta},\lambda_{1:M}} = (J_{\widetilde{\Theta},\lambda_{1:M}}^T J_{\widetilde{\Theta},\lambda_{1:M}})^{-1}. \tag{4}$$

3) Next-Best-View Configuration Selection: To reduce the uncertainty in the calibration parameters with each subsequent measurement set, we seek a locally optimal mechanism configuration, λ^*, which will minimize the uncertainty of the estimation parameters. Suppose we have an arbitrary mechanism configuration for the next-best-view, $\hat{\lambda}$, then the resulting measurement Jacobian matrix, which includes the measurement from $\hat{\lambda}$, has the form

$$J_{\widetilde{\Theta},\eta}(\hat{\lambda}) = \begin{bmatrix} J_{\widetilde{\Theta},\lambda_{1:M}} & J_{\widetilde{\Theta},\hat{\lambda}} \end{bmatrix}^T, \tag{5}$$

where $\eta = \{\lambda_{1:M}, \hat{\lambda}\}$ denotes the set of actuator inputs from $\lambda_{1:M}$ and the optimal next-best-view configuration, $\hat{\lambda}$. Using (5), the parameter covariance for the estimation parameters can be approximated by

$$\Sigma_{\widetilde{\Theta},\eta}(\hat{\lambda}) = (J_{\widetilde{\Theta},\eta}(\hat{\lambda})^T J_{\widetilde{\Theta},\eta}(\hat{\lambda}))^{-1}. \tag{6}$$

Note that (6) is an approximation to the true parameter uncertainty when measurements from $\hat{\lambda}$ are included, as $\widetilde{\Theta}$ is computed using the configurations from $\lambda_{1:M}$, and does not include $\hat{\lambda}$. The accuracy of this approximation will degrade according to the error between $\widetilde{\Theta}$ and the true estimation parameters, however, in our experiments, we have seen promising results with $\widetilde{\Theta}$ being initialized according

to the process described in Section IV-B.1. Further, the approximation improves as the calibration process proceeds and converges to an accurate set of parameters.

Our next-best-view configuration is determined by formulating a cost function using the covariance matrix given in (6). Suppose we have an actuated mechanism with L joints. Then, we shall define a cost $\Omega : \mathbb{R}^L \mapsto \mathbb{R}$ which is given as,

$$\Omega(\hat{\lambda}) = h_e(\Sigma_{\widetilde{\Theta},\eta}(\hat{\lambda})). \tag{7}$$

The cost function in (7) maps a next-best-view mechanism input, $\hat{\lambda}$, to the expected entropy of the parameter covariance matrix from (6). Although the cost defined in (7) uses the entropy of the covariance matrix in order to quantify the parameter uncertainty, it is possible to also use other metrics, such as the Frobenius norm or the trace of the covariance matrix. Entropy of the covariance matrix was selected in this case, as that metric has been shown to work well in related applications, such as key-frame selection for visual SLAM [20]. Complete exploration and analysis of these alternative uncertainty metrics is left as an area of future work. In order to find the optimal next-best-view, λ^*, the cost function from (7) is optimized over the feasible configurations of the actuated mechanism,

$$\begin{aligned} \min \quad & \Omega(\hat{\lambda}) \\ \text{subject to} \quad & \lambda^l < \hat{\lambda} < \lambda^u, \end{aligned} \tag{8}$$

where λ^l and λ^u are the upper and lower bounds, respectively, of the mechanism input angles.

C. Optimization with Successive Next-Best-View Measurements

Once the next-best-view configuration, λ^*, is determined, the actuated mechanism is moved and a measurement set, Z_{λ^*}, is collected from the corresponding optimal configuration. The measurement set is then appended such that

$$\widetilde{Z}_{1:M+1} \leftarrow \widetilde{Z}_{1:M} \cup Z_{\lambda^*}, \quad M = M + 1. \tag{9}$$

The estimation parameters are optimized using the updated measurement sets, in order to recompute $\widetilde{\Theta}_{1:M}$ using the additional measurements from Z_{λ^*}. Then, the next-best-view selection procedure described in Section IV-B.3 is performed again. The process of selecting the next-best-view, then recalculating the estimation parameters, $\widetilde{\Theta}_{1:M}$, is repeated until the entropy score from (7) reaches a user selected threshold, or a maximum number of views is selected. Figure 1 visualizes the next-best-view cost from (7) for a two degree of freedom (DOF) mechanism, and also illustrates the NBV optimization and selection process over 4 measurement collections.

Note that our next-best-view approach performs a continuous optimization over the mechanism's configuration space, and also takes into account the viewpoint of the fiducial target implicitly in the formulation through the kinematic and projection Jacobians. Thus, our approach does not require discretization of the configuration space, or predefined motion paths over a finite set of target positions.

V. Experimental Validation

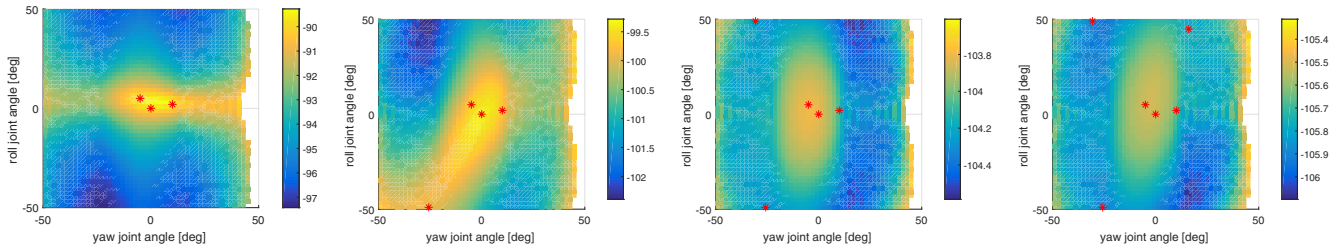
The proposed next-best-view approach is validated using two sets of experiments. In the first set, the calibration of a one, two, and three degree-of-freedom (DOF) actuated mechanism is performed in simulation using a realistic camera lens model which includes radial and tangential distortion, and realistic pixel noise values. In the second set, our autonomous next-best-view calibration is executed on a physical dynamic camera cluster consisting of two Ximea xIQ cameras which operate at 60fps and 900×600 resolution. To build the DCC, one camera is statically mounted and the other is fitted to a 3-DOF gimbal that is typically used for aerial photography applications. We use OpenCV to detect the location of a chessboard target with respect to the camera, however, any fiducial target with known scale is suitable for this application. The experimental set-up is depicted in Figure 2.

A. Simulation Experiments

To validate our next-best-view approach, we generate a 1-,2-, and 3-DOF mechanism in simulation, and perform the DCC calibration using measurements from a simulated fiducial target. In all three cases, we generate a static camera which is fixed in the world, and a moving camera which is attached to the end effector of each actuated mechanism. The 3-DOF mechanism simulates a 3-axis gimbal, and allows the camera to perform yaw, pitch, and roll, motions. The 2-DOF mechanism allows for yaw and roll motions of the dynamic camera, and finally the 1-DOF mechanism simply allows for yaw motions of the dynamic camera.

In order to demonstrate how our automatic next-best-view approach can be used in a field calibration setting, we compare our method to two other view-point selection strategies which could easily be performed by a non-expert human operator in the field. The first competing strategy simply selects random viewpoints within the bounds of the configuration space of the mechanism. The second competing strategy discretizes the configuration space of the mechanism using a linear spacing. For example, suppose we wish to collect measurements from a 2-DOF mechanism with angle bounds of -20 to 20 degrees for each axis. Using a spacing of 20 degrees, the set of M desired viewpoint angles generated using this linear spacing approach is $\lambda_{1:M} = \{(-20,-20), (-20,0), (-20,20), (0,-20), (0,0), (0,20), (20,-20), (20,0), (20,20)\}$. Note that linear spacing based selections are sampled from the joint angle space as opposed to the two dimensional image space, as the former offers a richer set of configurations to consider for maximum covariance reduction. In all cases, the number of collected viewpoints was selected to be the same, in order to compare the entropy reduction for each approach after collecting the same number of measurements.

Figure 3 presents the entropy of the covariance matrix, as computed using (4), for each mechanism. It is evident that for all tested mechanisms, our next-best-view approach provides the lowest covariance matrix entropy score over all collected measurements. Our approach is able to provide



(a) Initial Surface (b) 1 NBV Addition (c) 2 NBV Additions (d) 3 NBV Additions

Fig. 1. Progression of the NBV cost surface for a two degree-of-freedom mechanism, as measurements are added from the NBV configurations. In this case, the autonomous calibration is initialized using 3 configurations, then 3 additional measurements are added using our proposed NBV approach. The red asterisks denote configurations from which the collected measurements were used for parameter estimation. (a) shows the cost surface after initialization, while (b)-(d) depict the changing cost surface as NBV measurements are added. Note that the plots are coloured according to the covariance entropy in nats.

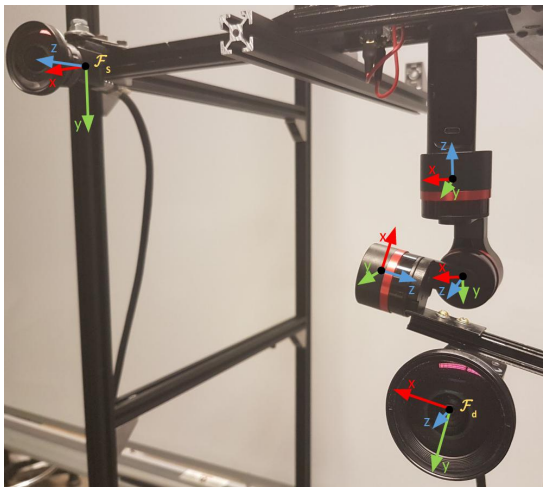


Fig. 2. Dynamic Camera Cluster consisting of one static camera and one dynamic camera mounted on a 3-DOF gimbal.

lower parameter uncertainty, when compared to the random and linear spacing strategy, as the next-best-view method seeks the measurement configurations at each iteration that directly improve the uncertainty of the estimation parameters. Although the random sampling and linear spacing techniques are also able to provide a decrease in parameter uncertainty, those approaches do not evaluate the selected viewpoint, and therefore may include measurements which do not provide significant entropy reduction. This behaviour is evident in Figure 3(b), where between measurements 3 and 4, a viewpoint is selected that does not provide any improvement of the covariance matrix entropy.

From Figure 3, it is also evident that for a desired target covariance entropy, the next-best-view approach is able to achieve the target using fewer viewpoints compared to the random and linear spacing strategies. For example, in Figure 3(c), a target covariance entropy of -133 nats is achieved by the next-best-view approach using 10 views, while the linear spacing and random strategies required 17 and 23 views, respectively, to achieve the same calibration quality. Our results demonstrate that for any parameter entropy target, our next best view approach will be able to achieve the target with fewer collected measurement views compared to the competing strategies.

TABLE I

SUMMARY STATISTICS FOR PIXEL REPROJECTION OF VALIDATION SET

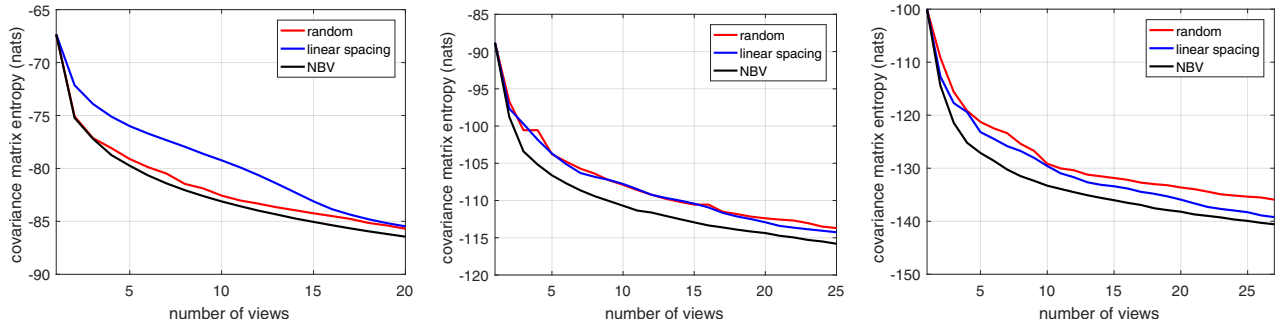
	RMS Pixel Error
NBV	**4.31**
random	6.47
linear spacing	6.32

B. 3-DOF Gimbal Calibration

Our next-best-view approach is also validated using a physical, 3-DOF gimbal, that allows for yaw, pitch, and roll motions of the dynamic camera, which is mounted to the end-effector of the gimbal mechanism. We perform a similar experiment to that presented in Section V-A, where we compare the next-best-view approach to the linear spacing and random sampling viewpoint selection methods. Figure 4 depicts the parameter covariance matrix entropy versus the number of collected views for each of the three strategies tested on the 3-DOF gimbal. We see that the next-best-view approach maintains the lowest entropy over all of the collected viewpoints, and shows a rapid rate of decreasing entropy over the first few collected images.

In order to further verify the calibration quality, we also collect a *validation set*, which consists of 20 independently collected measurement sets, that were not used in the calibration process. The 3D points from the validation set are transformed according to the estimation parameters, into the observing camera, then projected into the image. Table I presents the RMS re-projection errors of the validation set, using the calibration parameters that were computed using the tested approaches For each case, the calibration parameters are computed using 5 collected measurements, in order to illustrate the efficacy of the next-best-view approach.

As seen in Table I, the next-best-view strategy achieves the lowest RMS pixel error, after using 5 viewpoints. Note that in our formulation does not currently account for offset or scaling errors in the joint angle feedback. As such, measured encoder angles add additional error to the transformation between the static and dynamic camera, and subsequently increase the RMS pixel error of the validation set. This can be mitigated by adding additional offset and scaling variables for each joint axis as part of the optimization, and is left as an area of future work.



(a) 1-DOF mechanism (b) 2-DOF mechanism (c) 3-DOF mechanism

Fig. 3. Comparison of the parameter covariance entropy versus the number of collected views, for the random sampling, linear spacing, and next-best-view approaches. (a)-(c) present the results for the 1-,2-,and 3-DOF cases, respectively, where it is evident that our next-best-view approach provides the greatest uncertainty reduction across all collected measurements.

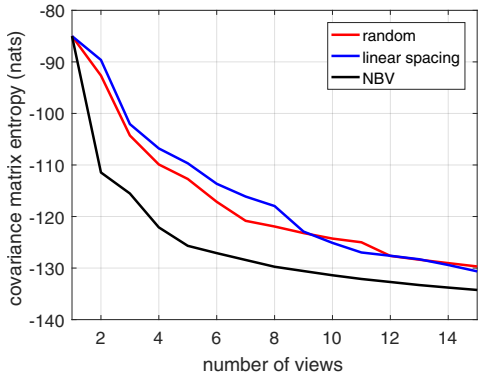


Fig. 4. Comparison of the parameter covariance entropy versus the number of collected views, for the random sampling, linear spacing, and next-best-view approaches, using the hardware set-up depicted in Figure 2.

VI. Conclusion

This work presents an autonomous calibration method for dynamic camera clusters, which uses a next-best-view approach to determine a locally optimal configuration from which to collect measurements, in order to directly reduce the parameter uncertainty. Our proposed approach generates the next best view by solving a continuous optimization over the entropy of the parameter covariance matrix, and selecting the next measurement configuration that minimizes the expected parameter uncertainty. Our results demonstrate that the next-best-view approach is able to autonomously generate accurate calibrations, but with fewer measurement sets when compared to manual or sampling based viewpoint selection methods. Our future work will perform degeneracy and sensitivity analysis of the optimization, and test the approach on a wider range of actuated mechanisms for various applications.

References

[1] A. Das and S. L. Waslander, "Calibration of a dynamic camera cluster for multi-camera visual SLAM," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Daejeon, South Korea, October 2016.

[2] R. Bajcsy, "Active perception," in *Proceedings of IEEE*, vol. 76, no. 8, 1988.

[3] C. I. Connolly, "The determination of next best views," vol. 2, 1985.

[4] S. Frintrop and P. Jensfelt, "Attentional landmarks and active gaze control for visual SLAM," *IEEE Transactions on Robotics, special Issue on Visual SLAM*, vol. 24, no. 5, Oct. 2008.

[5] A. Richardson, J. Strom, and E. Olson, "AprilCal: Assisted and repeatable camera calibration," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, November 2013.

[6] Y. F. Z. Chichyang Chen, "A new robotic hand/eye calibration method by active viewing of a checkerboard pattern," in *IEEE Transactions on Robotics and Automation*. IEEE, 1993.

[7] M. D. Kevin Nickels, Eric Huber, "Hand-eye calibration using active vision," in *IEEE Aerospace Conference*. Pasadena, CA : Jet Propulsion Laboratory, National Aeronautics and Space Administration, 2007.

[8] S. Remy, M. Dhome, J. Lavest, and N. Daucher, "Hand-eye calibration," in *Proceedings of the 1997 IEEE/RSJ International Conference on Intelligent Robot and Systems. Innovative Robotics for Real-World Applications. IROS '97, September 7-11, 1997, Grenoble, France*, 1997, pp. 1057–1065.

[9] V. Pradeep, K. Konolige, and E. Berger, "Calibrating a multi-arm multi-sensor robot: A bundle adjustment approach," in *International Symposium on Experimental Robotics (ISER)*, 12/2010 2010.

[10] Y.-J. C. Jwu-Sheng Hu, "Automatic calibration of hand-eye-workspace and camera using hand-mounted line laser," vol. 18. IEEE, 2013, pp. 1778–1786.

[11] R. Y. Tsai and R. K. Lenz, "A new technique for fully autonomous and efficient 3d robotics hand/eye calibration," in *IEEE Transactions on Robotics and Automation*, vol. 5, no. 3, 1989, pp. 345–358.

[12] D. Bennett, D. Geiger, and J. Hollerbach, "Autonomous robot calibration for hand-eye coordination," *International Journal of Robotics Research*, vol. 10, no. 5, pp. 550–559, 10 1991.

[13] S. Wenhardt, B. Deutsch, J. Hornegger, H. Niemann, and J. Denzler, "An Information Theoretic Approach for Next Best View Planning in 3-D Reconstruction," in *The 18th International Conference on Pattern Recognition*, Y. Tang, S. Wang, G. Lorette, D. Yeung, and H. Yan, Eds., vol. 1, 2006, pp. 103–106.

[14] E. Dunn and J.-M. Frahm, "Next best view planning for active model improvement." in *BMVC*. British Machine Vision Association, 2009.

[15] F. S., A. G., and T. C., "Towards plant monitoring through next best view," Nov 2010.

[16] R. S. Hartenberg and J. Denavit, *Kinematic Synthesis of Linkages*. McGraw-Hill, 1964.

[17] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York, NY: Wiley-Interscience, 2006.

[18] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, Jun. 1981.

[19] S. Thompson, *Sampling*, ser. CourseSmart. Wiley, 2012. [Online]. Available: https://books.google.ca/books?id=-sFtXLIdDiIC

[20] A. Das and S. L. Waslander, "An entropy based approach to keyframe selection for multi-camera parallel tracking and mapping," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Hamburg, Germany, October 2015.