Calibration of a Dynamic Camera Cluster for Multi-Camera Visual SLAM

Arun Das* and Steven L. Waslander†

Abstract— **Multi-camera clusters used for visual SLAM assume a fixed calibration between the cameras, which places many limitations on its performance, and directly excludes all configurations where a camera in the cluster is mounted to a moving component. In this work, we present a calibration method for *dynamic* multi-camera clusters, where one or more of the cluster cameras is mounted to an actuated mechanism, such as a gimbal or robotic manipulator. Our calibration approach parametrizes the actuated mechanism using the Denavit-Hartenberg convention, then determines the calibration parameters which allow for the estimation of the time varying extrinsic transformations between camera frames. We validate our calibration approach using a dynamic camera cluster consisting of a static camera and a camera mounted to a pan-tilt unit, and demonstrate that the dynamic camera cluster can provide accurate tracking when used to perform SLAM.**

I. Introduction

The use of multi-camera clusters (MCCs) has emerged as a popular choice for performing robust and accurate visual SLAM. The ability of the MCC to take measurements over a wide field of view (FOV) improves camera localization robustness by better constraining the motion solution, and prevents feature starvation by consistently tracking features over longer durations and over large viewpoint changes. Furthermore, so long as the extrinsic calibration is known, multi-camera systems do not require overlap in the FOV to resolve the scale of the solution [1].

Although capable of performing accurate localization in a variety of environments, a major disadvantage of all multi-camera systems to date is that they require a fixed calibration between cameras to provide the solution at the correct scale, which places many limitations on MCC performance. First, any camera cluster must be re-calibrated if a new configuration is required. Second, since the MCC is fixed to the vehicle frame, the observation viewpoints of the cameras are highly dependent on the vehicle motion, which is especially problematic if the vehicle undergoes degenerate motions, or if the camera cluster observes areas of the environment where only poor feature measurements are possible. Finally, many systems, such as UAVs, cannot use the existing gimballed camera payload to assist with the visual navigation, as the extrinsic camera calibration for such systems is time varying.

In this work, we propose a calibration procedure which will enable the estimation of time varying extrinsic calibrations for multi-camera clusters, which can then easily be used in existing vision based tracking and SLAM systems. We formulate the calibration process for a *dynamic* multi-camera cluster, in which some or all of the cameras in the cluster are mounted to mechanisms which allow them to move independently of each other. Our calibration procedure then determines the parameters of the system, such that the transformation between cameras can be computed using only the control inputs to the mechanisms. We experimentally demonstrate our approach on a camera mounted to a pan-tilt unit, and show that a high quality calibration is achievable, and can be easily integrated into existing visual SLAM methods. The calibrated pan-tilt dynamic camera cluster is able to track the camera cluster motion in an indoor space approximately 15m×10m×15m in size, with an average accuracy of 2.3cm, which is comparable to the performance of a fixed cluster with the same number of cameras. This is, to the best of our knowledge, the first dynamic camera cluster system used for visual SLAM.

II. Related Works

Current camera-to-camera calibration approaches typically use fiducal markers to generate common observations between cameras [1], [2], [3], though unsupervised methods which use natural features in the environment from pre-existing maps or online SLAM solutions have also provided good results [4]. Although there has been significant work done in the area of camera to camera calibration, we have not found any existing results for camera to camera calibration through an actuated mechanism.

The hand-eye calibration problem, from the field of robotic manipulators, consists of computing the relative position and orientation between the motion frame of a mechanism, and a sensor which is rigidly mounted to the mechanism. The main focus for the hand-eye problem is simultaneously estimating the relative translation between a camera mounted to a robotic manipulator and the manipulator's end effector frame, as well as the transformation between the manipulator's base frame and the camera's motion base frame [5]. Although the hand-eye problem is similar to the dynamic MCC calibration problem, the hand-eye calibration assumes that the parameters of the mechanism's forward kinemetics (such as the DH parameters), are known, whereas our dynamic MCC calibration requires estimation of these parameters.

The class of calibration methods related to our problem is known as *kinematic calibration*, and seeks to refine the forward kinematic parameters of robotic manipulators in order to improve overall end effector positioning performance. Generally, the kinematic parameters are optimized by comparing the motion of the end effector to the predicted motion of the mechanism given the forward kinematic parameters and the joint angles. External measurement of the end effector can be collected using co-ordinate measurement machines (CMM)[6] and externally mounted theodolites [7],

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

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

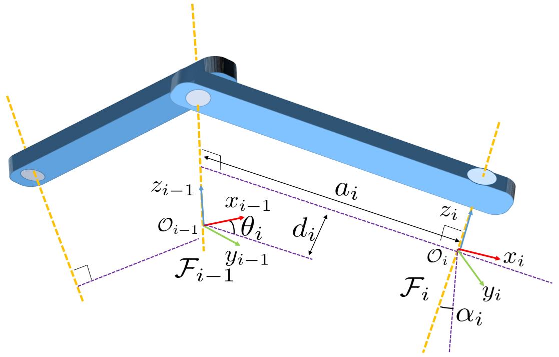

Fig. 1. Example of DH convention between two rotational joints

however the cost of such measurement equipment is typically quite high, which has motivated the use of low-cost camera based solutions for kinematic calibration.

Camera based kinematic calibration for manipulators consists of collecting relative measurements between a camera mounted on the manipulator and a fiducial target in the environment, or mounting the target on the manipulator and placing a static camera in the environment [8]. The work most similar to ours uses a laser pointer as a target and estimates the DH-parameters of two pan-tilt units with attached cameras, mounted to a manipulator end effector [9]. Although such approaches use the camera and fiducal marker to perform the calibration, the estimated parameters only describe the forward kinematics of the manipulator with respect to the *robot base*, whereas calibration of the dynamic MCC, for use in a SLAM problem, requires knowledge of the *camera to camera* calibration, which only include the mechanism's kinematic parameters as part of the total transformation between camera co-ordinate frames.

III. Background

General Rigid Body Transformation: Let a point in 3D be denoted as $\mathbf{p} \in \mathbb{R}^3$. In order to transform points between co-ordinate frames, we will define the rigid body transformation between frames a and b as $\mathbf{T}_\tau^{a:b} \in \mathbb{SE}(3)$, where $\mathbf{T}_\tau^{a:b} : \mathbb{R}^3 \mapsto \mathbb{R}^3$ and $\tau \in \mathbb{R}^6$ is a parameter vector which is used to construct $\mathbf{T}_\tau^{a:b}$. The parameter vector, $\tau = [r_x\, r_y\, r_z\, t_x\, t_y\, t_z]^T$ is composed of three rotation parameters, r_x, r_y, r_z, which denote the 3-2-1 Euler angle rotations, and three translation parameters, t_x, t_y, t_z, which denote the translations along the x, y, and z axes of frame a, respectively. We shall also define the function $\nu(\mathbf{T}) : \mathbb{SE}(3) \mapsto \mathbb{R}^6$, which computes the transformation parameters from the transformation matrix.

Denavit-Hartenburg Parameterization: The Denavit-Hartenburg (DH) convention is a widely used method to assign co-ordinate frames to the links of a robotic manipulator. Here, we will provide a brief overview of the DH approach for a serial manipulator with rotational joints [10], [11].

Suppose co-ordinate frame, $\mathcal{F}_i$, is attached to the i^{th} link of a robotic manipulator. The DH convention uses four independent parameters to define the transformation between adjacent links, as depicted in Figure 1.

Consider the two adjacent co-ordinate frames $\mathcal{F}_{i-1}$ and $\mathcal{F}_i$ from Figure 1. In order to construct co-ordinate frame i using the DH convention, the z axis of the frame is placed co-incident with the joint angle. Then, a common normal direction between z_{i-1} and z_i can be determined as $n_i = {}^{z_{i-1}\times z_i}\!/\!_{\|z_{i-1}\times z_i\|}$.

Using the common normal, the x_i axis is placed along n_i and points from z_{i-1} to z_i, and the intersection of the x_i and z_i axes define the origin, $\mathcal{O}_i$, of frame $\mathcal{F}_i$. With the x_i and z_i axes defined, the y_i axis is constructed on the frame according to the right-hand rule. Typically, frames are assigned in a sequential fashion, starting from the end effector frame and ending at the base frame of the mechanism.

With the i^{th} frame constructed, the transformation between $i-1^{\text{th}}$ and i^{th} frames can be performed using a four step sequence with the DH parameters: First, $\mathcal{F}_{i-1}$ is rotated about z_{i-1} by θ_i, then is translated along z_{i-1} by d_i, then is translated along x_i by a_i, and is finally rotated about x_i by α_i. After applying the transformations with the four parameters, frames $\mathcal{F}_{i-1}$ and $\mathcal{F}_i$ are co-incident.

In this work, we shall denote the DH parameters which describe the transformation between frames $i-1$ and i on an actuated mechanism as $\omega_i = [\theta_i, d_i, a_i, \alpha_i]^T \in \mathbb{R}^4$. Using the DH parameters, a homogeneous rigid body transformation, $\mathbf{T}_{\omega_i}^{i-1:i} \in \mathbb{SE}(3)$, can be computed [10], [11].

Image Projections of 3D Points: The projection function, which maps a point, $\mathbf{p}_i$, to a pixel location on the 2D image plane is defined as

$$\begin{aligned} \Psi(\mathbf{p}_i) &: \mathbb{R}^3 \mapsto \mathbb{P}^2 \\ \Psi(\mathbf{p}_i) &= [u_i\, v_i]^T, \end{aligned} \tag{1}$$

where u_i and v_i are the pixel co-ordinates of the projected point along the u and v image directions, respectively. In this work, we do not assume a specific type of camera model, however, it is important to consider the decrease in sensitivity of the measurement Jacobian for wide angle camera models such as the Taylor model [12]. In such cases, image measurements of the points seen near the boundary of the lens' field of view are less sensitive to perturbations of the point position in 3D, thus degrading the information quality required for precise localization of the camera [13]. In this work, the sensitivity issue can be mitigated by ensuring sufficient point measurements are collected over the entire image plane.

IV. Problem Formulation

This section will describe the calibration process for a dynamic MCC where some or all of the cameras are non-static. First, we will formulate the calibration process between a single static camera and a camera mounted to an actuated mechanism, which will be referred to as the *actuated camera*. Second, we will describe an extension of the static-to-actuated camera calibration case which will allow for calibration of the actuated-to-actuated camera case. Using the static-to-actuated and actuated-to-actuated camera calibration techniques, the extrinsics of any arbitrary dynamic camera cluster can be calibrated in a pair-wise fashion.

A. Static-to-Actuated Camera Calibration

The aim of the calibration process is to determine the rigid body transformation between the static camera frame, and the actuated camera frame, $\mathbf{T}^{s:m}_{\Theta,\lambda}$, where $\Theta \in \mathbb{R}^n$ is the set of *estimated* parameters which is used to build the rigid body transform, and $\lambda \in \mathbb{R}^d$ is the set of *measured* parameters which are used to build the transformation, available from either known inputs to the mechanism, or measured joint feedback. For the calibration, the transformation between cameras has the form $\mathbf{T}^{s:m}_{\Theta,\lambda} = \mathbf{T}^{e:m}_{\tau_m}\mathbf{T}^{b:e}_{\omega,\lambda}\mathbf{T}^{s:b}_{\tau_s}$, where $\mathbf{T}^{s:b}_{\tau_s}$ defines the transformation between the static camera and mechanism base frame, $\mathbf{T}^{b:e}_{\omega,\lambda}$ defines the transformation from the base frame of the mechanism to the end effector frame, and $\mathbf{T}^{e:m}_{\tau_m}$ defines the transformation from the end effector frame to the actuated camera frame. Note that $\mathbf{T}^{b:e}_{\omega,\lambda}$ 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.

In order to perform the calibration between the static camera and one of the actuated cameras, a fiducial marker is used to collect feature measurements in both cameras. Note that any marker, such as a chess board, is suitable, so long as the scale of the points can be determined. Measurements of the marker are taken from both cameras simultaneously, which requires that the two cameras share an overlapping field of view over a subset of the actuated mechanism's possible configurations. Using the measurements and known scale of the fiducial marker, it is possible for the observing camera to compute its 3D pose relative to the marker frame using a simple bundle adjustment approach [14]. Given the pose of the observing camera relative to the marker frame, we can determine the position of marker points relative to the camera frame.

For each instance where both the actuated and static camera capture measurements to the fiducial marker, we can define the i$^{\text{th}}$ *measurement set* as $Z_i = \{P^s_i, P^m_i, Q^s_i, Q^m_i, \lambda_i\}$, where P^s_i and P^m_i is the set of marker points defined in the frames of the static and actuated cameras, respectively, Q^s_i and Q^m_i is the set of measurements to the marker points, as observed by the static and actuated cameras, respectively, and λ_i is the set of joint inputs for the mechanism at snapshot i. Note that the measurement sets only include corresponding points visible in both cameras, so consequently $|P^s_i| = |P^m_i| = |Q^s_i| = |Q^m_i|$. In order to produce a high quality calibration, multiple measurement sets need to be collected, while ensuring sufficient excitation of the joint inputs by collecting measurements from many different configurations of the actuated camera.

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

$$e^m_j(\Theta, \lambda_i) = z^m_j - \Psi^m(\mathbf{T}^{s:m}_{\Theta,\lambda_i}\mathbf{p}^s_j) \tag{2}$$

where $z^m_j \in Q^m_i$ is the measurement of point j, from measurement set Q^m_i, observed in the actuated camera, and $\mathbf{p}^s_j \in P^s_i$ is the 3D position of point j, from the point set P^s_i, 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 $e^s_j(\Theta, \lambda_i) = z^s_j - \Psi^s((\mathbf{T}^{s:m}_{\Theta,\lambda_i})^{-1}\mathbf{p}^m_j)$, where $z^s_j \in Q^s_i$ is the measurement of point j, from measurement set Q^s_i, observed in the static camera, and $\mathbf{p}^m_j \in P^m_i$ is the 3D position of point j, from the point set P^m_i, as observed from the actuated camera. The total squared reprojection 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

$$\begin{aligned}\Lambda(\Theta) = \sum_{Z_i \in \Gamma} \sum_{j=1}^{|P^s_i|} & e^m_j(\Theta, \lambda_i)^T e^m_j(\Theta, \lambda_i) \\ & + e^s_j(\Theta, \lambda_i)^T e^s_j(\Theta, \lambda_i).\end{aligned} \tag{3}$$

Finally, to perform the calibration and determine the optimal parameters, Equation (3) is optimized in order to find the parameters which minimize the total reprojection error.

Since both the static and actuated cameras view the same fiducial marker, a constraint on estimated transformation, $\mathbf{T}^{s:m}_{\Theta,\lambda_i}$, is present as a result of a closed loop transformation chain. As illustrated in Figure 2, for each measurement set, we have that the transform, $\mathbf{T}^{f:f}_{\Theta,\lambda_i}$, which travels from the fiducial frame $\mathcal{F}_f$, through both camera frames, $\mathcal{F}_s$ and $\mathcal{F}_m$, and back to the fiducial frame, can be written as $\mathbf{T}^{f:f}_{\Theta,\lambda_i} = (\mathbf{T}^{f:m})^{-1}\mathbf{T}^{s:m}_{\Theta,\lambda_i}\mathbf{T}^{f:s}$. As the transformation is closed loop, we also require that $\mathbf{T}^{f:f}_{\Theta,\lambda_i} = \mathbf{I}$, where $\mathbf{I}$ is the identity matrix. Since the loop constraint is present for all collected measurement sets, we can now reformulate the problem as a constrained optimization over all measurement sets,

$$\begin{aligned} \min \quad & \Lambda(\Theta) \\ \text{subject to} \quad & \mathbf{T}^{f:f}_{\Theta,\lambda_i} = \mathbf{I} \quad \text{for } i = 1, \ldots |\Gamma|. \end{aligned} \tag{4}$$

The decision of using the constrained or unconstrained optimizations is dependent on the accuracy with which the camera-to-fiducial transforms, $\mathbf{T}^{f:s}$ and $\mathbf{T}^{f:m}$, can be computed, as the presence of noise in these transformations make it difficult to exactly enforce the constraint when solving (4).

B. Actuated-to-Actuated Camera Calibration

Compared to the static-to-actuated calibration, the calibration of an actuated-to-actuated camera pair requires the calculation of an additional transform between the base frames of the each mechanism, $\mathbf{T}^{b_1:b_2}$, as depicted in Figure 3. Suppose the camera pair consists of two cameras, Camera 1 and Camera 2. To determine the unknown transform, the camera pair is calibrated by first holding Camera 1 stationary using the static control input, $\bar{\lambda}_1$, and performing the static-to-actuated calibration by moving Camera 2, thereby estimating $\Theta_2 = [\tau_2, \gamma_2]^T$. The static-to-actuated calibration is performed again, except now holding Camera 2 stationary

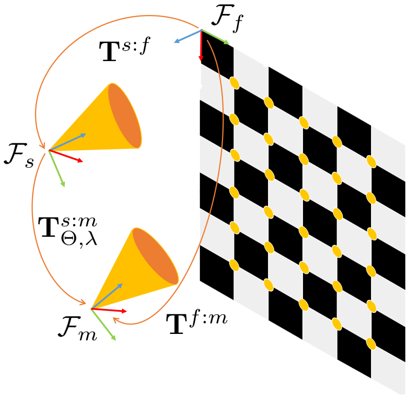

Fig. 2. Example transformation loop for a measurement set.

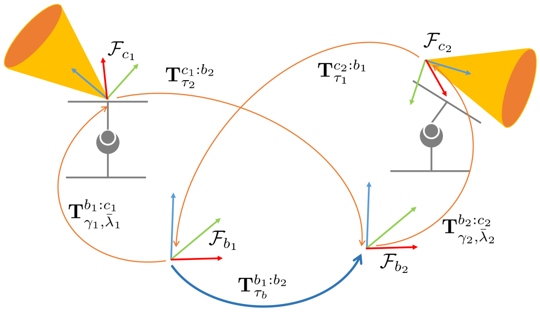
$\mathcal{F}_{c_1}$
$\mathbf{T}^{c_1:b_2}_{\tau_2}$
$\mathbf{T}^{c_2:b_1}_{\tau_1}$
$\mathcal{F}_{c_2}$
$\mathbf{T}^{b_1:c_1}_{\gamma_1,\bar{\lambda}_1}$
$\mathcal{F}_{b_1}$
$\mathbf{T}^{b_2:c_2}_{\gamma_2,\bar{\lambda}_2}$
$\mathcal{F}_{b_2}$
$\mathbf{T}^{b_1:b_2}_{\tau_b}$

Fig. 3. Frame transformations for actuated-to-actuated camera calibration case. Note the unknown transformation, $\mathbf{T}^{b_1:b_2}_{\tau_b}$, depicted with the blue arrow.

using the static control input, $\bar{\lambda}_2$, then moving Camera 1 and estimating $\Theta_1 = [\tau_1, \gamma_1]^T$.

As illustrated in Figure 3, we can now define two equivalent transformation loops using the estimated parameters from the static-to-dynamic calibrations and the static control inputs $\bar{\lambda}_1$ and $\bar{\lambda}_2$,

$$\begin{aligned}\mathbf{T}^{c_2:c_2}_{\tau_b} &= \mathbf{T}^{b_2:c_2}_{\gamma_2,\bar{\lambda}_2}\mathbf{T}^{b_1:b_2}_{\tau_b}\mathbf{T}^{c_2:b_1}_{\tau_1}\\ \mathbf{T}^{c_1:c_1}_{\tau_b} &= \mathbf{T}^{b_1:c_1}_{\gamma_1,\bar{\lambda}_1}(\mathbf{T}^{b_1:b_2}_{\tau_b})^{-1}\mathbf{T}^{c_1:b_2}_{\tau_2},\end{aligned} \tag{5}$$

where $\tau_b \in \mathbb{R}^6$ are the parameters describing the transformation between frames $\mathcal{F}_{b_1}$ and $\mathcal{F}_{b_2}$. We can now estimate the parameters of the unknown base to base transformation, τ_b, such that $\mathbf{T}^{c_2:c_2}_{\tau_b} = \mathbf{T}^{c_1:c_1}_{\tau_b} = \mathbf{I}$. Let us define the error vector for the transformation loop for Camera 1 and Camera 2 as $e_{c_1}(\tau_b) = \nu(\mathbf{T}^{c_1:c_1}_{\tau_b})$ and $e_{c_2}(\tau_b) = \nu(\mathbf{T}^{c_2:c_2}_{\tau_b})$, respectively. Then, the cost function, $\Omega(\tau_b) : \mathbb{R}^6 \mapsto \mathbb{R}$, which penalizes the error in the loop transformations from (5), is given as, $\Omega(\tau_b) = [e_{c_1}(\tau_b)\ e_{c_2}(\tau_b)][e_{c_1}(\tau_b)\ e_{c_2}(\tau_b)]^T$, which can be optimized using any unconstrained nonlinear method in order to find the optimal parameters for the unknown transformation, $\mathbf{T}^{b_1:b_2}_{\tau_b}$.

C. Estimated Parameter Reduction

Given that the joint angle, θ_i, for each joint in the mechanism can be estimated, or in many cases, directly measured using rotary encoder feedback, the number of estimated DH parameters required for the calibration of each link is 3. Thus, for the general camera-to-camera calibration problem through a mechanism with N joints, we will have N measured parameters and $12 + 3N$ estimated parameters. It should be noted, however, that some of the degrees of freedom are contained within the transformations from the static camera to the mechanism base, and from the mechanism's end effector to the moving camera frame. In the DH convention, the z axis of each joint's frame is along the axis of revolution for the joint, thus, without loss of generality, the mechanism base frame's z axis can be aligned with the axis of rotation of the first joint, and the origin of the base frame can be placed on the common normal defined by the frame of the first link. Such an approach means the link offset parameter, d_1, between the base frame and first link is zero, and thus no longer needs to be estimated.

For a mechanism with N joints, the N^{th} set of DH parameters will define the transformation between the N^{th} joint and the end effector frame. However, the physical placement of the end effector frame is not required, as we estimate an additional six degree of freedom transformation between the end effector frame and the camera frame. Thus, without loss of generality, the parameters d_N, a_N, and α_N can be set to zero and removed from the set of estimated parameters. Therefore the number of estimated parameters required is actually $8 + 3N$. Note that in the case of a single link there is only one link offset parameter which can be removed ($d_1 = d_N$), thus when $N = 1$, the number of estimation parameters required for the calibration is 12.

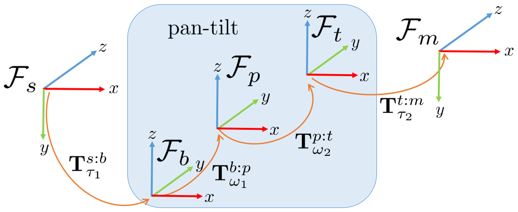

Fig. 4. Frame transformations for a general pan-tilt unit with two degrees of freedom.

V. Calibration of a Pan-Tilt Unit

We shall now apply the general static-to-actuated formulation described in Section IV-A to a dynamic MCC consisting of one static camera and one camera mounted to a pan-tilt unit with two degrees of freedom, with direct measurements of the joint angles available. Figure 4 depicts the co-ordinate frames and associated transforms for the dynamic MCC. Since measurements are available to the pan and tilt angle DH parameters of the mechanism, we shall define the measured parameters as $\lambda = [\theta_1, \theta_2]^T$. Using the parameter reduction method described in Section IV-C, the estimated parameters can be defined as $\Theta = [\tau_1, \hat{\omega}_1, \tau_2]^T \in \mathbb{R}^{14}$, where $\tau_1 \in \mathbb{R}^6$ are the six parameters for the transformation between Camera 1 and the pan-tilt base frame, $\hat{\omega}_1 = [a_1, \alpha_1]^T$ are the DH parameters for the pan-tilt unit, and $\tau_2 \in \mathbb{R}^6$ are the six parameters for the transformation between the pan-tilt end-effector frame and the frame of Camera 2.

VI. Experimental Validation

The proposed calibration approach is experimentally validated using a dynamic camera cluster consisting of two Ximea xIQ cameras, capturing 900×600 resolution images at 60 frames per second. The pan-tilt unit uses stepper motors to determine the joint angles, which are reported at a 30Hz update rate. Using the pan-tilt based dynamic MCC, we perform two sets of experiments. First, we collect 3D point data using a chessboard marker and show that it is possible to achieve a high quality calibration. Second, we integrate the calibrated cluster into the multi-camera parallel tracking and mapping (MCPTAM) [1] pipeline [15], [1], and demonstrate the effectiveness of the dynamic MCC in a SLAM scenario.

A. Calibration of the Pan-Tilt Unit

Existing marker based MCC calibration relies on relative motion between the marker and camera rig to collect 3D point and image measurement information from multiple viewpoints, whereas a dynamic MCC is able to observe the marker from different viewpoints by actuating the camera. To that end, we perform and compare two off-line calibrations. In the first, excitation of the image input is provided by moving both the marker and the actuated camera through various configurations (dual-excitation), and in the second, excitation of the image input is accomplished by only moving the camera (single-excitation) and keeping the target stationary. In both cases, approximately 130 measurement sets are collected with a chessboard target consisting of 35 points. Furthermore, we test both the dual-excitation and single-excitation cases using the unconstrained and constrained methods for calibration described in section IV-A.

Verification of the calibrations are performed with respect to an independently collected *validation set*, which consists of an additional 130 measurement sets, with excitation of the marker and pan and tilt angles. For the nonlinear optimization, the initial rotation parameters are approximated by sight, and the translation parameters are initialized to zero. Note that in practice, even with such a rough initialization of the estimated parameters, the optimization was able to converge to the correct solution in all of our trials. Summary statistics of the pixel reprojection errors are provided in Table I.

In general, all methods, with the exception of the constrained dual-excitation approach, produced good calibrations, as the mean pixel reprojection error was less than 1. However, the unconstrained dual-excitation method provided a calibration with the lowest error mean and covariance, due to the richness of the input images collected by moving both the marker and actuated camera. Note that the constrained dual-excitation approach produced a less accurate calibration compared to the unconstrained dual-excitation method, which can be attributed to noise or error in the fiducial to camera transforms which could be present from observing the marker with large skew angles, or near edges of the image where pixel sensitivity is reduced.

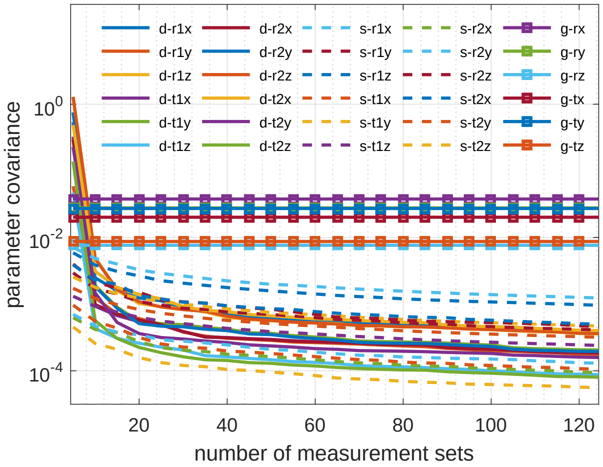

Fig. 5. Parameter uncertainty plotted as a function of the number of measurement sets used to determine the calibration. Entries with the prefixes ‘d-’, ‘s-’, and‘g-’denote the dual-excitation, single-excitation, and degraded-stereo cases, respectively.

The effectiveness of the single-excitation approach is further verified in Figure 5, which captures the uncertainty in the calibration parameters as a function of the measurement sets. We compare the unconstrained dual-excitation and constrained single-excitation cases, as they provided the best calibration for the dual-excitation and single-excitation cases, respectively, and a *degraded stereo* (d-stereo) configuration. For the degraded stereo, the dynamic MCC was held static in a forward facing configuration, and a six degree of freedom transformation between the camera frames was estimated using a single image of the fiducial target. For the single-excitation case, the degraded stereo provides a baseline to illustrate the improvement in calibration quality that is possible when one of the cameras is actuated.

To study how the calibration quality is affected by the number of measurement sets, the calibration method is performed repeatedly, first using 5 randomly selected measurement sets, then adding an additional 5 unique and randomly selected measurement sets to every subsequent optimization epoch. For each epoch, we estimate the solution covariance with $\Sigma = (J_e^T J_e + \sigma \operatorname{diag}(J_e^T J_e))^{-1}$, where J_e is the Jacobian matrix of the stacked error equations defined in (2), σ is a regularization factor, and $\operatorname{diag}(J_e^T J_e)$ defines the regulation term. Figure 5 displays the individual parameter uncertainty (the diagonal elements of Σ). Both the single- and dual-excitation cases have comparable improvements in the solution confidence as the number of measurements sets used increases, and the single-excitation case performs significantly better than the degraded stereo case. Although some of the individual parameters for the single-excitation case are not as well estimated when compared to the dual-excitation case, in practice the calibration may still perform well when implemented in the SLAM system.

[1] The MCPTAM code is available on GitHub: https://github.com/ wavelab/mcptam

TABLE I

Summary statistics for pixel reprojection of validation set

	mean (cam 1)	**mean (cam 2)**	**covariance (cam 1)**	**covariance (cam 2)**
uncon. dual	0.9713	0.9318	0.2076	0.2584
ucon. single	0.9818	1.0601	0.3028	0.2089
con. dual	1.5866	1.6418	0.3103	0.5069
con. single	0.9938	0.9622	0.2446	0.2605

TABLE II

Average error statistics for SLAM trajectory

	trans. err. [cm]	**rot. err. [deg]**
fixed calibration	2.341	0.6850
dual-excitation	2.503	0.7831
multi-excitation	2.419	0.7606
degraded-stereo	4.303	1.1748

B. Use of a Dynamic MCC in MCPTAM

At the time of each image acquisition, the extrinsic camera transformation is updated using the calibrated cluster parameters determined in Section VI-A, and the reported pan and tilt angles. Although we demonstrate SLAM results using the MCPTAM method, the dynamic MCC can be integrated into any multi-camera SLAM system using a similar approach.

For the SLAM experiment, we move the pan-tilt MCC through an indoor environment approximately 15m×10m×15m in size. We perform four separate motion trials, each approximately 150m in length, and compare the dual-excitation case, the single-excitation case, the d-stereo case, and a baseline fixed calibration (fixed-cal) performed using the extrinsic calibrator included with MCPTAM, with the pan-tilt cluster held static at zero pan and tilt angles. For each motion trial, the pan-tilt unit is moved around the indoor environment at walking speed, with the fixed camera maintaining view of the textured walls of the room. For the dual-excitation, and single-excitation cases, the pan and tilt joints are actuated in a periodic trajectory (± 45 degrees from the zero angle for pan, and to +40 degrees for tilt). Note that we did not produce joint trajectories with negative tilt angles, as the resulting camera viewpoint of the textureless floor would pose a difficulty in tracking not present in the fixed-cal or d-stereo cases. Ground truth of the motion is captured by an OptiTrack indoor positioning system (IPS). To compare the motion tracks, the SLAM and IPS motion solutions are aligned using an off-line calibration method [1].

Table II presents the median translation and rotation errors for the four tested cases. It is evident that the dual-excitation and single-excitation approaches achieve comparable performance to the fixed calibration case, while the degraded stereo case performs the worst. The slight increase in the error of the dynamic MCC cases compared to the fixed-cal case is due to high frequency vibrations from the tilt axis stepper motor causing image distortion of the actuated camera. Vibration issues can be mitigated through the use of brushless or direct drive servo motors in the actuated mechanism, as is commonly used on most UAV gimbals where vibration mitigation is important.

VII. Conclusion

This work presents the calibration of a dynamic MCC, which allows for the use of time varying camera-to-camera extrinsic transformations in multi-camera visual SLAM. The unknown parameters of the actuated mechanism are parameterized using the DH convention, and calibrated using a fiducial marker of known scale. We experimentally demonstrate the validity of the calibration on a pan-tilt based dynamic cluster using the MCPTAM method. Our future work includes performing degeneracy and sensitivity analysis for the general dynamic camera cluster, testing our approach on a wider class of complex mechanisms, integrating dynamic MCCs into other existing SLAM methods, and performing active gaze selection on the actuated camera to help improve localization accuracy.

References

[1] A. Tribou, Michael J. and. Harmat, D. Wang, I. Sharf, and S. L. Waslander, "Multi-camera parallel tracking and mapping with non-overlapping fields of view," *International Journal of Robotics Research*, vol. 34, no. 12, pp. 1480–1500, December 2015.

[2] P. Lbraly, E. Royer, O. Ait-Aider, C. Deymier, and M. Dhome, "Fast calibration of embedded non-overlapping cameras," in *IEEE International Conference on Robotics and Automation (ICRA)*, Shanghai, China, May 2011, pp. 221–227.

[3] B. Li, L. Heng, K. Koser, and M. Pollefeys, "A multiple-camera system calibration toolbox using a feature descriptor-based calibration pattern," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Tokyo, Japan, Nov 2013, pp. 1301–1307.

[4] L. Heng, M. Burki, G. H. Lee, P. Furgale, R. Siegwart, and M. Pollefeys, "Infrastructure-based calibration of a multi-camera rig," in *2014 IEEE International Conference on Robotics and Automation (ICRA)*, Hong Kong, China, 2014, pp. 4912–4919.

[5] K. Daniilidis, "Hand-eye calibration using dual quaternions," *The International Journal of Robotics Research*, vol. 18, no. 3, pp. 286–298, 1999.

[6] M.-S. Kim, H.-S. Yoo, S.-W. Cho, H.-S. Chang, and G. Spur, "A new calibration method," *CIRP Annals-Manufacturing Technology*, vol. 39, no. 1, pp. 421–424, 1990.

[7] D. Whitney, C. Lozinski, and J. M. Rourke, "Industrial robot forward calibration method and results," *Journal of dynamic systems, measurement, and control*, vol. 108, no. 1, pp. 1–8, 1986.

[8] D. C.-C. Lu, "Kinematic calibration of serial manipulators using relative measurements," Master's thesis, Ottawa-Carleton Institute for Mechanical and Aerospace Engineering Department, Jan 2014.

[9] C. S. Gatla, R. Lumia, J. Wood, and G. Starr, "Calibrating pan-tilt cameras in robot hand-eye systems using a single point," in *IEEE International Conference on Robotics and Automation (ICRA)*, Roma, Italy, April 2007, pp. 3186–3191.

[10] M. W. Spong and M. Vidyasagar, *Robot dynamics and control*. John Wiley & Sons, 2008.

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

[12] D. Scaramuzza, A. Martinelli, and R. Siegwart, "A flexible technique for accurate omnidirectional camera calibration and structure from motion," in *IEEE International Conference on Computer Vision Systems (ICVS)*. New York, NY: IEEE, Jan. 2006, pp. 45–45.

[13] 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.

[14] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*. Cambridge University Press, 2003.

[15] A. Harmat, M. Trentini, and I. Sharf, "Multi-camera tracking and mapping for unmanned aerial vehicles in unstructured environments," *Journal of Intelligent & Robotic Systems*, vol. 78, no. 2, pp. 291–317, 2014.