Multi-Capture Dynamic Calibration of Multi-Camera Systems

Avinash Kumar
Intel Labs

Manjula Gururaj
Intel Labs

Kalpana Seshadrinathan
Intel Labs

Ramkumar Narayanswamy

Abstract

Multi-camera systems have seen an emergence in various consumer devices enabling many applications e.g. bokeh (Apple IPhone), 3D measurement (Dell Venue 8) etc. An accurately calibrated multi-camera system is essential for proper functioning of these applications. Usually, a onetime factory calibration with technical targets is done to accurately calibrate such systems. Although accurate, factory calibration does not hold over the life time of the device as normal wear and tear, thermal effects, device usage etc. can cause calibration parameters to change. Thus, a dynamic or self-calibration based on multi-view image features is required to refine calibration parameters. One of the important factors governing the accuracy of dynamic calibration is the number and distribution of feature points in the captured scene. A dense feature distribution enables better sampling of the 3D scene, while avoiding degenerate situations (e.g. all features on one plane), thus sufficiently modeling the forward imaging process for calibration. But, single real life images with dense feature distribution are difficult or nearly impossible to capture e.g. texture-less indoor or occluded scenes.

In this paper, we propose a new multi-capture paradigm for multi-camera dynamic calibration where multiple multi-view images of different 3D scenes (thus varying feature point distribution) are jointly used to calibrate the multi-camera system. We present a new optimality criteria to select the best set of candidate images from a pool of multi-view images, along with their order, to use for multi-capture dynamic calibration. We also propose a methodology to jointly model calibration parameters of multiple multi-view images. Finally, we show improved performance of multi-capture dynamic calibration over single-capture dynamic calibration in terms of lower epipolar rectification and 3D measurement error.

1. Introduction

The past few years have seen an emergence of multi-camera system based devices, e.g. Dell Venue 8 7000 (3 cameras), IPhone (2 cameras), Facebook 360 (14 cameras) to enable various computational photography applications for consumer use. An accurately calibrated multi-camera system is essential for proper functioning of these applications. Multi-camera calibration entails estimating intrinsic parameters like focal length and principal point of individual cameras and the extrinsic parameters of relative rotation and translation between all pairs of cameras. These parameters can be used to accurately compute metric 3D reconstruction of the imaged scene. This is a key component driving many of the computational photography applications e.g. 3D measurement, depth based blurring/bokeh. An out-of-calibration camera can result in inaccurate 3D reconstruction and thus affect the performance of many of these applications. Thus, being able to calibrate multiple view camera systems accurately is essential.

While the current industry practice is to have a one-time calibration done in the factory floor as part of the device manufacturing process, a pre-calibrated device is bound to go out of calibration over time due to various factors like heat, mechanical stress, moving auto-focus lens etc. These effects render a one-time calibration inadequate over the life-time of the device. Thus there is a need for methods to re-estimate the calibration parameters which adapt to these changes. The traditional technical target based methods for calibration [6, 18] are not practical at the consumer end due to requirement of buying accurate technical targets and thereafter collecting calibration data. A more convenient method is to have a **dynamic/self calibration** method which can use multi-view images of natural scenes as input to calibrate the multi-camera to its most recent geometric configuration. Henceforth, a single capture of multi-view images from a multi-camera system will be denoted as an *image frame set*.

Typically, high accuracy technical target calibration requires densely sampling the camera's field of view in captured calibration images. This is because some of the parameters like image distortion are dependent on calibration features being present at the image corners and at varying scene depths [13]. Extrapolating this observation to single image frame set based dynamic calibration means that an ideal natural scene for dynamic calibration should be the one with feature-points densely distributed. However, capturing such ideal scenes is challenging and may require multiple attempts on the part of the user to get the best image frame set. In-fact, occlusion in scenes will hide objects in the back, thus never allowing to capture a scene with uniformly distributed dense set of features.

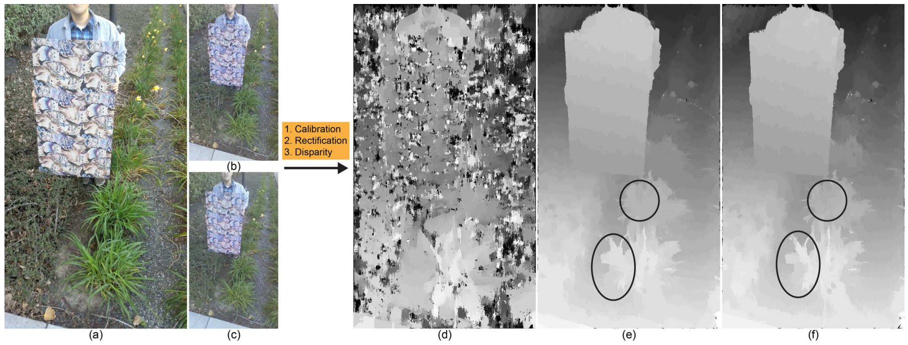

Figure 1. (a, b, c) Three-view images with horizontal (b, c) and vertical parallax (a w.r.t b and c) parallax, (d, e, f) Disparity estimate using factory calibration, single-capture and multi-capture dynamic calibration parameters. The disparity estimates are much smoother using dynamic calibration parameters. The dark circles show regions where multi-capture dynamic calibration resulted in better resolution of disparity as compared to single-capture.

Our innovation in this paper lies in removing the constraint a single "ideal" image frame set altogether, thus making dynamic calibration much more adaptable for the user. We achieve this relaxation by allowing dynamic calibration to use cumulative feature points from multiple image frame sets of completely different scenes. The accruing of feature points from different scenes allows to create a dense distribution of feature points which can be jointly used for dynamic calibration. This paradigm of **multi-capture multi-camera dynamic calibration** is non-trivial due to two major reasons. **First**, multiple image frame sets are independently captured over an extended period of time. Thus, the calibration parameters associated with them may be different. How do we jointly model the different calibration parameters? and **Second**, given a pool of candidate image frame sets, not all of them would be equally effective in modeling a single-capture "ideal" image frame set. What is the best set of image frame sets to select? How do we select them without exhaustive search? In what order should they be selected?

In this paper, we answer all of these questions and in doing so present a complete framework for optimal multi-capture multi-camera dynamic calibration and its benefits. **In summary, our main contributions in this paper are:**

1. We propose a new framework of multi-capture multi-camera dynamic calibration.
2. We propose a method to share the parameters among the multiple image frame sets allowing us to leverage the benefit of accrued feature points from these images (Sec. 4.1).
3. We propose an optimality criteria, which takes feature distribution as one of the factors, to select the best set of image frame sets and their sequence to use for multi-capture dynamic calibration. We show accuracy in calibration parameter estimates which are comparable to those obtained after time-inefficeint exhaustive greedy search for the best images (Sec. 4.2).

Fig. 1 is a visual representation of our results. The images shown in Fig. 1(a,b,c) are a set of three-view images. Fig. 1(d) shows the disparity obtained using factory calibration. As can be seen, the device has lost its factory calibration settings resulting in incorrect rectified images and thus noisy disparity map. Fig. 1(e, f) are the disparities obtained after single and multi-capture dynamic calibration. As can be seen, both single and multi-capture dynamic calibration are able to restore calibration leading to better disparity estimates.

Sec. 2 presents related work. Sec. 3 gives an overview of why dynamic calibration is required. Sec. 4 presents our proposed framework of multi-capture multi-camera dynamic calibration. Finally, Sec. 5 presents quantitative and qualitative results on real images to show the performance of the presented framework.

2. Previous Work

Dynamic calibration of multi-camera systems has been a well researched topic in computer vision mainly due to its similarity to techniques of simultaneous localization and mapping (SLAM) and structure from motion (SfM). Most of the previous work in this area (feature based calibration) can broadly be divided into three classes:

Class 1: Single camera with sequence of images: If the intrinsic parameters are known with high accuracy, then the problem of extrinsic estimation maps to the problem of SLAM. If the intrinsic parameters are constant but unknown, then a number of methods exist to do a linear estimation of intrinsic parameters under relaxed constraints and then jointly refine the intrinsic and extrinsic parameters [17, 9].

Class 2: Multi-camera with single image: If the multi-camera system is homogeneous, i.e. all the cameras have same intrinsics, then intrinsic calibration can be solved by methods of Case 1 by treating multiple views as views from a moving camera, followed by extrinsic estimation based on computed intrinsic parameters and E-matrix computation [16]. Of-course these methods assume undistorted images. If the distortion was unknown, then under the assumption of pure radialdistortion Fitzgibbon [10] have proposed method to compute radial fundamental matrix and the distortion parameters. For heterogeneous cameras with varying intrinsic and distortion parameters, this problem was solved by Barreto [4].

Class 3: Multi-camera with sequence of images of the same scene: This case has recently become popular in robotics and autonomous driving systems. Assuming known intrinsics, Carerra [8] proposed a feature-based non-overlapping extrinsic-only calibration method based on visual SLAM algorithm upto scale. Heng [12] proposes to do a metric estimation by adding a calibrated stereo pair to their multi-camera system. They also assume that the intrinsic parameters are accurately known based on target calibration. Similar methods exist for multi-cameras in autonomous driving [11, 14].

In this paper, we consider heterogeneous multi-camera systems, thus methods in Class 1 are inapplicable to our work. The methods in Class 2 and Class 3 are mostly based on SLAM and SfM in some form, causing their accuracy to be heavily dependent on the quality and distribution of feature points in the imaged scene. Also, most methods in Class 1 and Class 3 consider constant calibration parameters for captured frames possibly to enable them to relate tracked features across frames. But it may not always hold for longer tracked image frames.

Comparatively, our proposed method of utilizing images of completely different scenes for dynamic calibration while also not assuming that the calibration parameters are constant doesn't fall in any of the above categories. This motivates us to propose a new class of dynamic calibration methods: **Class 4: Multi-camera with sequence of completely different scenes**

3. Need for Dynamic Calibration

A factory calibrated multi-camera system can go out of calibration due to external factors e.g.

1. Thermal heat generated due to continuous use can cause the camera module components to temporarily expand leading to physical change in camera focal length or CMOS/CCD sensor expansion causing captured images to not to confirm to factory calibration.

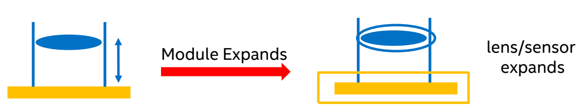

Figure 2. Thermal heat changes individual camera geometry.

2. Mechanical stress generated due to everyday use e.g. fall and transportation can cause the printed circuit board (PCB) connecting cameras to bend. It can modify relative pose between the cameras.

3. Camera module non-rigidity can cause the camera optics to change temporarily, e.g. tilting the device downward can cause the non-rigid auto-focus lens to move due to gravity. This can cause additional magnification in captured images.

Figure 3. Mechanical stress modifies relative camera pose.

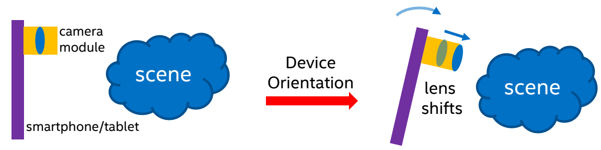

Figure 4. Camera non-rigidity causes movement of components leading to change in calibration.

The following methods could be employed to re-calibrate the system:

1. Factory-like calibration routine at regular intervals (accurate but expensive and cumbersome) but buying a specific technical target could be expensive and extensive data capture requirement makes it impractical.

2. Send back to manufacturer (accurate but not scalable) who re-calibrates the camera or replaces partial camera modules but its not scalable.

3. Build mechanically robust systems (accurate but expensive) which are verse to the effects mentioned above. But, designing them may be expensive due to specific requirements on material types, module designs and robust housing of the modules.

4. Dynamic calibration (accurate, scalable, inexpensive) which requires everyday images as the sole input. The proposed multi-capture approach will also utilize the existing pre-captured images from the image library. Sec. 4 presents the details of the proposed multi-capture dynamic calibration method.

4. Multi-Capture Dynamic Calibration

We define a reference image frame set as the one for which the best calibration parameters are required to be computed using single or multi-capture dynamic calibration. We also assume that a *library* of previously captured image frame set exists from which the images for multi-capture dynamic calibration will be selected. Sec. 4.1 presents a method to jointly model calibration parameters corresponding to different image sets. Sec. 4.2 presents our optimality measure to select the best image frame set and its sequence from the pool.

4.1. Joint Calibration Parameter Modeling

Each of the multiple image frame sets in our library images a different scene, thereby capturing different feature point distributions, at different time instants. They also have their own unique capture setting of intrinsic and extrinsic parameters. As there is no shared information (parameters+3D points) between different image frame sets, a simple joint calibration of multiple image frame sets will only be equivalent to performing individual single-capture dynamic calibration on each image frame set. Thus, in order to truly leverage the advantage of multiple image frame sets in the form of accrued 2D feature points, we propose to relax the assumption of uniqueness of calibration parameters in each capture and assume that some of them are sharable by all the multiple image frame sets. For example, the 3-camera Dell Venue 8 tablet (Fig. 8) used in all of our experiments has an auto-focus camera with moving lens module. Gravity effects on the lens module as the orientation of the tablet changes while capturing data can cause intrinsic parameters of the camera to be unique. But, the relative pose of the 3-cameras do not tend to vary among different captures.

Thus, we divide the different calibration parameter types (intrinsic + extrinsic) into two sets: **(1) low-frequency** parameters which are expected to remain same for different image frame sets e.g. pose between cameras. **(2) high-frequency** parameters which are likely to vary for each image frame set e.g. focal length of auto-focus camera. As part of multi-capture dynamic calibration implementation, this classification is used to optimize only one instance of low-frequency parameter type for all image frame sets, while high-frequency parameter types have their own instances in the optimization corresponding to each of the multiple image frame sets.

4.2. Optimal Selection Criteria

There are two methods to sample the pool of candidate image frame sets for multi-capture dynamic calibration:

Greedy: This is an exhaustive search approach where all images from the pool are sequentially selected and jointly used with the reference image frame set for multi-capture dynamic calibration. If the obtained calibration parameters result in reducing some measure of accuracy e.g. mean rectification error on reference image frame set, then that image frame set is selected. The above procedure is then repeated for all the remaining images in the pool, each time adding a new one. Although accurate, this method has very high runtime-complexity requiring $\mathtt{O}(\mathtt{N}^2)$ multi-capture dynamic calibrations for a pool size of $\mathtt{N}$ image frame sets.

Optimality Criteria: This is our second key innovation where we design an optimality criteria to find the best image frame set to add to current multi-capture image frame set without an exhaustive greedy search, thus leading to faster runtime of $\mathtt{O}(\mathtt{N})$ multi-capture dynamic calibrations. As explained in Sec. 1, our goal is to give priority to images of feature rich scenes with uniform feature distribution e.g. dense vegetation; followed by images which are partially feature-rich, e.g. urban scenes with have no/less features in the sky; and finally sparse feature scenes e.g. indoors. In order to quantify this metric on a given image frame set, one of its component images is selected. This image is divided into 2D bins of size $\mathtt{b} * \mathtt{b}$ ($\mathtt{b} = 5$ in this paper). Also, for all feature points in the image, their z-depth after single capture dynamic calibration (Fig. 6) is recorded. The z-depth is assumed to be divided into $\mathtt{d}$ blocks. In this work $\mathtt{d} = 3$ with ranges: $[\mathtt{0m} - \mathtt{4m}), [\mathtt{4m} - \mathtt{8m}), [\mathtt{8m} - \infty)$. Thus, the scene being imaged is divided into $\mathtt{B} = \mathtt{b} * \mathtt{b} * \mathtt{d}$ bins and each feature point with pixel location $(\mathtt{I}, \mathtt{J})$ and depth $\mathtt{z}$ is assigned to one of the $\mathtt{B}$ bins. An example histogram of this 3D distribution on two scenes is shown in Fig. 5. The

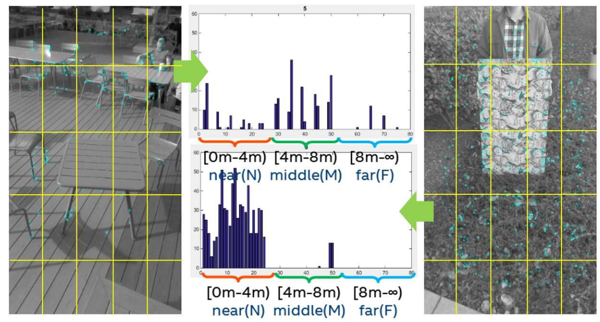

Figure 5. Feature distribution histogram (center top) for left and (center bottom) for right image.

left image has features across a wide range of 3D depths but each depth has less features. Right image has dense features but those are located only in the nearby depth range of $[\mathtt{0m} - \mathtt{4m})$. Thus, each image have their own feature distribution. A combined histogram of a collection of images will tend to have dense and uniform distribution.

Based on the above analysis, we propose our optimality criteria in Eq. 1. It has three components. The first component counts the number of feature points in all the candidate images. The second component takes into account the mean rectification error in each of the candidate image frame sets computed using their own single-capture dynamic calibration parameters. The mean rectification error is a measure of the goodness of single-capture calibration for that particular candidate image frame set favoring lower mean rectification error. Lastly, the third component takes into account the 3D feature point distribution in the scene with an even distribution being favored as compared to a skewed spatial distribution of points. Thus, optimality measure $\mathtt{M}$ selects the best image frame set $\mathtt{I}_{\mathtt{opt}}$ from a candidate pool $\mathtt{I}_{\mathtt{i}}$ and adds to the current set of multi-capture image frame set $\mathtt{S}$:

$$\begin{aligned}\mathtt{I}_{\mathtt{opt}} = \underset{\mathtt{I}_\mathtt{i} \in \{\mathtt{I}_1, \ldots, \mathtt{I}_\mathtt{n}\} \notin \mathtt{S}}{\arg\max} &\Big[\alpha * \#\mathtt{keypoints}(\mathtt{I}_\mathtt{i}) + \beta * \frac{1}{\mathtt{E}_\mathtt{i}(\mathtt{I}_\mathtt{i})} \\ &+ \gamma * \mathtt{stddev}((\mathtt{h}(\mathtt{S}) + \mathtt{h}(\mathtt{I}_\mathtt{i})) > 0)\Big] \quad (1)\end{aligned}$$

Here, $\mathtt{E}_\mathtt{i}(\mathtt{I}_\mathtt{i})$ denotes the rectification error for image frame set $\mathtt{I}_\mathtt{i}$ using calibration done on image frame set $\mathtt{I}_\mathtt{i}$. $\mathtt{h}(\mathtt{S})$ and $\mathtt{h}(\mathtt{I}_\mathtt{i})$ denote the combined and individual 3D feature point histogram for $\mathtt{S}$ and $\mathtt{I}_\mathtt{i}$ respectively. The function $\mathtt{stddev}()$ computes the standard deviation of the index of the location of non-zero entries in the histograms. The parameters (α, β, γ) are weights which are obtained through regression on a test set of images, where the best sequence of images to add is based on the greedy approach.

4.3. Multi-Capture Dynamic Calibration Algorithm

In this section, we present the multi-capture dynamic calibration algorithm. Fig. 6 presents the single-capture dynamic calibration algorithm. It primarily consists of standard techniques from SfM, namely: (1) multi-view image acquisition; (2) feature detection (AKAZE [3]); (3) feature matching (2-nearest neighbor, ratio-test [15], symmetry-test, fundamental matrix based RANSAC to remove outliers, all-pair feature matching); (4) 3D triangulation based on factory intrinsics and pose from 5-point estimation [16]; (5) bundle-adjustment (CERES [2]); (6) guided feature matching [5] to increase feature matches along epipolar lines; (7) validation metrics (epipolar rectification error, 3D measurement error, disparity).

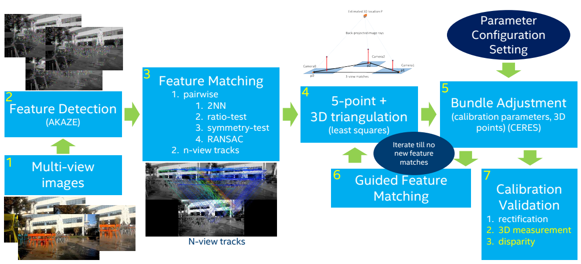

Figure 6. Flowchart of single-capture dynamic calibration (DynCal).

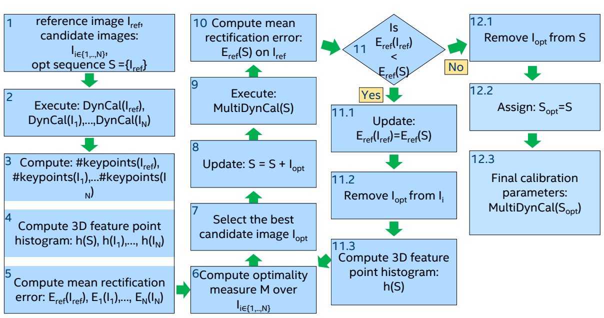

Figure 7. Flowchart of multi-capture dynamic calibration (MultiDynCal).

Fig 7 presents our proposed multi-capture dynamic calibration pipeline which includes single-capture dynamic calibration as a block. The details of each of the blocks are explained below. We denote the reference image as $\mathtt{I_{ref}}$, which may be the most current captured image frame set, as the data for which the dynamic calibration parameters are needed to be estimated. We denote the pool of $\mathtt{N}$ candidate image frame sets as $\mathtt{I_{i\in\{1,\ldots,N\}}}$.

1. Set $\mathtt{I_{ref}}$ as reference image and $\mathtt{I_{i\in\{1,\ldots,N\}}}$ as candidate image frame sets. Let the current optimal sequence be a singleton set $\mathtt{S} = \mathtt{I_{ref}}$.

2. Execute single-capture dynamic calibration on each candidate image set $\mathtt{I_{i\in\{1,\ldots,N\}}}$ as well as $\mathtt{I_{ref}}$ to get the base performance metric of mean rectification error $\mathtt{E_{ref}(I_{ref})}$.

3. Store the number of keypoints used for single-capture dynamic calibration for $\mathtt{I_{i\in\{1,\ldots,N\}}}$.

4. Compute the 3D feature point histogram $\mathtt{h(S), h(I_1), \ldots, h(I_N)}$ [Sec. 4.2]

5. Compute mean rectification error $\mathtt{E_{ref}(I_{ref}), E_{I_1}(I_1), \ldots, E_{I_N}(I_N)}$ on each image frame-set from their own single-capture based (Step 2) dynamic calibrated parameters.

6. Compute optimality measure $\mathtt{M}$ for each candidate image frame set in $\mathtt{I_{i\in\{1,\ldots,N\}}}$.

7. Select the image with highest measure $\mathtt{M}$ and assign it as $\mathtt{I_{opt}}$ and the best image to add to sequence $\mathtt{S}$.

8. Update the optimal sequence $\mathtt{S} = \mathtt{S} + \mathtt{I_{opt}}$.

9. Execute multi-capture dynamic calibration $\mathtt{MultiDynCal(S)}$ on multi-capture sequence $\mathtt{S}$ based on parameter modeling of Sec. 4.1, where only one instance of low-frequency parameters is optimized while high-frequency parameters have their own instances for all the image frame sets.

10. Compute mean rectification error $\mathtt{E_ref}$ on reference image set $\mathtt{I_{ref}}$ using calibration parameters from $\mathtt{MultiDynCal(S)}$.

11. If the mean rectification error on $\mathtt{I_{ref}}$ reduces then:

    11.1. Update reference rectification error to correspond to one obtained from $\mathtt{DynCal(S)}$.

    11.2. Remove $\mathtt{I_{opt}}$ from candidate set $\mathtt{I_i}$.

    11.3. Compute joint 3D feature point distribution histogram of accumulated feature points in $\mathtt{S}$ and goto Step 6.

12. Else adding $\mathtt{I_{opt}}$ didnt help in reducing mean rectification error:

    12.1. Remove $\mathtt{I_{opt}}$ from optimal sequence $\mathtt{S}$.

    12.2. Assign optimal sequence $\mathtt{S}$ as $\mathtt{S_{opt}}$.

    12.3. Output the calibration parameters obtained from $\mathtt{DynCal(S)}$ in Step 9 as final result.

5. Results and Comparison

5.1. Data Collection

For all results in this paper, a three-camera system commercially knowns as Dell Venue 8 7000 [1] is used to capture three-view images of a scene. The multi-camera system consists of one 8MP auto-focus camera and two 2MP fixed-focus cameras arranged in a triangular orientation. The baseline between the 8MP and each of the 2MP cameras is $\mathtt{43mm}$, while the baseline between the 2MP cameras is $\mathtt{72mm}$. See Fig. 8 for details on the multi-camera configuration.

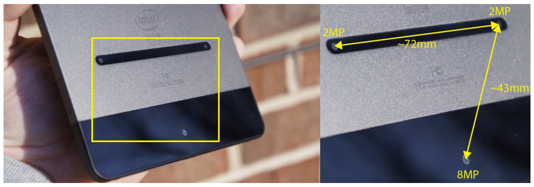

Figure 8. (left) Multi-capture data capture device: Dell Venue 8 7000 [1]. (right) Detailed view of the three-camera system.

5.2. Test Data Design

A set of 3 Dell Venue 8 7000 devices were used to capture 42 scenes each, resulting in a total of 126 scenes or reference image sets. Each of these scenes was constrained to be captured under the following four variations, thus resulting in a wide variability in our test data-set:

1. Four lighting conditions (measured with light meter): 150 lux (indoor hallway), 400 lux (indoors with fluorescent lighting), 1000 lux (outdoors evening) and 5000 lux (outdoors with clear sky). See Fig. 9.

Figure 9. Light variation and 3D measurement target distance: (a) 150 lux,5 m (b) 400 lux,3 m (c) 1000 lux,7 m (d) 5000 lux,3 m in reference scenes.

2. Three depths of 3m, 5m and 7m at which a textured planar board was placed along with a subject for 3D measurement validation.
3. Two orientation modes: landscape and portrait.
4. Three variations in feature point density: low-level (indoors), mid-level (outdoors with featureless walkway) and high-level (dense vegetation). See Fig. 10.

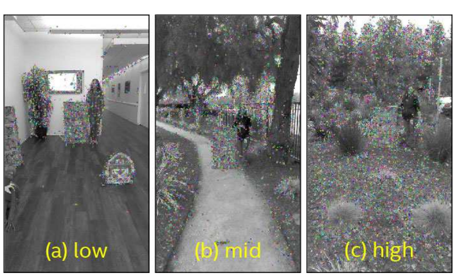

Figure 10. (a) low feature distribution (b) mid-level feature distribution (c) high feature distribution in reference scenes.

Figure 11. Multi-capture dynamic calibration candidate scenes.

All the 126 (42 sets *3 devices) image frame sets captured above will each be used as a reference image frame set. The technique of multi-capture dynamic calibration requires an additional pool of candidate scenes (image frame sets). A total of 27 such scenes were captured using each of the same 3 devices used to capture the reference image frame sets. Each of the 27 scenes were captured under the same variations of feature point density, light and orientation as mentioned above. Thus, in total, 81 scenes were captured from 3 devices. An example set of these candidate images is shown in Fig. 11.

5.3. Performance of Optimality Criteria

In this section, the performance of our optimality criteria `M` (Eq. 1) is analyzed as compared to an exhaustive greedy approach in for image frame set selection with respect to mean rectification error over 8 different reference image frame sets. The pool of candidate image frame sets is kept same for both the methods. Fig. 12 shows our results with each of the 8 plots corresponding to 8 different image frame sets. In each plot, the mean rectification error is plotted against the current size of the multi-capture image frame set (`S` in Fig. 7) used for doing multi-capture dynamic calibration and generating the calibration parameters for rectification error. As can be seen, the mean rectification error obtained from selections based on optimality criteria are quite close to those obtained using exhaustive greedy approach and also have better time complexity.

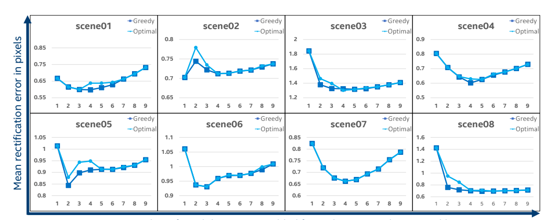

Figure 12. Comparison of mean-rectification error (along Y-axis) on 8 different reference image sets computed from multi-capture dynamic calibration method using greedy (deep blue with square markers) and optimal approach (light blue with round small markers) on a candidate pool of size 9 (best viewed in color).

5.4. Mean Rectification Error

The intrinsic and extrinsic calibration parameters can be used rectify pair-wise images in a multi-camera system. An accurately rectified image will typically have parallel epipolar lines in the rectified images. Incorrect calibration can result in shifting of epipolar lines, thereby resulting in a vertical error (in pixel units) between an epiploar line and feature point correspondence as shown in Fig. 13. This can be used as a metric to evaluate calibration as it is independent of pixel-reprojection error which is the criteria used to optimize in dynamic calibration. The average of all pair-wise rectification errors is called as the mean rectification error. Fig. 14 shows the mean rectification error performance with respect to four methods: factory, single-capture, optimal and greedy multi-capture for 12 randomly selected reference image sets. An error of < 1 pxiels is desirable as disparity estimation methods can handle that. Both approaches of multi-capture dynamic calibration perform better than single-capture and factory, thus showing their efficacy. Since greedy approach is exhaustive, it performs the best but the optimal criteria approach is also close.

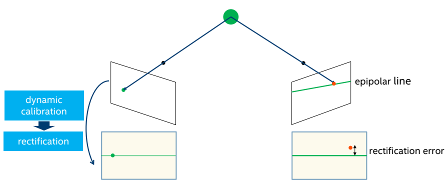

Figure 13. Geometry of rectification error for stereo cameras.

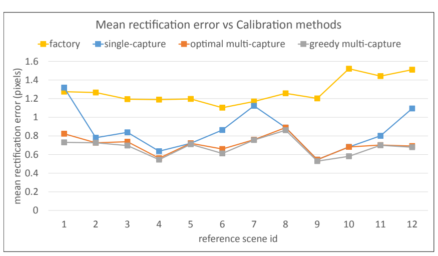

Figure 14. Mean rectification error (< 1 pixel is desired).

5.5. 3D Measurement Error

In this section, we compute the percentage accuracy of 3D measurement using calibration parameters computed from different calibration methods: factory, single-capture and multi-capture under variations of different factors of feature density, light and the distance of the measurement object from the system. The calibration parameters are first used to rectify all the cameras in a multi-camera system to a reference camera, followed by pairwise disparity computation [7], merging of different disparity estimates and finally 3D measurement on the merged disparity map. Thus, any error in calibration propagates to final disparity as well as in converting disparity to metric units. In all our reference image frame sets, there is a planar texture board on which two measurements are done and there is a person whose height measurement from top of the head to the middle of the feet is done, thereby resulting in 3 measurements per scene.

5.5.1 Scene Feature Density Variation

Fig. 15 shows the percentage accuracy with actual values in Table 3 for all 126 reference image frame sets as a function of feature density variation. It is expected that both single and multi-capture dynamic calibration will work best on high feature density images as is evident from 95% and 94%accuracy for both methods. Multi-capture dynamic calibration performs better in low (94%) and mid (95%) feature images as compared to single-capture method being 91% and 92% accurate. This makes it more generic to adopt as most of everyday images fall in these two categories. The performance of factory calibration is worse for all the three cases being 87% (low), 89% (mid) and 87% (high) accurate.

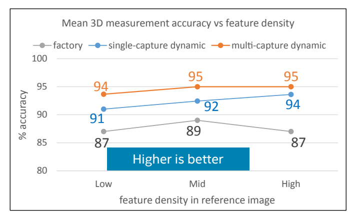

Figure 15. 3D measurement accuracy w.r.t feature point distribution in the reference image. Higher is better.

Feature density	Factory calibration		Single-capture dynamic calibration		Multi-capture dynamic calibration	
	Mean accuracy (%)	Std. dev. (%)	Mean accuracy (%)	Std. dev. (%)	Mean accuracy (%)	Std. dev. (%)
Low	87	7	91	8	**94**	**6**
Mid	89	5	92	6	**95**	**7**
High	87	20	94	5	**95**	**5**

Figure 16. (μ, σ) of 3D measurement accuracy in Fig. 15.

5.5.2 Distance of 3D measurement object

Fig. 17 and Fig. 18 show the average percentage 3D measurement accuracy as a function of distance of the planar texture pattern on which measurements are made averaged over all lighting conditions and reference image feature point distribution. It can be seen that the multi capture performs better than factory calibration and is slightly better or equal when compared to single-capture calibration. The accuracy of all methods goes down for measurements which are done far from the multi-camera system.

5.5.3 Light Level Variations

Fig. 19 and Fig. 20 show the percentage accuracy of 3D measurement with varying light conditions. The performance of factory calibration is worse. Multi-capture cali-

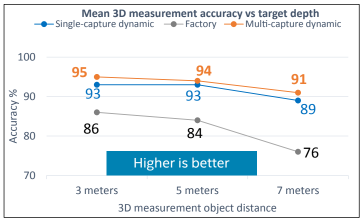

Figure 17. 3D measurement accuracy w.r.t distance of the object in the reference image. Higher is better.

Target distance	Factory calibration		Single-capture dynamic calibration		Multi-capture dynamic calibration	
	Mean accuracy (%)	Std. dev. (%)	Mean accuracy (%)	Std. dev. (%)	Mean accuracy (%)	Std. dev. (%)
3 meters	86	8	93	5	**95**	**8**
5 meters	84	14	93	6	**94**	**7**
7 meters	76	18	89	10	**91**	**10**

Figure 18. (μ, σ) of 3D measurement accuracy in Fig. 17.

bration method performs the best and is typically independent of the lighting condition averaging around $93\% - 94\%$ accuracy. For case of low light reference image frames sets $(150, 400$ lux), multi-capture method basically selects well lit candidate images as they will have better distribution of feature points, thereby obtaining improved results.

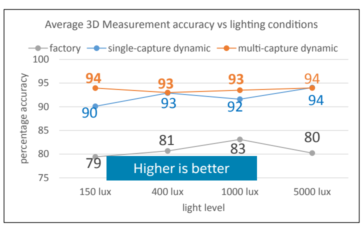

Figure 19. 3D measurement accuracy w.r.t light conditions in the reference image. Higher is better.

light level (lux)	Factory calibration		Single-capture dynamic calibration		Multi-capture dynamic calibration	
	Mean accuracy (%)	Std. dev. (%)	Mean accuracy (%)	Std. dev. (%)	Mean accuracy (%)	Std. dev. (%)
150 lux	79	10	90	10	**94**	**5**
400 lux	81	8	93	7	**93**	**7**
1000 lux	83	9	92	7	**93**	**6**
5000 lux	80	10	94	5	**94**	**5**

Figure 20. (μ, σ) of 3D measurement accuracy in Fig. 19.

5.6. Image Undistortion: Single vs Multi-Capture

Fig. 21 shows the effect of improved undistortion when using a multi-capture dynamic calibration method. The reference image does not have any feature points in the lower part of the image (see the inset image in Fig. 21(left)). Single-capture dynamic calibration results in parameters which results in over-curving of the image corners as evident in Fig. 21(middle) where the straight light fixture on the ceiling curves. When applying the multi-capture dynamic calibration method, the additional image (seen in inset of Fig. 21(right)) used for calibration has many features in the regions where the reference image lacked features. The overall spread in the feature points on 2D image space as well as 3D depth results in better calibration parameters leading to expected undistortion of the light fixture, where it remains straight.

Figure 21. Undistortion performance: (a) reference (b) undistorted image using single-capture and (c) multi-capture dynamic calibration parameters.

5.7. Run-time Performance

All our results are based on a C++ code running on Intel(R) Core(TM) i7-5775C CPU @ 3.30GHZ (4 cores) with 8GB RAM and uses OpenMP (8 threads) and SSE (Streaming SIMD Extensions) optimizations. The average runtime over our dataset for single-capture dynamic calibration is 2.97 secs. The proposed multi-capture dynamic calibration (10 image frame sets) takes 11.48 secs as compared to greedy approach which takes 570.65 secs making it 50 times faster.

6. Conclusion

In this paper, we have presented a new paradigm of multi-capture multi-camera calibration which accrues feature points from image of completely different scenes for dynamic calibration. We have shown methods to jointly model calibration parameters along with an optimality criteria to select the best set of scenes to use as part of joint calibration. We have shown better performance of the multi-capture calibration parameters over factory and single-capture parameters with respect to various validation metrics.

References

[1] Dell venue 8 7000. https://www.cnet.com/products/dell-venue-8-7000/. 5, 6

[2] S. Agarwal, K. Mierle, and Others. Ceres solver. http://ceres-solver.org. 5

[3] P. F. Alcantarilla, J. Nuevo, and A. Bartoli. Fast explicit diffusion for accelerated features in nonlinear scale spaces. In *British Machine Vision Conference (BMVC)*, 2013. 5

[4] J. P. Barreto and K. Daniilidis. Fundamental matrix for cameras with radial distortion. In *IEEE International Conference on Computer Vision (ICCV)*, volume 1, pages 625–632 Vol. 1, Oct 2005. 3

[5] P. Beardsley, P. Torr, and A. Zisserman. *3D model acquisition from extended image sequences*, pages 683–695. 1996. 5

[6] J.-Y. Bouguet. Camera calibration toolbox for matlab. Website, 2000. http://www.vision.caltech.edu/bouguetj/calib_doc/. 1

[7] G. Bradski and A. Kaehler. *Learning OpenCV: Computer Vision in C++ with the OpenCV Library*. O'Reilly Media, Inc., 2nd edition, 2013. 7

[8] G. Carrera, A. Angeli, and A. J. Davison. Slam-based automatic extrinsic calibration of a multi-camera rig. In *IEEE International Conference on Robotics and Automation (ICRA)*, pages 2652–2659, May 2011. 3

[9] O. D. Faugeras, Q. T. Luong, and S. J. Maybank. *Camera self-calibration: Theory and experiments*, pages 321–334. Springer Berlin Heidelberg, Berlin, Heidelberg, 1992. 2

[10] A. W. Fitzgibbon. Simultaneous linear estimation of multiple view geometry and lens distortion. In *IEEE Computer Vision and Pattern Recognition (CVPR)*, volume 1, 2001. 2

[11] L. Heng, M. Brki, G. H. Lee, P. Furgale, R. Siegwart, and M. Pollefeys. Infrastructure-based calibration of a multi-camera rig. In *IEEE International Conference on Robotics and Automation (ICRA)*, pages 4912–4919, May 2014. 3

[12] L. Heng, G. H. Lee, and M. Pollefeys. Self-calibration and visual slam with a multi-camera system on a micro aerial vehicle. *Autonomous Robots*, 39(3):259–277, Oct 2015. 3

[13] A. Kumar and N. Ahuja. On the equivalence of moving entrance pupil and radial distortion for camera calibration. In *IEEE International Conference on Computer Vision (ICCV)*, December 2015. 1

[14] 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)*, pages 1301–1307, Nov 2013. 3

[15] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, Nov. 2004. 5

[16] D. Nistér. An efficient solution to the five-point relative pose problem. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(6):756–777, June 2004. 2, 5

[17] M. Pollefeys, R. Koch, and L. V. Gool. Self-calibration and metric reconstruction in spite of varying and unknown internal camera parameters. In *IEEE International Conference on Computer Vision (ICCV)*, pages 90–95, Jan 1998. 2

[18] Z. Zhang. A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000. 1