14th IAPR International Conference on Machine Vision Applications (MVA)
May 18-22, 2015. Miraikan, Tokyo, Japan

Tsai Camera Calibration Enhanced

Trevor Gee[1], Patrice Delmas[1], Nick Stones-Havas[1], Chris Sinclair[1], Wannes Van Der Mark[1], Wei Li[2], Heide Friedrich[2], and Georgy Gimel'farb[1]

[1]Department of Computer Science, The University of Auckland, New Zealand
[2]Department of Civil and Environmental Engineering, The University of Auckland, New Zealand

Abstract

The rise of Unmanned Aerial Vehicles for surveying and sensing tasks have created new challenges for quick calibration of sensing systems, which we feel is a critical issue. In this context, calibration is performed often and needs to be achieved as quickly as possible. An approach with minimal user-interaction, which preserves sensing accuracy would be ideal. We propose a version of Tsai's camera calibration with an improved distortion model and two non-linear optimization phases to calibrate our UAV equipped with a stereo-camera system. We trialled our proposed calibration approach against three known reconstruction pipelines: an uncalibrated pipeline, a pipeline calibrated using Zhang's approach and a pipeline calibrated using Tsai's original approach. Our findings indicate that our approach has competitive accuracy, while requiring far less user-interaction than Zhang's approach.

1 Introduction

Traditionally, remote controlled quadcopter drones are armed with a single camera system. Thus, in terms of image processing, camera pose needs to be estimated based on feature point matching or optic flow across image frames. Unfortunately, since single camera systems have no notion of depth, single image processing algorithms need to make many assumptions that often lead to errors. We decided to arm our quadcopter drone with a stereo camera system instead.

With a stereo system, comes the need to calibrate cameras. In our research, this calibration needs to happen often, due to the volatile nature of drone flights and the frequent exchange of camera rigs. Zhang's calibration is an obvious first choice due to it's wide accessibility (the MATLAB toolkit for example) and it's robust performance. In practice, we found Zhang's approach problematic due to the image acquisition phase. Zhang calibration requires 15 or more images covering the entire field of view (FOV). Due to outdoor constraints, this acquisition needs to be rushed (limited battery life), resulting in blurred images, or insufficient numbers of images that cover only a small portion of the FOV.

Tsai's calibration is an attractive alternative to Zhang's approach, since it only requires a single image per camera. Also, as neither the cameras nor the calibration object move during the image acquisition phase, issues such as synchronization and blur are no longer problems.

Therefore we address of accuracy of Tsai's calibration, particularly in the context of 3D reconstruction for a UAV. Based on the observation that the traditional Tsai approach is far less optimized than the Zhang approach, we pose the question: Given further optimization, are we able to improve the performance of Tsai calibration?

Consequently, we propose a three stage optimized calibration algorithm based on Tsai's approach: We first use the traditional Tsai linear equations to acquire a basic best-fit linear calibration model. This is then used to initialize a second phase, where the Levenberg-Marquardt algorithm optimizes 17 calibration parameters, including an upgraded distortion model, to find a second best-fit calibration model. These calibration parameters are then inputs to initialize yet another Levenberg-Marquardt optimization to refine a model that describes the relationship between left and right images of our stereo camera system. This final model can then, in turn, be used as the basis for rectification of stereo image pairs.

In the next section, we briefly review related literature. In section 3, we describe some of the equipment that was used during our experimentation. After that, section 4 covers the 3D pipelines featured in this paper. Next, section 5 introduces our massively optimized Tsai algorithm. Finally, we close with experimental results and a discussion of conclusions in section 6 and section 7 respectively.

2 Literary Review

Drones or unmanned aerial vehicles (UAV), traditionally only have a *single* camera attached to them. It is for this reason that many articles [7] focus on a single camera approach for these vehicles. Many of these approaches are very similar to the uncalibrated pipeline presented in this work.

As calibrated rectification tends to outperform uncalibrated rectification, there is a lot of research exploring ways to reduce the difference between the two approaches [2].

Of the calibrated approaches, two important approaches stand out. Tsai's calibration [10] was one of the top calibration algorithms in 1987, and continued to dominate over a number of years [8]. Later, when Zhang's calibration [11] was published, several studies were performed to compare the these two algorithms [9]. Typically these studies found Zhang's approach to be more complete and more accurate than Tsai's approach. However, in 2014, Wei Li et al. [4] found these differences not to be that significant in the context of 3D reconstruction.

3 Equipment used

The equipment used in the experiments within this paper is shown in Fig. 1.

Figure 1. Zhang calibration Object (Top Left), GoPro HERO3+ Cameras (Top Right), Phantom quadcopter (Bottom Middle)

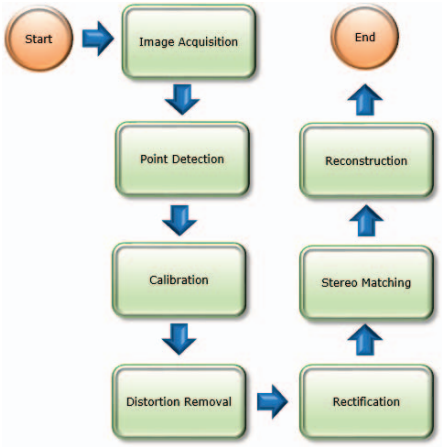

Figure 2. Calibrated Stereo Vision Pipeline

The first item is the Tsai calibration object. Our calibration object was constructed out of perspex (top left of Fig. 1). A checkerboard pattern was attached to each of the two surfaces, to facilitate automatic point extraction [4].

The UAV featured is the DJI Phantom quadcopter (bottom middle of Fig. 1). It is capable of lifting a payload of 300 grams. It's top flying speed is 10 $m.s^{-1}$. The UAV has been upgraded with carbon fibre propellers and propeller guards to improve it's performance and robustness.

The payload of our UAV is a stereo box housing twin GoPro HERO3+ cameras (top right of Fig.. 1). Our motivation for choosing these cameras for our experiments is that they are relatively cheap, robust to crashes and give good quality images. However, it should be noted that GoPro HERO3+ cameras possess a significant amount of barrel distortion which must be accounted for.

4 Pipelines

4.1 A calibration 3D reconstruction pipeline

The calibrated pipeline, is shown in Fig. 2. The steps of the pipeline are exactly the same for both calibration techniques.

The first step of the process is image acquisition. The goal of the image acquisition step is to acquire

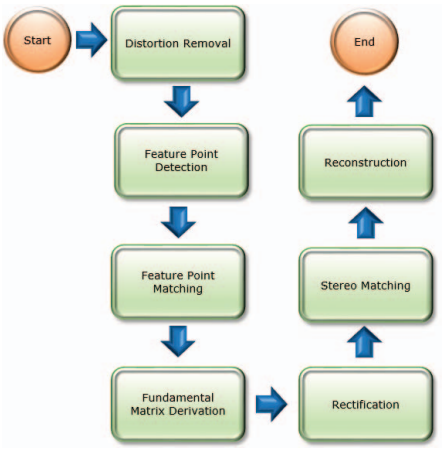

Figure 3. Uncalibrated Stereo Vision Pipeline

data to drive the calibration process. During Tsai's calibration, only a single image is required for each camera. That image must be of a calibration object with two orthogonal planes such as the one in Fig. 1. Zhang's calibration requires 15 or more images from each camera. The Zhang's calibration object needs to be planar, and is orientated in different ways for each image, so that the entire body of images collectively cover camera's field of view.

Once the images have been acquired, feature points need to be extracted from those images. In our experiments, we attach chessboard patterns to our calibration objects so that we are able use the approach of Arca et al. [1] for feature point detection and Wei's [4] approach for feature point extraction.

The extracted set of feature points forms the input to the calibration algorithm. We used OpenCV's version of Zhang's calibration, and wrote our own version of Tsai's calibration.

Both calibration algorithms produce a model of radial distortion within the image. In both cases, we applied this model, along with bilinear interpolation, to perform distortion removal from images.

Next, we rectified stereo pairs using the approach of Fusiello et al. [3].

Following rectification is stereo matching. We used the guided stereo dynamic programming algorithm developed by Nguyen et al [6] for improved accuracy at a reasonable processing cost. This is a hybrid algorithm that uses a sub-sampled graph-cut stereo algorithm to guide a dynamic programming stereo algorithm.

Stereo matching produces a disparity map that we use, along with the implementation of Nguyen et al [6], to generate a 3D reconstruction.

4.2 A 3D reconstruction pipeline without calibration

Our uncalibrated 3D reconstruction pipeline rectifies images based on a sparse set of matching feature points and epipolar geometry. This pipeline is illustrated in Fig. 3.

The first step in the pipeline is distortion removal, as GoPro HERO3+ cameras have large barrel distortion. In order to remove this distortion, we needed to derive a distortion model. We began by manually locating an image within the sequence of images that contains a

known straight line. Edge detection is used to extract the pixels which belong to the line. We approximated the level of line curvature by determining the aggregated deviation between the line pixels and the line of best-fit. Finally, we used the first order radial distortion model (Eq. 1) and the bisection minimization algorithm to refine κ_1. This approach may be applied to each camera in the camera system, and as distortion should not vary across images in a sequence for a fixed focal length, the resultant model may be used to remove distortion from an entire sequence, including those frames that do not contain any known straight lines.

$$x_u = x_d(1 + \kappa_1 r^2), y_u = y_d(1 + \kappa_1 r^2) \quad (1)$$

After the removal of distortion, the next step is to find features within the scene. We used OpenCV's implementation of SIFT. Once feature points have been discovered, they need to be matched across images. We used OpenCV's *BFMatcher* to perform this function.

The 3 by 3 fundamental matrix maps epipolar lines to points across stereo pairs. To calculate the fundamental matrix, we chose to use the random sample consensus (RANSAC) algorithm to reduce sensitivity to outliers. Once the fundamental matrix was acquired, we performed an uncalibrated rectification as per OpenCV.

After rectification, the remaining steps are the same as the ones in the calibrated pipeline. See section 4.1 for details.

5 Optimized Tsai

The traditional Tsai approach is a two-step approach. Firstly, image acquisition mechanics are described using a very simple linear pin-hole camera. Next, the non-linear distortion model described by Eq. 1 is added to the model. An iterative non-linear solver such as the Levenberg-Marguardt algorithm is then used to optimize and refine the focal length f, the z-component of translation T_z, and the first order radial distortion κ_1.

5.1 An enhanced distortion model

The first modification we made to the standard Tsai algorithm was an extension of the distortion model. Our motivation for doing this was the notion that a more sophisticated model of distortion could lend more flexibility to the camera model, allowing for a closer fit to the actual mechanics of the image acquisition process. To achieve this, we modified our Tsai implementation to use Zhang's distortion model, including principle point decentering.

$$\begin{aligned} x_d &= x_u(1 + \kappa_1 r^2 + \kappa_2 r^4 + \kappa_3 r^6) \\ &\quad + 2P_1 x_u y_u + P_2(r^2 + 2x_u^2) \\ y_d &= y_u(1 + \kappa_1 r^2 + \kappa_2 r^4 + \kappa_3 r^6) \\ &\quad + P_1(r^2 + 2y_u^2) + 2P_2 x_u y_u \end{aligned} \quad (2)$$

The Eq. 2 depicts the Zhang distortion model. The distorted point (x_d, y_d) is mapped to the undistorted point (x_u, y_d). Coefficients κ_1, κ_2, κ_3 are the first, second and third order radial distortion coefficients, while P_1 and P_2 are the first and second tangential distortion coefficients.The variable r is the radial displacement of the undistorted point from the principle point.

5.1.1 First Optimization

Next, we focused on the non-linear optimization of Tsai's algorithm. Instead of optimizing only three calibration parameters as is normally done in Tsai, we optimized all 5 distortion coefficients, x and y coordinates of the principle point, the focal length f, the three components of the translation vector and three rotation components in Rodrigues' form (since it was found experimentally that this representation of rotation produced the best results). The optimization itself was performed with the Levenberg-Marquardt algorithm from the cminpack open source library.

5.1.2 Second Optimization

Standard Tsai calibration derives a model that describes the relationship between the calibration object and the image. However, with calibrating for rectification in mind, our focus is on the relationship between different cameras. Therefore we add a second optimization to minimize inter-camera error.

$$\begin{aligned} {P_1}^T &= R_1 O^T + T_1 \\ {P_2}^T &= R_2 O^T + T_2 \end{aligned} \quad (3)$$

In order to perform this second operation, we used the extrinsic model depicted in Eq. 3. Here ${P_1}^T$ and ${P_2}^T$ are two camera coordinate systems that have been transformed from the calibration object coordinates O^T. The rotations R_1 and R_2 and the translations T_1 and T_2 are determined by the first optimization of the algorithm. We define the relationship between the two cameras as shown in Eq. 4. This relationship allows us to initialize the rotation between cameras as ${R_1}^{-1}R_2$ and the translation as $T_2 - {R_1}^{-1}R_2T_1$. After initialization, optimization takes place using Levenberg-Marquardt guided by the reprojection errors of calibration points.

$${P_2}^T = ({R_1}^{-1}R_2){P_1}^T + (T_2 - {R_1}^{-1}R_2T_1) \quad (4)$$

6 Results

The first series of experiments were performed indoors with controlled lighting. We implemented a pipeline for Zhang, Tsai, Optimized Tsai and an uncalibrated stereo pipeline. Results can be seen in Table 1. The first step was point extraction (labelled Pnt in Table 1). The quality of this extraction was evaluated by the amount of outliers removed during calibration based on reprojection error. This amounted to 10% of measured points across the different pipelines. The consistency of this number was due to same algorithm being used across approaches. Calibration quality (Calib in Table 1) was measured via reprojection error and is reported in pixels. The quality of distortion removal (Dist in Table 1) was approximated using a grid image and measured by the amount of deviation from the each line of best-fit. The fully distorted image we used had a typical deviation of about 63 pixels. Rectification accuracy (Rect in Table 1) was based

on the average deviation per pixel from the horizontal epipolar line. Experiments were repeated 10 times, each time capturing a new set of calibration images and running through the steps of each pipeline.

Table 1. Accuracy (pixels) in stages of of the pipelines based on 10 calibration cycles

Type	Pnt(%)	Calib	Dist	Rect
Zhang [11]	0.1 ± 0.1	0.3 ± 0.2	9.1 ± 2.3	0.2 ± 0.1
Tsai [10]	0.1 ± 0.1	1.1 ± 0.6	11.8 ± 3.6	0.9 ± 0.3
OTsai	0.1 ± 0.1	0.5 ± 0.2	10.2 ± 2.9	0.5 ± 0.2
UnCal [6]	N/A	N/A	16.4 ± 5.2	1 ± 0.5

The same experiments were repeated outdoors where Fig. 4 are some of our rectified images. We noticed that in general uncalibrated and Tsai calibration techniques produce the most warped images, while the Optimized Tsai (see OTsai in Table 1) and Zhang produced less distorted images. As our stereo box keeps our cameras roughly in a canonical epipolar configuration, one would expect the most accurate rectification to only have minor warping.

Figure 4. Rectification: Uncalibrated (top left), Tsai (top right), Optimized Tsai (bottom right), Zhang (bottom left)

Fig. 5 depicts disparity maps produced after Zhang and optimized Tsai calibration, rectification and stereo matching [5]. While the Tsai disparity map is a little more noisy, the obtained disparity map produces a reasonable 3D reconstruction as can be seen in Fig. 6.

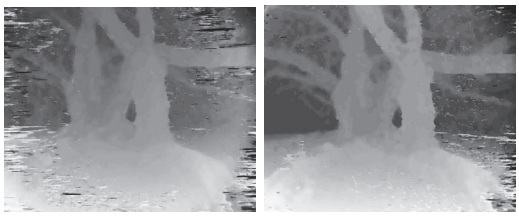

Figure 5. Optimized Tsai Disparity Map (left) Zhang Disparity Map (right)

7 Conclusions

We managed to significantly improve on the performance of the original Tsai algorithm. Our optimized Tsai algorithm produces rectification with errors within 0.3 ± 0.2 of a pixel to that of the Zhang algorithm. The importance of this becomes obvious when we consider the fact that Tsai's algorithm only requires a single image per camera while the Zhang's approach requires at least 15. As image acquisition tends to be a manual process, using Tsai's approach is a significant saving. This fact, in conjunction with the assurance that using Tsai's approach does not result in much reduction in accuracy, makes our optimized Tsai algorithm an attractive proposition when calibrating drone cameras for 3D reconstruction tasks.

References

[1] S. Arca, E. Casiraghi, and G. Lombardi. Corner localization in chessboards for camera calibration. *IADAT-micv2005*, 2005.

[2] A. Fusiello and L. Irsara. Quasi-euclidean uncalibrated epipolar rectification. In *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*, pages 1–4. IEEE, 2008.

[3] A. Fusiello, E. Trucco, and A. Verri. A compact algorithm for rectification of stereo pairs. *Machine Vision and Applications*, 12(1):16–22, 2000.

[4] W. Li, T. Gee, H. Friedrich, and P. Delmas. A practical comparison between zhang's and tsai's calibration approaches. *IVCNZ 2014*, 2014.

[5] M. Nguyen, Y.-H. Chan, P. Delmas, and G. Gimel'farb. Symmetric dynamic programming stereo using block matching guidance. *IEEE IVCNZ'13 Proc.*, pages 88–93, 2013.

[6] M. Nguyen, G. Gimel'farb, and P. Delmas. Web-based on-line computational stereo vision. *IEEE IVCNZ'08 Proc.*, pages 1–6, 2008.

[7] F. Remondino, L. Barazzetti, F. Nex, M. Scaioni, and D. Sarazzi. Uav photogrammetry for mapping and 3d modeling–current status and future perspectives. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 38(1):C22, 2011.

[8] J. Salvi, X. Armangué, and J. Batlle. A comparative review of camera calibrating methods with accuracy evaluation. *Pattern recognition*, 35(7):1617–1635, 2002.

[9] W. Sun and J. Cooperstock. Requirements for camera calibration: Must accuracy come with a high price? In *IEEE 2005 Workshops on Application of Computer Vision*, pages 356–361, 2005.

[10] R. Tsai. A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf tv cameras and lenses. *IEEE Journal of Robotics and Automation*, 3(4):323–344, 1987.

[11] Z. Zhang. A flexible new technique for camera calibration. *IEEE PAMI*, 22(11):1330–1334, 2000.

Figure 6. Two 3D models using optimized Tsai with the stereo system of Nguyen et al. [6].