Calibration for High-Definition Camera Rigs with Marker Chessboard

Jianhui Chen Karim Benzeroual
Robert S. Allison
Centre for Vision Research, Department of Computer Science and Engineering
York University, Toronto Canada
{jianhui,karim,allison}@cse.yorku.ca

Abstract

The geometrical calibration of a high-definition camera rig is an important step for 3D film making and computer vision applications. Due to the large amount of image data in high-definition, maintaining execution speeds appropriate for on-set, on-line adjustment procedures is one of the biggest challenges for machine vision based calibration methods. Our aims are to provide a low-cost, fast and accurate system to calibrate both the intrinsic and extrinsic parameters of a stereo camera rig. We first propose a novel calibration target that we call marker chessboard to speed up the corner detection. Then we develop an automatic key frame selection algorithm to optimize frames used in calibration. We also propose a bundle adjustment method to overcome the geometrical inaccuracy of the chessboard. Finally we introduce an online stereo camera calibration system based on the above improvements.

1. Introduction

In 3D film making, one of the main problems is stereo camera calibration which aims to match the two camera images in their optical characteristics and to align them optomechanically. Typically this includes compensating for differences in focal length, distortion, position and orientation leaving the separation of the cameras that produces the intended stereo baseline (and possibly their convergence) as the only difference between the camera pair. Calibration and alignment are critical since only tenths of millimeters of difference in a camera setting can cause many pixels displacement on the screen. Spurious disparities due to misalignment can degrade fusion[3], stereoscopic depth perception [12], and comfort [17, 16] leading to a poor stereoscopic experience for viewers.

On set camera calibration needs to be performed effectively and efficiently to avoid costly delays in production. Though there are many commercial systems available for stereo calibration, they may be too expensive for independent film-makers. For instance the Stereo Image Processor (3ality Technica, Burbank CA) uses hardware image processing and closed loop camera motor control to estimate and automatically adjust stereo rig alignment. Such a system is far beyond the budget of independent and modest productions, which rely on manual adjustment of the rig parameters to achieve alignment. Recently, systems such as the Stereo3D CAT (Dashwood Cinema Solutions, Toronto Canada)[1] have been introduced to aid the manual alignment of stereo rigs on set with accurate localization of markers in a test chart. Currently rig alignment on a film production set still relies on either expensive custom hardware or highly experienced operators to align the cameras aided by custom alignment charts and tools.

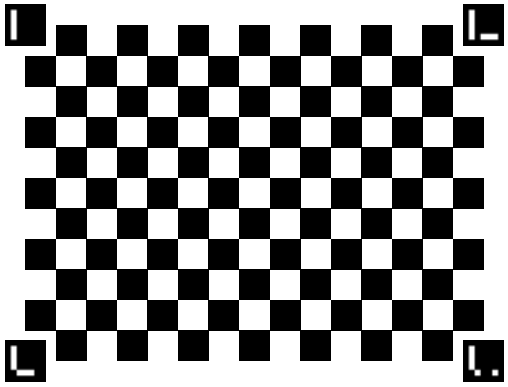


Figure 1: Marker chessboard.

On the other hand, camera calibration is a mathematically well-defined problem in machine vision. Researchers first use 3D or 2D objects to estimate intrinsic parameters of the camera [22, 14, 23]. The camera intrinsic parameters describe the optical properties of the camera such as the focal length and lens distortion. They then obtain extrinsic parameters of the stereo camera rig [13, 19] which describe the relative orientation and position of each camer-

978-1-4673-1612-5/12/$31.00 ©2012 IEEE

a in the rig relative to a base frame (typically corresponding either one of the cameras or to a frame of reference fixed to the scene). This is accomplished by analyzing and applying geometrical constraints to stereoscopic images or image sequences. The process is usually facilitated by imaging a target with known geometry, such as a chessboard.

These techniques potentially provide an affordable solution to professional camera calibration. In this paper we describe an extension of computer vision algorithms for camera calibration to the alignment of a stereoscopic film camera rig. Our goals for the system are as follows. First, it should be affordable to independent 3D film makers both in hardware and software. Second, it should provide competitive accuracy with commercial systems and machine vision standards.

However, there are several challenges to achieving these goals by straightforwardly applying machine vision techniques to film-making. The first challenge comes from high image resolution. Manual alignment of a rig on set requires interactive execution speeds to allow for adjustments while monitoring responses. Furthermore the camera chosen to mount on the rig, lens selection or zoom, interaxial distance, and other parameters change frequently from shot to shot eliminating the possibility of 'locking down' the rig after a single time consuming calibration. Current software implementations of standard techniques are much too slow. Specifically, the corner detection processing in the most popular technique [5] is not fast enough for interactive response. Second, the film production crew must do the selection of suitable image frames for the calibration process on set. A skilled cinematographer, camera operator or stereographer can be assumed but familiarity with computer vision techniques can not. This is further complicated by the fact that some cameras can only record video instead of an image. Manually selecting frames for calibration is time consuming and may be inaccurate. Finally, a wide variety of lighting conditions, focal lengths, camera distances and environments must be accommodated due to the large variety of scenes in even a single film production. This removes the possibility of relying on a single highly precise calibration object. Planar chessboards are relatively easy to construct and modify making them an attractive and flexible solution. However a chessboard printed and mounted on a flat substrate using economical techniques available on set may not be geometrically accurate. To overcome these difficulties, we propose three improvements which are introduced in section 4.

To summarize, the contribution of our paper are as follows:

- A novel design called the marker chessboard that greatly speeds up corner detection.
- An automatic key frame selection algorithm that liberates people from repetitive and time consuming work.
- An easily implemented bundle adjustment method to improve the accuracy of intrinsic calibration.
- A low-cost online stereo camera calibration system.

2. Related work

There are two steps to recover the relative pose of a stereo camera in Euclidean space. The first step is to compute the camera matrix and distortion parameters usually by analyzing images of a target with known geometry. These parameters are called the intrinsic parameters since they are relatively stable for a given camera. In a film production context these are often fixed (with the exception of focus) for a given camera-lens pair for a group of shots or often even for an entire production (with a variety of cameras or rigs used across shots). The second step is to compute the rotation matrix and translation vector between two cameras. These parameters are called extrinsic parameters since they are changed by the relative pose of cameras. In a film production scenario the extrinsic parameters of a rig vary from shot to shot or occasionally within a shot as cinematographers vary the interaxial distance and convergence to achieve the desired stereoscopic effects. Thus monitoring and alignment of the extrinsic parameters is a frequent and routine requirement. It is worth noting that, embedded in the estimation of the camera intrinsic parameters, another set of extrinsic parameters is involved, but it describes the relative pose between the target and its original (canonical) pose rather than the relationship between two cameras.

For intrinsic calibration, choices of calibration target, lens distortion model, and optimization method are three of the most important factors impacting accuracy. For the calibration target, we chose a planar chessboard. Such a target could even be manufactured on set using standard office printing equipment and flat stock materials typically available on set, allowing for flexibility to address unforeseen issues with scale, access, illumination or other factors. Another advantage is that a significant number of research have deeply investigated this configuration and have led to impressive accuracy improvements. For example, Datta *et al*. [9] proposed an iterative refinement algorithm to increase the accuracy of control point localization and calibration results. Albarelli *et al*. [2] proposed a bundle adjustment based algorithm and obtained a very accurate result. For the lens distortion model, we chose three radial and two tangential distortion coefficients proposed by Brown [7]. Such a model is applicable to most film camera-lens combinations and in our experiments, they have shown to be accurate enough for stereo rig calibration.

For the optimization method, we use the implementation of [23] in Open CV [6] for its free availability and efficiency.

For extrinsic calibration, there are also many algorithms available in the computer vision literature. The eight point [13] and five point [19] algorithms are the two standard ones and can be performed with the given number of stereo correspondences in a general scene. However, it is not possible to recover the camera position displacement from a given fundamental matrix based on the eight points algorithm. Also, image noise will influence the accuracy of the five point algorithm in practice. Our method obtains the relative pose from a chessboard pattern as we do in the intrinsic calibration. Camera crew are used to relying on calibration charts for colour, geometry and lighting control and thus this will be a standard procedure for them. We first get the relative pose between each camera and the target, then we obtain the relative pose of cameras by rotating and translating the right camera to the left camera's coordinate frame. In this method, we can take advantage of known correspondence between many features in the target (corners in the chessboard) and reduce the influence of noise.

3. System overview

Our system has two parts. The first part computes intrinsic parameters for different lenses. Typically on a film set, focal length is controlled by using a set of high-quality fixed focal length 'prime' lenses but occasionally, particularly on low budget productions, a zoom lens may be used. For our testing this was convenient and we calculated intrinsic parameters for a range of different zoom levels. At a specified zoom level, we first record a video in which an operator intermittently changes the position and orientation of the chessboard target. Then our system automatically selects proper frames from the video to estimate the intrinsic parameters. We call these frames as key frames in this paper. The second part computes extrinsic parameters of a stereo camera rig. We mounted the stereo camera pair in a mirror rig. Our system provides the relative pose of the cameras at an interactive rate based on the known intrinsic parameters and the stereo correspondences in the captured frames.

The rest of paper is organized as follows. Section 4 describes three improvements for calibrating intrinsic parameters from a video. Section 5 reports performance of the extrinsic calibration system. Finally, we draw conclusions and discussion in section 6.

4. Intrinsic calibration

4.1. Marker chessboard

The chessboard pattern is widely used in calibration and its corner positions can be automatically detected in Open CV. The corner detection algorithm works in two steps. The first step is to obtain rough corner positions. The algorithm first finds all squares as connected components, then fits a polygon to each connected component. If a resulting polygon has four vertices, it is a qualified square. Then it orders qualified squares into a grid until the pattern is found. The second step is to refine the resulting corner position to sub-pixel localization. However, there are two problems with this algorithm. First, the speed decreases rapidly with increase in image resolution. For a 1920×1080 image, it takes about 200ms per frame. Second, processing speed severely degrades when the image quality suffers from motion blur which frequently happens when the operator moves the target. Through the experiments, we found the first step costs most of execution time. If we can replace it with a time efficient method, we can greatly speed up the corner detection processing.

Our method is inspired by the manual corner detection method in the camera calibration toolbox [5]. In their method, the user needs to select four corner positions that are located in the outermost level of the chessboard with mouse clicks. Then it computes a homography matrix from the corresponding physical corners to these four corners, and uses the homography matrix to obtain the rough position of the remaining corners within the chessboard. If we can automatically detect four special corners in the chessboard, this manual method can be automated. This was our motivation for adding four markers to the traditional chessboard. Figure 1 shows our new chessboard which we refer to as the marker chessboard. Both markers and chessboard have been used in corner detection for camera calibration [10]. The combination of these techniques have not, to our knowledge, been proposed in previous research.

Compared with the traditional "black" and "white" chessboard, the marker chessboard has four easily identifiable markers in the four corners. Each marker is made a little bit larger than the square of the chessboard for easy detection (edge length ratio between them is 2:1.5). Without loss of generality, for a chessboard with a size of $w' \times h'$ (w' and h' are the number of inside corners along the chessboard width and height, respectively, with $w' \geq h'$), we set origin at the top-left inside corner and set the small square edge length as the unit length. Relative to this coordinate frame, then the marker centers are placed at the coordinate of $(-1,-1), (w',-1), (w',h'), (-1,h')$ respectively and their edges are parallel with the chessboard square edges. We denote the marker centers and inside corner as p_m and p_c in the physical board, and p'_m and p'_c in the image and express them in homogeneous coordinates. For an image with small distortion, they are mapped by a homography matrix H in the perspective transformation: $p'_m = Hp_m$ and $p'_c = Hp_c$ respectively. The algorithm

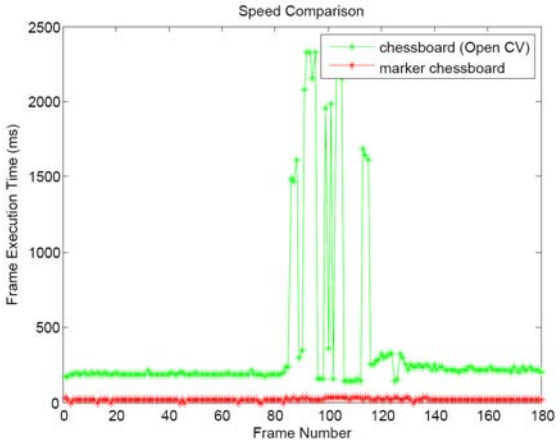
of marker chessboard detection is as follows:

1. Detect or track four marker centers in the image. If less than four markers are found, stop corner detection in the current frame immediately.
2. Compute the homography matrix H from physical marker centers p_m to image marker centers p'_m.
3. Compute the rough position of corners by $p'_c = Hp_c$.
4. Refine the corners p'_c with sub-pixel accuracy.

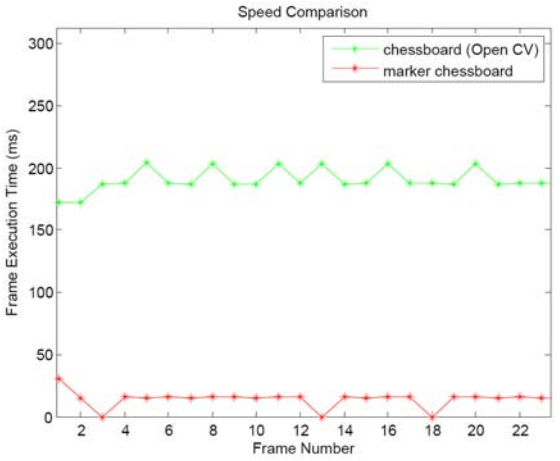
We choose the marker detection method in ArUco [20] for its unambiguous marker coding. One modification of their marker detection method is that we use adaptive thresholding of the perimeters of potential markers. Since the whole chessboard should be shot in the image, the maximal edge length of marker is $l_m = max(w/w', h/h')$, where w and h are the width and height of image with $w \geq h$. We heuristically set the maximum perimeter of a candidate marker as $6 \times l_m$, and also set the minimum threshold of it as l_m to filter short edges in the image. Moreover, a marker position should have small displacement in consecutive frames, therefore the algorithm can track the markers in a small region of interest in the image (searching 10 pixels beyond the bounding box of the marker corners works in practice) to further improve detection speed.

Figure 2 shows a speed comparison between the traditional chessboard using Open CV method and the marker chessboard using our method. The test video has 180 frames during which the chessboard is changed from one pose to another. As we can see, our method can greatly speed up corner detection without losing accuracy. Figure 2(b) shows a detailed comparison over the first 23 frames. Our method is so fast that the execution time is only a few milliseconds in some frames when marker tracking is near perfect (little or no motion between frames). Even without marker tracking, our method takes only about 30ms per frame (see the first frame in the Figure 2(b)). In contrast, the Open CV method takes about 200ms per frame.

Table 1 shows the average time cost for detected, un-detected and all frames for the same video. Detected frames are frames where the corners could be extracted from the frame, un-detected where the frame was discarded from further analysis and the 'all' conditions considers both detected and undetected frames. As we can see, our method speeds up the corner detection process by more than 10 times. The number of detected frames is smaller than the Open CV method since our method automatically discards the blurred images in the process of marker detection. On the contrary, the Open CV method spends a lot of execution time processing blurred images which provide little useful data but require significant computational effort (see the peaks of execution time for the chessboard curve between frames 80 and 120 in figure 2(a)).



(a)



(b) Detailed view of (a) over the first 23 frames.

Figure 2: Corner detection speed for a 180 frame calibration sequence.

Frame Statistic	Open CV	Our method
Detected number	151	128
Avg detected (ms)	213	15
Avg un-detected (ms)	1146	22
Avg all (ms)	363	17

Table 1: Statistics of corner detection.

4.2. Key frame selection

Some digital video cameras can only record a video instead of capturing a single image. So people need to manually select key frames that are used in calibration. It is a tedious and time consuming job. For example, for a

4-minute video, it usually takes more than 5 minutes for a well-trained operator to select the key frames. Since the selected frames may suffer from blur, have similar orientation or miss some particular orientations, they often need to repeat the process several times.

As suggested by [23] and our experiments, the key frames should be blur-free and have a sufficiently broad range of target orientation to allow for robust calibration. We developed an automatic key frame selection method to achieve these two demands. It first selects an initial key frame set by dividing the video into sub-sequences, then refines the set by iteratively checking each frame. The algorithm is as follows:

1. Detect the corners in each frame and mark blur free frames (and their corners) as candidate frames.
2. Divide the whole frame sequence into sub-sequences so that in each sub-sequence the corner displacement of consecutive frames is smaller than a threshold θ_a.
3. Randomly select one frame from each sub-sequence as initial key frame.
4. Calibrate the camera from the key frames.
5. Compute the unit normal of the target plane for each key frame in world coordinates, and build a kd-tree with these unit normals.
6. For each candidate frame, compute the target plane unit normal, and find its nearest neighbor distance in the kd-tree. If the distance is larger than a threshold θ_b, add it to the key frames and rebuild kd-tree.
7. Find the nearest neighbor of each key frame in the kd-tee (exclude itself). If its distance is less than the threshold θ_b, discard the one with larger blur.
8. If there is new frame added or discarded, rebuild kd-tree and go to step 4. Otherwise finish the selection.

In the candidate frame selection, we measure the motion blur using the corner displacement between the current frame and the preceding and subsequent frame. If the corner displacement is less than certain threshold (we set it empirically as 0.5 pixels), the frame is regarded as a blur free candidate frame. Only these candidate frames are used in key frame selection.

The algorithm divides the whole frame sequence into sub-sequences based on an observation: when the chessboard is changed from one pose to next pose, the corner position has a large displacement. So a displacement jump (θ_a) indicates that the chessboard may have been placed in a different pose in the subsequent frames. In our implementation, we set θ_a as $0.02 \times max(w, h)$.

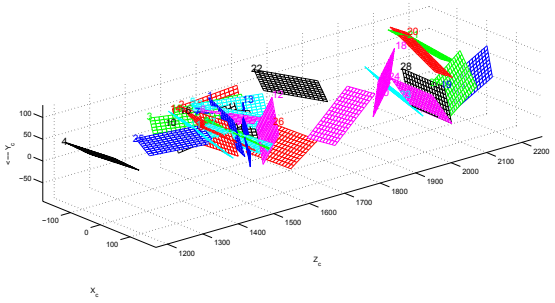


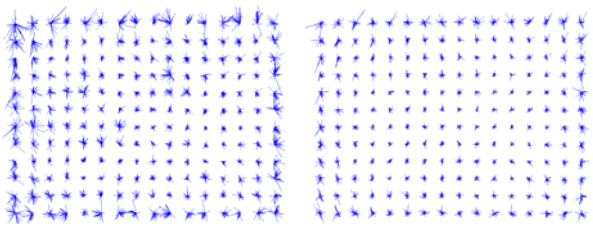
Figure 3: Key frame orientations.

We randomly select one frame from each sub-sequence as initial key frame for two reasons: first, there is, by construction, only a small corner displacement within a sub-sequence; second, every frame is guaranteed to be blur-free by step 1.

We use the unit normal of the target plane to describe the chessboard pose. Given a (roughly) calibrated camera in step 4, we compute the unit normal from 3D-2D correspondences for each frame. The unit normals are distributed over a unit sphere and their relative distance is defined as the geodesic distance in the sphere. We use the computationally more expedient calculation of the Euclidean distance to query the nearest neighbor in the kd-tree [4] since the Euclidean distance is approximately equal to the geodesic distance when two normals are close enough. We use the kd-tree based ANN (approximate nearest neighbor) searching method instead of a general cluster method such as mean shift [8] since the former approximately provides the key frames without computing the accurate cluster centers.

We check each candidate frame to make sure no key frame is missing. This process may add frames to the kd-tree. Also since a large displacement in the image plane does not guarantee a large orientation difference in 3D space, *e.g.* two parallel frames could differ in depth or location, we re-check each key frame with its nearest neighbor in the kd-tree. This process may discard key frames. Since every new selected key frame has a distance larger than the θ_b from its neighbours, and every discarded frame increases the minimum distance between the remaining key frames in the tree, the whole process converges rapidly. In our experiment, it usually takes 2-3 iterations. The threshold θ_b depends on the field of view of the camera and we empirically set it as 0.08.

Figure 3 shows 30 key frames obtained by our algorithm. They are automatically selected from 2621 frames with 634 candidate frames. As we can see, all of them have different orientations.



(a) before (b) after

Figure 4: Re-projection error before and after bundle adjustment (magnified by a factor of 40).

Item	Open CV	Our method
Corner detection (min)	35	3
Key frame selection (sec)	N/A	21
Calibration (sec)	4.3	16.7
Re-projection error	0.35	0.22

Table 2: Calibration of intrinsic parameters from a 4-minute video.

4.3. Bundle adjustment

The precision of intrinsic calibration relies on many factors [21]. In the experiment, we found that inaccuracy in the construction of the chessboard has a large impact on our system. Unfortunately, a highly accurate 2D chessboard requires very expensive material and needs a remarkable amount of effort for maintenance. To overcome this problem, we developed a bundle adjustment method.

$$s \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = R \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + T \qquad (1)$$

In equation 1, $[x, y, 1]$ is an undistorted corner position in camera coordinates and $[X, Y, Z]$ is the corresponding corner position in the chessboard. In a calibrated camera, we assume intrinsic parameters and target pose to be correct. Then we can produce a more accurate description of the chessboard corners in 3D space, $[X, Y, Z]$, by using equation 1 and a least square optimization procedure. Finally we re-calibrate the camera using the refined chessboard geometry.

Our method is inspired by [2], however there are two differences between our implementation and theirs. First, we drop the assumption that all corners lay in a plane which does not hold in our case (allowing for manufacturing tolerances effecting the z direction such as rippling of the printed material or variation in adhesion to the substrate). Second, we assume the extrinsic parameters R and T are accurate enough to be used in the bundle adjustment to simplify the optimization process. In the implementation, we only do one iteration of the bundle adjustment since we found that the changes of re-projection error and the intrinsic parameters are very small after the first iteration.

Figure 4 shows the re-projection error distribution before (0.35 pixels) and after (0.22 pixels) bundle adjustment. Considering the image resolution, this error is acceptable for the system in its current stage.

4.4. Intrinsic parameter estimation performance

Table 2 shows an example of the result of the entire process of intrinsic parameter estimation from a 4-minute video. Comparing with 35 minutes of Open CV method, our corner detection method only requires 3 minutes. Also the key frame selection algorithm automatically selects 30 key frames from the video and the bundle adjustment provides more accurate intrinsic parameters with only a small increment in processing time.

5. Extrinsic calibration

Figure 5 gives us an overview of the system and table 3 lists the equipments used. The marker chessboard was made by printing out the marker chessboard pattern on an A4 sheet of paper and stuck to a flat board. Two Canon XF105 cameras were synchronized by an AJA Gen10 sync generator. The software processed synchronous images from the cameras through the capture card.

Our software is based on Open CV which provides efficient fundamental image processing and camera calibration algorithms. The software frame rate is about 1 fps. For every processed frame, it first detects the markers and inside corner positions in both the left and right images. Then it provides two kinds of feedback to the camera operator to indicate the stereo camera position and orientation.

The first kind of feedback is image-based measurement which is based on the relative location of the four marker centers in each image. The four marker centers form a rectangle on the target plane and a quadrilateral in the left and right image respectively. The algorithm provides zoom level ratio by comparing the perimeters of these two quadrilaterals. Also it provides vertical and horizontal offset by averaging the offset of marker center positions in the horizontal and vertical direction respectively. The rotation is obtained by comparing the angles formed by each quadrilateral's diagonal and horizontal line. The image based measurement has its merits on simplicity and stability. However, it is not possible to differentiate the effects between the interaxial distance and the convergence angle directly. Operators need to place the chessboard at different depth to adjust these parameters. Similar to other stereo alignment systems such as [15], the software

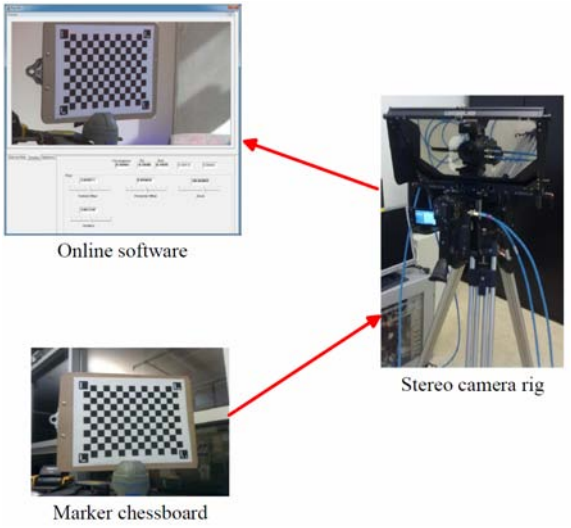


Figure 5: System overview: the arrows indicate the data flow.

Item	Equipment
Calibration target	Marker chessboard
Stereo camera	Two Canon XF105
Sync Generator	AJA Gen10
Capture Card	Black Magic Extreme 3D+
Computer	Duo Core 2.4GHz, 2.0 GB memory

Table 3: Extrinsic calibration test bed.

provides an image difference to quickly reveal coarse misalignments such as a vertical offset.

The second kind of feedback is a 3D-based measurement which is based on the inside corner positions of each square within the calibration grid (14×10) and the intrinsic parameters of the left and right cameras. We use these points to estimate the relative pose in form of the rotation matrix between the two cameras. Then we compute convergence, tilt, and roll angles from the rotation matrix. These three angles provide additional information of relative pose of two cameras, especially the convergence and tilt angles (the roll angle can be roughly estimated by rotation in the image-based measurement). In the experiment, we can adjust the tilt and roll angle to less than about 0.2° . However, we found that the convergence angle estimated is noisier and the error can vary between about -1.0° to 1.0°. Our analysis is that: although 5 or 8 points can provide minimal solution for relative camera pose, accurate result relies on more points or multiple chessboard orientations since the corner detection suffers from noise. The main contributions of this paper are to intrinsic camera calibration for stereo rigs and the extrinsic

Figure 6: An anaglyph image after camera alignment, best viewed with Red (left)/Cyan (right) glasses.

Figure 7: SIFT feature matchings that have less than 1 pixel vertical offset.

calibration implementation is preliminary. We plan to improve the accuracy and robustness of the system in future work.

Figure 6 shows an anaglyph image captured by our system. In the experiment, we use SIFT feature [18] matchings to estimate the vertical offset and verify the alignment guidance provided by our system. We first detect the SIFT features in the left and right undistorted images, then match them by estimating the fundamental matrix in a RANSAC [11] loop. For figure 6, we obtained 381 matching pairs with an average vertical offset of 1.84 pixels. Among them, there are 121 matching pairs (shown in Figure 7) whose vertical offset is less than 1 pixel.

6. Conclusion and discussion

In this paper, we propose a low cost, fast and accurate high-definition camera calibration system. The system can provide the relative camera orientation at an interactive rate. The novel design of the marker chessboard, the key frame selection algorithm and the bundle adjustment enable our system to automatically obtain accurate camera intrinsic parameters in a short time. Also our extrinsic calibration system provides both 2D and 3D feedback to help operators set the stereo camera to a desired orientation.

Though the marker chessboard shows good performance compared to the traditional chessboard, there are two aspects that a user should pay attention to. The first is that the homography mapping may fail in the case of images with very large distortion, *e.g.* the images taken from

fisheye cameras. The second is that the whole chessboard should be captured without any occlusion, otherwise the corner position can not be estimated correctly. In practice this is not an extreme constraint as the operator places the target in the scene. Alternatively, variants of the intrinsic estimation technique could be developed that are robust to occlusion.

There are many applications that can benefit from our research directly. For example, the chessboard can have more inside corners, which will make the calibration more accurate. And we can use multiple marker chessboards in the 3D calibration target by changing the marker code. Also, the online intrinsic calibration becomes feasible because the corner detection is faster than the camera frame rate.

There are several improvements or extensions that could be made to our system. First, the camera model now used does not consider the influence of rig mirror which reflects or refracts light to the cameras. Second, the frame rate of extrinsic calibration needs to be increased for faster response. Last but not least, we plan to use general feature points instead of chessboard corners to get the relative camera pose in the future.

Acknowledgements

The support of the Ontario Media Development Corporation and the Ontario Centres of Excellence is greatly appreciated. The authors also would like to thank Ali Kazimi for loan of equipment used in the experiments.

References

[1] Dashwood stereo 3d chart. http://www.dashwood3d.com/3dchart.php. 1

[2] A. Albarelli, E. Rodolà, and A. Torsello. Robust camera calibration using inaccurate targets. In *Proceedings of BMVC*, pages 16.1–16.10, 2010. 2, 6

[3] R. S. Allison, I. P. Howard, and X. Fang. Depth selectivity of vertical fusional mechanisms. *Vision Research*, 40(21):2985–2998, 2000. 1

[4] S. Arya, D. M. Mount, N. S. Netanyahu, R. Silverman, and A. Y. Wu. An optimal algorithm for approximate nearest neighbor searching fixed dimensions. *J. ACM*, 45(6):891–923, 1998. 5

[5] J. Y. Bouguet. Camera calibration toolbox for matlab. http://www.vision.caltech.edu/bouguetj/calib_doc/. 2, 3

[6] A. Bradski. *Learning OpenCV*. O'Reilly Media, 2008. 3

[7] D. C. Brown. Close-range camera calibration. *Photogrammetric Enginnering*, 37(8):855–866, 1971. 2

[8] D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(5):603–619, 2002. 5

[9] A. Datta, J. Kim, and T. Kanade. Accurate camera calibration using iterative refinement of control points. In *Workshop on Visual Surveillance (VS)*, 2009. 2

[10] M. Fiala and C. Shu. Self-identifying patterns for plane-based camera calibration. *Mach. Vision Appl.*, 19(4):209–216, 2008. 3

[11] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24(6):381–395, 1981. 7

[12] K. Fukuda, L. M. Wilcox, R. S. Allison, and I. P. Howard. A reevaluation of the tolerance to vertical misalignment in stereopsis. *Journal of Vision*, 9(2), 2009. 1

[13] R. I. Hartley. In defense of the eight-point algorithm. *IEEE Trans. Pattern Anal. Mach. Intell.*, 19(6):580–593, 1997. 1, 3

[14] J. Heikkilä. Geometric camera calibration using circular control points. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(10):1066–1077, 2000. 1

[15] S. Heinzle, P. Greisen, D. Gallup, C. Chen, D. Saner, A. Smolic, A. Burg, W. Matusik, and M. Gross. Computational stereo camera system with programmable control loop. *ACM Trans. Graph.*, 30(4):94:1–94:10, 2011. 6

[16] D. Kane, R. T. Held, and M. S. Banks. Visual discomfort with stereo 3D displays when the head is not upright. *Proceedings of SPIE*, 8288(1):828814:1–828814:10, 2012. 1

[17] F. L. Kooi and M. Lucassen. Visual comfort of binocular and 3D displays. *Proceedings of SPIE*, 4299(1):586–592, 2001. 1

[18] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004. 7

[19] D. Nistér. An efficient solution to the five-point relative pose problem. *IEEE Trans. Pattern Anal. Mach. Intell.*, 26(6):756–777, 2004. 1, 3

[20] R. M. Salinas. Aruco: a minimal library for augmented reality applications based on opencv. http://www.uco.es/investiga/grupos/ava/node/26, 2012. 4

[21] W. Sun and R. Cooperstock. An empirical evaluation of factors influencing camera calibration accuracy using three publicly available techniques. *Mach. Vision Appl.*, 17(1):51–67, 2006. 6

[22] R. Tsai. A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses. *IEEE Journal on Robotics and Automation*, 3(4):323–344, 1987. 1

[23] Z. Zhang. A flexible new technique for camera calibration. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(11):1330–1334, 2000. 1, 3, 5