

AUTOMATIC DETECTION AND LABELLING OF PHOTOGRAMMETRIC CONTROL POINTS IN A CALIBRATION TEST FIELD

D. Jarron¹, M. Shahbazi¹, D. Lichti¹, R. Radovanovic²

¹ Dept. of Geomatics Engineering, University of Calgary, T2N 1N4 Calgary AB, Canada - (dmjarron, mozhdeh.shahbazi, ddlichti)@ucalgary.ca

² McElhanney Geomatics Engineering Ltd., T2G 0Y4 Calgary, AB, Canada - rradovanovic@mcclhanney.net

Commission II, WG II/1

KEY WORDS: camera calibration, automatic target detection, automatic target labelling

ABSTRACT:

In this work, a new method is developed for the automatic and accurate detection and labelling of signalized, un-coded circular targets for the purpose of automated camera calibration in a test field. The only requirements of this method are the approximate height of the camera, an approximate range of orientations of the camera, and the object-space coordinates of the targets. In each image, circular targets are detected using adaptive thresholding and robust ellipse fitting. Labelling of those targets is performed next. First, the exterior orientation parameters of the image are estimated using a one-point pose-estimation approach, where a list of possible orientation and target labels are used, along with height, to calculate the camera position. The estimated position and orientation of the camera combined with the interior orientation parameters (IOPs) are then used to back-project the known object-space coordinates of the targets into the image space. These targets are then matched against the targets detected in the image, and the list entry with the best fit is chosen as the solution. This resolves both the detection and labelling of the targets, without the need for any coded targets or their associated software packages, and each image is solved independently allowing for parallel processing. This process accurately labels 92-97% of images, with average accuracy rates of 97% or better, and average completeness rates of 70-95% in imagery from the three cameras tested. The cameras were calibrated using observations from the detection and labelling process, which resulted in sub-pixel root mean square (RMS) values determined for the pixel space residuals.

1. INTRODUCTION

1.1 Literature Review

Automatic detection and labelling of imagery is an essential step in the process of calibrating a camera or a multi-camera system. Calibration requires a large amount of imagery, and manual target detection and labelling are time consuming. Coded targets are currently the standard type of targets applied in photogrammetric calibration. There are many styles of coded targets such as: unique patterns of circles on a target, where the layout of the circles encodes the target (Ahn et al., 2001; Hattori et al., 2002; Knyaz and Sibiryakov, 1998), centripetal encoding where unique sections of a disk surrounding a central circular target are used to encode the target (Niederöst and Maas, 1997; Schneider et al., 1992), algorithms that augment these styles by using colour information (Cronk et al., 2006; Moriyama et al., 2008). A newer algorithm developed by Shortis and Seager (2014) works similarly to the centripetal targets, using straight lines on the boundaries of the targets and converting them to the polar space to encode them. It can only encode 124 possible targets and uses low-cost materials to make this system easy to implement.

Systems that use coded targets need to reconstruct the scene in 3D for onsite self-calibration. This can be done by first resecting an External Orientation (EO) device or Autobar (a device that has multiple distinguishable targets with known relative coordinates) to determine model space coordinates of the cameras. Then, the coordinates of coded targets that appear in multiple images are determined by spatial intersection. Alternatively, the process is done simultaneously with a bundle adjustment (Fraser, 1997; Ganci and Handley, 1998). This process is repeated for each

image until all are processed. This means that the detection and labelling of the targets are partially dependent on the accurate determination of previous images in the data set. Newer methods avoid EO devices in favour of more robust procedures such as simulating camera positions through the Monte Carlo method (Shortis and Seager, 2014).

1.2 Contributions

The method proposed in this paper takes a different approach to automatic detection and labelling of targets in the sense that no coded targets are required. The targets used in the experiments are simple, (non-retroreflective) white circular targets on black backing. This method is designed for a test-field calibration, not an onsite calibration like many coded target systems, and thus requires the object-space coordinates of the targets to be known through some method of surveying, and knowledge of the approximate height of the camera at each image. This method is generalized for multiple applications. One of these applications is metrology, where images are typically under-exposed and utilizes retro-reflective targets (Shortis and Seager, 2014). In this scenario there is very little texture in the imaging scene, precluding the use of structure-from-motion techniques. This method requires more external information than coded targets but is designed primarily for test-field calibration where this is not an issue. This method requires approximate knowledge of the camera's height, which can be simply achieved via a tape measure. With the proposed approach, each image is processed independently of other images, which makes it suitable for parallel computing. Additionally, there is no strict requirement regarding the size of the simple targets; they can be much smaller than that of a typical coded target. For example, the targets used in Shortis and Seager (2014) need to be seven times larger than a

standard target. This large size can be inconvenient when using a low-resolution camera. For instance, the required size for a coded target would become greater than 1 m in diameter in some of the experiments performed in this paper. Moreover, many current coded target systems require expensive product licenses to operate the system such as V-STARS or iWitness (Cronk et al., 2006; Ganci and Handley, 1998), which this method avoids.

2. METHODOLOGY

2.1 Target Detection and Ellipse Fitting

The proposed target-detection approach first converts the true-colour imagery to greyscale. The image is then converted to a binary image using a threshold determined using adaptive local thresholding (Figure 1b). The threshold is determined for each pixel based on the weighted average intensity of its neighbourhood. The neighbourhood size used is the maximum possible diameter of the target in pixels. This thresholding technique is useful for poorly lit, or unequally exposed images where a global threshold would allow many targets to fall below the threshold and be left undetected.

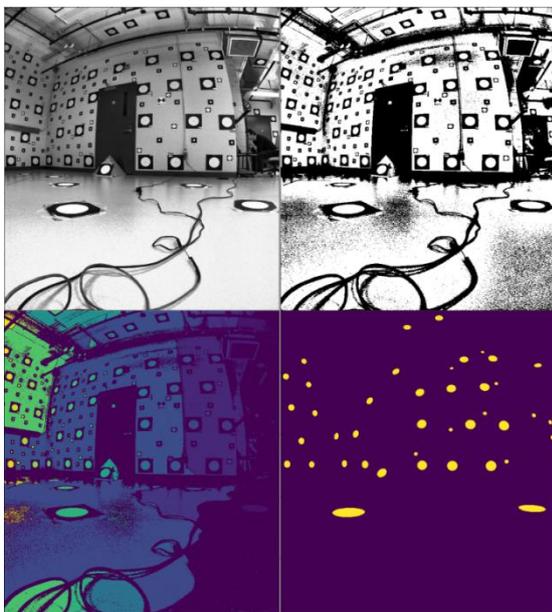


Figure 1a) Top left, original image. 1b) Top right, binary image after threshold. 1c) Bottom left, image after connected component analysis, with each colour designating a separate component. 1d) Bottom right, components that were successfully fit to an ellipse

Connected component analysis (Figure 1c) is used to label each separate region in the binary image and find the contour/edges of the region (Burger and Burge, 2009). The edge points are passed into a 5-point random sample consensus (RANSAC) algorithm to fit an ellipse to the edges using the general equation for a conic section (Rosin, 1999). The approximate parameters from the RANSAC algorithm are passed into a least squares algorithm for further refining the estimated centre of the target. A series of tests are utilized to ensure the connected component is an elliptical shape. First, the ratio between the semi-minor and semi-major axes is examined. A ratio greater than 0.3 allows distinguishing circular targets from stretched shapes. This also removes targets imaged with very high perspective distortion since they fit poorly in bundle adjustments. Second, the target data must fit both the RANSAC and LS algorithm models to some pre-defined

threshold, which is determined by accuracy requirements and image quality. Third, the semi-major and semi-minor axes must be within the bounds set by the user for the image set. The ellipse axis bounds are approximate estimates determined by the user which are the largest and smallest possible radii for the semi-minor and semi-major ellipses of the circular targets in the images. This can be performed for a camera once for a full set of imagery in a calibration test field. These estimates in the experiments performed were based on the observed sizes of the ellipses in the imagery. A future improvement to the algorithm would be to base the radii on projections of the circular target size into the image at the approximate maximum and minimum range of the test field. The centres of the fitted ellipses are stored to be matched against the list of labels in the labelling process. Figure 1 provides an example of the target detection process. The connected components to which ellipses are fit successfully are shown in Figure 1d. Figure 2 shows a close up of the ellipses fit to the targets.

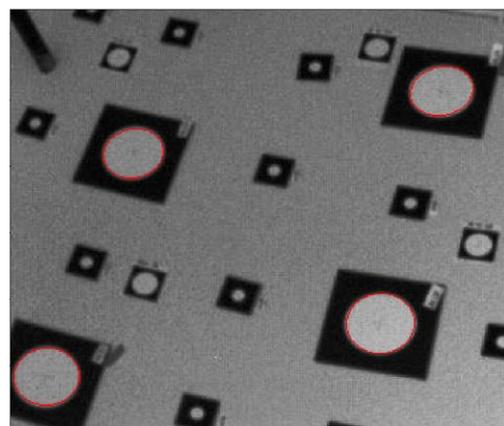


Figure 2. Ellipses fit to the targets

2.2 Target Labelling

The targets have been detected but, without the context provided by a label, not enough information has been gathered to generate observations for a camera calibration bundle adjustment. While images are taken, the approximate height of each image is recorded. A reasonable set of orientation constraints are determined for each image or sets of similar images by the user. This process could be replaced by an inexpensive IMU device to measure approximate rotation angles, which would be converted to approximate external orientation angle parameters. This requires knowledge of the 3D object space coordinates of the target field so that the rotation matrix between the object frame and the camera frame can be approximated. The level of uncertainty of the camera's orientation will determine how tight the constraints on the approximate angles are. For example, if a camera is set level, the two tilt angles can be constrained to a specific angle. If the camera was only approximately level, then a looser constraint can be used, such as a 10° buffer around the tilt angles of a level camera. The angles that lie within the constraints are discretized based on an angular resolution chosen by the user. The smaller the step used, the longer the image labelling takes, and vice-versa.

The target centres determined from the detection step are now transformed by correcting the distortion of the camera determined by Equations 1 and 2. These are referred to as the rectified target positions. The distortion parameters do not need to be known accurately. Thus, they can come from approximate prior knowledge or previous calibrations of the camera. The

boundary pixels of the image are also rectified, to create a new bounding area for possible target positions on the image since radial distortions will change the size and shape of the entire image when corrected. These bounds are used later to determine whether a back-projected target lies within the image or not.

$$d_x = x * r^2 * (k_1 + r^2(k_2 + r^2k_3)) \quad (1)$$

$$d_y = y * r^2 * (k_1 + r^2(k_2 + r^2k_3)) \quad (2)$$

where r = target radial distance from the principal point
 x, y = image coordinates
 k_1, k_2, k_3 = radial lens distortion coefficients
 d_x, d_y = radial lens distortion along x and y axes

The radial distance from the centre of the image to each of the non-rectified positions is determined, and four targets with the largest radial distance, one from each quadrant of the image, are chosen as seed targets. A list of all possible combinations of camera height, seed targets, seed target labels, and discrete orientation angles is generated.

Equation 3 describes the intrinsic calibration matrix, which transforms between sensor space, centred at the top left of the image with pixel units, and camera space centred at the perspective centre expressed in millimetre units. Equation 4 projects object space coordinates into sensor space. Equations 5 – 7 show how Equation 4 is rearranged to solve for the object space coordinates of the camera, using the sensor space observations, rotation angles, IOPs and camera height. Each seed target's rectified position is tested against every target label in the target field, using Equations 7. For each image, each seed point is tested against every possible target label for each combination of possible discrete angles. For each entry in this list, the X and Y camera position is calculated (Equation 7), giving a complete set of exterior and interior orientation parameters. These parameters are used to back-project targets into the image to find which set of exterior orientation parameters (EOPs) is closest to the true EOPs. It is important that the seed target and camera do not have the same, or very similar heights, as this generates a critical configuration in Equation 7, leading to incorrect EOPs, which degrades the back-projection quality. To avoid this problem, multiple seed points are chosen from the different quadrants in the image, reducing the likelihood of such a critical configuration.

$$\mathbf{K} = \begin{bmatrix} \frac{1}{p_s} & 0 & \frac{(-\frac{C_n}{2} + 0.5) * p_s - x_p}{p_s * f} \\ 0 & -\frac{1}{p_s} & \frac{-(-\frac{R_n}{2} + 0.5) * p_s - y_p}{p_s * f} \\ 0 & 0 & -\frac{1}{f} \end{bmatrix} \quad (3)$$

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \mathbf{K} [\mathbf{R}_g^i \quad -\mathbf{R}_g^i \mathbf{C}] \begin{bmatrix} X_{tar} \\ Y_{tar} \\ Z_{tar} \\ 1 \end{bmatrix} \quad (4)$$

$$\mathbf{R}_i^g \mathbf{K}^{-1} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} X_{tar} - C_x \\ Y_{tar} - C_y \\ Z_{tar} - C_z \end{bmatrix} \quad (5)$$

$$\mathbf{A} = \mathbf{R}_i^g \mathbf{K}^{-1} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (6)$$

$$\begin{bmatrix} C_x \\ C_y \end{bmatrix} = \begin{bmatrix} X_{tar} - A_x \left(\frac{Z_{tar} - B_z}{A_z} \right) \\ Y_{tar} - A_y \left(\frac{Z_{tar} - B_z}{A_z} \right) \end{bmatrix} \quad (7)$$

where \mathbf{R}_g^i = rotation matrix from object to image space
 x, y = rectified image coordinates of the target centre in pixel space
 \mathbf{C} = object space coordinates of the perspective centre
 \mathbf{K} = intrinsic calibration matrix of the camera
 C_n, R_n = width and height of the image in pixels
 $X_{tar}, Y_{tar}, Z_{tar}$ = object space coordinates of the target label
 p_s = pixel size
 f = principal distance of the camera

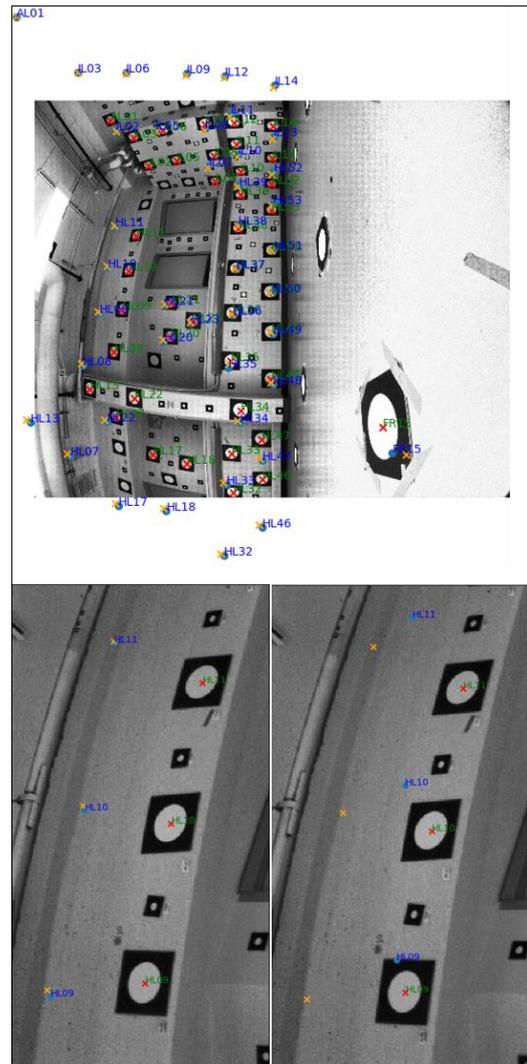


Figure 3. a) Top: Example of matching back-projected targets to rectified target positions. Red x's are detected target centres. Blue circles are back-projected target centres, orange x's are rectified detected target centres. Labels are included to show matching back projected targets and their matched target centres. b) Bottom left: close-up of matching results from list entry with the best fitness measure. c) Bottom right: close-up of matching results from list entry with the 10th best fitness measure

The EOPs, combined with the nominal IOPs, allow for all points in the target field to be back projected into the image. The back-projection used is for a perspective camera. A future improvement to the algorithm would be to have an alternate back-projection to use with fish-eye lens cameras. Any targets that are back projected behind the camera or outside the bound of the

rectified image are discarded. The back-projected target positions are compared to the rectified positions of the detected target centres, an example of which can be seen in Figure 3a. A simple process is used where the closest rectified detected target position to each back projected target is found. The mean and standard deviation of the distances between the detected and back-projected targets is determined, and any matches that are outside the 95% confidence interval are removed. This procedure is performed for each set of EOPs, and the list of labelled targets along with the mean distance between the detected and back-projected target positions are stored. The number of back projections that match to targets, i.e. the number of successfully labelled targets, must be greater than some pre-defined percentage of the number of detected targets. This ratio is determined based on the outlier likelihood during the target detection stage. Using the fitness measure defined in Equation 8, the entry list is sorted such that an entry with a low mean distance between detected and back-projected targets and a large number of matches is at the top of the list. Figure 3b shows an example of the entry with the smallest fitness measure and Figure 3c shows the entry with the 10th smallest fitness measure. It can be observed that the back-projected target positions and the rectified detected target centres in Figure 3b are much closer to one another compared to Figure 3c. The fitness measure of Figure 3b is 0.317 while the fitness measure of Figure 3c is 1.68, and their rotation angles differ by 10°.

$$a = \mu/l \quad (8)$$

where a = fitness measure for determining best EOPs combination
 μ = mean distance between detected and back-projected targets
 l = number of matches between detected and back-projected targets

Once all the detected targets have a label, the algorithm moves on to the next image. Each image is processed independently, meaning that the software can utilize multi-processor parallel computing techniques to detect and label multiple images simultaneously. The number of parallel processes is dependent

To solve this problem, the algorithm can be iteratively run, such that the first iterations' observations are used in a camera calibration bundle adjustment to determine distortion parameter estimates. These estimates can be used in the second iteration of labelling to compensate for the distortion, allowing for more accurate labelling, and potentially further iterations until results are satisfactory.

3. EXPERIMENTS

The algorithm described in the methodology section was applied to three different image sets, captured by three different cameras. The three cameras used were a Ladybug5 (FLIR Systems, Oregon, USA), a GoPro Hero5, and a Canon Rebel T3i. These camera systems were used in the same calibration space, which is a room of dimensions approximately 11 m x 11 m x 4 m with a total of 232 targets of 125 mm radius made from 4 mm thick BubbleX plastic, and 59 paper targets of 40 mm radius, which cover walls, ceiling, and floor of the calibration space. Smaller targets also exist in the calibration space, but they are ignored in these experiments. A panoramic view of the calibration space can be seen in Figure 4, a plan view with the camera imaging locations is shown in Figure 5. Before data acquisition with the cameras was performed, the targets in the calibration space were imaged with a laser scanner, and the centre coordinates of each target were extracted by fitting a circle to the edge points.

| Camera | Principal Distance (mm) | Sensor Size (pixels x pixels) |
|-----------------|-------------------------|-------------------------------|
| Ladybug5 | 4.4 | 2048 x 2448 |
| GoPro Hero5 | 3 | 4000 x 3000 |
| Canon Rebel T3i | 50 | 5184 x 3456 |

Table 1. Specifications of the cameras used in the experiments

The Ladybug5 is an omnidirectional camera system composed of 6 wide angle lens cameras with sensor sizes of 2048x2448 and principal distances of 4.4 mm. From 6 different positions, and many orientations in the room, 336 images were taken in total, from the 6 cameras. The angular resolution used to define the



Figure 4. Panoramic view of calibration space

on the computer being used. Once the labelling is completed for all images, the approximate EOPs and the image space observations are generated and can be used to run a self-calibration bundle adjustment. This can be used to estimate the IOPs, distortion parameters, and EOPs of the camera. For cameras with large amounts of unknown radial lens distortion, targets at the perimeter of an image are unlikely to be labelled.

EOP entry list was 3°, and tilt angles of the camera were constrained to a single angle since the images were captured in levelled portrait rotations. The GoPro Hero5 is a very-wide-angle lens camera with a 3mm principal distance and a sensor size of 4000x3000 pixels. This camera was used to take 48 images, from a variety of angles and heights, and using a smaller section of the calibration space including only 160 targets. The angular

resolution used was 5° and tilt angle constraints assume the camera was levelled to $\pm 20^\circ$.

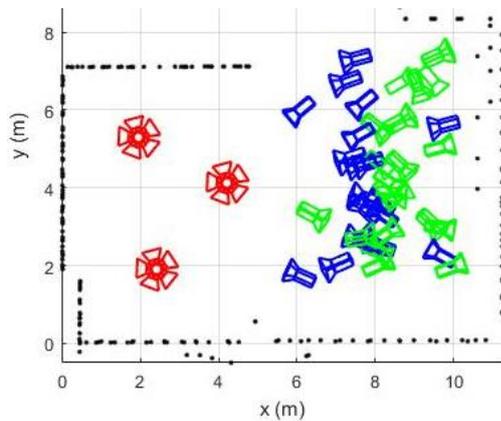


Figure 5. Plan view of the calibration space and camera network configurations in the 3 experiments. Red cameras are the 6 Ladybug5 cameras, green is the GoPro Hero5, blue is the Canon Rebel T3i, the black dots are the targets¹.

For the final experiment using the Canon Rebel T3i. The Canon has a sensor size of 5184×3456 pixels and was fitted with a 50 mm principal distance lens. This camera was used to take 29 images from various heights and orientations, capturing 102 unique targets in the calibration space. The angular resolution used was 5° and tilt angle constraints were the level angle $\pm 10^\circ$.

For each experiment, the height of the camera from the floor was recorded at each imaging location. The images were then run through the algorithm to detect and label all targets within them. To quantify the quality of the algorithm results, each target in the images was manually examined to determine the accuracy and completeness of the labelling. In addition, the observations generated were put through a bundle adjustment camera calibration to determine the residual fit of the pixel space observations, and the estimated object-space target position

precision of the detected targets. The time to complete detection and labelling for each data set is also considered. In addition, the required accuracy of the height determination of the camera was tested. To this end, the algorithm was run with deliberately and increasingly inaccurate height measurements to determine the approximate accuracy required to find a correct labelling solution.

4. RESULTS

4.1 Labelling Accuracy

The first experiment, performed with the Ladybug5, accurately detected and labelled 318 out of 336 images, or 95%. The 318 accurately labelled images had an average labelling accuracy of 99.4% with a standard deviation of 5.5%. These images have an average completeness of 77.3% with a standard deviation of 14.3%. The average accuracy is the average percentage of targets in accurately labelled images that were correctly identified. Examples of labelled images can be seen in Figure 6. The RMS of the residuals of the pixel-space observations from the bundle adjustment was 0.25 pixels, and the mean estimated object-space target position precision was 1.0 mm. The object-space target position precision is calculated based on Equation 8.

$$\sigma_p = \sqrt{\sigma_x^2 + \sigma_y^2 + \sigma_z^2} \quad (8)$$

The second experiment, with the GoPro Hero5, accurately detected and labelled 44 out of 48 images, or 92%. The 44 accurately labelled images had an average labelling accuracy of 97.2% with a standard deviation of 5.5%. These images have an average completeness of 73.1% with a standard deviation of 16.3%. The RMS of the residuals of the pixel-space observations from the bundle adjustment was 0.38 pixels, and the mean estimated object-space target position precision was 1.4 mm. Examples of labelled images can be seen in Figure 7. This experiment had the lowest accuracy and completeness (Table 2). One possible reason for this is that a collinearity model with distortions was used, and for a camera with such large amounts of radial distortion, a fisheye model potentially yield

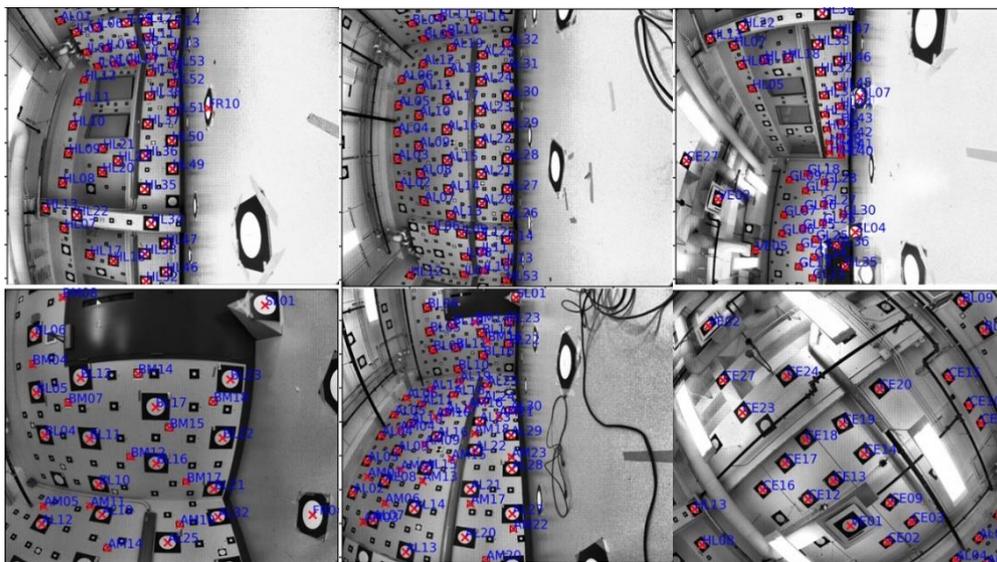


Figure 6. Selection of images from the Ladybug5 with detected and labelled targets

¹ Ceiling and floor targets not included for sake of visibility

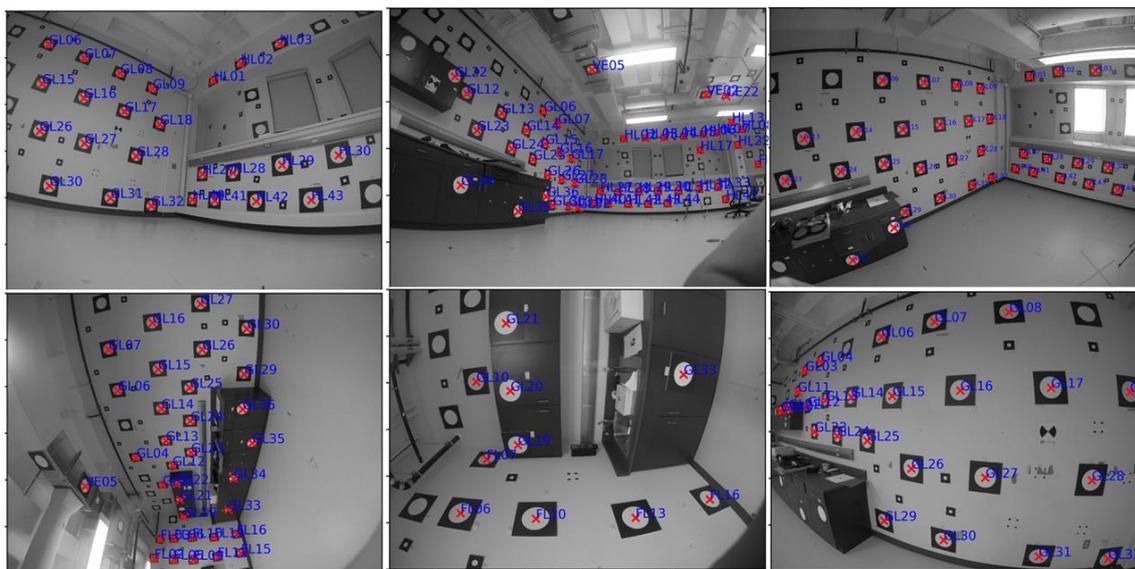


Figure 7. Selection of images from GoPro Hero5 with detected and labelled targets



Figure 8. Selection of images from Canon Rebel T3i with detected and labelled targets

better accuracy, especially at the image edges, where radial distortion is at its highest

| Camera | Ladybug5 | GoPro Hero5 | Canon Rebel T3i |
|---------------------------------|----------|-------------|-----------------|
| Average Accuracy | 99.4% | 97.2% | 98.9% |
| Accuracy Standard Deviation | 4.3% | 5.5% | 2.7% |
| Average Completeness | 77.3% | 73.1% | 96.8% |
| Completeness Standard Deviation | 14.3% | 16.3% | 4.5% |

Table 2. Average accuracy and completeness of labelling for all 3 experiments.

The third experiment with the Canon Rebel T3i accurately labelled 28 out of the 29 images, or 97%. The 28 accurately labelled images had an average labelling accuracy of 98.9% with a standard deviation of 2.7%. These images have an average completeness of 96.8% with a standard deviation of 4.5%. The RMS of the residuals of the pixel-space observations from the bundle adjustment was 0.39 pixels, and the mean estimated object-space target position precision was 1.9 mm. Examples of labelled images can be seen in Figure 8. The long principal distance of the Canon meant that the images had minimal radial lens distortion and had a relatively narrow Field of View (FOV).

This likely contributes to the very high completeness percentage of this camera compared to the other two cameras. The images tended to observe few targets, and the targets tended to be of similar sizes and perspective distortions. This made both the detection and labelling portions of the algorithm more effective, and the smaller distortions meant that targets at the edges of the images were easier to label.

4.2 Relative Execution times of the Algorithm

As part of the evaluation of the algorithm, the time required to process large datasets is considered. The computer used to process the data described in the experiments section has an Intel® Core™ i7-8700k 3.7GHz CPU with 6 cores, and 64 GB of RAM and a 64-bit OS. The language used to implement the algorithm was Python 3, with the Spyder integrated development environment. Using only one processor to run the algorithm, the time required to process the image sets is many hours, as can be seen in Table 3. However, when utilizing multiple cores of the processor, the time required to process large sets of data is drastically reduced. It is also important to note that the number of list entries per image is the main factor of the time required per image (number of list entries defined in Equation 9). The Ladybug5 dataset has fewer list entries per image (due to its mounting apparatus, all the imagery acquired was approximately level, allowing for tilt angles constrained to one discrete angle) which means that even though it has almost 7 times the number of images of the GoPro dataset, it only took twice as long to complete. Many of the GoPro images were taken with off-level

orientations. This meant that a wider range of angles needed to be tested to assure accurate results, as indicated by the GoPro having the largest number of list entries. The difference in the size of list entries are primarily generated by the number of angles that need to be tested. The number of combinations of rotations and target labels is described in Equation 9.

$$L = n_{\omega} * n_{\kappa} * n_{\phi} * n_{labels} * n_{seed} \quad (9)$$

where L = number of entries in the list
 n_{ω} = number of discrete omega angles tested
 n_{κ} = number of discrete kappa angles tested
 n_{ϕ} = number of discrete phi angles tested
 n_{labels} = number of targets in the test field
 n_{seed} = number of seed points used

The time difference between the algorithm when run using parallel processing, and when not, emphasizes the importance of parallel computing as it drastically reduces the runtime of the algorithm.

| Camera | L | # of images | Runtime on 1 processor (H:M:S) | Runtime on 6 processors (H:M:S) |
|-----------------|---------|-------------|--------------------------------|---------------------------------|
| Ladybug5 | 139680 | 336 | 10:42:23 | 1:59:13 |
| GoPro Hero5 | 3732480 | 50 | 7:15:16 | 1:09:12 |
| Canon Rebel T3i | 1454400 | 30 | 4:03:11 | 0:42:35 |

Table 3. Run-time of experiments using different numbers of processors

4.3 Required Height Accuracy

Using data from the Ladybug5, the accuracy of the approximate camera height required for the labelling procedure was tested. As can be seen in Figure 9, the accuracy of the solution is the same at both 0 and 5 cm of error, with all the targets accurately labelled. At 10 cm of error, there is one incorrectly labelled target (highlighted by the yellow circle in Figure 9), and past 15 cm of error a correct solution is no longer found.

This can be observed in Figure 9 as all the target labels at a height error of 15 cm are different from the previous, correctly labelled images within Figure 9 at smaller height errors. The position where these images were taken from was approximately 4 m from the wall with the targets. A similar simulation was performed with Ladybug5 imagery taken 2.5 m from the targets. In this simulation, the labelling was accurate when the height error was within 5 cm of the true height, and when the height error was 10 cm, the labelling was entirely incorrect as can be seen in Figure 10. This demonstrates that the height used in the algorithm does not need to be more accurate than 5 cm, and simply using a tape measure, or other approximate should be effective in determining the height. It also demonstrates that the closer the camera to the targets, the more refined the approximate height must be. This is also true of the chosen angular resolution for EOPs. For camera positions closer to the targets, a finer angular resolution is required.

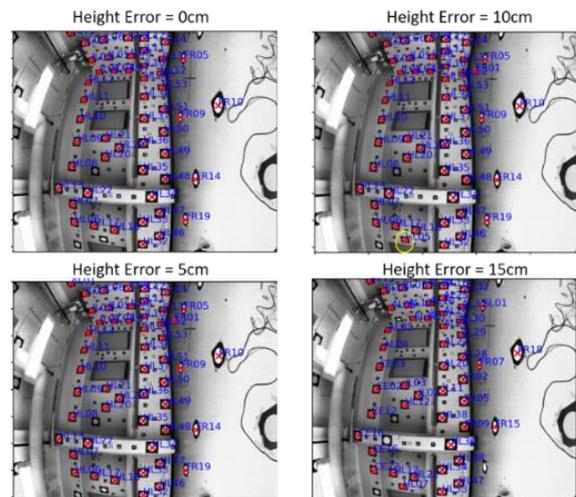


Figure 9. Height measurement error simulation for targets at 4 metres

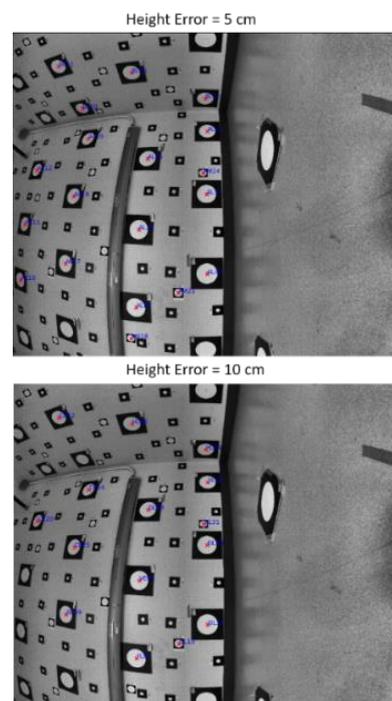


Figure 10. Height measurement error simulation for targets at 2.5 metres

5. DISCUSSION

When an accurate set of EOPs is used for the back projection, the points usually match very accurately, with very few incorrect matches being found. This is exemplified in the very high accuracy of the labelling in all 3 experiments (Table 2). A future improvement to the algorithm would be to automatically identify failed images and remove them from the dataset automatically. Another future improvement is to determine if a labelled target is behind any other targets. Due to the complex shape of the calibration space, (Figure 5) it is possible to have targets on a surface within the FOV of the camera but occluded by a wall or other surface, meaning that the detected target might be labelled as a target that is behind it in reality.



Figure 11. Section of an image from GoPro Hero5, highlighting incorrect labelling caused by radial lens distortion and perspective distortion clustering targets in the image

Determining this using the surface normals of the targets and ray intersection could be implemented to ensure no targets are mislabelled as targets that are behind them. There are other possible situations which could negatively affect the algorithm's ability to determine an accurate labelling solution. A target field that is symmetrical will introduce a high likelihood for incorrect solutions, due to the algorithm relying on the target space to have a distinct geometric configuration in the 2D image, from different viewpoints in the 3D object-space. A similar problem could occur if a target field has repetitive patterns, such as a fixed grid placement that targets adhere to as this can cause the algorithm to mistake those targets for other targets that hold a similar pattern.

A problem that occurred in processing the imagery when the image was taken from a view point that causes a very few targets to appear in the image. In this case, there is a high likelihood of finding an incorrect labelling solution because there are many possible combinations of angles and positions that could approximately fit a small number of targets (less than 8). The configuration that leads to the largest number of missed and incorrectly identified targets are images with high perspective distortion. For instance, viewing a surface with a high volume of targets from an extreme angle clusters many of the targets together in the image space. This can confuse the labelling process because as the targets become closer together in the image space, the associated increase in the fitness measure of mislabelling a target as one of its neighbours becomes smaller. This problem is exacerbated by cameras with high amounts of positive radial lens distortion, which causes targets at the edge of images to distort away from the centre of the image towards the edges, such as in Figure 11.

6. CONCLUSION

In this work, a new method has been developed for automatic and accurate detection and labelling of circular photogrammetry targets for the purpose of automated camera calibration. This new method works using an approximate camera height and known object space coordinates of the targets to calibrate a camera or camera system in a test field environment. The detection and labelling of targets is automated and allows each image to be processed independently of one another. This independent processing means that large batches of images can be resolved by processing them in parallel. This process is very accurate with 92%, or greater, of the images being accurately labelled, and

greater than 97% average accuracy for all cameras tested. This is true even for cameras with high levels of radial lens distortion. The detected and labelled targets serve as observations for camera-calibration bundle adjustment. When used to calibrate the cameras tested, sub-pixel image-space residuals were achieved.

REFERENCES

- Ahn, S.J., Rauh, W., Kim, S.I., 2001. Circular coded target for automation of optical 3d-measurement and camera calibration. *Int. J. Patt. Recogn. Artif. Intell.* 15, 905–919. <https://doi.org/10.1142/S0218001401001222>
- Burger, W., Burge, M.J., 2009. Principles of Digital Image Processing: Core Algorithms, Undergraduate Topics in Computer Science. Springer-Verlag, London, pp. 5-26
- Cronk, S., Fraser, C., Hanley, H., 2006. Automated metric calibration of colour digital cameras. *The Photogrammetric Record* 21, 355–372. <https://doi.org/10.1111/j.1477-9730.2006.00380.x>
- Fraser, C.S., 1997. Digital camera self-calibration. *ISPRS Journal of Photogrammetry and Remote Sensing* 52, 149–159. [https://doi.org/10.1016/S0924-2716\(97\)00005-1](https://doi.org/10.1016/S0924-2716(97)00005-1)
- Ganci, G., Handley, H., 1998. Automation in Videogrammetry. *International Archives of Photogrammetry and Remote Sensing*, XXXII-5, 53-58
- Hattori, S., Akimoto, K., Fraser, C., Imoto, H., 2002. Automated Procedures with Coded Targets in Industrial Vision Metrology. *Photogrammetric Engineering and Remote Sensing* 68, 441-446
- Knyaz, V.A., Sibiryakov, R.V., 1998. The Development Of New Coded Targets For Automated Point Identification And Non -, in: 3D Surface Measurements, *International Archives of Photogrammetry and Remote Sensing*, Vol. XXXII-5, 80–85.
- Moriyama, T., N. Kochi, N., Yamada, M., Fukaya, N., 2008. Automatic Target-Identification with the Color-Coded-Targets. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XXXVII-5B, 39-44
- Niederöst, M., Maas, H.-G., 1997. Automatic deformation measurement with a digital still video camera. Presented at the Optical 3-D Measurement Techniques IV, ETH Zurich, Institute of Geodesy and Photogrammetry. <https://doi.org/10.3929/ethz-a-004332760>
- Rosin, P.L., 1999. Further Five-Point Fit Ellipse Fitting. *Graphical Models and Image Processing* 61-5, 245–259. <https://doi.org/10.1006/gmip.1999.0500>
- Schneider, D.C.-T., Sinnreich, K., Photogrammetry, A.-I., 1992. Optical 3-D-Measurement Systems For Quality Control In Industry. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XXIX-B5, 56–59.
- Shortis, M.R., Seager, J.W., 2014. A practical target recognition system for close range photogrammetry. *The Photogrammetric Record* 29, 337–355. <https://doi.org/10.1111/phor.12070>