ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume II-5, 2014
ISPRS Technical Commission V Symposium, 23 – 25 June 2014, Riva del Garda, Italy

AUTOMATED TARGET-FREE NETWORK ORIENTATION AND CAMERA CALIBRATION

C. Stamatopoulos *, C.S. Fraser

Cooperative Research Centre for Spatial Information
Department of Infrastructure Engineering, University of Melbourne, Victoria 3010, Australia
xstamatopoulos@gmail.com, c.fraser@unimelb.edu.au

Commission V, WG V/1

KEY WORDS: Target-Free Network Orientation, Camera Calibration, Feature-based Matching, Relative Orientation

ABSTRACT:

Automated close-range photogrammetric network orientation and camera calibration has traditionally been associated with the use of coded targets in the object space to allow for an initial relative orientation (RO) and subsequent spatial resection of the images. However, over the last decade, advances coming mainly from the computer vision (CV) community have allowed for fully automated orientation via feature-based matching techniques. There are a number of advantages in such methodologies for various types of applications, as well as for cases where the use of artificial targets might be not possible or preferable, for example when attempting calibration from low-level aerial imagery, as with UAVs, or when calibrating long-focal length lenses where small image scales call for inconveniently large coded targets. While there are now a number of CV-based algorithms for multi-image orientation within narrow-baseline networks, with accompanying open-source software, from a photogrammetric standpoint the results are typically disappointing as the metric integrity of the resulting models is generally poor, or even unknown. The objective addressed in this paper is target-free automatic multi-image orientation, maintaining metric integrity, within networks that incorporate wide-baseline imagery. The focus is on both the development of a methodology that overcomes the shortcomings that can be present in current CV algorithms, and on the photogrammetric priorities and requirements that exist in current processing pipelines. This paper also reports on the application of the proposed methodology to automated target-free camera self-calibration and discusses the process via practical examples.

1. INTRODUCTION

Automated generation of high density point clouds from images has nowadays become a relatively standard procedure. An important step, however, that precedes such 3D reconstruction, is camera calibration and network orientation, which are required for the subsequent step of dense stereo matching. Until recently, there were only a few software packages that fully supported such workflows. However, the advent of UAVs has attracted a lot of attention in both industry and academia and this has resulted in an increase in the availability of software systems for automated orientation and digital surface model generation.

For aerial photogrammetric datasets, the image capture process tends to be more structured. In many cases the interior orientation (IO) is already known and the positional part of the exterior orientation (EO) can be obtained from a GPS device. This allows for a much more flexible way of processing, since already a number of EO parameters values, as well as information with regard to image overlap, is readily available. Nonetheless, even in the case where no information is available, the imagery of aerial photogrammetric datasets is more straightforward to process compared to the typical case of close-range photogrammetric networks that exhibit different viewpoint angles, camera orientations and scales, as well as higher convergence between the images.

Tools for the automated network orientation are available, with the majority emanating mainly from the CV community. In most cases, the process involves the selection of images, with the end result being both a fully oriented multi-image network and a 3D model in the form of either a point cloud or a mesh. These software packages utilise a feature descriptor matching approach for the image point correspondence determination, which is further refined by a RANSAC methodology based either on the fundamental or homography matrices (e.g. Barazzetti, 2011a, Remondino et al., 2006, Abdel-Wahab et al., 2011). This procedure is commonly referred to as Structure from Motion (SfM).

However, SfM approaches are typically subject to accuracy limitations and also to constraints upon image network geometry due to the adoption of linear algorithms. As an example, a common feature of CV algorithms is that they perform optimally on narrow-baseline, low-convergence imagery. Yet, convergent imaging configurations are generally a prerequisite for reliable camera self-calibration and thus any multi-view stereo approach must accommodate wide as well as narrow baselines (image-to-image separation). Thus, in order to become more applicable to close-range photogrammetric measurement tasks, automated target-free network orientation needs to accommodate the multi-image, convergent network configurations that characterise high-accuracy (say sub-pixel) close-range measurement operations. Moreover, it is essential that rigorous quality measures in relation to accuracy, precision and reliability for parameters of both image and object space are generated,. This is generally not the case for CV-based algorithms and associated software tools. In the absence of quality metrics, the most basic being the estimated precision of

* Corresponding Author

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
doi:10.5194/isprsannals-II-5-339-2014

object point coordinate determination, there is limited utility beyond visualisation in CV-based 3D model reconstruction. Nonetheless, the adoption of SfM and multi-view stereo approaches in close-range photogrammetry has been rapid, especially in UAV applications. However, there has been less attention paid to the prospect of utilising these same approaches for stand-alone camera calibration, though the prospect has been recognised (eg Barazzetti et al., 2011b; Stamatopoulos & Fraser, 2013). While targeting continues to be widely used, especially in high-accuracy industrial and engineering measurement where precise positioning of specific object points is central to many dimensional metrology tasks, the use of artificially signalised points and especially coded targets, is nevertheless inconvenient in an increasing number of applications. Two examples are networks of images captured from UAVs and measurements over longer distances with lenses of long focal length, the required coded targets being inconveniently large in both instances.

This paper discusses the development of a processing pipeline designed to facilitate automatic photogrammetric network orientation and camera calibration in convergent, multi-image and possibly multi-camera networks in the absence of targets. An important development objective has been optimization of algorithmic efficiency in recognition of the computationally expensive nature of SfM methodologies. In pursuit of the objective of minimising computation time, selected algorithms have been implemented to run in parallel either in the CPU or GPU.

2. AUTOMATED NETWORK ORIENTATION

The proposed five-stage methodology for automatic orientation of convergent multi-image and multi-camera networks is laid out in Figure 1, with the stages being:

1. Initial point correspondence determination,
2. Point correspondence refinement,
3. Identification of additional point correspondences,
4. Epipolar mismatch filtering, and
5. Network orientation.

2.1 Initial Point Correspondence Determination

Point correspondence determination is arguably the most important step in fully automated network orientation. Typically in photogrammetry, this is achieved via arrays of coded or other types of artificial targets that are placed in the object space to facilitate the identification of common points between the images. In target-free networks, the point correspondence determination commences with the familiar feature extraction stage, via algorithms such as SIFT (Lowe, 1999) or SURF (Bay et al., 2008), something that is commonly known as feature based matching (FBM). Remondino (2006) presents an overview of various feature detector and descriptor algorithms that have potential use in photogrammetric applications. Nowadays there is a plethora of new detection and/or descriptor algorithms that can be adopted for point correspondence determination. Examples are BRIEF (Calonder, 2010), ASIFT (Yu, 2011) ORB (Rublee, 2011), BRISK (Leutenegger, 2011), FREAK (Alahi, 2012) and A-KAZE (Alcantarilla, 2012, 2013).

The image points detected in this first step of the proposed FBM process are salient features of the linear and sometimes non-linear scale-space of the image (depending upon the chosen detector). They are described based on the distribution of image gradients and/or image intensities (depending upon the chosen descriptor algorithm). The extracted feature points along with their descriptors are also known as keypoints. Feature point descriptors have been traditionally represented as a vector-based feature with a dimension of either 64 or 128 and they are utilised to match points in different images. However, nowadays there is a growing interest in binary-valued features, e.g. those produced via BRIEF, ORB and FREAK, since they have the advantages of being faster to compute and more compact to store. Also, their comparison is more efficient (Muja, 2012).

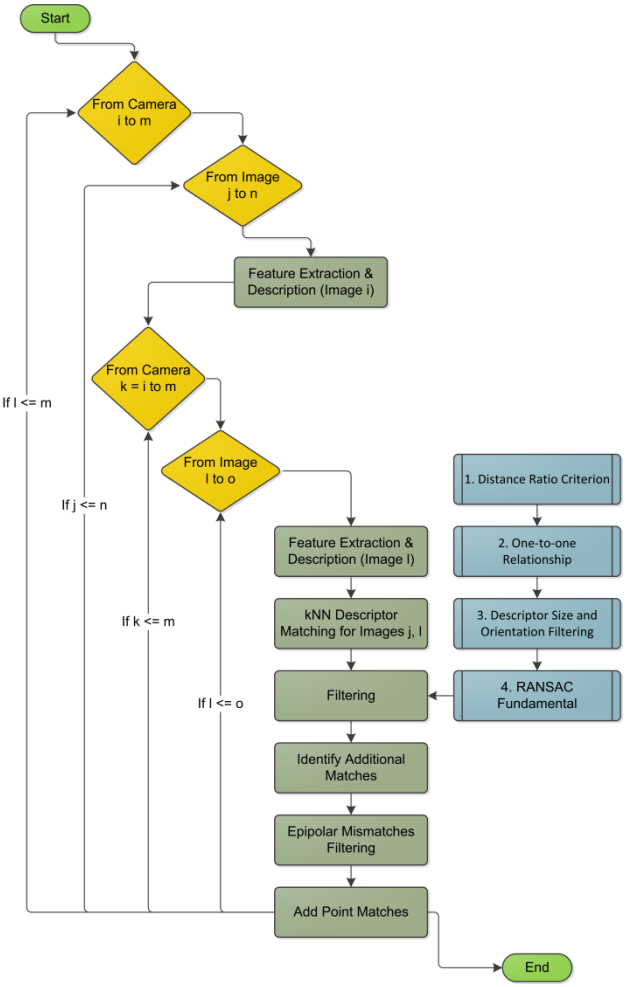


Figure 1. Flowchart of the proposed methodology

Regardless of the descriptor algorithm used, the aim is to describe a scale and rotation invariant feature with the purpose of determining matching points in different images. For vector-based features, the Euclidean length of the descriptor vector is used to quantify the level of correspondence. When binary-valued features are used, the Hamming distance is employed as a measure. Thus, the process of finding point correspondences for a pair of images, known as a nearest neighbour (NN) search, involves the calculation and comparison of all the descriptor lengths or Hamming distances in order to find the best possible matches.

For vector-based features, especially, the computation of all NN within a dataset with high dimensionality is a difficult task. For this reason, optimised data structures have been employed in order to speed up the search. Various researchers, e.g Lowe (1999) and Barazzetti (2011a), have proposed the use of a kd-tree to organize the data and optimise the calculation of NN searches. The computational cost of the NN search, after

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
doi:10.5194/isprsannals-II-5-339-2014

building the kd-tree, is O(logN) compared to O(N) when a simple linear search is performed. While the kd-tree algorithm offers various advantages it can also suffer from what is known as the curse of dimensionality. For a large number of points, and especially for high dimensionality cases of descriptor lengths of 64 or 128, the speed of the algorithm is only slightly better than a linear search of all points. For this reason, approximate NN search has been preferred to exact NN search. In the algorithm developed for the reported method, the Fast Library for Approximate Nearest Neighbours, FLANN (Muja and Lowe, 2009) has been used. The FLANN library is implemented in such a way that it allows data to be stored in a randomised forest of kd-trees in order to improve the effectiveness of the calculations for high-dimension cases. Additionally, the search operation is performed in parallel within available CPU cores. The kd-tree is a highly scalable algorithm and the availability of each extra CPU core can halve the search time. Alternatively, a NN search running on the GPU can be used, as proposed by Stamatopoulos et al. (2012).

While binary-based feature distances are much faster to compute and match, linear search is only practical in small datasets. For large datasets, it becomes impractical and in a similar way to the vector-based descriptors an approximate matching algorithm has to be used to speed up the process. Most of the approximate nearest search algorithms for binary features are based on hashing algorithms, e.g. locality sensitive hashing (LSH) (Rublee, 2011), semantic hashing (Salakhutdinov, 2009) and min-hash (Zitnick, 2010). However, Muja (2012) proposed a novel algorithm based on hierarchical decomposition of the search space that outperforms most hashing algorithms. This algorithm is also available as part of the FLANN library.

The use of coded targets precludes mismatching of image points, but this is not so for FBM. NN matching methods will always return a match, that being the closest match. In addition, with repetitive patterns and similarities that are present in the image space false matches are inevitable. Thus, this procedure produces both valid point matches and a large percentage of outliers. The prime objective is then to develop a methodology to effectively filter out erroneous 2D feature point matches within image pairs, such that the remaining valid matches can be used to perform a robust RO, followed by successive resections, spatial intersections and bundle adjustments to achieve an automated network orientation to sub-pixel accuracy.

2.2 Point Correspondence Refinement

An advantage of optimised structures such as the kd-tree is their ability to also perform k-nearest neighbor searches (kNN). This allows the calculation of not only the best match for a point, but potentially the k best matches as well. This information can be used for the point correspondence calculation stage to filter similar descriptor vectors. The filtering can be achieved by setting a criterion so that the difference in the descriptor length or Hamming distance, of the first and second best match for each point is more than 70%, for example. It should be noted, however, that this does not ensure that a keypoint is uniquely matched, and consequently an additional filtering process has to be performed so that a one-to-one relationship between the feature point matches is ensured.

For the next step, a histogram is built using the keypoint matches that have passed the first two stages of filtering. More specifically, the size and rotation of the descriptors are used to create a 2D histogram, essentially a 3D plot. The matched features whose scale and rotation do not agree with the majority are then removed from the list of valid matches. In this plot the x axis represents the bins of the keypoint scales and the y axis the bins of the rotations scales. The z axis is the number of counts for each combination of the scale and rotation bins.

Next, a RANSAC approach utilizing the fundamental matrix is employed with the purpose of further filtering the point correspondences for each image pair, the algorithm employed being the 5-point algorithm of Nistér (2004). RANSAC algorithms perform well when at least 50% of the input is inliers, which explains the reason for all the filtering steps leading up to this stage. After the RANSAC process, a least-squares adjustment involving all the inliers is performed in order to improve the accuracy of the calculated fundamental matrix.

2.3 Identification of Additional Point Correspondences

Under the assumption that a successful initial image point correspondence solution has been obtained and the geometry of the image pair can be described to an extent by the fundamental matrix, it is now possible and desirable to try to identify additional matches. While this step may not be needed in many cases where there is a large overlap between the images, for networks where there is less overlap and wider baselines it is helpful to increase the number of keypoint matches to achieve a more robust network orientation.

At the initial stage of the previous filtering procedure, a large number of matches are filtered out due to the distance ratio criterion. However, there is a real prospect that among those correspondences there are many that are valid. Having already calculated the fundamental matrix it is possible to test the validity of all the points matched during the kNN search by re-projection of each point correspondence and a check of the resulting residuals from the epipolar line. This calculation not only involves the first best match, but also the second nearest neighbor match as well, as there is a possibility that it could actually be the valid one. A point correspondence is accepted as valid when only one of the two neighbours is valid. Otherwise, they are both discarded since it is impossible to identify the one that is valid. It can be shown experimentally that this step can significantly increase the number of keypoint matches for pairs of images, with the associated computation cost being insignificant.

It should be noted that the use of the fundamental matrix at this as well as previous stages implies that the matching is not performed in 2D space, as with the homography matrix. It instead involves the calculation of a distance to an epipolar line. While the results of this stage can be used to successfully perform a RO using the coplanarity equation model, a number of epipolar mismatches generally remain among the point correspondences. Such mismatches may not pose a problem for the RO of an image pair, but they can be quite problematic at a later stage when more images are involved, especially at the subsequent image resection stage. The problem can be mitigated to a considerable extent via adoption of a RANSAC resection algorithm. However, it is preferable that any outliers are dealt with at an earlier rather than later stage, as will now be explained.

2.4 Epipolar Mismatch Filtering

At this final filtering stage, the problem of the remaining epipolar mismatches is addressed. These mismatches need to be

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
doi:10.5194/isprsannals-II-5-339-2014

eliminated because, firstly, they will result in erroneous 3D points and in the absence of additional imaging rays they will not be detectable as blunders in later network orientations. Secondly, the complexity of subsequent procedures is reduced when erroneous matches are eliminated (Stamatopoulos et al., 2012). Within the developed filtering/refinement procedure, the feature points of an image are clustered using a theoretical grid. For each cell of the grid, a projective transformation is performed between the 2D points and their corresponding matches in the second image. In this least-squares projective transformation, point matches with high residuals are flagged for later inspection because they may in fact still be valid correspondences. The same procedure is then repeated, the difference being that the second image is now considered as the 'base' and consequently the projective transformation from cells of the second to the first image are calculated. The flagged point matches from the two transformations are then examined and those common to both are deemed mismatches and are rejected. The reason for this two-way procedure is to avoid rejection of correct matches for complex 3D structures. Additionally, in order to ensure the integrity of the local transformations performed, criteria such as minimum number of points per cluster (grid cell) are enforced. A smart cluster merging algorithm is applied to avoid the elimination of cells containing only a small number of points.

2.5 Network Orientation

With the procedure described above, correct 2D point correspondences among pairs of images can be found. The orientation of an arbitrary number of images is thus reduced to the same process as orienting image networks that employ targets. Since all correspondences are known prior to the network orientation phase, it is possible to find an optimal starting point for the RO utilising additional information from the matching process. For example, the RO can start with the two images containing the largest number of multi-ray points. This will allow the maximum possible number of images to be resected before the need for associated spatial intersection.

In the proposed methodology, additional metrics are kept for each of the processed pairs of images. These include information such as number of matches, area coverage of the matched keypoints and convergence angle between the two images. By combining this additional information it is possible to ensure that the automated orientation will be performed with optimal robustness. Selection of an appropriate starting pair of images is important because the remaining images are sequentially added via a series of resections and spatial intersections. A bad starting point can preclude a complete network orientation, which will necesitate an alternative starting image pair for the RO. Additionally, in order to ensure convergence when new images are added to the network, a bundle adjustment with less stringent criteria for convergence and outlier detection is run. The noise-free network of 23 images and 80,000 points shown in Figure 2 highlights the capabilities of the proposed target-free automated orientation procedure.

3. EXPERIMENTAL EVALUATION

A number of experimental tests have been conducted to evaluate the proposed automated target-free network orientation methodology. New software was developed in C++ and integrated into the iWitnessPRO software package (Photometrix, 2014) to allow for operational testing. For this paper, two multi-image photogrammetric networks, the first being a close-range image configuration and the second an aerial dataset captured via a UAV are showcased. Additionally, two self-calibrations were carried out as part of the experimental testing in order to demonstrate the prospect of fully automated camera calibration without the need for artificial targets.

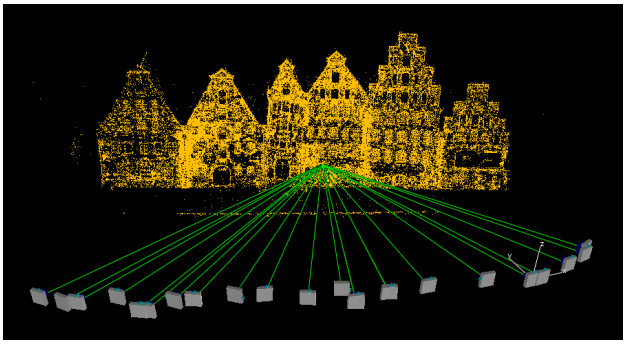


Figure 2. Configuration for the 80000-point, 23-image Salzspeicher network.

3.1 Automated Network Orientation

3.1.1 Close-Range Dataset

For the close-range photogrammetric network, the camera was a 12 mpixel point-and-shoot Nikon CoolPix P300. Figure 3 shows the network configuration employed to reconstruct a heritage building in Luebeck, Germany, namely the 15th Century Holstentor. The network is comprised of 46 images, taken from object distances ranging from about 15m to 60m, with various orthogonal camera roll angles so as to allow for camera self-calibration. The final photogrammetric network consisted of 225,000 3D points with a mean relative accuracy in object space of 1:8,500, the RMS image coordinate misclosure value being 0.35 pixels. It is noteworthy that 3D points with a maximum of 18 rays were present in the network, the maximum intersection angle being 120^o.

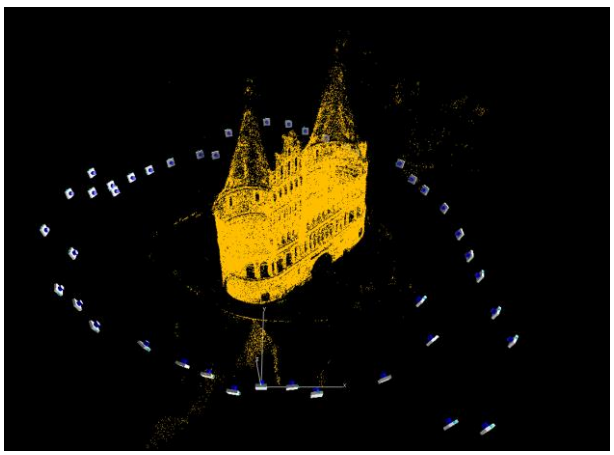


Figure 3. Configuration for 225,000-point Holstentor network.

3.1.2 Aerial UAV Network

In the case of the UAV network, a Canon IXUS 100 IS consumer camera with an integrated zoom lens was used, with the focal length being set to 5.9mm. This camera was deployed within a UAV flying at a mean flying height of 200m in a 23-image block configuration of near-nadir imagery. The images used for the calibration testing of the Canon IXUS formed a sub-block of a larger network established to investigate the

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
doi:10.5194/isprsannals-II-5-339-2014

accuracy potential of DEM determination from UAV imagery (Cramer, 2013). This data set was kindly provided to the authors by the Institute for Photogrammetry, University of Stuttgart.

The scene was a vineyard area of approximately 320 x 250 m with a height range of 60m and it was imaged in a configuration of three strips of nominally 90% forward and 70% side overlap, with an additional cross strip, captured at an image scale of close to 1:33,000 (GSD of 5cm). The configuration of the 23-image block is shown in Figure 4. The scene was feature-rich, with some 23,000 feature points being matched, the number of imaging rays per matched point varying from a maximum of 19 down to an assigned minimum of four. This yielded an image coordinate misclosure value of 0.43 pixels RMS and a relative accuracy in object space of 1:7,500 or 4.5 cm. In order to scale and transform the photogrammetric network to a global coordinate system 6 control points were used, the RMS of the shape-invariant transformation being 2.5 cm

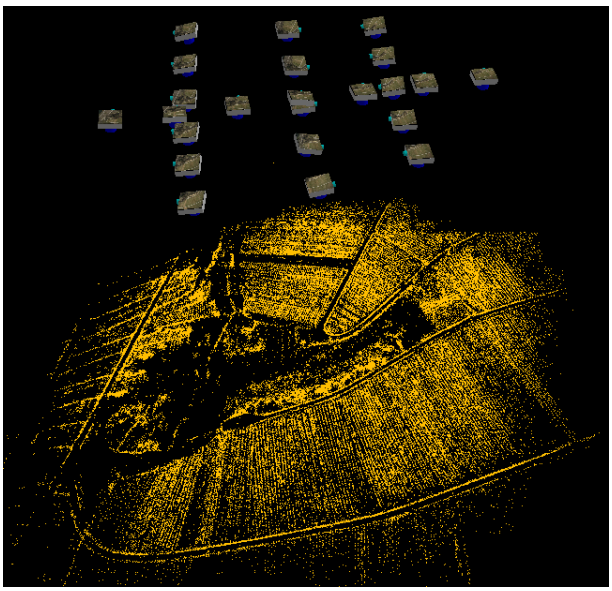


Figure 4. Configuration of the 23,000-point Hessigheim UAV network.

3.2 Camera Calibration

Two target-free camera self-calibrations were conducted to validate the proposed processing pipeline. The camera used for both tests was an off-the-shelf 10.2 mpixel Nikon D200 DSLR with a 17mm unifocal lens. The only concession made is regard to rendering this camera metric was to tape the lens barrel so as to ensure a fixed focus (of nominally 5m). The camera was to be hand-held.

The aim was to establish two self-calibration networks, with there being a requirement that these would be suitable for automated exterior orientation via both coded targets and 'natural' feature points extracted via feature point detectors. It was anticipated that image measurement and correspondence determination to 0.1 pixel accuracy would be achieved in the targeted networks, where the image point correspondence determination was via the codes. On the other hand, an accuracy of close to 0.25 – 0.4 pixels was to be expected for the feature-based matching of 'natural' feature points.

The plan was to utilize the same imagery for the targeted and untargeted cases in the calibrations of the Nikon D200 camera, with the network geometry needing to support a camera parameter recovery of maximum fidelity. To maximise the prospect of recovering a scene independent calibration, while minimising projective coupling between interior and exterior orientation parameters, the following geometric characteristics were adopted:

- A highly convergent imaging configuration comprising 20 or more camera stations, albeit with limited camera station separation in the vertical direction (only a short ladder was available).
- A wide diversity of orthogonal camera roll angles.
- An object that was three dimensional, such that the target and feature point fields were non-coplanar.
- An object with rich texture suited to feature point detection algorithms. The artificial coded targets comprised simply printed 11mm-diameter white dots on a black background.

The two networks for the Nikon DSLR calibration are shown in Figures 5a and 6a. Both comprised sections of wall from two buildings, one of light sandstone, Figure 5a. The texture here was not 'rich', but moderate, however the texture on the brick paving in front of the wall proved to be very rich and very favourable for feature point matching. There was also some vegetation, namely ivy, across the base of the 6m section of wall. The second building, Figure 6a, was red brick, and this exhibited moderate texture. Again, the brick paving in front of the 5m section of wall provided more feature points than originally anticipated, which enhanced the 3D nature of the object point array in the target-free calibrations.

The photogrammetric networks for the Nikon D200 calibration are shown in Figure 5b, for the sandstone building, and 6b, for the red brick building. In the former, 25 codes (200 points) and 55,500 feature points were recorded within the 27 images. In the latter, 31 images covered 24 codes (192 points) and 26,500 feature points. In both cases the set-back distance from the target array was approximately 5.5m, leading to an average image scale of 1: 300.

3.2.1 Results

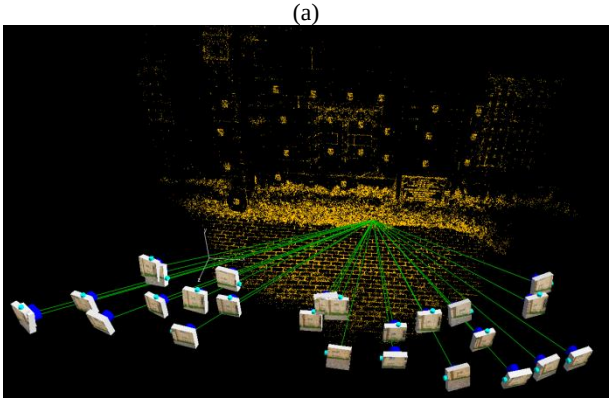
Table 1 provides a summary of the results of the self-calibrating bundle adjustments for the two Nikon D200 networks. In the table, the coded target networks are referred to as T and the untargeted as U. Listed are the adjusted values of the interior orientation parameters, focal length c, and principal point offsets x_p, y_p, along with their estimated standard errors. Also listed are radial distortion correction values at three selected radial distances, and two decentring distortion profile values listed for two radial distances. The reason for reporting lens distortion in this manner is that it provides a more easily interpretable indicator of the repeatability of the computed distortion profiles than would be the case if polynomial coefficients only were listed. The RMS value of image coordinate residuals and the number of object points in each network adjustment are also shown in the table.

Two aspects of the results listed in Table 1 are worthy of note prior to any discussion of the quality of the camera self-calibrations. The first is that in view of the fact that the Nikon D200 could not be considered a truly metric camera, given that no effort had been made to stabilise the lens assembly, expectations of high repeatability in interior orientation parameters between the two networks should be modest. However, expectations of repeatability between the targeted and untargeted cases, which used the same images, should be high. The second point to recall here is that the 'true' calibration

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
doi:10.5194/isprsannals-II-5-339-2014

values are not known, and the quality and fidelity of the self-calibrations can only be assessed via internal means. Here, two measures are relied upon, namely the precision and repeatability of recovery of the calibration parameters, and the resulting discrepancies in object space coordinates when these parameters are subsequently applied. In the following paragraphs the former measure is discussed, whereas the latter measure will be addressed in the following section.

(a)



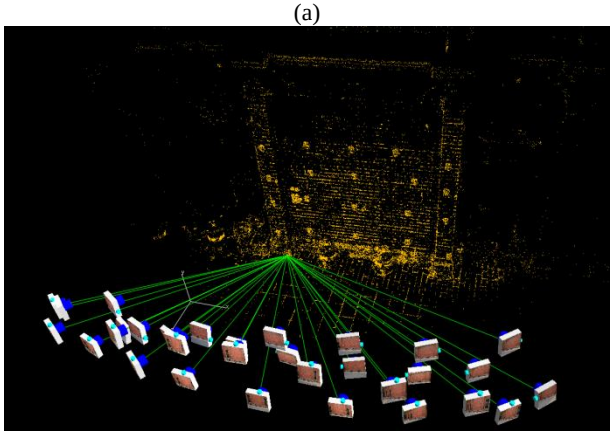
(b)

Figure 5. Configuration for 55,500-point 'Sandstone' self-calibration network.

As anticipated, there is a 2 – 3 times discrepancy between the accuracy of image coordinate measurement in the targeted and untargeted cases, with the RMS v_{xy} values being close to 0.1 pixel for the targeted cases and 0.25 pixels for the feature-based matching cases. It is interesting to note that there were no common points between the two cases. Although the same images were used, none of the coded target 'dots' were extracted by the feature point detectors. On the basis of the difference in triangulation closure alone, it could be anticipated that the precision of recovery of calibration parameters would be better for the targeted case. However, the precision of recovery of calibration parameters was in fact superior for the untargeted network adjustments, simply because there were so many more matched feature points than coded targets, there being more than 250 untargeted points for every artificial target for the Sandstone network and more than 130 for the Brick Wall network.

From a practical point of view, this is a noteworthy finding because it illustrates that the feature-based matching approach coupled with very dense point fields of thousands of points can yield camera calibration parameters to higher precision than from targeted arrays comprising a few hundred points. The same phenomenon was experienced with the development of image matching-based relative orientation on photogrammetric workstations back in the 1990s (eg Heipke, 1996).

(a)



(b)

Figure 6. Configuration for 26,500-point 'Brick Wall' self-calibration network.

The repeatability between the targeted and untargeted cases was very high for both the lens distortion parameters and the interior orientation elements in the Sandstone network, being within 1.5 μm (0.25 pixel) for all values except the principal distance, where the discrepancy was 5μm. The repeatability of interior orientation parameters was also high for the Brick Wall case, with the largest discrepancies between the two networks being 2μm for the principal distance and 5μm for the y_p offset parameter.

A contributing factor to the camera parameter discrepancies found in the test networks is the fact that the image points for the untargeted case covered a significantly greater area of the image format than for the targeted case. The radially dependent lens distortion functions were thus modelled with greater fidelity and the scale variation within images in the untargeted network was greater, which is a desirable attribute for self-calibration networks employing highly convergent imaging configurations.

On the subject of convergent imaging, there was one noteworthy surprise. Whereas it could be anticipated that accurate centroiding on high-contrast circular targets would be possible to incidence angles of 30 degrees to the target plane

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
doi:10.5194/isprsannals-II-5-339-2014

and potentially lower, there was not the same level of confidence that descriptor-based matching of feature points would accommodate moderately high convergence angles. This would suggest that the imaging geometry of the targeted array might be stronger, due to points having a higher number of imaging rays over a wider diversity of viewing angles. However, as mentioned in section 3.1.1, the proposed methodology is able to accommodate relatively wide baseline configurations, resulting in many points having effective convergence angles between imaging rays of up to 90 degrees. The success with the multi-ray matching of features also resulted in object points which were better distributed in three dimensions than the targeted points.

Table 1. Results of self-calibrations of the Nikon D200 camera for targeted and untargeted cases.

	Sandstone		**Brick Wall**	
	T	U	T	U
Focal Length, c / σ_c (mm)	17.632 / 0.0010	17.627 / 0.0008	17.633 / 0.0007	17.635 / 0.0004
x_p / σ_{xp} (mm)	-0.038 / 0.0007	-0.036 / 0.0005	-0.054 / 0.0005	-0.054 / 0.0003
y_p / σ_{yp} (mm)	-0.193 / 0.0007	-0.193 / 0.0005	-0.196 / 0.0006	-0.191 / 0.0003
Δr @ r=8mm (μm)	121.8	120.9	121.6	121.4
Δr @ r=10mm (μm)	217.2	216.2	217.0	216.9
Δr @ r=12mm (μm)	332.3	333.1	332.7	333.9
P(r) @ r=10mm (μm)	5.7	5.3	5.9	5.7
P(r) @ r=12mm (μm)	8.1	7.6	8.5	8.1
RMS v_{xy} / No. of points	0.09 pl / 200	0.25 pl / 55,500	0.10 pl / 192	0.25 pl / 26,500

3.2.2 Quality Evaluation in the Object Space

In order to assess the impact on object space point determination of the variation in the calibration parameters computed in the untargeted and targeted cases for the Nikon D200, a simple test was carried out. From the two sets of 30-odd images, two subsets were selected, 10 from the Sandstone and eight from the Brick Wall data set. Standard bundle adjustments employing the camera parameters from the self-calibrations were then carried out for each of these networks, the aim being to determine the accuracy of object point determination - for the coded targets only - as quantified by the XYZ coordinate discrepancies.

Listed in Table 2 are the RMS values of XYZ object point coordinate standard errors obtained in the network adjustments. These show anticipated accuracies (RMS 1-sigma) of 0.13 to 0.17 mm in the depth direction (Z), and between 0.05 and 0.09 mm in the dominant plane of each target array (XY). There is little distinction in quality as judged by the RMS discrepancy values, S_{XYZ}. In both networks, the accuracy is marginally better for the calibration obtained using the coded targets (Cal. 1 in Table 2) than via feature-based matching (Cal. 2 in Table 2). On the basis of S_{XYZ} values there is minimal distinction between the four network adjustments, and thus by implication, no practical difference between the individual calibrations.

A further means to assess calibration quality is via an examination of the difference in the computed shape of the object point fields determined from the different sets of calibration parameters, for both the Sandstone and Brick Wall test fields. The computed RMS coordinate discrepancy values S_{XYZ} are listed in Table 3, where it can be seen that the overall agreement in coordinate determination is at the level of 0.065mm or 1:80,000 of the size of the object field for the Sandstone network, and 0.027mm or 1:150,000 for the Brick Wall network. This impressive level of agreement is reasonably consistent with standard error estimates listed in Table 2. The shape and orientation of both networks can be considered identical for all practical applications, considering that the RMS discrepancy is only 65μm for the Sandstone and 27μm for the Brick Wall.

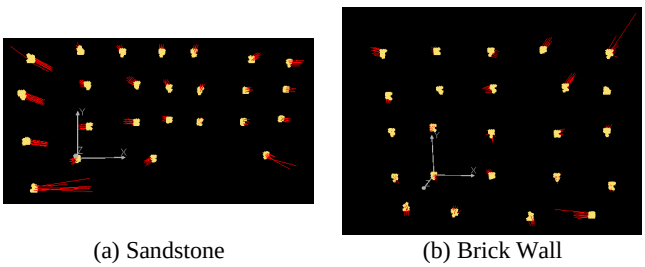
Table 2. Object point precision (σ) and RMS coordinate discrepancies (S_{XYZ}) for adjustments run with different calibration parameter sets.

	Sandstone		**Brick Wall**	
Object Point Standard Errors (mm)	Cal. 1 (targets)	Cal. 2 (FBM)	Cal. 1 (targets)	Cal. 2 (FBM)
σ_X	0.067	0.073	0.057	0.060
σ_Y	0.083	0.091	0.058	0.063
σ_Z (depth)	0.155	0.166	0.126	0.138
S_{XYZ} (pixels / μm)	0.08 / 0.51	0.09 / 0.55	0.08 / 0.50	0.09 / 0.53

Nevertheless, it is interesting to note that the differences in object point XYZ coordinates obtained with the different camera calibration parameter sets show clearly systematic trends. This is illustrated in Figure 7, where the coordinate differences for the coded target clusters of 8 individual code 'nuggets' show consistent magnitude and orientation.

Table 3. RMS coordinate discrepancy values: Calibration 1 versus Calibration 2 (mm)

	S_X	S_Y	S_Z	S_{XYZ}
Sandstone	0.075	0.039	0.076	0.065 mm or 1:80,000
Brick	0.032	0.025	0.022	0.027 mm or 1:150,000



(a) Sandstone (b) Brick Wall

Figure 7. Object point discrepancies from network solutions using different calibration parameter sets.

4. CONCLUSION

It has been demonstrated that the proposed target-free automatic network orientation approach can yield a comprehensive and effective elimination of outliers in the initial point

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
doi:10.5194/isprsannals-II-5-339-2014

correspondence determination stage, which leads to largely noise-free networks. Moreover, the experimental test results show that the approach can accommodate the convergent imaging configurations that characterise moderate-to-high accuracy close-range photogrammetric measurement, and which are necessary for reliable network orientation and self-calibration. Through the target-free camera calibration tests conducted for the Nikon D200 cameras it has been demonstrated that camera calibration parameters of greater precision and of equal accuracy can be achieved, as far as could be assessed, compared to today's 'standard' automatic self-calibration approach which involves the use of targets. The poorer image point measurement accuracy of descriptor-based feature point matching is more than offset by the provision of potentially 100-fold more object points within the photogrammetric network. Within the close-range measurement context, both approaches fit well into automatic data processing pipelines, as exemplified by the iWitnessPRO software system, which accommodates both the targeted and targetless cases. Also, from a practical standpoint, if the scene or object being imaged is texture rich and conducive to the target-free approach, then this is arguably the more flexible automated camera calibration option, yet there is limited time distinction between the two processes since the time to position 50 or so coded targets to form an object point array, can be similar to the extra computation time associated with the target-free calibration.

5. REFERENCES

Alahi, A., Ortiz, R., Vandergheynst, P., 2012. FREAK: Fast Retina Keypoint. In: IEEE Conference on Computer Vision and Pattern Recognition, Fiorenze, Italy.

Alcantarilla, P.F., Bartoli, A., Davison, A.J.: KAZE features. In: *ECCV 2012*, Part VI. LNCS, vol. 7577, pp. 214–227. Springer, Heidelberg.

Alcantarilla, P.F., Nuevo, J., Bartoli, A., 2013. Fast Explicit Diffusion for Accelerated Features in Nonlinear Scale Spaces. In: *British Machine Vision Conference (BMVC)*, Bristol, UK.

Abdel-Wahab, M., Wenzel, K., Fritsch, D., 2011. Orientation and Geometry from Large Unordered Image. Datasets for Low Cost Applications. In: *Proceedings LC3D Workshop*, Berlin.

Barazzetti, L., 2011a. Automatic tie point extraction from markerless image blocks in close-range photogrammetry. Ph.D. thesis, Politecnico di Milano.

Barazzetti, L., Mussio, L., Remondino, F., Scaioni, M., 2011b. Targetless camera calibration. In: *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, Trento, Vol. XXXVIII-5 /W16, pp. 335-342.

Bay H., Ess A., Tuytelaars T., Gool L.V., 2008. SURF: Speeded Up Robust Features. In: *Computer Vision and Image Understanding (CVIU)*, Vol. 110, No. 3, pp. 346-359.

Calonder, M., Lepetit, V., Strecha, C., Fua, P., 2010. BRIEF: Binary Robust Independent Elementary Features. In: *Proc. of the European Conf. on Computer Vision (ECCV)*.

Cramer, M., 2013. The UAV@LGL BW project – a NMCA case study. Proceedings of Photogrammetric Week '13 (Ed. D. Fritsch). Wichmann, pp. 165-179.

Heipke C., 1996. Automation of interior, relative and absolute orientation. In: *International Archives of Photogrammetry & Remote Sensing*, Vienna, XXXI(B3): 297-311.

Leutenegger, S., Chli, M., Siegwart, R., 2011. BRISK: Binary Robust Invariant Scalable Keypoints. In: *ICCV 2011*, pp. 2548-2555.

Lowe, D.G., 1999. Object recognition from local scale-invariant features. In: *Proceedings of the International Conference on Computer Vision*, pp. 1150–1157, Corfu, Greece.

Muja M. & Lowe D. G., 2009. Fast Approximate Nearest Neighbours with Automatic Algorithm Configuration. In: *International Conference on Computer Vision Theory and Applications (VISAPP)*, Lisbon, pp. 331-340.

Muja, M., Lowe, D.G., 2012. Fast matching of binary features. In: *Conference on Computer and Robot Vision (CRV)*, Toronto, Canada.

Nistér D. 2004. An efficient solution to the five-point relative pose problem. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 26(6): 756-770.

Photometrix, 2014.
http://www.photometrix.com.au (13/02/2014).

Remondino, F., 2006. Detectors and descriptors for photogrammetric applications. In: *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. 36(3), pp. 49-54.

Rublee, E., Rabaud, V., Konolige, K., Bradski, G.R., 2011. ORB: An efficient alternative to SIFT or SURF. In: *Proceedings of the IEEE Intl. Conf. on Computer Vision (ICCV)*, Vol. 13, pp. 2564-2571, Barcelona.

Salakhutdinov, R., Hinton, G., 2009. Semantic hashing. In: *International Journal of Approximate Reasoning*, Col. 50(7), pp. 969-978.

Stamatopoulos, C., Chuang, T.Y., Fraser, C.S., Lu, Y.Y., 2012. Fully automated image orientation in the absence of targets. In: *International Archives of Photogrammetry, Remote Sensing & Spatial Information Sciences*, Vol. XXXIX-B5, pp. 303-308.

Stamatopoulos, C., Fraser, C. S., 2013. Target-free automated image orientation and camera calibration in close-range photogrammetry. In: *Proceedings of ASPRS Annual Conference*, Baltimore, Maryland, March 24-28, 8 Pages.

Yu, G., Morel, J.M, ASIFT: An Algorithm for Fully Affine Invariant Comparison. In: *Image Processing On Line*, vol. 2011, http://dx.doi.org/10.5201/ipol.2011.my-asift.

Zitnick, C., 2010. Binary coherent edge descriptors. In: *European Conference on Computer Vision*, pp 1-14.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
doi:10.5194/isprsannals-II-5-339-2014