

MULTI-CAMERA GEOMETRIC CALIBRATION: PRE-CALIBRATION AND ON-THE-JOB CALIBRATION OF THE MAIA MULTISPECTRAL SYSTEM

P. Garieri ^{1*}, F. Diotri ¹, G. Forlani ¹, U. Morra di Cella ², R. Roncella ¹

¹ Department of Engineering and Architecture, Area delle Scienze 181/A, 43124 Parma, Italy - (pietro.garieri, gianfranco.forlani, riccardo.roncella)@unipr.it, fabrizio.diotri@gmail.com

² CIMA Research Foundation, Via A. Magliotto, 2, 17100 Savona, Italy - umberto.morradicella@cimafoundation.org

Commission II, WG II/7

KEY WORDS: Geometric Calibration, UAV Remote Sensing, Accuracy, Sensor Co-registration, MAIA

ABSTRACT:

Though with a less dramatic growth compared to photogrammetry, remote sensing from multispectral imagery taken by UAV (Unmanned Aerial Vehicles) platforms is applied to vegetation health monitoring, crop management, water quality assessments, geological inspections and much more, with a sizeable number of multispectral cameras now available on the market. As for satellite images, a key point in remote sensing is calibration, both geometric and radiometric, and the modelling of disturbances, to get well co-registered reflectance data. Leaving aside radiometric calibration, this paper focuses on how to best georeference the different bands one with respect to the other. This is normally achieved by the so-called band-to-band registration (BBR). Here, a straightforward approach is proposed, that exploits the multi-camera geometry and, unlike BBR, all the information contents of the bands, as inter-band matches are searched (and possibly found) for every pair of bands and not only between a reference and a slave band. Tests on images taken with the 9-band MAIA S2 camera are presented, discussing the pro and cons of pre-calibration and on-the-job calibration of the camera parameters of each sensor. The results found show that the proposed method is at least as good as the BBR ones.

1. INTRODUCTION

Multispectral as well as hyperspectral cameras have been hosted, since decades, on aerial and (above all) on satellite platforms. However, the dramatic push to the diffusion of photogrammetry brought by UAVs is now affecting more and more also remote sensing, as several multispectral cameras are now available, purposely designed for UAV systems.

Indeed, sensor miniaturization allowed manufacturers to build very compact systems, further extending the appeal of UAVs, already a fundamental platform for imagery in the visible spectrum. Building on UAV versatility, remote sensing observations with increased spatial and temporal resolution opened new perspectives environmental studies and monitoring. To be really useful for remote sensing classification, however, multispectral imaging needs extensive efforts in calibration, both geometric and radiometric and modelling of disturbances, to get to accurate georeferenced reflectance data.

Putting all required bands on a single multispectral camera is therefore a great step towards simplification of the processing chain, especially with simultaneous imagery acquisition. In this paper, leaving aside everything about radiometric calibration, we focus on how to best georeference the different bands one with respect to the other.

In most cases, the multispectral cameras developed for installation on UAVs are in fact multi-cameras. In other words, each band is acquired by a different camera (optics and sensor), with its own projective centre, geometric lens distortion and possibly different viewing angles. As spectral firm interpretation requires the best possible co-registration of the different bands, a procedure is to be developed to this aim, either in image space or in object space.

The former approach, by far the most frequently used, is also known as band-to-band registration (BBR): one of the cameras (bands) is chosen as reference (master) while the others act as

slave ones (Anuta et al, 1984, Goforth, 2006; Holtkamp and Goshtasby, 2009). A possible alternative is the creation of a virtual image plane (Ladstädter et al, 2010). Features are extracted and matched between the master and each slave image and used to estimate an image transformation that resamples each slave over the master image frame.

In the latter approach to co-registration, which we propose in this paper, the pixels of each band are projected to the ground after image block orientation: in other words, the (georeferenced) orthoimage of each band is generated. If each band has been treated separately in an image block, the co-registration will merely depend on the overall orientation quality of each block. If, to the contrary, all bands participate simultaneously to the Structure-from-Motion orientation procedure, matches are found also across bands and contribute to the overall determination of the Interior (IO) and Exterior (EO) Orientation parameters. In case of synchronous acquisition by a multi-camera system, this process could conceivably be further constrained if the relative orientation parameters between the different cameras are enforced.

As anticipated, most previous work that we found on the topic follows the former approach. In (Laliberté et al, 2011) a procedure in image space is presented where matches between bands are computed based on the phase correlation method (Kuglin and Hines, 1975) applied to relatively large image patches to sample the local variations of the field representing the mis-registration.

After outlier filtering, the mis-registration field is smoothed with a local weighted mean transform and interpolated bilinearly.

In (Turner et al., 2014) key points between bands are selected with the Scale Invariant Feature Transform (SIFT) and then matched with a matching algorithm. Using the matched key points image coordinates, a Delaunay triangulation is applied to model the transformation, resampling the slave images with the

*Corresponding author

nearest neighbour. The authors notice the alignment quality to be highly dependent on the camera-to-object distance, as they aimed at using a single parameter set for all images of a project. In (Jhan et al, 2016) a mixed approach has been used, partly similar to the one we implemented. To improve band co-registration of the multi-lens multi-spectral MiniMCA (Miniature Multiple Camera Array) first an indoor laboratory calibration is performed, to estimate camera model parameters and the relative orientation parameters w.r.t. a reference camera. These data are then used to compute the coefficients of a modified projective transformation, inspired to the epipolar image normalization method of (Cho et al., 1992), between the reference and the slave image planes to resample all bands to the same image plane. The reported accuracy of the co-registration is 0.6 pixels.

In (Banerjee et al, 2018) the performance of the co-registration between the bands of a 15-bands hyperspectral cameras was studied comparing different keypoints selection methods. Among the six descriptors tested, SIFT performed best. Uniformity of distribution in keypoints selection was sought with a maximal total mutual distance criterion; an affine transformation model has been used in resampling.

The primary objective of this paper is to evaluate multispectral sensors geometric accuracy in object space. Though primarily developed to acquire radiometric information, a geometric calibration is also necessary for such cameras to model distortion and to generate correctly geo-referenced ortho-images. Tests have been carried out to determine if the camera can be exploited as a regular multi-camera system in a photogrammetric workflow. Indeed, the capability to produce digital terrain models (DTMs) and acquire radiometric data during the same UAV flight would bring time and cost savings.

A secondary objective of this paper is to evaluate the quality of the latter approach to band co-registration i.e., one that takes place in object space. Under the hypothesis of dealing with a multi-camera system with synchronous acquisition, the framework of the method is quite straightforward, as it exploits the capabilities of commercial software in multi-camera orientation with Structure-from-Motion (SfM). In other words, rather than selecting as reference a single band, all bands are treated equally. This is obtained by simultaneous orientation of all images of all bands in a SfM process. Depending on the parameters selected as unknown, in principle all camera model parameter for each band, as well as all exterior orientation parameters, can be computed. If an a-priori calibration is performed, either the camera model parameters or the relative orientation parameters between the cameras (in this case a master camera is selected) part of the parameters can be enforced as known data.

The method has been applied to close-range and UAV images taken with the multispectral camera MAIA S2.

To compare both a self-calibrated and a pre-calibrated approach to the estimation of the camera parameters, as well as the offsets and rotations between the sensors, a tri-dimensional close-range calibration test-field has been used. The investigation of the calibration parameters sets quality has been carried out both on the test-field with images in different light conditions and block geometry as well as on MAIA images taken by UAV in an 80 m above ground level flight. To this aim, the accuracy in object space has been evaluated. Finally, after generation of the orthoimage of each band, the co-registration quality has been assessed finding homologous points in the different bands by matching keypoints extracted with different detectors and descriptors.

2. MATERIALS AND METHODS

2.1 MAIA S2 camera system

MAIA S2 (Figure 1) is a multispectral camera empowered by nine different CMOS sensors, which are arranged as a 3x3 array on the camera body. The centre-to-centre distance for two consecutive sensors on the array is 25 mm. Each sensor is equipped with a different band-pass filter that allows the camera to record nine different spectral bands. The original MAIA model leverages a RGB sensor and eight different monochrome ones while the MAIA S2 model considered in this work is provided with nine different monochrome sensors, which record the same spectral bands as Copernicus Sentinel-2 satellite. This feature could be exploited to selectively increase information resolution of Sentinel-2 remotely sensed data in specific (small size) regions. The camera is explicitly designed for being equipped in UAV systems.

Each sensor, with dimensions of (3.6 x 4.8 mm²), is equipped with a fixed nominal focal length of 7.5 mm and delivers a resolution of 1.2 Mpixel (1280 x 960 pixels), with a pixel size of 3.75 µm and a radiometric resolution up to 12 bit per sample.



Figure 1. MAIA S2 camera system

MAIA S2 is equipped with an Incident Light Sensor (ILS) aimed at measure the level of ambient lighting during exposures and correct the evaluated reflectance values of each single band. Each sensor is equipped with a global shutter and allows synchronous image capture for the nine different bands. Global shutter technology allows the camera to register the electromagnetic charge at the same time in each photo-sensitive cell of the sensors. This avoids image crawling distortions, spatial and temporal aliasing, and other types of artifacts like wobbling artifacts which are witnessed in rolling shuttered sensors at high acquisition speeds. The different spectral bands acquired by the system are reported in Table 1.

<i>Band</i>	<i>Start WL (nm)</i>	<i>Stop WL (nm)</i>	<i>CWL (nm)</i>	<i>FWHM (nm)</i>	<i>Colour</i>
B1	433	453	443	20	VIOLET
B2	457.5	522.5	490	65	BLUE
B3	542.5	577.5	560	35	GREEN
B4	650	680	665	30	RED
B5	697.5	712.5	705	15	RED EDGE 1
B6	732.5	747.5	740	15	RED EDGE 2
B7	773	793	783	20	NIR 1
B8	784.5	899.5	842	115	NIR 2
B9	855	875	865	20	NIR 3

Table 1. MAIA S2 spectral resolution specifications.

Each single camera (i.e., sensor with its optics) is calibrated in laboratory before the MAIA unit is sold and a calibration certificate is provided. In (Nocerino et al., 2017) specific information of the MAIA calibration procedure is provided. The post-processing software provided with the MAIA system allows for geometric correction and inter-band co-registration of the acquired data, based on the provided initial laboratory calibration parameters. For band co-registration, since the nine sensors are not co-axial, the software requires the user to specify an approximate average altitude/distance from the object.

However, it is well known that such small and compact camera systems might suffer of optical instabilities in its lifecycle due to mechanical (e.g., vibrations) and thermal stresses. Best practices suggest performing a periodical calibration procedure or compute an on-the-job calibration to implement up-to-date calibration parameters. The aim of the experiments presented in this work is therefore to evaluate the influence of different calibration strategies and workflows to assess the actual accuracies achievable, more generally, in a multi-camera optical system, at least for those systems with a sensor layout similar to the MAIA.

2.2 Calibration test-fields

For the geometrical characterization of the MAIA S2 sensors, a 3D test-field (Figure 2) similar to the one proposed by (Nocerino et al., 2017) with a volume of $5 \times 5 \times 4 \text{ m}^3$ has been set up, with 34 coded targets homogeneously deployed. The targets were uniformly distributed on three mutually orthogonal planar surfaces, while seven additional targets were placed on vertical and horizontal elements inside the scene to increase the depth variability of known points inside the calibration volume. The targets have been surveyed from two different points with a Total Station: the estimated precision of their coordinates is ca. 0.2 mm (front vertical plane) and 0.5 mm (depth). Camera-to-object distance ranges approximately from 5 to 7 m.

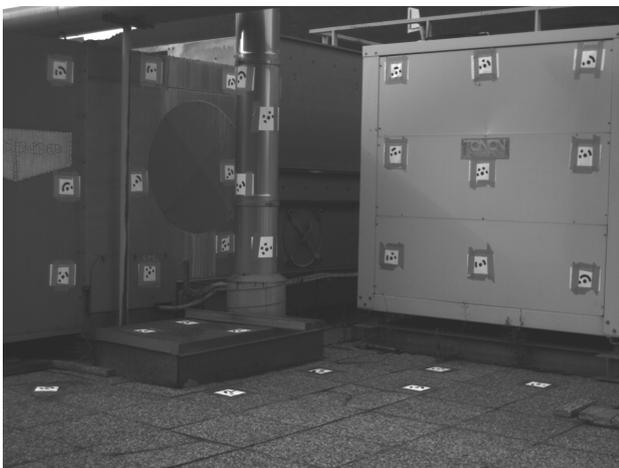


Figure 2. Close-range test-field

For this first close-range calibration field three different MAIA image blocks were considered. A 45-images calibration block (Block 1), made of three different strips at different heights, has been acquired with a Ground Sampling Distance (GSD) of about 2.5 mm. Each strip, composed by 5 stations arranged along a curved path, features convergent camera axes. At every station, 3 images have been taken, one in normal position and two rotated $\pm 90^\circ$ around the camera axis. Block-1 was used in the experiment for full-field calibration, estimating both interior orientation and distortion parameters of each optics, as well as

the relative spatial orientation between the different sensors. Such parameters were considered as a fixed pre-calibrated set to be used in subsequent experiments with other image blocks. A second image acquisition (Block 2) consisted of 15 images, split in three different strips, acquired approximately from the same location of Block 1 but without rotating the camera by $\pm 90^\circ$ around the optical axis. Block 2 should therefore be considered as a sort of subset of Block 1, with an imaging geometry similar to the ones usually implemented in a real-world survey, but also with a weaker control over possible parameters correlations due to the missing rotated photos. Finally, Block 3 has been acquired under slightly different lighting conditions and is made of 12 straight photos arranged over two strips, with a slightly larger base-length compared to Block 2.

To test the actual quality of the close-range pre-calibration, and provide additional data for evaluating on-the-job calibration performances, a test-field on a much wider area (Figure 3) has been considered as well (Block 4). This second test-field is located in Verrayes ($45^\circ 45' 37'' \text{ N}$, $7^\circ 32' 26'' \text{ E}$ - Valle d'Aosta region, Italy) and was used previously in other experiments (see for instance (Forlani et al., 2020) for further details). The test-field covers an area of approximately $250 \times 180 \text{ m}^2$ and consists of a gently undulated terrain (total height range of about 20 m), mostly covered by short grass with some buildings (5 m to 10 m high) and thicker vegetation on the edges. The area was surveyed mounting the MAIA camera system on a DJI Matrice 200 UAV. Nine ground control points and checkpoints have been used for assessing the orientation solution accuracy.



Figure 3. Verrayes (UAV) test-field

The image block consists of 8 strips (with 60% sidelap and 80% overlap) for a total of 189 images acquired at a relative altitude of ca. 80 m (average GSD is ca. 39 mm/pixel).

2.3 Calibration workflow

To find out the best calibration workflow, multiple combinations of self-calibration on the camera model parameters, offsets and rotations have been processed. All the blocks considered have been processed using Agisoft Metashape Pro v.1.8.0. (Agisoft, 2022). The MAIA camera system saves the acquired images in a proprietary raw format that can be subsequently processed using the MAIA image processing software package. The user can choose to generate multi-layer Tiff images, with one layer for every single spectral band (bit depth up to 12 bits/sample) or export a false RGB (3 bands) or an Indexed (1 band) image, applying or not radiometric and geometric correction and, in the latter case, co-registering or not the different bands. To limit the influence of user choices in this stage, the images used for calibration were

obtained converting from the raw data to a multi-layer Tiff without applying any correction and preserving the original (12 bits/sample) bit depth.

When such data is injected in Metashape, the software allows two different approaches: the first, in the following called “multi-camera”, considers the different Tiff layers (i.e., the ones corresponding to the nine spectral bands) as geometrically connected i.e., assuming the nine layer-images coming from sensors whose relative orientations are fixed and do not change over time. In other words, for every multi-layer image, a single set of exterior orientation parameters is considered and the relative orientation between each spectral sensor is used to determine the individual orientation of each layer frame. More specifically, the software considers one of the nine bands (band 1 as default) as a sort of master (or reference) image and the exterior orientation parameters of the multi-camera refer to its orientation in object space. The exterior orientation of any other (slave) layer is obtained from the master one applying, for each slave, a translation vector (in the following “Tx”), that represents the relative position between the current slave and the master projection centres, and a rotation (in the following “Rx”) that represents the relative pose between master and current slave image spaces. On the contrary, in the second approach, every Tiff layer is treated as an individual image, whose exterior orientation parameters are considered independent from the other layers.

During Structure from Motion stages, regardless of the selected approach, all the layers are considered as individual images i.e., tie points are extracted also between different bands of the same multi-layer image (even when the “multi-camera” approach is adopted). Since the base-lengths between the sensors are usually very small if compared with the camera-to-object distance (the maximum distance between two sensors is ca. 75 mm), this may cause unwanted behaviour. Indeed, when camera-to-object distance is large compared to camera-to-camera distance as in UAV surveys, some object points coordinates could be extremely inaccurate if the intersection is made only of corresponding rays that are almost parallel. This is the case, for instance, when correspondences are found only among the bands of a single multi-camera shot. However, from a theoretical point of view, such corresponding points (i.e., the ones common only to different bands of the same image, in the following called “inter-band” tie-points) should strengthen the estimation of the relative pose “Rx” of the different sensors. To evaluate their actual influence on the calibration procedure, the extracted image features were filtered to obtain three different calibration configurations. The first, indicated with “All points” considers all the correspondences found in the SfM stage. In the second, indicated as “inter-band filtered”, the extracted tie points, whose corresponding image-points come all from the same multi-layer image, are discarded. Finally, in the “base-length filtered” configuration, the tie points whose corresponding image-points are all extracted between images with short base-length (in the experiments a threshold of 0.5 m was used) are discarded. In the latter, in other words, also the corresponding points found only between $\pm 90^\circ$ rotated images are discarded.

As already pointed out, Block 1 was used to perform a full-field calibration (in the following “Pre-calibration”). The estimation of the calibration parameters has been carried according to the Brown camera-model (Brown, 1971) for IO and distortion parameters (i.e., principal distance “f”, principal point location “cx and cy”, radial distortion “k1, k2, k3”, and tangential distortion “p1 and p2”).

For Block 2, 3 and 4 also an on-the-job calibration was performed keeping fixed all the inter-band relative orientations (“IO and distortion”), keeping fixed the pre-calibrated IO and

distortion parameters and re-estimating only the translation vector between master and slaves band (“Tx”) or only the rotation between them (“Rx”) or both (“Tx + Rx”) or, finally, estimating all the parameters (“All”). The influence of the offsets should be accurately assessed as there may be a direct relationship with the distance between the acquisition station and the object to be reconstructed.

2.4 Calibration quality evaluation

To check whether a particular calibration strategy provides better performances than the others, two different methods were considered. For the close-range image blocks (Blocks 2 and 3) half of the known points were used as check points and their final, bundle-adjusted, coordinates were compared with the ones obtained by the total station survey. For the UAV image block (Block 4), as just few ground points were available, restricting the accuracy evaluation only to check point residuals was deemed not appropriate. In fact, in an operational scenario of the “object space” band registration procedure, it is more important to assess whether the Digital Surface Model (DSM) generated using the MAIA data is appropriate to support orthoimage generation and, finally, to assess the accuracy of the bands co-registration. Therefore, two additional procedures were considered.

The first consisted in comparing the MAIA-generated Digital Surface Model (DSM) with each calibration parameters set, with the DSM obtained in a high-resolution UAV survey using a DJI Phantom 4 (RTK) system (Figure 4). The reference Phantom 4 survey has a better image scale than the MAIA image block (2 cm/pixel GSD vs 3.9 cm/pixel) and was considered as a reference for evaluating the MAIA DSM accuracy. Its DSM was derived acquiring more than 180 million points (1650 points/m² on average) from which a TIN (Triangulated Irregular Network) mesh was derived. The MAIA point clouds, in comparison have a much lower point density: ca. 400 points/m². Worst results might be expected on the image block boundaries, due for instance to the lower number of overlapping images and the higher uncertainties of exterior orientation estimates (Dall’Asta et al. 2015). For this reason, all the comparisons were conducted considering only the central part of the image block, on an area approximately 140x100 m, well inscribed inside the GCP area. For the comparisons, the software Cloud Compare (CloudCompare, 2022) version 2.11.3 was used. In a first stage the MAIA point cloud was finely registered on the reference DSM using an ICP (Iterative Closest Point) algorithm, in order to remove unwanted systematic shifts that might have been caused by the orientation procedure of both the image blocks. After the registration, the C2M (Cloud to Mesh) distance algorithm was used to compute the difference between the MAIA point cloud and the reference model.

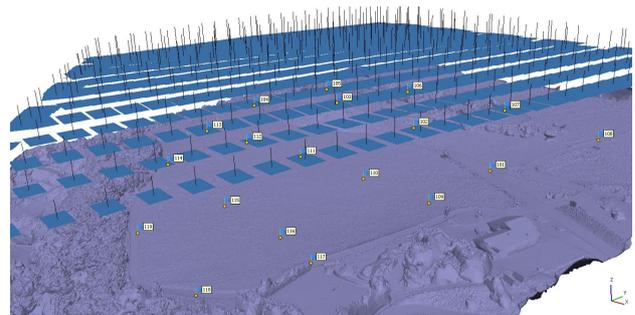


Figure 4. Phantom 4 DSM of the test area

The second test considered the accuracy achievable by the different calibration strategies comparing the quality of

		Pre-calibrated	Self-calib. (IO and distortion)	Self-calib. (Tx)	Self-calib. (Rx)	Self-calib. (Tx + Rx)	Self-calibrated (All)
		RMSE (mm)	RMSE (mm)	RMSE (mm)	RMSE (mm)	RMSE (mm)	RMSE (mm)
BLOCK 2	No-filtering	3.0	1.5	1.2	1.5	1.2	1.2
	Inter-band filtering	2.7	12.4	1.9	11.8	1.8	1.8
	Base-length filtering	3.5	38.7	1.7	37.6	1.7	2.5
BLOCK 3	No-filtering	3.2	36.3	2.1	35.9	2.5	5.2
	Inter-band filtering	4.3	14.3	3.2	13.3	3.3	3.5
	Base-length filtering	4.7	50.2	2.3	47.4	2.7	5.1
BLOCK 4	No-filtering	48.5	32.4	36.2	36.3	36.2	32.2

Table 2. RMSE (in mm) of the residuals of the CP.

orthophotos band co-registration at the end of the photogrammetric process. Unfortunately, at least as far as authors software knowledge is correct, Metashape does not allow to generate a multi-layer orthophoto from a “multi-camera” project. The orthophoto generation routine implemented in the software package, after tessellating the extent of the area acquired by the image block in smaller tiles, determines which image is more appropriate (i.e., has better image scale and frames the tile in the centre of the image) for a specific tile, and uses that image to orthorectify the area inside the tile. However, as stated before, in a “multi-camera” image block, every band is considered in Metashape as an independent image (although its orientation parameters are derived from the ones of the corresponding reference band). These imply that, in orthophoto generation, it is not possible to specify which specific spectral band must be used. In other words, at the end of the process, the final orthophoto is a sort of a patchwork where each tile might have been obtained from a different band. To overcome this limitation, leveraging the scripting possibility offered by the software package, an ad-hoc orthophoto generation procedure was developed by the authors. At the end of the DSM generation (which uses all the available bands), all the images exterior and interior parameters are exported. From each multi-layer image all the single bands are extracted and saved separately. Then, a new image block, considering one spectral band at a time, is generated assigning to each image its corresponding exterior and interior orientation parameters and importing the DSM obtained in the previous step. In this way, the user can obtain an orthophoto for each spectral band of the MAIA camera.

To evaluate the co-registration quality of the obtained orthophotos the following procedure was implemented: for each band-to-band combination (having 9 different bands a total of 36 combinations can be considered) a feature-based matching algorithm is applied to extract corresponding points. The choice of preferring a feature-based to an area-based method, which would have allowed a pixel-to-pixel comparison between the images, was made (i) to avoid those radiometric differences, due to different reflectance of the object at the different wavelengths, could be erroneously considered as displacements, and (ii) to calculate displacements mainly on well-textured areas.

For feature extraction and matching, five different algorithms implemented in OpenCV libraries (vers. 4.5.5) were tested, including three blobs extractor (KAZE (Alcantarilla et al., 2012), AKAZE (Alcantarilla et al., 2013) and SIFT (Lowe, 2004)), and two corner detectors: Brisk (Leutenegger, 2011) and ORB (Rublee et al., 2011), each one with its corresponding feature descriptor. Ideally, corresponding features should be identified, excluding the errors due to feature extraction

algorithm, exactly in the same position. Therefore, the distance between the matched features represents, at least statistically, the co-registration error. To remove matching outliers, a threshold on this distance has been set and matches between points having a distance greater than 0.20 m (five times the GSD) have been removed from the statistics. The Root Mean Square (RMS) of the distances is taken as co-registration performance index. In the tested dataset, one feature-based matching method performed much better than the others. In fact, as far as the number of extracted matches is concerned, Brisk outperforms the other algorithms, in some cases acquiring 50-100 times more features than the other methods. On the other hand, although the other algorithms find less features (on average ten times less than Brisk), they usually achieve a similar matching accuracy. For this reason, in the following, only the results obtained using Brisk detector will be presented. All the orthophotos have been generated with a ground resolution of 4 cm/pixel: since the average GSD of the images is ca. 3.9 cm, their content is not significantly up- or down-sampled and the residuals obtained in the co-registration test can easily be scaled to any different configuration.

3. RESULTS AND DISCUSSION

Table 2 summarizes the results obtained with (in columns) the different calibration strategies for the three tested image blocks. For Block 2 and Block 3, being the distance-to-object quite small (ca. 5 metres), the structure from motion stage does not exclude tie-points extracted only on different bands of the same multi-image, nor excludes tie-points extracted only on the images acquired from the same position (+/-90° rotated). As stated in section 2.3, these tie-points results in very inaccurate object point coordinates, due to the very small corresponding ray intersection angle. In Table 2, the rows “inter-band filtering” and “base-length filtering” refer, therefore, to bundle-adjustments where such points were removed. Likewise, as far as “Pre-calibration” strategy is concerned, three different calibration sets (for Block 2 and 3), in which the different filtering methods were applied, are considered. The same issue does not show up in Block 4 (Verrayes), since in this case the distance from the object is much higher (approximately 80 m) and Metashape automatically excludes inter-band points (base-length to distance ratio is too small) during SfM.

It is worth noting that removing tie-points with base-length filtering i.e., removing points between rotated images, always makes the results less accurate. However, for some calibration strategy, in particular in pre-calibration or as far as the translation vector between the master and the slave sensors is self-calibrated, the removal of such tie-points does not compromise excessively the quality of the reconstruction. Pre-

calibrated solutions show much less variability and, at least for Block 2, the inter-band filtering seems to provide slightly better results.

It is important to note that according to the results, even if pre-calibration seems the most fail-safe strategy, it never provides the best outcome. This seems to indicate that, in real-world scenarios, it is advisable to perform an on-the-job calibration of the MAIA camera, since its camera model parameters tend to change to some extent over time. At the same time, the geometrical robustness of the image block, as can be expected, is a relevant feature for ensuring that on-the-job calibration provides reliable results. Comparing Block 2 and Block 3 it can be noted that even if the two blocks are not that different (i.e., Block 2 has only an additional image strip) the estimation of a wider set of camera model parameters during bundle adjustment can be numerically unstable. Considering, for instance, only the unfiltered results, it can be noted that the estimation of the full camera model parameters set i.e., both interior and distortion parameters as well as the relative orientation parameters of each single sensor, provides the best results only for Block2. At the same time, for Block 2, comparable results can be obtained using the pre-calibrated parameters of the Brown model (IO and distortion) and limiting the estimation to Tx and/or Rx relative orientation parameters. The very same is true for Block 3 even if, in this case, the estimation of Tx provides the overall best results.

It might seem surprising that in Block 4, the behaviour is the opposite: the best results are obtained including the Brown model's parameters in the bundle adjustment estimation, while considering the additional degrees of freedom represented by Tx and Rx parameters does not achieve optimal results (although still better than the ones using a pure pre-calibrated solution). This, however, can be easily explained considering the actual influence of the estimated position of the camera centre on image blocks with large camera-to-object distances. As long as the distance from the object is small (as in Block 2 and 3), the estimation of sensor-to-sensor translation vector (Tx) provides good precisions and its correlations with other parameters, in particular with the principal point position, should be small. At the same time (and for the same reasons) the influence of such parameters is critical for evaluating the intersection of corresponding image rays and, consequently, the object point position. On the contrary, with large camera-to-object distances, the impact (as well as the estimation precision) of the Tx parameters becomes less important. It is worth noting that, in these contexts, at least in the experiments conducted, Metashape structure from motion routines filter out more inter-band tie-points (i.e., corresponding points found only on different bands of the same multi-layer image) at bigger distances: for instance, in Block 2 and 3 respectively the average number of such points over the total tie-points extracted was 51.8%, while for Block 4 the same ratio is lower, 41.6%. This reduction, in the authors' opinion, is not enough to motivate the negligible influence of Rx estimation in the process and the fact that considering these additional degrees of freedom seems not having any impact on the final CP accuracy. One motivation might be that, at greater distances from the object, Tx and Rx parameters are much more correlated and estimating one set or the other (or both) does not change the final outcome.

As far as Block 4 results are concerned, Figure 5 shows the results obtained in DSM comparisons, which are in very good agreement with the Check Point residuals, although a higher level of discrepancies (between 70 mm and 80 mm i.e., ca. two times the GSD) is found. It should be noted, however, that in these cases the actual differences between the two models are partly due also to the reconstruction errors of the reference DSM. The best results are achieved, as before, implementing

on-the-job calibration strategies that involves the estimation of the Brown's model parameters, while considering Tx and Rx might improve the performances of pre-calibration but, alone, provides sub-optimal reconstruction.

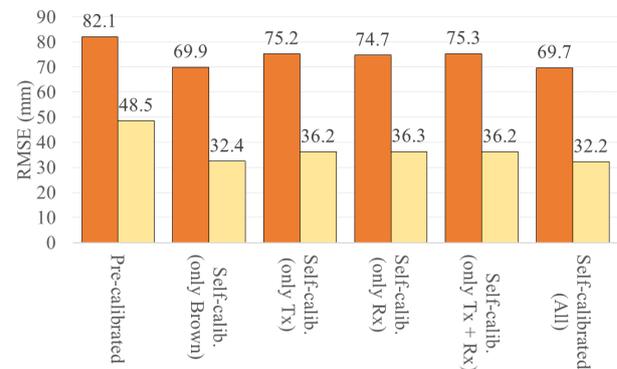


Figure 5. DSM comparison results (orange bars) and RMSE of Check Points (yellow bars) obtained with the different calibration strategies in Block 4.

Finally, the results of orthophoto co-registration are considered. For consistency, for each calibration strategy, the corresponding DSM is used for orthorectification. Although, according to previous results, some strategies obtain more accurate reconstructions than others, the differences (see figure 5) are limited and the use of a DSM rather than another should not influence significantly the results.

As illustrated in section 2.4, only the feature matching results obtained using Brisk will be presented in the following, being the ones with the highest number of matching features between the different band combinations and, at the same time, provides matching scores similar (although not always better) than the other algorithms.

Table 3 reports, as an example, the number of matched features considering the different band combination using the pre-calibration strategy.

	B2	B3	B4	B5	B6	B7	B8	B9
B1	7640	1210	4413	875	14	12	12	5
B2		2190	6628	1246	14	11	11	3
B3			1298	1894	18	16	23	10
B4				1027	4	-	5	-
B5					34	17	19	13
B6						598	431	191
B7							579	196
B8								187

Table 3. number of inter-band matched features for ortho-images generated using the pre-calibration strategy.

It is worth noting that band combinations that involves visible or RedEdge1 bands result in a high number of correspondences (in particular, B1 with B2 and B2 with B4). The lower right section of Table 3 also shows a good correlation (i.e., ability to find a good number of corresponding features) between RedEdge2 and NIR bands.

It could be interesting to compare the results presented in Table 3 with the ones in Table 4, which shows how many tie-points (average per image) are extracted between the different band combinations during the Metashape structure from motion process.

	B2	B3	B4	B5	B6	B7	B8	B9
B1	1326	861	1070	871	15	12	13	12
B2		906	1182	931	20	17	17	16
B3			750	1252	34	29	29	29
B4				848	16	14	13	12
B5					37	31	32	31
B6						2193	2187	2051
B7							2328	2185
B8								2252

Table 4. number of inter-band matched tie-points during the Metashape structure from motion process.

It can be seen that the Metashape procedure shows a similar behaviour with bands B1-B5 well-connected and correlated, as within the bands B6-B9. Actually, in the latter band combinations a higher number of tie-points are extracted with respect to the visible and RedEdge1 bands: in the orthophoto co-registration experiment happened the opposite.

However, the two sets are in both experiments unconnected, with RedEdge1 (B5) having very few correspondences with its “neighbour” RedEdge2 band (and even fewer with NIR bands). This issue does not invalidate the following results, but readers must be advised that the evaluation of co-registration quality between these bands is based on a statistically poor population. Table 5 summarizes the results obtained considering the different calibration strategies. It is worth noting that different strategies produce different number of matches, which in authors opinion is somehow surprising: being the image data always the same, an almost identical number of matched features was expected. However, at the time being, a solid explanation of this fact has not yet been found.

For each strategy the best co-registration result (usually coming from the combinations B3-B5, B6-B7, B2-B3), the worst (combinations B1-B4 and, for pre-calibration, strategy B6-B7) and the average only for the combinations including the first five bands (indicated with “Visible + RedEdge1”) and the average only for the last three bands (indicated with “RedEdge2 + NIR”) are presented.

	Nr. of matches	Best	Worst	Avg. Vis. + RE1	Average (RE2 + NIR)
Pre-calibrated	30844	0.64	1.28	0.80	0.89
Brown only	34106	0.34	0.60	0.50	0.38
Tx only	35580	0.36	0.63	0.51	0.40
Rx only	34777	0.33	0.61	0.51	0.39
Tx + Rx	34419	0.35	0.62	0.50	0.38
All	33613	0.34	0.59	0.50	0.38

Table 5. Orthophoto co-registration residuals, expressed in pixel, for the different calibration strategies (in rows).

The results confirm that the best calibration strategies are the ones that involve estimating the Brown’s parameters. In all the cases, however, the differences between the alternative strategies are indeed limited, with the exception of the pre-calibrated set which provides co-registration results significantly worse than the others.

4. CONCLUSIONS

Accurate co-registration of spectral bands is in most cases performed in image space. Here a simple and straightforward alternative procedure in object space has been presented, that exploits the multi-camera features and the scripting capabilities of commercial software in a general bundle block adjustment of all images and all bands. This approach exploits in a better way the information contents of the bands, as matches are searched (and possibly found) for every pair of bands and not only between the reference and the slave bands as in BBR.

The accuracy of band registration from the bundle-block adjustment was investigated using different combinations of on-the-job calibrated and pre-calibrated camera model and offsets and rotation parameters (Tx and Rx) with respect to a reference sensor. Since the proposed procedure relies on a correct reconstruction of both interior and exterior orientation parameters, sensor calibration plays a critical role in providing good results. This is particular true as far as object point accuracy is concerned, where the geometrical robustness of the image block, the camera-to-object distance and the use of all the structure from motion extracted tie-points, rather than a filtered subset, might influence (in some cases drastically the final outcome). On the contrary, orthophoto co-registration results (Table 5), seem indicating that there is not a clear winning calibration strategy. It is, in this case, advised to perform an on-the-job calibration. In fact, as far as the alternative between using pre-calibrated parameters or let the bundle adjustment to estimate them all, it has been found that, generally speaking, the latter approach yields more accurate results.

Overall, the accuracy figures are well in line or even slightly better than those reported in literature.

The outcome of the analysis is that Brown’s camera model parameters seem the most important parameter set, so they should preferably be left as unknown. Only at very close ranges (a few metres) we found that using the pre-calibrated offsets significantly affects the accuracies (so the offsets should be estimated) while using or not the pre-calibrated values of the rotations hardly influence the accuracy.

As far as the self-sufficiency of the data to complete the procedure, the quality of the DSM generated from the multi-band images seems appropriate (with residuals in the order of two times the GSD). This is another relevant outcome of the experiments: even if the MAIA camera (as other similar multispectral systems currently available) has sensors with low resolution (i.e., 1.2 Mpixel), the achieved accuracy, although not as good as the ones obtainable with higher resolution systems, still reaches a satisfactory level for many environmental survey scenarios, with the undeniable benefit of producing DTM data and acquire multispectral information during the same UAV flight.

The outcome of these tests may be extendable to most of the multispectral multi-camera systems, like MAIA S2, where the optical axes are roughly parallel. Therefore, the ideal imaging geometry for the calibration is similar to that for a single camera. To the contrary, canonical multi-camera equipment features optical axes in a wide range of orientation as they are designed to capture different portions of the scene with a pre-set image overlap. The imaging geometry of the calibration network, therefore, must be adapted accordingly, so the conclusions reported in this paper might not be suitable for standard multi-camera systems.

The procedure has been tested on a single UAV block over a single terrain type (grass field): more tests should therefore be performed to better appreciate its performance in “average” conditions.

ACKNOWLEDGMENTS

The authors acknowledge Cima Research Foundation for providing the MAIA S2 multispectral camera system.

REFERENCES

- Agisoft, 2022. Agisoft Metashape Official Web site. <https://www.agisoft.com/> (last accessed 02.02.2022).
- Alcantarilla, P. F., Bartoli, A., & Davison, A. J., 2012. KAZE features. In *European conference on computer vision* (pp. 214-227). Springer, Berlin, Heidelberg.
- Alcantarilla, P. F., Nuevo, J. & Bartoli, A., 2011. Fast explicit diffusion for accelerated features in nonlinear scale spaces. *IEEE Trans. Patt. Anal. Mach. Intell.*, 34(7), 1281-1298.
- Anuta, P. E., Bartolucci, L. A., Dean, M. E., Lozano, D. F., Malaret, E., McGillem, C. D., ... & Valenzuela, C. R., 1984. Landsat-4 MSS and Thematic Mapper data quality and information content analysis. *IEEE Transactions on Geoscience and Remote Sensing*, (3), 222-236.
- Banerjee, B. P., Raval, S. A., & Cullen, P. J., 2018. Alignment of uav-hyperspectral bands using keypoint descriptors in a spectrally complex environment. *Remote Sensing Letters* 9(6), 524–533.
- Brown, C.D., 1971. Close-range camera calibration. *Photogramm. Eng.*, 37(8), 855-866.
- Cho, W., Schenk, T., & Madani, M., 1993. Resampling digital imagery to epipolar geometry. *Int. Arch. Photogramm. Remote Sens.*, 29, 404-404.
- CloudCompare, 2022. Cloud Compare Official Web site. <http://www.cloudcompare.org/> (last accessed 21.03.2022)
- Dall'Asta, E., Thoeni, K., Santise, M., Forlani, G., Giacomini, A., & Roncella, R., 2015. Network design and quality checks in automatic orientation of close-range photogrammetric blocks. *Sensors*, 15(4), 7985-8008.
- Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2), 91-110.
- Nocerino, E., Dubbini, M., Menna, F., Remondino, F., Gattelli, M., and Covi, D., 2017. Geometric calibration and radiometric correction of the MAIA multispectral camera. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLII-3/W3, 149–156.
- Forlani, G., Diotri, F., Di Cella, U. M., & Roncella, R. (2020). UAV block georeferencing and control by on-board GNSS data. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, 43, 9-16.
- Goforth, M.A., 2006. Sub-pixel registration assessment of multispectral imagery. Proc. SPIE 2006, 6302, 63020S.
- Holtkamp, D. J., Goshtasby, A. A., 2009. Precision registration and mosaicking of multicamera images. *IEEE Transactions on Geoscience and Remote Sensing*, 47(10), 3446-3455.
- Jhan, J. P., Rau, J. Y., Huang, C. Y., 2016. Band-to-band registration and ortho-rectification of multilens/multispectral imagery: A case study of MiniMCA-12 acquired by a fixed-wing UAS. *ISPRS Journal of Photogrammetry and Remote Sensing*, 114, 66-77.
- Kuglin, C.D., Hines, D.C., 1975. "The phase correlation image alignment method", Proc. Int. Conf. Cybernetics Society, 163-165.
- Ladstädter, R., Gruber, M., Wiechert, A., 2010. Monolithic Stitching: One Sensor Geometry for Multiple Sensor Cameras. ASPRS, San Diego, CA, 7 pp.
- Laliberte, A. S., Goforth, M. A., Steele, C. M., Rango, A., 2011. Multispectral remote sensing from unmanned aircraft: Image processing workflows and applications for rangeland environments. *Remote Sens.*, 3(11), 2529-2551.
- Leutenegger, S., Chli, M., & Siegwart, R. Y., 2011. BRISK: Binary robust invariant scalable keypoints. In *Computer Vision (ICCV) 2011. IEEE International conference on*, 2548-2555.
- Rublee, E., Rabaud, V., Konolige, K., & Bradski, G., 2011. ORB: An efficient alternative to SIFT or SURF. In *Computer Vision (ICCV) 2011. IEEE International conference on*, 2564-2571.
- Turner, D., Lucieer, A., Malenovsky, Z., King, D.H., Robinson, S.A., 2014. Spatial Co-Registration of Ultra-High Resolution Visible, Multispectral and Thermal Images Acquired with a Micro-UAV over Antarctic Moss Beds. *Remote Sens.*, 6, 4003-4024.