

SEAMLESS CO-REGISTRATION OF IMAGES FROM MULTI-SENSOR MULTISPECTRAL CAMERAS

M. Shahbazi^{1,*}, C. Cortes²

¹ Centre de géomatique du Québec, Québec, Canada - mshahbazi@cgg.qc.ca

² Dept. of Geomatic Engineering, University of Calgary, Alberta, Canada - cecortes@ucalgary.ca

Commission II

KEY WORDS: multispectral images, fisheye geometry, data fusion, camera calibration, trifocal tensor, co-registration

ABSTRACT:

Small-format, consumer-grade multi-camera multispectral systems have gained popularity in recent years. This is specifically due to the simplicity of their integration onboard platforms with limited payload capacity, such as Unmanned Aerial Vehicles (UAVs). Commercially available photogrammetric software can process the image data collected by these cameras to create multispectral ortho-rectified mosaics. However, misalignments of several pixels between spectral bands have been observed to be a common issue when employing these solutions, which can undermine the spectral and geometric integrity of the data. Besides, in advanced processing workflows such as object detection and classification with deep learning algorithms, band-to-band co-registered images are needed rather than one mosaic. We propose a two-fold solution for seamless band-to-band registration of images captured by five cameras integrated into a miniature multispectral camera system. This approach consists of 1) a robust self-calibration of the multispectral camera system to accurately estimate the intrinsic calibration parameters and relative orientation parameters of all cameras; 2) a single capture, band-to-band co-registration method based on trifocal constraints. This approach differs from existing literature since it is fully automatic, does not make any assumptions about the scene, does not use any best-fit projective or similarity transformations, and does not attempt cross-spectral feature-point matching. Our experiments confirm that the proposed co-registration method can accurately fuse multispectral images from a miniature multi-camera system and is invariant to large depth-variations in the captured scene.

1. INTRODUCTION

With the advancement of digital imaging technologies, multispectral cameras have been developed in small formats at prices affordable by a wide range of end-users. These cameras have, thus, become popular in applications such as wildlife monitoring, forestry, precision agriculture, heritage documentation, disaster management, and even visual odometry (Albetis et al., 2019; Beauvisage et al., 2017; Candiago et al., 2015; Chen et al., 2018; Crutsinger et al., 2016; Dash et al., 2017; Del Pozo et al., 2017; Hassan et al., 2019; Munawar et al., 2019; Wang and Thomasson, 2019).

Based on their structural design, these sensors are classified into single-camera systems with multiple filters (e.g. ADC Lite by Tetracam) and multi-camera systems with multiple lenses and sensor arrays (e.g. ALTUM by MicaSense). Single-camera systems are very similar to conventional RGB cameras in the sense that an array of optical filters cover a single sensor array; e.g. replacing the red filter with a near-infrared filter allows capturing images in blue, green, and near-infrared bands. Besides being light-weight, the main advantage of these systems is that there are only one lens and one sensor. Therefore, there are no misalignments between the created images; that is, they are band-to-band co-registered and ready for application (Sulik and Long, 2016). However, the number of available spectral bands in these cameras is limited.

Multi-camera systems consist of multiple integrated cameras, each of which captures an image in a different spectral band. However, images are captured simultaneously, and the relative orientation parameters between the cameras are considered rigid over their stability period. Each camera in such a system has its

specific interior orientation parameters and distortion patterns. The cameras are also not perfectly aligned and are shifted with respect to one another. Ignoring these differences leads to significant band-to-band misregistration errors (Jhan et al., 2017). Therefore, fusing images of multiple spectral bands is still a topic of ongoing research.

Existing methods to achieve co-registration of multispectral imagery can be divided into single image registration and multiple image ortho-mosaicking (Jhan et al., 2018). Multi-image ortho-mosaicking approaches first use all the images of one camera (a.k.a. the reference camera) to create the dense point cloud and the digital surface model (DSM) of the scene. Then, the DSM is used to create ortho-rectified images of other cameras. The result is a multi-band orthorectified mosaic (Zhang et al., 2016). Commercial solutions, such as MetaShape by AgiSoft, include such processing tools. The pre-requisites for this approach to function well include: 1) accurate knowledge of the intrinsic calibration parameters of all cameras; 2) accurate knowledge of the relative orientation parameters between each camera and the reference one; 3) proper geometric configuration of the reference images to be able to apply photogrammetric 3D reconstruction techniques. While the two first requirements can be achieved via rigorous calibrations, the third one might not be easily accessible in all applications; for instance, when highly oblique images are captured that do not have adequate overlap either. Besides, many applications do not even require 3D reconstruction of the scene, and thus, this approach is a computational burden in such applications. For instance, for image classification using deep learning approaches, numerous single multi-band images are needed rather than one large ortho-mosaic. An overall ortho-rectified

* Corresponding author

mosaic of the scene does also not preserve the original quality of single images in terms of geometric resolution. Small errors in the process of structure-from-motion may also impact the accuracy of this ortho-mosaic, which eventually result in misalignments between the bands.

On the other hand, single image registration methods aim to perform band-to-band registration of images taken by the multispectral cameras at each capture instance (Jhan et al., 2016). This group of approaches avoids the complexities of the previous category. They are useful for creating real-time vegetation index maps and implementation of multispectral image classification techniques (Ampatzidis and Partel, 2019; Knyaz and Zheltov, 2018). The main solution in this approach is to use a global transformation to model the geometric correspondence of all pixels between the bands. Affine transformation (Kelcey and Lucieer, 2012), projective transformation (Jhan et al., 2017), first-order and second-order transformation functions (Yang et al., 2014) are examples of the transformations used in the literature. These approaches either use the calibrated relative orientation parameters between the cameras and/or automatically matched features across the bands to determine the geometric transformation. The primary assumption made in these approaches is that there are inconsiderable depth variations in the scene (compared to the distance of the camera to the scene), and thus the scene can be almost considered as a planar or piece-wise planar object. Therefore, all such approaches fail in the presence of considerable variations in depth, e.g. in close-range imagery (Jhan et al., 2018). These approaches, however, are useful for UAV-based precision agriculture since, compared to the distance of the drone from the scene, the vegetation field can be assumed planar. Some studies attempt sparse, feature-based matching between images in order to determine refined local transformations, instead of a fixed global transformation, for registering the images (Laliberte et al., 2011; Li et al., 2015). These methods require to find sufficient matching points across the bands, e.g. between blue and near-infrared bands. However, the spectral similarities between some bands can be minimal, and also distinct feature points might not be present in the scene. Thus, this requirement may never be fulfilled. Regardless of the approach, the precise knowledge of the relative orientation parameters (ROPs) of the cameras and the intrinsic calibration parameters of each camera are still two critical factors for robust band-to-band registration (Jhan et al., 2018).

The contributions of this paper are twofold. First, we propose a robust multi-camera calibration approach that integrates relative orientations between cameras as additional constraints in the bundle adjustment and handles both fisheye geometry and central perspective geometry at the same time. This is particularly interesting for a camera system, such as Parrot Sequoia, which consists of both fisheye and rectilinear lenses. Second, we propose an approach for single image band-to-band registration that does neither make any assumptions about the scene structure nor attempt cross-spectral feature matching. Instead, it uses dense-matching between two similar bands, e.g. red-edge and near-infrared, to register two of the images. Then, it applies the knowledge of the relative orientation parameters from the calibration step to determine a set of trifocal constraints that register these images to all the other bands automatically.

The rest of this paper is organized as follows. First, self-calibration and co-registration methodologies are described. Second, experiments and results are discussed. Finally, conclusions and future-work suggestions are presented.

2. PROPOSED METHODOLOGY

The proposed method aims to seamlessly co-register images of a multi-sensor multispectral camera system. To achieve this objective, two key processes take place: 1) self-calibration that solves for all the geometric parameters of the integrated cameras, 2) band-to-band registration of multispectral images using trifocal constraints. A Parrot Sequoia multispectral camera is used in the experiments (Figure 1). The camera includes four sensors to capture red (R), green (G), red-edge (REG), and near-infrared (NIR) images. These cameras are equipped with fisheye lenses to provide a larger field-of-view on a smaller sensor. The fifth camera captures regular RGB images and has a rectilinear lens.



Figure 1. a) Parrot Sequoia multispectral camera with its sunshine sensor (source: parrot.com); b) Multi-depth test field used for self-calibration (bottom)

2.1 Self-Calibration

We adopt a self-calibration method that uses a central perspective geometry model for the RGB camera, and a fisheye equidistant projection model for other cameras (G, R, REG, and NIR). Equation 1 describes the augmented co-linearity equations used to describe the central perspective projection geometry.

$$\begin{aligned}
 x_{ij} - c_x + \delta_{x_{ij}} + f \frac{u_{ij}}{w_{ij}} &= 0 \\
 y_{ij} - c_y + \delta_{y_{ij}} + f \frac{v_{ij}}{w_{ij}} &= 0
 \end{aligned}
 \tag{1}$$

where

$$\begin{bmatrix} u_{ij} & v_{ij} & w_{ij} \end{bmatrix}^T = \begin{bmatrix} \mathbf{R}_o^j & -\mathbf{R}_o^j \mathbf{C}_j \end{bmatrix} \tilde{\mathbf{x}}_i$$

i : tie-point index
 j : image index
 o : denoting object coordinate system

In Equation 1, $\mathbf{R}_o^j(\omega_j, \phi_j, \kappa_j)$ is the rotation matrix, \mathbf{C}_j is the position of the image perspective center, $\tilde{\mathbf{X}}_i$ is the homogeneous object coordinates of a tie-point and (x_{ij}, y_{ij}) are the observations of that point in the image coordinate system. Interior orientation parameters of the camera include the principal point offsets and the principal distance, (c_x, c_y, f) . The distortion corrections $(\delta_{x_{ij}}, \delta_{y_{ij}})$ are defined by Equation 2.

$$\begin{aligned}\delta_{x_{ij}} &= (x_{ij} - c_x)(k_1 r_{ij}^2 + k_2 r_{ij}^4 + k_3 r_{ij}^6) \\ &\quad + p_1(r_{ij}^2 + 2(x_{ij} - c_x)^2) + 2p_2(x_{ij} - c_x)(y_{ij} - c_y) \\ &\quad + s_1(x_{ij} - c_x) + s_2(y_{ij} - c_y) \\ \delta_{y_{ij}} &= (y_{ij} - c_y)(k_1 r_{ij}^2 + k_2 r_{ij}^4 + k_3 r_{ij}^6) \\ &\quad + p_2(r_{ij}^2 + 2(y_{ij} - c_y)^2) + 2p_1(x_{ij} - c_x)(y_{ij} - c_y)\end{aligned}\quad (2)$$

where:

$$r_{ij} = \sqrt{(x_{ij} - c_x)^2 + (y_{ij} - c_y)^2}$$

k_1, k_2, k_3 : radial lens distortion parameters

p_1, p_2 : decentering lens distortion parameters

s_1, s_2 : sensor affine distortion parameters

Fisheye geometry cannot be precisely described by Equation 1, even by using higher orders of polynomials to model radial distortions (Jarron et al., 2019). In fisheye cameras, the incidence angle of the projection ray is different from its reflection angle. The physical/mathematical relation between these two angles defines the type of the fisheye projection: equidistant projection, equisolid-angle projection, orthographic projection, and stereographic projection (Schneider et al., 2009). Equation 3 describes the equidistant projection model used in this study to calibrate single-band images of the Sequoia camera system.

$$\begin{aligned}x_{ij} - c_{x_m} + \delta_{x_{ij}} + f_m \frac{u_{ij} \tan^{-1} \left(\frac{\sqrt{u_{ij}^2 + v_{ij}^2}}{w_{ij}} \right)}{\sqrt{u_{ij}^2 + v_{ij}^2}} &= 0 \\ y_{ij} - c_{y_m} + \delta_{y_{ij}} + f_m \frac{v_{ij} \tan^{-1} \left(\frac{\sqrt{u_{ij}^2 + v_{ij}^2}}{w_{ij}} \right)}{\sqrt{u_{ij}^2 + v_{ij}^2}} &= 0\end{aligned}\quad (3)$$

where:

$$m \in \{R, G, REG, NIR\}$$

In addition, relative orientation constraints between the reference camera, in this study set to RGB, and the other four cameras are considered. These relations are described in Equation 4, where \mathbf{R}_{RGB}^m is the 3D relative rotation matrix from RGB image to another band and $\mathbf{r}_{RGB \rightarrow m}^{RGB}$ is the translation vector from the perspective center of the RGB camera to the perspective center of another band resolved in the coordinate system of the RGB camera.

$$\begin{aligned}\mathbf{R}_{RGB}^m &= \mathbf{R}_o^m \mathbf{R}_{RGB}^o \\ \mathbf{r}_{RGB \rightarrow m}^{RGB} &= \mathbf{R}_o^{RGB} (\mathbf{C}_m - \mathbf{C}_{RGB})\end{aligned}\quad (4)$$

where:

$$m \in \{R, G, REG, NIR\}$$

A sparse free-network bundle adjustment solution with Gauss Helmert model is developed that simultaneously solves all the internal orientation parameters, additional distortion coefficients, and relative orientation parameters. The implementation details for this solution can be found in our previous work (Shahbazi et al., 2017).

2.2 Co-registration using Trifocal Constraints

Once the calibration is completed, images can be rectified for their distortions and fisheye effects. Thus, they are transformed to straight-line preserving pinhole camera models. Since the orientations of all cameras with respect to the reference camera are known, their “relative” perspective projection matrices can be expressed as follows.

$$\mathbf{P}_m = \mathbf{K}_m \begin{bmatrix} \mathbf{R}_{RGB}^m & -\mathbf{R}_{RGB}^m \mathbf{r}_{RGB \rightarrow m}^{RGB} \end{bmatrix}\quad (5)$$

where:

\mathbf{K}_m : linear intrinsic calibration matrix

$$m \in \{R, G, RGB, REG, NIR\}$$

Thus, a trifocal constraint can be established between any of the three images. In this study, it will be considered as REG and NIR cameras with R, G, and RGB cameras. The trifocal constraint is simply the coplanarity condition (epipolar constraint) expanded from two views to three views. In a similar way that the fundamental matrix (a 3x3 homogeneous matrix) expresses the epipolar constraint to relate corresponding points in two images, the trifocal tensor (a 3x3x3 matrix) sets the constraints to relate corresponding lines in three images as well as the constraints relating the corresponding points that are on those corresponding lines. Equations 6-10 illustrate how a trifocal tensor (\mathbf{T}) can be derived from the perspective projection matrices corresponding to the three cameras in question.

$$\mathbf{P}_{REG} = [\mathbf{A}_1 \quad \mathbf{B}_1 \quad \mathbf{C}_1] \quad \mathbf{P}_{NIR} = [\mathbf{A}_2 \quad \mathbf{B}_2 \quad \mathbf{C}_2] \quad \mathbf{P}_m = [\mathbf{A}_3 \quad \mathbf{B}_3 \quad \mathbf{C}_3] \quad (6)$$

$$\mathbf{T}_{REG} = - \begin{bmatrix} \det([\mathbf{B}_1, \mathbf{C}_1, \mathbf{A}_2, \mathbf{A}_3]) & \det([\mathbf{B}_1, \mathbf{C}_1, \mathbf{A}_2, \mathbf{A}_3]) & \det([\mathbf{B}_1, \mathbf{C}_1, \mathbf{A}_2, \mathbf{A}_3]) \\ \det([\mathbf{B}_1, \mathbf{C}_1, \mathbf{B}_2, \mathbf{A}_3]) & \det([\mathbf{B}_1, \mathbf{C}_1, \mathbf{B}_2, \mathbf{A}_3]) & \det([\mathbf{B}_1, \mathbf{C}_1, \mathbf{B}_2, \mathbf{A}_3]) \\ \det([\mathbf{B}_1, \mathbf{C}_1, \mathbf{C}_2, \mathbf{B}_3]) & \det([\mathbf{B}_1, \mathbf{C}_1, \mathbf{C}_2, \mathbf{B}_3]) & \det([\mathbf{B}_1, \mathbf{C}_1, \mathbf{C}_2, \mathbf{B}_3]) \end{bmatrix} \quad (7)$$

$$\mathbf{T}_{NIR} = - \begin{bmatrix} \det([\mathbf{C}_1, \mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3]) & \det([\mathbf{C}_1, \mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3]) & \det([\mathbf{C}_1, \mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3]) \\ \det([\mathbf{C}_1, \mathbf{A}_1, \mathbf{B}_2, \mathbf{A}_3]) & \det([\mathbf{C}_1, \mathbf{A}_1, \mathbf{B}_2, \mathbf{A}_3]) & \det([\mathbf{C}_1, \mathbf{A}_1, \mathbf{B}_2, \mathbf{A}_3]) \\ \det([\mathbf{C}_1, \mathbf{A}_1, \mathbf{C}_2, \mathbf{B}_3]) & \det([\mathbf{C}_1, \mathbf{A}_1, \mathbf{C}_2, \mathbf{B}_3]) & \det([\mathbf{C}_1, \mathbf{A}_1, \mathbf{C}_2, \mathbf{B}_3]) \end{bmatrix} \quad (8)$$

$$\mathbf{T}_m = - \begin{bmatrix} \det([\mathbf{A}_1, \mathbf{B}_1, \mathbf{A}_2, \mathbf{A}_3]) & \det([\mathbf{A}_1, \mathbf{B}_1, \mathbf{A}_2, \mathbf{A}_3]) & \det([\mathbf{A}_1, \mathbf{B}_1, \mathbf{A}_2, \mathbf{A}_3]) \\ \det([\mathbf{A}_1, \mathbf{B}_1, \mathbf{B}_2, \mathbf{A}_3]) & \det([\mathbf{A}_1, \mathbf{B}_1, \mathbf{B}_2, \mathbf{A}_3]) & \det([\mathbf{A}_1, \mathbf{B}_1, \mathbf{B}_2, \mathbf{A}_3]) \\ \det([\mathbf{A}_1, \mathbf{B}_1, \mathbf{C}_2, \mathbf{B}_3]) & \det([\mathbf{A}_1, \mathbf{B}_1, \mathbf{C}_2, \mathbf{B}_3]) & \det([\mathbf{A}_1, \mathbf{B}_1, \mathbf{C}_2, \mathbf{B}_3]) \end{bmatrix} \quad (9)$$

$$\mathbf{T} = [\mathbf{T}_{REG} \cdot \mathbf{T}_{NIR} \cdot \mathbf{T}_m] \quad (10)$$

If the trifocal tensor between three images and the correspondences of a point in two images are known, then the

point correspondence in the third image can be determined via Equation 11 (Förstner and Wrobel, 2016). We exploit this relationship by first performing dense matching between the REG and NIR images to obtain the initial corresponding points between them. These bands are chosen for this step due to their closeness in wavelengths, which makes their images spectrally comparable to each other and very simple to match. A fast approach of dense matching based on intrinsic curves is used for this purpose. The implementation details of this dense matching approach can be found in our previous work (Shahbazi et al., 2018). Once the disparity map is generated between the REG and NIR images, any gaps in the disparity map (mainly happening due to either lack of spectral correspondence or actual occlusions) are filled by inward interpolations. These regions are very small in area and are seamlessly filled with this interpolation. Once the point matches between these two images are obtained, the corresponding points on the other images are found utilizing the trifocal tensor relating them (Equation 11). This method does not assume the scene to be planar and is, therefore, invariant to depth variations. This makes it suitable for applications where objects of interest might be present both near to and far from the point of view.

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \tilde{x}_{i,NIR_j} \\ \tilde{x}_{i,REG_j} \end{bmatrix}_x \tilde{\mathbf{T}}(\tilde{x}_{i,REG_j}) \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \tilde{x}_{i,m_j} \\ \tilde{x}_{i,REG_j} \end{bmatrix}_x^T = 0$$

where:

$$m \in \{R, G, RGB\} \quad (11)$$

$$\tilde{x}_{i,m_j} = [u_{i,m_j}, v_{i,m_j}, w_{i,m_j}]^T : \text{homogeneous representation}$$

of the coordinates of point i in image j of camera m

$$\tilde{\mathbf{T}}(\tilde{x}_{i,REG_j}) = u_{i,REG_j} \mathbf{T}_{REG} + v_{i,REG_j} \mathbf{T}_{NIR} + w_{i,REG_j} \mathbf{T}_m$$

3. EXPERIMENTS AND RESULTS

3.1 Self-Calibration

To perform the calibration, a multi-depth test field (Figure 1) containing a checkerboard pattern (with 72 corners) and 12 circular targets was used. A total of 24 poses were taken, resulting in 120 images. A total of 9342 image points were observed in these images. During the self-calibration of the multispectral camera, the RGB sensor was used as the reference camera to determine the ROPs of the remaining cameras within the bundle adjustment, as summarized in Table 1. The relative translations and rotations were estimated with standard deviations of 1 mm and 0.016 degrees, respectively.

Image	Relative Translation (cm)			Relative Rotation (deg.)		
	x	y	z	x	y	z
G	-2.681	-0.818	-0.194	-0.080	-0.053	0.177
R	-2.712	0.743	-0.327	-0.164	0.097	0.117
REG	-1.199	-0.724	-0.490	-0.025	0.229	0.095
NIR	-1.217	0.720	-0.334	-0.456	0.114	0.141

Table 1. Results of relative orientation parameters from self-calibration bundle adjustment

To assess the precision of the calibration, the residuals of all observations at radial and tangential directions were derived. The standard deviation and the root mean square of these residuals are all shown to be in the sub-pixel level, indicating a reliable solution of the calibration parameters (Table 2).

Statistics (in pixels)	tangential residual	radial residual
Std. Dev.	0.048	0.053
RMS	0.075	0.053

Table 2. Image residual from the calibration results

3.2 Co-registration

The proposed method was tested on images of an outdoor scene since the noise of these sensors increases under indoor illumination conditions. Figure 2 shows individual images captured by the Sequoia camera. One can notice the spectral resemblance of red-edge and near-infrared images, which justifies our choice of these two images as references for trifocal constraints. The results of rectification, to turn camera models to straight-line preserving, are shown in Figure 3. It can be seen that all nonlinear distortions are removed; for example, notice the flooring tiles. We chose to keep the original size of the images; this causes marginal edges of the images captured by the fisheye cameras to be removed after rectification. Please note that the staircase in these images has small physical curves, and it should not be misinterpreted as remaining radial distortions on rectified images.



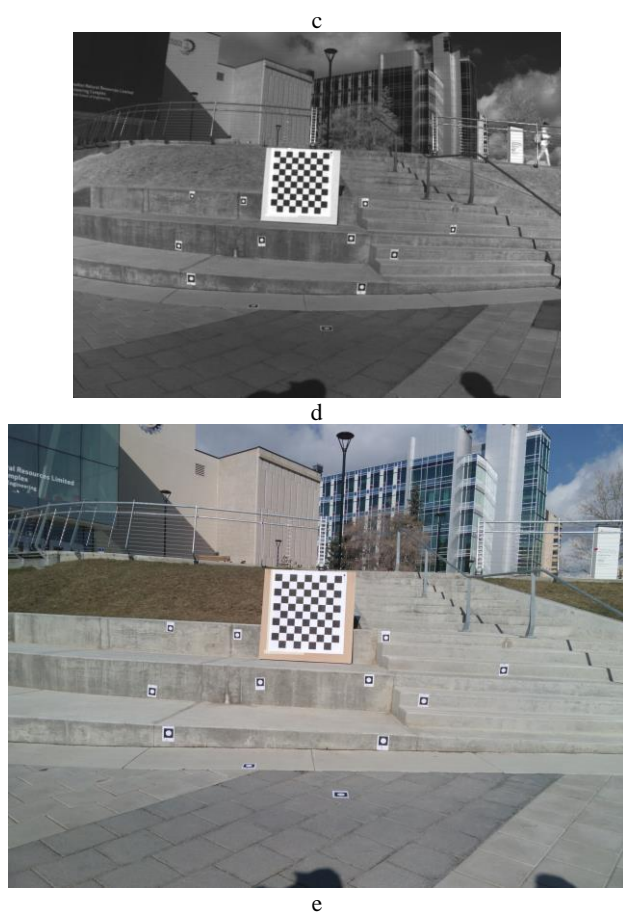


Figure 2. raw images captured by the camera; a) Red band; b) Green band; c) Red-edge band; d) Near-infrared band; e) RGB image

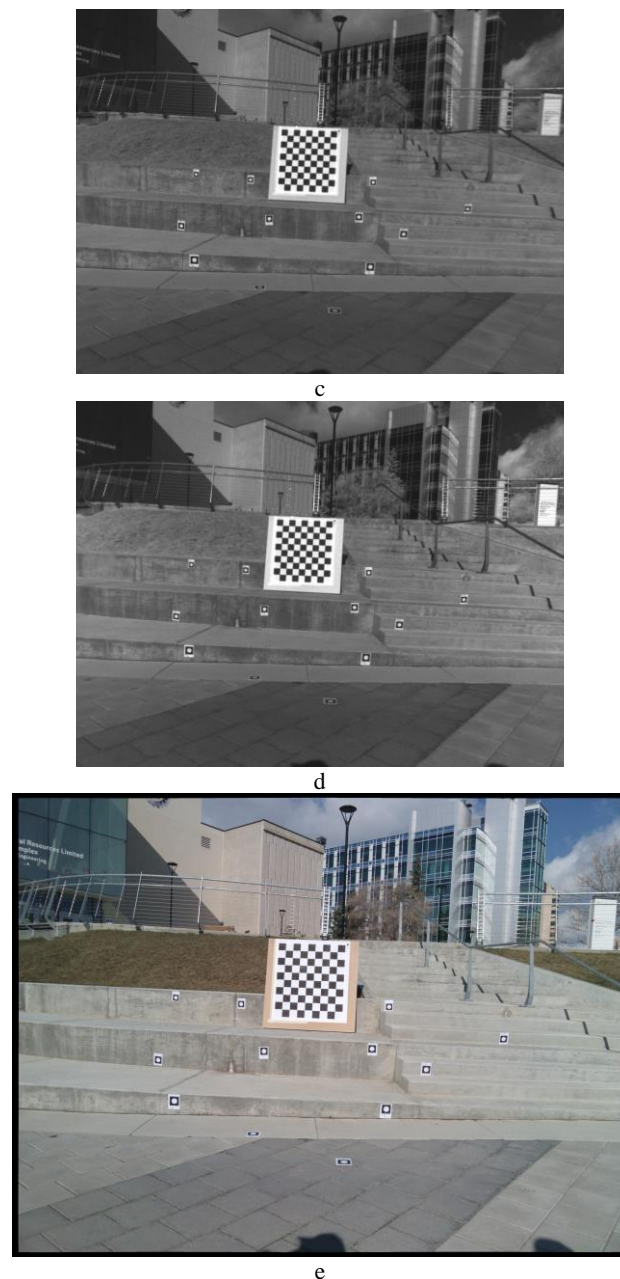
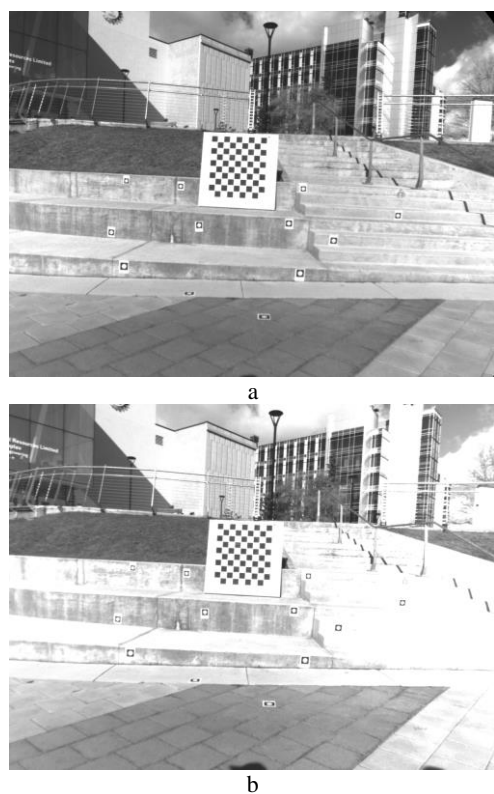


Figure 3. Rectified images; a) Red band; b) Green band; c) Red-edge band; d) Near-infrared band; e) RGB image

Figure 4 shows the epipolar resampled REG and NIR images, as well as the disparity map, resulted from their dense matching.

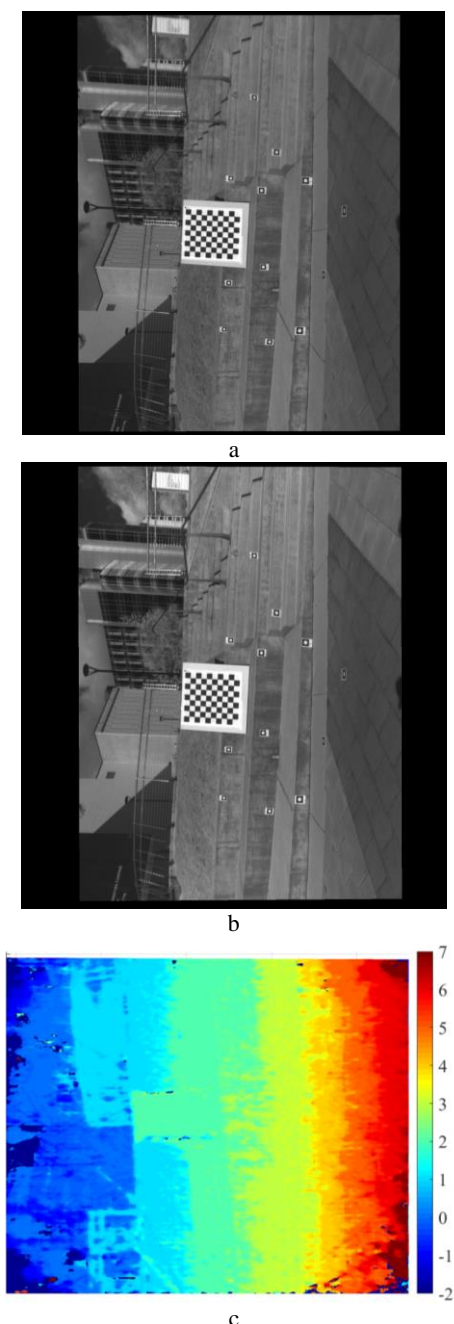


Figure 4. a) REG image after epipolar resampling; b) NIR image after epipolar resampling; c) disparity map (unit: pixels) generated by dense matching

Figure 5 shows a zoomed-in area of an image, before and after co-registration, which captures objects in the scene at close-range (checkerboard), mid-range (handrail), and long-range (background wall) distances from the sensor. For clarity purposes, the images before and after co-registration were compared between two bands at a time, with the REG band as an anchor: REG-G (Figure 5.a), REG-NIR (Figure 5.b), and REG-R (Figure 5.c). As Figure 5 illustrates, the ghosting effect is quite evident in all three image-pairs before co-registration (left side of Figure 5). On the contrary, after co-registration (right side of Figure 5), the images appear to overlay seamlessly.

To quantify the accuracy of the proposed method, band intensity profiles across well-defined edges of objects near and far from

the camera system were studied. The edge between contrasting surfaces in the image is defined for each spectral band as the pixel value intercepting the mean intensity across the one-dimensional profile. A mean edge and standard deviation were calculated based on the edges pertaining to each of the spectral bands. Figure 6 shows the results from a section across the checkerboard (top) and the background wall (bottom) with misalignments between spectral bands of only 0.15 and 0.3 pixels, respectively. It is worth noting that, in general, it is very difficult to quantify the band-to-band co-registration errors. For example, along a white-black edge, there is always some missing spectral information in each of the bands compared to another, which may be confused with a lack of accurate registration.

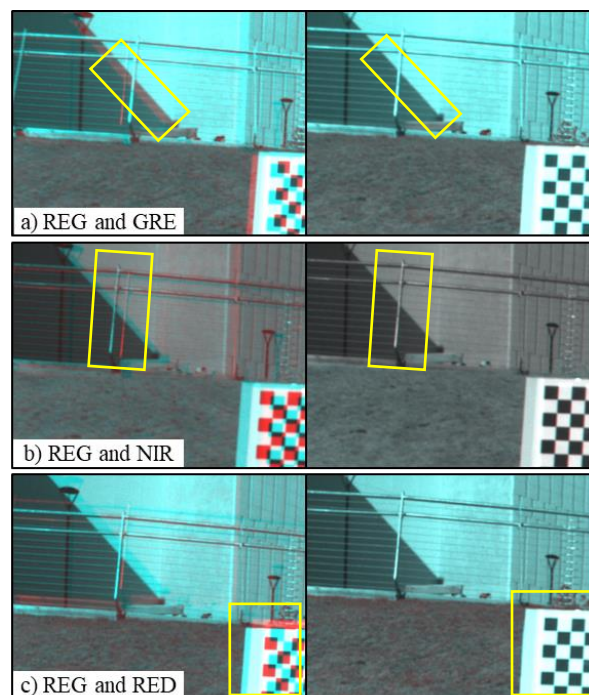


Figure 5. Images showing before (left) and after (right) co-registration between the REG band (in red) and the other three bands (in cyan)

Another example of applying this technique is shown in the images captured from an agriculture field (Figure 7). Similarly, it can be noticed that the objects at far-depth, like side trees, or the ones closer to the camera, like the herbs on the ground, are equally well co-registered through the bands. This success is the result of the robustness of our co-registration approach to the scene structure. These band-to-band registered images can be readily used for real-time, single-shot vegetation index mapping or as input to various classification workflows, e.g. convolutional neural networks.

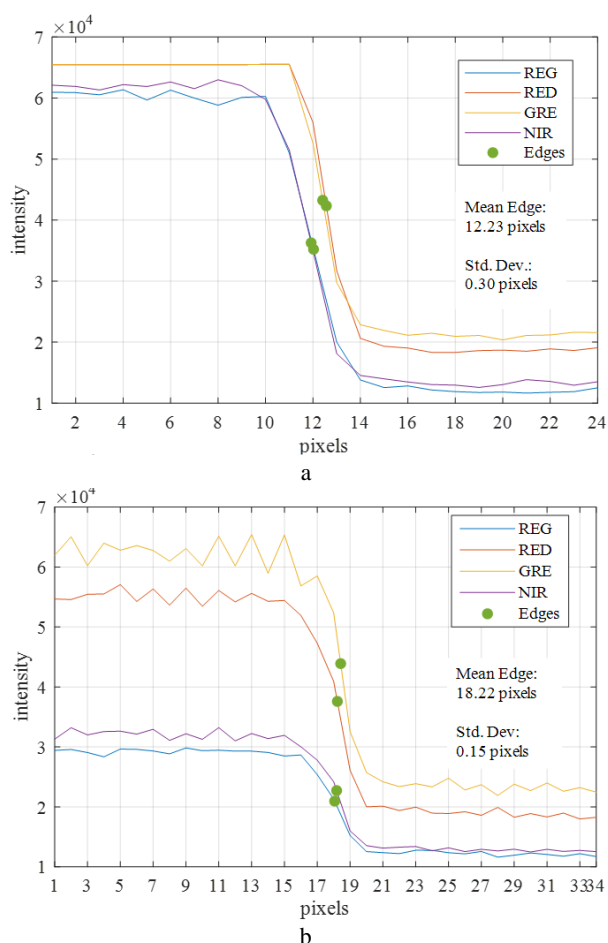
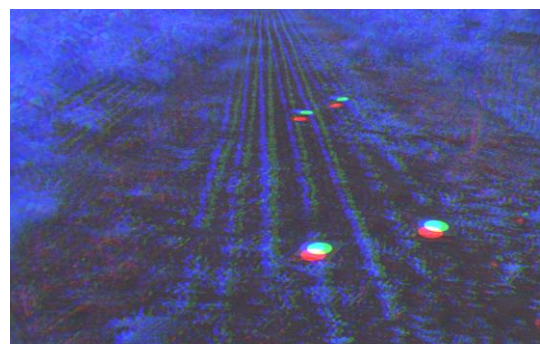


Figure 6: a) Intensity profiles of the spectral bands across a section of the checkerboard pictured in Figure 5c; b) Intensity profiles of the spectral bands across a section of the background wall in Figure 5a (transitioning from sunny to shaded areas)



a



b



c

Figure 7. Example of band-to-band multispectral image registration; a) raw RGB image; b) false-color image generated by overlaying R, G, NIR bands before co-registration; c) false-color image generated by co-registered R, G, NIR bands

4. CONCLUSIONS

This paper proposes a method to co-register images from a Parrot Sequoia multispectral camera based on trifocal tensor estimations through a rigorous calibration procedure. Initial results indicate that this co-registration method is adaptable to both close-range and aerial applications with an accuracy higher than 0.3 pixels. It is also fully automatic and does not require any inputs from the user. Previous methods use different forms of transformations (e.g. homographies) to describe the co-registration model in a global way. However, such models best describe projections from a projective space to another; i.e. the world space is assumed (quasi-)planar. This assumption is generally correct for close-to-nadir airborne imagery of low-relief terrains. However, they do fail to perform co-registration for close-range imagery, oblique imagery, and aerial imagery of high-relief terrains (Jhan et al., 2018). In the literature, some “correction” methods are introduced to reduce these co-registration errors. However, these methods depend on matching features from the reference image against all the other images. This cross-spectral matching introduces a problem since features visible in one band may not even appear in another band. In contrast, the proposed approach does not make any assumptions about the scene structure and only requires dense matching between two spectrally similar bands. Therefore, it is applicable to imagery captured from all types of scenes regardless of their geometric or radiometric properties.

5. REFERENCES

Albetis, J., Jacquin, A., Goulard, M., Poilvé, H., Rousseau, J., Clenet, H., Dedieu, G., Duthoit, S., 2019: On the potentiality of

- UAV multispectral imagery to detect Flavescence dorée and Grapevine Trunk Diseases. *Remote Sensing*, 11(1), 23-49.
- Ampatzidis, Y., Partel, V., 2019: UAV-based high throughput phenotyping in citrus utilizing multispectral imaging and artificial intelligence. *Remote Sensing*, 11(4), 410-429.
- Beauvisage, A., Aouf, N., Courtois, H., 2017. Multi-spectral visual odometry for unmanned air vehicles. *IEEE Proceedings of International Conference on Systems, Man, and Cybernetics*, 1994-1999. doi.org/10.1109/SMC.2016.7844533.
- Candiago, S., Remondino, F., De Giglio, M., Dubbini, M., Gattelli, M., 2015: Evaluating multispectral images and vegetation indices for precision farming applications from UAV images. *Remote Sensing*, 7(4), 4026-4047.
- Chen, F., Song, Y., Zhu, S., Li, J., Wang, C., Zhu, X., 2018. Preliminary Comparison of the Multispectral Cameras Onboard UAV Platform for Environment Monitoring. *IEEE Proceedings of International Conference on Geoinformatics*. 1-4, doi.org/10.1109/GEOINFORMATICS.2018.8557041.
- Crutsinger, G.M., Short, J., Sollenberger, R., 2016: The future of UAVs in ecology: an insider perspective from the Silicon Valley drone industry. *Journal of Unmanned Vehicle Systems*, 4(3), 161-168.
- Dash, J.P., Watt, M.S., Pearse, G.D., Heaphy, M., Dungey, H.S., 2017: Assessing very high resolution UAV imagery for monitoring forest health during a simulated disease outbreak. *ISPRS Journal of Photogrammetry and Remote Sensing*, 131(2017), 1-14.
- Del Pozo, S., Rodríguez-González, P., Sánchez-Aparicio, L.J., Muñoz-Nieto, A., Hernández-López, D., Felipe-García, B., González-Aguilera, D., 2017. Multispectral imaging in cultural heritage conservation, in: *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLII-2/W5, 155-162. doi.org/10.5194/isprs-archives-XLII-2-W5-155-2017.
- Förstner, W., Wrobel, B., 2016: *Photogrammetric Computer Vision*. Springer Nature, Cham.
- Hassan, M.A., Yang, M., Rasheed, A., Yang, G., Reynolds, M., Xia, X., Xiao, Y., He, Z., 2019: A rapid monitoring of NDVI across the wheat growth cycle for grain yield prediction using a multi-spectral UAV platform. *Plant science*, 282(2019), 95-103.
- Jarron, D., Shahbazi, M., Lichti, D., Radovanovic, R., 2019. Modelling wide-angle lens cameras for metrology and mapping applications. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, IV-2/W7, 79–86. doi.org/10.5194/isprs-annals-iv-2-w7-79-2019.
- Jhan, J.P., Rau, J.Y., Haala, N., 2018: Robust and adaptive band-to-band image transform of UAS miniature multi-lens multispectral camera. *ISPRS Journal of Photogrammetry and Remote Sensing*, 137(2018), 47-60.
- Jhan, J.P., Rau, J.Y., Haala, N., Cramer, M., 2017. Investigation of parallax issues for multi-lens multispectral camera band co-registration. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLII-2/W6, 157-163. doi.org/10.5194/isprs-archives-XLII-2-W6-157-2017.
- Kelcey, J., Lucieer, A., 2012: Sensor correction of a 6-band multispectral imaging sensor for UAV remote sensing. *Remote Sensing*, 4(5), 1462-93.
- Knyaz, V.A., Zheltov, S., 2018. Deep learning object recognition in multi-spectral UAV imagery. *Proceedings of Optics, Photonics, and Digital Technologies for Imaging Applications*, 1067920(2018). doi.org/10.1117/12.2307661.
- 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 Sensing*, 3(11), 2529-2551.
- Li, H., Zhang, A., Hu, S., 2015: A multispectral image creating method for a new airborne four-camera system with different bandpass filters. *Sensors*, 15(7), 17453-17469.
- Munawar, H.S., Zhang, J., Li, H., Mo, D., Chang, L., 2019: Mining Multispectral Aerial Images for Automatic Detection of Strategic Bridge Locations for Disaster Relief Missions. *Lecture Notes in Computer Science*, 11607, 189-200.
- Schneider, D., Schwalbe, E., Maas, H.G., 2009: Validation of geometric models for fisheye lenses. *ISPRS Journal of Photogrammetry and Remote Sensing*, 64(3), 259-266.
- Shahbazi, M., Sohn, G., Théau, J., 2018: High-density stereo image matching using intrinsic curves. *ISPRS journal of photogrammetry and remote sensing*, 146, 373-388.
- Shahbazi, M., Sohn, G., Théau, J., Ménard, P., 2017: Robust Structure-from-Motion Computation: Application to Open-Pit Mine Surveying from Unmanned Aerial Images. *Journal of Unmanned Vehicle Systems*, 5(4), 126-145.
- Sulik, J.J., Long, D.S., 2016: Spectral considerations for modeling yield of canola. *Remote Sensing of Environment*, 184(2016), 161-174.
- Wang, T., Thomasson, J.A., 2019. Plant-by-plant level classifications of cotton root rot by UAV remote sensing. *Proceedings of Autonomous Air and Ground Sensing Systems for Agricultural Optimization*, 110080N. doi.org/10.1117/12.2519394.
- Yang, C., Westbrook, J.K., Suh, C.P.C., Martin, D.E., Hoffmann, W.C., Lan, Y., Fritz, B.K., Goolsby, J.A., 2014: An airborne multispectral imaging system based on two consumer-grade cameras for agricultural remote sensing. *Remote Sensing*, 6(6), 5257-5278.
- Zhang, J., Yang, C., Song, H., Hoffmann, W.C., Zhang, D., Zhang, G., 2016: Evaluation of an airborne remote sensing platform consisting of two consumer-grade cameras for crop identification. *Remote Sensing*, 8(3), 257-280.