International Journal of Applied Earth Observations and Geoinformation 97 (2021) 102274

Contents lists available at ScienceDirect

International Journal of Applied Earth Observations and Geoinformation

journal homepage: www.elsevier.com/locate/jag

A framework for registering UAV-based imagery for crop-tracking in Precision Agriculture

Alfonso López *, Juan M. Jurado, Carlos J. Ogayar, Francisco R. Feito

Department of Computer Science, University of Jaén, EPS Jaén, 23071, Spain

ARTICLE INFO

Keywords:
Image registration
Multispectral imagery
Thermal imagery
Tree recognition

ABSTRACT

Multiple types of images provide useful information about a crop, but image fusion is still a challenge in Precision Agriculture (PA). We describe a framework which manages a multi-layer registration model of heterogeneous images obtained by an unmanned aerial vehicle (UAV) by proposing pair-to-pair steps through a registration method invariant to intensity differences, allowing us to connect different aerial images with significant differences. Correction of deformed images is treated as a first step to end up with our registration algorithms. These methods conform the base of more advanced systems that combine 2D and spatial information, therefore it represents the link of several types of images. The evaluation shows the flexibility of our framework when dealing with different requirements. Effectiveness of the Enhanced Correlation Coefficient method is proved and thus shown as a suitable method for the registration of heterogeneous images.

1. Introduction

Precision agriculture ensures the development of accurate farming management practices. Through these activities, crop yield is maximised, and environmental impact is reduced by optimising the usage of pesticides and fertilisers (Pablo et al., 2014). The monitoring of crops maintains control of many indicators, such as the state of health, the amount of water and fertilisers, and possible infections (Zhang and Kovacs, 2012).

Remote sensing techniques have been used for decades as a part of PA practices for crop and soil monitoring. It was recently enhanced by the reduction in price of sensors (RGB, multispectral, hyperspectral, thermal). The development of Unmanned Aerial vehicles (UAV) is also playing a key role in PA monitoring (Tang et al., 2020). The outcome of computations based on several sources of information frequently consists of orthomosaic maps, vegetation indices or point clouds.

Regarding UAV-based multispectral and hyperspectral imaging, their applications in PA cover the detection of diseases, classification of weeds, and estimation of water, biomass or chlorophyll (Lu et al., 2020). These optical Remote Sensing (RS) methods are known to be slower at differentiating properties such as stress levels unless visual symptoms are noticeable. On the other hand, surface temperature obtains more accurate results on stress levels and faster results (Khanal et al., 2017). Even though the application of thermal imaging has been limited, it is considered as a promising tool in PA. The combination of both types of information yields a more complete view of a crop. For instance, complementarity of multispectral or hyperspectral data with thermal data helps to discriminate infected areas from healthy and low leaf areas through vegetation indices such as NDVI (Normalised Difference Vegetation Index) (Maes and Steppe, 2019). This complementarity is also extended for RGB images, which are mostly used for weed and logging detection.

By combining images from up to four sensors (RGB, multispectral, hyperspectral and thermal) we can cover most of the applications of UAV remote sensing. Registration of sensor data is a common topic in PA, even for multi-temporal applications. Moreover, 3D information is frequently combined with sensor imagery. Most of the works found merge at most two types of UAV-based information. Regarding RGB and multispectral imaging, (Garcia et al., 2020) estimates corn grain yield from vegetation indices, plant density and canopy cover using multispectral and RGB images along with a neural network model. Some applications rely on ground control points (GCP) and computing of separate orthomosaic maps to estimate nitrogen accumulation (Zheng et al., 2018). Other works decompose RGB images and apply wavelet transformations which end up affecting their colour information (Barrero and Perdomo, 2018). Additionally, several studies manage to map RGB point clouds with spectral information (Jurado et al., 2020) through the registration of high-resolution RGB and multispectral

* Corresponding author.
E-mail address: allopezr@ujaen.es (A. López).

https://doi.org/10.1016/j.jag.2020.102274
Received 14 July 2020; Received in revised form 15 November 2020; Accepted 24 November 2020
Available online 12 January 2021

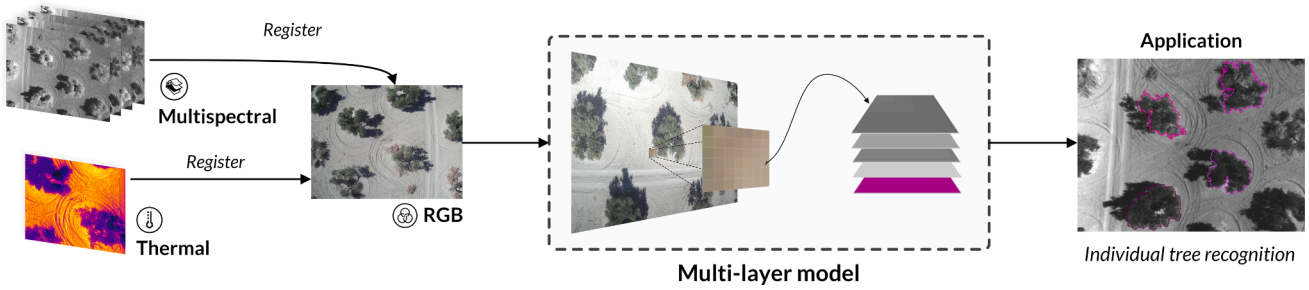

Fig. 1. Overview of the framework methodology as well as the multi-layer model composed of RGB, multispectral and thermal imagery.

imagery, along with photogrammetry.

Multispectral images do not lack differences either, as they are retrieved by separate sensors which do not share viewing angles or perspective centres. Some works can be found in this research area, either using multispectral or hyperspectral imagery. (Shen et al., 2014) introduces an image-matching solution based on a descriptor that establishes dense pixel correspondence in input images. (Jhan et al., 2016) relies on image metadata to fix the misalignment, while (Jhan et al., 2017) also uses parameter estimating methods such as Random Sample Consensus (RANSAC) to minimize the error induced by calibration uncertainty. Finally, (Hakim et al., 2018) combines image-matching and metadata approaches in a supervised model.

Registration of RGB and thermal imagery is not a trivial task either (Bavirisetti and Dhuli, 2015). In this scope, most of the PA works frequently avoid merging them. (Liu et al., 2018; Tucci et al., 2019) relies on ground control points and external software, such as Pix4D-mapper or Photoscan Agisoft. Then, a registration process is exchanged for an overlapping of orthomosaic maps. Development of specific platforms which integrates thermal and RGB modules is also well suited to monitor a crop, as the differences between images are previously known and constant (Osroosh et al., 2018). Literature related to RGB and thermal registration shows that there exist studies which combine multispectral, temperature and RGB information, but they are mainly focused on the extraction of separate orthomosaic maps. Thus, the registration algorithm lacks scientific interest as it is already solved by professional software (Santini et al., 2019; Matese and Di Gennaro, 2018).

We can conclude that PA is an active research area which involves many groups, mainly focused on the search of new applications and the development of algorithms which extract useful information from a crop. Most of them are not developed from scratch and they are also fed with data which has been previously processed by other software solutions. Although accuracy and fine-grain details are relevant for PA applications, many of the cited articles rely on the overlapping of multiple orthomosaic maps, extracted from the fusion of a large number of images. Also, these solutions depend on software capacity to generate an orthomosaic map from any type of image. Moreover, some of the methods found in the literature need ground control points to guide the registration process. Thus, a fully automatic registration algorithm which does not rely on previous solutions is highly suitable for PA applications involving multiple data sources.

In this article, we present an automatic framework to register several types of UAV-based images that do not need to be preprocessed. Crops are not registered using overlapped orthomosaic maps but through the registration of pairs of images, as shown in Fig. 1. The solution takes advantage of an algorithm invariant to photometric distortions that is well suited to register heterogeneous images. Therefore, our framework works with low-level implementations that correct and register the original UAV-based images. Finally, we propose a naive application that uses the result of a registration process.

2. Material and methods

This section describes UAV sensors, acquired datasets, and methods developed for fusing images from multiple sources. The methodology is based on three main stages: (1) registration of multispectral images, (2) registration of multispectral and RGB imagery and (3) registration of thermal and RGB imagery. Once the framework is fully described, an application is presented through the registration of multispectral images.

2.1. Study area and data acquisition

Our method takes input data from the monitoring of an olive plantation, which is located in Jaén, a southern region of Spain. The study area covers two hectares of olive trees where the proposed methodology has been tested and optimised. Fig. 2 presents a general overview of the study area.

In this research, four datasets have been acquired using multiple UAV-based sensors in different stages of the olive cycle. The first two datasets contain multispectral images which capture the reflected light in visible and near-infrared ranges. On the other hand, third and fourth datasets include high-resolution RGB imagery and thermal images which are widely used to assess both healthy and under stress olive trees. The third dataset also includes multispectral images. Multispectral, RGB

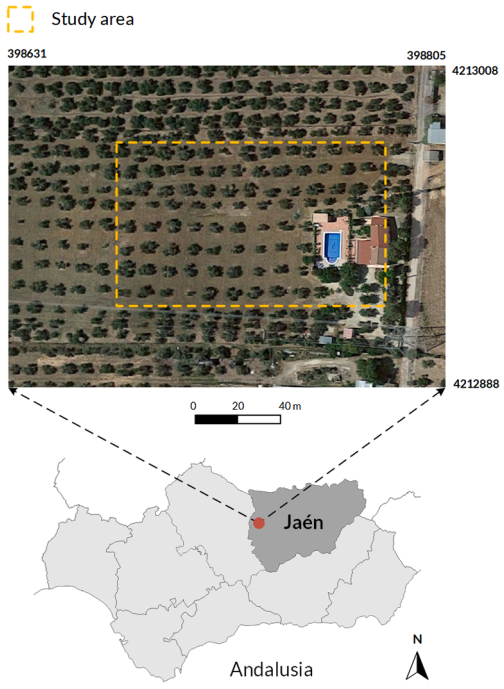

Fig. 2. An overview of the study area. Coordinates are given in UTM (Universal Transverse Mercator) coordinate system.

Table 1
Wavelength range for each multispectral band retrieved by a Parrot Sequoia device.

Multispectral band	Wavelength (nm)
Green (GRE)	[530, 570]
Red (RED)	[640, 680]
Red-edge (REG)	[730, 740]
Near-infrared (NIR)	[770, 810]

and thermal images have been captured with a frontal overlap of 90% and a side overlap of 80%, and the flight height is set as 30 meters. Missions have been planned using DroneDeploy (California, United States of America) in a portable device considering a single-grid configuration.

Regarding acquisition technology, a multispectral sensor (Parrot Sequoia) and a thermal sensor (DJI Zenmuse XT2) are mounted on board of a drone. The multispectral device captures four spectral bands (Table 1) with a focal length of 4 mm. These wide-angle lenses can cover a large area of terrain to capture more plants in a single image. However, these images present high geometric distortions following a fisheye model. The four spectral bands are shown in Fig. 3.

The fifth lens of the multispectral device takes RGB photos (15.9 megapixels) with a focal length of 4.9 mm. This RGB sensor is mounted with a rolling shutter and the quality of resulting images is low due to vibrations associated with the drone flight and the rolling shutter effect. Consequently, these images are not used in this work. Regarding the thermal sensor, it is characterised by two lenses which capture thermal and high-resolution RGB images. The thermal lens has a focal length of 19 mm, which avoids previous deformations and obtains rectilinear images. The second lens has a focal length of 8 mm and it also retrieves high-resolution RGB images. A summary of image dimensions and focal length sizes with which they are taken is shown in Table 2.

2.2. Image correction

The wide-angle lens of many of our sensors allows us to cover a large area of terrain but also causes a high visual deformation known as fisheye distortion (Fig. 4). The correction process is illustrated using multispectral images as input data, but it can be applied to the rest of images which present the fisheye distortion. This effect is mostly visible on the corners of an image, where objects take a circular shape. The correction algorithm consists of a transformation matrix which is used to create a new image with the same size, where its pixels $p_{x,y}$ take the colour of $p_{i,j}$ from the original image. i,j are not necessarily integer values; in fact, $i,j \in \mathbb{R}$.

This relation can also be presented in terms of angles (Fig. 5). A lens forms an angle α between its optical axis and a segment that goes from the source point of the lens, c_o, to any point $p_{x,y}$. Both segments share their origin (c_o), which is known as the optical point. This angle is modified once the image is corrected. β turns into α, and $p_{i,j}$ is translated to $p_{x,y}$. $(i,j) \neq (x,y)$, but it remains on the segment that goes from the principal point, c_p, to $p_{x,y}$. A triangle-similarity can be easily observed in Fig. 5. Pulling the image corners would be an appropriate example for illustrating the correction process. Thus, a smaller translation is applied to those points which are closer to c_p.

The correction process can be performed through two functions, $f(x)$ and $f(y)$, which return the values i and j respectively. Therefore, the angle α between the optical axis and a segment $c_o p_{x,y}$ can be defined as shown in the Eq. 1, where a simple trigonometrical operation is applied.

Table 2
Dimensions in pixels and size of focal length for each type of image.

Sensor	Dimensions (px)	Focal length (mm)
Multispectral	1280 × 960	4
Multispectral (RGB)	4608 × 3456	4.9
Thermal	640 × 512	19 mm
Thermal (RGB)	4000 × 3000	8

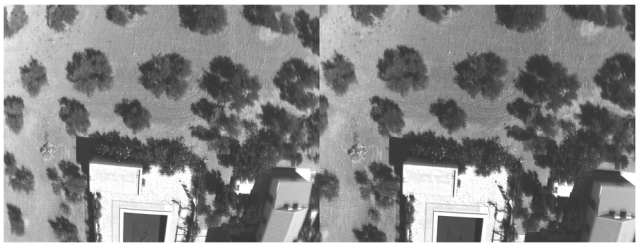

Fig. 4. (a) Original image with fisheye distortion, (b) Corrected image.

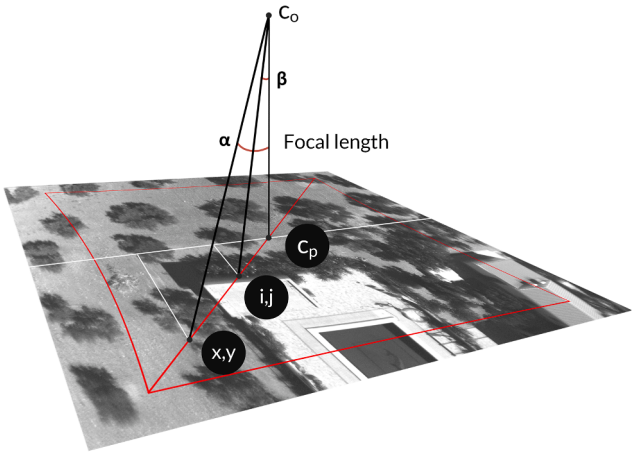

Fig. 5. Representation of fisheye distortion parameters in a multispectral image. The surface of the corrected image is drawn as a deformed red-bordered shape. The relation between a pixel from the corrected image, $p_{x,y}$, and a position from the original image, (i,j), is also presented.

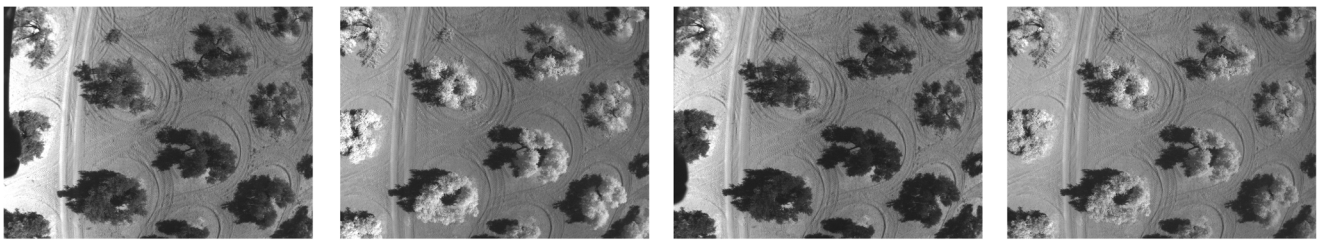

Fig. 3. Multispectral bands of a single capture. (a) Green, (b) Near-infrared, (c) Red, (d) Red-edge.

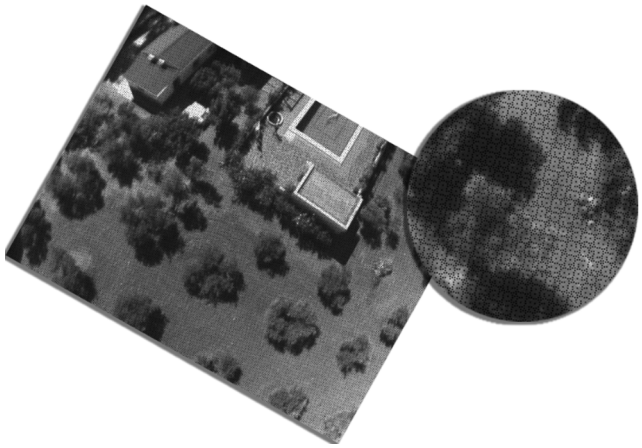

Fig. 6. Pattern obtained when multiple pixels from the first image are mapped into the same position of the resulting image, leaving pixels with null values.

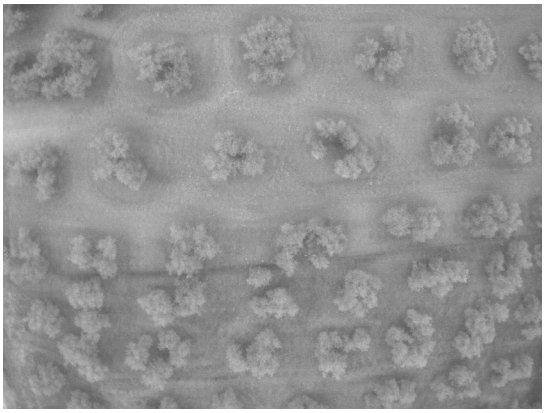

Fig. 7. Effect of image ghosting when overlapping multispectral images using the alpha channel.

Finally, the angle α is normalised in [0, 1]:

$$\alpha = \frac{2}{\pi} * \tan^{-1}\left(\frac{\sqrt{(x - c_{p_x})^2 + (y - c_{p_y})^2}}{f}\right) = \frac{2}{\pi} * \tan^{-1}\left(\frac{r_1}{f}\right) \quad (1)$$

Up to four distortion coefficients (k_1, k_2, k_3, k_4) can be retrieved from the image metadata. These coefficients can be used to compute the angle β between the optical axis and the segment $c_o p_{i,j}$ (Eq. 2).

$$\beta = k_1 + k_2 * \alpha + k_3 * \alpha^2 + k_4 * \alpha^3 \quad (2)$$

Therefore, we obtain a new image with the same size, where $p_{x,y}$ takes the colour of a position $p_{i,j}$ computed using previous equations. This process guarantees that every pixel of the corrected image receives a colour. Otherwise, the inverse mapping (from $p_{i,j}$ to $p_{x,y}$) leaves null values, as two pixels from the original image can be mapped to the same position $x, y \in \mathbb{N}$ (Fig. 6). As a solution, a pixel $p_{i,j}$ would need to modify multiple pixels from the resulting image, where its contribution is given by the distance to pixel centres. However, the process here described only requires a writing operation and multiple reads from neighbour pixels.

The last step of this method computes the colour for each pixel $p_{x,y}$, as it is mapped to a real position, $i, j \in \mathbb{R}$. Hence, a bilinear interpolation algorithm is used to solve this problem.

2.3. Registration of multispectral imagery

Four spectral bands are captured from our multispectral device (Green (GRE), Red (RED), Red Edge (REG) and Near Infrared (NIR)). All observed bands are shown in Fig. 3. Even though these images are taken from the same device, there are significant differences between them, which are mostly visible through an overlapped composition (Fig. 7):

a. There exists a translation since they have been taken by different lenses of the multispectral sensor. Each lens has a static position along the device.
b. Each image has its own optical axis and therefore, its own principal point. The optimal scenario considers that axes are parallel and as a result, a translation solves the misalignment since we know their physical distance. However, the metadata of multispectral imagery shows non-parallel axes through the definition of a camera ring (Rig Relatives and Rig Index attributes), i.e., a set of cameras which are connected and defined by geometric constraints. The origin of the ring is one of the lenses (master), defined by a position T_m and its rotation in the world, R_m, and rest of lenses (secondary cameras) defines their position, T_s, and rotation, R_s, with respect to the master lens.
c. Each lens takes a photo with different timestamps. The time difference can be up to a few milliseconds. Consequently, this delay involves both rotation and translation transformations considering the drone movement.

Registration of multispectral imagery can be performed through metadata, but it relies on calibration and accuracy of the sensor. Some of the previously cited articles use this approach and try to avoid errors with parameter estimating algorithms. The approach here proposed does not rely on image metadata but only on the colour information. Traditional image-matching approaches, such as SIFT (Scale-invariant feature transform), are not suitable for spectral bands or heterogeneous images (e.g. RGB and thermal images), as the colour of a material may not be constant over different images. Although there exists derived solutions from SIFT method to work under photometric differences (Park et al., 2008), they do not seem as robust as the algorithm which is next presented.

Further than SIFT, other image-matching algorithms are invariant to photometric distortions. One of them is Enhanced Correlation Coefficient (ECC) (Evangelidis and Psarakis, 2008). This is a highly suitable algorithm for many reasons. First, it uses up to four motion models: translation, euclidean, affine and homography, from lower to higher computational complexity. Therefore, the motion model can be selected by considering which transformations are the minimum the misalignment responds to, i.e., translation, rotation and scale. Even with the most complex motion model, the algorithm remains linear ($O(n)$). Furthermore, it is possible to select how fine-grained the process is through a precision factor.

Although homography is the safest motion model to register a pair of images, a euclidean model is proved to be enough for the registration of multispectral bands. Also, algorithmic complexity is reduced, as the matrix size is 2x3 instead of 3x3. As described in a previous paragraph, differences between bands can be expressed by means of translation (physical distance), rotation (drone movement) and scaling (due to different perspective views). Translation is the most visible difference, while rotation and scale allow to reach a more accurate registration.

Despite the desirable behaviour of ECC for registering heterogeneous images, the algorithm can be improved by first registering those images which are more similar in terms of intensity. Therefore, we consider a hierarchical process that guarantees success in the registration of multispectral bands. The outcome is given by four images of the same size which do not suffer from the image ghosting effect when overlapped.

ECC yields a motion model, i.e. a matrix, which can be used to transform an image. We can multiply several motion models and apply the result to reach a hierarchical alignment. However, the results contain areas with null values after a transformation. Once again, we can use motion models and apply those matrices to image corners in order to

Table 3
Normalised correlation coefficient between original pairs of multispectral images. This table is also intended to support comparisons with the final results of our methods.

Bands	Dataset 1 ρ	Dataset 2 ρ	Dataset 4 ρ	Avg.
GRE-NIR	0.9182	0.9175	0.9308	**0.9221**
GRE-RED	0.9438	0.9488	0.9516	**0.9480**
GRE-REG	0.9433	0.9464	0.9567	**0.9488**
NIR-RED	0.8904	0.8978	0.9438	**0.9106**
NIR-REG	0.9514	0.9629	0.9630	**0.9591**
RED-REG	0.9028	0.9222	0.9405	**0.9218**

retrieve the part of the image to be discarded. Furthermore, the dimensions of the transformed images might be different, so the minimum non-null area needs to be computed. The final area is defined as follows (Eqs. (3)–(6)), where M_i is the composite transformation for a multispectral band:

$$min_x = max_x \forall i \in [0,3]\{M_i{*}[0,0,1]^\top, M_i{*}[0,h-1,1]^\top, min_x\} \forall i \in [0,3] \quad (3)$$

$$min_y = max_y\{M_i{*}[0,0,1]^\top, M_i{*}[w-1,0,1]^\top, min_y\} \forall i \in [0,3] \quad (4)$$

$$max_x = min_x\{M_i{*}[w-1,0,1]^\top, M_i{*}[w-1,h-1,1]^\top, max_x\} \forall i \in [0,3] \quad (5)$$

$$max_y = min_y\{M_i{*}[0,h-1,1]^\top, M_i{*}[w-1,h-1,1]^\top, max_y\} \forall i \in [0,3] \quad (6)$$

The pairs of images to be registered are selected by using any technique which analyses image similarity, such as normalised correlation coefficient (CC). This method calculates a value $\rho \in [0,1]$ which represents the similarity between two grayscale images, $f(x,y), f'(x,y)$. As intensity values highly differ between bands, CC is used as a measure to guide the process, instead of an exact similarity value. Therefore, we can compute CC between the six pairs of images. For sake of simplicity, GRE image is considered as the master image, just as our device model does with the camera ring.

Three multispectral datasets are evaluated in Table 3 to calculate the average CC between any pair of multispectral bands. As a result, RED and REG are very similar to the master image, while NIR band needs a composite matrix to get registered with GRE image. For that purpose, REG band is aligned with GRE, and its matrix is used to register NIR with GRE. The hierarchy of multispectral registration is defined as shown in Fig. 9.

The minimum area cannot be computed just from lower left and upper right corners, $[0,0]$ and $[w-1,h-1]$, since M_i includes a rotation, i.e., it is not enough to check two non-adjacent corners.

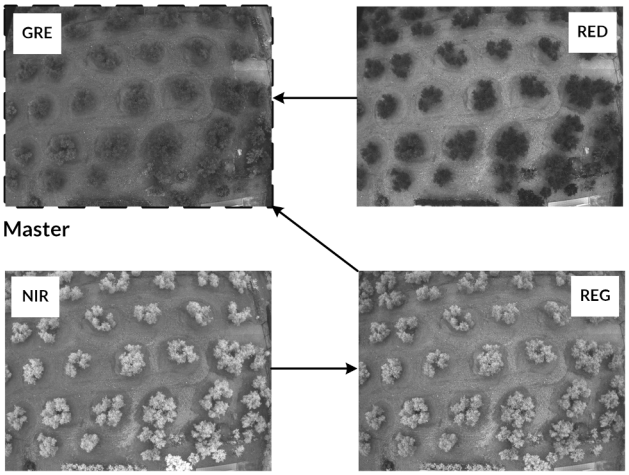

Fig. 9. Hierarchy of multispectral images when registering. NIR band needs 2 steps to get the result.

2.4. Registration of multispectral and RGB images

This section aims to describe the steps which are needed to register RGB images and the previously corrected multispectral captures. The main contribution of RGB imagery to a multi-layer model is the addition of GPS information, beyond any application which was shown in the Introduction. Even though this data is not necessary for our framework, we need to highlight the relevance of this registration step. as the final multi-layer model is georeferenced through this process.

Multispectral and RGB images are not taken from the same device, thus the differences between both types are logically more relevant. Both images are also taken with different timestamps as the devices are not synchronised. Therefore, the time difference may scale to seconds. Also, RGB images are taken from a device with a longer focal length, but it still shows a visual distortion.

The images to be registered are selected by their temporal distance, i.e., for each multispectral capture (composed of four images) we search for the RGB image whose timestamp is closer. This behaviour guarantees our framework registers overlapping images.

As occurred with previous registrations, the ECC algorithm is well suited for this scenario, where images of different intensities are compared. Nevertheless, we need to decide which multispectral band needs to be registered with RGB images. In this case, GRE spectral image is the most appropriate image in terms of intensity (Table 5). However, this decision is also supported by the algorithm efficiency as GRE does

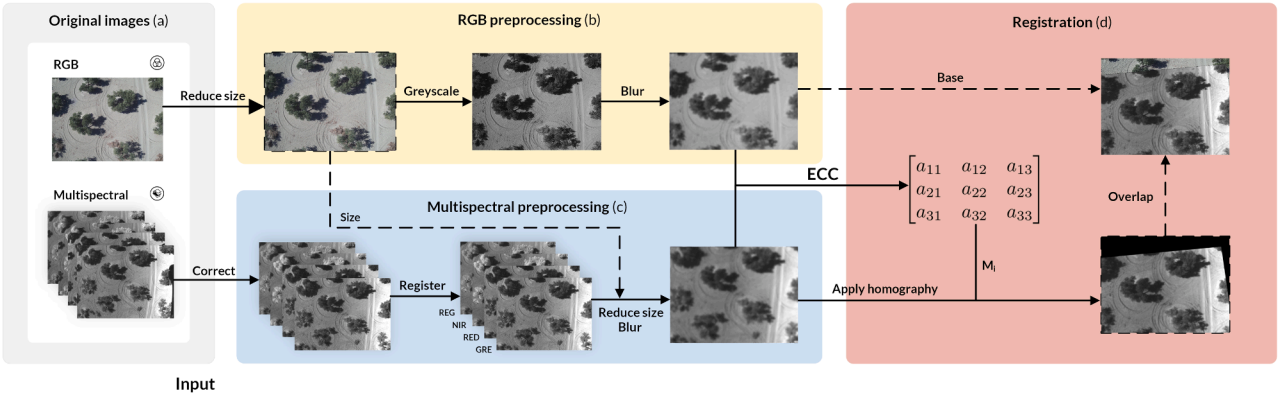

Fig. 8. Registration methodology for RGB and multispectral images. Blur size was exaggerated for visualisation purposes.

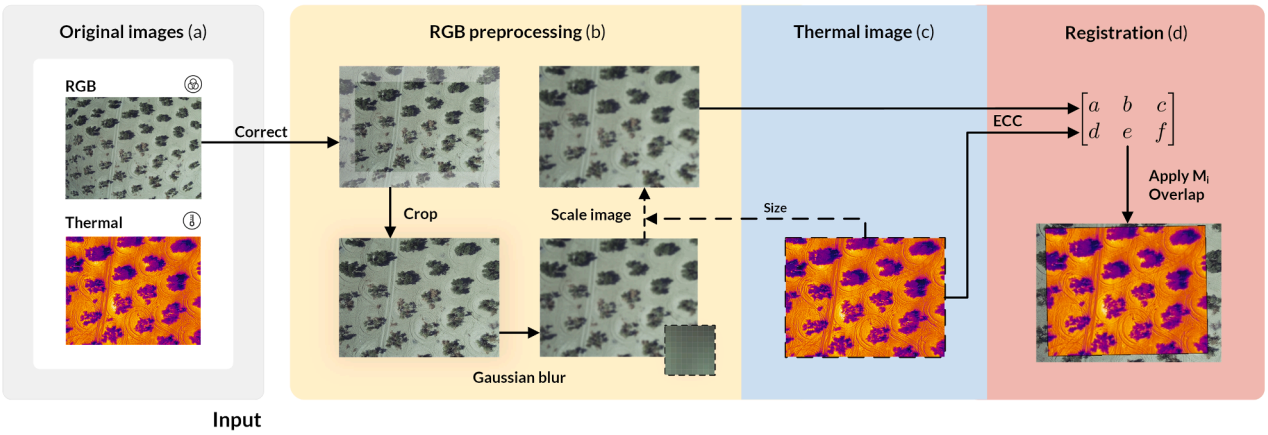

Fig. 10. Overview of the registration process for RGB and thermal images. Green channel of RGB images was emphasized to improve the understanding of this scheme.

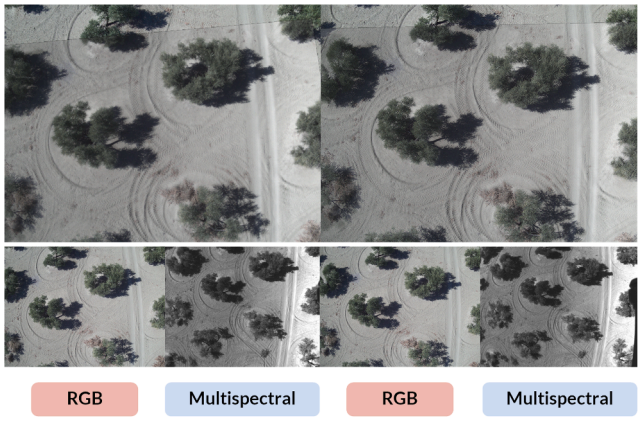

Fig. 11. Registration of multispectral and RGB images. The unregistered images are shown below the results.

not need to be registered with other image previously. Furthermore, no transformations are needed on that image but a fisheye correction.

RGB imagery typically have higher resolution than multispectral images, thus RGB images are resized to match the size of the second type. However, the ECC algorithm can be applied at a sub-pixel level and parameters can be selected to compute a fine-grain registration. Although this configuration is slower, a faster convergence is obtained by applying a previous gaussian blur filter in both images. The complete process is shown in Fig. 8, where the algorithm parameters and their values are given by a blur mask of size 3, a high number of iterations, $N = 400$, and a precision factor P of 1^{-60}. The values of these parameters are further discussed in a later chapter. Fig. 11 presents a result of this registration process.

2.5. Thermal and RGB registration

Registration of thermal and RGB imagery is the last step to complete the proposed multi-layer model, where RGB images are used as a link between multispectral and thermal imagery. This scenario avoids some of the previous problems, as both types of images are taken by the same device. Therefore, it is not necessary to find the image which is more close in time, since it is already known.

However, there exists a noticeable difference in focal length sizes as well as the consequent visual deformations. RGB lens captures a wider area of terrain, but it also involves a fisheye distortion. Although both images are taken from the same device, the aspect factor between focal lengths is not enough to calculate which part of the RGB image is visible

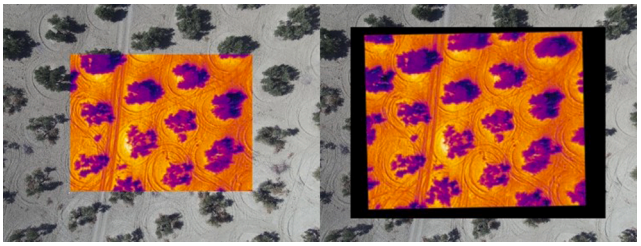

Fig. 12. Registration of RGB and thermal images: (a) using an RGB area bigger than thermal image (b) using an RGB area smaller than thermal image.

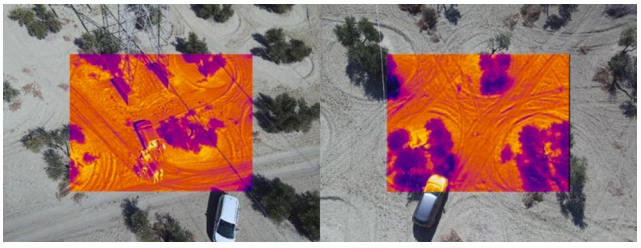

Fig. 13. Registration of images with objects not related to a crop.

in a thermal image, since other parameters such as sensor width also affect the result. There also exists a translation related to the physical distance between lenses, as well as a rotation derived from a small distance in milliseconds between captures, even though they are synchronized.

A thermal image is not located at the centre of an RGB image neither we know with certainty which part of the image do they both share. However, the affine motion model of the ECC algorithm includes translation, rotation and scale transformations. Translation and rotation are supposed to fix the differences between RGB and thermal imagery, while scale allows to search for an appropriate size where both images get an accurate overlapping.

At least two approaches can be used to register a pair of RGB-thermal images: (a) select a bigger area of RGB which fully covers the thermal image (the visual result includes null values) or (b) select a smaller area which does not necessarily cover the thermal image, improving the result in terms of graphic visualization (Fig. 12). In any case, the result of this process is a matrix M_i that allows to register both images. Therefore, these two approaches (shown in Fig. 12) obtain the same off-screen result. The complete process is shown in Fig. 10.

Regarding the dimensions of images, RGB is scaled down to thermal size, while precision and number of iterations are adjusted to provide a

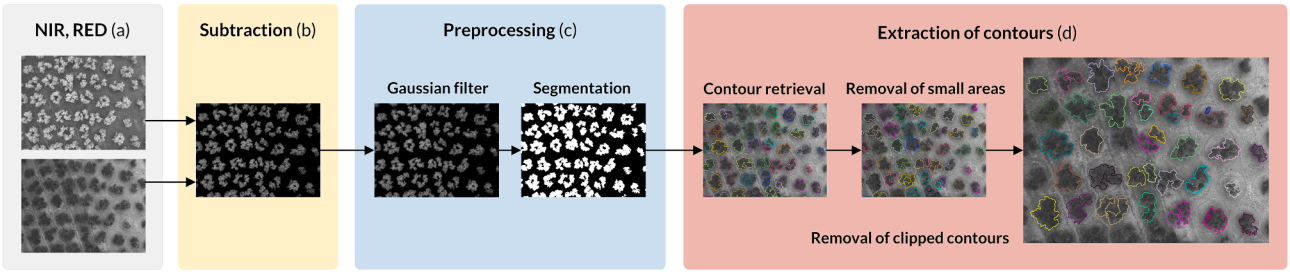

Fig. 14. Overview of the processing of multispectral images to distinguish vegetation of interest from crop soil.

good sub-pixel level registration. Consequently, execution time also decreases. Some results of this step are shown in Fig. 13. Further than a crop, the registration algorithm is tested against images which include objects not related to a crop, where it also returns good results.

2.6. Segmentation of individual trees

This framework allows to register multiple heterogeneous images, thus it can be used to extract RGB, multispectral and temperature information from a crop. It can also be tracked over time to retrieve the same information at several time frames. This section aims to provide an example of a tracking application where trees are identified through multispectral information. Points inside the extracted contours, which are known to belong to vegetation, can be further analysed to extract indices such as NDVI.

After registering previous multi-source datasets, a new method is proposed to identify individual trees. In this way, meaningful data for each crop can be obtained from all studied layers (multispectral, thermal and RGB). For this purpose, multispectral images are highly suitable for the recognition of individual trees, as the reflectance function for vegetation, $f(\theta)$, presents some peaks where the reflectance distance is maximised. This contrast allows to differentiate soil and canopy. Although we do not have access to the complete spectrum of $f(\theta)$, there exists (at least) relative maximum and minimum points which are visible at some of our four spectral bands.

In this case, NIR and RED include relative maximum and minimum reflectance values for vegetation, respectively. An assumption to be considered here is that soil reflectance keeps constant across multispectral images or difference is low enough to be filtered. Therefore, the operation NIR - RED returns the result shown in Fig. 14.

Beyond an operation of image subtraction, the following steps are also proposed:

1. Gaussian blur filter, to smooth the colour and remove noise.
2. Image thresholding, so we retrieve the structure of trees which are visible in the scene.
3. Extraction of the hierarchy of contours. Contours within another are not of interest, thus they must be discarded. Neither are of interest those small contours which are mostly related to low vegetation or any other object placed in a crop. Finally, the contours which are not completely visible in the image are also discarded. They can be identified using their position (close to image boundaries) and fully horizontal or vertical edges.

Therefore, the identification of individuals trees can be adjusted through the size of a blur mask (constant across images), the maximum area to be discarded (calculated with the image size) and a threshold value high enough to avoid merging tree contours (depends on reflectance function). Despite this method yields accurate contours, most of them contain hundreds of points. Nevertheless, this level of detail could not be necessary when storing this data in a database. Using this approach, contours can be simplified into convex hulls with an algorithm such as the Sklansky method (Sklansky, 1982) or into fixed-length

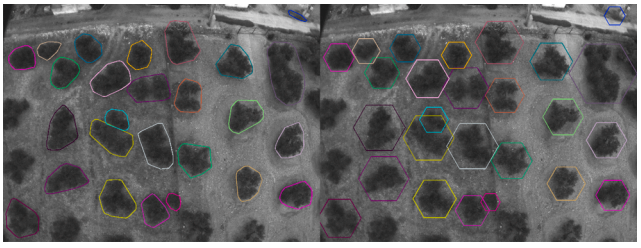

Fig. 15. Simplified contours: (a) Convex hulls (b) Fixed-length polygons.

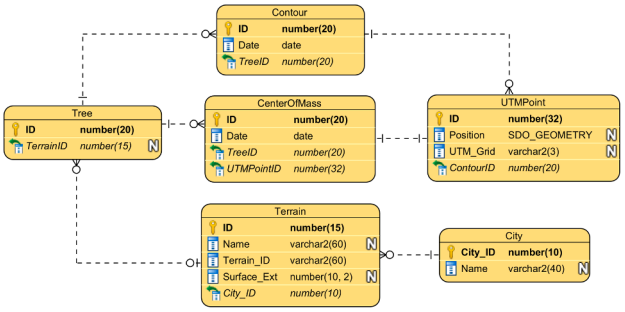

Fig. 16. Entity-Relationship model of a database for crown-tracking.

polygons (e.g. a hexagon) (Figure 15).

Finally, contours can be saved in a geospatial database (Fig. 16). Each tree is identified by a centre of mass computed from the retrieved polygon. Although this point may not be inside the polygon, it represents a unique identifier for a tree. Also, this point may be translated over time, therefore a distance threshold is needed, which in many geospatial databases is already implemented by the own system.

The mass centre points and their UTM coordinates are stored in a geospatial database. Here we propose a database model which allows us to track trees over time to analyze their growth speed and changes in their shape. Also, these trees are wrapped in a crop entity. Fig. 16 presents a basic approach that could be extended with additional information, such as the health state of a tree or the amount of water.

3. Validation

Our framework is evaluated with four datasets which were previously described. Several tests are considered by modifying number of iterations and precision factor of the ECC algorithm. The purpose is not only to demonstrate the accuracy of our algorithms but also to show different configurations which can be more appropriate to certain applications (real-time vs deferred analysis).

Measurements were performed on a PC with Intel Core i7.6700 3.4 GHz, 16 GB RAM, NVIDIA GTX 1070 with 8 GB RAM and Windows 10 OS. The implementation of the Enhanced Correlation Coefficient algorithm is provided by the image processing library OpenCV. The application has been developed in C++ within a graphical interface built

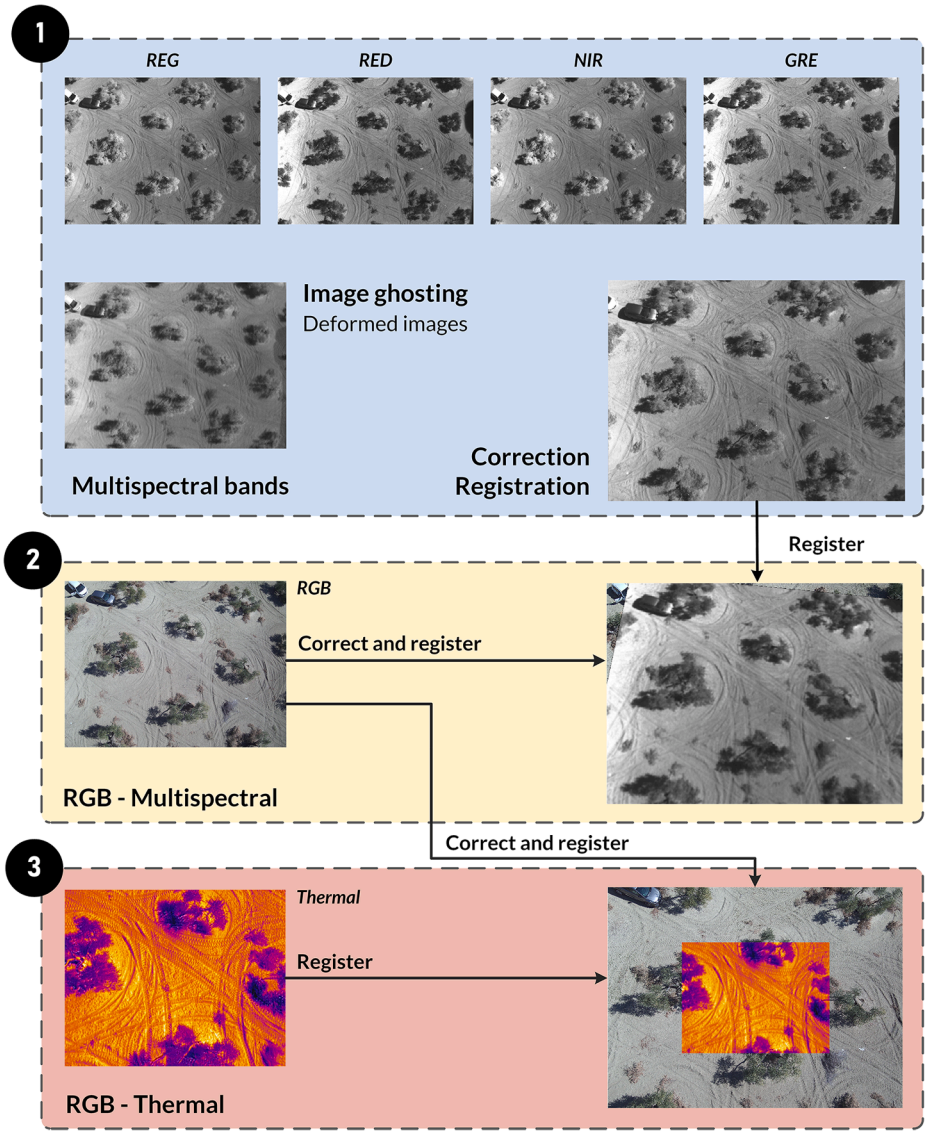

Fig. 17. Example of a complete registration process for multispectral, thermal and RGB imagery.

with the Qt framework.

The accuracy of the registration process is measured through a normalised correlation coefficient (CC) defined by Eq. 7. CC is a coefficient $\rho \in [0,1]$ that measures the linear relation between two images I and T, source and template. Although this value gives us a valid estimation, it is not an exact measurement of the alignment accuracy, since it considers the intensity difference between both images.

The images to be compared are selected randomly from our four datasets. The number of images from each one depends on the dataset size. Furthermore, not all of them contains all the types of images that our multi-layer model contains.

$$C = \frac{\sum_{x',y'} T(x',y') * I(x+x',y+y')}{\sqrt{\sum_{x',y'} T(x',y')^2 * \sum_{x',y'} I(x+x',y+y')^2}} \tag{7}$$

Up to three parameters are tested, as they all affect response time and registration accuracy:

- Precision to converge. Values near zero. The closer it gets to zero, the harder is for the algorithm to converge.
- Number of iterations. A high amount of iterations might not improve the result since the algorithm can converge earlier. Thus, it needs to be adjusted.
- Size of images. Smaller dimensions do not necessarily obtain worse results since ECC can work at a sub-pixel level. In that case, execution time also decreases as the linear search involves a smaller number of pixels.

4. Results and discussion

Multispectral registration. Four tests are executed to prove the accuracy of the registration methodology using the four multispectral bands. The correlation coefficient for unregistered images is presented in Table 3 to show the improvement of our four tests from the first scenario. In the case of multispectral images, their size is small (1280 × 960) in comparison with RGB images (4000 × 3000), therefore they do not need to be reduced for optimisation purposes. Test_1 uses parameters which can be near the optimal configuration, despite the execution time is relatively high. Table 6 shows that for images that present other objects the correlation coefficient drops dramatically due to differences in colour intensity (although by visual inspection images are well aligned). Test_2 is run to prove that the execution time can be reduced in exchange for a decrease in the correlation coefficient. Test_3 tries to get an intermediate configuration which reaches high values of CC in a reduced execution time. Test_4 proves what occurs when image size is reduced (by three in this case): the CC value increases while the execution time

decreases. However, by visual inspection $Test_4$ returns worse results.

RGB and multispectral registration. Table 7 shows the results of three tests in terms of CC and execution time. As the differences between images are now higher, the execution time increases dramatically. $Test_2$ reaches an optimal state with a reduced time, while $Test_1$ gets the same result with an increased build time. This scenario proves the convergence of the ECC algorithm; a safer approach would pick good enough values despite the execution time. However, real-time approaches can take a greater risk using values which reduce the execution time and still return good results (either they are optimal results or not). Both tests use images with a size equal to multispectral dimensions divided by two. This size allows to remove some details and improve the registration process (although a small blur filter is still needed). $Test_3$ uses the original multispectral size and RGB images are scaled to such size. It returns worse results than previous tests since the number of iterations and precision are reduced to avoid high execution times.

Thermal and RGB registration. Up to three tests are executed with the same parameters as before (Table 5). In contrast with previous tests, the size reduction does not have a great impact since thermal images do not present well-defined contours. CC values are also expected to be lower than in previous tests as the results include null values due to the selection of a bigger area of RGB image. Therefore, $Test_1$ and $Test_3$ show what occurs when both images are fitted into thermal image dimensions divided by two. Both reach the best correlation coefficient found, while $Test_3$ proves again the convergence of ECC as it yields the same results than $Test_1$ in a lower time. $Test_2$ uses the original size of thermal images and obtains almost the same CC average at the cost of a slightly higher build time. In contrast with multispectral registration case, $Test_1$ and $Test_3$ use smaller sizes but do not return visually incorrect alignments. In conclusion, using the original size is not needed in this case.

Regarding the whole framework, a graphic result is shown in Fig. 17, where multispectral, thermal and RGB images are registered to illustrate the multi-layer model. The RGB image is aligned with the GRE image, which was previously registered with its own multispectral bands. Also, the thermal image is aligned with the RGB image. GRE band is used as a link only for multispectral imagery, while RGB image acts as a link for both multispectral and thermal imagery.

5. Conclusions

We described a framework for registering heterogeneous UAV-based images into a multi-layer model. We have also exploited the ECC algorithm to show that it is a suitable algorithm to register images which highly differ on their intensity values. The process here described shows that heterogeneous images can be registered to integrate a homogeneous model with a few steps. Through a case study, we have also proposed a method to detect individual trees in multispectral images. This method could be further used to enable multi-temporal tracking of trees (growth speed, health state, etc).

From this work, we aim to create an orthomosaic map which includes multispectral, RGB and temperature data. For such task, we need to rely on accurate geolocation data, mostly provided by RGB images or even the UAV. Also, this multi-layer model can be extended with further multispectral information (hyperspectral), or even it could be translated to 3D space by registering all this data with a point cloud (e.g. generated by photogrammetry).

In the case of aerial images, the differences between images are typically smaller than in terrestrial images, even though we have needed affine motion models to merge images. Therefore, the computation process with terrestrial images is more time-consuming and requires more iterations. From this framework, we could develop an adaptive process which starts from a reduced size and advances to the original size. With a reduced size, we can obtain an approximate transformation in a lower execution time. This approximate transformation could be applied so that the distance between images in their original size is clearly reduced, and so is the complexity of the matching algorithm with greater dimensions.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work has been partially supported by the Spanish Ministry of Science and Innovation and the European Union (via ERDF funds) through the research project TIN2017-84968-R.

Appendix A

Tables 4–7.

Table 4
Normalised correlation coefficient retrieved from some multispectral captures in their default state.

	Base correlation		
Multispectral	ρ(GRE-RED)	ρ(GRE-NIR)	ρ(GRE-REG)
Dataset 1: $Image_1$	0.8954	0.8777	0.9306
Dataset 1: $Image_4$	0.9777	0.9381	0.9618
Dataset 1: $Image_6$	0.9756	0.9520	0.9727
Dataset 2: $Image_{48}$	0.9265	0.8918	0.9316
Dataset 2: $Image_{58}$	0.9146	0.8878	0.9242
Dataset 2: $Image_{53}$	0.9777	0.9381	0.9618
Dataset 2: $Image_{37}$	0.9286	0.9100	0.9371
Dataset 3: $Image_{35}$	0.9578	0.9453	0.9739
Dataset 3: $Image_{107}$	0.9602	0.9537	0.9686
Dataset 3: $Image_{136}$	0.9755	0.9612	0.9692
Dataset 3: $Image_{78}$	0.9707	0.9504	0.9672
Dataset 3: $Image_{142}$	0.9728	0.9375	0.9656
Dataset 3: $Image_{144}$	0.9731	0.9316	0.9622
Dataset 3: $Image_9$	0.9669	0.9545	0.9714
Dataset 3: $Image_{141}$	0.9732	0.9551	0.9744
Average	**0.9564**	**0.9323**	**0.9581**

Table 5
Normalised correlation coefficient and execution time retrieved from the registration of a random subset of RGB and thermal images. • Test 1. Image dimensions: $thermal_{size}/2$. Maximum iterations: 300. Precision: 1^{-60} • Test 2. Image dimensions: thermal. Maximum iterations: 100. Precision: 1^{-20} • Test 3. Image dimensions: $thermal_{size}/2$. Maximum iterations: 60. Precision: 1^{-20}.

	Test 1		Test 2		Test 3	
RGB-thermal pair	ρ	Total time (ms)	ρ	Total time (ms)	ρ	Total time (ms)
Dataset 3: $Image_{394}$	0.9756	2779	0.9771	3204	0.9756	1419
Dataset 3: $Image_{626}$	0.9728	2834	0.9745	3320	0.9728	1389
Dataset 3: $Image_{726}$	0.9728	2951	0.9741	3211	0.9728	1411
Dataset 3: $Image_{928}$	0.9756	2761	0.9770	3341	0.9756	1401
Dataset 3: $Image_{467}$	0.9416	2731	0.9385	3134	0.9417	1381
Dataset 3: $Image_{643}$	0.9511	2789	0.9494	3137	0.9511	1404
Dataset 3: $Image_{963}$	0.9444	2881	0.9425	3177	0.9444	1403
Dataset 3: $Image_{475}$	0.9396	2946	0.9358	3257	0.9396	1390
Dataset 4: $Image_{922}$	0.9674	2791	0.9690	3200	0.9674	1396
Dataset 4: $Image_{554}$	0.9786	2790	0.9803	3267	0.9786	1402
Dataset 4: $Image_{405}$	0.9266	2716	0.9265	3187	0.9266	1400
Dataset 4: $Image_{607}$	0.9511	2794	0.9502	3236	0.9511	1382
Average	**0.9581**	**2813.58**	**0.9579**	**3222.58**	**0.9581**	**1398.16**

Table 6
Normalised correlation coefficient and execution time retrieved from the registration of a random subset of multispectral captures which have been corrected and registered. • Test 1. Maximum iterations: 30. Precision: 1^{-10} • Test 2. Maximum iterations: 15. Precision: 1^{-2} • Test 3. Maximum iterations: 15. Precision: 1^{-10} • Test 4. Maximum iterations: 150. Precision: 1^{-60}.

	Test 1				Test 2			
Multispectral	ρ(GRE-RED)	ρ(GRE-NIR)	ρ(GRE-REG)	Total time (ms)	ρ(GRE-RED)	ρ(GRE-NIR)	ρ(GRE-REG)	Total time (ms)
D1: $Image_1$	0.9545	0.9186	0.9403	3627	0.9545	0.9181	0.9395	1128
D1: $Image_4$	0.9888	0.9473	0.9681	3576	0.9814	0.9473	0.9681	950
D1: $Image_6$	0.9861	0.9591	0.9772	3622	0.9817	0.9591	0.9767	1079
D2: $Image_{48}$	0.9654	0.9107	0.9368	3580	0.9654	0.9107	0.9368	1282
D2: $Image_{58}$	0.9546	0.9088	0.9321	3607	0.9547	0.9088	0.9321	1224
D2: $Image_{53}$	0.9888	0.9473	0.9681	3573	0.9814	0.9473	0.9681	955
D2: $Image_{37}$	0.9628	0.9265	0.9461	3569	0.9463	0.9265	0.9460	1100
D3: $Image_{35}$	0.9973	0.9654	0.9785	3543	0.9973	0.9645	0.9766	1090
D3: $Image_{107}$	0.9973	0.9621	0.9776	3558	0.9973	0.9621	0.9776	1356
D3: $Image_{136}$	0.9973	0.9677	0.9767	3577	0.9781	0.9623	0.9760	581
D3: $Image_{78}$	0.9957	0.9640	0.9731	3575	0.9956	0.9615	0.9696	1353
D3: $Image_{142}$	0.9962	0.9632	0.9749	3567	0.9772	0.9632	0.9748	989
D3: $Image_{144}$	0.9964	0.9580	0.9715	3586	0.9781	0.9580	0.9715	994
D3: $Image_9$	0.9967	0.9713	0.9810	3585	0.9967	0.9713	0.9810	1471
D3: $Image_{141}$	0.9966	0.9746	0.9791	3518	0.9967	0.9713	0.9810	1110
Average	**0.9849**	**0.9496**	**0.9654**	**3577.53**	**0.9788**	**0.9488**	**0.965**	**1110.8**

	Test 3				Test 4			
Multispectral	ρ(GRE-RED)	ρ(GRE-NIR)	ρ(GRE-REG)	Total time (ms)	ρ(GRE-RED)	ρ(GRE-NIR)	ρ(GRE-REG)	Total time (ms)
D1: $Image_1$	0.9545	0.9186	0.9403	1933	0.9540	0.9181	0.9394	1797
D1: $Image_4$	0.9866	0.9474	0.9682	1900	0.9883	0.9473	0.9665	1780
D1: $Image_6$	0.9838	0.9592	0.9772	1927	0.9849	0.9610	0.9767	1818
D2: $Image_{48}$	0.9654	0.9107	0.9367	1884	0.9651	0.9113	0.9368	1746
D2: $Image_{58}$	0.9546	0.9088	0.9321	1994	0.9651	0.9113	0.9368	1774
D2: $Image_{53}$	0.9866	0.9474	0.9682	1921	0.9883	0.9473	0.9665	1760
D2: $Image_{37}$	0.9628	0.9265	0.9461	1895	0.9618	0.9266	0.9455	1774
D3: $Image_{35}$	0.9973	0.9654	0.9785	1905	0.9972	0.9659	0.9778	1805
D3: $Image_{107}$	0.9973	0.9621	0.9776	1920	0.9970	0.9618	0.9762	1788
D3: $Image_{136}$	0.9966	0.9645	0.9766	1900	0.9969	0.9671	0.9752	1761
D3: $Image_{78}$	0.9957	0.9640	0.9731	1903	0.9948	0.9633	0.9714	1798
D3: $Image_{142}$	0.9941	0.9632	0.9749	1909	0.9961	0.9632	0.9740	1773
D3: $Image_{144}$	0.9909	0.9580	0.9715	1898	0.9963	0.9580	0.9702	1774
D3: $Image_9$	0.9967	0.9713	0.9810	1911	0.9964	0.9713	0.9806	1789
D3: $Image_{141}$	0.9966	0.9746	0.9791	1873	0.9965	0.9750	0.9786	1789
Average	**0.9839**	**0.9493**	**0.9654**	**1904.86**	**0.9852**	**0.9499**	**0.9648**	**1781.73**

Table 7
Normalised correlation coefficient and execution time retrieved from the registration of a random subset of multispectral and RGB images. • Test 1. Image dimensions: $multispectral_{size}/2$. Maximum iterations: 400. Precision: 1^{-60} • Test 2. Image dimensions: $multispectral_{size}/2$. Maximum iterations: 150. Precision: 1^{-30} • Test 3. Image dimensions: $multispectral_{size}$. Maximum iterations: 100. Precision: 1^{-20}.

		Test 1		Test 2		Test 3	
Multispectral	RGB	ρ	Total time (ms)	ρ	Total time (ms)	ρ	Total time (ms)
D4: $Image_1$	D4: $Image_{916}$	0.9683	8090	0.9683	2810	0.9569	7405
D4: $Image_{148}$	D4: $Image_{918}$	0.9664	7882	0.9664	2839	0.9408	7022
D4: $Image_{142}$	D4: $Image_{904}$	0.9711	7506	0.9711	2757	0.9453	7058
D4: $Image_{137}$	D4: $Image_{892}$	0.9761	7221	0.9761	2718	0.9747	7448
D4: $Image_{144}$	D4: $Image_{906}$	0.9690	7266	0.9690	2724	0.9680	7087
Average		**0.9701**	**7593**	**0.9701**	**2769.6**	**0.9571**	**7204**

Appendix B. Supplementary material

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.jag.2020.102274.

References

Barrero, O., Perdomo, S.A., 2018. Rgb and multispectral uav image fusion for gramineae weed detection in rice fields. Precision Agric. 19 (5), 809–822.

Bavirisetti, D.P., Dhuli, R., 2015. Fusion of infrared and visible sensor images based on anisotropic diffusion and karhunen-loeve transform. IEEE Sens. J. 16 (1), 203–209.

Evangelidis, G.D., Psarakis, E.Z., 2008. Parametric image alignment using enhanced correlation coefficient maximization. IEEE Trans. Pattern Anal. Mach. Intell. 30 (10), 1858–1865.

Garcia, H., Flores, H., Ascencio-Hernández, R., Khalil-Gardezi, A., Tijerina-Chávez, L., Mancilla, O., Peña, V.M., 2020. Corn grain yield estimation from vegetation indices, canopy cover, plant density, and a neural network using multispectral and rgb images acquired with unmanned aerial vehicles. Agriculture 10.

Hakim, P., Syafrudin, A., Utama, S., Jayani, A., 2018. Band coregistration modeling of lapan-a3/ipb multispectral imager based on satellite attitude. IOP Conf. Ser.: Earth Environ. Sci. 149, 1–7.

Jhan, J., 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 Inform. Sci. 42, 157.

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 J. Photogramm. Remote Sens. 114, 66–77.

Jurado, J.M., Ortega, L., Cubillas Mercado, J., Feito, F., 2020. Multispectral mapping on 3d models and multi-temporal monitoring for individual characterization of olive trees. Remote Sens. 12.

Khanal, S., Fulton, J., Shearer, S., 2017. An overview of current and potential applications of thermal remote sensing in precision agriculture. Comput. Electron. Agric. 139, 22–32.

Liu, T., Li, R., Zhong, X., Jiang, M., Jin, X., Zhou, P., Liu, S., Sun, C., Guo, W., 2018. Estimates of rice lodging using indices derived from uav visible and thermal infrared images. Agric. For. Meteorol. 252, 144–154.

Lu, B., Dao, P.D., Liu, J., He, Y., Shang, J., 2020. Recent advances of hyperspectral imaging technology and applications in agriculture. Remote Sens. 12 (16), 2659.

Maes, W.H., Steppe, K., 2019. Perspectives for remote sensing with unmanned aerial vehicles in precision agriculture. Trends Plant Sci. 24 (2), 152–164.

Matese, A., Di Gennaro, S.F., 2018. Practical applications of a multisensor uav platform based on multispectral, thermal and rgb high resolution images in precision viticulture. Agriculture 8 (7), 116.

Osroosh, Y., Khot, L.R., Peters, R.T., 2018. Economical thermal-rgb imaging system for monitoring agricultural crops. Comput. Electron. Agric. 147, 34–43.

Pablo, J., Zarco-Tejada, N.H., Loudjani, P., 2014. Precision agriculture: an opportunity for EU farmers: potential support with the CAP 2014-2020, p. 56.

Park, J., Park, K., Baeg, S.-H., Baeg, M.-H., 2008. pi-sift: A photometric and scale invariant feature transform. In: ICPR 2008 19th International Conference on Pattern Recognition. IEEE Computer Society.

Santini, F., Kefauver, S.C., Resco de Dios, V., Araus, J.L., Voltas, J., 2019. Using unmanned aerial vehicle-based multispectral, rgb and thermal imagery for phenotyping of forest genetic trials: A case study in pinus halepensis. Ann. Appl. Biol. 174 (2), 262–276.

Shen, X., Xu, L., Zhang, Q., Jia, J., 2014. Multi-modal and Multi-spectral Registration for Natural Images. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (Eds.), Computer Vision – ECCV 2014, Lecture Notes in Computer Science. Springer International Publishing, Cham, pp. 309–324.

Sklansky, J., 1982. Finding the convex hull of a simple polygon. Pattern Recogn. Lett. 1 (2), 79–83.

Tang, Q., Zhang, R., Chen, L., Xu, G., Deng, W., Ding, C., Xu, M., Yi, T., Wen, Y., Li, L., 2020. High-accuracy, high-resolution downwash flow field measurements of an unmanned helicopter for precision agriculture. Comput. Electron. Agric. 173, 105390.

Tucci, G., Parisi, E.I., Castelli, G., Errico, A., Corongiu, M., Sona, G., Viviani, E., Bresci, E., Preti, F., 2019. Multi-sensor uav application for thermal analysis on a dry-stone terraced vineyard in rural tuscany landscape. ISPRS Int. J. Geo-Inform. 8 (2), 87.

Zhang, C., Kovacs, J.M., 2012. The application of small unmanned aerial systems for precision agriculture: a review. Precision Agric. 13 (6), 693–712.

Zheng, H., Cheng, T., Li, D., Zhou, X., Yao, X., Tian, Y., Cao, W., Zhu, Y., 2018. Evaluation of rgb, color-infrared and multispectral images acquired from unmanned aerial systems for the estimation of nitrogen accumulation in rice. Remote Sens. 10 (6), 824.