Registration of visible and near infrared unmanned aerial vehicle images based on Fourier-Mellin transform

Gilles Rabatel, S. Labbe

Author-produced version of the article published in Precision Agriculture, 2016, N°17, p.564-587.
The original publication is available at http://link.springer.com/article/10.1007%2Fs11119-016-9437-x
Doi: 10.1007/s11119-016-9437-x

Registration of Visible and Near Infrared Unmanned Aerial Vehicle Images Based on Fourier-Mellin Transform

Gilles Rabatel [1,]*, Sylvain Labbé [2]

[1] Irstea, UMR ITAP, 361 rue J-F Breton, BP 5095, 34196 Montpellier Cedex 5, France;

[2] Irstea, UMR TETIS, 500 rue Jean-François Breton, 34093 Montpellier, France

* Corresponding author: gilles.rabatel@irstea.fr; Tel.: +33-4-67-04-63-58

Abstract: The combination of aerial images acquired in the visible and near infrared spectral ranges is particularly relevant for agricultural and environmental survey. In unmanned aerial vehicle (UAV) imagery, such a combination can be achieved using a set of several embedded cameras mounted close to each other, followed by an image registration step. However, due to the different nature of source images, usual registration techniques based on feature point matching are limited when dealing with blended vegetation and bare soil patterns. Here, another approach is proposed based on image spatial frequency analysis. This approach, which relies on the Fourier-Mellin transform, has been adapted to homographic registration and distortion issues. It has been successfully tested on various aerial image sets, and has proved to be particularly robust and accurate, providing a registration error below 0.3 pixels in most cases.

Keywords: multimodal image registration; Fourier-Mellin transform; homographic transformation; near infrared; NDVI

Introduction

In the context of aerial imagery for agricultural and environmental monitoring purposes, the acquisition of radiance data in the near-infrared domain is particularly important. This is due to the spectral response of green vegetation material, which is characterized by a very sharp transition between low reflectance in the visible domain (typically 10 to 20%), and high reflectance in the near-infrared domain (typically 50 to 80%). Indeed, since the early times of remote sensing (RS), observation satellites (such as Landsat and SPOT) have been equipped with multi-spectral sensors including a near-infrared band (typically around 800 nm) associated with visible ones.

One of the most popular usages of this multi-spectral combination is the computation of the normalized difference vegetation index or NDVI, introduced by Rouse et al (1973). NDVI is defined as the ratio (NIR-R)/(NIR+R), where NIR and R are respectively the near-infrared and red bands. As far as the radiance measurements are corrected for lighting and atmospheric effects (reflectance correction), the NDVI can be considered as a quantitative index for crop assessment (Myneni et al, 1995). In other cases, it remains a robust tool for the discrimination of vegetation areas in various lighting conditions (Torres-Sanchez et al, 2013; Rabatel et al, 2014).

Other vegetation indices have also been proposed in various soil and vegetation conditions, with respect to their relationship to agronomic indices such as biomass and LAI (leaf area index) (Huete et al, 1997; Jindong et al, 2007;

Author-produced version of the article published in Precision Agriculture, 2016, N°17, p.564-587.
The original publication is available at http://link.springer.com/article/10.1007%2Fs11119-016-9437-x
Doi: 10.1007/s11119-016-9437-x

Zhengwei et al, 2009). However, it is important to notice that most of these indices (such as ratio normalized difference vegetation index (RNDVI), soil-adjusted vegetation index (SAVI) and transformed normalized difference vegetation index (TNDVI)) rely on the same red and near-infrared bands.

In satellite and airborne technologies, image acquisition is usually based on push-broom sensors, combined with the regular motion of the vehicle to build 2D images. Separate spectral bands can thus be collected using parallel line sensors (Petrie, 2005). It is quite different with UAV, for which the cost of an accurate inertial unit to control the motion regularity is rarely affordable. Therefore UAV imagery generally relies on 2D image sensors. While Color Filter Arrays (CFA) are nearly universally employed for standard color cameras, CFA including a near-infrared band are presently not available. As an alternative, some camera manufacturers have proposed multi-CCD devices including a near-infrared channel, e.g. the MS-4100 (Geospatial Systems Inc., West Henrietta, NY, USA), or the AD-080 (JAI AS, Copenhagen, Denmark). However, such devices are characterized by high cost and limited image resolution.

On the other hand, UAV applications to agricultural monitoring are currently increasing dramatically, thanks to their specific advantages compared to aerial or satellite approaches: ease of use, flexibility, low cost, and very short revisit time. Also, the low flight elevation of UAV provides now access to imagery with centimetric spatial resolution, opening the door to new kinds of applications (e.g. plant counting or adventice detection). However, such applications require low cost acquisition devices to fit with the overall system cost, with high image resolution (typically more than 10 million pixels) in order to compensate for their limited flight duration (low energy autonomy). As a consequence, despite increasing demand, there is still no commercial multi-spectral solution that entirely fulfills the UAV requirements, i.e. combining low cost, low footprint, low weight and high image resolution. Operators usually implement general use still color cameras, either accepting limitation to standard color acquisition, or trying to adapt these cameras to their particular needs.

Hunt et al, (2010) proposed the use of a standard color camera where the internal near-infrared-blocking has been removed and replaced by an external red-light-blocking filter, in order to get the G, B and NIR bands instead of R,G,B. Similarly, Rabatel et al (2014) obtained R and NIR bands using an external red long-pass filter and a linear combination of the resulting raw channels. However, in both cases, the original color image of the scene is no longer available, and the spectral quality of the bands obtained is limited.

An alternative solution consists in coupling two still cameras, one providing a standard color image, the other being devoted to the near-infrared band. For this purpose, the internal near-infrared-blocking filter of the second camera is replaced with an external near-infrared band-pass filter (Lebourgeois et al, 2008). Despite the payload increase due to the implementation of two cameras, this approach remains particularly attractive, as being able to provide a full four-channel multi-spectral image including R,G, B and NIR bands. This kind of solution is also currently proposed by some camera manufacturers or UAV service companies involved in agricultural survey applications. They propose multi-spectral cameras based on 4 to 12 image sensors mounted close to each other, at the price of lower image resolution (Micro-MCA, Tetracam Inc. Chatworth, CA, US; Agrosensor, Airinov, Paris, France). However, multi-camera approaches require a post-processing step to accurately establish the relationship between the pixel co-ordinates of every camera. As detailed in the following, this processing step, known as image registration, remains a challenge in the case of near-infrared and visible image pairs (Dare, 2008; Hunt et al, 2010).

Indeed, image registration has been widely studied in the literature. Basically, it consists in finding local correspondences between two images, in order to identify an image transformation model. According to Zitova and Flusser (2003), there are two main approaches to find these local correspondences: i) feature-based methods, which search for salient characteristics in both images (such as edges, lines and corners) and then try to match them in a separate matching

Author-produced version of the article published in Precision Agriculture, 2016, N°17, p.564-587.
The original publication is available at http://link.springer.com/article/10.1007%2Fs11119-016-9437-x
Doi: 10.1007/s11119-016-9437-x

step ii) area-based methods, which directly attempt to estimate the correspondence between predefined windows (or the entire images) using integrative characteristics (such as correlation, Fourier spectrum and mutual information).

The choice between these two approaches mainly depends on the nature of the images. For *unimodal registration* (i.e. when images are obtained from the same sensor with different viewpoints), very efficient algorithms like the scale invariant feature transform (SIFT) (Lowe, 1999), or the more recent speeded up robust features (SURF) (Bay et al, 2008), have now definitely established the pre-eminence of the feature-based methods: these algorithms are known for their robustness with respect to scale and luminosity changes between images, and are widely used in image mosaicking, photogrammetry, stitching, or mobile robotics.

On the other hand, *multimodal registration* deals with the registration of images captured from sensors with different characteristics. A typical area is medical imagery, where images are captured from sensors using very different physical principles (magnetic resonance, ultrasonic echoes, positron emission tomography (PET)) can be combined (Oliveira and Tavares, 2014). Moreover, multi-temporal image acquisition and non-rigid transformations often add supplementary issues. Therefore, multimodal image registration remains an open field for research. While some authors have attempted to adapt feature point methods (Sargent et al, 2009; Zhi et al, 2009), area-based methods associated with various similarity measures have been widely investigated. Mutual information (MI), a measure derived from joint entropy computation, is popularly used in medical imagery (Pluim et al, 2003). It can be combined with variational approaches, well adapted to non-rigid transformations (Hermosillo et al, 2002).

Registration of visible and NIR images is a kind of multimodal registration. By the way, in the case of vegetation scenes, the feature point matching is no longer reliable, due to contrast inversion in the pixel digital counts (see next section for more detail). Figure 1 shows an illustration of this concern, when using a standard SIFT matching tool (VLFeat, available at www.vlfeat.org/applications/sift-mosaic-code.html; last accessed: 29 April 2015) on a typical field scene. Clearly, matching links are limited to bare soil areas, indicating a problem when vegetation is mixed with other material.

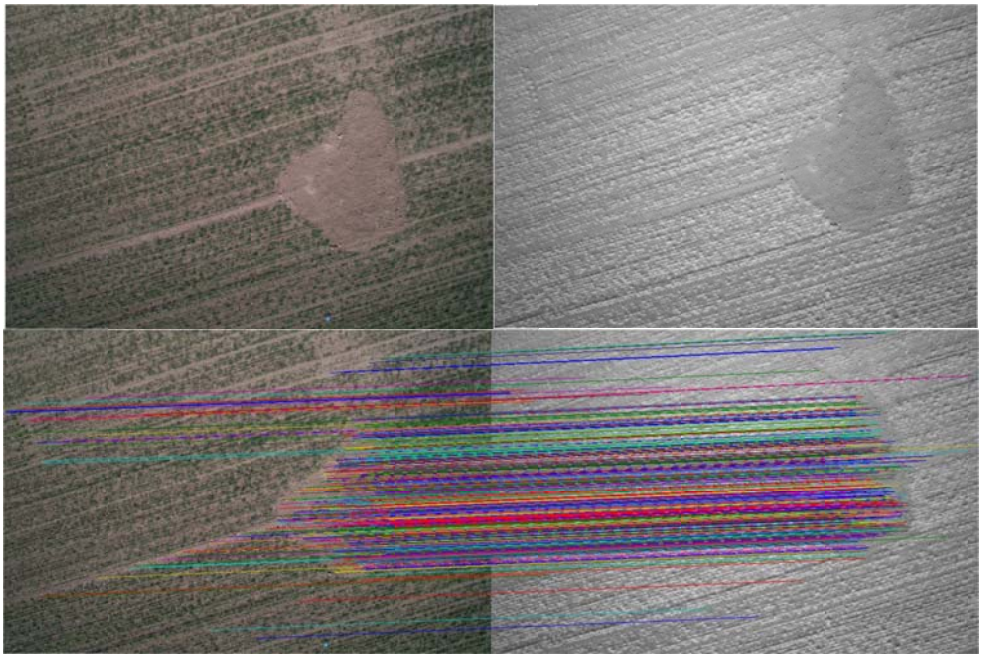


Figure 1. Example of SIFT point matching on a wheat crop scene.

Author-produced version of the article published in Precision Agriculture, 2016, N°17, p.564-587.
The original publication is available at http://link.springer.com/article/10.1007%2Fs11119-016-9437-x
Doi: 10.1007/s11119-016-9437-x

Top: original visible and NIR images. Bottom: matching result using VLFeat SIFT tool (www.vlfeat.org/applications/sift-mosaic-code.html)

Some authors have proposed a SIFT descriptor variant where the sign of image gradients is not taken into account (Firmenich et al, 2011). However the corresponding improvement did not appear to be sufficient for high resolution aerial images.

A more robust solution could be to investigate some of the numerous multimodal registration algorithms mentioned above. However, the complexity of such algorithms, mainly driven by medical imagery, is not well adapted to the specific context of UAV imagery, for which the availability of user friendly and fully automated software tools is a priority. Indeed, such a requirement for user-friendly tools is currently met in the field of photogrammetry (mosaicking and 3D reconstruction based on unimodal registration): numerous commercial or open source tools based on SIFT point matching (Photoscan[1], ERDAS Imagine[2], MicMac[3], VisualSFM[4]) are able to process automatically hundreds of UAV images in a couple of hours. Therefore, most UAV users dealing with multi-spectral image registration still try to use the same photogrammetric tools for multimodal registration as well either by:

i) registering directly the different bands using the embedded feature point registration algorithm with more or less success and accuracy depending on the image content

or by:

ii) computing a geo-referenced and ortho-rectified mosaic for both sets of images and then overlaying them based on their geographic co-ordinates, leading to a very limited registration quality.

The objective of the present paper is to propose an intermediate approach for the registration of visible and near-infrared images in the context of UAV imagery, able to provide robust and high quality registration of vegetation scenes while remaining fast and fully automated. The proposed approach involves spatial frequency analysis through the Fourier-Mellin transform (FMT). FMT is a well-known tool that has been widely used for image registration before the emergence of the SIFT algorithm (Reddy and Chatterji. 1996; Marcel et al, 1997). Because Fourier analysis integrates spatially characteristic features over the whole image, it is robust to contrast inversion (see figure 3), and efficiently registers images of very different kinds. Recently, Jaud et al (2013), for instance, have successfully used the FMT to match radar maps with aerial images in the visible domain.

As a counterpart, the main drawback of the FMT approach is that it is only applicable to register images linked through a transformation limited to translation, rotation and scale change. However, as shown further, it remains usable in aerial imaging, as long as the distance between the cameras is negligible compared to the scene distance, and the camera optical axis is nearly vertical. For this purpose, the initial method has been adapted to large size images and homographic transformations, using an iterative implementation based on image partitioning. Lens distortion aspects have been taken into account as well.

[1] Photoscan, Agisoft, St. Petersburg, Russia (www.agisoft.com)
[2] ERDAS Imagine, GEOSYSTEMS France SARL, Montigny-le-Bretonneux, France. (www.geosystems.fr)
[3] MicMac, IGN, France (http://logiciels.ign.fr/?-Micmac,3-)
[4] VisualSFM (http://ccwu.me/vsfm/)

Author-produced version of the article published in Precision Agriculture, 2016, N°17, p.564-587.
The original publication is available at http://link.springer.com/article/10.1007%2Fs11119-016-9437-x
Doi: 10.1007/s11119-016-9437-x

Theory of operation

Rotation-translation-scale identification

This sub-section addresses the problem of registering images that are linked by a similarity transformation, i.e. a transformation that exclusively includes translation, rotation and homothetic scaling. Extension to homographic transformation will be considered further.

Translation identification by Fourier Transform

Consider two grey-level images $I_1(x,y)$ and $I_2(x,y)$ where I_2 is an image obtained from I_1 through a translation (T_x, T_y). Formally, it can be written:

$$I_2(x,y) = I_1(x,y) * \delta(T_x - x, T_y - y) \quad (1)$$

where * is the convolution operator and $\delta(T_x\text{-}x, T_y\text{-}y)$ is the Dirac distribution centered on (T_x, T_y).

The 2D Fourier Transforms (FT) of the two images are thus linked by the relation:

$$F_2(u,v) = F_1(u,v).e^{-2i\pi(T_x.u+T_y.v)} \quad (2)$$

$F_1(u,v)$ and $F_2(u,v)$ only differ through a complex multiplicative term which is the FT of the Dirac distribution $\delta(T_x\text{-}x, T_y\text{-}y)$. Thus, $\delta(T_x\text{-}x, T_y\text{-}y)$ can be directly recovered through the inverse Fourier transform of the ratio $F_2(u,v)/F_1(u,v)$, i.e. by computing the image:

$$ImT(x,y) = \mathrm{FT}^{-1}\left(\frac{F_2(u,v)}{F_1(u,v)}\right) = \mathrm{FT}^{-1}\left(e^{-2i\pi(T_x.u+T_y.v)}\right) = \delta(T_x - x, T_y - y) \quad (3)$$

As illustrated in Figure 2, the computed image $ImT(x,y)$ appears as a black image with a unique amplitude peak at (T_x, T_y). The position of this peak can be retrieved by searching for the maximum amplitude in the image, and completely identifies the translation between I_1 and I_2.

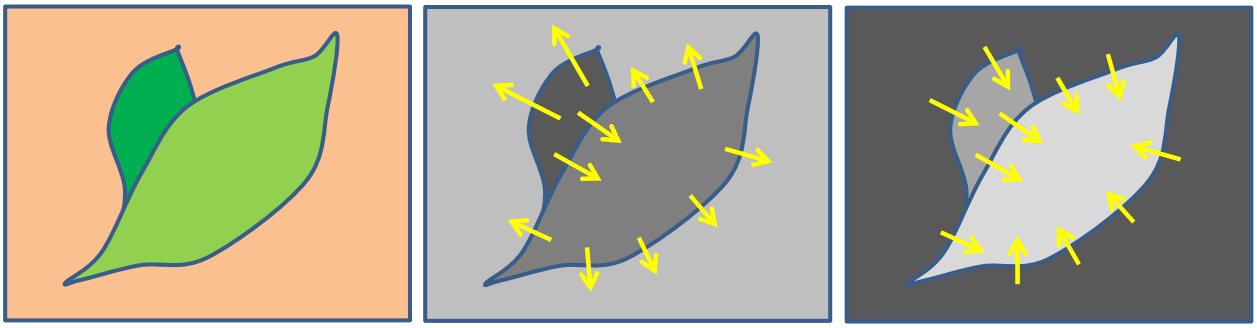
Figure 2. Example of translation identification.
From left to right: images I_1 & I_2 and translation peak (2D & 3D representation)

Author-produced version of the article published in Precision Agriculture, 2016, N°17, p.564-587.
The original publication is available at http://link.springer.com/article/10.1007%2Fs11119-016-9437-x
Doi: 10.1007/s11119-016-9437-x

Notice that Equation (2) also indicates that $F_1(u,v)$ and $F_2(u,v)$ have the same modulus, as the modulus of the multiplicative term $e^{-2\,i\,\pi\,(Tx.u\,+\,Ty.v)}$ is equal to one. This property will be used further.

Fourier Transform versus Feature Point in vegetation scenes

As stated above, feature point approaches like SIFT are largely used for registering images of the same nature. Now, let consider visible and NIR images of a vegetation scene. As already mentioned, in the NIR band, vegetation reflects almost all the incident light. Thus, vegetation appears much brighter than bare soil in the NIR image, while it will often appear darker in the visible one. A direct consequence, illustrated in Figure 3, is that the direction of grey level gradients will be inverted from one image to the other at the boundaries between vegetation and soil, while remaining the same within one kind of object.



Visible NIR

Figure 3. Contrast inversion in a vegetation scene.

SIFT descriptors are based on gradient direction histograms collected over a large neighborhood around the feature point (16x16 pixels). If the vegetation and soil are so blended that most feature point neighborhoods include both of them, gradient inversions will occur within these neighborhoods, and the descriptors will not match between visible and NIR images.

On the other side, the FT identification approach proceeds by enhancing the spatial frequency shifts that are coherent all over the image area (i.e. that correspond to the same translation). In that sense, it accumulates area-extended information. Within a given kind of object (vegetation or soil), the Fourier transform $F_2(u,v)$ in the second image will be affected by a multiplicative factor compared to $F_1(u,v)$, due to a different brightness, but the phase shift information will be preserved. Finally, both kinds of objects will contribute to the translation identification peak, proportionally to their area. Only edges between them will introduce some perturbation in the Fourier spectra, but this edge contribution will remain limited compared to the image area (1D versus 2D accumulation), and should not jeopardize the translation identification.

Rotation and scale identification by Fourier-Mellin Transform

The Fourier-Mellin Transform (FMT) was first introduced by Casasent and Psaltis (1977) in the context of optical image processing. The idea was to combine the Fourier Transform and the Mellin Transform (MT, which can be assimilated to a Fourier transform after logarithmic co-ordinate scaling), in order to get both rotation and scaling invariance. FT and MT are known to have their modulus invariant respectively to translation and scale. By considering a polar representation $f(r,\theta)$ of an image and by applying respectively FT on the θ axis and MT on the r axis, the FMT is obtained, which is formally defined as:

Author-produced version of the article published in Precision Agriculture, 2016, N°17, p.564-587.
The original publication is available at http://link.springer.com/article/10.1007%2Fs11119-016-9437-x
Doi: 10.1007/s11119-016-9437-x

$$\mathrm{Mf}(k,iv) = \frac{1}{2\pi} \int_0^{\infty} \int_0^{2\pi} f(r,\theta).e^{-ik\theta}.r^{-iv} \frac{dr}{r} d\theta \quad (4)$$

The FMT is an invertible transform. Its module is invariant to θ shift and to r scaling, i.e:

$$\forall (a,b) \in \Re^2 \quad \left|\mathrm{FMT}(f(a.r, \theta + b)\right| = \left|\mathrm{FMT}(f(r,\theta)\right| = \left|\mathrm{Mf}(k,iv)\right| \quad (5)$$

FMT was initially designed to compare images that differ by scaling and rotation, looking at their FMT module, but not to identify this difference. In a registration context, it is important to notice that FMT can be assimilated to the TF of a log-polar representation of the image, as illustrated in Figure 4.

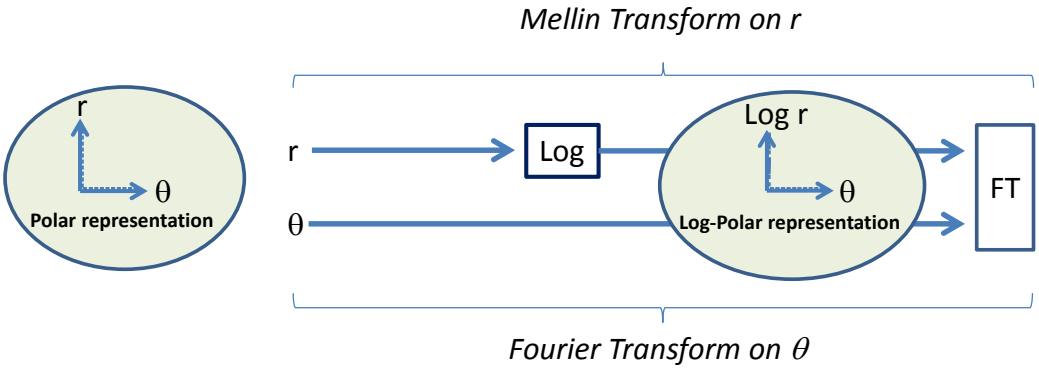


Figure 4. The Fourier-Mellin Transform scheme

Now, consider two images $I_1(r, \theta)$ and $I_2(r, \theta)$ in polar representation, which only differ by a centered rotation $\Delta\theta$ and a scaling factor k. When considering the log-polar representation, the rotation corresponds to an image shift equal to $\Delta\theta$ along the θ axis, and the scaling factor k corresponds to an image shift equal to Log(k) along the Log r axis. So the two log-polar images will only differ by a 2D translation ($\Delta\theta$, Log(k)). In that sense, the same approach that was described for translation identification can be applied: it consists in computing the ratio of the FT of the two log-polar representations, i.e. the ratio of the FMT of images $I_1(r, \theta)$ and $I_2(r, \theta)$, and then in searching for the maximum of its inverse Fourier Transform.

Notice that this approach implicitly supposes that the image scaling is isotropic, i.e. that the scaling factor k is independent of the angle θ (in other words, the same scaling factor is applied on x and y axes in the original images).

General scheme

Comparing the polar representations of two images through FMT supposes that these polar representations refer to the same image center, which is not the case when a translation is involved. Fortunately, original images can be replaced by the modulus of their Fourier transform, centered on frequency (0,0). As seen above, the FT modulus is invariant to translation. On the other hand, it preserves rotation and scaling information: a rotation θ and a scaling factor k applied to a given image generate a rotation θ and a scaling factor $1/k$ of its FT spectrum.

Author-produced version of the article published in Precision Agriculture, 2016, N°17, p.564-587.
The original publication is available at http://link.springer.com/article/10.1007%2Fs11119-016-9437-x
Doi: 10.1007/s11119-016-9437-x

According to this modification, already proposed by Casasent and Psaltis (1977), the general computing scheme illustrated in Figure 5 is obtained, allowing recovering completely the similarity transform between two images: rotation and scaling factor are recovered first using the FMT of the FT modulus, and are used to reverse rotate and scale the second image so that only the translation is remaining. This translation is then recovered.

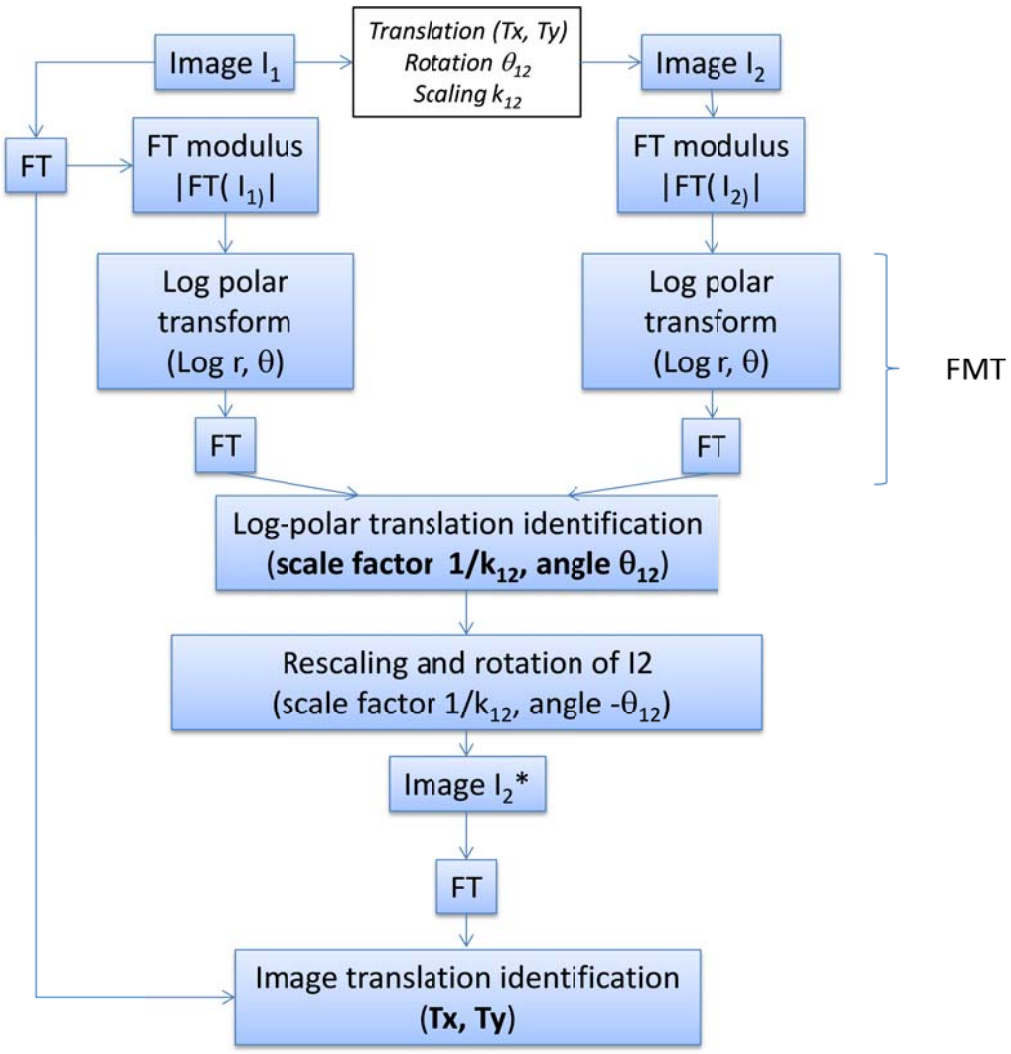


Figure 5. General scheme for rotation, scaling and translation identification

Homographic registration of large size images

Homographic transformation

Homographic transformation is of primary importance in computer vision calibration issues, because it is directly related to the perspective projection associated with image acquisition. Consider two cameras observing the same planar scene, and providing respectively images $I_1(x,y)$ and $I_2(x,y)$ of this scene (Figure 6). It can be shown that the co-ordinates (x_1,y_1) and (x_2,y_2) of the same point in both images are related by the homographic relation:

Author-produced version of the article published in Precision Agriculture, 2016, N°17, p.564-587.
The original publication is available at http://link.springer.com/article/10.1007%2Fs11119-016-9437-x
Doi: 10.1007/s11119-016-9437-x

$$\begin{bmatrix} u.x_2 \\ u.y_2 \\ u \end{bmatrix} = \mathbf{H}.\begin{bmatrix} .x_1 \\ y_1 \\ 1 \end{bmatrix} \quad (6)$$

where **H** is a 3x3 matrix, and *u* is a non-null scalar.

Taking the left member of equation (6), (x_2,y_2) co-ordinates can be recovered by dividing the two first vector components by the 3rd one. The actual value of *u* has no importance, which means that the matrix **H** can be defined up to a multiplicative factor.

Figure 6. Example of homographic transformation

Homographic transformation versus similarity transformation

Because the homographic matrix can be defined up to a scale factor, consider the following equations, where **H** is constrained to have its term H_{33} equal to 1:

$$\mathbf{H} = \begin{bmatrix} RS_{11} & RS_{12} & T_x \\ RS_{21} & RS_{22} & T_y \\ \varepsilon_x & \varepsilon_y & 1 \end{bmatrix} \quad (7)$$

Now, consider the matrix **S** obtained from **H** by setting ε_x and ε_y equal to zero:

$$\mathbf{S} = \begin{bmatrix} RS_{11} & RS_{12} & T_x \\ RS_{21} & RS_{22} & T_y \\ 0 & 0 & 1 \end{bmatrix} \quad (8)$$

In that case, the *u* term of equation (6) is always equal to one, so that **S** corresponds to a simple similarity, transforming $(x_1, y1)$ co-ordinates in $(x_2, y2)_S$ co-ordinates defined as:

$$\begin{bmatrix} x_2 \\ y_2 \end{bmatrix}_S = \mathbf{RS}.\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} + \begin{bmatrix} T_x \\ T_y \end{bmatrix} \quad (9)$$

where **RS**= $[RS_{ij}]$ is a rotation and scaling 2D sub-matrix, and (T_x, T_y) is a translation vector.

Author-produced version of the article published in Precision Agriculture, 2016, N°17, p.564-587.
The original publication is available at http://link.springer.com/article/10.1007%2Fs11119-016-9437-x
Doi: 10.1007/s11119-016-9437-x

Thus, applying the homographic transformation **H** to (x_1,y_1):

$$(x_2,y_2) = (x_2,y_2)_S \ /u = (x_2,y_2)_S \ /(1 + \varepsilon_x.x_1 + \varepsilon_y.y_1) \quad (10)$$

As a conclusion, a homographic transformation can be considered as an extension of a similarity transformation where a supplementary scale factor $1/u = 1/(1 + \varepsilon_x.x_1 + \varepsilon_y.y_1)$ *is applied, which depends on the position* (x_1, y_1) *in the initial image.*

Application to homographic transformation identification

At this stage, some hypotheses must be introduced concerning the homographic transformation to identify:

- The values of terms ε_x and ε_y are small compared to 1, which means that the homographic transformation is close to a similarity transformation.
- The similarity scaling factor is nearly isotropic (same homothetic factors on *x* and *y* axes).
- The image translation is small compared to the image size, in terms of pixels.

Fortunately, all these hypotheses are compatible with usual aerial imaging. More precisely, the first two hypotheses are both related to the angle between the camera optical axis and the vertical. The more the camera orientation will be close to the nadir, the more they will be matched. Of course, it also requires that the viewed scene is nearly planar (compared to the camera-scene distance), otherwise homographic transformations are not applicable at all. Once again, it will be usually verified except in the case of very low flight altitude or very rough or mountainous terrain. The last hypothesis about translation amplitude will be usually matched in a multi-spectral acquisition context where coupled cameras are mounted close to each other.

The first hypothesis $\varepsilon_x << 1$ and $\varepsilon_y << 1$ implies that the homographic factor $1/(1 + \varepsilon_x.x_1 + \varepsilon_y.y_1)$ will be slowly varying across the images, so that simple similarity transformations can be considered over large image sub-portions. The second isotropic hypothesis implies that in such sub-portions, the rotation-translation-scale identification scheme described earlier can be applied. Finally, the hypothesis on limited translation amplitude means that matching sub-portions between the two images can be easily defined.

As a consequence, the following scheme can be implemented for complete large size image registration (Figure 7):

- Image partitioning in a set of *p* small rectangular regions *Ri*
- Rotation-Translation-Scale identification on each region *Ri*
- Definition of a set of *p* matching points (C_{1i}, C_{2i}), where C_{1i} and C_{2i} are respectively the center of region *Ri* in image I_1, and the corresponding point in image I_2 using the Rotation-Translation-Scale transformation model obtained for this region *Ri*.
- Identification of the global homographic matrix **H** from the set of points (C_{1i}, C_{2i}).

Author-produced version of the article published in Precision Agriculture, 2016, N°17, p.564-587.
The original publication is available at http://link.springer.com/article/10.1007%2Fs11119-016-9437-x
Doi: 10.1007/s11119-016-9437-x

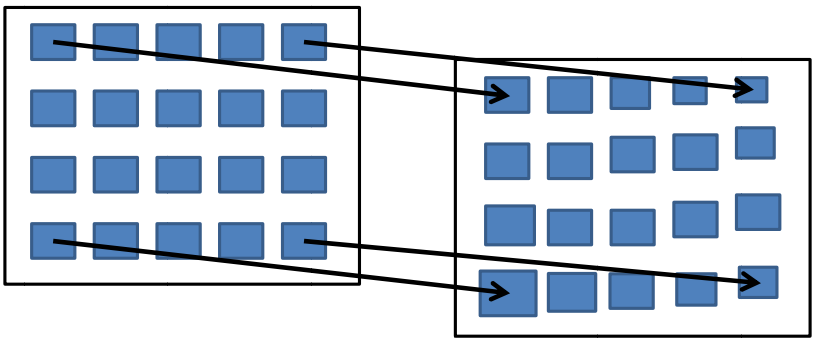


Figure 7. Large size image registration scheme

2.3. Distortion correction

Image acquisition by means of a still camera does not match exactly the projective projection model mentioned above. The image projection on the plane CCD sensor through the camera lens induces a non-linear deformation referred as "lens distortion". In most cases, lens distortion can be modeled as a radial distortion, where the (x^*, y^*) co-ordinates on the sensor can be obtained from the theoretical perspective projection model co-ordinates (x,y) as follows:

$$r^* = r.(1 + a_1.r^2 + a_2.r^4 + ... + a_n.r^{2n}) \quad (11)$$

where r and r^* are respectively the radius of the polar representation of initial and corrected co-ordinates (x, y) and (x^*, y^*), centered on the optical axis of the camera.

Thus, radial distortion can be identified using n+2 parameters:

- cx, cy : optical center co-ordinates on the CDD sensor (pixels)
- a_1, …, a_n: distortion coefficients

In order to identify both camera distortions and to refine homographic coefficients accordingly, a set of parameters **P** = (**D**$_1$, **H**, **D**$_2$) is considered, which represent respectively the distortion parameters for image I_1, the homographic matrix and the distortion parameters for image I_2. A cost function involving all the p matching points (C_{1i} , C_{2i}) is then defined as:

$$\mathrm{CF}(\mathbf{P}) = \sum_{i=1}^{p} (CEst_{2i} - C_{2i})^2 \quad (12)$$

where $CEst_{2i}$ is the estimated position in image I_2 computed from C_{1i} using the set of parameters **P** (Figure 8):

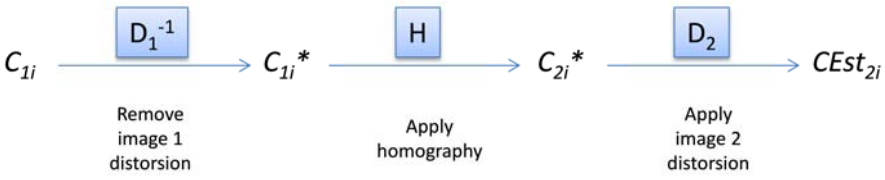


Figure 8. Position estimation on image I_2 from image I_1

The cost function CF(**P**) is minimized iteratively using the Gauss-Newton method, starting from the initial homographic matrix **H** issued from the previous identification step.

Author-produced version of the article published in Precision Agriculture, 2016, N°17, p.564-587.
The original publication is available at http://link.springer.com/article/10.1007%2Fs11119-016-9437-x
Doi: 10.1007/s11119-016-9437-x

3. Material and methods

Image acquisition

Aerial images of wheat were acquired using a couple of Sigma DP2 Merril still cameras (Sigma, Kawasaki, Japan) mounted on a UAV AR 200 (AirRobot® GmbH, Arnsberg, Germany) (see Figure 9). The images were acquired in Arganda, Madrid, May 2013 (40.315421° N, 3.483774 E). Late wheat was specially prepared to be at early stage at this period, for experimentation purposes. One of the cameras was modified for NIR acquisition, i.e. internal NIR blocking filter was removed and replaced by an external NIR band-pass filter. Image size was 4704x3136 pixels. Two kinds of flight have been considered: 60 m and 120 m altitudes, leading respectively to spatial resolutions of 10 mm per pixel and 20 mm per pixel.

Figure 9. Two Sigma DP2 Merril cameras mounted on a UAV

Images were shot in manual exposure mode with the following parameters:

- Speed 1/800 , aperture F5.0, 100 ISO for visible images
- Speed 1/800, aperture F7.1, 100 ISO for NIR images

Images were then converted from raw image format X3F to TIFF Adobe RGB 16 bits, using Sigma proprietary software (SIGMA Photo Pro 5.0). Green channel of the visible TIFF image and red channel of the near infrared TIFF image were respectively selected as 16 bits grey level images for registration input.

For practical reasons, all image pairs are referred to in the following according to the original names of images I_1 and I_2. As an example, SDIM0861_C2__SDIM0989_C1 indicates the 2nd channel (green) of the visible image SDIM0861.tif and the 1st channel (red) of the NIR image SDIM0989.tif.

Image registration

All the registration processes have been implemented according to the theory of operation described above with the following features:

- Before each image matching, the NIR image was rotated by 180° (to compensate for the opposite mounting of the cameras as shown in Figure 9).
- Before every Fourier transform, a hanning window (Blackman et Tukey, 1959) was applied to both images, in order to remove image edge artefacts.

Author-produced version of the article published in Precision Agriculture, 2016, N°17, p.564-587.
The original publication is available at http://link.springer.com/article/10.1007%2Fs11119-016-9437-x
Doi: 10.1007/s11119-016-9437-x

- Before every inverse Fourier transform for maximum peak search, a Gaussian filtering envelop was introduced in order to smooth the peak and allow its sub-pixel location.
- For every peak search, the peak amplitude was compared to the average amplitude of the whole spectrum, and rejected if the ratio was lower than a threshold (set equal to 10).
- In order to overcome sub-image identification failures due to large initial translations in the images (typically more than 100 pixels), the identification process was organized in three steps:

  i) an initial FMT registration applied to entire images, after reducing them by a factor of 10, leading to a rough similarity transformation,

  ii) an initial partitioning in sub-regions of 600x600 pixels, leading to a first set of about 40 matching points, according to the procedure described in the previous section

  iii) a second partitioning in sub-regions of 200x200 pixels, leading to a second set of about 300 to 340 matching points, according to the procedure described in the previous section

In the 2nd and 3rd steps, the partitioning was made according to the transformation model obtained in the previous step: centers of sub-regions in image I_1 were regularly distributed, and centers of the sub-regions in image I_2, were defined as the transformation of centers of the sub-regions in image I_1.

- In the 2nd and 3rd steps, the homographic matrix was computed from the set of matching points using a RANSAC algorithm (Fischler and Bolles, 1981) with 5000 iterations and an error tolerance of respectively 20 pixels on the 2nd step and 5 pixels on the 3rd step. Images with a number of inliers less than 50% of the number of matching points were rejected.
- The distortion model was limited to one distortion parameter (n=1), and a loop of 10 iterations of the Gauss-Newton minimization was systematically applied.

As the final result of the whole process, a complete transformation model $\mathbf{P} = (\mathbf{D_1}, \mathbf{H}, \mathbf{D_2})$ was provided (see previous section). This model was then used to geometrically transform the original NIR image, so that it can be exactly matched with the visible one as a new layer.

CUDA implementation

All computations related to Fourier transform and spatial spectrum analyses have been implemented on NVIDIA® GPU using CUDA (Compute Unified Device Architecture) C++ extension language, and the FFT computation library provided in the CUDA package.

Results and discussion

The procedure described above has been successfully applied to the two sets of wheat crop images: 39 image pairs with 10 mm spatial resolution (60 m elevation), and 20 image couples with 20 mm spatial resolution (120 m elevation). No parameter tuning has been made from one set to the other, or within a given set. For illustration, the detailed results obtained on a typical pair of visible and NIR images are first presented. More general results for all image pairs are then provided. Finally, a different example with a stronger homographic transformation is also presented.

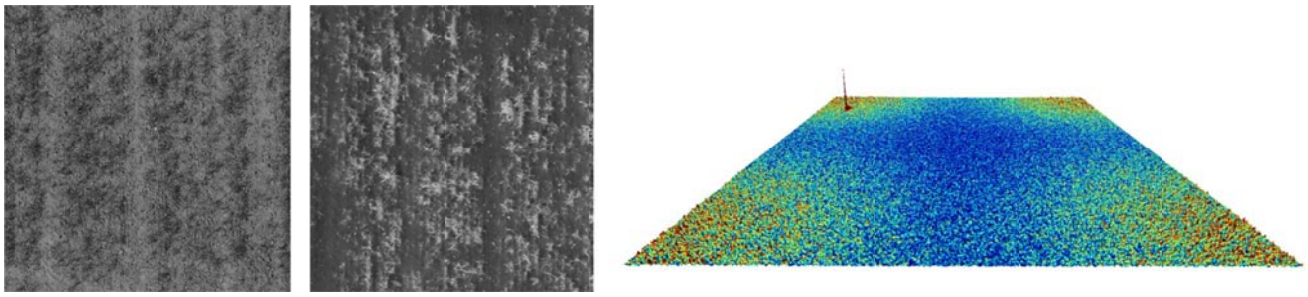
Detailed results for one image pair

Author-produced version of the article published in Precision Agriculture, 2016, N°17, p.564-587.
The original publication is available at http://link.springer.com/article/10.1007%2Fs11119-016-9437-x
Doi: 10.1007/s11119-016-9437-x

The first image pair of the 10 mm resolution image set is considered here, referred to as SDIM0861_C2__SDIM0989_C1, in order to examine some details of the registration process.

Matching point determination

The first fundamental concern in applying the Fourier-Mellin approach for registration was its ability to deal with images of different nature, i.e. visible and near-infrared. As an illustration, Figure 10 shows an example of translation identification, on a pair of 200x200 windows extracted from the pair under study (after transforming the NIR image according to the results of registration, in order to limit this example to translation). The translation peak is much less sharp than in Figure 1, but is still usable for a robust identification.

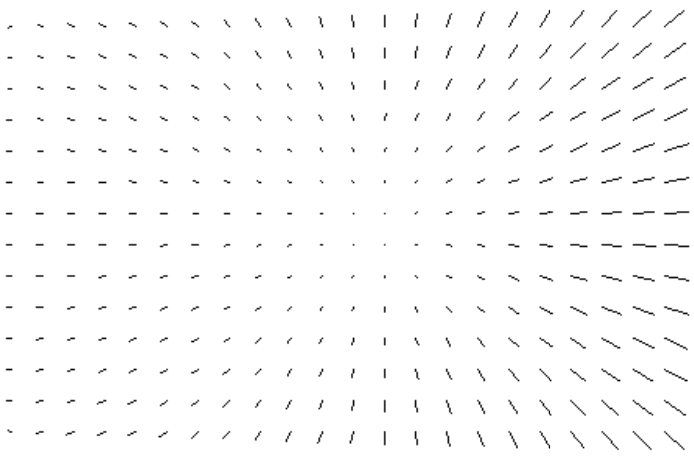


Visible image NIR image Translation peak (3D view)

Figure 10. Example of translation identification between visible and NIR images

As explained above, the identification procedure has been applied in three steps. In the second step, the image partitioning in 40 windows of size 600x600 has allowed determination of a first homographic matrix **H**, with translation terms $(T_x, T_y) =$ (133.267, -14.3521) (pixels). Large translations limit the number of windows in the visible image that have a counterpart entirely included in the NIR image. In the present case, 308 matching windows of size 200x200 have been determined for the third step.

Among those 308 matching windows, no identification procedure has failed, leading to 308 pairs of matching points. Figure 11 shows the repartition of these matching points in the visible image, as well as the disparity of their co-ordinate translations in the NIR image. It visually confirms that the transformation between visible and NIR images is not a simple translation, and that the radial distortions play an important part in it.



Author-produced version of the article published in Precision Agriculture, 2016, N°17, p.564-587.
The original publication is available at http://link.springer.com/article/10.1007%2Fs11119-016-9437-x
Doi: 10.1007/s11119-016-9437-x

Figure 11. Transformation disparity for image pair SDIM0861_C2__SDIM0989_C1.
Segments indicate the 308 difference vectors ($T_{12} - T_{12}c$) in the visible image, where T_{12} is the translation vector between matching point co-ordinates in NIR and visible images, and $T_{12}c$ is the T_{12} value closest to the image center. Segments are magnified (x10) for visibility.

Transformation model computation

In the 3rd step, using the 308 pairs of matching points, a new homographic matrix has been computed by the RANSAC algorithm with a 5 pixel tolerance, giving no outliers. Then the distortion parameters have been adjusted as described earlier.

For each couple (C_{1i} , C_{2i}), where C_{1i} and C_{2i} are matching points respectively in visible and NIR images, the error of a transformation model TM has been defined as the distance in the NIR image between C_{2i} and its estimated position $CEst_{2i}$ = TM(C_{1i}). In the present case, the following error figures have been obtained:

	Number of links	RMS error	Maximal error
Step 2 (200x200 windows)	308	1.2626	3.1953
Step 3 (distortion adjustment)	308	0.205765	1.01606

As expected from Figure 11, the distortion adjustment dramatically reduces the root mean square (RMS) error. In Figure 12, the residual error for every matching point has been plotted as a function of the distance to the image center. It shows that the largest errors only concern a few points at the borders of the image, while for others the error remains below 0.4 pixels.

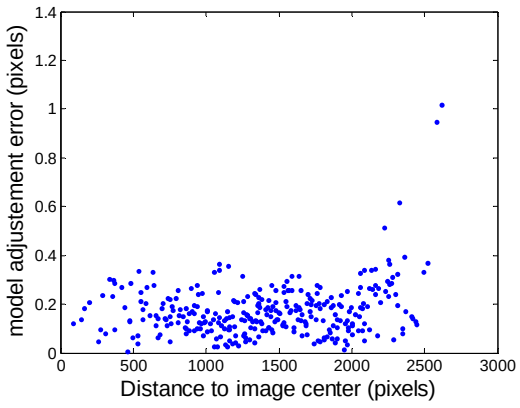


Figure 12. Final registration error repartition for image pair SDIM0861_C2__SDIM0989_C1

The final transformation parameters for the image pair SDIM0861_C2__SDIM0989_C1 are given below:

$$\mathbf{H} = \begin{bmatrix} 0.988860 & -0.000585 & 141.303881 \\ -0.005695 & 0.997577 & -8.8490590 \\ -3.608560\,10^{-6} & -2.036398\,10^{-7} & 1 \end{bmatrix}$$

Distortion parameters (center: *cx*, *cy* ; radial distortion: $1 + a_1.r^2 + a_2.r^4 + a_3.r^6$):

Author-produced version of the article published in Precision Agriculture, 2016, N°17, p.564-587.
The original publication is available at http://link.springer.com/article/10.1007%2Fs11119-016-9437-x
Doi: 10.1007/s11119-016-9437-x

Visible image :	cx =2331.56	cy =1593.41	a_1=-5.47682 10^{-9} ; $a2$=$a3$=0
NIR image :	cx =2351.76	cy =1568	a_1=-5.69993 10^{-9} ; $a2$=$a3$=0

As expected, the ε_x and ε_y values of the homographic matrix are very low (less than 10^{-5}). However, according to the image size, they lead to a scale variation of about 1.7 % from top-left to bottom-right of the image.

Finally, Figure 13 shows a general view of the original images, as well as the registration result in false color (NIR, R, B), on which the translation between images can be observed. Detail in false color shows the quality of registration, compatible with NDVI computation at 10 mm spatial resolution.

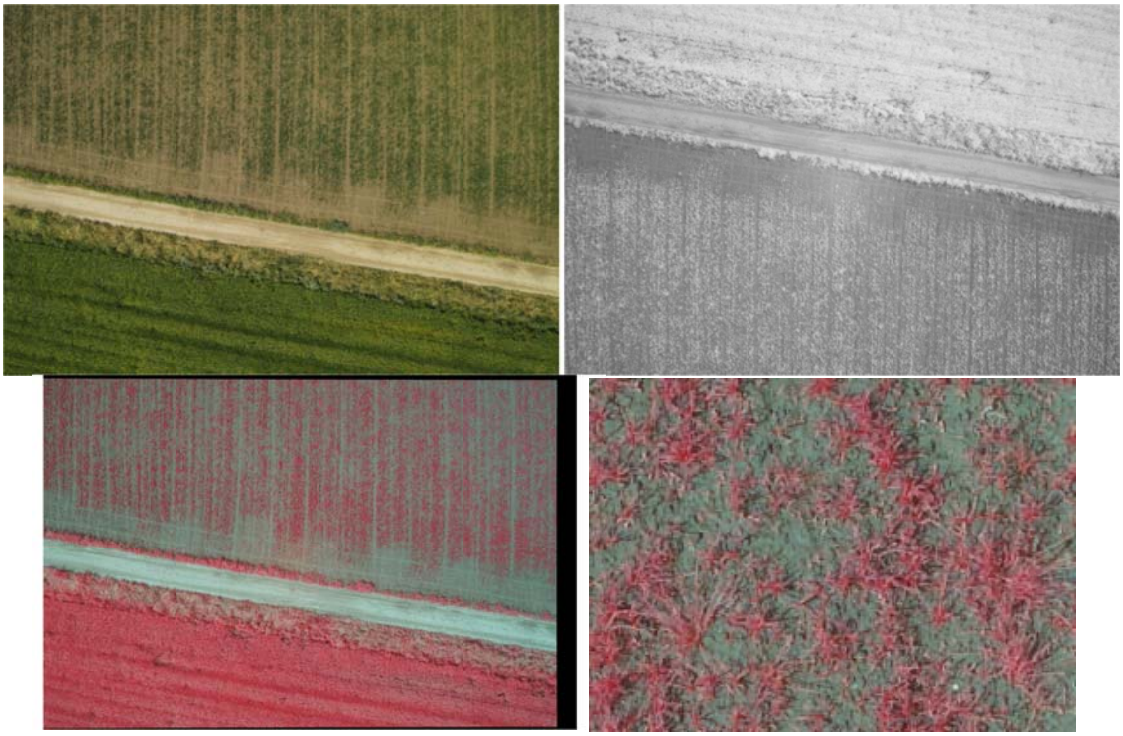


Figure 13. Registration of pair SDIM0861-C2__SDIM0989_C1. Top: original images (visible + NIR). Bottom: false color (NIR, R, G) representation of the registration result (general view and detail)

Comparison with a feature-based approach

For comparison purposes, the registration result obtained for the same image pair with a commercial registration tool is presented here: Imagine AutoSync (ERDAS Imagine), which is based on feature point matching.

Figure 14 shows the matching points that have been retained by the software. As already discussed, most of them are concentrated in homogeneous zones in terms of material content, i.e. the bare soil pathway and the dense vegetation area in the top of the picture. Because of a lack of matching points, registration inaccuracy can be expected in the sparse vegetation zone.

Author-produced version of the article published in Precision Agriculture, 2016, N°17, p.564-587.
The original publication is available at http://link.springer.com/article/10.1007%2Fs11119-016-9437-x
Doi: 10.1007/s11119-016-9437-x

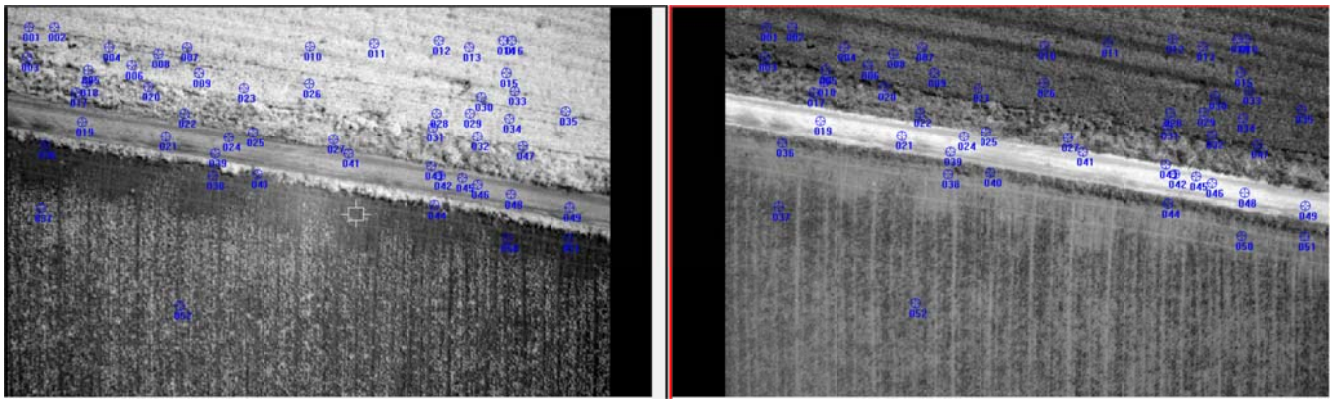


Figure 14. Set of matching points obtained for the pair SDIM0861-C2__SDIM0989_C1 with ERDAS Imagine Autosync.

In Figure 15, a detail of the resulting false color image reconstruction in the sparse vegetation zone is presented. A slight blurring effect can be observed compared to the same reconstruction using FMT registration, which is due to a less accurate correspondence between NIR and visible radiometric values at a given spatial position. Such a positioning error, with an order of magnitude of about one pixel, could be considered as visually negligible. However, it becomes important in the case of multi-spectral processing. In Figure 16, the results for a simple NDVI computation are compared: NDVI has been computed from the NIR and R components after inverse gamma correction (according to the Adobe RGB standard) and then linearly converted from [-1, 1] range to [0, 255] range for image representation. An automatic thresholding based on the Otsu algorithm (Otsu, 1979) has then been applied to simulate vegetation segmentation (Figure 17). It can be observed that the segmentation result obtained with the feature point approach presents more noisy patterns, which indicate less accurate multi-spectral information, and thus confirm a less accurate NIR-visible registration.

Finally, a quantitative comparison has been made using the tri-dimensional histograms of sub-images in Figures 15.a and 15.b, and computing in each case the mutual information MI (Pluim, 2003) between R and NIR channels. The results are given in Table 1.

Table 1: comparison of R-NIR mutual information

	Entropy R	Entropy NIR	Joint Entropy	**MI**
FMT registration	4.549	4.657	9.055	**0.150**
Feature point registration	4.557	4.653	9.126	**0.084**

The MI obtained with FMT registration is nearly the double of the one obtained using the feature point approach, confirming the previous visual assessment (Figures 15 to 17).

Author-produced version of the article published in Precision Agriculture, 2016, N°17, p.564-587.
The original publication is available at http://link.springer.com/article/10.1007%2Fs11119-016-9437-x
Doi: 10.1007/s11119-016-9437-x

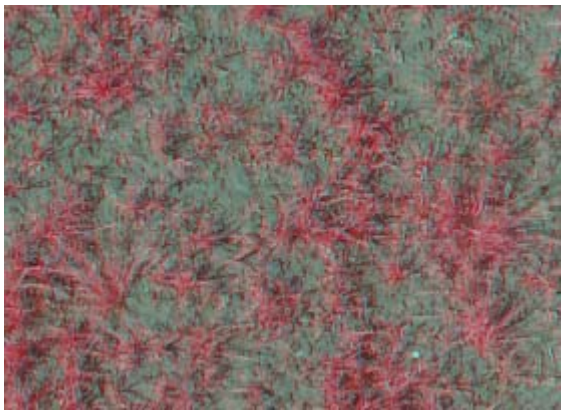
a  b 

Figure 15: false color representation of registration results

a) Feature point approach b) FMT approach

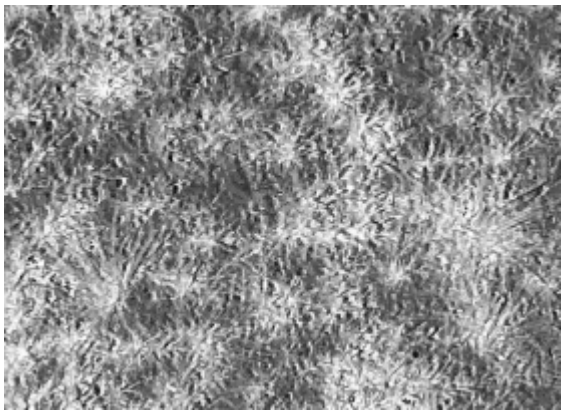
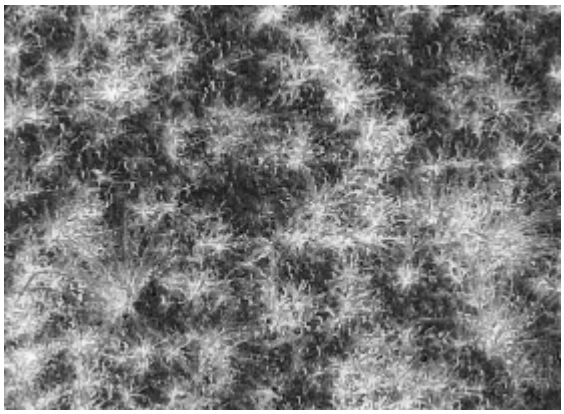
a  b 

Figure 16: NDVI computation results

a) Feature point approach b) FMT approach

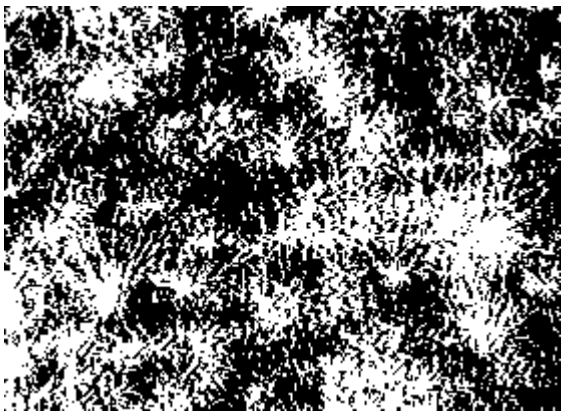
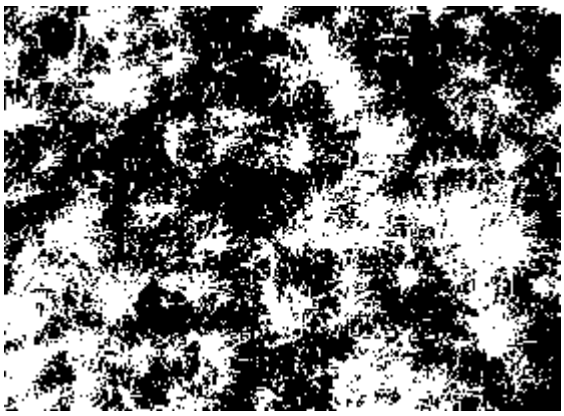
a  b 

Figure 17: NDVI thresholding

a) Feature point approach b) FMT approach

General results for all image pairs

The more significant figures obtained for the two sets of images are reported in Table 2 and Table 3 respectively. These figures are:

Author-produced version of the article published in Precision Agriculture, 2016, N°17, p.564-587.
The original publication is available at http://link.springer.com/article/10.1007%2Fs11119-016-9437-x
Doi: 10.1007/s11119-016-9437-x

- The number of 200x200 windows (NbW) that have been used for the second step of identification. Because the window distribution in the second image depends on the homographic computation issued from the first step, this number can vary from one image pair to another. It ranges between 294 and 345.
- The number of inliers after RANSAC computation of the homographic matrix in step 3.
- The root mean square error (RMS) and maximal error obtained at the end of the registration process. These values are computed considering all the inliers.
- The translation between visible and NIR images according to the transformation model (terms T_x and T_y of the homographic matrix)

The set of images at 10 mm resolution provides particularly good results. Only two sub-window identifications have failed (for SDIM0873_C2__SDIM1001_C1 and SDIM0898_C2__SDIM1025_C1) over about 12000 (39 image pairs). RMS error remains between 0.2 and 0.44 pixels. Maximal error is up to 1.81 pixel (as seen before, it corresponds to the borders of the images).

It is worth noticing that translation values T_x and T_y are very different from one image pair to the other. It means that most of the translations observed are due to random shooting latency of the two cameras combined with UAV attitude instability, rather than to their mechanical positioning.

Table 2. Main results for 10 mm resolution image set

Couple	NbW	Inliers	RMS (pixels)	Max error (pixels)	Tx	Ty
SDIM0861_C2__SDIM0989_C1	308	308	0.205765	1.01606	141.304	-8.84906
SDIM0862_C2__SDIM0990_C1	310	310	0.268648	1.61977	206.65	7.2018
SDIM0863_C2__SDIM0991_C1	309	309	0.248254	1.75487	179.991	6.23115
SDIM0864_C2__SDIM0992_C1	308	308	0.213607	1.03461	134.235	-70.6822
SDIM0865_C2__SDIM0993_C1	330	330	0.246042	1.55204	164.918	35.8471
SDIM0866_C2__SDIM0994_C1	308	308	0.260177	1.01122	216.79	-23.4694
SDIM0867_C2__SDIM0995_C1	308	308	0.205254	1.08457	127.934	-56.9324
SDIM0868_C2__SDIM0996_C1	308	308	0.199852	0.993497	135.468	-19.6565
SDIM0869_C2__SDIM0997_C1	330	330	0.245332	1.49931	172.553	12.098
SDIM0870_C2__SDIM0998_C1	345	345	0.225745	1.63805	74.4094	21.4359
SDIM0871_C2__SDIM0999_C1	329	329	0.253018	1.78237	167.997	26.0862
SDIM0872_C2__SDIM1000_C1	308	308	0.201389	1.01772	112.148	-2.6374
SDIM0873_C2__SDIM1001_C1	**308**	**307**	0.277907	1.20625	190.863	-51.4321
SDIM0874_C2__SDIM1002_C1	308	308	0.203207	1.13215	140.025	-3.1145
SDIM0875_C2__SDIM1003_C1	308	308	0.202374	0.657403	-99.231	-2.28731
SDIM0876_C2__SDIM1004_C1	330	330	0.343585	1.32893	-121.629	20.3692
SDIM0877_C2__SDIM1005_C1	330	330	0.293634	1.18788	-139.211	0.988007
SDIM0878_C2__SDIM1006_C1	330	330	0.364554	1.64644	-94.0046	20.0194
SDIM0879_C2__SDIM1007_C1	308	308	0.263832	1.00428	-195.768	-51.1397
SDIM0880_C2__SDIM1008_C1	313	313	0.255508	0.869559	-170.886	-28.6676
SDIM0881_C2__SDIM1009_C1	294	294	0.280621	1.02383	-228.698	-130.32
SDIM0882_C2__SDIM1010_C1	308	308	0.246351	1	-164.136	-83.1163
SDIM0883_C2__SDIM1011_C1	325	325	0.282107	1.19546	-154.086	-13.6132
SDIM0884_C2__SDIM1012_C1	330	330	0.415529	1.70528	-60.063	25.6828
SDIM0886_C2__SDIM1013_C1	330	330	0.233672	1.6979	106.579	23.3675
SDIM0887_C2__SDIM1014_C1	330	330	0.235286	1.64073	130.636	42.0943
SDIM0888_C2__SDIM1015_C1	330	330	0.24275	1.64965	147.942	63.0498
SDIM0889_C2__SDIM1016_C1	330	330	0.256439	1.75686	151.126	13.9706
SDIM0890_C2__SDIM1017_C1	330	330	0.239517	1.47383	126.01	105.774
SDIM0891_C2__SDIM1018_C1	330	330	0.250685	1.69992	109.195	42.9601

Author-produced version of the article published in Precision Agriculture, 2016, N°17, p.564-587.
The original publication is available at http://link.springer.com/article/10.1007%2Fs11119-016-9437-x
Doi: 10.1007/s11119-016-9437-x

SDIM0892_C2__SDIM1019_C1	312	312	0.443181	1.63157	121.565	7.33859
SDIM0893_C2__SDIM1020_C1	330	330	0.270487	1.68809	110.432	67.1179
SDIM0894_C2__SDIM1021_C1	343	343	0.246417	1.81263	78.996	40.0072
SDIM0895_C2__SDIM1022_C1	308	308	0.37076	1.37542	128.205	227.749
SDIM0896_C2__SDIM1023_C1	330	330	0.247288	1.74025	155.986	26.5518
SDIM0897_C2__SDIM1024_C1	316	316	0.261298	1.77994	176.179	6.57355
SDIM0898_C2__SDIM1025_C1	**335**	**334**	0.227009	1.73867	91.7082	29.1066
SDIM0899_C2__SDIM1026_C1	345	345	0.235825	1.72096	40.737	61.905
SDIM0900_C2__SDIM1027_C1	330	330	0.309022	1.4055	144.271	59.9648

The set of images at 20 mm resolution provides comparable results except in two cases (pairs SDIM0904_C2__SDIM1031_C1 and SDIM0907_C2__SDIM1034_C1) where respectively 53 and 59 sub-window identifications have failed. For these pairs, maximal errors are up to 5 pixels. When examining the corresponding images, it can be observed that these images are blurred due to camera motion (Figure 18).

Table 3. Main results for 20 mm resolution image set

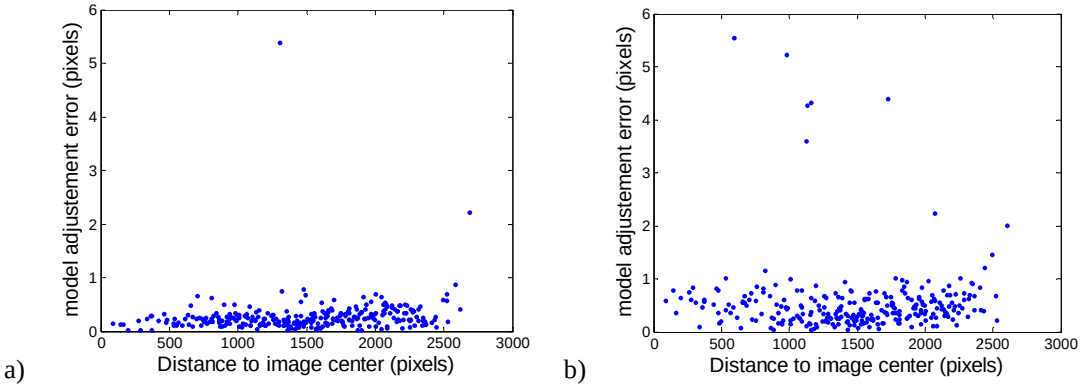
Couple	NbW	Inliers	RMS (pixels)	Max error (pixels)	Tx	Ty
SDIM0901_C2__SDIM1028_C1	345	345	0.238231	1.62974	19.837	33.0571
SDIM0902_C2__SDIM1029_C1	340	340	0.252882	1.81972	17.4474	54.3257
SDIM0903_C2__SDIM1030_C1	320	320	0.232668	1.29357	-84.5478	-10.915
SDIM0904_C2__SDIM1031_C1	**345**	**292**	**0.449388**	**5.40087**	**19.6816**	**17.9849**
SDIM0905_C2__SDIM1032_C1	345	345	0.236375	1.70266	15.3072	38.9054
SDIM0906_C2__SDIM1033_C1	332	332	0.219682	1.55367	2.05891	11.721
SDIM0907_C2__SDIM1034_C1	**330**	**271**	**0.877862**	**5.55471**	**-64.2983**	**27.52**
SDIM0908_C2__SDIM1035_C1	330	330	0.209042	0.938976	1.72416	8.94636
SDIM0909_C2__SDIM1036_C1	341	341	0.214602	1.52815	10.616	25.1422
SDIM0910_C2__SDIM1037_C1	336	336	0.212402	1.60552	12.199	45.3367
SDIM0911_C2__SDIM1038_C1	315	315	0.375712	1.21472	-16.1354	-15.9621
SDIM0912_C2__SDIM1039_C1	330	330	0.193718	1.15938	2.19996	36.8913
SDIM0913_C2__SDIM1040_C1	345	345	0.21571	1.55553	32.6721	49.4492
SDIM0914_C2__SDIM1041_C1	330	329	0.254166	1.14625	-13.2549	30.7771
SDIM0915_C2__SDIM1042_C1	330	330	0.358858	1.44805	-55.6044	38.0781
SDIM0916_C2__SDIM1043_C1	345	344	0.33042	1.84099	15.7268	50.7469
SDIM0917_C2__SDIM1044_C1	340	340	0.714914	2.1673	8.88959	21.606
SDIM0918_C2__SDIM1045_C1	330	330	0.276748	1.28785	-28.2551	36.5391
SDIM0919_C2__SDIM1046_C1	329	329	0.273745	1.60227	-27.7507	-2.04273
SDIM0920_C2__SDIM1047_C1	308	308	0.223654	1.38067	-26.9073	-45.8858

Author-produced version of the article published in Precision Agriculture, 2016, N°17, p.564-587.
The original publication is available at http://link.springer.com/article/10.1007%2Fs11119-016-9437-x
Doi: 10.1007/s11119-016-9437-x

SDIM0901 SDIM0904 SDIM0907

Fig. 18. Zoomed detail of 20 mm resolution images for sharpness assessment (the blue objects are geo-referencing targets of 30 mm diameter distributed around the crop field)

Moreover, these large error values are not limited to image borders, as indicated by Figure 19.



a) b)

Figure 19. Registration error repartition for image pairs SDIM0904_C2__SDIM1031_C1 (a) and SDIM0907_C2__SDIM1034_C1 (b)

Thanks to the GPU implementation, the processing time has been reduced from about 3 min with a conventional CPU implementation (Intel Xeon CPU, 3.19 GHz, 8 Go RAM) to 30 s per image pair, even using a low-end NVIDIA Quadro 400 graphic card with 500 Mbytes of memory.

Example of stronger homographic transformation

In the image sets presented previously, the homographic transformation models were close to affine transformation, due to a rigorous image acquisition protocol and camera alignment. Indeed, the first processing step (initial registration on reduced images) was optional. The registration result for a pair of images acquired with the same sensors in a different context (south of France, May 2015) is presented in figure 20. Visible and NIR images were captures during successive flights. Due to the time delay between images (tree shadow is different in both images), the registration has a lower accuracy. However, it illustrates the ability of the registration procedure to deal with stronger homographic transformations. In the present case, the three successive steps were mandatory.

NbInliers	RMS error	Max error
236	0.638701	3.08678

Author-produced version of the article published in Precision Agriculture, 2016, N°17, p.564-587.
The original publication is available at http://link.springer.com/article/10.1007%2Fs11119-016-9437-x
Doi: 10.1007/s11119-016-9437-x

Transformation matrix:

$$\mathbf{H} = \begin{bmatrix} 1.0288 & -0.047 & 219.1279 \\ 0.03087 & 0.962 & -54.7878 \\ 1.3939\ 10^{-5} & -1.807\,10^{-5} & 1 \end{bmatrix}$$

Figure 20. Registration of a strong homographic transformation

5. Conclusion

The assembly of several cameras close to each other is a relevant solution for multi-spectral image acquisition with a UAV. However, it requires a very accurate registration between bands to further compute multi-spectral indices such as NDVI. In this context, the visible-NIR registration procedure proposed here has proved to be particularly efficient for agricultural scenes in which vegetation and bare soil areas are deeply nested. Moreover, this procedure is very robust: for the set of data under study, no parameter tuning was necessary from one image pair to another, and (though it was no presented in this study), it has been also observed that the choice of the visible band (blue, green or red) had no influence.

An error below 0.3 pixels was obtained in most cases, except at the image borders. Error at the image border is due to an incomplete modeling of the sensor distortion. Starting from the set of matching points obtained from sub-window processing, a more sophisticated camera modeling, e.g. issued from laboratory calibration, could probably lead to still better results. However, this issue is not specific to the Fourier-Mellin approach, and would be similar with a registration process based on feature points. Another possible way of improvement could be to introduce a last step of local refinement, e.g. using mutual information minimization on each sub-region (associated with a flexible transformation model).

Author-produced version of the article published in Precision Agriculture, 2016, N°17, p.564-587.
The original publication is available at http://link.springer.com/article/10.1007%2Fs11119-016-9437-x
Doi: 10.1007/s11119-016-9437-x

The main limitation, compared to feature point approaches, is related to some geometric assumptions that are mandatory, i.e. large camera-scene distance and camera orientation close to the nadir. Nevertheless, these assumptions are usually reached in the context of aerial imagery.

Because it is well adapted to images of different nature, beyond the acquisition of high quality NDVI data, the proposed approach could certainly be used in many other registration problem involving several image acquisitions in different optical bands, either simultaneously (multi-sensor cameras) or sequentially (multiple flights).

The registration of thermal and optical images is a particular case that could also be investigated. It would require some software adaptations, in order to deal with images of different native size and resolution.

Acknowledgements

The research leading to these results has received funding from the European Union's Seventh Framework Program [FP7/2007-2013] under grant agreement n°245986.

Conflicts of Interest

The authors declare that they have no conflict of interest.

References

Bay, H., Ess A., Tuytelaars T., & Van Gool L.(2008). SURF: Speeded Up Robust Features. *Computer Vision and Image Understanding*, 110(3), 346-359.

Blackman, R. B. & Tukey, J. W. (1959). B.5: Particular Pairs of Windows. In *The Measurement of Power Spectra, From the Point of View of Communications Engineering*. New York: Dover, pp.499-504.

Casasent, D. & Psaltis, D. (1977). New Optical Transforms for Pattern Recognition. *Proceedings of the IEEE*, 65 (1), 77-84.

Dare, P.M. (2008). Small format digital sensors for aerial imaging applications. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. Vol. XXXVII. Part B1.

Firmenich, D., Brown, M. &Susstrunk, S. (2011). Multispectral Interest Points for RGN-NIR Image Registration. In: Proc ICIP 2011: IEEE International Conference on Image Processing,, pp181 - 184

Fischler M.A. & Bolles R.C. (1981). Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Communications of the ACM*, 24 (6), pp. 381–395.

Author-produced version of the article published in Precision Agriculture, 2016, N°17, p.564-587.
The original publication is available at http://link.springer.com/article/10.1007%2Fs11119-016-9437-x
Doi: 10.1007/s11119-016-9437-x

Hermosillo, G., ChefdHotel, C., & Faugeras, O. (2002). Variational methods for multimodal image matching. *International Journal of Computer Vision*, 50(3), 329–343..

Huete, A.R., Liu, H.Q., Batchily, KL., van Leeuwen, W. (1997). A comparison of vegetation indices over a global set of TM images for EOS-MODIS. *Remote Sensing of Environment*, 59(3), 440-451.

Hunt, E. R, Jr., Hively, W. D., Fujikawa, S.J., Linden, D. S., Daughtry, C.S.T., & McCarty, G.W. (2010). Acquisition of NIR-Green-Blue Digital Photographs from Unmanned Aircraft for Crop Monitoring. *Remote Sensing*. 2(1), 290-305.

Jaud, M., Rouveure, R., Faure, P., & Monod, M.O. (2013). Methods for FMCW radar map georeferencing. *ISPRS Journal of Photogrammetry and Remote Sensing*, 84, 33-42.

Jindong, W., Dong, W. & Bauer, M.E.(2007).. Assessing broadband vegetation indices and QuickBird data in estimating leaf area index of corn and potato canopies. *Field crops research*, 102(1), 33-42.

Lebourgeois V, Bégué A, Labbé S, Mallavan B, Prévot L, & Roux B. (2008). Can commercial digital cameras be used as multispectral sensors? A crop monitoring test. *Sensors*, 8, 7300-7322.

Lowe, David G. (1999). Object recognition from local scale-invariant features. *In: Proceedings of the 7th International Conference on Computer Vision*, Ed. Bob Werner, IEEE Computer Society, Los Alamitos, CA, US, pp. 1150–1157.

Marcel, B., Briot, M., & Murrieta, R. (1997). Calcul de translation et rotation par la transformation de Fourier (Translation and rotation computation by Fourier Transform). *Traitement du Signal*, 14(2), 135-149..

Myneni, R.B., Hall, F.G , Sellers, P.J., & Marshak, A.L. (1995). The interpretation of spectral vegetation indexes. *IEEE Transactions on Geoscience and Remote Sensing*, 33, 481-486.

Oliveira, F.P., & Tavares, J.M. (2014). Medical image registration: a review. *Computer Methods in Biomechanics and Biomedical Engineering*,17(2),73-93.

Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man and Cybernetics*, 9(1), 62-66 .

Petrie, G. (2005) Airborne Pushbroom Line Scan; An Alternative to Digital Frame Cameras. GEOInformatics Magazine, January-February 2005, pp 50-57.

Pluim, J.P.W, Maintz,J.B.A., & Viergever M.A. (2003). Mutual-information-based registration of medical images: a survey. *IEEE Transactions on Medical Imaging* , 22 (8), 986-1004.

Rabatel, G, Gorretta, N. & Labbé, S. (2014). Getting simultaneous red and near-infrared band data from a single digital camera for plant monitoring applications: theoretical and practical study. *Biosystems Engineering,* 117(1), pp. 2-14.

Reddy, B. S. & B. N. Chatterji. (1996). An FFT-based technique for translation, rotation, and scale-invariant image registration. *IEEE Transactions on Image Processing*. 5(8), 1266-1271.

Rouse, J. W., Haas, R. H., Schell, J. A. & Deering D. W. (1973). Monitoring vegetation systems in the Great Plains with ERTS. In: *Proceedings of the 3rd ERTS Symposium*, , Vol. 1, 309-317.

Sargent, D., Chen C-I, Tsai, C-M, Wang Y-F, & Koppel, D. (2009). Feature Detector and Descriptor for Medical Images. In: *Proc. SPIE 7259, Medical Imaging 2009: Image Processing*, 72592Z ([page nos?]

Torres-Sanchez J, Lopez-Granados F, De Castro AI, & Pena-Barragan JM. (2013). Configuration and Specifications of an Unmanned Aerial Vehicle (UAV) for Early Site Specific Weed Management. *PLoS ONE*, 2013, 8(3): e58210.

Zhi Li-jia, Z., Zhang Shao-min, Z., Da-zhe, Z., Zhao, H., & Shu-kuan, L. (2009). Medical Image Retrieval Using SIFT Feature. In: *Qiu, P., Yiu, C., Zhang, H., Wen, X., Proc. of CISP '09. 2nd International Congress on Image and Signal Processing, IEEE, New-York, USA,* 1-4.

Zitova, B., & Flusser, J. (2003). Image registration methods: a survey. *Image and Vision Computing* , 21(11), 977–1000.

Zhengwei, Y., Hu Zhao, Liping Di, Yu, G. A (2009). Comparison of Vegetation Indices for Corn and Soybean Vegetation Condition Monitoring. In: *Harold Annegarn*, *Geoscience and Remote Sensing Symposium, IGARSS 2009*, *IEEE, New-York, USA*, IV-801 - IV-804.