

Two-step multi-spectral registration via key-point detector and gradient similarity. Application to agronomic scenes

Jehan-Antoine VAYSSADE¹ ^a, Gawain Jones¹ ^b, Jean-Noel Paoli¹ ^c and Christelle Gee¹

¹Agroécologie, AgroSup Dijon, INRA, Univ. Bourgogne-Franche-Comté, F-21000 Dijon, France
jehan-antoine.vayssade@inra.fr, {gawain.jones, jean-noel.paoli, christelle.gee}@agrosupdijon.fr

Keywords:

Registration, Multi-spectral imagery, Precision farming, Feature descriptor

Abstract:

The potential of multi-spectral images is growing rapidly in precision agriculture, and is currently based on the use of multi-sensor cameras. However, their development usually concerns aerial applications and their parameters are optimized for high altitudes acquisition by drone (UAV ≈ 50 meters) to ensure surface coverage and reduce technical problems. With the recent emergence of terrestrial robots (UGV), their use is diverted for nearby agronomic applications. It is possible to explore new agronomic applications, maximizing specific traits extraction (spectral index, shape, texture ...) which requires high spatial resolution.

The problem with these cameras is that all sensors are not aligned and the manufacturers' methods are not suitable for close-field acquisition, resulting in offsets between spectral images and degrading the quality of extractable information. We therefore need a solution to align the different images accurately in such condition. The objective of this article is to describe and evaluate a method that allows to define the ideal conditions for matching multi-spectral images from a multi-sensor camera at low heights of use applied on agronomic scenes.


In this study we propose a two-step method applied to the six-bands Airphen multi-sensors camera with (i) affine correction using pre-calibrated matrix at different heights, the closest transformation can be selected via internal GPS and (ii) perspective correction to refine the previous one, using key-points matching between enhanced gradients of each spectral bands. Nine types of key-point detection algorithms (ORB, GFTT, AGAST, FAST, AKAZE, KAZE, BRISK, SURF, MSER) with three different modalities of parameters was evaluated for their benchmark and their corresponding best reference spectra.


The results show that GFTT is the most suitable methods for key-point extraction using our enhanced gradients, and the best spectral reference was identified to be the 570 nm for this one. According to the results the initial error is about 62 px, with our methods, the remaining residual error is less than 1 px, where the manufacturer's involves distortions and loss of information with an estimated residual error of ≈ 12 px.


1 INTRODUCTION

Modern agriculture is changing towards a system that is less dependent on pesticides [1] (herbicides remain the most difficult pesticides to reduce) and digital tools are of great help in this matter. The development of imaging and image processing have made it possible to characterize

an agricultural plot [2] (crop health status or soil characteristics) using non-destructive agronomic indices [3, 4, 5] replacing traditional destructive and time-consuming methods. In recent years, the arrival of miniaturized multi-spectral and hyper-spectral cameras on Unmanned Aerial Vehicles (UAVs) has allowed spatio-temporal field monitoring. These vision systems have been developed for precise working conditions (flight height 50 m). Although, very practical to use, they are also used for proxy-sensing applications. However, the algorithms offered by manufactur-

^a  <https://orcid.org/0000-0002-7418-8347>

^b  <https://orcid.org/0000-0002-5492-9590>

^c  <https://orcid.org/0000-0002-0499-9398>

ers to co-register multiple single-band images at different spectral range, are not optimal for low heights. It thus requires a specific close-field image registration.

Image registration is the process of transforming different images of one scene into the same coordinate system. The spatial relationships between these images can be rigid (translations and rotations), affine (shears for example), homographic, or complex large deformation models (due to the difference of depth between ground and leafs) [6]. The main difficulty is that multispectral cameras have low spectral coverage between bands, resulting in a loss of characteristics between them. Which is caused by (i) plant leaves have different aspect depending on the spectral bands (ii) there are highly complex and self-similar structures in our images [7]. It therefore affects the process of detecting common characteristics between bands for image registration. There are two types of registration, feature based and intensity based [8]. Feature based methods work by extracting point of interest and use feature matching, in most cases a brute-force matching is used, making those techniques slow. Fortunately these features can be filtered on the spatial properties to reduce the matching cost. A GPGPU implementation can also reduce the comparisons cost. Intensity-based automatic image registration is an iterative process, and the metrics used are sensitive to determine the numbers of iteration, making such method computationally expensive for precise registration. Furthermore multi-spectral implies different metrics for each registered bands which is hard to achieve.

Different studies of images alignment using multi-sensors camera can be found for acquisition using UAV at medium (50 – 200 m) and high (200 – 1000 m) distance. Some show good performances (in term of number of key-points) of feature based [9, 10] with strong enhancement of feature descriptor for matching performances. Other prefer to use intensity based registration [7] on better convergence metrics [11] (in term of correlation), which is slower and not necessarily robust against light variability and their optimization can also fall into a local minimum [12].

Traditional approach to multi-spectral image registration is to designate one channel as the target channel and register all the others on the

selected one. Currently, only [9] show a method for selecting the best reference, but there is no study who as defined the best spectral reference in agronomic scene. In all cases NIR (850nm) or middle range spectral reference are conventionally used without studying the others on precision agriculture. In addition those studies mainly propose features matching without large methods comparison [9] (less than 4) of their performance (time/precision), without showing the importance of the spectral reference and the interest of normalized gradients transformation (like in Intensity-based methods).

However, despite the growing use of UGVs and multi-spectral imaging, the domain is not very well sourced, and no study has been found under agricultural and external conditions in near field of view (less than 10 meter) for multi-spectral registration.

Thus, this study propose a benchmark of popular feature extractors inside normalized gradients transformation and the best spectral reference was defined for each of them. Moreover a pre-affine registration is used to filter the feature matching, evaluated at different spatial resolutions. So this study shows the importance of the selection of the reference and the features extractor on normalized gradients in such registration.

The results of this study show that GFTT has the best overall performances in both computation time and accuracy in all modalities. Additionally the 570 nm is the best reference for GFTT and 710 nm for most of the others. At the lower height 1.6 m we have an initial error of ≈ 62 px, the manufacturer’s methods show an error of ≈ 12 px when our methods has < 1 px of error.

2 MATERIAL AND METHOD

2.1 Material

2.1.1 Camera

The multi-spectral imagery is provided by the six-band multi-spectral camera Airphen developed by HiPhen society. Airphen is a scientific multi-spectral camera developed by agronomists for agricultural applications. It can be embedded in different types of platforms such as UAV,

phenotyping robots, etc.

AIRPHEN is highly configurable (bands, fields of view), lightweight and compact. The camera was configured using interferential filter centered at 450/570/675/710/730/850 nm with FWHM¹ of 10nm, the position of each bands is referenced on figure 1. The focal lens is 8 mm for all wavelength. The raw resolution for each spectral band is 1280×960 px with 12 bit of precision. Finally the camera also provides an internal GPS antenna that can be used to get the distance from the ground.



Figure 1: Disposition of each bands on the Airphen multi-sensors camera

2.1.2 Datasets

Two datasets were taken at different heights with images of a chessboard and of an agronomic scene. We have used a metallic gantry for positioning the camera at different heights. The size of the gantry is $3m \times 5m \times 4m$. Due to the size of the chessboard (57×57 cm with 14×14 square of 4 cm), the limited focus of the camera and the gantry height, we have bounded the acquisition height from 1.6 m to 5 m with 20 cm steeps, which represent 18 acquisitions.

The first dataset is for the calibration. A chessboard is taken at different heights. The second one is for the alignment verification under real conditions. One shot of an agronomic scene is taken at different heights with a bias of $\pm 10cm$ to be in the worst case.

2.2 Methods

Alignment is refined in two stages, with (i) affine registration approximately estimated and (ii) perspective registration for the refinement and precision. As example the figure 2 shows each correction step, where the first line is for the (i) affine correction (section 2.2.1), the second is for (ii)

perspective correction. More precisely the second are per-channel pre-processing where feature detectors are used to detect key points (section 2.2.2). And the last line each channel key points are associated to compute the perspective correction through homography, to the selected spectral band (section 2.2.2). Those steps are explained on specific subsections.

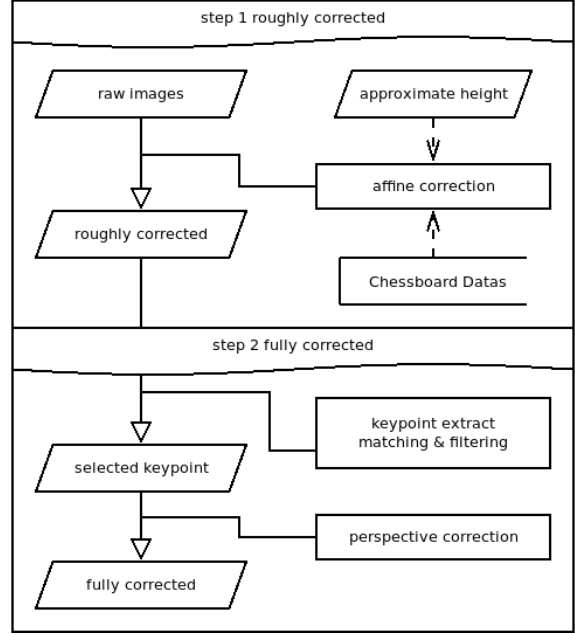


Figure 2: Each step of the alignment procedure, with (1) roughly corrected from affine correction and (2) enhancement via key-points and perspective

2.2.1 Affine Correction

We make the assumption that closer we take the snapshot, the bigger is the distance of the initial Affine Correction. On the other hand at a distance superior or equals to 5 m, the initial affine correction become stable. A calibration is used to build a linear model based on that assumption, which will allow the affine correction to work at any height. The main purpose of this step is to reduce the offset between each spectral band, which allows the similarity to be spatially delimited within a few pixels, making feature matching more effective.

Calibration Based on that previous assumption a calibration is run over the first dataset (i.e the chessboard). We detect the chessboard using the opencv calibration toolbox [13] on each spectral image (normalized by $I = (I - \min(I)) / \max(I)$)

¹Full Width at Half Maximum

where I is the spectral image) at different heights (from 1.6 m to 5 m). The function `findChessboardCorners` attempts to determine whether the input image is a view of the chessboard pattern and locate the internal chessboard corners. The detected coordinates are roughly approximated, and to determine their positions accurately we use the function `cornerSubPix` as explained in the documentation ².

Linear model Using all the points detected for each spectral band, we calculate the centroid grid (each point average). The affine transform from each spectral band to this centroid grid is estimated. Theoretically, the rotation and the scale (A, B, C, D) do not depend on the distance to the ground, but only on the translation (X, Y) depends on the height. Thus a Levenberg-Marquardt curve fitting algorithm with linear least squares regression [14] can be used to fit an equation for each spectral band against X and Y independently to the centroid grid. We adjust the following curve $t = \alpha h^3 + \beta h^2 + \theta h + \gamma$ where h is the height, t is the resulted translation and factors $\alpha, \beta, \theta, \gamma$ are the model parameters.

Correction Based on the model estimated on the chessboard dataset, we transpose them to the agronomic dataset. To make the affine matrix correction, we used the rotation and scale factors at the most accurate height (1.6 m where the spatial resolution of the chessboard is higher), because it does not theoretically depend on the height. For the translation part, the curve model is applied for each spectral bands at the given height provided by the user. Each spectral band are warped using the corresponding affine transformation. Finally, all spectral bands are cropped to the minimal bounding box (minimal and maximal translation of each affine matrix). This first correction is an approximation, it enable to get some spatial properties that we will use on the second stage.

2.2.2 Perspective correction

Each spectral band has different properties and values by nature but we can extract the corresponding similarity by transforming each spectral band into its absolute derivative, to find similarities in gradient break among them. As we can

²https://docs.opencv.org/2.4/modules/calib3d/doc/camera_calibration_and_3d_reconstruction.html

see in the figure 3 gradients can have opposite direction depending on the spectral bands, making the absolute derivative an important step for matching between different spectral band.

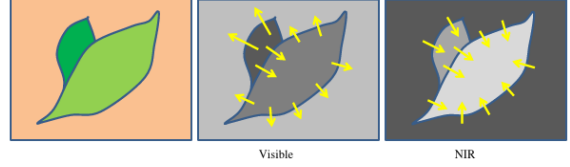


Figure 3: Gradient orientation in spectral band [15], orientation of the gradient is not the same depending to the spectral band,

The affine correction attempts to help the feature matching by adding properties of epipolar lines (close). Thus, the matching of extracted features can be spatially bounded, (i) we know that the maximum translation is limited to a distance of a few pixels (less than 10px), and (ii) the angle between the initial element and the matched one is limited to $[-1, 1]$ degree.

Computing the gradient To compute the gradient of the image with a minimal impact of the light distribution (shadow, reflectance, specular, ...), each spectral band is normalized using Gaussian blur [16], the kernel size is defined by $\text{next_odd}(\text{image_width}^{0.4})$ (19 in our case) and the final normalized images are defined by $I/(G+1)*255$ where I is the spectral band and G is the Gaussian blur of those spectral bands. This first step minimizes the impact of the noise on the gradient and smooths the signal in case of high reflectance. Using this normalized image, the gradient $I_{grad}(x, y)$ is computed with the sum of absolute Sharr filter [17] for horizontal S_x and vertical S_y derivative, noted $I_{grad}(x, y) = \frac{1}{2}|S_x| + \frac{1}{2}|S_y|$. Finally, all gradients $I_{grad}(x, y)$ are normalized using CLAHE [18] to locally improve their intensity and increase the number of key points detected.

Key-points Extractor A key point is a point of interest. It defines what is important and distinctive in an image. Different types of key point extractors are available and the following are tested :

(ORB) Oriented FAST and Rotated BRIEF [19], (AKAZE) Fast explicit diffusion for accelerated features in nonlinear scale spaces [20], (KAZE) A novel multi-scale 2D feature detection and description algorithm in nonlinear scale spaces [21], (BRISK) Binary robust invariant

scalable key-points [22], (AGAST) Adaptive and generic corner detection based on the accelerated segment test [23], (MSER) maximally stable extremal regions [24], (SURF) Speed-Up Robust Features [25], (FAST) FAST Algorithm for Corner Detection [26] and (GFTT) Good Features to Track [27].

These algorithms are largely summarized across multiple studies [9, 28, 29, 30], they are all available and easily usable in OpenCV, thus we have studied them by varying the most influential parameters for each of them with three modalities, the table 1 in appendix shows all modalities and methods. All the results can be found in “figures/*”.

Key-point detection We use one of the key-point extractors mentioned above between each spectral band gradients (all extractors are evaluated). For each detected key-point, we extract a descriptor using ORB features. We match all detected key-points to a reference spectral band (all bands are evaluated). All matches are filtered by distance, position and angle, to eliminate a majority of false positives along the epipolar line. Finally we use the function findHomography between the key points detected/filtered with RANSAC [31], to determine the best subset of matches to calculate the perspective correction.

Correction The perspective correction between each spectral band to the reference is estimated and applied. Finally, all spectral bands are cropped to the minimum bounding box, the minimum and maximum points are obtained by applying a perspective transformation to each corner of the image.

3 RESULTS AND DISCUSSION

Firstly the results will focus on affine corrections and then on the effects of the perspective correction. Figure 4 shows a closeup inside at 1.6 m (4a) unregistered image, (4c & 4d) registered image of each correction steps and (4b) the manufacturer results.

3.1 Affine correction

The affine correction model is based on the calibration dataset (where the chessboard are acquired). The 6 coefficients (A, B, C, D, X, Y) of the

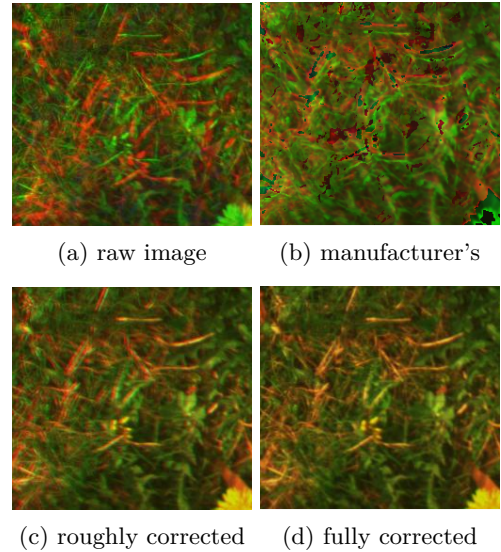


Figure 4: Example of each correction and the manufacturer's results

affine matrix were studied according to the height of the camera in order to see their stability. It appears that the translation part (X, Y), depends on the distance of the field (appendix figure 5) according to the initial assumption. On this part the linear model is used to estimate the affine correction from an approximated height.

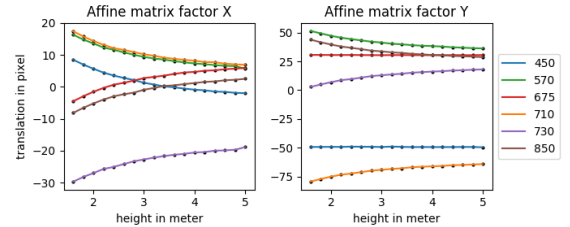


Figure 5: Affine matrix value by height

Rotation and scale do not depend on the ground distance (figure 6) according to the theory. These factors (A, B, C, D) are quite stable and close to identity, as expected (accuracy depends on the spatial resolution of the board). As result, single calibration can be used for this part of the matrix, and the most accurate are used (i.e where the chessboard has the best spatial resolution).

After the Affine correction, the remaining residual distances has been extracted, it is computed using the detected, filtered and matched key-point to the reference spectral band, the figure 7 show an example using 570 nm as reference before the perspective correction. The remaining distance between each spectral band to the reference varies according to the distance between the

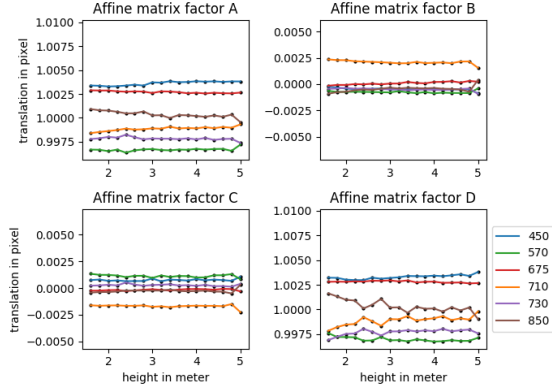


Figure 6: Affine matrix value by height

real height and the nearest selected (through linear model). Remember that a bias of $\pm 10\text{cm}$ was initially set to show the error in the worst case, so the difference of errors between each of them are due to the difference of sensors position in the array to the reference and the provided approximate height.

3.2 Perspective correction

The figures 8 shows the numbers of key-points after filtering and homographic association (minimum of all matches) as well as the computation time and performance ratio (matches/time) for each method. The performance ratio is used to compare methods between them, bigger he is, greater is the method (balanced between times and accuracy), making lower of them unsuitable.

All these methods offer interesting results, the choice of method depends on application needs between computation time and accuracy, three methods stand out in all of there modality:

- GFTT shows the best overall performance in both computation time and number of matches
- FAST and AGAST1 are quite suitable too, balanced in time and with greater matches performances.

The other ones did not show improvement in term of time or matches (especially compared to GFTT), some of them show a small number of matches which can be too small to ensure the precision. Increasing the number of key points matched allows a slightly higher accuracy [9]. For example, switching from SURF (30 results) to FAST (130 results) reduces the final residual distances from ≈ 1.2 to ≈ 0.9 pixel but increases

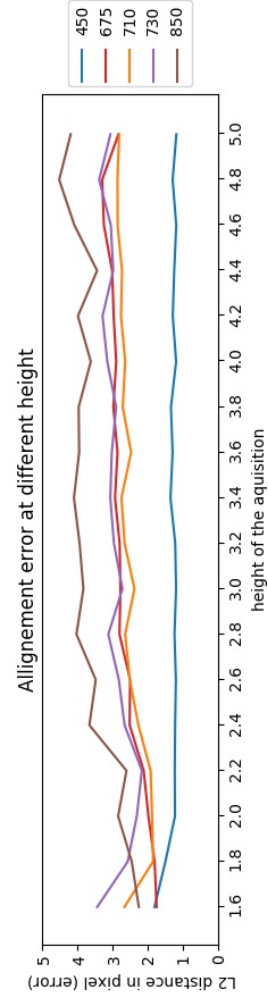


Figure 7: The mean distance of detected key-point before perspective correction with 570 nm as spectral reference

the calculation time from ≈ 5 to ≈ 8 seconds.

All methods show that the best spectral band is 710 nm (in red), with the exception for SURF and GFTT which is 570 nm. The figure 9 shows the minimum number of matches between each reference spectrum and all the others, for each relevant methods and modalities (KAZE, AGAST, FAST GFTT). Choosing the right spectral reference is important, as we can see, no correspondence is found in some cases between 675-850 nm, but correspondences are found between 675-710 nm and 710-850 nm, making the 710 nm more appropriate, the same behavior can be observed for the other bands and 570 nm. This is visible on the figure for all methods, 570 nm and 710 nm have the best minimum number of matches where

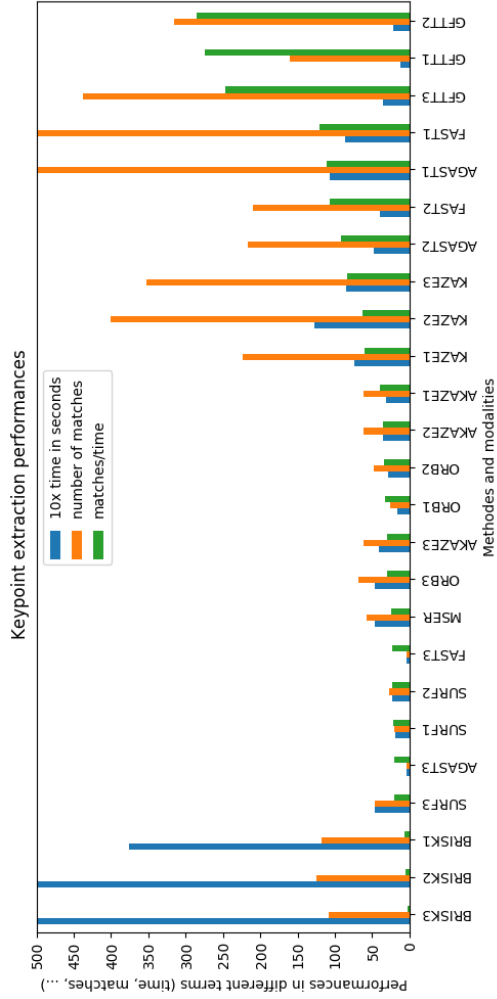


Figure 8: features extractor performances after filtering and homography association

all the other are quite small.

Residues of the perspective correction show that we have correctly registered each spectral band, the figure 10 shows the residual distance at different ground distances. In comparisons the affine correction error are between $[1.0 - 4.8]$ pixel where with the combination of affine and perspective the residual error are between $[0.7 - 1.0]$ pixel. On average the perspective correction enhance the residual error by $(3.5 - 0.9)/3.5 \approx 74\%$.

3.3 General discussion

Even if the relief of the scene is not taken into account due to the used deformation model, in our case, with flat ground, no difference arise. However, more complex deformation models

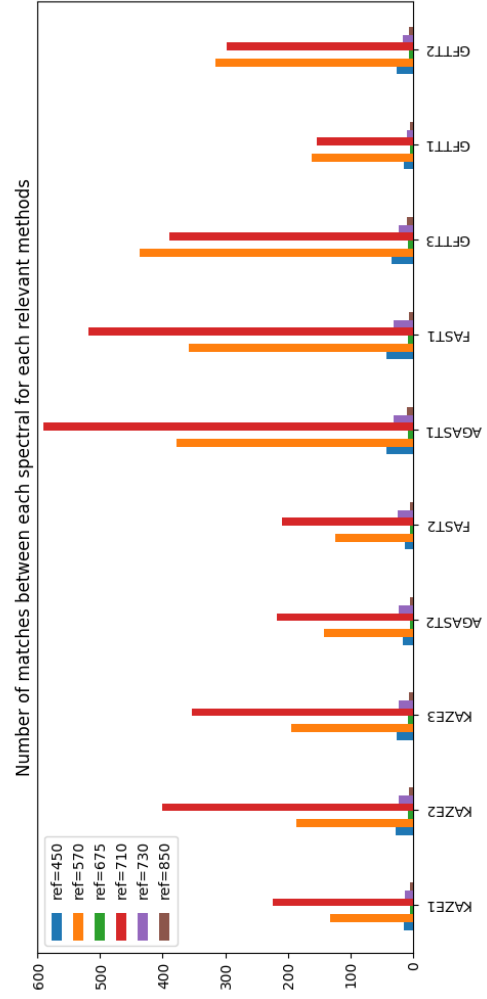


Figure 9: key-point extractor performances

[32, 33] can be used to improve the remaining error. But can also, in some cases, create large angular deformations caused by the proximity of key-points, of course, it's possible to filter these key-points, which also reduces the overall accuracy.

Further researches can be performed on each parameter of the feature extractors, for those who need specific performance (time/precision), We invite anyone to download the dataset and test various combinations. Otherwise feature matching can be optimized, at this stage, we use brute-force matching with post filtering, but a different implementation that fulfill our spatial properties should greatly enhance the number of matches by reducing false positives.

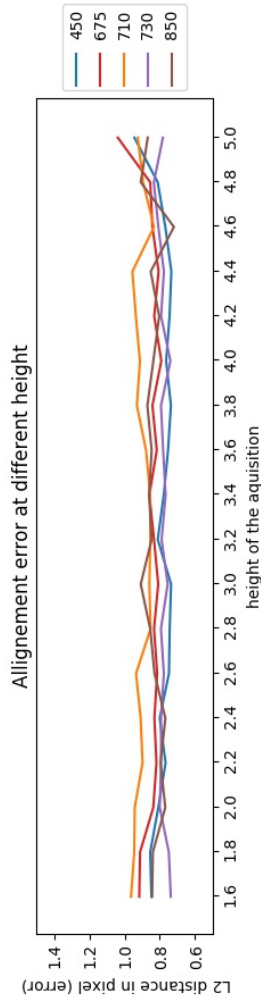


Figure 10: Perspective re-projection error with GFTT using the first modality and 570 nm as reference

4 CONCLUSION

In this work, the application of different techniques for multi-spectral image registration was explored using the Airphen camera. We have tested nine type of key-points extractor (ORB, GFTT, AGAST, FAST, AKAZE, KAZE, BRISK, SURF, MSER) at different heights and the number of control points obtained. As seen in the method, the most suitable method is the GFTT (regardless of modalities 1, 2 or 3) with a significant number of matches 150 – 450 and a reasonable calculation time 1.17 s to 3.55 s depending on the modality.

Furthermore, the best spectral reference was defined for each method, for example 570 nm for

GFTT. We have observed a residual error of less than 1 px, supposedly caused by the difference of sensors nature (spectral range, lens).

ACKNOWLEDGMENTS

The authors acknowledge support from European Union through the project H2020 IWMPRAISE³ (Integrated Weed Management: PRActical Implementation and Solutions for Europe) and from ANR Challenge ROSE through the project ROSEAU⁴ (RObotics SEnsorimotor loops to weed AUtonomously).

We would like to thanks Jones Gawain, Combaluzier Quentin, Michon Nicolas, Savi Romain and Masson Jean-Benoit for the realization of the metallic gantry that help us positioning the camera at different heights.

SUPPLEMENTARY MATERIAL

The additional data and source code associated with this article can be found in the online version at the following address gitlab.com/phd-thesis-adventice/phd-airphen-alignment the access is limited, and we invite you to send an email to the author for full access.

REFERENCES

- [1] Martin Lechenet, Vincent Bretagnolle, Christian Bockstaller, François Boissinot, Marie-Sophie Petit, Sandrine Petit, and Nicolas M. Munier-Jolain. Reconciling pesticide reduction with economic and environmental sustainability in arable farming. *PLOS ONE*, 9(6):1–10, 06 2014.
- [2] Sindhuja Sankaran, Lav R. Khot, Carlos Zúñiga Espinoza, Sanaz Jarolmasjed, Vidyasagar R. Sathuvalli, George J. Vandemark, Phillip N. Miklas, Arron H. Carter, Michael O. Pumphrey, N. Richard Knowles, and Mark J. Pavék. Low-altitude, high-resolution aerial imaging systems for row and field crop phenotyping: A review. *European Journal of Agronomy*, 70:112 – 123, 2015.

³<https://iwmpraise.eu/>

⁴<http://challenge-rose.fr/en/projet/roseau-2/>

- [3] A. Bannari, D. Morin, F. Bonn, and A. R. Huete. A review of vegetation indices. *Remote Sensing Reviews*, 13(1-2):95–120, 1995.
- [4] I Filella, L Serrano, JI Serra, and J Penuelas. Evaluating wheat nitrogen status with canopy reflectance indices and discriminant analysis. *Crop Science*, 35(5):1400–1405, 1995.
- [5] Xiu-liang Jin, Wan-ying Diao, Chun-hua Xiao, Fang-yong Wang, Bing Chen, Ke-ru Wang, and Shao-kun Li. Estimation of wheat agronomic parameters using new spectral indices. *PLOS ONE*, 8(8):1–9, 08 2013.
- [6] Emna Kamoun. Image registration: From sift to deep learning. july 2019.
- [7] Clément Douarre, Carlos F Crispim-Junior, Anthony Gelibert, Laure Tougne, and David Rousseau. A strategy for multimodal canopy images registration. In *7th International Workshop on Image Analysis Methods in the Plant Sciences*, Lyon, France, July 2019.
- [8] Barbara Zitová and Jan Flusser. Image registration methods: A survey. *Image and Vision Computing*, 21:977–1000, 10 2003.
- [9] Jocival Dantas Dias Junior, Andre Backes, and Maurício Escarpinati. Detection of control points for uav-multispectral sensed data registration through the combining of feature descriptors. pages 444–451, 01 2019.
- [10] Maria Vakalopoulou and Konstantinos Karantzalos. Automatic descriptor-based co-registration of frame hyperspectral data. *Remote Sensing*, 6, 03 2014.
- [11] S. Chen, H. Shen, C. Li, and J. H. Xin. Normalized total gradient: A new measure for multispectral image registration. *IEEE Transactions on Image Processing*, 27(3):1297–1310, March 2018.
- [12] Jean-Baptiste Vioix. Conception et réalisation d’un dispositif d’imagerie multispectrale embarqué : du capteur aux traitements pour la détection d’adventices. 2004.
- [13] Jean-Yves Bouguet. Camera calibration toolbox for matlab. 2001.
- [14] Jorge J. Moré. The levenberg-marquardt algorithm: Implementation and theory. In G.A. Watson, editor, *Numerical Analysis*, volume 630 of *Lecture Notes in Mathematics*, pages 105–116. Springer Berlin Heidelberg, 1978.
- [15] G. Rabatel and S. Labbe. Registration of visible and near infrared unmanned aerial-vehicle images based on Fourier-Mellin transform. *Precision Agriculture*, 17(5):564–587, 2016.
- [16] D. Sage and M. Unser. Teaching image-processing programming in java. *IEEE Signal Processing Magazine*, 20(6):43–52, November 2003. Using “Student-Friendly” ImageJ as a Pedagogical Tool.
- [17] Hanna Seitz. Contributions to the minimum linear arrangement problem, 2010.
- [18] Karel Zuiderveld. Contrast limited adaptive histogram equalization. In *Graphics gems IV*, pages 474–485. Academic Press Professional, Inc., 1994.
- [19] Ethan Rublee, Vincent Rabaud, Kurt Konolige, and Gary Bradski. Orb: An efficient alternative to sift or surf. In *Proceedings of the 2011 International Conference on Computer Vision, ICCV ’11*, pages 2564–2571, Washington, DC, USA, 2011. IEEE Computer Society.
- [20] Pablo F Alcantarilla and T Solutions. Fast explicit diffusion for accelerated features in nonlinear scale spaces. *IEEE Trans. Patt. Anal. Mach. Intell*, 34(7):1281–1298, 2011.
- [21] Alvaro Ordonez, Francisco Arguello, and Dora B. Heras. Alignment of hyperspectral images using kaze features. *Remote Sensing*, 10(5), 2018.
- [22] Stefan Leutenegger, Margarita Chli, and Roland Siegwart. Brisk: Binary robust invariant scalable keypoints. In *2011 IEEE international conference on computer vision (ICCV)*, pages 2548–2555. Ieee, 2011.
- [23] Elmar Mair, Gregory D Hager, Darius Burschka, Michael Suppa, and Gerhard Hirzinger. Adaptive and generic corner detection based on the accelerated segment test. In *European conference on Computer vision*, pages 183–196. Springer, 2010.
- [24] Michael Donoser and Horst Bischof. Efficient maximally stable extremal region (mscr) tracking. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’06)*, volume 1, pages 553–560. Ieee, 2006.
- [25] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. Surf: Speeded up robust features. In *European conference on computer vision*, pages 404–417. Springer, 2006.

- [26] Miroslav Trajković and Mark Hedley. Fast corner detection. *Image and vision computing*, 16(2):75–87, 1998.
- [27] Jianbo Shi et al. Good features to track. In 1994 Proceedings of IEEE conference on computer vision and pattern recognition, pages 593–600. IEEE, 1994.
- [28] Shaharyar Ahmed Khan Tareen and Z. Saleem. A comparative analysis of sift, surf, kaze, akaze, orb, and brisk. 2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET), pages 1–10, 2018.
- [29] Hongmou Zhang, Jürgen Wohlfeil, and Denis Griebbach. Extension and evaluation of the agast feature detector. 2016.
- [30] Farman Ali, Sajid Ullah Khan, Muhammad Zarrar Mahmudi, and Rahmat Ullah. A comparison of fast, surf, eigen, harris, and mser features. *International Journal of Computer Engineering and Information Technology*, 8(6):100, 2016.
- [31] Martin A. Fischler and Robert C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24(6):381–395, June 1981.
- [32] Herve Lombaert, Leo Grady, Xavier Pennec, Nicholas Ayache, and Farida Cheriet. Spectral demons – image registration via global spectral correspondence. In Andrew Fitzgibbon, Svetlana Lazebnik, Pietro Perona, Yoichi Sato, and Cordelia Schmid, editors, *Computer Vision – ECCV 2012*, pages 30–44, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.
- [33] F. L. Bookstein. Principal warps: thin-plate splines and the decomposition of deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(6):567–585, June 1989.

APPENDIX

ABRV	modality 1	modality 2	modality 3
ORB	nfeatures=5000	nfeatures=10000	nfeatures=15000
GFTT	maxCorners=5000	maxCorners=10000	maxCorners=15000
AGAST	threshold=71	threshold=92	threshold=163
FAST	threshold=71	threshold=92	threshold=163
AKAZE	nOctaves=1, nOctaveLayers=1	nOctaves=2, nOctaveLayers=1	nOctaves=2, nOctaveLayers=2
KAZE	nOctaves=4, nOctaveLayers=2	nOctaves=4, nOctaveLayers=4	nOctaves=2, nOctaveLayers=4
BRISK	nOctaves=0, patternScale=.1	nOctaves=1, patternScale=.1	nOctaves=2, patternScale=.1
SURF	nOctaves=1, nOctaveLayers=1	nOctaves=2, nOctaveLayers=1	nOctaves=2, nOctaveLayers=2
MSER	None	None	None

Table 1: list of algorithms with 3 modalities of their parameters