Vine disease detection in UAV multispectral images using optimized image registration and deep learning segmentation approach

Mohamed Kerkech, Adel Hafiane, Raphaël Canals

Vine disease detection in UAV multispectral images using optimized image registration and deep learning segmentation approach

Mohamed Kerkech *, Adel Hafiane *, Raphael Canals **

* *INSA-CVL, Univ. Orléans, PRISME, EA 4229, F18022, Bourges, France*

** *Univ. Orléans, INSA-CVL, PRISME, EA 4229, F45072, Orléans, France*

Abstract

One of the major goals of tomorrow's agriculture is to increase agricultural productivity but above all the quality of production while significantly reducing the use of inputs. Meeting this goal is a real scientific and technological challenge. Smart farming is among the promising approaches that can lead to interesting solutions for vineyard management and reduce the environmental impact. Automatic vine disease detection can increase efficiency and flexibility in managing vineyard crops, while reducing the chemical inputs. This is needed today more than ever, as the use of pesticides is coming under increasing scrutiny and control. The goal is to map diseased areas in the vineyard for fast and precise treatment, thus guaranteeing the maintenance of a healthy state of the vine which is very important for yield management. To tackle this problem, a method is proposed here for Mildew disease detection in vine field using a deep learning segmentation approach on Unmanned Aerial Vehicle (UAV) images. The method is based on the combination of the visible and infrared images obtained from two different sensors. A new image registration method was developed to align visible and infrared images, enabling fusion of the information from the two sensors. A fully convolutional neural network approach uses this information to classify each pixel according to different instances, namely, shadow, ground, healthy and symptom. The proposed method achieved more than 92% of detection at grapevine-level and 87% at leaf level, showing promising perspectives

for computer aided disease detection in vineyards.

Keywords: Unmanned aerial vehicle (UAV); image registration; convolutional neural network; precision agriculture; disease mapping.

1. Introduction

Several studies have been carried out on the overuse of crop pesticides and their negative effects on human health [1, 2, 3]. Like other crops, vines are very vulnerable to viruses, bacteria and fungi. This vulnerability favours their contamination by several types of disease that are harmful and destructive [4], such as *Esca* [5], *Flavescence dore* [6] and Mildew [7]. Vine contamination generally reduces productivity [8], which implies economic losses for the winegrower. To deal with this situation, winegrowers have to frequently check the state of the vine leaves. However, this traditional procedure is laborious and expensive, since it involves several experts for many days [9]. To reduce economic loss and the environmental impact, remote sensing methods are a promising approach for effective vineyard monitoring.

Remote sensing of agricultural crops [10] has evolved considerably over the past decade. Applications such as calculating fertilizer rates [11], monitoring biomass production [12], weed detection [13], detecting defective crops [14] or disease detection [15, 16, 17] have been proposed. These applications are constantly progressing as technology advances, especially with the evolution of Unmanned Aerial Vehicles (UAV) which have opened up further research opportunities thanks to their low manufacturing costs.

UAVs are increasingly used in many fields, such as urban remote sensing, but also in a wide range of agricultural applications [18]. Previous studies have shown the importance of both the visible and the infrared spectrum for disease detection [19]. Combining these two imaging modalities would therefore ensure better detection. In UAV imaging systems, usually two separate sensors are used, one for each modality. However, an acquisition by two sensors generates a spatial shift between the visible and infrared image which makes it difficult

to process the information from the two sensors simultaneously. Therefore, multimodal alignment or registration [20, 21, 22, 23] is required to merge the information from the two sensors with deep learning.

Deep learning techniques have enabled great progress in the computer vision field thanks to the convolutional neural networks (CNNs) approach [24, 25, 26, 27, 28, 29]. As in several fields of application, these technologies are increasingly used in the remote sensing domain for agriculture [30, 31, 32, 33, 34, 35, 36, 37, 38]. Most research on agricultural applications uses CNNs with the sliding window technique, which generally leads to fuzzy boundaries of the image regions. On the other hand, crop disease detection can be seen as an image segmentation problem. Therefore, one can benefit from the deep learning segmentation approach to detect crop disease with a better boundary precision compared to the sliding window technique. Several segmentation architectures have been developed, such as SegNet [39], DeconvNet [40] and U-Net [41]. SegNet is the most popular architecture for semantic segmentation [42, 43, 44, 45]. It has shown a very good performance in solving problems related to semantic segmentation for several applications [46, 47, 48]. So far, very little attention has been paid to the role of deep learning segmentation for vine disease detection.

This paper presents a new methodology for Mildew disease detection in vineyards from aerial images, using multispectral information. The aim was to develop algorithms and methods in order to investigate the possibility of detecting vine diseases, using the potential of a deep learning segmentation architecture. The problem was addressed by the semantic segmentation approach in order to identify classes such as shadow, ground, healthy and symptomatic vines. The method consists in two main steps. The first one deals with the problem of multispectral image registration, and proposes a new method to effectively align images from the visible and infrared spectra. The second one uses the SegNet architecture to delineate semantic areas in each image type separately, then a fusion procedure was applied to the segmentation outputs. Data were collected under real conditions on two vineyard plots. Several experimental schemes were set up to show the contribution of different elements of

the proposed method. The study provides important insight into the potential of recent machine learning approaches for disease mapping using UAV remote sensing technology.

The main contributions of this work are: optimization of the image registration algorithm, introduction of a semi-automatic method for pixel-wise image labelling and SegNet for vine disease detection in the visible and infrared ranges.

The article is organized as follows: related work is presented in section 2, the study areas and materials are described in section 3, the proposed methods are detailed in section 4, the experiments and results are presented in section 5, the proposed system is discussed in section 6 and section 7 concludes the paper.

2. Related work

This section summarizes the main studies carried out on image registration, and disease detection in vineyards, plants or crops.

2.1. Image registration

In the literature, the problem of image registration dates back to the 1980s. Since then, several methods have been implemented. The work accomplished in various fields has been surveyed in: medicine [49], computer vision [50], remote sensing [51] and various applications [52]. In all areas studied, it is concluded that image registration is based on two main methods: the area-based method, and the feature-based method.

The area-based method: This method is not widely used in the remote sensing field, because most of the algorithms are highly sensitive to several uncontrollable parameters, such as variations in brightness, image noise, etc. The method is therefore generally applicable only to non-rigid problems. However, Wang et al. [53] implemented an algorithm (An automatic cross-correlation (ACC)) insensitive to the lighting conditions and applied it to multimodal images (visible and infrared). The authors concluded that this algorithm performs better than

other area-based algorithms and is suitable for multimodal images. Another registration method applied to precision agriculture was proposed by Erives et al. [54]. This method that processes multispectral images is based on the phase correlation algorithm (PC). The results obtained indicate that this algorithm is robust to modality change, brightness difference, noise, rotation and translation. Zhuang et al. [55] performed a multimodal registration based on mutual information with a combination of Particle Swarm Optimization (PSO) and Powell search algorithms. The proposed method was found to be faster in terms of runtime and more accurate in terms of results compared to traditional methods.

The feature-based method: Lakshmi et al. [56] and Javadi et al. [57] worked on natural terrain and city video frames, acquired by a UAV for the creation of an orthorectified image. In these studies, the standard registration method based on the SURF algorithm was used. In [56], this method was compared with the Cross-Correlation algorithm (Area-based method) which failed to register the aerial images, whereas the feature-based method using the variants of the SIFT algorithm outperformed the other methods in terms of results and in terms of runtime. Tsai et al. [22] conducted a similar study to compare the SIFT algorithm with ABRISK, and concluded that ABRISK was up to 312 times faster than SIFT, and had a lower mean error. In another study, Onyango et al. [58] used the AKAZE algorithm to match oblique building images with images of cities taken by a UAV. The study concluded that the AKAZE algorithm outperformed other algorithms of the same type. More recently, image matching algorithms based on deep learning have appeared [59, 21]. The deep learning architecture is used as a feature extractor to create a correspondence between the two images. Wang et al. [21] worked on remote sensing images using a supervised architecture, while Yang et al. [59] used an unsupervised architecture to recalibrate multi-temporal images. The latter showed better accuracy than the SIFT type algorithms, but these results only correspond to multi-temporal images, of the same modality and on low resolution images.

2.2. Disease detection

A comprehensive review of the literature was conducted by Mahlein [60]. The survey lists several studies on disease detection by multispectral imaging. Among others, Oerke et al. [61] showed by using a thermal camera that the leaves of plants in the early stages of a disease have a more heterogeneous temperature than when the disease progresses over time. The potential of multispectral information in the early detection of plant disease is attracting more and more interest in the remote sensing field. Two studies on the hyperspectral reflectance of vine leaves diseased by the *Esca* complex have been carried out. The first one [62] was performed at the leaf level and the second one [63] at the vineyard level (UAV images). Both showed a difference between the reflectance of a healthy and a diseased leaf.

In a first study, Albetis et al. [64] investigated the detection of *Flavescence dore* in UAV images. The study was carried out on plots of white and red cultivars. The results obtained indicate the feasibility of disease detection using aerial images. In a second study, Albetis et al. [65] examined the potential of multispectral imaging by UAV for the detection of symptomatic and asymptomatic vines. In addition to the first study, a larger dataset was acquired and used to test 24 variables calculated from this new dataset. The best results were obtained by the red-green index (RGI) and the red-green vegetation index (GRVI).

In Al-Saddik et al. [66], a first study on hyperspectral images at the vine leaf scale was carried out. The aim of the study was to develop spectral disease indices capable of detecting and identifying *Flavescence dore* disease in vines, and achieved 90% classification accuracy. A second study by Al-Saddik et al. [67] on disease identification at the vine leaf level was carried out to differentiate between yellowing leaves, and leaves infected with *Esca* disease through a neural network classifier. The best results were obtained when textural and spectral data were combined. A third study by the same authors [68] consisted in defining the most significant spectral channels for *Flavescence dore* disease detection.

Ranon et al. [69] carried out a similar study on *Esca* disease detection in

vineyards. The imaging system was mounted on a small vehicle that passed between the vinerows for image acquisition. Two methods for detecting *Esca* disease were used: Scale Invariant Feature Transform (SIFT) encoding, and the deep learning MobileNet network. The authors concluded that the deep learning method was better than the SIFT encoded method.

In our previous study [38], a new method for detecting *Esca* disease in UAV RGB images was proposed. The method used LeNet5 CNN architecture and good results were obtained achieving 95% disease detection accuracy in the visible range.

To the best of our knowledge, no research has been conducted so far on the combination of visible and infrared UAV images for vine disease detection by a deep learning segmentation approach.

3. Study areas and materials

3.1. Study areas

This study was carried out on two parcels of vines located in the Center Val de Loire region in France. The first plot (P1) can be seen in Figure 1a and the second one (P2) in Figure 1b. P1 and P2 are at an altitude of 110 and 114 meters respectively, with a surface of 1.8 hectares for P1 and 1.5 hectares for P2, and are positioned on a silty sand soil for P1 and sandy loam soil for P2. The ground of P1 is slightly inclined (7% of slope) (Figure 2) and flat for P2. The Centre Val de Loire region is characterized by a moderate temperature variation between 1 to $26°C$, and the average annual rainfall reaches 700 mm. Table 1 gives more details about the plots.

In order to carry out this study and obtain healthy and diseased samples, part of the P1 plot was treated with phytosanitary products (to protect the vine against diseases), and the other part remained untreated in order to allow the development of the disease. During this time, Mildew disease spread to all

untreated areas of the plot. The P2 plot was treated globally against disease and remained totally healthy. It was used in this study for qualitative validation.

Table 1: Description of the study vineyard.

Types of information	Description (P1)	Description (P2)
Surface	1.8 hectares	1.5 hectares
Altitude	110 meter	114 meter
Planting year	1976	1991
Variety	Malbec	Sauvignon
Vine plant spacing	1 meter	1 meter
Interline distance	1.5 meter	1.5 meter
Slope	7%	0%
Exposure	North-South	Northwest-Southeast
Coarse fraction	25%	17%
Soil Organic Carbon	4.46%	3.94%
Soil type	Silty-Sand	Sandy-Loam

3.2. Materials

The UAV used in the data acquisition process was a Quadcopter drone (Figure 3a) manufactured by Scanopy. This drone embeds two camera sensors MAPIR Survey2 (Figure 3b). The first is a visible light sensor (RGB) set to automatic lighting, and the second is an infrared light sensor (Near Infrared (NIR), Red and Normalized Difference Vegetation Index (NDVI)). For the latter, the near infrared wavelength is 850 nm. Both image sensors have a high-resolution of 16 megapixels (a size of 4608 $\times$ 3456).

Data acquisition was carried out by the UAV, which flew over the plot at an altitude of 25 meters and with an average speed of 10 km/h. At this altitude, the ground resolution is $1cm^2/pixel$. An image was taken every 2 seconds without stopping the UAV. Each image taken overlaps the previous image by

Figure 1: The study vineyard seen by satellite. Plot (P1) in picture (a) and plot (P2) in picture (b).

Figure 2: The plot (P1) seen from the ground.

over 70%. The drone has an average energy autonomy of 20 minutes. The climatic conditions of the acquisition were moderate, i.e. low winds and optimal lighting (acquisition times were between 11:30 and 13:30 to avoid the shadow

of the vinerow). The acquisition was carried out in summer 2018.

Figure 3: The acquisition equipment used in this study. The quadcopter UAV drone (a) and the high-resolution Survey2 sensor (b).

4. Methods

Using the multispectral and standard RGB images, automatic processing and analysis methods were developed to extract relevant information and correlate it with the ground truth results. Deep learning segmentation was applied on the two types of images to automatically delineate different regions (healthy, symptomatic, etc.). This generated a disease map of the vineyard comprising different segmented regions, which was used to monitor the vineyards.

The method comprised three main steps (see Figure 4). The first one consisted in image registration between the images acquired in the visible and infrared range. This step is essential for the third step, as it enables the pixel-wise superposition of the two images and thus allows segmentation fusion. Once the registration of the two images had been performed, the next stage consisted in segmenting the plot in the visible and infrared range using the SegNet architecture. The two segmented images were merged in the third step, to generate a disease map.

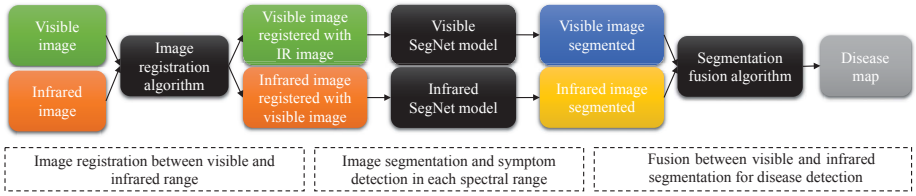

Figure 4: Overview of disease detection system in grapevine fields.

4.1. Image registration

The objective of the registration algorithm is to realign and geometrically correct the shift [70] between the visible and infrared images at the pixel-wise level. Generally, the UAV image is partially distorted due to the UAV vibrations, read in rolling shutter mode of the sensors and optics. Therefore, the rigid alignment model (translation and rotation) is not appropriate in our case. Hence, the proposed algorithm used the non-rigid model. However, even if the registration model is correct it is not sufficient to align the two types of images perfectly. Moreover, matching image points between the visible and infrared bands is difficult, since the key points do not necessarily have the same visual properties in the two spectral bands. In order to improve the accuracy of the image alignment, an iterative process based on minimization of the registration error was implemented.

The registration algorithm proposed in this study is based on the Accelerated-KAZE (AKAZE) algorithm [71]. AKAZE is an algorithm used in computer vision for detecting objects or similarities in two images. Its principle is comparable to that of the Scale Invariant Feature Transform (SIFT) [72], Speeded Up Robust Features (SURF) [73], Features from Accelerated Segment Test (FAST), Binary Robust Independent Elementary Features (BRIEF), Oriented FAST and Rotated BRIEF (ORB) [74], and KAZE [75] algorithms. However, AKAZE is much more efficient in detection robustness, description and in the speed of calculation as it was created with high-performance algorithms in a pyramidal framework comparable to other algorithms of the same type. Even if AKAZE

follows the same pyramidal steps as the other algorithms, the method used is very different. AKAZE integrates Fast Explicit Diffusion (FED) [76] systems for accelerated feature detection in a non-linear scale space. Moreover, a modified version of Local Difference Binary (M-LDB) [71] has been integrated into AKAZE. Unlike the old version, this modified version of LDB [77] is a rotationally invariant, scaled descriptor and can exploit gradient information from the non-linear scale space.

The proposed registration method is shown in Figure 5. The first step is to extract the green channel (G) from the visible image, and the near infrared channel (NIR) from the infrared image (these channels were selected for their vegetation texture information). By using the normalization equation (1), the second phase normalizes the two spectral channels to improve their contrast.

$$I_{Normalized} = 255 \times \frac{I - min(I)}{max(I) - min(I)} \tag{1}$$

where $I_{Normalized}$ is a normalized source image I (visible or infrared), $min(I)$ and $max(I)$ are respectively the minimum and maximum grey level value of the image I.

The third step is to extract the points of interest and calculate their features from the two channels (by the AKAZE algorithm). Based on the features of the interest points, the fourth stage is to map each point of interest extracted in the G channel to the corresponding point in the NIR channel. In the fifth step, to eliminate some outliers, a first thresholding algorithm performs a preselection of inlier matching points, then a final selection is performed by the RANdom SAmple Consensus (RANSAC) algorithm [78]. To obtain the best setting for algorithms that contribute to removing the outliers, a dynamic setting (by threshold variation) method based on homographic matrix analysis is integrated into the registration system. This dynamic method (dynamic threshold) very significantly reduces the number of registration failures. Note that X is the coordinates of the source infrared image (x, y), X' is the coordinates of the

registered infrared image (x', y') and H is the homographic matrix 2. In order to avoid the problems that lead to misregistration, a dynamic threshold regulation of the RANSAC algorithm was used, and the distance-based algorithm. For each given value, a homographic matrix is estimated. The viability of this matrix is then tested by equation 3. This test is performed by projecting the end coordinates of the source image into the new space of the registered image. If the result found is coherent (the distance between the old coordinates and the new coordinates does not exceed a certain threshold), this implies the end of the dynamic adjustment procedure, and the matrix tested is used to register the image. Otherwise, a new setting is made to repeat the same procedure. Once this step is finished, the pre-registered infrared image can be obtained by the homographic matrix validated by the dynamic adjustment procedure.

$$H = \begin{pmatrix} 1 + h_{00} & h_{01} & h_{02} \\ h_{10} & 1 + h_{11} & h_{12} \\ h_{20} & h_{21} & 1 \end{pmatrix} \tag{2}$$

$$X' = HX \tag{3}$$

where $(1 + h_{00})$ *and* $(1 + h_{11})$ *are stable scale factors respectively in the X and Y direction only.* h_{01} *and* h_{10} *are scale factors respectively in the X direction relative to the Y distance from the origin and Y direction relative to the X distance from the origin.* h_{02} *and* h_{12} *are respectively translation in the X and Y direction.*h_{20} *and* h_{21} *are relative scale factors X and Y respectively as a function of X and Y.*

In order to reduce the registration error, an iterative method was implemented. After the pre-registration has been completed, an iterative phase starts from the result obtained, calculates the error by the Root Mean Squared Error (RMSE) [79] between matched points which are calculated on the X coordinates with equation (4), and on the Y coordinates with equation (5), and then calculates their module by equation (6). If the error decreases, another iteration is

performed, if not, the iterative process stops and the best result, which is the result with the minimum RMSE, is kept.

$$RMSE_x = \sqrt{\frac{\sum_{i=0}^{N}(x_{VISi} - x_{IRi})^2}{N}} \tag{4}$$

$$RMSE_y = \sqrt{\frac{\sum_{i=0}^{N}(y_{VISi} - y_{IRi})^2}{N}} \tag{5}$$

$$RMSE = \sqrt{RMSE_x^2 + RMSE_y^2} \tag{6}$$

where (x_{VISi}, y_{VISi}) *and* (x_{IRi}, y_{IRi}) *are respectively the "i" th coordinates of matching points between the visible and infrared images.* N *is the number of matches found between the visible and infrared images.*

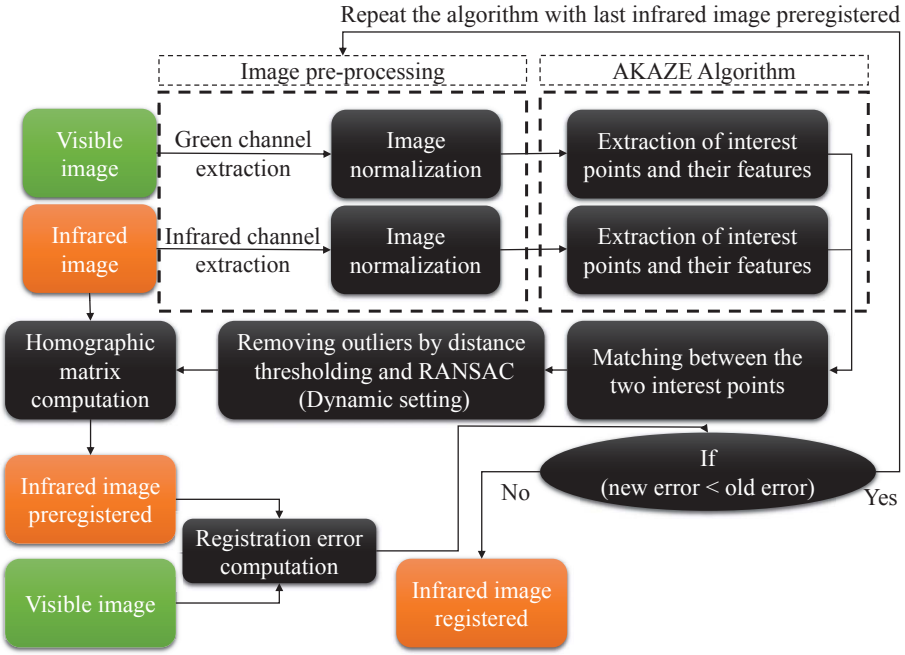

Figure 5: Proposed method for non-rigid registration of multimodal (visible and infrared) image.

4.2. Dataset

This subsection presents the description of different classes, the labelling method, and the data generation using data augmentation.

4.2.1. Classes description

The dataset was composed of visible and infrared range images. In our case, the visible sensor was used to detect the presence or absence of chlorophyll in the crop, i.e. to detect any anomalies in the vegetation related to its discoloration during flowering and fruit set period. The infrared sensor used operates in the range of 850 nm. This wavelength was chosen for its sensitivity to changes in the different states of vegetation. The high reflectance generated by vegetation in the near infrared spectrum also makes it relevant for plant analysis.

The visible and infrared datasets were labelled based on four dominant classes in the visible and infrared ranges, fusion by intersection, and fusion by union. The first class represents the shaded areas in the vine and on the ground (all the dark areas). This class does not reflect light, which means that no relevant information can be drawn from it. The second class represents the ground; it can be an area of weeds, grass or any type of soil. The third class represents the healthy vegetation of the vine. This class has a green color in the visible range and has a high reflectance level compared to the previous classes in the infrared range. Finally, the fourth class represents the symptomatic vine. This class is generally characterized by yellow or brown color of the foliage in the visible range (in the case of an advanced symptom). This discoloration results from a problem with chlorophyll production. In the infrared range, the symptomatic class is characterized by a variation in reflectance between a diseased leaf and its neighbors. Visually it appears as specific texture, with which this class can be distinguished from the others. After merging the segmented visible and infrared images, two classes were created a "symptomatic intersection" class when symptoms were detected in the same location of visible and infrared areas (AND operator), and a "symptomatic union" class when two symptoms were detected either in the visible or in the infrared spectra (OR operator).

4.2.2. Semi-automatic data labelling

Due to the large amount of data and the difficulty of achieving accurate labelling, the data labelling process was performed by a semi-automatic method (see Figure 6). First, a sort of weak learning procedure was used to roughly detect the potential diseased areas using a sliding window over the image. Each block was then classified by LeNet5 network for pre-labelling. Then, manual labelling was performed on the basis of the ground truth provided by technicians in the field.

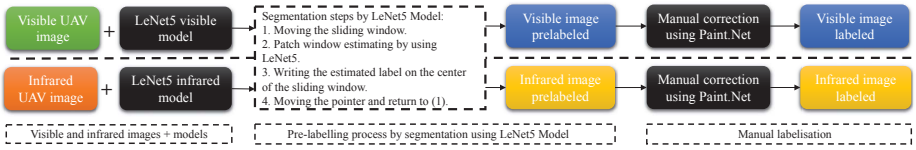

Figure 6: Overview of semi-automatic pixel-wise labelling for visible and infrared images.

To achieve the automatic labelling step, two LeNet5 datasets were created for visible and infrared images (from 480 samples of each range) using the translation and rotation data augmentation method (see Figure 7). Each dataset was organized into 4 classes (shadow, ground, healthy and symptomatic class), and contained $70,560$ patches of size 32×32 pixels ($17,640$ samples for each class, among these samples, there were $14,994$ samples for training and $2,646$ samples for validation). Visible and infrared LeNet5 models were generated from these two datasets (see Figure 8).

Once LeNet5 models had been generated, the sliding window method was used to perform automatic labelling on UAV images with a size of 4608×3456 pixels. To obtain the best possible accuracy, the sliding window was set with a displacement step of 2×2. The automatic labelling operation required more than an hour and a half of runtime to label only a single UAV image.

Finally, manual correction was performed by the free software "Paint.Net", which consists in correcting possible misclassification generated by automatic labelling. Technicians observed vines at the ground and reported all the dis-

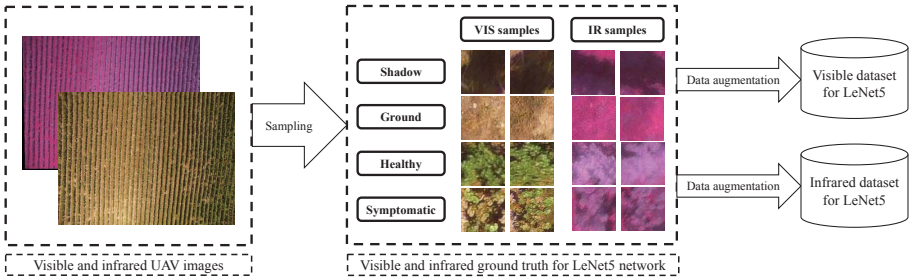

Figure 7: Making datasets for LeNet5 network. VIS and IR mean visible and infrared respectively.

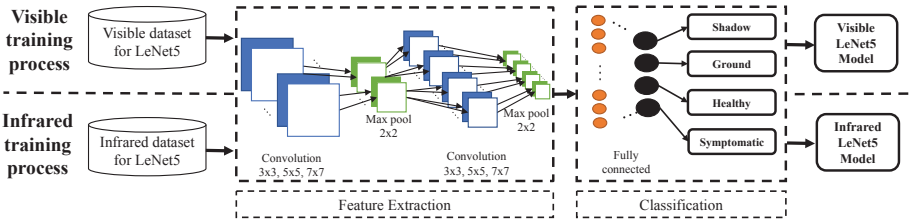

Figure 8: LeNet5 training scheme using visible and infrared image datasets.

eased ones. By referring to the ground truth, each vine was referenced with its coordinates in the field. Then, this information was used to label the areal images. Figure 9 shows an example of the semi-automatic labelling steps.

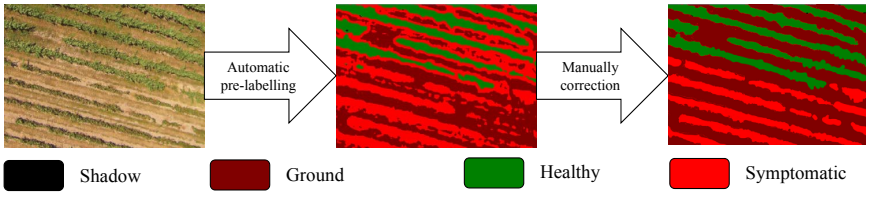

Figure 9: Example of semi-automatic labelling. The pre-labeling was carried out by the LeNet5 model using the sliding window method for segmentation. Manual correction was performed using Paint.Net software.

4.2.3. Data augmentation

Due to the huge amount of data required to train a deep learning network, the lack of data (specially the symptomatic sample), and the difficulty of labelling images, several methods of data augmentation [80] were used (see Figure 10). First, two registered images acquired by UAV with a size of 4608 × 3456 pixels were labelled by a semi-automatic method. Then each of these images underwent automatic data augmentation to generate patches of 360 × 480 pixels labelled, used in SegNet network training.

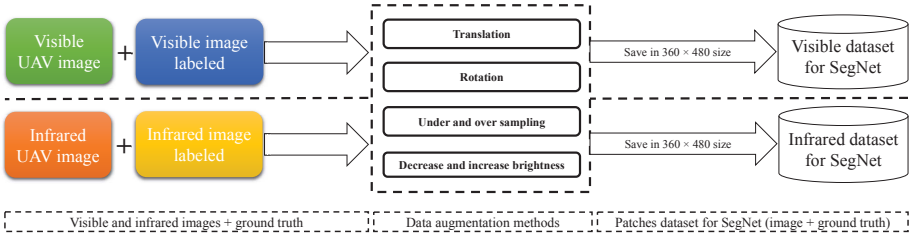

Figure 10: Data augmentation process of the visible and infrared datasets for training the SegNet architecture.

In this process, several data augmentation techniques were combined. The first transformation consisted in horizontal shift by an overlap of 50% to enable the network to learn the transition areas. Rotation was also used, since it enables the network to learn the different orientations of the vinerows. Several rotation values ranging between 0° and 180° with a step of 30° were used. The third transformation consisted in under and over sampling (scale change) in order to enable the network to learn that the thickness of the vinerows could change, to reduce misclassification in the case of a scale change, such as a change in altitude of the UAV. For that purpose, a sub-sampling between 0.5 to 1 of the real scale, and an oversampling between 1 to 1.5 was performed with a step of 0.25. The fourth technique of data augmentation consisted in modifying the brightness, to make the network less sensitive to light changes. As the image acquisition was done outdoors, the brightness parameters were uncontrollable

because of changes in the weather. Thus, coefficients between 0.8 to 1.2 with a step of 0.1 were multiplied by the grey levels of the images to generate dark, normal, and bright effects.

4.2.4. Training and testing data

By using data augmentation methods, a dataset of $105,515$ and $98,895$ patches (360×480 pixels for each patch) was generated for the visible and infrared range respectively. This dataset was used to train and validate the SegNet models. Among the dataset patches, 85% ($89,688$ for the visible range, and $84,061$ for the infrared range) was randomly selected for training and the remaining ones (15%) were used for validation. To evaluate and test the SegNet models, 200 patches for each range (without undergoing data augmentation), taken from another area at two different periods of time were used to test the visible and infrared models, and the fusion between them.

4.3. Segmentation and fusion

This subsection presents the visible and infrared image segmentation system, the deep learning architecture used in this process, the labelling and training method, and finally, the overall operation.

4.3.1. Deep learning segmentation

The CNN architecture has been very successful in the pattern recognition and computer vision fields. Since then, there has been a continuous evolution of CNN architectures. The new architectures have become deeper, but also new types of architectures have emerged that directly segment an image, such as the SegNet [39] architecture, used in the present study. To segment an image, the SegNet architecture (Figure 11) operates through two opposite phases, an information encoding phase and a decoding and classification phase (Table 2). The encoding phase is in fact a classic CNN architecture, usually with a VGG-16 [27] architecture. The coding phase consists of three types of processing, namely convolution layers, ReLU layers (ReLU is an activation function for removing the negative values that result from convolution and deconvolution. It is commonly

used in deep learning networks, because it performs better than other activation functions) and MaxPooling layers (non linear sub-sampling function). Decoding consists of the same types of processing except for the convolution layers which are replaced by deconvolution layers, and the MaxPooling layers which are replaced by Upsampling layers. It is in the decoding part of the network that the segmented image is formed, until the final decoding layer is reached. At this level, a pixel-wise segmentation is performed by the Softmax function.

In the present study, generation of the disease map of a vineyard field was considered as a four-class segmentation problem in the visible and infrared range. The objective was to build a SegNet model capable of differentiating between shadow, ground, healthy and symptomatic (visible and infrared) classes in the two spectral bands. The distinction between each class is mainly based on variations in color, texture, spectral information and spatial relative position of each class. This important information must be extracted by the SegNet network during the encoding phase, also called the feature extraction phase, then rebuilt and segmented by the decoding phase.

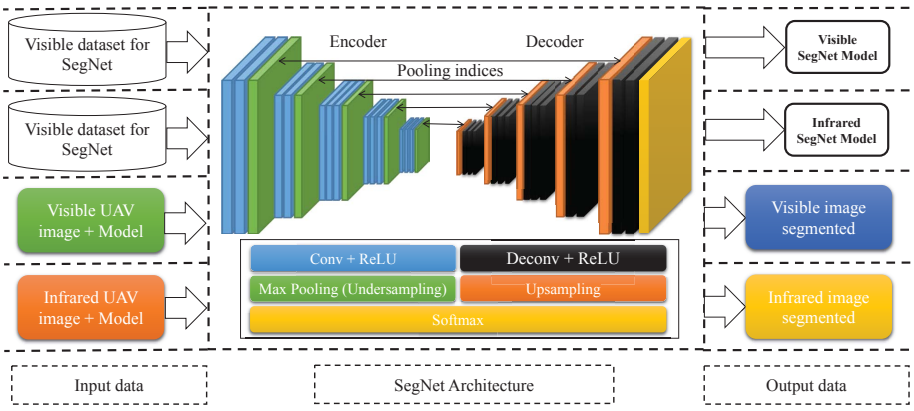

Figure 11: Visible and infrared image modeling and SegNet segmentation system.

Table 2: SegNet setting.

Phase	Number	Layer type	Filter size	Number of feature maps
Encoder	1	Conv1-1	3×3	64
		Conv1-2	3×3	64
	2	Conv2-1	3×3	128
		Conv2-2	3×3	128
	3	Conv3-1	3×3	256
		Conv3-2	3×3	256
		Conv3-3	3×3	256
	4	Conv4-1	3×3	512
		Conv4-2	3×3	512
		Conv4-3	3×3	512
	5	Conv5-1	3×3	512
		Conv5-2	3×3	512
		Conv5-3	3×3	512
Decoder	5	Deconv5-3	3×3	512
		Deconv5-2	3×3	512
		Deconv5-1	3×3	512
	4	Deconv4-3	3×3	512
		Deconv4-2	3×3	512
		Deconv4-1	3×3	512
	3	Deconv3-3	3×3	256
		Deconv3-2	3×3	256
		Deconv3-1	3×3	256
	2	Deconv2-2	3×3	128
		Deconv2-1	3×3	128
	1	Deconv1-2	3×3	64
		Deconv1-1	3×3	64

4.3.2. Fusion of multimodal image segmentation

The fusion of segmentations was performed in order to obtain a disease map with more robust results. To generate a disease map, each pixel of the image segmented in the visible range was compared with the pixel of the same position in the infrared range. Here, three main cases are considered. The first one is that the two pixels represent the symptomatic class. In this case, the result is a symptom intersection class. The second case is when the pixel is classified as symptomatic in the infrared range, and healthy in the visible range. In this case, the resulting class is symptomatic infrared (This may be a case where the disease has not yet affected the visible range by discoloration of the leaves). The third case is when the pixel is classified as healthy in the infrared range, and symptomatic in the visible range. The resulting class is visible symptomatic.

To evaluate the disease map, two cases were assessed. The first one is the case described in the previous paragraph. It is named fusion by intersection and symbolized "Fusion AND"; the AND operator means that the symptom is considered to be detected if it is present in both visible and infrared images. The second case is named fusion by union and is symbolized "Fusion OR". As the name indicates, this case unites visible and infrared range detections with the OR operator.

5. Experimentation

This section details the experiments, testing, validation and interpretation of the results. The algorithms were developed under the Python 2.7 development environment using the Tensorflow 1.8.0, NumPy 1.16.2 and OpenCV 3.0.0 libraries. To run and evaluate the runtime of our algorithms, we used the following hardware; an Intel Xeon (R) W-2123 (a) 3.60 GHz×8 processor (CPU) with 32 GB of RAM, and a graphics processing unit (GPU) NVidia GTX 1080Ti with an internal RAM of 11 GB under the Linux operating system Ubuntu 16.04 LTS (64 bits).

The experimental section is divided into two subsections. The first one

concerns the visible and infrared image registration, and the second one details the segmentation and fusion steps.

5.1. Evaluations of image registration algorithm

Performance measurement was carried out by computing the RMSE Eq. 4, 5 and 6 between the points of interest matched, in pixel units. RMSE provides information on the geometric correction and the shift between the visible and infrared images.

Table 3: Statistical performance results for standard and optimized registration for a dataset of 150 images. Mean, Min and Max are the average, minimum and maximum number of the statistical results, respectively.

Methods / Measure	Standard registration			Optimized registration		
	Mean	Min	Max	Mean	Min	Max
RMSE "Pixels"	3.29 ± 1.57	1.13	9.75	$\mathbf{2.43 \pm 1.26}$	**0.87**	**9.02**
Runtime "Seconds"	$\mathbf{92 \pm 19}$	**49**	**129**	139 ± 40	67	238
Number of iterations	-	-	-	3.12 ± 1.48	1	7

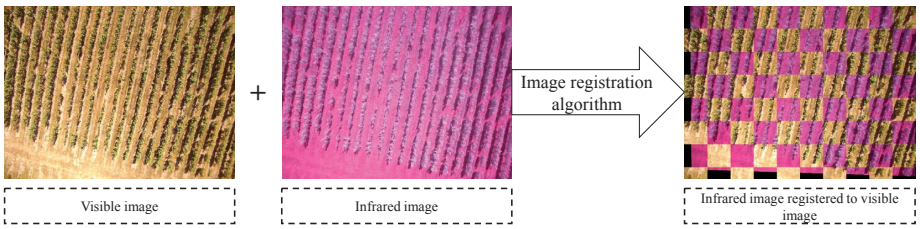

Figure 12: Result on image registration.

5.2. Experiment on image segmentation and fusion

5.2.1. Training and testing procedure

Training of the SegNet model was carried out using data from the first area of P1. During this phase, the network uses patches of size 360 $\times$ 480 pixels. Each

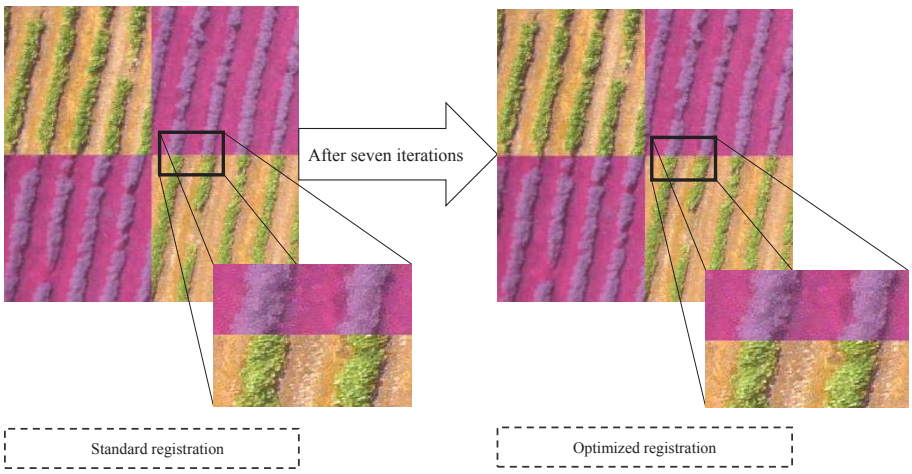

Figure 13: Correction of the shift by seven iterations.

batch was composed of 5 patches and with a number of iterations of 100,000. Lastly, two SegNet models were generated, the first one for the visible and the second one for the infrared. These models were evaluated by quantitative and qualitative methods.

The quantitative evaluation was carried out on two different times from the second area of P1 (not used for training). The results were compared to the ground truth, to evaluate the method.

In order to show the segmentation quality, a qualitative evaluation was carried out on some images from the second area of P1 (Figure 14), and some images from P2 (Figure 15). To realize this step, the UAV image was cut into patches of size 360 $\times$ 480 pixels. These patches were segmented by the SegNet network, then stitched together to find the original size of the UAV image (4608 $\times$ 3456 pixels).

5.2.2. Performance measurement

The segmentation performance was measured by two methods. The first one (presented in Table.4) is based on the leaf-level (pixel-wise) computation of

the Recall (eq.7), precision (eq.8), F1-Score (eq.9)/Dice coefficient (eq.10) and Accuracy (eq.11) for each class (shadow, ground, healthy and symptomatic). The second evaluation is based on the grapevine-level (presented in Table.5). The pixel-wise segmentation measurement does not provide information about detection at the grapevine level; it merely indicates whether a given grapevine is infected or not. This measurement uses a sliding window with a size of 64×64 pixels (corresponding to the average size of a grapevine in the studied plots). Inside this window, only the dominant class in the ground truth is evaluated. If there is a match between the ground truth and the SegNet estimation, it is considered as true positive, otherwise it is false positive. At the end of this process, the same measurement is computed.

$$Recall = \frac{TP}{TP+FN} \tag{7}$$

$$Precision = \frac{TP}{TP+FP} \tag{8}$$

$$F1 - Score = 2\frac{Recall \times Precision}{Recall+Precision} = \frac{2TP}{FP+2TP+FN} \tag{9}$$

$$Dice = \frac{2|X \cap Y|}{|X|+|Y|} = \frac{2(TP)}{(FP+TP)+(TP+FN)} = \frac{2TP}{FP+2TP+FN} \tag{10}$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{11}$$

where TP, TN, FP and FN are the number of samples for "True Positive", "True Negative", "False Positive" and "False Negative". For the Dice equation, X is the set of ground truth pixels and Y is the set of pixels classified by the SegNet classifier.

6. Discussion

The first question of this study was to determine the ability of multispectral drone imaging to map grapevine Mildew symptoms using machine learning

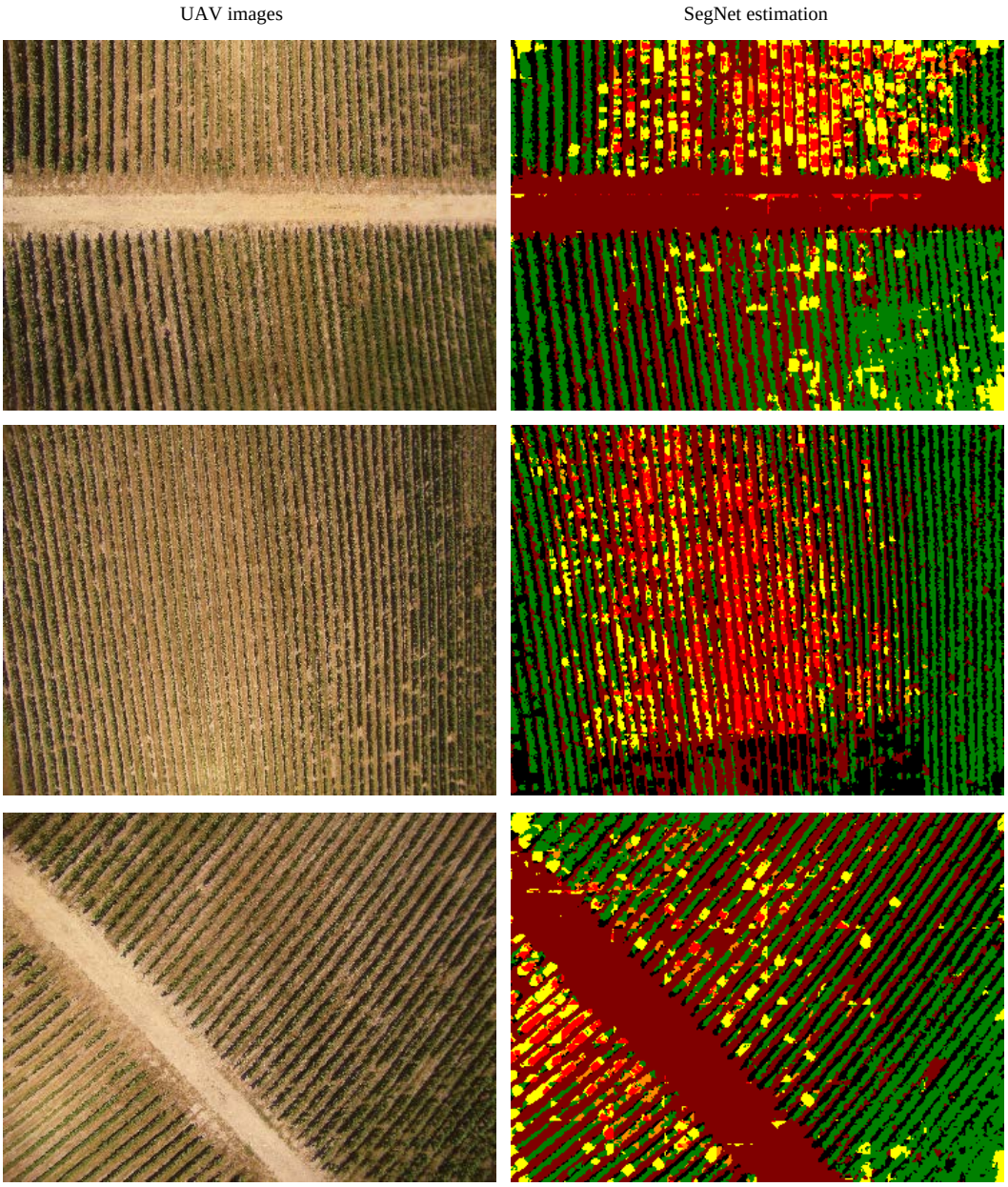

Figure 14: Qualitative segmentation results by the SegNet method for the P1 vineyard. On the left, images from UAV on the right, the segmentation result of these images. The color code of the segmentation is: Black: Shadow, Brown: Ground, Green: Healthy, Yellow: Visible symptom, Orange: Infrared symptom, Red: Symptom intersection.

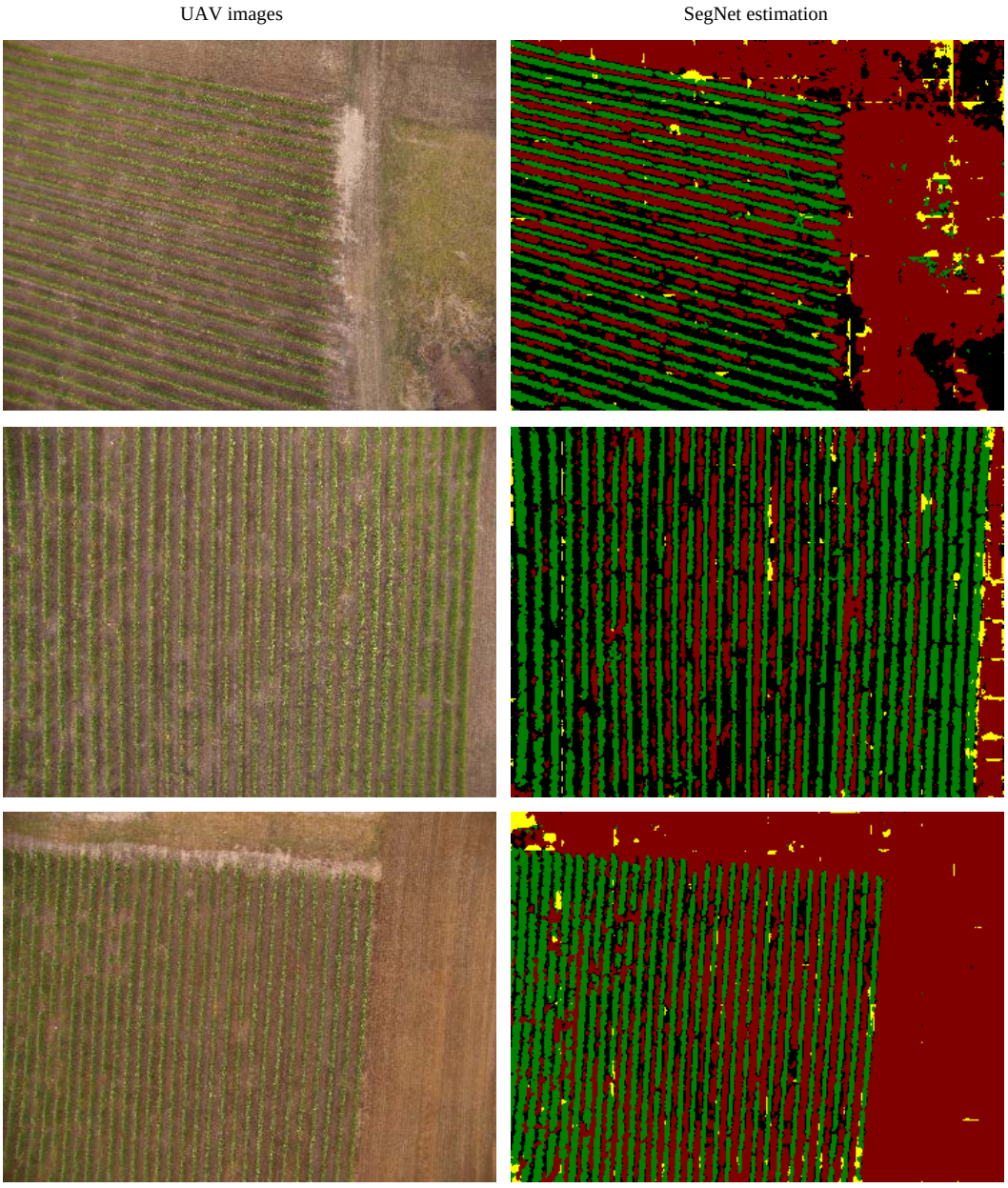

Figure 15: Qualitative segmentation results by the SegNet method for the P2 vineyard. On the left, images from UAV on the right, the segmentation result of these images. The color code of the segmentation is: Black: Shadow, Brown: Ground, Green: Healthy, Yellow: Visible symptom, Orange: Infrared symptom, Red: Symptom intersection.

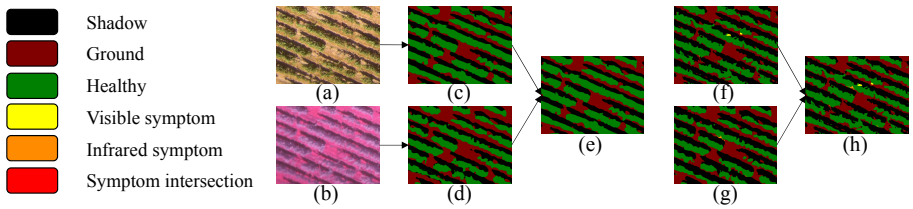

Figure 16: Example of segmentation and fusion of a healthy area. (a): Visible image, (b): Infrared image, (c): Visible ground truth, (d): Infrared ground truth, (e): Fusion ground truth, (f): Visible SegNet estimation, (g): Infrared SegNet estimation, (h): Fusion of segmentation results.

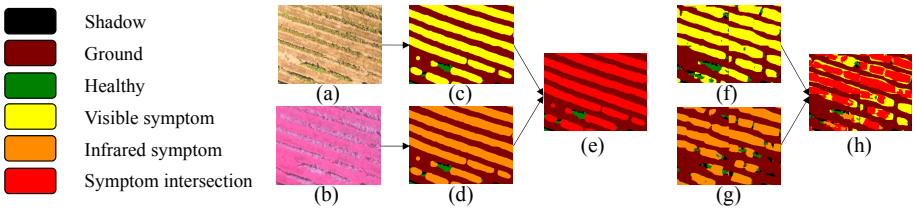

Figure 17: Example of segmentation and fusion of an area contaminated by Mildew. (a): Visible image, (b): Infrared image, (c): Visible ground truth, (d): Infrared ground truth, (e): Fusion ground truth, (f): Visible SegNet estimation, (g): Infrared SegNet estimation, (h): Fusion of segmentation results.

Table 4: The leaf-level (pixel-wise) average result on two temporal tests by measuring Recall (Rec.), Precision (Pre.), F1-Score/Dice (F1/D.) and Accuracy (Acc.) on the performance of visible, infrared image segmentation and fusion (values presented in percent). Note that: "Fusion AND" and "Fusion OR" are the cases where their symptomatic classes are respectively the intersection and the union of the symptomatic visible and infrared classes.

Class name	Shadow			Ground			Healthy			Symptomatic			Total
Measure	Rec.	Pre.	F1/D	Rec.	Pre.	F1/D	Rec.	Pre.	F1/D	Rec.	Pre.	F1/D	Acc.
Visible	76.31	87.25	81.05	91.37	95.95	93.51	86.86	66.89	75.31	80.22	77.99	78.72	85.13
Infrared	84.25	72.25	77.69	87.74	91.33	89.42	73.81	50.18	58.58	59.02	85.06	69.55	78.72
Fusion AND	87.84	86.78	87.30	95.73	95.95	95.84	83.73	69.09	75.60	53.70	94.02	67.93	82.20
Fusion OR	87.84	86.78	87.30	95.73	95.95	95.84	82.12	72.30	76.55	84.07	90.47	87.12	90.23

Table 5: The grapevine-level average result on two temporal tests by measuring Recall (Rec.), Precision (Pre.), F1-Score/Dice (F1/D.) and Accuracy (Acc.) on the performance of visible, infrared image segmentation and fusion (values presented in percent). Note that: "Fusion AND" and "Fusion OR" are the cases where their symptomatic classes are respectively the intersection and the union of the symptomatic visible and infrared classes.

Class name	Shadow			Ground			Healthy			Symptomatic			Total
Measure	Rec.	Pre.	F1/D	Rec.	Pre.	F1/D	Rec.	Pre.	F1/D	Rec.	Pre.	F1/D	Acc.
Visible	94.00	93.42	93.63	97.39	97.94	97.66	95.16	85.20	89.91	90.15	92.97	91.50	94.41
Infrared	97.53	79.97	87.55	97.30	95.41	96.32	93.72	69.19	79.19	70.49	96.92	81.66	89.16
Fusion AND	94.01	86.62	89.96	97.41	97.89	97.65	93.81	87.55	90.56	66.92	73.12	68.03	88.14
Fusion OR	94.00	94.03	93.95	97.39	97.94	97.66	93.81	89.65	91.68	92.91	92.78	92.81	95.02

Table 6: Results on runtime performance (expressed in seconds) for the entire system as a function of the average runtime of standard and optimized registration.

Step	Registration	SegNet seg.	Fusion	Total
Standard registration	92	140×2	2	374
Optimized registration	139	140×2	2	421

approaches. This led to studying imaging modalities in the visible and infrared spectral domains, since several studies have shown the interest of these domains for symptom detection. As our system consisted of two cameras for each modality, image alignment was required. We were, therefore, led to develop an

algorithm for image registration and then to use deep learning segmentation to detect the affected surfaces in the vineyard. Another question of this research was how, through the deep learning approach, we can combine the two types of images to delineate symptomatic areas as precisely as possible. Thus, the rate of affected areas can be obtained at the leaf scale or at the vine plant scale. The following sections first discuss the results of the registration and then those of the image segmentation.

6.1. Image registration

Figure 12 shows the qualitative result of the registration. On the left and in the middle are the visible and infrared images respectively. The images were taken at a very short time interval (less than one second) and in the same field of view. After registration, two black areas can be observed on the left and at the bottom of the image of the right. The two areas were captured only by the visible sensor, which explains why there are no equivalents in the infrared image registered at the moment of superposition. There are also areas captured only by the infrared sensor, but due to the fact that the infrared image is registered to the visible image, these areas are not displayed on the result.

An example of how the proposed registration method operates is shown in Figure 13. In the image on the left (image registered with the standard method) there is a certain shift between the vinerows present in the visible and infrared images. This is due to a lack of matched points between the two images. However, after each iteration, some new correspondence points are detected thanks to the dynamic threshold, and gradually, seven iterations later, a good alignment of the vinerows is achieved as shown in the image on the right.

The quantitative results obtained during the registration experiments are presented in Table 3. This table reports a statistical study on a dataset of 150 images, for a comparison between the standard registration method used in [56, 57] and the proposed optimized registration method. The values obtained are expressed in "pixels" for the RMSE error measurement, in "seconds" for the runtime and in "times" for the iteration. For each line, the figures in bold show

the best results obtained. For standard registration, an average error of 3.29 pixels was obtained (this result corresponds to the range of results obtained in [58] by the same algorithm), compared to 2.43 pixels for the optimized method, i.e. an average error reduction of 0.86 pixels over the entire dataset. This error reduction can be explained by the appearance of new correspondence points, in subsequent iterations, between the two images. These points are used to calculate a new homographic matrix, which corrects the registration slightly, and reduces the RMSE error. However, there are special cases where the optimized method does not reduce the error, in which the result of the optimized method is cancelled and the output of the standard registration is kept. The best score obtained by the standard method yields an error of 1.13 pixels, whereas with the optimized method the minimum error is 0.87 pixels. However, both methods can produce larger errors, as we can find errors up to 9.75 pixels with the standard method and 9.02 pixels with the optimized method. Note that our disease detection system performs better with an RMSE value ≤ 5; RMSE values between 5 and 10 may in some cases reduce the accuracy of disease localization.

The runtime measurement is shown in Table 3. For the standard registration, an average runtime of 92 seconds was obtained, versus 139 seconds with the optimized method. This increase in runtime is due to the additional processing performed by the optimized method (3.12 iterations on average). This is not the only reason, however: with the standard method, there is a very significant difference (80 seconds) between the minimum and the maximum values. The difference is due to the number of interest points detected in an image by the AKAZE algorithm: the larger the number of interest points detected, the longer the processing time. In other words, there is a direct relation between these two parameters.

The proposed algorithm based on the AKAZE detector was tested by other feature extraction algorithms such as SIFT, SURF, ORB and KAZE. The results obtained by SIFT showed a slightly higher error than those obtained by AKAZE, and the runtime was between 2 to 10 times longer than AKAZE, confirming the study by Tareen et al. [81]. For the other algorithms, several cases of

failure were identified. They are mainly due to the lack of correspondence between the two images, which implies a sensitivity of these algorithms to the differences of modality between the two images. Another algorithm developed by Yang et al. [59], which is based on feature extraction through a deep learning architecture was tested, and several problems were observed. Not only did the deep learning method use a large amount of RAM memory, making it very difficult to register high resolution images, but even with a decrease in the resolution of the images, the algorithm cannot find good matches, which leads to the failure of the registration process. This is due to the difference in image modalities.

Some other Area-based registration algorithms were also tested, such as: Normalized Cross-Correlation (NCC) [53] and Phase-Correlation (PC) [54]. But due mainly to the presence of some deformations on the image dataset used, these algorithms did not manage to correctly register most of the images. Another disadvantage of these algorithms is that they are very time consuming, and were therefore discarded.

6.2. Image segmentation and fusion

The qualitative results given in figures 14 and 15 represent the results on the vineyards of P1 and P2 respectively. For the P1 plot (Figure 14), a large part of the untreated plot was contaminated by Mildew. This presence of disease is well detected by segmentation. As can be observed, in most cases, the symptoms are better detected in the visible range (coded in yellow) than in the infrared range (coded in orange). This hypothesis was confirmed by the quantitative results in Table 4. Of course, there are in some cases false detections in both areas. However, when symptoms are detected in both ranges it is likely to be true detection.

For the P2 plot (Figure 15), the vineyard is healthy, and the SegNet estimation generally matches the ground truth. However, in this example of 3 images, some misclassification of symptomatic areas in the visible range can be observed. This generally occurs when there is a gap in the vegetation (a plant

missing in a row, for example) and the color of the soil is similar to that of a symptomatic plant (brown or golden). Also, it is usually explained by the edge effects of the sliding window, because the information is not complete on some borders (these effects are not always present in the outputs). In other cases, the symptom class may be assigned when yellow is mixed with green, as this case is usually comparable to that of symptomatic leaves. To counter this problem, it is necessary to check the neighboring images that cover this area. If the symptom is detected in all or most of the images, there may be a real symptom, otherwise, it is a false detection. Misclassification of the other classes can also occur. Indeed, in these examples, some misclassification of the grass can be observed which is detected as healthy vine or symptom class (in the 1st example of Figure 15). Some confusion between the ground and the shadow class can also occur because of the low ground brightness (in the 2nd and 3rd examples of Figure 15).

Figures 16 and 17 show an example of SegNet segmentation and the fusion compared with the ground truth. In Figure 16 the area is healthy, so there are no samples of the symptom class. In this case, it can be observed that the fusion is identical to the visible and infrared estimation (idem for the ground truth). In Figure 17, in contrast, which is an area almost completely contaminated by Mildew, it can be seen that the ground truth is identical in both spectra, apart from the color code which is different. But for the SegNet output, the result for the symptomatic class detection is not identical in the two ranges, and consequently the symptomatic visible and infrared classes are merged. In addition, fusion by intersection is generated by the AND operator between the two segmentation outputs.

The quantitative results obtained for the visible, infrared, fusion by intersection and union segmentation experiments are presented in Table 4 and 5, respectively for leaf-level and grapevine-level. They show the results obtained in terms of the Recall, Precision, F1-Score/Dice coefficient and Accuracy measures, expressed in percentages.

As shown in the "Accuracy" column of Table 4, the different classes were

generally better detected from the visible image (accuracy of 85.13%) than from the infrared image (accuracy of 78.72%). This difference in result is due to the fact that the visible image provides a better colorimetric description than the infrared image for the different classes studied. The fusion by union gave a result of 90.23%. This result is better than the visible image, because the method takes the best of visible and infrared information. It can be noticed that symptoms appear at different locations in the visible and infrared spectral ranges. One interesting finding is that detection in the two ranges is complementary, since fusion by union increases the detection performance. The fusion by intersection yielded a score of 82.20%. This result is less than that of the visible range, because the method is conditioned by the intersection of the visible and infrared, and in this case, it is the infrared result that decreases the result. The fusion by intersection provides important information about the position of the joint symptom detection in the visible and infrared images. This finding can be used to strengthen the robustness of Mildew detection, where detection can be considered reliable if it is observed in both types of images. Besides, the fusion by union gives an idea about the quantitative detection.

In addition to measuring the affected areas at leaf level, the second type of assessment consists in testing the detection at vine plant level, as this helps to better manage certain operations in the vineyard. The results presented in Table 5 give better insight into the detection of symptoms at the grapevine-level. The results show that the detection of symptoms in fusion by union is much better (detection exceeds 92.81%) compared to the fine scale detection (leaf level), followed by the visible range (91.50%), then the infrared range (81.66%). As can be seen, the results of infrared and fusion by intersection are less than the pixel-wise evaluation (Table 4). On the other hand, an increase in precision for the cases of fusion by union and the visible range can be observed.

Results obtained by the proposed method are likely to be consistent with several studies in the field of remote sensing [47, 82, 83, 84, 44, 85] using the SegNet network. Indeed, the overall accuracy range obtained by these studies is between 70% and 90%. In addition, it has been observed that when the surface

area is large, the detection result is better. Conversely, the smaller the surface area, the more difficulty the SegNet network has in correctly detecting the diseased area. A possible explanation for this might be the loss of the resolution information of small areas during the downsampling and upsampling operations in the SegNet network. Another reason why the results are limited is the difficulty of realizing an accurate ground truth for training and testing the network, but also the difficulty identifying an area in a low resolution ($1cm/pixel$) (in the case of images taken by a UAV at high altitude). Other remote sensing applications with large database studies that have tested and compared several types of deep learning architecture [86, 84] such as FCN, U-Net, DeepLab, PSPNet, etc., obtained the best semantic segmentation results with the SegNet network.

In the proposed system, the fusion by intersection of the two modalities indicates the locations where symptoms were located at the same position in the visible and infrared images. In other words, fusion provides important information about areas where the system confidence is higher for disease detection. Moreover, it also provides information about areas where the disease has been detected only in one range (visible or infrared). Therefore, even after the establishment of the method for disease detection, fusion by intersection remains a more reliable class than the symptomatic classes detected in the visible or infrared range.

RGB and multispectral images were combined as one input then fed to the SegNet. The results obtained were poorer. This is due to the registration of the visible and infrared images. In fact, there is always a small random shift between the multispectral images, which implies that the same pixels are never aligned exactly in the same position.

6.3. Runtime system

Table 6 presents the runtime results of all stages of the disease detection system. For the image registration step, the average runtime value was used to evaluate the overall system, because the image registration runtime is variable from one pair of images to another. For the SegNet segmentation step, the

runtime was multiplied by two ($\times$ 2) because the process must segment both images (visible and infrared) and the GPU can only handle one process at a time, unlike registration and fusion operations, where the processing is joint for the two images (visible and infrared). Unlike image registration, the runtime of the SegNet on a UAV image is constant, at 140 seconds. This value is the same for both visible and infrared images. The fusion between the two segmented images takes less than 2 seconds; this runtime value includes the computation and saving the fusion file. The runtime of the disease detection system varies according to the image registration method chosen. This implies that to improve the accuracy of the results, an additional average processing of 47 seconds per image is necessary.

7. Conclusion

In this study, a new method based on optimized image registration and a deep learning segmentation method has been proposed for detecting vine disease using multimodal UAV images (visible and infrared ranges). The method comprises three steps. The first one is the image alignment, where an iterative algorithm based on an interest points detector has been developed. The second step is the segmentation of visible and infrared images based on the SegNet architecture to identify four classes: shadow, ground, healthy and symptomatic vine. Lastly, the third step consists in generating a disease map by fusion of the segmentations obtained from the visible and infrared images. This study showed that the proposed method enables the detection of vine symptoms using information from images of visible and infrared spectra. It provides a framework for the exploration of earlier detection and mapping of vine diseases. One of the limitations of this research is the small size of the training sample which reduced the performance of the deep learning segmentation. In future work, the following improvements could be made. The segmentation method can be improved by enriching the dataset (diversification of disease samples), and also by

testing other deep learning architectures for segmentation. Another possibility is the use of 3D information from the vine canopy, thus reducing false detection and improving the accuracy of image registration.

Acknowledgment

This work is part of the VINODRONE project supported by the Region Centre-Val de Loire (France). We gratefully acknowledge Region Centre-Val de Loire for its support.

References

[1] W. Aktar, D. Sengupta, A. Chowdhury, Impact of pesticides use in agriculture: Their benefits and hazards, Interdisciplinary Toxicology 2 (1) (2009) 1–12. doi:10.2478/v10102-009-0001-7.

[2] D. Pimentel, L. McLaughlin, A. Zepp, B. Lakitan, T. Kraus, P. Kleinman, F. Vancini, W. J. Roach, E. Graap, W. S. Keeton, G. Selig, Environmental and economic effects of reducing pesticide use in agriculture, Agriculture, Ecosystems and Environment 46 (1-4) (1993) 273–288. doi:10.1016/0167-8809(93)90030-S.

[3] M. Hernández, A. Margalida, Pesticide abuse in Europe: Effects on the Cinereous vulture (Aegypius monachus) population in Spain, Ecotoxicology 17 (4) (2008) 264–272. doi:10.1007/s10646-008-0193-1.

[4] S. D. Tyerman, R. K. Vandeleur, M. C. Shelden, J. Tilbrook, G. Mayo, M. Gilliham, B. N. Kaiser, Water transport & aquaporins in grapevine, 2009. doi:10.1007/978-90-481-2305-6_4.

[5] V. Hofstetter, B. Buyck, D. Croll, O. Viret, A. Couloux, K. Gindro, What if esca disease of grapevine were not a fungal disease?, Fungal Diversity 54 (2012) 51–67. doi:10.1007/s13225-012-0171-z.

[6] J. Chuche, D. Thiéry, Biology and ecology of the Flavescence dorée vector Scaphoideus titanus: A review, Agronomy for Sustainable Development 34 (2) (2014) 381–403. doi:10.1007/s13593-014-0208-7.

[7] C. Gessler, I. Pertot, M. Perazzolli, Plasmopara viticola: A review of knowledge on downy mildew of grapevine and effective disease management, Phytopathologia Mediterranea 50 (1) (2011) 3–44. doi:10.14601/Phytopathol_Mediterr-9360.

[8] S. L. MacDonald, M. Staid, M. Staid, M. L. Cooper, Remote hyperspectral imaging of grapevine leafroll-associated virus 3 in cabernet sauvignon vineyards, Computers and Electronics in Agriculture 130 (2016) 109–117. doi:10.1016/j.compag.2016.10.003.
URL http://linkinghub.elsevier.com/retrieve/pii/S016816991630237X

[9] F. C. H. Al-Saddik, JC. Simon, O. Brousse, Damav : Un Projet Interregional De Detection De Foyers Infectieux De Flavescence Doree Par Imagerie De Drone, Journée technique VITINNOV Viticulture de précision : les capteurs à la loupe DAMAV (2016) 32–35.

[10] M. Teke, H. S. Deveci, O. Haliloglu, S. Z. Gurbuz, U. Sakarya, A short survey of hyperspectral remote sensing applications in agriculture, RAST 2013 - Proceedings of 6th International Conference on Recent Advances in Space Technologies (June) (2013) 171–176. doi:10.1109/RAST.2013.6581194.

[11] A. G. Schut, P. C. Traore, X. Blaes, R. A. de By, Assessing yield and fertilizer response in heterogeneous smallholder fields with UAVs and satellites, Field Crops Research 221 (February) (2018) 98–107. doi:10.1016/j.fcr.2018.02.018.
URL https://doi.org/10.1016/j.fcr.2018.02.018

[12] M. Karpina, M. Jarzabek-Rychard, P. Tymków, A. Borkowski, Uav-based automatic tree growth measurement for biomass estimation, In-

ternational Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives 41 (July) (2016) 685–688. doi:10.5194/isprsarchives-XLI-B8-685-2016.

[13] M. D. Bah, A. Hafiane, R. Canals, Weeds detection in UAV imagery using SLIC and the hough transform, Proceedings of the 7th International Conference on Image Processing Theory, Tools and Applications, IPTA 2017 2018-January (2018) 1–6. doi:10.1109/IPTA.2017.8310102.
URL http://ieeexplore.ieee.org/document/8310102/

[14] M. D. Altnba, T. Serif, Detecting Defected Crops: Precision Agriculture Using Haar Classifiers and UAV, in: I. Awan, M. Younas, P. Ünal, M. Aleksy (Eds.), 13th International Conference, MobiWIS 2016 Vienna, Austria, August 2224, 2016, Vol. 11673 of Lecture Notes in Computer Science, Springer International Publishing, Cham, 2019, pp. 27–40. doi:10.1007/978-3-030-27192-3_3.
URL http://link.springer.com/10.1007/978-3-030-27192-3http://link.springer.com/10.1007/978-3-030-27192-3{_}3

[15] S. D. Khirade, A. B. Patil, Plant disease detection using image processing, Proceedings - 1st International Conference on Computing, Communication, Control and Automation, ICCUBEA 2015 (2015) 768–771doi:10.1109/ICCUBEA.2015.153.
URL http://ieeexplore.ieee.org/document/7155951/

[16] L. S. Pinto, A. Ray, M. U. Reddy, P. Perumal, P. Aishwarya, Crop disease classification using texture analysis, 2016 IEEE International Conference on Recent Trends in Electronics, Information and Communication Technology, RTEICT 2016 - Proceedings (2017) 825–828doi:10.1109/RTEICT.2016.7807942.

[17] N. Schor, A. Bechar, T. Ignat, A. Dombrovsky, Y. Elad, S. Berman, Robotic Disease Detection in Greenhouses: Combined Detection of Powdery Mildew and Tomato Spotted Wilt Virus, IEEE Robotics and Automation Letters

1 (1) (2016) 354–360. doi:10.1109/LRA.2016.2518214.
URL http://ieeexplore.ieee.org/document/7383246/

[18] J. G. A. Barbedo, A Review on the Use of Unmanned Aerial Vehicles and Imaging Sensors for Monitoring and Assessing Plant Stresses, Drones 3 (2) (2019) 40. doi:10.3390/drones3020040.

[19] W. Zhu, H. Chen, I. Ciechanowska, D. Spaner, Application of infrared thermal imaging for the rapid diagnosis of crop disease, IFAC-PapersOnLine 51 (17) (2018) 424–430. doi:10.1016/j.ifacol.2018.08.184.
URL https://doi.org/10.1016/j.ifacol.2018.08.184

[20] F. Peñaranda, V. Naranjo, R. Verdú-Monedero, G. R. Lloyd, J. Nallala, N. Stone, Multimodal registration of optical microscopic and infrared spectroscopic images from different tissue sections: An application to colon cancer, Digital Signal Processing: A Review Journal 68 (2017) 1–15. doi:10.1016/j.dsp.2017.04.014.
URL http://dx.doi.org/10.1016/j.dsp.2017.04.014

[21] S. Wang, D. Quan, X. Liang, M. Ning, Y. Guo, L. Jiao, A deep learning framework for remote sensing image registration, ISPRS Journal of Photogrammetry and Remote Sensing 145 (2018) 148–164. doi:10.1016/j.isprsjprs.2017.12.012.
URL https://doi.org/10.1016/j.isprsjprs.2017.12.012

[22] C. H. Tsai, Y. C. Lin, An accelerated image matching technique for UAV orthoimage registration, ISPRS Journal of Photogrammetry and Remote Sensing 128 (2017) 130–145. doi:10.1016/j.isprsjprs.2017.03.017.
URL http://dx.doi.org/10.1016/j.isprsjprs.2017.03.017

[23] F. Inostroza, S. Díaz, J. Cárdenas, S. E. Godoy, M. Figueroa, Embedded registration of visible and infrared images in real time for noninvasive skin cancer screening, Microprocessors and Microsystems 55 (January) (2017) 70–81. doi:10.1016/j.micpro.2017.09.006.
URL http://dx.doi.org/10.1016/j.micpro.2017.09.006

[24] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, Proceedings of the IEEE 86 (11) (1998) 2278–2323. doi:10.1109/5.726791.

[25] B. Monien, R. Preis, S. Schamberger, Approximation algorithms for multilevel graph partitioning, Handbook of Approximation Algorithms and Metaheuristics (2007) 60–1–60–16doi:10.1201/9781420010749.

[26] M. D. Zeiler, R. Fergus, Visualizing and Understanding Convolutional Networks BT - Computer VisionECCV 2014, European Conference on Computer Vision (ECCV) 8689 (Chapter 53) (2014) 818–833.
URL http://link.springer.com/10.1007/978-3-319-10590-1{_}53{%}5Cnpapers2://publication/doi/10.1007/978-3-319-10590-1{_}53

[27] J. Zhang, H. Zhu, An empirical-theoretical calculation method and full-scale verification test for the ultimate bearing capacity of an expanded pedestal uplift pile, Modern Tunnelling Technology 51 (6) (2014) 50–57. doi:10.13807/j.cnki.mtt.2014.06.009.

[28] G. Zeng, Y. He, Z. Yu, X. Yang, R. Yang, L. Zhang, Inception-Net/GoogLeNet - Going Deeper with Convolutions, Cvpr 91 (8) (2016) 2322–2330. arXiv:1409.4842, doi:10.1002/jctb.4820.

[29] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2016-December (2016) 770–778. arXiv:arXiv:1512.03385v1, doi:10.1109/CVPR.2016.90.

[30] A. Kamilaris, F. X. Prenafeta-Boldú, Deep learning in agriculture: A survey, Computers and Electronics in Agriculture 147 (February) (2018) 70–90. doi:10.1016/j.compag.2018.02.016.

[31] C. H. Chen, H. Y. Kung, F. J. Hwang, Deep learning techniques for

agronomy applications, Agronomy 9 (3) (2019) 1–5. doi:10.3390/agronomy9030142.

[32] S. Sladojevic, M. Arsenovic, A. Anderla, D. Culibrk, D. Stefanovic, Deep Neural Networks Based Recognition of Plant Diseases by Leaf Image Classification, Computational Intelligence and Neuroscience 2016. doi:10.1155/2016/3289801.

[33] Y. Lu, S. Yi, N. Zeng, Y. Liu, Y. Zhang, Identification of rice diseases using deep convolutional neural networks, Neurocomputing 267 (2017) 378–384. doi:10.1016/j.neucom.2017.06.023.
URL http://dx.doi.org/10.1016/j.neucom.2017.06.023

[34] Q. Yang, L. Shi, J. Han, Y. Zha, P. Zhu, Deep convolutional neural networks for rice grain yield estimation at the ripening stage using UAV-based remotely sensed images, Field Crops Research 235 (February) (2019) 142–153. doi:10.1016/j.fcr.2019.02.022.
URL https://doi.org/10.1016/j.fcr.2019.02.022

[35] A. Fuentes, S. Yoon, S. C. Kim, D. S. Park, A robust deep-learning-based detector for real-time tomato plant diseases and pests recognition, Sensors (Switzerland) 17 (9). doi:10.3390/s17092022.

[36] S. S. Lee, Y. N. Jeong, S. R. Son, B. K. Lee, A self-predictable crop yield platform (SCYP) based on crop diseases using deep learning, Sustainability (Switzerland) 11 (13). doi:10.3390/su11133637.

[37] M. D. Bah, E. Dericquebourg, A. Hafiane, R. Canals, Deep learning based classification system for identifying weeds using high-resolution UAV imagery, Advances in Intelligent Systems and Computing 857 (July) (2019) 176–187. doi:10.1007/978-3-030-01177-2_13.

[38] M. Kerkech, A. Hafiane, R. Canals, Deep learning approach with colorimetric spaces and vegetation indices for vine diseases detection in UAV images,

Computers and Electronics in Agriculture 155 (October) (2018) 237–243. doi:10.1016/j.compag.2018.10.006.

[39] V. Badrinarayanan, A. Kendall, R. Cipolla, SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation, IEEE Transactions on Pattern Analysis and Machine Intelligence 39 (12) (2017) 2481–2495. arXiv:arXiv:1511.00561v3, doi:10.1109/TPAMI.2016.2644615.

[40] H. Noh, S. Hong, B. Han, Learning deconvolution network for semantic segmentation, Proceedings of the IEEE International Conference on Computer Vision 2015 International Conference on Computer Vision, ICCV 2015 (2015) 1520–1528. doi:10.1109/ICCV.2015.178.

[41] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 9351 (2015) 234–241. arXiv:arXiv:1505.04597v1, doi:10.1007/978-3-319-24574-4_28.

[42] A. Kendall, V. Badrinarayanan, R. Cipolla, Bayesian SegNet: Model Uncertainty in Deep Convolutional Encoder-Decoder Architectures for Scene UnderstandingarXiv:arXiv:1511.02680v2, doi:10.5244/c.31.57.

[43] I. Alonso, A. C. Murillo, EV-SegNet: Semantic Segmentation for Event-based CamerasarXiv:1811.12039.
URL http://arxiv.org/abs/1811.12039

[44] N. Audebert, B. Le Saux, S. Lefèvre, Segment-before-detect: Vehicle detection and classification through semantic segmentation of aerial images, Remote Sensing 9 (4). doi:10.3390/rs9040368.

[45] H. Guo, G. Wei, J. An, Dark spot detection in SAR images of oil spill using segnet, Applied Sciences (Switzerland) 8 (12). doi:10.3390/app8122670.

[46] H.-D. Nguyen, I.-S. Na, S.-H. Kim, Hand Segmentation and Fingertip Tracking from Depth Camera Images Using Deep Convolutional Neural Network and Multi-task SegNetarXiv:arXiv:1901.03465v2.

[47] Z. Du, Training SegNet for Cropland Classification of High Resolution Remote Sensing Images, 21st International Conference on Geographic Information Science (AGILE 2018) (2018) 1–6.

[48] P.-a. Wei, C.-c. Tsai, F.-c. Tai, Autonomous Navigation of an Indoor Mecanum-Wheeled Omnidirectional Robot Using Segnet, Proceedings of 2019 National Symposium on System Science and Engineering (2019) 2–4.

[49] E. Ferrante, N. Paragios, Slice-to-volume medical image registration: A survey, Medical Image Analysis 39 (2017) 101–123. doi:10.1016/j.media.2017.04.010.
URL http://dx.doi.org/10.1016/j.media.2017.04.010

[50] B. Zitová, J. Flusser, Image registration methods: A survey, Image and Vision Computing 21 (11) (2003) 977–1000. doi:10.1016/S0262-8856(03)00137-9.

[51] S. Dawn, V. Saxena, B. Sharma, I. Technology, Remote Sensing Image Registration Techniques : Survey, Image and Signal Processing. Springer Berlin Heidelberg 6134 (c) (2010) 103–112.

[52] S. Nag, Image Registration Techniques: A SurveyarXiv:1712.07540, doi:10.17605/OSF.IO/RV65C.
URL http://arxiv.org/abs/1712.07540{%}0Ahttp://dx.doi.org/10.17605/OSF.IO/RV65C

[53] X. Wang, W. Yang, A. Wheaton, N. Cooley, B. Moran, Efficient registration of optical and IR images for automatic plant water stress assessment, Computers and Electronics in Agriculture 74 (2) (2010) 230–237. doi:10.1016/j.compag.2010.08.004.
URL http://dx.doi.org/10.1016/j.compag.2010.08.004

[54] H. Erives, G. J. Fitzgerald, Automated registration of hyperspectral images for precision agriculture, Computers and Electronics in Agriculture 47 (2) (2005) 103–119. doi:10.1016/j.compag.2004.11.016.

[55] Y. Zhuang, K. Gao, X. Miu, L. Han, X. Gong, Infrared and visual image registration based on mutual information with a combined particle swarm optimization - Powell search algorithm, Optik 127 (1) (2016) 188–191. doi:10.1016/j.ijleo.2015.09.199.
URL http://dx.doi.org/10.1016/j.ijleo.2015.09.199

[56] M. T. L. Sreenu, T. Teja, Control of Brushless DC Motor with Direct Torque and Indirect Flux using SVPWM Technique, Indian Journal of Science and Technology, 8 (November) (2015) 507–515. doi:10.17485/ijst/2015/v8i.

[57] M. S. Javadi, Z. Kadim, H. H. Woon, K. M. Johari, N. Samudin, An Automatic Robust Image Registration Algorithm for Aerial Mapping, International Journal of Image and Graphics 15 (02) (2015) 1540002. doi:10.1142/s0219467815400021.

[58] F. A. Onyango, F. Nex, M. S. Peter, P. Jende, Accurate estimation of orientation parameters of UAV images through image registration with aerial oblique imagery, International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives 42 (1W1) (2017) 599–605. doi:10.5194/isprs-archives-XLII-1-W1-599-2017.

[59] Z. Yang, T. Dan, Y. Yang, Multi-temporal remote sensing image registration using deep convolutional features, IEEE Access 6 (2018) 38544–38555. doi:10.1109/ACCESS.2018.2853100.

[60] A. K. Mahlein, Plant disease detection by imaging sensors Parallels and specific demands for precision agriculture and plant phenotyping, Plant Disease 100 (2) (2016) 241–254. doi:10.1094/PDIS-03-15-0340-FE.
URL http://apsjournals.apsnet.org/doi/10.1094/PDIS-03-15-0340-FE

[61] E. C. Oerke, P. Fröhling, U. Steiner, Thermographic assessment of scab disease on apple leaves, Precision Agriculture 12 (5) (2011) 699–715. doi:10.1007/s11119-010-9212-3.

[62] A. H. Junges, J. R. Ducati, C. Scalvi Lampugnani, M. A. K. Almança, Detection of grapevine leaf stripe disease symptoms by hyperspectral sensor, Phytopathologia Mediterranea 57 (3) (2018) 399–406. doi:10.14601/Phytopathol_Mediterr-22862.

[63] S. F. di Gennaro, E. Battiston, S. di Marco, O. Facini, A. Matese, M. Nocentini, A. Palliotti, L. Mugnai, Unmanned Aerial Vehicle (UAV)-based remote sensing to monitor grapevine leaf stripe disease within a vineyard affected by esca complex, Phytopathologia Mediterranea 55 (2) (2016) 262–275. doi:10.14601/Phytopathol_Mediterr-18312.

[64] J. Albetis, S. Duthoit, F. Guttler, A. Jacquin, M. Goulard, H. Poilvé, J. B. Féret, G. Dedieu, Detection of Flavescence dorée grapevine disease using Unmanned Aerial Vehicle (UAV) multispectral imagery, Remote Sensing 9 (4) (2017) 308. doi:10.3390/rs9040308.
URL http://www.mdpi.com/2072-4292/9/4/308

[65] J. Albetis, A. Jacquin, M. Goulard, H. Poilvé, J. Rousseau, H. Clenet, G. Dedieu, S. Duthoit, On the potentiality of UAV multispectral imagery to detect Flavescence dorée and Grapevine Trunk Diseases, Remote Sensing 11 (1) (2019) 0–26. doi:10.3390/rs11010023.

[66] H. Al-Saddik, J. C. Simon, F. Cointault, Development of spectral disease indices for flavescence dorée' grapevine disease identification, Sensors (Switzerland) 17 (12). doi:10.3390/s17122772.

[67] H. Al-Saddik, A. Laybros, B. Billiot, F. Cointault, Using image texture and spectral reflectance analysis to detect Yellowness and Esca in grapevines at leaf-level, Remote Sensing 10 (4). doi:10.3390/rs10040618.

[68] H. Al-saddik, Assessment of the optimal spectral bands for designing a sensor for vineyard disease detection : the case of Flavescence dorée ', Precision Agriculture (0123456789). doi:10.1007/s11119-018-9594-1.
URL https://doi.org/10.1007/s11119-018-9594-1

[69] F. Rançon, L. Bombrun, B. Keresztes, C. Germain, Comparison of SIFT encoded and deep learning features for the classification and detection of esca disease in Bordeaux vineyards, Remote Sensing 11 (1) (2019) 1–26. doi:10.3390/rs11010001.

[70] R. Szeliski, Image Alignment and Stitching: A Tutorial, Springer US, Boston, MA, 2006. doi:10.1007/0-387-28831-7.
URL http://link.springer.com/10.1007/0-387-28831-7

[71] P. F. Alcantarilla, J. Nuevo, A. Bartoli, Fast explicit diffusion for accelerated features in nonlinear scale spaces, BMVC 2013 - Electronic Proceedings of the British Machine Vision Conference 2013doi:10.5244/C.27.13.

[72] D. G. Lowe, Distinctive Image Features from Scale-Invariant Keypoints, International Journal of Computer Vision 60 (2) (2004) 91–110. doi:10.1023/B:VISI.0000029664.99615.94.
URL http://link.springer.com/10.1023/B:VISI.0000029664.99615.94

[73] M. Jakab, Planar object recognition using local descriptor based on histogram of intensity patches, Proceedings of the 17th Central European Seminar on ....
URL http://www.cescg.org/CESCG-2013/papers/Jakab-Planar{_}Object{_}Recognition{_}Using{_}Local{_}Descriptor{_}Based{_}On{_}Histogra
pdf

[74] M. Hasenbusch, A. Pelissetto, E. Vicari, The critical behavior of 3D Ising spin glass models: Universality and scaling corrections, Journal of Statistical Mechanics: Theory and Experiment 2008 (2) (2008) 2564–2571. doi:10.1088/1742-5468/2008/02/L02001.

[75] A. B. Pablo Fernández Alcantarilla, A. J. Davison, LNCS 7577 - KAZE Features (2012) 1–14.
URL papers3://publication/uuid/4EFB9AB0-0CA9-4814-9A06-8590445894E7

[76] S. Grewenig, J. Weickert, A. Bruhn, From box filtering to fast explicit diffusion, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 6376 LNCS (2010) 533–542. doi:10.1007/978-3-642-15986-2_54.

[77] X. Yang, K. T. T. Cheng, Local difference binary for ultrafast and distinctive feature description, IEEE Transactions on Pattern Analysis and Machine Intelligence 36 (1) (2014) 188–194. doi:10.1109/TPAMI.2013.150.

[78] M. a. Fischler, R. C. Bolles, Paradigm for Model, Communications of the ACM 24 (6) (1981) 381–395. doi:10.1145/358669.358692.

[79] K. Ross, Geopositional Statistical Methods Lockheed Martin Integrated Systems & Solutionsdoi:https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20080021615.pdf.
URL https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20080021615.pdf

[80] R. Dellana, K. Roy, Data augmentation in CNN-based periocular authentication, Proceedings of the 6th International Conference on Information Communication and Management, ICICM 2016 (2016) 141–145doi:10.1109/INFOCOMAN.2016.7784231.

[81] S. A. K. Tareen, Z. Saleem, A comparative analysis of SIFT, SURF, KAZE, AKAZE, ORB, and BRISK, 2018 International Conference on Computing, Mathematics and Engineering Technologies: Invent, Innovate and Integrate for Socioeconomic Development, iCoMET 2018 - Proceedings 2018-January (2018) 1–10. doi:10.1109/ICOMET.2018.8346440.

[82] J. Jiang, C. Lyu, S. Liu, Y. He, X. Hao, RWSNet: a semantic segmentation network based on SegNet combined with random walk for remote sensing,

International Journal of Remote Sensing 00 (00) (2019) 1–19. doi:10.1080/01431161.2019.1643937.
URL https://doi.org/10.1080/01431161.2019.1643937

[83] D. Marmanis, K. Schindler, J. D. Wegner, S. Galliani, M. Datcu, U. Stilla, Classification with an edge: Improving semantic image segmentation with boundary detection, ISPRS Journal of Photogrammetry and Remote Sensing 135 (2018) 158–172. doi:10.1016/j.isprsjprs.2017.11.009.
URL https://doi.org/10.1016/j.isprsjprs.2017.11.009

[84] D. Chai, S. Newsam, H. K. Zhang, Y. Qiu, J. Huang, Cloud and cloud shadow detection in Landsat imagery based on deep convolutional neural networks, Remote Sensing of Environment 225 (September 2018) (2019) 307–316. doi:10.1016/j.rse.2019.03.007.
URL https://doi.org/10.1016/j.rse.2019.03.007

[85] A. Milioto, P. Lottes, C. Stachniss, Real-Time Semantic Segmentation of Crop and Weed for Precision Agriculture Robots Leveraging Background Knowledge in CNNs, Proceedings - IEEE International Conference on Robotics and Automation (2018) 2229–2235arXiv:arXiv:1709.06764v2, doi:10.1109/ICRA.2018.8460962.

[86] X. Ma, X. Deng, L. Qi, Y. Jiang, H. Li, Y. Wang, X. Xing, Fully convolutional network for rice seedling and weed image segmentation at the seedling stage in paddy fields, PLoS ONE 14 (4) (2019) 1–13. doi:10.1371/journal.pone.0215676.