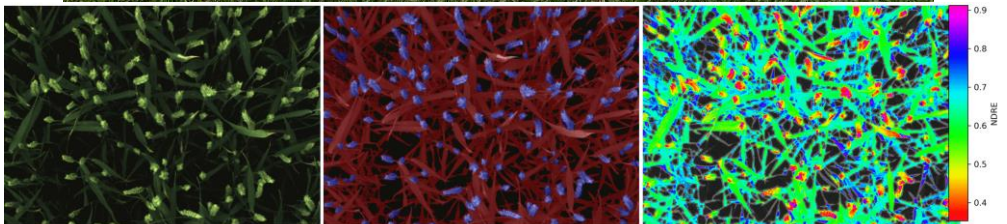# Dynamics of wheat organs by close-range multimodal machine vision

Sébastien Dandrifosse

COMMUNAUTÉ FRANÇAISE DE BELGIQUE
UNIVERSITÉ DE LIÈGE – GEMBLOUX AGRO-BIO TECH

DYNAMICS OF WHEAT ORGANS BY CLOSE-RANGE MULTIMODAL MACHINE VISION

Sébastien DANDRIFOSSE

Dissertation originale présentée en vue de l'obtention du grade de docteur en sciences agronomiques et ingénierie biologique

Promoteurs : Benoît MERCATORIS et Benjamin DUMONT

Année civile : 2022

Abstract

High-throughput plant phenotyping has a key role to play in crop improvement and crop adaptation to climate change. While it works routinely in greenhouses or controlled facilities, it is less advanced in the field. Instead of bringing potted-plants to the sensors, the sensors must be carried above the crop, dealing with the wind and the changing sunlight for image capture. It constraints the sensor types, the scanning possibilities and the choice of the view point. The measured signal comes from a whole cover instead of individuals. That signal is influenced by the background and does not provide traits separately for the organs (leaves, stems, ears). This research aimed to develop and evaluate imaging methods to characterise the wheat morphology and physiology at the organ scale, and study their dynamics in the field.

Nadir images were acquired in six field trials by two RGB cameras, a multispectral camera array and a thermal camera. The height of the wheat organs was computed by stereo vision from the RGB images. Using deep learning, the ears were counted with a F1 score of 0.93 and segmented with a F1 score of 0.86, then the plant height was used to compute the image footprint at ear height and measure the ear density. Regarding the multispectral images, the data from an incident light sensor was used to convert digital values to reflectance factor. To fuse the information from the multiple cameras, the RGB, thermal and multispectral images were registered using a two-steps approach: a global transformation based on the relative position of the sensors followed by a local transformation based on the mutual image content. The registration error was 2 mm. In the end, the image fusion provided the dynamics of reflectance factors, vegetation indices and 3D traits at the organ scale, as well as the ratios of organs, green surface and foliar damage in the images. Those traits were exploited to derive nitrogen, water or disease stress indices. PLS regressions estimated leaf area index, above-ground dry matter, above-ground nitrogen concentration, above-ground nitrogen content and nitrogen nutrition index respectively with external validation RMSE of 0.95 m²/m², 2.19 t/ha, 0.57 %N, 32.14 kgN/ha and 0.13. Grain yield was estimated with a cross-validation RMSE of 0.32 t/ha.

The main limitations of the system were the complexity to acquire proper data with all the sensors simultaneously, and the amount of processing steps and computing power necessary. While trying to improve the system operability, future works should also seek to generate interoperable data able to contribute to the global phenotyping efforts.

Résumé

Le phénotypage haut débit est un élément clé de l'amélioration culturale et l'adaptation des cultures aux changements climatiques et environnementaux. La discipline est cependant bien moins avancée au champ que dans les serres ou les installations contrôlées. Au lieu d'apporter des plantes en pot aux capteurs, ceux-ci doivent être transportés au-dessus de la culture, en tenant compte du vent et de l'ensoleillement changeant pour l'acquisition d'images. Cela contraint les types de capteurs, les possibilités de scan et le choix du point de vue. Le signal mesuré provient d'une canopée plutôt que d'individus ou d'organes isolés. Ce signal est influencé par le sol et ne fournit pas de traits distincts pour les organes (feuilles, tiges, épis, …). Cette étude visait à développer et évaluer des méthodes d'imagerie pour caractériser la morphologie et la physiologie du froment d'hiver à l'échelle de l'organe, et étudier leur dynamique au champ.

Les images nadir ont été acquises lors de six essais sur le terrain par deux caméras RGB, une matrice multi-spectrale de caméras et une caméra thermique. La hauteur des organes du froment a été calculée par vision stéréoscopique à partir des images RGB. Les épis ont été comptés par deep learning avec un F1 score de 0,93 et segmentés avec un F1 score de 0,86, puis la hauteur des plantes a été utilisée pour calculer l'empreinte de l'image à hauteur des épis et ainsi obtenir le nombre d'épis par mètre carré. Au niveau des images multi-spectrales, les données d'un capteur de lumière incidente ont été utilisées pour convertir les valeurs numériques en facteur de réflexion. Pour fusionner les informations des différentes caméras, les images RGB, thermiques et multi-spectrales ont été alignées en utilisant une approche en deux étapes : une transformation globale basée sur la position relative des capteurs, suivie d'une transformation locale basée sur le contenu mutuel des images. L'erreur d'alignement était de 2 mm. Au final, la fusion d'images a fourni la dynamique des facteurs de réflexion, des indices de végétation et des traits 3D pour les différents organes du froment, ainsi que la proportion des organes, de la surface verte et des dommages foliaires dans les images. Ces traits ont été exploités pour dériver des indices de stress azotés, hydriques ou de maladie. Des régressions PLS ont estimé l'indice de surface foliaire, la matière sèche aérienne, la concentration d'azote aérien, la teneur en azote aérien et l'indice de nutrition azotée respectivement avec une RMSE en validation externe de 0,95 m²/m², 2,19 t/ha, 0,57 %N, 32,14 kgN /ha et 0,13. Le rendement en grains a été estimé avec une RMSE en validation croisée de 0,32 t/ha.

Les principales limites du système sont la complexité d'acquérir des données appropriées avec tous les capteurs en même temps, ainsi que la quantité d'étapes de traitement et la puissance de calcul nécessaires. Tout en essayant d'améliorer l'opérabilité du système, les travaux futurs devraient également chercher à générer des données interopérables capables de contribuer aux efforts mondiaux de phénotypage.

Remerciements

Enfin, je remercie profondément ma famille, ainsi que tous les amis qui n'ont cessé de me soutenir et m'encourager pendant ces quatre années, même bien avant, et qui à présent se tournent avec moi vers l'avenir.

Table of content

I

III

List of figures

List of tables

List of abbreviations

ABG   Above-ground

A-KAZE   Accelerated-KAZE – image registration algorithm

BRF   Bi-directional Reflectance Factor

CIREDE   Red edge chlorophyll index

CMOS   Complementary Metal-Oxide Semiconductor

CWSI   Crop Water Stress Index

DAS   Days After Sowing

DDTM   Distance Dependent Transformation Matrix

DN   Digital number

ECC   Enhanced Correlation Coefficient

ExR   Excess Red – Vegetation index

GDD   Growing Degree Day

GWHD   Global Wheat Head Detection (dataset)

ILS   Incident Light Sensor

IoU   Intersection over Union

HFOV   Horizontal Field Of View

LAI   Leaf Area Index

LiDAR   Light Detection And Ranging (device)

mAP   Mean Average Precision

NDRE   Normalised Difference Red Edge Index

NDVI   Normalised Difference Vegetation Index

NIR   Near-infrared

NMI   Normalised Mutual Information

NNI   Nitrogen Nutrition Index

ORB   Oriented FAST and rotated BRIEF – image registration algorithm

PLSR   Partial Least Squares Regression

RGB   Red Green Blue – the three channels of a colour image

RMSE   Root Mean Square Error

SfM	Structure from Motion
SGBM	Semi-Global Block Matching – stereo vision algorithm
SIFT	Scale Invariant Feature Transform – image registration algorithm
SURF	Speed Up Robust Features – image registration algorithm
SVM	Support Vector Machine
SWIR	Shortwave Infrared
TIR	Thermal Infrared
UAV	Unmanned Aerial Vehicle
VFOV	Vertical Field Of View
VIS	Visible light
WLS	Weighted Least Squares – used to designate an image filtering method

I

General introduction

1. High-throughput crop phenotyping

Crop phenotyping consists in observing plant traits, i.e. morphological, physiological or phenological features, in order to assess their health, study their response to environmental and agronomic factors, improve plant cultivars, … In a context of climate changes, reduction of the use of fertilisers and pesticides, and of constantly evolving crop diseases, it is more than ever important to monitor plant traits. Crop phenotyping is the key to adapt our management practices and to select varieties adapted to local conditions. Overall, this is however a tedious task. It necessitates measuring traits in, often dense, crop canopies and on numerous plant individuals. Historically, it has been performed by human hands and eyes, implicating that the measure was influenced by the operator. Most of the time, established quantitative phenotyping methods necessitate destroying the crop to sample plants (e.g. to measure biomass or to quantify nutrients in plant organs). Additionally, crop phenotyping concerns a wide diversity of organs at different scales (plant cover, leaf, tillers, roots, flowers, fruits, grains, …). Destructive measurements often prevent studying time series integrating all those organs. To summarise this problem, crop phenotyping is a complex and challenging discipline, implying to measure many different plant traits as objectively and as efficiently as possible, striving not to destroy the crop.

During the last decades, the discipline evolved to what is called high-throughput phenotyping. The development of sensors (cameras, spectrometers, …) and their remote control allows to measure plant traits in an automatic, objective, fast and non-destructive way. Sensors provide information that is not directly perceptible by human eyes such as the light reflected by the leaves or the plant temperature. The ever-increasing computing power of our machines and the development of data processing methods (image analysis, machine learning, deep learning, …) provide means to extract useful plant traits from sensor data. Moreover, the evolving data storage capacity and the new possibilities to share data (clouds, …) make possible the use of huge amounts of numerical data. For those reasons, high-throughput phenotyping is a booming research topic and its use in agronomic research areas is spreading all around the world.

High-throughput phenotyping is performed in controlled chambers, in greenhouses, or in the field. Depending on the phenotyping environment, different technologies and measurement pipelines are exploited. In the field, the scale of the measurement can

range from plant organs to a whole plot, depending if the sensors are carried by human operators, ground vehicles, aerial vehicles or satellites. Measurements at small scale provide great level of details about the crop but necessitate time and can be confounded by environment variations during the measurement timespan (Smith et al., 2021). Aerial measurements avoid these issues but may lack resolution to capture traits of interest. Moreover, they can be influenced by viewing geometry and weather, and fail to account for the environmental variations from one day to another. Whatever the scale, high-throughput phenotyping technologies have the potential to generate tsunamis of data, which needs to be recorded, managed, processed, standardised and centralised so that they can be converted to agronomic knowledge.

High-throughput phenotyping is not only about measuring crop traits but also about recording associated environmental and agronomic factors. A phenotype (P) is the expression of a genotype (G) for specific environmental (E) and crop management (M) conditions. Each phenotype may be expressed as the interaction: $P = G \times E \times M$. The $E \times M$ conditions are referred to as a scenario. Creating knowledge necessitates to associate the measured plant traits with a certain number of data and metadata of the scenario. In that problem, the G term can be provided at a high-throughput and low-cost thanks to the development of DNA sequencing technologies (Berkman et al., 2012). The P term, the phenotyping, is for its part repeatedly presented as the bottleneck for crop genetic improvement (Houle et al., 2010; Araus et al., 2018). The challenges of plant phenotyping are i) to acquire, process and store data about P, E and M and ii) to create knowledge from those data. This second step is particularly challenging because, once P and G have been characterised for a wide range of $E \times M$ scenarios, the question is to disentangle the P effects resulting from the $E \times M$ interaction. Only after this step, models can be created to simulate plant development (Tardieu et al., 2017).

Additionally, the exploration of the phenotyping plasticity of a genotype implies to study its expression regarding intra-site factors (fertilisation, presence of disease, irrigation, …) but also regarding inter-site factors (climate, soil, …). In other words, each scenario has a scene and actors. The scene is the site of the plot and the actors are the intra-site factors. To generate knowledge, the same actors should be playing the scenario on contrasted scenes. For those reasons, national and international networks of phenotyping facilities are being created and data standards are emerging (Pieruschka et al., 2019). The International Plant Phenotyping Network (IPPN) gathers public and private stakeholders of the plant phenotyping community. At the European level, EMPHASIS is an organisation aiming at developing multi-scale crop

phenotyping, supporting development of phenotyping infrastructures and increasing synergy between phenotyping facilities.

Given the current rate of high-throughput phenotyping development, the portrait presented here is likely to evolve quickly. This thesis is part of this evolution by bringing original elements regarding the integration of multiple close-range cameras in the field. It especially provides methods to combine their information with the aim to extract morphological and physiological plant traits up to the plant organ scale.

2. Frame of the thesis

Scientific research can rarely be general and this thesis is no exception to the rule. All the developments presented in this paper have been established in a specific frame that is important to delimit at this point. In one sentence, this frame could be formulated as: the proxy-sensing of above-ground winter wheat organs in a context of Belgian field trials. These elements are detailed in the following paragraphs.

The studied crop was winter wheat, *Tritium aestivum L.*, grown in field trials, i.e. plots divided into separated micro-plots to investigate the effects of agronomic management practices. Phenotyping in such field trials must be distinguished from phenotyping in controlled or semi-controlled conditions (greenhouses). In the field, the crops present some variability due to factors such as soil properties or topography. The natural conditions are also susceptible to influence image acquisition. Belgian springs are especially characterised by light conditions that change quickly due to the fast alternation of clouds and direct sunlight. It is also to note that the landscape is mainly made of open fields. The absence of hedges and trees causes an important exposure to wind, which impacts image capture. Besides, the context influences the studied $E \times M$ scenarios. For wheat, nitrogen is by far the most limiting nutrient and at the heart of the fertilisation practices. Regarding the environment, the wet climate associated with positive temperatures favours the development of fungal diseases. Fungi are a major threat and a key factor limiting yield. They may cause up to 70 % of yield loss in the event of serious infection on sensible varieties (Livre Blanc Céréales – website, available at: https://www.livre-blanc-cereales.be/). For the years 2014 to 2018, yield losses caused by diseases on unprotected plots were estimated on average to 17.2 % (Bodson, 2019). Conversely, drought is less problematic than in further south regions.

Our research focused on close-range imagery technologies to get a high level of detail regarding the plant organs and the canopy structure. The sensors were

embedded on a mobile platform that brought them to a few metres of the plants. Satellite images and aerial vehicles were not exploited. Additionally, the study was defined by a focus on low-cost custom material. The sensors had to be compact, affordable and allow fast acquisition in natural conditions. At the current maturity level of technologies, this excluded expensive and scanning-dependent devices such as line-scan hyperspectral imager or Light Detection and Ranging (LiDAR) scanners. The focus was set on cameras that were easy to integrate on a mobile platform and able to quickly acquire information.

3. Objectives and organisation of the thesis.

3.1 Issue

Due to the difficulties caused by the natural conditions, phenotyping in the field is less advanced than in greenhouses or indoor controlled conditions (Deery et al., 2021). Instead of bringing potted-plants to the sensors, these must be carried above the crop at high-throughput, dealing with the wind and the sunlight while acquiring data. It adds constraints regarding the types of sensors, the scanning possibilities and the choice of the view point. The measured signal comes from a whole canopy instead of particular individuals or plant organs. That signal is influenced by the background and, without further processing, do not allow to extract plant traits separately for the organs (stems, leaves, fruits, …). There is a need to account for the spatial information in the canopy. This could especially go along with the need of studies integrating multiple stress factors in the field. Most of the previous research was limited to experimental designs reproducing only one environmental or management factor. In these conditions, the variation of a high-throughput canopy trait was linked to a single factor, but such trait could very well be influenced by other factors, and even by the interaction of multiple E × M factors. To take a simple example, a more yellowish colour of the canopy can be due to nitrogen deficiency or the presence of a fungal disease. Trying to disentangle the effects of those factors on the plant phenotype may necessitate extracting plant traits at the organ scale, considering the light reflected from the healthy parts of the leaves separately from the disease spots.

Having an image of the canopy may not be sufficient to go to the organ scale. In the example of thermal imagery, the temperature map alone is not sufficient to isolate the temperature of the leaves, or of sunny parts of the leaves. Consequently, thermal cameras have to work along with other cameras. And even when getting leaf

temperature, or more interestingly stomatal conductance (Costa et al., 2013), using this information alone will not provide certainties about the biotic or abiotic nature of the stress, or the combination of both (Pineda et al., 2020).

The current challenge is the joint processing of multi-sensor data to extract an ensemble of standard and robust traits, from the canopy scale to the organ scale. This challenge is all the more difficult for wheat. The canopy is dense and complex, the leaves are thin and erected, and all this structure greatly varies with the development stage.

3.2 Goal and research question

Given a set of low-cost sensors – a multispectral camera array, a stereoscopic pair of RGB cameras, a thermal camera and an incident light spectrometer –, the goal of this research was to propose and evaluate methodologies to characterise the wheat morphology and physiology at the organ scale, and study their dynamics in the field throughout the growing season, from tillering to maturity.

The research question can be formulated as follows: how can a close-range multi-sensor vision system enrich and standardise the morphological and physiological characterisation of a wheat crop?

3.3 Structure of the thesis

The structure of the thesis is summarised in **Figure 1**.

Chapter II presents a state of the art of the imaging techniques used at close-range for the phenotyping of cereal crops in field conditions. It develops the advantages and disadvantages of each sensor, and why their combination has such a potential to unlock robust and standard plant trait extraction methods.

Chapter III briefly introduces winter wheat dynamics before detailing the field trials that were exploited for this research, from 2018 to 2021. It encompasses the description of all the reference manual measurements and observations, the collection of environmental data, and the protocols of image acquisition, including the presentation of the phenotyping platforms and of the sensors used.

Chapter IV is dedicated to the automatic measurement of plant height. It explains how two RGB cameras can be exploited in a process called stereo vision to derive height maps of the crop. Those height maps can be processed to extract a variety of morphological wheat traits. Moreover, the distance between the cameras and the plant organs is a prerequisite to the fusion of the images from the multi-sensor system.

Chapter V tackles the challenge of wheat ear detection. While the plants and the background were not too complicated to separate using near-infrared (NIR) channels, a robust segmentation of the upper organs – the ears and the leaves – was more difficult to obtain, considering their range of variations through the development stages of the crop. It is explained how a deep learning solution was used to solve the segmentation issue, and how deep learning ear detection was combined with the estimation of the image footprint by stereo vision (from Chapter IV) to automatically measure the ear density, which is one of the most important agronomic parameters.

In **Chapter VI**, several multimodal image registration methods are investigated to align pixel to pixel the images from the different cameras. For the first time, an automatic solution is proposed to this problem in the context of close-range RGB, multispectral and thermal wheat images. This is especially possible thanks to the knowledge of the distance between the wheat and the camera, as computed in Chapter IV.

Chapter VII is dedicated to the conversion of the pixel digital numbers of the multispectral images to values of bi-directional reflectance factor, a standard measure to account for the proportion of light reflected by the plant organs, in relation with their inclination and physiology. An incident light spectrometer was exploited to avoid the need of a bulky reference panel in the captured scenes and increase the throughput of the data acquisitions. And this is where all the pieces of the puzzle of the multi-sensor fusion come together: the mask with the organ separation (Chapter V), registered with all the multispectral reflectance maps (Chapter VI), can be used to extract the reflectance of the organs separately in each spectral channel.

Chapter VIII demonstrates how all the plant traits extracted from the multi-sensor system help to generate useful agronomic information. It gathers exploratory analyses regarding the dynamics of wheat morphology (leaf area, dry matter) and physiology (vegetation indices, nitrogen status), the biotic and abiotic stresses, and the yield estimation.

Finally, **Chapter IX** sums up the findings, highlights the limitations of the work, and takes a step back to position it in the global context of agronomic innovation.
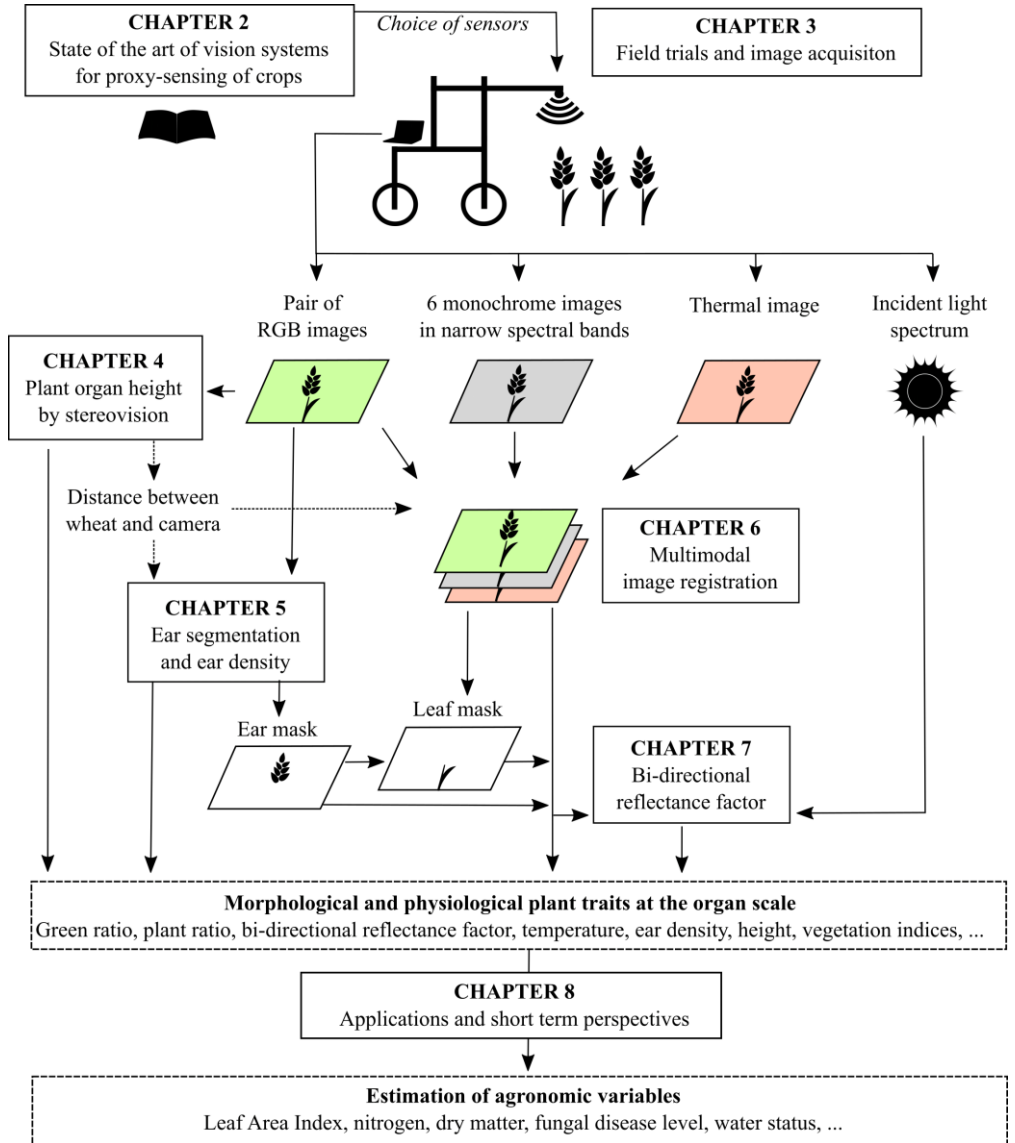
Figure 1. Structure of the thesis.

3.4 Publications associated with the thesis

The chapters presenting the main results of this thesis – Chapter IV, Chapter V, Chapter VI and Chapter VII – are associated with peer-reviewed international scientific publications.

- **Chapter IV.** Dandrifosse S., Bouvry A., Leemans V., Dumont B. & Mercatoris B., 2020. Imaging wheat canopy through stereo vision : overcoming the challenges of the laboratory to field transition for morphological features extraction. *Front. Plant Sci*. **11**(February), 1–15.
- **Chapter V.** Dandrifosse S., Ennadifi E., Carlier A., Gosselin B., Dumont B. & Mercatoris B., 2022. Deep learning for wheat ear segmentation and ear density measurement : From heading to maturity. *Comput. Electron. Agric.* **199**(June).
- **Chapter VI**. Dandrifosse S., Carlier A., Dumont B. & Mercatoris B., 2021. Registration and Fusion of Close-Range Multimodal Wheat Images in Field Conditions. *Remote Sens.* **13**(7), 1380.
- **Chapter VII.** Dandrifosse S., Carlier A., Dumont B. & Mercatoris B., 2022. In-Field Wheat Reflectance: How to Reach the Organ Scale? *Sensors* **22**(9), 3342.

II

State of the art of vision systems for proxy-sensing of crops

1. Synopsis

This chapter aims to present a state of the art of the machine vision systems used for the proxy-sensing of crops. By vision, it is understood that the measured signal is an image, i.e. a representation of the crop in a two, or more, dimensional plane based on the spatialized measurement of electromagnetic radiations coming from the scene. Any device capable of recording such an image is called a camera. Several types of cameras can be distinguished, depending especially on the spectral region of measurement (**Figure 2**).
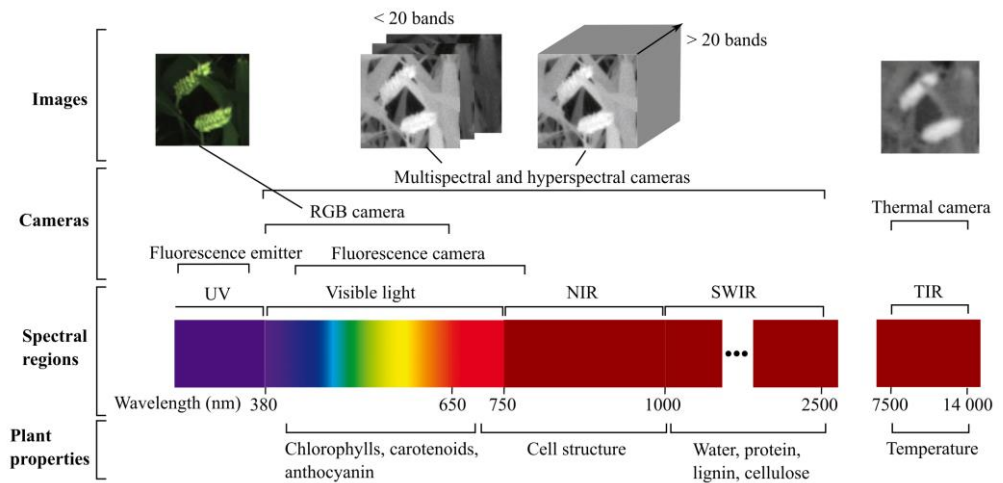
Figure 2. Regions of electromagnetic spectrum exploited for crop image acquisition. NIR stands for Near-Infrared, SWIR for Shortwave Infrared, and TIR for Thermal Infrared. For each region, it is indicated which properties or constituents of the plants drive their spectral signature (Pauli et al., 2016; Mishra et al., 2017).

This state of the art reviews the use of RGB, multispectral, hyperspectral and thermal camera for crop phenotyping, with a focus on winter wheat. The goal is not to go deep in the details of how each technology works, but rather to provide an overview of the plant traits they allow to measure, their limitations and the opportunities of sensor combination to face these individual limitations. The review focuses on imaging in field conditions. It means that the measured spectral information not only depends on the plant properties highlighted in **Figure 2**, as it could be for isolated leaves in the laboratory, but also on environmental factors, canopy architecture and sensor position (Pauli et al., 2016).

This state of the art does not include fluorescence imaging due to the complexity to implement it in field conditions and its limited throughput (Pauli et al., 2016; Deery et al., 2021). The state of the art does also not contain a comparative approach with non-imaging methods (LiDAR, ultrasonic sensor, spectrometer, …) because several very good and recent reviews already exist (Li et al., 2014; Vázquez-arellano et al., 2016; Qiu et al., 2018; Roitsch et al., 2019; Deery et al., 2021). Although great benefit can be obtained from the fusion of LiDAR and imagery (Roitsch et al., 2019), this method remains hard to implement in the field and was not studied in this research. Therefore, it falls outside the scope of this thesis.

2. RGB vision

Colour cameras, or RGB cameras, are probably the most widely used technology to characterise crops, at close-range but also massively from unmanned aerial vehicles (UAV). The sensor, a matrix of photosensitive elements, is covered with a matrix of red, green and blue filters. They let the light through in these three wide spectral bands: around 400-550 nm, 450-600 nm and 550-650 nm. Beyond 650 nm, an infrared-cut filter prevents the light from reaching the sensor. These cameras are low-cost and widely available. They are the ones we use in everyday life, in our cameras or in our smartphones. Even the consumer-grade devices possess a wide image size and an excellent resolution. Moreover, they record the light in the visible spectrum and reproduce the colours in a way similar to the human eyes. It allows operators to easily interpret the images and label the features of interest (Deery et al., 2021).

An important trait they allow to measure is the canopy cover, i.e. the proportion of vegetation in the image. This widely used trait, however, should be better defined. We suggest using the terms green ratio and plant ratio. The plant ratio is the proportion of all the plant parts in the image while the green ratio is only the proportion of the green parts of the plants. Those traits are commonly measured using cameras in nadir position. However, there could be some interest in tilting the camera. Using a zenith angle of 57.5° would improve the estimation of Green Area Index (Baret et al., 2010).

Colour analysis easily provides the green ratio but RGB colours alone cannot provide an accurate measure of the plant ratio in all the circumstances. The images are likely to contain yellowish or brownish leaf parts that sometimes present the exact same colours as the soil or the wheat ears. An interesting insight into colour analysis was written by Golzarian et al. (2012). The three RGB components can undergo several simple mathematical operations, called vegetation indices, to enhance the

vegetation identification (Meyer et al., 1998). The vegetation indices often require converting the RGB colour space to the normalised RGB colour space, for which the digital value of each component is divided by the sum of the values of the three components. Additionally, it is possible to convert the RGB colour space to other colour spaces such as the frequently used Hue-Saturation-Value (HSV) (Hamuda et al., 2017), Hue-Saturation-Intensity (HSI) or Lab colour spaces (Wang, Zhang, et al., 2019). Then, the vegetation segmentation can be performed by a simple thresholding approach or by machine learning techniques, taking as input a more elaborated set of features. For green ratio measurement, a colour threshold is sufficient (Kipp, Mistele, Baresel, et al., 2014). No need to use a bazooka to kill a fly. To automatically find a proper value for the threshold the Otsu's algorithm can be used (Otsu, 1979). For plant ratio, it may be necessary to add textural features and employ machine learning, especially to account for complex illumination conditions (Hamuda et al., 2016). It is also to note that interesting traits derive from the dynamics of such ratios: the green ratio can especially be used to assess early plant vigour (Kipp, Mistele, Baresel, et al., 2014) or senescence (Walter et al., 2019).

RGB cameras were used to measure leaf nitrogen or chlorophyll content. The estimation could be based on segmented leaf colour (Tewari et al., 2013; Baresel et al., 2017; Tavakoli et al., 2019), on canopy colour (Jia et al., 2004, 2009) or on canopy cover (Prey et al., 2018; Tavakoli et al., 2019). Such features from the RGB images were also correlated with biomass (Prey et al., 2018) and even water content (Tavakoli et al., 2019). Looking at these results, it may appear that RGB cameras are a powerful tool to extract some of the most important agronomic traits. One could however question the generalisation and the robustness of these correlations, established for a few dates and plots, and in very specific scenarios and varieties. Given the limitations of the RGB cameras – only three wide spectral bands and the incapacity to deal with 3D information –, it appears not likely that a RGB camera alone can quantify in a robust way traits such as biomass, water content or nitrogen status.

A realistic expectation is the measurement of all the traits based on the detection of specific features on the plants or in the canopy. RGB cameras especially offer the possibility to identify and quantify some diseases. On isolated leaves, Xu et al. (2017) graded wheat leaf rust. In the field, Jay et al. (2020) identified cercospora spots on sugar beet leaves and Walter et al. (2019) evaluated the presence of septoria tritici blotch on wheat. Both approaches were based solely on a colour analysis and could probably not be generalised for all background conditions and the presence of other elements such as other diseases, marks of abiotic stress or wheat ears. At this level,

the recent deep learning techniques based on convolutional neural networks offer new opportunities to obtain robust disease identification. Such approaches were developed to detect fusarium head blight (Su et al., 2021) and stripe rust (Schirrmann et al., 2021). The challenge will probably be greater for septoria tritici blotch because this disease first appears on the lower leaves, not visible in the image, before climbing in the canopy through rain-splash (Suffert et al., 2011). But the disease spots are not the only features that can be detected in the RGB images. Sadeghi-Tehran et al. (2017) worked on the detection of the heading and flowering development stages based on the presence of emerged ears and anthers in the images. The number of ears per square metre, called ear density, is also a trait of great importance. A full chapter of this thesis is dedicated to the processing of RGB images to segment the ears and estimate the ear density. The reader should refer to this part (Chapter V) for an exhaustive state of the art.

Finally, it is worth mentioning that some researchers developed modified RGB cameras (Rabatel et al., 2011; Dworak et al., 2013). The idea was to overcome their spectral limitations by replacing one or two of the channels by a NIR channel. One of the main interests of the NIR is to easily segment the vegetation. This can be useful to compute the plant ratio. Such a modified camera was used by Grieder et al. (2015) to follow the canopy cover of wheat during winter.

3. Stereo vision

Stereopsis is the perception of depth and 3D information from the shift in the position of objects between two points of view. This binocular perception mechanism has been proved to be deployed by humans and various animals such as the owl, the falcon, the horse, the toad or the praying mantis (Nityananda et al., 2017). By using two cameras horizontally spaced at eye-to-eye width, it is possible to exploit that kind of vision for the machines. This is called stereo vision. Most of the time, RGB cameras are selected, because of their low-cost and availability, but stereo vision algorithms require only one channel, so monochrome cameras are well suited for it as well. Explanations on how to implement stereo vision for crop phenotyping are provided in the Chapter IV of this thesis. This vision method can provide 3D information about the crop at a low cost. Briefly, it is important to mention that stereo vision can be passive or active. Passive methods rely only on the ambient light to illuminate the scene. They can be easily implemented with any set of two similar cameras. Active methods use the projection of a light pattern on the scene to ease the search of matches

between the two images (O'Riordan et al., 2019). They generally face difficulties under natural light conditions.

Plant phenotyping based on stereo vision has been mainly developed in the laboratory (He et al., 2003; Andersen et al., 2005; Biskup et al., 2007; Lin et al., 2011; Tilneac et al., 2012). Only few in-field approaches have been established. Ivanov et al. (2003) applied stereo vision to study leaf angle and leaf area in maize. Kise et al. (2008) used stereo vision for crop row detection. Müller-Linow et al. (2015) tested the method on sugar beet in field conditions. Salas Fernandez et al. (2017) exploited stereo cameras to acquire side views of tall crops. Xiang et al. (2020) used close-range stereo vision to measure the diameter of sorghum stems. For wheat and barley, stereo vision is challenging because of homogeneous leaf texture and complex canopy structure made of thin and long overlapping leaves. Leemans et al. (2013) introduced a method for wheat LAI and leaf angle computation.

Alternatives to binocular stereo vision are interesting to mention. Increasing the number of cameras could result in more accurate distance estimates and less visual occlusions (Kaczmarek, 2015, 2017). These systems, called multi-camera, feature multiple cameras which use roughly the same point of view. They must be distinguished from the multi-view systems, where one or several cameras acquire images all around an object to reconstruct in 3D. Such multi-view configurations are well suited for indoor phenotyping but difficult to reproduce in the field, unless using a vehicle equipped with a robotic arm such as the one of Shafiekhani et al. (2017), used to reconstruct 3D model of maize and sorghum plants in a field and measure traits such as height and LAI. Another alternative to stereo vision is structure from motion (SfM): the different images are acquired using the displacement of a single camera. This is the classical approach to obtain 3D information using the series of overlapping images acquired during a UAV flight. It was exploited at close-range in the field by Jay et al. (2014) to estimate leaf area for several vegetable species.

4. Multispectral and hyperspectral visions

The spectrum of the light reflected by the crop responds not only to change in illumination but also to changes in canopy architecture and composition of plants in water, pigments and other plant constituents. Depending on the crop property to assess, some wavelengths are of particular interest. It is also relevant to isolate the light from the canopy parts of interest, hence the use of spectral cameras, able to couple the spatial and the spectral characterisation of the crop.

Multispectral and hyperspectral cameras provide images in different wavelength bands, usually ranging in the visible (VIS) from 380 to 700 nm, the near infrared (NIR) from 700 to 1100 nm and the Shortwave Infrared (SWIR) from 1100 to 2500 nm. Multispectral vision consists of a limited number of spaced spectral bands (< 20). At the opposite, hyperspectral vision exploits a high number of narrow bands covering the spectrum in a continuous way. Both kinds of cameras are more adapted to compute vegetation indices than the RGB cameras, thanks to their higher number of spectral bands and the possibility to exploit narrow bands that can be used to extract specific physiological or biochemical traits (Deery et al., 2021). Overall, they however provide images with a lower size than the RGB images. It is also to be specified that the same multispectral or hyperspectral camera usually cover either the VIS-NIR or the SWIR regions, as different sensor types are required.

In this state of the art, the hyperspectral and multispectral cameras are classified according to the main image acquisition methods because these methods greatly impact the way cameras are implemented in the field and plant traits are extracted. Those main methods are point scanning, line scanning, spectral scanning, snapshot and camera array. Other more exotic acquisition approaches exist (Genser et al., 2020), but they are marginal and, to our knowledge, not encountered in the field of plant sciences.

Point scanning and line scanning methods belong to the field of hyperspectral imagery. In those methods, the light is split into spectral components through a spectrograph and projected on an optical sensor. Each spectral band intensity is associated with a position. To associate each spectral band with each position, a relative movement between the scene and the sensor is necessary. That movement is performed in one direction (line scanning) or two directions (point scanning). Those methods are costly (Mishra et al., 2017), and well adapted for indoor applications. In field conditions, the acquisition is more challenging because of the scanning time. Images may be impacted by the wind and the illumination changes (Behmann et al., 2018) although some acquisition platforms or rovers allow controlled conditions (Busemeyer et al., 2013). Line scanning hyperspectral cameras have been implemented in the field to estimate the canopy cover (Busemeyer et al., 2013), measure the leaf nitrogen content (Vigneau et al., 2011), detect and quantify diseases (Lowe et al., 2017; Behmann et al., 2018; Whetton et al., 2018; Cheshkova, 2022), estimate photosynthetic variables such as the maximum rate of carboxylation and the maximum rate of electron transport (Fu et al., 2020), and classify grapevine varieties (Gutiérrez et al., 2018).

Spectral scanning acquires the 2D spatial information one spectral band at the time. Thus, the scan is performed on the spectral bands instead of the spatial location. This method does not necessitate a relative movement between the scene and the sensor. It can be performed by hyperspectral tunable filters, i.e. filters whose properties change to allow different spectral bands to pass (Genser et al., 2020). A low cost multispectral alternative is the use of a filter wheel: a wheel of filters that revolve in front of an optical sensor. However, wheels are bulky, the filters turn-over is slower than for tunable filters and the number of spectral bands is limited, although filters can easily be changed. Such filter wheels were used in field conditions to estimate wheat leaf nitrogen concentration (Leemans et al., 2017) or detect wheat diseases (Bebronne et al., 2020). Both wheels and tunable filters present the same drawbacks that line scanning cameras regarding wind and illumination changes.

The snapshot multispectral cameras are able to acquire spatial and spectral information at the same time. They rely on a multispectral filter array attached to a unique sensor. Filters as disposed on a mosaic that can be seen as an extension of a Bayer matrix. Thus, the number of spectral bands is limited and the addition of filters comes at the expense of spatial resolution (Genser et al., 2020). That loss of spatial resolution may become limiting when imaging thin leaves. There is also an issue with filter patterns located at leaf edges for which some filters may receive light from the background and others from the leaves. Filters of such cameras cannot be changed. The method has been employed in greenhouse to study the reflectance of sugar beet in response to several stresses (Khanna et al., 2019).

The multispectral camera arrays are matrices of cameras, typically between four and nine cameras, each equipped with a different pass-band optical filter. Most multispectral camera arrays include bands in the 380 – 1000 nm interval, which is the range of sensibility of conventional digital imaging sensors. Those arrays are able to acquire both the spatial and the spectral information at the same time. There is no degradation of the spatial information as for the so-called snapshot cameras. However, the spatial information is limited to the scene that is commonly observed by the cameras of the array. It is also to note that the number of spectral channels is limited, and that each new channel necessitates buying a filter and a camera, which can quickly increase the cost and the size of the system. A number of companies developed multispectral camera arrays: the Tetracam ADC Snap cameras (Tetracam, Inc., Gainesville, FL, USA), the MAIA camera (SAL Engineering, Russi, Italy), the Sequoia camera (Parrot SA, Paris, France), the Airphen camera (Hiphen, Avignon, France) and the MicaSense RedEdge-M (MicaSense, Inc., Seattle, WA, USA)

(Tripodi et al., 2018; Kim, 2020; Deery et al., 2021). Those devices were conceived for UAV surveys. We found no tracks of attempts of the use of such multispectral arrays for proxy-sensing, probably because of the difficulty to align the images at close-range. Nevertheless, the task is possible and a solution is described in Chapter VI of this thesis. Camera arrays offer the advantage to quickly acquire images in difficult field conditions, especially fast changing ambient light. Such systems may be relatively low-cost and easy to install on a phenotyping platform compared to hyperspectral scanning cameras. It should be preferred over multispectral snapshot cameras in case the spatial resolution is an important criterion. It also offers more customization possibilities than commercial snapshot cameras. Nevertheless, attention should be paid to image synchronisation to remove the effect of wind, for example by relying on a controller to send a trigger to all the cameras.

Regarding the extraction of plant properties, multispectral and hyperspectral cameras often rely on vegetation indices. These indices can be defined as transformations of two or more spectral bands that aim to enhance the spectral data to retrieve vegetation properties. Although some vegetation indices can be built using the three channels of a RGB camera (Section 2), most of the indices necessitates more narrow bands in specific regions of the spectrum. Vegetation indices have been used extensively to estimate morphological wheat traits such as LAI (Ajayi et al., 2016), or physiological traits such as nitrogen (Cammarano et al., 2014), light-use efficiency (Tayade et al., 2022) or fungal diseases (Ashourloo et al., 2014; Wan et al., 2022). They can be determined from close-range images, spectroscopy data, aerial or satellite images. Many of them exist, as demonstrated by the paper of Xue et al. (2017) that reviews more than hundred vegetation indices. A single index is sometimes used to derive a wide variety of traits for a large number of plant species and at measurement scales ranging from close-range to space-born. It happens that the same transformation of the spectral bands has been named differently by different authors. It also happens that various transformations are associated with the same name of vegetation index, especially using bands of slightly different width and/or central wavelength. From all of this, a great confusion arises… The users should be aware of the original purpose of each vegetation index, although many other purposes may appear. A particular attention should be paid to the scale for which the index was designed. The easiest way to develop a new vegetation index and relate it to a physiological property of the crop is to work in a laboratory at the scale of a single leaf. Indices developed at this scale could not be suitable for measurement at the scale of the canopy, where the signal is influenced by sun-sensor geometry, leaf angle, LAI and background, that

change greatly during the crop development (Cammarano et al., 2014). There are also indices specifically developed for satellite applications, that integrate atmospheric corrections useless if the index is used at a closer range.

A challenge with spectral imaging, and particularly hyperspectral methods, can be the huge number of vegetation indices or spectral bands to analyse (Benelli et al., 2020; Sarić et al., 2022). This requires the help of advanced multivariate analysis techniques and machine learning algorithms.

5. Thermal vision

Each object at a temperature above 0 K emits electromagnetic radiation depending on its temperature. At temperatures on the Earth surface, objects emit in the infrared (IR) domain. Thermal cameras provide images of those radiations in the 8-14 μm range. In comparison to optical cameras, they are more expensive and present smaller image sizes (e.g. in 2020, a thermal camera of 640 x 480 pixels cost 7000 to 20 000 euros). In agriculture, thermal cameras allow measurements of canopy temperature. However, that temperature is influenced by a number of environmental factors: air temperature, air humidity, wind speed and solar irradiance. In field conditions, those factors are likely to change rapidly. Additionally, apparent temperature may also be influenced by canopy architecture (overlapping leaves and variations in leaf angle). As a result, the canopy temperature by itself is not an interesting measurement. Most past research focused on converting canopy temperature to stomatal conductance or stress index. This state of the art is divided into four paragraphs: i) issues regarding the measurement of canopy temperature, ii) conversion of temperature to water stress indices, iii) conversion of temperature to stomatal conductance and iv) other agronomic information derived from canopy temperature.

In the seventies, non-contact thermal measurements from infrared thermometers were used to assess canopy temperature. Those tools provided punctual measurements, not images. A main issue was the effect of gaps in the plant cover, so that soil radiance influenced the measurement (Kimes, 1980). Fuchs (1990) concluded the need to collect digital images of the crop temperature. Tracks of the use of a thermal camera to study leaves date back to the indoor experiment of Hashimoto et al. (1984) but plant researches involving thermal only increased in the 2000s (Costa et al., 2013). Jones (2002) highlighted that the development of thermal cameras allowed to get not only an average temperature of ground and canopy but also to isolate leaf temperature and to obtain a frequency distribution of those temperatures.

The automatic segmentation of leaves from the background was still an issue. Applying an arbitrary temperature threshold on the images was not a satisfying solution. Jones (2002) studied the use of dry and wet reference paper to threshold the image and isolate the leaves between the two reference limits. However, from a high-throughput phenotyping perspective, placing references in the imaged scenes is not a practical solution.

Numerous efforts have been made to derive water stress information from non-contact canopy temperature measurements. Those researches started before the use of thermal cameras, while average canopy temperature was still recorded through infrared thermometers. Idso et al. (1977) developed the stress-degree-day concept, exploiting the difference between canopy and air temperatures as a stress index. That approach has the advantage of being simple to implement. The effect of environmental factors (wind, air humidity and net radiation) is supposed to be reflected in the temperature difference. Idso et al. (1981) and Jackson et al. (1981) introduced the Crop Water Stress Index (CWSI) to add the influence of the air humidity. This index necessitates establishing a "non-stressed" baseline by associating temperature and humidity measurements. The CWSI was adapted by Jones (1999) by adding temperature of wet and dry references imaged in the same environmental conditions as the crop. This modified CWSI does not need a baseline and accounts for all environmental effects. The CWSI has been widely used until now (Costa et al., 2013; Khanal et al., 2017). This approach is based on the average temperature of leaves. Fuchs (1990) stated that the distribution of leaf temperatures is a better indicator of water stress than this mean value because the stress-induced temperature rise only concerns a small fraction of the total area. This interesting track should be considered taking into account that sunlit leaves present higher variability than shaded leaves (Jones, 2002). At the opposite of this idea, Grant et al. (2007) found that variation of temperature within the canopy (side view of grapevines) was not a reliable stress predictor. At the plot scale, Clawson et al. (1982) suggested that water stress may be detected by looking at the variability between average temperatures of several zones. That spatial response to a water stress could be due to the non-homogeneity of soil properties.

Alongside that stress-based approach, researches were carried out to measure the stomatal conductance. The first obvious approach is to study the relationship between the stress indices and the conductance. To directly obtain the stomatal conductance from the leaf temperature, there is a need to measure wind speed, net absorbed radiation, air temperature and air humidity at the same time. Stomatal conductance is

then computed from the well-known Penman-Monteith equation of energy balance (Monteith, 1973) in which wind speed is used to estimate resistance to water vapour and heat transfer (Thom et al., 1977). A dry reference can be used to avoid the need to measure net absorbed radiation. A wet and a dry reference can be used to avoid measuring both net absorbed radiation and air humidity (Jones, 1999). Leinonen et al. (2006) stated that the preferred method to measure stomatal conductance should be the use of a dry reference because high uncertainties come from absorbed radiation due to the impact of leaf orientation. Measuring all the terms of the energy balance equation is also a valuable method although estimation of conductance could be less accurate.

Canopy temperature could also be related to other stresses or agronomic parameters. It could be a more suitable tool than optical remote sensing to detect stresses prior to the visual symptoms (Khanal et al., 2017). An increase of temperature could early indicate the presence of a soil-borne disease affecting water supply (Prashar et al., 2014). Overall, modifications of canopy temperature can be related to a diversity of stresses but by observing a change of temperature alone we cannot conclude on the nature of the stress. The interaction between stresses also impacts the thermal response (Stoll et al., 2008). In the light of those studies considering different stresses, we may even doubt that the proposed methods to detect water stress cannot confuse the need of water with other stresses. In terms of agronomic parameters, Banerjee et al. (2018) demonstrated that thermal images allow a more accurate estimate of wheat canopy cover than RGB images, especially because thermal images get rid of shadows and illumination issues.

To conclude, it appears that the segmentation of leaves and ground in proxy-sensing thermal images is not always as clear as in the study of Banerjee et al. (2018). Thermal imaging allows to detect stresses but the temperature information alone is not sufficient to return a verdict on the stress nature. Registration of thermal images with RGB and multispectral images has the potential to solve both issues. The additional spectral information would help to segment leaves from the background and bring new inputs to determine the stress nature and improve its quantification. Additionally, there is a need to consider canopy structure – shaded leaves, sunlit leaves, leaf angle and foliar floors – when analysing thermal images (Jerbi et al., 2015). Registration with some 3D information such as a height map or a point cloud would help at this level. The registration approach of images from multiple sensors is a key point of this study and will be discussed in Chapter VI.

6. Conclusions

The vision methods are complementary in the plant traits they measure. Moreover, some traits necessitate the combination of several types of images. The most obvious example is the temperature of the leaves, which requires a thermal image and another image, RGB or NIR, to provide the mask of the leaves. Overall, masks built from some types of images will allow to move from traits at the canopy scale to traits at the scale of specific organs or plant parts. For example, the 3D information from stereo vision can be used to study only the upper leaves, or the detection of shadows from RGB images can be exploited to extract a trait only on the sunny organs. All those traits, at different scales, will be combined to characterise as fully as possible the plant phenotype. As stated by Prashar et al. (2014): "Unfortunately, a single sensor approach will always have limitations as stress is a complex trait and is not just influenced by one physiological or morphological component. In order to properly define the stress, a multi-sensor approach is needed." And this is that multi-sensor approach that we intend to develop in this thesis.

III

Field trials and image acquisition

1. Winter wheat

1.1 Importance

Winter wheat, *Triticum aestivum L.*, is an annual hexaploid monocot of the Poaceae family originating from the Near East. This cereal is currently the main source of human food on the planet (Le Stum et al., 2017). The main producers are India, European countries, China, Russia and the United States of America (Le Stum et al., 2017). However, winter wheat is grown on all continents. The production is dedicated to human food, animal feed, food industry – for gluten, starch, alcohol, etc. –, specific industries such as cosmetics or biodegradable plastics, and crop multiplication (Le Stum et al., 2017).

1.2 Growth and development stages

Winter wheat growth dynamics can be described as a succession of development stages. Several codes aim at describing those stages. The most famous and complete ones are the Zadoks scale (Zadoks et al., 1974) and the BBCH scale for cereals (Lancashire et al., 1991). Both are very similar because the BBCH is inspired from the Zadoks scale. For this thesis, the BBCH scale has been used because it fits into a more general framework, valid for other types of crops. This scale describes the growth of the crop as a succession of hundred stages coded from 00 (dry seed) to 99 (harvest). Those stages can be summarised in ten principal stages coded from 0 to 9. Each principal stage gathers ten stages from the detailed scale. They are illustrated in **Figure 3**. In Belgium, the plots are sowed in autumn or at the beginning of winter. The seedlings produce a few leaves and tillers before a phase of winter dormancy. In spring, they reach their maximum number of tillers, straighten and start to elongate their stems. At the end of this phase they have produced their upper leaf, called the flag leaf. This leaf is of great importance because, due to its youth and position, it will contribute to a large part of the plant's photosynthesis. During this time, the future ears have developed within each stem, and they are finally ready to emerge: those are the booting and heading stages. After that, the ears rise above the canopy and start producing flowers. What we call the ear is actually a spike-like inflorescence made of several spikelets. Each spikelet contains two to five flowers, which have three stamens, an ovary and two styles extended with a feathery stigma to hold pollen, except for a few unisexual female flowers (Willenborg et al., 2008). The fruit will be a caryopsis, an indehiscent dry fruit containing a single seed. That grain starts to grow.

At first, it is still green and milky inside. Once it reaches its final size, its content becomes doughy, it continues to harden and finish to ripe. It is ready for harvest.

Grain yield is determined by three components: the number of ears per square metre, the number of grains per ear and the weight of the grains, whose measurement unit is the average weight of a thousand grains. The development of those yield components is a dynamic process that is influenced at each development stage (**Figure 3**). Nitrogen being the main limiting factor, fertilisation is a requirement to achieve a good yield. The fertiliser is usually brought at three key stages: at tillering (BBCH 28), at the beginning of stem elongation (BBCH 30) and at flag leaf stage (BBCH 39). The first input favours the tiller number, the second input allows the tillers to grow and develop their ear, and the third input plays on the ear fertility, the grain filling and the grain quality. Strategies relying on two nitrogen inputs may also be interesting. As for a fourth input, in some conditions it can increase the protein content of the grain for bread-making varieties.
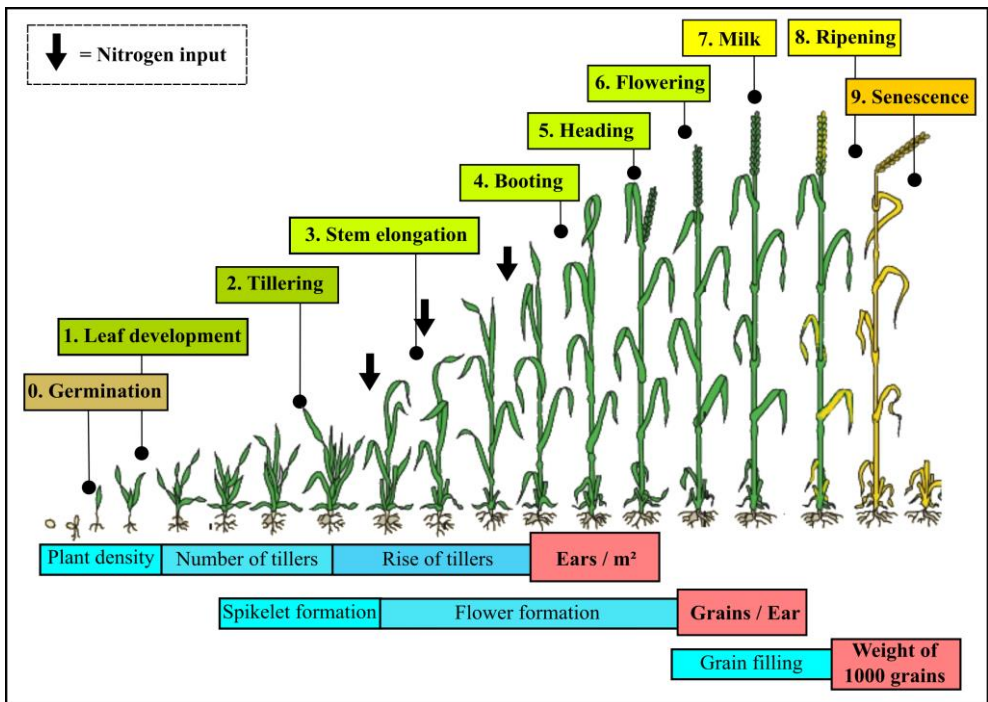
Figure 3. Wheat development stages and yield components. Each principal stage of the BBCH scale is indicated by a number from 0 to 9. The three yield components appear in the red boxes below the drawings. The blue boxes contain the elements influencing the elaboration of these yield components. The black down arrows represent nitrogen inputs according to a conventional fertilisation strategy.

2. Field trials

Trials were conducted from 2018 to 2021 in fields located in the Hesbaye area, Belgium (around 50° 33' N and 4° 44' E) on homogenous silty soil and a temperate climate. They were managed by the ASBL Ferme expérimentale of Gembloux Agro-Bio Tech. The characteristics of the trials are summarised in **Table 1**. To facilitate the descriptions, each trial was named with a specific ID. The first two numbers of the ID indicate the year and the letters detail the type of trial. For example, 18-F means a fertilisation trial harvested in 2018. The letter F denotes a fertilisation trial and the letters FP denote a trial dedicated to both fertilisation and fungicide protection. The trials F were kept under total fungicide protection to avoid the unwanted impact of fungal diseases. All the trials also underwent weeding and the application of a growth regulator, to avoid non studied stresses and accidents.

Table 1. Summary of the field trials.

Trial IDs	Type	Variety	Sowing date	Sowing (grains/m²)	Previous crop
18-F	Fertilisation	Edgar	2017-10-13	250	Potato
19-FP	Fertilisation × Protection	Smart	2018-10-23	250	Spinach
20-F	Fertilisation	Mentor	2019-11-07	250	Spinach
20-FP	Fertilisation × Protection	Vertikal	2019-11-05	250	Spinach
21-FP	Fertilisation × Protection	Vertikal	2020-10-27	300	Potato
21-F	Fertilisation	Mentor	2020-10-20	275	Potato
21-T	A single micro-plot	Dorset	2020-11-13	400	Wheat

Each trial F or FP consisted of a grid of micro-plots of 1.95×6 m, with a wheat row spacing of 0.146 m. For all the trials, the fertilisation strategy was based on three times of nitrogen input: at tillering (BBCH 28), at the beginning of stem elongation (BBCH 30) and at flag leaf stage (BBCH 39). The fertiliser was 27 % ammonium nitrate.

The main purpose of the fertilisation trials (F) was to generate contrasted canopy architectures to test our imaging methods. This is the reason why the trials included fertilisation scenarios not recommended for farmers, such as no fertilisation of three inputs of 105 kgN/ha, which is likely an excess fertilisation. The contrasted nitrogen inputs were also thought to study the dynamics of carbon and nitrogen allocation in

the plant organs. This was planned to be studied by both imaging at the organ scale and destructive manual measurements (detailed in Section 3.1). This thesis only focuses on the development of the imaging method at the organ scale, but future research activities will tackle the allocation question.

The trial 18-F was a randomised complete block design with four or eight replications of eleven fertilisation levels (**Table 2**). The number of replications was variable because only some scenarios were replicated for destructive measurements. The same kind of design was reproduced for trials 20-F and 21-F, but with only eight fertilisation levels (**Table 3**).

Table 2. Fertilisation levels of trial 18-F. The scenarios studied by destructive measurements were replicated eight times, and the others four times.

Nitrogen input (kgN/ha) at:			Destructive measurements
Tillering	**Stem elongation**	**Flag leaf**	
0	0	0	✓
0	0	60	
0	60	0	
60	0	0	
0	60	60	
60	0	60	
60	60	0	
60	60	60	✓
30	90	60	✓
30	60	90	✓
90	30	30	✓

Table 3. Fertilisation levels of trials 20-F and 21-F.The scenarios studied by destructive measurements were replicated eight times, and the others four times.

Nitrogen inputs (kgN/ha) at:			Destructive measurements
Tillering	**Stem elongation**	**Flag leaf**	
0	0	0	✓
30	30	30	✓
40	40	40	
50	40	55	
60	60	60	✓
80	40	60	✓
90	30	60	
105	105	105	✓

The fertilisation + fungicide protection trials (FP) were dedicated to study the possible interactions between nitrogen inputs and fungal diseases. The strategy to obtain fungal diseases was not the inoculation, but rather to rely on their natural occurrence. It was implemented by choosing disease-susceptible varieties and playing on the number of fungicide inputs during the season: from zero to three inputs. It was interesting to acquire images in those trials because one of the advantages sought by the image fusion approach developed in this thesis is to disentangle co-occurring biotic and abiotic stresses. A first approach to do so is presented in Section 3.2 of Chapter VIII.

The trial 19-FP was a randomised complete block design with four replications of nine experimental scenarios (**Table 4**). No destructive measurements were performed. The trial 20-FP was a randomised complete block design with four replications of sixteen experimental scenarios (**Table 5**). No additional micro-plots were dedicated to destructive measurements. They were performed on three of the four replications, on limited parts of the micro-plots. The trial 21-FP implemented fifteen of the sixteen objects of the 20-FP. The removed scenario was the one with fertiliser inputs of 80-60-60 kgN/ha and fungicide at BBCH stages 32 and 55. The other scenarios were replicated eight times, instead of four as for the 20-FP trial. This way, four of the eight replicates could be dedicated to destructive measurements during the season.

Table 4. Fertilisation and fungicide protection levels of trial 19-FP.

Nitrogen inputs (kgN/ha) at:			Fungicide at BBCH stages:	Destructive measurements
Tillering	Stem elongation	Flag leaf		
40	30	30	/	
60	50	50	/	
100	80	80	/	
40	30	30	39	
60	50	50	39	
100	80	80	39	
40	40	40	32 - 39 - 65	
60	50	50	32 - 39 - 65	
100	80	80	32 - 39 - 65	

Table 5. Fertilisation and fungicide protection levels of trial 20-FP.

Nitrogen inputs (kgN/ha) at:			Fungicide at BBCH stages:	Destructive measurements
Tillering	Stem elongation	Flag leaf		
40	40	40	/	✓
60	60	60	/	
80	60	60	/	
100	80	80	/	✓
40	40	40	39	✓
60	60	60	39	✓
80	60	60	39	
100	80	80	39	✓
40	40	40	32 - 55	
60	60	60	32 - 55	
80	60	60	32 - 55	
100	80	80	32 - 55	
40	40	40	32 - 39 - 65	✓
60	60	60	32 - 39 - 65	
80	60	60	32 - 39 - 65	
100	80	80	32 - 39 - 65	✓

The last trial, called 21-T, was not really a trial. It consisted of a single zone of a single micro-plot. The T in the ID means "temporal". That micro-plot was used to acquire images at the same place throughout the day on several dates. The plot was fertilised three times with 60-60-60 kgN/ha and protected with a fungicide at BBCH stage 39.

3. Agronomic measurements

3.1 Destructive measurements: LAI, dry matter and nitrogen

The Leaf Area Index (LAI), the nitrogen concentration of the plants and the above-ground dry matter were measured several times during the season based on samples harvested at key development stages (**Table 6**). The management scenarios concerned by those destructive measurements are indicated in **Table 2**, **Table 3** and **Table 5**. The measurements were made on four replicates of each scenario for trials 18-F, 20-F and 21-F, and on three replicates for trials 20-FP and 21-FP.

Table 6. BBCH stages of destructive measurements. The table indicates for each trial the stages at which LAI and dry matter were measured. Nitrogen concentration was always measured together with dry matter except for the trial 18-F for which nitrogen concentration was not assessed.

Trial	Trait	BBCH stage						
		30	32	39	55	65	75	89
18-F	LAI	✓	✓					
	Dry matter	✓	✓		✓			
20-F	LAI		✓	✓		✓	✓	
	Dry matter		✓	✓		✓	✓	✓
21-F	LAI	✓	✓	✓		✓	✓	
	Dry matter	✓	✓	✓		✓	✓	✓
20-FP	LAI			✓		✓		
	Dry matter			✓		✓		✓
21-FP	LAI			✓		✓		
	Dry matter			✓		✓	✓	✓

To measure the dry matter, wheat plants were harvested on a length of 50 cm on the three central rows of the micro-plot, to avoid a border effect. The plants were cut using a pruner at the basis of the tillers to collect all the above-ground biomass. In the laboratory, the organs were manually separated. Before flag leaf stage (BBCH39), the two pools were the leaves and the tillers. At flag leaf stage, the flag leaves were separated from the other leaves to create a third pool. After heading, the ears constituted a fourth pool. Each pool was weighted to obtain its wet mass, then they

were dried at 60°C for two to five days depending on the load of the oven (shared with other researchers) and the size of the samples. Once dry, they were weighted again. After that, the dry samples were sent to an external laboratory (ASBL Requasud, Laboratoire d'analyses agricoles, La Hulpe) to measure their nitrogen concentration through the Dumas method. The nitrogen concentration is expressed as a mass percentage (%N). Multiplied by the dry matter for a given area, it provides the nitrogen content (kgN/ha). From the nitrogen concentration and the dry matter, the Nitrogen Nutrition Index (NNI) was computed using the formulas from Lemaire et al. (1997) with the coefficients for wheat determined by Justes et al. (1994) (**Eq.1**).

$$\text{NNI} = \frac{N_{act}}{N_C} \; ; \; N_C = 5.35 \, (\text{DM})^{-0.442} \qquad \text{(Eq.1)}$$

where N_{act} is the actual nitrogen concentration (%N) and N_C is the critical nitrogen concentration (%N) corresponding to the crop dry matter: DM (t/ha).

To measure the LAI, the wheat plants were harvested on 50 cm of a row, aside the three rows sampled for the dry matter. All the leaves were separated from tillers, weighted, glued on white paper thanks to a transparent adhesive sheet and scanned. The Otsu segmentation method (Otsu, 1979) was applied on the scanned images to isolate the leaves from the white background. The leaf surface was computed as the sum of the areas of the scanned paper sheets multiplied by the proportion of pixels segmented as leaf. The LAI was finally obtained by dividing the leaf area by the sampling area (50 cm $\times$ 14.6 cm of row spacing). That tedious protocol was however not employed for all the LAI measurements. A method was developed to diminish the number of samples and lighten the work in the laboratory. At each measurement date, only five micro-plots, with contrasted fertilisations, were selected to manually measure the LAI. To get the results for the other micro-plots the trick consisted of relying on the relation between the LAI and the mass of the leaves. Our hypothesis was that this relation could be different at each date. Therefore, the five samples collected at each date were used to build a linear regression between the LAI and the wet mass. This relation was used to compute the LAI of the other micro-plots, for which the mass of the leaves had been measured. For almost all of the dates, determination coefficients (R^2) of 0.98 or higher were obtained. However, a later analysis revealed that our hypothesis was too conservative: a same relation could have been used for several dates. **Figure 4** shows the relation between the wet mass of the leaves and their area for all the samples harvested in 2020 and 2021. Interestingly, all the points lie on the same line, except for the measurement made at the end of tillering stage (BBCH 30). Consequently, only two relations could have been used: a linear

relation with an intercept of -14.01 and a slope of 31.48 for the measurements at tillering ($R^2 = 0.996$) and a linear relation with an intercept of 60.11 and a slope of 46.67 for all the other measurements ($R^2 = 0.984$).
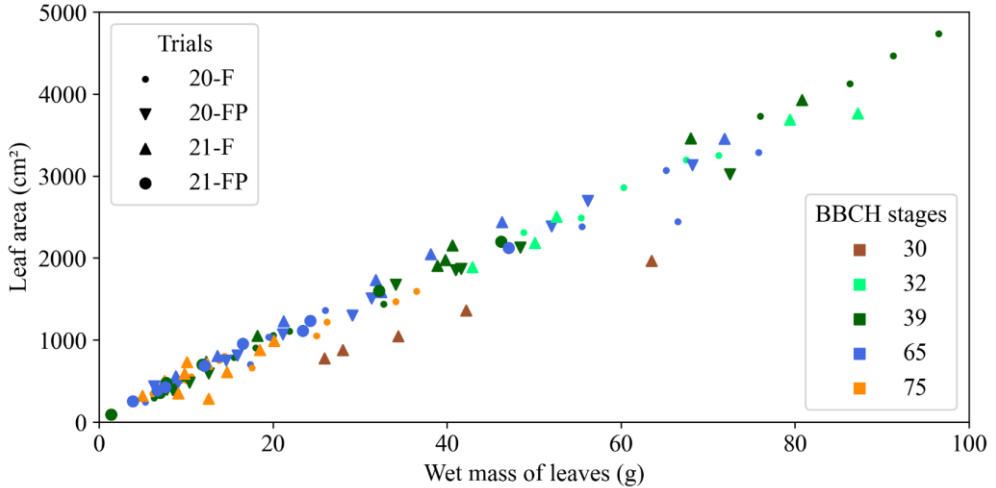
Figure 4. Relation between the wet mass of sampled leaves and their area. The samples were harvested in trials 20-F, 20-FP, 21-F and 21-FP at BBCH stages 30, 32, 39, 65 and 75.

3.2. Lodging, pests and fungal diseases

At each acquisition date, the field trials were observed to evaluate potential stress factors or damage. The presence of diseases or pests was recorded for the entire trials, except if some micro-plots underwent severe damage, in which case they were individually reported. Overall, the fertilisation trials – 18-F, 20-F and 21-F – were quite healthy, despite the inevitable presence of septoria tritici blotch on the lowest leaves and some upper leaves often attacked by beetle larvae. Those damages, however, were considered having a limited impact on the plants. A more concerning accident was the lodging of half of the micro-plots in July 2021, due to heavy rains. For the trials 21-F and 21-FP, it prevented image acquisition from BBCH 85 to maturity.

In addition to those general observations, the fungal diseases were graded for trial 19-FP and 21-FP. The trial 19-FP was observed on May 23rd and June 4th, respectively before and after spraying fungicide at BBCH 39 for the micro-plots concerned with that treatment. That period was characterised by an important development of stripe rust, while the other diseases were almost not present. The trial 21-FP was observed at BBCH stages 62, 71, 75 and 87. At this period, moderate disease levels were

recorded for septoria tritici blotch and stripe rust. Brown rust was also observed. The visual scoring was performed by two operators at the same time, each quickly examining the whole micro-plot and parting the foliage at three or four random spots. The scale used for scoring is presented in **Table 7**. For stripe rust, the scale was used with additional conditions: the score was minimum 4 in case of an active outbreak zone of disease, minimum 5 in case of two outbreaks and minimum 6 in case of three outbreaks.

Table 7. Scale for the visual scoring of wheat fungal diseases. The scale is based on three criteria: i) the affected foliar floor (L1 refers to the flag leaf, L2 to the second upper leaf, …), ii) the average intensity of the infection on a leaf (Lo = low, M = medium and Hi = high), and iii) the repartition of the disease in the micro-plot.

		Affected foliar floor											
		L4 and <			L3			L2			L1		
Average intensity		Lo	M	Hi	Lo	M	Hi	Lo	M	Hi	Lo	M	Hi
Repartition	< 3 plants	1	1	1	1	1	1	1	1	1	1	1	1
	< 10 plants	2	2	2	2	2	2	2	2	2	2	2	82
	10 < x < 50 plants	2	2	2	2	2	2	3	3	3	4	4	4
	1 in 10 plants	2	2	3	3	3	4	4	5	6	6	7	8
	1 in 2 plants	2	2	3	3	4	4	5	6	6	7	8	8
	All the plants	2	2	3	3	4	5	5	6	7	7	8	9
	All the leaves	3	3	4	4	5	5	6	7	7	8	9	9

4. Environmental measurements

The ICOS station of Lonzée (https://www.icos-belgium.be/ESLonzee.php) was located at less than one kilometre of the field trials, at coordinates 50°33'5.8"N - 4°44'46.5"E. It provided all the year, at a half hour time step, measures of air temperature, air humidity, wind speed, precipitations, incident radiation in the plant photosynthetic spectral range (400-700 nm), the shortwave IR (300-2800 nm) and the long wave IR (4500-42000 nm). For the image acquisition dates, those data were available at a time step of twenty seconds.

Data from the Sombreffe meteorological station (CRA-W/Agromet.be) were also exploited. This station was located in another village, at five kilometres, but it was

useful to fill the gaps in the meteorological data. It recorded at a time step of one hour the air temperature, the air humidity and the precipitations.

5. Image acquisition

5.1 State of the art of field proxy-sensing platforms

As crop phenotyping is moving from the laboratory to the field, multi-sensors field phenotyping platforms are being developed worldwide. The challenge is not any more to bring the plants to the sensors, as for indoor phenotyping, but to bring the sensors to the plants. Natural light conditions, wind and the architecture of crops themselves (dense crops, shadows, overlapping leaves) also pose issues that need to be considered.

A first category of platforms are fixed installations: towers (Naito et al., 2017; Shafiekhani et al., 2017) or sensors displaced above the canopy by means of rails, cables or automated gantries (Kirchgessner et al., 2017; Virlet et al., 2017; Bai et al., 2019; Beauchêne et al., 2019). Some of those platforms are turnkey solutions proposed by companies, such as: the LemnaTec GmbH Field Scanalyzer (Sadeghi-Tehran et al., 2017; Virlet et al., 2017) and the Phenosphex Field Scan. Fixed platforms present excellent repeatability performances and may reach a high level of automation, which allows to spare workforce. Images are acquired at the same spots with high spatial and temporal resolution. This makes them a precious tool to build time series. They are able to acquire data without destroying the crop and whatever the soil conditions. However, their cost and the impossibility to move them to another field are major drawbacks. Towers are especially very limited regarding the number of studied micro-plots, and a compromise needs to be made between the spatial extent of the measurement and the spatial resolution.

A second category comprises the manned vehicles. They can be as diverse as one-wheel platforms (Crain et al., 2016), manned wheeled platforms (Bai et al., 2016; Hasan et al., 2018; Fu et al., 2020), self-propelled electric platforms (Jimenez-Berni et al., 2018; Pérez-Ruiz et al., 2020), towed platforms (Busemeyer et al., 2013), buggies (Deery et al., 2014) and modified tractors (Comar, Burger, et al., 2012; Andrade-Sanchez et al., 2014; Jiang et al., 2018; Walter et al., 2019).

The third category comprises autonomous vehicles and robots. Asefpour et al. (2012) developed a robot able to advance between spaced crop rows to acquire side images. Salas Fernandez et al. (2017) presented an auto-guided tractor equipped with

stereo cameras. The Robotanist of Mueller-Sim et al. (2017) is capable of navigating between corn rows. Shafiekhani et al. (2017) combined their tower (Vinoculer) with a robot (Vinobot). Gao et al. (2018) studied a phenotyping platform able to deploy several lightweight robots in the field. The French institutes ARVALIS and INRAE conceived an autonomous vehicle called "The Phénomobile", which is guided between micro-plots and takes advantage of a robotic arm to place the measurement head above the zones of interest (Madec et al., 2017). Most of those autonomous vehicles use a Real-Time Kinematic (RTK) positioning system. It relies on a station to enhance the satellite positioning of a mobile unit, providing position at a centimetre accuracy.

Finally, it is inevitable to compare those proxy-sensing platforms with unmanned aerial vehicles (UAV). Especially since UAV can fly very low (at a few metres above the canopy) and provide measurements with a spatial resolution similar to that of proxy-sensing platforms. UAVs unquestionably have the advantage in terms of spatial extent and speed. They can fly whatever the soil conditions while vehicles may not access the field if the soil is too wet. As they do not enter the canopy, there is no risk of damaging the crop and no need to have paths in the crops. Nevertheless, UAVs present a limited payload. Big and expensive UAVs are necessary to carry heavy and bulky sensors, or carry numerous sensors at a time. Those big UAV's require pilot training and flying permits.

5.2 Our multi-sensor approach

This section describes that multi-sensor system that was employed to follow all the trials of 2020 and 2021.The sensor pod was composed of three types of cameras: two RGB cameras, a thermal camera and a multispectral camera array (**Figure 5**). In 2019, the thermal camera was not yet implemented and the position of the cameras was slightly different. In 2018, the only sensors were the two RGB cameras. That 2018 system is presented in Chapter IV, dedicated to the stereo vision tests performed that year.
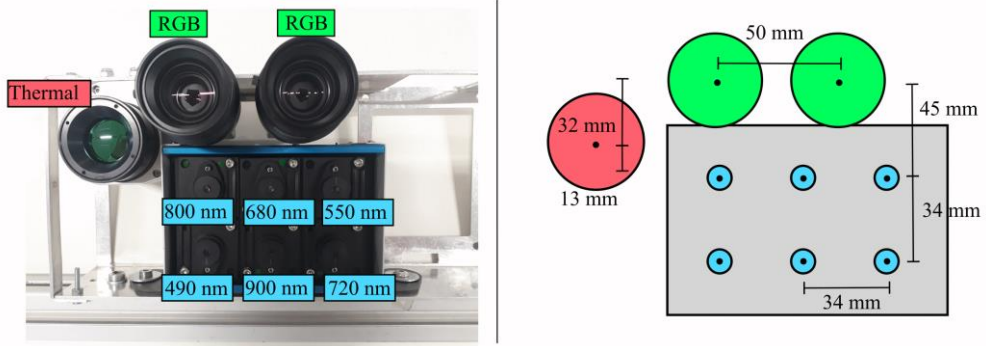
Figure 5. The cameras and their relative position. The cameras of the multispectral array are indicated with the central wavelength of their optical filter.

The two RGB cameras were GO-5000C-USB cameras (JAI A/S, Copenhagen, Denmark). They aimed at forming a stereoscopic camera pair. The distance between the centres of the two sensors, called the baseline, was 50 mm. Each camera was equipped with a 2560×2048 pixels CMOS sensor and a LM16HC objective (Kowa GmbH, Düsseldorf, Germany). Their focal length was 16 mm. The HFOV and VFOV were 44.3° and 33.6°, respectively. The aperture was set to f/4.0. The multispectral camera array was a Micro-MCA (Tetracam Inc., Gainesville, FL, USA). It consisted of six monochrome cameras equipped with 1280×1024 pixels CMOS sensors. The optical filters were narrow bands centred at 490, 550, 680, 720, 800 and 900 nm. The width of each band-pass filter was 10 nm except for the 900 nm filter that had a width of 20 nm. The lenses had a focal length of 9.6 mm and an aperture of f/3.2. The horizontal field of view (HFOV) was 38.26° and the vertical field of view (VFOV) was 30.97°. The thermal camera was a PI640 thermal camera (Optris GmbH, Berlin, Germany). It was equipped with a sensor of 640×480 pixels and covered a spectral range from 7.5 to 13 μm. The focal length was 18.7 mm. The HFOV and VFOV were 33° and 25°, respectively. At a distance of 1.6 m, the footprint of the frames was 1.26 m² for the RGB cameras, 0.98 m² for the cameras of the multispectral array and 0.67 m² for the thermal camera.

The optical axes of all the cameras were theoretically parallel but small deviations were possible due to mechanical imperfections. Each of those vision systems was individually calibrated to remove the geometrical distortions induced by the lenses. The multispectral array was geometrically calibrated using thirty images of a 10×7 chessboard of 24-mm squares for each camera. That calibration provided intrinsic camera parameters and coefficients to correct image distortion. The average

reprojection error varied between 0.11 and 0.12 pixels depending on the camera. The RGB stereo pair was calibrated using twenty-five images of a 10×7 chessboard of 26-mm squares. That calibration provided not only intrinsic camera parameters and distortion coefficients but also extrinsic parameters allowing rectification of the images in a context of stereo vision. The average error for the camera pair was 0.4 pixels. For the thermal camera, it was not possible to use the chessboard for calibration. A dedicated thermal target was built, inspired from Liu et al. (2018). That target consisted of a 36×28 mm white Forex® plate including twelve removable black-painted disks of 4-cm diameter. The disks were arranged in three rows at regular intervals. The distances between the disks themselves and between the disks and the borders of the plate were 4 cm. Before calibration the plate was stored for 15 minutes in a freezer at -18°C and the disks were placed on a radiator. Twenty-three images were acquired during the 10 minutes after reassembling the target. The algorithm segmented the disks thanks to the temperature differences. The key points used for calibration (equivalent to the corners of the chessboard in the conventional method) were the centroids of the disk objects. This method was robust to heat diffusion because, regardless of the diameter of the detected hot disks, the centroids were always at the same positions. As for the multispectral cameras array, the calibration provided intrinsic parameters of the camera, including distortion coefficients. The average reprojection error was 0.24 pixels.

In addition to the cameras, the multi-sensor system comprised in 2020 and 2021 an AvaSpec-ULS2048 incident light spectrometer equipped with a cosine corrector (Avantes, Apeldoorn, Netherlands). Its signal-to-noise ratio was 200:1. The irradiance calibration was carried out in the factory on March 23, 2020.

5.3 Our proxy-sensing platforms

Over the years, three platforms were used to carry the sensor pod (**Figure 6**).

Figure 6. Image acquisition platforms.

In 2018, the platform, called manned platform I, was carried by two human operators. The only implemented sensors were the RGB cameras, kept at a distance around one metre above the canopy. The main problem was the presence in the image of shadows due to the bars of the platform itself. This problem was solved for the manned platform II, used in 2019, 2020 and 2021. The sensor pod was installed on a cantilever, oriented so that shadows from the bars never appeared in the images. Nevertheless, the cantilever beam tended to vibrate when the operators posed the platform, which necessitated waiting a few seconds before capturing the images. The cameras were oriented to capture nadir images of the plot. This choice of a nadir view was made to ease the plant height computation and the image fusion. It was also the easiest option to detect the ears and compute the number of ears per square metre, as it will develop in Chapter V. From 2020, the cameras were kept at a distance around

1.6 m above the canopy instead of 1 m. The reason for this increased distance compared to 2018 and 2019 was to ease the alignment of the images from the different cameras. This point will be further developed in Chapter VI, dedicated to close-range multimodal image registration. At this distance of 1.6 m, the footprint of the frames was 0.98 m² for the cameras of the multispectral array, 0.67 m² for the thermal camera and 1.26 m² for the RGB cameras.

In 2021, a mobile platform arrived. It was still under development so it was not available at all the dates and never totally replaced the manned platform II. This mobile platform however counted numerous improvements compared to his manned ancestors. The four motorised and independent wheels allowed great manoeuvrability in the field. It spared human forces while slightly improving the throughput of acquisitions. Considering four images per micro-plot, which was thought a good trade-off between measurement time and capturing the micro-plot heterogeneity, the manned platform II was able to measure around sixty adjacent micro-plots per hour. The wheeled platform could increase that performance from a few plots, but it was still necessary to stop the platform and wait for image capture, which explains the limited throughput improvement. When dealing with non-adjacent micro-plots, the wheeled platform was however far more effective than the manned platform II, because the operators could not carry it easily on long distances.

A remote control allowed the operator to move the wheeled platform and trigger image acquisition. Later, additional cameras were added at the front and the back of the platform to detect the micro-plots limits and allow autonomous navigation in the field. The sensors were still located on a cantilever beam but a tensioner was added to prevent vibrations (not visible in **Figure 6**). Thanks to a sliding system, the height of the cantilever beam was easy to adjust, contrary to the old manned platforms where it was necessary to disassemble and reassemble the bars. This mobile platform was also thought to be compact and portable. All the parts could easily be unscrewed and stored in the trunk of a small van. The main issues encountered were derailments of wheel chains and, more critically, deformations of weak mechanical parts. It is also to note that, despite the possibility of a continuous displacement of the platform, it was still necessary to stop a few seconds to capture images.

5.4 Acquisition parameters

In 2018, the only sensors were the RGB cameras. Images were recorded with a 12-bit colour resolution. The software trigger did not allow a perfect synchronisation between the cameras. The exposure was managed using the camera auto-exposure algorithm. It caused image saturation on some wheat organs.

In 2019, the sensors were the RGB cameras and the multispectral camera array. The RGB images were recorded with a 12-bit colour resolution and the multispectral images with an 8-bit colour resolution. The exposure settings were not yet optimal. For the multispectral camera array, the default exposure settings were kept during all the season. It did not cause saturation but some images were too dark. For the RGB cameras, a custom auto-exposure algorithm, based on a dichotomous search, was designed to set image saturation at 1 % of the pixels. It was thought to be a good compromise between too dark or too bright images.

In 2020 and 2021, the whole multi-sensor system had been implemented. The cameras were asked by the software to capture images at the same time but only the two RGB cameras were triggered perfectly together thanks to an external trigger from an Arduino Uno micro-controller. Considering all the cameras, the absence of a common external trigger and the different needs in terms of integration time resulted in images acquired with a slight temporal shift. The maximum shift was less than a second. The RGB images were recorded with a 12-bit colour resolution and the multispectral images with a 10-bit colour resolution. Care was taken to avoid image saturation. For that, the auto-exposure algorithm of the RGB cameras was modified. A dichotomous search was performed to find the highest exposure time for which no saturation was detected in the image. The two limits of the search were the exposure time computed by the manufacturer algorithm and this time divided by five. The ISO was kept to base value and not allowed to change. For the multispectral camera array, the factory auto-exposure algorithm computed the exposure time for a master camera and multiplied this time by pre-defined coefficients to obtain the exposure times for the other cameras. The images from the master camera were never over-exposed but the images from the slave cameras could have been too dark or too bright at some acquisition dates if the multiplication coefficients were not tuned. Indeed, the proportion of light reflected from the scene in the different wavebands of the multispectral device is strongly affected by the development stage of the crop. To ensure proper exposure, the coefficients determining the exposure time of the slave cameras were adapted three or four times during a season. The first set of coefficients accounted for the early development stages where a lot of light was reflected from

bare soil. The second set of coefficients suited for most of the growing season but a third and a fourth sets were necessary to account for the maturity of the ears.

In 2020 and 2021, a spectrum of the incident sunlight was recorded at each image acquisition using a 16-bit resolution. Each recorded spectrum was the average of three consecutive measurements. It was corrected for dark noise and non-linearity of pixel response to exposure time. Thanks to the factory calibration, digital values were converted to irradiance data. Each acquisition of images and their associated solar spectrum took only a few seconds. It corresponded to the time necessary to average the spectra and ensure a proper exposure time for all the cameras.

All the images were saved in the TIFF format to avoid information loss. Regarding the colour resolution, it is to note that, except for the stereovision process of 2018 (Chapter IV, Section 3), all colour resolutions of cameras were reduced to 8 bits per pixel as a first step of image post-processing. This was required because several Python libraries only worked with 8-bits images.

5.5 Dates of acquisition and imaged scenes

Images were acquired on all the replicates and for all the scenarios mentioned in Section 2 of this chapter. Most of the time, four images were acquired on each micro-plot, at regular spacing on the length. This number was a good compromise between taking into account the intra-plot variability and limiting the data volume and the acquisition time.

For the trial 18-F, images were acquired by the manned platform I. For the trial 19-FP, images were acquired by the manned platform II. Unlike future uses of this acquisition platform, the distance between the cameras and the canopy was 1 m instead of 1.6 m.

For the trials 20-F and 21-F, four images were captured per micro-plot for four replicates of each fertilisation scenario. For the scenarios with additional replicates dedicated to destructive measurements, only two images were acquired per micro-plot. This choice was made to keep constant the number of acquired images, because on those micro-plots it would not have been possible to acquire images at four different spots after the destructive measurements. For the trial 20-F and 20-FP, images were captured by the manned platform II. For the trial 21-FP and 21-F, images were acquired by the manned platform II or the mobile platform.

For the trials 21-F and 21-FP, in addition to the images of the micro-plots, a panel of known reflectance was imaged at the beginning and at the end of the acquisition series. That reference panel was the MAPIR camera reflectance calibration target V2

(MAPIR Inc., San Diego, CA, USA). It consisted of four targets of a felt-like material: the black, dark grey, light grey and white targets. Each target was provided with a reflectance spectrum. According to the manufacturer there was no bi-directional anisotropy because of the diffuse properties of the material.

The trial 21-T was different from the other trials. It aimed to record time series on a single micro-plot. Images were acquired by the manned platform II that stayed at the same position from 9 a.m. to 5 p.m. Every quarter of an hour, four acquisitions were performed at 10 s intervals: one where the reference panel was placed above the canopy and three of the canopy without the reference panel. This protocol was repeated on six dates.

The dates of all the acquisitions and the destructive measurements are detailed in **Figure 7**. To make the measurement periods comparable from one year to another in terms of crop development, the timeline is expressed in Photo-Vernalo-Thermic Units (°C-days). They correspond to growing degree days corrected for vernalisation and photo-period, as detailed in the paper of Duchene et al. (2021). To compute the growing degree days, the base temperature was 0°C, the minimal, optimal and maximum temperature for crop growth were respectively 0, 35 and 45 °C. In our implementation, the computation of growing degree days started from sowing. To compute the photoperiod factor, base photoperiod was 6 hours and saturation photoperiod was 20 hours. To compute the cold requirement factor, the range of temperature for vernalisation was 10 °C.
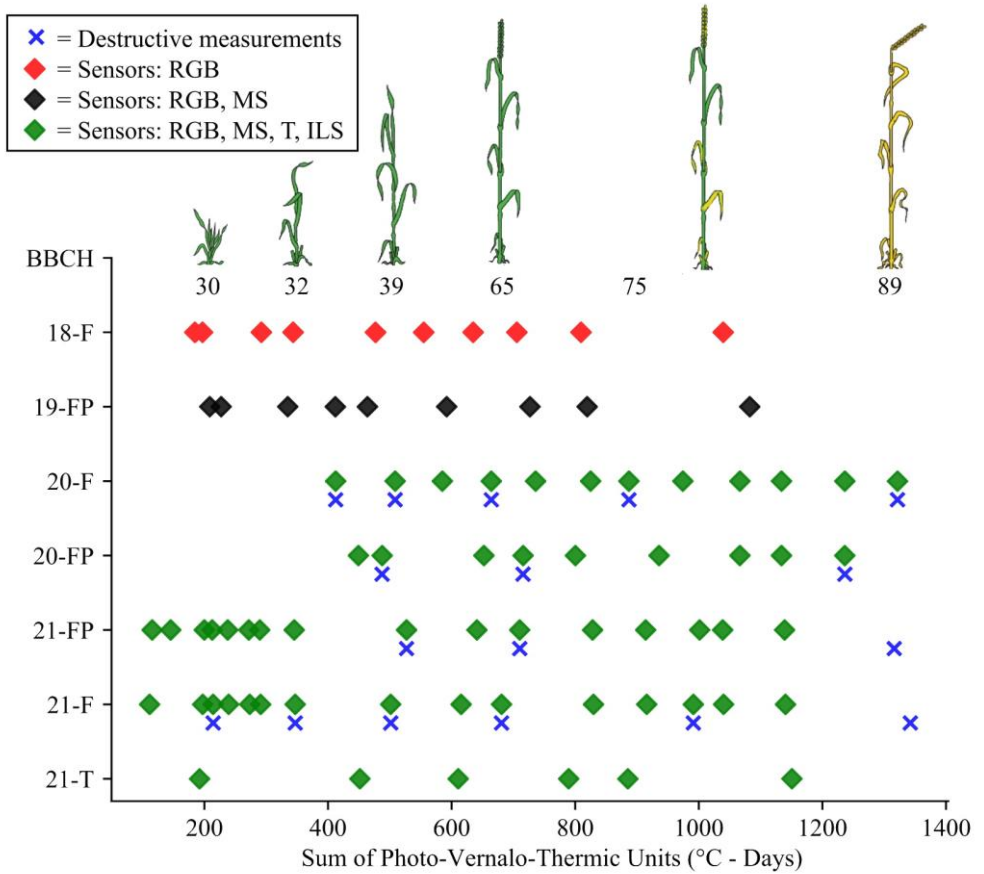
Figure 7. Dates of multi-sensor acquisitions and destructive measurements.
MS = Multispectral camera array, T = thermal camera and ILS = Incident Light sensor.

IV

Plant organ height by stereo vision

1. Synopsis

This chapter is largely based on the research paper of Dandrifosse et al. (2020):

Dandrifosse S., Bouvry A., Leemans V., Dumont B. & Mercatoris B., 2020. Imaging wheat canopy through stereo vision : overcoming the challenges of the laboratory to field transition for morphological features extraction. *Front. Plant Sci*. **11**(February), 1–15. DOI:10.3389/fpls.2020.00096.

It presents the automatic measurement of wheat height using stereo vision. But, more than height, the technique can provide height maps of vegetation, i.e. maps that associate each pixel in a colour image to a height value. Those maps can be further processed to extract a number of relevant wheat traits: some statistical descriptors of the heights of the leaves, the height of the ears, the average leaf angle, the leaf angle distribution and other geometrical features. The analysis of the colour images associated with the height maps identifies the specific areas, such as the plant organs, for which the height is studied.

The height maps are also of high interest for multi-sensor image fusion. The distance between the canopy and the sensor pod is used for image registration (Chapter VI), which is the operation of aligning pixel to pixel the images from all the sensors. That distance is also used to compute the image footprint at canopy height, that is necessary to compute area-based plant traits, such as the number of ears per square metre (Chapter V). Finally, that height information at the pixel scale can be fused with signals from multispectral or thermal cameras, for example to extract an indicator of plant physiology only on the lower or the upper leaves.

2. Goal and structure

The goal was to develop and validate an automatic method to obtain height maps of wheat canopies.

Over the years, the method was implemented twice: a first time in 2018 using the Matlab programming language and a second time in 2020 using the Python language. We make the distinction between these two implementations because the libraries containing the stereo vision algorithms do not propose the same functionalities for the two languages. The results presented in this chapter in Sections 5.1 and 5.2 were generated in 2018 using the first implementation and detailed in the research paper (Dandrifosse et al., 2020). They relied on images from the trial 18-F. Afterwards, the

stereo vision process was implemented in open-source thanks to Python. In the Section 5.3 of this chapter, the two implementations are compared on the images from the trial 18-F.

The image acquisition in the trial 18-F is described in Chapter III. The distance between the canopy and the cameras was 1 m. At this distance, the footprint of one image was 0.5 m².

3. Method: first implementation

3.1 Calibration

The two cameras were the GO-5000C-USB cameras (JAI A/S, Copenhagen, Denmark) presented in Chapter III in the frame of the multi-sensor approach. The mechanical support to position them was different but the distance between the centres of the two sensors was also 50 mm. In 2018, the camera pair was calibrated using a 9×6 chessboard of square side of 40 mm. The chessboard images were processed by Matlab Stereo Camera Calibrator App according to the method proposed by Zhang (2000). The calibration error was 0.32 pixels.

3.2 Depth map

All image treatments were realised with Matlab R2016a. The first step was the rectification: the images were aligned so that the same point of the scene appeared at the same y-coordinate in the two images. This was performed by Bouguet's algorithm thanks to the calibration parameters of the system (Bradski & Kaehler, 2008). This rectification algorithm also relied on calibration parameters to account for radial lens distortion. The rectified images were converted to grayscale. In order to reduce the effect of noise on ulterior 3D computations, the grayscale image size was reduced to 1280×1024 pixels by averaging the pixel values on each 2×2 square. The second main step was the stereo matching which consisted in finding corresponding pixels in right and left images. The difference of x-coordinate of corresponding pixels gave the disparity between pixels. Stereo matching was performed with the Semi-Global Block Matching algorithm (SGBM) proposed by Hirschm (2007). The principle is to detect corresponding pixels by means of similar neighbourhoods. The two most important parameters are the matching window size, which is the size of a side of investigated neighbourhoods, and the disparity range, which corresponds to the maximal possible disparity. Matching window sizes of 5, 9, 15 and 19 pixels were tested. A window

size of 15 is the default value, while 5 is the minimum value. The two other values were chosen to test other configurations, one between the default and the minimum value (9) and the other greater than the default value (19). Disparity range was automatically adjusted for each image pair if disparities peaked at the maximum allowed value. The disparity estimation was also controlled by post filtering based on minimum uniqueness value, set to 5, to remove false matches (Bradski & Kaehler, 2008). The complex texture of images acquired in natural conditions resulted in incomplete disparity maps which were filled by the method proposed by Yun (2012). This method performed interpolation only if reliable information was available in the neighbourhood. It will be named in this paper: Yun's method.

The last step consisted in computing depths, which are inversely proportional to disparities. For each pixel in the left frame of a pair of stereo images, the depth to the camera is given by **Eq.2**.

$$z = \frac{b\,f}{d} \qquad\qquad (Eq.2)$$

where b is the baseline (m), f is the focal length (pixel), d is the disparity (pixel) and z is the distance (m) between the observed object and the camera, commonly referred to as depth. The result of this whole step of image processing was a depth map, showing distances between objects in the scene and the cameras.

3.3 Segmentation of the 2018 images

Images acquired in the trial 18-F before ear emergence stage, May 24th, were separated into two classes: soil and leaves. The segmentation method was based on a Support Vector Machine (SVM) classifier trained with the components of RGB and HSV colour spaces. According to Hamuda et al. (2017), the addition of the HSV colour space increases the segmentation robustness to natural light conditions. The use of machine learning helps to deal with complex situations containing enlightened and shadowed canopy elements. To train and evaluate the classifier, 10000 pixels were selected in a set of images representative of the different acquisition dates and conditions. The selected pixels were split so that 70 % were dedicated to training and 30 % to validation. The last step of the process consisted in median filtering with a window of 5×5 pixels to remove segmentation noise on the resulting binary image.

Images acquired at flowering stage, June 5th, contained ears and were segmented into three classes: soil, leaves and ears. SVM providing binary outputs, three classifiers were combined according to the "Error Correcting Output Codes" principle

(Dietterich et al., 1994). But colour information was not sufficient to distinguish ears at their early development stages because they were as green as leaves. For this reason, in addition to RGB and HSV components, height and texture predictors were also used to train the SVM. Texture predictors of a pixel were i) the average of pixel intensities over a 7×7 square centred on the considered pixel and ii) the average of the squared differences of intensities between each pixel and the central pixel of the neighbourhood. These parameters aimed at taking into account the differences between the grainy texture of ears and the smooth texture of leaves. To be independent of the camera-ground distance, the considered height predictor for each pixel was the difference between the 95th percentile of heights and the height of this pixel. To train and evaluate the classifier, 5000 pixels were selected with 70 % dedicated to training and 30 % to validation.

The method performed soil-leaf segmentation with an accuracy of 98.5 % on the validation dataset. By adding depth and texture information, the method separated soil, leaves and ears with an accuracy of 99.8 % on the validation dataset. Such performances were however overestimated due to pixel saturation. Those pixels were either soil, leaf or ear but, as the intensity values peaked, their classification based on colour was impossible. For the sake of properly training the classifier, training zones in saturated areas have only been selected for the most commonly saturated class: leaves for the soil-leaf classification and ears for the soil-leaf-ear classification. As a result, badly classified pixels in saturated zones – *e.g.* saturated soil pixels classified as leaf pixels – could not be taken into account to compute the classification error, leading to an overestimated accuracy. We also attract the attention of the reader that the soil-leaf-ear segmentation method was only validated for one date, and the soil-leaves segmentation for six dates and only one wheat variety. Although this machine learning method worked well in 2018, more recent findings showed that other options may be more interesting. In Chapter VII, exploiting the multi-sensor system, we detail a robust and simple soil-leaf segmentation method based on thresholds in RGB and NIR images and on a cloudiness index derived from the incident light spectrometer. In Chapter V, we detail a robust deep learning approach to isolate the ears at all the development stages of the crop.

3.4 Height map

To extract only depth of plant objects, the segmentation mask was applied on the depth map. Ground-wheat distances, called plant heights, were computed on the basis of this plant depth map. Plant heights were the difference between camera-ground and

camera-wheat distances. A unique height value for the canopy could be obtained from the heights of the map elements using different statistical descriptors such as the median, the 75th percentile or the 95th percentile. **Figure 8** and **Figure 9** show the image treatment pipeline from colour images to height maps of plant elements.
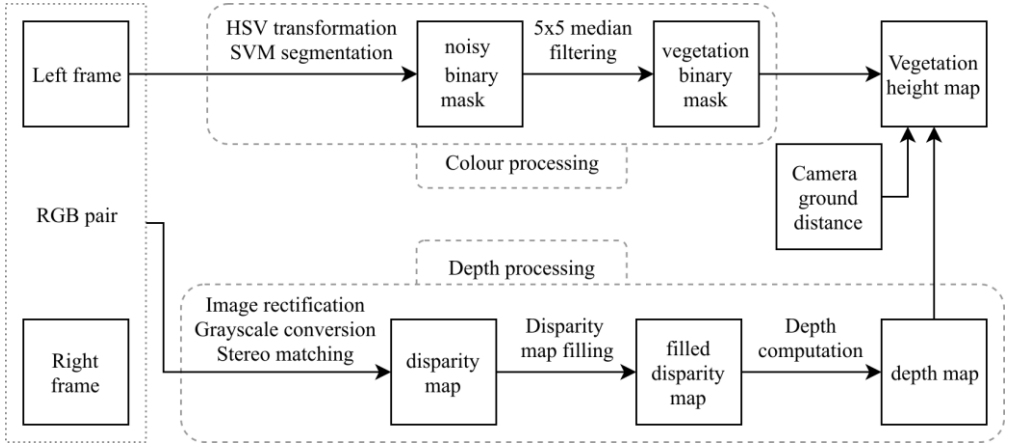
Figure 8. Pipeline to build a plant height map from a pair of RGB images.
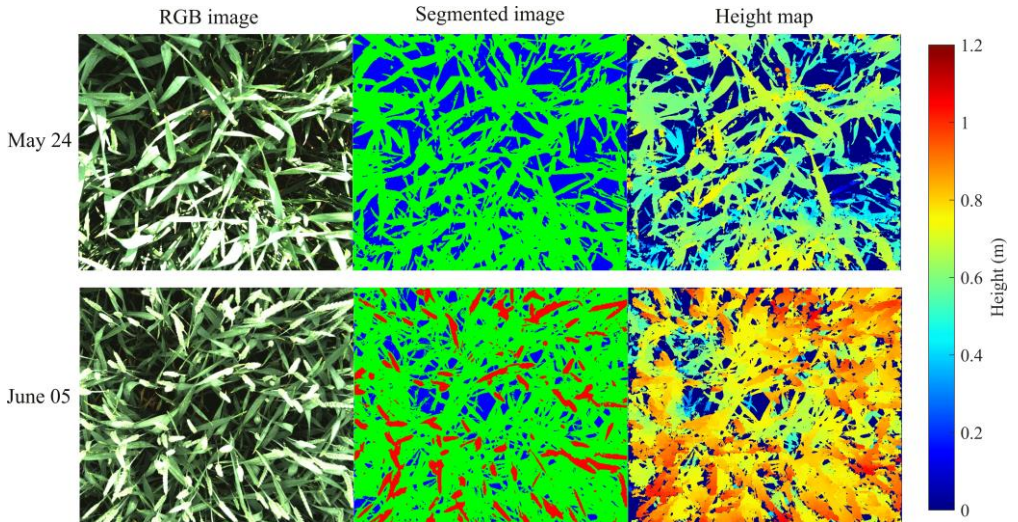
Figure 9. RGB image, segmented image and height map. In the segmented image, soil is blue, leaves are green and ears are red. The two rows illustrate wheat before heading (May 24th) and wheat after heading (June 5th).

3.5 Manual height measurements

May 24th, 2018, when the ears did not emerge yet, heights were manually measured at the insertion of flag leaf for thirty-six tillers per micro-plot. The micro-plots were systematically divided into twelve zones in which three tillers were randomly selected. Insertion of flag leaf was chosen to perform repeatable height measurements. Such measurements on the tiller were independent from leaf orientation and did not necessitate stretching leaves. The reason for this measurement before heading was to assess the ability of manual measurements to record plant height at a vegetative stage, although wheat height is conventionally measured on ears (Pask et al., 2012). June 5th, heights were manually measured at ear tops for thirty-six ears per micro-plot.

4. Method: second implementation

The Python implementation of the stereo vision process relied on the OpenCV-Python library (Bradski et al., 2008). The same steps and the same matching algorithm as for the Matlab version were used but some elements could not be reproduced between the two libraries. OpenCV-Python (version 4.5.3.56) did not allow to perform stereo matching on images with a colour resolution superior to 8 bits. However, OpenCV-Python proposes a post-filtering algorithm, the Weighted Least Squares (WLS) filter (Min et al., 2014), to smooth the disparity map and fill the gaps. **Figure 10** illustrates a raw height map and the associated height maps obtained with the Yun's and WLS filling methods, and **Figure 11** shows the histograms for those maps. Both gap filling methods slightly reduced the mean height. The WLS method filled more gaps than the Yun's approach. However, from our observations, the WLS is only effective to fill the height map when there are already a lot of good matches as illustrated in the figure. When there were only a few good matches, the method could extrapolate too much.
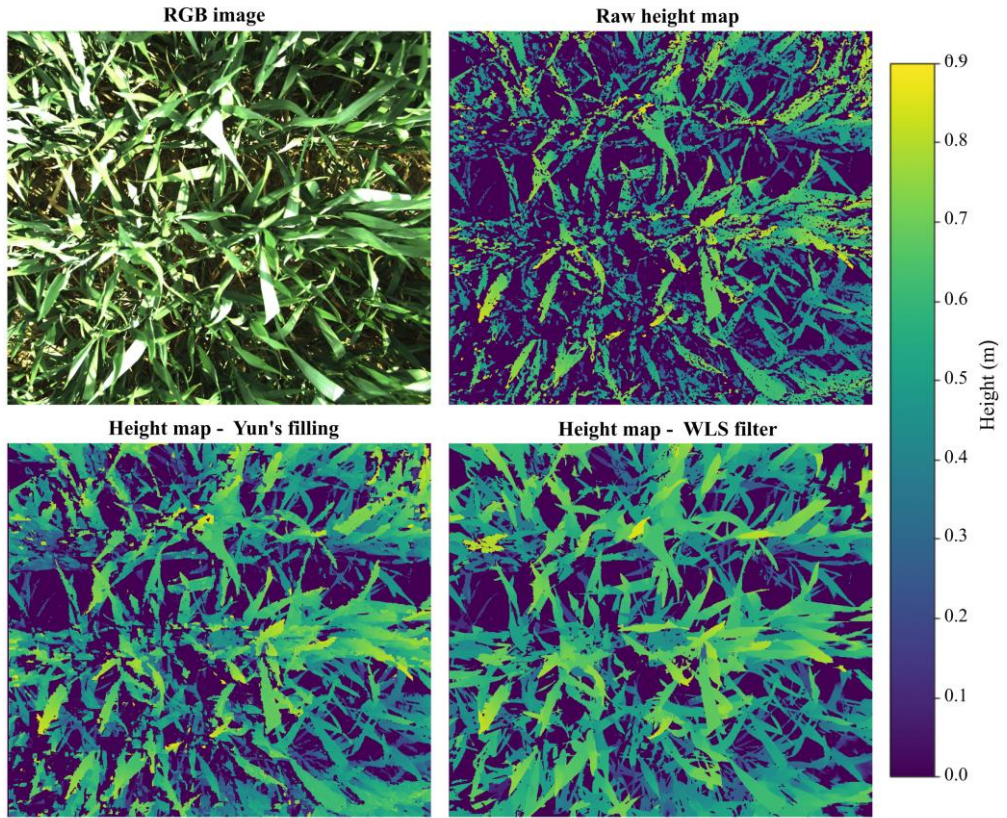
Figure 10. Comparison of filling algorithms for height maps. The scene was imaged on May 16th on a micro-plot of the 18-F trial. This figure shows the RGB image cropped to the zone commonly observed by the two cameras, the raw height map, the height map filled with the algorithm of Yun (2012), and the height map filled with the WLS method. The dark blue zones, at 0 on the scale, corresponded either to soil or to plant pixels for which the height could be computed.
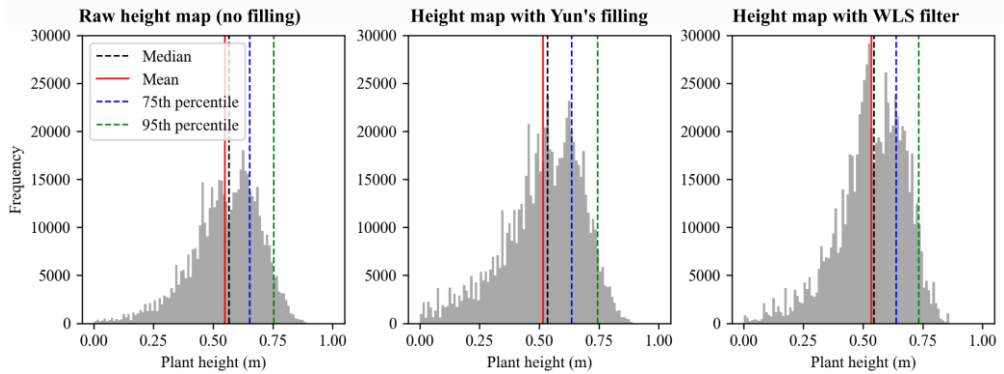
Figure 11. Histograms of a height map with different filling algorithms. The scene is the same as the one in **Figure 10**. This figure shows the histogram of the raw height map, the height map filled with the algorithm of Yun (2012), and the height map filled with the WLS method. The height data are presented using 100 bins of 1 cm.

The soil-leaf segmentation approach was simplified. Instead of relying on machine learning, the Excess Red (ExR) vegetation index (Meyer et al., 2008) was used, with a threshold of 0.05. This method retrieved the green and healthy leaves (**Figure 12**). For the vegetative development stages, it was considered sufficient to derive plant organ heights.
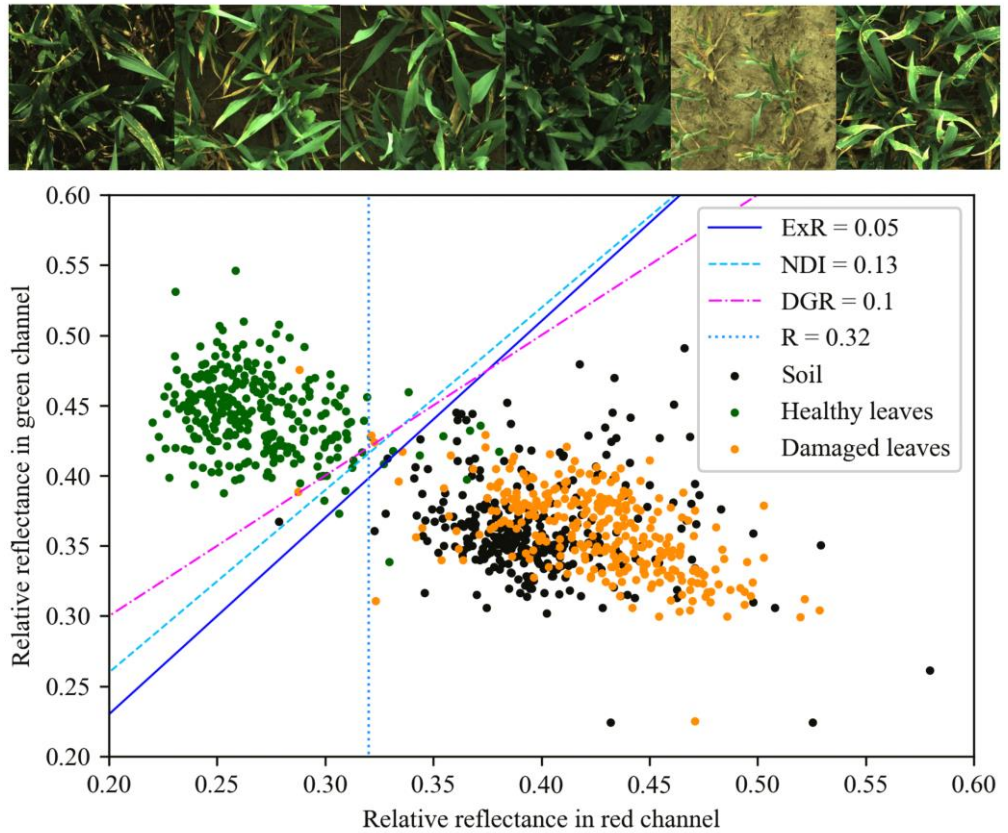
Figure 12. The best RGB vegetation indices to segment the healthy leaves. The relative reflectance factors in red and green channels, respectively, R and G were computed based on the digital values, the lens aperture, the exposure time and the transmission factor of the lens in the channels. Each point in the graph corresponds to a pixel manually sampled in one of the six sub-images above. The indices are: $ExR = 1.4\,R - G$; $NDI = \frac{G-R}{G+R}$ and $DGR = G - R$ (Golzarian et al., 2012).

5. Results and discussion

5.1 The stereo matching algorithm

Stereo matching is a challenging task. The projections of an object on two different optical planes are not necessarily represented by the same number of pixels. This results in incomplete pixel-to-pixel matching for depth computation. Visual occlusions can also prevent full depth mapping of stereo-images.

The Matlab implementation of the SGBM algorithm was firstly assessed on the Middlebury dataset, which contains reference images provided with dense disparity maps (Scharstein et al., 2002, 2003, 2007). The *cones* and *teddy* reference images were firstly considered due to their complex scenes with contrasted objects. The stereo matching was performed with errors of 7.4 % for cones image and 9.5 % for teddy image. For less complex reference images, the error significantly decreased to 2 %. Finally, the algorithm was tested on the *Aloe* reference image that is the most representative image of vegetation and led to an error of 8.4 %. It is noticed that this error represents the number of pixels for which disparities differ from at least one pixel. It means that disparities differing from 1 pixel contributed to the error, with the same weight as a more important error. In comparison with the literature (Scharstein et al., 2002) and more particularly with an optimised stereo matching algorithm leading to errors of 2.9 % and 7 % for *cones* and *teddy* images respectively (Li et al., 2017), the performances of the SGBM algorithm on the Middlebury dataset were considered as sufficient.

Secondly, stereo matching performances were evaluated for the specific case of winter wheat canopy by studying the effects of image size, pixel colour resolution, disparity map filling and matching window size on images acquired at four dates (**Figure 13**). Since no reference maps were available for the canopy images, an indicator based on the plausible height percentage was introduced to assess the matching quality. This indicator expressed the proportion of plant pixels for which the computed height value ranges between the ground and 0.6 m below the stereoscopic device, even though this height value could be inaccurate. This choice was based on the hypothesis that the highest plants could have been found 0.6 m below the cameras, considering that the average camera-wheat distance was approximately 1 metre but that some plants were taller than the average canopy level. As shown in **Figure 13**, this sensitivity analysis revealed that the best stereo matching performances were obtained for an image size of 1280 × 1024 pixels with a colour resolution of 12 bits.

Moreover, the computation time to extract the disparity map was roughly ten times higher for 2560×2048 pixel images than for 1280×1024 pixel images. The absolute value of the computation time depends on the hardware. As an order of magnitude, the average time to compute a disparity map for 1280×1024 pixel images was around 0.8 s on a Windows computer with a 2.8 GHz Intel Core I5-4200H processor. This computation time was not significantly influenced by colour resolution. Applying an interpolation-based filling algorithm helped to complete the disparity map. Regarding the matching window size, little effect was observed for 12-bit images. Overall, a window size of 15 pixels provided the best results regardless of the image size. For 8-bit images, the choice of a proper matching window size was more decisive. The optimum for wheat images was found for a size of 9 pixels. It is noted that the stereo matching variability increased at the last two dates. This could be explained by more contrasted canopy architectures generated by the different fertilisation practices.
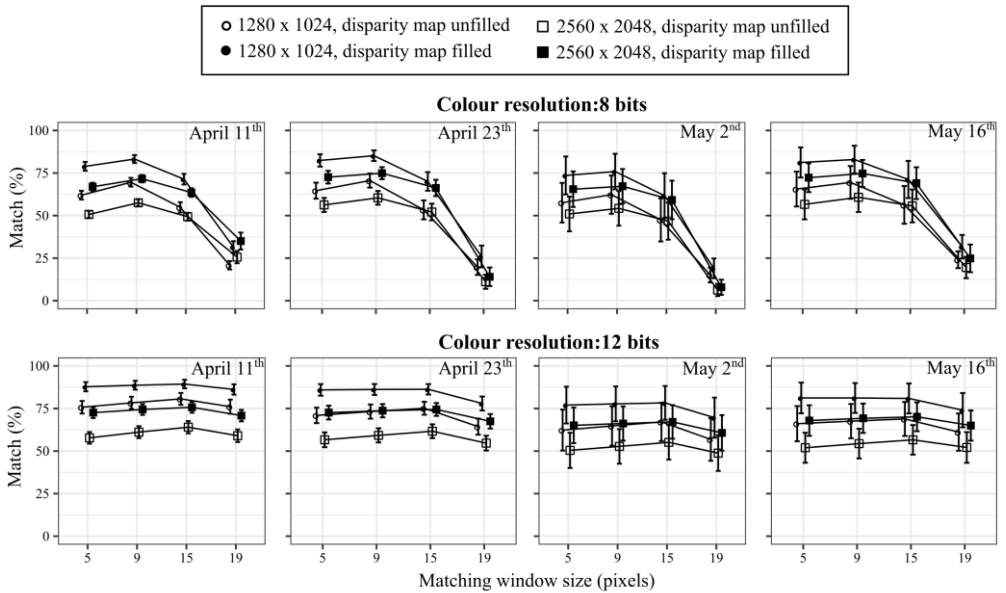
Figure 13. Stereo matching performances. The match percentage is the proportion of plant pixels for which the matching algorithm found a plausible height, i.e. a height between the ground and 0.6 m below the cameras. It was computed for all the images acquired in the 18-F trial on April 11th, April 23th, May 2nd and May 16th.

5.2 Comparison of stereo and manual height measurements

For data recorded before ear emergence, stereo-based and manual measurements with a metre stick provided non-equivalent indicators to describe canopy height. Both present advantages and inconveniences and should be used for different purposes. Manual measurements have the advantage that the operator directly chooses the point of interest – ear tip, flag leaf tip, flag leaf insertion, last node, ... – which is convenient to study specific vegetative organs. On the contrary, the image-based height measurement of specific points is a complex, and sometimes impossible, task due to the difficulties encountered to automatically detect such points, often hidden by the leaves. For manual height measurement, numerous repetitions were necessary to obtain a robust estimation, which can be seen as the main drawback of this method. On the contrary, the stereo-based method allows acquiring height descriptors of a zone of several plants in a simultaneous way. As demonstrated by Cai et al. (2018), the height map yields complete height distribution which provides much more information on canopy development than a manual height measurement. Several statistical descriptors of the height deduced from the height map are proposed in this study. The suggested descriptors are the median, the 75th percentile and the 95th percentile of the plant heights. **Figure 14** compares the different height descriptors from stereo vision and the manual height measurements. Pearson's correlation coefficients between the manual measures and the median, the percentile 75 and the percentile 95 of automatic measures are respectively 0.66, 0.58 and 0.39.
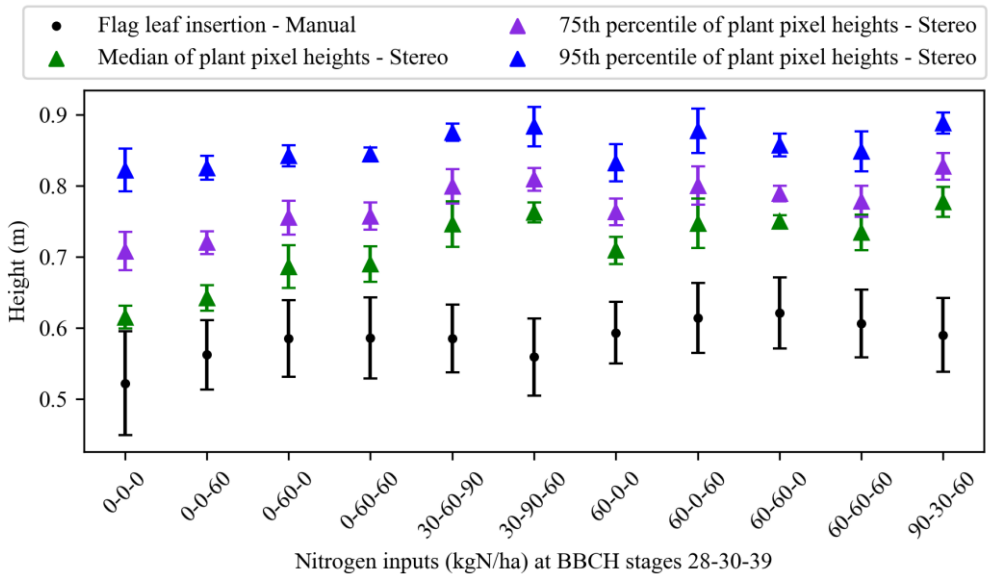
Figure 14. Comparison of automatic and manual measurements of canopy height at flag leaf stage. The measurements were performed on May 24th on eleven micro-plots from the 18-F trial, each in a different fertilisation scenario. The vertical bars indicated standard deviations.

A note on height measurement at the vegetative stages concerns the manual reference method. A metre stick, as used for this study, is not the only possibility to record crop height. Using a rising plate – a plate of known weight attached to a rule – as described by Barmeier et al. (2016), would provide a weighted plant height. This measurement is considered to be more representative and objective than a measurement at a specific point. Moreover, as the rising plate measures a weighted height on a zone and not height at a point, it could be better suited to provide a reference for stereo vision. It would even be possible to design the rising plate with a size similar to the captured zone. A rising plate must however be adjusted to account for various degrees of stem stiffness, depending on growth stage or cultivar.

For data recorded after ear emergence, the relevant height descriptor is the mean height of ear tips both for manual and imaging methods. For the automatic measurement, the height of each ear object was the 95th percentile of heights, so that the mean height of ear tips for one image was the mean of those 95th percentiles. This trait was measured for two blocks of micro-plots both manually and by stereo vision on June 5th (**Figure 15**).
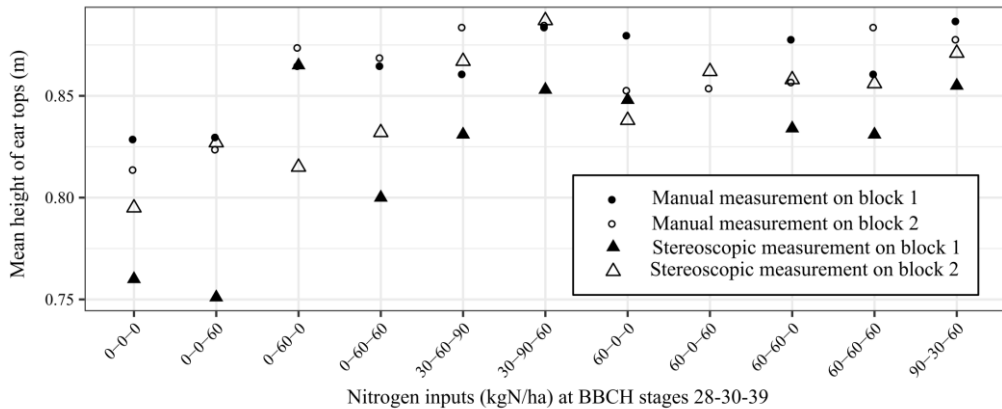
Figure 15. Comparison of automatic and manual measurement of the average height of ear tops. Measurements were performed on June 5th for two replicates of the eleven fertilisation scenarios of trial 18-F: block 1 and block 2.

Considering the manual measurements as a reference, mean ear top heights were measured by stereo vision with an accuracy of 97.9% for block 2 and 95.2% for block 1. For block 1, the automatic measurement systematically underestimated the mean height but the evolution of height according to the fertilisation practices followed the same trend as for manual determination, with a Pearson's correlation coefficient of 0.89. For block 2, the correlation coefficient was 0.70.

The systematic error observed for block 1 was probably not due to the accuracy of camera-ear distance measurements but may be due to some other issues such as saturated leaves badly classified as ears or the value of camera-ground distance. Those issues represent challenges inherent to field acquisition. As suggested above, a custom auto-exposure algorithm should help to deal with important image saturation. The second issue is more challenging. The camera-ground distance was not constant due to soil surface irregularities induced for instance by tractor passage. To avoid that, an estimation of the camera-ground distance can be deduced from the soil pixel depth. However, for dense and high canopies, the estimation of ground depth was not reliable due to the lack of visible soil spots rendering stereo matching troubles. Finally, because of an imperfect positioning of the acquisition device in the field, the cameras were not exactly perpendicular to the ground resulting in non-constant real camera-ground distance on the stereo-image.

To conclude, stereo-based height measurement in a complex canopy offered an easy way to compare canopy height and average ear top height of different micro-plots.

However, the absolute height of micro-plots remained uncertain because of difficulties to automatically get camera-ground distance at each point of the area of interest. Manual measurements are useful to measure the height of specific plant elements that would be difficult to spot on images.

5.3 Comparison of the two implementations

The 95th percentile of leaf pixel heights computed by the Matlab and the Python implementations were compared for the five acquisitions dates before heading on trial 18-F. The average height difference per image was 3.0 cm. Those heights are illustrated for three fertilisation scenarios in **Figure 16**.
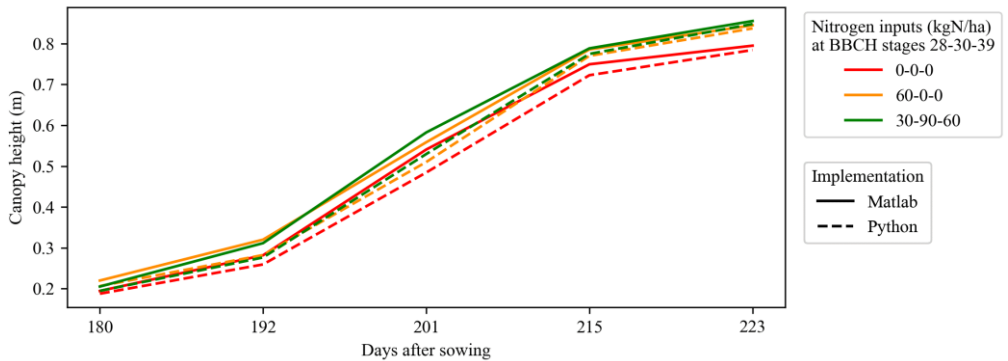
Figure 16. Canopy height for the Matlab and the Python implementations of the stereo vision method. Yun's filling was applied on the height maps. Canopy height was computed as the 95th percentile of leaf pixel heights for the wheat images acquired before heading on trial 18-F. For graph clarity sake, the comparison is presented only for three fertilisation scenarios.

A small height shift was observed between the two implementations. This may be due to the different soil-leaf segmentation approaches. Maybe the Python segmentation tended to include more of the low and dark leaves, which resulted in lowering the average height. But the shift could also be due to the different implementations of the registration method or the semi-global block matching. The matching quality was evaluated for the two implementations using the plausible height percentage defined in Section 5.1. For the Matlab implementation, the score was 61.8 % without height map filling and 78.2 % with Yun's filling algorithm. For the Python implementation, the score was 52.7 % without filling, 76.9 % with Yun's filling and 78.0 % with the WLS filling. The matching quality difference could

explain the canopy height difference, especially if matching was best for one method over the other for the high or low leaves.

Despite the height shift, the comparison is satisfying: two different implementations from scratch yielded similar height curves. The three fertilisation scenarios in **Figure 16** show the same trends. The unfertilised plants presented on average the lowest height at all the development stages. The canopies fertilised with 60 kgN/ha at tillering were taller at first than the canopies fertilised with 30 kgN/ha but this trend was reversing when those second canopies received 90 kgN/ha in addition. What is important to understand is that the canopy heights presented here account for all the plant pixels in the image, and not just a single point at the top of a tiller. Consequently, those height measurements better describe the overall vegetative development, which is linked to the nitrogen nutrition level. In other words, if the nitrogen inputs do not favour taller tillers but favour the development of more numerous high leaves, this will be observed in such image-based canopy height indicators.

V

Ear segmentation and ear density

1. Synopsis

This chapter is largely based on the research paper of Dandrifosse et al. (2022):

It tackles the automatic detection of wheat ears in the field, which provides two main outputs: the ear density and the ear mask.

The ear density, i.e. the number of ears per square metre, is one of the three yield components of wheat, which makes it an important agronomic trait to record. Its automatic measurement necessitates both the ability to count the ears in an image and to estimate the area corresponding to that ear count, called the image footprint. In this chapter, it is demonstrated that the ear density can be measured at high throughput using a combination of two RGB cameras. A deep learning approach can detect and count the ears in a robust way based on a RGB image, while the stereo vision method presented in Chapter IV can be exploited to compute the image footprint.

The ear mask is a map of the pixels corresponding to the ear positions in a wheat canopy image. It is not by itself an element of high agronomic relevance. However, it constitutes a crucial step in the process of the extraction of plant traits at the organ scale. Thanks to the registration method presented in Chapter VI, the images from multiple cameras can be aligned pixel to pixel, and therefore the ear mask computed from a RGB image can be applied on multispectral or thermal images. There is a high interest to remove the ears if we want to study the physiology of the leaves, but the ears can also deliver interesting traits, for example regarding their morphology or the presence of an ear disease such as fusarium head blight.

This chapter is also part of a success story in international data sharing. As all deep learning models, the method to automatically detect the ears in the images required a lot of learning data. Our model could be trained in a robust way because of the creation of an international dataset: the Global Wheat Head Detection (GWHD) dataset, gathering labelled ears from all around the world (David et al., 2020). We contributed to this initiative, by providing images for the second version of the dataset (David et al., 2021).

2. State of the art

The traditional measurement of ear density in the field relies on manual counting along a wooden stick. The method is slow, subject to human bias and may necessitate the operator to work in an uncomfortable position, leaning forward to isolate the tillers in a row. Large uncertainties on that manual method were reported (Madec et al., 2019). For this reason, a major avenue of improvement in wheat phenotyping is to develop imaging methods to automate and objectify the ear counting task. To our knowledge, the first attempt of wheat ear detection in the field dates back to 1995 (Germain, 1995), followed by a period of sparse innovations dominated by classic image analysis and machine learning approaches (Chopinet et al., 2006; Cointault, Guerin, et al., 2008; Cointault, Journaux, et al., 2008; Cointault et al., 2012; Zhu et al., 2016). From 2018, an impressive intensification of the wheat ear detection research was observed. Classic image analysis approaches continued to be proposed, but none of those studies was able to demonstrate a method robust for many wheat cultivars, development stages and light conditions. Indeed, most studies were validated on images from only one acquisition date (Alharbi et al., 2018; Zhou, Liang, Yang, Xu, et al., 2018; Zhou, Liang, Yang, Yang, et al., 2018; Dandrifosse et al., 2020; Tan et al., 2020). Fernandez-Gallego et al. (2018) acquired images on two sites and at three development stages, but they discarded images acquired under non optimal light conditions and development stages. Similarly, Fernandez-Gallego et al. (2020) acquired images at five dates but were able to validate their approach only on the images from the two dates of diffuse sky conditions. In parallel, the use of 3D sensors to detect ears has been studied to overcome the issue of illumination conditions, but only a few papers concern their implementation in the field (Saeys et al., 2009; Velumani et al., 2017), and these methods required expensive sensors.

In recent years, the field of image analysis has evolved due to the fast development of deep learning. Regarding wheat phenotyping, these algorithms were identified as the solution to propose a robust ear detection solution based on RGB images. A first deep learning approach was used by Pound et al. (2017) on wheat ears in controlled conditions. Then, in-field ear counting methods were proposed by Cao et al. (2020), Hasan et al. (2018), Lu et al. (2020), Madec et al. (2019), Xiong et al. (2019), Xu et al. (2020), Yang et al. (2019) and Zhao et al. (2021). Each of those methods focused on a single or a few growth stages. In parallel, a major contribution occurred in 2020 when David et al. (2020) released their open access dataset: the Global Wheat Head Detection dataset (GWHD). Although other open datasets were released before, for

example the SPIKE (Hasan et al., 2018) and the WED (Madec et al., 2019) datasets, the GWHD dataset was identified as the tipping point towards a new era of international data regrouping. This first version of the dataset gathered data from ten different places and obtained huge visibility thanks to an online competition. It was exploited for counting tasks by Ayalew et al. (2020), Fourati et al. (2021), Gong et al. (2020), Li et al. (2021), Wang et al. (2021), Wu et al. (2020) and Yang et al. (2021). A second version of the dataset was released to study the robustness of ear detection with respect to the development stage (David et al., 2021). New images were added and the development stages were organised according to four grades: post-flowering, filling, filling-ripening and ripening. Thanks to deep learning methods, the community started to master the ear counting task. The ability to count the ears in images is however not sufficient to get the ear density in the field, which is the variable of agronomic interest. To estimate the ear density, it is necessary to know the footprint of the image at ear height, as underpinned by David et al. (2020). As this height may vary from micro-plot to micro-plot the task is not so easy. Only a few partial solutions were proposed. Fernandez-Gallego et al. (2018) computed ear densities using the camera-canopy distance but is not detailed how this distance was measured. Madec et al. (2019) exploited a Light Detection And Ranging device (LiDAR) to measure that distance but without digging into the question of the reference height to compute the image footprint. Sadeghi-Tehran et al. (2019) placed a sheet of known dimensions in the image. This technique eliminates the camera-canopy distance issue but adds constraints to the acquisition system. Most of the other studies concerning ear detection did not include a way to measure the ear density. One of the reasons to explain this lack is that they often relied solely on existing datasets, such as the GWHD.

The interest of detecting ears goes beyond the counting and the estimation of density. It also implies the possibility to segment the ears in the images and build a mask that can be used for the extraction of further plant traits at the organ scale. Provided a mask of the ears, it is possible to study the wheat reflectance, related to its physiology, independently for the leaves and the ears. It is especially conceivable to exploit a mask built from RGB images to extract organ specific reflectance from registered multispectral images or even organ temperature from registered thermal images. Moreover, Li, Jiang, et al. (2021) demonstrated the impact of wheat ears on canopy reflectance, which supports the need for wheat ear segmentation methods. But ear counting and segmentation are two different tasks. The knowledge of ear bounding boxes, used for counting, does not directly provide the masks of the ears in those

boxes. None of the previously mentioned ear counting studies tackled the segmentation step. However, as for counting, deep learning algorithms have been pointed out as a robust segmentation solution. A first deep learning segmentation was proposed on rice panicles by Xiong et al. (2017). For wheat, segmentation solutions were developed by Ma, Li, Du, et al. (2020), Ma, Li, Liu, et al. (2020) and Wang et al. (2019). Methods allowing both counting and segmentation were detailed by Grbović et al. (2019), Sadeghi-Tehran et al. (2019) and Su et al. (2021). The limitation of all those studies is that each of them was validated for a few development stages. However, masks allowing to separate the ears from the leaves are needed throughout the season, especially if they aim at the extraction of physiological plant traits such as nitrogen status, senescence dynamics or water stress. Ear masks can also be used at various growth stages to look for the presence of diseases (Su et al., 2021). Additionally, they may provide ear morphological information such as ear width or length. In the laboratory, Pound et al. (2017) showed that it was even feasible to detect the spikelets. The measurement of such morphological information would probably exploit the ear masks only at late development stages, when the ears are bent from the weight of the grains, and thus imaged along their length.

3. Goal and structure

This chapter firstly proposes a robust deep learning approach able to count and segment the wheat ears. The performances for both tasks are evaluated at all the key development stages from heading to maturity. Secondly, it is detailed how the ear count in an image can be converted to an ear density in the field. Ear densities derived from the images are studied from heading to maturity and compared with i) density measurements performed in the field at one date by three human operators and ii) reference density measurements performed on harvested wheat plants.

4. Method

4.1 Ear bounding box detection

4.1.1 Model

The deep learning model selected for the detection of ear bounding boxes was YOLOv5 (DOI: 10.5281/zenodo.3908559). YOLOv5 is the last born from the YOLO (You Only Look Once) family (Redmon et al., 2016). That model was picked because

i) it counted among the most recent and popular deep learning solutions, ii) it is faster than other classical models such as models from the R-CNN family (Ren et al., 2017), and iii) it already demonstrated good performances for wheat ear detection tasks (Yang et al., 2021; Zhao et al., 2021). The method to decide the final predictions among all the bounding box propositions was the weighted boxes fusion approach (Solovyev et al., 2021).

4.1.2 Image pre-processing and labelling

The images of the trials 20-F and 20-FP were considered for this study. They were complemented with the already labelled images of another dataset: the GWHD V2 (David et al., 2021). The images from the GWHD V2 were in a square format (1024×1024 pixels), as required by the YOLOv5 algorithm. Since our images were acquired in the format of 2560×2048 pixels, they required some pre-processing steps. Considering the footprint of the image, if they were directly converted to 1024×1024 pixels the ears would have been too small compared to those in the GWHD V2. For this reason, the images were divided into four sub-images of 1280×1024 pixels, and then each sub-image was resized to 1024×1024 pixels. To label these sub-images, the LabelImg tool was used (https://github.com/tzutalin/labelImg). Two labelled sets were built from the 2020 sub-images: a set dedicated to validation and a set that aimed to complement the GWHD V2 for the model training. To build the validation set, at each acquisition date and for each trial, around forty sub-images were randomly selected and labelled, leading to a total of 64091 labelled ears in 566 sub-images. Regarding the training set, images were picked at growth stages not present or under-represented in the GWHD V2: heading and maturity. One hundred and three sub-images were randomly picked among two dates at heading stage (June 3$^{\text{rd}}$ for trial 20-F, June 2$^{\text{nd}}$ for trial 20-FP) and thirty-two sub-images were randomly selected among three dates at maturity stage (July 29$^{\text{th}}$ for trial 20-F, July 13$^{\text{th}}$ and July 22$^{\text{nd}}$ for trial 20-FP). It led to a total of 13566 labelled ears in 135 sub-images. All the labelled data, called the Contrasted-Fertilisation Wheat Ear Dataset 2020 (CFWED2020), are available online for future use in the community (10.5281/zenodo.5709821).

4.1.3 Semi-supervised training

The training consisted of four steps. Firstly, the model was trained on the GWHD V2. Secondly, the trained model was used to predict ear bounding boxes on all the images acquired in 2020. All the predicted boxes were saved as pseudo-labels. Thirdly, both the labels from the GWHD V2 and the pseudo-labels were exploited to train the model again. Finally, the model was improved using transfer learning thanks

to the labelled training set built from our 2020 images. That way, the last update of the model weights relied on trustworthy data.

4.1.4 Validation of bounding box detection

The performances of the bounding box detection were evaluated on the labelled validation set built from data acquired on trial 20-F and 20-FP in 2020 (detailed in Section 4.1.2). The predictions of the model were compared with the labels to obtain a number of true positives, false positives and false negatives. A true positive (TP) is an ear that was correctly detected, a false positive (FP) is the detection of an ear that is not an ear, and a false negative (FN) is an ear that was not detected by the model but should have been. However, in that kind of object detection task, it is quite rare that the detected bounding box perfectly matches the labelled bounding box. For this, the definition of a correct or incorrect detection was based on the notion of Intersection over Union (IoU), i.e. the ratio between the area formed by the overlap of the detected box and the labelled box and the area formed by the set of these two boxes (**Figure 17**). The TP, FP and FN are determined by choosing an IoU threshold. A value of 0.5 was chosen. This value is considered as the standard to evaluate a detection model. Moreover, it was pointed out as the optimal choice in the study of Madec et al. (2019).
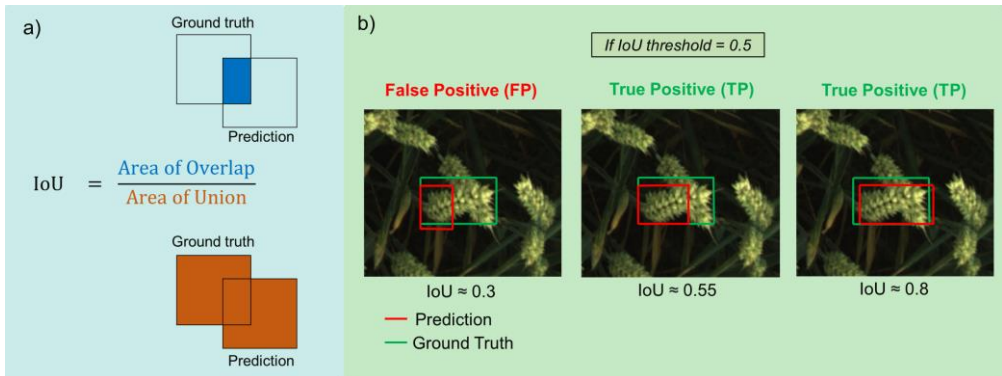
Figure 17. Concept of Intersection over Union (IoU). a) Illustration of the definition of the IoU. b) IoU applied to differentiate false positives and true positives in the frame of bounding box detection.

From the TP, FP and FN, several other meaningful and widely-used indicators were built. The precision (**Eq.3**) measures the fraction of correct detections among all the detections. The recall (**Eq.4**) measures the fraction of correct detections among all the objects that should have been detected. The accuracy (**Eq.5**) is an obvious performance metric. Its general formula contains the true negatives (TN) but in such

object detection tasks, there is no TN. The F1 score (**Eq.6**) is the harmonic mean of precision and recall, which provides a robust model performance assessment.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{Eq.3}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{Eq.4}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{Eq.5}$$

$$\text{F1 score} = 2\,\frac{\text{Precision}\ \text{Recall}}{\text{Precision} + \text{Recall}} \tag{Eq.6}$$

Another useful indicator is the Average Precision (AP), which corresponds to the area under the precision-recall curve (PRC) (**Eq.7**). This curve is obtained by plotting the precision versus the recall for various confidence levels of the network prediction. The PRC represents the influence of this confidence level on the relation between recall and precision.

$$\text{AP@}\alpha = \int_{0}^{1} \text{Precision(Recall) dRecall} \tag{Eq.7}$$

where α is the IoU threshold for which precision and recall are determined, and dRecall is the differential of the recall. By averaging the AP obtained for each class of the object detection task, the mean average precision (mAP) is obtained (**Eq.8**). However, in the case of this study, there was only one class and therefore the AP was identical to the mAP.

$$\text{mAP@}\alpha = \frac{1}{n} \sum_{i=1}^{n} \text{AP}_i \tag{Eq.8}$$

where n is the number of classes in the object detection problem.

Two mAP metrics were considered: the mAP@0.5 and the mAP@0.5:0.75. The mAP@0.5 is the AP with an IoU threshold of 0.5. The mAP@0.5:0.75 is the mean of

AP values for thresholds ranging from 0.5 to 0.75 with a step of 0.05. One of the interests of building so many various indicators is to increase the possibility to compare on a common ground the performances with other studies.

4.2 Ear segmentation

4.2.1 Automatic segmentation using DeepMAC

The state-of-the-art DeepMAC (Deep Mask-heads Above CenterNet) neural network model (Birodkar et al., 2022) was used to segment the ears in the bounding boxes. The huge advantage of this approach is that it did not require manual construction of training masks, which is even more arduous and time-consuming work than building bounding boxes. Thanks to the strong generalisation ability of the pre-trained DeepMAC model, no specific training was needed to segment the wheat ears in their bounding boxes. Ear masks were generated for each square sub-image (**Figure 18**). Then, the four sub-masks obtained associated with an image were transformed back to the original sub-image size of 1280×1024 pixels and brought together to form a mask of 2560×2048 pixels, having the same format as the original RGB image. The ears cut at the sub-images junctions were reconstituted when assembling the sub-masks (**Figure 19**). However, it happened that the parts of such cut ears did not match to the pixel, especially when the mask predicted by DeepMAC did not extend to the edge of the image. To solve this issue, pixels lying in a 10 pixel range between two ear pixels from either side of a junction were considered as belonging to the ear mask. This filling algorithm has visually proven its usefulness. The risk was not excluded to regroup ear parts that did not belong to the same ear but the probability of such errors was judged small enough to be neglected.
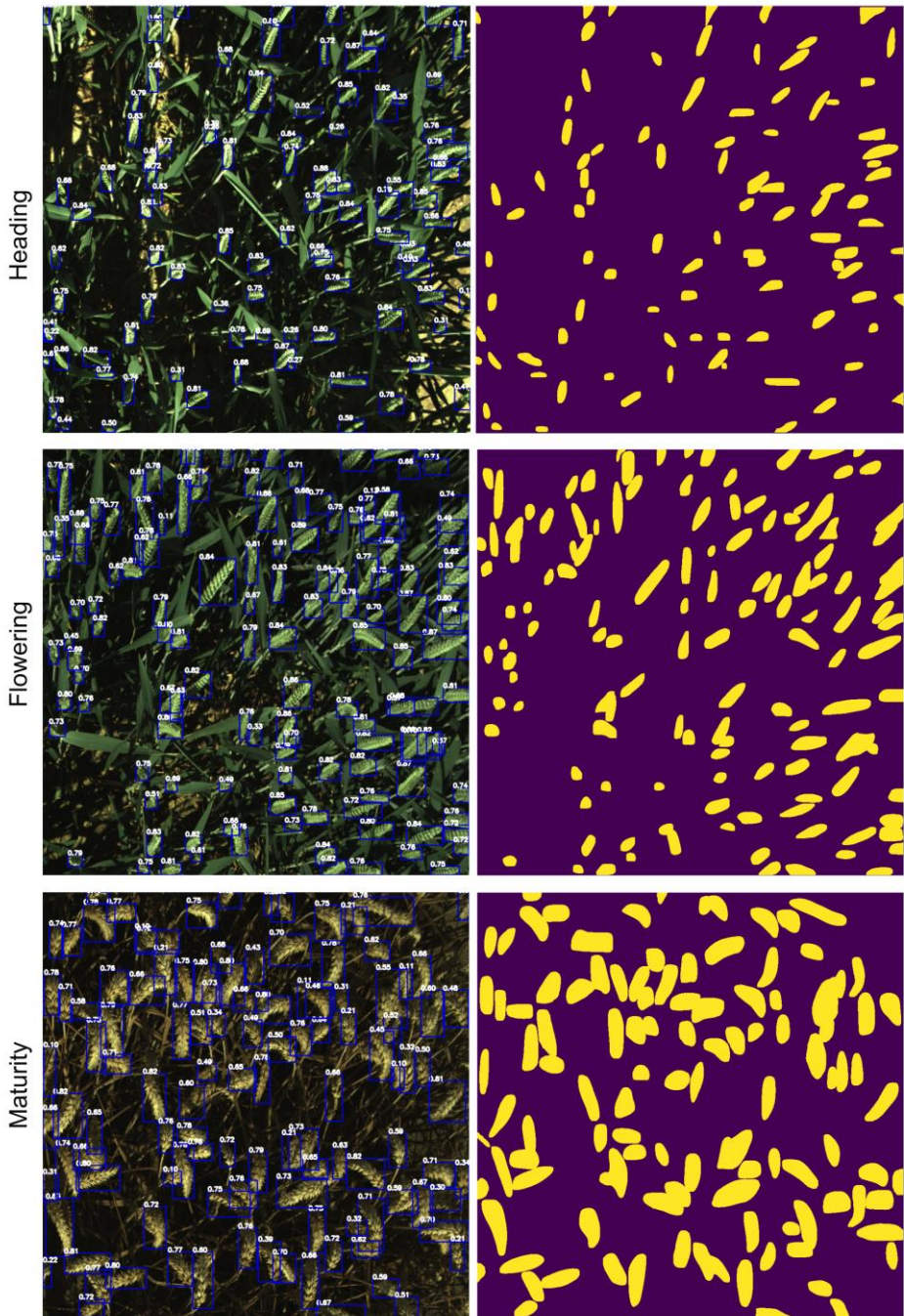
Figure 18. Ear bounding box detection and segmentation. The presented sub-images were randomly selected among the 20-F images from three dates: June 3rd (heading), June 15th (flowering) and July 29th (maturity). The size of the sub-images is 1024×1024 pixels.
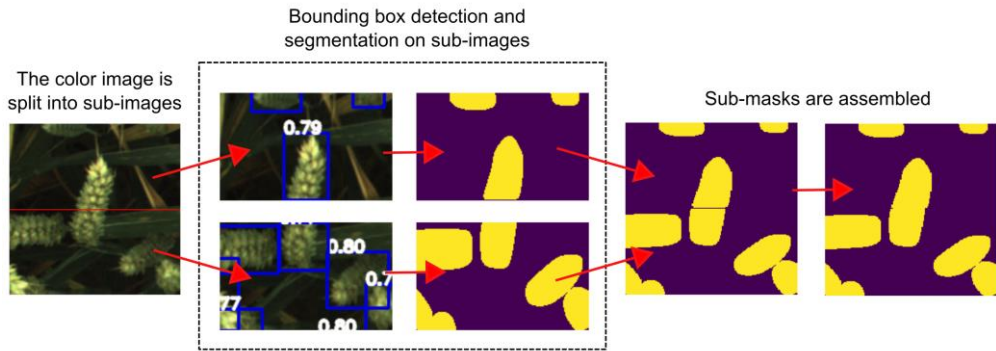
Figure 19. Pipeline of sub-image treatment. The process is illustrated for a zone crossed by a sub-image delimitation line. For the steps illustrated in the dotted rectangle, the size of the sub-images is slightly different because they had been resized to a square format

4.2.2 Validation of segmentation

A custom annotation tool was created (Carlier et al., 2022). The method consisted in annotating eighteen pixels at fixed coordinates in each RGB image. The tool automatically zoomed on each pixel. Then, the operator pushed one of three buttons to attribute a class to this pixel: class 1 for the background (soil, leaves, stems, …), class 2 for the ears, and class 3 if it was not possible to decide between class 1 or 2, for example if the operator did not distinguish well what the pixel represented or if the pixel was located at the edge between an ear and the background. That procedure was executed for all the images of all the acquisition dates.

4.3 Ear counts and density

4.3.1 Automatic method

The sum of the bounding boxes in each of the four sub-images provided a biased number of ears for the whole image because some ears were cut between two sub-images and thus counted twice. Those problematic ears were however reconstituted when gathering the four sub-masks. The excess ear parts were counted by the difference between the sum of the objects in the sub-masks and in the whole mask. The total number of ears was corrected by subtracting the excess ear parts.

The ear density, expressed in ears per square metre, is the ratio of the number of ears in the image to the footprint of the image at ear height. That footprint was obtained by **Eq.9**.

$$\text{footprint}_{ears} = 4 \, (z + 0.05)^2 \tan\left(\frac{\text{HFOV}}{2}\right) \tan\left(\frac{\text{VFOV}}{2}\right) \qquad \text{(Eq.9)}$$

where footprint_{ears} (m²) is the footprint at ear height, HFOV (°) is the horizontal field of view of the camera, VFOV (°) is the vertical field of view of the camera and z (m) is the distance between the camera and the tops of the ears. That distance was automatically measured by stereo vision, exploiting the shift between the images from the two RGB cameras. The stereo vision method is detailed in Chapter IV. The second implementation of the method, including the WLS filter, was used. The ear mask from the segmentation step (Section 4.2.1) was applied on the depth map obtained by stereo vision to produce a map of ear depths. As the ears were vertical most of the time, it was considered that the depth points were located at the tops of the ears. The distance z in **Eq.9** was the median of ear depths. It was increased by 0.05 m to account for the size of the ears, and thus estimate the image footprint in the middle of the ear layer.

4.3.2 Manual methods

Ears were counted by humans on July 6th, 2020 for the 64 micro-plots of trial 20-F. All the micro-plots were assessed by three human operators. For each micro-plot, an operator performed three counts, at about a quarter, half and three-quarter the length of the micro-plot. The row and the exact spot of the count were randomly selected by blindly dropping a wooden stick. This selection never occurred in the two first rows of the plot in which the ear density could have been influenced by a border effect. The stick had a length of 50 cm and was positioned along a row, at the basis of the tillers. All fully grown tillers on the row were considered to carry an ear, and counted. Ears were also counted in the laboratory on wheat samples from trial 20-F harvested on July 29th. Each sample was collected over 50 cm along three rows located halfway across the width of the plot. Knowing the row spacing, both in-field and post-harvest counts were converted to ear densities.

4.3.3 Statistical analyses of differences in ear density

A two-way analysis of variance with repeated measures was used for each trial to investigate the effect of date and fertilisation on the ear density measured by the automatic method. The pingouin Python library (version 0.5.0) was used for the implementation (Vallat, 2018). The within-subject factor was the date and the between-subject factor was the fertilisation. The interaction between the factors was significant so a one-way analysis of variance was performed for each date to compare the fertilisation scenarios. Post-hoc Tukey HSD tests were used.

A two-way analysis of variance was also led to investigate the effect of the measurement method on the ear density of trial 20-F. The two factors were the fertilisation scenario and the measurement method. This last factor had five levels: the automatic method, the reference manual method and the three human observations in the field. The date was not a factor as each manual measurement was performed at one date, and the automatic measurement was considered only for July 7th, which is the date closest to the human field observations. The interaction between the factors was significant so a one-way analysis of variance was performed for each fertilisation scenario to compare the measurement methods. Post-hoc Tukey HSD tests were used. All the tests in this study had a significance level of 0.05.

As the fields were quite homogeneous and the disease pressure was weak (in particular no fusarium head blight was spotted), random blocks and fungicide treatments were not considered as factors for the analyses. The purpose of those analyses was to get some clues to evaluate the ear density measurement methods, not to explain the smallest variations of ear density in our fields.

5. Results and discussion

5.1 Evaluation of ear bounding box detection

The performances of ear bounding box detection for the YOLOv5 model are detailed in **Table 8**. Depending on the date, the mAP@0.5 varied between 0.79 and 0.94 and the F1 score between 0.89 and 0.97, with the best performances reached for both trials at flowering stage. One detail that could ease detection at that stage is the presence of anthers on the ears.

Table 8. Evaluation of ear bounding box detection. The evaluation is based on all the labelled images of trials 20-F and 20-FP.The indicators are presented for each acquisition date, expressed as days after heading (DAH). The heading date for both trials was May 28th.

DAH	BBCH stage	Trial	TP (%)	FP (%)	FN (%)	Precision	mAP@ 0.5	mAP@ 0.5:0.75	Accuracy	F1 score
5	55: Heading	20-FP	84.67	5.93	9.4	0.93	0.84	0.79	0.85	0.92
6	59: Heading	20-F	87.53	3.76	8.71	0.96	0.87	0.86	0.88	0.93
12	65: Flowering	20-FP	94.25	2.24	3.51	0.98	0.94	0.93	0.94	0.97
14	67: Flowering	20-F	92.99	2.5	4.51	0.97	0.93	0.93	0.93	0.96
18	69: Flowering	20-F	90.95	2.92	6.13	0.97	0.91	0.90	0.91	0.95
19	69: Flowering	20-FP	89.54	4.17	6.29	0.96	0.89	0.86	0.90	0.94
21	71: Watery ripe	20-F	91.47	3.35	5.18	0.96	0.92	0.91	0.91	0.96
26	75: Medium milk	20-F	87.09	4.23	8.68	0.95	0.87	0.85	0.87	0.93
29	77: Late milk	20-FP	91.68	3.31	5.01	0.97	0.92	0.88	0.92	0.96
40	83: Early dough	20-F	87.81	4.04	8.15	0.96	0.87	0.85	0.88	0.94
40	83: Early dough	20-FP	87.66	4.13	8.21	0.96	0.88	0.87	0.88	0.93
46	85: Soft dough	20-FP	84.84	4.75	10.41	0.95	0.84	0.82	0.85	0.92
55	89: Maturity	20-FP	79.73	6.42	13.85	0.93	0.79	0.76	0.8	0.89
62	89: Maturity	20-F	81.61	5.35	13.04	0.94	0.81	0.79	0.82	0.9

5.2 Segmentation quality

The performances of ear segmentation are detailed in **Table 9**. As for bounding box detection (Section 5.1), the lowest performances were observed at heading and maturity stages, but even in these cases the F1 scores were close or superior to 0.75, which is considered good. The performances in this study were superior to the ones announced by Carlier et al. (2022) using a non deep learning segmentation technique on a part of the same dataset.

The DeepMAC model, thanks to its strong generalisation ability, performed ear segmentation without having been trained for this specific task. There is however a downside to this ability: sometimes the algorithm segmented another object in the bounding box, rather than the ear. This phenomenon is illustrated in **Figure 20**. To avoid those issues, future ear segmentation pipelines should include some training specific to the ear detection task, or at least some criteria to exclude the wrong bounding box segmentations.
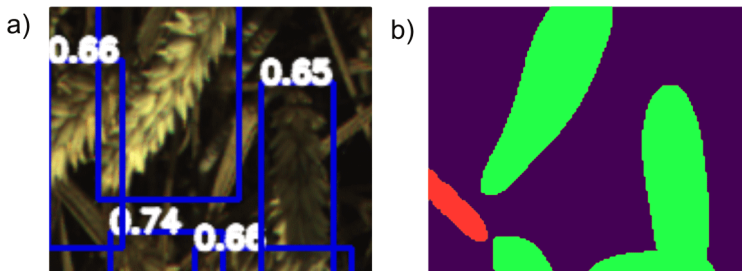
Figure 20. Example of wrong ear segmentation. a) RGB image with the detected bounding boxes. b) Ear mask with correct mask parts in green and erroneous parts in red. In the bounding box on the left, DeepMAC isolated a leaf rather than the ear.

Table 9. Evaluation of ear segmentation. The evaluation is based on all the images of trials 20-F and 20-FP. The indicators are presented for each acquisition date, expressed as days after heading (DAH). The heading date for both trials was May 28th.

DAH	BBCH stage	Trial	TP (%)	FP (%)	FN (%)	TN (%)	Precision	Accuracy	F1 score
5	55: Heading	20-FP	4.01	0.78	1.80	93.41	0.84	0.97	0.76
6	59: Heading	20-F	4.73	0.72	2.50	92.05	0.87	0.97	0.75
12	65: Flowering	20-FP	7.14	0.72	0.18	91.95	0.91	0.99	0.94
14	67: Flowering	20-F	7.60	0.78	0.24	91.38	0.91	0.99	0.94
18	69: Flowering	20-F	10.98	1.39	1.07	86.57	0.89	0.98	0.90
19	69: Flowering	20-FP	9.27	0.71	1.81	88.21	0.93	0.97	0.88
21	71: Watery ripe	20-F	10.91	0.94	1.50	86.66	0.92	0.98	0.90
26	75: Medium milk	20-F	9.04	1.47	1.01	88.48	0.86	0.98	0.88
29	77: Late milk	20-FP	10.46	2.94	0.39	86.21	0.78	0.97	0.86
40	83: Early dough	20-F	15.41	3.45	0.67	80.48	0.82	0.96	0.88
40	83: Early dough	20-FP	17.21	0.86	2.46	78.98	0.95	0.97	0.91
46	85: Soft dough	20-FP	9.21	5.83	0.79	84.17	0.61	0.93	0.74
55	89: Maturity	20-FP	26.08	2.40	7.58	63.94	0.92	0.90	0.84
62	89: Maturity	20-F	22.01	2.40	4.61	70.98	0.90	0.93	0.86

5.3 From ear count to ear density: a non-trivial conversion

The ear density was computed by dividing the number of ears detected by the image footprint at ear height. The temporal evolution of the measured ear density is presented in **Figure 21** for trials 20-F and 20-FP. The two-way analysis of variance showed for both trials an interaction between the two factors: date and fertilisation scenario. For that reason, the fertilisation scenarios were statistically compared separately for each date. Almost all of the corresponding one-way variance analyses showed a significant effect of the fertilisation on ear density. The results of the post-hoc Tukey HSD test are presented in **Table 10** and **Table 11**, respectively for trial 20-F and 20-FP.
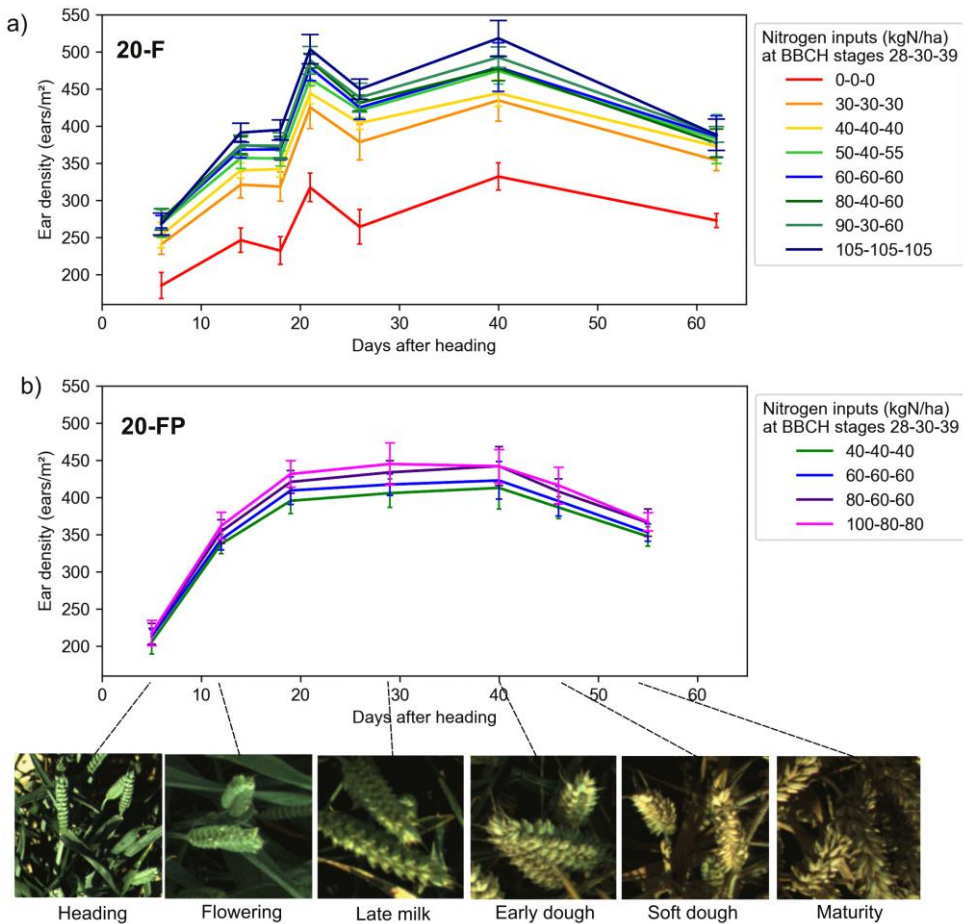
Figure 21. Dynamics of the measured ear density for trial 20-F and 20-FP. The standard deviation is indicated on each point. Various cap sizes are used to differentiate the standard deviation bars from different fertilisation scenarios.

Table 10. Tukey HSD test to investigate the effects of the fertilisation scenarios of trial 20-F on the wheat ear density measured by the automatic method. The test was performed for each date, expressed in days after heading (DAH), corresponding to one row in the table. On each row, fertilisation scenarios marked with at least the same letter are not considered statistically different. The significance level is 0.05.

DAH	Nitrogen inputs (kgN/ha) at BBCH stages 28-30-39							
	0-0-0	30-30-30	40-40-40	50-40-55	60-60-60	80-40-60	90-30-60	105-105-105
6	c	b	ab	a	a	a	a	a
14	e	d	cd	bc	b	ab	ab	a
18	e	d	cd	bc	b	ab	ab	a
21	e	d	cd	bc	ab	ab	ab	a
26	e	d	cd	bc	abc	ab	ab	a
40	e	d	cd	bc	b	bc	ab	a
62	b	b	ab	ab	ab	ab	a	a

Table 11. Tukey HSD test to investigate the effects of the fertilisation scenarios of trial 20-FP on the wheat ear density measured by the automatic method. The test was performed for each date, expressed in days after heading (DAH), corresponding to one row in the table. On each row, fertilisation scenarios marked with at least the same letter are not considered statistically different. The significance level is 0.05.

DAH	Nitrogen inputs at BBCH stages 28-30-39			
	40-40-40	60-60-60	80-60-60	100-80-80
5	a	a	a	a
12	c	bc	a	a
19	c	bc	ab	a
29	c	bc	ab	a
40	b	ab	a	a
46	c	bc	ab	a
55	c	c	ab	a

The measured ear densities showed significant increases with the fertilisation level of the crop (**Table 10** and **Table 11**), which is an expected behaviour (Oscarson, 2000): the first nitrogen input, at tillering, favours the tiller number while the second nitrogen input, at stem elongation, allows the tillers to grow and develop the ear. This seems to prove that, at almost all of the growth stages, the method is useful to highlight relative ear density differences between fertilisation scenarios. Nevertheless, for a given fertilisation object, the measured ear density varied between the dates (**Figure 21**). Yet, no incident such as lodging was observed in the field and thus the number of ears was not supposed to change. That said, some of the ear density variations were expected. The smaller densities measured at heading can be explained by the fact that not all ears did emerge from their sheaths yet. The smaller densities measured at maturity can be explained by the change of ear morphology and position: at this stage, the ears were larger and bent from the weight of the grains, which increased the number of ears totally hidden in the image. In addition, the bounding box detection at those stages was not as good as for the others (**Table 8**), probably because of many ear overlaps. That lower performance is translated by a number of ears that were not detected by the model, leading to an underestimation of the ear density. A more troubling point for trial 20-F is the variation of the measured ear density at the other development stages, where the ears were well visible and the bounding box detection presented excellent performances with mAP@0.5 close or superior to 0.9.

A first hypothesis to explain those ear density differences across dates is the effect of scene lighting. This may be hard to believe since the deep learning algorithm succeeded well in detecting the ears in both sunny and cloudy conditions. However, the bounding box detection performances could be a misleading track. Direct sunlight induces shadows and strong contrasts in wheat canopy images. In those conditions, some low ears could stand in dark areas and be missed at both the human annotation and the automatic detection step. An observation of this kind was made by Madec et al. (2019). Looking at the trial 20-F data in **Figure 21**, the ear densities measured at 26 days after heading (June 23rd) in sunny conditions are lower than the densities measured at the dates after and before, in cloudy conditions. The ear densities measured at 18 days after sowing are also lower than expected, and this coincides one again with sunny acquisitions. A solution to avoid ears missed by both human image annotators and the deep learning algorithm would be to deal with the high contrasts responsible for image zones either too dark or saturated. It could be achieved by acquiring the same scene using various exposure times to build high dynamic range images.

A second hypothesis to explain the ear density differences across dates is the variability inherent to the image footprint estimation. That footprint depends on the distance between the cameras and the ears (**Figure 22**), which may vary depending on the wheat varieties, the fertilisation level or even the depth of the ruts between the micro-plots in which the image acquisition platform was placed. For this reason, the distance needs to be measured for each imaged wheat canopy zone. But, as highlighted by David et al. (2020), the definition of this distance is still an open question. Which depth in the ear layer should be considered to compute the image footprint? And how to measure that depth? In this study, it was considered the median depth of the tops of the ears plus 0.05 m, but this choice was quite arbitrary and did not account for the possible inclination of the ears. The curve in **Figure 22** (b) illustrates the variations of image footprint expected for a modification of the distance between the cameras and the ears.
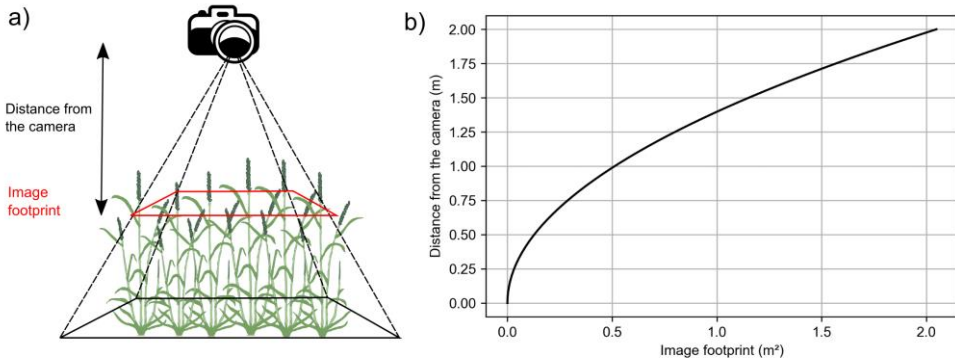
Figure 22. Notion of image footprint. a) Illustration of mage footprint at ear height. b) Relation between the image footprint at ear height and the distance between the ears and the camera.

5.4 Human and deep learning measurements of ear density

The comparison between the ear density measured by the algorithms, the human operators in the field and the reference measurement from samples counted in the laboratory is presented in **Figure 23** for the eight fertilisation scenarios of trial 20-F. The highest ear densities were recorded by the reference method. For all the scenarios, operator 3 recorded the lowest values. The two-ways variance analysis showed an interaction between the two factors: measurement method and fertilisation scenario. For that reason, the methods were statistically compared separately for each fertilisation level. Almost all of the corresponding one-way variance analyses showed a significant effect of the measurement method. The results of the post-hoc Tukey HSD tests are presented in Table 5. In most of the scenarios, operators 1 and 3 were significantly different from the reference, while operator 2 was only different from the reference in one scenario. The significant differences between the operators demonstrated an operator bias. Deep learning was never judged different from operator 2 or the reference. Both deep learning and operator 2 were reliable but, looking at their mean in each scenario (**Figure 23**) and the number of scenarios in which they were assumed equal to the reference (Table 7), deep learning was considered a better estimator.

Considering the ear densities averaged by fertilisation scenario, the correlation coefficients between the reference method and operator 1, operator 2, operator 3 and deep learning were respectively 0.88, 0.94, 0.83 and 0.90. Using this indicator, operator 2 slightly outperformed deep learning.
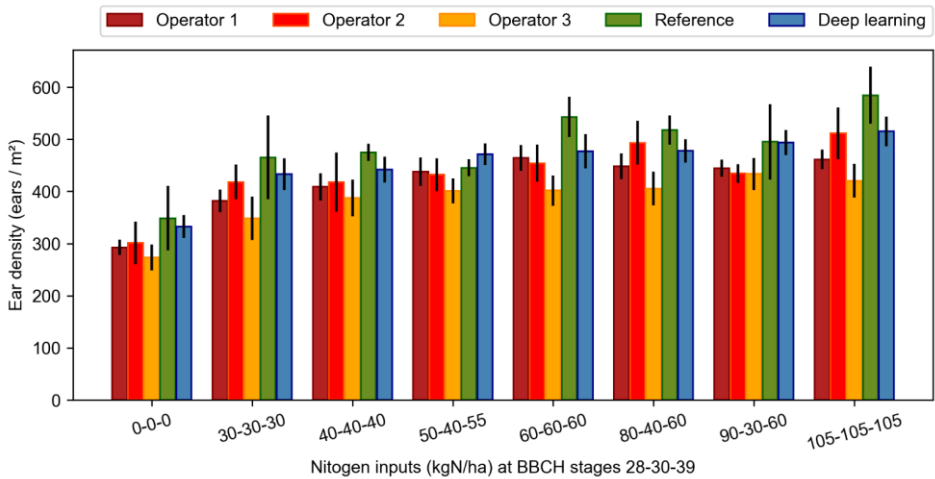
Figure 23. Comparison of five estimations of the ear densities of trial 20-F. The five approaches were: observations made by three human operators in the field (July 6th), deep learning (images from July 7th) and the reference measurement carried out on harvested wheat samples (July 29th). The standard deviation is indicated on each bar.

Table 12. Tukey HSD test to investigate the effect of the measurement method on wheat ear density. The test was performed for each fertilisation level, corresponding to one row in the table. On each row, measurement methods marked with at least the same letter are not considered statistically different. The significance level is 0.05.

Nitrogen inputs (kgN/ha) at BBCH stages 28-30-39	Measurement method				
	Operator 1	Operator 2	Operator 3	Deep learning	Reference
0-0-0	ab	ab	b	a	a
30-30-30	ab	a	b	a	a
40-40-40	bc	abc	c	ab	a
50-40-55	ab	ab	b	a	ab
60-60-60	b	bc	c	ab	a
80-40-60	bc	ab	c	ab	a
90-30-60	a	a	a	a	a
105-105-105	bc	ab	c	ab	a

For operator 1, operator 2, operator 3, deep learning and reference method, the average coefficients of variation for a fertilisation scenario were respectively 0.05,

0.09, 0.08, 0.06 and 0.09. All those coefficients were close, but deep learning and operator 1 had the advantage with the smallest values. For deep learning, this may be explained by the ability of the method to better deal with the heterogeneity within the micro-plots. Both the counting in the field and the counting on harvested plants were based on small zones in the micro-plots and could have been performed on areas not representative of the average ear density of the scenario. On the contrary, the images allowed measurements on larger zones and all along the micro-plots. To allow a better comparison between the deep learning and reference measurements, Madec et al. (2019) suggested working on the exact same zones in the micro-plots. They also proposed to increase the size of the sampled reference zones, which is a laborious task.

A point of attention would be the scenarios with low fertilisation. Some tillers may develop less and cause more ears located lower in the canopy, and thus difficult to spot on images. In those conditions, the manual measurements could be advantaged over images. It was not observed at the date of the comparison between the methods (July 7th, at early dough stage), where deep learning was not judged significantly different from the reference for the scenario with 0 nitrogen input. This date was chosen because it was close to the counting made by the human operators (July 6th), but also because it was considered as the most reliable date: performances of ear detection were excellent, the sky was cloudy and ear densities measured agreed with other reliable measurements sooner in the season. The detection issue with low ears could however happen earlier in the season, or in other conditions, for example in the presence of shadows and highly reflective leaves due to direct sunlight.

For deep learning, a critical question is the choice of the acquisition date to obtain the ear density values. In this study, the method did not yield the same ear densities at all the dates. July 7th, at early dough stage, was used for comparison with human measurements. Nevertheless, values from another date would have modified the comparison. This latter point highlights that, if human operator measurements can show ear density discrepancies on a same micro-plot, so does the automatic method. In the future, it will be important to work on the weak points of the method: the estimation of image footprint and the ears non visible because of shadows or overlaps. As only one trustworthy ear density measurement is needed for a season, the camera operators have the luxury to select convenient conditions or growth stages.

6. Conclusions

The combination of the YOLOv5 bounding box detection model and the DeepMAC segmentation model brought an innovative solution for both counting and segmenting the wheat ears. Thanks to the semi-supervised learning of the bounding box detection model, the use of a rich existing dataset and the strong generalisation ability of the segmentation model, few efforts were necessary to train the whole method and adapt it to reach robust performances at all development stages of wheat from heading to maturity. Most of the labelling work has been dedicated to build a strong validation dataset. All the labels were made available to the community. The average F1 score of bounding box detection was 0.93 and the average F1 score of segmentation was 0.86. A limitation of this study is the diversity of the acquired data set. It covered all the development stages of the crops from heading to maturity and included contrasted fertilisation scenarios, but it was limited to one growing season and two varieties. An improvement would be to acquire similar time series on a larger diversity of wheat varieties, especially including ears with awns. It would also be interesting to investigate the detection performances on ears from other cereal crops such as barley or rye.

Counting the ears in an image is not sufficient to provide exploitable agronomic information. A method was proposed to convert ear counts to ear density, i.e. the number of ears per square metre, which is the variable of agronomic interest and which has been widely neglected in previous studies dedicated to ear detection. The method relied on a second RGB camera to determine the distance between the cameras and the ears by stereo vision and thus compute the image footprint at ear height. The most reliable deep learning ear density values outperformed observations from human operators in the field. They had the advantage to better capture the spatial heterogeneity of the micro-plots compared to human observations and reference harvests that were very localised. Further studies should not focus solely on ear counting methods based on existing datasets but also set field experiments and investigate the retrieval of ear density. A particular focus should be put to establish a robust reference distance between the camera and the ears to compute the image footprint.

VI

Multimodal image registration

1. Synopsis

This chapter is largely based on the research paper of Dandrifosse et al. (2021):

It tackles the automatic registration of multimodal wheat canopy images, i.e. the pixel to pixel alignment of those images. The task is complex, due to the depth of the studied scene and the close distance to the cameras, depending on the canopy architecture.

Overcoming this issue is the key to the fusion of multimodal images acquired by modern field phenotyping platforms. It will open new doors to the joint-processing of multimodal images and allow the extraction of plant traits at the organ scale (Chapters 7 and 8).

2. State of the art

In recent years, close-range multi-sensors platforms and vehicles have been developed for crop phenotyping in natural conditions. The obvious interest of multi-sensor approaches lies in the ability to measure an increased number of pertinent traits. This is especially crucial when studying plant stresses whose symptoms are often complex and not determined by a single physiological or morphological component. For this reason, the philosophy for most modern field phenotyping platforms is to measure both physiological and morphological traits. This requires several types of sensors. On the one hand, spectrometers and 2D imagers provide plant reflectance (visible, NIR, thermal IR, …) relative to physiological information. On the other hand, 3D cameras and LiDAR devices provide morphological information. Platforms combining such sensors are described by Kirchgessner et al. (2017), Shafiekhani et al. (2017), Virlet et al. (2017), Jiang et al. (2018), Bai et al. (2019), Beauchêne et al. (2019) and Pérez-Ruiz et al. (2020). Each sensor of the platform provides a number of plant traits related to the observed scene. Then, analyses exploit traits from the different sensors to generate agronomic knowledge. This is what is habitually called "data fusion". In this generic pipeline, the fusion of information from the different sensors takes place after the extraction of plant traits. However, the complementary nature of the information from the different sensors may also be exploited before that

step of traits extraction. This is where the process of image fusion comes into play, as illustrated in **Figure 24**. Instead of considering separately the images of different cameras (RGB, monochrome, thermal, depth, …), those images could be fused at the pixel level to enrich the available information (Jiang et al., 2018). Such a fusion would allow to segment more finely the images and extract plant traits at a finer spatial scale. Instead of separating only leaves and background, the fusion of data for each pixel may allow to identify upper leaves, lower leaves, sick tissues, wheat ears, … Then, each trait could be computed for those different organs instead of for the whole canopy. This would for example solve a well-known issue of close-range thermal imaging: isolating leaves of interest for water status assessment (Leinonen et al., 2004). Image fusion would also allow to disentangle the effects of leaf morphology and physiology on light reflection. This could be obtained by fusing leaf angles from depth map and reflectance maps. Such orientation-based reflectance has been suggested to improve thermal imaging by Jerbi et al. (2015). It is to notice that this chapter only envisions the fusion of images, implying that 3D information is provided as a depth map and not as a 3D point cloud.
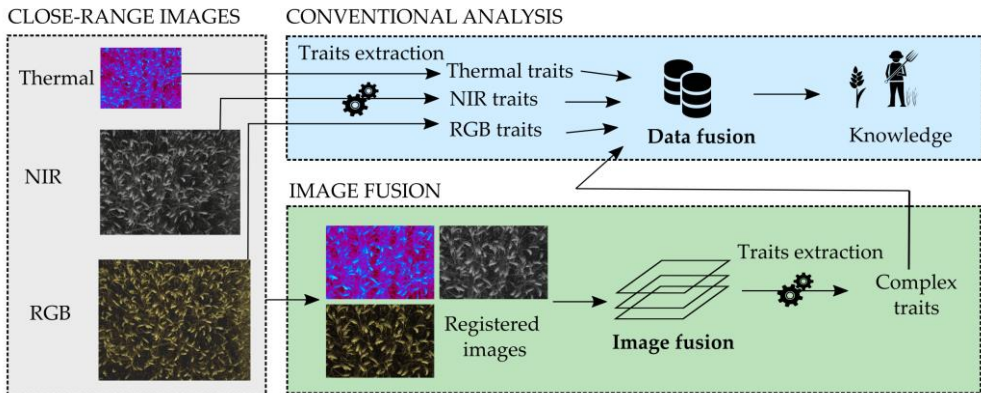
Figure 24. Difference between conventional data fusion and image fusion.

In the context of phenotyping platforms already equipped with different types of cameras, multimodal images fusion may be an asset that does not demand supplementary material investment. It offers the possibility to fully benefit from the spatial information brought by cameras, in comparison to non-imaging devices such as thermometers or spectrometers. It also overcomes the disadvantages of mono-sensor multispectral and hyperspectral cameras. Despite those many benefits,

multimodal image fusion was rarely implemented in close-range systems, probably due to the difficulty to overlay the images from the different sensors.

This alignment step is called image registration. Considering two images of a same scene acquired by two cameras, the registration consists in geometrically transforming one image – the slave – so that the objects of the scene overlay the same objects in the other image – the master. The registration can be divided into two main steps: the matching between the slave and the master images and the transformation of the slave image. In general, multimodal images registration is a complex problem because the cameras present different spatial positions, different fields of view and different image sizes. Additionally, the multimodal nature of the images implies that they present different intensity patterns, which complicates the matching. In the domain of in-field plant phenotyping, registration is even more challenging due to i) the nature of the crops (wheat leaves are complex overlapping objects arranged in several levels) and ii) the natural conditions (sunlight generates shadows and wind induces leaf movement). Most of the studies that included close-range plants registration concerned thermal and RGB images. It is to note that some commercial cameras are able to acquire both RGB and thermal images that are roughly aligned (Jiménez-Bello et al., 2011). However, plant researchers relied most of the time on separated cameras and had to deal themselves with the registration step. The most basic approaches were the manual selection of matching points in the slave and master images (Leinonen et al., 2004; Möller et al., 2007) or directly the empirical choice of an unique transformation for all the images (Bai et al., 2019). An automatic method was developed by Wang et al. (2010) to align thermal and RGB images of side-viewed grapevines. In another study, Jerbi et al. (2015) solved the problem of RGB-thermal registration for maize images. They validated the method using a heated chessboard. The error measured on a simple pattern such as a chessboard may however not be representative of the error occurring in a complex crop canopy image. Matching is far more complex on plant structures. Even assuming an optimal matching, the different points of view of the cameras may lead to additional errors caused by parallax and visual occlusions. This implies that measuring the registration errors is a difficult task and that the distortion between the images is often complex. None of the plant registration approaches presented here above succeeded in taking into account local distortion in the images. Indeed, those approaches relied on global transformations, i.e. functions for which the mapping parameters are the same for the entire image (Zitová et al., 2003). Yet, the parallax effect alone makes the distortion dependent on the distances of the objects. When acquiring multimodal images from aerial vehicles,

this effect is negligible because of the huge distance between the cameras and the scene compared to the displacement of the optical centres and the distance between the objects themselves (Rabatel et al., 2016). At close-range, local distortion between the images, and especially parallax effect, may have a significant impact on registration quality. A possible track to solve the issue would be to use a local transformation, i.e. a transformation that is able to locally warp the slave image. The use of such transformations is very scarce in the field of plant sciences but common in other fields such as multimodal medical imaging (Klein et al., 2010; Sotiras et al., 2013). Local transformations on images of potted-plants were used by De Vylder et al. (2012) and Raza et al. (2015). To our knowledge, no study provided a solution for the registration of close-range wheat canopy images. Nevertheless, Henke et al. (2019a, 2019b) studied wheat image registration under controlled conditions and on isolated potted-plants. They tested three matching methods on side-viewed wheat to align fluorescence and RGB images. Their study stuck to a global transformation.

3. Goal and structure

The main objective of this chapter is to solve the challenge of automatic registration of close-range multimodal wheat canopy images in field conditions, assuming no targets or markers on the plants.

The method is detailed in Section 4. Eight registration approaches were studied to relate the slave and master images: one approach based on a calibration accounting for the cameras system geometry and seven approaches based on the content of the images. Both global and local transformations were investigated. A rigorous validation of the methods was performed. In Section 5, the best methods are highlighted regarding several scenarios and some solutions are advanced to deal with the remaining alignment errors. Section 6 discusses the performances of matching algorithms and the choice of the transformation model. It also provides a deeper look on the different natures of distortion between images of a same scene. Finally, it expands on the challenges of registration quality evaluation.

4. Method

4.1 Choice of the master camera

The multi-sensor system was composed of two RGB cameras, a multispectral camera array and a thermal camera. It is detailed in Chapter III, Section 5. Prior to any registration attempt, it was necessary to determine which camera was the master, i.e. the camera providing reference images. The other cameras are considered as slaves. The goal of registration is to find the transformations to apply to slave images so that they are aligned with the master image. The camera of the multispectral array equipped with the filter centred at 800 nm, called the 800-nm camera, was chosen as the master. This choice was made because i) the camera occupied a central position on the sensor pod, ii) the filter allowed to segment the images and provide plant masks, which are crucial to extract plant traits and iii) the 800-nm filter clearly highlighted leaves, which could have been important to favour matching. Concerning the two RGB cameras, only one of the two cameras was considered as a slave of the 800-nm master camera: the one that was the closest to it. The images of this RGB camera were cropped to remove the zone not seen by the second RGB camera.

4.2 Wheat canopy images

The registration methods were evaluated on the images acquired in trial 20-F on May 7th, May 14th, May 20th, May 27th, June 2nd, June 11th, June 23rd, July 7th and July 29th and in trial 20-FP on May 12th, May 18th, June 2nd, June 9th, June 16th, June 26th, July 13th, July 22nd.

4.3 Calibration-based registration method

That first image registration method is based on the hypothesis that, for a given configuration of cameras, the best global transformation to register images only depends on the distance between the objects of interest in the scene and the cameras. It relies on a calibration step to establish the distance dependent transformation matrix (DDTM) between the images. This DDTM allows to express the coefficients of a global transformation as distance-dependent functions (Berenstein et al., 2015). To be as general as possible, the considered approach was the global transformation with the most degrees of freedom, i.e. a homography (**Eq.10**). This transformation takes into account rotation, translation, shear and scale. Moreover, the scale factors depend on

the pixel position in the image which allows to deal with perspective differences (Rabatel et al., 2016).

$$\begin{bmatrix} x_2 \\ y_2 \\ 1 \end{bmatrix} = DDTM(d) \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \; ; \; DDTM = \begin{bmatrix} h_{00}(d) & h_{01}(d) & h_{02}(d) \\ h_{10}(d) & h_{11}(d) & h_{12}(d) \\ h_{20}(d) & h_{21}(d) & 1 \end{bmatrix} \qquad (Eq.10)$$

where x_2 and y_2 are the registered coordinates of a pixel of coordinates x and y in a slave image, d is the distance of the object of interest and h_{ij} are the eight independent coefficients of the transformation matrix.

The calibration step was performed in the laboratory using the thermal calibration target described in Chapter III, Section 5.2. This target was made of heated black disks inserted on a cooled white plate. It was captured at distances ranging from 1 m to 2.2 m by steps of 0.05 m. As the removable disks were of different temperature and of different colour that the main body of the target, it was possible to detect the centroids of the disks in images from all the cameras. Those centroids served as key points to determine the best transformation at each distance (**Figure 25**).
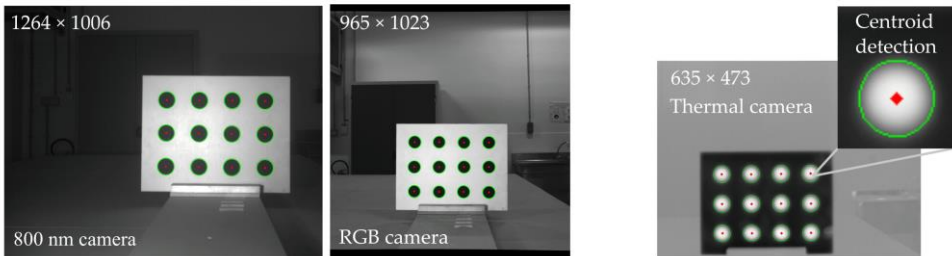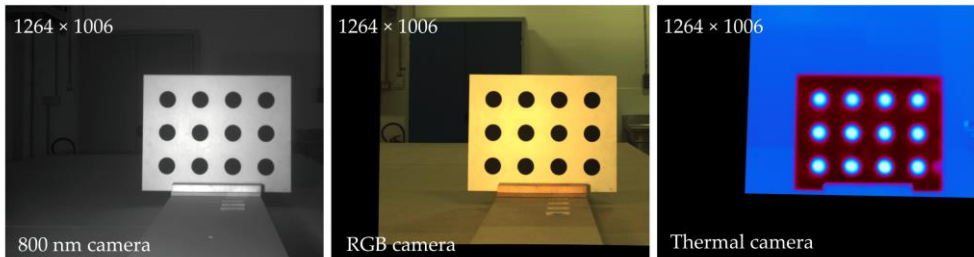
Figure 25. Registration of NIR (800 nm), RGB and thermal images of the calibration target. The 800-nm image is the master that is used as a reference to align the other images.

For each camera, the values of the eight coefficients providing the best transformations were related to the distance of the target using linear regressions. An example of such a relation is given in **Figure 26** for the DDTM that linked the thermal images to the master images. The same trend was observed for all the cameras. Only h_{02} and h_{12} significantly varied with distance. For the h_{02} regressions corresponding to the 490, 550, 680, 720, 900-nm, RGB and thermal cameras, the determination coefficients (R^2) were respectively 0.00 (no change with distance), 0.85, 0.65, 0.86, 0.79, 0.01 and. 0.94. The RMSE were respectively 5.2, 8.4, 7.1, 7.7, 4.8, 8.6 and 4.2 pixels. For the h_{12} regressions, R^2 were respectively 0.75, 0.03, 0.0, 0.84, 0.86, 0.81 and 0.90. The RMSE were respectively 5.9, 8.0, 8.1, 5.1, 4.5, 7.4 and 2.7 pixels. The other coefficients were approximated to constants by considering the median of the measured values. h_{20} and h_{21} of all the matrices were close to 0, which is the case in an affine transformation matrix. This implies that the affine transformation model could have been used.
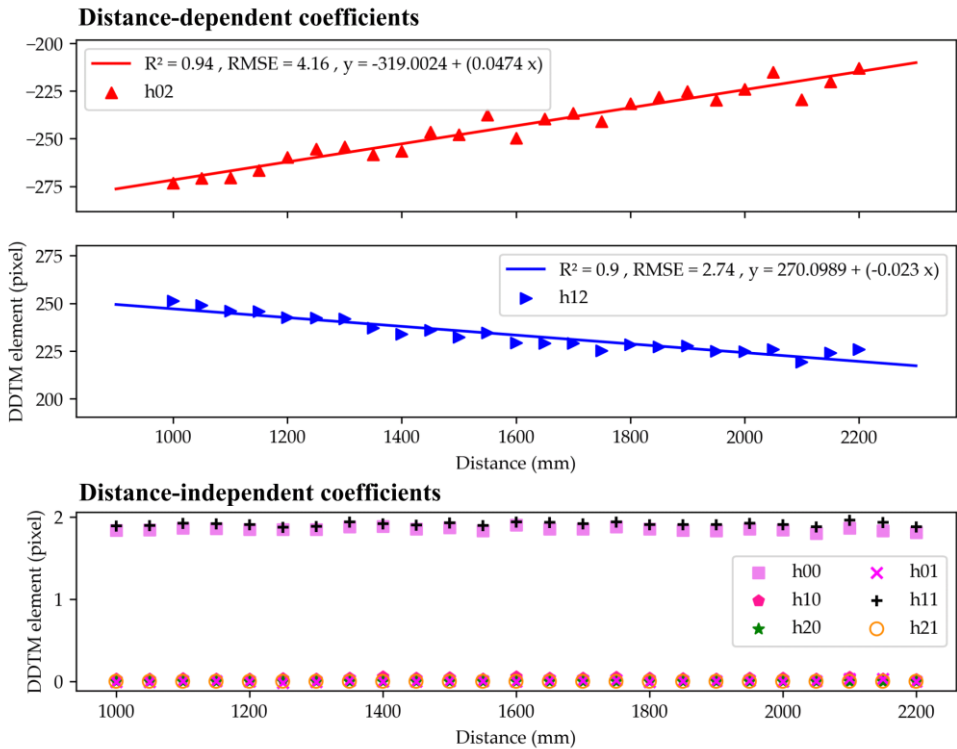
Figure 26. Relation between the distance of the target and the coefficients of the transformation matrix of a homography.

Applying the DDTM method to wheat canopy images required measurements of median distances of wheat elements. For the images acquired before heading stage (BBCH 50), that distance was measured by stereo vision, using the second implementation of the process as described in Chapter IV. For the images acquired after heading stage, that distance was approximated based on manual measurements using a stick metre. In that scenario, instead of computing the median distance of wheat for each micro-plot, the approximation was used for all the plots. This choice was made in 2020 because stereo vision was judged unreliable for several dates. Afterwards, we discovered it was because of a bad choice of calibration parameters. The method could have relied on stereo vision measurements of wheat distance at all the dates, and thus be fully automatic.

4.4 Image-based registration methods

Instead of deducing the transformation from the relative positions of the cameras, those methods exploit similarities between the contents of the slaves and the master images to find the best transformation. They do not need a calibration, although some of them need prior information on the transformation or initial alignment. They allow to take into account the nature of the scene. The diversity of images-based registration methods is described in existing reviews (Zitová et al., 2003; Xiong et al., 2010). An overview of the registration pipeline and of the different methods is presented in **Figure 27**. The first step of registration is called the matching (**Figure 27**. Step I). The aim is to detect corresponding zones in the master and the slave images. That correspondence may be features-based or area-based. In features-based methods, the goal is to identify a set of features – points, lines or patterns. The sets of features of the slave and the master images are compared to find matches. Popular methods exploit point features that are robust to scale and rotation changes. In the area-based methods, no features are detected. All efforts are put to the matching between the images, or two windows from those images, through the maximisation of a similarity metric such as cross-correlation coefficient or mutual information. After establishing a correspondence, the second step of registration is to determine the geometric transformation to apply to the slave image (**Figure 27**. Step II). Transformations are divided into global and local methods. Global methods use the same mapping parameters for the entire image while the local methods are various techniques designed to locally warp the image. If there is no distortion between the images, rotation and translation are sufficient to align them. Otherwise, hypotheses on the distortion should be established to select either another global transformation –

similarity, affine, homography – or a local transformation. For complex distortion, a possible approach is to begin with a global transformation and then to refine the registration using one or several local methods. Once coordinates in the slave image have been remapped, the last step consists in resampling the image to compute the new intensities (**Figure 27**. Step III). It involves convolutional interpolation algorithms such as nearest neighbours, bilinear (based on four neighbours) or cubic (based on sixteen neighbours) (Xiong et al., 2010). Despite the development of more complex resampling approaches slightly outperforming the traditional ones, it is often sufficient to stick to the simple bilinear or cubic algorithm (Zitová et al., 2003). Registration approaches are mainly differentiated by the choice of the matching method and of the transformation model.
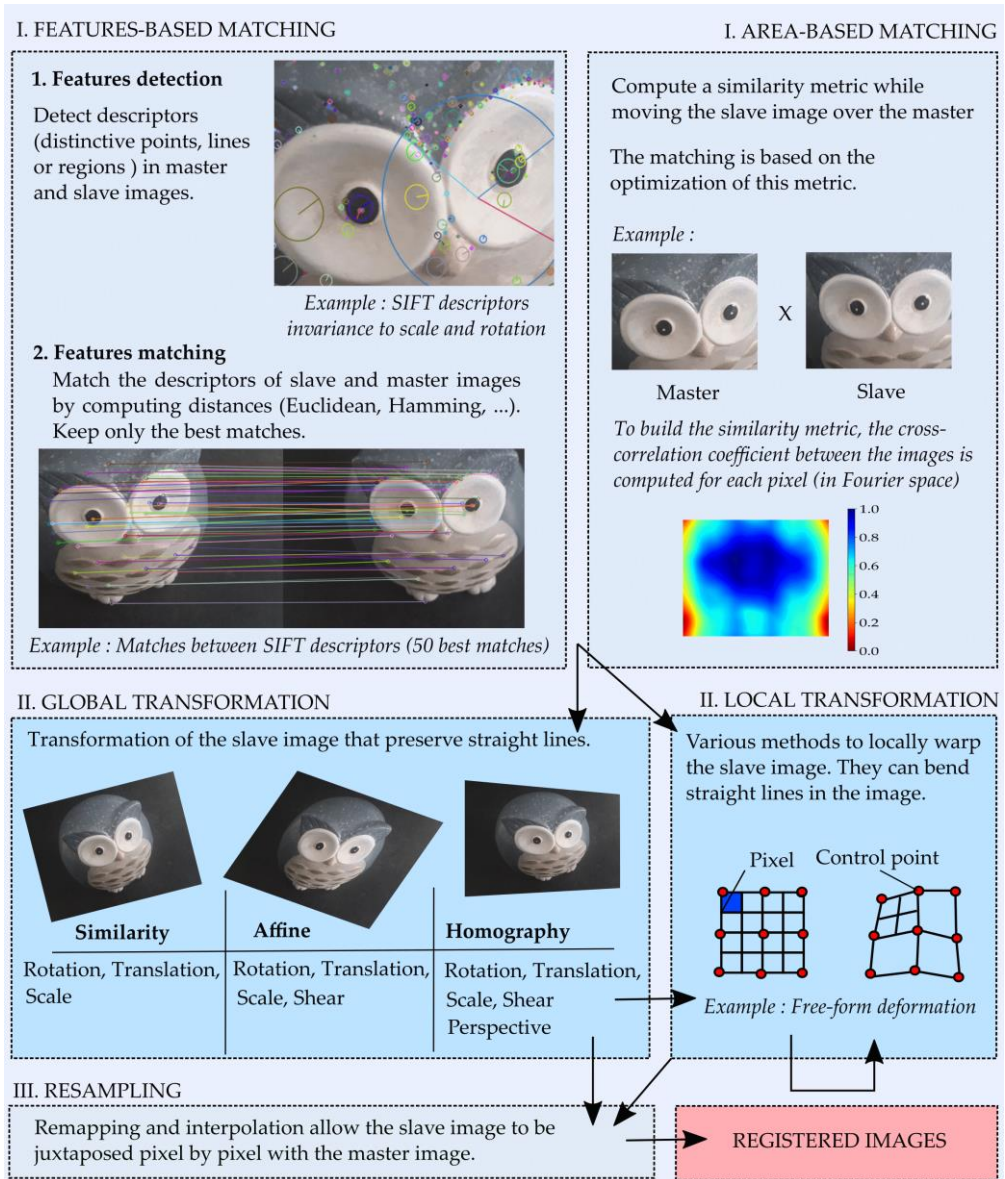
Figure 27. Overview of image-based registration pipeline.

The registration methods tested in the frame of this study are summarised in **Table 13**, also including the DDTM calibration-based method. The idea was to test methods that rely on open-source algorithms and libraries so that they can be easily implemented by all plant sciences stakeholders. The programming language was Python 3.7. Four popular methods relying on features-based matching and global

transformations were tested from the famous OpenCV library (Bradski et al., 2008) (version 4.1.0.25). Those methods were SIFT (Lowe, 2004), SURF (Bay et al., 2008), ORB (Rublee et al., 2011) and A-KAZE (Alcantarilla et al., 2013). Default parameters were used for feature detection. Then, the matches were sorted by score and only the best matches were kept to compute the transformation. That proportion of valid matches was considered as a sensitive parameter and a sensitivity study was led to identify the best value for each method and each camera. In addition to those features-based methods, three area-based methods were tested. The first method, referred as DFT, exploited a discrete Fourier transform to compute a correlation metric in the frequency domain (Reddy et al., 1996). It was implemented using the imreg_dft Python library (version 2.0.0). The second method, named ECC, relied on a similarity metric built using an enhanced correlation coefficient (Evangelidis et al., 2008). It was implemented using the Python OpenCV library (version 4.1.0.25). The third area-based method, called B-SPLINE, used a Normalised Mutual Information (NMI) metric and differentiated itself from all the others by performing a local transformation of the slave image. That method was implemented using the Elastix library, initially developed for medical applications (Klein et al., 2010). For Python, the library wrapper was pyelastix (version 1.2). This allowed a local transformation based on a 3rd-order (cubic) B-spline model (Rueckert et al., 1999). In addition, the NMI metric is recognized to be particularly suitable for multimodal images registration (Studholme et al., 1999; Zitová et al., 2003; Klein et al., 2010; Sotiras et al., 2013; Keszei et al., 2017). However, the main drawback of area-based methods is that they may necessitate an initial alignment if the slave images underwent transformations such as huge rotation or scaling. For this reason, the calibration-based DDTM method was exploited to provide roughly registered images before applying the DFT, ECC and B-SPLINE methods. In the end, aligned slave images were cropped to save images of 855×594 pixels that were limited to the commonly aligned zone. Considering the cameras at 1.6 m height, that zone represented an area of 0.38 m².

Table 13. Summary of tested registration methods. For all the methods, the resampling was bilinear.

Method	Matching	Transformation	Library	Origin
SIFT	Features	Homography	OpenCV	(Lowe, 2004)
SURF	Features	Homography	OpenCV	(Bay et al., 2008)
ORB	Features	Homography	OpenCV	(Rublee et al., 2011)
A-KAZE	Features	Homography	OpenCV	(Alcantarilla et al., 2013)
DDTM	/	Homography	/	(Berenstein et al., 2015)
DFT	Area	Similarity	Imreg_dft	(Reddy et al., 1996)
ECC	Area	Homography	OpenCV	(Evangelidis et al., 2008)
B-SPLINE	Area	B-spline	Elastix	(Rueckert et al., 1999)

4.5 Validation of registration methods

The evaluation of registration performances is a difficult task and each method has its drawbacks. For this reason, three different indicators were employed:

• The percentage of plausible alignments. This indicator assessed the number of images that seemed visually aligned. It was computed by a human operator examining one by one the registered images in a viewer, beside their master image. Bad automatic registrations were characterised by aberrant global transformations that were easy to identify (**Figure 28**). For the local transformation, alignments were considered aberrant in case of the apparition of deformed black borders in the frame or illogical warping of objects such as leaves curving in complete spirals. That indicator was computed for all the acquired images, i.e. a total of 3968 images for each camera.

• The average distance between control points in aligned slave and master images (control point error) (Wang et al., 2010). The control points were visually selected on the leaves and ears by a human operator. The points had to be selected on recognisable pixels. Attention was paid to select them in all image regions, at all canopy floors and at different positions on the leaves – edges, centre, tips. It was supposed that registration performances may differ depending on the scene content: only leaves or leaves + ears. Thus, two validation image sets were created. The vegetative set consisted of twelve images from trials 20-F and 20-FP acquired at the six dates before ear emergence. The ears set consisted of twelve images from trials 20-F and 20-FP acquired at the twelve dates after ear emergence. Ten control points were selected for

each image. Firstly, this indicator was only computed for the 900-nm images, as their intensity content was close enough to the 800-nm master image to allow human selection of control points. Additionally, the other types of images would not have allowed to quantify errors for all the registration methods because some of those methods generated aberrant alignments. Secondly, the control point error indicator was also computed for the RGB images, but only for the ECC and B-SPLINE methods. Those methods were chosen because they were the two best methods for the 900-nm images and because they provided plausible alignments for all of the RGB images of the two validation sets.

• The overlaps between the plant masks in registered slaves and master images. Contrarily to the two other indicators, this one could be automatically computed. However, it necessitated to isolate plants from background in the slaves and master image. A comparable segmentation could only be obtained for the 900-nm slave and the 800-nm master. The segmentation algorithm relied on a threshold at the first local minimum in the intensity range 20 – 60 of the image histogram. Then, plants masks were compared to compute the percentage of plant pixels in the aligned slave image that were not plant pixels in the master image (plant mask error). That plant mask error indicator was computed for all acquired images. For the presentation of the results, averaged scores are presented for the two sets of images acquired before and after ear emergence.
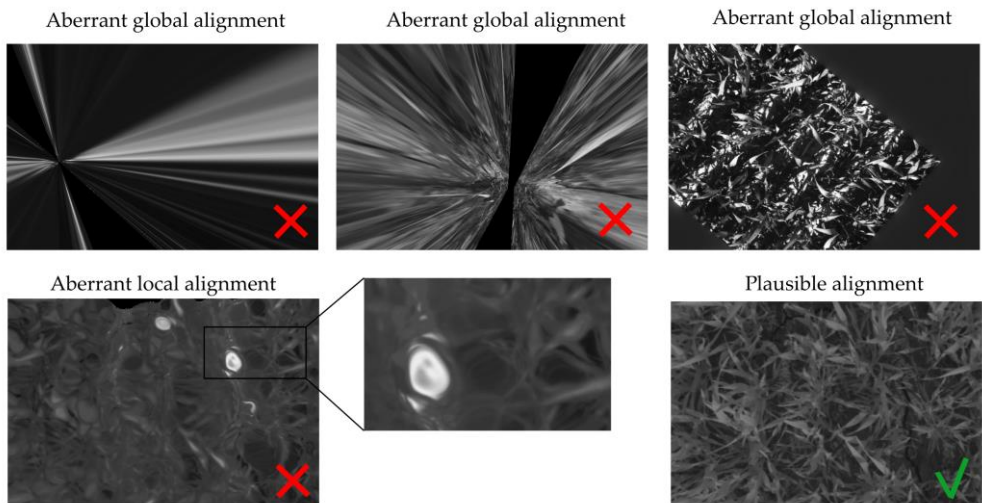
Figure 28. Examples of plausible and aberrant alignments of images from various cameras.

5. Results

5.1 Plausible alignments

The results of plausible alignments percentages for all slave cameras are presented in **Figure 29**. The DDTM method is also included although by its nature this method always yields a plausible alignment. For the other methods, the score depended on the camera type. The 900-nm images, whose intensity content was close to the master images, were well aligned by all methods. On the contrary, the thermal images were difficult to align and most image-based methods yielded aberrant alignments. Concerning the comparison of registration methods, the four features-based approaches – SIFT, SURF, ORB and AKAZE – failed to align all the images. The DFT method reached higher scores but similarly appeared as non-reliable to align 100 % of the images. Only the ECC and B-SPLINE methods succeeded in aligning almost all the images for all the cameras, except for the thermal camera. The few failures of the B-SPLINE were less problematic than the failures of other methods. In those cases, the images were still properly aligned and only some elements underwent local aberrant warps. For the thermal cameras, the ECC method reached 100 % of aligned images at most of the dates. On the contrary, the B-SPLINE was not reliable for thermal images.
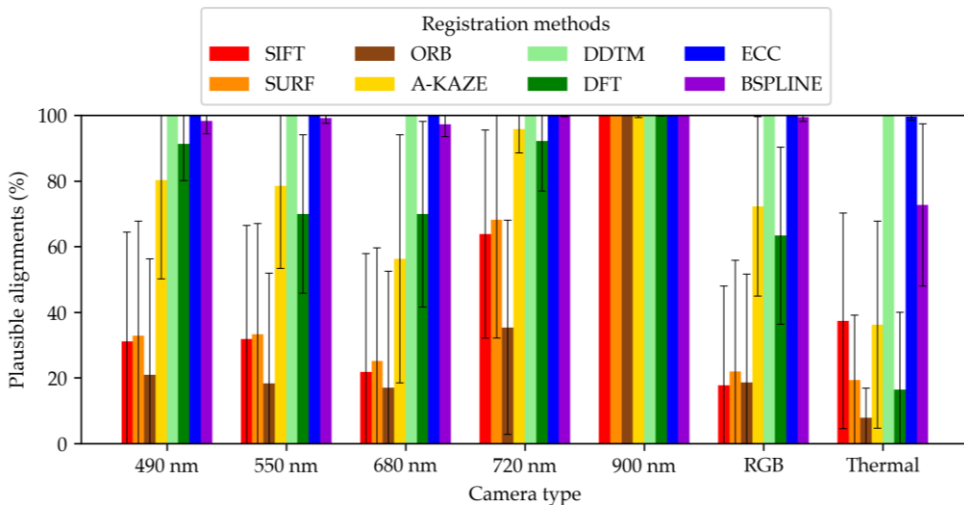
Figure 29. Mean and standard deviation of plausible alignment percentages compared for each camera and registration method. Those results were computed using all the images of trial 20-F and 20-FP mentioned in Section 4.2, i.e 3776 images per camera.

5.2 Registration accuracy and computation time

Table 14 presents average computation time, control point error and plant mask error for the different registration methods. The computation time was the average time to register one image of all the cameras using a 3.2 GHz Intel I7-8700 processor. The average was computed for the six dates before ear emergence. The computation times for the DFT, ECC and B-SPLINE methods included the pre-registration performed by the DDTM. As justified in Section 4.5, control point error and plant mask error were only computed for 900-nm and RGB images. Errors were computed independently for the dates before and after ear emergence. For both indicators, the smallest errors were obtained for the B-SPLINE method. However, the computation time was much higher than the other methods and would make it more difficult to use for real-time applications.

Table 14. Comparison of the registration methods. The comparison is based on three criteria: the average computation time to register one image of all the cameras, the control point error and the plant mask error. Errors are averaged for dates before and after ear emergence. For the RGB images, some methods were discarded (NA values) because they did not yield a plausible alignment for all the test images.

Method	Average time (s)	Control point error (mm)				Plant mask error (%)	
		900 nm	RGB	900 nm	RGB	900 nm	900 nm
SIFT	4.0	3.7	NA	3.4	NA	9.7	9.7
SURF	6.2	3.6	NA	3.4	NA	9.5	9.5
ORB	1.0	5.5	NA	3.6	NA	10.6	10.3
A-KAZE	2.7	3.4	NA	3.7	NA	10.1	9.7
DDTM	2.6	5.2	NA	4.2	NA	11.7	10.5
DFT	41.3	3.9	NA	4.1	NA	9.7	9.8
ECC	21.9	3.2	3.0	3.0	3.0	9.8	9.7
B-SPLINE	176.7	1.9	2.0	2.0	1.6	7.0	6.5

5.3 Parametrisation of the B-SPLINE method

The B-SPLINE method had proven to be the most accurate method. That test was performed using the default parameters. At the light of those first results, a particular attention was paid to tune the parameters of the method in order to further increase its performance. An important parameter is the final grid spacing, which defines the spacing between the grid points. The term final is used because the registration starts with a coarse points grid to warp large structures and then refine it in several steps until reaching the final grid spacing (Klein et al., 2010). A fine grid offers the possibility to account for fine-scale deformations but may also cause more aberrant warps. **Figure 30** illustrates the difference of deformation fields for final grid spaces of 16 and 2 pixels.
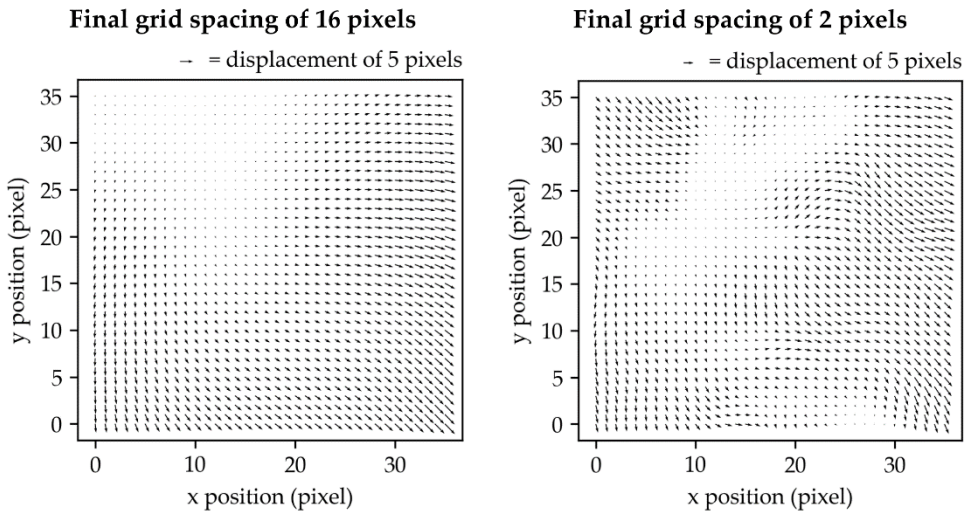
Figure 30. Example of slave image deformation fields for the B-SPLINE method with final grid spacing of 16 or 2 pixels. The deformation fields were extracted for an image region of 36×36 pixels. The lengths of the arrows are proportional to the pixel displacement.

Default parameters used for this study implied an initial grid of 128 pixels refined to 64 and 32 pixels to reach a final grid spacing of 16 pixels. Additional trials were carried out for the vegetative validation images set by reducing the final grid spacing to 8, 4, 2 and 1 pixels. Trials on the 900-nm images showed that it was important to gradually warp the images. To reach a final grid spacing of 1 pixel, the steps were grids of 128, 64, 32, 16, 8, 4 and 2 pixels. Burning steps by directly reducing the grid

from 32 to 1 pixel, for example, caused aberrant deformations. Concerning the other image types – RGB, thermal, 490, 550, 680 and 720 nm –, all grid refining levels led to aberrant deformations. Grid refining only worked on the 900-nm images thanks to their intensity content close to the 800-nm master images. **Figure 31** details the effect of grid refining for the 900-nm images on control point error and computation time. The smallest average error was obtained for a final grid spacing of 2 pixels. However, the computation time was eight times higher than for a grid spacing of 16 pixels. Results also demonstrated that refining the final grid spaces to 1 pixel had no interest. Not only was the error higher than for a 2-pixel spacing but also the computation was extremely slow. It is also to note that for the 1-pixel spacing, some aberrant deformations were visually noticed.
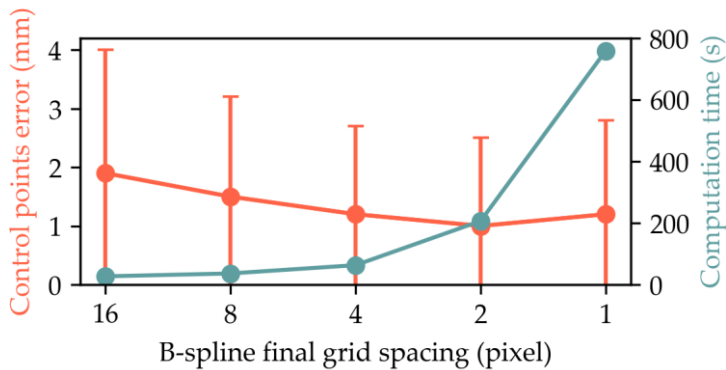
Figure 31. Effect of B-SPLINE final grid spacing on control point error and average computation time. Those indicators were computed for twelve 900-nm images of wheat canopy acquired at six different dates before ear emergence.

5.4 Plant mask erosion

Image fusion consists in exploiting a plant mask to extract and combine information from the wheat organs in the different images. However, even with the best registration method, close-range image registration inevitably leads to errors that are an issue for image fusion. This is especially problematic at leaf edges. A slight shift of a leaf edge between one of the aligned images and the common plant mask may lead in some background to being considered as leaf. To overcome that issue, a solution is to erode the common plant mask so that the remaining plant mask pixels comprise scene plant zones in all the aligned images. An example is provided in **Figure 32** where plant mask of the 900-nm slave image is considered as reference for fusion. By eroding this mask, it is possible to reduce it to pixels that represent plants

in both master and aligned slave images. In other words, this had the effect to remove pixels of the slave plant mask that do not represent plants in the master image.
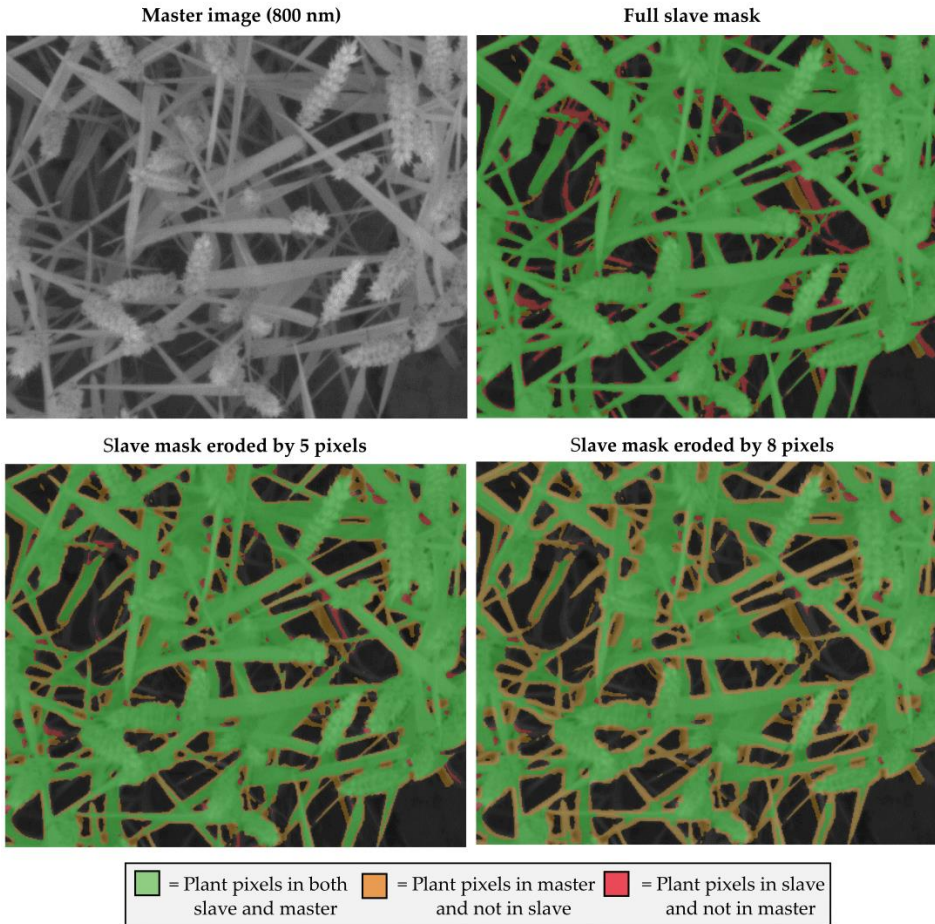
Figure 32. Example of plant mask erosion. Plant mask of the aligned slave image (900 nm) is considered as the reference for image fusion. In that case, eroding the mask avoids selecting pixels from the master image that do not represent plants (the red pixels in this example).

Erosion of 900-nm slave plant mask was tested for erosion values of 0 to 12 pixels. The impact on the remaining plant area and the plant mask error is presented in **Figure 33** for four registration methods of interest. It shows that the error tended to reach an asymptote while the exploitable plant area continued to decrease. The asymptote of plant mask error was close to zero. For an erosion of 12 pixels, error values were 1.2, 0.8, 0.7 and 0.5 %, respectively for the DDTM, ECC, B-SPLINE coarse final grid (16

pixels) and B-SPLINE fine final grid (2 pixels) methods. The remaining errors may have been artificial errors due to a difference of plant segmentation between the 800 and 900-nm images. Those curves imply that erosion of the common plant mask should be adapted to the quality of the registration method. Theoretically, for a perfect registration, no erosion would be necessary. On the opposite, for huge erosion values, the quality of the registration would have less importance. This is well illustrated in **Figure 33**. As the erosion value increased, the difference between the registration methods decreased. For each method, it is possible to assess the added value of a greater erosion by looking at the slope of the error curve (red curve) in **Figure 33**. For DDTM, an erosion of more than 8 pixels would be advised. For ECC, an erosion between 6 and 8 pixels would be a good compromise. For B-SPLINE, erosion could be limited to 6 for a coarse final grid or to 3 pixels for a fine final grid. Those values are however only valid for our imaging set-up. They should be adapted if the scale of the image changes. The erosion could also be adapted in case of very thin leaves, for example at an early development stage, to avoid eroding too much of the leaf surface.
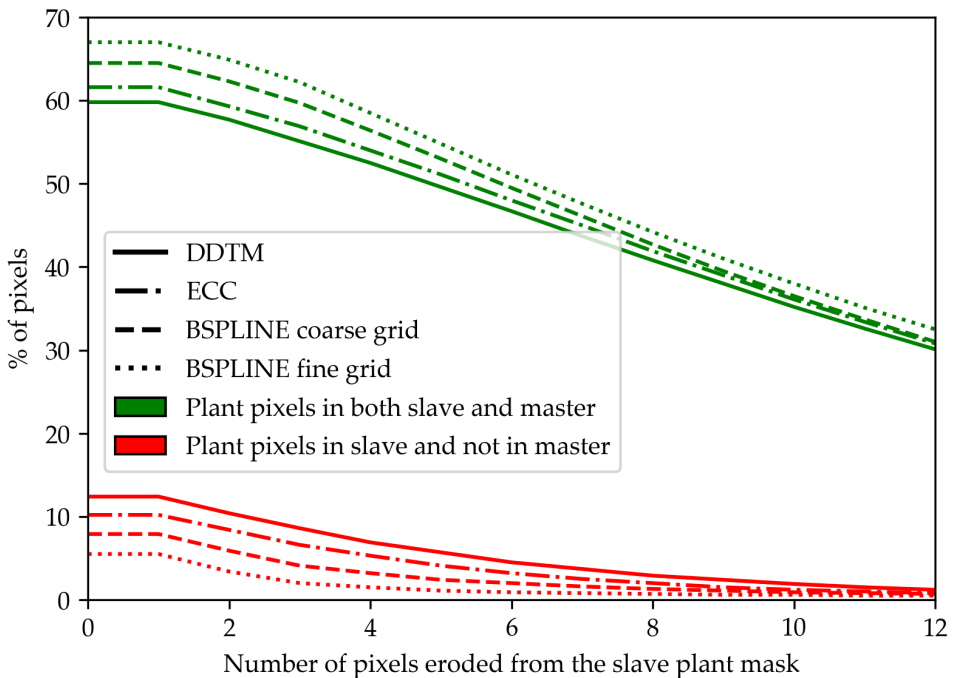
Figure 33. Impact of plant mask erosion for 900-nm images. The considered images were the twelve of the vegetative validation set aligned using the DDTM, ECC, B-SPLINE coarse grid (final grid spacing of 16 pixels) or B-SPLINE fine grid (final grid spacing of 2 pixels).

5.5 Suggested registration – fusion strategies

The results of this experiment have shown that the same registration method could not be used for all cameras and in all circumstances. At first glance, the B-SPLINE method seemed an obvious choice because of its accuracy. Even if the local transformation failed in some image zones for alignments judged as aberrant, the rest of the regions of those images were still properly registered. However, the choice of this method no longer held for thermal images or if the computation time was crucial. For real-time applications, only the DDTM method could satisfy the need of an almost instantaneous registration. Indeed, the only time-consuming steps of DDTM are related to stereo vision to automatically get the distance of the objects of interest. If this distance is provided, as for the image acquired after ear emergence, the registration is performed nearly instantaneously. For a compromise between an acceptable computation time and a small error, the ECC method would be the best choice. This method would be also recommended for the thermal images as a substitute of B-SPLINE. In addition, the results of plant mask erosion have shown that registration and fusion should not be considered as independent steps. The quality of the registration conditioned the processes necessary before fusion: plant mask erosion and pixel intensities averaging. A rougher registration such as the DDTM method necessitated more corrections before image fusion.

Four registration – fusion strategies are proposed in **Table 15**. In addition to registration and plant mask erosion, it is suggested that plant pixel intensities should be averaged after registration to counterbalance the possible local intensity shifts. It is also suggested that plant traits extracted from fused images should preferably rely on median intensity values (rather than average values) to prevent the scenario in which a few background values would still have slipped into the plant mask.

The choice of the strategy should be envisioned in relation to the final application and the nature of the available cameras. If all plant organs need to be measured, the ACCURATE method or the HIGHLY ACCURATE method should be employed to get rid of the need for plant mask erosion. Likewise, if the application implies the measurement of tiny details such as fungal spores on leaves, those methods should be preferred to limit small local shifts of particular intensities representing those details. The choice of the strategy should also be considered taking into account the whole set of cameras. It may not be a good idea to fuse a thermal image registered with DDTM and a NIR image registered with a B-SPLINE fine grid, especially if a common plant mask is used to extract plant features. Finally, it is necessary to clarify now that those strategies are results based on the data from this study. They are suggestions that

should be validated in future works and do not claim to cover all the diversity of registration and fusion approaches that could be applied to close-range wheat images.

Table 15. Proposed registration – fusion strategies.

Strategy name	REAL-TIME	FAST	ACCURATE	HIGHLY ACCURATE
Registration	DDTM	DDTM + ECC	DDTM + B-SPLINE (coarse grid)	DDTM + B-SPLINE (fine grid)
Plant mask erosion	Wide	Medium	Medium	Tiny or none
Intensity averaging	Wide window	Medium window	Small window	None
Computation time	Instantaneous	Moderated	Slow	Extremely slow
Scope	Suitable for multi-modal images	Suitable for multi-modal images	Not suitable for thermal images (if the master is NIR)	Limited to mono-modal images

6. Discussion

6.1 Considerations on the matching step

The main issue of features-based matching algorithms in a multimodal framework was their incapacity to yield plausible alignments for all the images. When they worked, they were however able to provide images as well registered as by global area-based methods. Observations realised in this study suggested that the reliability of features-based methods could be impacted by environmental factors such as wind (the acquisitions were not perfectly synchronous) and cloudiness. The clearest trend was however the impact of the nature of the observed scene. All the image modalities showed increased matching performances when the scene contained wheat ears. The hypothesis is that the structure of wheat ears was more suitable than that of leaves for point feature detection. Another element to explain bad feature detection performances is the phenomenon of gradient inversion observed by Rabatel et al.

(2016) for visible and NIR images. An avenue for improving features-based methods would be to perform feature detection after some background removal pre-processing (Henke et al., 2019b). Another hint related to features would be to filter images using an edges detector prior registration (Yang et al., 2009, 2012). This approach would combine the use of robust features (leaf edges) and area-based matching. The detection of similar wheat leaves boundaries in all images seems however a challenging task. Considering the area-based matching metrics such as NMI or ECC, they showed robust performances for multimodal plant images registration as already highlighted in the literature (Studholme et al., 1999; Zitová et al., 2003; Evangelidis et al., 2008; Sotiras et al., 2013; Keszei et al., 2017).

6.2 Nature of distortion and choice of the transformation model

The choice of the transformation model depends on the type of distortion between the two images. It is important to notice that the notion of distortion between images differs from the commonly used "image distortion" term that usually refers to optical distortion of images from a single camera. The possible distortions between images are:

- Differences of optical distortion between the images. The two types of optical distortion are radial and tangential distortions. Radial distortion is due to the spherical shape of the lenses. Tangential distortion is due to misalignment between lens and image plane. If the images are acquired by two different cameras with different optical distortions, it causes a distortion between the images.

- Differences of perspective. Those differences appear if the cameras that acquired the two images are at different distances. For the same distance of the cameras, images present the same perspective, whatever the lens. However, two cameras with different fields of view (determined by focal length and sensor size) necessitate being at different distances to capture the same scene. For this reason, differences of fields of view are intuitively perceived as responsible for differences in perspective distortion.

- Differences of point of view. The cameras that acquire the slave and the master image are not at the same position. It results in two different effects. Firstly, due to the relief of the scene, some elements may be observed in an image and not in the other one. This is called the occlusion effect. Secondly, the relative position of the objects becomes distance-dependent (that property is especially exploited for stereo vision). This is referred to as the

parallax effect. It is greater when the distance between the cameras increases compared to the distance between the cameras and the objects of interest.

- Scene motion. If the acquisition of the images is not perfectly synchronous, a relative displacement of scene objects with respect to the sensors causes distortion between the images. Objects such as wheat leaves are liable to be moved by the wind.
- Differences in scene illumination.
- Differential impact of heat waves (some images may be blurry).

For the close-range wheat images acquired in this study, differences of optical distortion were very limited thanks to the calibration of all the cameras to remove the optical distortions. As the cameras were located almost at the same distance of the scene, the perspective effect was also negligible. The main source of distortion between the images was attributed to the difference of point of view. Additionally, the acquisitions were not perfectly synchronous and some wind-induced movement may have impacted a few leaves. Theoretically, only a local transformation could handle such multiple and complex distortions. Even considering the difference of point of view as the only source of distortion, the parallax effect implied that a global transformation could only register without errors the objects lying on a same plane (perpendicular to the camera optical axes). Nevertheless, investigating global methods was compulsory because: i) those methods are the simplest and the most common, ii) a global transformation is a preliminary step before any local refinement and iii) it was chimerical to imagine a close-range registration without any error on a scene as challenging as a wheat canopy. The complexity of local methods could have been a disadvantage, leading to higher errors than those obtained with simpler approaches. Concerning the choice of the global transformation, the homography was preferred to be as general as possible. In this study, there were no significant perspective differences and it is stated that affine transformation models could have been employed. The proof is that the elements h_{20} and h_{21} (**Eq.10**) of all the homographic transformation matrices obtained after calibration were close to 0 (**Figure 26**). This simplification of the homography was also observed by Berenstein et al. (2015).

6.3 Critical look on the validation methods

Validation and error quantification of registration methods are always a difficult topic because ground truth maps of pixels are not available. This is especially challenging in case of multimodal images because pixel intensities cannot be compared. Different approaches encountered in the literature are:

- To visually assess the success of registration: aligned slave and master images look similar (Wang et al., 2010).
- To verify that the values of the transformation parameters fall in the range of plausible values (Henke et al., 2019b). This method can be assimilated to the previous one but presents the advantage to be automatic.
- To test the algorithm on a target of known pattern (Jerbi et al., 2015).
- To manually select control points and assess the distances between their positions in aligned slave and master images (Wang et al., 2010).
- To segment objects in the scene and study the overlaps between those objects in aligned slave and master images (Raza et al., 2015; Henke et al., 2019a).
- To use a similarity metric as a proxy of registration quality (Raza et al., 2015; Rabatel et al., 2016).

Among those methods, the ground truth target was discarded because it does not help to estimate real errors occurring on plant canopy images registration. In a certain way, this evaluation was however performed during the calibration of the DDTM method. The similarity metric was also discarded for the reasons detailed by Rohlfing (2012). They especially stated that the validation should be as independent as possible from the registration itself. The other methods listed here above were used for our study. The choice to rely on three very different methods, including two human validations, was judged as a strength of this study compared to existing research in the plant domain for which only one approach was usually chosen. This is especially true because each approach has advantages and weaknesses.

The number of visually plausible alignments (number of successes) is a first way to reject unreliable methods and it can be applied to all types of images. In this study, it was however a laborious task because performed on a huge number of images. Moreover, it does not help to quantify the errors. This method should be used as a first test but followed by quantitative methods.

The control point error (manually selected control points) presents the advantage of being totally independent from the registration process. It provides an error in pixel

or physical distance units. The method is however very time-consuming. It is limited to slave images types where it is possible to visually, precisely and without any doubts select the same control points than in the master images. It is also subject to human bias. Operators could select non representative control points sets. They are also imprecise in the selection of control points. The imprecision, i.e. the fact that the operator does not always select the same pixel for a same scene element, was quantified in this study by repeating three times the marking of same control points for three different master images (10 points per images). The average error between two repetitions was 0.6 pixels, which corresponded to 0.5 mm in the scene. Additionally, it was sometimes difficult to identify to the pixel level the same scene elements in slave and master images because of the intensity resampling effect.

The third indicator, the plant mask error, presents the advantage of being fully automated. At the opposite of the control point error, it accounts for most of the image pixels. It is also goal-oriented, in the sense that a common plant mask is the key element to extract plant traits by image fusion. This overlap-based validation is criticised by Rohlfing (2012) in the frame of medical tissue imaging. In the plant canopy context, the situation may be different. In this study, the plant mask indicator was relatively coherent with the other indicators. It was especially useful to study the impact of plant mask erosion to mitigate registration errors before images fusion. A concern is raised about the quality of the plant segmentation based on histograms thresholds. That simplistic segmentation approach may have led some barely visible leaves to be included in the aligned slave mask and not in the master mask, for example. To build our plant mask error, it was arbitrary decided that the error would be the percentage of pixels considered as plant in the aligned slave image and not in the master image. Those pixels would be problematic for image fusion in case the plant mask is provided by the slave image. Another option would have been to focus on plants pixels in the master images that were not plants in the slave image. Those pixels would be problematic for image fusion in case the plant mask is provided by the master image. In practice, a perfect plant mask for fusion would necessitate to combine information from both slave and master images. Thus, no approach makes more sense than the other. Neither of the two could perfectly estimate the error that would occur in the final fusion pipeline where the plant mask would be built by a combination of already aligned images.

6.4 Visualisation of successful image registrations

Visually demonstrating the good quality of a registration method is a non-trivial task. In the plant literature, some papers present the aligned images side by side (Möller et al., 2007; Rabatel et al., 2016). Those figures show the success of the registration but do not highlight the small misalignments of the plant organs. Others rely on the superposition of aligned images at a certain level of transparency (Wang et al., 2010; Yang et al., 2012; Henke et al., 2019a), which can yield pretty readable figures or confuse representations depending on the imaged scene and the figure realisation. Another option is to exploit a colour code to show the plant mask overlaps (Raza et al., 2015). Jerbi et al. (2015) exploited a chess-like mosaic made of squares from both aligned slave and master images. In **Figure 34**, we propose an alternative visualisation method that allows the operator to clearly compare the alignments of plant organs. To take the best from that cross method, the size of the observed images regions should be chosen so that the scene details are big enough on the figure.

7. Conclusions

The calibration-based registration method (DDTM) provided a solution to roughly align multimodal images, especially as a first step before a more advanced registration. Among seven image-based registration methods, the one relying on an area-based matching (mutual information) and a local transformation model (B-SPLINE) aligned the visible and NIR images with the highest accuracy. The local transformation model was not reliable for thermal images for which a global transformation was necessary. To mitigate registration errors, it was suggested to erode the plant mask used for the fusion of aligned images. Consequently, different registration – fusion strategies were identified. On the one hand, slow local transformations techniques provide accurate registrations and allow to keep the majority of the plant surface. On the other hand, for real time applications it is possible to combine fast but less accurate registration techniques with a high level of plant mask erosion to counterbalance their errors.
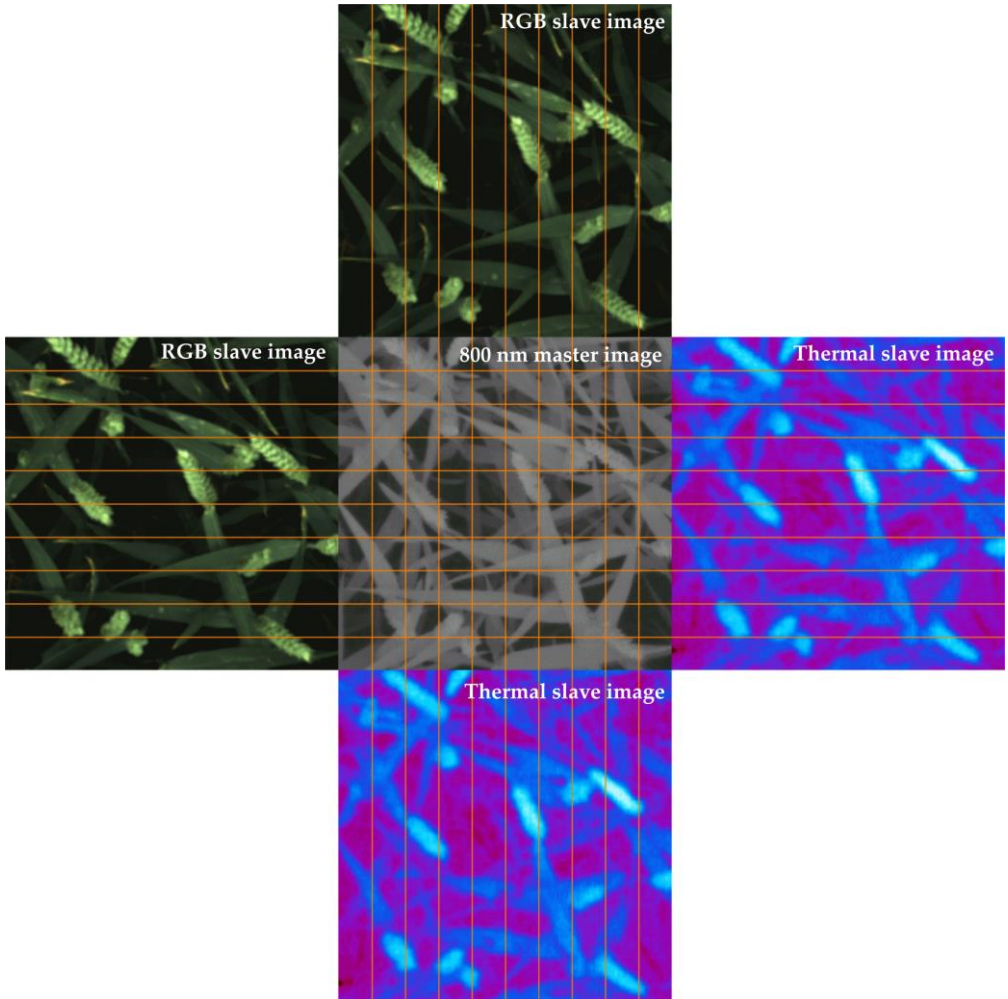
Figure 34. Visualisation of image registration quality for RGB and thermal slave images aligned using the ECC method. Orange lines are dedicated to ease image comparison. Colour and contrast of the master image have been adjusted for this figure to increase plant organ visibility.

VII

Bi-directional reflectance factor

1. Synopsis

This chapter is largely based on the publication of Dandrifosse et al. (2022):

The spectrum of the light reflected from a wheat canopy to a multispectral camera array is influenced not only by the canopy properties but also by the spectrum of the incident sunlight. This chapter proposes a high-throughput method to compute, for each band of the multispectral device, a standard measure to account for the incident sunlight: the bi-directional reflectance factor.

This is where all the pieces of the puzzle of the multi-sensor fusion come together: the mask with the organ separation (Chapter V), registered with all the multispectral bi-directional reflectance factor maps (Chapter VI), can be used to extract that bi-directional reflectance factor of the wheat organs separately in each spectral channel. Thanks to that, all the plant traits related to the spectrum of the reflected light could be computed at the scale of the wheat organs, and not only at the scale of the whole canopy as for previous studies in the field.

2. State of the art

The term « reflectance » refers to the proportion of light that is reflected by an object. It generally varies with the wavelength of the incident light. It is an absolute quantity that is comparable between different irradiance levels and sensors. In field phenotyping, spectral reflectance is used to derive the architecture or physiology of the crops. The sensors can be carried by UAVs, ground vehicles and platforms or human operators (Reynolds et al., 2019). Satellites and aircrafts are generally used in precision agriculture, where it is not necessary to distinguish experimental micro-plots, but they could be useful in large-scale phenotyping approaches (Smith et al., 2021). Extracting information at the plant organ scale requires close-range vectors such as ground-based approaches or low altitude UAVs. Nevertheless, in practice, reflectance at the organ scale has almost exclusively been studied in controlled conditions, where the reflectance of single isolated organs could be directly related to their anatomy and physiology. It has for example been exploited to detect nitrogen content (Vigneau et al., 2011) and diseases (Yuan et al., 2012; Ashourloo et al., 2014;

Mahlein, 2016). In the field, reflectance measured at the canopy scale depends not only on organ anatomy and physiology but also on sun-sensor view geometry, canopy architecture and soil background (Ranson et al., 1985; Shibayama et al., 1985; Camacho-de-Coca et al., 2001; Chakraborty et al., 2005, 2015; Barman et al., 2010; Comar, Baret, et al., 2012). To further complicate the problem, all those factors have a different impact depending on the development stage of the crop (Lunagaria et al., 2017). One approach to extract crop properties based on canopy reflectance is to exploit vegetation indices (Cammarano et al., 2014), i.e. mathematical operations combining reflectance from several wavebands. These indices are sometimes complex, and supposed to be sensitive to crop variables of interest while being relatively robust to the confounding factors. To extract useful agronomic information from reflectance measured at the canopy scale, another approach is to exploit models. Many types of models can be built (Goel, 1988), and one of the most widely-used is PROSAIL (Baret et al., 1992; Jacquemoud et al., 2009; Berger et al., 2018) made of two components : PROSPECT (Jacquemoud et al., 1990) and SAIL (Verhoef, 1984). This model integrates characteristics of the incident light as well as canopy architecture to link canopy reflectance to leaf biochemical content. Several versions of the two components of the model have been developed and many researchers have discussed them (Jacquemoud et al., 2009; Lu et al., 2021). Wheat canopies are composed of several types of organs: stems, leaves and ears. A model such as PROSAIL works on leaves but ears contain unwanted information. Li, Jiang et al. (2021) showed that canopies with or without ears presented different reflectance properties, and Danner et al. (2019) observed that the proportion of ears in the nadir-viewed canopy elements correlated with deviations between PROSAIL spectral outputs and field spectral measurements. These studies highlight a broader issue: even at close-range, the current in-field approaches provide reflectance values for entire crop canopies, without distinction between plants and soil and between plant organs. To understand the reason for this lack, it is necessary to examine the reflectance measurement methods that have been used for crop phenotyping.

To compute a reflectance value, one needs to possess information on not only the light reflected from the plants but also the incident light. Active sensors solve this issue by including a light emitter. Therefore, the incident light possesses known characteristics. The reflectance sensors of this category are non-imaging devices. There are many commercial models such as the GreenSeeker (Trimble Navigation Ltd., Sunnyvale, CA, USA) (Tremblay et al., 2009; Erdle et al., 2011; Kipp, Mistele, & Schmidhalter, 2014; Barmeier et al., 2016; Qiu et al., 2018; Souza et al., 2021), the

N-Sensor ALS (YARA international ASA, Oslo, Norway) (Tremblay et al., 2009; Kipp, Mistele, & Schmidhalter, 2014; Qiu et al., 2018), the CropCircle (Holland Scientific Inc., Lincoln, NE, USA) (Erdle et al., 2011; Kipp, Mistele, & Schmidhalter, 2014; Qiu et al., 2018; Souza et al., 2021) and the CM 1000 (SDEC, Reignac-sur-Indre, France). Those farmer-oriented sensors rely on reflectance measurement but only provide vegetation indices values. Moreover, their wavebands are limited by the light source. They can be used on a specific organ to measure its reflectance but they require a human or a machine to position them.

Passive reflectance sensors, on the other hand, necessitate a way to measure the incident sunlight. Whether these sensors are imagers or not, two main methods are used. The first method to account for the incident light relies on a reference panel of known reflectance spectrum and properties. It was used along with non-imaging sensors by Rodriguez et al. (2006), Anderegg et al. (2019) and Zheng et al. (2019). The panel can be imaged before and after each canopy spectrum acquisition, but this does not account for changing sunlight conditions between those reference measurements. If the sensor is an imager, it is also possible to include the panel in the observed canopy (Fu et al., 2020). This has the advantage of being able to collect the reference reflectance at the exact time of the canopy image capture. However, it also reduces the size of the studied plant zone and limits the throughput of the system. Most studies exploiting a camera to measure reflectance used such a panel of known reflectance (Moshou et al., 2004; Verger et al., 2014; Haghighattalab et al., 2016; Naito et al., 2017; Behmann et al., 2018; Jay et al., 2018; Fu et al., 2020; Li, Jiang, et al., 2021). The second method to account for the incident light deploys a sensor facing the sky in addition to the sensor facing the canopy. At close range, it was mainly used for spectrometers. The up-facing probe was exploited to compute reflectance (Barmeier et al., 2016; Bai et al., 2019; Pérez-Ruiz et al., 2020) or to compensate for variations in sunlight if the reflectance is derived from a reference panel acquired before or after the canopy spectra (Tavakoli et al., 2019). That principle of up and down-facing probes has been integrated in a diversity of commercial devices such as the HandySpec Field (tec5, Oberusel, Germany) used by Elsayed et al. (2018), Prey et al. (2018), and Tavakoli et al. (2019). Concerning imaging devices, several commercial solutions have coupled an incident light sensor (ILS) and a multispectral camera array: the ILS of the MAIA camera (SAL Engineering, Russi, Italy), the smart ILS of the MCA and ADC cameras (Tetracam Inc., Gainesville, FL, USA) (Raymond Hunt et al., 2019), the SEQUOIA sunshine sensor (Parrot SA, Paris, France) and the MicaSense DLS 2 (MicaSense Inc., Seattle, WA, USA). Reflectance computations

are possible with proprietary software or using custom scripts. The Micasense company uploaded detailed code (https://github.com/micasense/imageprocessing). A documented UAV approach was proposed by Hakala et al. (2018) for forest reflectance measurements.

It appears that most of the wheat reflectance measurements have been carried out by non-imaging devices such as spectrometers or by high altitude cameras with limited spatial resolution. The reflectance measurements performed with close-range imagers rely on bulky reference panels and the image processing step often does not include the separation of organs. A method to isolate the reflectance of shaded and sunlit wheat leaves was used by Bebronne et al. (2020), but their homemade reference panel, of unknown reflectance, only permitted relative reflectance measurements. In addition, the positioning of the panel in the images turned out to be complicated, especially because of the shadows caused by their acquisition platform. Sadeghi-Tehran et al. (2021) exploited a hyperspectral camera to extract the reflectance of wheat leaves and ears, shaded or sunlit, but the imaging of the reference panel added a constraint and reference reflectance measurement was only performed every 15 minutes, while natural lighting conditions can change much faster. For other crops, respectively tobacco and sugar beet, Fu et al. (2020) and Jay et al. (2017) managed to get the reflectance of leaves separately from the soil, using also a hyperspectral camera and a reference panel.

3. Goal and structure

The goal of this study was to propose a clear and simple method based on a multispectral camera array and an incident light sensor to compute the reflectance of wheat organs separately and without the need for a reference panel in each image. The reference panel is only used once to build the response curves of the cameras, and once at the beginning and once at the end of each field trip. During a field trip, the incident light sensor and the camera response curves allow for the computation of the reflectance. To validate the method, the evolution of the reflectance measured on a reference panel and on wheat organs was studied on the same wheat area throughout the day at six wheat growing stages. Finally, the approach was tested in two fertilisation trials all over the growing season to assess its consistency for two different varieties and the contrasted architectures generated by the fertilisation gradient.

4. Method

4.1 Data acquisition

4.1.1 All-day-long acquisitions

The purpose of those acquisitions was to study the evolution of wheat reflectance throughout the day on a defined wheat zone. The experiment was repeated at several dates during the 2021 season. At all the concerned dates, the manned platform II (Chapter III, Section 5.3) was positioned on the micro-plot 21-T (Chapter III Section 2). Images from the multiple cameras (Chapter III, Section 5.2) and incident light spectra were acquired at six important BBCH crop development stages and in a diversity of sky conditions (**Table 16**). The acquisitions were performed between 9 a.m. and 5 p.m. Every quarter of an hour, four acquisitions were performed at 10 s intervals: one where the reference panel was placed above the canopy and three of the canopy without the reference panel.

Table 16. All-day-long acquisition in trial 21-T

Date	Days after sowing	BBCH	Sky conditions	
			Morning	**Afternoon**
April 13th	157	29: Tillering	Sunny	Cloudy
May 28th	202	39: Flag leaf	Sunny	Sunny
June 10th	215	65: Flowering	Sunny	Sunny
June 23rd	228	69: End of flowering	Heavy clouds	Heavy clouds
July 1st	236	77: Late milk	Heavy clouds	Sunny
July 22nd	257	87: Hard dough	Sunny	Sunny

4.1.2 Acquisitions in fertilisation trials

Images and incident light spectra were acquired during the 2021 season in the trials 21-F and 21-FP. Contrary to the all-day-long acquisitions, the platform was moved over all the micro-plots. The aim was to test the reflectance measurement method in a diversity of canopy development scenarios. Two images of the reference panel were acquired at each field trip in a trial, one at the beginning and one at the end of the data acquisition.

4.2 Reflectance computation

4.2.1 Theoretical basis

The reflectance of a plant surface involves a directional aspect, depending on the azimuth and zenith angles of the light source and the camera relative to the surface. The bi-directional reflectance factor (BRF) was defined by Nicodemus et al. (1977) as "the ratio of the radiant flux actually reflected by a sample surface to that which would be reflected into the same reflected-beam geometry by an ideal (lossless) perfectly diffuse (lambertian) standard surface irradiated in exactly the same way as the sample". It is a function of the wavelengths of the light source and the direction of the camera and the light source (azimuth and zenith angles). Mathematically, the BRF can be written as (**Eq.11**):

$$\rho(\omega_i, \omega_r, \lambda) = \frac{L(\omega_r, \lambda)\,\pi}{E(\omega_i, \lambda)} \tag{Eq.11}$$

where L is the reflected radiance (W/(m^2 sr)), E is the incident irradiance (W/m^2), λ is the wavelength, ω_i is the direction of the light source, and ω_r is the direction of the camera. The BRF, as formulated here, is a physical quantity that cannot be measured, also known as a conceptual quantity. To be defined as "directional", the formula considers infinitesimal solid angles, which cannot include measurable amounts of light flux. The measurable reflectance quantities necessitate a conical or hemispherical geometry (Schaepman-Strub et al., 2006). In natural conditions, the incident sunlight is a combination of direct and diffuse irradiance, so it integrates the hemispherical aspect. From a strictly theoretical point of view, the quantity measured in this paper was actually an approximated BRF, and more accurately a hemispherical-conical reflectance factor. In practice, as for previous studies in the plant domain, the term BRF is used.

4.2.2 Camera Response Curves

The response curve of a camera is the relation between the digital numbers (DNs) in the images and the corresponding exposures (J/m^2) at the sensor. That relation depends on camera electronics. It may not be perfectly linear. Considering that one exposure value is associated with each pixel, the term exposure map is used. The exposure map at the sensor is (**Eq.12**):

$$H = E_{img}\, t \tag{Eq.12}$$

where t is the exposure time (s) and E_{img} is the image irradiance map (W/m^2) (**Eq.13**):

$$E_{img} = \frac{L \pi T}{4 N^2} \qquad\qquad (Eq.13)$$

where L is the radiance map from the scene (W/(m^2 sr)), T is the transmission factor of the objective and the optical filter (between 0 and 1), and N is the aperture (f-number). The transmission factor depends on the wavelength.

Eq.12 and **Eq.13** demonstrate a proportional relation between the exposure and the radiance maps, given a fixed configuration of the camera. Consequently, it is possible to build a response curve integrating not only the effects of camera electronics but also the effects of the optics. For this, we define the exposure map at the entrance of the objective (**Eq.14**):

$$H_{lens} = L\,t \qquad\qquad (Eq.14)$$

The modified response curve expresses the relation between the DN and that exposure at the entrance of the objective. To build that curve, it is necessary to obtain H_{lens} values for a range of DN values. For this, **Eq.11** and **Eq.14** are combined (**Eq.15**):

$$H_{lens} = \frac{\rho\,E\,t}{\pi} \qquad\qquad (Eq.15)$$

where ρ is the known reflectance of a reference panel, E is measured by the ILS, and t is recorded at image acquisition. The method to build the modified camera response curve is represented in **Figure 35**. For each camera of the multispectral array, fifty images of the panel were acquired under natural light. A different exposure time value was set for each image. That way, while exploiting the four reflectance targets, it was possible to generate a wide range of DN values. The position of the four targets was automatically extracted in each image thanks to an ArUco marker in the panel suitcase. A spectrum of the incident light was recorded at the time of each image acquisition. That spectrum was integrated over the wavelength bands corresponding to the camera filters, considering the transmission of an ideal filter. Similarly, the reference reflectance spectrum of the panel was integrated in the wavelength bands of the filters. Using those values and **Eq.15**, it was possible to compute the H_{lens}. Then, the H_{lens} values and the DNs extracted from the images were used to build the response curve. The DN-intercept was imposed at 0 for fitting the model. This choice

was made to avoid posterior issues when exploiting the curve in case DN values were smaller than a non-zero DN-intercept.
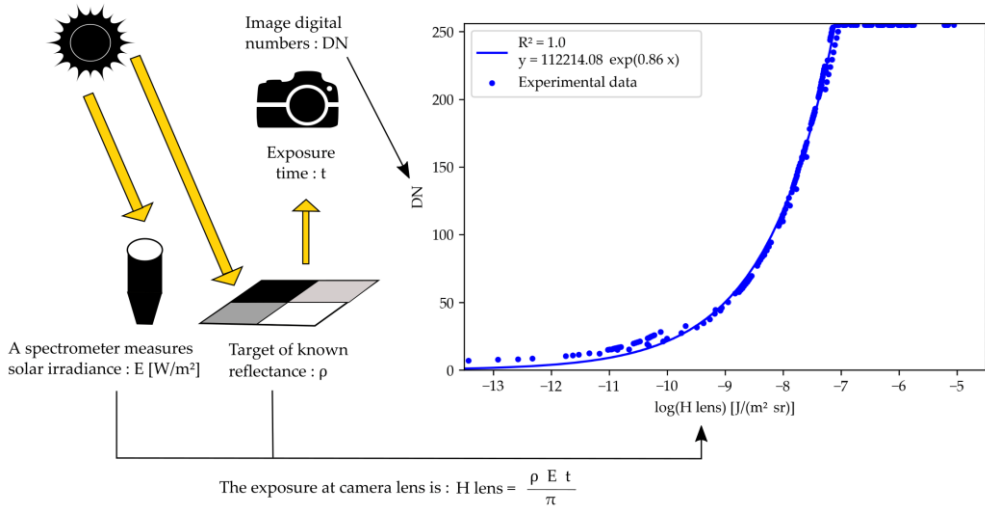
Figure 35. Method to build a camera response curve. Such a curve links the image digital numbers and the exposure at camera lens. The example is the response curve of the 490-nm camera.

4.2.3 Wheat organ bi-directional reflectance factor

For each channel, average DN values were computed for the segmented wheat organs. Those DNs were converted to exposure values through the camera response curves and the recorded exposure times. The sunlight irradiance spectra measured by the ILS were integrated over the wavelength bands of the cameras filters and combined with exposure values to compute the reflectance of the plant organs. The method is represented in **Figure 36**.
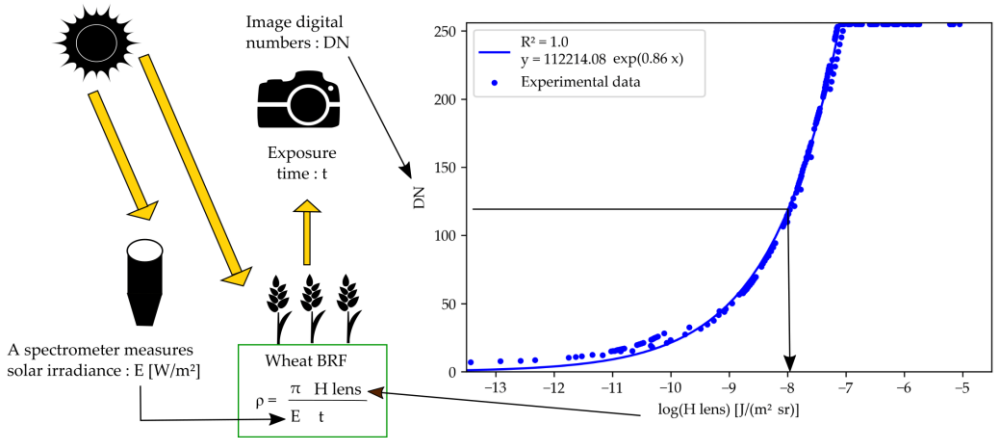
Figure 36. Method to compute the bi-directional reflectance factor (BRF). The method exploits the camera response curve, the ILS readings and the image digital numbers. H_{lens} is the exposure at camera lens.

4.3 Image segmentation at the organ scale

As a prerequisite for segmentation, the images from the multispectral camera array and the left RGB camera were registered using the B-SPLINE method described in Chapter VI. Three classes were considered for segmentation: the background (the soil), the leaves plus the few stems visible in the images, and the ears. The segmentation process was performed in one or two steps, depending on whether the image contained ears. The first step was the segmentation between the background and all the wheat organs. The separation used a threshold in the 800-nm channel, which was automatically determined for each image based on the first local minimum of the histogram. That simple method worked because of the significant reflectance difference of the wheat and the soil in the near-infrared. However, in some cases where direct sunlight reached the soil, a few soil pixels could be confused with the low and shaded leaves. To avoid this error, two additional thresholds were applied for the pixels of low values in the 800-nm channel: a threshold in the blue channel of the multispectral camera array and a threshold on the Excess Red index (Meyer et al., 1998) computed from the green and red channels of one RGB camera. The need to add these thresholds was determined thanks to a cloudiness index derived from the ILS data at the time of image acquisition. The index was formulated as (**Eq.16**):

$$C_T = 1 - \frac{E}{E_0 \cos(z)} \qquad\qquad (Eq.16)$$

where E is the solar irradiance (W/m²) in the spectral measurement range of the ILS, E_0 is the solar constant (1360 W/m²), and z is the sun zenith angle. An empirically found threshold of 0.90 was exploited to determine the conditions of strong direct sunlight.

The second segmentation step, when ears were in the imaged scene, was used to separate the ears and the rest of the image. That segmentation was performed using the deep learning approach detailed in Chapter V. An example of segmented BRF map is presented in **Figure 37**.
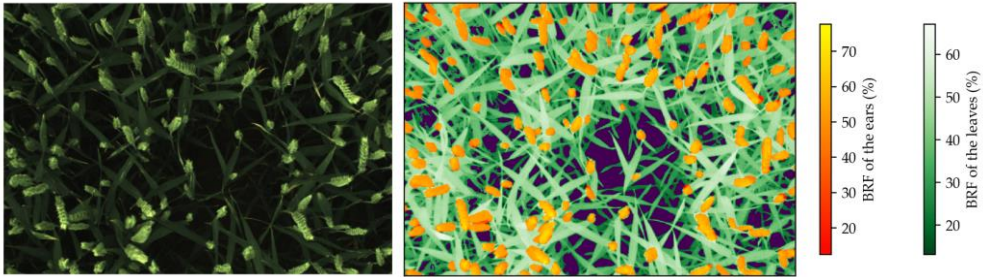
Figure 37. BRF map at the organ scale. To the left, a registered RGB image. To the right, the corresponding registered and segmented bi-directional reflectance factor (BRF) map in the 800 nm channel. The colour bars indicate the BRF values separately for ears and leaves. The dark purple regions in the image represent the background.

5. Results and discussion

5.1 Camera response curves

The response curves of all the monochrome cameras are presented in **Figure 38**. The determination coefficients demonstrate that the curves were well-adjusted on the measurement points for all the cameras. The differences between the curves can be explained by the relative sensitivity of the sensors to the wavelength and the different transmissivities of the optical filters.
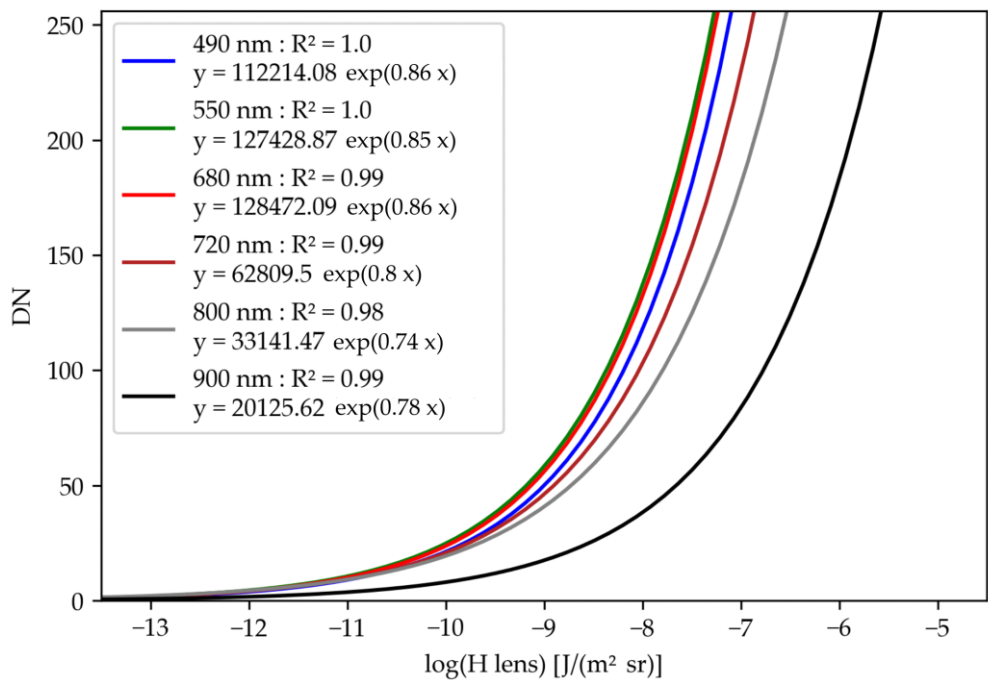
Figure 38. Response curves of the monochrome cameras. Each camera is designated using the central wavelength of the narrow-band optical filter. The response curves model the relation between image digital numbers (DNs) and exposure at camera lens (H_{lens}).

5.2 Bi-directional reflectance factor of the reference panel

Figure 39 presents the evolution of the BRF measured throughout the day on the dark grey reference target. The evolution was studied for the six filters and the six all-day-long acquisition dates.
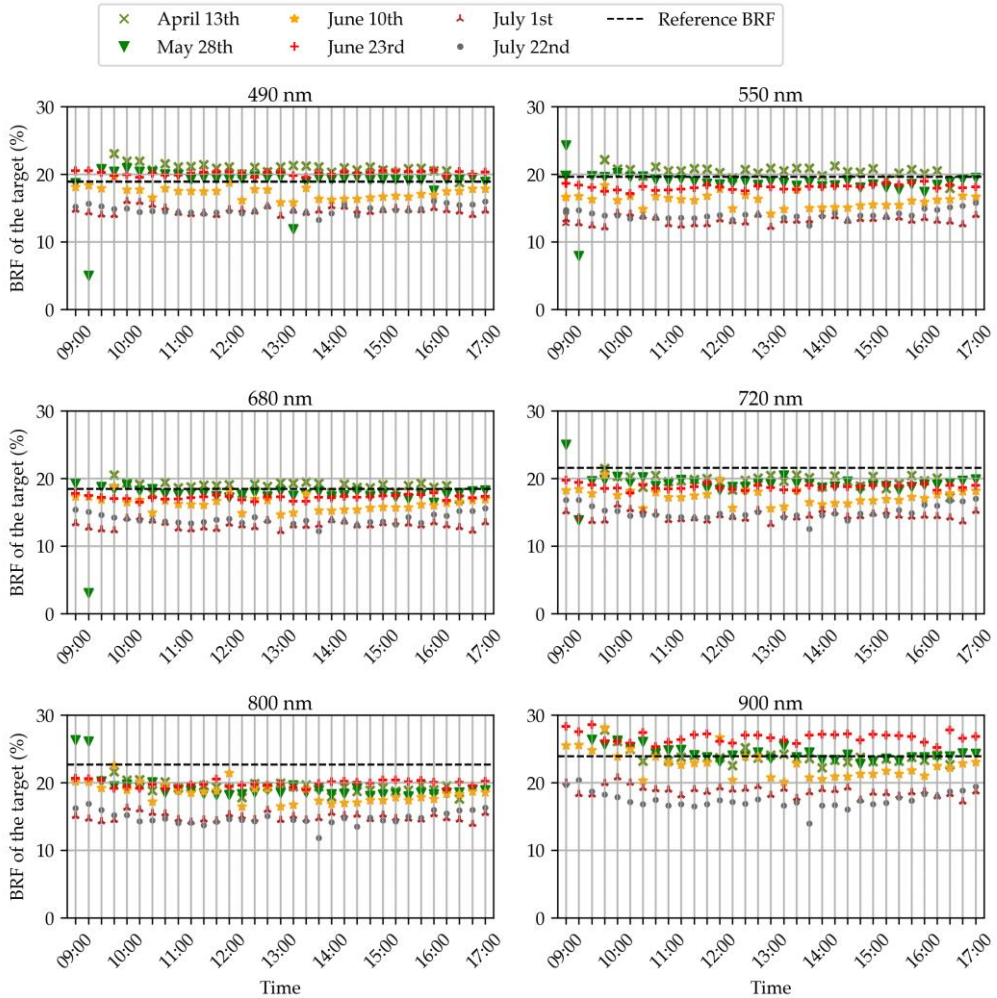
Figure 39. Average bi-directional reflectance factor (BRF) measured on the dark grey target throughout the day. Each subplot is dedicated to a camera of the multispectral array, designated by the central wavelength of its optical filter. The evolution of the measured BRF is represented for six different acquisition dates, indicated by a colour and symbol code on each subplot. The dotted line corresponds to the theoretical BRF of the target that was provided by the manufacturer.

A first important observation is that, in all the scenarios, the measured BRF was roughly constant throughout the day. This proves that the method based on the spectrometer was able to compensate for the relative changes in scene illumination despite major variations on the concerned days. The few outliers (May 28th at 9 and 9.15 a.m. in **Figure 39**) can be attributed to errors in the automatic detection of the reference panel. The reflectance was influenced by some pixels not belonging to a target of the panel but considered as such in the segmentation process.

A second important observation is that the measured BRF was not always equal to the reference value provided by the manufacturer. This highlights that the method cannot reliably retrieve accurate absolute BRF values. The differences from the reference are clearly related to the acquisition date, and sensor drift is unlikely because the importance of the variations is not related to a chronological order. The first obvious explanation could be the illumination conditions. In the literature, Guo et al. (2019) measured the reflectance spectra of reference targets and pointed out differences between three days of different weather conditions. For our study, the response curves of the cameras were established under very diffuse light conditions. This could explain an underestimation of the BRF on sunny days such as June 10th and July 22nd. Nevertheless, variations of sky conditions were also observed within the same date, e.g., on July 1st when the light was almost totally diffuse in the morning but the sky cleared up in the afternoon, and no major variations of BRF occurred within those dates. For this reason, our hypothesis is that the differences between dates were generated by other factors than the illumination conditions. We suspect that the acquisition system configuration has a role. Potential sources of variations were the orientation of the cosine corrector of the spectrometer; the connection between the optical fibre, the cosine corrector, and the spectrometer; and the orientation of the cameras.

Similar observations were made for the light grey and black targets, except that for the black target the differences were mainly due to slight overestimations. The white target could not be used because it appeared saturated in some images. This saturation was due to the auto-exposure algorithm of the Micro-MCA camera array. The exposures from the five slave cameras were computed according to the auto-exposure of the 800 nm camera, referred as the master. The relations between the master and the slaves, called relative exposures, were tuned to suit the light reflected from the wheat in the different wavelengths, not from the target.

5.3 Organ bi-directional reflectance factor throughout the day

Figure 40 presents the evolution of the BRF measured throughout the day on wheat leaves. The evolution was studied for the six filters and the six all-day-long acquisition dates. The goal was to study the robustness of leaf BRF measurements. It was assumed that no change of BRF would occur during the day because of plant physiology or canopy architecture variations. For most dates, the BRF gradually decreased in the morning and then was roughly constant in the afternoon. We hypothesise that the sun zenith angle had an influence. Indeed, the effect was the strongest on April 13th, and at this period of the year, the sun was really low on the horizon in the morning. On the contrary, no change of BRF was observed on June 23rd, which was the moment in the year where the sun was the highest in the morning and specifically very cloudy day. Similarly, July 1st in the morning was also cloudy, which may explain why the sun zenith angle had no effect. On the concerned dates, solar noon varied from 13:38 to 13:47 h. The best acquisition time to obtain comparable BRF measurements seems to be in the hours close to solar noon. Moreover, these acquisition hours minimise the impact of shadows. Rather than only focusing on hours, the sun zenith angle should be examined. Periods when the days are longer and the sun is higher would permit a wider range of acquisition hours. The data suggest that a zenith angle above 55° could secure comparable wheat BRF measurements for a camera in the nadir position. It also seems that the azimuth had no impact on the BRF of wheat leaves, which was expected because i) the crop was dense with barely visible rows and ii) the cameras captured the crop from nadir. For the hours when the sun was high enough, the BRF showed almost no variations, which proves the strong quality of the method and confirms the hypothesis that no variations due to physiology or canopy architecture occurred during these hours. For June 10th, June 23rd, July 1st, and July 22nd, the same study was carried out on the ears in addition to the leaves. The curves showed the same trends except on June 10th, when the ear BRF curve was bowl-shaped, with a minimum around solar noon.
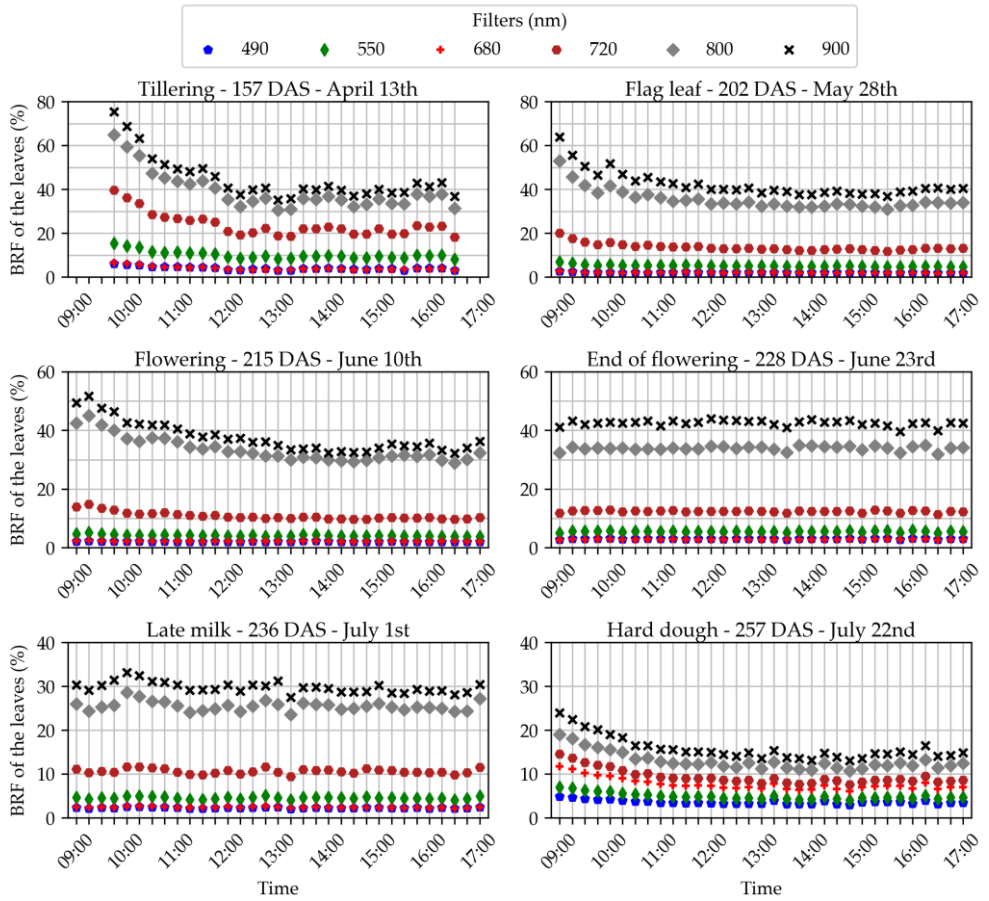
Figure 40. Average bi-directional reflectance factor (BRF) measured on the leaves throughout the day. Each subplot is dedicated to an acquisition date, designated by the development stage of the crop and the number of days after sowing (DAS). The evolution of the measured BRF is represented for the six different optical filters of the cameras, indicated by a colour and symbol code on each subplot.

5.4 Organ bi-directional reflectance factor in fertilisation trials

The results from Sections 5.2 and 5.3 indicate that the computed wheat BRF values were comparable within the same date if the sun was high enough or if the sky was cloudy. However, different dates presented slight variations in absolute BRF values, probably because of the configuration of the acquisition system. To deal with this effect, the values of the dark grey target on the reference panel were used to correct the wheat organ average BRF. Each wheat BRF value was multiplied by the ratio of the BRF measured on the target at this date to the theoretical reference BRF of the

target. In this multiplication factor, the BRF measured on the target was the mean of the values measured at the beginning and the end of the acquisition of all wheat micro-plots. In the rare cases where the difference between the BRF of the target at the beginning and at the end was larger than 3%, only the value closest to the theoretical BRF was considered. The furthest value was considered to be unreliable. The differences of reference BRF values within the same date were probably due to the positioning of the panel in the field. Less care was used during this step than in the all-day-long experiments. Here, the reference panel was held by human hands or dropped above the canopy. The two parts of the open suitcase containing the panel and the ArUco marker for panel segmentation could have been folded, modifying the relative position of the panel and the marker, thus leading to a panel segmentation error. A wheat leaf could also have been moved above the panel at the time of acquisition.

Figure 41 and **Figure 42** present the dynamics of the average BRF measured in the fertilisation trials and corrected using the BRF of the reference panel measured at the beginning and the end of the acquisition. For both trials, the curves show a coherent evolution, i.e., the points could be distributed on a smooth line. The same trends were observed for all the fertilisation objects and the BRF values stratified along with the fertilisation level. Those elements validate the consistency of the method. In **Figure 41**, a suspected increase in BRF can be observed for the 105-105-105 kgN/ha fertilisation object at 261 days after sowing (DAS) for the visible and red edge wavebands. This spike is probably due to the mass lodging on the micro-plots of that fertilisation object. Almost all the plants were lying on the ground, so the images had to be taken in the borders of the micro-plots, where their nutrition and development could have been different. Some plants were also tilted in those images.

Figure 43 shows the spectral dependence of the BRF at lag leaf and late milk stages for trial 21-F. It is a different way to present a part of the information from **Figure 41**. Nevertheless, it seemed important to include that wavelength-based vision. It highlights for example that, at flag leaf stage, the fertilisation objects can best be differentiated in the near infrared between 800 and 900 nm. At the previous vegetative stages this effect is much less marked and other wavelengths may be more interesting (**Figure 41**).

A final note is that those measurements concerned average BRF including both sunlit and shaded leaves. For further research it could be interesting to study the potential irradiance bias introduced by low and shaded leaves. This can be performed using image segmentation techniques able to separate sunlit and shaded leaves.
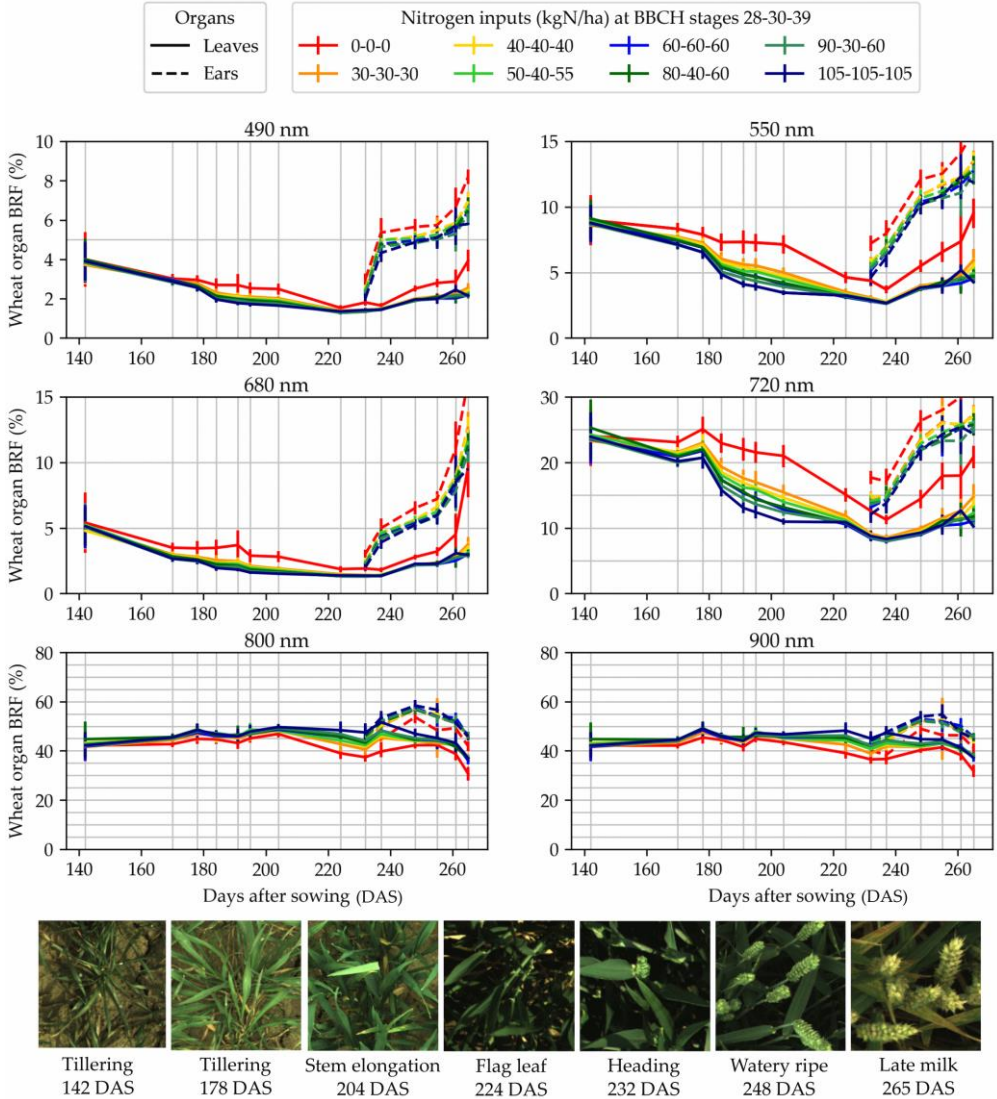
Figure 41. Dynamics of the measured bi-directional reflectance factor (BRF) during the measurement campaign in trial 21-F. The evolution is presented for the six monochrome cameras, designated by the central wavelength of their narrow-band optical filter.
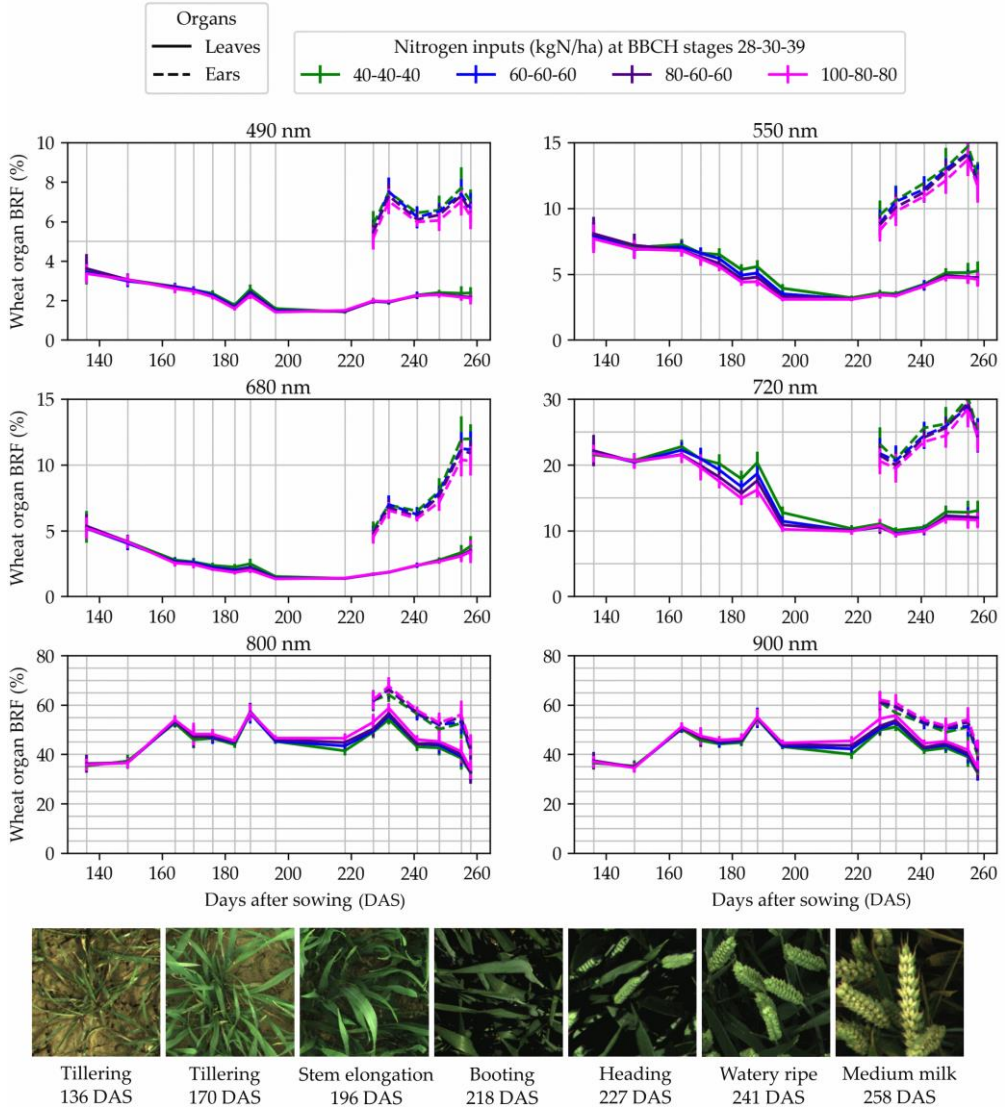
Figure 42. Dynamics of the measured bi-directional reflectance factor (BRF) during the measurement campaign in trial 21-FP. The evolution is presented for the six monochrome cameras, designated by the central wavelength of their narrow-band optical filter.
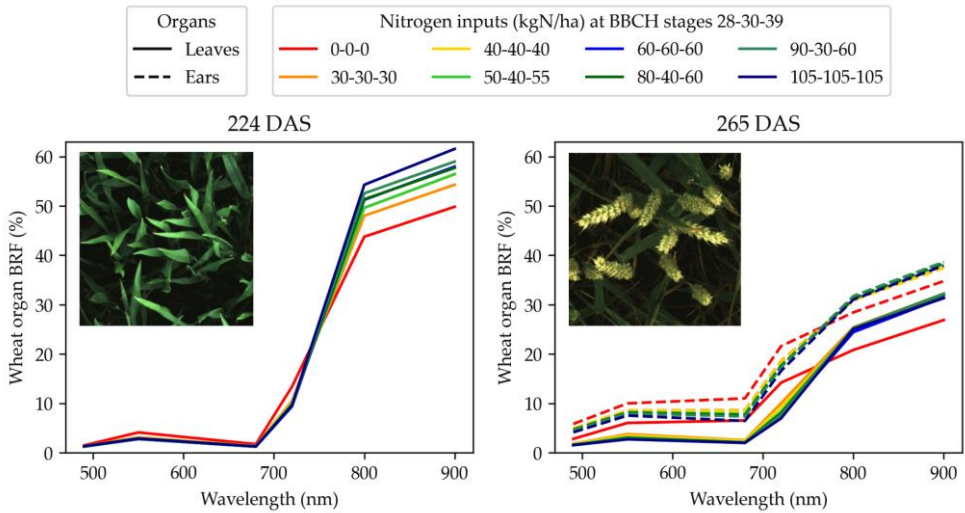
Figure 43. BRF of the trial 21-F fertilisation objects according to the wavelength. The relation is presented for two different wheat development stages: flag leaf stage (BBCH 39) and soft dough (BBCH 85), respectively 224 and 265 DAS. The pictures above the curves have been acquired by a RGB camera at the concerned stages.

6. Conclusions

This chapter detailed an automatic method to compute the bi-directional reflectance factor (BRF) of wheat while disentangling the contribution of leaves and ears. The method was found to allow for high-throughput measurements and did not necessitate the positioning of a reference panel in all imaged scenes. Multispectral images were acquired at the same time as incident light spectra. First, image digital numbers were converted to BRF values by exploiting camera response curves and the incident light integrated in the wavelength bands corresponding to the camera filters. Second, the BRF maps were segmented to separate the background, the leaves, and the ears.

The BRF measured on reference targets was robust throughout the day but showed variations with the acquisition date. As a consequence, when measuring the BRF of wheat micro-plots at several dates, the reference panel needs to be imaged at least once during each date. In this study, the panel was captured at the beginning and the end of the acquisitions, and it was used to correct the absolute BRF values. All over the wheat growing season, the method yielded consistent BRF values that were in good agreement with the fertilisation levels. Regarding the BRF of leaves and ears throughout the day, constant values were recorded in cloudy conditions or close to

solar noon. However, the organ BRF gradually changed in the morning under sunny or partially cloudy skies. This suggests that images for robust BRF measurement should preferably be acquired under cloudy conditions or close to solar noon.

In a broad context, this method will ease the high-throughput measurement of spectral features of wheat corrected for variations in the sunlight spectrum and promote their extraction at the organ scale in field conditions.

VIII

Applications and short term perspectives

1. Synopsis

The previous chapters described how plant traits at the organ scale could be extracted from a close-range multimodal imaging system. A number of technical challenges have been solved, and we believe it unlocks many possibilities to finely characterise the crop health and development. This chapter aims to provide an overview of the agronomic information that can be obtained from such plant traits. The explored topics are the dynamics of wheat morphology and physiology (Section 2), the biotic and abiotic stresses (Section 3) and the grain yield (Section 4). The reference agronomic measurements presented in Chapter III are used to validate the imaging techniques. Contrary to the previous chapters, this one is not based on peer-reviewed published results. It gathers exploratory analyses performed at the end of the research period, dedicated to highlight applications and short-term research perspectives.

2. Dynamics of wheat morphology and physiology

2.1 Dynamics of plant ratio

The plant ratio is the proportion of plant pixels in a nadir image. It comprises all the visible plant elements: the tillers, the green leaves, the damaged leaves and the ears. As explained in Chapter VII, in cloudy conditions it was computed by a threshold in the 800-nm image. When the ILS determined sunny conditions, thresholds in the 490-nm and RGB channels were added for the pixels of low values at 800-nm, which could be either soil or shaded leaves. An example of plant ratio dynamics is presented in **Figure 44** for the trial 21-F. The nadir plant ratio alone is limited in its capacity to represent the leaf area or the biomass because of the leaf overlaps and the 2D nature of the measure. It was however possible to highlight significant differences between fertilisation scenarios.
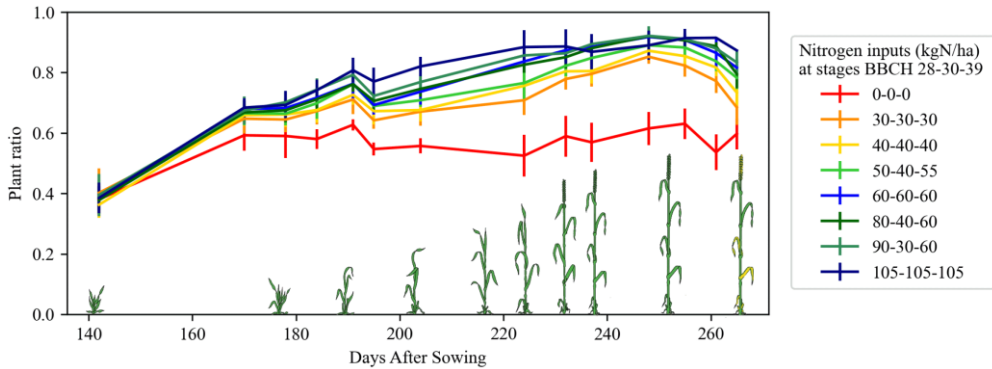
Figure 44. Dynamics of plant ratio for trial 21-F. The bar on each point indicates the standard deviation.

2.2 Dynamics of green ratio

The green ratio is the proportion of green pixels in an image, it is supposed to represent the proportion of healthy plant organs that contribute to photosynthesis. Those green pixels were identified using a threshold of 0.05 on the Excess Red vegetation index (Meyer et al., 2008) computed from the RGB images. An example of green ratio dynamics is presented in **Figure 45** for the trial 19-FP, from flag leaf stage to maturity. This example is particularly interesting because the crop was severely attacked by stripe rust. The curves show that the micro-plots that were protected with fungicide treatments were able to keep their green surface longer. Looking at the unprotected micro-plots, the smaller green surfaces were observed for the micro-plots that received the most nitrogen. It suggests that an over-fertilisation could have favoured the development of the disease, and end up with a reduction of the healthy green surface. This seems to demonstrate an interaction between the factors disease and fertilisation, as already noticed by Devadas et al. (2014).
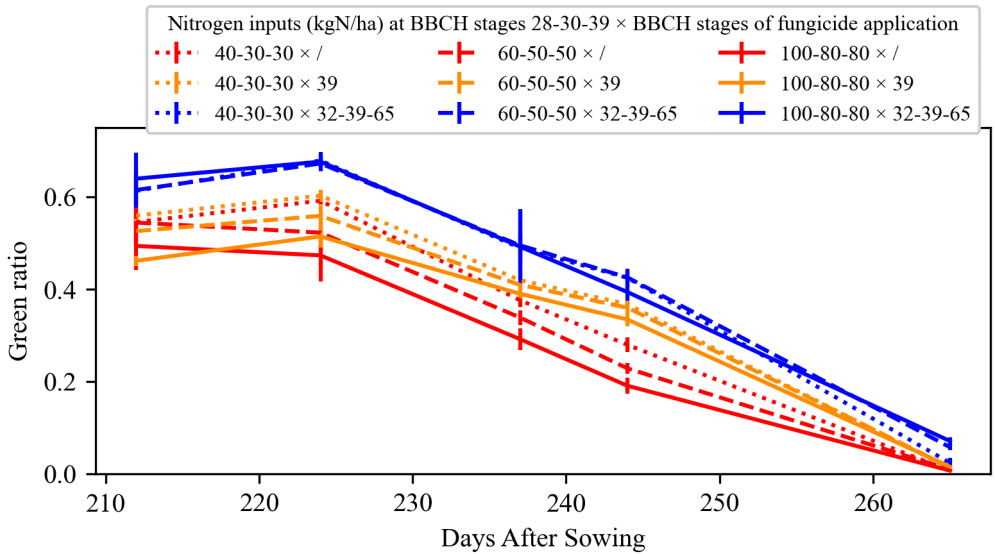
Figure 45. Dynamics of green ratio for the trial 19-FP from flag leaf stage to maturity.

In this case study, the curves of green surface were integrated from 212 to 265 DAS. Those integrated values were good predictors of the grain yield (**Figure 46**) for this trial, but similar relations were not observed for the other trials. The reduction of the photosynthetic green cover induced by the disease may have directly limited the grain filling, but the correlation between the integrated green cover and the yield does not necessarily imply a causal relationship.
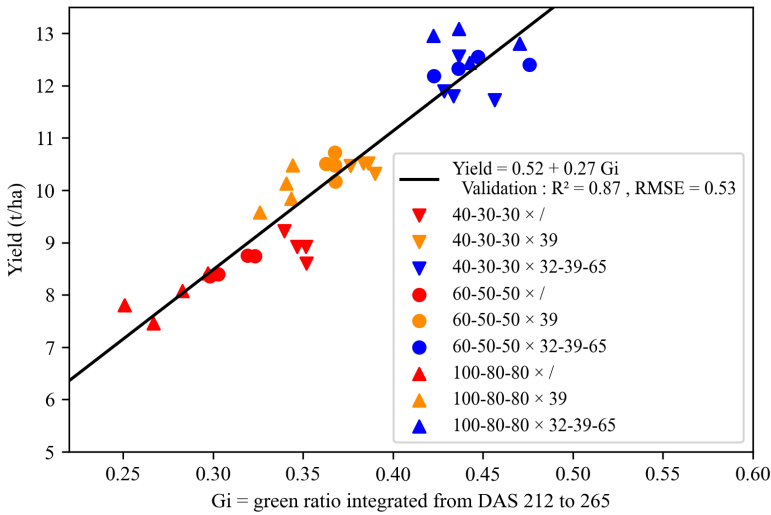
Figure 46. Relation between the grain yield and the integrated green ratio for trial 19-FP.

2.3 Dynamics of vegetation indices

While most of the studies used vegetation indices at the scale of the canopy, the approach developed in this work also provides indices at the scale of the organs. The indices can be computed for only the leaves or the ears, or even for specific canopy parts such as the sunny/shady leaves if a shadow detection algorithm is used, the lowest/highest leaves if the 3D information is exploited or from any other mask that could be derived from the registered multimodal images. Considering the six spectral bands of the multispectral camera array, the huge number of indices that can be computed from them and the multiple scales for which the values of the index in the image can be averaged, the multi-sensor system presented in this study is able to generate hundreds of vegetation indices of potential interest. In this chapter, for the purpose of demonstration, only three vegetation indices will be exploited. The Normalised Difference Vegetation Index (NDVI) (**Eq.17**) (Rouse et al., 1973) is one of the most widely-used of all (Araus et al., 2022). It exploits a NIR band that highlights the vegetation and a Red band that highlights the soil or the damage on the vegetation. It can be used at all the scales as it translates both the vigour and the quantity of vegetation. It is mainly influenced by the quantity of vegetation at the canopy scale, but when extracted at the leaf scale, it can provide information on the leaf health and nutrition status. The Normalised Difference Red Edge index (NDRE) (**Eq.18**) (Barnes et al., 2000) is based on the same construction as the NDVI but uses

a Red Edge band instead of a Red band. This index is especially known for its use to build the Canopy Chlorophyll Content Index (CCCI) (Fitzgerald et al., 2006; Cammarano et al., 2011). According to Fitzgerald et al. (2006) the chlorophyll content, proxy of the nitrogen content, induces a shift of the leaf reflectance in the Red Edge. The Red Edge Chlorophyll Index (CIREDE) (**Eq.19**) (Gitelson et al., 2003) is based on the similar hypothesis that chlorophyll content can best be assessed exploiting the Red Edge region of the spectrum.

$$\text{NDVI} = \frac{\text{BRF}_{800} - \text{BRF}_{680}}{\text{BRF}_{800} + \text{BRF}_{680}} \qquad \text{(Eq.17)}$$

$$\text{NDRE} = \frac{\text{BRF}_{800} - \text{BRF}_{720}}{\text{BRF}_{800} + \text{BRF}_{720}} \qquad \text{(Eq.18)}$$

$$\text{CIREDE} = \frac{\text{BRF}_{800}}{\text{BRF}_{720}} - 1 \qquad \text{(Eq.19)}$$

where BRF_{λ} is the bi-directional reflectance factor in a band centred at λ nm.

The **Figure 47** illustrates the kind of map obtained after computing a vegetation index on registered images. Exploiting the ear and the leaf mask, the average value of the index can be extracted for each type of organ. Then, the dynamics of the index can be studied throughout the growing season (**Figure 48**).
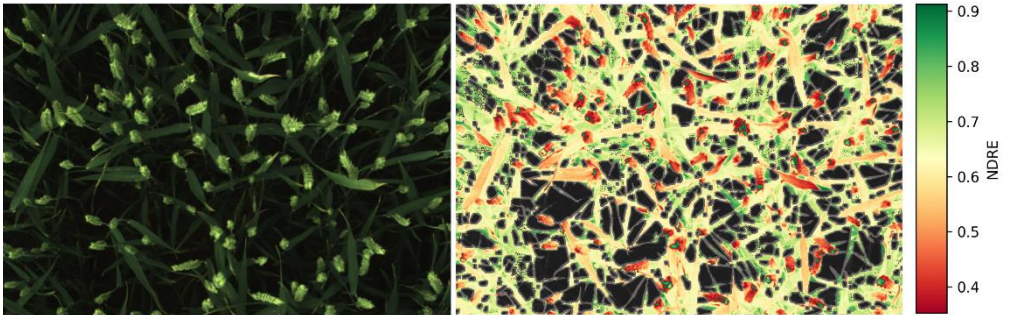
Figure 47. Map of NDRE vegetation index. The left frame is a registered RGB image and the right frame is the corresponding NDRE map. On that map, the colour scale indicates the NDRE for the wheat organs. The grey pixels are the parts of the wheat mask eroded during image registration (Chapter VI, Section 5.4). The black pixels represent the background.
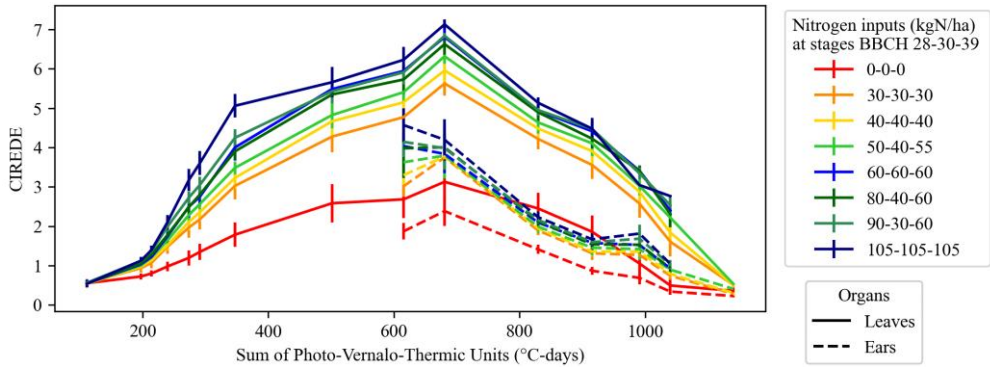
Figure 48. Dynamics of CIREDE vegetation index for the trial 21-F. The timeline is expressed in Photo-Vernalo-Thermic Units as detailed in Section 5.5 of Chapter III.

2.4 Estimation of LAI, dry matter and nitrogen status

Above-ground dry matter, LAI and nitrogen status of the crop are key agronomic parameters. They cannot be directly measured from nadir cameras but they can be estimated using a combination of traits measured by the cameras. Three main categories of predictors can be identified: the 2D ratios such as plant ratio or green ratio, the 3D traits such as plant height or leaf angle, and all the traits computed from the spectral reflectance. Most of the previous studies exploited only one category of predictor at a time, constrained by the camera type. Several advances demonstrate however the interest of combining the predictors from different categories. Liu et al. (2017) showed that adding the height profile from LiDAR to the green ratio improved the Green Area Index (GAI) estimation. Zou et al. (2017) demonstrated the interest of adding mean leaf angle to vegetation indices from aerial spectroscopy for the LAI estimation. Schirrmann et al. (2016) found an improvement of dry matter estimation by combining plant ratio and crop height. At this stage, crop height was measured manually, but they concluded that the fusion of height sensor and camera was a promising research direction. Tilly et al. (2015) investigated with some success the fusion of height and spectral measurements for dry matter estimation.

The multi-sensor approach developed in this work provides predictors from the three categories. A large wealth of predictors can be obtained. 2D ratios are green, leaf, ear or plant ratios. 3D descriptors include all the statistical descriptors of height points but also the mean leaf angle computed as described using the local fitting method described by Dandrifosse et al. (2020). The spectral predictors are the most numerous. They count the average BRF values in the six channels of the multispectral

camera as well as all the vegetation indices that can be calculated. Those spectral indicators can be extracted for the whole image or for some plant organs or specific parts of the canopy, for example only the sunny leaves or the upper leaves. Hundreds of predictors, even thousands, could be used.

While advanced feature selection methods could be investigated, the human logic and the crop knowledge already help to navigate through all these variables. For example, it seems unnecessary to include hundreds of vegetation indices based on the same spectral bands. A selection of twenty-one predictors was performed based on previous knowledge, tests and intuition. Those predictors were investigated through the lens of a principal component analysis. **Figure 49** shows the selected predictors and their relation with the first two principal components. In such a loading plot, the length of an arrow expresses how much weight the variable has on the principal components. When two arrows are close, the associated variables are strongly positively correlated. Here, the wide distribution of the arrows tends to prove the good selection of the predictors. It also highlights some redundant predictors. For example, the percentile 95 and the mean of plant height are strongly correlated. Maybe it is not useful to include these two height descriptors. Similarly, the BRFs at 800 and 900 nm are really close. Both bands belong to the NIR region. As already observed, they bring similar information about the wheat canopy. Two other variables relatively close are the BRF of leaves at 490 nm (blue) and 680 nm (red), maybe because both correspond to pigment absorption bands. Principal components can also be exploited to highlight trends and patterns in the data. Negative values of principal component two are clearly related to plant height traits. The other trends are less obvious. Negative values of principal component one seem linked to indices that increase with the health and the development of the canopy: NDVI, CIREDE, NDRE, green ratio, plant ratio. Interestingly, the NDVI-related variables, whose values decrease with the red reflectance, are opposed to the red reflectance variable (BRF of leaves at 680 nm). Similarly, the CIREDE and NDRE-related variables, whose values decrease with the red edge reflectance, are opposed to the red edge reflectance variable (BRF of leaves at 720 nm).
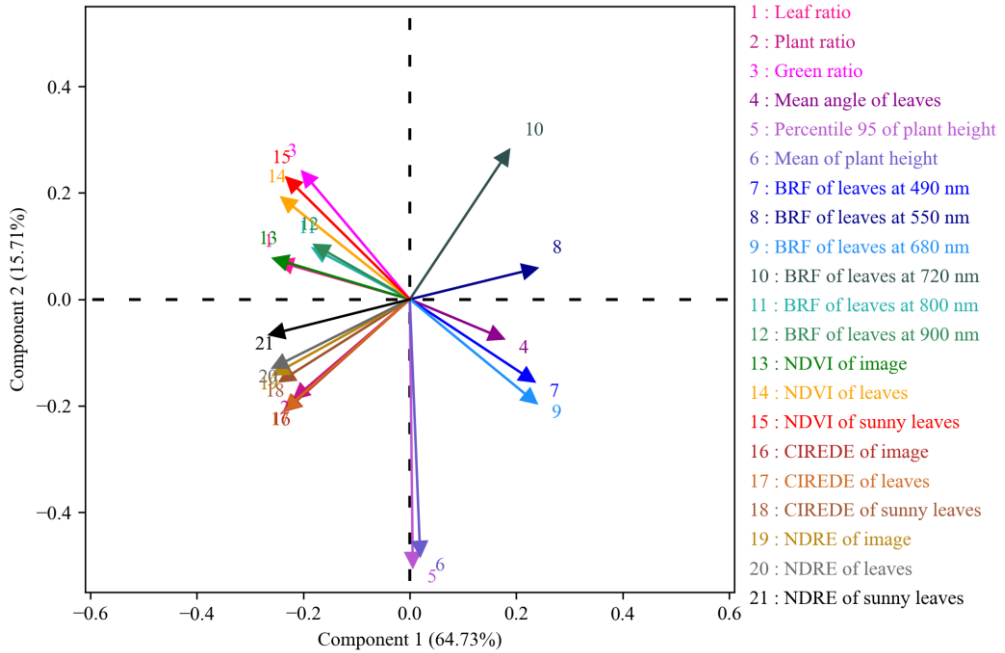
Figure 49. Loading plot expressing the relation between a set of plant traits of interest and their first two principal components. The analysis was performed using all the images of trials 20-F, 20-FP, 21-F and 21-FP.

From those traits, partial least square regression (PLSR) models were built to estimate the following agronomic variables: the LAI (m²/m²), the above-ground dry matter (t/ha), the leaf and above-ground nitrogen contents (kgN/ha), the leaf and above-ground nitrogen concentrations (%N) and the NNI. The reference measurements of those variables are detailed in Chapter III, Section 3.1. As those reference measurements and the image measurements were not performed on the same areas in the micro-plots, analyses were made on the values averaged by scenario of fertilisation and fungicide protection to diminish the impact of heterogeneities within the micro-plots. The PLSR model was chosen for its capacity to deal with numerous correlated predictors. It was implemented using the Python scikit-learn library (version 0.22.1). For the estimation of each agronomical variable of interest, the predictor traits were firstly selected based on human knowledge. For example, to estimate the LAI the 2D leaf ratio was thought as a good predictor, complemented with leaf angle and plant height to account for the 3D architecture of the canopy. The predictor sets were then empirically adjusted based on external validation performances. The final selected predictors are presented in **Table 17**.

Table 17. Traits selected for PLSR models to estimate agronomic variables. Variables were: LAI, above-ground dry matter (DM), leaf and above-ground (ABG) N content, leaf and above-ground N concentration, and nitrogen nutrition index (NNI).

	LAI	DM	N content		N concentra.		NNI
			Leaf	ABG	Leaf	ABG	
Leaf ratio	✓	✓	✓	✓			✓
Plant ratio	✓	✓			✓	✓	
Green ratio	✓	✓			✓	✓	
Ear ratio				✓			✓
Mean angle of leaves	✓	✓					
Percentile 95 of plant height		✓	✓	✓	✓	✓	✓
Mean of plant height	✓	✓					
BRF of leaves at 490 nm		✓			✓	✓	
BRF of leaves at 550 nm		✓			✓	✓	
BRF of leaves at 680 nm		✓			✓	✓	
BRF of leaves at 720 nm		✓			✓	✓	
BRF of leaves at 800 nm		✓			✓	✓	
BRF of leaves at 900 nm		✓					
NDVI of image	✓	✓	✓	✓			✓
NDVI of leaves		✓			✓	✓	
NDVI of sunny leaves		✓					
CIREDE of image		✓		✓			✓
CIREDE of leaves		✓	✓		✓	✓	
CIREDE of sunny leaves							
NDRE of image		✓		✓		✓	✓
NDRE of leaves		✓	✓		✓		✓
NDRE of sunny leaves							

Two different approaches were used to validate the models. Firstly, the models were trained on data from the trials 20-F, 20-FP, 21-F and 21-FP and validated using a 10-fold cross-validation with a test size of 30 %. The data from the four trials were judged of importance to build robust models, so it was pertinent to use cross-validation in order to evaluate the performances in training scenarios gathering data from all the

wheat varieties and climate years available. Secondly, the models were trained on data from the trials 20-F, 20-FP and 21-FP, and evaluated on the data from trial 21-F. It was the closest possible test to an external validation, but since the wheat variety of trial 21-F was the same as for 20-F, and that the climate year was the same as for 21-FP, this was not really an external validation. This second test was nevertheless a good way to evaluate the robustness of the models and the possibility to use them for new wheat canopies. Performances are presented in **Table 18**.

Table 18. Performances of PLSR to estimate agronomic variables. Variables were: LAI, above-ground (ABG) dry matter (DM), leaf and above-ground N content (kgN/ha), leaf and above-ground N concentration (%N), and nitrogen nutrition index. Performance indicators were the determination coefficient (R^2) and the Root Mean Square Error (RMSE). The range indicates for each variable the min and max observed values. Performances were yielded for i) cross-validation: training and cross-validation on trials 20-F, 20-FP, 21-F, 21-FP; ii) external- training: training and validation on trials 20-F, 20-FP, 21-FP; and iii) external – validation: training on trials 20-F, 20-FP, 21-FP and validation on 21-F. All the values were averages per management scenario for one date.

Estimated trait	Range (min-max)	Cross validation		External : training		External : validation	
		R^2	RMSE	R^2	RMSE	R^2	RMSE
LAI (m²/m²)	0 – 7.2	0.63	0.79	0.67	0.72	0.66	0.95
ABG DM (t/ha)	0.9 - 21.2	0.97	0.95	0.97	0.76	0.86	2.19
ABG N (kgN/ha)	27 - 310	0.81	24.11	0.84	19.78	0.8	32.14
Leaf N (kgN/ha)	5 - 134	0.83	12.40	0.89	11.0	0.61	15.12
ABG N (%N)	0.77 - 4.31	0.68	0.34	0.83	0.2	0.65	0.57
Leaf N (%N)	0.48 - 4.34	0.91	0.29	0.97	0.18	0.75	0.42
N nutrition index	0.46 - 1.31	0.58	0.11	0.75	0.09	0.59	0.13

Good models were obtained for all the agronomic variables of interest, but the lower performances for external validation suggest that more years of data would be required to build more robust models. Nevertheless, a few easy solutions could already be deployed to improve the performances.

The most obvious solution is probably the exploration of automatic feature selection methods. The feature selection algorithm could explore the hundreds of image traits and unearth key predictors that have not been selected by humans. Attention should be paid that the algorithms do not select features too correlated with each other.

Moreover, human intuition is probably the best way to avoid selecting traits that would perform well for the available data but reveal poor adaptability to future conditions.

Another solution, that requires only a few supplementary data, would be to ensure that the models are trained and validated for all the development stages. This was not true for this work. At tillering stage, images were acquired along with reference measurements only for trial 21-F. This could explain the degradation of performances when that trial was used for validation but not for training.

The classic image analysis approach, necessitating tedious feature selection from multi-sensor data, may also be questioned regarding the rise of deep learning. Some recent studies demonstrate the possibility to estimate the key agronomic variables from only RGB images using deep learning models (Ma et al., 2019; Castro et al., 2020; de Oliveira et al., 2021).

3. Biotic and abiotic stresses

3.1 Water stress

The multi-sensor approach provided a mask of the leaves (Chapter VII) to extract their temperature in the registered thermal images. Furthermore, the mask of the leaves was segmented into sunny and shady leaves using an Otsu threshold (Otsu, 1979) followed by a 3×3 median blur applied in the 800-nm image. The variable of interest to study the water stress was the average temperature of sunny leaves, because they are supposedly more sensitive to plant water status than shady areas (Jerbi et al., 2015). That leaf temperature was combined with environmental measurements (Chapter III, Section 4) to build the Crop Water Stress Index (CWSI) (Jackson et al., 1981; Leinonen et al., 2004) using the equations of energy balance described by Jones (1999) to obtain the temperatures of the wet and dry surfaces. The equations are not detailed here for brevity but the details can be found in the referenced papers. All the environmental variables were measured with a time step of 20 s but the values used in the equations were the average in a 5 min period before the related image acquisition. That choice was made to account for the response time of stomata, from 1 to 10 min according to Lawson et al. (2014). The air temperature was directly provided by the sensors. The vapour pressure deficit was obtained from the relative humidity and the air temperature. The net isothermal radiation was considered equal to the absorbed short ave radiation (Leinonen et al., 2006). It was computed using the sun zenith angle, the direct and diffuse shortwave radiations measured by the sensors

and the shortwave leaf absorbance – set to a value of 0.7. The measured wind speed was reduced by a factor of 0.55 to account for its change in the canopy, at the leaf boundary layer (Leinonen et al., 2006). Finally, the characteristic length of a leaf was estimated at 5 cm (Schuepp, 1993).

The dynamics of CWSI is presented in **Figure 50** for the trial 20-FP. This trial has been chosen because the spring 2020 was characterised by drought conditions, unlike the very rainy 2021. On the micro-plots of the trial 20-FP, leaf rolling was visually observed, indicating water stress. This trial was also selected because images were always acquired at the same hours – between 10 a.m. and 12 a.m. Surely, it would have been better to acquire the images in the afternoon, when the effects of the sun and air temperature on the wheat water status are the most important.

Unfortunately, except the rain events, no data were available to establish a water balance of the plot. The rain events could however indicate when the crop received water, and so when the stress index could have diminished. Although it is rather a qualitative information, it is at least an element to interpret the dynamics of the CWSI. For this reason, the rain events were added in **Figure 50**.
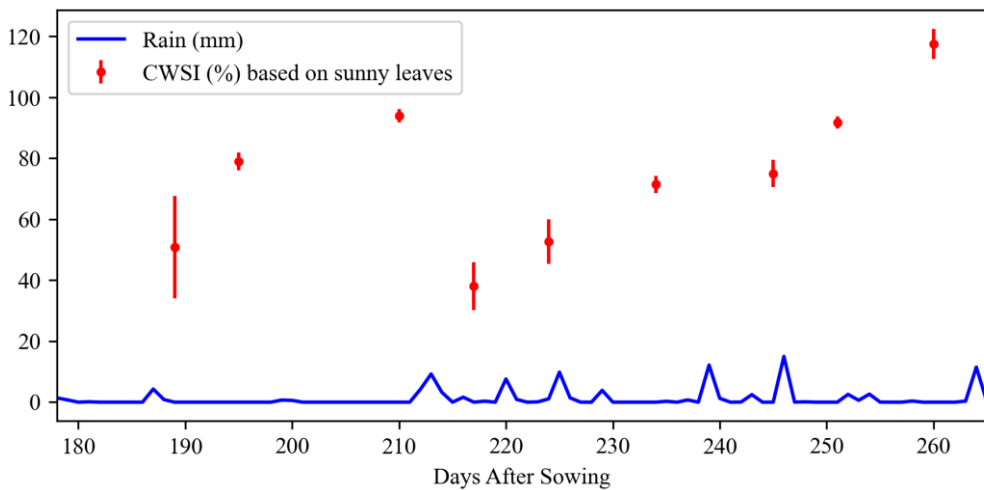
Figure 50. Dynamics of CWSI regarding the rain events. The values of CWSI are the mean of all the fungicide and fertilisation treatments of trial 20-FP. The bar on each point indicates the standard deviation between the treatments. Higher values of CWSI represent a higher level of stress.

For clarity, the CWSI are presented as the mean of all the fungicide and fertilisation treatments. Indeed, few differences were observed between the treatments and no

trend emerged. The interaction between water stress and other stresses such as nitrogen or diseases is an interesting topic that should be investigated later, after more rigorous validation of the CWSI obtained with our system and using dedicated experimental designs.

The dynamics of CWSI, illustrated in **Figure 50**, is logical regarding the drought from 190 to 210 DAS. Later, the CWSI increases as the season advances, which is also logical, especially because the rains were limited. This seems to demonstrate the success of the method, at least in a relative way. In terms of absolute value, the index exceeded 100% at the end of the season, which may indicate that the parameters used yield an overestimation of the CWSI.

Those results, although promising, should be carefully considered. The method counts numerous improvement points. The environmental measurements should be performed close to the studied micro-plots, instead of coming from a weather station located 1 km away. For the assessment of water stress, the rain events should be complemented with soil water content measured in the micro-plot of interest. Another idea to validate the method would be to introduce crop water supplies as a factor in the experimental design, for example using controlled irrigation and removable greenhouses to prevent the rain. A popular validation solution is also to measure the stomatal conductance of wheat leaves using a porometre and compare that reference measurement to the stomatal conductance that it is possible to estimate with thermal imagery, using similar equations and variables than for the CWSI.

Basing the method on the full energy balance is another questionable choice. It necessitates not only the measurement of wind speed, air temperature, air humidity and net radiation absorbed by the leaf, but also the estimation of the leaf characteristic length and the wind speed reduction factor in the canopy, which can be complex to determine, and should probably be computed for each development stage. All those measures and estimations can introduce errors. As explored by Leinonen et al. (2006), the method based on the full energy balance may generate less accurate estimates of stomatal conductance than a method that would use reference targets in the thermal images. Even if reference targets present a major flaw, that is the need to position the targets in the observed scene, it seems the best option to improve the method.

The classical method to derive CWSI with targets is to include in the thermal image a dry and a wet reference (Jones, 1999; Leinonen et al., 2006; Costa et al., 2013). Those references pose some conception issues. A possible approach is to use real leaves. The wet reference is obtained by spraying water on a leaf, and the dry reference is obtained by covering a leaf with a Vaseline-like substance to prevent transpiration

(Costa et al., 2013). These manipulations are however laborious to implement in the field and not in agreement with high-throughput acquisitions. For this reason, many artificial references have been proposed (Pou et al., 2014; Maes et al., 2016). To prepare the next measurement campaign, we built hemispherical reference targets inspired from the ones described by Apolo-Apolo et al. (2020). Those references were easy to realise using a 3D printer and some green cloths (**Figure 51**).
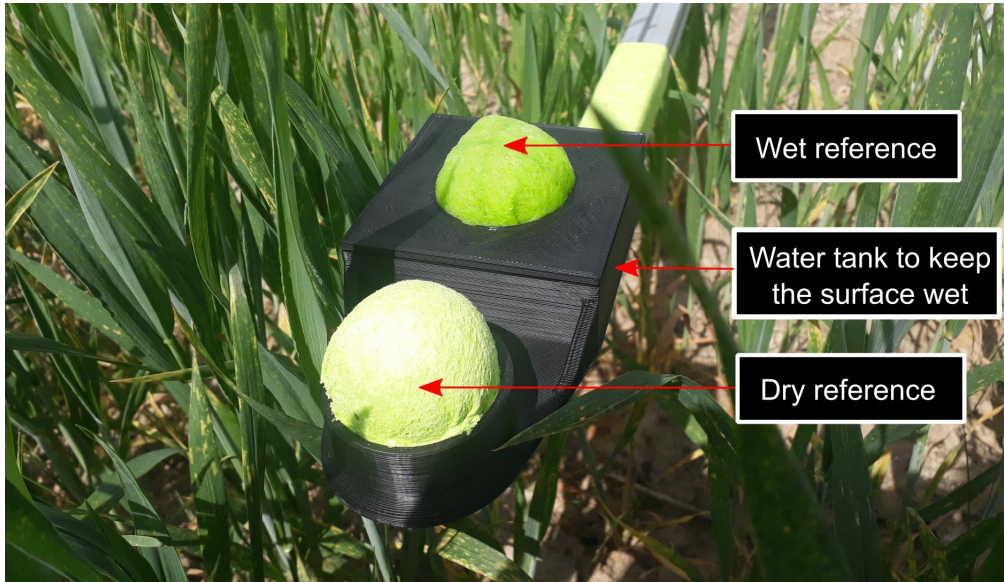
Figure 51. Wet and dry reference surfaces for thermal imaging.

3.2 Disentangle biotic damage and nitrogen stress

Previous studies have demonstrated that imaging methods could detect in the field nitrogen deficiency (Baresel et al., 2017; Fernández et al., 2019) or the presence of fungal diseases (Zheng et al., 2019; Bebronne et al., 2020; Dehkordi et al., 2020). Nevertheless, those tests were realised in specific experimental designs where only one of the two types of stress was generated. To our knowledge, no method was developed to quantify the two stresses when both are present. Regarding the possible interaction between the plant nitrogen and the fungal diseases (Simón et al., 2020), this would be of high interest.

The proposed approach relies on a double hypothesis:

A) The symptoms of a biotic attack appear as localised strong change of reflectance on the leaves. This corresponds to the spots of fungal diseases or the damage caused by insects.

B) The symptoms of a nitrogen deficiency occur as a global small change of reflectance of the leaves.

The method would require to build a mask of the leaf damage, it would include the biotic damage and the senescent leaf parts. It is supposed that basic image segmentation techniques would not allow to distinguish those several types of damage, but that they would allow to separate them from healthy green leaf areas. Once the damage mask is established, the damage could be quantified by determining the proportion of damage pixels in the image. The nitrogen deficiency could be quantified by computing the mean of vegetation indices on the non-damaged leaf area. Other traits such as the 2D vegetation ratios or the 3D information should also be taken into account, as demonstrated in Section 2.4.

First tests were realised to build a mask of the leaf damage. Empirical trials showed that a threshold of 0.73 on the NDVI vegetation index applied on the leaf pixels was a good choice to isolate the damage (**Figure 52**). The percentage of damage measured from the images was compared to human scoring of disease (Chapter III Section 3.2) for three data acquisition dates of trial 21-FP for which fungal diseases were observed (**Figure 53**). No other biotic damage was recorded for those dates.
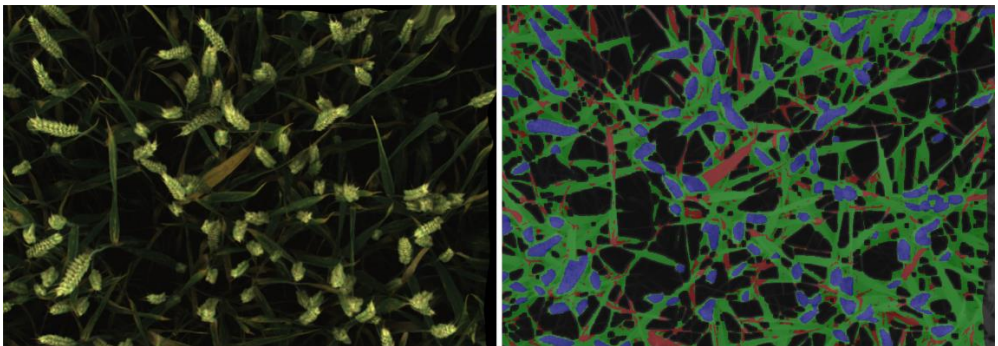
Figure 52. Map of foliar damage. The left frame is a registered RGB image and the right frame is the corresponding damage map. On that map, foliar damage appears in red, ears in blue and other wheat elements in green. The grey pixels are the parts of the wheat mask eroded during image registration (Chapter VI, Section 5.4). The black pixels represent the background.
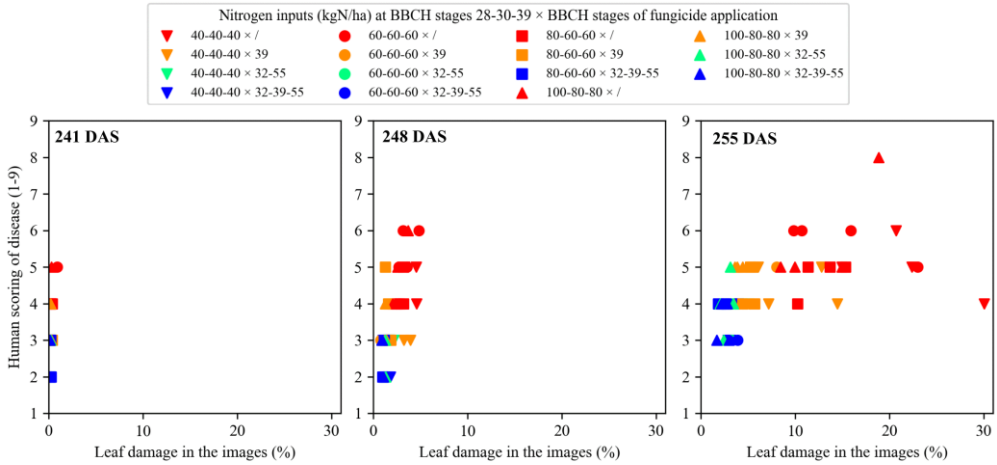
Figure 53. Dynamics of leaf damage measured by imagery and human scoring of disease. Data were recorded on trial 21-FP at 241, 248 and 255 days after sowing (DAS). The human score for a micro-plot is the maximum of the scores given for septoria tritici blotch, stripe rust and brown rust.

At 241 DAS, almost no damage was detected in the images but humans observed the presence of diseases. They could spot the spores by manually moving wheat leaves to look for the lowest leaves. At 248 DAS, the micro-plots for which the highest percentages of damage were detected were also the ones receiving high scores from human operators. The same conclusion was drawn from the results at 255 DAS, but the relation between image damage and human scores was different. As the diseases reached the highly positioned leaves, far more damage was visible in the images.

From those results, it appears that human observations are well adapted to assess the early presence of diseases, while nadir imagery is suitable to quantify the damage on the highly positioned leaves due to the disease activity. Both approaches are complementary and may serve different purposes: determining the need for fungicide intervention, evaluating the behaviour of the crop or predicting the impacts of the disease on crop development and yield. The results here do not indicate that multispectral nadir imagery would be suitable for the early detection of disease but other studies propose approaches in this direction (Bravo et al., 2003; Schirrmann et al., 2021).

The main limit of the biotic damage detection method developed in this section is the use of a unique threshold based on a single vegetation index. The index and the threshold value may not be the most adequate. Tillers, included in the "leaf" mask, were sometimes labelled as damage. Moreover, if some yellow ears were missed

during the ear segmentation, and thus included in the leaf mask, they were also marked as damage. Finally, all senescent elements were counted as damage. This could be a concerning issue regarding the possibility to disentangle the biotic and the nitrogen stress.

Further efforts should focus on validating a robust method to detect the areas of biotic damage, whether they result from fungal diseases or insect bite. Deep learning based on RGB images appears as a promising way to detect those areas without confusion with healthy plant parts, and even to determine the type of damage, looking at both the shape and the colour of the damage. In wheat fields, deep learning already showed results for the detection of fusarium head blight (Su et al., 2021) and stripe rust (Schirrmann et al., 2021). Genaev et al. (2021) succeeded in identifying the main wheat diseases based on individual leaves captured in the field. It would also be interesting to combine nadir image acquisition with mechanical systems able to spread the plants, so that lower leaves are included in the images.

4. Yield estimation

Regarding yield estimation, much relies on crop models, or crop models fed with remote sensing data (Basso et al., 2013). The topic is huge and the point of this section is not to review it. We wanted to investigate the possibility to estimate yield based solely on data from a close-range multi-sensor system.

Previous attempts for yield estimation from close-range images include indices of green surface from RGB images (Fernandez-Gallego et al., 2019; Gracia-Romero et al., 2019) or all the methods to estimate the ear density (Chapter V), but it was found no approach taking full advantage of the close-range information. The ideal method should be able to account for the three yield components: the ear density, the number of grains per ear and the weight of the grains. But only the ear density can be directly measured by imagery (Chapter V). To our knowledge, no high-throughput machine vision method exists to measure the number of grains and the weight of the grains in the field. Therefore, other plant traits must be used to account for these components. The elements and the development stages determining each of the three components are represented in **Figure 3**. For the weight of the grains, a good solution may be to integrate the green surface during the grain filling stages as explained in Section 2.2 of this chapter. The number of grains per ear, on its side, is established over a long period, from tillering to flowering. It would probably be interesting to integrate the green surface during this period as well. Nevertheless, this component is probably the

most difficult to predict using traits from in field imagery. Yet, it is important to include predictors related to the three yield components, especially because of compensation mechanisms.

For this study, the yield estimation was investigated in trials 20-F and 20-FP. The trials 21-F et 21-FP were not considered because of important lodging at the end of the season due to heavy rains and storms. The selected predictors were the ear density, the 95th percentile of plant heights on the date it was the maximum, a series of traits integrated from tillering to flowering and the same traits integrated from flowering to ripening. The integrated traits were the green ratio, the NDVI of image, the NDRE of image, the CIREDE of image, the NDVI of leaves, the NDRE of leaves and the CIREDE of leaves. The integration was based on the sum of Photo-Vernalo-Thermic Units, i.e. growing degree-day corrected for photoperiod and vernalisation (Duchene et al., 2021). The first period, influencing the ear fertility, was considered from 412 to 725 °C-days and the second period, influencing the grain filling, was considered from 725 to 1139 °C-days, based on the observation of wheat phenology in the field during the 2020 season. To improve the robustness to the number of measurement points and the precision of the measurements, the dynamics of the traits of interest were smoothed using a spline of degree three computed by the function UnivariateSpline from the scipy python module (version 1.7.3) (**Figure 54**).
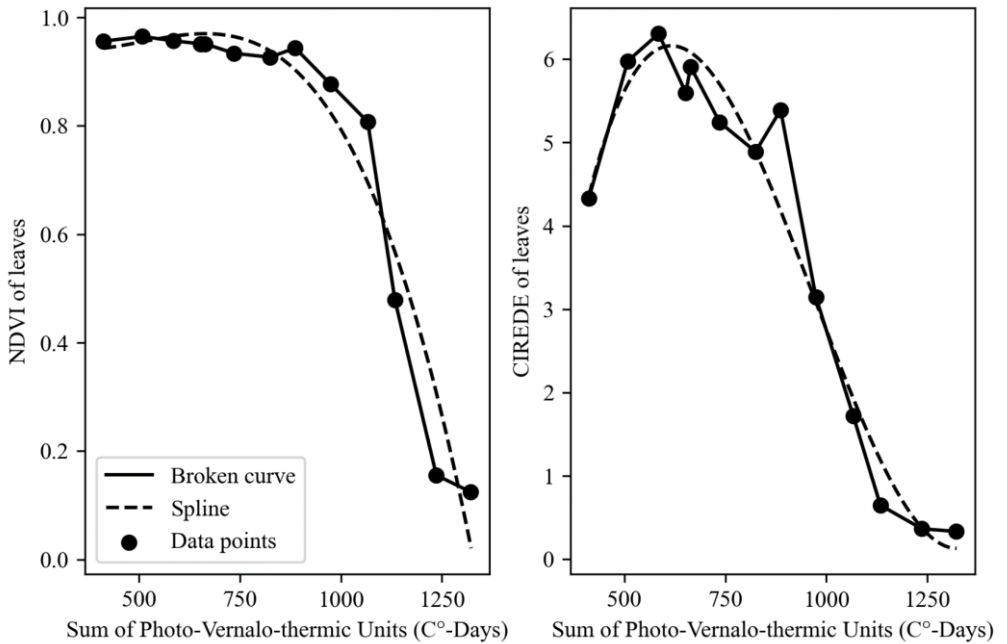
Figure 54. Examples of splines to smooth the measured dynamics of traits. The figure illustrates the dynamics of the average NDVI and CIREDE vegetation indices for the micro-plots of trial 20-F fertilised with 60-60-60 kgN/ha at BBCH stages 28-30-39.

Figure 55 illustrates the relations between grain yield and several predictors. For some predictors – the green ratio integrated over the grain filling period, the max plant height –, the relation between the yield and the trait is different for the two field trials. The best univariate model was built using the NDVI of image integrated over the grain filling period. The linear relation was well adapted for both trials. It was validated by a 10-fold cross-validation with a test size of 30 %. The R^2 was 0.71 and the RMSE 0.32 t/ha. The same RMSE and a R^2 of 0.72 were obtained for a PLSR model trained with all the mentioned predictors. Nevertheless, it is not likely that a single predictor would be robust in all the conditions. More field trials should be studied to highlight a robust approach. It is also to note that this analysis was limited to a set of predictors, thought to be of interest. A more in-depth study could identify other useful yield predictors.
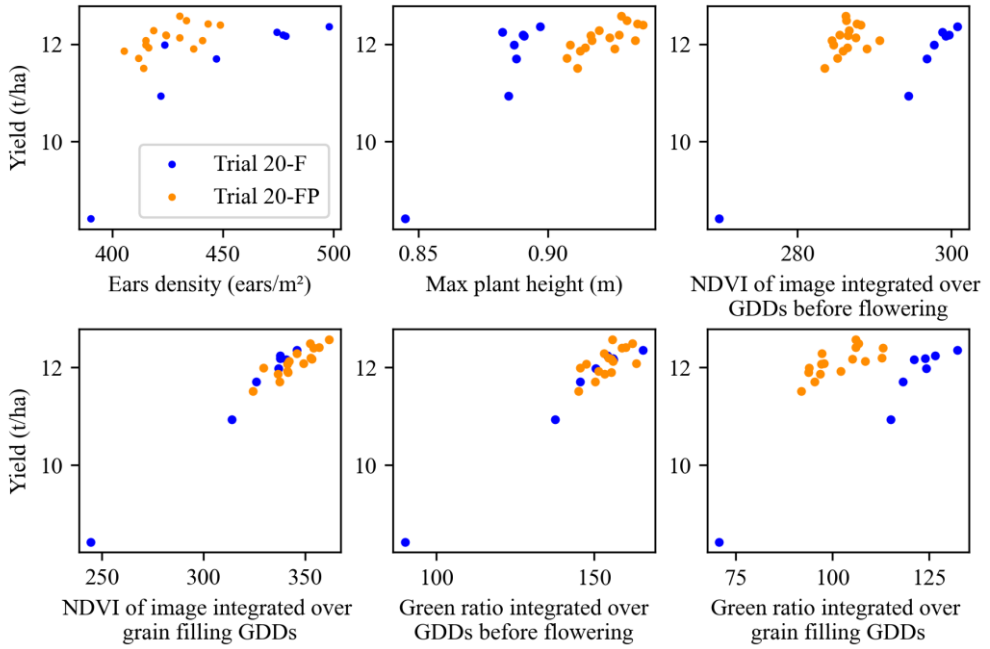
Figure 55. Relations between grain yield and different predictors. The measures from the trials 20-F and 20-FP are represented by different colours.

IX

General conclusions and discussion

1. Main outcomes

This study developed and investigated a multi-sensor pipeline to extract morphological and physiological wheat traits in field conditions, at the organ scale, and for all the crop development stages.

- In Chapter IV, a stereo vision approach was proposed to compute the height of wheat organs from a pair of RGB cameras. The height of ear tops was retrieved with an accuracy of 96.5%. But the height maps generated by stereo vision brought far more information than manual measurements, given that height values were associated with most of the plant elements. Moreover, the distance between the plants and the cameras was a prerequisite for further analyses on the images from the multi-sensor system.

- In Chapter V, RGB deep learning algorithms were exploited to obtain a robust segmentation of the wheat ears for all the development stages. The average accuracy was 96 %. Deep learning also yielded the number of ears in images, that was divided by the image footprint computed by stereo vision to obtain the ear density, i.e. the number of ears per square metre.

- In Chapter VI, eight image registration approaches were investigated to align pixel to pixel the images from the RGB cameras, a multispectral camera array and a thermal camera. The best method exploited local image deformations and succeeded in registering the images with an average error of 2 mm. For thermal images, the method was not reliable so a global transformation of the image was preferred, with an error around 3 mm.

- In Chapter VII, using the incident light spectrum, the opto-electronic response curves of the multispectral camera and image segmentation, a method was proposed to convert the registered multispectral images to bi-directional reflectance factor (BRF) maps at the organ scale.

- In Chapter VIII, it is developed how the plant traits extracted by the multi-sensor system can best be used to derive useful agronomic information. The leaf area index, the above-ground dry matter, the above-ground nitrogen concentration, the above-ground nitrogen content and the nitrogen nutrition index were estimated respectively with external validation RMSE of 0.95 m²/m², 2.19 t/ha, 0.57 %N, 32.14 kgN/ha and 0.13. The human and

machine evaluation of foliar disease damage were compared. And the grain yield was estimated with a cross-validation RMSE of 0.32 t/ha.

2. Limitations

This work sought to demonstrate how a combination of cameras could enrich the characterisation of a wheat canopy more than the addition of their individual contributions, and could allow to extract traits at the organ scale. However, these four years of work also showed how such a multi-sensor approach can be tough to implement.

The acquisitions in the field did not always go as planned. A first thing to consider is that increasing the complexity of the acquisition system also increases the probability of an issue. As the acquisition platforms were re-assembled at each field trip, some parts could be badly positioned or screwed, for example the camera connectors or the cosine corrector of the incident light spectrometer. Some acquisition parameters could also have been set to an unsuitable value. It was especially a problem for the micro-MCA multispectral camera array because the images were stored in SD cards and could not be checked in real time. It was not possible to verify if the exposure was satisfying. Other issues could be the unwanted movement of the focus or aperture screw, the presence of an insect on a camera lens or simply forgetting to remove the protection cap of a lens. For the field missions, it seems reasonable to consider that a sensor had a probability of 3 % of not acquiring the data correctly. If only one sensor is used, then the probability of failure of the mission is 3%. For our multi-sensor approach, it was necessary that all the sensors record correct data. Considering five sensors, each with a failure probability of 3 %, the probability that at least one sensor did not record proper data was 14 %. This result is obtained by **Eq.20**.

$$P_{1failure} = 1 - P_{Nsuccess} = 1 - (P_{success})^N \qquad (Eq.20)$$

Where $P_{1failure}$ is the probability to obtain at least one failure, N is the number of sensors (in this example 5), $P_{Nsuccess}$ is the probability to obtain a success for the N sensors, and $P_{success}$ is the probability of one sensor to obtain a success (in this example 0.97).

The second difficulty regarding data acquisition was the necessity to trigger all the sensors simultaneously. The control of the sensors could not rely on the proprietary software. It required to master the Software Development Kits (SDK) of the devices

and integrate them into a Python script coding for a custom graphical user interface. Such a script rapidly became complex. Moreover, the Python SDKs were sometimes poorly documented or did not include all the camera parameters of interest.

Another complexity of our multi-sensor approach is the need for calibrations accounting for the relative position of the cameras: a calibration for the stereo pair of RGB cameras (Chapter IV, Section 3.1) and a calibration for the registration of the images from all the cameras (Chapter VI, Section 4.3). If the cameras move, even slightly, it is necessary to perform the calibrations again.

The multi-sensor system comprised very specific devices. If one of them had reached the end of its life, it would have jeopardised the whole pipeline. The exact same device would have been necessary to replace it without changing all the acquisition and data processing scripts.

And that data processing was also complex, as illustrated throughout this work. Many steps were necessary and often the output of one step is the input of another. The length of that pipeline increased the possibility of errors or exceptions. It also necessitated expert operators to deal with the huge amount of code lines and manage the different processing scripts and their input parameters. The process was divided into several main steps. Some of those steps necessitated huge computing power: especially the local method for image registration and the deep learning approach to obtain the mask of the ears.

Regarding the storage of the data, raw data from all the devices with their maximum colour resolution, and considering one hour of measurement (sixty micro-plots, two hundred forty imaged scenes), necessitated 14 Go. That number climbed to 19.3 Go when converting the raw multispectral images to a usable TIF format. Recording images with the highest possible colour resolution was however not necessary, since the algorithms could not exploit them. Considering a classic colour resolution of 8 bits for all the images, the raw data set for one hour could be reduced to 4.3 Go, or 6.1 Go with the TIF images. To this, it must be added all the intermediary images that the user would save: the registered images, the height maps, the organ masks, …

Considering all those limitations, the multi-sensor set-up and the processing pipeline as implemented for this work would be hard to manipulate for users who would not have spent four years of their life on it. End of the story? Not at all. Now that a number of technical and scientific challenges have been completed, future developments will be able to focus on improving the operability of the system.

First important advances have already been made to improve the hardware. We developed a low-cost camera array gathering six monochrome cameras equipped with

narrow optical filters and two RGB cameras aiming to perform stereo vision (**Figure 56**). Both RGB and monochrome cameras were part of AR0135 modules (Arducam Technology Co., Ltd, China). Those modules can easily be triggered through a Python SDK, there is a software trigger to improve the synchronisation of the frames from the different cameras. The design of the device has been made so that all the pieces can easily be interchanged and replaced. Modularity is the key. Optical filters can be replaced in a few seconds, which can be useful to investigate new wavebands without buying a new camera or using a hyperspectral device. The design was also thought so that everything fits in a closed box, to prevent the dust from damaging the system.

A GPS-RTK should also complement the hardware of the system. It would allow to locate the images and automate the acquisition at the same positions. That way, it would be possible to study the same zone within a micro-plot throughout the season.
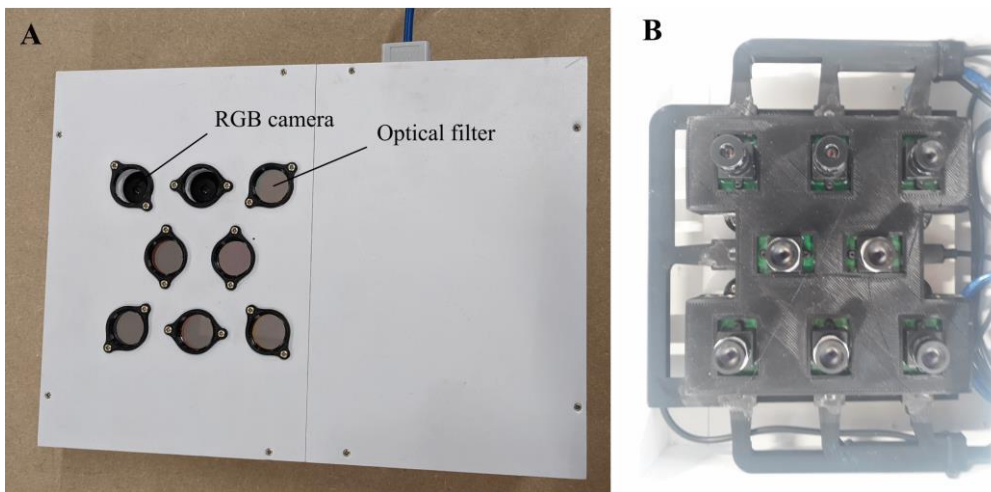
Figure 56. Development of a custom multimodal camera array. A) The box. B) The inside of the box.

Improvements should also be made at the software level. This multi-sensor pipeline has been built component after component as the research progressed. The programming architecture could be simplified and optimised if the pipeline is redesigned from the beginning with the whole picture in mind. To extract the plant traits of interest from the raw images, the final user should run only one script, with an associated configuration file, or benefit from a graphical user interface.

3. About high-throughput phenotyping adoption

For a non-expert, high-throughput phenotyping methods can appear attractive. Those methods indeed take advantage of the enthusiasm for new technologies such as robotics, advanced image analysis and machine learning techniques. But they can also appear frightening. For crop physiologists and breeders accustomed to traditional phenotyping approaches, the investment and the need to master new technologies may be a brake. They may also be confused by a huge amount of technical terms specific to the world of sensors and image processing. In addition, plant traits and indices extracted from images are often different from the information provided by traditional phenotyping campaigns, which increases the gap between the two approaches. Currently, crop science stakeholders still rely massively on the sole grain yield traits, a few human visual scores and laborious destructive sampling methods (Reynolds et al., 2020).

High-throughput phenotyping has tremendous potential but it must be channelled in user-friendly applications. Several levels of application should be identified, from simple handed sensors to multi-cameras systems mounted on UAVs or ground platforms and vehicles (Reynolds et al., 2020). The system described in this thesis fits in the last category: it offers numerous possibilities for a precise characterisation of the crop but it is also the most difficult to share to the end users. To ease this transition, multidisciplinary teams are required, or turnkey high-throughput phenotyping solutions with a user-friendly display and control, and vocabulary that is relevant to crop science stakeholders.

Close-range high-throughput phenotyping systems should be developed with the downstream use of the measurement in mind. At some point, the measures must feed functional structural plant models (Saint Cast et al., 2022), crop models, decision support systems, evaluation or breeding programs (van Eeuwijk et al., 2019; Deery et al., 2021) or plant research in general (Furbank et al., 2011). The development of phenotyping methods should be driven by the end-users rather than by the available technology (Deery 2021).

Moreover, to generate agronomic knowledge from phenotyping data, it should be accounted for the diversity of intra-site and inter-site factors, as explained in the Section 1 of Chapter I. A single team cannot explore all those factors, so international data sharing will become necessary, especially to benefit from multi-site field trials in contrasted pedo-climatic regions (Morisse et al., 2022). It implies standards for the data (Krajewski et al., 2015). Plant traits should be provided with environmental and

crop management data allowing to replicate the experiment and exploit the data in larger research. At this level, the Minimum Information About a Plant Phenotyping Experiment (MIAPPE) framework has been developed (Papoutsoglou et al., 2020). Data should be Findable, Accessible, Interoperable and Reusable (FAIR) (Wilkinson et al., 2016) and the plant traits should be standardised. Solving those challenges is crucial for the sustainability of the phenotyping efforts. This is the development of workable standards, and not the technology itself, that will eventually lead to a wide adoption of high-throughput phenotyping (Casto et al., 2021).

4. Broad perspectives

The multi-sensor techniques developed in this study have the potential to serve many purposes other than wheat phenotyping. Most of the methods can be extended to characterise other crops or even to studies involving plant covers in non-agronomic contexts. For example, our image registration approach was already used to align multispectral nadir images of young pea plants in the field or to align side-view images of a tropical forest canopy, which was needed because the phenology camera was many times moved by the wind, human operators or wild animals. The developed imaging methods can also serve to pilot robotic agricultural vehicles and for the real time precision management of agricultural operations (sowing, weeding, spraying, harvesting, …). From the other side, the wide-spread use of autonomous field robots for those operations will provide many new opportunities to bring the sensors to the plants and gather crop characterisation data from embedded cameras.

The measurement techniques themselves could be revolutionised, whether at the level of image acquisition or image analysis. Current state-of-the art technologies that could enrich the crop characterisation with our system would be the hyperspectral images to benefit from the full reflected spectrum or the LiDAR to improve the 3D characterisation of the crop and, for example, account for the angle of individual leaves while extracting reflectance information. The rapid technological progress could lower the cost of those technologies and make it easier to integrate them in fast image acquisition systems, especially by removing the need for a scanning time and decreasing their sensitivity to wind and sunlight. But maybe we should look further than the existing sensors and platforms, and imagine crazy new phenotyping technologies. What about a ground vehicle that could carry not only sensors above the canopy but also within the canopy? It could exploit a robotic system to spread the plants and make visible the lowest canopy floors.

As a final word, while many technical and technological improvements can be brought to the multi-sensor system we developed, the most important and needed development stays to give the means to this local advanced characterisation of the crop to contribute to crop breeding, smart farming, and to the generation of agronomic knowledge at a global scale.

X

References

Ajayi S., Reddy S.K., Gowda P.H., Xue Q., Rudd J.C., Pradhan G., Liu S., Stewart B.A., Biradar C. & Jessup K.E., 2016. Spectral Reflectance Models for Characterizing Winter Wheat Genotypes. *J. Crop Improv.* **30**(2), 176–195, DOI:10.1080/15427528.2016.1138421.

Alcantarilla P.F., Nuevo J. & Bartoli A., 2013. Fast explicit diffusion for accelerated features in nonlinear scale spaces. *In*: *BMVC 2013 - Electronic Proceedings of the British Machine Vision Conference 2013*. UK: BMVA Press.

Alharbi N., Zhou J. & Wang W., 2018. Automatic Counting of Wheat Spikes from Wheat Growth Images. *In*: *Proceedings of the 7th International Conference on Pattern Recognition Applications and Methods*. SCITEPRESS - Science and Technology Publications, 346–355.

Anderegg J., Hund A., Karisto P. & Mikaberidze A., 2019. In-Field Detection and Quantification of Septoria Tritici Blotch in Diverse Wheat Germplasm Using Spectral–Temporal Features. *Front. Plant Sci.* **10**(October), 1–19, DOI:10.3389/fpls.2019.01355.

Andersen H.J., Reng L. & Kirk K., 2005. Geometric plant properties by relaxed stereo vision using simulated annealing. *Comput. Electron. Agric.* **49**, 219–232, DOI:10.1016/j.compag.2005.02.015.

Andrade-Sanchez P., Gore M.A., Heun J.T., Thorp K.R., Carmo-Silva A.E., French A.N., Salvucci M.E. & White J.W., 2014. Development and evaluation of a field-based high-throughput phenotyping platform. *Funct. Plant Biol.* **41**(1), 68, DOI:10.1071/FP13126.

Apolo-Apolo O.E., Martínez-Guanter J., Pérez-Ruiz M. & Egea G., 2020. Design and assessment of new artificial reference surfaces for real time monitoring of crop water stress index in maize. *Agric. Water Manag.* **240**(June), 106304, DOI:10.1016/j.agwat.2020.106304.

Araus J.L. & Kefauver S.C., 2018. Breeding to adapt agriculture to climate change: affordable phenotyping solutions. *Curr. Opin. Plant Biol.* **45**(Figure 2), 237–247, DOI:10.1016/j.pbi.2018.05.003.

Araus J.L., Kefauver S.C., Vergara-Díaz O., Gracia-Romero A., Rezzouk F.Z., Segarra J., Buchaillot M.L., Chang-Espino M., Vatter T., Sanchez-Bragado R., Fernandez-Gallego J.A., Serret M.D. & Bort J., 2022. Crop phenotyping in a context of global change: What to measure and how to do it. *J. Integr. Plant Biol.* **64**(2), 592–618, DOI:10.1111/jipb.13191.

Asefpour K. & Massah J., 2012. Design, development and performance evaluation of a robot to early detection of nitrogen deficiency in greenhouse cucumber (cucumis sativus) with machine vision. *Int. J. Agric. Res. Rev.* **2**(4), 448–454.

Ashourloo D., Mobasheri M.R. & Huete A., 2014. Developing two spectral disease

indices for detection of wheat leaf rust (Pucciniatriticina). *Remote Sens.* **6**(6), 4723–4740, DOI:10.3390/rs6064723.

Ayalew T.W., Ubbens J.R. & Stavness I., 2020. Unsupervised Domain Adaptation for Plant Organ Counting. *In*: *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 330–346.

Bai G., Ge Y., Hussain W., Baenziger P.S. & Graef G., 2016. A multi-sensor system for high throughput field phenotyping in soybean and wheat breeding. *Comput. Electron. Agric.* **128**, 181–192, DOI:10.1016/j.compag.2016.08.021.

Bai G., Ge Y., Scoby D., Leavitt B., Stoerger V., Kirchgessner N., Irmak S., Graef G., Schnable J. & Awada T., 2019. NU-Spidercam: A large-scale, cable-driven, integrated sensing and robotic system for advanced phenotyping, remote sensing, and agronomic research. *Comput. Electron. Agric.* **160**, 71–81, DOI:10.1016/j.compag.2019.03.009.

Banerjee K., Krishnan P. & Mridha N., 2018. Application of thermal imaging of wheat crop canopy to estimate leaf area index under different moisture stress conditions. *Biosyst. Eng.* **166**, 13–27, DOI:10.1016/j.biosystemseng.2017.10.012.

Baresel J.P., Rischbeck P., Hu Y., Kipp S., Hu Y., Barmeier G. & Mistele B., 2017. Use of a digital camera as alternative method for non-destructive detection of the leaf chlorophyll content and the nitrogen nutrition status in wheat. *Comput. Electron. Agric.* **140**, 25–33, DOI:10.1016/j.compag.2017.05.032.

Baret F., de Solan B., Lopez-Lozano R., Ma K. & Weiss M., 2010. GAI estimates of row crops from downward looking digital photos taken perpendicular to rows at 57.5° zenith angle: Theoretical considerations based on 3D architecture models and application to wheat crops. *Agric. For. Meteorol.* **150**(11), 1393–1401, DOI:10.1016/j.agrformet.2010.04.011.

Baret F., Jacquemoud S., Guyot G. & Leprieur C., 1992. Modeled analysis of the biophysical nature of spectral shifts and comparison with information content of broad bands. *Remote Sens. Environ.* **41**(2–3), 133–142, DOI:10.1016/0034-4257(92)90073-S.

Barman D., Sehgal V.K., Sahoo R.N. & Nagarajan S., 2010. Relationship of bidirectional reflectance of wheat with biophysical parameters and its radiative transfer modeling using Prosail. *J. Indian Soc. Remote Sens.* **38**(1), 35–44, DOI:10.1007/s12524-010-0010-1.

Barmeier G. & Schmidhalter U., 2016. High-throughput phenotyping of wheat and barley plants grown in single or few rows in small plots using active and passive spectral proximal sensing. *Sensors (Switzerland)* **16**(11), 1–14, DOI:10.3390/s16111860.

Barnes E.M., Clarke T.R., Richards S.E., Colaizzi P.D., Haberland J., Kostrzewski

M., Waller P., Choi C. R.E., Thompson T., Lascano R.J., Li H. & Moran M.S., 2000. Coincident detection of crop water stress, nitrogen status and canopy density using ground based multispectral data. *In*: *Proceedings of the 5th International Conference on Precision Agriculture*. Madison, WI, USA: American Society of Agronomy.

Basso B., Cammarano D. & Carfagna E., 2013. Review of Crop Yield Forecasting Methods and Early Warning Systems. *First Meet. Sci. Advis. Comm. Glob. Strateg. to Improv. Agric. Rural Stat.* 1–56, DOI:10.1017/CBO9781107415324.004.

Bay H., Ess A., Tuytelaars T. & Gool L. Van, 2008. Speeded-Up Robust Features (SURF). *In*: *Computer Vision and Image Understanding*. Springer, Berlin/Heidelberg, Germany, 346–359.

Beauchêne K., Leroy F., Fournier A., Huet C., Bonnefoy M., Lorgeou J., de Solan B., Piquemal B., Thomas S. & Cohan J.-P., 2019. Management and Characterization of Abiotic Stress via PhénoField®, a High-Throughput Field Phenotyping Platform. *Front. Plant Sci.* **10**(July), 1–17, DOI:10.3389/fpls.2019.00904.

Bebronne R., Carlier A., Meurs R., Leemans V., Vermeulen P., Dumont B. & Mercatoris B., 2020. In-field proximal sensing of septoria tritici blotch, stripe rust and brown rust in winter wheat by means of reflectance and textural features from multispectral imagery. *Biosyst. Eng.* **197**, 257–269, DOI:10.1016/j.biosystemseng.2020.06.011.

Behmann J., Acebron K., Emin D., Bennertz S., Matsubara S., Thomas S., Bohnenkamp D., Kuska M.T., Jussila J., Salo H., Mahlein A.K. & Rascher U., 2018. Specim IQ: Evaluation of a new, miniaturized handheld hyperspectral camera and its application for plant phenotyping and disease detection. *Sensors (Switzerland)* **18**(2), DOI:10.3390/s18020441.

Benelli A., Cevoli C. & Fabbri A., 2020. In-field hyperspectral imaging: An overview on the ground-based applications in agriculture. *J. Agric. Eng.* **51**(3), 129–139, DOI:10.4081/jae.2020.1030.

Berenstein R., Hočevar M., Godeša T., Edan Y. & Ben-Shahar O., 2015. Distance-dependent multimodal image registration for agriculture tasks. *Sensors (Switzerland)* **15**(8), 20845–20862, DOI:10.3390/s150820845.

Berger K., Atzberger C., Danner M., D'Urso G., Mauser W., Vuolo F. & Hank T., 2018. Evaluation of the PROSAIL Model Capabilities for Future Hyperspectral Model Environments: A Review Study. *Remote Sens.* **10**(2), 85, DOI:10.3390/rs10010085.

Berkman P.J., Lai K., Lorenc M.T. & Edwards D., 2012. Next-generation sequencing applications for wheat crop improvement. *Am. J. Bot.* **99**(2), 365–371, DOI:10.3732/ajb.1100309.

Birodkar V., Lu Z., Li S., Rathod V. & Huang J., 2022. The surprising impact of mask-

head architecture on novel class segmentation. *arXiv* **abs/2104.0**, 6995–7005, DOI:10.1109/iccv48922.2021.00693.

Biskup B., Scharr H., Schurr U. & Rascher U., 2007. A stereo imaging system for measuring structural parameters of plant canopies. *Plant, Cell Environ.* **30**(10), 1299–1308, DOI:10.1111/j.1365-3040.2007.01702.x.

Bodson B., 2019. *Livre Blanc des Céréales - Edition de février 2019*, 364.

Bradski G. & Kaehler A., 2008. *Learning OpenCV*, Newton, MA, USA: O'Reilly Media, Inc, 571.

Bravo C., Moshou D., West J., McCartney A. & Ramon H., 2003. Early disease detection in wheat fields using spectral reflectance. *Biosyst. Eng.* **84**(2), 137–145, DOI:10.1016/S1537-5110(02)00269-6.

Busemeyer L., Mentrup D., Möller K., Wunder E., Alheit K., Hahn V., Maurer H.P., Reif J.C., Würschum T., Müller J., Rahe F. & Ruckelshausen A., 2013. Breedvision - A multi-sensor platform for non-destructive field-based phenotyping in plant breeding. *Sensors (Switzerland)* **13**(3), 2830–2847, DOI:10.3390/s130302830.

Cai J., Kumar P., Chopin J. & Miklavcic S.J., 2018. Land-based crop phenotyping by image analysis: Accurate estimation of canopy height distributions using stereo images. *PLoS One* **13**(5), 1–21, DOI:10.1371/journal.pone.0196671.

Camacho-de-Coca F., Gilabert M.A. & Meliá J., 2001. Bidirectional reflectance factor analysis from field radiometry and HyMap data. *In*: *The Digital Airborne Spectrometer Experiment (DAISEX)*. Paris, France: European Space Agency, 163–175.

Cammarano D., Fitzgerald G., Basso B., Leary G.O., Chen D., Grace P. & Fiorentino C., 2011. Use of the Canopy Chlorophyl Content Index (CCCI) for Remote Estimation of Wheat Nitrogen Content in Rainfed Environments (Ccci), 1597–1603, DOI:10.2134/agronj2011.0124.

Cammarano D., Fitzgerald G.J., Casa R. & Basso B., 2014. Assessing the Robustness of Vegetation Indices to Estimate Wheat N in Mediterranean Environments. *Remote Sens.* **6**(4), 2827–2844, DOI:10.3390/rs6042827.

Cao L., Zhang X., Pu J., Xu S., Cai X. & Li Z., 2020. The Field Wheat Count Based on the Efficientdet Algorithm. *In*: *2020 IEEE 3rd International Conference on Information Systems and Computer Aided Education (ICISCAE)*. IEEE, 557–561.

Carlier A., Dandrifosse S., Dumont B. & Mercatoris B., 2022. Wheat Ear Segmentation Based on a Multisensor System and Superpixel Classification. *Plant Phenomics* **2022**, 1–10, DOI:10.34133/2022/9841985.

Casto A.L., Schuhl H., Tovar J.C., Wang Q., Bart R.S., Fahlgren N. & Gehan M.A., 2021. Picturing the future of food. *Plant Phenome J.* **4**(1), 1–17,

DOI:10.1002/ppj2.20014.

Castro W., Junior J.M., Polidoro C., Osco L.P., Gonçalves W., Rodrigues L., Santos M., Jank L., Barrios S., Valle C., Simeão R., Carromeu C., Silveira E., Jorge L.A. de C. & Matsubara E., 2020. Deep learning applied to phenotyping of biomass in forages with uav-based rgb imagery. *Sensors (Switzerland)* **20**(17), 1–18, DOI:10.3390/s20174802.

Chakraborty A., Remote N. & Centre S., 2005. Spectral reflectance anisotropy of wheat canopy in assessing biophysical parameters. *J. Agric. Phys.* **5**(1), 1–10.

Chakraborty D., Sehgal V.K., Sahoo R.N., Pradhan S. & Gupta V.K., 2015. Study of the anisotropic reflectance behaviour of wheat canopy to evaluate the performance of radiative transfer model PROSAIL5B. *J. Indian Soc. Remote Sens.* **43**(2), 297–310, DOI:10.1007/s12524-014-0411-7.

Cheshkova A.F., 2022. A review of hyperspectral image analysis techniques for plant disease detection and identification. *Vavilovskii Zhurnal Genet Sel.* **26**(2), 202–213, DOI:10.18699/VJGB-22-25.

Chopinet B. & Cointault F., 2006. Colour-texture image analysis for in-field wheat head counting. *In*: *Second International Symposium on Communications, Control, and Signal Processing (ISCCSP 2006)*. 13--15.

Clawson K.L. & Blad B.L., 1982. Infrared Thermometry for Scheduling Irrigation of Corn 1. *Agron. J.* **74**(2), 311–316, DOI:10.2134/agronj1982.00021962007400020013x.

Cointault F., Guerin D., Guillemin J.P. & Chopinet B., 2008. In-field triticum aestivum ear counting using colour-texture image analysis. *New Zeal. J. Crop Hortic. Sci.* **36**(2), 117–130, DOI:10.1080/01140670809510227.

Cointault F., Journaux L., Miteran O. & Destain M.F., 2008. Improvements of image processing for wheat ear counting. *In*: *Agricultural and Biosystems Engineering for a Sustainable World. International Conference on Agricultural Engineering*.

Cointault F., Journaux L., Rabatel G., Germain C., Ooms D., Destain M.-F., Gorretta N., Grenier G., Lavialle O. & Marin A., 2012. Texture, Color and Frequential Proxy-Detection Image Processing for Crop Characterization in a Context of Precision Agriculture. *In*: Aflakpui, G. ed. *Agricultural Science*. London: INTech, 49–70.

Comar A., Baret F., Viénot F., Yan L. & de Solan B., 2012. Wheat leaf bidirectional reflectance measurements: Description and quantification of the volume, specular and hot-spot scattering features. *Remote Sens. Environ.* **121**, 26–35, DOI:10.1016/j.rse.2011.01.028.

Comar A., Burger P., De Solan B., Baret F., Daumard F. & Hanocq J.F., 2012. A semi-automatic system for high throughput phenotyping wheat cultivars in-field conditions: Description and first results. *Funct. Plant Biol.* **39**(11), 914–924, DOI:10.1071/FP12065.

Costa J.M., Grant O.M. & Chaves M.M., 2013. Thermography to explore plant-environment interactions. *J. Exp. Bot.* **64**(13), 3937–3949, DOI:10.1093/jxb/ert029.

Crain J.L., Wei Y., Barker J., Thompson S.M., Alderman P.D., Reynolds M., Zhang N. & Poland J., 2016. Development and deployment of a portable field phenotyping platform. *Crop Sci.* **56**(3), 965–975, DOI:10.2135/cropsci2015.05.0290.

D.X. He, Y. Matsuura, T. Kozai & K.C. Ting, 2003. A binocular stereovision system for transplant growth variables analysis. *Appl. Eng. Agric.* **19**(5), 183–194, DOI:10.13031/2013.15308.

Dandrifosse S., Bouvry A., Leemans V., Dumont B. & Mercatoris B., 2020. Imaging wheat canopy through stereo vision : overcoming the challenges of the laboratory to field transition for morphological features extraction. *Front. Plant Sci.* **11**(February), 1–15, DOI:10.3389/fpls.2020.00096.

Dandrifosse S., Carlier A., Dumont B. & Mercatoris B., 2021. Registration and Fusion of Close-Range Multimodal Wheat Images in Field Conditions. *Remote Sens.* **13**(7), 1380, DOI:10.3390/rs13071380.

Dandrifosse S., Carlier A., Dumont B. & Mercatoris B., 2022. In-Field Wheat Reflectance: How to Reach the Organ Scale? *Sensors* **22**(9), 3342, DOI:10.3390/s22093342.

Dandrifosse S., Ennadifi E., Carlier A., Gosselin B., Dumont B. & Mercatoris B., 2022. Deep learning for wheat ear segmentation and ear density measurement : From heading to maturity. *Comput. Electron. Agric.* **199**(June), DOI:10.1016/j.compag.2022.107161.

Danner M., Berger K., Wocher M., Mauser W. & Hank T., 2019. Fitted PROSAIL parameterization of leaf inclinations,water content and brown pigment content for winter wheat and maize canopies. *Remote Sens.* **11**(10), DOI:10.3390/rs11101150.

David E., Madec S., Sadeghi-Tehran P., Aasen H., Zheng B., Liu S., Kirchgessner N., Ishikawa G., Nagasawa K., Badhon M.A., Pozniak C., de Solan B., Hund A., Chapman S.C., Baret F., Stavness I. & Guo W., 2020. Global Wheat Head Detection (GWHD) Dataset: A Large and Diverse Dataset of High-Resolution RGB-Labelled Images to Develop and Benchmark Wheat Head Detection Methods. *Plant Phenomics* **2020**, 1–12, DOI:10.34133/2020/3521852.

David E., Serouart M., Smith D., Madec S., Velumani K., Liu S., Wang X., Pinto F., Shafiee S., Tahir I.S.A., Tsujimoto H., Nasuda S., Zheng B., Kirchgessner N., Aasen H., Hund A., Sadhegi-Tehran P., Nagasawa K., Ishikawa G., Dandrifosse S., Carlier A., Dumont B., Mercatoris B., Evers B., Kuroki K., Wang H., Ishii M., Badhon M.A., Pozniak C., LeBauer D.S., Lillemo M., Poland J., Chapman S., de Solan B., Baret F., Stavness I. & Guo W., 2021. Global Wheat Head

Detection 2021: An Improved Dataset for Benchmarking Wheat Head Detection Methods. *Plant Phenomics* **2021**(May), 1–9, DOI:10.34133/2021/9846158.

de Oliveira G.S., Junior J.M., Polidoro C., Osco L.P., Siqueira H., Rodrigues L., Jank L., Barrios S., Valle C., Simeão R., Carromeu C., Silveira E., Jorge L.A. de C., Gonçalves W., Santos M. & Matsubara E., 2021. Convolutional neural networks to estimate dry matter yield in a guineagrass breeding program using uav remote sensing. *Sensors* **21**(12), 1–19, DOI:10.3390/s21123971.

De Vylder J., Douterloigne K., Vandenbussche F., Van Der Straeten D. & Philips W., 2012. A non-rigid registration method for multispectral imaging of plants. *Sens. Agric. Food Qual. Saf. IV* **8369**(March 2014), 836907-836907–8, DOI:10.1117/12.918752.

Deery D., Jimenez-Berni J., Jones H., Sirault X. & Furbank R., 2014. Proximal Remote Sensing Buggies and Potential Applications for Field-Based Phenotyping. *Agronomy* **5**, 349–379, DOI:10.3390/agronomy4030349.

Deery D.M. & Jones H.G., 2021. Field Phenomics: Will It Enable Crop Improvement? *Plant Phenomics* **2021**, 1–16, DOI:10.34133/2021/9871989.

Dehkordi R.H., Jarroudi M. El, Kouadio L. & Meersmans J., 2020. Monitoring Wheat Leaf Rust and Stripe Rust in Winter Wheat Using High-Resolution UAV-Based Red-Green-Blue Imagery 1–21, DOI:10.3390/rs12223696.

Devadas R., Simpfendorfer S., Backhouse D. & Lamb D.W., 2014. Effect of stripe rust on the yield response of wheat to nitrogen. *Crop J.* **2**(4), 201–206, DOI:10.1016/j.cj.2014.05.002.

Dietterich T.G. & Bakiri G., 1994. Solving Multiclass Learning Problems via Error-Correcting Output Codes. *J. Artif. Intell. Res.* **2**, 263–286, DOI:10.1.1.72.7289.

Duchene O., Dumont B., Cattani D.J., Fagnant L., Schlautman B., DeHaan L.R., Barriball S., Jungers J.M., Picasso V.D., David C. & Celette F., 2021. Process-based analysis of Thinopyrum intermedium phenological development highlights the importance of dual induction for reproductive growth and agronomic performance. *Agric. For. Meteorol.* **301**–**302**(July 2020), 108341, DOI:10.1016/j.agrformet.2021.108341.

Dworak V., Selbeck J., Dammer K.H., Hoffmann M., Zarezadeh A.A. & Bobda C., 2013. Strategy for the development of a smart NDVI camera system for outdoor plant detection and agricultural embedded systems. *Sensors (Switzerland)* **13**(2), 1523–1538, DOI:10.3390/s130201523.

Elsayed S., Barmeier G. & Schmidhalter U., 2018. Passive reflectance sensing and digital image analysis allows for assessing the biomass and nitrogen status of wheat in early and late tillering stages. *Front. Plant Sci.* **9**(October), 1–15, DOI:10.3389/fpls.2018.01478.

Erdle K., Mistele B. & Schmidhalter U., 2011. Comparison of active and passive spectral sensors in discriminating biomass parameters and nitrogen status in

wheat cultivars. *F. Crop. Res.* **124**(1), 74–84, DOI:10.1016/j.fcr.2011.06.007.

Evangelidis G.D. & Psarakis E.Z., 2008. Parametric image alignment using enhanced correlation coefficient maximization. *IEEE Trans. Pattern Anal. Mach. Intell.* **30**(10), 1858–1865, DOI:10.1109/TPAMI.2008.113.

Fernandez-Gallego J.A., Kefauver S.C., Gutiérrez N.A., Nieto-Taladriz M.T. & Araus J.L., 2018. Wheat ear counting in-field conditions: High throughput and low-cost approach using RGB images. *Plant Methods* **14**(1), 1–12, DOI:10.1186/s13007-018-0289-4.

Fernandez-Gallego J.A., Kefauver S.C., Vatter T., Aparicio Gutiérrez N., Nieto-Taladriz M.T. & Araus J.L., 2019. Low-cost assessment of grain yield in durum wheat using RGB images. *Eur. J. Agron.* **105**(March), 146–156, DOI:10.1016/j.eja.2019.02.007.

Fernandez-Gallego J.A., Lootens P., Borra-Serrano I., Derycke V., Haesaert G., Roldán-Ruiz I., Araus J.L. & Kefauver S.C., 2020. Automatic wheat ear counting using machine learning based on RGB UAV imagery. *Plant J.* **103**(4), 1603–1613, DOI:10.1111/tpj.14799.

Fernández E., Gorchs G. & Serrano L., 2019. Use of consumer-grade cameras to assess wheat N status and grain yield. *PLoS One* **14**(2), 1–18, DOI:10.1371/journal.pone.0211889.

Fitzgerald G.J., Rodriguez D., Christensen L.K., Belford R., Sadras V.O. & Clarke T.R., 2006. Spectral and thermal sensing for nitrogen and water status in rainfed and irrigated wheat environments. *Precis. Agric.* **7**(4), 233–248, DOI:10.1007/s11119-006-9011-z.

Fourati F., Mseddi W.S. & Attia R., 2021. Wheat Head Detection using Deep, Semi-Supervised and Ensemble Learning. *Can. J. Remote Sens.* **0**(0), 1–13, DOI:10.1080/07038992.2021.1906213.

Fu P., Meacham-Hensold K., Guan K., Wu J. & Bernacchi C., 2020. Estimating photosynthetic traits from reflectance spectra: A synthesis of spectral indices, numerical inversion, and partial least square regression. *Plant Cell Environ.* **43**(5), 1241–1258, DOI:10.1111/pce.13718.

Fuchs M., 1990. Infrared measurement of canopy temperature and detection of plant water stress. *Theor. Appl. Climatol.* **42**(4), 253–261, DOI:10.1007/BF00865986.

Furbank R.T. & Tester M., 2011. Phenomics - technologies to relieve the phenotyping bottleneck. *Trends Plant Sci.* **16**(12), 635–644, DOI:10.1016/j.tplants.2011.09.005.

Gao T., Emadi H., Saha H., Zhang J., Lofquist A., Singh A., Ganapathysubramanian B., Sarkar S., Singh A. & Bhattacharya S., 2018. A Novel Multirobot System for Plant Phenotyping. *Robotics* **7**(4), 61, DOI:10.3390/robotics7040061.

Genaev M.A., Skolotneva E.S., Gultyaeva E.I., Orlova E.A., Bechtold N.P. &

Afonnikov D.A., 2021. Image-based wheat fungi diseases identification by deep learning. *Plants* **10**(8), DOI:10.3390/plants10081500.

Genser N., Seiler J. & Kaup A., 2020. Camera Array for Multi-Spectral Imaging. *IEEE Trans. Image Process.* **29**, 9234–9249, DOI:10.1109/tip.2020.3024738.

Germain C., 1995. Non destructive counting of wheatear with picture analysis. *In*: *Fifth International Conference on Image Processing and Its Applications*. Stevenage, UK: IEE, 435–439.

Gitelson A.A., Vina A., Arkebauer T.J., Rundquist D.C., Keydan G. & Leavitt B., 2003. Remote estimation of leaf area index and green leaf biomass in maize canopies. *Geophys. Res. Lett.* **30**(5), 4–7, DOI:10.1029/2002gl016450.

Goel N.S., 1988. Models of vegetation canopy reflectance and their use in estimation of biophysical parameters from reflectance data. *Remote Sens. Rev.* **4**(1), 1–212, DOI:10.1080/02757258809532105.

Golzarian M.R., Lee M.K. & Desbiolles J.M.A., 2012. Evaluation of color indices for improved segmentation of plant images. *Trans. ASABE* **55**(1), 261–273.

Gong B., Ergu D., Cai Y. & Ma B., 2020. Real-Time Detection for Wheat Head Applying Deep Neural Network. *Sensors* **21**(1), 191, DOI:10.3390/s21010191.

Gracia-Romero A., Kefauver S.C., Fernandez-Gallego J.A., Vergara-Díaz O., Nieto-Taladriz M.T. & Araus J.L., 2019. UAV and ground image-based phenotyping: A proof of concept with durum wheat. *Remote Sens.* **11**(10), DOI:10.3390/rs11101244.

Grant O.M., Tronina Ł., Jones H.G. & Chaves M.M., 2007. Exploring thermal imaging variables for the detection of stress responses in grapevine under different irrigation regimes. *J. Exp. Bot.* **58**(4), 815–825, DOI:10.1093/jxb/erl153.

Grbović Z., Panić M., Marko O., Brdar S. & Crnojević V., 2019. Wheat Ear Detection in RGB and Thermal Images Using Deep Neural Networks. *In*: *International Conference on Machine Learning and Data Mining, MLDM 2019*. Fockendorf, Germany: ibai.

Grieder C., Hund A. & Walter A., 2015. Image based phenotyping during winter: A powerful tool to assess wheat genetic variation in growth response to temperature. *Funct. Plant Biol.* **42**(4), 387–396, DOI:10.1071/FP14226.

Guo Y., Senthilnath J., Wu W., Zhang X., Zeng Z. & Huang H., 2019. Radiometric calibration for multispectral camera of different imaging conditions mounted on a UAV platform. *Sustain.* **11**(4), DOI:10.3390/su11040978.

Gutiérrez S., Fernández-Novales J., Diago M.P. & Tardaguila J., 2018. On-the-go hyperspectral imaging under field conditions and machine learning for the classification of grapevine varieties. *Front. Plant Sci.* **9**(July), 1–11, DOI:10.3389/fpls.2018.01102.

Haghighattalab A., González Pérez L., Mondal S., Singh D., Schinstock D., Rutkoski J., Ortiz-Monasterio I., Singh R.P., Goodin D. & Poland J., 2016. Application of unmanned aerial systems for high throughput phenotyping of large wheat breeding nurseries. *Plant Methods* **12**(1), 1–15, DOI:10.1186/s13007-016-0134-6.

Hakala T., Markelin L., Honkavaara E., Scott B., Theocharous T., Nevalainen O., Näsi R., Suomalainen J., Viljanen N., Greenwell C. & Fox N., 2018. Direct reflectance measurements from drones: Sensor absolute radiometric calibration and system tests for forest reflectance characterization. *Sensors (Switzerland)* **18**(5), DOI:10.3390/s18051417.

Hamuda E., Glavin M. & Jones E., 2016. A survey of image processing techniques for plant extraction and segmentation in the field. *Comput. Electron. Agric.* **125**, 184–199, DOI:10.1016/j.compag.2016.04.024.

Hamuda E., Mc Ginley B., Glavin M. & Jones E., 2017. Automatic crop detection under field conditions using the HSV colour space and morphological operations. *Comput. Electron. Agric.* **133**, 97–107, DOI:10.1016/j.compag.2016.11.021.

Hasan M.M., Chopin J.P., Laga H. & Miklavcic S.J., 2018. Detection and analysis of wheat spikes using Convolutional Neural Networks. *Plant Methods* **14**(1), 1–13, DOI:10.1186/s13007-018-0366-8.

Hashimoto Y., Ino T., Kramer P.J., Naylor A.W. & Strain B.R., 1984. Dynamic analysis of water stress of sunflower leaves by means of a thermal image processing system. *Plant Physiol.* **76**(1), 266–269, DOI:10.1104/pp.76.1.266.

He D.X., Matsuura Y., Kozai T. & Ting K.C., 2003. A binocular stereovision system for transplant growth variables analysis. *Appl. Eng. Agric.* **19**(5), 611–617.

Henke M., Junker A., Neumann K., Altmann T. & Gladilin E., 2019a. Comparison of feature point detectors for multimodal image registration in plant phenotyping. *PLoS One* **14**(9), 1–16, DOI:10.1371/journal.pone.0221203.

Henke M., Junker A., Neumann K., Altmann T. & Gladilin E., 2019b. Comparison and extension of three methods for automated registration of multimodal plant images. *Plant Methods* **15**(1), 1–15, DOI:10.1186/s13007-019-0426-8.

Hirschmüller H., 2007. Stereo Processing by Semi-Global Matching and Mutual Information. *In*: *IEEE Transactions on Pattern Analysis and Machine Intelligence*. Manhattan, NY, USA: IEEE, 328–341.

Houle D., Govindaraju D.R. & Omholt S., 2010. Phenomics: The next challenge. *Nat. Rev. Genet.* **11**(12), 855–866, DOI:10.1038/nrg2897.

Idso S.B., Jackson R.D., Pinter P.J., Reginato R.J. & Hatfield J.L., 1981. Normalizing the stress-degree-day parameter for environmental variability. *Agric. Meteorol.* **24**(C), 45–55, DOI:10.1016/0002-1571(81)90032-7.

Idso S.B., Jackson R.D. & Reginato R.J., 1977. Remote-Sensing of Crop Yields. *Science (80-.).* **196**(4285), 19–25.

Jackson R.D., Idso S.B., Reginato R.J. & Pinter P.J., 1981. Canopy temperature as a crop water stress indicator. *Water Resour. Res.* **17**(4), 1133–1138, DOI:10.1029/WR017i004p01133.

Jacquemoud S. & Baret F., 1990. PROSPECT: A model of leaf optical properties spectra. *Remote Sens. Environ.* **34**(2), 75–91, DOI:10.1016/0034-4257(90)90100-Z.

Jacquemoud S., Verhoef W., Baret F., Bacour C., Zarco-Tejada P.J., Asner G.P., François C. & Ustin S.L., 2009. PROSPECT + SAIL models: A review of use for vegetation characterization. *Remote Sens. Environ.* **113**(SUPPL. 1), S56–S66, DOI:10.1016/j.rse.2008.01.026.

Jay S., Baret F., Dutartre D., Malatesta G., Héno S., Comar A., Weiss M. & Maupas F., 2018. Exploiting the centimeter resolution of UAV multispectral imagery to improve remote-sensing estimates of canopy structure and biochemistry in sugar beet crops. *Remote Sens. Environ.* (April), 110898, DOI:10.1016/j.rse.2018.09.011.

Jay S., Comar A., Benicio R., Beauvois J., Dutartre D., Daubige G., Li W., Labrosse J., Thomas S., Henry N., Weiss M. & Baret F., 2020. Scoring Cercospora Leaf Spot on Sugar Beet: Comparison of UGV and UAV Phenotyping Systems. *Plant Phenomics* **2020**, 1–18, DOI:10.34133/2020/9452123.

Jay S., Gorretta N., Morel J., Maupas F., Bendoula R., Rabatel G., Dutartre D., Comar A. & Baret F., 2017. Estimating leaf chlorophyll content in sugar beet canopies using millimeter- to centimeter-scale reflectance imagery. *Remote Sens. Environ.* **198**, 173–186, DOI:10.1016/j.rse.2017.06.008.

Jay S., Rabatel G., Gorretta N., Jay S., Rabatel G. & In-field N.G., 2014. In-field crop row stereo-reconstruction for plant phenotyping. *In*: *Second International Conference on Robotics and Associated High-Technologies and Equipment for Agriculture and Forestry (RHEA-2014)*. Madrid, ES, 10.

Jerbi T., Wuyts N., Cane M.A., Faux P.-F. & Draye X., 2015. High resolution imaging of maize (Zea maize) leaf temperature in the field: the key role of the regions of interest. *Funct. Plant Biol.* **42**(9), 858, DOI:10.1071/FP15024.

Jia L., Chen X., Li M., Cui Z. & Zhang F., 2009. Comparsion of multispectral reflectance with digital color image in assessing the winter wheat nitrogen status. *IFIP Int. Fed. Inf. Process.* **294**, 1161–1170, DOI:10.1007/978-1-4419-0211-5-43.

Jia L., Chen X., Zhang F., Buerkert A. & Römheld V., 2004. Use of digital camera to assess nitrogen status of winter wheat in the Northern China Plain. *J. Plant Nutr.* **27**(3), 441–450, DOI:10.1081/PLN-120028872.

Jiang Y., Li C., Robertson J.S., Sun S., Xu R. & Paterson A.H., 2018. GPhenoVision:

A ground mobile system with multi-modal imaging for field-based high throughput phenotyping of cotton. *Sci. Rep.* **8**(1), 1–15, DOI:10.1038/s41598-018-19142-2.

Jiménez-Bello M.A., Ballester C., Castel J.R. & Intrigliolo D.S., 2011. Development and validation of an automatic thermal imaging process for assessing plant water status. *Agric. Water Manag.* **98**(10), 1497–1504, DOI:10.1016/j.agwat.2011.05.002.

Jimenez-Berni J.A., Deery D.M., Rozas-Larraondo P., Condon A.T.G., Rebetzke G.J., James R.A., Bovill W.D., Furbank R.T. & Sirault X.R.R., 2018. High throughput determination of plant height, ground cover, and above-ground biomass in wheat with LiDAR. *Front. Plant Sci.* **9**(February), 1–18, DOI:10.3389/fpls.2018.00237.

Jones H.G., 1999. Use of thermography for quantitative studies of spatial and temporal variation of stomatal conductance over leaf surfaces. *Plant, Cell Environ.* **22**(9), 1043–1055, DOI:10.1046/j.1365-3040.1999.00468.x.

Jones H.G., 2002. Use of infrared thermography for monitoring stomatal closure in the field: application to grapevine. *J. Exp. Bot.* **53**(378), 2249–2260, DOI:10.1093/jxb/erf083.

Justes E., Mary B., Meynard J.-M., Machet J.-M. & Thelier-Huche L., 1994. Determination of a Critical Nitrogen Dilution Curve for Winter Wheat Crops. *Ann. Bot.* **74**(4), 397–407, DOI:10.1006/anbo.1994.1133.

Kaczmarek A.L., 2015. Improving depth maps of plants by using a set of five cameras. *J. Electron. Imaging* **24**(2), 023018, DOI:10.1117/1.JEI.24.2.023018.

Kaczmarek A.L., 2017. Stereo vision with Equal Baseline Multiple Camera Set (EBMCS) for obtaining depth maps of plants. *Comput. Electron. Agric.* **135**, 23–37, DOI:10.1016/j.compag.2016.11.022.

Keszei A.P., Berkels B. & Deserno T.M., 2017. Survey of Non-Rigid Registration Tools in Medicine. *J. Digit. Imaging* **30**(1), 102–116, DOI:10.1007/s10278-016-9915-8.

Khanal S., Fulton J. & Shearer S., 2017. An overview of current and potential applications of thermal remote sensing in precision agriculture. *Comput. Electron. Agric.* **139**, 22–32, DOI:10.1016/j.compag.2017.05.001.

Khanna R., Schmid L., Walter A., Nieto J., Siegwart R. & Liebisch F., 2019. A spatio temporal spectral framework for plant stress phenotyping. *Plant Methods* **15**(1), 1–18, DOI:10.1186/s13007-019-0398-8.

Kim J.Y., 2020. Roadmap to High Throughput Phenotyping for Plant Breeding. *J. Biosyst. Eng.* **45**(1), 43–55, DOI:10.1007/s42853-020-00043-0.

Kimes D.S., 1980. Effects of vegetation canopy structure on remotely sensed canopy temperatures. *Remote Sens. Environ.* **10**(3), 165–174, DOI:10.1016/0034-

4257(80)90020-6.

Kipp S., Mistele B., Baresel P. & Schmidhalter U., 2014. High-throughput phenotyping early plant vigour of winter wheat. *Eur. J. Agron.* **52**, 271–278, DOI:10.1016/j.eja.2013.08.009.

Kipp S., Mistele B. & Schmidhalter U., 2014. The performance of active spectral reflectance sensors as influenced by measuring distance, device temperature and light intensity. *Comput. Electron. Agric.* **100**, 24–33, DOI:10.1016/j.compag.2013.10.007.

Kirchgessner N., Liebisch F., Yu K., Pfeifer J., Friedli M., Hund A. & Walter A., 2017. The ETH field phenotyping platform FIP: A cable-suspended multi-sensor system. *Funct. Plant Biol.* **44**(1), 154–168, DOI:10.1071/FP16165.

Kise M. & Zhang Q., 2008. Development of a stereovision sensing system for 3D crop row structure mapping and tractor guidance. *Biosyst. Eng.* **101**(2), 191–198, DOI:10.1016/j.biosystemseng.2008.08.001.

Klein S., Staring M., Murphy K., Viergever M. a. & Pluim J., 2010. Elastix : A Toolbox for Intensity-Based Medical Image Registration. *IEEE Trans. Med. Imaging* **29**(1), 196–205, DOI:10.1109/TMI.2009.2035616.

Krajewski P., Chen D., Ćwiek H., Van Dijk A.D.J., Fiorani F., Kersey P., Klukas C., Lange M., Markiewicz A., Nap J.P., Van Oeveren J., Pommier C., Scholz U., Van Schriek M., Usadel B. & Weise S., 2015. Towards recommendations for metadata and data handling in plant phenotyping. *J. Exp. Bot.* **66**(18), 5417–5427, DOI:10.1093/jxb/erv271.

Lancashire P.D., Bleiholder H., Boom T. Van Den, Langelüddeke P., Stauss R., Weber E. & Witzenberger A., 1991. A uniform decimal code for growth stages of crops and weeds. *Ann. Appl. Biol.* **119**(3), 561–601, DOI:10.1111/j.1744-7348.1991.tb04895.x.

Lawson T. & Blatt M.R., 2014. Stomatal size, speed, and responsiveness impact on photosynthesis and water use efficiency. *Plant Physiol.* **164**(4), 1556–1570, DOI:10.1104/pp.114.237107.

Le Stum H., Bonjean A., Rush O. & Bouquery J.-M., 2017. *Le Blé*. France Agricole ed., France Agricole, 270.

Leemans V., Dumont B. & Destain M.-F., 2013. Assessment of plant leaf area measurement by using stereo-vision. *In*: *2013 International Conference on 3D Imaging (IC3D)*. Manhattan, NY, USA: IEEE, 1–5.

Leemans V., Marlier G., Destain M.-F., Dumont B. & Mercatoris B., 2017. Estimation of leaf nitrogen concentration on winter wheat by multispectral imaging. *In*: *Hyperspectral Imaging Sensors: Innovative Applications and Sensor Standards 2017*. Bellingham, WA, USA: SPIE - International Society for Optics and Photonics.

Leinonen I., Grant O.M., Tagliavia C.P.P., Chaves M.M. & Jones H.G., 2006. Estimating stomatal conductance with thermal imagery. *Plant, Cell Environ.* **29**(8), 1508–1518, DOI:10.1111/j.1365-3040.2006.01528.x.

Leinonen I. & Jones H.G., 2004. Combining thermal and visible imagery for estimating canopy temperature and identifying plant stress. *J. Exp. Bot.* **55**(401), 1423–1431, DOI:10.1093/jxb/erh146.

Lemaire G. & Gastal F., 1997. N Uptake and Distribution in Plant Canopies. *In*: *Diagnosis of the Nitrogen Status in Crops*. Berlin, Heidelberg: Springer Berlin Heidelberg, 3–43.

Li D., Xu L., Tang X., Sun S., Cai X. & Zhang P., 2017. 3D Imaging of Greenhouse Plants with an Inexpensive Binocular Stereo Vision System. *Remote Sens.* **9**(12), 508, DOI:10.3390/rs9050508.

Li J., Li C., Fei S., Shi J., Xiao Z., Ma C., Chen W., Ding F., Wang Y. & Li Y., 2021. Wheat ear recognition based on retinanet and transfer learning. *Sensors* **21**(14), DOI:10.3390/s21144845.

Li L., Zhang Q. & Huang D., 2014. A review of imaging techniques for plant phenotyping. *Sensors (Switzerland)* **14**(11), 20078–20111, DOI:10.3390/s141120078.

Li W., Jiang J., Weiss M., Madec S., Tison F., Philippe B. & Comar A., 2021. Impact of the reproductive organs on crop BRDF as observed from a UAV. *Remote Sens. Environ.* **259**(April), 1–14, DOI:10.1016/j.rse.2021.112433.

Lin T.-T., Lai T.-C., Liu, Chang-Chih L. & Cheng Y.-C., 2011. A Three-Dimensional Imaging Approach for Plant Feature Measurement Using Stereo Vision. *Tarım Makinaları Bilim. Derg. (Journal Agric. Mach. Sci.* **7**(2), 153–158.

Liu R., Zhang H. & Scherer S., 2018. Multiple Methods of Geometric Calibration of Thermal Camera and A Method of Extracting Thermal Calibration Feature Points.

Liu S., Baret F., Abichou M., Boudon F., Thomas S., Zhao K., Fournier C., Andrieu B., Irfan K., Hemmerlé M. & Solan B. de, 2017. Estimating wheat green area index from ground-based LiDAR measurement using a 3D canopy structure model. *Agric. For. Meteorol.* **247**(November 2016), 12–20, DOI:10.1016/j.agrformet.2017.07.007.

Lowe A., Harrison N. & French A.P., 2017. Hyperspectral image analysis techniques for the detection and classification of the early onset of plant disease and stress. *Plant Methods* **13**(1), 1–12, DOI:10.1186/s13007-017-0233-z.

Lowe D.G., 2004. Distinctive Image Features from Scale-Invariant Keypoints. *Int. J. Comput. Vis.* **60**(2), 91–110, DOI:10.1023/B:VISI.0000029664.99615.94.

Lu B., Proctor C. & He Y., 2021. Investigating different versions of PROSPECT and PROSAIL for estimating spectral and biophysical properties of photosynthetic

and non-photosynthetic vegetation in mixed grasslands. *GIScience Remote Sens.* **58**(3), 354–371, DOI:10.1080/15481603.2021.1877435.

Lu H. & Cao Z., 2020. TasselNetV2+: A Fast Implementation for High-Throughput Plant Counting From High-Resolution RGB Imagery. *Front. Plant Sci.* **11**(December), 1–15, DOI:10.3389/fpls.2020.541960.

Lunagaria M.M. & Patel H.R., 2017. Changes in reflectance anisotropy of wheat crop during different phenophases. *Int. Agrophysics* **31**(2), 203–218, DOI:10.1515/intag-2016-0045.

Ma J., Li Y., Chen Y., Du K., Zheng F., Zhang L. & Sun Z., 2019. Estimating above ground biomass of winter wheat at early growth stages using digital images and deep convolutional neural network. *Eur. J. Agron.* **103**(December 2018), 117–129, DOI:10.1016/j.eja.2018.12.004.

Ma J., Li Y., Du K., Zheng F., Zhang L., Gong Z. & Jiao W., 2020. Segmenting ears of winter wheat at flowering stage using digital images and deep learning. *Comput. Electron. Agric.* **168**(November 2019), 105159, DOI:10.1016/j.compag.2019.105159.

Ma J., Li Y., Liu H., Du K., Zheng F., Wu Y. & Zhang L., 2020. Improving segmentation accuracy for ears of winter wheat at flowering stage by semantic segmentation. *Comput. Electron. Agric.* **176**(July), 105662, DOI:10.1016/j.compag.2020.105662.

Madec S., Baret F., De Solan B., Thomas S., Dutartre D., Jezequel S., Hemmerlé M., Colombeau G. & Comar A., 2017. High-throughput phenotyping of plant height: Comparing unmanned aerial vehicles and ground lidar estimates. *Front. Plant Sci.* **8**, 2, DOI:10.3389/fpls.2017.02002.

Madec S., Jin X., Lu H., De Solan B., Liu S., Duyme F., Heritier E. & Baret F., 2019. Ear density estimation from high resolution RGB imagery using deep learning technique. *Agric. For. Meteorol.* **264**(May 2018), 225–234, DOI:10.1016/j.agrformet.2018.10.013.

Maes W.H., Baert A., Huete A.R., Minchin P.E.H., Snelgar W.P. & Steppe K., 2016. A new wet reference target method for continuous infrared thermography of vegetations. *Agric. For. Meteorol.* **226**–**227**, 119–131, DOI:10.1016/j.agrformet.2016.05.021.

Mahlein A.-K., 2016. Plant disease detection by imaging sensors - Parallels and specific demands for precision agriculture and plant phenotyping. *Plant Dis.* **100**(2), 1–11, DOI:10.1007/s13398-014-0173-7.2.

Meyer G.E., Hindman T. & Laksmi K., 1998. Machine Vision Detection Parameters for Plant Species Identification. *In*: *Precision Agriculture and Biological Quality*. Bellingham, WA, USA: SPIE - International Society for Optics and Photonics, 327–335.

Meyer G.E. & Neto J.C., 2008. Verification of color vegetation indices for automated

crop imaging applications. *Comput. Electron. Agric.* **63**(2), 282–293, DOI:10.1016/j.compag.2008.03.009.

Min D., Choi S., Lu J., Ham B., Sohn K. & Do M.N., 2014. Fast Global Image Smoothing Based on Weighted Least Squares. *IEEE Trans. Image Process.* **23**(12), 5638–5653, DOI:10.1109/TIP.2014.2366600.

Mishra P., Asaari M.S.M., Herrero-Langreo A., Lohumi S., Diezma B. & Scheunders P., 2017. Close range hyperspectral imaging of plants: A review. *Biosyst. Eng.* **164**, 49–67, DOI:10.1016/j.biosystemseng.2017.09.009.

Möller M., Alchanatis V., Cohen Y., Meron M., Tsipris J., Naor A., Ostrovsky V., Sprintsin M. & Cohen S., 2007. Use of thermal and visible imagery for estimating crop water status of irrigated grapevine. *J. Exp. Bot.* **58**(4), 827–838, DOI:10.1093/jxb/erl115.

Monteith J.L., 1973. *Principles of Environmental Physics*, London, UK: Edward Arnold Ltd, 241.

Morisse M., Wells D.M., Millet E.J., Lillemo M., Fahrner S., Cellini F., Lootens P., Muller O., Herrera J.M., Bentley A.R. & Janni M., 2022. A European perspective on opportunities and demands for field-based crop phenotyping. *F. Crop. Res.* **276**, DOI:10.1016/j.fcr.2021.108371.

Moshou D., Bravo C., West J., Wahlen S., McCartney A. & Ramon H., 2004. Automatic detection of "yellow rust" in wheat using reflectance measurements and neural networks. *Comput. Electron. Agric.* **44**(3), 173–188, DOI:10.1016/j.compag.2004.04.003.

Mueller-Sim T., Jenkins M., Abel J. & Kantor G., 2017. The Robotanist: A ground-based agricultural robot for high-throughput crop phenotyping. *In*: *Proceedings - IEEE International Conference on Robotics and Automation*. 3634–3639.

Müller-Linow M., Pinto-Espinosa F., Scharr H. & Rascher U., 2015. The leaf angle distribution of natural plant populations: Assessing the canopy with a novel software tool. *Plant Methods* **11**(1), 1–16, DOI:10.1186/s13007-015-0052-z.

Naito H., Ogawa S., Valencia M.O., Mohri H., Urano Y., Hosoi F., Shimizu Y., Chavez A.L., Ishitani M., Selvaraj M.G. & Omasa K., 2017. Estimating rice yield related traits and quantitative trait loci analysis under different nitrogen treatments using a simple tower-based field phenotyping system with modified single-lens reflex cameras. *ISPRS J. Photogramm. Remote Sens.* **125**, 50–62, DOI:10.1016/j.isprsjprs.2017.01.010.

Nicodemus F.E., Richmond J.C., Hsia J.J., Ginsberg I.W. & Limperis T., 1977. Geometrical Considerations and Nomenclature for Reflectance., Final Report National Bureau of Standards, Washington, WA, USA, 1–52.

Nityananda V. & Read J.C.A., 2017. Stereopsis in animals: evolution, function and mechanisms. *J. Exp. Biol.* **220**(14), 2502–2512, DOI:10.1242/jeb.143883.

O'Riordan A., Newe T., Dooly G. & Toal D., 2019. Stereo vision sensing: Review of existing systems. *In*: *Proceedings of the International Conference on Sensing Technology (ICST)*. Manhattan, NY, USA: IEEE, 178–184.

Oscarson P., 2000. The strategy of the wheat plant in acclimating growth and grain production to nitrogen availability. *J. Exp. Bot.* **51**(352), 1921–1929, DOI:10.1093/jexbot/51.352.1921.

Otsu N., 1979. A Threshold Selection Method from Gray-Level Histograms. *IEEE Trans. Syst. Man. Cybern.* **9**(1), 62–66, DOI:10.1109/TSMC.1979.4310076.

Papoutsoglou E.A., Faria D., Arend D., Arnaud E., Athanasiadis I.N., Chaves I., Coppens F., Cornut G., Costa B. V., Ćwiek-Kupczyńska H., Droesbeke B., Finkers R., Gruden K., Junker A., King G.J., Krajewski P., Lange M., Laporte M.A., Michotey C., Oppermann M., Ostler R., Poorter H., Ramírez-Gonzalez R., Ramšak Ž., Reif J.C., Rocca-Serra P., Sansone S.A., Scholz U., Tardieu F., Uauy C., Usadel B., Visser R.G.F., Weise S., Kersey P.J., Miguel C.M., Adam-Blondon A.F. & Pommier C., 2020. Enabling reusability of plant phenomic datasets with MIAPPE 1.1. *New Phytol.* **227**(1), 260–273, DOI:10.1111/nph.16544.

Pask A., Pietragalla J., Mullan D. & Reynolds M., 2012. *Physiological Breeding II A Field Guide to Wheat Phenotyping ons*, Mexico, MX: CIMMYT, 132.

Pauli D., Chapman S.C., Bart R., Topp C.N., Lawrence-Dill C.J., Poland J. & Gore M.A., 2016. The quest for understanding phenotypic variation via integrated approaches in the field environment. *Plant Physiol.* **172**(2), 622–634, DOI:10.1104/pp.16.00592.

Pérez-Ruiz M., Prior A., Martinez-Guanter J., Apolo-Apolo O.E., Andrade-Sanchez P. & Egea G., 2020. Development and evaluation of a self-propelled electric platform for high-throughput field phenotyping in wheat breeding trials. *Comput. Electron. Agric.* **169**(September 2019), 105237, DOI:10.1016/j.compag.2020.105237.

Pieruschka R. & Schurr U., 2019. Plant Phenotyping: Past, Present, and Future. *Plant Phenomics* **2019**, 1–6, DOI:10.34133/2019/7507131.

Pineda M., Barón M. & Pérez-Bueno M.-L., 2020. Thermal Imaging for Plant Stress Detection and Phenotyping. *Remote Sens.* **13**(1), 68, DOI:10.3390/rs13010068.

Pou A., Diago M.P., Medrano H., Baluja J. & Tardaguila J., 2014. Validation of thermal indices for water status identification in grapevine. *Agric. Water Manag.* **134**, 60–72, DOI:10.1016/j.agwat.2013.11.010.

Pound M.P., Atkinson J.A., Wells D.M., Pridmore T.P. & French A.P., 2017. Deep learning for multi-task plant phenotyping. *In*: *Proceedings - 2017 IEEE International Conference on Computer Vision Workshops, ICCVW 2017*. 2055–2063.

Prashar A. & Jones H., 2014. Infra-Red Thermography as a High-Throughput Tool

for Field Phenotyping. *Agronomy* **4**(3), 397–417, DOI:10.3390/agronomy4030397.

Prey L., von Bloh M. & Schmidhalter U., 2018. Evaluating RGB imaging and multispectral active and hyperspectral passive sensing for assessing early plant vigor in winter wheat. *Sensors (Switzerland)* **18**(9), DOI:10.3390/s18092931.

Qiu R., Wei S., Zhang M., Li H., Sun H., Liu G. & Li M., 2018. Sensors for measuring plant phenotyping: A review. *Int. J. Agric. Biol. Eng.* **11**(2), 1–17, DOI:10.25165/j.ijabe.20181102.2696.

Rabatel G., Gorretta N. & Labbé S., 2011. Getting NDVI spectral bands from a single standard RGB digital camera: A methodological approach. *In*: *Advances in Artificial Intelligence: 14th Conference of the Spanish Association for Artificial Intelligence, CAEPIA 2011*. Berlin, DE: Springer, 333–342.

Rabatel G. & Labbé S., 2016. Registration of visible and near infrared unmanned aerial vehicle images based on Fourier-Mellin transform. *Precis. Agric.* **17**(5), 564–587, DOI:10.1007/s11119-016-9437-x.

Ranson K.J., Daughtry C.S.T., Biehl L.L. & Bauer M.E., 1985. Sun-view angle effects on reflectance factors of corn canopies. *Remote Sens. Environ.* **18**(2), 147–161, DOI:10.1016/0034-4257(85)90045-8.

Raymond Hunt E. & Stern A.J., 2019. Evaluation of incident light sensors on unmanned aircraft for calculation of spectral reflectance. *Remote Sens.* **11**(22), DOI:10.3390/rs11222622.

Raza S.E.A., Sanchez V., Prince G., Clarkson J.P. & Rajpoot N.M., 2015. Registration of thermal and visible light images of diseased plants using silhouette extraction in the wavelet domain. *Pattern Recognit.* **48**(7), 2119–2128, DOI:10.1016/j.patcog.2015.01.027.

Reddy B.S. & Chatterji B.N., 1996. An FFT-based technique for translation, rotation, and scale-invariant image registration. *IEEE Trans. Image Process.* **5**(8), 1266–1271, DOI:10.1109/83.506761.

Redmon J., Divvala S., Girshick R. & Farhadi A., 2016. You only look once: Unified, real-time object detection. *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* **2016-Decem**, 779–788, DOI:10.1109/CVPR.2016.91.

Ren S., He K., Girshick R. & Sun J., 2017. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **39**(6), 1137–1149, DOI:10.1109/TPAMI.2016.2577031.

Reynolds D., Baret F., Welcker C., Bostrom A., Ball J., Cellini F., Lorence A., Chawade A., Khafif M., Noshita K., Mueller-Linow M., Zhou J. & Tardieu F., 2019. What is cost-efficient phenotyping? Optimizing costs for different scenarios. *Plant Sci.* **282**(December 2017), 14–22, DOI:10.1016/j.plantsci.2018.06.015.

Reynolds M., Chapman S., Crespo-Herrera L., Molero G., Mondal S., Pequeno D.N.L., Pinto F., Pinera-Chavez F.J., Poland J., Rivera-Amado C., Saint Pierre C. & Sukumaran S., 2020. Breeder friendly phenotyping. *Plant Sci.* **295**(July 2019), 110396, DOI:10.1016/j.plantsci.2019.110396.

Rodriguez D., Fitzgerald G.J., Belford R. & Christensen L.K., 2006. Detection of nitrogen deficiency in wheat from spectral reflectance indices and basic crop eco-physiological concepts. *Aust. J. Agric. Res.* **57**(7), 781–789, DOI:10.1071/AR05361.

Rohlfing T., 2012. Image Similarity and Tissue Overlaps as Surrogates for Image Registration Accuracy: Widely Used but Unreliable. *IEEE Trans. Med. Imaging* **31**(2), 153–163, DOI:10.1109/TMI.2011.2163944.Image.

Roitsch T., Cabrera-Bosquet L., Fournier A., Ghamkhar K., Jiménez-Berni J., Pinto F. & Ober E.S., 2019. Review: New sensors and data-driven approaches—A path to next generation phenomics. *Plant Sci.* **282**(December 2018), 2–10, DOI:10.1016/j.plantsci.2019.01.011.

Rouse J.W., Hass R.H., Schell J.A. & Deering D.W., 1973. Monitoring vegetation systems in the great plains with ERTS. *In*: *Third Earth Resources Technology Satellite (ERTS) Symposium*. NASA SP-351, 309–317.

Rublee E., Rabaud V., Konolige K. & Bradski G., 2011. ORB: An efficient alternative to SIFT or SURF. *In*: *Proceedings of the IEEE International Conference on Computer Vision*. Manhattan, NY, USA: IEEE, 2564–2571.

Rueckert et al. D., 1999. Nonrigid Registration Using Free-Form Deformations: Application to Breast MR Images. *IEEE Trans. Med. Imaging* **18**(8), 712–721.

Sadeghi-Tehran P., Virlet N., Ampe E.M., Reyns P. & Hawkesford M.J., 2019. DeepCount: In-Field Automatic Quantification of Wheat Spikes Using Simple Linear Iterative Clustering and Deep Convolutional Neural Networks. *Front. Plant Sci.* **10**(September), 1–16, DOI:10.3389/fpls.2019.01176.

Sadeghi-Tehran P., Virlet N. & Hawkesford M.J., 2021. A neural network method for classification of sunlit and shaded components of wheat canopies in the field using high-resolution hyperspectral imagery. *Remote Sens.* **13**(5), 1–17, DOI:10.3390/rs13050898.

Sadeghi-Tehran P., Virlet N., Sabermanesh K. & Hawkesford M.J., 2017. Multi-feature machine learning model for automatic segmentation of green fractional vegetation cover for high-throughput field phenotyping. *Plant Methods* **13**(1), 1–16, DOI:10.1186/s13007-017-0253-8.

Saeys W., Lenaerts B., Craessaerts G. & De Baerdemaeker J., 2009. Estimation of the crop density of small grains using LiDAR sensors. *Biosyst. Eng.* **102**(1), 22–30, DOI:10.1016/j.biosystemseng.2008.10.003.

Saint Cast C., Lobet G., Cabrera-Bosquet L., Couvreur V., Pradal C., Tardieu F. & Draye X., 2022. Connecting plant phenotyping and modelling communities:

lessons from science mapping and operational perspectives. *in silico Plants* **4**(1), 1–13, DOI:10.1093/insilicoplants/diac005.

Salas Fernandez M.G., Bao Y., Tang L. & Schnable P.S., 2017. A High-Throughput, Field-Based Phenotyping Technology for Tall Biomass Crops. *Plant Physiol.* **174**(4), 2008–2022, DOI:10.1104/pp.17.00707.

Sarić R., Nguyen V.D., Burge T., Berkowitz O., Trtílek M., Whelan J., Lewsey M.G. & Čustović E., 2022. Applications of hyperspectral imaging in plant phenotyping. *Trends Plant Sci.* **27**(3), 301–315, DOI:10.1016/j.tplants.2021.12.003.

Schaepman-Strub G., Schaepman M.E., Painter T.H., Dangel S. & Martonchik J. V., 2006. Reflectance quantities in optical remote sensing-definitions and case studies. *Remote Sens. Environ.* **103**(1), 27–42, DOI:10.1016/j.rse.2006.03.002.

Scharstein D. & Pal C., 2007. Learning conditional random fields for stereo. *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* DOI:10.1109/CVPR.2007.383191.

Scharstein D. & Szeliski R., 2002. A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms. *Int. J. Comput. Vis.* **47**(1/3), 7–42, DOI:10.1023/A:1014573219977.

Scharstein D. & Szeliski R., 2003. High-Accuracy Stereo Depth Maps Using Structured Light. *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* **1**(June), I-195-I–202, DOI:10.1109/CVPR.2003.1211354.

Schirrmann M., Hamdorf A., Garz A., Ustyuzhanin A. & Dammer K.H., 2016. Estimating wheat biomass by combining image clustering with crop height. *Comput. Electron. Agric.* **121**, 374–384, DOI:10.1016/j.compag.2016.01.007.

Schirrmann M., Landwehr N., Giebel A., Garz A. & Dammer K.-H., 2021. Early Detection of Stripe Rust in Winter Wheat Using Deep Residual Neural Networks. *Front. Plant Sci.* **12**(March), DOI:10.3389/fpls.2021.469689.

Schuepp P.H., 1993. Leaf Boundary Layers. *New Phytol.* **125**(3), 477–507.

Shafiekhani A., Kadam S., Fritschi F. & DeSouza G., 2017. Vinobot and Vinoculer: Two Robotic Platforms for High-Throughput Field Phenotyping. *Sensors* **17**(12), 214, DOI:10.3390/s17010214.

Shibayama M. & Wiegand C.L., 1985. View azimuth and zenith, and solar angle effects on wheat canopy reflectance. *Remote Sens. Environ.* **18**(1), 91–103, DOI:10.1016/0034-4257(85)90040-9.

Simón M.R., Fleitas M.C., Castro A.C. & Schierenbeck M., 2020. How Foliar Fungal Diseases Affect Nitrogen Dynamics, Milling, and End-Use Quality of Wheat. *Front. Plant Sci.* **11**(November), 1–23, DOI:10.3389/fpls.2020.569401.

Smith D.T., Potgieter A.B. & Chapman S.C., 2021. Scaling up high-throughput phenotyping for abiotic stress selection in the field. *Theor. Appl. Genet.* **134**(6),

1845–1866, DOI:10.1007/s00122-021-03864-5.

Solovyev R., Wang W. & Gabruseva T., 2021. Weighted boxes fusion: Ensembling boxes from different object detection models. *Image Vis. Comput.* **107**, DOI:10.1016/j.imavis.2021.104117.

Sotiras A., Davatzikos C. & Paragios N., 2013. Deformable medical image registration: A survey. *IEEE Trans. Med. Imaging* **32**(7), 1153–1190, DOI:10.1109/TMI.2013.2265603.

Souza R.D., Buchhart C., Heil K., Plass J., Padilla F.M. & Schmidhalter U., 2021. Effect of Time of Day and Sky Conditions on Different Vegetation Indices Calculated from Active and Passive Sensors and Images Taken from UAV. *Remote Sens.* **13**(9), 1–21, DOI:10.3390/rs13091691.

Stoll M., Schultz H.R. & Berkelmann-Loehnertz B., 2008. Thermal sensitivity of grapevine leaves affected by Plasmopara viticola and water stress. *Vitis - J. Grapevine Res.* **47**(2), 133–134.

Studholme C., Hill D.L.G. & Hawkes D.J., 1999. An overlap invariant entropy measure of 3D medical image alignment. *Pattern Recognit.* **32**(1), 71–86, DOI:10.1016/S0031-3203(98)00091-0.

Su W.H., Zhang J., Yang C., Page R., Szinyei T., Hirsch C.D. & Steffenson B.J., 2021. Automatic evaluation of wheat resistance to fusarium head blight using dual mask-rcnn deep learning frameworks in computer vision. *Remote Sens.* **13**(1), 1–20, DOI:10.3390/rs13010026.

Suffert F., Sache I. & Lannou C., 2011. Early stages of septoria tritici blotch epidemics of winter wheat: Build-up, overseasoning, and release of primary inoculum. *Plant Pathol.* **60**(2), 166–177, DOI:10.1111/j.1365-3059.2010.02369.x.

Tan C., Zhang P., Zhang Y., Zhou X., Wang Z., Du Y., Mao W., Li W., Wang D. & Guo W., 2020. Rapid Recognition of Field-Grown Wheat Spikes Based on a Superpixel Segmentation Algorithm Using Digital Images. *Front. Plant Sci.* **11**(March), 1–13, DOI:10.3389/fpls.2020.00259.

Tardieu F., Cabrera-Bosquet L., Pridmore T. & Bennett M., 2017. Plant Phenomics, From Sensors to Knowledge. *Curr. Biol.* **27**(15), R770–R783, DOI:10.1016/j.cub.2017.05.055.

Tavakoli H. & Gebbers R., 2019. Assessing Nitrogen and water status of winter wheat using a digital camera. *Comput. Electron. Agric.* **157**(October 2018), 558–567, DOI:10.1016/j.compag.2019.01.030.

Tayade R., Yoon J., Lay L., Khan A.L., Yoon Y. & Kim Y., 2022. Utilization of Spectral Indices for High-Throughput Phenotyping. *Plants* **11**(13), 1712, DOI:10.3390/plants11131712.

Tewari V.K., Arudra A.K., Kumar S.P., Pandey V. & Chandel N.S., 2013. Estimation of plant nitrogen content using digital image processing. *Agric. Eng. Int. CIGR*

J. **15**(2), 78–86.

Thom A.S. & Oliver H.R., 1977. On Penman's equation for estimating regional evaporation. *Quaterly J. R. Meteorol. Soc.* **103**(436), 345–357.

Tilly N., Aasen H. & Bareth G., 2015. Fusion of Plant Height and Vegetation Indices for the Estimation of Barley Biomass. *Remote Sens.* **7**(9), 11449–11480, DOI:10.3390/rs70911449.

Tilneac M., Dolga V., Grigorescu S. & Bitea M.A., 2012. 3D stereo vision measurements for weed-crop discrimination. *Elektron. ir Elektrotechnika* **123**(7), 9–12, DOI:10.5755/j01.eee.123.7.2366.

Tremblay N., Wang Z., Ma B.L., Belec C. & Vigneault P., 2009. A comparison of crop data measured by two commercial sensors for variable-rate nitrogen application. *Precis. Agric.* **10**(2), 145–161, DOI:10.1007/s11119-008-9080-2.

Tripodi P., Massa D., Venezia A. & Cardi T., 2018. Sensing Technologies for Precision Phenotyping in Vegetable Crops: Current Status and Future Challenges. *Agronomy* **8**(4), 57, DOI:10.3390/agronomy8040057.

Vallat R., 2018. Pingouin: statistics in Python. *J. Open Source Softw.* **3**(31), 1026, DOI:10.21105/joss.01026.

van Eeuwijk F.A., Bustos-Korts D., Millet E.J., Boer M.P., Kruijer W., Thompson A., Malosetti M., Iwata H., Quiroz R., Kuppe C., Muller O., Blazakis K.N., Yu K., Tardieu F. & Chapman S.C., 2019. Modelling strategies for assessing and increasing the effectiveness of new phenotyping techniques in plant breeding. *Plant Sci.* **282**(December 2017), 23–39, DOI:10.1016/j.plantsci.2018.06.018.

Vázquez-arellano M., Griepentrog H.W., Reiser D. & Paraforos D.S., 2016. 3-D Imaging Systems for Agricultural Applications — A Review. *Sensors* **16**(5), 618–641, DOI:10.3390/s16050618.

Velumani K., Oude Elberink S., Yang M.Y. & Baret F., 2017. Wheat Ear Detection in Plots by Segmenting Mobile Laser Scanner Data. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **4**(2W4), 149–156, DOI:10.5194/isprs-annals-IV-2-W4-149-2017.

Verger A., Vigneau N., Chéron C., Gilliot J.M., Comar A. & Baret F., 2014. Green area index from an unmanned aerial system over wheat and rapeseed crops. *Remote Sens. Environ.* **152**, 654–664, DOI:10.1016/j.rse.2014.06.006.

Verhoef W., 1984. Light scattering by leaf layers with application to canopy reflectance modeling: The SAIL model. *Remote Sens. Environ.* **16**(2), 125–141, DOI:10.1016/0034-4257(84)90057-9.

Vigneau N., Ecarnot M., Rabatel G. & Roumet P., 2011. Potential of field hyperspectral imaging as a non destructive method to assess leaf nitrogen content in Wheat. *F. Crop. Res.* **122**(1), 25–31, DOI:10.1016/j.fcr.2011.02.003.

Virlet N., Sabermanesh K., Sadeghi-Tehran P. & Hawkesford M.J., 2017. Field

Scanalyzer: An automated robotic field phenotyping platform for detailed crop monitoring. *Funct. Plant Biol.* **44**(1), 143–153, DOI:10.1071/FP16163.

Walter J., Edwards J., Cai J., McDonald G., Miklavcic S.J. & Kuchel H., 2019. High-throughput field imaging and basic image analysis in a wheat breeding programme. *Front. Plant Sci.* **10**(April), 1–12, DOI:10.3389/fpls.2019.00449.

Wan L., Li H., Li C., Wang A., Yang Y. & Wang P., 2022. Hyperspectral Sensing of Plant Diseases: Principle and Methods. *Agronomy* **12**(6), 1451, DOI:10.3390/agronomy12061451.

Wang A., Zhang W. & Wei X., 2019. A review on weed detection using ground-based machine vision and image processing techniques. *Comput. Electron. Agric.* **158**(November 2018), 226–240, DOI:10.1016/j.compag.2019.02.005.

Wang D., Fu Y., Yang G., Yang X., Liang D., Zhou C., Zhang N., Wu H. & Zhang D., 2019. Combined Use of FCN and Harris Corner Detection for Counting Wheat Ears in Field Conditions. *IEEE Access* **7**, 178930–178941, DOI:10.1109/ACCESS.2019.2958831.

Wang X., Yang W., Wheaton A., Cooley N. & Moran B., 2010. Efficient registration of optical and IR images for automatic plant water stress assessment. *Comput. Electron. Agric.* **74**(2), 230–237, DOI:10.1016/j.compag.2010.08.004.

Wang Y., Qin Y. & Cui J., 2021. Occlusion Robust Wheat Ear Counting Algorithm Based on Deep Learning. *Front. Plant Sci.* **12**(June), 1–14, DOI:10.3389/fpls.2021.645899.

Whetton R.L., Waine T.W. & Mouazen A.M., 2018. Hyperspectral measurements of yellow rust and fusarium head blight in cereal crops: Part 2: On-line field measurement. *Biosyst. Eng.* **167**, 144–158, DOI:10.1016/j.biosystemseng.2018.01.004.

Wilkinson M.D., Dumontier M., Aalbersberg Ij.J., Appleton G., Axton M., Baak A., Blomberg N., Boiten J.-W., da Silva Santos L.B., Bourne P.E., Bouwman J., Brookes A.J., Clark T., Crosas M., Dillo I., Dumon O., Edmunds S., Evelo C.T., Finkers R., Gonzalez-Beltran A., Gray A.J.G., Groth P., Goble C., Grethe J.S., Heringa J., Hoen P.A.. 't, Hooft R., Kuhn T., Kok R., Kok J., Lusher S.J., Martone M.E., Mons A., Packer A.L., Persson B., Rocca-Serra P., Roos M., van Schaik R., Sansone S.-A., Schultes E., Sengstag T., Slater T., Strawn G., Swertz M. a., Thompson M., van der Lei J., van Mulligen E., Velterop J., Waagmeester A., Wittenburg P., Wolstencroft K., Zhao J. & Mons B., 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **3**, 160018, DOI:10.1038/sdata.2016.18.

Willenborg C.J. & Van Acker R.C., 2008. The biology and ecology of hexaploid wheat (Triticum aestivum L.) and its implications for trait confinement. *Can. J. Plant Sci.* **88**(5), 997–1013, DOI:10.4141/CJPS07144.

Wu Y., Hu Y. & Li L., 2020. BTWD: Bag of Tricks for Wheat Detection. *In*: *Lecture*

Notes in Computer Science. New-York, NY, USA: Springer International Publishing, 450–460.

Xiang L., Tang L., Gai J. & Wang L., 2020. PhenoStereo: a high-throughput stereo vision system for field-based plant phenotyping-with an application in sorghum stem diameter estimation. *2020 ASABE Annu. Int. Virtual Meet. July 13-15, 2020* DOI:10.13031/aim.202001190.

Xiong H., Cao Z., Lu H., Madec S., Liu L. & Shen C., 2019. TasselNetv2: In-field counting of wheat spikes with context-augmented local regression networks. *Plant Methods* **15**(1), 0–14, DOI:10.1186/s13007-019-0537-2.

Xiong X., Duan L., Liu L., Tu H., Yang P., Wu D., Chen G., Xiong L., Yang W. & Liu Q., 2017. Panicle-SEG: A robust image segmentation method for rice panicles in the field based on deep learning and superpixel optimization. *Plant Methods* **13**(1), 1–15, DOI:10.1186/s13007-017-0254-7.

Xiong Z. & Zhang Y., 2010. A critical review of image registration methods. *Int. J. Image Data Fusion* **1**(2), 137–158, DOI:10.1080/19479831003802790.

Xu P., Wu G., Guo Y., Chen X., Yang H. & Zhang R., 2017. Automatic Wheat Leaf Rust Detection and Grading Diagnosis via Embedded Image Processing System. *Procedia Comput. Sci.* **107**, 836–841, DOI:10.1016/j.procs.2017.03.177.

Xu X., Li H., Yin F., Xi L., Qiao H., Ma Z., Shen S., Jiang B. & Ma X., 2020. Wheat ear counting using K-means clustering segmentation and convolutional neural network. *Plant Methods* **16**(1), 1–13, DOI:10.1186/s13007-020-00648-8.

Xue J. & Su B., 2017. Significant remote sensing vegetation indices: A review of developments and applications. *J. Sensors* **2017**, DOI:10.1155/2017/1353691.

Yang B., Gao Z., Gao Y. & Zhu Y., 2021. Rapid Detection and Counting of Wheat Ears in the Field Using YOLOv4 with Attention Module. *Agronomy* **11**(6), 1202, DOI:10.3390/agronomy11061202.

Yang W., Wang X., Moran B., Wheaton A. & Cooley N., 2012. Efficient registration of optical and infrared images via modified Sobel edging for plant canopy temperature estimation. *Comput. Electr. Eng.* **38**(5), 1213–1221, DOI:10.1016/j.compeleceng.2012.05.014.

Yang W.P., Wang X.Z., Wheaton A., Cooley N., Moran B. & Ieee, 2009. Automatic Optical and IR Image Fusion for Plant Water Stress Analysis. *In*: *12th International Conference on Information Fusion*. Manhattan, NY, USA: IEEE, 1053–1059.

Yang Y., Huang X., Cao L., Chen L. & Huang K., 2019. Field Wheat Ears Count Based on YOLOv3. *In*: *Proceedings - 2019 International Conference on Artificial Intelligence and Advanced Manufacturing, AIAM 2019*. Manhattan, NY, USA: IEEE, 444–448.

Yuan L., Zhang J., Zhao J., Huang W., Wang J., Yuan L., Zhang J. & Wang J., 2012.

Discrimination of yellow rust and powdery mildew in wheat at leaf level using spectral signatures. *In*: *First International Conference on Agro-Geoinformatics*. Manhattan, NY, USA: IEEE, 1–5.

Yun L., 2012. Traitement et analyse d'images stéréoscopiques avec les approches du calcul générique sur un processeur graphique.

Zadoks J., Chang T. & Konzak C., 1974. A decimal growth code for the growth stages of cereals. *Weed Res.* **14**(14), 415–421.

Zhang Z., 2000. A flexible new technique for camera calibration. *IEEE Trans. Pattern Anal. Mach. Intell.* **22**(11), 1330–1334, DOI:10.1109/34.888718.

Zhao J., Zhang X., Yan J., Qiu X., Yao X., Tian Y., Zhu Y. & Cao W., 2021. A Wheat Spike Detection Method in UAV Images Based on Improved YOLOv5. *Remote Sens.* **13**, 1–16.

Zheng Q., Huang W., Cui X., Dong Y., Shi Y., Ma H. & Liu L., 2019. Identification of wheat yellow rust using optimal three-band spectral indices in different growth stages. *Sensors (Switzerland)* **19**(1), DOI:10.3390/s19010035.

Zhou C., Liang D., Yang X., Xu B. & Yang G., 2018. Recognition of wheat spike from field based phenotype platform using multi-sensor fusion and improved maximum entropy segmentation algorithms. *Remote Sens.* **10**(2), DOI:10.3390/rs10020246.

Zhou C., Liang D., Yang X., Yang H., Yue J. & Yang G., 2018. Wheat ears counting in field conditions based on multi-feature optimization and TWSVM. *Front. Plant Sci.* **9**(July), DOI:10.3389/fpls.2018.01024.

Zhu Y., Cao Z., Lu H., Li Y. & Xiao Y., 2016. In-field automatic observation of wheat heading stage using computer vision. *Biosyst. Eng.* **143**, 28–41, DOI:10.1016/j.biosystemseng.2015.12.015.

Zitová B. & Flusser J., 2003. Image registration methods: A survey. *Image Vis. Comput.* **21**(11), 977–1000, DOI:10.1016/S0262-8856(03)00137-9.

Zou X. & Mõttus M., 2017. Sensitivity of common vegetation indices to the canopy structure of field crops. *Remote Sens.* **9**(10), 1–15, DOI:10.3390/rs9100994.